

Homework 1

Instructions:

1. Answer all the questions
 2. The number represented in parenthesis, e.g. (2pt) represents two points for the question
 3. Follow the sub points carefully
 - i) Objects (datasets) in this document appear in **bold** letters
 - ii) Variables in the datasets appear *italicized*
 - iii) Any values of the variable in the dataset are represented in red color
 - iv) Any code is represented in blue color
 4. Do not submit any R project code files (.R extension files). You may wish to code in R and check the correctness/error free operations by executing them. Once you finalize the code, copy the code into a word or pdf file
 5. Please submit only pdf file for the group in canvas by their due date
 6. Only one member of the group makes the submission on canvas. List the group name and members' names in the submitted file itself
-

1. Basic R operations (Total 36pt)

Write command/code for the following:

- a) A numeric vector with values 1 to 30 and 50 to 80 (1pt)
- b) A numeric vector from 0 to 1000 where the difference between successive elements is 50 (2pt)
- c) A sequence where the vector `c(1, 3)` is repeated ten times (2pt)
- d) A sequence where each element of the vector `c(3, 30)` is repeated ten times (2pt)
- e) Sort the numeric vector `c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)` in descending order (1pt)
- f) Reverse the numeric vector `c(10, 9, 8, 7, 6, 5, 4, 3, 2, 1)` (1pt)
- g) Unique values of the character vector `c("a", "b", "b", "c")` (1pt)
- h) Retrieve the 3rd and 6th element of the vector `a = c(12, 4, 6, 73, 67, 84, 45, 74, 5, 52, 35)` (2pt)
- i) Retrieve all elements except the 3rd and 6th element of the vector `a = c(12, 4, 6, 73, 67, 84, 45, 74, 5, 52, 35)` (2pt)
- j) Retrieve elements which are less than 10 in vector `a = c(12, 4, 6, 73, 67, 84, 45, 74, 5, 52, 35)` (2pt)
- k) Write a "while" loop which prints values 1 to 10 (3pt)

Homework 1

- l) Write a conditional if statement which prints the following (3pt)
 - “Yes” if the value of variable “i” is equal to 3
 - “No” if the value of variable “i” is not equal to 3
 - m) What is the class of the object **a** in the code **a = letters** ? (1pt)
 - n) Convert the character vector to **c(“1”, “3”, “6”)** to a numeric vector (1pt)
 - o) Convert the numeric vector to **c(6, 3, 1)** to a character vector (1pt)
 - p) Generate a vector where each of its element represents rank of each element in the vector **c(12, 4, 6, 73, 67, 84, 45, 74, 5, 52, 35)** (2pt)
 - q) Variance of the elements in the numeric vector **c(12, 4, 6, 73, 67, 84, 45, 74, 5, 52, 35)** (1pt)
 - r) List all variables in the environment window (1pt)
 - s) Remove all the variables in the environment window (1pt)
 - t) Create a 3 by 3 matrix with values 1 to 9 (2pt)
 - u) Join three multiple vectors (say, **X**, **Y** and **Z**) separated by an underscore symbol (**_**) (1pt)
 - v) Join elements of a vector (say, **X**) (1pt)
 - w) Display the number of characters in a string (1pt)
 - x) Convert a string from lower case to uppercase (1pt)
2. Download the Riding Mowers.csv data from canvas (Assignments → Homework 1) which has information on 24 households. The three columns (or variables) in the data are: *Income* (in thousands of dollars), *Lot_Size* (in thousands of square foot) and *Ownership* status (**Owner** or **Nonowner**). Write codes for each of the following questions and if required present the output. (Total 36pt)
- a) Read the Riding Mowers.csv data into an dataset/object name “**ridmov**” (1pt)
 - b) Command to display the first 10 rows of the “**ridmov**” data (1pt)
 - c) Summarize the variables in the “**ridmov**” data. For numeric variables *Income* and *Lot_Size*, calculate their mean, median, variance and standard deviation. For the character variable - *Ownership*, tabulate the count of **Owner** and **Nonowner** (3pt)
 - d) Calculate the mean, standard deviation of *Income* and *Lot_Size* by *Ownership* variable (2pt)
 - e) Filter observations with *Income* less than 50 and *Ownership* status equal to **Owner** (2pt)
 - f) Filter observations with *Income* greater than equal to 50 or *Lot_Size* greater than equal to 10 (2pt)
 - g) Create a dataset which has distinct rows of “**ridmov**” data (2pt)

Homework 1

- h) Filter observations whose row number is multiple of 3 (2pt)
- i) Sort the data in the ascending order of *Income* variable (2pt)
- j) Sort the data in the ascending order of *Income* and descending order of *Lot_Size* (2pt)
- k) Create two new variables – *Income* in units of thousands of Chinese Yuan and *Lot_Size* in units of thousands of square yards and name them “*Income_Cy*” and “*Lot_Size_Sy*” (3pt)
- l) Create two new variables which represent cumulative sum and cumulative mean of *Income* (2pt)
- m) Create a two datasets, one only with variable *Income* and the other dataset only with variable *Lot_Size*. Then, create a new dataset which combine the two datasets by column (4pt)
- n) Create two data sets, one with *Ownership* status being **Owner** and other dataset with status **Nonowner**. Then, create a new dataset which combine the two datasets by row (4pt)
- o) Create a new variable *Household_Number* in the **ridmov** data which represents the row number (2pt)
- p) Create a new dataset **temp1** which contains first 16 rows, variables *Household_number* and *Income* from the **ridmov** data (4pt)
- q) Create a new dataset **temp2** which contains last 16 rows, variables *Household_number* and *Lot_Size* from the **ridmov** data (4pt)
- r) Execute a joining operation of datasets **temp1** and **temp2** by *Household_number* which results in an dataset containing only rows of **temp1**, but you have to use the command `right_join(.....)` (4pt)
- s) Execute a joining operation of objects **temp1** and **temp2** by *Household_number* which results in dataset containing only rows of **temp2**, but you have to use the command `left_join(.....)` (4pt)
- t) Execute a joining operation of objects **temp1** and **temp2** by *Household_number* which results in a dataset containing common rows of **temp1** and **temp2** (2pt)
- u) Execute a joining operation of objects **temp1** and **temp2** by *Household_number* which results in a dataset containing all rows of **temp1** and **temp2** (2pt)
- v) In the dataset obtained from question **u**), replace the missing values of *Income* by the mean of its non-missing values, *Lot_Size* by the mean of its non-missing values (4pt)
- w) Draw a scatter plot with *Income* as X-axis and *Lot_Size* as Y-axis. In the plot, name the X-axis as “Income (\$000s)” and Y-axis as “Lot_Size (000 ft²)” (3pt)

Homework 1

- x) Draw a box plot of *Income* by *Ownership*. The first box has to represent for **Nonowner** status while the second box **Owner** status. In the plot, name the X-axis as “Ownership of Riding Mower” and Y-axis as “Income (\$000s)” (3pt)