

Package ‘mosaic’

March 1, 2014

Type Package

Title Project MOSAIC (mosaic-web.org) statistics and mathematics teaching utilities

Version 0.8-18

Date 2014-03-01

Depends R (>= 3.0.0), grid, lattice (>= 0.20-21), car

Imports Hmisc, methods, utils, splines

Suggests

MASS, reshape2, ggplot2, plyr, maptools, manipulate, vcd, RCurl, testthat, knitr, tools, parallel

VignetteBuilder knitr

Author Randall Pruim <rpruim@calvin.edu>, Daniel Kaplan
<kaplan@macalester.edu>, Nicholas Horton <nhorton@amherst.edu>

Maintainer Randall Pruim <rpruim@calvin.edu>

Description Data sets and utilities from Project MOSAIC (mosaic-web.org) used to teach mathematics, statistics, computation and modeling. Funded by the NSF, Project MOSAIC is a community of educators working to tie together aspects of quantitative work that students in science, technology, engineering and mathematics will need in their professional lives, but which are usually taught in isolation, if at all.

License GPL (>= 2)

LazyLoad yes

LazyData yes

NeedsCompilation no

Repository CRAN

Date/Publication 2014-03-01 23:43:34

R topics documented:

mosaic-package	4
.polyExp	5
.qdata_old	6
adapt_seq	8
aggregatingFunction1	9
aggregatingFunction2	9
Alcohol	10
as.xtabs	11
bargraph	12
binom.test	13
Births78	14
Broyden	15
Cards	16
cdist	16
CIsim	17
coef.function	18
columns	19
compareMean	19
compareProportion	20
confint.numeric	21
CoolingWater	22
Countries	23
CPS85	24
cross	25
D	25
deg2rad	27
deltaMethod	28
derivedFactor	29
dfapply	31
do	32
dotPlot	33
ediff	34
evalFormula	35
evalSubFormula	35
expandFun	36
fav_stats	37
fetchData	37
fetchGapminder1	39
fetchGoogle	39
findZeros	40
findZerosMult	42
fitModel	43
fitSpline	44
freqpolygon	45
FunctionsFromData	47
Galton	48

Gestation	49
getVarFormula	50
googleMap	51
HeatX	52
Heightweight	53
HELP	53
HELPfull	55
HELPmiss	82
HELPrct	84
inferArgs	86
integrateODE	86
interval	87
is.wholenumber	88
joinFrames	89
KidsFeet	90
ladd	91
linear.algebra	92
logical2factor	93
logit	93
maggregate	94
makeColorscheme	95
makeFun	96
Marriage	97
mean	98
Mites	100
mm	101
modelVars	103
mosaic.options	103
mosaic_formula	104
mPlot	105
ntiles	106
numD	107
oddsRatio	109
panel.levelcontourplot	110
panel.lmbands	111
panel.plotFun	112
panel.plotFun1	113
parse.formula	114
perctable	115
plotCumfreq	116
plotDist	117
plotFun	118
plotPoints	120
project	121
prop	122
prop.test	123
r.squared	125
RailTrail	125

rand	126
read.file	127
repeater-class	128
resample	129
rflip	130
rfun	131
rgeo	132
rkintegrate	133
rsquared	134
SAT	135
SnowGR	136
sp2df	137
statTally	138
surround	139
SwimRecords	140
symbolicD	140
symbolicInt	141
t.test	142
tally	143
TenMileRace	144
theme.mosaic	145
TukeyHSD.lm	146
Utilities	147
Utilities2	148
Whickham	149
xchisq.test	150
xhistogram	150
xpnorm	152
xqqmath	153
xyz2latlon	154
Index	156

mosaic-package

mosaic: the Project MOSAIC package

Description

mosaic

Details

Data sets and utilities from Project MOSAIC (mosaic-web.org) used to teach mathematics, statistics, computation and modeling. Funded by the NSF, Project MOSAIC is a community of educators working to tie together aspects of quantitative work that students in science, technology, engineering and mathematics will need in their professional lives, but which are usually taught in isolation, if at all.

Author(s)

Randall Pruim (<rpruim@calvin.edu>), Daniel Kaplan (<kaplan@macalester.edu>), Nicholas Horton (<nhorton@smith.edu>)

References

<http://mosaic-web.org>

.polyExp	<i>Takes a call and returns its polynomial coefficients</i>
----------	---

Description

Takes a call and returns its polynomial coefficients

Takes a call and returns its polynomial coefficients as numerics.

Method for putting a polynomial together given the coefficients and power from .polyExp()

Usage

.polyExp(tree, .x., params, iterate = 1)

.polyExp.num(tree, .x.)

.makePoly(form, poly)

Arguments

tree	A call that will be parsed and simplified recursively
.x.	the variable name with respect to which the polynomial should be most simplified
params	All names of free variables. If there are no free variables, the value should be ""
iterate	The number of times the call is nested. Default and proper value when called from the outside is 1
poly	output of .polyExp()
form	original formula - provides information on which variable the polynomial was reduced with respect to.

Details

Will work on any call as long as it can be reduced to a polynomial with respect the the variable and each of the parameters. Operates recursively, reducing each of the coefficients with respect to the extra parameters in turn. Calls .polyExp.num when all remaining coefficients are numeric to reduce the expression more fully.

works with the same structure as .polyExp() but will return only if all coefficients reduce to numeric values.

Value

A list containing a list, `coeffs`, of coefficients ordered high to low (i.e. the list (2,3,4) would correspond to the polynomial $2*x^2+3*x+4$) and value, `pow`, indicating the order of the polynomial. If the expression is not a polynomial, this method returns an empty list or an error.

A list containing a list, `coeffs`, of coefficients ordered high to low (i.e. the list (2,3,4) would correspond to the polynomial $2*x^2+3*x+4$) and value, `pow`, indicating the order of the polynomial. If the expression is not a polynomial, this method returns an empty list or an error.

A formula whose left hand side is a polynomial that fits the description given with the input `poly`.

Examples

```
mosaic:::polyExp(lhs((2*x+x^3+1)^3+x~x), "x", "")
mosaic:::polyExp(lhs((3+a*x)^2+a^2*(x+2)~x), "x", "a")
mosaic:::polyExp(lhs((a+b*x)^3~x), "x", c("a", "b"))
```

.qdata_old

The Data Distribution

Description

Density, distribution function, quantile function, and random generation from data.

`cdata` is a wrapper around `qdata` and determines endpoints of central probabilities rather than tail probabilities.

`pdata` computes cumulative probabilities from data.

`rdata` randomly samples from data. It is a wrapper around `sample` that unifies syntax.

`ddata` computes a probability mass function from data.

Usage

```
.qdata_old(p, vals, data = NULL, ...)
```

```
qdata_v(x, p = seq(0, 1, 0.25), ...)
```

```
qdata_f(x, ..., data, groups = NULL, ..fun.. = qdata_v)
```

```
qdata(p, vals, data = NULL, ...)
```

```
.cdata_old(p, vals, data = NULL, ...)
```

```
cdata_v(x, p = 0.95, ...)
```

```
cdata_f(x, ..., data, groups = NULL, ..fun.. = cdata_v)
```

```
cdata(p, vals, data = NULL, ...)
```

```
pdata_v(x, q, lower.tail = TRUE, ...)  
pdata_f(x, ..., data, groups = NULL, ..fun.. = pdata_v)  
pdata(q, vals, data = NULL, ...)  
.pdata_old(q, vals, data = NULL, lower.tail = TRUE, ...)  
rdata(n, vals, data = NULL, replace = TRUE, ...)  
ddata(x, vals, data = NULL, log = FALSE, ...)
```

Arguments

<code>p</code>	a vector of probabilities
<code>vals</code>	a vector containing the data
<code>data</code>	a data frame in which to evaluate vals
<code>groups</code>	a grouping variable, typically the name of a variable in data
<code>..fun..</code>	a function. Most users will not need to change the default value.
<code>...</code>	additional arguments passed to <code>quantile</code> or <code>sample</code>
<code>q</code>	a vector of quantiles
<code>lower.tail</code>	a logical indicating whether to use the lower or upper tail probability
<code>n</code>	number of values to sample
<code>replace</code>	a logical indicating whether to sample with replacement
<code>x</code>	a vector of quantiles
<code>log</code>	a logical indicating whether the result should be log transformed

Details

`qdata` is a wrapper around [quantile](#) that makes the syntax more like the syntax for quantiles from theoretical distributions

Value

For `qdata`, a vector of quantiles
For `pdata`, a vector of probabilities
For `rdata`, a vector of values sampled from `vals`
For `ddata`, a vector of probabilities (empirical densities)

Examples

```
data(iris)  
qdata(.5, Sepal.Length ~ Species, data=iris)  
qdata(.5, ~Sepal.Length, groups=Species, data=iris)  
qdata(.5, iris$Sepal.Length)
```

```

qdata(.5, Sepal.Length, data=iris)
qdata(.5, Sepal.Length, groups=Species, data=iris)
data(iris)
cdata(.5, iris$Sepal.Length)
cdata(.5, Sepal.Length, data=iris)
cdata_f(~Sepal.Length, data=iris, p=.5)
cdata_f(~Sepal.Length | Species, data=iris, p=.5)
data(iris)
pdata(3:6, iris$Sepal.Length)
pdata(3:6, Sepal.Length, data=iris)
data(iris)
rdata(10, iris$Species)
rdata(10, Species, data=iris)
data(iris)
ddata('setosa', iris$Species)
ddata('setosa', Species, data=iris)

```

adapt_seq

Adaptively generate sequences in an interval

Description

adapt_seq is similar to seq except that instead of selecting points equally spaced along an interval, it selects points such that the values of a function applied at those points are (very) roughly equally spaced. This can be useful for sampling a function in such a way that it can be plotted more smoothly, for example.

Usage

```

adapt_seq(from, to, length.out = 100, f = function(x, ...) { 1 },
          args = list())

```

Arguments

from	start of interval
to	end of interval
length.out	desired length of sequence
f	a function
args	arguments passed to f

Value

a numerical vector

Examples

```

adapt_seq(0, pi, 25, sin)

```

aggregatingFunction1 *1-ary Aggregating functions*

Description

aggregatingFunction1 creates statistical summaries of one numerical vector that are formula aware.

Usage

```
aggregatingFunction1(fun, input.multiple = FALSE, output.multiple = FALSE,  
  envir = parent.frame())
```

Arguments

fun a function that takes a numeric vector and computes a summary statistic, returning a numeric vector of length 1.

output.multiple a boolean indicating whether `..fun..` returns multiple values

input.multiple a boolean indicating whether `..fun..` can accept 2 vectors (e.g., `var`)

envir an environment in which evaluation takes place.

Value

a function that generalizes `fun` to handle a formula/data frame interface.

Examples

```
foo <- aggregatingFunction1( base::mean )  
foo( ~length, data=KidsFeet )  
base::mean(KidsFeet$length)  
foo( length ~ sex, data=KidsFeet )
```

aggregatingFunction2 *2-ary Aggregating functions*

Description

aggregatingFunction2 creates statistical summaries of two numerical vectors that are formula aware.

Usage

```
aggregatingFunction2(fun)
```

Arguments

`fun` a function that takes two numeric vectors and computes a summary statistic, returning a numeric vector of length 1.

Value

a function that generalizes `fun` to handle a formula/data frame interface.

Examples

```
foo <- aggregatingFunction2( stats::cor)
foo( length ~ width, data=KidsFeet )
stats::cor( KidsFeet$length, KidsFeet$width )
```

Alcohol

Alcohol Consumption per Capita

Description

These data provide per capita alcohol consumption values for many countries in 2005 and 2008. There are also a few countries for which there are data in other years.

Usage

```
data(Alcohol)
```

Format

A data frame with 411 observations on the following variables.

- country country name
- year year
- alcohol per capita alcohol consumption

Source

Gapminder (<http://www.gapminder.org/>)

References

<http://www.fvgreenway.org/pdfs/Northampton-Bikepath-Volume-Counts>

Examples

```
data(Alcohol)
# There are only a few observations in years other than 2005 and 2008
subset(Alcohol, ! year %in% c(2005,2008))
```

as.xtabs *Convert objects to xtabs format*

Description

Convert a data frame or a matrix into an xtabs object.

Usage

```
as.xtabs(x, ...)  
  
## S3 method for class 'data.frame'  
as.xtabs(x, rowvar = NULL, colvar = NULL, labels = 1,  
        ...)  
  
## S3 method for class 'matrix'  
as.xtabs(x, rowvar = NULL, colvar = NULL, ...)
```

Arguments

x	object (typically a data frame) to be converted to xtabs format
...	additional arguments to be passed to or from methods.
rowvar	name of the row variable as character string
colvar	name of the column variable as character string
labels	column of data frame that contains the labels of the row variable.

Details

The intended use is to convert a two-way contingency table stored in a data frame or a matrix into an xtabs object.

Value

An xtabs object.

Examples

```
# example from example(fisher.test)  
df <- data.frame( X=c('Tea', 'Milk'), Tea=c(3,1), Milk=c(1,3) )  
xt <- as.xtabs(df, rowvar="Guess", colvar="Truth"); xt  
if (require(vcd)) { mosaic(xt) }
```

bargraph	<i>Create bar graphs from raw data</i>
----------	--

Description

[barchart](#) from the `lattice` package makes bar graphs from pre-tabulated data. Raw data can be tabulated using [xtabs](#), but the syntax is unusual compared to the other `lattice` plotting functions. `bargraph` provides an interface that is consistent with the other `lattice` functions.

Usage

```
bargraph(x, data = parent.frame(), groups, horizontal = FALSE, origin = 0,
         ylab = ifelse(horizontal, "", "Frequency"), xlab = ifelse(horizontal,
         "Frequency", ""), subset, ...)
```

Arguments

<code>x</code>	a formula describing the plot
<code>data</code>	a data frame in which the formula <code>x</code> is evaluated
<code>groups</code>	a variable or expression used for grouping. See barchart .
<code>horizontal</code>	a logical indicating whether bars should be horizontal
<code>...</code>	additional arguments passed to barchart
<code>origin</code>	beginning point for bars. For the default behavior used by barchart set <code>origin</code> to <code>NULL</code> , but 0 is often a better default. If 0 is not good, perhaps you should use a different kind of plot as the results may be misleading.
<code>subset</code>	a vector used to subset data. This may be an expression that will be evaluated within data.
<code>ylab</code>	a character vector of length one used for the y-axis label
<code>xlab</code>	a character vector of length one used for the x-axis label

Value

a trellis object describing the plot

See Also

[barchart](#)

Examples

```
data(HELPrct)
bargraph( ~ substance, data=HELPrct)
bargraph( ~ substance, data=HELPrct, horizontal=TRUE)
bargraph( ~ substance | sex, groups=homeless, auto.key=TRUE, data=HELPrct)
bargraph( ~ substance, groups=homeless, auto.key=TRUE, data=HELPrct, subset=sex=="male")
```

Description

The `binom.test` function performs an exact test of a simple null hypothesis about the probability of success in a Bernoulli experiment from summarized data or from raw data. The mosaic `binom.test` provides wrapper functions around the function of the same name in **stats**. These wrappers provide an extended interface (including formulas).

Usage

```
binom.test(x, n, p = 0.5, alternative = c("two.sided", "less", "greater"),
  conf.level = 0.95, ...)

## S4 method for signature 'ANY'
binom.test(x, n, p = 0.5, alternative = c("two.sided",
  "less", "greater"), conf.level = 0.95, ...)

## S4 method for signature 'formula'
binom.test(x, n, p = 0.5, alternative = c("two.sided",
  "less", "greater"), conf.level = 0.95, success = NULL, data.name, data,
  ...)

## S4 method for signature 'numeric'
binom.test(x, n, p = 0.5, alternative = c("two.sided",
  "less", "greater"), conf.level = 0.95, success = NULL, data.name, ...)

## S4 method for signature 'character'
binom.test(x, n, p = 0.5, alternative = c("two.sided",
  "less", "greater"), conf.level = 0.95, success = NULL, data.name, ...)

## S4 method for signature 'logical'
binom.test(x, n, p = 0.5, alternative = c("two.sided",
  "less", "greater"), conf.level = 0.95, success = NULL, data.name, ...)

## S4 method for signature 'factor'
binom.test(x, n, p = 0.5, alternative = c("two.sided",
  "less", "greater"), conf.level = 0.95, success = NULL, data.name, ...)
```

Arguments

<code>x</code>	count of successes, length 2 vector of success and failure counts, a formula, or a character, numeric, or factor vector containing raw data.
<code>n</code>	sample size (successes + failures) or a data frame (for the formula interface)
<code>p</code>	probability for null hypothesis

<code>alternative</code>	type of alternative hypothesis
<code>conf.level</code>	confidence level for confidence interval
<code>success</code>	level of variable to be considered success. All other levels are considered failure.
<code>data.name</code>	name for data. If missing, this is inferred from variable names.
<code>data</code>	a data frame (if missing, n may be a data frame)
<code>...</code>	additional arguments (often ignored)

Details

This is a wrapper around `binom.test` from the base package to simplify its use when the raw data are available, in which case an extended syntax for `binom.test` is provided.

Value

an object of class `htest`

See Also

[prop.test](#), [binom.test](#)

Examples

```
# Several ways to get a confidence interval for the proportion of Old Faithful
# eruptions lasting more than 3 minutes.
data(faithful)
binom.test(faithful$eruptions > 3)
binom.test(97, 272)
binom.test(c(97, 272-97))
faithful$long <- faithful$eruptions > 3
binom.test(faithful$long)
binom.test(~ long, faithful)
```

Births78

US Births in 1978

Description

A day by day record of the number of births in the United States in 1978.

Usage

```
data(Births78)
```

Format

A data frame with 365 observations on the following variables.

- date date in 1978
- births number of US births
- dayofyear sequential number of days from 1 to 365

Examples

```
data(Births78)
xyplot(births ~ dayofyear, Births78)
xyplot(births ~ dayofyear, Births78, groups=dayofyear%%7)
```

Broyden

Multi-Dimensional Root Finding

Description

Implementation of Broyden's root finding function to numerically compute the root of a system of nonlinear equations

Usage

```
Broyden(system, vars, x = 0, tol = .Machine$double.eps^0.4,
maxiters = 10000)
```

Arguments

system	A list of functions
vars	A character string list of variables that appear in the functions
x	A starting vector
tol	The tolerance for the function specifying how precise it will be
maxiters	maximum number of iterations.

Cards *Standard Deck of Cards*

Description

A character vector with two or three character representations of each card in a standard 52-card deck.

Usage

Cards

Details

The 2 of clubs is represented as "2C", while the 10 of diamonds is "10D".

Examples

```
deal(Cards, 13)      # bridge hand
deal(Cards, 5)       # poker hand
shuffle(Cards)      # shuffled deck
```

cdist *Central portion of a distribution*

Description

This function determines the critical values for isolating a central portion of a distribution with a specified probability. This is designed to work especially well for symmetric distributions, but it can be used with any distribution.

Usage

```
cdist(dist, p, ..., tail = c("upper", "lower"), warn = TRUE)
```

Arguments

dist	a character string naming a distribution family (e.g., "norm"). This will work for any family for which the usual d/p/q functions exist.
p	the proportion to be in the central region, with equal proportions in either "tail".
...	additional arguments passed to the distribution functions. Typically these specify the parameters of the particular distribution desired. See the examples.
tail	one of "upper" or "lower" specifying whether the lower or upper critical value is returned.
warn	a logical indicating whether a warning should be given when using a distribution that is not symmetric.

Note

This function is still experimental and changes the input or output formats are possible in future versions of the package.

Examples

```

cdist( "norm", .95)
cdist( "t", c(.90, .95, .99), df=5)
cdist( "t", c(.90, .95, .99), df=50)
cdist( "t", .95, df=c(3,5,10,20) )
cdist( "norm", .95, mean=500, sd=100 )
cdist( "chisq", c(.90, .95), df=3 )
cdist( "chisq", c(.90, .95), df=3, tail="lower" )

```

CIsim

*Compute confidence intervals from (multiple) simulated data sets***Description**

This function automates the calculation of coverage rates for exploring the robustness of confidence interval methods.

Usage

```

CIsim(n, samples = 100, rdist = rnorm, args = list(), estimand = 0,
      conf.level = 0.95, method = t.test, method.args = list(),
      interval = function(x) { do.call(method, c(list(x, conf.level =
      conf.level), method.args))$conf.int }, estimate = function(x) {
      do.call(method, c(list(x, conf.level = conf.level), method.args))$estimate },
      verbose = TRUE)

```

Arguments

n	size of each sample
samples	number of samples to simulate
rdist	function used to draw random samples
args	arguments required by rdist
estimand	true value of the parameter being estimated
conf.level	confidence level for intervals
method	function used to compute intervals. Standard functions that produce an object of class htest can be used here.
method.args	arguments required by method
interval	a function that computes a confidence interval from data. Function should return a vector of length 2.
estimate	a function that computes an estimate from data
verbose	print summary to screen?

Value

A data frame with variables lower, upper, estimate, cover ('Yes' or 'No'), and sample is returned invisibly. See the examples for a way to use this to display the intervals graphically.

Examples

```

CIsim(10,1000) # 1000 95% intervals using t.test; population is N(0,1)
CIsim(10,1000, rdist=rexp, estimand=1) # this time population is Exp(1)
if (require(Hmisc)) {
  xYplot(Cbind(estimate,lower,upper) ~ sample,
         data=CIsim(10,100, rdist=rexp, estimand=1),
         par.settings=col.mosaic(), groups=cover)
  ladd(panel.abline(h=1))
}

```

coef.function	<i>Extract coefficients from a function</i>
---------------	---

Description

coef will extract the coefficients attribute from a function. Functions created by applying link{makeFun} to a model produced by [lm](#), [glm](#), or [nls](#) store the model coefficients there to enable this extraction.

Usage

```

## S3 method for class 'function'
coef(object, ...)

```

Arguments

object	a function
...	ignored

Examples

```

model <- lm( width ~ length, data=KidsFeet)
f <- makeFun( model )
coef(f)

```

columns	<i>return a vector of row or column indices</i>
---------	---

Description

return a vector of row or column indices

Usage

```
columns(x, default = c())
```

```
rows(x, default = c())
```

Arguments

x	an object that may or may not have any rows or columns
default	what to return if there are no rows or columns

Value

if x has rows or columns, a vector of indices, else default

Examples

```
dim(HELPrct)
columns(HELPrct)
rows(HELPrct)
columns(NULL)
columns("this doesn't have columns")
```

compareMean	<i>Compare means between 2 groups</i>
-------------	---------------------------------------

Description

A function to calculate the difference between the means two of groups for a continuous outcome variable

Usage

```
compareMean(formula, data = NULL, ...)
```

Arguments

formula	a formula
data	a data frame in which x is evaluated if x is a formula.
...	other arguments

Value

the difference in means between the second and first group

See Also

[do](#), [compareProportion](#) and [shuffle](#)

Examples

```
data(HELPrct)
# calculate the observed difference
mean(age ~ sex, data=HELPrct)
obs <- compareMean(age ~ sex, data=HELPrct); obs
# calculate the permutation distribution
nulldist <- do(100) * compareMean(age ~ shuffle(sex),
  data=HELPrct)
histogram(~ result, groups=(result >= obs), nulldist,
  xlab="difference in means")
```

compareProportion	<i>Compare proportions between 2 groups</i>
-------------------	---

Description

A function to facilitate 2 group permutation tests for a categorical outcome variable

Usage

```
compareProportion(formula, data = NULL, ...)
```

Arguments

formula	a formula
data	a data frame in which x is evaluated if x is a formula.
...	other arguments

Value

the difference in proportions between the second and first group

See Also

[do](#), [compareMean](#) and [shuffle](#)

Examples

```

data(HELPrct)
# calculate the observed difference
mean(homeless=="housed" ~ sex, data=HELPrct)
obs <- compareProportion(homeless=="housed" ~ sex, data=HELPrct); obs
# calculate the permutation distribution
nulldist <- do(100) * compareProportion(homeless=="housed" ~ shuffle(sex), data=HELPrct)
histogram(~ result, groups=(result >= obs), nulldist,
  xlab="difference in proportions")

```

confint.numeric

Confidence interval methods for output of resampling

Description

Methods for confint to compute confidence intervals on numerical vectors and numerical components of data frames.

Usage

```

## S3 method for class 'numeric'
confint(object, parm, level = 0.95, ...,
  method = "stderr", margin.of.error = "stderr" %in% method == "stderr")

## S3 method for class 'do.data.frame'
confint(object, parm, level = 0.95, ...,
  method = "stderr", margin.of.error = "stderr" %in% method)

## S3 method for class 'data.frame'
confint(object, parm, level = 0.95, ...)

```

Arguments

object	The data frame or numerical vector.
parm	not used – for compatibility with other confint methods
level	confidence level (default 0.95)
...	additional arguments (currently ignored)
method	either "stderr" (default) or "quantile". ("se" and "percentile" are allowed as aliases) or a vector containing both.
margin.of.error	if true, report intervals as a center and margin of error.

Value

When applied to a data frame, returns a data frame giving the confidence interval for each variable in the data frame using `t.test` or `binom.test`, unless the data frame was produced using `do`, in which case it is assumed that each variable contains resampled statistics that serve as an estimated sampling distribution from which a confidence interval can be computed using either a central proportion of this distribution or using the standard error as estimated by the standard deviation of the estimated sampling distribution. When applied to a numerical vector, returns a vector.

Examples

```
s <- do(500)*mean( age ~ sex, data=resample(HELPrct) )
confint(s)
confint(s, method="quantile")
confint(s, margin.of.error=TRUE)
confint(s, margin.of.error=TRUE, level=0.99 )
s2 <- do(500)*mean( resample(1:10) )
confint(s2)
```

CoolingWater

CoolingWater

Description

Temperature of a mug of water as it cools

Usage

```
data(CoolingWater)
```

Format

A data frame with 222 observations of the following variables.

- `time` time in minutes
- `temp` temperature in Celsius

Details

The water was poured into a mug and a temperature probe inserted into the water with a few seconds of the pour.

Source

These data were collected Stan Wagon to help his mathematical modeling students explore Newton's Law of Cooling and the ways that the law is really only an approximation. More about Stan: <http://stanwagon.com>.

Examples

```
data(CoolingWater)
xyplot(temp~time, data=CoolingWater)
```

Countries

Countries

Description

A data frame containing country names as used by GapMinder and the maps package to facilitate conversion between the two.

Usage

```
data(Countries)
```

Format

A data frame with 258 observations on the following variables.

- worldmap region name <http://mappinghacks.com/> data sets
- gapminder country name in GapMinder data sets
- maps region name in maps data sets

Details

The "countries" in the maps data include several other geographic regions (bodies of water, islands belonging to other countries, Hawaii, etc.) that are not countries. Furthermore, the maps countries do not include many of the countries that have been created since ca. 2000. The mapping is therefore many-to-many, and also includes some NAs when there is no appropriate mapping. Bodies of water in the maps data, for example, are not assigned a country in the GapMinder.

Examples

```
data(Countries)
subset(Countries, maps=="Yugoslavia") # Where has Yugoslavia gone?
subset(Countries, is.na(gapminder))   # Things from maps with no GapMinder equivalent
subset(Countries, is.na(maps))        # Things from GapMinder with no maps equivalent
```

CPS85

Data from the 1985 Current Population Survey (CPS85)

Description

The Current Population Survey (CPS) is used to supplement census information between census years. These data consist of a random sample of persons from the CPS85, with information on wages and other characteristics of the workers, including sex, number of years of education, years of work experience, occupational status, region of residence and union membership.

Usage

```
data(CPS85)
```

Format

A data frame with 534 observations on the following variables.

- wage wage (US dollars per hour)
- educ number of years of education
- race a factor with levels NW (nonwhite) or W (white)
- sex a factor with levels F M
- hispanic a factor with levels Hisp NH
- south a factor with levels NS S
- married a factor with levels Married Single
- exper number of years of work experience (inferred from age and educ)
- union a factor with levels Not Union
- age age in years
- sector a factor with levels clerical const manag manuf other prof sales service

Details

Data are from 1985. The data file is recoded from the original, which had entirely numerical codes.

Source

Data are from http://lib.stat.cmu.edu/datasets/CPS_85_Wages.

References

Berndt, ER. *The Practice of Econometrics* 1991. Addison-Wesley.

Examples

```
data(CPS85)
```

cross *Factor cross products*

Description

Construct a product of factors.

Usage

```
cross(..., sep = ":", drop.unused.levels = FALSE)
```

Arguments

... factors to be crossed.
 sep separator between levels
 drop.unused.levels should levels that do not appear in cross product be dropped?

Value

a factor

Examples

```
x <- letters[1:3]
y <- c(1,2,1,1,3,1,3)
cross(x, y)
cross(x, y, drop.unused.levels=TRUE)
```

D *Derivative and Anti-derivative operators*

Description

Operators for computing derivatives and anti-derivatives as functions.

Usage

```
D(formula, ..., .hstep = NULL, add.h.control = FALSE)
antiD(formula, ..., lower.bound = 0, force.numeric = FALSE)
makeAntiDfun(.function, .wrt, from, .tol = .Machine$double.eps^0.25)
numerical_integration(f, wrt, av, args, vi.from, ciName = "C", .tol)
```

Arguments

formula	A formula. The right side specifies the variable(s) with which to carry out the integration or differentiation. On the left side should be an expression or a function that returns a numerical vector of the same length as its argument. The expression can contain unbound variables.
...	Default values to be given to unbound variables in the expression <code>expr</code> . See examples.#' Note that in creating anti-derivative functions, default values of "from" and "to" can be assigned. They are to be written with the name of the variable as a prefix, e.g. <code>y.from</code> .
<code>.hstep</code>	horizontal distance between points used for secant slope calculation in numerical derivatives.
<code>add.h.control</code>	logical indicating whether the returned derivative function should have an additional parameter for setting <code>.hstep</code> . Meaningful only for numerical derivatives.
<code>lower.bound</code>	for numerical integration only, the lower bound used
<code>force.numeric</code>	If TRUE, a numerical integral is performed even when a symbolic integral is available.
<code>.function</code>	function to be integrated
<code>.wrt</code>	character string naming the variable of integration
<code>from</code>	default value for the lower bound of the integral region
<code>f</code>	a function
<code>wrt</code>	character string naming a variable: the var. of integration
<code>av</code>	a list of the arguments passed to the function calling this
<code>args</code>	default values (if any) for parameters
<code>vi.from</code>	the the lower bound of the interval of integration
<code>ciName</code>	character string giving the name of the symbol for the constant of integration
<code>.tol</code>	Numerical tolerance. See <code>stats::integrate</code>

Details

D attempts to find a symbolic derivative for simple expressions, but will provide a function that is a numerical derivative if the attempt at symbolic differentiation is unsuccessful. The symbolic derivative can be of any order (although the expression may become unmanageably complex). The numerical derivative is limited to first or second-order partial derivatives (including mixed partials). `antiD` will attempt simple symbolic integration but if it fails it will return a numerically-based anti-derivative.

`antiD` returns a function with the same arguments as the expression passed to it. The returned function is the anti-derivative of the expression, e.g., `antiD(f(x)~x) -> F(x)`. To calculate the integral of `f(x)`, use `F(to) - F(from)`.

Value

For derivatives, the return value is a function of the variable(s) of differentiation, as well as any other symbols used in the expression. Thus, `D(A*x^2 + B*y ~ x + y)` will compute the mixed

partial with respect to x then y (that is, $\frac{d^2 f}{dy dx}$). The returned value will be a function of x and y, as well as A and B. In evaluating the returned function, it's best to use the named form of arguments, to ensure the order is correct.

a function of the same arguments as the original expression with a constant of integration set to zero by default, named "C", "D", ... depending on the first such letter not otherwise in the argument list.

Note

numerical_integration is not intended for direct use. It packages up the numerical anti-differentiation process so that the contents of functions produced by antiD look nicer to human readers.

Examples

```
D(sin(t) ~ t)
D(A*sin(t) ~ t)
D(A*sin(2*pi*t/P) ~ t, A=2, P=10) # default values for parameters.
f <- D(A*x^3 ~ x + x, A=1) # 2nd order partial -- note, it's a function of x
f(x=2)
f(x=2,A=10) # override default value of parameter A
g <- D(f(x=t, A=1)^2 ~ t) # note: it's a function of t
g(t=1)
gg <- D(f(x=t, A=B)^2 ~ t, B=10) # note: it's a function of t and B
gg(t=1)
gg(t=1, B=100)
f <- makeFun(x^2~x)
D(f(cos(z))~z) #will look in user functions also
antiD( a*x^2 ~ x)
antiD( A/x~x )
F <- antiD( A*exp(-k*t^2 ) ~ t, A=1, k=0.1)
F(t=Inf)
one = makeFun(1~x&y)
by.x = antiD( one(x=x, y=y) ~x)
by.xy = antiD(by.x(x=sqrt(1-y^2), y=y)~y)
4*by.xy(y=1) #area of quarter circle
```

deg2rad

Convert between degrees and radians

Description

Facilitates conversion between degrees and radians.

Usage

```
deg2rad(x)
```

```
rad2deg(x)
```

Arguments

x a numeric vector

Value

a numeric vector

See Also

[latlon2xyz](#), [googleMap](#), and [rgeo](#).

Examples

```
deg2rad(180)
rad2deg(2*pi)
```

deltaMethod

Delta method on data frames

Description

An expansion of the capabilities of [deltaMethod](#) from the **car** package.

Usage

```
## S3 method for class 'data.frame'
deltaMethod(object, g, uncertainties,
  estimates = measurements, func = g, constants = c(),
  measurements = NULL, vcov., ...)
```

Arguments

object a data frame containing measured quantities

g a quoted string that describes the function of the parameter estimates to be evaluated; see [deltaMethod](#) for details.

uncertainties a data frame with the same dimension as object or numeric vector of length `ncol(object)` containing the uncertainties on each measured value in object or a matrix providing a variance-covariance matrix for the uncertainties. Alternatively, if estimates is not NULL, then uncertainties may be a vector of names or integers used to select columns from object. There is one potentially ambiguous case: It is not possible to specify the uncertainties as a vector of integers if estimates is not NULL – such integers will be treated as column numbers for subsetting. If uncertainties is not a matrix, independence is assumed and the variance-covariance matrix is created under that assumption. Matching of uncertainties to measured values is by position, so names are irrelevant. Uncertainties will be converted into a covariance matrix assuming independence.

vcov.	a covariance matrix or a list of covariance matrices. Only one of vcov. and uncertainties may be defined.
func	a quoted string used to annotate output. The default of func = g is usually appropriate.
constants	This argument is a named vector whose elements are constants that are used in the f argument. This is needed only when the function is called from within another function to comply to R scoping rules.
estimates	a vector of column names or column numbers used to specify a subset of object containing the measured/estimated quantities.
measurements	an alternative name for estimates
...	additional arguments passed through to deltaMethod in the car package.

See Also

deltaMethod in the **car** package.

Examples

```
C_p <- 4.182 / 60 # / 60 because measureing m in L/min
exprforQ <- "(T.cold.out - T.cold.in) * C_p * m.cold"
deltaMethod( HeatX[, c("T.cold.in", "T.cold.out", "m.cold")], exprforQ, c(1,1,.5) )
# This is just wordier in this example, but would allow the uncertainties to vary
# from row to row.

HeatX3 <- transform(HeatX,
                    u.cold.in=1, u.cold.out=1, u.hot.in=1, u.hot.out=1,
                    u.m.cold=0.5, u.m.hot=0.5)
deltaMethod( HeatX3[, c("T.cold.in", "T.cold.out", "m.cold")], exprforQ,
            HeatX3[, c("u.cold.in", "u.cold.out", "u.m.cold")])
# Rather than specifying two data frames, we can use subsetting instead
deltaMethod( HeatX3, exprforQ,
            estimates=c("T.cold.in", "T.cold.out", "m.cold"),
            uncertainties=c("u.cold.in", "u.cold.out", "u.m.cold"))
# Can also specify vcov. as a matrix or list of matrices:
deltaMethod(HeatX[, c("T.cold.in", "T.cold.out", "m.cold")], exprforQ,
            vcov. = diag(c(1,1,.5)^2) )
deltaMethod(HeatX[, c("T.cold.in", "T.cold.out", "m.cold")], exprforQ,
            vcov. = list( diag(c(1,1,.5)^2), diag(c(1,2,.8)^2) ) )
```

derivedFactor

Create factors from logicals

Description

A utility function for creating new factors from logicals describing the levels

Usage

```
derivedFactor(..., .ordered = FALSE, .method = c("unique", "first", "last"),
  .debug = c("default", "always", "never"), .sort = c("given", "alpha"),
  .default = NULL)
```

Arguments

<code>...</code>	named logical "rules" defining the levels of the factor.
<code>.ordered</code>	a logical indicating whether the resulting factored should be ordered
<code>.method</code>	one of "unique", "first", and "last". If "unique", exactly one rule must be TRUE for each position. If "first", the first TRUE rule defines the level. If "last", the last TRUE rule defines the level.
<code>.debug</code>	one of "default", "always", and "never", indicating whether debugging information should be printed. If "default", debugging information is printed only when multiple rules give conflicting definitions for some positions.
<code>.sort</code>	One of "given" (the default) or "alpha" or a vector of integers the same length as the number of levels indicating the order in which the levels should appear in the resulting factor.
<code>.default</code>	character vector of length 1 giving name of default level or NULL for no default.

Details

Each logical "rule" corresponds to a level in the resulting factor. If `.default` is defined, an implicit rule is added that is TRUE whenever all other rules are FALSE. When there are multiple TRUE rules for a slot, the first or last such is used or an error is generated, depending on the value of `method`.

`derivedFactor` is designed to be used with `transform` to add new factor variables to a data frame. See the examples.

Examples

```
Kf <- transform(KidsFeet, biggerfoot2=derivedFactor(
  dom = biggerfoot == domhand,
  nondom = biggerfoot != domhand)
)
tally( ~biggerfoot + biggerfoot2, data=Kf)
tally( ~biggerfoot + domhand, data=Kf)

# Three equivalent ways to define a new variable
# Method 1: explicitly define all levels
modHELP <- transform(HELPrct, drinkstat = derivedFactor(
  abstinent = i1 == 0,
  moderate = (i1>0 & i1<=1 & i2<=3 & sex=='female') |
    (i1>0 & i1<=2 & i2<=4 & sex=='male'),
  highrisk = ((i1>1 | i2>3) & sex=='female') |
    ((i1>2 | i2>4) & sex=='male'),
  .ordered = TRUE)
)
tally( ~drinkstat, data=modHELP )
```

```

# Method 2: Use .default for last level
modHELP <- transform(HELPrct, drinkstat = derivedFactor(
  abstinent = i1 == 0,
  moderate = (i1<=1 & i2<=3 & sex=='female') |
    (i1<=2 & i2<=4 & sex=='male'),
  .ordered = TRUE,
  .method = "first",
  .default = "highrisk")
)
tally( ~drinkstat, data=modHELP )

# Method 3: use TRUE to catch any fall through slots
modHELP <- transform(HELPrct, drinkstat = derivedFactor(
  abstinent = i1 == 0,
  moderate = (i1<=1 & i2<=3 & sex=='female') |
    (i1<=2 & i2<=4 & sex=='male'),
  highrisk=TRUE,
  .ordered = TRUE,
  .method = "first"
)
)
tally( ~drinkstat, data=modHELP )

```

dfapply

apply-type function for data frames

Description

An apply-type function for data frames.

Usage

```
dfapply(data, FUN, select = is.numeric, ...)
```

Arguments

data	data frame
FUN	a function to apply to (some) variables in the data frame
select	function used to select variables to which FUN is applied. See examples.
...	arguments passed along to FUN

See Also

[apply](#), [sapply](#), [tapply](#), [lapply](#)

Examples

```
dfapply(iris, favstats)
dfapply(HELPrct, table, select=is.factor)
```

do *Do Things Repeatedly*

Description

do() provides a natural syntax for repetition tuned to assist with replication and resampling methods.

Usage

```
do(n = 1L, cull = NULL, mode = "default", algorithm = 1,
   parallel = TRUE)
```

```
## S4 method for signature 'repeater'
print(x, ...)
```

```
## S4 method for signature 'repeater,ANY'
e1 * e2
```

Arguments

n	number of times to repeat
cull	function for culling output of objects being repeated. If NULL, a default culling function is used. The default culling function is currently aware of objects of types lme, lm, htest, table, cointoss, and matrix.
mode	target mode for value returned
algorithm	a number used to select the algorithm used. Currently numbers below 1 use an older algorithm and numbers ≥ 1 use a newer algorithm which is faster in some situations.
parallel	a logical indicating whether parallel computation should be attempted using the parallel package (if it is installed).
e1	an object (in cases documented here, the result of running do)
e2	an object (in cases documented here, an expression to be repeated)
x	an object used to select a method.
...	further arguments passed to or from other methods.

Value

do returns an object of class repeater which is only useful in the context of the operator *. See the examples.

Author(s)

Daniel Kaplan (<kaplan@macalaster.edu>) and Randall Pruim (<rpruim@calvin.edu>)

See Also[replicate](#)**Examples**

```
do(3) * rnorm(1)
do(3) * "hello"
do(3) * lm(shuffle(height) ~ sex + mother, Galton)
do(3) * anova(lm(shuffle(height) ~ sex + mother, Galton))
do(3) * 1:4
do(3) * mean(rnorm(25))
do(3) * c(sample.mean = mean(rnorm(25)))
do(3) * tally( ~sex|treat, data=resample(HELPrct))
```

dotPlot

*Dotplots***Description**

A high level function and panel function for producing a variant of a histogram called a dotplot.

Usage

```
dotPlot(x, breaks, ..., panel = panel.dotPlot)
```

```
panel.dotPlot(x, breaks, equal.widths = TRUE, groups = NULL, nint = if
  (is.factor(x)) nlevels(x) else round(1.3 * log2(length(x)) + 4), pch, col,
  lty = trellis.par.get("dot.line")$lty,
  lwd = trellis.par.get("dot.line")$lwd,
  col.line = trellis.par.get("dot.line")$col,
  alpha = trellis.par.get("dot.symbol")$alpha, cex = 1, type = "count",
  ...)
```

Arguments

x	a vector of values or a formula
nint	the number of intervals to use
panel	a panel function
breaks, equal.widths, groups, pch, col, lty, lwd, col.line, type, alpha	as in histogram
cex	a ratio by which to increase or decrease the dot size
...	additional arguments

Value

a trellis object

See Also[histogram](#)**Examples**

```
dotPlot( ~ age, data = HELPrct)
dotPlot( ~ age, nint=42, data = HELPrct)
dotPlot( ~ height | voice.part, data = singer, nint = 17,
         endpoints = c(59.5, 76.5), layout = c(2,4), aspect = 1,
         xlab = "Height (inches)")
```

ediff*Lagged Differences with equal length*

Description

Often when creating lagged differences, it is awkward that the differences vector is shorter than the original. `ediff` pads with NAs to make its output the same length as the input.

Usage

```
ediff(x, lag = 1, differences = 1, frontPad = TRUE, ...)
```

Arguments

<code>x</code>	a numeric vector or a matrix containing the values to be differenced
<code>lag</code>	an integer indicating which lag to use
<code>differences</code>	an integer indicating the order of the difference
<code>frontPad</code>	logical indicating whether padding is on the front (default) or back end.
<code>...</code>	further arguments to be passed to or from methods

See Also

[diff](#) since `ediff` is a thin wrapper around `diff`.

Examples

```
ediff(1:10)
ediff(1:10, 2)
ediff(1:10, 2, 2)
x <- cumsum(cumsum(1:10))
ediff(x, lag = 2)
ediff(x, differences = 2)
ediff(.leap.seconds)
Men <- subset(SwimRecords, sex=="M")
Men <- transform(Men, change=ediff(time), interval=ediff(year))
head(Men)
```

evalFormula	<i>Evaluate a formula</i>
-------------	---------------------------

Description

Evaluate a formula

Usage

```
evalFormula(formula, data = parent.frame(), subset, ops = c("+", "&"))
```

Arguments

formula	a formula ($y \sim x \mid z$) to evaluate
data	a data frame or environment in which evaluation occurs
ops	a vector of operator symbols allowable to separate variables in rhs
subset	an optional vector describing a subset of the observations to be used. Currently only implemented when data is a data frame.

Value

a list containing data frames corresponding to the left, right, and condition slots of formula

Examples

```
data(CPS85)
cps <- CPS85[1:6,]
cps
evalFormula(wage ~ sex & married & age | sector & race, data=cps)
```

evalSubFormula	<i>Evaluate a part of a formula</i>
----------------	-------------------------------------

Description

Evaluate a part of a formula

Usage

```
evalSubFormula(x, data = parent.frame(), ops = c("+", "&"))
```

Arguments

x	an object appearing as a subformula (typically a call)
data	a data fram or environment in which things are evaluated
ops	a vector of operators that are not evaluated as operators but instead used to further split x

Value

a data frame containing the terms of the evaluated subformula

Examples

```
data(CPS85)
cps <- CPS85[1:6,]
cps
evalSubFormula( rhs( ~ married & sector), data=cps )
```

expandFun

Expand the left-hand side of a formula

Description

Expands the contents of functions used in a formula.

Usage

```
expandFun(formula, ...)
```

Arguments

formula A mathematical expression (see examples and [plotFun](#))
... additional parameters

Value

A list with the new expanded formula and the combined formals

Examples

```
f=makeFun(x^2~x)
expandFun(f(z)~z) #Returns z^2~z
```

fav_stats	<i>Some favorite statistical summaries</i>
-----------	--

Description

Likely you mean to be using `favstats`. Each of these computes the mean, standard deviation, quartiles, sample size and number of missing values for a numeric vector, but `favstats` can take a formula describing how these summary statistics should be aggregated across various subsets of the data.

Usage

```
fav_stats(x, ..., na.rm = TRUE)
```

Arguments

<code>x</code>	numeric vector
<code>na.rm</code>	boolean indicating whether missing data should be ignored
<code>...</code>	additional arguments (currently ignored)

Value

A vector of statistical summaries

Examples

```
fav_stats(1:10)
fav_stats(faithful$eruptions)
favstats(length ~ sex, data=KidsFeet) # Note: this is favstats() rather than fav_stats()
```

fetchData	<i>A Web and Library Data-Loading Facility</i>
-----------	--

Description

`fetchData` provides a means for students and others to locate and load data sets and R commands provided by instructors. Data can be pre-loaded for off-line sessions, can be positioned on identified web sites, or loaded from packages. `fetchData` also will load local files using a complete path name (relative to the current working directory) or, if no name is given, via a dialog box.

Usage

```
fetchData(name = NULL, show.path = FALSE, add.to.path = NULL,
          drop.from.path = NULL, cache = FALSE, verbose = TRUE)
```

Arguments

name	a character string naming a data set. This will often end in .csv for reading in a data set. If it ends in .r or .R, the file will be "sourced".
show.path	If TRUE, causes the current search path to be returned
add.to.path	Name of a web directory (ending in /), which should be pre-pended to the search path.
drop.from.path	Name of a web directory to be deleted from the path.
cache	If TRUE, indicates that a data set is to be pre-loaded into the cached library. This allows, for instance, users to pre-load on-line data to be used when they are off-line.
verbose	a logical indicating whether additional status messages (e.g., indicating where the dataset was located) should be printed.

Details

There are two major purposes for this function. One is to provide a consistent interface to reading data: a file name is given and a data frame is returned, which can be assigned to an object as the user desires. This differs from the behavior of `data`, which doesn't return a value but instead creates an object without explicit assignment.

The other purpose is to allow instructors or other group leaders to post data and R code on web sites that can be searched as naturally as if the data were on the users' own machines. For instance, an instructor might want to post a new data set just before class, enabling her students to access it in class.

To support this, `fetchData` allows new web sites to be added to the web search path. Typically, the command to add a site would be in a script file that is provided to the student that could be run automatically at start up or sourced over the web. That is, an instructor might create a script file stored on a website and, using a web page, provide students with the text of the command to source it.

Currently, https addresses are changed to http

Value

a data frame.

Examples

```
## Not run: dome <- fetchData("Dome.csv")
## Not run: carbon <- fetchData("CO2")
## Not run: fetchData(show=TRUE)
## Not run: fetchData(add.to.path="http://www.macalester.edu/~kaplan/ISM/datasets/")
## Not run: fetchData(drop.from.path="http://www.macalester.edu/~kaplan/ISM/datasets/")
## Not run: dome <- fetchData("Dome.csv", cache=TRUE)
```

fetchGapminder1	<i>Fetch Gapminder data</i>
-----------------	-----------------------------

Description

Fetch data originally obtained from Gapminder.

Usage

```
fetchGapminder1(name, value.name = NULL)
fetchGapminder(..., all.cases = TRUE, all.vars = FALSE)
```

Arguments

name	a character vector of length 1
value.name	name of variable in resulting data frame
all.cases	a logical indicating whether all cases should be included or only complete cases
all.vars	a logical, if TRUE, then ... is ignored and all available variables are fetched.
...	character strings naming desired variables

Value

A data frame

fetchGoogle	<i>Fetch data from a web service</i>
-------------	--------------------------------------

Description

Read a data set generated from a web service such as Google Docs.

Usage

```
fetchGoogle(URL, key = NULL)
```

Arguments

URL	the URL to retrieve a CSV file from the service
key	for convenience, just the "key" part of the Google link

Details

Web services such as Google Docs allow you to store spreadsheets "in the cloud". By setting permissions in the service, you can arrange to make the data set public, so that anyone with an appropriate URL can access the data. Reading such data into R can be done simply if the service supports exporting the data in a CSV format via URL link. For instance, Google Spreadsheets can be set up to publish a spreadsheet via a URL. Unfortunately, the `read.csv()` function, although able to read URLs pointing to a file, cannot handle the protocol needed to talk to services such as Google Docs. `fetchGoogle()` allows you to do this. `fetchGoogle()` derives its functionality from the RCurl package, which must be installed for the function to work. RCurl will be loaded automatically if it is installed. Generating the URL from the web service will, of course, depend on how that service is set up. For Google Spreadsheets, you, the owner of a spreadsheet, can (1) open the spreadsheet in a browser (2) select the File/Publish to the Web menu item (3) in the resulting dialog box, press "Start publishing" (4) under "Get a link to the published data", select CSV format (5) copy the `https://docs.google.com/spreadsheet/pub?...` link and post it where your users can get to it.

Note

The URL must instruct the service to generate a CSV file. The URLs from Google Docs are very long and contain random-looking sequences. You may want to post the URL on a web page whence it can be cut and paste as part of the command. The `key=` argument is provided as a convenience so that a shorter character string can be used to refer to a Google document. Use URL rather than key if you are using a non-Google service or if the Google interface changes. `fetchData()` expects the spreadsheet to be in a straightforward rectangular spreadsheet format.

Author(s)

Daniel Kaplan (<kaplan@macalester.edu>)

Examples

```
## Not run: s = fetchGoogle(key="0Am13enSa1074dEVzMGJSMU5TbTc2eW1WakppQ1pjcGc")
```

findZeros

Find zeros of functions

Description

Compute numerically zeros of a function or simultaneous zeros of multiple functions.
Solve an equation

Usage

```
findZeros(expr, ..., xlim = c(near - within, near + within), near = 0,
  within = Inf, nearest = 10, npts = 1000, iterate = 1,
  sortBy = c("byx", "by", "radial"))
```



```
## S3 method for class 'formula'
solve(form, ..., near = 0, within = Inf, nearest = 10,
      npts = 1000, iterate = 1, sortBy = c("byx", "byy", "radial"))
```

Arguments

expr	A formula. The right side names the variable with respect to which the zeros should be found. The left side is an expression, e.g. $\sin(x) \sim x$. All free variables (all but the variable on the right side) named in the expression must be assigned a value via ...
...	Formulas corresponding to additional functions to use in simultaneous zero finding and/or specific numerical values for the free variables in the expression.
xlim	The range of the dependent variable to search for zeros. Inf is a legitimate value, but is interpreted in the numerical sense as the non-Inf largest floating point number. This can also be specified replacing x with the name of the variable. See the examples.
near	a value near which zeros are desired
within	only look for zeros at least this close to near. near and within provide an alternative to using xlim to specify the search space.
nearest	the number of nearest zeros to return. Fewer are returned if fewer are found.
iterate	maximum number of times to iterate the search. Subsequent searches take place with the range of previously found zeros. Choosing a large number here is likely to kill performance without improving results, but a value of 1 (the default) or 2 works well when searching in $c(-\text{Inf}, \text{Inf})$ for a modest number of zeros near near.
npts	How many sub-intervals to divide the xlim into when looking for candidates for zeros. The default is usually good enough. If Inf is involved, the intervals are logarithmically spaced up to the largest finite floating point number. There is no guarantee that all the roots will be found.
sortBy	specifies how the zeros found will be sorted. Options are 'byx', 'byy', or 'radial'.
form	Expression to be solved

Details

Searches numerically using uniroot.

Uses findZerosMult of findZeros to solve the given expression

Value

A dataframe of zero or more numerical values. Plugging these into the expression on the left side of the formula should result in values near zero.

a dataframe with solutions to the expression.

Author(s)

Daniel Kaplan (<kaplan@macalester.edu>)

Cecylia Bocovich

Examples

```

findZeros( sin(t) ~ t, xlim=c(-10,10) )
# Can use tlim or t.lim instead of xlim if we prefer
findZeros( sin(t) ~ t, tlim=c(-10,10) )
findZeros( sin(theta) ~ theta, near=0, nearest=20)
findZeros( A*sin(2*pi*t/P) ~ t, xlim=c(0,100), P=50, A=2)
# Interval of a normal at half its maximum height.
findZeros( dnorm(x,mean=0,sd=10) - 0.5*dnorm(0,mean=0,sd=10) ~ x )
# A pathological example
# There are no "nearest" zeros for this function. Each iteration finds new zeros.
f <- function(x) { if (x==0) 0 else sin(1/x) }
findZeros( f(x) ~ x, near=0 )
# Better to look nearer to 0
findZeros( f(x) ~ x, near=0, within=100 )
findZeros( f(x) ~ x, near=0, within=100, iterate=0 )
findZeros( f(x) ~ x, near=0, within=100, iterate=3 )
# Zeros in multiple dimensions (not run: these take a long time)
# findZeros(x^2+y^2+z^2-5~x&y&z, nearest=3000, within = 5)
# findZeros(x*y+z^2~z&y&z, z+y~x&y&z, npts=10)
solve(3*x==3~x)

# plot out sphere (not run)
# sphere = solve(x^2+y^2+z^2==5~x&y&z, within=5, nearest=1000)
# cloud(z~x+y, data=sphere)

```

findZerosMult

Find the zeros of a function of two or more variables

Description

Compute numerically zeros of a function of two or more variables. All free variables (all but the variable on the right side) named in the expression must be assigned a value via . . .

Usage

```
findZerosMult(..., npts = 10, rad = 5, near = 0, sortBy = "byx")
```

Arguments

...	arguments for values NOTE: if the system has more than one equation and the rhs variables do not match up, there will be an error.
npts	number of desired zeros to return
rad	radius around near in which to look for zeros
near	center of search for zeros
sortBy	options for sorting zeros for plotting. Options are 'byx', 'byy' and 'radial'. The default value is 'byx'.

Details

sorts points in the domain according to the sign of the function value at respective points. Use continuity and uniroot to find zeros between points of opposite signs. Returns any number of points which may be sorted and plotted according to x, y, or radial values.

Value

A data frame of numerical values which should all result in a value of zero when input into original function

Author(s)

Cecylia Bocovich

Examples

```
findZerosMult(a*x^2-8~a&x, npts = 50)
findZerosMult(a^2+x^2-8~a&x, npts = 100, sortBy='radial')
## Not run: findZerosMult(a^2+x^2-8~a&x, npts = 1000, sortBy='radial')
```

fitModel

Fit a nonlinear least squares model

Description

Allows you to specify a formula with parameters, along with starting guesses for the parameters. Refines those guesses to find the least-squares fit.

Usage

```
fitModel(formula, data = parent.frame(), start = list(), ...)
```

```
model(object, ...)
```

```
## S3 method for class 'nlsfunction'
model(object, ...)
```

```
## S3 method for class 'nlsfunction'
summary(object, ...)
```

```
## S3 method for class 'nlsfunction'
coef(object, ...)
```

Arguments

formula	formula specifying the model
data	dataframe containing the data to be used
start	passed as start to nls . If an empty list, a simple starting point is used (thus avoiding the usual warning message).
...	additional arguments passed to nls
object	an R object (typically a the result of fitModel)

Details

Fits a nonlinear least squares model to data. In contrast to linear models, all the parameters (including linear ones) need to be named in the formula. The function returned simply contains the formula together with pre-assigned arguments setting the parameter value. Variables used in the fitting (as opposed to parameters) are unassigned arguments to the returned function.

Value

a function

Note

This doesn't work with categorical explanatory variables.

See Also

[linearModel](#), [nls](#)

Examples

```
f <- fitModel(temp ~ A+B*exp(-k*time), data=CoolingWater, start=list(A=50,B=50,k=1/20))
f(time=50)
coef(f)
summary(f)
model(f)
```

fitSpline

Fit splines to data

Description

These functions create mathematical functions from data, using splines.

Usage

```
fitSpline(formula, data = parent.frame(), df = NULL, knots = NULL,
          degree = 3, type = c("natural", "linear", "cubic", "polynomial"), ...)
```

Arguments

formula	a formula. Only one quantity is allowed on the left-hand side, the output quantity
data	a data frame in which formula is evaluated.
type	type of splines to use; one of "linear", "cubic", "natural" (cubic with linear tails, the default), or "polynomial".
degree	parameter for splines when type is "polynomial". 1 is locally linear, 2 is locally quadratic, etc.
df	degrees of freedom (used to determine how many knots should be used)
knots	a vector of knots
...	additional arguments passed to spline basis functions (ns and bs).

Value

a function of the explanatory variable

See Also

[bs](#) and [ns](#) for the bases used to generate the splines.

Examples

```
f <- fitSpline( weight ~ height, data=women, df=5 )
xyplot( weight ~ height, data=women )
plotFun(f(height) ~ height, add=TRUE)

g <- fitSpline( height ~ weight, Heightweight, type='natural', df=5 )
h <- fitSpline( height ~ weight, Heightweight, type='linear', df=5 )
xyplot( height ~ weight, Heightweight, col='gray70', pch=16)
plotFun(g, add=TRUE, col='navy')
plotFun(h, add=TRUE, col='red')
```

freqpolygon

Frequency Polygons

Description

Frequency polygons are an alternative to histograms that make it simpler to overlay multiple distributions.

Usage

```
freqpolygon(x, ..., panel = panel.freqpolygon)
```

```
panel.freqpolygon(x, plot.points = "jitter", ref = FALSE, groups = NULL,
  weights = NULL, jitter.amount = 0.01 * diff(current.panel.limits())$ylim),
  type = "density", breaks = NULL, nint = 1.5 * log2(length(x) + 1),
```

```
center = NULL, width = NULL,
gcol = trellis.par.get("reference.line")$gcol,
glwd = trellis.par.get("reference.line")$glwd, h, v, ...,
identifier = "density")
```

Arguments

x	a formula or a numeric vector
...	additional arguments passed on to histogram and panel .
panel	a panel function
plot.points	one of TRUE, FALSE, "jitter", or "rug" indicating how points are to be displayed
gcol	color of guidelines
glwd	width of guidelines
groups, weights, jitter.amount, identifier	as in densityplot or histogram
type	one of 'density', 'percent', or 'count'
breaks	a vector of breaks for the frequency polygon bins
nint	an approximate number of bins for the frequency polygon
center	center of one of the bins
width	width of the bins
h, v	a vector of values for additional horizontal and vertical lines
ref	a logical indicating whether a horizontal reference line should be added (roughly equivalent to h=0)

Details

These functions are still under development. Future improvements may be forthcoming.

Value

a trellis object

Examples

```
freqpolygon(~age | substance, HELPrct, v=35, fit='normal')
freqpolygon(~age, HELPrct, labels=TRUE, type='count')
freqpolygon(~age, HELPrct, groups=cut(age, seq(10,80,by=10)))
freqpolygon(~age, HELPrct, groups=sex, stripes='horizontal')
freqpolygon(~racegrp, HELPrct, groups=substance, auto.key=TRUE)
## comparison of histogram and frequency polygon
histogram(~eruptions, faithful, type='density', width=.5)
ladd( panel.freqpolygon(faithful$eruptions, width=.5 ))
```

FunctionsFromData *Create function from data*

Description

These functions create mathematical functions from data, by smoothing, splining, or linear combination (fitting). Each of them takes a formula and a data frame as an argument

Usage

```
spliner(formula, data = NULL, method = "fmm", monotonic = FALSE)
connector(formula, data = NULL, method = "linear")
smoother(formula, data, span = 0.5, degree = 2, ...)
linearModel(formula, data, ...)
```

Arguments

formula	a formula. Only one quantity is allowed on the left-hand side, the output quantity
data	a data frame
method	a method for splining. See spline .
monotonic	a TRUE/FALSE flag specifying whether the spline should respect monotonicity in the data
span	parameter to smoother. How smooth it should be.
degree	parameter to smoother. 1 is locally linear, 2 is locally quadratic.
...	additional arguments to loess or lm

Details

These functions use data to create a mathematical, single-valued function of the inputs. All return a function whose arguments are the variables used on the right-hand side of the formula. If the formula involves a transformation, e.g. `sqrt(age)` or `log(income)`, only the variable itself, e.g. `age` or `income`, is an argument to the function. `linearModel` takes a linear combination of the vectors specified on the right-hand side. It differs from `project` in that `linearModel` returns a function whereas `project` returns the coefficients. NOTE: An intercept term is not included unless that is explicitly part of the formula with `+1`. This conflicts with the standard usage of formulas as found in `lm`. `spliner` and `connector` currently work for only one input variable.

See Also

[project](#) method for formulas

Examples

```

data(CPS85)
f <- smoother(wage ~ age, span=.9, data=CPS85)
f(40)
df <- D(f(age) ~ age)
df(40)
g <- linearModel(log(wage) ~ age+educ+1, data=CPS85)
g(age=40, educ=12)
dgdeduc <- D(g(age=age, educ=educ) ~ educ)
dgdeduc(age=40, educ=12)
x<-1:5; y=c(1, 2, 4, 8, 8.2)
f1 <- spliner(y ~ x)
f1(x=8:10)
f2 <- connector(x~y)

```

Galton

*Galton's dataset of parent and child heights***Description**

In the 1880's, Francis Galton was developing ways to quantify the heritability of traits. As part of this work, he collected data on the heights of adult children and their parents.

Usage

```
data(Galton)
```

Format

A data frame with 898 observations on the following variables.

- family a factor with levels for each family
- father the father's height (in inches)
- mother the mother's height (in inches)
- sex the child's sex: F or M
- height the child's height as an adult (in inches)
- nkids the number of adult children in the family, or, at least, the number whose heights Galton recorded.

Details

Entries were deleted for those children whose heights were not recorded numerically by Galton, who sometimes used entries such as "tall", "short", "idiotic", "deformed" and so on.

Source

The data were transcribed by J.A. Hanley who has published them at <http://www.medicine.mcgill.ca/epidemiology/hanley/galton/>

References

"Transmuting" women into men: Galton's family data on human stature. (2004) *The American Statistician*, 58(3):237-243.

Examples

```
data(Galton)
```

Gestation

Data from the Child Health and Development Studies

Description

Birth weight, date, and gestational period collected as part of the Child Health and Development Studies in 1961 and 1962. Information about the baby's parents — age, education, height, weight, and whether the mother smoked is also recorded.

Usage

```
data(Gestation)
```

Format

A data frame with 1236 observations on the following variables.

- id identification number
- plurality 5 = single fetus
- outcome 1 = live birth that survived at least 28 days
- date birth date where 1096=January 1, 1961
- gestation length of gestation (in days)
- sex infant's sex (1=male, 2=female)
- wt birth weight (in ounces)
- parity total number of previous pregnancies (including fetal deaths and still births)
- race mother's race: 0=white 6=mex 7=black 8=asian 9=mixed
- age mother's age in years at termination of pregnancy
- ed mother's education: 0= less than 8th grade, 1 = 8th -12th grade - did not graduate, 2= HS graduate—no other schooling, 3= HS+trade, 4=HS+some college, 5=College graduate, 6=Trade school, 7=HS unclear
- ht mother's height in inches to the last completed inch
- wt. 1 mother's prepregnancy weight (in pounds)
- drace father's race (a factor with levels equivalent to mother's race)
- dage father's age (in years)
- ded father's education (same coding as mother's education)

- dht father's height in inches to the last completed inch
- dwt father's weight (in pounds)
- marital marital status: 1=married, 2=legally separated, 3=divorced, 4=widowed, 5=never married
- inc family yearly income in \$2500 increments: 0=under 2500, 1=2500-4999, ..., 8=12,500-14,999, 9=15000+
- smoke does mother smoke? 0=never, 1=smokes now, 2=until current pregnancy, 3=once did, not now
- time time since quitting smoking: 0=never smoked, 1=still smokes, 2=during current preg, 3=within 1 yr, 4=1 to 2 years ago, 5= 2 to 3 yr ago, 6= 3 to 4 yrs ago, 7=5 to 9yrs ago, 8=10+yrs ago, 9=quit and don't know
- number number of cigs smoked per day for past and current smokers 0=never, 1=1-4, 2=5-9, 3=10-14, 4=15-19, 5=20-29, 6=30-39, 7=40-60, 8=60+, 9=smoke but don't know

Details

The data were presented by Nolan and Speed to address the question of whether there is a link between maternal smoking and the baby's health.

Source

The book by Nolan and Speed describes the data in more detail and provides an Internet site for accessing them: <http://www.stat.berkeley.edu/users/statlabs/>

References

D Nolan and T Speed. *Stat Labs: Mathematical Statistics Through Applications* (2000), Springer-Verlag.

Examples

```
data(Gestation)
```

```
getVarFormula
```

Extract data from a data frame using a formula interface

Description

Uses the full model syntax.

Usage

```
getVarFormula(formula, data = parent.frame(), intercept = FALSE)
```

Arguments

data	a data frame
formula	a formula. The right-hand side selects variables; the left-hand side, if present, is used to set row names. A . on the right-hand side indicates to use all variables not in the LHS.
intercept	a logical indicating whether to include the intercept in the model default: FALSE (no intercept)

Examples

```
getVarFormula(~wt + mpg, data=mtcars)
```

googleMap	<i>Display a point on earth on a Google Map</i>
-----------	---

Description

Creates a URL for Google Maps for a particular latitude and longitude position.

Usage

```
googleMap(latitude, longitude, position = NULL, zoom = 12,
  maptype = c("roadmap", "satellite", "terrain", "hybrid"), mark = FALSE,
  radius = 0, browse = TRUE, ...)
```

Arguments

latitude, longitude	vectors of latitude and longitude values
position	a data frame containing latitude and longitude positions
zoom	zoom level for initial map (1-20)
maptype	one of 'roadmap', 'satellite', 'terrain', and 'hybrid'
mark	a logical indicating whether the location should be marked with a pin
radius	a vector of radii of circles centered at position that are displayed on the map
browse	a logical indicating whether the URL should be browsed (else only returned as a string)
...	additional arguments passed to browseURL

Value

a string containing a URL. Optionally, as a side-effect, the URL is visited in a browser

See Also

[deg2rad](#), [latlon2xyz](#) and [rgeo](#).

Examples

```
## Not run:
googleMap(40.7566, -73.9863, radius=1) # Times Square
googleMap(position=rgeo(2), radius=1) # 2 random locations

## End(Not run)
```

HeatX

Data from a heat exchanger laboratory

Description

These data were collected by engineering students at Calvin College. The apparatus consists of concentric pipes insulated from the environment so that as nearly as can be managed the only heat exchange is between the hot and cold water.

Usage

```
data(HeatX)
```

Format

A data frame with 6 observations on the following variables.

- trial trial number
- T.cold.in temperature (C) of the cold water as it enters the apparatus
- T.cold.out temperature (C) of the cold water as it leaves the apparatus
- m.cold flow rate (L/min) of the cold water
- T.hot.in temperature (C) of the hot water as it enters the apparatus
- T.hot.out temperature (C) of the hot water as it leaves the apparatus
- m.hot flow rate (L/min) of the hot water

Examples

```
# We can test for heat exchange with the environment by check to see if the
# heat gained by the cold water matches the heat lost by the hot water.
C_p <- 4.182 / 60 # / 60 because measuring m in L/min
HeatX2 <- transform(HeatX,
                    Q.cold = m.cold * C_p * (T.cold.out - T.cold.in),
                    Q.hot = m.hot * C_p * (T.hot.out - T.hot.in)
)
HeatX2 <- transform(HeatX2, Q.env = Q.cold + Q.hot)
stripplot( ~ Q.env, data=HeatX2, alpha=.6, cex=2, jitter.data=TRUE, factor=4)
t.test( ~Q.env, data = HeatX2 )
```

Heightweight

Height and Weight

Description

The height and weight data collected from 57 males and 24 females for the purpose of exploring how the weight of a person is related to his or her height.

Usage

```
data(Heightweight)
```

Format

A data frame with 81 observations on the following variables.

- female 0 = male, 1 = female
- gender a factor with levels male or female
- height subject height (in inches)
- weight subject weight (in pounds)

References

Part of the Carnegie Mellon University Online Learning Initiative datasets.

HELP

Health Evaluation and Linkage to Primary Care

Description

The HELP study was a clinical trial for adult inpatients recruited from a detoxification unit. Patients with no primary care physician were randomized to receive a multidisciplinary assessment and a brief motivational intervention or usual care, with the goal of linking them to primary medical care.

Usage

```
data(HELP)
```

Format

Data frame with 453 observations on the following variables.

- age subject age at baseline (in years)
- anysub use of any substance post-detox: a factor with levels no yes
- cesd Center for Epidemiologic Studies Depression measure of depressive symptoms at baseline (higher scores indicate more symptoms)
- d1 lifetime number of hospitalizations for medical problems (measured at baseline)
- daysanysub time (in days) to first use of any substance post-detox
- dayslink time (in days) to linkage to primary care
- drugrisk Risk Assessment Battery drug risk scale at baseline
- e2b number of times in past 6 months entered a detox program (measured at baseline)
- female 0 for male, 1 for female
- sex a factor with levels male female
- g1b experienced serious thoughts of suicide in last 30 days (measured at baseline): a factor with levels no yes
- homeless housing status: a factor with levels housed homeless
- i1 average number of drinks (standard units) consumed per day, in the past 30 days (measured at baseline)
- i2 maximum number of drinks (standard units) consumed per day, in the past 30 days (measured at baseline)
- id subject identifier
- indtot Inventory of Drug Use Consequences (InDUC) total score (measured at baseline)
- linkstatus post-detox linkage to primary care (0 = no, 1 = yes)
- link post-detox linkage to primary care: no yes
- mcs SF-36 Mental Component Score (measured at baseline), higher scores are better
- pcs SF-36 Physical Component Score (measured at baseline), higher scores are better
- pss_fr perceived social support by friends (measured at baseline)
- racegrp race/ethnicity: levels black hispanic other white
- satreat any BSAS substance abuse treatment at baseline: no yes
- sexrisk Risk Assessment Battery sex risk score (measured at baseline)
- substance primary substance of abuse: alcohol cocaine heroin
- treat randomized to HELP clinic: no yes

Details

Eligible subjects were adults, who spoke Spanish or English, reported alcohol, heroin or cocaine as their first or second drug of choice, resided in proximity to the primary care clinic to which they would be referred or were homeless. Patients with established primary care relationships they planned to continue, significant dementia, specific plans to leave the Boston area that would prevent

research participation, failure to provide contact information for tracking purposes, or pregnancy were excluded.

Subjects were interviewed at baseline during their detoxification stay and follow-up interviews were undertaken every 6 months for 2 years. A variety of continuous, count, discrete, and survival time predictors and outcomes were collected at each of these five occasions.

This dataset is equivalent to the less confusingly named [HELPrct](#) dataset.

Source

<http://www.math.smith.edu/help>

References

Samet JH, Larson MJ, Horton NJ, Doyle K, Winter M, and Saitz R. Linking alcohol and drug-dependent adults to primary medical care: A randomized controlled trial of a multi-disciplinary health intervention in a detoxification unit. *Addiction*, 2003; 98(4):509-516.

See Also

[HELPrct](#).

Examples

```
data(HELP)
```

HELFull

Health Evaluation and Linkage to Primary Care

Description

The HELP study was a clinical trial for adult inpatients recruited from a detoxification unit. Patients with no primary care physician were randomized to receive a multidisciplinary assessment and a brief motivational intervention or usual care, with the goal of linking them to primary medical care.

Usage

```
data(HELFull)
```

Format

A data frame with 1472 observations on the following variables.

- ID Subject ID
- TIME Interview time point
- NUM_INTERVALS Number of 6-month intervals from previous to current interview
- INT_TIME1 # of months from baseline to current interview
- DAYS_SINCE_BL # of days from baseline to current interview

- INT_TIME2 # of months from previous to current interview
- DAYS_SINCE_PREV # of days from previous to current interview
- PREV_TIME Previous interview time
- DEAD a numeric vector
- A1 Gender (1=Male, 2=Female)
- A9 Years of education completed
- A10 Marital Status (1=Married, 2=Remarried, 3=Widowed, 4= Separated, 5=Divorced, 6=Never Married)
- A11A Do you currently have a living mother? (0=No, 1= Yes)
- A11B Do you currently have a living father? (0=No, 1=Yes)
- A11C Do you currently have siblings? (0=No, 1=Yes)
- A11D Do you currently have a partner (0=No, 1=Yes)
- A11E Do you currently have children? (0=No, 1=Yes)
- A12B Hollingshead categories (1=Major profess, 2= Lesser profess, 3=Minor profess, 4=Clerical/sales, 5=Skilled manual, 6=Semi-skilled, 7=Unskilled, 8= Homemaker, 9=No occupation)
- A13 Usual employment pattern in last 6 months (1=Full time, 2= Part time, 3=Student, 4=Unemployed, 5=Control envir)
- A14A Lived alone-last 6 mos (0=No, 1=Yes)
- A14B Lived w/a partner-last 6 mos (0=No, 1=Yes)
- A14C Lived with parent(s)-last 6 mos (0=No, 1=Yes)
- A14D Lived w/children-last 6 mos (0=No, 1=Yes)
- A14E Lived w/other family-last 6 mos (0=No, 1=Yes)
- A14F Lived w/friend(s)-last 6 mos (0=No, 1=Yes)
- A14G Lived w/other-last 6 mos (0=No, 1=Yes)
- A14G_T a factor with levels 1/2 WAY HOUSE 3/4 HOUSE ANCHOR INN ARMY ASSOCIATES BOARDERS BOYFRIENDS MOM CORRECTIONAL FACILIT CRACK HOUSE DEALER ENTRE FAMILIA FENWOOD GAVIN HSE GIRLFRIENDS DAUGHTE GIRLFRIENDS SON GIRLFRIENDS CHILDREN GIRLFRIENDS DAUGHTER GROUP HOME HALF-WAY HOUSE HALFWAY HOUSE HALFWAY HOUSES HALFWAY HSE HOLDING UNIT HOME BORDER HOMELESS HOMELESS SHELTER IN JAIL IN PROGRAMS INCARCERATED JAIL JAIL HALFWAY HOUSE JAIL, SHELTER JAIL, STREET JAIL/PROGRAM JAIL/SHELTER JAILS LANDLADY LANDLORD LODGING HOUSE MERIDIAN HOUSE NURSING HOME ON THE STREET PARTNERS MOTHER PARTNERS CHILD PARTNERS CHILDREN PRDGRAMS PRISON PROGRAM PROGRAM MTHP PROGRAM ROOMMATES PROGRAM SOBER HOUSE PROGRAM-RESIDENTIAL PROGRAM/HALFWAY HOUS PROGRAM/JAIL PROGRAM/SHELTER PROGRAM/SHELTERS PROGRAMS PROGRAMS SUBSTANCE PROGRAMS/SHELTER PROGRAMS/SHELTERS PROGRAMS/SHELTERS/DE PROJECT SOAR RESIDENTIAL FACILITY RESIDENTIAL PROGRAM ROOMING HOUSE ROOMING HOUSE (RELIG ROOMMATE ROOMMATES ROOMMATES AT TRANSIT RYAN HOUSE SALVATION ARMY SHELTER SHELTER/HALFWAY HSE SHELTER/HOTEL SHELTER/PROGRAM SHELTERS SHELTERS/HOSPITALS SHELTERS/JAIL SHELTERS/PROGRAMS SHELTERS/STREETS SOBER HOUSE SOBER HOUSING SOUTH BAY JAIL STEPSON STREET STREETS SUBSTANCE ABUSE TREA TRANSITIONAL HOUSE VA SHELTER
- A15A #nights in ovrnight shelter-last 6 mos

- A15B # nights on street-last 6 mos
- A15C #months in jail-last 6 mos
- A16A # months in overnight shelter-last 5 yrs
- A16B #months on street-last 5 yrs
- A16C #months in jail-last 5 yrs
- A17A Received SSI-past 6 mos (0=No, 1=Yes)
- A17B Received SSDI-past 6 mos (0=No, 1=Yes)
- A17C Received AFDC-past 6 mos (0=No, 1=Yes)
- A17D Received EAEDC-past 6 mos (0=No, 1=Yes)
- A17E Received WIC-past 6 mos (0=No, 1=Yes)
- A17F Received unemployment benefits-past 6 mos (0=No, 1=Yes)
- A17G Received Workman's Comp-past 6 mos (0=No, 1=Yes)
- A17H Received Child Support-past 6 mos (0=No, 1=Yes)
- A17I Received other income-past 6 mos (0=No, 1=Yes)
- A17I_T a factor with levels DISABLED VETERAN EBT (FOOD STAMPS) EMERGENCY FOOD STAMP FOOD STAMP FOOD STAMPS FOOD STAMPS/VETERAN FOOD STAMPS/VETERANS INSURANCE SETTLEMENT PENSION CHECK SECTION 8 SERVICE CONNECTED DI SOCIAL SECURITY SSDI FOR SON SURVIVORS BENEFITS TEMPORARY DISABILITY VA BENEFITS-DISABILI VA COMPENSATION VA DISABILITY PENSIO VETERAN BENEFITS VETERANS SERVICES VETERANS AFFAIRS
- A18 Most money made in any 1 year-last 5 yrs (1=<5000, 2=5000-10000, 3=11000-19000, 4=20000-29000, 5=30000-39000, 6=40000-49000, 7=50000+)
- B1 In general, how is your health (1=Excellent, 2=Very Good, 3=Good, 4=Fair, 5=Poor)
- B2 Comp to 1 yr ago, how is your health now (1=Much better, 2=Somewhat better, 3=About the same, 4=Somewhat worse, 5=Much worse)
- B3A Does health limit you in vigorous activity (1=Limited a lot, 2=Limited a little, 3=Not limited)
- B3B Does your health limit you in moderate activity (1=Limited a lot, 2=Limited a little, 3=Not limited)
- B3C Does health limit you in lift/carry groceries (1=Limited a lot, 2=Limited a little, 3=Not limited)
- B3D Hlth limit you in climb sev stair flights (1=Limited a lot, 2=Limited a little, 3=Not limited)
- B3E Health limit you in climb 1 stair flight (1=Limited a lot, 2=Limited a little, 3=Not limited)
- B3F Health limit you in bend/kneel/stoop (1=Limited a lot, 2=Limited a little, 3=Not limited)
- B3G Does health limit you in walking >1 mile (1=Limited a lot, 2=Limited a little, 3=Not limited)
- B3H Hlth limit you in walking sevrl blocks (1=Limited a lot, 2=Limited a little, 3=Not limited)
- B3I Does health limit you in walking 1 block (1=Limited a lot, 2=Limited a little, 3=Not limited)
- B3J Hlth limit you in bathing/dressing self (1=Limited a lot, 2=Limited a little, 3=Not limited)
- B4A Cut down wrk/act due to phys hlth-1st 4 wks (0=No, 1=Yes)

- B4B Accomplish less due to phys hlth-1st 4 wks (0=No, 1=Yes)
- B4C Lim wrk/act type due to phys hlth-1st 4 wks (0=No, 1=Yes)
- B4D Diff perf work due to phys hlth-1st 4 wks (0=No, 1=Yes)
- B5A Cut wrk/act time due to emot prbs-1st 4 wks (0=No, 1=Yes)
- B5B Accomplish ess due to emot probs-1st 4 wks (0=No, 1=Yes)
- B5C <carefl w/wrk/act due to em prb-1st 4 wks (0=No, 1=Yes)
- B6 Ext phys/em intf w/norm soc act-1st 4 wk (1=Not at all, 2=Slightly, 3=Moderately, 4=Quite a bit, 5=Extremely)
- B7 Amount of bodily pain-past 4 wks (1=None, 2=Very mild, 3= Mild, 4=Moderate, 5= Severe, 6= Very severe)
- B8 Amt pain interf with norm work-last 4 wks (1=Not at all, 2=A little bit, 3=Moderately, 4=Quite a bit, 5=Extremely)
- B9A Did you feel full of pep-past 4 wks (1=All of the time, 2=Most of the time, 3 = Good bit of the time, 4=Some of the time, 5=A little of time, 6=None of the time)
- B9B Have you been nervous-past 4 wks (1=All of the time, 2=Most of the time, 3 = Good bit of the time, 4=Some of the time, 5=A little of time, 6=None of the time)
- B9C Felt nothing could cheer you-1st 4 wks (1=All of the time, 2=Most of the time, 3 = Good bit of the time, 4=Some of the time, 5=A little of time, 6=None of the time)
- B9D Have you felt calm/peaceful-past 4 wks (1=All of the time, 2=Most of the time, 3 = Good bit of the time, 4=Some of the time, 5=A little of time, 6=None of the time)
- B9E Did you have a lot of energy-past 4 wks (1=All of the time, 2=Most of the time, 3 = Good bit of the time, 4=Some of the time, 5=A little of time, 6=None of the time)
- B9F Did you feel downhearted-past 4 wks (1=All of the time, 2=Most of the time, 3 = Good bit of the time, 4=Some of the time, 5=A little of time, 6=None of the time)
- B9G Did you feel worn out-past 4 wks (1=All of the time, 2=Most of the time, 3 = Good bit of the time, 4=Some of the time, 5=A little of time, 6=None of the time)
- B9H Have you been a happy pers-past 4 wks (1=All of the time, 2=Most of the time, 3 = Good bit of the time, 4=Some of the time, 5=A little of time, 6=None of the time)
- B9I Did you feel tired-past 4 wks (1=All of the time, 2=Most of the time, 3 = Good bit of the time, 4=Some of the time, 5=A little of time, 6=None of the time)
- B10 Amyphys/em prb intf w/soc act-1st 4 wks (1All of the time, 2=Most of the time, 3=Some of the time, 4= A lttle of time, 5= Non of the time)
- B11A I seem to get sick easier than oth peop (1=Definitely true, 2=Mostly True, 3=Don't know, 4=Mostly false, 5=Definitely false)
- B11B I am as healthy as anybody I know (1=Definitely true, 2=Mostly true, 3=Don't know, 4=Mostly false, 5=Definitely False)
- B11C I expect my health to get worse (1=Definitely true, 2=Mostly true, 3=Don't know, 3=Mostly false, 5=Definitely false)
- B11D My health is excellent (1=Definitely true, 2=Mostly true, 3=Don't know, 4=Mostly false, 5=Definitely false)
- C1A Tolf by MD had seix, epil, convuls (0=No, 1=Yes)

- C1B Told by MD had asth, emphys, chr lung dis (0=No, 1=Yes)
- C1C Told by MD had MI (0=No, 1=Yes)
- C1D Told by MD had CHF (0=No, 1=Yes)
- C1E Told by MD had other heart dis (req med) (0=No, 1=Yes)
- C1F Told by MD had HBP (0=No, 1=Yes)
- C1G Told by MD had chronic liver disease (0=No, 1=Yes)
- C1H Told by MD had kidney failure (0=No, 1=Yes)
- C1I Told by MD had chronic art, osteoarth (0=No, 1=Yes)
- C1J Told by MD had peripheral neuropathy (0=No, 1=Yes)
- C1K Ever told by MD had cancer (0=No, 1=Yes)
- C1L Ever told by MD had diabetes (0=No, 1=Yes)
- C1M Ever told by MD had stroke (0=No, 1=Yes)
- C2A1 Have you ever had skin infections (0=No, 1=Yes)
- C2A2 Have you had skin infections-past 6 mos (0=No, 1=Yes)
- C2B1 Have you ever had pneumonia (0=No, 1=Yes)
- C2B2 Have you had pneumonia-past 6 mos (0=No, 1=Yes)
- C2C1 Have you ever had septic arthritis (0=No, 1=Yes)
- C2C2 Have you had septic arthritis-past 6 mos (0=No, 1=Yes)
- C2D1 Have you ever had TB (0=No, 1=Yes)
- C2D2 Have you had TB-last 6 mos (0=No, 1=Yes)
- C2E1 Have you ever had endocarditis (0=No, 1=Yes)
- C2E2 Have you had endocarditis-past 6 mos (0=No, 1=Yes)
- C2F1 Have you ever had an ulcer (0=No, 1=Yes)
- C2F2 Have you had an ulcer-past 6 mos (0=No, 1=Yes)
- C2G1 Have you ever had pancreatitis (0=No, 1=Yes)
- C2G2 Have you had pancreatitis-past 6 mos (0=No, 1=Yes)
- C2H1 Ever had abdom pain req overnt hosp stay (0=No, 1=Yes)
- C2H2 Abdom pain req ovrnt hosp stay-1st 6 mos (0=No, 1=Yes)
- C2I1 Have you ever vomited blood (0=No, 1=Yes)
- C2I2 Have you vomited blood-past 6 mos (0=No, 1=Yes)
- C2J1 Have you ever had hepatitis (0=No, 1=Yes)
- C2J2 Have you had hepatitis-past 6 mos (0=No, 1=Yes)
- C2K1 Ever had blood clots in legs/lungs (0=No, 1=Yes)
- C2K2 Blood clots in legs/lungs-past 6 mos (0=No, 1=Yes)
- C2L1 Have you ever had osteomyelitis (0=No, 1=Yes)
- C2L2 Have you had osteomyelitis-past 6 mos (0=No, 1=Yes)
- C2M1 Chst pain using cocaine req ER/hosp (0=No, 1=Yes)

- C2M2 Chst pain using coc req ER/hosp-1st 6 mos (0=No, 1=Yes)
- C2N1 Have you ever had jaundice (0=No, 1=Yes)
- C2N2 Have you had jaundice-past 6 mos (0=No, 1=Yes)
- C2O1 Lower back pain > 3mos req med attn (0=No, 1=Yes)
- C2O2 Lwr bck pain >3mos req med attn-last 6 mos (0=No, 1=Yes)
- C2P1 Ever had seizures or convulsions (0=No, 1=Yes)
- C2P2 Had seizures or convulsions-past 6 mos (0=No, 1=Yes)
- C2Q1 Ever had drug/alc overdose req ER attn (0=No, 1=Yes)
- C2Q2 Drug/alc overdose req ER attn (0=No, 1=Yes)
- C2R1 Have you ever had a gunshot wound (0=No, 1=Yes)
- C2R2 Had a gunshot wound-past 6 mos (0=No, 1=Yes)
- C2S1 Have you ever had a stab wound (0=No, 1=Yes)
- C2S2 Have you had a stab wound-past 6 mos (0=No, 1=Yes)
- C2T1 Ever had accid/falls req med attn (0=No, 1=Yes)
- C2T2 Had accid/falls req med attn-past 6 mos (0=No, 1=Yes)
- C2U1 Ever had fract/disloc to bones/joints (0=No, 1=Yes)
- C2U2 Fract/disloc to bones/joints-past 6 mos (0=No, 1=Yes)
- C2V1 Ever had injury from traffic accident (0=No, 1=Yes)
- C2V2 Had injury from traffic accid-past 6 mos (0=No, 1=Yes)
- C2W1 Have you ever had a head injury (0=No, 1=Yes)
- C2W2 Have you had a head injury-past 6 mos (0=No, 1=Yes)
- C3A1 Have you ever had syphilis (0=No, 1=Yes)
- C3A2 # times had syphilis
- C3A3 Have you had syphilis in last 6 mos (0=No, 1=Yes)
- C3B1 Have you ever had gonorrhea (0=No, 1=Yes)
- C3B2 # times had gonorrhea
- C3B3 Have you had gonorrhea in last 6 mos (0=No, 1=Yes)
- C3C1 Have you ever had chlamydia (0=No, 1=Yes)
- C3C2 # of times had Chlamydia
- C3C3 Have you had chlamydia in last 6 mos (0=No, 1=Yes)
- C3D Have you ever had genital warts (0=No, 1=Yes)
- C3E Have you ever had genital herpes (0=No, 1=Yes)
- C3F1 Have you ever had other STD's (not HIV) (0=No, 1=Yes)
- C3F2 # of times had other STD's (not HIV)
- C3F3 Had other STD's (not HIV)-last 6 mos (0=No, 1=Yes)
- C3F_T a factor with levels 7 CRABS CRABS - TRICHONOMIS CRABS, HEP B DOESNT KNOW NAME HAS HAD ALL 3 ABC HEP B HEP B, TRICAMONAS HEP. B HEPATITIS B HEPATITS B TRICHAMONAS VAGINALA TRICHAMONIS TRICHOMONAS TRICHOMONIASIS TRICHOMONIS TRICHOMONIS VAGINITI TRICHOMORAS TRICHONOMIS

- C3G1 Have you ever been tested for HIV/AIDS (0=No, 1=Yes)
- C3G2 # times tested for HIV/AIDS
- C3G3 Have you been tested for HIV/AIDS-1st 6 mos (0=No, 1=Yes)
- C3G4 What was the result of last test (1=Positive, 2=Negative, 3=Refused, 4=Never got result, 5=Inconclusive)
- C3H1 Have you ever had PID (0=No, 1=Yes)
- C3H2 # of times had PID
- C3H3 Have you had PID in last 6 mos (0=No, 1=Yes)
- C3I Have you ever had a Pap smear (0=No, 1=Yes)
- C3J Have you had a Pap smear in last 3 years (0=No, 1=Yes)
- C3K Are you pregnant (0=No, 1=Yes)
- C3K_M How many mos pregnant
- D1 \$ of times hospitalized for med probs
- D2 Take prescr med regularly for phys prob (0=No, 1=Yes)
- D3 # days had med probs-30 days bef detox
- D4 How bother by med prob-30days bef detox (0=Not at all, 1=Slightly, 2=Moderately, 3=Considerably, 4=Extremely)
- D5 How import is trtmnt for these med probs (0=Not at all, 1=Slightly, 2= Moderately, 3= Considerably, 4= Extremely)
- E2A Detox prog for alc or drug prob-1st 6 mos (0=No, 1=Yes)
- E2B # times entered a detox prog-1st 6 mos
- E2C # nights overnight in detox prg-1st 6 mos
- E3A Holding unit for drug/alc prob-1st 6 mos (0=No, 1=Yes)
- E3B # times in holding unity=1st 6 mos
- E3C # total nights in holding unit-1st 6 mos
- E4A In halfway hse/resid facil-1st 6 mos (0=No, 1=Yes)
- E4B # times in hlfwy hse/res facil-1st 6 mos
- E4C Ttl nites in hlfwy hse/res fac-last 6 mos
- E5A In day trtmnt prg for alcohol/drug-1st 6 mos (0=No, 1=Yes)
- E5B Total # days in day trtmnt prg-1st 6 mos
- E6 In methadone maintenance prg-1st 6 mos (0=No, 1=Yes)
- E7A Visit outpt prg subst ab couns-1st 6 mos (0=No, 1=Yes)
- E7B # visits outpt prg subst ab couns-1st 6 mos
- E8A1 Saw MD/H care wkr re alcohol/drugs-1st 6 mos (0=No, 1=Yes)
- E8A2 Saw Prst/Min/Rabbi re alcohol/drugs-1st 6 mos (0=No, 1=Yes)
- E8A3 Employ Asst Prg for alcohol/drug prb-1st 6 mos (0=No, 1=Yes)
- E8A4 Oth source cnsl for alcohol/drug prb-1st 6 mos (0=No, 1=Yes)
- E9A AA/NA/slf-hlp for drug/alcohol/emot-1st 6 mos (0=No, 1=Yes)

- E9B How often attend AA/NA/slf-hlp-1st 6 mos (1=Daily, 2=2-3 Times/week, 3=Weekly, 4=Every 2 weeks, 5=Once/month)
- E10A have you been to med clinic-1st 6 mos (0=No, 1=Yes)
- E10B1 # x visit ment hlth clin/prof-1st 6 mos
- E10B2 # x visited med clin/priv MD-1st 6 mos
- E10C19 Visited private MD-last 6 mos (0=No, 1=Yes)
- E11A Did you stay ovrnite/+ in hosp-1st 6 mos (0=No, 1=Yes)
- E11B # times ovrnight/+ in hosp-last 6 mos
- E11C Total # nights in hosp-last 6 mos
- E12A Visited Hosp ER for med care-past 6 mos (0=No, 1=Yes)
- E12B # times visited hosp ER-last 6 mos
- E13 Tlt # visits to MDs-1st 2 wks bef detox
- E14A Recd trtmt from acupuncturist-last 6 mos (0=No, 1=Yes)
- E14B Recd trtmt from chiropractor-last 6 mos (0=No, 1=Yes)
- E14C Trtd by hol/herb/hom med prac-1st 6 mos (0=No, 1=Yes)
- E14D Recd trtmt from spirit healer-1st 6 mos (0=No, 1=Yes)
- E14E Have you had biofeedback-last 6 mos (0=No, 1=Yes)
- E14F Have you underwent hypnosis-1st 6 mos (0=No, 1=Yes)
- E14G Received other treatment-last 6 mos (0=No, 1=Yes)
- E15A Tried to get subst ab services-1st 6 mos (0=No, 1=Yes)
- E15B Always able to get subst ab servies (0=No, 1=Yes)
- E15C1 I could not pay for services (0=No, 1=Yes)
- E15C2 I did not know where to go for help (0=No, 1=Yes)
- E15C3 Couldn't get to services due to transp prob (0=No, 1=Yes)
- E15C4 The offie/clinic hrs were inconvenient (0=No, 1=Yes)
- E15C5 Didn't speak/understnd English well enough (0=No, 1=Yes)
- E15C6 Afraid other might find out about prob (0=No, 1=Yes)
- E15C7 My substance abuse interfered (0=No, 1=Yes)
- E15C8 Didn't have someone to watch my children (0=No, 1=Yes)
- E15C9 I did not want to lose my job (0=No, 1=Yes)
- E15C10 My insurance didn't cover services (0=No, 1=Yes)
- E15C11 There were no beds available at the prog (0=No, 1=Yes)
- E15C12 Other reason not get sub ab services (0=No, 1=Yes)
- E16A1 I cannot pay for services (0=No, 1=Yes)
- E16A2 I am not eligible for free care (0=No, 1=Yes)
- E16A3 I do not know where to go (0=No, 1=Yes)
- E16A4 Can't get to services due to trans prob (0=No, 1=Yes)

- E16A5 a numeric vector Office/clinic hours are inconvenient (0=No, 1=Yes)
- E16A6 I don't speak/understand enough English (0=No, 1=Yes)
- E16A7 Afraid others find out about my health problem (0=No, 1=Yes)
- E16A8 My substance abuse interferes (0=No, 1=Yes)
- E16A9 I don't have someone to watch my children (0=No, 1=Yes)
- E16A10 I do not want to lose my job (0=No, 1=Yes)
- E16A11 My insurance doesn't cover charges (0=No, 1=Yes)
- E16A12 I do not feel I need a regular MD (0=No, 1=Yes)
- E16A13 Other reasons don't have regular MD (0=No, 1=Yes)
- E18A I could not pay for services (0=No, 1=Yes)
- E18B I did not know where to go for help (0=No, 1=Yes)
- E18C Couldn't get to services due to transport problem (0=No, 1=Yes)
- E18D The office/clinic hours were inconvenient (0=No, 1=Yes)
- E18F Afraid others might find out about problem (0=No, 1=Yes)
- E18G My substance abuse interfered (0=No, 1=Yes)
- E18H Didn't have someone to watch my children (0=No, 1=Yes)
- E18I I did not want to lose my job (0=No, 1=Yes)
- E18J My insurance didn't cover services (0=No, 1=Yes)
- E18K There were no beds available at the program (0=No, 1=Yes)
- E18L I do not need substance abuse services (0=No, 1=Yes)
- E18M Other reason not get substance abuse services (0=No, 1=Yes)
- F1A Bothered by things not generally bothered by (0=Rarely/never, 1=Some of the time, 2=Occas/moderately, 3=Most of the time)
- F1B My appetite was poor (0=Rarely/never, 1=Some of the time, 2=Occas/moderately, 3=Most of the time)
- F1C Couldn't shake blues even with family and friends help (0=Rarely/never, 1=Some of the time, 2=Occas/moderately, 3=Most of the time)
- F1D Felt I was just as good as other people (0=Rarely/never, 1=Some of the time, 2=Occas/moderately, 3=Most of the time)
- F1E Had trouble keeping mind on what doing (0=Rarely/never, 1=Some of the time, 2=Occas/moderately, 3=Most of the time)
- F1F I felt depressed (0=Rarely/never, 1=Some of the time, 2=Occas/moderately, 3=Most of the time)
- F1G I felt everything I did was an effort (0=Rarely/never, 1=Some of the time, 2=Occas/moderately, 3=Most of the time)
- F1H I felt hopeful about the future (0=Rarely/never, 1=Some of the time, 2=Occas/moderately, 3=Most of the time)
- F1I I thought my life had been a failure (0=Rarely/never, 1=Some of the time, 2=Occas/moderately, 3=Most of the time)

- F1J I felt fearful (0=Rarely/never, 1=Some of the time, 2=Occas/moderately, 3=Most of the time)
- F1K My sleep was restless (0=Rarely/never, 1=Some of the time, 2=Occas/moderately, 3=Most of the time)
- F1L I was happy (0=Rarely/never, 1=Some of the time, 2=Occas/moderately, 3=Most of the time)
- F1M I talked less than usual (0=Rarely/never, 1=Some of the time, 2=Occas/moderately, 3=Most of the time)
- F1N I felt lonely (0=Rarely/never, 1=Some of the time, 2=Occas/moderately, 3=Most of the time)
- F1O People were unfriendly (0=Rarely/never, 1=Some of the time, 2=Occas/moderately, 3=Most of the time)
- F1P I enjoyed life (0=Rarely/never, 1=Some of the time, 2=Occas/moderately, 3=Most of the time)
- F1Q I had crying spells (0=Rarely/never, 1=Some of the time, 2=Occas/moderately, 3=Most of the time)
- F1R I felt sad (0=Rarely/never, 1=Some of the time, 2=Occas/moderately, 3=Most of the time)
- F1S I felt that people dislike me (0=Rarely/never, 1=Some of the time, 2=Occas/moderately, 3=Most of the time)
- F1T I could not get going (0=Rarely/never, 1=Some of the time, 2=Occas/moderately, 3=Most of the time)
- G1A Diff contr viol beh for sig time per evr (0=No, 1=Yes)
- G1A_30 Diff contr viol beh-sig per lst 30 days (0=No, 1=Yes)
- G1B Ever had thoughts of suicide (0=No, 1=Yes)
- G1B_30 Had thoughts of suicide-lst 30 days (0=No, 1=Yes)
- G1C Attempted suicide ever (0=No, 1=Yes)
- G1C_30 Attempted suicide-lst 30 days (0=No, 1=Yes)
- G1D Prescr med for pst/emot prob ever (0=No, 1=Yes)
- G1D_30 Prescr med for psy/emot prob-lst 30 days (0=No, 1=Yes)
- H1_30 # days in past 30 bef detox used alcohol
- H1_LT # yrs regularly used alcohol
- H1_RT Route of administration use alcohol (0=N/A. 1=Oral, 2=Nasal, 3=Smoking, 4=Non-IV injection, 5=IV)
- H2_30 #days in 3- bef detox use alc to intox
- H2_LT # yrs regularly used alcohol to intox
- H2_RT Route of admin use alcohol to intox (0=N/A. 1=Oral, 2=Nasal, 3=Smoking, 4=Non-IV injection, 5=IV)
- H3_30 # days in past 30 bef detox used heroin
- H3_LT # yrs regularly used heroin
- H3_RT Route of administration of heroin (0=N/A. 1=Oral, 2=Nasal, 3=Smoking, 4=Non-IV injection, 5=IV)

- H4_30 # days used methadone-1st 30 bef detox
- H4_LT # yrs regularly used methadone
- H4_RT Route of administration of methadone (0=N/A. 1=Oral, 2=Nasal, 3=Smoking, 4=Non-IV injection, 5=IV)
- H5_30 # days used opi/analg-1st 30 bef detox
- H5_LT # yrs regularly used oth opiates/analg
- H5_RT Route of admin of oth opiates/analg (0=N/A. 1=Oral, 2=Nasal, 3=Smoking, 4=Non-IV injection, 5=IV)
- H6_30 # days in past 30 bef detox used barbit
- H6_LT # yrs regularly used barbiturates
- H6_RT Route of admin of barbiturates (0=N/A. 1=Oral, 2=Nasal, 3=Smoking, 4=Non-IV injection, 5=IV)
- H7_30 # days used sed/hyp/trnq-1st 30 bef det
- H7_LT # yrs regularly used sed/hyp/trnq
- H7_RT Route of admin of sed/hyp/trnq (0=N/A. 1=Oral, 2=Nasal, 3=Smoking, 4=Non-IV injection, 5=IV)
- H8_30 # days in 1st 30 bef detox used cocaine
- H8_LT # yrs regularly used cocaine
- H8_RT Route of admin of cocaine (0=N/A. 1=Oral, 2=Nasal, 3=Smoking, 4=Non-IV injection, 5=IV)
- H9_30 # days in 1st 30 bef detox used amphet
- H9_LT # yrs regularly used amphetamines
- H9_RT Route of admin of amphetamines (0=N/A. 1=Oral, 2=Nasal, 3=Smoking, 4=Non-IV injection, 5=IV)
- H10_30 # days in 1st 30 bef detox used cannabis
- H10_LT # yrs regularly used cannabis
- H10_RT Route of admin of cannabis (0=N/A. 1=Oral, 2=Nasal, 3=Smoking, 4=Non-IV injection, 5=IV)
- H11_30 # days in 1st 30 bef detox used halluc
- H11_LT # yrs regularly used hallucinogens
- H11_RT Route of admin of hallucinogens (0=N/A. 1=Oral, 2=Nasal, 3=Smoking, 4=Non-IV injection, 5=IV)
- H12_30 # days in 1st 30 bef detox used inhalant
- H12_LT # yrs regularly used inhalants
- H12_RT Route of admin of inhalants (0=N/A. 1=Oral, 2=Nasal, 3=Smoking, 4=Non-IV injection, 5=IV)
- H13_30 # days used >1 sub/day-1st 30 bef detox
- H13_LT # yrs regularly used >1 subst/day
- H13_RT Route of admin of >1 subst/day (0=N/A. 1=Oral, 2=Nasal, 3=Smoking, 4=Non-IV injection, 5=IV)

- H14 Accord to interview w/c subst is main prob (0=No problem, 1=Alcohol, 2=Alcool to intoxic, 3=Heroin 4=Methadone, 5=Oth opiate/analg, 6=Barbituates, 7=Sed/hyp/tranq, 8=Cocaine, 9=Amphetamines, 10=Marij/cannabis)
- H15A # times had alchol DTs
- H15B # times overdosed on drugs
- H16A \$ spent on alc-1st 30 days bef detox
- H16B \$ spent on drugs-1st 30 days bef detox
- H17A # days had alc prob-1st 30 days bef det
- H17B # days had drug prob-1st 30 days bef det
- H18A How troubled by alc probs-1st 30 days (0=Not at all, 1=Slightly, 2=Moderately, 3=Considerably, 4=Extremely)
- H18B How troubled by drug probs-1st 30 days (0=Not at all, 1=Slightly, 2=Moderately, 3=Considerably, 4=Extremely)
- H19A How import is trtmnt for alc probs now (0=Not at all, 1=Slightly, 2=Moderately, 3=Considerably, 4=Extremely)
- H19B How importy is trtmnt for drug probs now (0=Not at all, 1=Slightly, 2=Moderately, 3=Considerably, 4=Extremely)
- I1 Avg # drinks in 1st 30 days bef detox
- I2 Most drank any 1 day in 1st 30 bef detox
- I3 On days used heroin, avg # bags used
- I4 Most bgs heroin use any 1 day-30 bef det
- I5 Avg \$ amt of heorin used per day
- I6A On days used cocaine, avg # bags used
- I6B On days used cocaine, avg # rocks used
- I7A Mst bgs cocaine use any 1 day-30 bef det
- I7B Mst rcks cocaine use any 1 day-30 bef det
- I8 Avg \$ amt of cocaine used per day
- J1 Evr don't stop using cocaine when should (0=No, 1=Yes)
- J2 Ever tried to cut down on cocaine (0=No, 1=Yes)
- J3 Does cocaine take up a lot of your time (0=No, 1=Yes)
- J4 Need use > cocaine to get some feeling (0=No, 1=Yes)
- J5A Get phys sick when stop using cocaine (0=No, 1=Yes)
- J5B Ever use cocaine to prevent getting sick (0=No, 1=Yes)
- J6 Ever don't stop using heroin when should (0=No, 1=Yes)
- J7 Ever tried to cut down on heroin (0=No, 1=Yes)
- J8 Does heroin take up a lot of your time (0=No, 1=Yes)
- J9 Need use > heroin to get some feeling (0=No, 1=Yes)
- J10A Get phys sick when stop using heroin (0=No, 1=Yes)
- J10B Ever use heroin to prevent getting sick (0=No, 1=Yes)

- K1 Do you currently smoke cigarettes (1=Yes-every day, 2=Yes-some days, 3=No-former smoker, 4=No-never>100 cigs)
- K2 Avg # cigarettes smoked per day
- K3 Considering quitting cigs w/in next 6 mo (0=No, 1=Yes)
- L1 How often drink last time drank (1=To get high/less, 2=To get drunk, 3=To pass out)
- L2 Often have hangovrs Sun or Mon mornings (0=No, 1=Yes)
- L3 Have you had the shakes when sobering (0=No, 1=Sometimes, 2=Alm evry time drink)
- L4 Do you get phys sick as reslt of drinking (0=No, 1=Sometimes, 2=Alm evry time drink)
- L5 have you had the DTs (0=No, 1=Once, 2=Several times)
- L6 When drink do you stumble/stagger/weave (0=No, 1=Sometimes, 2=Often)
- L7 D/t drinkng felt overly hot/sweaty (0=No, 1=Once, 2=Several times)
- L8 As result of drinkng saw thngs not there (0=No, 1=Once, 2=Several times)
- L9 Panic because fear not have drink if need it (0=No, 1=Yes)
- L10 Have had blkouts as result of drinkng (0=No, never, 1=Sometimes, 2=Often, 3=Alm evry time drink)
- L11 Do you carry bottle or keep close by (0=No, 1=Some of the time, 2=Most of the time)
- L12 After abstin end up drink heavily again (0=No, 1=Sometimes, 2=Almost evry time)
- L13 Passed out due to drinking-1st 12 mos (0=No, 1=Once, 2=More than once)
- L14 Had convuls following period of drinkng (0=No, 1=Once, 2=Several times)
- L15 Do you drink throughout the day (0=No, 1=Yes)
- L16 Aftr drinkng heavily was thinkng unclear (0=No, 1=Yes, few hrs, 2=Yes, 1-2 days, 3=Yes, many days)
- L17 D/t drinkng felt heart beat rapidly (0=No, 1=Once, 2=Several times)
- L18 Do you constntly think about drinkng/alc (0=No, 1=Yes)
- L19 D/t drinkng heard things not there (0=No, 1=Once, 2= Several times)
- L20 Had weird/fright sensations when drinkng (0=No, 1=Once or twice, 2=Often)
- L21 When drinkng felt things rawl not there (0=No, 1=Once, 2=Several times)
- L22 With respect to blackouts (0=Never had one, 1=Had for <1hr, 2=Had several hrs, 3=Had for day/+)
- L23 Ever tried to cut down on drinking & failed (0=No, 1=Once, 2=Several times)
- L24 Do you gulp drinks (0=No, 1=Yes)
- L25 After taking 1 or 2 drinks can you stop (0=No, 1=Yes)
- M1 Had hangover/felt bad aftr using alcohol/drugs (0=No, 1=Yes)
- M2 Felt bad about self because of alcohol/drug use (0=No, 1=Yes)
- M3 Missed days wrk/sch because of alcohol/drug use (0=No, 1=Yes)
- M4 Fam/frinds worry/compl about alcohol/drug use (0=No, 1=Yes)
- M5 I have enjoyed drinking/using drugs (0=No, 1=Yes)
- M6 Qual of work suffered because of alcohol/drug use (0=No, 1=Yes)

- M7 Parenting ability harmed by alcohol/drug use (0=No, 1=Yes)
- M8 Trouble sleeping/nightmares aftr alcohol/drugs (0=No, 1=Yes)
- M9 Driven motor veh while undr inf alcohol/drugs (0=No, 1=Yes)
- M10 Using alcohol/1 drug caused > use othr drugs (0=No, 1=Yes)
- M11 I have been sick/vomited aft alcohol/drug use (0=No, 1=Yes)
- M12 I have been unhappy because of alcohol/drug use (0=No, 1=Yes)
- M13 Lost weight/eaten poorly due to alcohol/drug use (0=No, 1=Yes)
- M14 Fail to do what expected due to alcohol/drug use (0=No, 1=Yes)
- M15 Using alcohol/drugs has helped me to relax (0=No, 1=Yes)
- M16 Felt guilt/ashamed because of my alc drug use (0=No, 1=Yes)
- M17 Said/done emarras thngs when on alcohol/drug (0=No, 1=Yes)
- M18 Personality changed for worse on alcohol/drug (0=No, 1=Yes)
- M19 Taken foolish risk when using alcohol/drugs (0=No, 1=Yes)
- M20 Gotten into trouble because of alcohol/drug use (0=No, 1=Yes)
- M21 Said cruel things while using alcohol/drugs (0=No, 1=Yes)
- M22 Done impuls thngs regret due to alcohol/drug use (0=No, 1=Yes)
- M23 Gotten in phys fights when use alcohol/drugs (0=No, 1=Yes)
- M24 My phys health was harmed by alcohol/drug use (0=No, 1=Yes)
- M25 Using alcohol/drug helped me have more + outlook (0=No, 1=Yes)
- M26 I have had money probs because of my alcohol/drug use (0=No, 1=Yes)
- M27 My love relat harmed due to my alcohol/drug use (0=No, 1=Yes)
- M28 Smoked tobacco more when using alcohol/drugs (0=No, 1=Yes)
- M29 <y phys appearance harmed by alcohol/drug use (0=No, 1=Yes)
- M30 My family hurt because of my alc drug use (0=No, 1=Yes)
- M31 Close relationsp damaged due to alcohol/drug use (0=No, 1=Yes)
- M32 Spent time in jail because of my alcohol/drug use (0=No, 1=Yes)
- M33 My sex life suffered due to my alcohol/drug use (0=No, 1=Yes)
- M34 Lost interst in activity due to my alcohol/drug use (0=No, 1=Yes)
- M35 Soc life> enjoyable when using alcohol/drug (0=No, 1=Yes)
- M36 Spirit/moral life harmed by alcohol/drug use (0=No, 1=Yes)
- M37 Not had kind life want due to alcohol/drug use (0=No, 1=Yes)
- M38 My alcohol/drug use in way of personal growth (0=No, 1=Yes)
- M39 My alcohol/drug use damaged soc life/reputat (0=No, 1=Yes)
- M40 Spent/lost too much \$ because alcohol/drug use (0=No, 1=Yes)
- M41 Arrested for DUI of alc or oth drugs (0=No, 1=Yes)
- M42 Arrested for offenses rel to alcohol/drug use (0=No, 1=Yes)
- M43 Lost marriage/love relat due to alcohol/drug use (0=No, 1=Yes)

- M44 Susp/fired/left job/sch due to alcohol/drug use (0=No, 1=Yes)
- M45 I used drugs moderately w/o having probs (0=No, 1=Yes)
- M46 I have lost a friend due to my alcohol/drug use (0=No, 1=Yes)
- M47 Had an accident while using alcohol/drugs (0=No, 1=Yes)
- M48 Phys hurt/inj/burned when using alcohol/drugs (0=No, 1=Yes)
- M49 I injured someone while using alcohol/drugs (0=No, 1=Yes)
- M50 Damaged things/prop when using alcohol/drugs (0=No, 1=Yes)
- N1A My friends give me the moral support I need (0=No, 1=Yes)
- N1B Most people closer to friends than I am (0=No, 1=Yes)
- N1C My friends enjoy hearing what I think (0=No, 1=Yes)
- N1D I rely on my friends for emot support (0=No, 1=Yes)
- N1E Friend go to when down w/o feel funny later (0=No, 1=Yes)
- N1F Frnds and I open re what thnk about things (0=No, 1=Yes)
- N1G My friends sensitive to my pers needs (0=No, 1=Yes)
- N1H My friends good at helping me solve probs (0=No, 1=Yes)
- N1I have deep sharing relat w/ a # of frnds (0=No, 1=Yes)
- N1J When confide in frnds makes me uncomf (0=No, 1=Yes)
- N1K My friends seek me out for companionship (0=No, 1=Yes)
- N1L Not have as int relat w/frnds as others (0=No, 1=Yes)
- N1M Recent good idea how to do somethng frm frnd (0=No, 1=Yes)
- N1N I wish my friends were much different (0=No, 1=Yes)
- N2A My family gives me the moral support I need (0=No, 1=Yes)
- N2B Good ideas of how do/make thngs from fam (0=No, 1=Yes)
- N2C Most peop closer to their fam than I am (0=No, 1=Yes)
- N2D When confide make close fam membs uncomf (0=No, 1=Yes)
- N2E My fam enjoys hearing about what I think (0=No, 1=Yes)
- N2F Membs of my fam share many of my intrsts (0=No, 1=Yes)
- N2G I rely on my fam for emot support (0=No, 1=Yes)
- N2H Fam memb go to when down w/o feel funny (0=No, 1=Yes)
- N2I Fam and I open about what thnk about thngs (0=No, 1=Yes)
- N2J My fam is sensitive to my personal needs (0=No, 1=Yes)
- N2K Fam memb good at helping me solve probs (0=No, 1=Yes)
- N2L Have deep sharing relat w/# of fam membs (0=No, 1=Yes)
- N2M Makes me uncomf to confide in fam membs (0=No, 1=Yes)
- N2N I wish my family were much different (0=No, 1=Yes)
- O1A # people spend tx w/who drink alc (1=None, 2= A few, 3=About half, 4= Most, 5=All)
- O1B # people spend tx w/who are heavy drinkrs (1=None, 2= A few, 3=About half, 4= Most, 5=All)

- 01C # people spend tx w/who use drugs (1=None, 2= A few, 3=About half, 4= Most, 5=All)
- 01D # peop spend tx w/who supprt your abstin (1=None, 2= A few, 3=About half, 4= Most, 5=All)
- 02 Does live-in part/spouse drink/use drugs (0=No, 1=Yes, 2=N/A)
- P1A Phys abuse/assaul by fam memb/pers know (0=No, 1=Yes, 7=Not sure)
- P1B Age first phys assaulted by pers know
- P1C Phys assaulted by pers know-last 6 mos (0=No, 1=Yes)
- P2A Phys abuse/assaul by stranger (0=No, 1=Yes, 7=Not sure)
- P2B Age first phys assaulted by stranger
- P2C Phys assaulted by stranger-last 6 mos (0=No, 1=Yes)
- P3 Using drugs/alc when phys assaulted (1=Don't know, 2=Never, 3=Some cases, 4=Most cases, 5=All cases, 9=Never assaulted)
- P4 Pers who phys assault you using alcohol/drugs (1=Don't know, 2=Never, 3=Some cases, 4=Most cases, 5=All cases, 9=Never assaulted)
- P5A Sex abuse/assual by fam memb/pers know (0=No, 1= Yes, 7=Not sure)
- P5B Age first sex assaulted by pers know
- P5C Sex assaulted by pers know-last 6 mos (0=No, 1=Yes)
- P6A Sex abuse/assaul by stranger (0=No, 1=Yes, 7=Not sure)
- P6B Age first sex assaulted by stranger
- P6C Sex assaulted by stranger-last 6 mos (0=No, 1=Yes)
- P7 Using drugs/alc when sex assaulted (1=Don't know, 2=Never, 3=Some cases, 4=Most cases, 5=All cases, 9=Never assaulted)
- P8 Person who sex assaulted you using alcohol/drugs (1=Don't know, 2=Never, 3=Some cases, 4=Most cases, 5=All cases, 9=Never assaulted)
- Q1A Have you ever injected drugs (0=No, 1=Yes)
- Q1B Have you injected drugs-1st 6 mos (0=No, 1=Yes)
- Q2 Have you shared needles/works-last 6 mos (0=No/Not shot up, 3=Yes)
- Q3 # people shared needles w/past 6 mos (0=No/Not shot up, 1=1 other person, 2=2-3 diff people, 3=4/+ diff people)
- Q4 How often been to shoot gall/hse-1st 6 mos (0=Never, 1=Few times or less, 2= Few times/month, 3= Once or more/week)
- Q5 How often been to crack house-last 6 mos (0=Never, 1=Few times or less, 2=Few times/month, 3=Once or more/week)
- Q6 How often shared rinse-water-last 6 mos (0=Nevr/Not shot up, 1=Few times or less, 2=Few times/month, 3=Once or more/week)
- Q7 How often shared a cooker-last 6 mos (0=Nevr/Not shot up, 1=Few times or less, 2=Few times/month, 3=Once or more/week)
- Q8 How often shared a cotton-last 6 mos (0=Nevr/Not shot up, 1=Few times or less, 2=Few times/month, 3=Once or more/week)

- Q9 How often use syringe to div drugs-1st 6 mos (0=Nevr/Not shot up, 1=Few times or less, 2=Few times/month, 3=Once or more/week)
- Q10 How would you describe yourself (0=Straight, 1=Gay/bisexual)
- Q11 # men had sex w/in past 6 months (0=0 men, 1=1 man, 2=2-3 men, 3=4+ men)
- Q12 # women had sex w/in past 6 months (0=0 women, 1=1 woman, 2=2-3 women, 3=4+ women)
- Q13 # times had sex In past 6 mos (0=Never, 1=Few times or less, 2=Few times/month, 3=Once or more/week)
- Q14 How often had sex to get drugs-last 6 mos (0=Never, 1=Few times or less, 2=Few times/month, 3=Once or more/week)
- Q15 How often given drugs to have sex-1st 6 mos (0=Never, 1=Few times or less, 2=Few times/month, 3=Once or more/week)
- Q16 How often were you paid for sex-1st 6 mos (0=Never, 1=Few times or less, 2=Few times/month, 3=Once or more/week)
- Q17 How often you pay pers for sex-1st 6 mos (0=Never, 1=Few times or less, 2=Few times/month, 3=Once or more/week)
- Q18 How often use condoms during sex=1st 6 mos (0=No sex/always, 1=Most of the time, 2=Some of the time, 3=None of the time)
- Q19 Condoms are too much of a hassle to use (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- Q20 Safer sex is always your responsibility (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1A I really want to hange my alcohol/drug use (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1B Sometimes I wonder if I'm an alcohol/addict (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1C Id I don't chng alcohol/drug probs will worsen (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1D I started making changes in alcohol/drug use (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1E Was using too much but managed to change (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1F I wonder if my alcohol/drug use hurting othrs (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1G I am a prob drinker or have drug prob (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1H Already doing thngs to chnge alcohol/drug use (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1I have changed use-trying to not slip back (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1J I have a serious problem w/ alcohol/drugs (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)

- R1K I wonder if I'm in contrl of alcohol/drug use (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1L My alcohol/drug use is causing a lot of harm (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1M Actively curring down/stopping alcohol/drug use (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1N Want help to not go back to alcohol/drugs (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1O I know that I have an alcohol/drug problem (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1P I wonder if I use alcohol/drugs too much (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1Q I am an alcoholic or drug addict (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1R I am working hard to change alcohol/drug use (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1S Some changes-want help from going back (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- S1A At interview pt obviously depressed/withdrawn (0=No, 1=Yes)
- S1B at interview pt obviously hostile (0=No, 1=Yes)
- S1C At interview pt obviouslt anx/nervous (0=No, 1=Yes)
- S1D Trouble w/real tst/thght dis/par at interview (0=No, 1=Yes)
- S1E At interview pt trbl w/ compr/concen/rememb (0=No, 1=Yes)
- S1F At interview pt had suicidal thoughts (0=No, 1=Yes)
- T1 Have used alc since leaving River St. (0=No, 1=Yes)
- T1B # days in row continued to drink
- T1C Longest period abstain-1st 6 mos (alc)
- T2 Have used heroin since leaving River St (0=No, 1=Yes)
- T2B # days in row continued to use heroin
- T2C Longest period abstain-1st 6 mos (heroin)
- T3 Have used cocaine since leaving River St (0=No, 1=Yes)
- T3B # days in row continued to use cocaine
- T3C Lngest period abstain-1st 6 mos (cocaine)
- U1 It is important to have a regular MD (1=Strongly agree, 2=Agree, 3=Uncertain, 4=Disagree, 5=Strongly Disagree)
- U2A I cannot pay for services (0=No, 1=Yes)
- U2B I am not eligible for free care (0=No, 1=Yes)
- U2C I do not know where to go (0=No, 1=Yes)
- U2D Can't get services due to transport probs (0=No, 1=Yes)
- U2E Office/clinic hours are inconvenient (0=No, 1=Yes)

- U2F I do not speak/understand English well (0=No, 1=Yes)
- U2G Afraid others discover hlth prb I have (0=No, 1=Yes)
- U2H My substance abuse interferes (0=No, 1=Yes)
- U2I I do not have a babysitter (0=No, 1=Yes)
- U2J I do not want to lose my job (0=No, 1=Yes)
- U2K My insurance does not cover services (0=No, 1=Yes)
- U2L Medical care is not important to me (0=No, 1=Yes)
- U2M I do not have time (0=No, 1=Yes)
- U2N Med staff do not treat me with respect (0=No, 1=Yes)
- U2O I do not trust my doctors or nurses (0=No, 1=Yes)
- U2P Often been unsatisfied w/my med care (0=No, 1=Yes)
- U2Q Other reason hard to get regular med care (0=No, 1=Yes)
- U2Q_T a factor with many levels
- U2R a factor with levels 7 A B C D E F G H I J K L M N O P Q
- U3A Has MD evr talked to you about drug use (0=No, 1=Yes)
- U3B Has MD evr talked to you about alc use (0=No, 1=Yes)
- U4 Is there an MD you consider your regular MD (0=No, 1=Yes)
- U5 Have you seen any MDs in last 6 mos (0=No, 1=Yes)
- U6A Would you go to this MD if med prb not emer (0=No, 1=Yes)
- U6B Think one of these could be your regular MD (0=No, 1=Yes)
- PCP_ID a numeric vector
- U7A What type of MD is your regular MD/this MD (1=OB/GYN, 2=Family medicine, 3=Pediatrician, 4=Adolescent medicine, 5=Internal medicine, 6=AIDS doctor, 7=Asthma doctor, 8=Pulmonary doctor, 9=Cardiologist, 10=Gastroen)
- U7A_T a factor with levels ARTHRITIS DOCTOR CHIROPRACTOR COCAINE STUDY DETOX DOCTOR DO EAR DOCTOR EAR SPECIALIST EAR, NOSE, & THROAT. EAR/NOSE/THROAT ENT FAMILY PHYSICIAN GENERAL MEDICINE GENERAL PRACTICE GENERAL PRACTITIONER GENERAL PRACTITIONER HEAD & NECK SPECIALIST HERBAL/HOMEOPATHIC/ACUPUNCTURE ID DOCTOR MAYBE GENERAL PRACTITIONER MEDICAL STUDENT NEUROLOGIST NURSE NURSE PRACTITIONER NURSE PRACTITIONER ONCOLOGIST PRENATAL PRIMARY PRIMARY CAEE PRIMARY CARE PRIMARY CARE DOCTOR PRIMERY CARE THERAPIST UROLOGIST WOMENS CLINIC BMC
- U8A Only saw this person once (=Only saw once)
- U8B Saw this person for <6 mos (1=<6 mos)
- U8C Saw tis person for 6 mos-1year (2=Betwn 6 mos & 1 yr)
- U8D Saw this person for 1-2 years (3=1-2 years)
- U8E Saw this person for 3-5 years (4=3-5 years)
- U8F Saw this person for more than 5 years (5=>5 years)
- U10A # times been to regular MDs office-pst 6 mos
- U10B # times saw regular MD in office-pst 6 mos

- U10C # times saw oth prof in office-pst 6 mos
- U11 Rate convenience of MD office location (1=Very poor, 2=Poor, 3=Fair, 4=Good, 5=Very good, 6=Excellent)
- U12 Rate hours MD office open for med appts (1=Very poor, 2=Poor, 3=Fair, 4=Good, 5=Very good, 6=Excellent)
- U13 Usual wait for appt when sick (unsched) (1=Very poor, 2=Poor, 3=Fair, 4=Good, 5=Very good, 6=Excellent)
- U14 Time wait for appt to start at MD office (1=Very poor, 2=Poor, 3=Fair, 4=Good, 5=Very good, 6=Excellent)
- U15A DO you pay for any/all of MD visits (0=No, 1=Yes)
- U15B How rate amt of \$ you pay for MD visits (1=Very poor, 2=Poor, 3=Fair, 4=Good, 5=Very good, 6=Excellent)
- U16A Do you pay for any/all of prescript meds (0=No, 1=Yes)
- U16B Rate amt \$ pay for meds/prescript trtmnts (1=Very poor, 2=Poor, 3=Fair, 4=Good, 5=Very good, 6=Excellent)
- U17 Ever skip meds/trtmnts because too expensive (1=Yes, often, 2=Yes, occasionally, 3=No, never)
- U18A Ability to reach MC office by phone (1=Very poor, 2=Poor, 3=Fair, 4=Good, 5=Very good, 6=Excellent)
- U18B Ability to speak to MD by phone if need (1=Very poor, 2=Poor, 3=Fair, 4=Good, 5=Very good, 6=Excellent)
- U19 How often see regular MD when have regular check-up (1=Always, 2=Almost always, 3=A lot of the time, 4=Some of the time, 5=Almost never, 6=Never)
- U20 When sick + go to MD how often see regular MD (1=Always, 2=Almost always, 3=A lot of the time, 4=Some of the time, 5=Almost never, 6=Never)
- U21A How thorough MD exam to check hlth prb (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)
- U21B How often question if MD diagnosis right (1=Always, 2=Almost always, 3=A lot of the time, 4=Some of the time, 5=Almost never, 6=Never)
- U22A Thoroughness of MD questions re symptoms (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)
- U22B Attn MD gives to what you have to say (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)
- U22C MD explanations of hlth prbs/trtmnts need (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)
- U22D MD instrcts re sympt report/further care (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)
- U22E MD advice in decisions about your care (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)
- U23 How often leave MD office w/unanswd quests (1=Always, 2=Almost always, 3=A lot of the time, 4=Some of the time, 5=Almost never, 6=Never)

- U24A Amount of time your MD spends w/you (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)
- U24B MDs patience w/ your questions/worries (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)
- U24C MDs friendliness and warmth toward you (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)
- U24D MDs caring and concern for you (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)
- U24E MDs respect for you (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)
- U25A Reg MD ever talked to you about smoking (0=No, 1=Yes)
- U25B Reg MD ever talked to you about alc use (0=No, 1=Yes)
- U25C Reg MD ever talk to you about seat belt use (0=No, 1=Yes)
- U25D Reg MD ever talked to you about diet (0=No, 1=Yes)
- U25E Reg Mdever talked to you about exercise (0=No, 1=Yes)
- U25F Reg MD ever talked to you about stress (0=No, 1=Yes)
- U25G Reg MD ever talked to you about safe sex (0=No, 1=Yes)
- U25H Reg MD ever talked to you about drug use (0=No, 1=Yes)
- U25I Reg MD ever talked to you about HIV testing (0=No, 1=Yes)
- U26A Cut/quit smoking because of MDs advice (0=No, 1=Yes)
- U26B Tried to drink less alcohol because of MD advice (0=No, 1=Yes)
- U26C Wore my seat belt more because of MDs advice (0=No, 1=Yes)
- U26D Changed diet because of MDs advice (0=No, 1=Yes)
- U26E Done more exercise because MDs advice (0=No, 1=Yes)
- U26F Relax/reduce stress because of MDs advice (0=No, 1=Yes)
- U26G Practiced safer sex because of MDs advice (0=No, 1=Yes)
- U26H Tried to cut down/quit drugs because MD advice (0=No, 1=Yes)"
- U26I Got HIV tested because of MDs advice (0=No, 1=Yes)"
- U27A I can tell my MD anything (1=Strongly agree, 2= Agree, 3= Not sure, 4=Disagree, 5=Strongly disagree)"
- U27B My MD pretends to know thngs if not sure (1=Strongly agree, 2= Agree, 3= Not sure, 4=Disagree, 5=Strongly disagree)"
- U27C I trust my MDs judgement re my med care (1=Strongly agree, 2= Agree, 3= Not sure, 4=Disagree, 5=Strongly disagree)"
- U27D My MD cares > about < costs than my hlth (1=Strongly agree, 2= Agree, 3= Not sure, 4=Disagree, 5=Strongly disagree)"
- U27E My MD always tell truth about my health (1=Strongly agree, 2= Agree, 3= Not sure, 4=Disagree, 5=Strongly disagree)"
- U27F My MD cares as much as I about my hlth (1=Strongly agree, 2= Agree, 3= Not sure, 4=Disagree, 5=Strongly disagree)"

- U27G My MD would try to hide a mistake in trtmt (1=Strongly agree, 2= Agree, 3= Not sure, 4=Disagree, 5=Strongly disagree)"
- U28 How much do you trust this MD (0=Not at all, 1=1, 2=2, 3=3, 4=4, 5=5, 6=6, 7=7, 8=8, 9=9, 10=Completely)"
- U29A MDs knowledge of your entire med history (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)"
- U29B MD knowldg of your respons-home/work/sch (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)"
- U29C MD knowldg of what worries you most-hlth (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)"
- U29D MDs knowledge of you as a person (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)"
- U30 MD would know what want done if unconsc (1=Strongly agree, 2=Agree, 3=Not sure, 4= Disagree, 5=Strongly disagree)"
- U31 Oth MDs/RNs who play roel in your care (0=No, 1=Yes)" *
- U32A Their knowledge of you as a person (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)
- U32B The quality of care they provide (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)
- U32C Coordination betw them and your regular MD (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)
- U32D Their expl of your hlth prbs/trtmts need (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)
- U32D_T N/A, only my regular MD does this
- U33 Amt regular MD knows about care from others (1=Knows everything, 2=Knows almost everything, 3=Knows some things, 4=Knows very little, 5=Knows nothing)
- U34 Has MD ever recommended you see MD sepcialist (0=No, 1=Yes)
- U35A How helpful MD in deciding on specialist (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)
- U35B How helpful MD getting appt w/specialist (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)
- U35C MDs involvmt when you trtd by specialist (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)
- U35D MDs communic w/your specialists/oth MDs (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)
- U35E MD help in explain what specialists said (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)
- U35F Quality of specialists MD sent you to (1=Very poor, 2= Poor, 3=Fair, 4=Good, 5= Very good, 6= Excellent)
- U36 How many minutes to get to MDs office (1=<15, 2=16-30, 3=31-60, 4=More than 60)
- U37 When sick+call how long take to see you (1=Same day, 2=Next day, 3=In 2-3 days, 4=In 4-5 days, 5=in >5 days)

- U38 How many minutes late appt usually begin (1=None, 2=<5 minutes, 3=6-10 minutes, 4=11-20 minutes, 5=21-30 minutes, 6=31-45 minutes, 7=>45 minutes)
- U39 How satisfied are you w/your regular MD (1=Completely satisfied, 2=Very satisfied, 3=Somewhat satisfied, 4=Neither, 5=Somewhat dissatisfied, 6=Very dissatisfied, 7=Completely dissatisfied)
- V1 Evr needed to drink much more to get effect (0=No, 1=Yes)
- V2 Evr find alc had < effect than once did (0=No, 1=Yes)
- Z1 Breath Alcohol Concentration:1st test
- Z2 Breath Alcohol Concentration:2nd test
- AGE Age in years
- REALM REALM score
- E16A_RT Barrier to regular MD: red tape (0=No, 1=Yes)
- E16A_IB Barrier to regular MD: internal barriers (0=No, 1=Yes)
- E16A_TM Barrier to regular MD: time restrictions (0=No, 1=Yes)
- E16A_DD Barrier to regular MD: dislike docs/system (0=No, 1=Yes)
- GROUP Randomization Group (0=Control, 1=Clinic)
- MMSEC MMSEC
- PRIM_SUB First drug of choice (0=None, 1=Alcohol, 3=Cocaine, 3=Heroin, 4=Barbiturates, 5=Benzos, 6=Marijuana, 7=Methadone, 8=Opiates)
- SECD_SUB Second drug of choice (0=None, 1=Alcohol, 3=Cocaine, 3=Heroin, 4=Barbiturates, 5=Benzos, 6=Marijuana, 7=Methadone, 8=Opiates)
- ALCOHOL 1st/2nd drug of choice=Alcohol (0=No, 1=Yes)
- COC_HER 1st/2nd drug of choice=cocaine or heroin (0=No, 1=Yes)
- REALM2 REALM score (dichotomous) (1=0-60, 2=61-66)
- REALM3 REALM score (categorical) (1=0-44), 2=45-60), 3=61-66)
- RACE Race (recode) (1=Afr Amer/Black, 2=White, 3=Hispanic, 4=Other)
- RACE2 Race (recode) (1=White, 2=Minority)
- BIRTHPLC Where born (recode) (0=USA, 1=Foreign)
- PRIMLANG First language (recode) (0=English, 1=Other lang)
- MD_LANG Lang prefer to speak to MD (recode) (0=English, 1=Other lang)
- HS_GRAD High school graduate (0=No, 1=Yes)
- MAR_STAT Marital status (recode) (0=Married, 1=Not married)
- A12B_REC Hollingshead category (recode) (0=Cat 1,2,3, 1=Cat 4,5,6, 2=Cat 7,8,9)
- UNEMPLOY Usually unemployed last 6m (0=No, 1=Yes)
- ALONE6M Usually lived alone past 6m y/n (0=No, 1=Yes)
- HOMELESS Homeless-shelter/street past 6 m (0=No, 1=Yes)
- JAIL_MOS Total months in jail past 5 years
- JAIL_5YR Any jail time past 5 years y/n (0=No, 1=Yes)

- GOV_SUPP Received governemtn support past 6 m (0=No, 1=Yes)
- A18_REC1 Most money made in 1 yr (recode) (0=\$19,000 or less, 1=\$20,000-\$49,000, 2=\$50,000 or more)
- A18_REC2 Most money made-continuous recode
- STD_EVER Ever had an STD y/n (0=No, 1=Yes)
- STD_6M Had an STD past 6m y/n (0=No, 1=Yes)
- CHR_SUM Sum chronic medican conds/HIV ever
- CHR_EVER Chronic medical conds/HIV-ever y/n (0=No, 1=Yes)
- EPI_SUM Sum episodic (C2A-C2O, C2R-C2U, STD)-6m
- EPI_6M Episodic (C2A-C2O,C2R-C2U, STD)-6m y/n (0=No, 1=Yes)
- EPI_6M2B Episodic(C2A-C2O)-6m y/n (0=No, 1=Yes)
- SER_INJ Recent (6m) serious injury y/n (0=No, 1=Yes)
- D3_REC Any medical problems past 30d y/n (0=No, 1=Yes)
- D4_REC Bothered by medical problems y/n (0=No, 1=Yes)
- D5_REC Medical trtmnt is important y/n (0=No, 1=Yes)
- ANY_INS Did you have health insurance past 6 m (0=No, 1=Yes)
- FRML_SAT Formal substance abuse treatment y/n (0=No, 1=Yes)
- E10B1_R Mental health treatment past 6m y/n (0=No, 1=Yes)
- E10B2_R Med clinic/private MD past 6m y/n (0=No, 1=Yes)
- ALT_TRT Alternative tratments y/n (0=No, 1=Yes)
- ANY_UTIL Amy recent health utilization (0=No, 1=Yes)
- NUM_BARR # of perceived barriers to linkage
- G1B_REC Suicidal thoughts past 30 days y/n (0=No, 1=Yes)
- G1D_REC Prescribed psych meds past 30 daus y/n (0=No, 1=Yes)
- PRIMSUB2 First drug of choice (no marijuana) (0=None, 1=Alcohol, 2=Cocaine, 3=Heroin, 4=Barbituates, 5=Benzos, 6=Marijuana, 7=Methadone, 8=Opiates)
- ALCQ_30 Total number drinks past 30 days
- H2_PRB Problem sub: alc to intox (0=No, 1=Yes)
- H3_PRB Problem sub: heroin (0=No, 1=Yes)
- H4_PRB Problem sub: methadone (0=No, 1=Yes)
- H5_PRB Problem sub: oth opiates/analg (0=No, 1=Yes)
- H6_PRB Problem sub: barbituates (0=No, 1=Yes)
- H7_PRB Problem sub: sedat/hyp/tranq (0=No, 1=Yes)
- H8_PRB Problem sub: cocaine (0=No, 1=Yes)
- H9_PRB Problem sub: amphetamines (0=No, 1=Yes)
- H10_PRB Problem sub: marijuana, cannabis (0=No, 1=Yes)
- H11_PRB Problem sub: hallucinogens (0=No, 1=Yes)
- H12_PRB Problem sub: inhalants (0=No, 1=Yes)

- POLYSUB Polysubstance abuser y/n (0=No, 1=Yes)
- SMOKER Current smoker (every/some days) y/n (0=No, 1=Yes)
- 01B_REC Family/friends heavy drinkers y/n (0=No, 1=Yes)
- 01C_REC Family/friends use drugs y/n (0=No, 1=Yes)
- 01D_REC Family/friends support abst. y/n (0=No, 1=Yes)
- 02_REC Live-in partner drinks/drugs y/n (0=No, 1=Yes)
- PHYABUSE Physical abuse-stranger or family (0=No, 1=Yes)
- SEXABUSE Sexual abuse-stranger or family (0=No, 1=Yes)
- PHSXABUS Any abuse (0=No, 1=Yes)
- ABUSE2 Type of abuse (0=No abuse, 1=Physical only, 2=Sexual only, 3=Physical and sexual)
- ABUSE3 Type of abuse (0=No abuse, 1=Physical only, 2=Sexual +/- physical (0=No, 1=Yes)
- CURPHYAB Current abuse-physical (0=No, 1=Yes)
- CURSEXAB Current abuse-sexual (0=No, 1=Yes)
- CURPHYSEXAB Current abuse-physical or sexual (0=No abuse, 1=Physical only, 2=Sexual +/- physical)
- FAMABUSE Family abuse-physical or sexual (0=No, 1=Yes)
- STRABUSE Stranger abuse-physical or sexual (0=No, 1=Yes)
- ABUSE Abuse-physical or sexual (0=No abuse, 1= Family abuse, 2= Stranger only abuse)
- RAWPF Raw SF-36 physical functioning
- PF SF-36 physical functioning (0-100)
- RAWRP Raw SF-36 role-physical
- RP SF-36 role physical (0-100)
- RAWBP Raw SF-36 pain index
- BP SF-36 pain index (0-100)
- RAWGH Raw SF-36 general health perceptions
- GH SF-36 general health perceptions (0-100)
- RAWVT Raw SF-36 vitality
- VT SF-36 vitality 0-100)
- RAWSF Raw SF-36 social functioning
- SF SF-36 social functioning (0-100)
- RAWRE Raw SF-36 role-emotional
- RE SF-36 role-emotional (0-100)
- RAWMH Raw SF-36 mental health index
- MH SF-36 mental health index (0-100)
- HT Raw SF-36 health transition item
- PCS Standardized physical component scale-00
- MCS Standardized mental component scale-00

- CES_D CES-D score, measure of depressive symptoms, high scores are worse
- CESD_CUT CES-D score > 21 y/n (0=No, 1=Yes)
- C_MS ASI-Composite medical status
- C_AU ASI-Composite score for alcohol use
- C_DU ASI-Composite score for drug use
- CUAD_C CUAD-Cocaine
- CUAD_H CUAD-Heroin
- RAW_RE SOCRATES-Rognition-Raw
- DEC_RE SOCRATES-Recognition-Decile
- RAW_AM SOCRATES-Ambivalence-Raw
- DEC_AM SOCRATES-Ambivalence-Decile
- RAW_TS SOCRATES-Taking steps-Raw
- DEC_TS SOCRATES-Taking steps-Decile
- RAW_ADS ADS score
- PHYS InDUC-2L-Physical-Raw
- PHYS2 InDUC-2L-Physical 9Raw (w/o M48)
- INTER InDUC-2L-Interpersonal-Raw
- INTRA InDUC-2L-Intrapersonal-Raw
- IMPUL InDUL-2L-Impulse control-Raw
- IMPUL2 InDUC-2L-Impulse control-Raw (w/0 M23)
- SR InDUC-2L-Social responsibility-Raw
- CNTRL InDUC-2L-Control score
- INDTOT InDUC-2LTotal drlnC sore-Raw
- INDTOT2 InDUC-2L-Total drlnC-Raw- w/o M23 and M48
- PSS_FR Perceived social support-friends
- PSS_FA Perceived social support-family
- DRUGRISK RAB-Drug risk total
- SEXRISK RAB-Sex risk total
- TOTALRAB RAB-Total RAB sore
- RABSCALE RAB scale sore
- CHR_6M Chronic medical conds/HIV-past 6m y/n (0=No, 1=Yes)
- RCT_LINK Did subject link to primary care (RCT)–This time point (0=No, 1=Yes)
- REG_MD Did subject report having regular doctor–This time point (0=No, 1=Yes)
- ANY_VIS # visits to regular doctor’s office–This time point
- ANY_VIS_CUMUL Cumulative # visits to regular doctor’s office
- PC_REC Primary care received: Linked & #visits (0=Not linked, 1=Linked, 1 visit, 2=Linked, 2+ visits)

- PC_REC7 Primary cared received: linked & # visits (0=Not linked, 1=Linked, 1 visit, 2=Linked, 2 visits, 3=Linked, 3 visits, 4=Linked, 4 visits, 5= Linked, 5 visits, 6=Linked, 6+visits)
- SATREAT Any BSAS substance abuse this time point (0=No, 1=Yes)
- DRINKSTATUS Drank alcohol since leaving detox-6m
- DAYSDRINK Time (days) from baseline to first drink since leaving detox-6m
- ANYSUBSTATUS Used alcohol, heroin, or cocaine since leaving detox-6m
- DAYSANYSUB time (days) from baseline to first alcohol, heroin, or cocaine since leaving detox-6m
- LINKSTATUS Linked to primary care within 12 months (by administrative record)
- DAYSLINK Time (days) to linkage to primary care within 12 months (by administrative record)

Details

Eligible subjects were adults, who spoke Spanish or English, reported alcohol, heroin or cocaine as their first or second drug of choice, resided in proximity to the primary care clinic to which they would be referred or were homeless. Patients with established primary care relationships they planned to continue, significant dementia, specific plans to leave the Boston area that would prevent research participation, failure to provide contact information for tracking purposes, or pregnancy were excluded.

Subjects were interviewed at baseline during their detoxification stay and follow-up interviews were undertaken every 6 months for 2 years. A variety of continuous, count, discrete, and survival time predictors and outcomes were collected at each of these five occasions.

This dataset is a superset of the HELPmiss and HELPrct datasets which include far fewer variables.

Source

<http://www.math.smith.edu/help>

References

Samet JH, Larson MJ, Horton NJ, Doyle K, Winter M, and Saitz R. Linking alcohol and drug-dependent adults to primary medical care: A randomized controlled trial of a multi-disciplinary health intervention in a detoxification unit. *Addiction*, 2003; 98(4):509-516.

See Also

[HELPrct](#), and [HELPMiss](#).

Examples

```
data(HELFull)
```

Description

The HELP study was a clinical trial for adult inpatients recruited from a detoxification unit. Patients with no primary care physician were randomized to receive a multidisciplinary assessment and a brief motivational intervention or usual care, with the goal of linking them to primary medical care.

Usage

```
data(HELPMiss)
```

Format

Data frame with 470 observations on the following variables.

- `age` subject age at baseline (in years)
- `anysub` use of any substance post-detox: a factor with levels `no` `yes`
- `cesd` Center for Epidemiologic Studies Depression measure of depressive symptoms at baseline (higher scores indicate more symptoms)
- `d1` lifetime number of hospitalizations for medical problems (measured at baseline)
- `daysanysub` time (in days) to first use of any substance post-detox
- `dayslink` time (in days) to linkage to primary care
- `drugrisk` Risk Assessment Battery drug risk scale at baseline
- `e2b` number of times in past 6 months entered a detox program (measured at baseline)
- `female` 0 for male, 1 for female
- `sex` a factor with levels `male` `female`
- `g1b` experienced serious thoughts of suicide in last 30 days (measured at baseline): a factor with levels `no` `yes`
- `homeless` housing status: a factor with levels `housed` `homeless`
- `i1` average number of drinks (standard units) consumed per day, in the past 30 days (measured at baseline)
- `i2` maximum number of drinks (standard units) consumed per day, in the past 30 days (measured at baseline)
- `id` subject identifier
- `indtot` Inventory of Drug Use Consequences (InDUC) total score (measured at baseline)
- `linkstatus` post-detox linkage to primary care (0 = no, 1 = yes)
- `link` post-detox linkage to primary care: `no` `yes`
- `mcs` SF-36 Mental Component Score (measured at baseline, higher scores are better)
- `pcs` SF-36 Physical Component Score (measured at baseline, higher scores are better)

- pss_fr perceived social support by friends (measured at baseline)
- racegrp race/ethnicity: levels black hispanic other white
- satreat any BSAS substance abuse treatment at baseline: no yes
- sexrisk Risk Assessment Battery sex risk score (measured at baseline)
- substance primary substance of abuse: alcohol cocaine heroin
- treat randomized to HELP clinic: no yes

Details

Eligible subjects were adults, who spoke Spanish or English, reported alcohol, heroin or cocaine as their first or second drug of choice, resided in proximity to the primary care clinic to which they would be referred or were homeless. Patients with established primary care relationships they planned to continue, significant dementia, specific plans to leave the Boston area that would prevent research participation, failure to provide contact information for tracking purposes, or pregnancy were excluded.

Subjects were interviewed at baseline during their detoxification stay and follow-up interviews were undertaken every 6 months for 2 years. A variety of continuous, count, discrete, and survival time predictors and outcomes were collected at each of these five occasions.

This dataset is a superset of the HELPrct data with 17 subjects with partially observed data on some of the baseline variables. This is a subset of the HELPfull data which includes 5 timepoints and many additional variables.

Source

<http://www.math.smith.edu/help>

References

Samet JH, Larson MJ, Horton NJ, Doyle K, Winter M, and Saitz R. Linking alcohol and drug-dependent adults to primary medical care: A randomized controlled trial of a multi-disciplinary health intervention in a detoxification unit. *Addiction*, 2003; 98(4):509-516.

See Also

[HELPrct](#) , and [HELPfull](#).

Examples

```
data(HELPmiss)
```

HELPrct

*Health Evaluation and Linkage to Primary Care***Description**

The HELP study was a clinical trial for adult inpatients recruited from a detoxification unit. Patients with no primary care physician were randomized to receive a multidisciplinary assessment and a brief motivational intervention or usual care, with the goal of linking them to primary medical care.

Usage

```
data(HELPrct)
```

Format

Data frame with 453 observations on the following variables.

- age subject age at baseline (in years)
- anysub use of any substance post-detox: a factor with levels no yes
- cesd Center for Epidemiologic Studies Depression measure at baseline (high scores indicate more depressive symptoms)
- d1 lifetime number of hospitalizations for medical problems (measured at baseline)
- daysanysub time (in days) to first use of any substance post-detox
- dayslink time (in days) to linkage to primary care
- drugrisk Risk Assessment Battery drug risk scale at baseline
- e2b number of times in past 6 months entered a detox program (measured at baseline)
- female 0 for male, 1 for female
- sex a factor with levels male female
- g1b experienced serious thoughts of suicide in last 30 days (measured at baseline): a factor with levels no yes
- homeless housing status: a factor with levels housed homeless
- i1 average number of drinks (standard units) consumed per day, in the past 30 days (measured at baseline)
- i2 maximum number of drinks (standard units) consumed per day, in the past 30 days (measured at baseline)
- id subject identifier
- indtot Inventory of Drug Use Consequences (InDUC) total score (measured at baseline)
- linkstatus post-detox linkage to primary care (0 = no, 1 = yes)
- link post-detox linkage to primary care: no yes
- mcs SF-36 Mental Component Score (measured at baseline, lower scores indicate worse status)

- pcs SF-36 Physical Component Score (measured at baseline, lower scores indicate worse status)
- pss_fr perceived social support by friends (measured at baseline, higher scores indicate more support)
- racegrp race/ethnicity: levels black hispanic other white
- satreat any BSAS substance abuse treatment at baseline: no yes
- sexrisk Risk Assessment Battery sex risk score (measured at baseline)
- substance primary substance of abuse: alcohol cocaine heroin
- treat randomized to HELP clinic: no yes

Details

Eligible subjects were adults, who spoke Spanish or English, reported alcohol, heroin or cocaine as their first or second drug of choice, resided in proximity to the primary care clinic to which they would be referred or were homeless. Patients with established primary care relationships they planned to continue, significant dementia, specific plans to leave the Boston area that would prevent research participation, failure to provide contact information for tracking purposes, or pregnancy were excluded.

Subjects were interviewed at baseline during their detoxification stay and follow-up interviews were undertaken every 6 months for 2 years. A variety of continuous, count, discrete, and survival time predictors and outcomes were collected at each of these five occasions.

This dataset is a subset of the [HELPrct](#) data which includes an additional 17 subjects with partially observed data on some of the baseline variables. This is also a subset of the [HELPrct](#) data which includes 5 timepoints and many additional variables for all subjects.

Note

The HELPrct data set was originally named HELP but has been renamed to avoid confusion with the help function.

Source

<http://www.math.smith.edu/help>

References

Samet JH, Larson MJ, Horton NJ, Doyle K, Winter M, and Saitz R. Linking alcohol and drug-dependent adults to primary medical care: A randomized controlled trial of a multi-disciplinary health intervention in a detoxification unit. *Addiction*, 2003; 98(4):509-516.

See Also

[HELPrct](#), and [HELPrct](#).

Examples

```
data(HELPrct)
```

inferArgs

Infer arguments

Description

The primary purpose is for inferring argument settings from names derived from variables occurring in a formula. For example, the default use is to infer limits for variables without having to call them `xlim` and `ylim` when the variables in the formula have other names. Other uses could easily be devised by specifying different variants.

Usage

```
inferArgs(vars, dots, defaults = alist(xlim = , ylim = , zlim = ),
          variants = c(".lim", "lim"))
```

Arguments

<code>vars</code>	a vector of variable names to look for
<code>dots</code>	a named list of argument values
<code>defaults</code>	named list or alist of default values for limits
<code>variants</code>	a vector of optional postfixes for limit-specifying variable names

Value

a named list or alist of limits. The names are determined by the names in `defaults`.
If multiple variants are matched, the first is used.

Examples

```
inferArgs(c('x', 'u', 't'), list(t=c(1,3), xlim=c(1,10), u=c(1,3), u.lim=c(2,4)))
inferArgs(c('x', 'u'), list(u=c(1,3)), defaults=list(xlim=c(0,1), ylim=NULL))
```

integrateODE

Integrate ordinary differential equations

Description

A formula interface to integration of an ODE with respect to "t"

Usage

```
integrateODE(dyn, ..., tdur)
```

Arguments

dyn	a formula specifying the dynamics, e.g. $dx \sim -a*x$ for $\$dx/dt = -ax\$$.
...	arguments giving additional formulas for dynamics in other variables, assignments of parameters, and assignments of initial conditions
tdur	the duration of integration. Or, a list of the form <code>list(from=5, to=10, dt=.001)</code>

Details

The equations must be in first-order form. Each dynamical equation uses a formula interface with the variable name given on the left-hand side of the formula, preceded by a d, so use $dx \sim -k*x$ for exponential decay. All parameters (such as k) must be assigned numerical values in the argument list. All dynamical variables must be assigned initial conditions in the argument list. The returned value will be a list with one component named after each dynamical variable. The component will be a spline-generated function of t.

Value

a list with splined function of time for each dynamical variable

Examples

```
soln = integrateODE(dx~r*x*(1-x/k), k=10, r=.5, tdur=20, x=1)
soln$x(10)
soln$x(30) # outside the time interval for integration
plotFun(soln$x(t)~t, tlim=range(0,20))
soln2 = integrateODE(dx~y, dy~-x, x=1, y=0, tdur=10)
plotFun(soln2$y(t)~t, tlim=range(0,10))
# SIR epidemic
epi = integrateODE(dS~-a*S*I, dI ~ a*S*I - b*I, a=0.0026, b=.5, S=762, I=1, tdur=20)
```

interval

Extract summary statistics

Description

Extract confidence intervals, test statistics or p-values from an htest object.

Usage

```
interval(object, parm, level = 0.95, ...)
```

```
## S3 method for class 'htest'
confint(object, parm, level, ...)
```

```
pval(x, ...)
```

```
## S3 method for class 'htest'
```

```

pval(x, digits = 4, verbose = FALSE, ...)

stat(x, ...)

## S3 method for class 'htest'
stat(x, ...)

```

Arguments

x	An object of class htest.
...	Additional arguments.
object	a fitted model object or an htest object.
parm	a specification of which parameters are to be given confidence intervals, either a vector of numbers or a vector of names. If missing, all parameters are considered.
level	the confidence level required.
verbose	a logical
digits	number of digits to display in verbose output

Value

the extracted p-value, confidence interval, or test statistic

Examples

```

confint(t.test(rnorm(100)))
pval(t.test(rnorm(100)))
stat(t.test(rnorm(100)))
confint(var.test(rnorm(10,sd=1), rnorm(20, sd=2)))
pval(var.test(rnorm(10,sd=1), rnorm(20, sd=2)))

data(HELPrct)
stat(t.test (age ~ shuffle(sex), HELPrct))
# Compare to test statistic computed with permuted values of sex.
do(10) * stat(t.test (age ~ shuffle(sex), HELPrct))

```

is.wholenumber	<i>Check for whole number values</i>
----------------	--------------------------------------

Description

Unlike [is.integer](#), which checks the type of argument is integer, this function checks whether the value of the argument is an integer (within a specified tolerance).

Usage

```
is.wholenumber(x, tol = .Machine$double.eps^0.5)
```


Arguments

x a vector
tol a numeric tolerance

Details

This function is borrowed from the examples for [is.integer](#)

Value

a logical vector indicating whether x has a whole number value

Examples

```
is.wholenumber(1)
all(is.wholenumber(rbinom(100,10,.5)))
is.wholenumber((1:10)/2)
```

joinFrames	<i>Join data frames</i>
------------	-------------------------

Description

Join data frames

Usage

```
joinFrames(...)  
joinTwoFrames(left, right)
```

Arguments

left, right data frames
... data frames to be joined

Value

a data frame containing columns from each of data frames being joined.

KidsFeet

Foot measurements in children

Description

These data were collected by a statistician, Mary C. Meyer, in a fourth grade classroom in Ann Arbor, MI, in October 1997. They are a convenience sample — the kids who were in the fourth grade.

Usage

```
data(KidsFeet)
```

Format

A data frame with 39 observations on the following variables.

- name a factor with levels corresponding to the name of each child
- birthmonth the month of birth
- birthyear the year of birth
- length length of longer foot (in cm)
- width width of longer foot (in cm)
- sex a factor with levels B G
- biggerfoot a factor with levels L R
- domhand a factor with levels L R

Details

Quoted from the source: “From a very young age, shoes for boys tend to be wider than shoes for girls. Is this because boys have wider feet, or because it is assumed that girls, even in elementary school, are willing to sacrifice comfort for fashion? To assess the former, a statistician measures kids’ feet.”

References

Mary C. Meyer (2006) “Wider Shoes for Wider Feet?” *Journal of Statistics Education* 14(1), www.amstat.org/publications/jse/v14n1/datasets.meyer.html

Examples

```
data(KidsFeet)
```

ladd	<i>Add to Lattice Plots</i>
------	-----------------------------

Description

Simplified lattice plotting by adding additional elements to existing plots.

Usage

```
ladd(x, col, row, highlight = FALSE, verbose = FALSE)
```

Arguments

x	callable graphical element to be added to a panel or panels in a lattice plot
col, row	identifies desired panel(s) in multi-panel plots. If missing, all columns or rows are used.
verbose	a logical indicating whether to display some information about modified panels.
highlight	a logical indicating whether to highlight panels as they are being modified.

Details

ladd is simply a wrapper around [trellis.focus](#) and [trellis.unfocus](#).

Author(s)

Randall Pruim (<rpruim@calvin.edu>)

Examples

```
p <- xyplot(rnorm(100) ~rnorm(100))
print(p)
ladd(panel.abline(a=0,b=1))
ladd(panel.abline(h=0,col='blue'))
ladd(grid.text('Hello'))
ladd(grid.text(x=.95,y=.05,'text here',just=c('right','bottom')))
q <- xyplot(rnorm(100) ~rnorm(100)|factor(rbinom(100,4,.5)))
q <- update(q, layout=c(3,2))
print(q)
ladd(panel.abline(a=0,b=1))
ladd(panel.abline(h=0,col='blue'))
ladd( grid.text("(2,1)",gp=gpar(cex=3,alpha=.5)), 2, 1)
print(q)
ladd( grid.text(paste(current.column(), current.row(),sep=', '), gp=gpar(cex=3,alpha=.5)) )
histogram( ~eruptions, data=faithful)
ladd(panel.densityplot(faithful$eruptions))
```

`linear.algebra`*Functions for teaching linear algebra.*

Description

These functions provide a formula based interface to the construction of matrices from data and for fitting. You can use them both for numerical vectors and for functions of variables in data frames. These functions are intended to support teaching basic linear algebra with a particular connection to statistics.

Usage

```
mat(A, data = parent.frame())
```

```
singvals(A, data = parent.frame())
```

```
dot(u, v)
```

Arguments

A	a formula. In <code>mat</code> and <code>singvals</code> , only the right-hand side is used.
u	a numeric vector
data	a data frame from which to pull out numerical values for the variables in the formula
...	additional arguments (currently ignored)
	<code>mat</code> returns a model matrix
	To demonstrate singularity, use <code>singvals</code> .
v	a numeric vector

Value

`mat` returns a matrix

`singvals` gives singular values for each column in the model matrix

`dot` returns the dot product of `u` and `v`

See Also

[project](#)

[linearModel](#), which returns a function.

Examples

```

a <- c(1,0,0); b <- c(1,2,3); c <- c(4,5,6); x <- rnorm(3)
dot(b,c) # dot product
# Formula interface
mat(~a+b)
mat(~a+b+1)
mat(~length+sex, data=KidsFeet)
singvals(~length*sex*width, data=KidsFeet)

```

logical2factor	<i>Turn logicals into factors; leave other things alone</i>
----------------	---

Description

Turn logicals into factors; leave other things alone

Usage

```

logical2factor(x, ...)

## Default S3 method:
logical2factor(x, ...)

## S3 method for class 'data.frame'
logical2factor(x, ...)

```

Arguments

x	a vector or data frame
...	additional arguments (currently ignored)

Value

If x is a vector either x or the result of converting x into a factor with levels TRUE and FALSE (in that order); if x is a data frame, a data frame with all logicals converted to factors in this manner.

logit	<i>Logit and inverse logit functions</i>
-------	--

Description

Logit and inverse logit functions

Usage

```
logit(x)
ilogit(x)
```

Arguments

x a numeric vector

Value

For logit the value is

$$\log(x/(1-x))$$

For ilogit the value is

$$\exp(x)/(1+\exp(x))$$

Examples

```
p <- seq(.1, .9, by=.10)
l <- logit(p); l
ilogit(l)
ilogit(l) == p
```

maggregate

Aggregate for mosaic

Description

Compute function on subsets of a variable in a data frame.

Usage

```
maggregate(formula, data = parent.frame(), FUN, subset,
  overall = mosaic.par.get("aggregate.overall"), method = c("default",
  "ddply"), drop = FALSE, multiple = FALSE, groups = NULL,
  .name = deparse(substitute(FUN)), ...)
```

Arguments

formula a formula. Left side provides variable to be summarized. Right side and condition describe subsets. If the left side is empty, right side and condition are shifted over as a convenience.

data a data frame

FUN a function to apply to each subset

subset a logical indicating a subset of data to be processed.

drop a logical indicating whether unused levels should be dropped.

method	used for aggregation. Choosing "ddply" requires that plyr is installed.
overall	currently unused
.name	a name used for the resulting object
groups	grouping variable that will be folded into the formula (if there is room for it). This offers some additional flexibility in how formulas can be specified.
multiple	a logical indicating whether FUN returns multiple values
...	additional arguments passed to FUN

Value

a vector

Examples

```
maggregate( cesd ~ sex, HELPrct, FUN=mean )
# using groups instead
maggregate( ~ cesd, groups = sex, HELPrct, FUN=sd )
# the next four all do the same thing
maggregate( cesd ~ sex & homeless, HELPrct, FUN=mean )
maggregate( cesd ~ sex | homeless, HELPrct, FUN=sd )
maggregate( ~ cesd | sex , groups= homeless, HELPrct, FUN=sd )
maggregate( cesd ~ sex, groups = homeless, HELPrct, FUN=sd )
# this is unusual, but also works.
maggregate( cesd ~ NULL , groups = sex, HELPrct, FUN=sd )
```

makeColorscheme	<i>Create a color generating function from a vector of colors</i>
-----------------	---

Description

Create a color generating function from a vector of colors

Usage

```
makeColorscheme(col)
```

Arguments

col	a vector of colors
-----	--------------------

Value

a function that generates a vector of colors interpolated among the colors in col

Examples

```
cs <- makeColorscheme( c('red','white','blue') )
cs(10)
cs(10, alpha=.5)
```

 makeFun

Create a function from a formula

Description

Provides an easy mechanism for creating simple "mathematical" functions via a formula interface.

Usage

```
makeFun(object, ...)

## S4 method for signature 'formula'
makeFun(object, ..., strict.declaration = TRUE,
        use.environment = TRUE, suppress.warnings = FALSE)

## S4 method for signature 'lm'
makeFun(object, ..., transform = identity)

## S4 method for signature 'glm'
makeFun(object, ..., type = c("response", "link"),
        transform = identity)

## S4 method for signature 'nls'
makeFun(object, ..., transform = identity)
```

Arguments

object	an object from which to create a function. This should generally be specified without naming.
...	additional arguments in the form <code>var = val</code> that set default values for the inputs to the function.
strict.declaration	if TRUE (the default), an error is thrown if default values are given for variables not appearing in the object formula.
use.environment	if TRUE, then variables implicitly defined in the object formula can take default values from the environment at the time <code>makeFun</code> is called. A warning message alerts the user to this situation, unless <code>suppress.warnings</code> is TRUE.
suppress.warnings	A logical indicating whether warnings should be suppressed.
transform	a function used to transform the response. This can be useful to invert a transformation used on the response when creating the model.
type	one of 'response' (default) or 'link' specifying scale to be used for value of function returned.

Details

The definition of the function is given by the left side of a formula. The right side lists at least one of the inputs to the function. The inputs to the function are all variables appearing on either the left or right sides of the formula. Those appearing in the right side will occur in the order specified. Those not appearing in the right side will appear in an unspecified order.

Value

a function

Examples

```
f <- makeFun( sin(x^2 * b) ~ x & y & a); f
g <- makeFun( sin(x^2 * b) ~ x & y & a, a=2 ); g
h <- makeFun( a * sin(x^2 * b) ~ b & y, a=2, y=3); h
model <- lm( log(length) ~ log(width), data=KidsFeet)
f <- makeFun(model, transform=exp)
f(8.4)
head(KidsFeet,1)
model <- lm(wage ~ poly(exper,degree=2), data=CPS85)
fit <- makeFun(model)
xyplot(wage ~ exper, data=CPS85)
plotFun(fit(exper) ~ exper, add=TRUE)
model <- glm(wage ~ poly(exper,degree=2), data=CPS85, family=gaussian)
fit <- makeFun(model)
xyplot(wage ~ exper, data=CPS85)
plotFun(fit(exper) ~ exper, add=TRUE)
model <- nls( wage ~ A + B * exper + C * exper^2, data=CPS85, start=list(A=1,B=1,C=1) )
fit <- makeFun(model)
xyplot(wage ~ exper, data=CPS85)
plotFun(fit(exper) ~ exper, add=TRUE)
```

Marriage

Marriage records

Description

Marriage records from the Mobile County, Alabama, probate court.

Usage

```
data(Marriage)
```

Format

A data frame with 98 observations on the following variables.

- bookpageID a factor with levels for each book and page (unique identifier)

- `appdate` a factor with levels corresponding to each of the dates on which the application was filed (in the form MO/DY/YY, e.g. 1/22/99 represents January 22, 1999)
- `ceremonydate` a factor with levels corresponding to the date of the ceremony
- `delay` number of days between the application and the ceremony
- `officialTitle` a factor with levels BISHOP CATHOLIC PRIEST CHIEF CLERK CIRCUIT JUDGE ELDER MARRIAGE OFFICIAL MINISTER PASTOR REVEREND
- `person` a factor with levels Bride Groom
- `dob` a factor with levels corresponding to the date of birth of the person
- `age` age of the person (in years)
- `race` a factor with levels American Indian Black Hispanic White
- `prevcount` the number of previous marriages of the person, as listed on the application
- `prevconc` the way the last marriage ended, as listed on the application
- `hs` the number of years of high school education, as listed on the application
- `college` the number of years College education, as listed on the application. Where no number was listed, this field was left blank, unless less than 12 years High School was reported, in which case it was entered as 0.
- `dayOfBirth` the day of birth, as a number from 1 to 365 counting from January 1
- `sign` the astrological sign, with levels Aquarius Aries Cancer Capricorn Gemini Leo Libra Pisces Saggitarius Scorpio Taurus Virgo

Details

The calculation of the astrological sign may not correctly sort people directly on the borders between signs. This variable is not part of the original record.

Source

The records were collected through <http://www.mobilecounty.org/probatecourt/recordssearch.htm>

Examples

```
data(Marriage)
```

mean

Aggregating functions

Description

The mosaic package makes several summary statistic functions (like mean and sd) formula aware.

Usage

```

mean(x, ..., data, groups = NULL, ..fun.. = base::mean)

median(x, ..., data, groups = NULL, ..fun.. = stats::median)

range(x, ..., data, groups = NULL, ..fun.. = base::range)

sd(x, ..., data, groups = NULL, ..fun.. = stats::sd)

max(x, ..., data, groups = NULL, ..fun.. = base::max)

min(x, ..., data, groups = NULL, ..fun.. = base::min)

sum(x, ..., data, groups = NULL, ..fun.. = base::sum)

IQR(x, ..., data, groups = NULL, ..fun.. = stats::IQR)

fivenum(x, ..., data, groups = NULL, ..fun.. = stats::fivenum)

iqr(x, ..., data, groups = NULL, ..fun.. = stats::IQR)

prod(x, ..., data, groups = NULL, ..fun.. = base::prod)

sum(x, ..., data, groups = NULL, ..fun.. = base::sum)

favstats(x, ..., data, groups = NULL, ..fun.. = fav_stats)

var(x, ..., data, groups = NULL, ..fun.. = stats::var)

cor(x, y = NULL, ..., data = parent.frame())

cov(x, y = NULL, ..., data = parent.frame())

```

Arguments

x	an object, often a formula
y	an object, often a numeric vector
..fun..	the underlying function used in the computation
groups	a grouping variable, typically a name of a variable in data
data	a data frame in which to evaluate formulas (or bare names)
...	additional arguments

Examples

```

mean( HELPrct$age )
mean( ~ age, data=HELPrct )
mean( age ~ sex + substance, data=HELPrct )

```

```

mean( ~ age | sex + substance, data=HELPrct )
mean( sqrt(age), data=HELPrct )
sum( ~ age, data=HELPrct )
sd( HELPrct$age )
sd( ~ age, data=HELPrct )
sd( age ~ sex + substance, data=HELPrct )
var( HELPrct$age )
var( ~ age, data=HELPrct )
var( age ~ sex + substance, data=HELPrct )
IQR( width ~ sex, data=KidsFeet )
iqr( width ~ sex, data=KidsFeet )
favstats( width ~ sex, data=KidsFeet )

cor( length ~ width, data=KidsFeet )
cov ( length ~ width, data=KidsFeet )

```

Mites

Mites and Wilt Disease

Description

Data from an experiment to test whether exposure to mites protects against Wilt Disease in cotton plants.

Usage

```
data(Mites)
```

Format

A data frame with 47 observations on the following variables.

- treatment a factor with levels mites and no mites
- outcome a factor with levels wilt and no wilt

Details

Researchers suspected that attack of a plant by one organism induced resistance to subsequent attack by a different organism. Individually potted cotton plants were randomly allocated to two groups: infestation by spider mites or no infestation. After two weeks the mites were dutifully removed by a conscientious research assistant, and both groups were inoculated with *Verticillium*, a fungus that causes Wilt disease. More information can be found at <https://www.causeweb.org/webinar/activity/2010-01/>.

Source

Statistics for the Life Sciences, Third Edition; Myra Samuels & Jeffrey Witmer (2003), page 409.

Examples

```
data(Mites)
tally(~ treatment + outcome, data=Mites)
tally(~ outcome | treatment, format="percent", data=Mites)
```

mm

*Construct a model based on groupwise means***Description**

Calculate groupwise means, presenting the result as a model in the style of `lm`.

Usage

```
mm(formula, data = parent.frame(), fun = mean, drop = TRUE, ...)

## S3 method for class 'groupwiseModel'
confint(object, parm, level = 0.95, ...,
        pooled = TRUE, margin = FALSE)

## S3 method for class 'groupwiseModel'
coef(object, ...)

## S3 method for class 'groupwiseModel'
print(x, ..., digits = max(3, getOption("digits") -
  3))

## S3 method for class 'groupwiseModel'
residuals(object, ...)

## S3 method for class 'groupwiseModel'
fitted(object, ...)

## S3 method for class 'groupwiseModel'
summary(object, ...)

## S3 method for class 'summary_groupwiseModel'
print(x, digits = max(3, getOption("digits")
  - 3), ...)
```

Arguments

formula	A formula. The left-hand side specifies the variable over which the mean will be taken. The right-hand side gives the grouping variables, separated by <code>&</code> .
data	A data frame to which the formula variables refer. If not specified, variables will be taken from the current environment.

fun	The function used to calculate the means. Default: mean.
drop	Logical flag indicating whether to drop unoccupied groups. Default TRUE. NOT YET IMPLEMENTED.
...	Additional arguments to be passed to the fun doing the calculation.
parm	Not used
level	The confidence level (e.g., 0.95)
pooled	Whether to use a pooled variance of residuals to compute the standard error. (This is what lm does.)
margin	Whether to present the margin of error rather than the lower and upper bounds
x	Object to be printed
digits	number of digits to display
object	groupwiseMean object from which to extract the residuals

Details

mm is a sort of training function for lm, meant to provide a basis for discussing inference and introducing resampling in a simple, intuitive setting of groupwise means. lm provides a better, more general facility. When using lm to recreate the results of mm, include all the interaction terms, that is, use * instead of &. See the examples.

Value

mm returns an object of class groupwiseModel. The functions fitted.values, residuals, coefficients, and summary are useful for extracting various features of the value returned by mm

See Also

[lm](#), [do](#)

Examples

```
mm( wage ~ sex, data=CPS85 )
mm( wage ~ sex & married, data=CPS85 )
lm( wage ~ sex*married-1, data=CPS85)
do(5) * mm( wage ~ sex & married, data=resample(CPS85))
mod <- mm( width ~ domhand, data=KidsFeet)
summary(mod)
resid(mod)
fitted(mod)
```

modelVars	<i>extract predictor variables from a model</i>
-----------	---

Description

extract predictor variables from a model

Usage

```
modelVars(model)
```

Arguments

model a model, typically of class lm or glm

Value

a vector of variable names

Examples

```
model <- lm( wage ~ poly(exper,degree=2), data=CPS85 )
modelVars(model)
```

mosaic.options	<i>Setting options for mosaic package functions</i>
----------------	---

Description

A mechanism for setting options in the mosaic package.

Usage

```
mosaic.options(...)
```

```
mosaic.getOption(name)
```

```
mosaic.par.set(name, value, ..., theme, warn = TRUE, strict = FALSE)
```

```
mosaic.par.get(name = NULL)
```

```
restoreLatticeOptions()
```

```
mosaicLatticeOptions()
```

Arguments

name	the name of the option being set
value	the value to which to set the option
theme	a list appropriate for a mosaic theme
warn	a logical. UNUSED at present.
strict	a logical or numeric.
...	additional arguments that are turned into a list if a list cannot be inferred from theme, name, and value.

Details

restoreLatticeOptions returns any lattice options that were changed when the mosaic package was loaded back to their pre-mosaic state.

mosaicLatticeOptions sets a number of defaults for lattice graphics.

mosaic_formula	<i>Convert formulas into standard shapes</i>
----------------	--

Description

These functions convert formulas into standard shapes, including by incorporating a groups argument.

Usage

```
mosaic_formula(formula, groups = NULL, envir = parent.frame(),
               max.slots = 3)
```

```
mosaic_formula_q(formula, groups = NULL, envir = parent.frame(),
                 max.slots = 3)
```

Arguments

formula	a formula
groups	a name used for grouping
max.slots	an integer specifying the maximum number of slots for the resulting formula. An error results from trying to create a formula that is too complex.
envir	the environment in which the resulting formula may be evaluated. May also be NULL, a list, a data frame, or a pairlist.

Details

mosaic_formula_q uses nonstandard evaluation of groups that may be necessary for use within other functions. mosaic_formula is a wrapper around mosaic_formula_q and quotes groups before passing it along.

Examples

```

mosaic_formula( ~ x | z )
mosaic_formula( ~ x, groups=g )
mosaic_formula( y ~ x, groups=g )
# this is probably not what you want for interactive use.
mosaic_formula_q( y ~ x, groups=g )
# but it is for programming
foo <- function(x, groups=NULL) {
  mosaic_formula_q(x, groups=groups, envir=parent.frame())
}
foo( y ~ x , groups = g)

```

mPlot

Interactive plotting

Description

These functions provide a menu selection system (via **manipulate**) so that different aspects of a plot can be selected interactively. The **ggplot2** or **lattice** command for generating the plot currently being displayed can be copied to the console, whence it can be copied to a document for later direct, non-interactive use.

Usage

```
mPlot(data, default = plotType, system = c("lattice", "ggplot2"),
      show = FALSE, title = "", ...)
```

```
mMap(data, default = "map", system = "ggplot2", show = FALSE,
     title = title, ...)
```

```
mScatter(data, default = c("scatter", "jitter", "boxplot", "violin"),
         system = c("lattice", "ggplot2"), show = FALSE, title = "")
```

```
mUniplot(data, default = c("histogram", "density", "frequency polygon"),
         system = c("lattice", "ggplot2"), show = FALSE, title = "")
```

Arguments

data	a data frame containing the variables that might be used in the plot. Note that for maps, the data frame must contain coordinates of the polygons comprising the map and a variable for determining which coordinates are part of the same region. See sp2df for one way to create such a data frame. Typically merge will be used to combine the map data with some auxiliary data to be displayed as fill color on the map, although this is not necessary if all one wants is a map.
default	default type of plot to create; one of "scatter", "jitter", "boxplot", "violin", "histogram", "density", "frequency polygon", "xyplot", or "map". Unique prefixes suffice.

system	which graphics system to use (initially) for plotting (ggplot2 or lattice). A check box will allow on the fly change of plotting system.
show	a logical, if TRUE, the code will be displayed each time the plot is changed.
title	a title for the plot
...	additional arguments

Details

Only `mPlot` is required by end users. The other plotting functions are dispatched based on the value of default.

Currently maps are only supported in **ggplot2** and not in **lattice**.

Value

Nothing. Just for side effects.

Examples

```
## Not run:
mPlot(HELPrct, "scatter")
mPlot(HELPrct, "density")

## End(Not run)
```

ntiles	<i>Create factor based on roughly equally sized groups</i>
--------	--

Description

Create factor based on roughly equally sized groups

Usage

```
ntiles(x, n = 3)
```

Arguments

x	a numeric vector
n	number of quantiles

Examples

```
ntiles(1:50, 4)
```

 numD

Numerical Derivatives

Description

Constructs the numerical derivatives of mathematical expressions

Usage

```
numD(formula, ..., .hstep = NULL, add.h.control = FALSE)
```

```
setInterval(C, wrt, h)
```

```
setCorners(C, var1, var2, h)
```

```
dfdx(.function, .wrt, .hstep)
```

```
d2fdxdy(.function, .var1, .var2, .hstep)
```

```
d2fdx2(.function, .wrt, .hstep)
```

```
numerical.first.partial(f, wrt, h, av)
```

```
numerical.second.partial(f, wrt, h, av)
```

```
numerical.mixed.partial(f, var1, var2, h, av)
```

Arguments

formula	a mathematical expression (see examples and plotFun)
...	additional parameters, typically default values for mathematical parameters
.hstep	numerical finite-difference step (default is 1e-6 or 1e-4 for first and second-order derivatives, respectively)
add.h.control	arranges the returned function to have a .hstep argument that can be used to demonstrate convergence and error
C	list of arguments for evaluating the function at the "center" point
wrt	character string naming the variable with respect to which differentiation is to be done
h	the finite-difference step size
var1	character string naming the first variable with respect to which differentiation is to be done
var2	character string naming the second variable with respect to which differentiation is to be done
.function	function to be differentiated

<code>.wrt</code>	character string naming the variable with respect to which differentiation is to be done
<code>.step</code>	the finite-difference step size
<code>.var1</code>	character string naming the first variable with respect to which differentiation is to be done
<code>.var2</code>	character string naming the second variable with respect to which differentiation is to be done
<code>f</code>	function to differentiate
<code>av</code>	arguments to the function calling this

Details

Uses a simple finite-difference scheme to evaluate the derivative. The function created will not contain a formula for the derivative. Instead, the original function is stored at the time the derivative is constructed and that original function is re-evaluated at the finitely-spaced points of an interval. If you redefine the original function, that won't affect any derivatives that were already defined from it. Numerical derivatives, particularly high-order ones, are unstable. The finite-difference parameter `.hstep` is set, by default, to give reasonable results for first- and second-order derivatives. It's tweaked a bit so that taking a second derivative by differentiating a first derivative will give reasonably accurate results. But, if taking a second derivative, much better to do it in one step to preserve numerical accuracy.

Value

a function implementing the derivative as a finite-difference approximation

Numerical partials

These functions are not intended for direct use. They just package up the numerical differentiation process to make functions returned by `numD` and `D` easier to read.

Note

WARNING: In the expressions, do not use variable names beginning with a dot, particularly `.f` or `.h`

Helper function for `numD` for unmixed partials

Helper function for `numD` for mixed partials

Helper function for `numD` for first-order derivs.

Helper function for `numD` for second-order mixed partials

Helper function for `numD` for second-order derivs

Not for direct use. This just packages up the numerical differentiation process to make functions returned by `numD` and `D` easier to read.

Not for direct use. This just packages up the numerical differentiation process to make functions returned by `numD` and `D` easier to read.

Author(s)

Daniel Kaplan (<kaplan@macalester.edu>)

See Also

[D](#), [symbolicD](#), [makeFun](#), [antiD](#), [plotFun](#)

Examples

```
g = numD( a*x^2 + x*y ~ x, a=1)
g(x=2,y=10)
gg = numD( a*x^2 + x*y ~ x&x, a=1)
gg(x=2,y=10)
ggg = numD( a*x^2 + x*y ~ x&y, a=1)
ggg(x=2,y=10)
h = numD( g(x=x,y=y,a=a) ~ y, a=1)
h(x=2,y=10)
f = numD( sin(x)~x, add.h.control=TRUE)
plotFun( f(3,.hstep=h)~h, hlim=range(.0000001,.00001))
ladd( panel.abline(cos(3),0))
```

oddsRatio

Odds Ratio for 2X2 Contingency Tables

Description

This function calculates the odds ratio for a 2 X 2 contingency table and a confidence interval (default conf.level is 95 percent) for the estimated odds ratio. x should be a matrix, data frame or table. "Successes" should be located in column 1 of x, and the treatment of interest should be located in row 2. The odds ratio is calculated as (Odds row 2) / (Odds row 1). The confidence interval is calculated from the log(OR) and backtransformed.

Usage

```
oddsRatio(x, conf.level = 0.95)
```

```
## S3 method for class 'oddsRatio'
print(x, digits = 4, ...)
```

Arguments

x	a 2 X 2 matrix, data frame or table of counts
conf.level	the confidence interval level
digits	number of digits to display
...	additional arguments

Value

p1, p2	Proportions for rows 1 and 2
o1, o2	Odds for rows 1 and 2
OR	Odds ratio
lower	the lower bound of the confidence interval
upper	the upper bound of the confidence interval
conf.level	the confidence interval level

Author(s)

Kevin Middleton (<kmm@csusb.edu>)

See Also

[chisq.test](#)

Examples

```
M1 <- matrix(c(14, 38, 51, 11), nrow = 2)
M1
oddsRatio(M1)

M2 <- matrix(c(18515, 18496, 1427, 1438), nrow = 2)
rownames(M2) <- c("Placebo", "Aspirin")
colnames(M2) <- c("No", "Yes")
M2
oddsRatio(M2)
```

panel.levelcontourplot

Lattice plot that draws a filled contour plot

Description

Used within plotFun

Usage

```
panel.levelcontourplot(x, y, z, subscripts = 1, at, shrink, labels = TRUE,
  label.style = c("mixed", "flat", "align"), contour = FALSE,
  region = TRUE, col = add.line$col, lty = add.line$lty,
  lwd = add.line$lwd, border = "transparent", ...,
  col.regions = regions$col, filled = TRUE, alpha.regions = regions$alpha)
```

Arguments

x	x on a grid
y	y on a grid
z	zvalues for the x and y
subscripts	which points to plot
at	cuts for the contours
shrink	what does this do?
labels	draw the contour labels
label.style	where to put the labels
contour	logical draw the contours
region	logical color the regions
col	color for contours
lty	type for contours
lwd	width for contour
border	type of border
...	dots additional arguments
col.regions	a vector of colors or a function (topo.colors by default) for generating such
filled	whether to fill the contours with color
alpha.regions	transparency of regions

panel.lmbands *show confidence and preciction bands on plots*

Description

show confidence and preciction bands on plots

Usage

```
panel.lmbands(x, y, interval = "confidence", level = 0.95, model = lm(y ~
x), band.col = c(conf = slcol[3], pred = slcol[2]), band.lty = c(conf =
slty[3], pred = slty[2]), band.show = TRUE, fit.show = TRUE,
band.alpha = 0.6, band.lwd = 1, npts = 100, ...)
```

Arguments

x,y	numeric vectors
interval	a vector subset of 'confidence' and 'prediction'
level	confidence level
model	model to be used for generating bands
band.col	a vector of length 1 or 2 giving the color of bands
band.lty	a vector of length 1 or 2 giving the line type for bands
band.show	logical vector of length 1 or 2 indicating whether confidence and prediction bands should be shown
fit.show	logical indicating whether the model fit should be shown
band.alpha	a vector of length 1 or 2 alpha level for bands
band.lwd	a vector of length 1 or 2 giving line width for bands
npts	resolution parameter for bands (increase to get better resolution)
...	additional arguments

panel.plotFun	<i>Panel function for plotting functions</i>
---------------	--

Description

Panel function for plotting functions

Usage

```
panel.plotFun(object, ..., type = "l", npts = NULL, zlab = NULL,
  filled = TRUE, levels = NULL, nlevels = 10, surface = FALSE,
  col.regions = topo.colors, alpha = NULL)
```

Arguments

object	an object (e.g., a formula) describing a function
npts	an integer giving the number of points (in each dimension) to sample the function
zlab	label for z axis (when in surface-plot mode)
filled	fill with color between the contours (TRUE by default)
levels	levels at which to draw contours
nlevels	number of contours to draw (if levels not specified)
surface	a logical indicating whether to draw a surface plot rather than a contour plot
col.regions	a vector of colors or a function (topo.colors by default) for generating such
type	type of plot ("l" by default)

alpha number from 0 (transparent) to 1 (opaque) for the fill colors
 ... additional arguments, typically processed by lattice panel functions such as `panel.xyplot` or `panel.levelplot`. Frequently used arguments include
 lwd line width
 lty line type
 col a color

See Also

plotFun

Examples

```
x <- runif(30,0,2*pi)
d <- data.frame( x = x, y = sin(x) + rnorm(30,sd=.2) )
xyplot( y ~ x, data=d )
ladd(panel.plotFun( sin(x) ~ x, col='red' ) )
xyplot( y ~ x | rbinom(30,1,.5), data=d )
ladd(panel.plotFun( sin(x) ~ x, col='red', lty=2 ) ) # plots sin(x) in each panel
```

panel.plotFun1 *Panel function for plotting functions*

Description

Panel function for plotting functions

Usage

```
panel.plotFun1(..f.., ..., x, y, type = "l",
  col = trellis.par.get("superpose.line")$col, npts = NULL, zlab = NULL,
  filled = TRUE, levels = NULL, nlevels = 10, surface = FALSE,
  alpha = NULL)
```

Arguments

..f.. an object (e.g., a formula) describing a function
 x,y ignored, but there for compatibility with other lattice panel functions
 col a vector of colors
 npts an integer giving the number of points (in each dimension) to sample the function
 zlab label for z axis (when in surface-plot mode)
 filled fill with color between the contours (TRUE by default)
 levels levels at which to draw contours
 nlevels number of contours to draw (if levels not specified)

surface	a logical indicating whether to draw a surface plot rather than a contour plot
type	type of plot ("1" by default)
alpha	number from 0 (transparent) to 1 (opaque) for the fill colors
...	additional arguments, typically processed by lattice panel functions such as panel.xyplot or panel.levelplot . Frequently used arguments include
	lwd line width
	lty line type
	col a color

See Also

plotFun

Examples

```
x <- runif(30,0,2*pi)
d <- data.frame( x = x, y = sin(x) + rnorm(30,sd=.2) )
xyplot( y ~ x, data=d )
ladd(panel.plotFun1( sin, col='red' ) )
xyplot( y ~ x | rbinom(30,1,.5), data=d )
ladd(panel.plotFun1( sin, col='red', lty=2 ) ) # plots sin(x) in each panel
```

parse.formula

Parse formulas

Description

utilities for extracting portions of formulas.

Usage

```
parse.formula(formula, ...)
```

```
rhs(x, ...)
```

```
lhs(x, ...)
```

```
condition(x, ...)
```

```
operator(x, ...)
```

```
## S3 method for class 'formula'
rhs(x, ...)
```

```
## S3 method for class 'formula'
lhs(x, ...)
```

```

## S3 method for class 'formula'
condition(x, ...)

## S3 method for class 'formula'
operator(x, ...)

## S3 method for class 'parsedFormula'
rhs(x, ...)

## S3 method for class 'parsedFormula'
lhs(x, ...)

## S3 method for class 'parsedFormula'
operator(x, ...)

## S3 method for class 'parsedFormula'
condition(x, ...)

```

Arguments

formula, a formula
 ... additional arguments, current ignored
 x, an object (currently a formula or parsedFormula)

Details

currently this is primarily concerned with extracting the operator, left hand side, right hand side (minus any condition) and the condition. Improvements/extensions may come in the future.

Value

an object of class `parsedFormula` from which information is easy to extract

perctable	<i>Cross tabulation displayed as percents or proportions</i>
-----------	--

Description

`perctable` and `proptable` use the cross-classifying factors to build a contingency table of the percents or proportions at each combination of factor levels.

Usage

```

perctable(...)

proptable(...)

```

Arguments

... arguments passed directly to `table`; typically one or more objects which can be interpreted as factors (including character strings), or a list (or data frame) whose components can be so interpreted.

Details

See `table`.

Value

a contingency table, an object of class "table", an array of percentage or proportion values. Note that unlike S the result is always an array, a 1D array if one factor is given.

Examples

```
perctable(rbinom(1000,10,.5))
with(airquality,
     perctable(OzHi=Ozone > 80, Month, useNA="ifany"))
with(airquality,
     perctable(OzHi=Ozone > 80, Month, useNA="always"))
```

plotCumfreq

Cumulative frequency plots

Description

A high-level function for producing a cumulative frequency plot using `lattice` graphics.

Usage

```
plotCumfreq(x, data, ...)

## S3 method for class 'formula'
plotCumfreq(x, data = NULL, subscripts, ...)

## Default S3 method:
plotCumfreq(x, ...)

prepanel.cumfreq(x, ...)

panel.cumfreq(x, type = c("smooth", "step"), groups = NULL, ...)
```

Arguments

x	a formula or numeric vector
data	a data frame in which x is evaluated if x is a formula.
...	other lattice arguments
subscripts	as in lattice plots
type	smooth or step-function?
groups	grouping variable

See Also

[histogram](#), [densityplot](#)

Examples

```
plotCumfreq(~eruptions, faithful, xlab = 'duration of eruptions')
```

plotDist

Plots of Discrete and Continuous Distributions

Description

Provides a simple way to generate plots of pdfs, probability mass functions, cdfs, probability histograms, and normal-quantile plots for distributions known to R.

Usage

```
plotDist(dist, params = list(), kind = c("density", "cdf", "qq",
    "histogram"), xlab = "", ylab = "", breaks = NULL, type,
    resolution = 5000, ...)
```

Arguments

dist	A string identifying the distribution. This should work with any distribution that has associated functions beginning with 'd', 'p', and 'q' (e.g. dnorm , pnorm , and qnorm). dist should match the name of the distribution with the initial 'd', 'p', or 'q' removed.
params	a list containing parameters for the distribution
kind	one of "density", "cdf", "qq", or "histogram" (or prefix of any of these)
xlab, ylab	as per other lattice functions
breaks	a vector of break points for bins of histograms, as in histogram
type	passed along to various lattice graphing functions
resolution	number of points to sample when generating the plots
...	other arguments passed along to lattice graphing routines

Details

plotDist determines whether the distribution is continuous or discrete by seeing if all the sampled quantiles are unique. A discrete random variable with many possible values could fool this algorithm and be considered continuous.

The plots are done referencing a data frame with variables x and y giving points on the graph of the pdf or pmf for the distribution. This can be useful in conjunction with the groups argument. See the examples.

Examples

```
plotDist('norm')
plotDist('norm', type='h')
plotDist('norm', kind='cdf')
plotDist('norm', params=list(mean=100, sd=10), kind='cdf')
plotDist('exp', kind='histogram')
plotDist('binom', params=list( 25, .25))
plotDist('binom', params=list( 25, .25), xlim=c(-1,26) )
plotDist('binom', params=list( 25, .25), kind='cdf')
plotDist('beta', params=list( 3, 10), kind='density')
plotDist('beta', params=list( 3, 10), kind='cdf')
plotDist( "binom", params=list(35,.25),
          groups= y < dbinom(qbinom(0.05, 35, .25), 35,.25) )
plotDist( "binom", params=list(35,.25),
          groups= y < dbinom(qbinom(0.05, 35, .25), 35,.25),
          kind='hist')
```

plotFun

Plotting mathematical expressions

Description

Plots mathematical expressions in one and two variables.

Usage

```
plotFun(object, ..., add = FALSE, xlim = NULL, ylim = NULL, npts = NULL,
        ylab = NULL, xlab = NULL, zlab = NULL, filled = TRUE, levels = NULL,
        nlevels = 10, labels = TRUE, surface = FALSE, groups = NULL,
        col = trellis.par.get("superpose.line")$col, col.regions = topo.colors,
        type = "l", alpha = NULL)
```

Arguments

object	a mathematical expression or a function "of one variable" which will converted to something intuitively equivalent to object(x) ~ x. (See examples)
add	if TRUE, then overlay an existing plot
xlim	limits for x axis (or use variable names, see examples)

<code>ylim</code>	limits for y axis (or use variable names, see examples)
<code>npts</code>	number of points for plotting.
<code>xlab</code>	label for x axis
<code>ylab</code>	label for y axis
<code>zlab</code>	label for z axis (when in surface-plot mode)
<code>col</code>	vector of colors for line graphs and contours
<code>filled</code>	fill with color between the contours (TRUE by default)
<code>levels</code>	levels at which to draw contours
<code>nlevels</code>	number of contours to draw (if <code>levels</code> not specified)
<code>labels</code>	if FALSE, don't label contours
<code>surface</code>	draw a surface plot rather than a contour plot
<code>col.regions</code>	a vector of colors or a function (<code>topo.colors</code> by default) for generating such
<code>type</code>	type of plot ("l" by default)
<code>alpha</code>	number from 0 (transparent) to 1 (opaque) for the fill colors
<code>groups</code>	grouping argument ala lattice graphics
<code>...</code>	additional parameters, typically processed by <code>lattice</code> functions such as <code>xyplot</code> , <code>levelplot</code> or their panel functions. Frequently used parameters include
	<code>main</code> main title for plot
	<code>sub</code> subtitle for plot
	<code>lwd</code> line width
	<code>lty</code> line type
	<code>col</code> a color

Additionally, these arguments can be used to specify parameters for the function being plotted and to specify the plotting window with natural names. See the examples for such usage.

Details

makes plots of mathematical expressions using the formula syntax. Will draw both line plots and contour/surface plots (for functions of two variables). In RStudio, the surface plot comes with sliders to set orientation. If the colors in filled surface plots are too blocky, increase `npts` beyond the default of 50, though `npts=300` is as much as you're likely to ever need. See examples for overplotting a constraint function on an objective function.

Value

a trellis object

Examples

```

plotFun( a*sin(x^2)~x, xlim=range(-5,5), a=2 ) # setting parameter value
plotFun( u^2 ~ u, ulim=c(-4,4) ) # limits in terms of u
# Note roles of ylim and y.lim in this example
plotFun( y^2 ~ y, ylim=c(-2,20), y.lim=c(-4,4) )
# Combining plot elements to show the solution to an inequality
plotFun( x^2 -3 ~ x, xlim=c(-4,4), grid=TRUE )
ladd( panel.abline(h=0,v=0,col='gray50') )
plotFun( (x^2 -3) * (x^2 > 3) ~ x, type='h', alpha=.1, lwd=4, col='lightblue', add=TRUE )
plotFun( sin(x) ~ x,
  groups=cut(x, findZeros(sin(x) ~ x, within=10)$x),
  col=c('blue','green'), lty=2, lwd=3, xlim=c(-10,10) )
## plotFun( sin(2*pi*x/P)*exp(-k*t)~x+t, k=2, P=.3)
f <- rfun( ~ u & v )
plotFun( f(u=v,v=v) ~ u & v, u.lim=range(-3,3), v.lim=range(-3,3) )
plotFun( u^2 + v < 3 ~ u & v, add=TRUE, npts=200 )
# display a linear model using a formula interface
model <- lm(wage ~ poly(exper,degree=2), data=CPS85)
fit <- makeFun(model)
xyplot(wage ~ exper, data=CPS85)
plotFun(fit(exper) ~ exper, add=TRUE, lwd=8)
# Can also just give fit since it is a "function of one variable"
plotFun(fit, add=TRUE, lwd=2, col='white')
# Attempts to find sensible axis limits by default
plotFun( sin(k*x)~x, k=0.01 )

```

plotPoints

Scatter plot of points

Description

Make or add a scatter plot in a manner coordinated with plotFun.

Usage

```

plotPoints(x, data = parent.frame(), add = FALSE, panelfun = panel.xyplot,
  plotfun = xyplot, ...)

```

Arguments

x	A formula specifying $y \sim x$ or $z \sim x \& y$
data	Data frame containing the variables to be plotted. If not specified, the variables will be looked up in the local environment
add	If TRUE, overlay the scatter plot on the current plot.
panelfun	Lattice panel function to be used for adding. Set only if you want something other than a scatter plot. Mainly, this is intended to add new functionality through other functions.

plotfun	Lattice function to be used for initial plot creation. Set only if you want something other than a scatter plot. Mainly, this is intended to add new functionality through other functions.
...	additional arguments

Value

A lattice graphics object (if add=FALSE)

See Also

[plotFun](#)

Examples

```
plotPoints( width ~ length, data=KidsFeet, groups=sex, pch=20)
f <- makeFun( lm( width ~ poly(length,2) * sex, data=KidsFeet))
plotFun( f(length=length,sex="G")~length, add=TRUE, col="pink")
plotFun( f(length=length,sex="B")~length, add=TRUE)
```

project

Projections

Description

Compute projections onto the span of a vector or a model space.

Usage

```
project(x, u, data = parent.env(), ...)

## S4 method for signature 'formula'
project(x, u = NULL, data = parent.frame(), ...)

## S4 method for signature 'numeric'
project(x, u = rep(1, length(x)), type = c("vector",
      "length"), ...)

## S4 method for signature 'matrix'
project(x, u, data = parent.frame())
```

Arguments

x	a vector or formula. Left-hand sides of formulas should be a single quantity
u	a vector
data	a data frame.
type	one of length or vector determining the type of the returned value
...	additional arguments

Details

project (preferably pronounced "pro-JECT" as in "projection") does either of two related things: (1) Given two vectors as arguments, it will project the first onto the second, returning the point in the subspace of the second that is as close as possible to the first vector. (2) Given a formula as an argument, will work very much like `lm()`, constructing a model matrix from the right-hand side of the formula and projecting the vector on the left-hand side onto the subspace of that model matrix. In (2), rather than returning the projected vector, `project()` returns the coefficients on each of the vectors in the model matrix. UNLIKE `lm()`, the intercept vector is NOT included by default. If you want an intercept vector, include `+1` in your formula.

Value

project returns the projection of `x` onto `u` (or its length if `u` and `v` are numeric vectors and `type == "length"`)

Examples

```
a <- c(1,0,0); b <- c(1,2,3); c <- c(4,5,6); x <- rnorm(3)
# projection onto the 1 vector gives the mean vector
mean(x)
project(x, 1)
# return the length of the vector, rather than the vector itself
project(x, 1, type='length')
project(a~b)
project(width~length+sex, data=KidsFeet)
project(log(width) ~ I(length^2)+sin(length)+sex, data=KidsFeet)
```

prop

Compute proportions, percents, or counts for a single level

Description

Compute proportions, percents, or counts for a single level

Usage

```
prop(x, data = parent.frame(), ..., level = NULL, long.names = TRUE,
     sep = ".", format = "proportion")

count(x, data = parent.frame(), ..., format = "count")

perc(x, data = parent.frame(), ..., format = "percent")
```

Arguments

x	an R object, usually a formula
data	a data frame in which x is to be evaluated
...	arguments passed through to tally
level	the level for which counts, proportions or percents are calculated
long.names	a logical indicating whether long names should be when there is a conditioning variable
sep	a character used to separate portions of long names
format	one of proportion, percent, or count, possibly abbreviated

Examples

```
prop( ~sex, data=HELPrct)
prop( ~sex, data=HELPrct, level='male')
count( ~sex | substance, data=HELPrct)
prop( ~sex | substance, data=HELPrct)
perc( ~sex | substance, data=HELPrct)
```

prop.test

Exact and Approximate Tests for Proportions

Description

The mosaic `prop.test` provides wrapper functions around the function of the same name in **stats**. These wrappers provide an extended interface (including formulas). `prop.test` performs an approximate test of a simple null hypothesis about the probability of success in a Bernoulli or multinomial experiment from summarized data or from raw data.

Usage

```
prop.test(x, n, p = NULL, alternative = c("two.sided", "less", "greater"),
  conf.level = 0.95, ...)

## S4 method for signature 'ANY'
prop.test(x, n, p = NULL, alternative = c("two.sided",
  "less", "greater"), conf.level = 0.95, ...)

## S4 method for signature 'formula'
prop.test(x, n, p = NULL, alternative = c("two.sided",
  "less", "greater"), conf.level = 0.95, success = NULL, data.name, data,
  groups = NULL, ...)

## S4 method for signature 'numeric'
prop.test(x, n, p = NULL, alternative = c("two.sided",
  "less", "greater"), conf.level = 0.95, success = NULL, data.name, ...)
```

```
## S4 method for signature 'character'
prop.test(x, n, p = NULL, alternative = c("two.sided",
    "less", "greater"), conf.level = 0.95, success = NULL, data.name, ...)

## S4 method for signature 'logical'
prop.test(x, n, p = NULL, alternative = c("two.sided",
    "less", "greater"), conf.level = 0.95, success = NULL, data.name, ...)

## S4 method for signature 'factor'
prop.test(x, n, p = NULL, alternative = c("two.sided",
    "less", "greater"), conf.level = 0.95, success = NULL, data.name, ...)
```

Arguments

x	count of successes, length 2 vector of success and failure counts, a formula, or a character, numeric, or factor vector containing raw data.
groups	when x is a formula, groups can be used to compare groups. (This can also be done using by placing both variables into the formula.) See the examples.
n	sample size (successes + failures) or a data frame (for the formula interface)
p	a vector of probabilities of success. The length of p must be the same as the number of groups specified by x, and its elements must be greater than 0 and less than 1.
alternative	character string specifying the alternative hypothesis, must be one of "two.sided" (default), "greater" or "less". You can specify just the initial letter. Only used for testing the null that a single proportion equals a given value, or that two proportions are equal; ignored otherwise.
conf.level	confidence level of the returned confidence interval. Must be a single number between 0 and 1. Only used when testing the null that a single proportion equals a given value, or that two proportions are equal; ignored otherwise.
success	level of variable to be considered success. All other levels are considered failure.
data.name	name for data. If missing, this is inferred from variable names.
data	a data frame (if missing, n may be a data frame)
...	additional arguments (often ignored)

Details

conf.level = 0.95, ...)

This is a wrapper around [prop.test](#) to simplify its use when the raw data are available, in which case an extended syntax for `prop.test` is provided.

Value

an `htest` object

See Also

[binom.test](#), [prop.test](#)

Examples

```
# Several ways to get a confidence interval for the proportion of Old Faithful
# eruptions lasting more than 3 minutes.
prop.test( faithful$eruptions > 3 )
prop.test(97,272)
prop.test(c(97,272-97))
faithful$long <- faithful$eruptions > 3
prop.test( faithful$long )
prop.test( ~long , faithful )
prop.test( homeless ~ sex, data=HELPrct )
prop.test( ~ homeless | sex, data=HELPrct )
prop.test( ~ homeless, groups= sex, data=HELPrct )
```

r.squared

Extract r-squared value

Description

Attempts to extract an r-squared value from a model or model-like object.

Usage

```
r.squared(x, ...)
```

Arguments

x	an object
...	additional arguments

RailTrail

Volume of Users of a Rail Trail

Description

The Pioneer Valley Planning Commission (PVPC) collected data north of Chestnut Street in Florence, MA for ninety days from April 5, 2005 to November 15, 2005. Data collectors set up a laser sensor, with breaks in the laser beam recording when a rail-trail user passed the data collection station.

Usage

```
data(RailTrail)
```

Format

A data frame with 90 observations on the following variables.

- hightemp daily high temperature (in degrees Fahrenheit)
- lowtemp daily low temperature (in degrees Fahrenheit)
- avgtemp average of daily low and daily high temperature (in degrees Fahrenheit)
- spring indicator of whether the season was Spring
- summer indicator of whether the season was Summer
- fall indicator of whether the season was Fall
- cloudcover measure of cloud cover (in oktas)
- precip measure of precipitation (in inches)
- volume estimated number of trail users that day (number of breaks recorded)
- weekday indicator of whether the day was a non-holiday weekday

Details

There is a potential for error when two users trigger the infrared beam at exactly the same time since the counter would only logs one of the crossings. The collectors left the motion detector out during the winter, but because the counter drops data when the temperature falls below 14 degrees Fahrenheit, there is no data for the cold winter months.

Source

Pioneer Valley Planning Commission

References

<http://www.fvgreenway.org/pdfs/Northampton-Bikepath-Volume-Counts>

Examples

```
data(RailTrail)
```

rand

Random Regressors

Description

A utility function for producing random regressors with a specified number of degrees of freedom.

Usage

```
rand(df = 1, rdist = rnorm, args = list(), nrow, seed = NULL)
```

Arguments

df	degrees of freedom, i.e., number of random regressors
rdist	random distribution function for sampling
args	arguments for rdist
nrow	number of rows in resulting matrix. This can often be omitted in the context of functions like <code>lm</code> where it is inferred from the data frame, if one is provided.
seed	seed for random number generation

Value

A matrix of random variates with `df` columns. In its intended use, the number of rows will be selected to match the size of the data frame supplied to `lm`

Examples

```
rand(2,nrow=4)
rand(2,rdist=rpois, args=list(lambda=3), nrow=4)
summary(lm( waiting ~ eruptions + rand(1), faithful))
```

read.file	<i>Read data files</i>
-----------	------------------------

Description

A wrapper around [read.table](#), [read.csv](#), and [load](#) to unify and simplify reading data from files.

Usage

```
read.file(file, header = T, na.strings = c("NA", "", ".", "na", "-"),
  comment.char = "#", ...)
```

Arguments

file	character: The name of the file which the data are to be read from. This may also be a complete URL or a path to a compressed file. If it does not contain an absolute path, the file name is relative to the current working directory, <code>getwd()</code> . Tilde-expansion is performed where supported. See read.table for more details.
header	logical; For <code>.txt</code> and <code>.csv</code> files, this indicates whether the first line of the file includes variables names.
na.strings	character: strings that indicate missing data.
comment.char	character: a character vector of length one containing a single character or an empty string. Use "" to turn off the interpretation of comments altogether.
...	additional arguments passed on to read.table , read.csv , or load .

Details

`read.file` uses the file extension to determine how to read data from the file. If file ends in `.Rdata`, then `load` is used to load the file. If file ends in `.csv`, then `read.csv` is used. Otherwise, `read.table` is used.

Value

A data frame, unless file ends in `.Rdata`, in which case arbitrary objects may be loaded and a character vector holding the names of the loaded objects is returned invisibly.

See Also

`read.table`, `read.csv`, `load`.

repeater-class

Repeater objects

Description

Repeater objects can be used with the `*` operator to repeat things multiple time using a different syntax and different output format from that used by, for example, `replicate`.

Slots

`n`: Object of class "numeric" indicating how many times to repeat something.

`cull`: Object of class "function" that culls the output from each repetition.

`mode`: Object of class "character" indicating the output mode ('default', 'data.frame', 'matrix', 'vector', or 'list'). For most purposes 'default' (the default) should suffice.

`algorithm`: an algorithm number.

`parallel`: a logical indicating whether to attempt parallel execution.

See Also

`do`

resample

*More Random Samples***Description**

These functions simplify and unify sampling in various ways.

Usage

```
resample(..., replace = TRUE)

deal(...)

shuffle(x, replace = FALSE, prob = NULL, groups = NULL,
        orig.ids = FALSE)

sample(x, size, replace = FALSE, ...)

## Default S3 method:
sample(x, size, replace = FALSE, prob = NULL,
       groups = NULL, orig.ids = FALSE, ...)

## S3 method for class 'data.frame'
sample(x, size, replace = FALSE, prob = NULL,
       groups = NULL, orig.ids = TRUE, fixed = names(x), shuffled = c(),
       invisibly.return = NULL, ...)

## S3 method for class 'matrix'
sample(x, size, replace = FALSE, prob = NULL,
       groups = NULL, orig.ids = FALSE, ...)

## S3 method for class 'factor'
sample(x, size, replace = FALSE, prob = NULL,
       groups = NULL, orig.ids = FALSE, drop.unused.levels = FALSE, ...)
```

Arguments

x	Either a vector of one or more elements from which to choose, or a positive integer.
size	a non-negative integer giving the number of items to choose.
replace	Should sampling be with replacement?
prob	A vector of probability weights for obtaining the elements of the vector being sampled.
groups	a vector (or variable in a data frame) specifying groups to sample within. This will be recycled if necessary.

orig.ids	a logical; should original ids be included in returned data frame?
...	additional arguments passed to <code>sample</code> or <code>sample</code> .
shuffled	a vector of column names. these variables are reshuffled individually (within groups if groups is specified), breaking associations among these columns. examples.
fixed	a vector of column names. These variables are shuffled en masse, preserving associations among these columns.
invisibly.return	a logical, should return be invisible?
drop.unused.levels	a logical, should unused levels be dropped?

Details

These functions are wrappers around `sample` providing different defaults and natural names.

Examples

```
# 100 Bernoulli trials -- no need for replace=TRUE
resample(0:1, 100)
tally(resample(0:1, 100))
Small <- sample(KidsFeet, 10)
resample(Small)
tally(~ sex, data=resample(Small))
tally(~ sex, data=resample(Small))
# fixed marginals for sex
tally(~ sex, data=Small)
tally(~ sex, data=resample(Small, groups=sex))
# shuffled can be used to reshuffle some variables within groups
# orig.ids shows where the values were in original data frame.
Small <- transform(Small,
  id1 = paste(sex, 1:10, sep=":"),
  id2 = paste(sex, 1:10, sep=":"))
resample(Small, groups=sex, shuffled=c("id1", "id2"))
deal(Cards, 13) # A Bridge hand
shuffle(Cards)
```

rflip

Tossing Coins

Description

These functions simplify simulating coin tosses for those (students primarily) who are not yet familiar with the binomial distributions or just like this syntax and verbosity better.

Usage

```
rflip(n = 1, prob = 0.5, quiet = FALSE, verbose = !quiet)

## S3 method for class 'cointoss'
print(x, ...)

nflip(n = 1, prob = 0.5, ...)
```

Arguments

n	the number of coins to toss
prob	probability of heads on each toss
quiet	a logical. If TRUE, less verbose output is used.
verbose	a logical. If TRUE, more verbose output is used.
x	an object
...	additional arguments

Value

for rflip, a cointoss object
 for nflip, a numeric vector

Examples

```
rflip(10)
rflip(10, prob=1/6, quiet=TRUE)
do(5) * rflip(10)
as.numeric(rflip(10))
nflip(10)
```

 rflin

Generate a natural-looking function

Description

Produce a random function that is the sum of Gaussian random variables
 rpoly2 generates a random 2nd degree polynomial (as a function)

Usage

```
rflin(vars = ~x & y, seed = NULL, n = 0)

rpoly2(vars = ~x & y, seed = NULL)
```

Arguments

vars	a formula; the LHS is empty and the RHS indicates the variables used for input to the function (separated by &)
seed	seed for random number generator, passed to <code>set.seed</code> .
n	the number of Gaussians. By default, this will be selected randomly.

Details

rfun is an easy way to generate a natural-looking but random function with ups and downs much as you might draw on paper. In two variables, it provides a good way to produce a random landscape that is smooth. Things happen in the domain -5 to 5. The function is pretty flat outside of that. Use seed to create a fixed function that will be the same for everybody

These functions are particularly useful for teaching calculus.

Value

a function with the appropriate number of inputs

a function defined by a 2nd degree polynomial with coefficients selected randomly according to a Unif(-1,1) distribution.

Examples

```
f <- rfun( ~ u & v)
plotFun(f(u,v)~u&v,u=range(-5,5),v=range(-5,5))
myfun <- rfun(~ u & v, seed=1959)
g <- rpoly2( ~ x&y&z, seed=1964)
plotFun(g(x,y,z=2)~x&y,xlim=range(-5,5),ylim=range(-5,5))
```

rgeo

Sample longitude and latitude on a sphere

Description

Randomly samples longitude and latitude on earth so that equal areas are (approximately) equally likely to be sampled. (Approximation assumes earth as a perfect sphere.)

Usage

```
rgeo(n = 1, latlim = c(-90, 90), lonlim = c(-180, 180), verbose = FALSE)
```

```
rgeo2(n = 1, latlim = c(-90, 90), lonlim = c(-180, 180),
      verbose = FALSE)
```

Arguments

n	number of random locations
latlim, lonlim	range of latitudes and longitudes to sample within, only implemented for rgeo.
verbose	return verbose output that includes Euclidean coordinates on unit sphere as well as longitude and latitude.

Details

rgeo and rgeo2 differ in the algorithms used to generate random positions. Each assumes a spherical globe. rgeo uses that fact that each of the x, y and z coordinates is uniformly distributed (but not independent of each other). Furthermore, the angle about the z-axis is uniformly distributed and independent of z. This provides a straightforward way to generate Euclidean coordinates using `runif`. These are then translated into latitude and longitude.

rgeo2 samples points in a cube by independently sampling each coordinate. It then discards any point outside the sphere contained in the cube and projects the non-discarded points to the sphere. This method must oversample to allow for the discarded points.

Value

a data frame with variables `long` and `lat`. If `verbose` is `TRUE`, then `x`, `y`, and `z` coordinates are also included in the data frame.

See Also

[deg2rad](#), [googleMap](#) and [latlon2xyz](#).

Examples

```
rgeo(4)
# sample from a region that contains the continental US
rgeo( 4, latlim=c(25,50), lonlim=c(-65,-125) )
rgeo2(4)
```

rkintegrate

A simple Runge-Kutte integrator

Description

Integrates ordinary differential equations using a Runge-Kutta method

Usage

```
rkintegrate(fun, x0, tstart = 0, tend = 1, dt = NULL)
```

Arguments

fun	the dynamical function with arguments state (a vector) and t.
x0	the initial condition, a vector with one element for each state variable
tstart	starting time
tend	ending time for integration
dt	step size for integration

Details

This is mainly for internal use by integrateODE.

Value

a list containing x , a matrix of the state with one row for each time step and a vector t containing the times of those steps.

Author(s)

Daniel Kaplan (<kaplan@macalester.edu>)

rsquared

Extract r-squared value

Description

Attempts to extract an r-squared value from a model or model-like object.

Usage

```
rsquared(x, ...)
```

Arguments

x	an object
...	additional arguments

SAT

State by State SAT data

Description

SAT data assembled for a statistics education journal article on the link between SAT scores and measures of educational expenditures

Usage

```
data(SAT)
```

Format

A data frame with 50 observations on the following variables.

- `state` a factor with names of each state
- `expend` expenditure per pupil in average daily attendance in public elementary and secondary schools, 1994-95 (in thousands of US dollars)
- `ratio` average pupil/teacher ratio in public elementary and secondary schools, Fall 1994
- `salary` estimated average annual salary of teachers in public elementary and secondary schools, 1994-95 (in thousands of US dollars)
- `frac` percentage of all eligible students taking the SAT, 1994-95
- `verbal` average verbal SAT score, 1994-95
- `math` average math SAT score, 1994-95
- `sat` average total SAT score, 1994-95

Source

<http://www.amstat.org/publications/jse/secure/v7n2/datasets.guber.cfm>

References

Deborah Lynn Guber, "Getting what you pay for: the debate over equity in public school expenditures" (1999), *Journal of Statistics Education* 7(2).

Examples

```
data(SAT)
xyplot(sat ~ expend, SAT)
xyplot(sat ~ expend, SAT,
panel=function(x,y){grid.text(abbreviate(SAT$state, 3), x, y, default.units='native')})
```

SnowGR

Snowfall data for Grand Rapids, MI

Description

Official snowfall data by month and season for Grand Rapids, MI, going back to 1893.

Usage

```
data(SnowGR)
```

Format

A data frame with 119 observations of the following variables.

- SeasonStart Year in which season started (July is start of season)
- SeasonEnd Year in which season ended (June is end of season)
- Jul Inches of snow in July
- Aug Inches of snow in August
- Sep Inches of snow in September
- Oct Inches of snow in October
- Nov Inches of snow in November
- Dec Inches of snow in December
- Jan Inches of snow in January
- Feb Inches of snow in February
- Mar Inches of snow in March
- Apr Inches of snow in April
- May Inches of snow in May
- Jun Inches of snow in June
- Total Inches of snow for entire season (July-June)

Source

These data were compiled by Laura Kapitula from data available at <http://www.crh.noaa.gov/grr/climate/data/grr/snowfall/>.

Examples

```

data(SnowGR)
histogram(~Total, data=SnowGR)
favstats(SnowGR$Total)
xyplot(Total ~ SeasonStart, SnowGR, type=c('p','smooth'))
if (require(reshape2)) {
  Snow2 <- melt(SnowGR, id=1:2)
  names(Snow2)[3:4] <- c('Time','Snow')
  bwplot(Snow ~ Time, Snow2)
}

```

sp2df

*Map Utilities***Description**

Some utilities for working with map data

Usage

```
sp2df(map, ...)
```

Arguments

map	a map object of class SpatialPolygonsDataFrame
...	other arguments, currently ignored

Examples

```

## Not run:
if(require(maptools)) {
  data(wrld_simpl)
  worldmap <- sp2df(wrld_simpl)
}

if ( require(ggplot2) && require(maptools) && require(ply) ) {
  data(wrld_simpl)
  World <- sp2df(wrld_simpl)
  World2 <- merge(World, Countries, by.x="NAME", by.y="maptools", all.y=FALSE)
  Mdata <- merge(Alcohol, World2, by.x="country", by.y="gapminder", all.y=FALSE)
  Mdata <- Mdata[order(Mdata$order),]
  qplot( x=long, y=lat, fill=ntiles(alcohol,5),
        data=subset(Mdata, year==2008), group = group,
        geom="polygon")
}

## End(Not run)

```

statTally	<i>Tally test statistics</i>
-----------	------------------------------

Description

Tally test statistics from data and from multiple draws from a simulated null distribution

Usage

```
statTally(sample, rdata, FUN, direction = NULL, alternative = c("default",
  "two.sided", "less", "greater"), sig.level = 0.1, center = NULL,
  stemplot = dim(rdata)[direction] < 201, q = c(0.5, 0.9, 0.95, 0.99),
  fun = function(x) x, xlim, ...)
```

Arguments

sample	sample data
rdata	a matrix of randomly generated data under null hypothesis.
FUN	a function that computes the test statistic from a data set. The default value does nothing, making it easy to use this to tabulate precomputed statistics into a null distribution. See the examples.
direction	1 or 2 indicating whether samples in rdata are in rows (1) or columns (2).
stemplot	indicates whether a stem plot should be displayed
q	quantiles of sampling distribution to display
fun	same as FUN so you don't have to remember if it should be capitalized
xlim	limits for the horizontal axis of the plot.
center	center of null distribution
alternative	one of default, two.sided, less, or greater
sig.level	significance threshold for wilcox.test used to detect lack of symmetry
...	additional arguments passed to histogram

Value

A lattice plot showing the sampling distribution.

As side effects, information about the empirical sampling distribution and (optionally) a stem plot are printed to the screen.

Examples

```
# is my spinner fair?
x <- c(10, 18, 9, 15) # counts in four cells
rdata <- rmultinom(1000, sum(x), prob=rep(.25, 4))
statTally(x, rdata, fun=max) # unusual test statistic
statTally(x, rdata, fun=var) # equivalent to chi-squared test
# Can also be used with test stats that are precomputed.
D <- diff(mean( age ~ sex, data=HELPrct)); D
nullDist <- do(1000) * diff( mean( age ~ shuffle(sex), data=HELPrct))
statTally( D, nullDist)
```

surround	<i>Format strings for pretty output</i>
----------	---

Description

Format strings for pretty output

Usage

```
surround(x, pre = " ", post = " ", width = 8, ...)
```

Arguments

x	a vector
pre	text to prepend onto string
post	text to postpend onto string
width	desired width of string
...	additional arguments passed to format

Value

a vector of strings padded to the desired width

Examples

```
surround(rbinom(10,20,.5), " ", " ", width=4)
surround(rnorm(10), " ", " ", width=8, digits = 2, nsmall = 2)
```

SwimRecords	<i>100 m Swimming World Records</i>
-------------	-------------------------------------

Description

World records for men and women over time from 1905 through 2004.

Usage

```
data(SwimRecords)
```

Format

A data frame with 62 observations of the following variables.

- time time (in seconds) of the world record
- year Year in which the record was set
- sex a factor with levels M and F

Examples

```
data(SwimRecords)
xyplot(time~year, data=SwimRecords, groups=sex)
```

symbolicD	<i>Symbolic Derivatives</i>
-----------	-----------------------------

Description

Constructs symbolic derivatives of some mathematical expressions

Usage

```
symbolicD(formula, ..., .order = NULL)
```

Arguments

formula	a mathematical expression (see examples and plotFun)
...	additional parameters, typically default values for mathematical parameters
.order	a number specifying the order of a derivative with respect to a single variable

Details

Uses the built-in symbolic differentiation function to construct a formula for the derivative and packages this up as a function. The .order argument is just for convenience when programming high-order derivatives, e.g. the 5th derivative w.r.t. one variable.

Value

a function implementing the derivative

Author(s)

Daniel Kaplan (<kaplan@macalester.edu>)

See Also

[D](#), [numD](#), [makeFun](#), [antiD](#), [plotFun](#)

Examples

```
symbolicD( a*x^2 ~ x)
symbolicD( a*x^2 ~ x&x)
symbolicD( a*sin(x)~x, .order=4)
symbolicD( a*x^2*y+b*y ~ x, a=10, b=100 )
```

symbolicInt

Find the symbolic integral of a formula

Description

Find the symbolic integral of a formula

Use recursion to find a symbolic antiderivative

Attempts symbolic integration of some mathematical/arithmetical forms

Attempts symbolic integration of some mathematical forms

Attempts symbolic integration of some mathematical forms using trigonometric substitution

Takes a call and returns its affine coefficients.

Usage

```
symbolicInt(form, ...)
```

```
symbolicAntiD(form, ...)
```

```
.intArith(form, ...)
```

```
.intMath(form, ...)
```

```
.intTrig(form, num, den, .x.)
```

```
.affine.exp(tree, .x.)
```

Arguments

form	an object of type formula to be integrated. Rhs of formula indicates which variable to integrate with respect to. Must only have one variable.
...	extra parameters
num	numerator
den	denominator
.x.	the variable name
tree	the expression to be analyzed

Details

This symbolic integrator recognizes simple polynomials and functions such as sin, cos, tan, sinh, cosh, tanh, sqrt, and exp.

It will not perform more complicated substitutions or integration by parts.

Value

symbolicInt returns a function whose body is the symbolic antiderivative of the formula. If this method does not recognize the formula, it will return an error.

a formula implementing giving symbolic anti-derivative. If the formula isn't found by the algorithm, an error is thrown.

An expression with the integral, or throws an error if unsuccessful.

An expression with the integral, or throws an error if unsuccessful.

An expression with the integral, or throws an error if unsuccessful.

A list with values of a and b satisfying $a \cdot x + b = \text{tree}$. If the expression is not affine, returns an empty list.

t.test

Student's t-Test

Description

Performs one and two sample t-tests. The mosaic t.test provides wrapper functions around the function of the same name in **stats**. These wrappers provide an extended interface that allows for a more systematic use of the formula interface.

rdname ttest

Usage

```
## S3 method for class 'test'  
t(x, ...)  
  
ttest(x, ...)  
  
## Default S3 method:  
ttest(x, ...)  
  
## S3 method for class 'formula'  
ttest(x, data = parent.frame(), groups = NULL, ...)
```

Arguments

x	an object (e.g., a formula or a numeric vector)
data	a data frame
groups	$x \sim \text{var}$, groups=g is equivalent to $x = \text{var} \sim g$.
...	additional arguments, see t.test in the stats package.

Details

This is a wrapper around [t.test](#) from the **stats** package to extend the functionality of the formula interface.

Value

an object of class htest

See Also

[prop.test](#), [t.test](#)

Examples

```
t.test( ~ age, data=HELPrct)  
t.test( age ~ sex, data=HELPrct)  
t.test( ~ age | sex, data=HELPrct)  
t.test( ~ age, groups=sex, data=HELPrct)
```

tally

Tabulate categorical data

Description

Tabulate categorical data

Usage

```
tally(x, data = parent.frame(), format = c("default", "count", "proportion",
    "percent"), margins = FALSE, quiet = TRUE, subset, ...)
```

Arguments

x	an object
data	a data frame or environment in which evaluation occurs
format	a character string describing the desired format of the results. One of 'default', 'count', 'proportion', or 'percent'. In case of 'default', counts are used unless there is a condition, in which case proportions are used instead.
subset	an expression evaluating to a logical vector used to select a subset of data
quiet	a logical indicating whether messages about order in which marginal distributions are calculated should be suppressed. See addmargins .
margins	a logical indicating whether marginal distributions should be displayed.
...	additional arguments passed to table

Examples

```
tally( ~ substance, data=HELPrct)
tally( ~ substance & sex , data=HELPrct)
tally( sex ~ substance, data=HELPrct) # equivalent to tally( ~ sex | substance, ... )
tally( ~ substance | sex , data=HELPrct)
tally( ~ substance | sex , data=HELPrct, format='count')
tally( ~ substance & sex , data=HELPrct, format='percent')
tally( ~ link, data=HELPrct, useNA="always")
```

TenMileRace

Cherry Blossom Race

Description

The Cherry Blossom 10 Mile Run is a road race held in Washington, D.C. in April each year. (The name comes from the famous cherry trees that are in bloom in April in Washington.) The results of this race are published. This data frame contains the results from the 2005 race.

Usage

```
data(TenMileRace)
```


Format

A data frame with 8636 observations on the following variables.

- state State of residence of runner.
- time Official time from starting gun to finish line.
- net The recorded time from when the runner crossed the starting line to when the runner crossed the finish line. This is generally less than the official time because of the large number of runners in the race: it takes time to reach the starting line after the gun has gone off.
- age Age of runner in years.
- sex A factor with levels F M.

Examples

```
data(TenMileRace)
xyplot(net ~ age, data=TenMileRace, groups=sex)
lm(net ~ age + sex, data=TenMileRace)
```

theme.mosaic

Lattice Theme

Description

A theme for use with lattice graphics.

Usage

```
theme.mosaic(bw = FALSE, lty = 1:7)
```

```
col.mosaic(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 `trellis.par.set()`.

Note

These two functions are identical. `col.mosaic` is named similarly to `col.whitebg`, but since more than just colors are set, `theme.mosaic` is a preferable name.

See Also

[trellis.par.set](#), [show.settings](#)

Examples

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

 TukeyHSD.lm

Additional interfaces to TukeyHSD

Description

TukeyHSD requires use of **aov**. Since this is a hinderence for beginners, wrappers have been provided to remove this need.

Usage

```
## S3 method for class 'lm'
TukeyHSD(x, which, ordered = FALSE, conf.level = 0.95, ...)

## S3 method for class 'formula'
TukeyHSD(x, which, ordered = FALSE, conf.level = 0.95,
  data = parent.frame(), ...)
```

Arguments

x an object, for example of class `lm` or `formula`

data a data frame. NB: This does not come second in the argument list.

which, ordered, conf.level, ...
just as in [TukeyHSD](#) from the base package

Examples

```
## These should all give the same results
model <- lm(age ~ substance, data=HELPrct)
TukeyHSD(model)
TukeyHSD( age ~ substance, data=HELPrct)
TukeyHSD(aov(age ~ substance, data=HELPrct))
```

Utilities

Utility bills

Description

Data from utility bills at a residence. [Utilities2](#) is a similar data set with some additional variables.

Usage

```
data(Utilities)
```

Format

A data frame containing 117 observations for the following variables.

- month month (coded as a number)
- day day of month on which bill was calculated
- year year of bill
- temp average temperature (F) for billing period
- kwh electricity usage (kwh)
- ccf gas usage (ccf)
- thermsPerDay a numeric vector
- billingDays number of billing days in billing period
- totalbill total bill (in dollars)
- gasbill gas bill (in dollars)
- elecbill electric bill (in dollars)
- notes notes about the billing period

Source

Daniel T. Kaplan, *Statistical modeling: A fresh approach*, 2009.

See Also

[Utilities2](#).

Examples

```
data(Utilities)
xyplot(gasbill ~ temp, Utilities)
```

Utilities2

Utility bills

Description

Data from utility bills at a private residence. This is an augmented version of [Utilities](#).

Usage

```
data(Utilities2)
```

Format

A data frame containing 117 observations for the following variables.

- month month (coded as a number)
- day day of month on which bill was calculated
- year year of bill
- temp average temperature (F) for billing period
- kwh electricity usage (kwh)
- ccf gas usage (ccf)
- thermsPerDay a numeric vector
- billingDays number of billing days in billing period
- totalbill total bill (in dollars)
- gasbill gas bill (in dollars)
- elecbill exectric bill (in dollars)
- notes notes about the billing period
- ccfpday average gas usage per day [Utilities2 only]
- kwphpday average electric usage per day [Utilities2 only]
- gasbillpday gas bill divided by billing days [Utilities2 only]
- elecbillpday electric bill divided by billing days a numeric vector [Utilities2 only]
- totalbillpday total bill divided by billing days a numeric vector [Utilities2 only]
- therms thermsPerDay * billingDays [Utilities2 only]
- monthsSinceY2K months since 2000 [Utilities2 only]

Source

Daniel T. Kaplan, *Statistical modeling: A fresh approach*, 2009.

See Also

[Utilities](#).

Examples

```
data(Utilities2)
xyplot(gasbillpday ~ temp, Utilities2)
```

Whickham

Data from the Whickham survey

Description

Data on age, smoking, and mortality from a one-in-six survey of the electoral roll in Whickham, a mixed urban and rural district near Newcastle upon Tyne, in the UK. The survey was conducted in 1972-1974 to study heart disease and thyroid disease. A follow-up on those in the survey was conducted twenty years later.

Usage

```
data(Whickham)
```

Format

A data frame with 1314 observations on women for the following variables.

- outcome survival status after 20 years: a factor with levels Alive Dead
- smoker smoking status at baseline: a factor with levels No Yes
- age age (in years) at the time of the first survey

Details

This dataset contains a subset of the survey sample: women who were classified as current smokers or as never having smoked. The data were synthesized from the summary description tables given in the Appleton et al al paper.

References

DR Appleton, JM French, MPJ Vanderpump. "Ignoring a covariate: an example of Simpson's paradox". (1996) *American Statistician*, 50(4):340-341.

Examples

```
data(Whickham)
```

xchisq.test	<i>Augmented Chi-squared test</i>
-------------	-----------------------------------

Description

This augmented version of [chisq.test](#) provides more verbose output.

Usage

```
xchisq.test(...)
```

Arguments

... Arguments passed directly to [chisq.test](#).

See Also

[chisq.test](#)

Examples

```
# Physicians' Health Study data
phs <- cbind(c(104,189),c(10933,10845))
rownames(phs) <- c("aspirin","placebo")
colnames(phs) <- c("heart attack","no heart attack")
phs
xchisq.test(phs)
```

xhistogram	<i>Augmented histograms</i>
------------	-----------------------------

Description

The **mosaic** package adds some additional functionality to [histogram](#), making it simpler to obtain certain common histogram adornments. This is done by resetting the default panel and prepanel functions used by histogram.

Usage

```
xhistogram(x, data = NULL, panel = panel.xhistogram, type = "density",
  center = NULL, width = NULL, ...)
```

```
xhistogramBreaks(x, center = NULL, width = NULL, nint, ...)
```

```
prepanel.xhistogram(x, breaks = xhistogramBreaks, ...)
```

```
panel.xhistogram(x, dcol = trellis.par.get("plot.line")$col, dlwd = 2,
  gcol = trellis.par.get("add.line")$col, glwd = 2,
  fcol = trellis.par.get("superpose.polygon")$col, dmath = dnorm,
  verbose = FALSE, dn = 100, args = NULL, labels = FALSE,
  density = FALSE, fit = NULL, start = NULL, type = "density", v, h,
  groups = NULL, center = NULL, width = NULL, breaks, nint = round(1.5 *
  log2(length(x)) + 1), stripes = c("vertical", "horizontal", "none"),
  alpha = 1, ...)
```

Arguments

x	a formula or a numeric vector
data	a data frame in which to evaluate x
panel	a panel function
type	one of 'density', 'count', or 'percent'
nint	approximate number of bins
breaks	break points for histogram bins, a function for computing such, or a method hist knows about given as a character string. When using the mosaic package defaults, xhistogramBreaks is used.
...	additional arguments passed to histogram and (by default when the mosaic package has been loaded) on to panel.xhistogram .
dcol	color of density curve
gcol	color of guidelines
fcol	fill color for histogram rectangles
dmath	density function for density curve overlay
verbose	be verbose?
dn	number of points to sample from density curve
dlwd, glwd	like lwd but affecting the density line and guide lines, respectively
args	a list of additional arguments for dmath
labels	should counts/densities/precents be displayed or each bin?
density	overlay density?
fit	a character string describing the distribution to fit. Known distributions include "exponential", "normal", "lognormal", "poisson", "beta", "geometric", "t", "weibull", "cauchy", "gamma", "chisq", and "chi-squared"
start	numeric value passed to fitdistr
center	center of one of the bins
width	width of the bins
groups	as per histogram
stripes	one of "vertical", "horizontal", or "none", indicating how bins should be striped when groups is not NULL
h, v	a vector of values for additional horizontal and vertical lines
alpha	transparency level

Value

xhistogramBreaks returns a vector of break points

Note

The use of xhistogram has been deprecated. Versions of **lattice** since 0.20-21 support setting custom defaults for breaks, panel, and prepanel used by histogram, so xhistogram is no longer needed. xhistogram will be retained temporarily to support users with older versions of **lattice**.

See Also

[histogram](#)

Examples

```

histogram(~age | substance, HELPrct, v=35, fit='normal')
histogram(~age, HELPrct, labels=TRUE, type='count')
histogram(~age, HELPrct, groups=cut(age, seq(10,80,by=10)))
histogram(~age, HELPrct, groups=sex, stripes='horizontal')
histogram(~racegrp, HELPrct, groups=substance,auto.key=TRUE)
xhistogramBreaks(1:10, center=5, width=1)
xhistogramBreaks(1:10, center=5, width=2)
xhistogramBreaks(0:10, center=15, width=3)
xhistogramBreaks(1:100, center=50, width=3)
xhistogramBreaks(0:10, center=5, nint=5)

```

xpnorm

Augmented versions of pnorm and qnorm

Description

These functions behave similarly to the functions with the initial x removed from their names but add more verbose output and graphics.

Usage

```

xpnorm(q, mean = 0, sd = 1, plot = TRUE, verbose = TRUE,
       invisible = FALSE, digits = 4, lower.tail = TRUE, log.p = FALSE,
       xlim = mean + c(-4, 4) * sd, ylim = c(0, 1.4 * dnorm(mean, mean, sd)),
       vlwd = 2, vcol = trellis.par.get("add.line")$col, rot = 45,
       manipulate = FALSE, ...)

xqnorm(p, mean = 0, sd = 1, plot = TRUE, verbose = TRUE, digits = 4,
       lower.tail = TRUE, log.p = FALSE, xlim, ylim, invisible = FALSE,
       vlwd = 2, vcol = trellis.par.get("add.line")$col, rot = 45, ...)

```


Arguments

p	probability
q	quantile
mean, sd	parameters of normal distribution.
plot	logical. If TRUE, show an illustrative plot.
verbose	logical. If TRUE, display verbose output.
invisible	logical. If TRUE, return value invisibly.
digits	number of digits to display in output.
lower.tail	logical. If FALSE, use upper tail probabilities.
log.p	logical. If TRUE, uses the log of probabilities.
xlim, ylim	limits for plotting.
vlwd, vcol	line width and color for vertical lines.
rot	angle of rotation for text labels.
manipulate	logical. If TRUE and in RStudio, then sliders are added for interactivity.
...	additional arguments.

See Also

[histogram](#), [chisq.test](#), [pnorm](#), [qnorm](#), [qqmath](#), and [plot](#).

Examples

```
xpnorm(650, 500, 100)
xqnorm(.75, 500, 100)
## Not run:
if (require(manipulate)) {
  manipulate( xpnorm(score, 500, 100, verbose=verbose),
             score = slider(200,800),
             verbose = checkbox(TRUE, label="Verbose Output")
           )
}
## End(Not run)
```

xqqmath

Augmented version of qqmath

Description

Augmented version of qqmath

Usage

```
xqqmath(x, data = NULL, panel = "panel.xqqmath", ...)

panel.xqqmath(x, qqmathline = !(fitline || idline), idline = FALSE,
  fitline = FALSE, slope = NULL, intercept = NULL, overlines = FALSE,
  groups = NULL, ..., col.line = trellis.par.get("add.line")$col,
  pch = 16, lwd = 2, lty = 2)
```

Arguments

`x, data, panel, xqqmath, ...`
as in [qqmath](#)

`qqmathline` a logical: should line be displayed passing through first and third quartiles?

`idline` a logical; should the line $y=x$ be added to the plot?

`fitline` a logical; should a fitted line be added to plot? Such a line will use slope and intercept if provided, else the standard deviation and mean of the data.

`slope` slope for added math line

`intercept` intercept for added math line

`overlines` a logical: should lines be on top of qq plot?

`groups, pch, lwd, lty`
as in lattice plots

`col.line` color to use for added lines

Value

a trellis object

Examples

```
xqqmath(rnorm(100))
```

xyz2latlon

Convert back and forth between latitude/longitude and XYZ-space

Description

Convert back and forth between latitude/longitude and XYZ-space

Usage

```
xyz2latlon(x, y, z)

latlon2xyz(latitude, longitude)
```

Arguments

x,y,z numeric vectors
latitude,longitude
 vectors of latitude and longitude values

Value

a matrix each row of which describes the latitudes and longitudes
a matrix each row of which contains the x, y, and z coordinates of a point on a unit sphere

See Also

[deg2rad](#), [googleMap](#), and [rgeo](#).

Examples

```
xyz2latlon(1, 1, 1)    # point may be on sphere of any radius  
xyz2latlon(0, 0, 0)    # this produces a NaN for latitude  
latlon2xyz(45, 45)
```

Index

- *Topic **calculus**
 - findZeros, 40
- *Topic **datasets**
 - Cards, 16
 - CoolingWater, 22
 - Countries, 23
 - Galton, 48
 - Gestation, 49
 - HeatX, 52
 - Heightweight, 53
 - HELP, 53
 - HELPfull, 55
 - HELPmiss, 82
 - HELPrct, 84
 - KidsFeet, 90
 - Marriage, 97
 - Mites, 100
 - SAT, 135
 - SnowGR, 136
 - SwimRecords, 140
 - TenMileRace, 144
 - Utilities, 147
 - Utilities2, 148
 - Whickham, 149
- *Topic **distribution**
 - .qdata_old, 6
 - rand, 126
- *Topic **geometry**
 - rgeo, 132
- *Topic **graphics**
 - dotPlot, 33
 - ladd, 91
 - plotCumfreq, 116
 - plotDist, 117
 - theme.mosaic, 145
- *Topic **inference**
 - CIsim, 17
 - interval, 87
 - statTally, 138
- *Topic **iteration**
 - compareMean, 19
 - compareProportion, 20
 - do, 32
- *Topic **manipulate**
 - as.xtabs, 11
 - cross, 25
 - perctable, 115
- *Topic **map**
 - rgeo, 132
- *Topic **package**
 - mosaic-package, 4
- *Topic **random**
 - rfun, 131
 - rgeo, 132
- *Topic **regression**
 - rand, 126
- *Topic **simulation**
 - CIsim, 17
- *Topic **stats**
 - binom.test, 13
 - compareMean, 19
 - compareProportion, 20
 - fav_stats, 37
 - interval, 87
 - oddsRatio, 109
 - plotDist, 117
 - prop.test, 123
- *Topic **util**
 - fetchData, 37
 - read.file, 127
- *, repeater, ANY-method (do), 32
- ... (prop.test), 123
- .affine.exp (symbolicInt), 141
- .cdata_old (.qdata_old), 6
- .intArith (symbolicInt), 141
- .intMath (symbolicInt), 141
- .intTrig (symbolicInt), 141
- .makePoly (.polyExp), 5

- .pdata_old (.qdata_old), 6
- .polyExp, 5
- .qdata_old, 6
- = (prop.test), 123
- 0.95, (prop.test), 123
- adapt_seq, 8
- addmargins, 144
- aggregatingFunction1, 9
- aggregatingFunction2, 9
- Alcohol, 10
- antiD, 109, 141
- antiD(D), 25
- aov, 146
- apply, 31
- as.xtabs, 11
- barchart, 12
- bargraph, 12
- binom.test, 13, 14, 125
- binom.test, ANY-method (binom.test), 13
- binom.test, character-method (binom.test), 13
- binom.test, factor-method (binom.test), 13
- binom.test, formula-method (binom.test), 13
- binom.test, logical-method (binom.test), 13
- binom.test, numeric-method (binom.test), 13
- Births78, 14
- Broyden, 15
- bs, 45
- card (Cards), 16
- Cards, 16
- cdata (.qdata_old), 6
- cdata_f (.qdata_old), 6
- cdata_v (.qdata_old), 6
- cdist, 16
- chisq.test, 110, 150, 153
- CIsim, 17
- coef (coef.function), 18
- coef.function, 18
- coef.groupwiseModel (mm), 101
- coef.nlsfunction (fitModel), 43
- col.mosaic (theme.mosaic), 145
- col.whitebg, 145
- columns, 19
- compareMean, 19, 20
- compareProportion, 20, 20
- condition (parse.formula), 114
- conf.level (prop.test), 123
- confint.data.frame (confint.numeric), 21
- confint.do.data.frame (confint.numeric), 21
- confint.groupwiseModel (mm), 101
- confint.htest (interval), 87
- confint.numeric, 21
- connector (FunctionsFromData), 47
- CoolingWater, 22
- cor (mean), 98
- count (prop), 122
- Countries, 23
- cov (mean), 98
- CPS85, 24
- cross, 25
- D, 25, 109, 141
- d2fdx2 (numD), 107
- d2fdxdy (numD), 107
- data, (prop.test), 123
- data.name, (prop.test), 123
- ddata (.qdata_old), 6
- deal (resample), 129
- deg2rad, 27, 51, 133, 155
- deltaMethod, 28, 28
- densityplot, 46, 117
- derivedFactor, 29
- dfapply, 31
- dfdx (numD), 107
- diff, 34
- dnorm, 117
- do, 20, 32, 102, 128
- dot (linear.algebra), 92
- dotPlot, 33
- ediff, 34
- evalFormula, 35
- evalSubFormula, 35
- expandFun, 36
- fav_stats, 37
- favstats, 37
- favstats (mean), 98
- fetchData, 37
- fetchGapminder (fetchGapminder1), 39

- fetchGapminder1, 39
- fetchGoogle, 39
- findZeros, 40
- findZerosMult, 42
- fitdistr, 151
- fitModel, 43
- fitSpline, 44
- fitted.groupwiseModel (mm), 101
- fivenum (mean), 98
- format, 139
- freqpolygon, 45
- FunctionsFromData, 47
- Galton, 48
- Gestation, 49
- getVarFormula, 50
- glm, 18
- googleMap, 28, 51, 133, 155
- groups=NULL, (prop.test), 123
- HeatX, 52
- Heightweight, 53
- HELP, 53
- HELPfull, 55, 83, 85
- HELPmiss, 81, 82, 85
- HELPprct, 55, 81, 83, 84
- hist, 151
- histogram, 33, 34, 46, 117, 138, 150–153
- histogram (xhistogram), 150
- ilogit (logit), 93
- inferArgs, 86
- integrateODE, 86
- interval, 87
- IQR (mean), 98
- iqr (mean), 98
- is.integer, 88, 89
- is.wholenumber, 88
- joinFrames, 89
- joinTwoFrames (joinFrames), 89
- KidsFeet, 90
- ladd, 91
- lapply, 31
- latlon2xyz, 28, 51, 133
- latlon2xyz (xyz2latlon), 154
- levelplot, 119
- lhs (parse.formula), 114
- linear.algebra, 92
- linearModel, 44, 92
- linearModel (FunctionsFromData), 47
- lm, 18, 47, 102
- load, 127, 128
- loess, 47
- logical2factor, 93
- logit, 93
- maggregate, 94
- makeAntiDfun (D), 25
- makeColorscheme, 95
- makeFun, 96, 109, 141
- makeFun, formula-method (makeFun), 96
- makeFun, glm-method (makeFun), 96
- makeFun, lm-method (makeFun), 96
- makeFun, nls-method (makeFun), 96
- Marriage, 97
- mat (linear.algebra), 92
- max (mean), 98
- mean, 98
- median (mean), 98
- merge, 105
- min (mean), 98
- Mites, 100
- mm, 101
- mMap (mPlot), 105
- model (fitModel), 43
- modelVars, 103
- mosaic (mosaic-package), 4
- mosaic-package, 4
- mosaic.getOption (mosaic.options), 103
- mosaic.options, 103
- mosaic.par.get (mosaic.options), 103
- mosaic.par.set (mosaic.options), 103
- mosaic_formula, 104
- mosaic_formula_q (mosaic_formula), 104
- mosaicGetOption (mosaic.options), 103
- mosaicLatticeOptions (mosaic.options), 103
- mPlot, 105
- mScatter (mPlot), 105
- mUniplot (mPlot), 105
- nflip (rflip), 130
- nls, 18, 44
- ns, 45
- ntiles, 106
- numD, 107, 141

- numerical.first.partial (numD), 107
- numerical.mixed.partial (numD), 107
- numerical.second.partial (numD), 107
- numerical_integration (D), 25
- oddsRatio, 109
- operator (parse.formula), 114
- panel.cumfreq (plotCumfreq), 116
- panel.dotPlot (dotPlot), 33
- panel.freqpolygon (freqpolygon), 45
- panel.levelcontourplot, 110
- panel.levelplot, 113, 114
- panel.lmbands, 111
- panel.plotFun, 112
- panel.plotFun1, 113
- panel.xhistogram, 151
- panel.xhistogram (xhistogram), 150
- panel.xqqmath (xqqmath), 153
- panel.xyplot, 113, 114
- parse.formula, 114
- pdata (.qdata_old), 6
- pdata_f (.qdata_old), 6
- pdata_v (.qdata_old), 6
- perc (prop), 122
- perctable, 115
- plot, 153
- plotCumfreq, 116
- plotDist, 117
- plotFun, 36, 107, 109, 118, 121, 140, 141
- plotPoints, 120
- pnorm, 117, 153
- prepanel.cumfreq (plotCumfreq), 116
- prepanel.xhistogram (xhistogram), 150
- print, repeater-method (do), 32
- print.cointoss (rflip), 130
- print.groupwiseModel (mm), 101
- print.oddsRatio (oddsRatio), 109
- print.summary_groupwiseModel (mm), 101
- prod (mean), 98
- project, 47, 92, 121
- project, formula, ANY-method (project), 121
- project, formula-method (project), 121
- project, matrix, ANY-method (project), 121
- project, matrix-method (project), 121
- project, numeric, ANY-method (project), 121
- project, numeric-method (project), 121
- prop, 122
- prop.test, 14, 123, 124, 125, 143
- prop.test, ANY-method (prop.test), 123
- prop.test, character-method (prop.test), 123
- prop.test, factor-method (prop.test), 123
- prop.test, formula-method (prop.test), 123
- prop.test, logical-method (prop.test), 123
- prop.test, numeric-method (prop.test), 123
- proptable (perctable), 115
- pval (interval), 87
- qdata (.qdata_old), 6
- qdata_f (.qdata_old), 6
- qdata_v (.qdata_old), 6
- qnorm, 117, 153
- qqmath, 153, 154
- quantile, 7
- r.squared, 125
- rad2deg (deg2rad), 27
- RailTrail, 125
- rand, 126
- range (mean), 98
- rdata (.qdata_old), 6
- read.csv, 127, 128
- read.file, 127
- read.table, 127, 128
- repeater-class, 128
- replicate, 33, 128
- resample, 129
- residuals.groupwiseModel (mm), 101
- restoreLatticeOptions (mosaic.options), 103
- rflip, 130
- rfun, 131
- rgeo, 28, 51, 132, 155
- rgeo2 (rgeo), 132
- rhs (parse.formula), 114
- rkintegrate, 133
- rows (columns), 19
- rpoly2 (rfun), 131
- rsquared, 134
- sample, 130
- sample (resample), 129

sapply, [31](#)
SAT, [135](#)
sd (mean), [98](#)
set.seed, [132](#)
setCorners (numD), [107](#)
setInterval (numD), [107](#)
show.settings, [145](#)
shuffle, [20](#)
shuffle (resample), [129](#)
singvals (linear.algebra), [92](#)
smoother (FunctionsFromData), [47](#)
SnowGR, [136](#)
solve.formula (findZeros), [40](#)
sp2df, [105](#), [137](#)
spline, [47](#)
spliner (FunctionsFromData), [47](#)
stat (interval), [87](#)
statTally, [138](#)
success=NULL, (prop.test), [123](#)
sum (mean), [98](#)
summary.groupwiseModel (mm), [101](#)
summary.nlsfunction (fitModel), [43](#)
surround, [139](#)
SwimRecords, [140](#)
symbolicAntiD (symbolicInt), [141](#)
symbolicD, [109](#), [140](#)
symbolicInt, [141](#)

t.test, [142](#), [143](#)
table, [116](#), [144](#)
tally, [123](#), [143](#)
tapply, [31](#)
TenMileRace, [144](#)
theme.mosaic, [145](#)
trellis.focus, [91](#)
trellis.par.set, [145](#)
trellis.unfocus, [91](#)
ttest (t.test), [142](#)
TukeyHSD, [146](#)
TukeyHSD.formula (TukeyHSD.lm), [146](#)
TukeyHSD.lm, [146](#)

Utilities, [147](#), [148](#)
Utilities2, [147](#), [148](#)

var (mean), [98](#)

Whickham, [149](#)

xchisq.test, [150](#)
xhistogram, [150](#)
xhistogramBreaks, [151](#)
xhistogramBreaks (xhistogram), [150](#)
xpnorm, [152](#)
xqnorm (xpnorm), [152](#)
xqqmath, [153](#)
xtabs, [12](#)
xyplot, [119](#)
xyz2latlon, [154](#)