RWorksheet_Gerona#3b

Mariel M. Gerona

1. Create a data frame using the table below

#1a. Create a data frame

```
df <- data.frame(
  Name = c("John", "Alice", "Peter", "Emma", "Mark"),
  Siblings = c(4, 5, 3, 6, 5),
  Sex = c("Male", "Female", "Male", "Female", "Male"),
  Father_Occupation = c("Farmer", "Driver", "Farmer", "Others", "Farmer"),
  Type_of_House = c("Wood", "Concrete", "Semi-Concrete", "Wood", "Concrete"))
```

#1b. Get structure and summary of the data

```
str(df)
```

```
## 'data.frame':    5 obs. of  5 variables:
##  $ Name             : chr  "John" "Alice" "Peter" "Emma" ...
##  $ Siblings         : num  4 5 3 6 5
##  $ Sex              : chr  "Male" "Female" "Male" "Female" ...
##  $ Father_Occupation: chr  "Farmer" "Driver" "Farmer" "Others" ...
##  $ Type_of_House    : chr  "Wood" "Concrete" "Semi-Concrete" "Wood" ...
```

```
summary(df)
```

```
##      Name              Siblings       Sex            Father_Occupation
##  Length:5           Min.   :3.0   Length:5           Length:5
##  Class :character   1st Qu.:4.0   Class :character   Class :character
##  Mode  :character   Median :5.0   Mode  :character   Mode  :character
##                     Mean   :4.6
##                     3rd Qu.:5.0
##                     Max.   :6.0
##  Type_of_House
##  Length:5
##  Class :character
##  Mode  :character
##
##
##
```

#1c. Check if the mean number of siblings is 5

```
mean_siblings <- mean(df$Siblings)
mean_siblings == 5
```

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

#1d. Extract the first two rows and all columns using the subsetting functions

```r
first_two_rows <- df[1:2, ]
print(first_two_rows)
```

```
##    Name Siblings    Sex Father_Occupation Type_of_House
## 1  John        4   Male            Farmer          Wood
## 2 Alice        5 Female            Driver      Concrete
```

#1e. Extract 3rd and 5th rows with 2nd and 4th columns

```r
selected_rows_cols <- df[c(3, 5), c(2, 4)]
print(selected_rows_cols)
```

```
##   Siblings Father_Occupation
## 3        3            Farmer
## 5        5            Farmer
```

#1f. Select the variable "Type of Houses" then store the vector that results as types_houses

```r
types_houses <- df$Type_of_House
```

#1g. Select only all male respondents whose father's occupation was farmer

```r
males_farmers <- df[df$Sex == "Male" & df$Father_Occupation == "Farmer", ]
print(males_farmers)
```

```
##    Name Siblings  Sex Father_Occupation Type_of_House
## 1  John        4 Male            Farmer          Wood
## 3 Peter        3 Male            Farmer Semi-Concrete
## 5  Mark        5 Male            Farmer      Concrete
```

#1h. Select only all female respondents that have greater than or equal to 5 number of siblings attending schools

```r
females_siblings <- df[df$Sex == "Female" & df$Siblings >= 5, ]
print(females_siblings)
```

```
##    Name Siblings    Sex Father_Occupation Type_of_House
## 2 Alice        5 Female            Driver      Concrete
## 4  Emma        6 Female            Others          Wood
```

2. Create an empty data frame and describe the result

```r
df_empty <- data.frame(
  Ints = integer(),
  Doubles = double(),
  Characters = character(),
  Logicals = logical(),
  Factors = factor(),
  stringsAsFactors = FALSE)
```

Display the structure of the empty dataframe

```r
print("Structure of the empty dataframe:")
```

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

```r
str(df_empty)
```

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

3. Save the data frame to CSV

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

a. Import the CSV file into the R environment

```r
df_import <- read.csv("HouseholdData.csv")
```

b. Convert the Sex into factor using factor() function and change it into integer.

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

```
##    Name Siblings Sex Father_Occupation Type_of_House
## 1  John        4   1            Farmer          Wood
## 2 Alice        5   2            Driver      Concrete
## 3 Peter        3   1            Farmer Semi-Concrete
## 4  Emma        6   2            Others          Wood
## 5  Mark        5   1            Farmer      Concrete
```

c. Convert the Type of Houses into factor and change it into integer.

```r
df <- data.frame(Type_of_Houses = c("Wood", "Concrete", "Semi-Concretes", "Wood", "Concrete"))
df$Type_of_Houses <- factor(df$Type_of_Houses, levels = c("Wood", "Concrete", "Semi-Concretes"), labels
df$Type_of_Houses <- as.integer(df$Type_of_Houses)
print(df)
```

```
##   Type_of_Houses
## 1              1
## 2              2
## 3              3
## 4              1
## 5              2
```

d. Convert "Father_Occupation" to factor with Farmer = 1, Driver = 2, Others = 3

```
df_import$Father_Occupation <- factor(df_import$Father_Occupation, levels = c("Farmer", "Driver", "Othe
print(df_import)
```

```
##    Name Siblings Sex Father_Occupation Type_of_House
## 1  John        4   1                 1          Wood
## 2 Alice        5   2                 2      Concrete
## 3 Peter        3   1                 1 Semi-Concrete
## 4  Emma        6   2                 3          Wood
## 5  Mark        5   1                 1      Concrete
```

e. Select all Female respondents whose father is a Driver

```
female_driver <- subset(df_import, Sex == 2 & Father_Occupation == 2)
print(female_driver)
```

```
##    Name Siblings Sex Father_Occupation Type_of_House
## 2 Alice        5   2                 2      Concrete
```

f. Select respondents with 5 or more siblings attending school

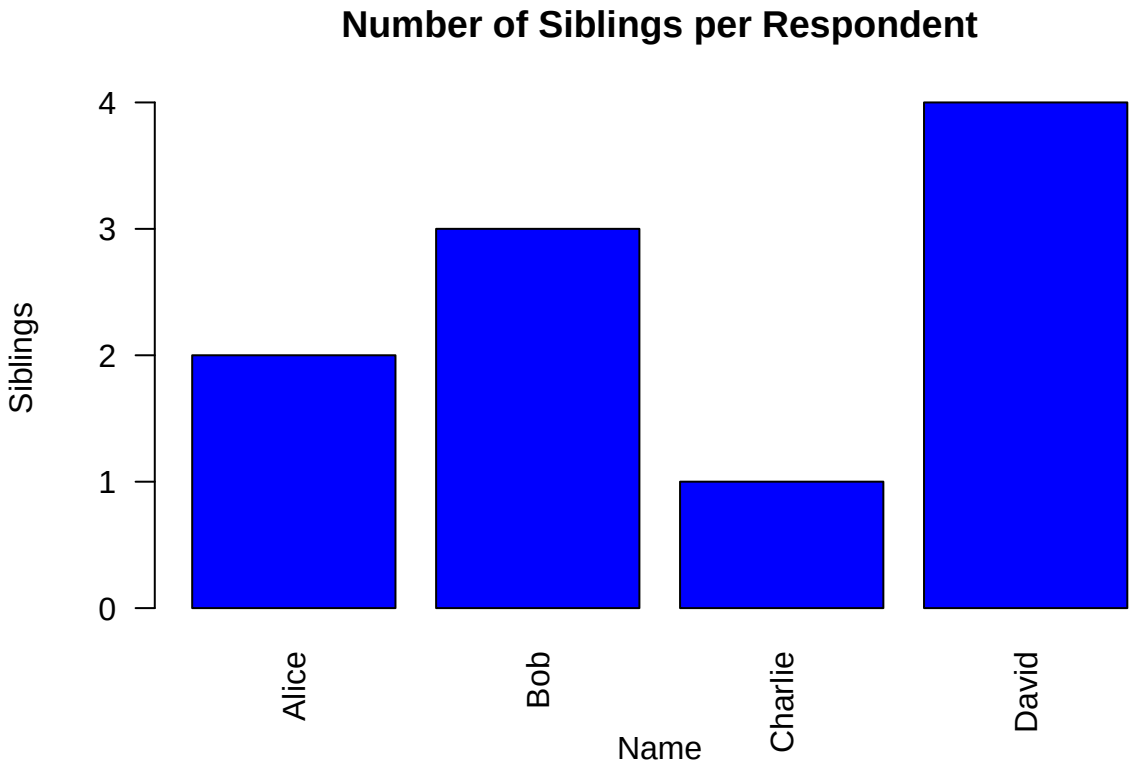
```
siblings_5_or_more <- subset(df_import, Siblings >= 5)
print(siblings_5_or_more)
```

```
##    Name Siblings Sex Father_Occupation Type_of_House
## 2 Alice        5   2                 2      Concrete
## 4  Emma        6   2                 3          Wood
## 5  Mark        5   1                 1      Concrete
```

4. Interpret the graph

```
df <- data.frame(
  Name = c("Alice", "Bob", "Charlie", "David"),
  Siblings = c(2, 3, 1, 4))

barplot(df$Siblings, names.arg = df$Name,
        main = "Number of Siblings per Respondent",
        xlab = "Name", ylab = "Siblings",
        col = "blue", las = 2)
```


Number of Siblings per Respondent
Siblings
0
1
2
3
4
Alice
Bob
Charlie
David
Name