

# 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_data <- data.frame(  
  Respondents = 1:20,  
  Sex = c(2, 2, 1, 2, 1, 2, 1, 2, 1, 1, 2, 2, 1, 1, 2, 1, 2, 1),  
  Fathers_Occupation = c(1, 3, 1, 3, 3, 1, 3, 2, 3, 1, 2, 1, 3, 1, 3, 1, 3, 1),  
  Person_at_Home = c(5, 7, 3, 8, 9, 6, 9, 6, 4, 3, 4, 5, 7, 8, 3, 7, 11, 7, 6, 6),  
  Siblings_at_School = c(6, 4, 4, 1, 1, 3, 3, 5, 3, 2, 4, 2, 3, 4, 3, 3, 5, 3, 2, 2),  
  Types_of_Houses = c(1, 2, 1, 1, 3, 3, 3, 2, 1, 3, 1, 2, 1, 3, 1, 3, 1, 3, 2, 2)  
)  
respondents_data
```

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

#The data contains information from 20 individuals regarding their families and residences. The majority of respondents are female, and their fathers are employed as farmers, drivers, or in various other occupations, with many falling into the "other" category. The number of residents in each household varies from 3 to 11, and most respondents report having 2 to 5 siblings currently in school. Their homes are primarily constructed of wood, semi-concrete, or concrete, with most living in concrete or semi-concrete structures. Overall, the data provides a clear overview of family size, housing types, and the occupations of their fathers.

```
str(respondents_data)  
summary(respondents_data)
```

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

Female-2

Farmer-1

Driver-2

Others-3

Wood-1

Semi-Concrete-2

Concrete-3

Figure 1: R Chunk

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

no, the mean of the number of siblings is 3.15

```
mean_siblings <- mean(respondents_data$Siblings_at_School)
mean_siblings == 5
mean_siblings
```

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

```
subset_data <- respondents_data[1:2,]
subset_data
```

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

```
subset_data2 <- respondents_data[c(3, 5), c(2, 4)]
subset_data2
```

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

```
types_houses <- respondents_data$Types_of_Houses
types_houses
[1] 1 2 1 1 3 3 3 2 1 3 1 2 1 3 1 3 1 3 2 2
```

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

```
male_farmers <- subset(respondents_data, Sex == 1 & Fathers_Occupation == 1)
male_farmers
```

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.

```
femalesibs <- subset(respondents_data, Sex == 2 & Siblings_at_School >= 5)
femalesibs
```

2. Write a R program to create an empty data frame. Using the following codes:

```
df = data.frame(Ids=integer(),
                Doubles=double(), Characters=character(),
                Logicals=logical(),
                Factors=factor(),
                stringsAsFactors=FALSE)

print("Structure of the empty dataframe:")
print(str(df))
```

a. Describe the results.

It is an empty data frame with 5 columns designated for integers, decimals, text, and boolean values (true or false). It currently has no rows or data, and the category column does not have any defined options. This serves merely as the structure of a data frame without any actual data included.

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-concrete
6	Female	2	4	3	Semi-concrete
7	Female	2	4	1	Wood
8	Male	3	2	2	Semi-concrete
9	Female	1	11	6	Semi-concrete
10	Male	3	6	2	Congrete

Figure 2: Figure 2: Sentiment Analysis

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

```
write.csv(household_data, "HouseholdData.csv", row.names = FALSE)
library(readr)
household_data <- read.csv("HouseHoldData.csv")
household_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.

```
household_data$Sex <- factor(household_data$Sex, levels = c("Male", "Female"), labels = c(1,2))
household_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.

```
household_data$Type_of_Houses <- factor(household_data$Types_of_Houses, levels =
c("Wood", "Concrete", "Semi-concrete"), labels = c(1, 2, 3))
household_data$Types_of_Houses
```

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

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

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

```
female_driver <- subset(household_data, Sex == 2 & Fathers_Occupation == "Driver")  
female_driver
```

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 <- subset(household_data, Siblings_at_School >= 5)  
siblings_5_or_more
```

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.

Sentiments Of Tweets Per Day

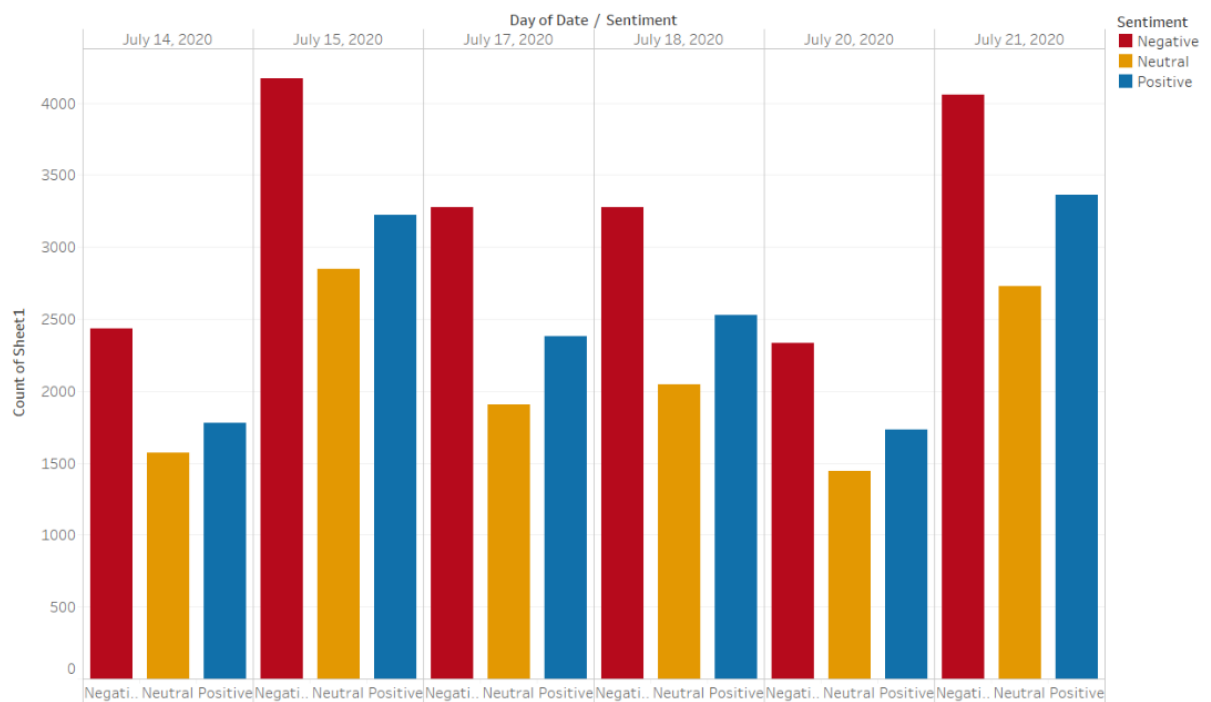


Figure 3: Figure 2: Sentiment Analysis