

# Worksheet 3b in R

Erl Syron R. Espadon

#1. Create a data frame from the table #a. codes

```
df <- data.frame(
  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, 1, 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)
)
```

#b. Describe the data

```
str(df)
```

```
## '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(df)
```

```
##   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 Types_of_houses
##   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. Calculate the mean number of siblings attending school

```
mean(df$Siblings_at_school)
```

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

#d. Extract the first two rows and all columns

```
df[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 the 3rd and 5th row with the 2nd and 4th column

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

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

#f. Select the variable 'Types\_of\_houses' and store it in a vector called 'types\_houses'

```
types_houses <- df$Types_of_houses
```

#g. Select only all Males respondent that their father occupation was farmer.

```
# Filter the data frame to select only male respondents (Sex == 1)
# whose fathers are farmers (Fathers_Occupation == 1)
males_farmer <- df[df$Sex == 1 & df$Fathers_Occupation == 1, ]
```

```
# Print the filtered data frame
print(males_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.

```
# Filter the data frame to select only female respondents (Sex == 2)
# who have 5 or more siblings attending school (Siblings_at_school >= 5)
females_siblings <- df[df$Sex == 2 & df$Siblings_at_school >= 5, ]
```

```
# Print the filtered data frame
print(females_siblings)
```

```
## 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
```

*#the code creates an empty data frame with columns pre-defined for different data types, ready to be fi*

#3 # Create a data frame from the table

```
df <- data.frame(
  Respondents = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10),
  Sex = c("Male", "Female", "Female", "Male", "Male", "Female", "Female", "Male", "Female", "Male"),
  Fathers_Occupation = c("Farmer", "Driver", "Others", "Others", "Farmer", "Driver", "Driver", "Others",
  Persons_at_Home = c(1, 2, 3, 3, 1, 2, 2, 3, 1, 3),
  Siblings_at_School = c(5, 7, 3, 8, 6, 4, 4, 2, 11, 6),
  Types_of_Houses = c("Wood", "Congrete", "Congrete", "Wood", "Semi-concrete", "Semi-concrete", "Wood",
)
```

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

```
write.csv(df, "HouseholdData.csv", row.names = FALSE)
```

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

```
df <- read.csv("HouseholdData.csv")
```

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.

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

```
## [1] 1 2 2 1 1 2 2 1 2 1
## Levels: 1 2
```

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.

```
df$Types_of_Houses <- factor(df$Types_of_Houses, levels = c("Wood", "Congrete", "Semi-concrete"), labels = c(1, 2, 3))
print(df$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?

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

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

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

```
females_driver_father <- df[df$Sex == 2 & df$Fathers_Occupation == 2, ]
print(females_driver_father)
```

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

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

```
respondents_5_siblings <- df[df$Siblings_at_School >= 5, ]
print(respondents_5_siblings)
```

```
## Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1          1    1                  1              1              5
## 2          2    2                  2              2              7
## 4          4    1                  3              3              8
## 5          5    1                  1              1              6
## 9          9    2                  1              1             11
## 10         10    1                  3              3              6
## Types_of_Houses
## 1          1
```

```
## 2          2
## 4          1
## 5          3
## 9          3
## 10         2
```

#4. Interpret the graph # Install the jpeg package if you haven't already `install.packages("jpeg")`

```
# Load the jpeg package
library(jpeg)
```

```
# Read the JPEG image
image <- readJPEG("C:\\Users\\Acer\\OneDrive\\Pictures\\Picturee.jpg")
```

```
# Display the image
plot(as.raster(image))
```

```
# Example interpretation text
```

```
interpretation <- "The graph displays the sentiment of tweets per day from July 14th to July 21st, 2017.
Overall Trend: The graph shows a general trend of negative sentiment dominating throughout the week, with a slight increase in positive sentiment towards the end.
Day-by-Day Breakdown:
```

```
July 14th: The majority of tweets were negative, followed by neutral and then positive.
```

```
July 15th: The number of negative tweets increased significantly, followed by neutral and then positive.
```

```
July 17th: The number of negative tweets remained high, followed by positive and then neutral.
```

```
July 18th: The number of negative tweets decreased slightly, followed by positive and then neutral.
```

```
July 20th: The number of negative tweets remained relatively low, followed by neutral and then positive.
```

```
July 21st: The number of negative tweets increased again, followed by positive and then neutral."
```

```
# Position the text (adjust coordinates as needed)
```

```
text(x = 1, y = 1, labels = interpretation, adj = c(.1, 0), cex = .5)
```

Sentiments Of Tweets Per Day

