R & S-Plus

Last Updated 2016-07-09

Help

- help(topic, ...) # Get help on various topics
- options(htmlhelp=FALSE/TRUE) # Use terminal or browser
- help.search(pattern, ...)
- library(help=*libName*)

Vector Indexing

NOTE: a<-1:10; names(a)<-letters(a)
NOTE: Subvectors have element names too)

- a[2] ↦ 2
- a[-2] ↦ 1 3 4 5 6 7 8 9 0
- a[c(1,3,4)] ↦ 1 3 4
- a[1:4] ↦ 1 2 3 4
- a[-c(1,3,4)] ↦ 2 5 6 7 8 9 0
- a[c(T,F,T,rep(F, 7))] ↦ 1 3
- a[c(T,F)] ↦ 1 3 5 7 9 (NOTE: Index vector is recycled)
- a["a"] ↦ 1
- a[c("a","b")] ↦ 1 2

Matrix (2D Array) Indexing

NOTE: a<-matrix(1:25, nrow=5) # Could use ncol too!
NOTE: rownames(a) <- letters[1:5]; colnames(a) <- letters[22:26]
NOTE: subarrays/vectors will have row/col names too

- a[2,] ↦ Row 2 — a vector
- a[,2] ↦ Column 2 — a vector
- a['b',] ↦ Row 2 — a vector
- a[,'w'] ↦ Column 2 — a vector
- a[,-2] ↦ All BUT column 2 — a matrix
- a[-2,] ↦ All BUT row 2 — a matrix
- a[c(1,3,4),] ↦ Row 1, 3, & 4 — a matrix
- a[,c(1,3,4)] ↦ Column 1, 3, & 4 — a matrix
- a[-c(1,3,4),] ↦ All BUT Row 1, 3, & 4 — a matrix
- a[,-c(1,3,4)] ↦ All BUT Column 1, 3, & 4 — a matrix
- a[1:6] ↦ 1 2 3 4 5 6 — a vector

  NOTE: Indexed as a column order vector

List Indexing

NOTE: a<-list(1:10); names(a)<-letters(1:10);
NOTE: All lists/vectors have element names

- a[[2]] ↦ 2 — a vector
- a[["b"]] ↦ 2 — a vector
- a$b ↦ 2 — a vector
- a$"b" ↦ 2 — a vector
- a[-2] ↦ 1 3 4 5 6 7 8 9 0 — a list
- a[c(1,3,4)] ↦ 1 3 4 — a list
- a[c(-1,-3,-4)] ↦ 2 5 6 7 8 9 0 — a list
- a[c(T,F,T,rep(F, 7))] ↦ 1 3 — a list
- a["a"] ↦ 1 — a list
- a[c("a","b")] ↦ 1 2 — a list

data.frames

NOTE: data.frames are basically a list of (column) vectors.

- Indexing may be done vector-style, list-style, or matrix-style
  - list-style: Requested column (vectors or factors) is returned
  - matrix: new sub-data frame is returned just like with a matrix
  - vector: new data frame with the given columns is returned
- Applying many functions, like sum or mean, to a data.frame will result in the functions being applied to each column.

Formulas

NOTE: The symbols in formulas can be the names of global variables or the names of columns in a given data.frame. The interpretation of formulas is context sensitive.

- General form to represent variable dependency relations: dependent variable ~ independent variable.
  Examples: plot(y~x, data=aFrame) <=> plot(aFrame$x, aFrame$y)
  lm(y~x, data=aFrame)

Factors

- Indexed vector-style, a sub-factor is returned with the given elements
  - The levels are the SAME even if the sub-factor doesn't need them all
- as.numeric(f) returns the numbers that code to each level
- as.vector(f) returns the original data as a character vector
  - Follow with as.numeric, as.POSIXct, etc... to transform type

Apply

- lapply/sapply

Apply function to ONE list/vector

- mapply — Apply function to SEVERAL vectors & lists
  Ex: mapply(sum, 1:4, 4:1) ↦ c(5,5,5,5)
- apply — Apply function to array slices
  Ex: apply(b, 1, sum) ↦ rowSums(b)
  Ex: apply(b, 2, sum) ↦ colSums(b)

Ragged Apply

- tapply — Apply function to 'ragged' array
  Breaks array into factor level groups, and applys function to groups
  Ex: tapply(dage, dsex, sum) # Total wt by sex
  Ex: tapply(dage, dsex, length) # Number of M/F in sample
- split — Break object into list of groups
- by — tapply for data.frames
- aggregate — Apply function to groups in data.frame

Time & Date

- as.POSIXct — Convert to UNIX-like time integer
  Use 'origin' when converting integers
  Ex: as.POSIXct(UNIX_DATE, origin='1970-01-01')
- as.POSIXlt — Convert to UNIX-like time components
  sec,min,hour,mday,mon,year,wday,yday,isdst
- strftime — Convert time types to strings.
- strptime — Convert strings into valid POSIXlt

  Format strings (strftime & strptime):
  - %a Abbreviated weekday name.
  - %A Full weekday name
  - %b Abbreviated month name.
  - %B Full month name.
  - %d Day of the month (01-31).
  - %j Day of year (001-366).
  - %H Hours (00-23).
  - %I Hours (01-12).
  - %y Year (00-99).
  - %Y Year with century.
  - %M Minute (00-59).
  - %S Second (00-61)
  - %p AM/PM indicator in the locale
  - %m Month (01-12).
  - %U Week of the year (00-53) first Sunday is day 1 of week 1.
  - %w Weekday as number (0-6, Sunday is 0).
  - %W Week of the year (00-53) first Monday is day 1 of week 1.
- Convert integers (UNIX time) to POSIXct with the following: class(anInt) <- Q "POSIXct"

Basic Stats

- max/min — Min/Max of vector/list/factor
- mean — Can trim
- weighted.mean — Can't trim, but has weights
- median — 50% quantile
- fivenum — 0%, 25%, 50%, 70%, 100% quantiles
- quantile — more flexible than fivenum
- summary — min, 1st Qu, Median, Mean, 3rd Qu, Max
- sum/prod — Arithmetic sum/product
- var/sd — Variance/Standard deviation
- cov/cor/cov.wt — Covariance/Correlation/Weighted covariance
- cumsum/cumprod — Cumulative sums/products
- cummax/cummin — Cumulative max/min
- mad — median absolute deviation
- IQR — interquartile range

Classical Tests

NOTE: See library(help=ctest) for a complete list

- ansari.test
- bartlett.test
- binom.test
- chisq.test
- cor.test
- fisher.test
- fligner.test
- friedman.test
- kruskal.test
- ks.test
- mantelhaen.test
- mcnemar.test
- mood.test
- oneway.test
- pairwise.prop.test
- pairwise.t.test
- pairwise.table
- pairwise.wilcox.test
- power.anova.test
- power.prop.test
- power.t.test
- print.power.htest
- prop.test
- prop.trend.test
- quade.test
- shapiro.test
- t.test
- var.test
- wilcox.test

Data Smoothing

- lowess — Cleveland lowess smoothing
- runmed — Moving median
- smooth — Tukey's smoothers
- smooth.spline — Fits a cubic smoothing spline to the supplied data.
- lowess — Cleveland's LOWESS smoother
- loess — Formula based version of lowess
- supsmu — Friedman's SuperSmoother
- convolve — General convolution
  Ex: convolve(data, rep(1,5), type='filter') ↦ 5pt moving average

Regression

- Compute Regression Curves
  - line — ROBUST regression line (EDA style)
  - lm — Linear regression
    Ex: lm(y~x, data=data.frame(x=1:10, y=2*(1:10)+3+rnorm(10)/5))
  - nls — Nonlinear Least Squares
    Ex: nls(y~a*x^2+b, data=data.frame(x=1:10, y=2*(1:10)^2+3+rnorm(10)/5), start=list(a=10, b=10), trace=TRUE)
- Fitting objects
  - summary — Summary of the fit (very useful)
  - resid — residuals from a line object
  - fitted — fitted points from a line object
  - coef — fitted coefficients
  - deviance — residual sum of squares
  - anova — sequential analysis of variance
  - predict — predict new values
  - plot — A sequence of diagnostic plots

Sub Samples

- sample — Sample n times from x without replacement
  - replace — sample with replacement
  - prob=vec — A vector of probabilities

Probability Distributions

- Using Distributions
  - d<dist> is density
  - p<dist> is distribution
  - r<dist> random number
  - q<dist> is the quantile
- Available Distributions

beta	beta	shape1, shape2
binomial	binom	size, prob
Cauchy	cauchy	location, scale
Chi-squared	chisq	df
exponential	f	rate
gamma	gamma	shape, rate
geometric	geom	prob
hypergeometric	hyper	m, n, k
log-normal	lnorm	meanlog, sdlong
logistic	logis	location, scale
negative binomial	nbinom	size, prob
normal	norm	mean, sd
Poisson	pois	lambda
T	t	df
uniform	unif	min, max
Weibull	weibull	shape, scale
Wilcoxon	wilcox	m, n

I/O

- read.table/read.csv Read in a table/CSV file into a data.frame
  - header — Header line (Boolean)
  - sep — Separator (string)
  - row.names — Vector of row names (strings)
  - col.names — Vector of col names (strings)
  - colClasses — Vector of col types (strings)
  - nrows — Number of rows (integer)
  - skip — Number of lines to skip (integer)
  - blank.lines.skip — Skip blank lines (Boolean)
  - comment.char — Skip comments (# for read.table)
- write — Write vectors and arrays
- source — Read in R code and evaluate it
- scan — Read in data in very flexible manner
- download.file — Download a file from the net
- sink — Send output to a file

R Options (options command)

- echo — controls whether input is echoed in non-interactive mode
- verbose — Report extra information on progress
- device — Default graphics device (generally x11)
- na.action — Default function to handle NAs
- width — controls the number of characters on a line.
- editor — sets the default text editor
- pager — Program used for displaying ASCII files
- browser — default HTML browser used by help.start()

High Level Plot Functions

Lattace	Traditional	Description
barchart	barplot	Barchart
bwplot	boxplot	Box-and-Wisker
densityplot	NA	Kernel Density
dotplot	dotchart	Dotplot
histogram	hist	Histograms
qqmath	qqnrom	Quantile-quantile vs. distribution
stripplot	stripchart	1d scatterplot
qq	qqplot	Quantile-quantile plot
xyplot	plot	Scatterplot
levelplot	image	level plots
contourplot	contour	Contour Plot
cloud	NA	3d Scatterplot
wireframe	persp	surface in 3d
splom	pairs	Scatterplot matrix
parallel	NA	Parallel coord plots

Traditional Graphics

- Draw an object in the plot region
  - abline, polygon, rect, text
    Ex: abline(b, m) ↦ Draw: y=m*x+b
    Ex: abline(h=b) ↦ Draw: y=b
    Ex: abline(v=a) ↦ Draw: x=a
- Draw several objects in the plot region
  - segments, arrows, points
- Draw objects in the margin space
  - axis, mtext
    Ex: mtext("hello", side=2, line=1)

Arithmetic Operators

- + - * — What you expect
- ^ — Exponent
- %% — Integer Remainder (mod)
- / — Floating point division
- %/% — Integer division
- %*% — Matrix multiplication

Strings

- sub, gsub — Pattern substitueion
  Ex: sub('1', 'a', '212121') ↦ "2a2121"
  Ex: gsub('1', 'a', '212121') ↦ "2a2a2a"
  Ex: gsub('.', 'a', '212121') ↦ "aaaaaa"
- substr — SUB-STRings
  Ex: substr('123456', 2, 3) ↦ "23"
- nchar — Chracter length
  Ex: nchar(c('1234', '123')) ↦ c(4,3)
- paste, strsplit — Split & join
  Ex: strsplit("a,b,cd,e;f", ',') ↦ c("a","b","cd","e;f")
  Ex: strsplit("a,b,cd,e;f", '[,;]') ↦ c("a","b","cd","e","f")
  Ex: paste("a", "b", sep='-') ↦ "a-b"
- tolower, toupper — Case conversion
- chartr — Like UNIX tr
  Ex: chartr('abc', 'ABC', 'aAbBcC') ↦ "AABBCC"
- grep, grepl — Searching for patterns
  Ex: grep("a.", c("bac", "a", "aa")) ↦ c(1, 3)
  Ex: grepl("a.", c("bac", "a", "aa")) ↦ c(T, F, T)

Graphics Devices

- Vector Files: postscript, pdf, cairo_pdf, cairo_ps
  Ex: pdf(file='foo.pdf', width=11, height=8.5) ↦ letter sized
- Raster Files: png, jpeg, tiff
  Ex: png(filename='foo.png', width=1024, height=768) ↦ XVGA sized
- Interactive: X11, quartz
  Ex: X11(display='foo.bar.com:0.0') ↦ Default size
- Generic Commands
  - dev.cur() — Current device
  - dev.off(which=dev.cur()) — Close current device
    Must close to write files