Project 1

Optional Project: The objective of this introductory project is to get you acquainted with the basic features of Rstudio (or R).

Task #1: Search on the WWW and compile a list of some 5 web sources of R (or RStudio) tutorials, wiki, and the likes. Your list should include the complete links to these sources.

Task #2: Read the document **Short Introduction to R** located in the file area of the course canvas and play around with R (Rstudio):

A) Do some arithmetic
  ➢ `Y=3*7^2+5 # simple calculation`
  ➢ `Y        # see the value of Y`

B) Do some computation on a 'vector'
  ➢ `X=c(2, 7, -1, 3, 8, 10, -5) # declaring the values of X`
  ➢ `sum(X)                 # compute the sum of X`
  ➢ `sum(X^2)               # compute the sum of` X^2
  ➢ `length(X)              # find the number of elements is X`
  ➢ `sum(X)/length(X)       # compute the average, or`
  ➢ `mean(X)                # same as the mean of X`
  ➢ `sum(X-mean(X))         # and compare it to`
  ➢ `sum((X-mean(X))^2)`

C) Define a function:

  ➢ `MyFunction<-function(x)`
    ```
    {
    Y<-3*x^2+5
    return(Y)
    }
    ```
D) Test your function:

  ➢ `MyFunction(7)`
  ➢ `MyFunction(7:11) # try it for x=7, 8, 9, 10, 11`

E) Plot your function:

  ➢ `xx<-seq(0, 7, length=200)`
  ➢ `plot(xx, MyFunction(xx), type="l", main="Plotting MyFunction",`
    `xlab="x-values", ylab="Y=f(x)")`

Task #2 – Refer to the MHW Data – see details below.

A) Retrieve the data file MHW.txt from the class site on Canvas and save it on your local work directory.

B) Read in the data to an appropriate data frame:

➢ `MHW<- read.table('MHW.txt', header=TRUE, sep=',')`

C) Print the top 6 lines of the data set

➢ `head(MHW)`

D) Print the first 10 rows of this data frame

➢ `MHW[1:10, ]`

E) Attach the data frame to be your working data set

➢ `attach(MHW)`

Task #3: Summarize the MHW data:

A) Obtain the five-number summary for the data:

➢ `summary(MHW)`

```
summary(MHW)
       r               c            grain            straw
 Min.   : 1.00   Min.   : 1    Min.   :2.730   Min.   :4.100
 1st Qu.: 5.75   1st Qu.: 7    1st Qu.:3.638   1st Qu.:5.878
 Median :10.50   Median :13    Median :3.940   Median :6.360
 Mean   :10.50   Mean   :13    Mean   :3.949   Mean   :6.515
 3rd Qu.:15.25   3rd Qu.:19    3rd Qu.:4.270   3rd Qu.:7.170
 Max.   :20.00   Max.   :25    Max.   :5.160   Max.   :8.850
```

➢ `summary(MHW[,3])` or `summary(MHW$grain)` or `summary(grain)`
➢ `summary(MHW[,4])` or `summary(MHW$straw)` or `summary(straw)`

B) Obtain the various 'statistics' for the `grain` yield and do the same for `straw`
➢ `min(grain)`
➢ `max(grain)`
➢ `mean(grain)`
➢ `median(grain)`
➢ `var(grain)`
➢ `sd(grain)`
➢ `quantile(grain)`

➢ IQR(grain)

C) Define and calculate a new variable 'yield.ratio' which is the ratio between `grain` and `straw` and obtain the summary statistics for it.

➢ yield.ratio=grain/straw

Task #4: Visualize the data:

A) Obtain a stem and leaf display of the `grain` and of the `straw` yields

➢ stem(MHW$grain) or stem(MHW[, 3]) or stem(grain)

B) Obtain a simple histogram of the `grain` and of the `straw` yields

➢ hist(grain)

C) Obtain a fancy histograms of these data

➢ hist(straw, nclass=30, col = "lightblue", border = "red", main = "The MHW Data", xlab = "Straw yield per plot")

➢ hist(grain, breaks = seq(2.6, 5.2, by = 0.1), col = "lightblue", border = "red", main = "The MHW Data", xlab = "Grain yield per plot")

- **THE MHW-Data—**
  In the early days of scientific agriculture, Mercer and Hall [1] were trying to determine the optimum plot size for agricultural yield trials: Plots that are too *small* will be too variable; and plots that are too *large* waste resources (land, labor, seed); if the land area is limited, the number of treatments will be unnecessarily small.

  So, they performed a very simple experiment: an apparently homogeneous field was selected, prepared as uniformly as possible and planted to the same variety of wheat. They attempted to treat all parts of the field exactly the same in all respects during subsequent farm operations. When the wheat had matured, the field was divided into 500 equally-size plots. Each plot was harvested separately. Both grain and straw were air-dried, then hand-threshed and weighed to a precision of 0.01 lb (= 4.54 g). The reported values are thus air-dry weight, in lb per plot.

  The field was a square of 1 acre, which is 0.40469 hectare or 4,046.9 m2, which was divided

into a 20 rows by 25 columns, giving 500 plots, each of 1/500 acre, which is about 8.09 m2 (3.30 m long x 2.45 m wide). We can assume that the rows ran W to E, with 25 plots in each row, beginning at 1 on the W and running to 25 at the E, so that columns run N to S with 20 plots in each, running from 1 at the N to 20 at the S. Thus the NW corner (1,1) is plot 1, the NE corner (1, 25) is plot 481, the SE corner (25, 20) is plot 500, and the SW corner (1, 20) is plot 20.

The CSV Data File: The data has been prepared as the comma-separated values ("CSV") file mhw.csv in a plain-text editor. The first line gives the four field names:
"r","c","grain","straw" standing for:
> **r** : Row number in the field
> **c** : Column number in the field
> **grain** : Grain yield, lbs per plot
> **straw** : Straw yield, lbs per plot

Each of the 500 lines in the data file represents a plot; the four data fields are separated by commas. For example, the first data line is: 1,1,3.63,6.37, the second data line is 2, 1, 407, 6.24 and so on…

r	c	grain	straw
1	1	3.63	6.37
2	1	4.07	6.24
3	1	4.51	7.05
4	1	3.9	6.91
5	1	3.63	5.93
6	1	3.16	5.59

[1] W B Mercer and A D Hall. The experimental error of field trials. *The Journal of Agricultural Science (Cambridge)*, 4:107–132, 1911.