# RWorksheet_Camarista#3b

## John Lyxton Camarista

## 2024-10-02

```r
# a. Write the codes.
dataFrame <- data.frame(
  Respondents = 1:20,
  Sex = c(2, 2, 1, 2, 2, 2, 2, 1, 2, 1, 2, 1, 2, 2, 1, 2, 1, 2, 1, 2),
  FathersOccupation = c(1, 3, 3, 3, 1, 2, 2, 3, 1, 3, 1, 3, 1, 2, 2, 3, 1, 3, 2, 1),
  PersonsAtHome = c(5, 7, 3, 8, 9, 6, 7, 8, 4, 6, 8, 5, 6, 9, 1, 6, 11, 7, 7, 6),
  SiblingsAtSchool = c(6, 4, 3, 1, 8, 5, 5, 2, 2, 3, 1, 2, 2, 1, 1, 6, 5, 1, 2, 2),
  TypesOfHouses = c(1, 3, 1, 3, 3, 2, 1, 3, 3, 2, 3, 2, 1, 3, 2, 3, 1, 3, 3, 2))
dataFrame
```

**1. Create a data frame using the table below.**

```
##    Respondents Sex FathersOccupation PersonsAtHome SiblingsAtSchool
## 1            1   2                 1             5                6
## 2            2   2                 3             7                4
## 3            3   1                 3             3                3
## 4            4   2                 3             8                1
## 5            5   2                 1             9                8
## 6            6   2                 2             6                5
## 7            7   2                 2             7                5
## 8            8   1                 3             8                2
## 9            9   2                 1             4                2
## 10          10   1                 3             6                3
## 11          11   2                 1             8                1
## 12          12   1                 3             5                2
## 13          13   2                 1             6                2
## 14          14   2                 2             9                1
## 15          15   1                 2             1                1
## 16          16   2                 3             6                6
## 17          17   1                 1            11                5
## 18          18   2                 3             7                1
## 19          19   1                 2             7                2
## 20          20   2                 1             6                2
##    TypesOfHouses
## 1              1
## 2              3
## 3              1
## 4              3
## 5              3
```

1

```
## 6               2
## 7               1
## 8               3
## 9               3
## 10              2
## 11              3
## 12              2
## 13              1
## 14              3
## 15              2
## 16              3
## 17              1
## 18              3
## 19              3
## 20              2
```

```r
#b. Describe the data. Get the structure or the summary of the data
str(dataFrame)
```

```
## 'data.frame':    20 obs. of  6 variables:
##  $ Respondents     : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Sex             : num  2 2 1 2 2 2 2 1 2 1 ...
##  $ FathersOccupation: num  1 3 3 3 1 2 2 3 1 3 ...
##  $ PersonsAtHome   : num  5 7 3 8 9 6 7 8 4 6 ...
##  $ SiblingsAtSchool : num  6 4 3 1 8 5 5 2 2 3 ...
##  $ TypesOfHouses   : num  1 3 1 3 3 2 1 3 3 2 ...
```

```r
summary(dataFrame)
```

```
##   Respondents          Sex        FathersOccupation PersonsAtHome
##  Min.   : 1.00   Min.   :1.00   Min.   :1.00      Min.   : 1.00
##  1st Qu.: 5.75   1st Qu.:1.00   1st Qu.:1.00      1st Qu.: 5.75
##  Median :10.50   Median :2.00   Median :2.00      Median : 6.50
##  Mean   :10.50   Mean   :1.65   Mean   :2.05      Mean   : 6.45
##  3rd Qu.:15.25   3rd Qu.:2.00   3rd Qu.:3.00      3rd Qu.: 8.00
##  Max.   :20.00   Max.   :2.00   Max.   :3.00      Max.   :11.00
##  SiblingsAtSchool TypesOfHouses
##  Min.   :1.00     Min.   :1.00
##  1st Qu.:1.75     1st Qu.:1.75
##  Median :2.00     Median :2.50
##  Mean   :3.10     Mean   :2.25
##  3rd Qu.:5.00     3rd Qu.:3.00
##  Max.   :8.00     Max.   :3.00
```

- the data frame has 20 rows and 6 columns.
  - There are 20 respondents
  - In sex, 1 is for male and 2 is for female
  - In father's occupation, 1 is for farmer, 2 for driver, and 3 for others
  - The Person's at home shows how many people living in the respondent's house
  - Siblings at schools shows how many siblings of the respondents are still attending school
  - In types of house, 1 is for wood, 2 is for semi-concrete ans 3 for concrete

```r
#c. Is the mean number of siblings attending is 5?
meanSiblings <- mean(dataFrame$SiblingsAtSchool)

print(paste(meanSiblings, "is the mean of Sibling attending school"))
```

```
## [1] "3.1 is the mean of Sibling attending school"
```

```r
#d. Extract the 1st two rows and then all the columns using the subsetting functions.
#Write the codes and its output.
subset(dataFrame[1:2, ])
```

```
##   Respondents Sex FathersOccupation PersonsAtHome SiblingsAtSchool
## 1           1   2                1             5                6
## 2           2   2                3             7                4
##   TypesOfHouses
## 1             1
## 2             3
```

```r
#e. Extract 3rd and 5th row with 2nd and 4th column. Write the codes and its result.
dataFrame[c(3, 5), c(2, 4)]
```

```
##   Sex PersonsAtHome
## 3   1             3
## 5   2             9
```

```r
#f. Select the variable types of houses then store the vector that results as types_houses.
#Write the codes.
types_houses <- dataFrame$TypesOfHouses
types_houses
```

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

```r
#g. Select only all Males respondent that their father occupation was farmer.
#Write the codes and its output.
male_farmer <- dataFrame[dataFrame$Sex == 1 & dataFrame$FathersOccupation == 1, ]
male_farmer
```

```
##    Respondents Sex FathersOccupation PersonsAtHome SiblingsAtSchool
## 17          17   1                1            11                5
##    TypesOfHouses
## 17             1
```

```r
#h. Select only all females respondent that have greater than or equal to 5 number of siblings attending
#Write the codes and its outputs.
female_siblings <- dataFrame[dataFrame$Sex == 2 & dataFrame$SiblingsAtSchool >= 5, ]
female_siblings
```

```
##   Respondents Sex FathersOccupation PersonsAtHome SiblingsAtSchool
## 1           1   2                1             5                6
## 5           5   2                1             9                8
```

```
## 6               6   2                  2              6                5
## 7               7   2                  2              7                5
## 16             16   2                  3              6                6
##      TypesOfHouses
## 1                1
## 5                3
## 6                2
## 7                1
## 16               3
```

#### 2. Write a R program to create an empty data frame. Using the following codes:

```r
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:"
```

```r
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 structure of the empty data frame shows no data but has predefined columns with their respective data types.

```r
#a. Import the csv file into the R environment. Write the codes.
HouseholdData <- read.csv("E:/Github/Data Science Worksheets/DataScience_Worksheets_Camarista/Worksheet
HouseholdData
```

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

```
##   Respondents    sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 1           1   Male                  1               5                  2
## 2           2 Female                  2               7                  3
## 3           3 Female                  3               3                  0
## 4           4   Male                  3               8                  5
## 5           5   Male                  1               6                  2
## 6           6 Female                  2               4                  3
```

```
## 7            7 Female                2              4                1
## 8            8   Male                3              2                2
## 9            9 Female                1             11                6
## 10          10   Male                3              6                2
##    Types.of.Houses
## 1             Wood
## 2         Concrete
## 3         Concrete
## 4             Wood
## 5    Semi-Concrete
## 6    Semi-Concrete
## 7             Wood
## 8    Semi-Concrete
## 9    Semi-Concrete
## 10        Concrete
```

```r
#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.
HouseholdData$sex <- factor(HouseholdData$sex, levels = c(1, 2), labels = c("Male", "Female"))
HouseholdData
```

```
##    Respondents  sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 1            1 <NA>                  1               5                  2
## 2            2 <NA>                  2               7                  3
## 3            3 <NA>                  3               3                  0
## 4            4 <NA>                  3               8                  5
## 5            5 <NA>                  1               6                  2
## 6            6 <NA>                  2               4                  3
## 7            7 <NA>                  2               4                  1
## 8            8 <NA>                  3               2                  2
## 9            9 <NA>                  1              11                  6
## 10          10 <NA>                  3               6                  2
##    Types.of.Houses
## 1             Wood
## 2         Concrete
## 3         Concrete
## 4             Wood
## 5    Semi-Concrete
## 6    Semi-Concrete
## 7             Wood
## 8    Semi-Concrete
## 9    Semi-Concrete
## 10        Concrete
```