RWorksheet_Parita#3b

2025-10-19

###1

```
#a. Write the codes
respondents <- c(1:20)
sex <- c(2, 2, 1, 2, 2, 2, 2, 2, 1, 2, 1, 2, 2, 2, 2, 2, 2, 2, 1, 2)
fathers_occ <- c(1, 3, 3, 3, 1, 2, 3, 1, 1, 1, 3, 2, 1, 3, 3, 1, 3, 1, 2, 1)
persons_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)
house_type <- c(1, 2, 3, 1, 1, 3, 3, 1, 2, 3, 2, 3, 2, 2, 3, 3, 3, 3, 3, 2)

info <- data.frame("Respondents" = respondents, "Sex" = sex,
"Fathers Occupation" = fathers_occ, "Persons at home" = persons_home,
"Siblings at school" = siblings_at_school, "Types of houses" = house_type)
```

```
#b. Describe the data. Get the structure or the summary of the data
# The data contains information about students' personal information
str(info)
```

```
## '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 2 1 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 ...
```

```
#c. Is the mean number of siblings attending 5?
mean(siblings_at_school)
```

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

```
#The answer is 2.95 so no.
```

```
#d. Extract the 1st two rows and then all the columns using the sub-setting
# functions. Write the codes and its output.
info_subset <- info[1:2,]

print(info_subset)
```

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

```r
info_subset2 <- info[c(3,5), c(2,4)]

print(info_subset2)
```

```
##   Sex Persons.at.home
## 3   1               3
## 5   2               5
```

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

types_houses <- info$Types.of.houses
```

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

```
##   Respondents Sex Fathers.Occupation Persons.at.home Siblings.at.school
## 9           9   1                  1               8                  1
##   Types.of.houses
## 9               2
```

```r
#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.
F_Siblings <- info[info$Sex == 2 & info$Siblings.at.school >= 5, ]
F_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

```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
```

```
#a. Describe the results.
# It says that there are 0 observations of 5 variables. Printing it results to
#NULL.
```

###3

```
#Create a .csv file of this. Save it as HouseholdData.csv
#a. Import the csv file into the R environment. Write the codes.
household <- read.csv("HouseholdData.csv")
print(household)
```

```
##    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         Congrete
## 3         Congrete
## 4             Wood
## 5    Semi-Congrete
## 6    Semi-Congrete
## 7             Wood
## 8    Semi-Congrete
## 9    Semi-Congrete
## 10        Congrete
```

```
#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$Sex <- factor(household$Sex, levels = c("Male", "Female"))
household$SexInt <- as.integer(household$Sex)
print(household[, c("Sex", "SexInt")])
```

```
##       Sex SexInt
## 1    Male      1
## 2  Female      2
## 3  Female      2
## 4    Male      1
## 5    Male      1
## 6  Female      2
## 7  Female      2
## 8    Male      1
## 9  Female      2
## 10   Male      1
```

```r
#c. Convert the Type of Houses into factor and change it into integer.
#[Legend: Wood = 1; Concrete = 2; Semi-Concrete = 3]. Write the R codes and its
#output.
household$Types.of.Houses <- factor(household$Types.of.Houses, levels =
                                      c("Wood", "Congrete", "Semi-Congrete"))
household$Types.of.HousesInt <- as.integer(household$Types.of.Houses)
print(household[, c("Types.of.Houses", "Types.of.HousesInt")])
```

```
##    Types.of.Houses Types.of.HousesInt
## 1             Wood                  1
## 2         Congrete                  2
## 3         Congrete                  2
## 4             Wood                  1
## 5    Semi-Congrete                  3
## 6    Semi-Congrete                  3
## 7             Wood                  1
## 8    Semi-Congrete                  3
## 9    Semi-Congrete                  3
## 10        Congrete                  2
```

```r
#d. On father's occupation, factor it as Farmer = 1; Driver = 2; and Others = 3.
#What is the R code and its output?
household$Fathers.Occupation <- factor(household$Fathers.Occupation, levels =
                            c(1, 2, 3), labels = c("Farmer", "Driver", "Other"))
household$Fathers.OccupationInt <- as.integer(household$Fathers.Occupation)
print(household[, c("Fathers.Occupation", "Fathers.OccupationInt")])
```

```
##    Fathers.Occupation Fathers.OccupationInt
## 1              Farmer                    1
## 2              Driver                    2
## 3               Other                    3
## 4               Other                    3
## 5              Farmer                    1
## 6              Driver                    2
## 7              Driver                    2
## 8               Other                    3
## 9              Farmer                    1
## 10              Other                    3
```

```r
#e. Select only all females respondent that has a father whose occupation is
#driver. Write the codes and its output.
filtered_a <- household[
  household$Sex == "Female" & household$Fathers.Occupation == "Driver",]
filtered_a
```

```
##   Respondents    Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 2           2 Female             Driver               7                  3
## 6           6 Female             Driver               4                  3
## 7           7 Female             Driver               4                  1
##   Types.of.Houses SexInt Types.of.HousesInt Fathers.OccupationInt
## 2        Congrete      2                  2                     2
## 6   Semi-Congrete      2                  3                     2
## 7            Wood      2                  1                     2
```

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

```r
filtered_b <- household[household$Siblings.at.School >= 5,]
filtered_b
```

```
##   Respondents    Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 4           4   Male              Other               8                  5
## 9           9 Female             Farmer              11                  6
##   Types.of.Houses SexInt Types.of.HousesInt Fathers.OccupationInt
## 4            Wood      1                  1                     3
## 9   Semi-Congrete      2                  3                     1
```

###4

```r
date_data <- c("July 14, 2020", "July 15, 2020", "July 17, 2020",
               "July 18, 2020", "July 20, 2020", "July 21, 2020")

negative <- c(2450, 4200, 3250, 3250, 2350, 4050)
neutral <- c(1575, 2800, 1900, 2050, 1450, 2700)
positive <- c(1725, 3200, 2400, 2525, 1700, 3375)

sentiments_tweets_df <- data.frame(
    Date = as.Date(date_data, format = "%B %d, %Y"),
    Negative = negative,
    Neutral = neutral,
    Positive = positive)

print(sentiments_tweets_df)
```

```
##         Date Negative Neutral Positive
## 1 2020-07-14     2450    1575     1725
## 2 2020-07-15     4200    2800     3200
## 3 2020-07-17     3250    1900     2400
## 4 2020-07-18     3250    2050     2525
## 5 2020-07-20     2350    1450     1700
## 6 2020-07-21     4050    2700     3375
```

```r
#The most interesting part about this tweet data are the spikes of negative
#tweets on the days July 14 and 21. Another thing is that the numbers are
#proportional on all days with negative tweets being the highest, neutral the
#lowest, and positive in between them.
```