
Intermediate R

MATTHEW DENNY

THURSDAY 19TH FEBRUARY, 2015



This workshop will introduce some basic programming concepts such as looping, user defined functions, and conditional statements, as well as a few other useful functions in base R. These concepts make up the vast majority of what you will gain in power and functionality over using a language like SAS, Stata, or SPSS. It is meant to serve as a follow up to the Intro to R workshop – materials are available here: [\[Download\]](#). We will be working through the attached R script but please take a look at some of the links below before coming to the workshop.

1 Web Resources

1. My favorite go-to guide for basic data management and statistics in R is the **Quick-R** website: <http://www.statmethods.net/>. Check out the following pages:
 - (a) **Operators in R:** <http://www.statmethods.net/management/operators.html>. These let you make your programs smart by letting them check to see if some condition is met and decide what to do based on what they find.
 - (b) **Built-in Functions:** <http://www.statmethods.net/management/functions.html>. These functions do lots of useful math related things like rounding, taking the mean etc., and are much faster than an equivalent function you could write using R code.
2. Here is a nice blog post on For loops in R: <http://paleocave.sciencesortof.com/2013/03/writing-a-for-loop-in-r/>
3. The R cookbook provides lots of helpful examples that you can adapt to do whatever you need to do: <http://www.cookbook-r.com/>
4. Hadley Wickham's excellent, thorough, well written ebook on advanced topics in R is available here: <http://adv-r.had.co.nz/>. The relevant chapters to check out for today's workshop are:

The current paper where we actually use the data is under review and there is not a public version available right now

 - (a) **Subsetting Data:** <http://adv-r.had.co.nz/Subsetting.html>. This chapter covers going and pulling out some part of your data.
 - (b) **Functions:** <http://adv-r.had.co.nz/Functions.html>. These let you automate some task or even a whole set of tasks.
5. The R graphics gallery is a really good place to go to find code for making whatever crazy graph you could think of. <http://rgraphgallery.blogspot.com/>
6. **ggplot2** is what all of the cool kids are using these days to make their graphics in R. It is harder than the graphing functionality in base R but once you feel comfortable with base R and want to make crazier graphics, check this site out: <http://docs.ggplot2.org/current/>
7. **magrittr** is a new package that some people may find makes writing functions much more intuitive, take a look at this blog post: <http://blog.rstudio.org/2014/12/01/magrittr-1-5/>

2 Useful R Packages

Below is a list of R packages (with short descriptions) that I use frequently:

1. **statnet** – The most feature-rich package for inferential social network analysis.
2. **igraph** – This package has a lot of nice functionality for network simulation and community detection.
3. **topicmodels** – A nice package for topic modeling in R.
4. **biglm** – A package that lets you run regression models when you do not have enough memory to hold your dataset.
5. **snowfall** – A package for creating your own mini-cluster to run analyses in parallel.
6. **scrapeR** – A package for automatically downloading webpages (web scraping).
7. **lme4** – mixed effects panel data models.
8. **Matching** – A package to do matching as pre-processing before running regression or calculating treatment effects.
9. **dplyr** – Useful for transforming and collapsing data.
10. **stringr** – Lots of functions for working with text strings.
11. **slam** – Nice for creating large sparse matrices that would otherwise be too large to hold in memory.
12. **corrplot** – makes amazing visualizations of correlation matrices.