RESOURCES for ENTM798V: Introduction to R for computation and analysis in ecology & evolutionary biology

Getting Started:

First, download the program for the platform you use:

From the R home page <http://www.r-project.org/>

Navigate to the R CRAN download page <http://cran.r-project.org/mirrors.html>

Find local mirror site (e.g. National Cancer Institute, Bethesda MD)
<http://watson.nci.nih.gov/cran_mirror/>

Download most recent compiled binary base package (to date, R v2.13.0) for:
windows* <http://watson.nci.nih.gov/cran_mirror/bin/windows/base/R-2.14.0-win.exe>
or mac os <http://watson.nci.nih.gov/cran_mirror/bin/macosx/R-2.14.0.pkg>
or linux <http://watson.nci.nih.gov/cran_mirror/> (navigate to ubuntu, red hat etc.)

Follow the menu options to install.

*For *Windows Vista or 7*, to preempt hassles from additional default security 'features' you may want to "Install as Administrator" and change the shortcut default to "Run as Administrator" (see R for Windows FAQ).

Also see:

Manuals, help tools, and tutorials from the folks at CRAN:
http://cran.r-project.org/manuals.html

References, Resources and Other Notables:

BOOKS <http://www.r-project.org/doc/bib/R-books.html>
- This is an incomplete list! More books are coming out all the time.
- Many are in stacks and some are on course reserve at McKeldin library, some are available online through UMCP library system [entire *Use R! series* and other Springer texts WERE online through 2011 - complain to your librarian if you want them back!]
  * indicates recommended resource (I have not consulted many of those lacking an asterisk)

Aitkin, M., B. Francis, J. Hinde, and R. Darnell. 2009. Statistical Modelling in R. Oxford University Press, Oxford, UK.

Albert, J. 2007. Bayesian Computation with R. Springer, New York. *(Use R! series)*

Albert, J., and M. Rizzo 2012. R by Example: Concepts to Code. Springer, New York. *(Use R! series)*

Andersen, P. K., and L. T. Skovgaard. 2008. Regression with Linear Predictors. Springer, New York. *(Statistics for Biology and Health series)*

*Beckerman, A. P. and O. L. Petchey. 2012. Getting Started with R: An Introduction for Biologists. Oxford University Press, Oxford, UK. <http://r4all.org/>

*Bivand, R. S., E. J. Pebesma, and V. Gómez-Rubio. 2008. Applied Spatial Data Analysis with R. Springer, New York. *(Use R! series)*

*Bolker, B. M. 2008. Ecological Models and Data in R. Princeton University Press, Princeton, NJ.

Borcard, D., F. Gillet, and P. Legendre. 2011. Numerical Ecology with R. Springer, New York. *(Use R! series)*

Braun, W. J., and D. J. Murdoch. 2009. A First Course in Statistical Programming with R. Cambridge University Press, Cambridge, UK.

Broman, K. W., and S. Sen. 2009. A Guide to QTL Mapping with R/qtl. Springer, New York. *(Statistics for Biology and Health series)*

Chambers, J. M. 2009. Software for Data Analysis. Springer, New York. *(Statistics and Computing series)*

Cook, D., and D. Swayne. 2007. Interactive and Dynamic Graphics for Data Analysis. Springer, New York. *(Use R! series)*

Cowpertwait, P. S. P., and A. V. Metcalfe. 2009. Introductory Time Series with R. Springer, New York. *(Use R! series)*

Crawley, M. J. 2005. Statistics: an Introduction using R. John Wiley & Sons, New York.

*Crawley, M. J. 2007. The R Book. John Wiley & Sons, Ltd, West Sussex, England.

*Dalgaard, P. 2008. Introductory Statistics with R. Springer, New York. *(Statistics and Computing series)*

Everitt, B. S., and T. Hothorn. 2006. A Handbook of Statistical Analyses Using R. Chapman & Hall/CRC, Boca Raton, Fl.

Everitt, B., and T. Hothorn. 2011. An Introduction to Applied Multivariate Analysis with R. Springer, New York. *(Use R! series)*

Ewens, W. J., and G. Grant. 2005. Statistical Methods in Bioinformatics: an Introduction. Springer, New York. *(Statistics for Biology and Health series)*

*Faraway, J. 2005. Linear Models with R. Taylor & Francis Group, Boca Raton, FL.

Faraway, J. 2006. Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. Taylor & Francis Group, Boca Raton, FL.

Foulkes, A. S. 2009. Applied Statistical Genetics with R. Springer, New York. *(Use R! series)*

Gentleman, R. 2008. R Programming for Bioinformatics. Chapman and Hall, Boca Raton, FL.

Gentleman, R., V. J. Carey, W. Huber, R. A. Irizarry, and S. Dudoit. 2005. Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Springer, New York. *(Statistics for Biology and Health series)*

Heiberger, R. M., and E. Neuwirth. 2009. R Through Excel: a Spreadsheet Interface for Statistics, Data Analysis, and Graphics. Springer, New York. *(Use R! series)*

Horgan, J. 2008. Probability with R: an Introduction with Computer Science Applications. John Wiley & Sons, New York.

Jones, O., R. Maillardet, and A. Robinson. 2009. Introduction to Scientific Programming and Simulation Using R. Chapman & Hall/CRC, Boca Raton, FL.

*Kéry, M. 2010. Introduction to WinBUGS for Ecologists: a Bayesian Approach to Regression, ANOVA, Mixed Models and Related Analyses. Academic Press, New York.

*Logan, M. 2010. Biostatistical Design and Analysis Using R: a Practical Guide. Wiley-Blackwell, Oxford, UK.

*Maindonald, J., and W. J. Braun. 2010. Data Analysis and Graphics Using R, 3rd edition. Cambridge University Press, Cambridge, UK.

Muenchen, R. A. 2009. R for SAS and SPSS Users. Springer, New York. *(Use R! series)*

*Murrell, P. 2005. R Graphics. Chapman & Hall/CRC, Boca Raton, FL.

*Paradis, E. 2006. Analysis of Phylogenetics and Evolution with R. Springer, New York. *(Use R! series)* [new edition in 2012]

Peng, R. D., and F. Dominici. 2008. Statistical Methods for Environmental Epidemiology with R. Springer, New York. *(Use R! series)*

Petris, G., S. Petrone, and P. Campagnoli. 2009. Dynamic Linear Models with R. Springer, New York. *(Use R! series)*

*Pinheiro, J. C., and D. M. Bates. 2000. Mixed-Effects Models in S and S-PLUS. Springer, New York. *(Statistics and Computing series)*

Ramsay, J., G. Hooker, and S. Graves. 2009. Functional Data Analysis with R and MATLAB. Springer, New York. *(Use R! series)*

Reimann, C., P. Filzmoser, R. Garrett, and R. Dutter. 2008. Statistical Data Analysis Explained: Applied Environmental Statistics With R. John Wiley & Sons, New York.

Ritz, C., and J. C. Streibig. 2009. Nonlinear Regression with R. Springer, New York. *(Use R! series)*

Rizzo, M. L. 2007. Statistical Computing with R. Chapman & Hall/CRC, Boca Raton, FL.

Robert, C. P., and G. Casella. 2010. Introducing Monte Carlo Methods with R. Springer, New York. *(Use R! series)*

*Sarkar, D. 2008. Lattice: Multivariate Data Visualization with R. Springer, New York. *(Use R! series)*

Sawitzki, G. 2009. Computational Statistics: an Introduction to R. Chapman and Hall, Boca Raton, FL.

Shahbaba, B. 2012. Biostatistics with R: An Introduction to Statistics Through Biological Data. Springer, New York. *(Use R! series)*

Soetaert, K., and P. M. J. Herman. 2009. A Practical Guide to Ecological Modelling: Using R as a Simulation Platform. Springer, New York.

*Spector, P. 2008. Data Manipulation with R. Springer, New York. *(Use R! series)*

*Stevens, M. H. 2009. A Primer of Ecology with R. Springer, New York. *(Use R! series)*

Suess, E. A., and B. E. Trumbo. 2010. Introduction to Probability Simulation and Gibbs Sampling with R. Springer, New York. *(Use R! series)*

Ugarte, M. D., A. F. Militino, and A. Arnholt. 2008. Probability and Statistics with R. Chapman & Hall/CRC, Boca Raton, FL.

Venables, W. N., and B. D. Ripley. 2002. Modern Applied Statistics with S, 4th edition. Springer, New York. *(Statistics and Computing series)*

Verzani, J. 2005. Using R for Introductory Statistics. Chapman & Hall/CRC, Boca Raton, FL.

Wickham, H. 2009. ggplot2: Elegant Graphics for Data Analysis. Springer, New York. *(Use R! series)*

*Zuur, A. F., E. N. Ieno, and E. Meesters. 2009a. A Beginner's Guide to R. Springer, New York. *(Use R! series)*

Zuur, A. F., E. N. Ieno, and G. M. Smith. 2007. Analysing Ecological Data. Springer, New York. *(Statistics for Biology and Health series)*

*Zuur, A. F., E. N. Ieno, N. J. Walker, A. A. Saveliev, and G. M. Smith. 2009b. Mixed Effects Models and Extensions in Ecology with R. Springer, New York. *(Statistics for Biology and Health series)*

R News, a newsletter with announcements on new packages and tools
http://cran.r-project.org/doc/Rnews/

 "The newsletter as a medium for communication intends to fill the gap between the R mailing lists and scientific journals: Compared with emails it is more persistent, one can cite articles in the newsletter and because the newsletter is edited it has better quality control. On the other hand, when compared to scientific journals, it is faster, less formal, and last but not least focused on R."

Examples of articles in _R News_:

> Bates, D. 2005. Fitting linear mixed models in R. _R News_ **5**:27-30.
> Bretz, F., T. Hothorn, and P. Westfall. 2002. On multiple comparisons in R. _R News_ **2**:14-17.
> Canty, A. J. 2002. Resampling methods in R: the boot package. _R News_ **2**:2-7.
> Dormann, C. F., B. Gruber, and J. Fründ. 2008. Introducing the bipartite package: analysing ecological networks. _R News_ **8**:8-11.
> Guha, R. 2006. Generating, using and visualizing molecular information in R. _R News_ **6**:28-33.
> Hankin, R. K. S. 2007. Very large numbers in R: introducing package Brobdingnag. _R News_ **7**:15-16.
> Ligges, U., and J. Fox. 2008. How can I avoid this loop or make it faster? _R News_ **8**:46-50.
> Pebesma, E. J., and R. S. Bivand. 2005. Classes and methods for spatial data in R. _R News_ **5**:9-13.
> Schwartz, M. 2003. An introduction to using R's base graphics. _R News_ **3**:2-6.
> Zeileis, A., and T. Hothorn. 2002. Diagnostic checking in regression relationships. _R News_ **2**:7-10.

The R Journal supersedes _R News_ (2008) http://journal.r-project.org/

The R Journal is the refereed journal of the R project for statistical computing. It features short to medium length articles covering topics that might be of interest to users or developers of R, including:

- Add-on packages: short introductions to or reviews of R extension packages.
- Programmer's Niche: hints for programming in R (or S).
- Hints for newcomers: Explaining aspects of R that might not be so obvious from reading the manuals and FAQs.
- Applications: examples of analyzing data with R.

Examples of articles in _The R Journal_:

> Graves, S., S. Dorai-Raj, and R. François. 2009. sos: searching help pages of R packages. _The R Journal_ **1**:56-59.
> Murrell, P. 2009. Drawing diagrams with R. _The R Journal_ **1**:15-21.

Poisot, T. 2011. The digitize package: extracting numerical data from scatterplots. *The R Journal* **3**:25-26.

Webpages, Blogs, and Online Resources

Agglomeration of R blogs: http://www.r-bloggers.com/ (this is a highly recommended aggregation site for news, tips, new packages, etc)

"R-Seek" – a specialized Google-esque search for R functions, examples etc. http://www.rseek.org/

Quick-R, designed especially for users of SAS, SPSS, Stata http://www.statmethods.net/index.html

R programming for those coming from other languages http://www.johndcook.com/R_language_for_programmers.html

Learning R blog: http://learnr.wordpress.com/

The R-wiki project: tips and tricks in wiki format: http://wiki.r-project.org/rwiki/doku.php

An Introduction to R: Software for Statistical Modelling & Computing, by Kuhnert, P. and B. Venables (2005). http://CRAN.R-project.org/doc/contrib/Kuhnert+Venables-R_Course_Notes.zip . (360 page PDF document of lecture notes in combination with the data sets and R scripts)

USGS training workshop in R, put together by Paul Geissler and Tom Philippi: http://www.fort.usgs.gov/BRDScience/LearnR.htm. This site contains a wealth of information and links to outside resources, as well as downloadable labs and exercises with accompanying audio files. Notes, slides, data, code and R files can be downloaded directly: ftp://ftpext.usgs.gov/pub/cr/co/fort.collins/Geissler/LearnR/

Ecological Models and Data in R, from Ben Bolker to support his book; contains code and practice exercises from the book, along with draft pdfs of each chapter http://www.zoology.ufl.edu/bolker/emdbook/

Ecology and epidemiology in the R programming environment by Karen Garrett et al.; provides introductory exercises along with more advances routines for programming analyses in the following topics: disease progress over time, modeling dispersal gradients, spatial analysis & disease forecasting http://www.apsnet.org/education/AdvancedPlantPath/Topics/RModules/default.html

A mini-course on using R for statistical analysis, by Dan Stoebel, http://life.bio.sunysb.edu/~dstoebel/R/

R & BioConductor Manual (for bioinformatics) http://manuals.bioinformatics.ucr.edu/home/R_BioCondManual

R Labs for Vegetation Ecologists (from the "Laboratory for Dynamic Synthetic Vegephenomenology"), includes tutorials, lab exercises, and example datasets for a course in quantitative analysis and multivariate statistics in vegetation ecology. http://ecology.msu.montana.edu/labdsv/R/

"Programming in R" by John Fox, author of a number of R packages – a short course towards getting started, with scripts and links http://socserv.socsci.mcmaster.ca/jfox/Courses/R-programming/index.html

"R By Example" by Ajay Shah, http://www.mayin.org/ajayshah/KB/R/index.html

"Looking at Data" – workshop organized by Di Cook & Hadley Wickham – explores the plyr, reshape packages for crunching data, and the ggobi program for visualizing complex data sets http://lookingatdata.com/

"A Quick and (Very) Dirty Intro to Doing Your Statistics in R" by Jarrett Byrnes http://homes.msi.ucsb.edu/~byrnes/rtutorial.html

Information on "Sweave": a utility with base installation that allows the creation of Latex documents embedded with R code scripts and output. "The purpose is to create dynamic reports, which can be updated automatically if data or analysis change. Instead of inserting a prefabricated graph or table into the report, the master document contains the R code necessary to obtain it." http://www.statistik.lmu.de/~leisch/Sweave/

And there is sooo much more……

Help, Forums, Listservs etc.

Reference cards (printable pdfs with most essential commands and shortcuts):
- Tom Short's: http://cran.r-project.org/doc/contrib/Short-refcard.pdf
- Jonathan Baron's: http://cran.r-project.org/doc/contrib/refcard.pdf

General R-Help listserv http://tolstoy.newcastle.edu.au/R/ (people here can be snippy if you do not first make an honest attempt to solve a problem and provide sample code with your question. Read up on etiquette before posting)

Ecology-specific list, more friendly…. https://stat.ethz.ch/mailman/listinfo/r-sig-ecology

…..and many others for interest groups   https://stat.ethz.ch/mailman/listinfo

R forum on Nabble: http://www.nabble.com/R-f13819.html (also find eco-sig on Nabble2: http://n2.nabble.com/r-sig-ecology-f471788.html)

Text Editors and Alternative Work Spaces

The Mac version of R includes a relatively rich suite of features for editing, including line numbers and syntax color codes. The windows version includes just the basics, and users may wish to use an editor that integrates with R – Rstudio and Tinn-R seem to be the most popular…

RStudio™ is a relatively new integrated development environment (IDE) for R that combines an intuitive user interface with powerful coding tools. http://www.rstudio.org/

Tinn-R is a free, simple, stand-alone but efficient replacement for the basic code editor provided by Rgui. Many people find this *very* helpful while learning the R syntax and basic commands http://www.sciviews.org/Tinn-R/

SciViews-R is a series of packages providing a GUI API on top of R, and integrating more completely with R http://www.sciviews.org/SciViews-R/index.html

For linux users, ESS + Emacs provide script editing tools ("Emacs Speaks Statistics: A Multiplatform, Multipackage Development Environment for Statistical Analysis") http://ess.r-project.org/

ESS can be implemented in windows through XEmacs, another open-source text editor http://socserv.mcmaster.ca/jfox/Books/Companion/ESS/ and http://www.xemacs.org/

Packages http://lib.stat.cmu.edu/R/CRAN/web/packages/

The R distribution comes with the following packages (a *package* is a contributed bundle of code and procedures, usually centered on a theme – aside from the core capabilities of the pprogram [often in C++] most R functions and datasets are stored and implemented in *packages*). This is not to be confused with a *library* which is a directory or sub-directory of packages. Only when a package is loaded are its contents available. This is done both for efficiency (the full list would require astounding memory and would take longer to search than a subset), and to aid package developers, who are protected from name clashes with other code. Most of the 'good stuff' can be found in these packages available on CRAN, Bioconductor, R-forge, etc. The best ones are actively managed and updated by their developers – those lacking recent updates are eventually deleted by CRAN.

Type: "library()" or "require()" with the package name inside the parentheses

base
   Base R functions (and datasets before R 2.0.0).
datasets
   Base R datasets (added in R 2.0.0).
grDevices
   Graphics devices for base and grid graphics (added in R 2.0.0).
graphics
   R functions for base graphics.
grid
   A rewrite of the graphics layout capabilities, plus some support for interaction.
methods
   Formally defined methods and classes for R objects, plus other programming tools, as described in the Green Book.
splines
   Regression spline functions and classes.
stats
   R statistical functions.
stats4
   Statistical functions using S4 classes.
tcltk
   Interface and language bindings to Tcl/Tk GUI elements.
tools
   Tools for package development and administration.
utils
   R utility functions.

Frequently used add-on packages (from CRAN; also others from Omegahat, Bioconductor…)

[as of 16-Aug-2012: the CRAN package repository features 3891 packages]

ade4
>Analysis of Ecological Data : Exploratory and Euclidean methods in Environmental sciences (also GUI: **ade4TkGUI**)

AICcmodavg
>Model selection and multimodel inference based on (Q)AIC(c)

BiodiversityR
>GUI for biodiversity and community ecology analysis

bbmle
>Tools for general maximum likelihood estimation

boot
>Bootstrap R (S-Plus) Functions (Canty)

car
>Companion to Applied Regression

ecodist
>Dissimilarity-based functions for ecological analysis

emdbook
>Ecological models and data (Bolker's book support)

geoR
>Analysis of geostatistical data

geoRglm
>a package for generalised linear spatial models

gplots
>Various R programming tools for plotting data

ggplot2
>An implementation of the Grammar of Graphics

lattice
>Lattice Graphics

lavaan
>Latent Variable Analysis

lme4
>Linear mixed-effects models using S4 classes

maanova
>Tools for analyzing Micro Array experiments

maps
>Draw Geographical Maps

matlab
>MATLAB emulation package

MuMIn
>Model selection and model averaging based on information criteria (AICc and alike)

nlme
>Linear and Nonlinear Mixed Effects Models

ouch
>Ornstein-Uhlenbeck models for phylogenetic comparative hypotheses.

PHYLOGR
>   Functions for phylogenetically based statistical analyses

picante
>   R tools for integrating phylogenies and ecology

plotrix
>   Various plotting functions

plyr
>   Tools for splitting, applying and combining data

qgen
>   Quantitative Genetics using R

R2WinBUGS
>   Running WinBUGS and OpenBUGS from R / S-PLUS

Rcmdr
>   A platform-independent basic-statistics GUI (graphical user interface) for R, based on the **tcltk** package.

RcolorBrewer
>   ColorBrewer palettes

RMySQL
>   R interface to the MySQL database

reshape and **reshape2**
>   Flexibly reshape data

scatterplot3d
>   3D Scatter Plot

sem
>   Structural Equation Models

spdep
>   Spatial dependence: weighting schemes, statistics and models

vegan
>   Community Ecology Package

zipcode
>   U.S. ZIP Code database

How do I find the package I need?
CRAN task views – narrative organization of packages around conceptual themes
http://cran.r-project.org/web/views/index.html

- **Bayesian** (Bayesian Inference)
  http://lib.stat.cmu.edu/R/CRAN/web/views/Bayesian.html

- **Environmetrics** (Analysis of ecological and environmental data)
  http://cran.r-project.org/web/views/Environmetrics.html

- **Graphics** (Graphic displays & dynamic graphics & graphicdevices & visualization)
  http://lib.stat.cmu.edu/R/CRAN/web/views/Graphics.html

- **Multivariate** (Multivariate statistics)
  http://lib.stat.cmu.edu/R/CRAN/

- **Spatial** (Analysis of spatial data)
  http://lib.stat.cmu.edu/R/CRAN/web/views/Spatial.html

(or use r-seek, or the forums & blogs above)