|  |  |
| --- | --- |
|  | **Faculty of Engineering and Technology**  **JGI Global Campus**  **Jakkasandra Post, Kanakapura Taluk, Ramanagara District, 562112** |

**Continuous Assessment Book**

**B.TECH (HONOURS) IN COMPUTER SCIENCE (DATA SCIENCE)**

Batch: 2018-2022

Academic Session: Jan – June 2020

|  |  |  |
| --- | --- | --- |
| **Course Code** | **Course** | **Semester** |
| **18CS4DS03L** | **R Programming Language Lab** | 4 |

**Faculty**

1. **Asst. Prof. Ashwini Kumar Mathur**

**LIST OF Experiments**

|  |  |
| --- | --- |
| **Exp. No** | **Experiment Name** |
| **LAB - 1** | 1. Install and configure R, set working directory. 2. Install Packages and calling installed packages 3. R studio environment and functionalities of R studio 4. Explore assignment, comments, and case sensitivity in R. 5. Use some built-in functions such as pi, sqrt, log, round, factorial, trigonometric functions, ect. From R. 6. Create R script files and call it from R console. 7. Explore for getting help, examples, and demonstrations on R. 8. Use R as calculator. 9. Explore different mathematical operators. |
| **LAB - 2** | 1. Explore assignment operator. 2. Create vectors using c(), seq(), rep(), colon operator. 3. Create different matrices using matrix() operator and explore its rows, columns, and diagonals. 4. Perform different basic operation of matrices on above created matrices. 5. Create single and multidimensional arrays with array() command. 6. Explore length(), dim(), ncol(), nrow() operators on above matrices and arrays. 7. Explore commands for Selecting and extracting elements from above matrices and arrays. 8. Explore logical operators from R programming language. 9. Remove elements from selected positions from a considered matrix.   **Lab -3** |
| **LAB - 3** | 1. Create some complex data structure variables such as list and data frames using list() and data.frame commands. 2. Create data using data.frames, lists, and tables. 3. Implement basic R operations (data input, missing values, Importing data into R using different formats : xlsx, CSV, Text files). 4. Explore data type conversions from one data structure to another with commands such as as.data.frame(), as.vector(), is.data.frame(), is.vector; and find the data type with class() command. 5. Explore function programming in R. 6. Explore loops in R programming such as if-else-ifelse, for, while, repeat-break, ect. |
| **LAB - 4** | 1. Calculate the interest earned after 5 years on an investment of $2000, assuming an interest rate of 3% compounded annually. 2. Use R to calculate the area of a circle with radius 7 cm. 3. Do you think there is a difference between 48:14ˆ2and 48:(14ˆ2)? 4. Usin grep() and seq()as needed, create the vectors:   0000011111222223333344444 and 1234512345123451234512345   1. Create the vector   ## [1]00011110001111000111100011110001111  and convert it to a factor. Identify the levels of the result, and then  change the level labels to obtain the factor:  ## [1] Male Male Male Female Female Female Female Male Male  ## [10] Male Female Female Female Female Male Male Male Female  ## [19] Female Female Female Male Male Male Female Female Female  ## [28] Female Male Male Male Female Female Female Female  ## Levels: Male Female   1. Use more.colors vector, rep() and seq() to create the vector   "red" "yellow" "blue" "yellow" "blue" "green"  "blue" "green" "magenta" "green" "magenta" "cyan" |
| **LAB - 5** | 1. Assume 4- binary digit accuracy for the following computations. 2. Write out the binary representation for the approximate value of 6/7. 3. Write out the binary representation for the approximate value of 1/7. 4. Add the two binary representations obtained above, and convert back to the decimal representation. 5. Compare the result of part (c) with the result from adding the binary representations of 6 and 1, followed by division by the binary representation of 7. 6. In R, evaluate the expressions   2^52 + k – 2^52  2^53 + k – 2^53  2^54 + k – 2^54,  for the cases where k = 1, 2, 3, 4. Explain what you observe. What could be done to obtain results in R which are mathematically correct?   1. The following are a sample of observations on incoming solar radiation at a greenhouse:   11.1 10.6 6.3 8.8 10.7 11.2 8.9 12.2   1. Assign the data to an object called solar.radiation. 2. Find the mean, median, range, and variance of the radiation observations. 3. Add 10 to each observation of solar.radiation, and assign the result to sr10. Find the mean, median, range, and variance of sr10. Which statistics change, and by how much? 4. Plot a histogram of the solar.radiation, sr10, and srm2. |
| **LAB - 6** | * + - 1. What happens when the following code is run?   gender <- c("M" , "M" , "F" , "F" , "F" )  whereF<- (gender == "F" )  gender[whereF] <- "Female"   * + - 1. Construct the data frame char-num in the following way:   char <- c("2" , "1" , "0" )  num<- 0:2  charnum<- data.frame(char, num);  Explore sampling distribution and central limit theorem in R   * + - 1. Use the inbuilt data car and uses the possible graphical plots using ggplot2 graphical packages.       2. Modify the code to generate the Fibonacci sequence in the following ways.  1. Change the first two elements to 2 and 2. 2. Change the first two elements to 3 and 2. 3. Change the update rule from summing successive elements to taking differences of successive elements. For example, the third element is defined as the second element minus the first element, and so on. |
| **LAB - 7** | In each of the following, determine the final value of answer. Check your result by running the code in R.   1. answer <- 0 for (j in 1:5) answer <- answer + j 2. answer <- NULL for (j in 1:5) answer <- c(answer, j) 3. answer <- 0 for (j in 1:5) answer <- c(answer, j) 4. answer <- 1 for (j in 1:5) answer <- answer \* j)   Does the Eratosthenes() function work properly if n is not an integer? Is an error message required in this case?  Use the idea of the Eratosthenes() function to prove that there are infinitely many primes. Hint: suppose all primes were less than m, and construct a larger value n that would not be eliminated by the sieve.  Write an R function called compound.interest() which computes this amount. Your function should have three arguments.  Consider the inbuilt data set “cars”.   1. Find Correlation between possible variables and pairwise correlation. 2. Find regression line between appropriate variables. 3. Display the summary statistics and comment on the results. |
| **LAB - 8** | 1. Plot a circle by (x,y) points, where x = r · cos(a) and y = r · sin(a), with a the angle, from 0 to 2π, and r the radius. 2. Using the same plot, add a number of graphical parameters that specify: 3. Rather than dots, the points should be connected by lines (type). 4. The line should be twice as wide as the default (lwd). 5. Give the appropriate x- and y-axes labels (xlab, ylab). 6. Give the axes and axes annotations (axes). 7. The graph has to be symmetrical, i.e. the x/y aspect ratio = 1 (asp). 8. Add the legend, that describe the cure. 9. Explore for plotting multiple figures in a single window with partition as par() command as follows   > par (mfrow=c(2,2))  > for ( i in 1:4) curve(sin(i\*pi\*x),0,1,main=i)   1. For the US, the population density in 1900 (N0) was 76.1 million; the population growth can be described with parameter values: a=0.02 yr−1, K = 500 million of people.   Actual population values are:  1900 1910 1920 1930 1940 1950 1960 1970 1980  76.1 92.4 106.5 123.1 132.6 152.3 180.7 204.9 226.5   1. Plot the population density curve as a thick line, using the US parameter values. 2. Add the measured population values as points. Finish the graph with titles, labels etc. |
| **LAB - 9** | 1. 1. famous inbuilt data set that is part of R is the ”iris” data set (Fisher, 1936). It gives measurements, in centimeters for sepal length and width and petal length and width, respectively, for 50 flowers of the species Iris setosa, Iris versicolor and Iris virginica. Have a look at the data: 2. What is the class of the data set? Why? 3. What are the dimensions of the data set? (number of rows, columns) 4. Produce a scatter plot of petal length against petal width; produce an informative title and labels of the two axes. 5. Repeat the same graph, using different symbol colors for the three species. 6. Add a legend to the graph. Copy-paste the result to a WORD document. If you do not have WORD, make a PDF file of the graph. 7. Create a box-and whisker plot for sepal length where the data values are split into species groups; use as template the first example in the ”boxplot” help file. 8. Now produce a similar box-and whisker plot for all four morphological measurements, arranged in two rows and two columns. First specify the graphical parameter that arranges the plots two by two. 9. Write a script file that solves the following system of ODEs.   for initial values x=300,y=10 and parameter values: a=0.05, K=500, b=0.0002, g=0.8, e=0.03.   1. Make three plots, one for x and one for y as a function of time, and one plot expressing y as a function of x (this is called a phase-plane plot). Arrange these plots in 2 rows and 2 columns. 2. Now run the model with other initial values (x=200, y=50); add the (x,y) trajectories to the phase-plane plot |
| **LAB - 10** | 1. Use R-function matrix to create the matrices called A and B: 2. Take the inverse of A and the transpose of A. 3. Multiply A with B. 4. Estimate the eigenvalues and eigenvectors of A. 5. For a matrix A, x is an eigenvector, and λ the eigenvalue of a matrix A, if A · x = λ · x. Test it! 6. What is the value of the largest eigenvalue (the so-called dominant eigenvalue) and the corresponding eigenvector? 7. Create a new matrix, T, which equals P, except for the first row, where the elements are 0. 8. Now estimate N= (I-T)-1, where I is the identity matrix. 9. Find the root of the equation *ex = 4x2* in the interval [0, 1]. And draw the function curve. 10. Solve the equations *1000 = y ∗ (3 + x) ∗ (1 + y)4* for y and with x varying over the range from 1 to 100. Plot the root as a function of x. |