list of some useful R functions

Charles DiMaggio

February 27, 2013

1 help

- *help()* opens help page (same as *?topic*)
- *apropos()*displays all objects matching topic (same as *??topic*)
- *library(help=packageName)* help on a specific package
- *example()* ; *demo()*
- *vignette(package="packageName")*; *vignette(package="topic")*
- RSiteSearch("packageName")
- ?NA - handling missing data values
- args() - arguments for a function
- *functionName* - just writing the name of the function returns the function source code
- help with math:
  - ?Control - Help on control flow statements (e.g. if, for, while)
  - ?Extract - Help on operators acting to extract or replace subsets of vectors
  - ?Logic - Help on logical operators
  - ?regex - Help on regular expressions used in R
  - ?Syntax - Help on R syntax and giving the precedence of operators

2 General

- append() - add elements to a vector
- cbind() - Combine vectors by row/column
- grep() - regular expressions

- identical() - test if 2 objects are exactly equal
- length() - no. of elements in vector
- ls() - list objects in current environment
- range(x) - minimum and maximum
- rep(x,n) - repeat the number x, n times
- rev(x) - elements of x in reverse order
- seq(x,y,n) - sequence (x to y, spaced by n)
- sort(x) - sort the vector x
- order(x) - list the sorted *element numbers* of x
- tolower(),toupper() - Convert string to lower/upper case letters
- unique(x) - remove duplicate entries from vector
- round(x), signif(x), trunc(x) - rounding functions
- getwd() - return working directory
- setwd() - set working directory
- choose.files() - get path to a file (useful for virtual machines)
- month.abb/month.name - abbreviated and full names for months
- pi,letters,(e.g. letters[7] = "g") LETTERS

3 Math

- sqrt(),sum()
- log(x),log10(),,exp(),sqrt()
- cos(),sin(),tan(),
- %% modulus
- %/% integer divisilln
- %*% matrix multiplication
- %o% outer product (a%o% equivalent to outer(a,b,"*"))
- union(),intersect(),setdiff(),setequal() - set operations
- eigen() - eigenvalues and eigenvectors
- deriv() - symbolic and algorithmic derivatives of simple expressions

- integrate() - adaptive quadrature over a finite or infinite interval.

4 Plotting

- plot() - generic R object plotting
- par() - set or query graphical parameters
- curve(equation,add=T) - plot an equation as a curve
- points(x,y) - add additional set of points to an existing graph
- arrows() - draw arrows
- abline() - ddd a straight line to an existing graph
- lines() - join specified points with line segments
- segments() - draw line segments between pairs of points
- hist() - histogram
- pairs() - plot matrix of scatter plots
- matplot() - plot columns of matrices
- persp() - perspective plot
- contour() - contour plot
- image() - plot an image file
- loess(), lowess() - scatter plot smoothing
- splinefun() - spline interpolation
- smooth.spline() - Fits a cubic smoothing spline
- jitter() - Add a small amount of noise to a numeric vector
- pdf()/ png() / jpeg() - send plot to .pdf / .png / .jpeg file

5 Statistics

- help(package=stats) - list all stats functions
- lm - fit linear model
- glm - fit generalized linear model
- cor.test() - correlation test

- cumsum() cumprod() - cumuluative functions for vectors
- density(x) - kernel density estimates
- ks.test() - one or two sample Kolmogorov-Smirnov tests
- mean(x), weighted.mean(x), median(x), min(x), max(x), quantile(x)
- rnorm(), runif() - generate random data with Gaussian/uniform distribution
- sd() - standard deviation
- summary(x) - a summary of x (mean, min, max)
- t.test() - Student's t-test
- var() - variance
- sample() - random samples
- qqplot() - quantile-quantile plot

6 regression

(Functions in italics, packages in quotation marks.)

- Linear models
  - *aov* ("stats"), *Anova()* ("car"): ANOVA models
  - *coef*: extract model coefficients ("stats")
  - *confint*: Computes confidence intervals for one or more parameters in a fitted model. ("stats")
  - *fitted*: extracts fitted values ("stats")
  - *lm*: fit linear models. ("stats")
  - *model.matrix*: creates a design matrix ("stats")
  - *predict*: predicted values based on linear model object ("stats")
  - *residuals*: extracts model residuals ("stats")
  - *summary* summary method for class "lm" (stats)
  - *vcov*: variance-covariance matrix of the main parameters of a fitted model object ("stats")
  - *AIC*: Akaike information criterion for one or several fitted model objects ("stats")
  - *extractAIC*: Computes the (generalized) Akaike An Information Criterion for a fitted parametric model ("stats")

  - *offset*: An offset is a term to be added to a linear predictor, such as in a generalised linear model

- Generalized Linear Models (GLM)
  - *glm*: is used to fit generalized linear models ("stats")

"family=" specify the details of the models used by *glm* ("stats")

  - *glm.nb*: fit a negative binomial generalized linear model ("MASS")

- Diagnostics
  - *cookd*: cook's distances for linear and generalized linear models ("car") "cooks.distance": Cooks distance ("stats")
  - *influence.measures*: suite of functions to compute regression (leave-one-out deletion) diagnostics for linear and generalized linear models ("stats")
  - *lm.influence*: provides the basic quantities used in diagnostics for checking the quality of regression fits ("stats")
  - *outlier.test*: Bonferroni outlier test ("car")
  - *rstandard*: standardized residuals ("stats")
  - *rstudent*: studentized residuals ("stats")
  - *vif*: variance inflation factor ("car")

- Graphics
  - *influence.plot*: regression influence plot ("car")
  - *leverage.plots*: regression leverage plots ("car")
  - *plot*: four residual plots ("stats")
  - *qq.plot*: quantile-comparison plots ("car")
  - *qqline*: adds a line to a normal quantile-quantile plot which passes through the first and third quartiles ("stats")
  - *qqnorm*: normal QQ plot of the values in y ("stats")
  - *reg.line*: plot regression line ("car")
  - *scatterplot*: scatterplots with boxplots ("car")

- Tests and Transformations
  - *durbin.watson*: Durbin-Watson Test for autocorrelated errors ("car")
  - *dwtest*: Durbin-Watson test ("lmtest")
  - *levene.test*: Levene's test ("car")
  - *lillie.test*: Lilliefors (Kolmogorov-Smirnov) test for normality ("nortest")

- *pearson.test*: Pearson chi-square test for normality ("nortest")
- *box.cox*: Box-Cox family of transformations ("car")
- *boxcox*: Box-Cox transformations for linear models ("MASS")

- Survival analysis
  - *anova.survreg*: ANOVA tables for survreg objects ("survival")
  - *clogit*: Conditional logistic regression ("survival")
  - *cox.zph*: Test the proportional hazards assumption of a Cox regression ("survival")
  - *coxph*: proportional hazards regression ("survival")
  - *coxph.detail*: details of a Cox model fit ("survival")
  - *coxph.rvar*: robust variance for a Cox model ("survival")
  - *ridge*: ridge regression ("survival")
  - *survdiff*: test survival curve differences ("survival")
  - *survexp*: compute expected survival ("survival")
  - *survfit*: compute a survival curve for censored data ("survival")
  - *survreg*: regression for a parametric survival model ("survival")

- Linear and nonlinear mixed effects models
  - *ACF*: autocorrelation function ("nlme")
  - *ACF.lme*: autocorrelation Function for lme Residuals ("nlme")
  - *intervals*: confidence intervals on coefficients ("nlme")
  - *intervals.lme*: confidence intervals on lme parameters ("nlme")
  - *lme*: linear mixed-effects models ("nlme")
  - *nlme*: nonlinear mixed-effects models ("nlme")
  - *predict.lme*: predictions from an lme object ("nlme")
  - *predict.nlme*: predictions from an nlme object ("nlme")
  - *qqnorm.lme*: normal plot of residuals or random effects from an lme object ("nlme")
  - *ranef.lme*: extract lme random effects ("nlme")
  - *residuals.lme*: extract lme residuals ("nlme")
  - *simulate.lme*: simulate lme models ("nlme")
  - *summary.lme*: summarize an lme object ("nlme")

- Structural Equation, Principal Components, Partial Least Squares Regression Models
  - *sem*: general structural equation models ("sem")
  - *systemfit*: fits a set of linear structural equations using ordinary least squares
  - *biplot.mvr*: biplots of PLSR and PCR Models ("pls")
  - *coefplot*: plot regression coefficients of plsr and pcr models ("pls")
  - *mvr*: partial least squares and principal components regression ("pls")
  - *scores*: extract scores and loadings from plsr and pcr models ("pls")
- Recursive Partitioning and Regression Trees
  - *cv.tree*: cross-validation for choosing tree complexity ("tree")
  - *deviance.tree*: extract deviance from a tree object ("tree")
  - *labels.rpart*: create split labels for an rpart object ("rpart")
  - *misclass.tree*: misclassifications by a classification tree ("tree")
  - *partition.tree*: plot the partitions of a simple tree model ("tree")
  - *path.rpart*: follow paths to selected nodes of an rpart object (rpart)
  - *plotcp*: plot a complexity parameter table for an rpart fit ("rpart")
  - *printcp*: displays cp table for fitted rpart object ("rpart")
  - *prune.misclass*: cost-complexity pruning of tree by error rate ("tree")
  - *rpart*: recursive partitioning and regression trees ("rpart")
  - *rsq.rpart*: plots the approximate r-square for the different splits ("rpart")
  - *tile.tree*: add class barplots to a classification tree plot ("tree")
  - *tree.control*: select parameters for tree (tree)
  - *tree.screens*: split screen for plotting trees ("tree")
  - *tree*: fit a classification or regression tree ("tree")

This list is based on material posted online by Alastair Sanderson and Vito Ricci.