ANLY 500 Laboratory #1 (Part 1) – Descriptive Statistics

Evans Chapter 1

“Performance Lawn Equipment Case Study” from Evans, **Business Analytics**

Contents

[Introduction 1](#_Toc476138323)

[Chapter 1 3](#_Toc476138324)

[Step 1 3](#_Toc476138325)

[Step 2 4](#_Toc476138326)

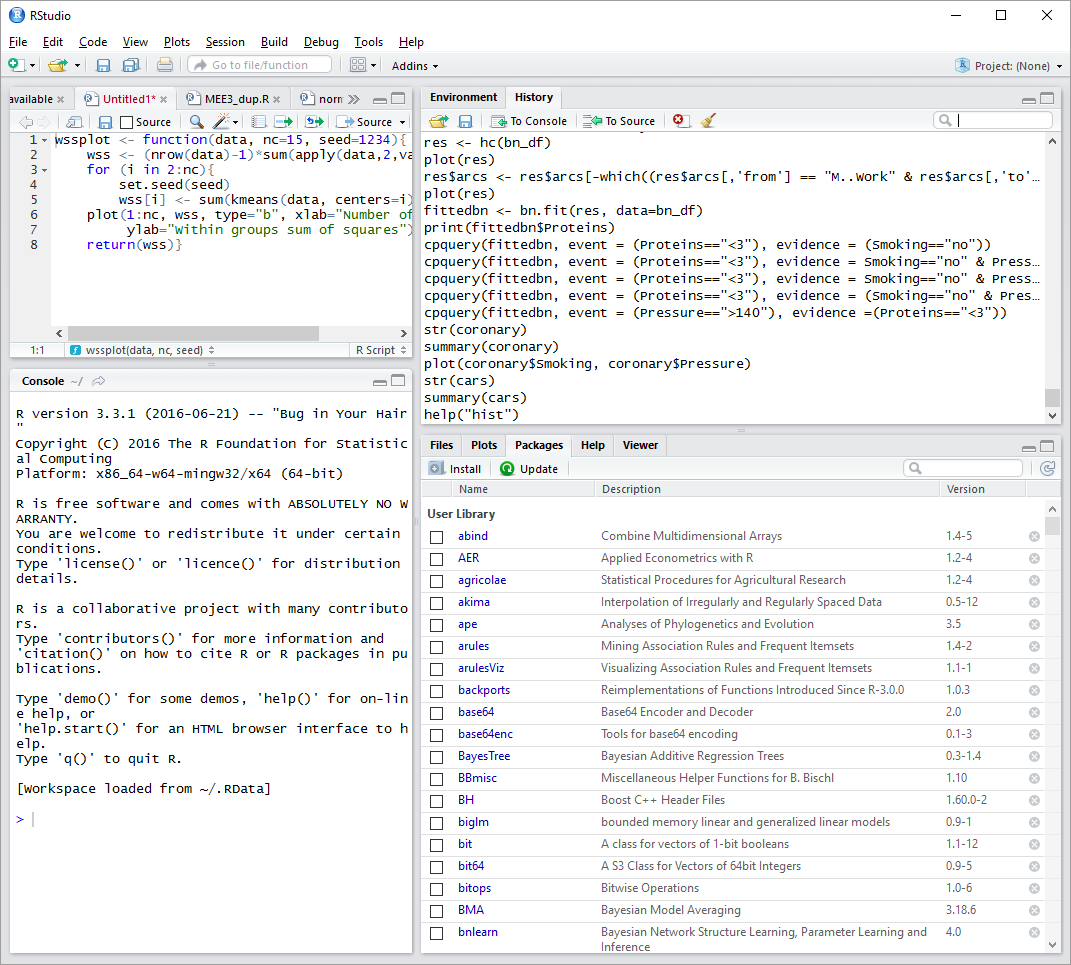
# Introduction

This laboratory follows the exercises in the book, specifically the Performance Lawn Equipment Case Study homework assigned exercises Chapters 1 through and including 7, except this laboratory requires that you use R to complete the exercises. That is, you should answer all questions in the textbook exercises and when necessary to complete computations use R. Each laboratory in ANLY 500 will build on the laboratories you have completed before. So, you will want to set-up a folder or file to keep your work in so that you can refer back to previous laboratories if necessary. If you have not used R before you should install R and RStudio on your computer or laptop. RStudio is a user interface for R that will make your life and work much easier. To get credit for completing this laboratory you must submit a report with your results on Moodle.

Once you have installed R and RStudio you may want to browse through some of the packages available for you. You can do that from the “Quick list of useful R packages” at <https://support.rstudio.com/hc/en-us/articles/201057987-Quick-list-of-useful-R-packages> or <https://cran.r-project.org/web/packages/>. Essentially what R does is use functions already coded in these packages to do the computations you want to perform. Each package will have an associated CRAN package website that provides all the information you need about any package. You can also do a Boolean search on Google or other browser to find additional information about packages or functions. If you need a specific package to complete an exercise you will be told which package that is as part of the exercise.

Unless told you will need to find a data set to use you will be provided data sets through Moodle. This is true for this laboratory, ANLY 500 Laboratory #1. For this laboratory there are a total of 23 data files in csv format. There is one data file for each spreadsheet in the Performance Lawn Equipment Excel Workbook that is also on Moodle. Your first task will be to load these data files into RStudio. However, before you can begin to read data into RStudio you will need to be able to move around the folders on your computer.

When you start RStudio you will see a number of frames in the RStudio window, going clockwise from the upper left: a frame showing contents of your R scripts or data objects; an Environment and History frame; an information frame including your files, plots, packages, help and viewer; and, your Console frame. This will look something like the figure below:



To find out what folder your default folder is set-up to be you can use the pull down menu “Tools” then look at your “Global Options”. If you click on “Browse” by the box for your “Default working directory” it will take you to the directory that RStudio goes to automatically when you start RStudio. If you want to change this directory just browse to the one you want to use and accept that change. I strongly suggest you set-up a separate directory just for your R/RStudio work. I have one I names “MyRWork” and within that folder I have a “data” folder and other folders for specific projects, etc.

The function to find out what directory you are actually in is getwd() which simply stands for get working directory. That is the syntax you need to use. The parentheses, which are empty, return the current directory. To change directory the function is setwd(). So, for example if I am in my default working directory “MyRWork” and want to go to my data directory I enter setwd(“data”) in the console frame. The quotation marks are necessary. R distinguishes between names with no quotation marks, single or double quotation marks and treats each entry differently. So, for the directory name use the double quotation marks around it. If things are not working quite right, e.g. RStudio isn’t reading files, chances are you are not in the correct directory.

Once you are in the directory where you’ve downloaded the data files you can load the data into RStudio. You can do this automatically using the pull down menu “Tools” then “Import Dataset”. Doing this you choose a “From Local File” or “From Web URL”. Since these will be on your computer or laptop choose “From Local File” then just go to the appropriate directory and select the data file you want to import. Or you can immediate begin using R’s functions for reading in data using the following command in the Console frame:

> BladeWeight <- read.csv("~/MyRWork/data/Evans/BladeWeight.csv")

You have extensive help files available through RStudio. To get help, in the Console use the help() function and the function name you need help with in double quotations, e.g.

> help("read.csv")

One thing to watch out for, whether you use the pull down menu to import or the command line, is that the column headings are recognized. For some reason when I read or import the Evans’ data files some files recognize headings and other do not. So, be careful about this.

# Chapter 1

## Step 1

Read the data files for Performance Lawn Equipment (PLE) into R/RStudio.

## Step 2

Determine the data type for each variable in the PLE data files.

It is easy to determine the data types for variables in R. Simply use the str() function with the data filename or data object in the parentheses. For example,

> str(BladeWeight)

'data.frame': 350 obs. of 2 variables:

$ Sample: int 1 2 3 4 5 6 7 8 9 10 ...

$ Weight: num 4.88 4.92 5.02 4.97 5 4.99 4.86 5.07 5.04 4.87 ...

For the BladeWeight data file the str() function returns the information that there are 350 observations of 2 variables, sample and weight. The sample variable is an integer variable. The weight variable is a numeric variable. You can use the str() function for each data file to determine the data type of all variables. Let’s consider one more of the data files in detail, i.e. the EmployeeRetention data file. Using the str() function we get:

> str(EmployeeRetention)

'data.frame': 40 obs. of 7 variables:

$ YearsPLE : num 10 10 10 10 9.6 8.5 8.4 8.4 8.2 7.9 ...

$ YrsEducation: int 18 16 18 18 16 16 17 16 18 15 ...

$ College.GPA : num 3.01 2.78 3.15 3.86 2.58 2.96 3.56 2.64 3.43 2.75 ...

$ Age : int 33 25 26 24 25 23 35 23 32 34 ...

$ Gender : Factor w/ 2 levels "F","M": 1 2 2 1 1 2 2 2 1 2 ...

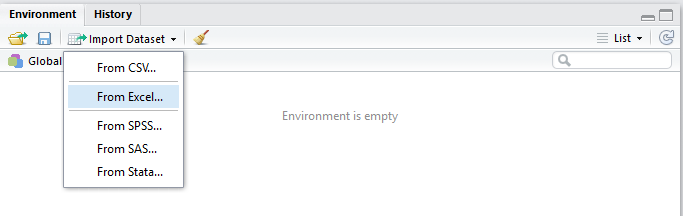
$ College.Grad: Factor w/ 2 levels "N","Y": 2 2 2 2 2 2 2 2 2 1 ...

$ Local : Factor w/ 2 levels "N","Y": 2 2 1 2 2 2 2 2 2 2 ...

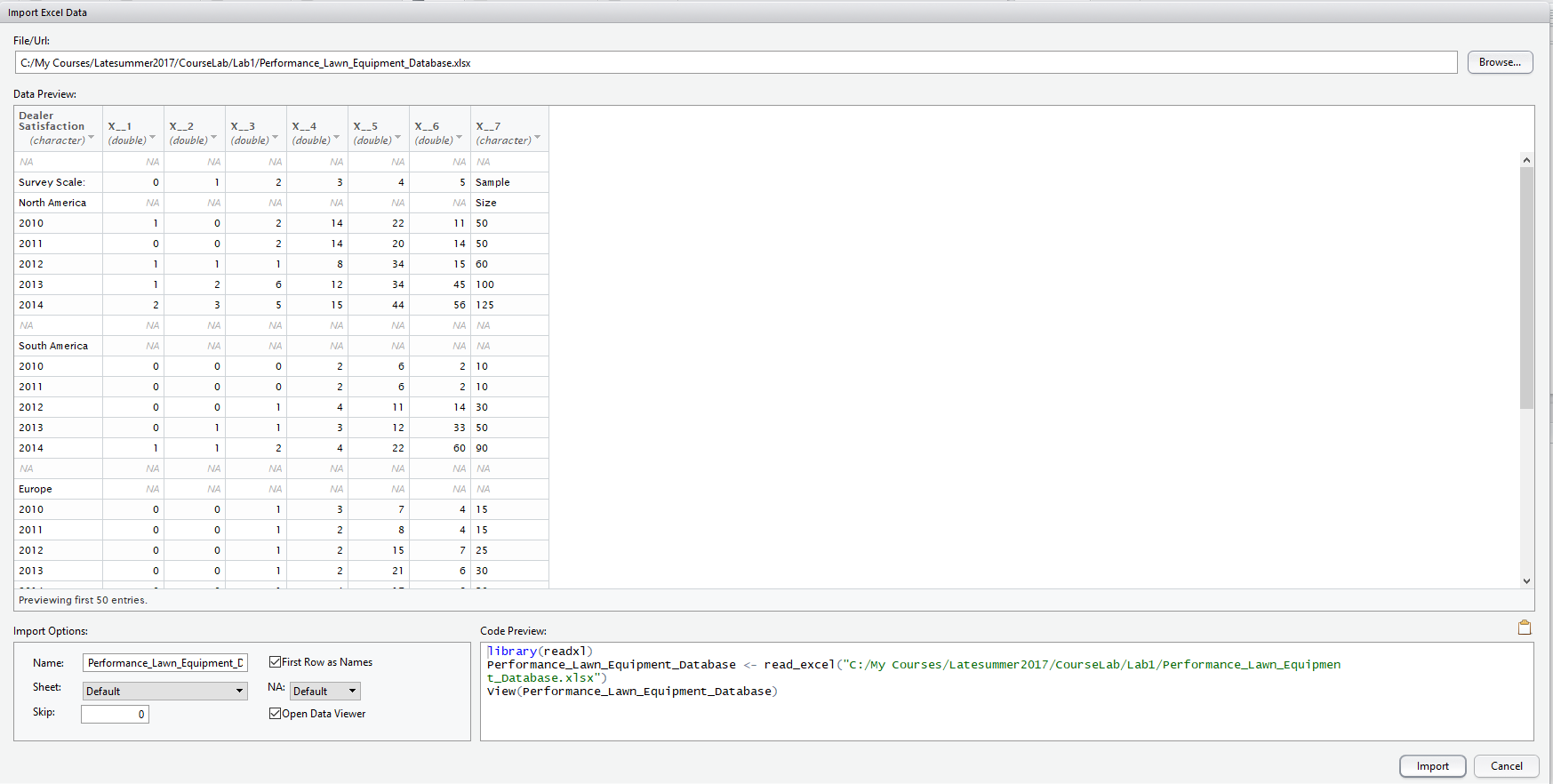
where we have 40 observations for the 7 listed variables. Gender, College Grad, and Local are all listed as “Factor” variables. This is the same as a categorical variable. If you are not familiar with different data types you should take some time to look into this. One source you can use is: <https://www.tutorialspoint.com/r/r_data_types.htm>.

*You may repeat these steps for 4 datasets.*

One other approach is to use import option in R. From the right top window, follow the following steps:

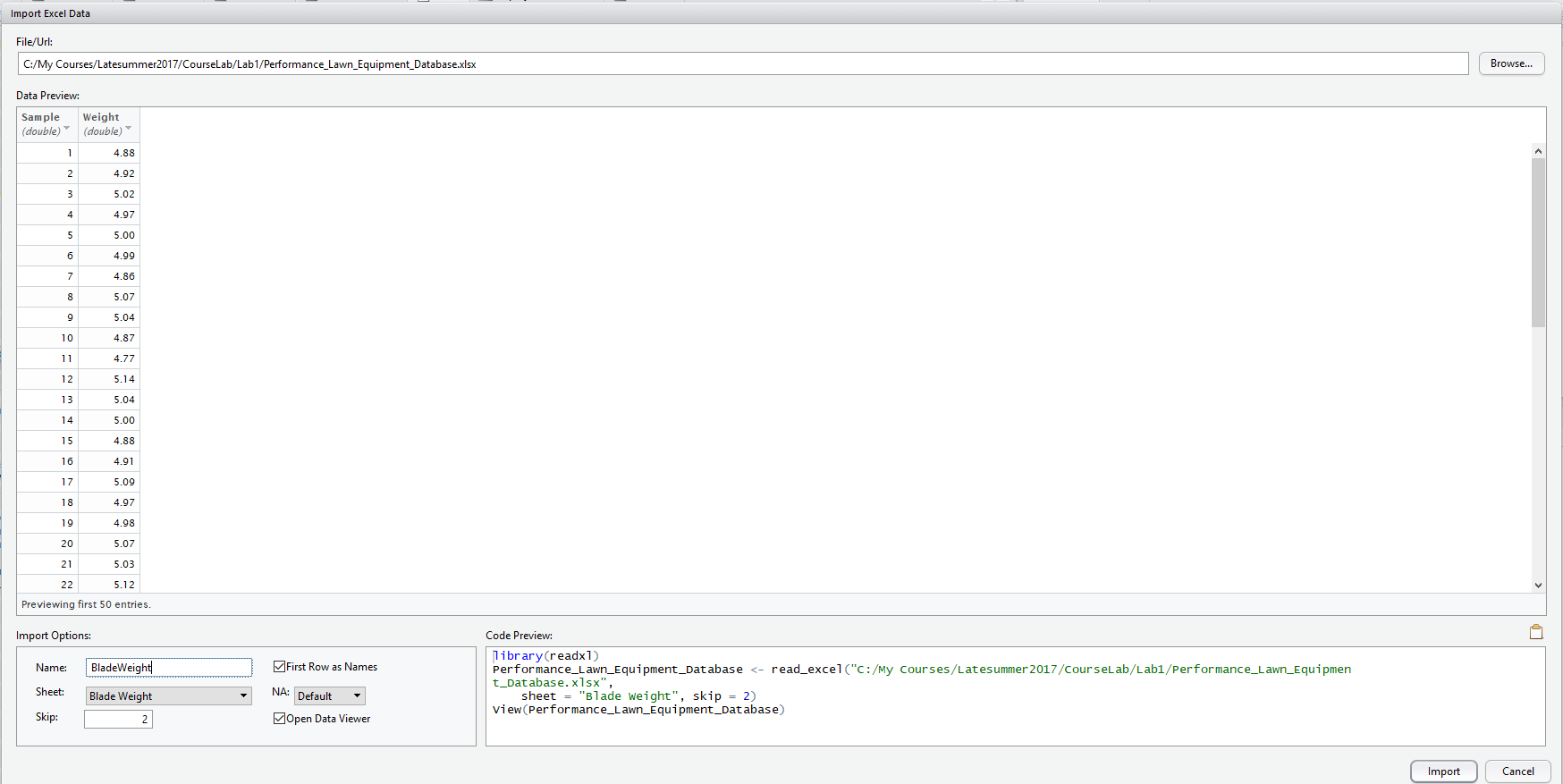


Now choose the provided excel file using Browse button



Choose the desired tab from here

Choose the desired tab and change the name using “Name” Box. As you see, the first couple of rows, which are descriptions, should be removed. Use “skip” box to choose how many rows should be eliminated. You can see the result of using skip=2 below for “BladeWeight” tab:



Now you can use “Import” button and you will see the database loaded into R.