R Workshop: Introduction

Jennifer Bufford
December 15, 2014

Useful Links

www.rseek.org (search engine for all things R)
http://cran.r-project.org/doc/contrib/Short-refcard.pdf (reference card)
http://rstudio.org/ (RStudio – IDE for R)

RStudio

4 panels

- *Scripts* - a place to write all the code you want to run in R, and to check and fix it before you run it
- *Console* - this is R, text entered here is run by R and any R outputs appear here (including error messages)
- *Environment/History* - shows all objects created and code run (respectively)
- *Files/Plots/Packages/Help* - multi-function panel
  - Files - search for files to import into R, set working directory (the folder R pulls files from); does not load files
    - Ellipses to the right opens a window to browse for files
  - Plots - shows any plots that R produces, save in various formats, including pdf
  - Packages - shows currently installed R packages, use menu to search for and install new packages, update packages and load or unload packages
  - Help - can search R documentation; any help/search queries conducted in R console open here

Projects

- Menu in upper right corner called “Project” allows you to group open scripts, environment, history, workspace for different projects/papers
  - To create a new project:
    - Project -> New Project -> Existing Directory -> select the folder you will be working from -> Create Project
    - This will automatically set the folder as your working directory
    - History and environment (workspace) will be saved to that folder
  - To switch projects use Project -> Open Project
    - Current project will be saved before it opens the new one

Other (advanced) features:

- Debugger
- Create headings with ##### Heading #####
- Tools for reproducibility including knitr, rmarkdown and sweave

Data Formats

Easy-to-read format:

- Variables in columns
- Replicates in rows
- Data as numbers whenever possible
- Eliminate "notes" or "observations"
- Use numbers and symbols appropriately
  - Coding binary data as 0 or 1 is OK (although not necessary)
  - Using "+1" as a note or observation could cause problems
- Descriptive but concise headers/labels (you could be typing them a lot!)

Save data as a .csv file for import into R

Loading Data

1) Use R code to load data

```
setwd("C:/...") #PC
setwd("/Users/…") #Mac
Data=read.csv("Data.csv")
```

2) Use RStudio menus to load data

Environment --> import data set

RStudio produces a menu-driven pop-up to format the data being loaded

- Can rename file, designate headers (but not row labels), data structure
- Again choose simple, descriptive names for clarity

Saving Data

Use the command "write" to save modified data

```
write.csv(Data, file = "DataName.csv")
```

Save environment (workspace) using

```
save(list=ls(), file=".RData") #save all objects
save.image(file=".RData") #save all objects
save(obj1, obj2, obj3, file=".RData") #save only objects 1, 2 and 3
```

- You can choose which objects to save, and can change the file name (but keep the .RData extension)
- Save the major data objects you're working with – it's much easier than always reading in from a .csv (keeps R formatting for factors, etc.)

Scripts

Scripts are a clean copy of your working code
- Anything you do to your data (permanently) should be in a script!!

Write code first in a script, then run it, then correct it in the script and run again and repeat

Save all your scripts with working code and appropriate metadata
- Project, author name(s), date last updated, description of what the code does, etc.

Break big projects into discrete tasks and make each task a separate script
- Consider using a master script to connect all the tasks together
- Run an entire script at once (even without opening it) using:
  - source("filename.R")
  - Replace “filename.R” with the name of your script

Scripts should be re-usable
- If your data changes, it’s easy to run processing/analysis again

Organization

Find a system that works for you, but you need a system!
- Keep data, code, output (e.g. model results, graphs, etc) together

Example system
- Project Folder (e.g. dissertation chapter 1)
  - Data – separate raw data and processed data
    - Save unaltered raw data!!
  - R Scripts – includes scripts and .RData files
  - Data Exploration – exploratory graphs and outputs
  - Data Analysis – graphs, model results, etc
  - MS – paper in progress
  - Old/Drafts – old code, code fragments or drafts

Document, document, document!
- Use additional files (or notebook) if necessary to document data sets
- Document data transformations, analyses, etc
- Use comments liberally (#comment) to clarify code
  - Code is a document of what you did (save before exiting!)
- Use scripts to create re-usable codes

Exercises

Intro Workshop Exercise Script.R has exercises in reading in data, examining the data, calculating exploratory stats, and making simple manipulations. Answers to the exercises can be found in *Intro Workshop Exercise Script Answers.R*