UNIT IV DATA EXPLORATION AND DATA VISUALIZATION IN R

PART - A Questions

1. What is meant by R scripts? (CO4 , K3)

A script is simply a series of commands that are run as a group rather than entering and running the code one line at a time from the Console window.

2. Why loops are not often used in R? (CO4 , K3)

Loops need not be used in R because of the built-in support for vectorization.

3. List down the panes in quadrant 1 in Rstudio. (CO4 , K3)

- Files Pane – (Q1)
- Plots Pane – (Q1)
- Packages Pane – (Q1)
- Help Pane – (Q1)
- Viewer Pane – (Q1)

4. What are the functionalities provided by files pane? (CO4 , K3)

i. Delete files and folders
ii. Create new folders
iii. Rename folders
iv. Folder navigation
v. Copy or move files
vi. Set working directory or go to working directory
vii. View files
viii. Import datasets

5. What is the need of viewer pane? (CO4,K3)

RStudio includes a Viewerpane that can be used to view local web content. For example, web graphics generated using packages like googleVis, htmlwidgets, and RCharts, or even a local web application created with Shiny.

UNIT IV DATA EXPLORATION AND DATA VISUALIZATION IN R

6. What does History pane display? (CO4, K3)

- The History pane displays a list of all commands that have been executed in the current session.
- This tab includes a number of useful functions including the ability to save these commands to a file or load historical commands from an existing file.
- It can also select specific commands from the History tab and send them directly to the console or an open script. It can also remove items from the History pane.

7. What is the function of Source pane? (CO4, K3)

- The Sourcepane in RStudio is used to create scripts and display datasets An R script is simply a text file containing a series of commands that are executed together.
- Commands can also be written line by line from the Console pane as well. When written from the Consolepane, each line of code is executed when you click the Enter (Return) key. However, scripts are executed as a group.

8. Why is terminal pane needed? (CO4, K3)

- The RStudio Terminalpane provides access to the system shell from within the RStudio IDE.
- It supports xterm emulation, enabling use of full-screen terminal applications (e.g. text editors, terminal multiplexers) as well as regular command-line operations with line editing and shell history.

9. What is meant by Ttidyverse package? (CO4, K3)

- Ttidyverse is a coherent system of packages for importing, tidying, transforming, exploring, and visualizing data.
- Tidyverse packages are intended to make statisticians and data scientists more productive by guiding them through workflows that facilitate communication, and result in reproducible work products.
- Fundamentally, the tidyverse is about the connections between the tools that make the workflow possible.

UNIT IV DATA EXPLORATION AND DATA VISUALIZATION IN R

10. What is the difference between tibbles and data frames? (CO4, K3)

- Tibbles never convert the data types of variables.
- Also, they never change the names of variables or create row names.
- Tibbles also have a refined print method that shows only the first 10 rows, and all columns that will fit on the screen.
- Tibbles also print the column type along with the name.

11. What is the function of ggplot2 package? (CO4, K3)

- The ggplot2 package can be used to create various types of charts and graphs from a data frame.
- The ggplot()function is used to define plots, and can be passed a number of parameters and joined with other functions to ultimately produce an output chart.

11. What is the geom_col() function? (CO4 , K3)

The geom_col() function is used to define the geometry of the graph (bar chart) and is passed a mapping parameter which is defined by calling the aes() function and passing in the columns for the x axis (Beat), and the y axis (n = count).

12. What is the need of dplyr package and mutate() function? (CO4 , K3)

- The dplyr package includes the ability to dynamically create new columns in a data frame through the manipulation of data from existing columns in the data frame.
- The mutate() function is used to create the new columns. Here the mutate() function will be used to extract the year from the CrimeDate column.

13. What will read.table() do? (CO4 , K3)

- The dplyr package includes the ability to dynamically create new columns in a data frame through the manipulation of data from existing columns in the data frame.
- The mutate() function is used to create the new columns. Here the mutate() function will be used to extract the year from the CrimeDate column.

UNIT IV DATA EXPLORATION AND DATA VISUALIZATION IN R

14. What are the five functions of dplyr? (CO4 , K3)

- filter(), arrange(), select(), mutate(), and summarize()

15. What is meant by arrange() function ? (CO4 , K3)

- The arrange() function in the dplyr package can be used to order the rows in a data frame. This function accepts a set of columns to order by with the default row ordering being in ascending order.

16. What is the advantage of using piping in R? (CO4 , K3)

- Piping makes your code more streamlined and easier to read and also takes away the need to create and populate variables that are only used as intermediate datasets.

17. What is meant by data tidying? (CO4 , K3)

Data tidying is a consistent way of organizing data in R and can be facilitated through the tidyr package found in the tidyverse ecosystem. There are three rules that we can follow to make a dataset tidy. First, each variable must have its own column. Second, each observation must have its own row, and finally, each value must have its own cell.

18. List the advantage of tidy data? (CO4 , K3)

- First, having a consistent, uniform data structure is very important. The other packages that are part of tidyverse, including dplyr and ggplot2 are designed to work with tidy data so ensuring that your data is uniform facilitates the efficient processing of your data.
- In addition, placing variables into columns allows for the easily facilitation of vectorization in R.

19. How spreading is different from gathering? (CO3 , K3)

- Spreading is the opposite of gathering and is used when an observation is spread across multiple rows.