**Experiment No. 2: R Data Preparation**

**Problem Statements**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Topic** | **Problem Statements** |
| **1** | **Creating data frames in R** | A college has conducted technical events for the students. It maintains the name of the participant and the score obtained in different events.   1. Create a data frame by considering 5 students and 4 events. Each event has a maximum score of 10. If a student participates in an event, its entry contains the score value and 0 otherwise. 2. View the contents of the data frame. 3. Find the total score of each participant. 4. Append a column to include the total score of the participants and view the data frame. 5. Find the maximum score and display the name of the participant who scored it. 6. Compute the average score of each events and append it as a new row in the data frame. 7. Store the details in a comma separated values (csv) file. Also suppress the row numbers. |
| **2** | **Indexing and Slicing data frames**  **(**Events.csv) | 1. Read the content of ‘Events.csv’ in a data frame and view it. 2. Access the scores of participants in event2 using the column name. 3. Use index number to retrieve the same data. 4. Extract the score of third participant in event3. 5. Extract the scores of the first and second participant in all the events. 6. Display the names and total scores of all participants. 7. Make the column “name” as the row index of the data frame. 8. Display the names of the students participated in event3. 9. Obtain the names whose total score is above its average. |
| **3** | **Basic Operations in data frame** | MASS package contains a data frame called ‘survey’ which contains the responses of 237 Statistics I students at the University of Adelaide to a number of questions.  The components of the data frame are:   * Sex The sex of the student. (Factor with levels "Male" and "Female".) * Wr.Hnd span (distance from tip of thumb to tip of little finger of spread hand) of writing hand, in centimetres. * NW.Hnd span of non-writing hand. * W.Hnd writing hand of student. (Factor, with levels "Left" and "Right".) * Fold “Fold your arms! Which is on top” (Factor, with levels "R on L", "L on R", "Neither".) * Pulse pulse rate of student (beats per minute). * Clap ‘Clap your hands! Which hand is on top?’ (Factor, with levels "Right", "Left", "Neither".) * Exer how often the student exercises. (Factor, with levels "Freq" (frequently), "Some", "None".) * Smoke how much the student smokes. (Factor, levels "Heavy", "Regul" (regularly), "Occas" (occasionally), "Never".) * Height height of the student in centimetres. M.I whether the student expressed height in imperial (feet/inches) or metric (centimetres/metres) units. (Factor, levels "Metric", "Imperial".) * Age age of the student in years.   Do the following:   1. Install the package MASS. 2. Import the package MASS. 3. Display the structure of the data survey. 4. Check the class and type of the data set survey in MASS. 5. Get the number of rows and columns of the survey data frame. 6. Get the dimension of the survey data frame. 7. Provide the statistical summary of the data frame. 8. Display the column names of the survey data frame 9. Retrieve the top 3 rows from the data frame. 10. Extract the bottom 2 rows from the data frame. |
| **4** | **Data Manipulation using dplyr package** | Use the newsurvey data obtained by cleaning ‘na’ values in survey data of MASS package to do the following:   1. Install the dplyr package and import it. 2. Filter all male left handers. 3. Display all female right handers who keep right on top while clapping. 4. Display all students who never exercise. 5. Display only the gender, age and writing hand of the students. 6. Display the age, pulse rate and writing hand span of female left handers.   Note: use dplyr::select while using select operator as dplyr package conflicts with MASS package |
| **5** | **More functions in dplyr package** | Use the newsurvey data obtained by cleaning ‘na’ values in survey data of MASS package to do the following:   1. Install the dplyr package and import it. 2. Arrange all male left handers according to descending order of their heights. 3. Introduce a new column hand\_span which contains the value as difference between the span of writing hand and non-writing hand and display it along with gender, writing hand and non-writing hand span. 4. Display the average writing span of male and female left handers. 5. Find the maximum pulse rate of male left and right handers. |