# RWorksheet_Tupaz#3b

## Lorie Mae Tupaz

## 2024-10-13

```
#1.a

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_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)

# Combine into a data frame
df <- data.frame(
  Respondents = respondents,
  Sex = sex,
  Fathers_Occupation = fathers_occupation,
  Persons_at_Home = persons_at_home,
  Siblings_at_School = siblings_at_school,
  Types_of_Houses = types_of_houses
)

# Display the data frame
print(df)
```

```
##    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
```

```r
#1.b
```

```r
# Get a summary of the data frame
summary(df)
```

```
##   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
```

```r
# Check the structure of the data
str(df)
```

```
## '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 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 ...
```

```
#1.c

# Calculate the mean number of siblings attending school
mean_siblings <- mean(df$Siblings_at_School)
mean_siblings
```

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

```
# Check if the mean is equal to 5
mean_check <- mean_siblings == 5
mean_check
```

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

```
#1.d

# Extract the first two rows
first_two_rows <- df[1:2, ]
print(first_two_rows)
```

```
##   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
```

```
#1.e

# Extract 3rd and 5th rows with 2nd and 4th columns
subset_rows_cols <- df[c(3, 5), c(2, 4)]
print(subset_rows_cols)
```

```
##   Sex Persons_at_Home
## 3   1               3
## 5   2               5
```

```
#1.f

# Select the variable Types of Houses and store it as types_houses
types_houses <- df$Types_of_Houses
print(types_houses)
```

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

```
#1.g

# Select all male respondents (Sex = 1) with father's occupation as farmer (1)
male_farmers <- df[df$Sex == 1 & df$Fathers_Occupation == 1, ]
print(male_farmers)
```

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

```r
#1.h

# Select all female respondents (Sex = 2) with 5 or more siblings attending school
female_siblings <- df[df$Sex == 2 & df$Siblings_at_School >= 5, ]
print(female_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
```

```r
#2

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
```

```r
#3.a

data <- read.csv("HouseholdData.csv")

print(data)
```

```
##    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
```

*#3.b*

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

print(data$Sex)
```

```
##  [1] 1 2 2 1 1 2 2 1 2 1
## Levels: 1 2
```

*#3.c*
```r
data$Types_of_Houses <- factor(data$Types_of_Houses, levels = c("Wood", "Congrete", "Semi-congrete"), la

print(data$Types_of_Houses)
```

```
##  [1] 1 2 2 1 3 3 1 3 3 2
## Levels: 1 2 3
```

*#3.d*
```r
data$Fathers_Occupation <- factor(data$Fathers_Occupation, levels = c(1, 2, 3), labels = c("Farmer", "D:

print(data$Fathers_Occupation)
```

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

*#3.e*
```r
female_driver <- data[data$Sex == 2 & data$Fathers_Occupation == "Driver", ]

print(female_driver)
```

```
##   Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 2           2   2             Driver               7                  3
## 6           6   2             Driver               4                  3
## 7           7   2             Driver               4                  1
##   Types_of_Houses
## 2               2
## 6               3
## 7               1
```

```
#3.f
siblings_5_or_more <- data[data$Siblings_at_School >= 5, ]

print(siblings_5_or_more)
```

```
##   Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 4           4   1             Others               8                  5
## 9           9   2             Farmer              11                  6
##   Types_of_Houses
## 4               1
## 9               3
```

```
#4

#The graph illustrates the sentiments expressed in tweets: negative sentiments are represented in red,
```