Guidelines for Introduction to R Programming (SEC)

Semester-4, B.Sc.(Hons.) Computer Science

Unit	Chapter	Reference	Hours
Unit 1 Introduction: R interpreter, Introduction to major R data structures like vectors, matrices, arrays, list and data frames, Control Structures, vectorized if and multiple selection, functions.	Ch-1	Ref 1	(Practical Only)
	Ch-2	Ref 1	02
	Ch-3	Ref 1	02
	Ch-4	Ref 1	03
	Ch-5	Ref 1	03
	Ch-6	Ref 1	02
	Ch-7 (up to Dropping factors)	Ref 1	02
	Ch-8	Ref 1	01
	Ch-9 (up to pg 135 before mapply)	Ref 1	02
	Ch-10	Ref 1	Practical Only
Unit 2 Installing, loading and using packages: Read/write data from/in	Ch-12 (Up to pg 174)	Ref 1	2

files, extracting data from websites, Clean data, Transform data by sorting, adding/removing new/existing columns, centering, scaling and normalizing the data values, converting types of values, using string in-built functions, Statistical analysis of data for summarizing and understanding data, Visualizing data using scatter plot, line plot, bar chart, histogram and box plot	Web API		01
	https://www.tylerclavelle.com/code/2017/randapis/		
	Ch-13 (up to Sorting on Pg 201-202 – excluding Functional Programming)	Ref 1	03
	Ch-14 Scatter Plot up to pg 214 Line plot pg 230 – 232 Histogram pg 238 – 241 BoxPlot pg 249 – 250 upto Fig. 14.38 Bar Charts Pg 252-255 Fig. 14.42	Ref 1	03
Unit 3 Designing GUI: Building interactive application and connecting it with database.	https://www.r-bloggers.com/2010/10/creating-guis-in-r-with-gwidgets/		Practical only
	Ch-12 (pg 185 – 187 (excluding ODBC database connection))	Ref 1	02
Unit 4 Building Packages.	Ch-17 (up to Your First Package; till Pg 317)	Ref 1	02

Reference

1. Cotton, R., Learning R: a step-by-step function guide to data analysis. 1st edition. O'reilly Media Inc.

2. Lawrence, M., & Verzani, J. (2016). Programming Graphical User Interfaces in R. CRC Press. (ebook)

List of Practicals

Introduction to R Programming (SEC)

B.Sc. (Hons.) Computer Science

Q1.   Write the R commands to do the following:

    a. Create a numeric, character, integer and logical vector.

    b. Create two numeric vectors x and y of different lengths. Next, observe the output of 2*x + y - 3.

    c. Create a sequence of all the odd numbers between 100 and 500.

    d. Calculate the mean and standard deviation of sequence created in part c.

    e. Create a list containing a vector, an array and a list.

    f. Create two multiplication compatible matrices M1 and M2 and display their product.

    g. Display row-wise and column-wise mean of the matrix M1.

    h. Combine the matrices M1 and M2 (as in previous part) using rbind/cbind function.

    i. Convert a 3x4 matrix into 2x4 matrix.

Q2.   Use the *airquality* dataset and write R commands to do the following:

    a. Display the structure of the dataset.

b. Display the number of observations and variables in this dataset.

c. Display the count of complete rows.

d. Display the data for the observation having *Temp* between 70 and 80.

e. Display the count of observations where *Month* = 5.

f. Display first 10 rows of the dataset.

g. Display summary of *Wind* variable.

h. Display the complete dataset sorted as per the decreasing order of *Temp* variable.

i. Add a new observation to this dataset.

Q3. Download the *Toy Dataset* CSV file available on Kaggle via the link given (

https://www.kaggle.com/carlolepelaars/toy-dataset). Write R commands to do the following:

a. Read the CSV into a data frame object.

b. Attach the data frame to the environment.

c. Count the number of rows with *City* = *New York City*.

d. Display rows that have top 5 income values.

e. Find out the number of rows with income greater than average income.

f. Find the highest salary for the female population.

g. Randomly select 10 rows from the given dataset.

h. Detach the data frame object from the environment.

Q4. Write R commands to do the following:

a. Create a sample STUDENT table in the MySQL database.

b. Insert 5 rows into the STUDENT table.

c. Display all the information stored in STUDENT table.

d. Delete all the rows from STUDENT table.

e. Drop the STUDENT table.

Q5. Using the *rainfall* data for the year 2015 available on the link https://data.world/rajanand/rainfall-in-india or https://www.kaggle.com/datasets/rajanand/rainfall-in-india

Write an R script to:

a. Read the first ten days of rainfall amounts from the CSV file.

b. Create a subset of the rainfall data where rain is larger than 20.

c. Find the mean rainfall for days where the rainfall was at least 4.

d. Subset the vector where rainfall is either exactly zero, or exactly 0.6.

e. Use a suitable plot to compare the rainfall of each day of every month, and to compare the summarized rainfall of every month.

Q6. Load the built-in *iris* dataset and write the R commands/script for the following:

a. Display the summary statistics for all the variables in the dataset.

b. Plot a scatter plot for Sepal length and Sepal width for the species "setosa".

c. Annotate the above graph (a) with appropriate X-axis, Y-axis labels and Main title.

d. Plot a histogram for the variable Petal.length.

e. Create a boxplot for Petal width of Species=versicolor.

Q7. Create a tab-delimited file containing the data as shown in the table below and then read this file into a data frame. Write an R script to do the following:

Beer	Wine	Spirit	Country	Year
5.24	2.86	1.81	Australia	1998
5.15	2.87	1.77	Australia	1999
5.06	2.94	1.88	NewZealand	2002
5.07	2.95	2.07	Australia	2001
4.8	2.91	1.81	NewZealand	1999
4.97	3.01	1.86	NewZealand	2000
4.68	3.07	2.06	Australia	2004
4.58	3.13	2.12	NewZealand	2003
4.57	3.11	2.15	Australia	2006
4.49	2.59	1.77	NewZealand	1998
4.26	2.65	1.64	NewZealand	2004

a) For each year, display the country with maximum Beer consumption and minimum Wine consumption.

b) Find the average alcohol consumption for each country.

c) Find the Beer consumption for New Zealand for the year(s) when its spirit consumption was more than the mean spirit consumption.

d) Find the average Beer consumption/Wine Consumption/Spirit consumption for each year.

e) Plot the mean Beer consumption/mean Wine consumption/mean Spirit consumption for each country using a suitable plot.

Q8. Create a package in R to perform certain basic statistics functions.

Q9. Create a database songs that contains the fields {song_name, mood, online_link_play_song}. Create an application where the mood of the user is given as input and the list of songs corresponding to that mood appears as the output. The user can listen to any song form the list via the online link given.