Lab Exercises

1. Given Orders.csv file in the folder datasets. Import the file into R and generate the following outputs:
   a. Display the meta-data of its data frame object

```
'data.frame':   70 obs. of  4 variables:
 $ Order.ID         : Factor w/ 70 levels "32 90 001","32 90 002",..: 1 2 3 4 5 6 7 8 9 10 ...
 $ Order.Date       : Factor w/ 67 levels "1-Jan-12","1-Mar-11",..: 63 66 31 37 42 2 64 35 44 59 ...
 $ Place.of.Shipment: Factor w/ 19 levels "Ahmednagar","Aurangabad",..: 12 11 1 10 6 3 11 12 7 19 ...
 $ Payment.Terms    : Factor w/ 3 levels "Cash","Cheque",..: 2 3 1 2 1 3 3 3 3 3 ...
```

   b. Descriptive Statistics:

```
      Order.ID      Order.Date   Place.of.Shipment Payment.Terms
32 90 001: 1   13-Jan-13: 2   Ratnagiri  : 8      Cash  :18
32 90 002: 1   14-Apr-13: 2   Dhule      : 5      Cheque:23
32 90 003: 1   19-Aug-11: 2   Raigad     : 5      Online:29
32 90 004: 1   1-Jan-12 : 1   Sindhudurga: 5
32 90 005: 1   1-Mar-11 : 1   Aurangabad : 4
32 90 006: 1   1-Mar-13 : 1   Nasik      : 4
(Other)  :64   (Other)  :61   (Other)    :39
```

2. Import the fileBollywood_2015.csv and generate the following outputs for variable BO_COLLECTION:

   a. Binning the data:

```
 [1] (0.78,67.1] (0.78,67.1] (67.1,133]  (0.78,67.1] (0.78,67.1] (0.78,67.1] (0.78,67.1] (0.78,67.1]
 [9] (0.78,67.1] (67.1,133]  (0.78,67.1] (0.78,67.1] (0.78,67.1] (67.1,133]  (0.78,67.1] (265,331]
[17] (0.78,67.1] (67.1,133]  (0.78,67.1] (0.78,67.1] (67.1,133]  (0.78,67.1] (0.78,67.1] (0.78,67.1]
[25] (0.78,67.1] (67.1,133]  (133,199]   (0.78,67.1] (67.1,133]  (67.1,133]  (0.78,67.1] (0.78,67.1]
[33] (0.78,67.1] (0.78,67.1] (0.78,67.1] (0.78,67.1] (0.78,67.1] (0.78,67.1] (0.78,67.1] (0.78,67.1]
[41] (0.78,67.1] (0.78,67.1] (0.78,67.1] (0.78,67.1] (0.78,67.1] (67.1,133]  (0.78,67.1] (0.78,67.1]
[49] (0.78,67.1] (265,331]   (0.78,67.1] (0.78,67.1]
Levels: (0.78,67.1] (67.1,133] (133,199] (199,265] (265,331]
```

   b. Generating the frequency counts

```
(0.78,67.1]  (67.1,133]  (133,199]  (199,265]  (265,331]
         40           9           1           0           2
```

   c. Precentages of corresponding frequency counts:

```
(0.78,67.1]  (67.1,133]  (133,199]  (199,265]  (265,331]
  76.923077   17.307692    1.923077    0.000000    3.846154
```

3. Draw a random sample with replacement of size 80 from the values of the variable Qty from dataset Ord_Details.csv

4. Load the file events.RData. Data frame object **events** will get loaded in the memory. The first 4 observations have been shown below :

	eventID	occur
1	1	12
2	2	14
3	3	15
4	4	34

You need to create a column named **dur** which will calculate the difference in two consecutive values in the variable **occur** as shown below:

	eventID	occur	dur
1	1	12	12
2	2	14	2
3	3	15	1
4	4	34	19

5. Create a function which accepts a number as temperature in degrees Fahrenheit and returns the temperature in Celsius

$$C = (F - 32) * \frac{5}{9}$$

6. Create a function which accepts a numeric variable(input) and outputs a data frame containing two calculated values namely, mean(input) – 2*SD(input) and mean(input) +2*SD(input). Take care that null values get removed before calculation

7. Create a function which accepts a numeric variable(input)and outputs its coefficient of variation with formula: (SD / mean) *100

8. Create a function which can impute the missing values in a numeric vector by mean. That means, the function should take a numeric vector as an argument and return a numeric vector in which NA values will be found imputed. e.g.

```
> g <- c(9.34,8.24,NA,1.345,0.56,0,NA,7.89)
> imputeMean(g)
[1] 9.3400 8.2400 4.5625 1.3450 0.5600 0.0000 4.5625 7.8900
```