**Module 3: Additional Exercises**

Download the “Financial Dataset”.

Course Content> Datasets > Financial.rda

Clear your R workspace using the following command

rm(list = ls())

Load the Financial data into R, using load() command e.g.

load("F:/Kent Teaching/Datasets/Financial.rda")

where F:/Kent Teaching/Datasets/ is a directory at which the downloaded file is located at. Depending on where you have downloaded this file on your computer, the path is going to be different for you.

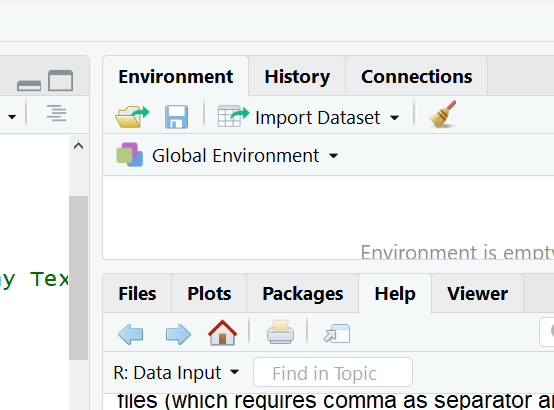
Note that that “/” is used instead of “\”. Alternatively you could use “//” i.e.

load("F:\\Kent Teaching\\Datasets\\Financial.rda")

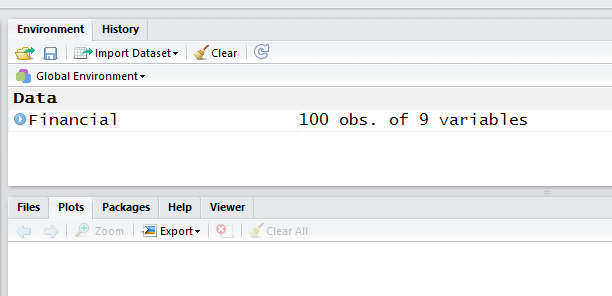
If you are using a Mac, download the file. Then right click on the file and click on “Get info” then in front of ‘where’ you get the path e.g.

load("/Users/Razavi/Desktop/Financial.rda")

if none of the above worked, you can use R-tudio graphical interface to locate and load “Financial.rda”. For that, simply click on the open icon on the R-studio envitonment section and locate the file,



Now you should see a Data frame named Financial in your Global Environment Section



This dataset contains the financial information about 100 companies including their type (type), revenue (rev), assets (assets), return of investment (roe), Earnings Per Share (eps), yield (yield), Dividend Per Share (dps) and finally their Property, Plant and Equipment (ppe).

Answer the following questions:

1. **What are the variable names in the “Financial” dataframe?**
2. **What is the mean, median and standard deviation of revenue across all 100 companies?**
3. **What is the highest revenue amongst all companies?**
4. **Which company has the highest revenue?**
5. **Which company has the lowest Return on Investment?**
6. **What are the top 5 companies with highest assets?**
7. **What is the standard deviation of “roe” values. Try to calculate the standard deviation without using the sd() command and by formula, then use the sd() command and compare the results.**
8. **Use the summary() command to get a summary of all variables in the Financial dataframe. By comparing the mean and median, can you say which variables are highly skewed? And in which direction (positive/right skewed versus negative/left skewed). Then use skewness() command to confirm that.**
9. **Calculate the skewness of “assets” without using the skewness() command. Then use the command to compare the results.**
10. **Examine the distribution of the “assets” and compare it with “dps”.**
11. **Plot the Boxplot of “Yeild” and “dps” (i.e. columns 7 and 8 of the Financial dataframe). Just by looking at the boxplot, can you say which variable has a larger spread i.e. standard deviation? Confirm this by comparing the standard deviation values.**
12. **Calculate and plot the correlation between all numerical variables in the “Financial” dataframe.**