MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL
(A constituent unit of MAHE, Manipal)

VI SEMESTER B.TECH
END SEMESTER EXAMINATIONS, APRIL/MAY 2019
SUBJECT: PROGRAM ELECTIVE - INTRODUCTION TO DATA SCIENCE
[CRA 4010]
REVISED CREDIT SYSTEM
(03/05/2019)

Time: 3 Hours   MAX. MARKS: 50

> **Instructions to Candidates:**
> ❖ Answer **ALL** the questions.
> ❖ Missing data if any, may be suitably assumed.

1A. Explain in detail the principles of Analytic Graphics. — 5

1B. List and explain the basic components of ggplot2. — 3

1C. Differentiate between the following exploratory graphs.

   i) Boxplot and Barplot

   ii) Histogram and Density plot — 2

2A. What is the significance of graphics devices in R? Explain vector and bitmap file devices in detail. — 5

2B. Discuss the significance of Principle Component Analysis and Single Value Decomposition in exploratory data analysis — 3

2C. Explain the following Base plotting function parameters.

   i) pch
   ii) lty
   iii) lwd
   iv) xlab — 2

3A. What is the difference between hierarchical clustering and k-means clustering? Explain how similarity/distance is defined between the data points in hierarchical clustering? State the significance of dendrogram in hierarchical clustering. — 5

3B. Give the five-number summary for the data having attribute age with values 22, 23, 22, 24, 25, 45, 33, 23, 44, 32, 56, 33, 21, 20, 46, 43, 95, 56, 43. — 3

3C. Explain the benefits of knitr package. — 2

4A. Suppose that the data mining task is to cluster points into two clusters, where the points are 2, 3, 4, 10, 11, 12, 20, 25, 30. The distance function is the Euclidian distance. Suppose initially we assign 4 and 12 as the center of each cluster, respectively. Use k-means algorithm to find final clusters and its centroids after convergence. — 5

4B. List and explain the checklists that we need to avoid in reproducible research.   3

4C. What are the basic guidelines for writing code in R?   2

5A. List and explain the hierarchy of information for:

  i)   Research paper

  ii) Email presentation   5

5B. Draw the schematic diagram of research pipeline.   3

5C. Write the Markdown syntax for:

  i)   Italics

  ii) bold   2