RStudio Tutorial 1

**Sourcing and Loading Data Into R**

**Objectives**

**In this exercise, you will explore the different types of data that can be loaded into R for processing.**

**Overview**

R has tremendous flexibility in reading data. This exercise will allow you to load internal data sets, .csv files, web pages (webscraping) and XML.

**1: Reviewing the first data set**

1. Open RStudio.
2. At the console prompt, type library() to view all loaded packages. A list of packages will appear in a window above the console window.

**library()**

1. Use the **data()** command to list all of the data sets available in these loaded packages.

**data()**

5. Load the AirPassengers data set into memory.

**data(AirPassengers)**

6. View the AirPassengers data set.

**AirPassengers**

*Which years are covered by the AirPassengers data set?*

**

1. Load some other of the available data sets into memory and view them. e.g., data(BJsales)
2. When you have loaded a number of data sets into memory, list all loaded data sets.

**ls()**

1. Remove the AirPassengers data set from memory and confirm that it has been removed.

**rm(AirPassengers)**

10. Remove all loaded data sets from memory.

**rm(list=ls())**

11. View data sets available in the rpart package.

**data(package="rpart")**

**2: Loading an external .csv file into R**

12. View the current working directory in R.

getwd()

13. View the structure of the following .csv file using Textpad:

RStudioTutorial 1 - weather.csv

When you have finished viewing the file, load it into memory in R, specifying appropriate parameters, as follows:

*Specify a separator symbol and headings, and don't forget to use double-backslashes when specifying the file location*

*myData <- read.csv("C:\\RStudio Tutorial 1-weather.csv", sep=",", header= TRUE);*

1. View the loaded data in a data viewer. When you have finished, remove the data from memory

View(myData)

rm(myData)

15. View the structure of the following .csv file using Textpad.

RStudio Tutorial 1 - weather2

When you have finished viewing the file, load it into memory in R, making the appropriate changes to the parameter(s).

*myData <- read.csv("C:\\* RStudio Tutorial 1 *weather2.csv", sep=",", header= FALSE);*

16. Give column headings to the first few columns:

names(myData) <- c("Date","MinTemp","MaxTemp")

**3: Loading a web page into R (webscraping)**

17. Load a web page into R and view it:

**file <- readLines("http://en.wikipedia.org/wiki/ Data\_science")**

18. View the loaded data:

**file**

19.

20.

Load an HTML table from a website into R:

**library(XML)**

**library(httr)**

**url <- "http://en.wikipedia.org/wiki/ List\_of\_countries\_by\_population"**

**population <- GET(url)**

**population <- readHTMLTable(rawToChar(population $content),stringAsFactors=F)**

View the loaded data (the 1st element)

**population[[1]]**