Useful R commands.

Basic Commands:

\[ x + y, x \cdot y, x^y, x/y \]  
Point-by-point addition, subtraction, multiplication, division of two vectors

If a is a constant,  
a + x, a*x, etc.  
Adds, multiplies, etc. each entry of vector x by the constant a

\[ x[n] \]  
Entry n of vector x

\[ x[-n] \]  
Remove entry n of vector x

\[ x = c(...) \]  
Concatenate listed data into vector x

\[ x = \text{read.csv}("address", \text{header}, \text{sep}) \]  
Read delimited data into vector x  
address = web address or file location or “clipboard”  
header = TRUE if data contains variable names in row 1  
sep is the symbol which separates entries  
(the default is comma, sep = ",")

\[ \text{sep = "/t", sep = " "} \]  
Use tab, empty space as separator

\[ x[m,n] \]  
Entry in row m, column n of data frame x

\[ x[\text{condition, }] \]  
Rows of data frame x satisfying listed condition

\[ \text{data.frame}(...) \]  
Creates a data frame using the listed vectors

\[ \text{table}(x,y) \]  
Creates a two-way table from vectors x and y

\[ x\$... \]  
A selected column of data frame x

\[ \text{mean}(x), \text{median}(x), \text{sum}(x), \text{length}(x) \]  
Mean, median, sum, length of vector x respectively

\[ \text{summary}(x) \]  
Gives 5-number summary of x plus its mean

\[ \text{sd}(x) \]  
Sample standard deviation of x

\[ \text{weighted.mean}(x, w) \]  
Weighted mean of data x, using weighting vector w

\[ \text{rep}(x,n) \]  
Repeats entry x a total of n times

Graphics:

\[ \text{stem}(...) \]  
Stem and leaf plot

\[ \text{boxplot}(...) \]  
Boxplot

\[ \text{hist}(x, \text{breaks}) \]  
Histogram of data x  
breaks defaults to Sturges rule, breaks = n  
creates n breakpoints for binning the data

\[ \text{plot}(x,y) \]  
Basic scatterplot of response variable y versus predictor variable x

Optional graphical arguments:

\[ \text{main = "title"} \]  
Quoted text appears as a title

\[ \text{xlab = "label", ylab = "label"} \]  
Labels x and y axes respectively with quoted text

\[ \text{col = "some color"} \]  
Makes various things in the plot the quoted color.

\[ \text{write.table}(\text{data}, \"\text{clipboard}\", \text{sep="\t"}) \]  
Copies data frame data onto clipboard so that it can be pasted into other applications.

Regression:

\[ \text{model = lm}(y \sim x_{1}+x_{2}+..., \text{data}) \]  
Creates linear model named “model”, with response variable y, predictor variables x1, x2,   
..., data is the data frame used

\[ \text{summary}(\text{model}) \]  
Provides information about linear model “model”, including regression coefficients, the   
coefficient of determination, and residuals.

\[ \text{cor}(y,x) \]  
Correlation coefficient between variables y and x.

\[ \text{predict}(\text{model, list(\ldots)}) \]  
Gives a point estimate of the value of the response variable in the linear model “model”,   
list(\ldots) specifies the values desired for the predictor variable(s).

\[ \text{predict}(\text{model, list(\ldots), interval, level}) \]  
Creates a prediction interval if interval=“predict”, with confidence level specified.
Probability distributions:

- `dbinom(v, size, prob)`
  Probability density of binomial distribution at point(s) given in vector v, size= # of trials, prob=success probability
- `pbinom(v, size, prob, lower.tail)`
  Tail (cumulative) probabilities in binomial distribution, lower.tail defaults to TRUE, lower.tail = FALSE gives upper tail probabilities, ignoring the point(s) in v.
- `rbinom(v, size, prob)`
  Simulate indicated number of binomial trials. Entries in v give the number of simulations desired.
- `dpois(v, lambda)`
  Similar commands for Poisson distribution.
- `ppois(v, lambda, lower.tail)`
- `rpois(v, lambda)`

- `pnorm(v, mean, sd, lower.tail)`
  Similar commands for normal distribution.
- `rnorm(v, mean, sd)`
- `qnorm(v, mean, sd)`
  Returns percentile(s) in v in the indicated normal distribution.
- `pchisq(v, df, lower.tail)`
  Similar commands for chi-square distribution
- `rchisq(v, df)`

Matrices:

- `matrix(v, nrow, ncol, byrow)`
  Creates a matrix from data v with indicated number of rows/columns. byrow=TRUE fills the matrix one row at a time, byrow=FALSE fills the matrix column by column.
- `M %*% N`
  Matrix multiplication of M and N
- `M %^% n`
  n-th power of square matrix M (with expm library loaded).
- `eigen(M)`
  Computes eigenvalues and corresponding eigenvectors of matrix M
- `eigen(M)$vectors[,n]`
  Selects indicated eigenvector of M.

Sampling:

- `sample(v, n, prob, replace)`
  Draws a random sample of n from objects in vector v. prob is a vector of probabilities that give the probability of drawing each member of v. If prob is missing, all entries in v are assumed equally likely. replace=TRUE samples with replacement.

Hypothesis Testing:

- `prop.test(x, n, p, alternative)`
  Compares a sample of x successes in n trials versus baseline (null hypothesis) p, alternative is either “less” (than null hypothesis expects), “greater”, or “two.sided”
- `t.test(data, mu, alternative)`
  Carries out one-sample t-test comparing data vs. baseline mean mu.
- `t.test(data1, data2, mu, alternative)`
  Compares two samples of data (data1 and data2), mu is expected difference in means, alternative = “less” uses mean1 < mean2 for the alternate, ...
- `chisq.test(counts, p)`
  Carries out chi-squared goodness-of-fit test. counts is a vector of data broken into categories, p is expected distribution of probabilities.

Analysis of variance:

- `aov(y~x, data)`
  Carries out (one-way) ANOVA to determine if significant differences in response variable y exist among the various groups given in vector x.
- `plot(y~x, data)`
  Produces boxplots of response variable y in each of the treatment groups
- `pairwise.t.test(y,x)`
  Gives the result of (post hoc) t-tests to detect differences in y among every possible pair of groups in x.

Miscellaneous:

- `help(any R command name)`
  Opens a window in a web browser with information about using the indicated command.
  Depending on the command, it could be kind of technical though, so don’t rely on this too much...