Assignment 1 – Data Analytics

0. Download and move the file named “heights_weights_genders2.csv” into your working directory. Open the file in an editor to see the structure of the file (does it have column names? if it has, don’t forget to set “header” argument True. What is the separator? Etc.)
1. Load the dataset into R and assign it to a variable named “df”.
   a. Show first 10 and last 10 rows of the dataset.
   b. How many observations (rows) and variables (columns) does the dataset have? (Hint. str(), dim(), nrow(), ncol())
   c. What are the types of variable of each column? (logical? Character? Numeric? Factor?)
   d. Use summary() function to see a simple quantitative statistics of the dataset. What are the maximum and minimum values of the “weight” and “height” columns?
   e. It seems there is another unwanted value than “female” and “male” in the “gender” column. Using function complete.cases() can you say which rows of the dataset have missing values? (Hint. Use which() to return the indices of rows that have missing values)
   f. Clean the dataset from any missing values and assign it to another variable. Now check the number of rows of the new variable. How many rows were removed?
   g. Sort the cleaned dataset by “height” column and use print() function to print the outcomes.
   h. [Bonus] Calculate Body Mass index for each observant using this formula

      BMI = (Weight in Pounds / (Height in inches) x (Height in inches)) x 703

      then add a new column to the dataset showing BMI of each person. (don’t use any loop!)
2. Use histogram to show distribution of the weights and heights of the dataset. Increase the number of breaks to have a plot with more bins. (Hint: use function Hist())
   a. Using plot() function, plot a two dimensional graph that its X-axis represents weights and Y-axis represents heights of observations. Do you see any relationship between these two?
   b. Calculate average heights of women and men separately and plot these two values by barplot() as two bar. ([Hint. df[df[,'Gender']=="Male",??])
   c. change the colors of these two bars to “blue” and “green”