# RWorksheet_MAMINTA#3b

## Ansarah Juliana Maminta

### 2025-11-03

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

#a. Write the codes.

```
matrix
```

```
## function (data = NA, nrow = 1, ncol = 1, byrow = FALSE, dimnames = NULL)
## {
##     if (is.object(data) || !is.atomic(data))
##         data <- as.vector(data)
##     .Internal(matrix(data, nrow, ncol, byrow, dimnames, missing(nrow),
##         missing(ncol)))
## }
## <bytecode: 0x5c47bfee9358>
## <environment: namespace:base>
```

```
survey <- data.frame( Respondents = 1:20, Sex = c(2,2,1,2,2,2,2,2,2,2,1,2,2,2,2,2,2,2,1,2), Fathers_Occu

print(survey)
```

```
##    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
```

```
##    Type_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
```

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

#The data has information about 20 people. For each person it is encoded the sex is either 1 for male and 2 for female , for the father's job (three types), how many people live in their home, how many siblings go to school, and what type of house they live in (three types). The number of people at home ranges from 3 to 11, and siblings at school range from 1 to 6.

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

#no, this will show what is the mean:

```
mean_siblings <- mean(survey$Siblings_at_School)
mean_siblings
```

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

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

```
subset_survey <- survey[1:2, ]

subset_survey
```

```
##   Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1           1   2                  1               5                  6
## 2           2   2                  3               7                  4
##   Type_of_Houses
## 1              1
## 2              2
```

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

```r
subset_surv <- survey[c(3,5), c(2,4)]

subset_surv
```

```
##   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.

```r
types_houses <- survey$Type_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.

```r
male_farmers <- subset(survey, Sex == 1 & Fathers_Occupation == 1)
print(male_farmers)
```

```
## [1] Respondents         Sex                 Fathers_Occupation Persons_at_Home
## [5] Siblings_at_School Type_of_Houses
## <0 rows> (or 0-length row.names)
```

#it showed nothing since there is no male respondent with a father with the occupation of a farmer.

#h. Select only all females respondent that have greater than or equal to 5 number ofsiblings attending school. Write the codes and its outputs.

```r
female_siblings <- subset(survey, Sex == 2 & Siblings_at_School == 5)
female_siblings
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 7            7   2                  3               6                  5
## 13          13   2                  1               4                  5
## 14          14   2                  3               7                  5
## 18          18   2                  1              11                  5
##    Type_of_Houses
## 7               3
## 13              2
## 14              2
## 18              3
```

#2.Write a R program to create an empty data frame.

```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:"
```

```
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 makes an empty table with five columns for different types of data like numbers, words, true or false, and categories. zThe table has no rows or data, but it's ready to hold information later.

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

```
survey <- data.frame( Respondents = 1:20, Sex = c(2,2,1,2,2,2,2,2,2,2,1,2,2,2,2,2,2,2,1,2), Fathers_Occu

write.csv(survey, file = "Householddata.csv", row.names = FALSE)
```

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

```
household_data <- read.csv("Householddata.csv")
head(household_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
##   Type_of_Houses
## 1              1
## 2              2
## 3              3
## 4              1
## 5              1
## 6              3
```

#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.

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


head(household_data$Sex)
```

```
## [1] Female Female Male   Female Female Female
## Levels: Male Female
```

#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.

```r
household_data$Type_of_Houses <- factor(household_data$Type_of_Houses,
levels = c(1, 2, 3),
labels = c("Wood", "Concrete", "Semi-Concrete"))


head(household_data$Type_of_Houses)
```

```
## [1] Wood          Concrete      Semi-Concrete Wood          Wood
## [6] Semi-Concrete
## Levels: Wood Concrete Semi-Concrete
```

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

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

head(household_data$Fathers_Occupation)
```

```
## [1] Farmer Others Others Others Farmer Driver
## Levels: Farmer Driver Others
```

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

```r
female_driver <- subset(household_data, Sex == "Female" & Fathers_Occupation == "Driver")
female_driver
```

```
##    Respondents    Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 6            6 Female             Driver               9                  1
## 12          12 Female             Driver               5                  2
##    Type_of_Houses
## 6   Semi-Concrete
## 12  Semi-Concrete
```

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

```
siblings <- subset(household_data, Siblings_at_School >= 5)
siblings
```

```
##    Respondents    Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1            1 Female             Farmer               5                  6
## 7            7 Female             Others               6                  5
## 13          13 Female             Farmer               4                  5
## 14          14 Female             Others               7                  5
## 18          18 Female             Farmer              11                  5
##    Type_of_Houses
## 1            Wood
## 7   Semi-Concrete
## 13       Concrete
## 14       Concrete
## 18  Semi-Concrete
```

#4. Interpret the graph.

#This graph shows how people felt in their tweets from July 14 to July 21, 2020. The red bars mean negative tweets, the yellow ones are neutral, and the blue ones are positive. Most of the tweets were negative, especially on July 15 and July 21, where there were more than 4,000 negative tweets. Neutral and positive tweets were fewer, but still showed up every day. This means that during that week, many people were posting tweets that showed bad or unhappy feelings.