

# 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.

```
df <- data.frame(
  Respondents = 1:20,
  Sex = c(1, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 1, 2, 2, 1, 2),
  Fathers_Occupation = c(1, 3, 3, 3, 1, 2, 3, 1, 1, 1, 3, 2, 1, 3, 3, 1, 3, 2, 2, 1),
  Persons_at_Home = c(5, 7, 3, 8, 5, 9, 6, 7, 8, 4, 7, 5, 4, 7, 8, 1, 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

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

```
str(df)
summary(df)
```

```
'data.frame':  0 obs. of  5 variables:
 $ Ints      : int
 $ Doubles   : num
 $ Characters: chr
 $ Logicals  : logi
 $ Factors   : Factor w/ 0 levels:
      Ints      Doubles   Characters   Logicals      Factors
Min.   : NA    Min.   : NA    Length:0    Mode:logical  NULL:
1st Qu.: NA    1st Qu.: NA    Class :character
Median : NA    Median : NA    Mode  :character
Mean   :NaN    Mean   :NaN
3rd Qu.: NA    3rd Qu.: NA
Max.   : NA    Max.   : NA
```

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?

```
mean(df$Siblings_at_School)
```

```
> mean(df$Siblings_at_School)
[1] 2.95
```

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

```
df[1:2, ]
```

	Respondents <int>	Sex <dbl>	Fathers_Occupation <dbl>	Persons_at_Home <dbl>	Siblings_at_School <dbl>	Types_of_Houses <dbl>
1	1	1	1	5	6	1
2	2	2	3	7	4	2

2 rows

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

```
df[c(3, 5), c(2, 4)]
```

	Sex <dbl>	Persons_at_Home <dbl>
3	1	3
5	2	5

2 rows

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
types_houses

[1] 1 2 3 1 1 3 3 1 2 3 2 3 2 2 3 3 3 3 2
```

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

```
subset(df, Sex == 1 & Fathers_Occupation == 1)
```

	Respondents <int>	Sex <dbl>	Fathers_Occupation <dbl>	Persons_at_Home <dbl>	Siblings_at_School <dbl>	Types_of_Houses <dbl>
1	1	1	1	5	6	1
13	13	1	1	4	5	2
16	16	1	1	1	1	3

3 rows

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.

```
subset(df, Sex == 2 & Siblings_at_School >= 5)
```

	Respondents <int>	Sex <dbl>	Fathers_Occupation <dbl>	Persons_at_Home <dbl>	Siblings_at_School <dbl>	Types_of_Houses <dbl>
7	7	2	3	6	5	3
14	14	2	3	7	5	2
18	18	2	2	11	5	3

3 rows

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.

The data frame has 0 observations and 5 variables (columns). Each column is labeled according to its type (int for integers, num for doubles, chr for characters, logi for logical, and Factor for factors with no levels yet). Since the data frame is empty, there are no actual data points (no rows), but the structure defines the types of data that can be stored in each column.

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.

```
imported_data <- read.csv("HouseholdData.csv")
print(imported_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.

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

Respondents	Sex	Fathers_Occupation	Persons_at_Home	Siblings_at_School	Types_of_Houses
<int>	<fctr>	<int>	<int>	<int>	<chr>
1	1	1	5	2	Wood
2	2	2	7	3	Concrete
3	2	3	3	0	Concrete
4	1	3	8	5	Wood
5	1	1	6	2	Semi-concrete
6	2	2	9	3	Semi-concrete
7	2	2	6	2	Wood
8	1	1	7	1	Semi-concrete
9	2	3	1	6	Semi-concrete
10	1	1	6	3	Concrete



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

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

print(imported_data)
```

Respondents <int>	Sex <fctr>	Fathers_Occupation <int>	Persons_at_Home <int>	Siblings_at_School <int>	Types_of_Houses <fctr>
1	1	1	5	2	1
2	2	2	7	3	2
3	2	3	3	0	2
4	1	3	8	5	1
5	1	1	6	2	3
6	2	2	9	3	3
7	2	2	6	2	1
8	1	1	7	1	3
9	2	3	1	6	3
10	1	1	6	3	2

1-10 of 10 rows

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

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

print(imported_data)
```

Respondents <int>	Sex <fctr>	Fathers_Occupation <fctr>	Persons_at_Home <int>	Siblings_at_School <int>	Types_of_Houses <fctr>
1	1	Farmer	5	2	1
2	2	Driver	7	3	2
3	2	Others	3	0	2
4	1	Others	8	5	1
5	1	Farmer	6	2	3
6	2	Driver	9	3	3
7	2	Driver	6	2	1
8	1	Farmer	7	1	3
9	2	Others	1	6	3
10	1	Farmer	6	3	2

1-10 of 10 rows

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

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

Respondents <int>	Sex <fctr>	Fathers_Occupation <fctr>	Persons_at_Home <int>	Siblings_at_School <int>	Types_of_Houses <fctr>
2	2	Driver	7	3	2
6	2	Driver	9	3	3
7	2	Driver	6	2	1

3 rows

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

```
siblings_school <- subset(imported_data, Siblings_at_School >= 5)
print(siblings_school)
```

Respondents <int>	Sex <fctr>	Fathers_Occupation <fctr>	Persons_at_Home <int>	Siblings_at_School <int>	Types_of_Houses <fctr>
4	1	Others	8	5	1
9	2	Others	1	6	3

2 rows

4. Interpret the graph.

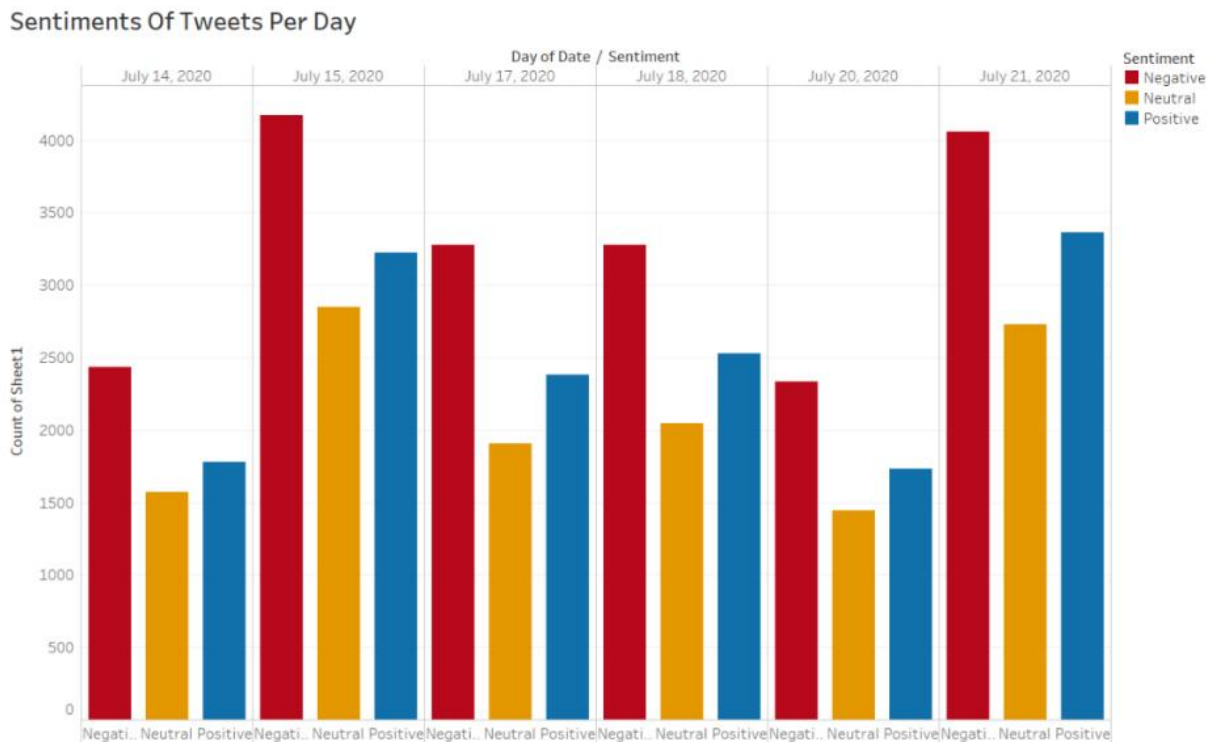


Figure 3: Figure 2: Sentiment Analysis

# The graph shows the daily distribution of tweet sentiments (Negative, Neutral, Positive) from July 14 to July 21, 2020. Negative sentiments (red) consistently dominate across all days, with notable peaks on July 15 and July 21. Positive sentiments (blue) are the second most frequent, showing a steady rise, particularly on July 21. Neutral sentiments (yellow) are generally the least represented and fluctuate more than the other categories. The chart indicates that during this period, the majority of tweets had negative sentiments, with positive sentiment increasing toward the end of the timeframe.