***Tutorial 8– Data Exploration***

1. Using *diamonds* data set is a data set in the *ggplot2* package which containing the prices and other attributes of almost 54,000 diamonds. Using this data set and *ggplot2* package do the following tasks:
   1. For the diamond colour (*color*), identify the most and least common values and their counts.
   2. For the weight of the diamond (*carat*), use the histogram to plot its distribution then identify the peak(s) and the outlier(s).
   3. For the weight of the diamond (*carat*), use the boxplot to plot its distribution then identify the outlier(s).
2. Create a new vector with the following data: 1,2,3,4,NA,6,7,8,NA,NA. NA means ‘Not Available’ / Missing Values. Use *min*, *max*, and *mean* functions to get the minimum, maximum, and average, respectively for this vector. Try using the argument *na.rm=TRUE* withthese three functionsand re-print the results.
3. Create a data frame with the following two columns:

minutes=c(1,1,1, 3,4, 4, 4,4,5,5,5,5)

sensorReading=c(10,NA,NA,5,NA,11,10,8,7,NA,8,9)

Replace NA by the mean/median.

1. What happens to missing values in a histogram and bar chart?