



Workshop Overview

Dates and Times: Monday and Wednesday (Dec 19th and 21st) 9:00am-4:00pm. The first day (see schedule of topics below) will be a general overview and application of the R language and working environment and is meant to provide an introduction to the fundamentals of working in R. My intent is to have each session consist of a brief presentation, followed by at least one “hands-on” example. For example, the presentation in the session on “Data import & export” would cover various text file formats (e.g., .csv, .dat, .txt), with simple examples. The second half of that session would include self-paced examples of loading data directly from these file types and examining the options available when doing so... The second day will introduce a few more topics, most of which are slightly more advanced, and has a primary focus on giving you additional experience opportunities doing analyses that are familiar (e.g., t-tests, regression) and task commonly used in exploratory data analyses (e.g., plotting data, building figures). Each of the two days will end with a block of time (about one hour) for you to practice the concepts and techniques learned using your own data (example datasets can be provided if needed).

Why R?

R is an open-source implementation of the S language for statistical computing. For over 20 years, applied statisticians have been submitting implementations of their new techniques to StatLib. When most of those implementations were written as libraries for the commercial S-plus implementation of the S language, statisticians were providing software for free, but users (including those same statisticians) had to pay a third party (Insightful Corp.) to be able to run the software. A very small group of statisticians took it upon themselves to write a complete open-source implementation of S that would run under most operating systems, which they called R. Since then, the vast majority of implementations of new statistical techniques have been made available as R packages, which include the code as a library of functions and at least some documentation. In fairness to S-Plus, while there are several GUI interfaces available for R, S-Plus provides a much more polished and complete GUI interface and user experience.

Because R is very useful for "computing with data", experts in many fields use it for their work. Because R is open source, many of those experts make their field-specific code and functions freely available as packages (currently over 9654 packages, see http://cran.stat.ucla.edu/web/packages/available_packages_by_date.html). For example, climate researchers use R with netCDF files, so there are packages for reading and writing netCDF files (netCDF, ncd4) as well as for generating standard climate diagrams, imputing missing weather data, downscaling from coarse data, etc. (climtol, clim.pact, seas, anm, zyp). Jari Oksanen (with help from others) provides package vegan for vegetation analysis (ordination, classification, analysis of similarity, and much more). There are several packages for species richness, diversity, rarefaction, etc..The key point is that by learning how to use R, at least to the level of writing code to reshape our data into the required structures and call the provided functions, we can leverage their efforts and expertise, and not reinvent those wheels. In order to produce informative and valid

results we still have to understand the topic (e.g., dendrochronology or wildlife population assessment), but we do not need to translate the approaches and equations in the literature into computer code, as experts have done that, tested it, and (to varying degrees) documented it.

Other Learning Options

We learn in very different ways, and we have a wide range of backgrounds. Even if it goes well, this workshop will not be the most efficient way for some of us to learn to use R in our work (it likely would not have been for me). You may do better simply working through the slides and examples associated with this workshop at your own (likely faster) pace. [Coursera](#) and [Lynda](#) have numerous offerings on data analysis with R. There are a number of paper books that provide an introduction to R and I would suggest the following resources:

- Many people who convert from SAS or SPSS recommend the R in Nutshell book from O'Reilly. It's good, but it's mostly a rehashing of available online help files: <http://oreilly.com/catalog/9780596801717>
- Perhaps the best compilation of online video/tutorial type resources for learning R is this collection compiled by Jeromy Anglim: <http://jeromyanglim.blogspot.com/2010/05/videos-on-data-analysis-with-r.html>
- As a former SAS user myself, I often suggest starting with Muenchen's book from Springer: <http://www.springer.com/statistics/computational+statistics/book/978-1-4614-0684-6> His accompanying website with examples is pretty useful tool: <http://r4stats.com/examples/>

As you progress, you'll quickly outgrow any single book and then I suggest using the true power of R, the available online community and resources. Here are a few suggestions:

- The [R Project homepage](#). It really should be bookmarked. This is the place to come for official news from the R Project, plus links to [documentation](#), [mailing lists](#), and the [official R FAQs](#)
- [StackOverflow](#). Have a question about R? Search for questions tagged with "r" and you'll probably find an answer. If not, post your question and I guarantee you'll have an answer before you know it....
- [R bloggers](#). This is the first "news" feed I check every morning. It's my go to for news, tips and articles related to R and is basically a blog aggregator for posts from dozens of R bloggers, including the awesome work from the team at [Revolution Analytics](#) it.

Pre-course Installation

We will start on Monday, December 19th with the expectation that everyone has R and RStudio installed and running on your computer if you so choose to bring it, otherwise we'll be working in the Graduate Computer Lab (Rm 112). I recommend that you either install R and RStudio yourself, following the directions below, or let your IT folks do the install if you don't have administrative rights or are uncomfortable installing software. My recommendations for a quick installation are as follows:

1. Download R from <http://cran.us.r-project.org/> (click on "Download R for Windows" > "base" > "Download R 3.3.2 for Windows"). Install R. Leave all default settings in the installation options. If you are running a 64-bit version of Windows, this version will install both 32-bit and 64-bit R, which is what you want.
2. When prompted, install R in c:/R/R-3.3.2 not c:/Program Files/R/... This is especially important if you don't have administrative rights on your computer and don't have permissions to install software to the C:/Program Files/ folder. You may be downloading additional packages, and they will need to write files under your R directory.
3. Download RStudio Desktop from <http://rstudio.org/download/desktop> (click on the Download link at the bottom of the column labeled "RStudio Desktop Personal License" and then click "RStudio 1.0.44 - Windows XP/Vista/7/8") and install it. Just select the default settings in the installation.

Session Topics

This schedule is tentative, and subject to change if the workshop is going too fast or too slow for the majority of the participants.

Day	Time	Topic
Monday Dec 19th	9:00-9:20am	Live Demo
	9:20-9:45	History of R and why you should learn R
	9:45-10:15	Overview: RStudio, Github, R's online community, & "packages"
	10:15-10:30	Break
	10:30-11:15	The R Language I: Data structures, objects, operators and functions
	11:15-12:00	The R Language II: Vectors, matrices, arrays, data frames, lists & factors
	12:00-1:00	Lunch
	1:00-1:30	Using R I: Data import & export
	1:30-2:00	Using R II: Subsetting, reshaping and combining dataframes
	2:00-2:15	Break
	2:15-2:45	Using R III: User-defined functions
	2:45-3:15	Using R IV: Manipulating data
	3:15-4:00	Begin working with your own data
Day	Time	Topic
Wednesday Dec 21th	9:00-10:00am	Advanced Importing: excel, access, shapefiles, and rasters
	10:00-10:15	Break
	10:15-11:00	Applied R I: t-tests, linear regression, and chi-squared tests
	11:00-12:00	Applied R II: one- & two-way ANOVA
	12:00-1:00	Lunch
	1:00-1:30	Applied R III: principal components analysis
	1:30-2:00	Visualizing data & ggplot
	2:00-2:15	Break
	2:15-3:00	Overview: putting it all together and other useful packages
	3:00-4:00	Continue working with your own data