Unique Paper Code	: 42343308
Name of the Course	: B.Sc. (Prog.) Physical Science/Mathematical Science
Name of the Paper	: Introduction to R Programming (SEC)
Semester	: III
Year of Admission	: 2019 and onwards

Duration: 3 Hours   Maximum Marks: 75

Attempt any four questions
All questions carry equal marks.

Q1.   Read the following CSV file (playerdata.csv) with header into a data frame in R (where first row representing a header).

playerdata.csv

Player_ID	Name	Age	Height (in cms)	Weight (in kg)	Club	Net potential
158023	L. Messi	28	170	72	FC Barcelona	95
20801	Cristiano Ronaldo	30	185	80	Real Madrid	93
9014	A. Robben	31	180	80	FC Bayern	90
167495	M. Neuer	29	193	92	FC Bayern	NA
176580	L. Suajrez	28	182	85	FC Barcelona	90
183277	E. Hazard	24	173	74	Chelsea	91
41236	Z. Ibrahimovia	33	195	95	Paris Saint-Germain	89

Write appropriate R commands to perform following operations on the obtained data frame:

- Calculate Pearson correlation between 'Height' and 'Weight' of the players.
- Display the names of the players who are associated with 'FC Barcelona' club.
- Write an SQL query in R to print ID of players ('Player_ID') whose 'Age' is in the range 30-34.
- Split the values in 'Name' column on the basis of '.'.
- Determine names of the 'Club' starting with letter 'F'.
- Create a Scatterplot for height and weight of the players titled "Weight vs Height".
- Draw a horizontal bar chart of age of the players with their names displayed on the Y axis.
- Remove missing values from the given data and determine player with highest 'Net potential'.

Q2.   There are twenty students enrolled in a course with marks (in percentage) given below:

90,91,91,92,72,82,83,83,83,93,74,74,74,95,94,84,58,54,30,89

  ▪ Store the given marks in a vector *S_Marks*.
  ▪ Create a user defined function in R to display the mode of the vector *S_Marks*.
  ▪ Use R commands to grade each student according to the following conditions:

        90% and above - Grade A
        80% to 89% - Grade B
        60% to 79% - Grade C
        Less than 60 % - Grade D

  ▪ Write a function in R to remove the duplicate elements from the vector *S_Marks*.
  ▪ Write R command/s to display Boxplot of the obtained marks and determine outlier (if any).

Q3.   Write R commands to create a: - vector *V* of names of months of a year; 2 X2 matrix *M;* and a function *F* to calculate natural log of a number n. Using *V*, *M* and *F*, create a list *L_Combined* in R and print its contents.

   Consider two matrices *X* and *Y* given as follows:
        X = matrix(1:12, ncol=3)
        Y = matrix(13:24, ncol=3)
   Write an R program to concatenate *X* and *Y* with same columns but different rows. Also, write R command/s to access the following: -
   ▪ the element at 3rd column and 2nd row of *X*
   ▪ the 3rd row of *X*
   ▪ 4th column of *Y*
   Convert matrix *X* into a list *L_New* and count the number of elements in *L_New* using R commands.

Q4.   Using *sample ()* function, simulate rolling of 6-sided dice (having numbers between 1 and 6) ten number of times. Calculate frequency of each value generated.

   Create 4 *X* 4 matrix M of first 16 natural numbers. Write a script in R to access and sum all the elements of matrix M using 'for' loop. Write R command to calculate transpose of the same matrix.

   Consider $m < - c(14, 26, 45, NA, NA, 4, 87, NA, 28, NA, 34)$. Which R-statement will count the number of NA values in m?

Q5.   Write an R command to store "Newtownabbey" in a string variable '*name_str'*. Count the number of vowels in '*name_str'* using available functions in *stringr* package.

   Consider $x < - c(8,9,15,17,8,9,2,3,1,10)$ and $y < - c(5,6,4,3,12,14,17,8,9,10)$.

   What will be the output of $pmin(x, y), pmax(x, y), cummin(x), cumsum(x), rank(x)$? Write R command/s to see how many values of vector x falls in each of the bins (0,4] (4,8], (8,12], (12,16], (16,20]. Determine the 5th highest value in vector y.

Q6. Write an R program to find the multiplicative factors of a given number N (e.g., multiplicative factors of N=6 are 1,2,3,6).

Consider the information given below about height (H) of students in (cms).

$H = c(152,175,188,176,153,142,123,112,152,70,154,123,110,132,176)$

Write R commands to answer the following:

- Convert 'H' into 2 factor levels of short and tall with 'H' <= 140 cms as short and 'H' >140 cms as tall.
- Arrange height of students ('H') in decreasing order.
- Create a contingency table of 'H'.
- Create a histogram for given height vector ('H') with 5 bins.
- Display five-point statistical summary of 'H'
- Differentiate between H%% 2 and H % / % 2.
- Convert 'H' into an array 'height_array' and access the 8th element of 'height_array'.