Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

(A constituent unit of MAHE, Manipal)

VI SEMESTER B.TECH. MAKE UP EXAMINATIONS, JUNE 2019
SUBJECT: PROGRAM ELECTIVE - DATA SCIENTIST'S TOOLBOX AND R PROGRAMING [CRA 4009]
REVISED CREDIT SYSTEM
(14/06/2019)

Time: 3 Hours MAX. MARKS: 50

Instructions to Candidates:

- Answer **ALL** the questions.
- Missing data, if any, may be suitably assumed.

1A. Write R code to perform the following: **5**

(a) Sum a matrix over all the columns using apply()
(b) Convert strings in a vector from lower case to upper case using lapply()
(c) Find number of characters in each string in a vector using sapply()
(d) Find the sum of first elements, second elements upto last elements across three vectors using mapply()
(e) In iris dataset, Species is a factor with 3 values namely Setosa, versicolor and virginica. Find the mean of sepal length of these 3 species(subsets) using tapply().

1B. Explain lubridate package used to work with dates in R. Write a function to convert x=c("1jan1960", "2jan1960"); to "1960-01-01" "1960-01-02". **3**

1C. Explain download.file() method to get data from internet in R. **2**

2A. List down the steps to convert raw data to tidy data. Explain the components of tidy data? **5**

2B. Define environment in R. Demonstrate the difference between lexical and dynamic scoping rules in R. **3**

2C. How can you specify variable number of arguments in the function definition? Illustrate with an example. **2**

3A. What are the properties of *dplyr* package? Briefly explain any five *dplyr* verbs. **5**

3B. Explain the following debugging tools with an example for each: **3**

(i) Browser()
(ii) debug()

3C. Explain system.time() method in code profiling? Give one example for each where *user time* > *elapsed time* and *user time* < *elapsed time* for computing tasks. **2**

4A. What are the different data types/objects in R? Explain each with suitable examples. **5**

4B. With an example explain the difference between sub() and gsub() methods used to edit text variables. **3**

4C. Write the commands to configure username and email in git. **2**

5A. Why is it essential to set the random number seed in probability distribution function? Explain with an example, various probability distributions functions in R. **5**

5B. What is the use of the following commands in R? Explain with an example.

(i) pwd

(ii) touch

(iii) rm

(iv) echo

(v) mv **3**

5C. Name and explain 4 parameters that can be passed to read.table() method used to load flat files. **2**