Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

(A constituent institution of MAHE, Manipal)

VI SEMESTER B.TECH END SEMESTER EXAMINATIONS, APRIL 2018

SUBJECT: INTRODUCTION TO DATA SCIENCE [CRA 4010]

REVISED CREDIT SYSTEM
(26/04/2018)

Time: 3 Hours MAX. MARKS: 50

Instructions to Candidates:

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

1A.	Write and Explain the Principles of Analytic Graphics.	5
1B.	Briefly explain the different plotting systems in R.	3
1C.	What are the two methods used for merging of points in Hierarchical clustering in R?	2
2A.	Explain K-Means Algorithm. Assume, you want to cluster 7 observations into 3 clusters using K-Means clustering algorithm. After first iteration clusters, C1, C2, C3 has following observations: C1: {(2,2), (4,4), (6,6)} C2: {(0,4), (4,0)} C3: {(5,5), (9,9)} What will be the Euclidean distance for observation (9, 9) from cluster centroid of C1, C2 and C3 in second iteration?	5
2B.	Explain the different types of palettes available for RColorBrewer Package.	3
2C.	Outline the uses of rgb function in R.	2
3A.	Write and explain salient features of qplot() used in ggplot2.	5
3B.	What are the ways you suggest to organize data for analysis that can be used to put things together in logical places and ultimately to ensure that data analysis is reproducible either by yourself or by others.	3
3C.	What are the limitations of original sweave system? How knitr package updates and improves on them? How is knitr system different from sweave?	2
4A.	List and explain briefly the steps in data analysis process. Illustrate with sample examples to boost your theory.	5
4B.	Explain the idea behind literate statistical programming? Mention the pros and cons of literate programming.	3
4C.	What are the coding standards that make code more readable and usable by others?	2

5A. Explain the cacher package for R. How cacher can be used as
   i) An author.
   ii) A reader. 5

5B. How do you distinguish between the words replication and reproducibility with their usage? 3

5C. What is reproducible research? What problems does this reproducibility solve? 2