DATA IMPORT AND FILE MANIPULATION NOTES

Jessica A. Turner & Matthew D. Turner
Department of Psychology, Georgia State University
jturner63@gsu.edu, mturner46@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.

3. Get the list of files in the working directory: **`list.files()`** or:

```
list.files(path=".")
```

4. Loading datafiles:

```
mydata <- read.csv('filename.csv')
mydata <- read.table('datafile.txt', header = T)
```

5. Clearing variables:

   Clear ALL variables: **`rm(list=ls())`**
   Clear variable 'V': **`rm(V)`**

6. Examining data
   a. **`summary(mydata)`**
   b. **`str(mydata)`**

7. Manipulating data
   a. Indexing: Use **`$label`** to get the part of the data called by that label, e.g. **`mydata$age`**

b. Selecting (the final comma is VERY important!):

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

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

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

9. 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.

10. 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/

11. 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.

12. Paired t-test and basic ANOVA examples
    a. `t.test(y, x, paired = TRUE)`
    b. `aov( DV ~ IV, data= mydata)`
    c. Note that **`summary()`** works on statistical analysis objects, too!

13. Introduction to plotting:
    a. Histogram: **`hist(y,x)`**
    b. **`plot(y, x)`** – this does a box plot or scatter plot depending on the variables type. To specifically force a boxplot, use **`boxplot()`**.

14. Libraries used:

    "**`foreign`**" for importing SPSS
    "**`xlsx`**" for importing Excel

15. Where to get more information:
    a. http://www.statmethods.net/
    b. https://www.r-bloggers.com/how-to-learn-r-2/

c. https://www.datacamp.com/community/tutorials/r-data-import-tutorial#gs.yiipYDA
d. http://stats.idre.ucla.edu/r/faq/how-to-input-data-into-r/
e. And web searches for specific commands, with the addition of "R" to the search often work, e.g. "summary in r" or "aov r".