

RWorksheet_SOCO-3_B

```
#1. Create a data frame using the table below.  
#a. Write the codes.  
#b. Describe the data. Get the structure or the summary of the data  
  
respondents <- c(1,2,3,4,5,6,7,8,9,10,  
                  11,12,13,14,15,16,17,18,19,20)  
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)  
  
data <- data.frame(respondents, sex, fathers_occupation,  
                    persons_at_home, siblings_at_school, types_of_houses)  
data  
  
##      respondents  sex fathers_occupation persons_at_home siblings_at_school  
## 1            1    2                  1          5                6  
## 2            2    2                  3          7                4  
## 3            3    1                  3          3                4  
## 4            4    2                  3          8                1  
## 5            5    2                  1          5                2  
## 6            6    2                  2          9                1  
## 7            7    2                  3          6                5  
## 8            8    2                  1          7                3  
## 9            9    2                  1          8                1  
## 10          10    2                  1          4                2  
## 11          11    1                  3          7                3  
## 12          12    2                  2          5                2  
## 13          13    2                  1          4                5  
## 14          14    2                  3          7                5  
## 15          15    2                  3          8                2  
## 16          16    2                  1          8                1  
## 17          17    2                  3          3                2  
## 18          18    2                  1          11               5  
## 19          19    1                  2          7                3  
## 20          20    2                  1          6                2  
##      types_of_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

```

#Getting the Structure and summary of the data.

```
str(data)
```

```

## 'data.frame':   20 obs. of  6 variables:
## $ respondents    : num  1 2 3 4 5 6 7 8 9 10 ...
## $ sex            : num  2 2 1 2 2 2 2 2 2 2 ...
## $ fathers_occupation: num  1 3 3 3 1 2 3 1 1 1 ...
## $ persons_at_home : num  5 7 3 8 5 9 6 7 8 4 ...
## $ siblings_at_school: num  6 4 4 1 2 1 5 3 1 2 ...
## $ types_of_houses  : num  1 2 3 1 1 3 3 1 2 3 ...

```

```
summary(data)
```

	respondents	sex	fathers_occupation	persons_at_home
## Min.	1.00	Min. :1.00	Min. :1.00	Min. : 3.0
## 1st Qu.	5.75	1st Qu.:2.00	1st Qu.:1.00	1st Qu.: 5.0
## Median	10.50	Median :2.00	Median :2.00	Median : 7.0
## Mean	10.50	Mean :1.85	Mean :1.95	Mean : 6.4
## 3rd Qu.	15.25	3rd Qu.:2.00	3rd Qu.:3.00	3rd Qu.: 8.0
## Max.	20.00	Max. :2.00	Max. :3.00	Max. :11.0
## siblings_at_school				
## Min.	:1.00	Min. :1.0		
## 1st Qu.	:2.00	1st Qu.:2.0		
## Median	:2.50	Median :2.5		
## Mean	:2.95	Mean :2.3		
## 3rd Qu.	:4.25	3rd Qu.:3.0		
## Max.	:6.00	Max. :3.0		

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

#No, since the mean based on the data is 2.95 as the display shown.

```
mean_siblings <- mean(data$siblings_at_school)
```

```
mean_siblings == 5
```

```
## [1] FALSE
```

```
mean_siblings
```

```
## [1] 2.95
```

#d. Extract the 1st two rows and all columns

```
data[1:2, ]
```

```
##   respondents sex fathers_occupation persons_at_home siblings_at_school
## 1           1    2                   1               5                  6
## 2           2    2                   3               7                  4
##   types_of_houses
## 1           1
## 2           2
```

#e. Extract 3rd and 5th rows with 2nd and 4th columns

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

```
##   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 and its output

```
types_houses <- data$types_of_houses
```

```
types_houses
```

```
## [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

```
male_farmer <- subset(data, sex == 1 & fathers_occupation == 1)
male_farmer
```

```
## [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 output

```
female_5plus <- subset(data, sex == 2 & siblings_at_school >= 5)
```

```
female_5plus
```

```
##   respondents sex fathers_occupation persons_at_home siblings_at_school
## 1           1    2                   1               5                  6
## 7           7    2                   3               6                  5
## 13          13   2                   1               4                  5
## 14          14   2                   3               7                  5
## 18          18   2                   1              11                  5
##   types_of_houses
## 1           1
## 7           3
## 13          2
## 14          2
## 18          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:")
```

[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.

The program creates an empty data frame with zero rows but five predefined columns, each assigned a specific data type. The str() output confirms the structure by showing the column names and their corresponding types even though no data is stored yet.

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

#a

```
data <- read.csv("HouseholdData.csv", stringsAsFactors = FALSE)
data
```

	Respondents	Sex	Fathers.Occupation	Persons.at.Home	Siblings.at.School
## 1	1	Male	1	1	5
## 2	2	Female	2	2	7
## 3	3	Female	3	3	3
## 4	4	Male	3	3	8
## 5	5	Male	1	1	6
## 6	6	Female	2	2	4
## 7	7	Female	2	2	4
## 8	8	Male	3	3	2
## 9	9	Female	1	1	11
## 10	10	Male	3	3	6

	Types.of.Houses
## 1	Wood
## 2	Congrete
## 3	Congrete
## 4	Wood
## 5	Semi-concrete

```

## 6      Semi-concrete
## 7          Wood
## 8      Semi-concrete
## 9      Semi-concrete
## 10     Congrete

#b

data$Sex <- factor(data$Sex, levels = c("Male","Female"))
as.integer(data$Sex)

## [1] 1 2 2 1 1 2 2 1 2 1

#c
data$Types.of.Houses <- factor(data$Types.of.Houses,
                                 levels = c("Wood","Congrete","Semi-concrete"))
as.integer(data$Types.of.Houses)

## [1] 1 2 2 1 3 3 1 3 3 2

#d
data$Fathers.Occupation <- factor(data$Fathers.Occupation,
                                     levels = c(1,2,3),
                                     labels = c("Farmer","Driver","Others"))
as.integer(data$Fathers.Occupation)

## [1] 1 2 3 3 1 2 2 3 1 3

#e
subset(data, Sex == "Female" & Fathers.Occupation == "Driver")

##   Respondents   Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 2           2 Female             Driver            2                  7
## 6           6 Female             Driver            2                  4
## 7           7 Female             Driver            2                  4
##   Types.of.Houses
## 2       Congrete
## 6   Semi-concrete
## 7       Wood

#f

subset(data, Siblings.at.School >= 5)

##   Respondents   Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 1           1 Male              Farmer            1                  5
## 2           2 Female            Driver            2                  7
## 4           4 Male              Others            3                  8
## 5           5 Male              Farmer            1                  6
## 9           9 Female            Farmer            1                 11

```

```
## 10          10   Male           Others        3          6
##   Types.of.Houses
## 1          Wood
## 2          Concrete
## 4          Wood
## 5          Semi-concrete
## 9          Semi-concrete
## 10         Concrete
```

#4. Interpret the graph.

*#The graph is the Sentiments of tweets per day. It shows the dates from July 14
#to July 21 2020 with the Sentiment of Positive, Neutral and Negative.
#The graph shows the date with the highest negative score are July 15 and July 21.*