

# RWorksheet\_Caoyonan#3b.Rmd

Leanie G. Caoyonan

2025-10-13

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

#a. Write the codes.
#b. Describe the data. Get the structure or the summary of the data

respondents <- c(1,2,3,4,5,6,7,8,9,10,
                 11,12,13,14,15,16,17,18,19,20)
sex <- c(2,2,1,2,2,2,2,2,2,
        1,2,2,2,2,2,2,2,1,2)
fathers_occupation <- c(1,3,3,3,1,2,3,1,1,1,
                         3,2,1,3,3,1,3,1,2,1)
persons_at_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)
types_of_houses <- c(1,2,3,1,1,3,3,1,2,3,
                     2,3,2,2,3,3,3,3,3,2)

data <- data.frame(respondents, sex, fathers_occupation,
                    persons_at_home, siblings_at_school, types_of_houses)
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
## 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
```

```

##      types_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

#Structure and Summary of the data.
str(data)

## 'data.frame':   20 obs. of  6 variables:
## $ respondents     : num  1 2 3 4 5 6 7 8 9 10 ...
## $ sex             : num  2 2 1 2 2 2 2 2 2 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 ...

summary(data)

##    respondents       sex   fathers_occupation persons_at_home
## Min.   : 1.00   Min.   :1.00   Min.   :1.00   Min.   : 3.0
## 1st Qu.: 5.75   1st Qu.:2.00   1st Qu.:1.00   1st Qu.: 5.0
## Median :10.50   Median :2.00   Median :2.00   Median : 7.0
## Mean   :10.50   Mean   :1.85   Mean   :1.95   Mean   : 6.4
## 3rd Qu.:15.25   3rd Qu.:2.00   3rd Qu.:3.00   3rd Qu.: 8.0
## Max.   :20.00   Max.   :2.00   Max.   :3.00   Max.   :11.0
##    siblings_at_school types_of_houses
## Min.   :1.00   Min.   :1.0
## 1st Qu.:2.00   1st Qu.:2.0
## Median :2.50   Median :2.5
## Mean   :2.95   Mean   :2.3
## 3rd Qu.:4.25   3rd Qu.:3.0
## Max.   :6.00   Max.   :3.0

#c. Is the mean number of siblings attending is 5?
#No, the mean based on the data is 2.95 as the display shown.
mean_siblings <- mean(data$siblings_at_school)
mean_siblings == 5

## [1] FALSE

```

```

mean_siblings

## [1] 2.95

#d. Extract the 1st two rows and then all the columns using the sub-setting functions. Write the codes
data[1:2, ]

##   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.
data [c(3, 5), c(2, 4)]

##   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. Write the code
types_houses <- data$types_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
male_farmer <- subset(data, sex == 1 & fathers_occupation == 1)
male_farmer

## [1] respondents      sex          fathers_occupation persons_at_home
## [5] siblings_at_school types_of_houses
## <0 rows> (or 0-length row.names)

#h. Select only all females respondent that have greater than or equal to 5 number of siblings attending school. Write the codes
female_5plus <- subset(data, sex == 2 & siblings_at_school >= 5)
female_5plus

##   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. 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:")

## [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.
#Describe: The result shows that the empty data frame has 0 rows and 5 columns with different data
#types.
#The str() output confirms the structure by showing the column names and their corresponding types
#even though no data is stored yet.

#3. Create a .csv file of this. Save it as HouseholdData.csv
#a. Import the csv file into the R environment. Write the codes.
data <- read.csv("HouseholdData.csv", stringsAsFactors = FALSE)
data

##      Sex Types_of_Houses Fathers_Occupation Num_Siblings_Attending
## 1    Male          Wood            Farmer                 4
## 2 Female        Concrete         Driver                 6
## 3    Male    Semi-Concrete       Others                 5
## 4 Female          Wood           Farmer                 3
## 5    Male        Concrete         Driver                 7

#b. Convert the Sex into factor using factor() function and change it into integer.[Legend: Male = 1 and
data$Sex <- factor(data$Sex, levels = c("Male", "Female"))
as.integer(data$Sex)

## [1] 1 2 1 2 1

#c.
# Correcting column reference
data$Types_of_Houses <- factor(data$Types_of_Houses,
                                levels = c("Wood", "Concrete", "Semi-concrete"))

# Converting the factor to integers
as.integer(data$Types_of_Houses)

## [1] 1 2 NA 1 2

#d.
data$Fathers_Occupation <- factor(data$Fathers_Occupation,
                                      levels = c(1,2,3),
                                      labels = c("Farmer", "Driver", "Others"))
as.integer(data$Fathers_Occupation)

```

```

## [1] NA NA NA NA NA
#e.
subset(data, Sex == "Female" & Fathers_Occupation == "Driver")

## [1] Sex           Types_of_Houses      Fathers_Occupation
## [4] Num_Siblings_Attending
## <0 rows> (or 0-length row.names)

# f.
subset(data, siblings_at_school >= 5)

##          Sex Types_of_Houses Fathers_Occupation Num_Siblings_Attending
## 1     Male        Wood            <NA>                  4
## NA    <NA>       <NA>            <NA>                 NA
## NA.1  <NA>       <NA>            <NA>                 NA
## NA.2  <NA>       <NA>            <NA>                 NA
## NA.3  <NA>       <NA>            <NA>                 NA

#4. Interpret the Graph
#The graph shows the sentiments of tweets per day. It shows the dates from July 14 to July 21, 2020
# with the sentiments of Positive, Neutral, and Negative.
#The graph shows the dates with the highest negative score, which are July 15 and July 21

```