# Worksheet-3b in R

**Instructions:**

- Use RStudio or the RStudio Cloud accomplish this worksheet.

- Save the R script as *RWorksheet_lastname#3b.R*.
- On your own *GitHub repository*, push the R script, the Rmd file, as well as this pdf worksheet 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.

| 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 |

a. | Write the codes.

```
dframe <- data.frame(
   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_Home = c(5,7,3,8,5,9,6,7,8,4,7,5,4,7,8,8,3,11,7,6),
   Siblings_school = c(6,4,4,1,2,1,5,3,1,2,3,2,5,5,2,1,2,5,3,2),
   Types_houses = c(1,2,3,1,1,3,3,1,2,3,2,3,2,2,3,3,3,3,3,2))
dframe
```

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

```
> summary(dframe)
 Respondents         Sex          Fathers_Occupation  Persons_Home     Siblings_school  Types_houses
Min.   : 1.00   Min.   :1.00   Min.   :1.00        Min.   : 3.0   Min.   :1.00   Min.   :1.0
1st Qu.: 5.75   1st Qu.:2.00   1st Qu.:1.00        1st Qu.: 5.0   1st Qu.:2.00   1st Qu.:2.0
Median :10.50   Median :2.00   Median :2.00        Median : 7.0   Median :2.50   Median :2.5
Mean   :10.50   Mean   :1.85   Mean   :1.95        Mean   : 6.4   Mean   :2.95   Mean   :2.3
3rd Qu.:15.25   3rd Qu.:2.00   3rd Qu.:3.00        3rd Qu.: 8.0   3rd Qu.:4.25   3rd Qu.:3.0
Max.   :20.00   Max.   :2.00   Max.   :3.00        Max.   :11.0   Max.   :6.00   Max.   :3.0
```

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

```
> mean(dframe$Siblings_school)
[1] 2.95
```

**No, the mean number of siblings attending is 2.95.**

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

```
> extract1 <- subset(dframe[c(1:2), ])
> extract1
  Respondents  Sex  Fathers_Occupation  Persons_Home  Siblings_school  Types_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.

```
> extract2 <- subset(dframe[c(3,5), c(2,4)])
> extract2
  Sex Persons_Home
3   1            3
5   2            5
```

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

```
> extract3 <- subset(dframe, select = Types_houses)
> extract3
   Types_houses
1             1
2             2
3             3
4             1
5             1
6             3
7             3
8             1
9             2
10            3
11            2
12            3
```

```
13        2
14        2
15        3
16        3
17        3
18        3
19        3
20        2
```

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

```
> extract4 <- subset(dframe,
+                Sex == 1 & Fathers_Occupation < 1,
+                select = c(Sex, Fathers_Occupation),
+                drop = FALSE)
> extract4
[1] Sex              Fathers_Occupation
<0 rows> (or 0-length row.names)
```

h. Select only all female respondent that have greater than or equal to 5 number of siblings attending school. Write the codes and its outputs.

```
> extract5 <- subset(dframe,
+                Sex == 2 & Siblings_school >= 5,
+                select = c(Sex, Siblings_school),
+                drop = FALSE)
> extract5
    Sex     Siblings_school
1        2                  6
7        2                  5
13       2                  5
14       2                  5
18       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))
```

```
> df = data.frame(Ints=integer(),
+
+              Doubles=double(), Characters=character(),
+              Logicals=logical(),
+              Factors=factor(),
```

```
+                stringsAsFactors=FALSE)
>
> print("Structure of the empty dataframe:")
[1] "Structure of the empty dataframe:"
> print(str(df))
'data.frame':    0 obs. of  5 variables:
 $ Ints          : int
 $ Doubles       : num
 $ Characters    : chr
 $ Logicals      : logi
 $ Factors       : Factor w/ 0 levels:
NULL
```

a. Describe the results.

**Based on the results of the program, since there are no inputted objects in each variable it resulted to a NULL.**

3. Interpret the graph.

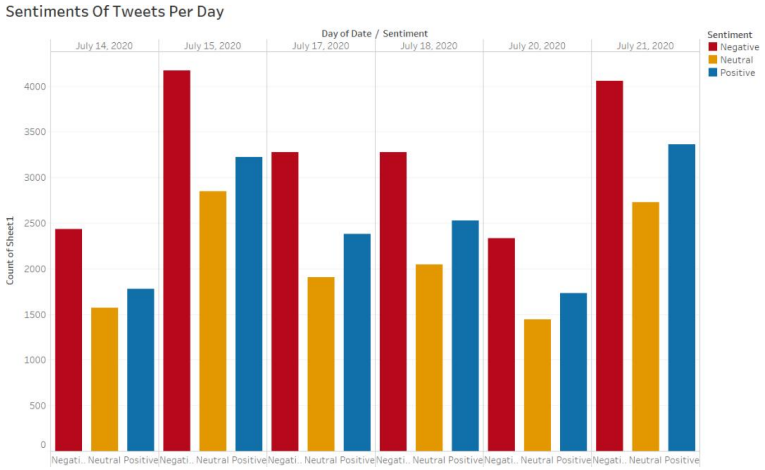


Figure 1: Sentiments of Tweets per day - Donald Trump

**The sentiments of tweets per day were shown as a bar graph. According to the graph, negative tweets had the highest tweet counts, particularly on July 15, 2020, and July 21, 2020, when there were over 4,000 negative tweets each.**