Data Import and File Manipulation

Jessica A. Turner

Department of Psychology, Georgia State University

jturner63@gsu.edu

# Introduction

This is the handout to accompany the second section of the workshop on data import to the R system. The script file to follow the commands is: **W02\_DataImport.R** – please see that for **much** more detail.

Important commands:

1. Finding out where you are, what the working directory is: **getwd()**
2. Set the working directory if you’re not there already:

**setwd(‘path’)**

where ‘path’ is the path to the directory you need.

1. Get the list of files in the working directory: **list.files()** or **list.files(path=”.”)**
2. Loading datafiles:

**mydata <- read.csv(‘filename.csv’)**

**mydata <- read.table(‘datafile.txt’, header = T)**

1. Clearing variables:

Clear ALL variables: **rm(list=ls())**

Clear variable ‘V’: **rm(V)**

1. Examining data
   1. **summary(mydata)**
   2. **str(mydata)**
2. Manipulating data
   1. Indexing: Use **$label** to get the part of the data called by that label, e.g. **mydata$age**
   2. Selecting (the final comma is VERY important!):

**mydata[mydata$Gender=="F" & mydata$StressReduction > 3),]**

1. Writing datafiles (basic form; there are many options):

**write.table(dataframe, file=”path/filename”, sep=”,”, row.names=FALSE)**

1. Data format:

In **tidy data**:

1. Each variable forms a column.
2. Each observation forms a row (if you have multiple observations per subject, you’ll have multiple rows per subject, one per observation).
3. Each type of observational unit forms a table.
4. More information available at:

<ftp://cran.r-project.org/pub/R/web/packages/tidyr/vignettes/tidy-data.html>

and at:

<http://www.cookbook-r.com/Manipulating_data/Converting_data_between_wide_and_long_format/>

1. Changing the organization of data in R is facilitated by the **reshape2** or **tidyr** libraries. See <https://blog.rstudio.org/2016/09/15/tidyverse-1-0-0/> for an article describing the related tidy packages, and <http://tidyverse.org/> which is the home for the related packages that work together.
2. Paired t-test and basic ANOVA examples
   1. **t.test(y, x, paired = TRUE)**
   2. **aov( DV ~ IV, data= mydata)**
   3. Note that **summary()** works on statistical analysis objects, too!
3. Introduction to plotting:
   1. Histogram: **hist(y,x)**
   2. **plot(y, x)** – this does a box plot or scatter plot depending on the variables type. To specifically force a boxplot, use **boxplot()**.
4. Libraries used:

“**foreign**” for importing SPSS

“**xlsx**” for importing Excel

1. Where to get more information:
   1. <http://www.statmethods.net/>
   2. <https://www.r-bloggers.com/how-to-learn-r-2/>
   3. <https://www.datacamp.com/community/tutorials/r-data-import-tutorial#gs.yi1pYDA>
   4. <http://stats.idre.ucla.edu/r/faq/how-to-input-data-into-r/>
   5. And web searches for specific commands, with the addition of “R” to the search often work, e.g. “summary in r” or “aov r”.

https://licensebuttons.net/l/by-nc-sa/4.0/88x31.png

The materials in this document are released under a **MIT License** (for code examples) and the Creative Commons for text.

SEPA 2017 R Workshop - R: From Startup to Statistics by Jessica A. Turner is licensed under a **Creative Commons Attribution–NonCommercial-ShareAlike 4.0 International License**. Based on a work at http://github.com/theEducationalCollective/SEPA2017-R-Workshop.