Using R for Statistics and Data Analysis

R is a versatile and powerful programming environment for statistics and data analysis, but it is far from user-friendly for the novice. First, you should install R on your computer. Versions for Windows, MacOS X, and Linux are available at http://cran.cnr.berkeley.edu/.

In these guides, R code will be shown as `red text`, example R output as `blue text`, and comments as `green text`, preceded by the R comment symbol `#`.

This isn't intended to be an exhaustive guide to R (if such a thing is even possible). The topics are chosen to address the material covered in my class, EART125 Statistics and Data Analysis in the Geosciences.

Topics:

R Basics: Data and Data Manipulation, Variables and Functions

- Importing Data and Data Entry
- Variables in R
- Manipulating Data
- Data Analysis Functions apply(), split() and sapply(), by()
- Graphing
- Writing Your Own Functions
- Installing and Loading Add-on Packages

Statistics in R

- Descriptive Statistics: mean, median, variance, standard deviation
- Univariate, Parametric Statistics: t-test, F test, ANOVA, Bartlett test
- Tests for Normality: Q-Q plots, Shapiro-Wilk test, Kolmogorov-Smirnov test
- Univariate, Non-Parametric Statistics: Wilcoxon rank-sum/Mann-Whitney U test, Kruskal-Wallis test, Levene test/Brown-Forsythe test
- Multivariate, Parametric Statistics: Hotelling T2 test, MANOVA
- Statistics for Categorical/Count Data: Exact binomial test, exact multinomial test, G-test, Fisher's exact test
- Correlation and Regression: Pearson r, Spearman rho, and Kendall tau, linear and logistic regression, autocorrelation and differencing
- Resampling Methods: sample, replicate, quantile, sort, order functions and for loops

Data Analysis in R

- Ordination: PCA, CA/DCA, PCoA, NMDS
- Cluster Analysis: distance matrices and cluster dendrograms
- Maximum Likelihood Estimation: parameter estimation and model comparison

Useful Functions for EART125 (copy and paste in the R window)

G-test for goodness-of-fit or independence

Levene test/Brown-Forsythe test

Reduced major axis (RMA) and Major axis (MA) regression