MECHELLE SIMPRON BSIT 2A

RWORKSHEET_3B

1. Create a data frame using the table below.
   a. Write the codes
      respondents <- c(1:20)
      sex <- c(2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 1, 2)
      fathers_occupation <- c(1, 3, 3, 3, 1, 2, 3, 1, 1, 1, 3, 2, 1, 3, 3, 1, 3, 1, 2, 1)
      persons_at_home <- c(5, 7, 3, 8, 5, 9, 6, 7, 8, 4, 7, 5, 4, 7, 8, 8, 3, 11,7, 6)
      siblings_at_school <- c(6, 4, 4, 1, 2, 1, 5, 3, 1, 2, 3, 2, 5, 5, 2, 1, 2, 5, 3, 2)
      types_of_houses <- c(1, 2, 3, 1, 1, 3, 3, 1, 2, 3, 2, 3, 2, 2, 3, 3, 3, 3, 3, 2)

      m_data <- data.frame(respondents, sex, fathers_occupation,persons_at_home,
      siblings_at_school,types_of_houses)
      m_data

   b. Describe the data. Get the structure or the summary of the data.
      The data show the number of respondents gathered together with the gender, the fathers
      occupation, persons at home, siblings at home and type of houses.

   c. Is the mean number of siblings attending is 5?
      NO because the mean of siblings at home is 2.95.
   d. Extract the 1st two rows and then all the columns using the subsetting functions. Write the
      codes and its output.
      Codes: m <- m_data[1:2, 1:6, drop =FALSE]
             m

      Output:

| respondents | sex | fathers_occupation | persons_at_home | siblings_at_school | types_of_houses |
|---|---|---|---|---|---|
| 1 | 2 | 1 | 5 | 6 | 1 |
| 2 | 2 | 3 | 7 | 4 | 2 |

   e. Extract 3rd and 5th row with 2nd and 4th column. Write the codes and its result.
      Codes: new_data <- m_data[c(3,5), c(2,4)]
             new_data

      Output:

| | sex | persons_at_home |
|---|---|---|
| 1 | 1 | 3 |
| 2 | 2 | 5 |

f.  Select the variable types of houses then store the vector that results as types_houses. Write the codes.
    Codes:  th <- types_of_houses
            th

    Output: [1] 1 2 3 1 1 3 3 1 2 3 2 3 2 2 3 3 3 3 3 2

g.  Select only all Males respondent that their father occupation was farmer. Write the codes and its output.
    Code: male_data <- data.frame(sex, fathers_occupation)
          subset(male_data, sex == 1 & fathers_occupation == 1)

    Output: sex     fathers_occupation

            <0 rows> (or 0-length row.names)

h.  Select only all females respondent that have greater than or equal to 5 number of siblings attending school. Write the codes and its outputs.
    Codes: girl_data <- data.frame(sex, siblings_at_school)
           subset(girl_data, sex == 2 & siblings_at_school >= 5)

    Output:
    sex siblings_at_school
     2          6
     2          5
     2          5
     2          5
     2          5

2.  Write a R program to create an empty data frame. Using the following codes:
    df = data.frame(Ints=integer(), Doubles=double(), Characters=character(), Logicals=logical(), Factors=factor(), stringsAsFactors=FALSE) print("Structure of the empty dataframe:") print(str(df))

    a. Describe the results.
    The result shows an empty data frame. It shows the structure of the empty data frame such as #int, num,chr, logic, and factors without levels that result as NULL.

3.  Interpret the graph
    The graph shows the sentiments of tweets per day where July 15 and 21 has the highest sentiments of negative and positive tweets.