Worksheet-3b in R

Instructions:

- Use RStudio or the posit(RStudio) Cloud accomplish this worksheet.

- Inside the folder worksheet#3, create an .Rmd (R Markdown) for this worksheet and saved it as *RWorksheet_lastname#3b.Rmd*
- **Knit to pdf** to render a pdf file.
- On your own *GitHub repository*, push the .Rmd file, as well as the pdf worksheet knitted to the repo you have created before.
- Do not forget to comment your Git repo on our VLE
- Accomplish this worksheet by answering the questions being asked and writing the code manually.

1. Create a data frame using the table below.

   a. Write the codes.

```
respondents <- 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)

df <- data.frame(
  Respondents = respondents,
  Sex = sex,
  Fathers_Occupation = fathers_occupation,
  Persons_at_Home = persons_at_home,
  Siblings_at_School = siblings_at_school,
  Types_of_Houses = types_of_houses
)

print(df)
```

   b. Describe the data. Get the structure or the summary of the data

```
 Respondents          Sex        Fathers_Occupation Persons_at_Home Siblings_at_School
Min.   : 1.00   Min.   :1.00   Min.   :1.00       Min.   : 3.0    Min.   :1.00
1st Qu.: 5.75   1st Qu.:2.00   1st Qu.:1.00       1st Qu.: 5.0    1st Qu.:2.00
Median :10.50   Median :2.00   Median :2.00       Median : 7.0    Median :2.50
Mean   :10.50   Mean   :1.85   Mean   :1.95       Mean   : 6.4    Mean   :2.95
3rd Qu.:15.25   3rd Qu.:2.00   3rd Qu.:3.00       3rd Qu.: 8.0    3rd Qu.:4.25
Max.   :20.00   Max.   :2.00   Max.   :3.00       Max.   :11.0    Max.   :6.00
Types_of_Houses
Min.   :1.0
1st Qu.:2.0
Median :2.5
Mean   :2.3
3rd Qu.:3.0
Max.   :3.0
```

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
3	1	3	3	4	3
4	2	3	8	1	1
5	2	1	5	2	1
6	2	2	9	1	3
7	2	3	6	5	3
8	2	1	7	3	1
9	2	1	8	1	2
10	2	1	4	2	3
11	1	3	7	3	2
12	2	2	5	2	3
13	2	1	4	5	2
14	2	3	7	5	2
15	2	3	8	2	3
16	2	1	8	1	3
17	2	3	3	2	3
18	2	1	11	5	3
19	1	2	7	3	3
20	2	1	6	2	2

Legend:

Male-1	Farmer-1	Wood-1
Female-2	Driver-2	Semi-Concrete-2
	Others-3	Concrete-3

Figure 1: R Chunk

c. Is the mean number of siblings attending is 5?

```
[1] FALSE
```

d. Extract the 1st two rows and then all the columns using the subsetting functions. Write the codes and its output.

```
first_two_rows <- df[1:2, ]
print(first_two_rows)
```

```
  Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
Types_of_Houses
1            1   2                 1              5                 6
1
2            2   2                 3              7                 4
2
```

e. Extract 3rd and 5th row with 2nd and 4th column. Write the codes and its result.

```
subset_rows_cols <- df[c(3, 5), c(2, 4)]
print(subset_rows_cols)
```

```
  Sex Persons_at_Home
3   1              3
5   2              5
```

f. Select the variable types of houses then store the vector that results as types_houses. Write the codes.

```
types_houses <- df$Types_of_Houses
print(types_houses)
```

g. Select only all Males respondent that their father occupation was farmer. Write the codes and its output.

```
male_farmers <- df[df$Sex == 1 & df$Fathers_Occupation == 1, ]
print(male_farmers)
```

```
[1] Respondents      Sex              Fathers_Occupation
Persons_at_Home
[5] Siblings_at_School Types_of_Houses
<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.

```
female_siblings <- df[df$Sex == 2 & df$Siblings_at_School >= 5, ]
print(female_siblings)
```

```
   Respondents Sex Fathers_Occupation Persons_at_Home
Siblings_at_School Types_of_Houses
1            1   2                 1              5                 6
1
7            7   2                 3              6                 5
3
13          13   2                 1              4                 5
2
14          14   2                 3              7                 5
2
18          18   2                 1             11                 5
3
```

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 output of the R program shows the structure of an empty data frame named `df`, which contains 0 observations (rows) and 5 variables (columns). The variables include **Ints** (integer), **Doubles** (numeric), **Characters** (character strings), **Logicals** (logical values), and **Factors** (categorical data) with 0 levels. This indicates that while the data frame is currently empty, it is set up to hold various data types for future use in analysis or data manipulation tasks.

3. Create a `.csv` file of this. Save it as `HouseholdData.csv`

Respondents	Sex	Fathers Occupation	Persons at Home	Siblings at School	Types of Houses
1	Male	1	5	2	Wood
2	Female	2	7	3	Congrete
3	Female	3	3	0	Congrete
4	Male	3	8	5	Wood
5	Male	1	6	2	Semi-congrete
6	Female	2	4	3	Semi-congrete
7	Female	2	4	1	Wood
8	Male	3	2	2	Semi-congrete
9	Female	1	11	6	Semi-congrete
10	Male	3	6	2	Congrete

Figure 2: Figure 2: Sentiment Analysis

a. Import the csv file into the R environment. Write the codes.

```
data <- read.csv("HouseholdData.csv")
print(data)
```

b. Convert the `Sex` into factor using `factor()` function and change it into integer.[Legend: Male = 1 and Female = 2]. Write the R codes and its output.

```
data$Sex <- factor(data$Sex, levels = c("Male", "Female"), labels = c(1, 2))
print(data$Sex)
```

c. Convert the `Type of Houses` into factor and change it into integer. [Legend: Wood = 1; Congrete = 2; Semi-Congrete = 3]. Write the R codes and its output.

```
data$Types_of_Houses <- factor(data$Types_of_Houses, levels = c("Wood", "Congrete", "Semi-congrete"), labels = c(1, 2, 3))

print(data$Types_of_Houses)
```

```
 [1] 1 2 2 1 3 3 1 3 3 2
Levels: 1 2 3
```

d. On father's occupation, factor it as Farmer = 1; Driver = 2; and Others = 3. What is the R code and its output?

```
data$Fathers_Occupation <- factor(data$Fathers_Occupation,
 levels = c(1, 2, 3), labels = c("Farmer", "Driver", "Others"))

print(data$Fathers_Occupation)
```

```
 [1] Farmer Driver Others Others Farmer Driver Driver Others
Farmer Others
Levels: Farmer Driver Others
```

e. Select only all females respondent that has a father whose occupation is driver. Write the codes and its output.

```
female_driver <- data[data$Sex == 2 & data$Fathers_Occupation == "Driver", ]

print(female_driver)
```

	Respondents	Sex	Fathers_Occupation	Persons_at_Home	Siblings_at_School	Types_of_Houses
2	2	2	Driver	7	3	2
6	6	2	Driver	4	3	3
7	7	2	Driver	4	1	1

f. Select the respondents that have greater than or equal to 5 number of siblings attending school. Write the codes and its output.

```
siblings_5_or_more <- data[data$Siblings_at_School >= 5, ]

print(siblings_5_or_more)
```

	Respondents	Sex	Fathers_Occupation	Persons_at_Home	Siblings_at_School	Types_of_Houses
4	4	1	Others	8	5	1
9	9	2	Farmer	11	6	3

4. Interpret the graph.

The graph illustrates the sentiments expressed in tweets: negative sentiments are represented in red, neutral in orange, and positive in blue. On certain days, such as July 15 and 21, there were a significant number of negative tweets, while other days showed a prevalence of positive or neutral sentiments. This indicates that people's opinions were varied and shifted in response to events occurring that week, reflecting their tweets.

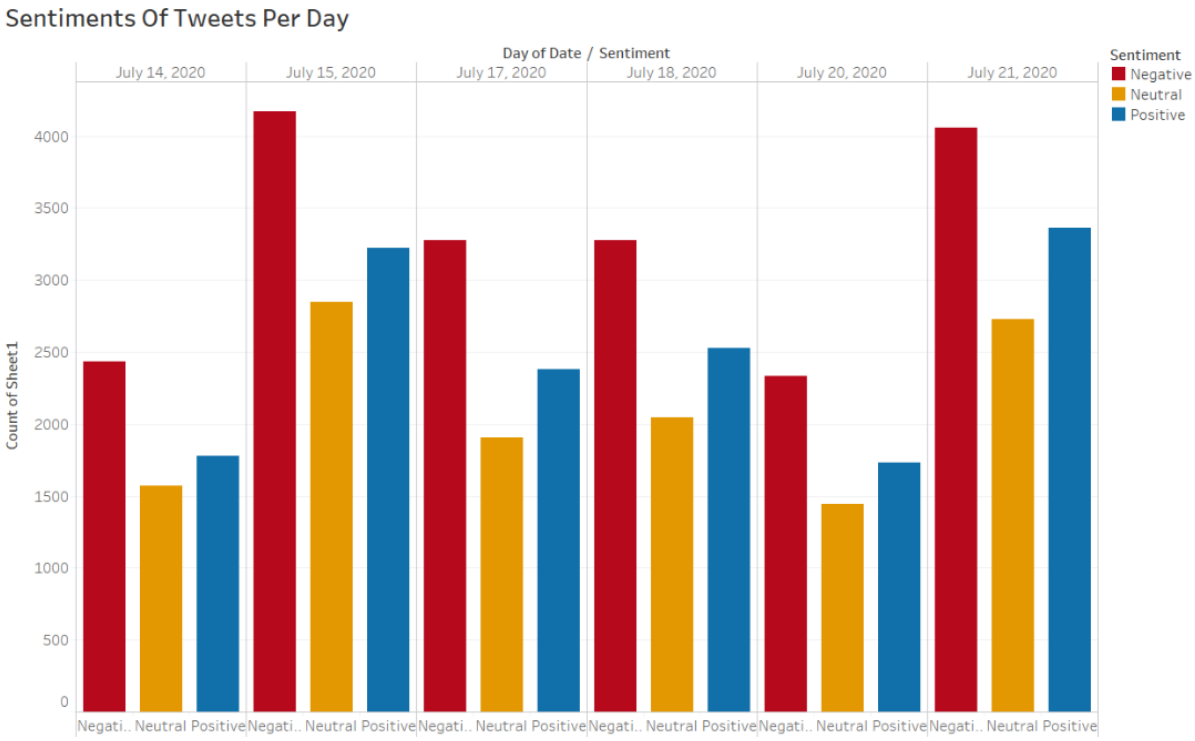

Figure 3: Figure 2: Sentiment Analysis