RWorksheet_Regacho#3b

1. Create a data frame using the table below.

a. Write the codes.

```
ID <- 1:6
Sex <- c("Male", "Female", "Male", "Female", "Female", "Male")
Type_of_House <- c("Wood", "Concrete", "Semi-Concrete", "Wood", "Concrete", "Wood")
Father_Occupation <- c("Farmer","Driver", "Other", "Farmer", "Driver", "Farmer")
Siblings_Attending <- c(3,5,2,6,5,4)

household <- data.frame(ID,Sex, Type_of_House, Father_Occupation, Siblings_Attending,
stringsAsFactors = FALSE)
household
```

```
##   ID    Sex Type_of_House Father_Occupation Siblings_Attending
## 1  1   Male          Wood            Farmer                 3
## 2  2 Female      Concrete            Driver                 5
## 3  3   Male Semi-Concrete             Other                 2
## 4  4 Female          Wood            Farmer                 6
## 5  5 Female      Concrete            Driver                 5
## 6  6   Male          Wood            Farmer                 4
```

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

```
str(household)
```

```
## 'data.frame':    6 obs. of  5 variables:
##  $ ID                : int  1 2 3 4 5 6
##  $ Sex               : chr  "Male" "Female" "Male" "Female" ...
##  $ Type_of_House     : chr  "Wood" "Concrete" "Semi-Concrete" "Wood" ...
##  $ Father_Occupation : chr  "Farmer" "Driver" "Other" "Farmer" ...
##  $ Siblings_Attending: num  3 5 2 6 5 4
```

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

```
mean_siblings <-
mean(household$Siblings_Attending)
mean_siblings
```

```
## [1] 4.166667
```

```r
mean_siblings == 5
```

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

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

```r
household[1:2,]
```

```
##   ID    Sex Type_of_House Father_Occupation Siblings_Attending
## 1  1   Male          Wood            Farmer                  3
## 2  2 Female      Concrete            Driver                  5
```

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

```r
household[c(3,5), c(2,4)]
```

```
##      Sex Father_Occupation
## 3   Male             Other
## 5 Female            Driver
```

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

```r
types_houses <- household$Type_of_House
types_houses
```

```
## [1] "Wood"          "Concrete"      "Semi-Concrete" "Wood"
## [5] "Concrete"      "Wood"
```

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

```r
male_farmers <- subset(household, Sex == "Male" & Father_Occupation == "Farmer")
male_farmers
```

```
##   ID  Sex Type_of_House Father_Occupation Siblings_Attending
## 1  1 Male          Wood            Farmer                  3
## 6  6 Male          Wood            Farmer                  4
```

```r
#subset() is readable and convenient
```

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.

```r
female_ge5 <- subset(household, Sex == "Female" & Siblings_Attending >=5)
female_ge5
```

```
##   ID    Sex Type_of_House Father_Occupation Siblings_Attending
## 2  2 Female      Concrete            Driver                  5
## 4  4 Female          Wood            Farmer                  6
## 5  5 Female      Concrete            Driver                  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)) a. Describe the results.

```
df= data.frame(Ints=integer(), Doubles= double(), Characters= character(), Logicals= logical(), Factors=
print("Structure of the empty dataframes:")
```

```
## [1] "Structure of the empty dataframes:"
```

```
str(df)
```

```
## 'data.frame':    0 obs. of  5 variables:
##  $ Ints      : int
##  $ Doubles   : num
##  $ Characters: chr
##  $ Logicals  : logi
##  $ Factors   : Factor w/ 0 levels:
```

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

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

Figure 2: Figure 2: Sentiment Analysis a. Import the csv file into the R environment. Write the codes.

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

```
##   ID    Sex Type_of_House Father_Occupation Siblings_Attending
## 1  1   Male          Wood            Farmer                  3
## 2  2 Female      Concrete            Driver                  5
## 3  3   Male Semi-Concrete             Other                  2
## 4  4 Female          Wood            Farmer                  6
## 5  5 Female      Concrete            Driver                  5
## 6  6   Male          Wood            Farmer                  4
```

```
# Save then import to demonstrate persistence/reproducibility. In RStudio make sure your working direct
```

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_imported$Sex <-
factor(household_imported$Sex, levels = c("Male", "Female"))
as_integer_sex <-
as.integer(household_imported$Sex)
data.frame(Sex = household_imported$Sex, Sex_int = as_integer_sex)
```

```
##      Sex Sex_int
## 1   Male       1
## 2 Female       2
## 3   Male       1
## 4 Female       2
## 5 Female       2
## 6   Male       1
```

```
#factor(..., levels =..) sets the internal codes - Male to 1, Female to 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.

```
household_imported$Type_of_House <-
factor(household_imported$Type_of_House, levels = c("Wood", "Concrete", "Semi-Concrete"))
household_imported$Type_of_House_int <-
as.integer(household_imported$Type_of_House)
household_imported
```

```
##   ID    Sex Type_of_House Father_Occupation Siblings_Attending
## 1  1   Male          Wood            Farmer                  3
## 2  2 Female      Concrete            Driver                  5
## 3  3   Male Semi-Concrete             Other                  2
## 4  4 Female          Wood            Farmer                  6
## 5  5 Female      Concrete            Driver                  5
## 6  6   Male          Wood            Farmer                  4
##   Type_of_House_int
## 1                 1
## 2                 2
## 3                 3
## 4                 1
## 5                 2
## 6                 1
```

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

```
household_imported$Father_Occupation <-
  as.integer(household_imported$Father_Occupation)
```

```
## Warning: NAs introduced by coercion
```

```
household_imported
```

```
##   ID    Sex Type_of_House Father_Occupation Siblings_Attending
## 1  1   Male          Wood                NA                  3
## 2  2 Female      Concrete                NA                  5
## 3  3   Male Semi-Concrete                NA                  2
## 4  4 Female          Wood                NA                  6
## 5  5 Female      Concrete                NA                  5
## 6  6   Male          Wood                NA                  4
```

```
##   Type_of_House_int
## 1                 1
## 2                 2
## 3                 3
## 4                 1
## 5                 2
## 6                 1
```

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

```
female_driver <- subset(household_imported, Sex == "Female" & Father_Occupation == "Driver")
female_driver
```

```
## [1] ID                 Sex                Type_of_House      Father_Occupation
## [5] Siblings_Attending Type_of_House_int
## <0 rows> (or 0-length row.names)
```

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

```
sib_ge5 <- subset(household_imported, Siblings_Attending >=5)
sib_ge5
```

```
##   ID    Sex Type_of_House Father_Occupation Siblings_Attending
## 2  2 Female      Concrete                NA                  5
## 4  4 Female          Wood                NA                  6
## 5  5 Female      Concrete                NA                  5
##   Type_of_House_int
## 2                 2
## 4                 1
## 5                 2
```

4. Interpret the graph.

#The bar plot shows that Concrete houses are the most common, followed by Wood. The Farmer occupation appears most often among fathers, and households with 5 or more siblings attending school are less frequent. This suggests a concentration of certain house types and occupations in the sample and a skew toward smaller numbers of attending siblings