

# Package ‘mosaic’

January 14, 2012

**Type** Package

**Title** Project MOSAIC (mosaic-web.org) math and stats teaching utilities

**Version** 0.3-2

**Date** 2012-01-12

**Depends** lattice,grid,methods,Hmisc,utils

**Suggests** MASS,reshape2,manipulate

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

**LazyLoad** yes

**LazyData** yes

**Collate**

'favstats.R''CIsim.R''DiffEQ.R''adapt\_seq.R''aggregating.R''binom.test.R''confint.R''cross.R''dfapply.R''distPlot.L  
algebra.R''logit.R''perctable.R''plt.R''pqrdata.R''prop.test.R''rand.R''read.file.R''resample.R''rfun.R''rgeo.R''setting  
package.R''datasets.R''themes.R''createFun.R'

## R topics documented:

mosaic-package . . . . .	3
.check_for_quant_input . . . . .	3
.createMathFun . . . . .	4
adapt_seq . . . . .	4
aggregating-methods . . . . .	5
binom.test . . . . .	6

Cards	8
CIsim	8
CPS	9
cross	10
D	11
deg2rad	12
dfapply	14
distPlot	15
do	16
dotPlot	17
favstats	18
findZeros	18
FunctionsFromData	19
Galton	20
Gestation	21
graphPaper	23
HELP	24
HELPfull	25
HELPmiss	50
HELPrct	51
interval	53
KidsFeet	54
ladd	55
linear-algebra	55
logit	57
makeFun	58
Marriage	58
mosaic.options	59
perctable	60
print.function	61
prop.test	62
qdata	63
rand	64
read.file	65
repeater-class	66
resample	66
rflip	67
rfun	68
SAT	69
SD	70
SnowGR	71
statTally	72
surround	73
TenMileRace	73
theme.mosaic	74
Utilities	75
Utilities2	75
Whickham	76
xchisq.test	77
xhistogram	78
xpnorm	79
xqqmath	80

**Index****82**

mosaic-package

*mosaic: the Project MOSAIC package***Description**

mosaic

**Details**

Data sets and utilities from Project MOSAIC ([mosaic-web.org](http://mosaic-web.org)) used to teach mathematics, statistics, computation and modeling. Funded by the NSF, Project MOSAIC is creating 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](mailto:rpruim@calvin.edu)>), Daniel Kaplan (<[kaplan@macalester.edu](mailto:kaplan@macalester.edu)>)

**References**

<http://mosaic-web.org>

---

`.check_for_quant_input`

*The Data Distribution*

---

**Description**

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

**Usage**

`.check_for_quant_input(x)`

**Arguments**

`x`                      a vector

**Value**

TRUE if quantitative; else stop

---

<code>.createMathFun</code>	<i>Internal function for handling formula-based expressions</i>
-----------------------------	---

---

### Description

`.createMathFun` processes an expression and assignments to produce a function and other information used by `D`, `antiD`, and `plotFun`

### Usage

```
.createMathFun(sexpr = NULL, ...)
```

### Arguments

<code>sexpr</code>	a formula containing a mathematical expression. The left-hand side has the expression, the right side contains the variables
<code>...</code>	additional assignments giving default values to symbolic parameters.

### Details

`.createMathFun` underlies user functions such as `makeFun`, `plotFun`, and `D`. It returns a structure that contains a function as well as information about the function needed for plotting, etc.

### Author(s)

Daniel Kaplan (<kaplan@macalester.edu>)

---

<code>adapt_seq</code>	<i>Adaptively generate sequences in an interval</i>
------------------------	---

---

### Description

`adapt_seq` is similar to `seq` except that instead of selecting point 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

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

**Value**

a numerical vector

**Examples**

```
adapt_seq(0, pi, 25, sin)
```

---

aggregating-methods	<i>Aggregating summary statistics</i>
---------------------	---------------------------------------

---

**Description**

These drop-in replacements and new summary statistics functions are formula-aware and allow the use of simple names within data frames. When given formulas, they call [aggregate](#) using the formula.

**Usage**

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

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

**Arguments**

x	a vector
na.rm	logical indicating whether NAs should be removed before calculating
...	additional arguments

**Details**

These methods are wrappers around functions and methods in the base and stats packages and provide additional interfaces.

The default value for `na.rm` is reversed from the functions in base and stats. Also, `na.rm`, `use`, and `trim` follow ... so must be named using their full names.

**Author(s)**

Randall Pruim ([rpruim@calvin.edu](mailto:rpruim@calvin.edu))

**See Also**

[aggregate](#), [sd](#), [var](#), [median](#), [mean](#), [max](#), [min](#), [sum](#)

**Examples**

```

mean(age, data=HELPrct)
mean(~age, data=HELPrct)
mean(age ~ ., data=HELPrct)
mean(age ~ 1, data=HELPrct)
mean(age ~ NULL, data=HELPrct)
mean(HELPrct$age)
mean(age ~ sex, data=HELPrct)
mean(age ~ sex + treat, data=HELPrct)
median(age, data=HELPrct)
median(~age, data=HELPrct)
median(age ~ ., data=HELPrct)
median(age ~ 1, data=HELPrct)
median(age ~ NULL, data=HELPrct)
median(HELPrct$age)
median(age ~ sex, data=HELPrct)
median(age ~ sex + treat, data=HELPrct)
sd(age, data=HELPrct)
sd(~age, data=HELPrct)
sd(age ~ ., data=HELPrct)
sd(age ~ 1, data=HELPrct)
sd(age ~ NULL, data=HELPrct)
sd(HELPrct$age)
sd(age ~ sex, data=HELPrct)
sd(age ~ sex + treat, data=HELPrct)
var(age, data=HELPrct)
var(~age, data=HELPrct)
var(age ~ ., data=HELPrct)
var(age ~ 1, data=HELPrct)
var(age ~ NULL, data=HELPrct)
var(HELPrct$age)
var(age ~ sex, data=HELPrct)
var(age ~ sex + treat, data=HELPrct)
min(age, data=HELPrct)
max(age, data=HELPrct)
max(~age, data=HELPrct)
max(age ~ ., data=HELPrct)
max(age ~ 1, data=HELPrct)
max(age ~ NULL, data=HELPrct)
max(HELPrct$age)
max(age ~ sex, data=HELPrct)
max(age ~ sex + treat, data=HELPrct)
count(sex, data=HELPrct)
count(sex, data=HELPrct, level='male')
count(HELPrct$sex)
prop(sex, data=HELPrct)
prop(sex, data=HELPrct, level='male')
prop(HELPrct$sex)

```

## Description

The mosaic `binom.test` provides wrapper functions around the function of the same name in **stats**. These wrappers provide an extended interface (including formulas). `binom.test` 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.

## 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, <code>n</code> 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`

## Author(s)

Randall Pruim (<rpruim@calvin.edu>)

## 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.
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 )
```

---

Cards	<i>Standard Deck of Cards</i>
-------	-------------------------------

---

### Description

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

### 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
```

---

CIsim	<i>Compute confidence intervals from (multiple) simulated data sets</i>
-------	---

---

### 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.

**Author(s)**

Randall Pruim (<rpruim@calvin.edu>)

**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)
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))

```

CPS

*Data from the Current Population Survey (CPS)***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 CPS, 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.

**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](http://lib.stat.cmu.edu/datasets/CPS_85_Wages).

**References**

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

**Examples**

```
data(CPS)
```

---

cross	<i>Factor cross products</i>
-------	------------------------------

---

**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

**Author(s)**

Randall Pruim (<rpruim@calvin.edu>)

**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(expr, ..., ..h.. = NULL, symbolic = TRUE,
   numerical = !symbolic, method = c("center", "right"))
```

```
antiD(expr, from = 0, to = NULL, ...)
```

**Arguments**

expr	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.
..h..	horizontal distance between points used for secant slope calculation in D(). This is used only if a symbolic derivative is not possible or if numerical=TRUE. The odd name, ..h.., is there to avoid conflicts with unbound variables in the formula f.
symbolic	a logical indicating whether symbolic differentiation should be attempted
numerical	opposite of symbolic available for convenience
method	For first-order numerical derivatives, whether to use a centered difference or a right-facing difference.
...	Default values to be given to unbound variables in the expression expr. See examples.
from	Default value for the lower bound of the interval of integration. This can be set at the time the integral function is invoked.
to	Default value for the upper bound of the interval of integration. This can be set at the time the integral function is invoked (and usually is).

**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 always returns a numerical integral.

For integrals, the return value is a function NOT of the variable (only one!) of integration, but of to and from, the upper and lower bounds of the interval of integration.

Since the return value is a function, the numerical value of the integral or derivative can be found by evaluating that function against its inputs.

**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,  $D(A \cdot x^2 + B \cdot y \sim x + y)$  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 make sure that you have the order right.

For anti-derivatives, a decorated function (eventually this will be an object of a new class).

**Note**

A major redesign of these functions is planned that will (1) refactor code so that the objects involved are S4 objects, and (2) allow `antiD` to return a single value if both `to` and `from` are specified.

**Author(s)**

Daniel Kaplan (<kaplan@macalester.edu>)

**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 <- antiD( A*exp(-k*t^2) ~ t, A=1, k=0.1)
F(from=-Inf, to=0)
F(from=-Inf, to=Inf)
```

---

deg2rad

---

*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**

```
deg2rad(x)

rad2deg(x)

xyz2latlon(x, y, z)

latlon2xyz(latitude, longitude)
```

```

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)

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

```

## Arguments

x	a numeric vector
y,z	numeric vectors
latitude,longitude	vectors of latitude and longitude values
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.
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

## Details

Some additional utility functions are documented here as well.

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 numeric vector  
a numeric vector  
a matrix each row of which contains a latitude and a longitude value  
a matrix each row of which contains the x, y, and z coordinates of a point on a unit sphere

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.

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.

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

### Author(s)

Randall Pruim (<rpruim@calvin.edu>)

### See Also

[googleMap](#)

### Examples

```
deg2rad(180)
rad2deg(2*pi)
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)
rgeo(4)
# sample from a region that contains the continental US
rgeo( 4, latlim=c(25,50), lonlim=c(-65,-125) )
rgeo2(4)
```

---

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 allong to FUN

### Author(s)

Randall Pruim (<rpruim@calvin.edu>)

### See Also

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

**Examples**

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

distPlot

*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**

```
distPlot(dist, params = list(),
  kind = c("density", "cdf", "qq", "histogram"),
  xlab = "", ylab = "", breaks, 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. <a href="#">dnorm</a> , <a href="#">pnorm</a> , and <a href="#">qnorm</a> ). 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	as per <a href="#">histogram</a>
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**

distPlot 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.

**Author(s)**

Randall Pruim (<[rpruim@calvin.edu](mailto:rpruim@calvin.edu)>)

**Examples**

```

distPlot('norm')
distPlot('norm', type='h')
distPlot('norm', kind='cdf')
distPlot('norm', params=list(mean=100, sd=10), kind='cdf')
distPlot('exp', kind='histogram')
distPlot('binom', params=list( 25, .25))
distPlot('binom', params=list( 25, .25), xlim=c(-1,26) )
distPlot('binom', params=list( 25, .25), kind='cdf')
distPlot('beta', params=list( 3, 10), kind='density')
distPlot('beta', params=list( 3, 10), kind='cdf')

```

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 = NULL)
```

**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

**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) * summary(lm(shuffle(height) ~ sex + mother, Galton))
do(3) * 1:4
do(3) * mean(rnorm(25))
do(3) * c(mean = mean(rnorm(25)))

```



dotPlot

*Dotplots***Description**

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

**Usage**

```
dotPlot(x,
  nint = if (is.factor(x)) nlevels(x) else round(1.3 * log2(length(x)) + 4),
  breaks, ..., panel = panel.dotPlot)

panel.dotPlot(x, breaks, equal.widths = TRUE,
  groups = NULL, nint = round(log2(length(x)) + 1),
  pch = if (is.null(groups)) trellis.par.get("dot.symbol")$pch else trellis.par.get("superpose"),
  col = if (is.null(groups)) trellis.par.get("dot.symbol")$col else trellis.par.get("superpose"),
  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**

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

**Value**

a trellis object

**Author(s)**

Randall Pruim (<rpruim@calvin.edu>)

**See Also**

[histogram](#)

**Examples**

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

---

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

---

**Description**

Computes mean, standard deviation, and quartiles of a data vector.

**Usage**

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

**Arguments**

x	numeric vector
na.rm	boolean indicating whether missing data should be ignored

**Value**

A vector of statistical summaries

**Examples**

```
favstats(1:10)
favstats(faithful$eruptions)
```

---

findZeros	<i>Find the zeros of a function</i>
-----------	-------------------------------------

---

**Description**

Compute numerically the zeros of a function.

**Usage**

```
findZeros(expr, ..., xlim = NULL, npts = 1000)
```

**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 variables named in the expression must be assigned values via ...
...	Specific numerical values for the unbound 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.
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. No guarantee that all the roots will be found.

**Details**

Searches numerically using uniroot.

**Value**

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

**Author(s)**

Daniel Kaplan (<kaplan@macalester.edu>)

**Examples**

```
findZeros( sin(x) ~ x, xlim=c(-10,10) )
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, xlim=c(-Inf,Inf))
```

---

FunctionsFromData	<i>Function from data.</i>
-------------------	----------------------------

---

**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 <a href="#">spline</a> .
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 <a href="#">loess</a> or <a href="#">lm</a>

## 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.

## Author(s)

Daniel Kaplan (<kaplan@macalester.edu>)

## See Also

[project](#) method for formulas

## Examples

```
cps <- fetchData("CPS.csv")
f <- smoother(wage~age,span=.9, data=cps)
f(40)
df <- D(f(age)~age)
df(40)
g <- linearModel(log(wage)~age+educ+1,data=cps)
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 <- spliner(x~y, monotonic=TRUE)
# f2(x=8:10)
f3 <- connector(x~y)
# f2(x=8:10)
```

---

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.

## 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.

### 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 admirable book by Deborah Nolan and Terry Speed describes the data in some 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)
```

---

graphPaper*Traditional graph-paper axes*

---

### Description

graphPaper plots out traditional graph-paper axes for making mathematical plots. Whereas standard R graphics put the axes at the edge of the plotting window, graphPaper puts them through the zeros and lets you set explicitly the location of tick marks and graph-paper rules. You can then plot over the axes by using functions such as lines, points, or funPlot (with add=TRUE).

### Usage

```
graphPaper(xticks = 0:5, yticks = -5:5,
  xlim = c(min(xticks), max(xticks)),
  ylim = c(min(yticks), max(yticks)), xlab = "x",
  ylab = "y", maxxlabels = 7, maxylabels = 7, ...)
```

### Arguments

xticks	numerical vector listing the position of x-axis ticks
yticks	like xticks but for y-axis
xlim	as in plot, sets x-axis limits (otherwise inferred from xticks)
ylim	similarly, set y-axis limits (otherwise inferred from yticks)
xlab	as in plot, sets the x-axis label
ylab	similarly, set the y-axis label
maxxlabels	limits the maximum number of tick marks on x-axis
maxylabels	limits the maximum number of tick marks on y-axis
...	additional arguments passed to plot

### Value

nothing

### Author(s)

Daniel Kaplan (<kaplan@macalester.edu>)

### Examples

```
graphPaper(xticks=-5:5, yticks=seq(-2, 2, by=.5))
plotFun(sin(x^2)~x, x=range(-4, 4), add=TRUE)
```

**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.

**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
- 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)
- pcs SF-36 Physical Component Score (measured at baseline)
- 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 subset of the HELPmiss data which includes an additional 17 subjects with partially observed data on some of the baseline variables. This is also a subset of the HELFull data which includes 5 timepoints and many additional variables for all subjects.

## 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

[HELP](#) , and [HELFull](#).

## 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.

## 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 Loved 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 DAUGHTER 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 overnight 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 #moths 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 activ (1=Limited a lot, 2=Limited a little, 3=Not limited)
- B3B Does your health limit you in moderate activ (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 sevr 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 d/t phys hlth-lst 4 wks (0=No, 1=Yes)
- B4B Accomplish less d/t phys hlth-lst 4 wks (0=No, 1=Yes)
- B4C Lim wrk/act type d/t phys hlth-lst 4 wks (0=No, 1=Yes)
- B4D Diff perf work d/t phys hlth-lst 4 wks (0=No, 1=Yes)
- B5A Cut wrk/act time d/t emot prbs-lst 4 wks (0=No, 1=Yes)
- B5B Accomplish ess d/t emot probs-lst 4 wks (0=No, 1=Yes)
- B5C <carefl w/wrk/act d/t em prb-lst 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 little 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 Told 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-lst 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-lst 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-lst 6 mos (0=No, 1=Yes)
- C3G4 What was the result of last test (1=Positive, 2=Negative, 3=Refued, 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 drg 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 drg/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 hlfrwy hse/res facil-1st 6 mos
- E4C Ttl nites in hlfrwy hse/res fac-last 6 mos
- E5A In day trtmnt prg for alc/drgu-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 alc/drugs-1st 6 mos (0=No, 1=Yes)
- E8A2 Saw Prst/Min/Rabbi re alc/drugs-1st 6 mos (0=No, 1=Yes)
- E8A3 Employ Asst Prg for alc/drg prb-1st 6 mos (0=No, 1=Yes)
- E8A4 Oth source cnsl for alc/drg prb-1st 6 mos (0=No, 1=Yes)
- E9A AA/NA/slf-hlp for drg/alc/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 overnite/+ in hosp-1st 6 mos (0=No, 1=Yes)
- E11B # times overnight/+ 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 trtmnt from acupuncturist-last 6 mos (0=No, 1=Yes)
- E14B Recd trtmnt 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 trtmnt 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 d/t 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 d/t trans prob (0=No, 1=Yes)
- E16A5 a numeric vectorOffice/clinic hours are inconvenient (0=No, 1=Yes)
- E16A6 I don't speak/understnd enough English (0=No, 1=Yes)
- E16A7 Afraid othrs find out about my hlth prob (0=No, 1=Yes)
- E16A8 My substance abuse interferes (0=No, 1=Yes)
- E16A9 I don't have someone to watch my childrn (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 d/t transp prob (0=No, 1=Yes)
- E18D The office/clinic hrs were inconvenient (0=No, 1=Yes)
- E18F Afraid others might find out about prob (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 prog (0=No, 1=Yes)
- E18L I do not need substance abuse services (0=No, 1=Yes)



- E18M Other reason not get sub ab services (0=No, 1=Yes)
- F1A Bothered by thngs not gen boethered by (0=Rarely/never, 1=Some of the time, 2=Occas/moderately, 3=Most of the time)
- F1B My appretite was poor (0=Rarely/never, 1=Some of the time, 2=Occas/moderately, 3=Most of the time)
- F1C Couldn't shake blues evn w/fam+frnds hlp (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 everthing 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 enoyed 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-1st 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-1st 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 lst 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 lst 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-lst 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 intrvwr w/c subst is main prob (0=No problem, 1=Alcohol, 2=Alcool to intox, 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-lst 30 days bef detox
- H16B \$ spent on drugs-lst 30 days bef detox
- H17A # days had alc prob-lst 30 days bef det
- H17B # days had drg prob-lst 30 days bef det
- H18A How troubled by alc probs-lst 30 days (0=Not at all, 1=Slightly, 2=Moderately, 3=Considerably, 4=Extremely)
- H18B How troubled by drug probs-lst 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 lst 30 days bef detox
- I2 Most drank any 1 day in lst 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 hangovers Sun or Mon mornings (0=No, 1=Yes)
- L3 Have you had the shakes when sobering (0=No, 1=Sometimes, 2=Alm evry time drnk)
- L4 Do you get phys sick as reslt of drinking (0=No, 1=Sometimes, 2=Alm evry time drnk)
- 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 bec 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 drnk)
- 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 d/t drinking-lst 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 dwn on drnkng & failed (0=No, 1=Once, 2=Several times)
- L24 Do you gulp drinks (0=No, 1=Yes)
- L25 After taking 1 or 2 drnks can you stop (0=No, 1=Yes)

- M1 Had hangover/felt bad aftr using alc/drugs (0=No, 1=Yes)
- M2 Felt bad about self bec of alc/drg use (0=No, 1=Yes)
- M3 Missed days wrk/sch bec of alc/drg use (0=No, 1=Yes)
- M4 Fam/frinds worry/compl about alc/drg use (0=No, 1=Yes)
- M5 I have enjoyed drinking/using drugs (0=No, 1=Yes)
- M6 Qual of work suffered bec of alc/drg use (0=No, 1=Yes)
- M7 Parenting ability harmed by alc/drg use (0=No, 1=Yes)
- M8 Trouble sleeping/nightmares aftr alc/drugs (0=No, 1=Yes)
- M9 Driven motor veh while undr inf alc/drugs (0=No, 1=Yes)
- M10 Using alc/1 drg caused > use othr drgs (0=No, 1=Yes)
- M11 I have been sick/vomited aft alc/drg use (0=No, 1=Yes)
- M12 I have been unhappy bec of alc/drg use (0=No, 1=Yes)
- M13 Lost weight/eaten poorly d/t alc/drg use (0=No, 1=Yes)
- M14 Fail to do what expected d/t alc/drg use (0=No, 1=Yes)
- M15 Using alc/drugs has helped me to relax (0=No, 1=Yes)
- M16 Felt guilt/ashamed bec of my alc drg use (0=No, 1=Yes)
- M17 Said/done emarras thngs when on alc/drg (0=No, 1=Yes)
- M18 Personality changed for worse on alc/drg (0=No, 1=Yes)
- M19 Taken foolish risk when using alc/drugs (0=No, 1=Yes)
- M20 Gotten into trouble bec of alc/drg use (0=No, 1=Yes)
- M21 Said cruel things while using alc/drugs (0=No, 1=Yes)
- M22 Done impuls thngs regret d/t alc/drg use (0=No, 1=Yes)
- M23 Gotten in phys fights when use alc/drugs (0=No, 1=Yes)
- M24 My phys health was harmed by alc/drg use (0=No, 1=Yes)
- M25 Using alc/drg helped me have more + outlook (0=No, 1=Yes)
- M26 I have had \$ probs bec of my alc/drg use (0=No, 1=Yes)
- M27 My love relat harmed d/t my alc/drg use (0=No, 1=Yes)
- M28 Smoked tobacco more when using alc/drugs (0=No, 1=Yes)
- M29 <y phys appearance harmed by alc/drg use (0=No, 1=Yes)
- M30 My family hurt bec of my alc drg use (0=No, 1=Yes)
- M31 Close relationsp damaged d/t alc drg use (0=No, 1=Yes)
- M32 Spent time in jail bec of my alc/drg use (0=No, 1=Yes)
- M33 My sex life suffered d/t my alc/drg use (0=No, 1=Yes)
- M34 Lost interst in activ d/t my alc/drg use (0=No, 1=Yes)
- M35 Soc life> enjoyable when using alc/drg (0=No, 1=Yes)
- M36 Spirit/moral life harmed by alc/drg use (0=No, 1=Yes)
- M37 Not had kind life want d/t alc/drg use (0=No, 1=Yes)
- M38 My alc/drg use in way of personal growth (0=No, 1=Yes)
- M39 My alc/drg use damaged soc life/reputat (0=No, 1=Yes)
- M40 Spent/lost too much \$ bec alc/drg use (0=No, 1=Yes)

- M41 Arrested for DUI of alc or oth drgs (0=No, 1=Yes)
- M42 Arrested for offenses rel to alc/drg use (0=No, 1=Yes)
- M43 Lost marriage/love relat d/t alc/drg use (0=No, 1=Yes)
- M44 Susp/fired/left job/sch d/t alc/drg use (0=No, 1=Yes)
- M45 I used drgs moderately w/o having probs (0=No, 1=Yes)
- M46 I have lost a friend d/t my alc/drg use (0=No, 1=Yes)
- M47 Had an accident while using alc/drugs (0=No, 1=Yes)
- M48 Phys hurt/inj/burned when using alc/drugs (0=No, 1=Yes)
- M49 I injured someone while using alc/drugs (0=No, 1=Yes)
- M50 Damaged things/prop when using alc/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 dwn w/o feel funny ltr (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 dwn 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 drnk/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 drgs/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 alc/drgs (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 drgs/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 assault you using alc/drgs (1=Don't know, 2=Never, 3=Some cases, 4=Most cases, 5=All cases, 9=Never assaulted)
- Q1A Have ypi ever omjected drugs (0=No, 1=Yes)
- Q1B Have you injected drugs-lst 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 oft been to shoot gall/hse-lst 6 mos (0=Never, 1=Few times or less, 2= Few times/month, 3= Once or more/week)
- Q5 How oft 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 oft 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 oft 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 oft 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 oft use syringe to div drgs-lst 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=1woman, 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 oft had sex to get drgs-last 6 mos (0=Never, 1=Few times or less, 2=Few times/month, 3=Once or more/week)
- Q15 How oft given drgs to have sex-1st 6 mos (0=Never, 1=Few times or less, 2=Few times/month, 3=Once or more/week)
- Q16 How oft 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 oft 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 oft 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 alc/drg use (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1B Sometimes I wonder if I'm an alc/addict (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1C Id I don't chng alc/drg probs will worsen (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1D I started making changes in alc/drg 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 alc/drg 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 alc/drg 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/ alc/drugs (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1K I wonder if I'm in contrl of alc/drg use (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1L My alc/drg use is causing a lot of harm (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1M Actively curring dwn/stopping alc/drg use (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1N Want help to not go back to alc/drugs (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)



- R10 I know that I have an alc/drg problem (1=Strongly disagree, 2=Disagree, 3= Agree, 4=Strongly agree)
- R1P I wonder if I use alc/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 alc/drg 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 intrvw pt obviously depressed/withdrawn (0=No, 1=Yes)
- S1B at intrvw pt obviously hostile (0=No, 1=Yes)
- S1C At intrvw pt obviouslt anx/nervous (0=No, 1=Yes)
- S1D Trbl w/real tst/thght dis/par at intrvw (0=No, 1=Yes)
- S1E At intrvw pt trbl w/ compr/concen/rememb (0=No, 1=Yes)
- S1F At intrvw 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 d/t 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 reg 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 reg 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 reg MD (0=No, 1=Yes)
- PCP\_ID a numeric vector
- U7A What type of MD is your reg 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 PRACTICIONER NURSE PRACTITIONER ONCOLOGIST PRENATAL PRIMARY PRIMARY CAAE 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 reg MDs office-pst 6 mos
- U10B # times saw reg 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 bec 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 oft see reg MD when have reg chck-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 oft see reg 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 oft 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 bec of MDs advice (0=No, 1=Yes)
- U26B Tried to drnk < alcohol bec of MD advice (0=No, 1=Yes)
- U26C Wore my seat belt more bec of MDs advice (0=No, 1=Yes)
- U26D Changed diet bec of MDs advice (0=No, 1=Yes)
- U26E Done more exercise bec MDs advice (0=No, 1=Yes)
- U26F Relax/reduce stress be of MDs advice (0=No, 1=Yes)
- U26G Practiced safer sex bec of MDs advice (0=No, 1=Yes)
- U26H Tried to cut dwn/quit drgs bec MD advice (0=No, 1=Yes)"
- U26I Got HIV tested bec 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 to you trst 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 reg 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 reg 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 mant 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 mant 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 reg MD (1=Completely satisfied, 2=Very satisfied, 3=Some-what satisfied, 4=Neither, 5=Somewhat dissatisfied, 6=Very dissatisfied, 7=Completely dissatisfied)
- V1 Evr needed to drnk much > 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 reg MD: red tape (0=No, 1=Yes)
- E16A\_IB Barrier to reg MD: internal barriers (0=No, 1=Yes)
- E16A\_TM Barrier to reg MD: time restrictions (0=No, 1=Yes)
- E16A\_DD Barrier to reg 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=Heroine, 4=Barbituates, 5=Benzos, 6=Marijuana, 7=Methadone, 8=Opiates)
- SECD\_SUB Second drug of choice (0=None, 1=Alcohol, 3=Cocaine, 3=Heroine, 4=Barbituates, 5=Benzos, 6=Marijuana, 7=Methadone, 8=Opiates)
- ALCOHOL 1st/2nd drug of coice=Alcohol (0=No, 1=Yes)
- COC\_HER 1st/2nd drug of choice=cocaine or heroine (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)
- O1B\_REC Family/friends heavy drinkers y/n (0=No, 1=Yes)
- O1C\_REC Family/friends use drugs y/n (0=No, 1=Yes)
- O1D\_REC Family/friends support abst. y/n (0=No, 1=Yes)
- O2\_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
- 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 subje 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 HELP 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

[HELP](#) , 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.

**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 at baseline
- 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)
- pcs SF-36 Physical Component Score (measured at baseline)
- 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 HELP data with 17 subjects with partially observed data on some of the baseline variables. This is a subset of the HELFull 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

[HELP](#) , and [HELFull](#).

## Examples

```
data(HELPrct)
```

---

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.

## 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
- 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)
- pcs SF-36 Physical Component Score (measured at baseline)
- 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 subset of the [HELPMiss](#) data which includes an additional 17 subjects with partially observed data on some of the baseline variables. This is also a subset of the [HELFull](#) 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**

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

**Examples**

```
data(HELPrct)
```

---

interval	<i>Extract summary statistics</i>
----------	-----------------------------------

---

**Description**

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

**Usage**

```
interval(x, ...)

## S3 method for class 'htest'
interval(x, verbose = FALSE, ...)

pval(x, ...)

## S3 method for class 'htest'
pval(x, digits = 4, verbose = FALSE,
     ...)

stat(x, ...)

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

**Arguments**

<code>x</code>	An object of class <code>htest</code> .
<code>...</code>	Additional arguments.
<code>verbose</code>	a logical
<code>digits</code>	number of digits to display in verbose output

**Author(s)**

Randall Pruim (<rpruim@calvin.edu>)

**Examples**

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

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

---

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.

**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](http://www.amstat.org/publications/jse/v14n1/datasets.meyer.html)

**Examples**

```
data(KidsFeet)
```

ladd

*Add to Lattice Plots***Description**

Simplified lattice plotting by adding additional elements to existing plots.

**Usage**

```
ladd(x, col = 1, row = 1)
```

**Arguments**

x	callable graphical element to be added to a panel in a lattice plot
row	row identifying desired panel in multi-panel plots.
col	column identifying desired panel in multi-panel plots.

**Details**

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

**Author(s)**

Randall Pruim (<rpruim@calvin.edu>)

**Examples**

```
print(xyplot(rnorm(100) ~ rnorm(100)))
ladd(panel.abline(a=0,b=1))
ladd(panel.abline(h=0,col='red'))
ladd(grid.text('Hello'))
ladd(grid.text(x=.95,y=.05,'text here',just=c('right','bottom'))))
print(xyplot( rnorm(100) ~ rnorm(100) | factor(rbinom(100,3,.5)), layout=c(2,2) ))
ladd( grid.text("(1,1)",gp=gpar(cex=3,alpha=.5)), 1,1)
ladd( grid.text("(2,1)",gp=gpar(cex=3,alpha=.5)), 2,1)
ladd( grid.text("(1,2)",gp=gpar(cex=3,alpha=.5)), 1,2)
ladd( grid.text("(2,2)",gp=gpar(cex=3,alpha=.5)), 2,2)
```

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.

**Usage**

```
project(x, u, ...)

singvals(A, data = NULL)

dot(u, v)
```

**Arguments**

A	a formula. In <code>mat()</code> and <code>singvals()</code> , only the right-hand side is used. In <code>project()</code> , both sides are used, but the left-hand side should be a single quantity
x	a numeric vector, formula, or matrix
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)
v	a numeric vector

**Details**

`mat()` enerates a model matrix from a formula while `project()` carries out the operation of least-squares fitting using a singular value method. This means that even when the matrix is singular, a solution, either exact or least-squares, will be found. To demonstrate singularity, use `singvals()`. NOTE: unlike the standard formula expansion in `lm()`, these linear algebra functions do NOT include an intercept by default. If you want an intercept, put `+1` as a term in your formula. (See the examples.)

**Value**

`mat` returns a matrix

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

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

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

**Author(s)**

Daniel Kaplan (<kaplan@macalester.edu>)

**See Also**

[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
# projection onto the 1 vector gives the mean vector
mean(x)
project(x, 1)
project(x, 1, type='length')
```



```

mat(~a+b)
mat(~a+b+1)
kids = fetchData("KidsFeet.csv")
mat(~length+sex,data=kids)
project(a~b)
project(width~length+sex,data=kids)
project(log(width)~I(length^2)+sin(length)+sex,data=kids)
singvals(~length*sex*width,data=kids)

```

---

logit

*Logit and inverse logit functions*


---

## 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

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

## Author(s)

Randall Pruim (<rpruim@calvin.edu>)

## Examples

```

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

```

---

makeFun

*Utilities for creating mathematical functions from expressions*


---

### Description

makeFun turns a mathematical expression (written as a formula) into the corresponding R function. It's purely a convenience that uses the same expression syntax as plotFun, D and antiD.

### Usage

```
makeFun(sexpr = NULL, ...)
```

### Arguments

sexpr	the mathematical expression written as a formula
...	additional arguments setting default numerical values for arguments or parameters

### Value

a function

### Author(s)

Daniel Kaplan (<kaplan@macalester.edu>)

### Examples

```
f = makeFun(sin(x^2) ~ x)
g = makeFun(a*x + b ~ x, a=2, b=4)
g(x=10)
g(x=10, a=5, b=2)
h = makeFun(sin(x)*y ~ x&y)
```

---

Marriage

*Marriage records*


---

### Description

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

### 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)
```

---

```
mosaic.options
```

---

```
Setting options for mosaic package functions
```

---

### 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)
```

**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.

perctable

*Cross tabulation displayed as percents or proportions***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 <a href="#">table</a> ; 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.

**Author(s)**

Randall Pruim (<rpruim@calvin.edu>)

**Examples**

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

---

print.function	<i>Printing Calculus Functions</i>
----------------	------------------------------------

---

**Description**

Printing Calculus Functions

**Usage**

```
## S3 method for class 'function'
print(x, useSource = TRUE,
      default = FALSE, ...)
```

**Arguments**

x	the function to be printed
default	use the standard function printing methods
useSource	passed along to the standard function printing methods
...	additional arguments

**Details**

The mosaic calculus operators such as `D` and `antiD` and `linearModel` sometimes create functions with an obscure numerical method as the internal contents. This print method detects such functions (through their `kind` attribute) and provides a more friendly display

**Value**

nothing

**Author(s)**

Daniel Kaplan (<kaplan@macalester.edu>)

**Examples**

```
F <- antiD(sin(x^2)~x)
F
print(F)
print(F, default=TRUE)
```

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.

**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>	a vector of probabilities of success. The length of <code>p</code> must be the same as the number of groups specified by <code>x</code> , and its elements must be greater than 0 and less than 1.
<code>alternative</code>	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.
<code>conf.level</code>	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.
<code>correct</code>	a logical indicating whether Yates' continuity correction should be applied where possible.
<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, <code>n</code> may be a data frame)
<code>...</code>	additional arguments (often ignored)

**Details**

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

**Author(s)**

Randall Pruim (<rpruim@calvin.edu>)

**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 )
```

---

qdata	<i>Compute quantiles from data</i>
-------	------------------------------------

---

## Description

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

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

pdata(q, vals, data = NULL, lower.tail = TRUE, ...)

rdata(n, vals, data = NULL, replace = TRUE, ...)

ddata(x, vals, data = NULL, log = FALSE, ...)
```

## Arguments

p	a vector of probabilities
vals	a vector containing the data
data	a data frame in which to evaluate vals
...	additional arguments passed to <code>quantile</code> or <code>sample</code>
q	a vector of quantiles
lower.tail	a logical indicating whether to use the lower or upper tail probability
n	number of values to sample
replace	a logical indicating whether to sample with replacement
x	a vector of quantiles
log	a logical indicating whether the result should be log transformed

**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, iris$Sepal.Length)
qdata(.5, Sepal.Length, data=iris)
data(iris)
pdata(3:6, iris$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)
```

---

rand	<i>Random Regressors</i>
------	--------------------------

---

**Description**

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

**Usage**

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

**Arguments**

df	degrees of freedom, i.e., number of random regressors
rdist	random distribution function for sampling
args	arguments for rdist
nr	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`

**Author(s)**

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



**Examples**

```
rand(2,nr=4)
rand(2,rdist=rpois, args=list(lambda=3), nr=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 simply 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 <a href="#">read.table</a> for more details.
header	logical; For .txt and .csv files, this indicates whether the first line of the file includes variables names.
na.strings	character: stings 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 <a href="#">read.table</a> , <a href="#">read.csv</a> , or <a href="#">load</a> .

**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.

**Author(s)**

Randall Pruim (<rpruim@calvin.edu>)

**See Also**

[read.table](#), [read.csv](#), [load](#).

---

repeater-class	<i>Repeater objects</i>
----------------	-------------------------

---

### 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](#).

### Details

Each object contains slots for

- `n` number of times to repeat something
- `cull` a function used to cull output
- `mode` the mode used for results (NULL or data frame or matrix)

### See Also

[do](#)

---

resample	<i>More Random Samples</i>
----------	----------------------------

---

### Description

These functions simplify and unify sampling in various ways.

### Usage

```
resample(x, size, replace = TRUE, prob = NULL,
  groups = NULL, orig.ids = FALSE, ...)

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

shuffle(x, replace = FALSE, prob = NULL, groups = NULL,
  orig.ids = 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	groups to sample within (works much like groups in lattice plots)
orig.ids	a logical; should original ids be included in returned data frame?
...	additional arguments passed to <a href="#">sample</a>
fixed	a vector of column names
shuffled	a vector of column names
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(c(0,1), 100)
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)
```

---

rfun

*Generate a natural-looking function*


---

**Description**

Produce a random function that is sum of gaussians  
 rpoly2 generates a random 2nd degree poly of 2 vars (as a function)

**Usage**

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

rpoly2(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 <a href="#">set.seed</a> .
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, 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

**Value**

a function with the appropriate number of inputs

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

**Author(s)**

Daniel Kaplan (<kaplan@macalester.edu>)

**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)
```

---

 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

**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')})
```

---

SD	<i>Compute standard deviation</i>
----	-----------------------------------

---

## Description

This computes the standard deviation as the square root of variance to avoid direct use of [sd](#).

## Usage

```
SD(x)
```

## Arguments

x                      a vector or formula

## Value

a numeric containing the standard deviation

## Note

The primary reason for this function is that [sd](#) generates warnings when used with [summary.formula](#) from the `lattice` package.

## See Also

[sd](#)

## Examples

```
x <- rnorm(10)
SD(x)
sd(x)
summary(age ~ substance, data=HELPrct, fun=SD)
```

SnowGR

*Snowfall data for Grand Rapids, MI***Description**

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

**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 start 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)
}
```

---

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,
  stemplot = dim(rdata)[direction] < 201,
  q = c(0.5, 0.9, 0.95, 0.99), fun, ...)
```

## 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.
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
...	additional arguments passed to <a href="#">xhistogram</a>

## Value

A lattice plot is returned invisibly but can be printed.

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

## Author(s)

Randall Pruim (<rpruim@calvin.edu>)

## Examples

```
# is my spinner fair?
x <- c(10,18,9,15) # counts in four cells
rdata <- rmultinom(1000, 4, prob=rep(.25,4))
print(statTally( x, rdata, fun=max )) # unusual test statistic
print(statTally( x, rdata, fun=var )) # equivalent to chi-squared test
```



---

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 <a href="#">format</a>

**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)
```

---

TenMileRace	<i>Cherry Blossom Race</i>
-------------	----------------------------

---

**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.

**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 in similarity to `col.whitebg`, but since more just colors are set, `theme.mosaic` seems like a preferable name.

## Author(s)

Randall Pruim (<rpruim@calvin.edu>)

## See Also

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

## Examples

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

---

Utilities	<i>Utility bills</i>
-----------	----------------------

---

**Description**

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

**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

**Source**

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

**Examples**

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

---

Utilities2	<i>Utility bills</i>
------------	----------------------

---

**Description**

Data from utility bills at a residence. This is a wider version of [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 exectric bill (in dollars)
- notes notes about the billing period
- ccfpday average gas usage per day [Utilities2 only]
- kwhpday 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.

### 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.

**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](#).

**Author(s)**

Randall Pruim (<rpruim@calvin.edu>)

**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

---

*Augmented histograms*

---

**Description**

xhisoram adds some additional functionality to [histogram](#) making it simpler to obtain certain common histogram adornments.

**Usage**

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

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,
  breaks, 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'
...	additional arguments passed to <a href="#">histogram</a> and on to <a href="#">panel.xhistogram</a>
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 <a href="#">fitdistr</a>
groups, breaks	as per <a href="#">histogram</a>

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**

a trellis object

**Examples**

```
xhistogram(~age | substance, HELPrct, v=35, fit='normal')
xhistogram(~age, HELPrct, labels=TRUE, type='count')
xhistogram(~age, HELPrct, groups=cut(age, seq(10,80,by=10)))
xhistogram(~age, HELPrct, groups=sex, stripes='horizontal')
```

---

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.

**Author(s)**

Randall Pruim (<rpruim@calvin.edu>)

**See Also**

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

**Examples**

```
xpnorm(650, 500, 100)
xqnorm(.75, 500, 100)
if (require(manipulate)) {
  manipulate( xpnorm(score, 500, 100, verbose=verbose),
    score = slider(200,800),
    verbose = checkbox(TRUE, label="Verbose Output")
  )
}
```

---

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.



<code>slope</code>	slope for added math line
<code>intercept</code>	intercept for added math line
<code>overlines</code>	a logical: should lines be on top of qq plot?
<code>groups, pch, lwd, lty</code>	as in lattice plots
<code>col.line</code>	color to use for added lines

**Value**

a trellis object

**Examples**

```
xqqmath(rnorm(100))
```

# Index

- \*Topic **Check**
  - .check\_for\_quant\_input, 3
- \*Topic **calculus**
  - D, 11
  - findZeros, 18
- \*Topic **datasets**
  - Cards, 8
  - CPS, 9
  - Galton, 20
  - Gestation, 21
  - HELP, 24
  - HELPfull, 25
  - HELPmiss, 50
  - HELPpct, 51
  - KidsFeet, 54
  - Marriage, 58
  - SAT, 69
  - SnowGR, 71
  - TenMileRace, 73
  - Utilities, 75
  - Utilities2, 75
  - Whickham, 76
- \*Topic **distribution**
  - .check\_for\_quant\_input, 3
  - rand, 64
- \*Topic **geometry**
  - deg2rad, 12
- \*Topic **graphics**
  - distPlot, 15
  - dotPlot, 17
  - ladd, 55
  - theme.mosaic, 74
- \*Topic **inference**
  - CIsim, 8
  - statTally, 72
- \*Topic **input**
  - .check\_for\_quant\_input, 3
- \*Topic **is**
  - .check\_for\_quant\_input, 3
- \*Topic **iteration**
  - do, 16
- \*Topic **manipulate**
  - cross, 10
  - perctable, 60
- \*Topic **methods**
  - aggregating-methods, 5
- \*Topic **package**
  - mosaic-package, 3
- \*Topic **quantitative**
  - .check\_for\_quant\_input, 3
- \*Topic **random**
  - deg2rad, 12
  - rfunc, 68
- \*Topic **regression**
  - rand, 64
- \*Topic **simulation**
  - CIsim, 8
- \*Topic **stats**
  - aggregating-methods, 5
  - binom.test, 6
  - distPlot, 15
  - favstats, 18
  - interval, 53
  - prop.test, 62
- \*Topic **teaching**
  - statTally, 72
- \*Topic **univar**
  - interval, 53
- \*Topic **util**
  - read.file, 65
- \*Topic **whether**
  - .check\_for\_quant\_input, 3
- \* (do), 16
- \*, repeater, ANY-method (do), 16
- .check\_for\_quant\_input, 3
- .createMathFun, 4
- adapt\_seq, 4
- aggregate, 5
- aggregating-methods, 5
- antiD (D), 11
- apply, 14
- binom.test, 6, 7, 62
- binom.test, ANY-method (binom.test), 6
- binom.test, character-method (binom.test), 6

- binom.test, factor-method (binom.test), 6
- binom.test, formula-method (binom.test), 6
- binom.test, logical-method (binom.test), 6
- binom.test, numeric-method (binom.test), 6
- card (Cards), 8
- Cards, 8
- chisq.test, 77, 80
- CIsim, 8
- col.mosaic (theme.mosaic), 74
- col.whitebg, 74
- connector (FunctionsFromData), 19
- count (aggregating-methods), 5
- count, ANY-method (aggregating-methods), 5
- count, data.frame-method (aggregating-methods), 5
- count, factor-method (aggregating-methods), 5
- count, formula-method (aggregating-methods), 5
- count, logical-method (aggregating-methods), 5
- CPS, 9
- cross, 10
- D, 11
- ddata (qdata), 63
- deal (resample), 66
- deg2rad, 12
- dfapply, 14
- distPlot, 15
- dnorm, 15
- do, 16, 66
- dot (linear-algebra), 55
- dotPlot, 17
- favstats, 18
- findZeros, 18
- fitdistr, 78
- format, 73
- FunctionsFromData, 19
- Galton, 20
- Gestation, 21
- googleMap, 14
- googleMap (deg2rad), 12
- graphPaper, 23
- HELP, 24, 25, 49, 51
- HELPfull, 25, 25, 51–53
- HELPMiss, 49, 50, 52, 53
- HELPrct, 51
- histogram, 15, 17, 78, 80
- ilogit (logit), 57
- interval, 53
- KidsFeet, 54
- ladd, 55
- lapply, 14
- latlon2xyz (deg2rad), 12
- linear-algebra, 55
- linearModel, 56
- linearModel (FunctionsFromData), 19
- lm, 19
- load, 65
- loess, 19
- logit, 57
- makeFun, 58
- Marriage, 58
- mat (linear-algebra), 55
- max, 5
- max (aggregating-methods), 5
- mean, 5
- mean (aggregating-methods), 5
- mean, ANY-method (aggregating-methods), 5
- mean, data.frame-method (aggregating-methods), 5
- mean, formula-method (aggregating-methods), 5
- mean, numeric-method (aggregating-methods), 5
- median, 5
- median (aggregating-methods), 5
- median, ANY-method (aggregating-methods), 5
- median, data.frame-method (aggregating-methods), 5
- median, formula-method (aggregating-methods), 5
- median, numeric-method (aggregating-methods), 5
- min, 5
- min (aggregating-methods), 5
- mosaic (mosaic-package), 3
- mosaic-package, 3
- mosaic.getOption (mosaic.options), 59
- mosaic.options, 59
- mosaic.par.get (mosaic.options), 59
- mosaic.par.set (mosaic.options), 59

- mosaicGetOption (mosaic.options), 59
- nflip (rflip), 67
- panel.dotPlot (dotPlot), 17
- panel.xhistogram, 78
- panel.xhistogram (xhistogram), 78
- panel.xqqmath (xqqmath), 80
- pdata (qdata), 63
- perctable, 60
- plot, 80
- pnorm, 15, 80
- print (do), 16
- print, repeater-method (do), 16
- print.cointoss (rflip), 67
- print.function, 61
- project, 20
- project (linear-algebra), 55
- project, formula, ANY-method (linear-algebra), 55
- prop (aggregating-methods), 5
- prop, ANY-method (aggregating-methods), 5
- prop, data.frame-method (aggregating-methods), 5
- prop, factor-method (aggregating-methods), 5
- prop, formula-method (aggregating-methods), 5
- prop, logical-method (aggregating-methods), 5
- prop.test, 7, 62, 62
- prop.test, ANY-method (prop.test), 62
- prop.test, character-method (prop.test), 62
- prop.test, factor-method (prop.test), 62
- prop.test, formula-method (prop.test), 62
- prop.test, logical-method (prop.test), 62
- prop.test, numeric-method (prop.test), 62
- proptable (perctable), 60
- pval (interval), 53
- qdata, 63
- qnorm, 15, 80
- qqmath, 80
- quantile, 63
- rad2deg (deg2rad), 12
- rand, 64
- rdata (qdata), 63
- read.csv, 65
- read.file, 65
- read.table, 65
- repeater-class, 66
- replicate, 16, 66
- resample, 66
- rflip, 67
- rfun, 68
- rgeo (deg2rad), 12
- rgeo2 (deg2rad), 12
- rpoly2 (rfun), 68
- sample, 67
- sample.data.frame (resample), 66
- sample.default (resample), 66
- sample.factor (resample), 66
- sample.matrix (resample), 66
- sapply, 14
- SAT, 69
- SD, 70
- sd, 5, 70
- sd (aggregating-methods), 5
- sd, ANY-method (aggregating-methods), 5
- sd, data.frame-method (aggregating-methods), 5
- sd, formula-method (aggregating-methods), 5
- sd, numeric-method (aggregating-methods), 5
- set.seed, 68
- show.settings, 74
- shuffle (resample), 66
- singvals (linear-algebra), 55
- smoother (FunctionsFromData), 19
- SnowGR, 71
- spline, 19
- spliner (FunctionsFromData), 19
- stat (interval), 53
- statTally, 72
- sum, 5
- summary.formula, 70
- surround, 73
- table, 60
- tapply, 14
- TenMileRace, 73
- theme.mosaic, 74
- trellis.focus, 55
- trellis.par.set, 74
- trellis.unfocus, 55
- Utilities, 75, 75
- Utilities2, 75, 75
- var, 5
- var (aggregating-methods), 5
- var, ANY, ANY, ANY, ANY, ANY-method (aggregating-methods), 5

`var, data.frame, ANY, ANY, ANY, ANY-method`  
    (aggregating-methods), [5](#)  
`var, formula, data.frame, ANY, ANY, missing-method`  
    (aggregating-methods), [5](#)  
`var, formula, missing, ANY, ANY, data.frame-method`  
    (aggregating-methods), [5](#)  
`var, matrix, ANY, ANY, ANY, ANY-method`  
    (aggregating-methods), [5](#)  
`var, numeric, ANY, ANY, ANY, ANY-method`  
    (aggregating-methods), [5](#)  
`var, numeric, numeric, ANY, ANY, ANY-method`  
    (aggregating-methods), [5](#)

Whickham, [76](#)

`xchisq.test`, [77](#)  
`xhistogram`, [72](#), [78](#)  
`xpnorm`, [79](#)  
`xqnorm(xpnorm)`, [79](#)  
`xqqmath`, [80](#)  
`xyz2latlon(deg2rad)`, [12](#)