Rworksheet_benjamin#3b

Mikee Lorraine S. Benjamin

2025-10-20

1. Create a data frame using the table below.

a. Write the codes.

```
# Create the data frame
df <- data.frame(
  Respondents = 1:20,
  Sex = c(2,2,1,2,2,2,2,1,2,2,1,2,2,2,2,2,2,1,2),
  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(6,4,4,1,2,1,5,3,1,2,3,2,5,5,2,2,2,5,3,2),
  Types_of_Houses = c(1,2,3,1,1,3,3,1,2,3,2,3,2,2,3,3,3,3,3,2)
)
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   1                  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                  2
## 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
```

b. Describe 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 1 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(df)
```

```
##   Respondents         Sex      Fathers_Occupation Persons_at_Home
##  Min.   : 1.00   Min.   :1.0   Min.   :1.00       Min.   : 3.0
##  1st Qu.: 5.75   1st Qu.:2.0   1st Qu.:1.00       1st Qu.: 5.0
##  Median :10.50   Median :2.0   Median :2.00       Median : 7.0
##  Mean   :10.50   Mean   :1.8   Mean   :1.95       Mean   : 6.4
##  3rd Qu.:15.25   3rd Qu.:2.0   3rd Qu.:3.00       3rd Qu.: 8.0
##  Max.   :20.00   Max.   :2.0   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   :3.00       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 school equal to 5?

```
mean(df$Siblings_at_School)
```

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

d. Extract the 1st two rows.

```
df[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.

```
df[c(3,5), c(2,4)]
```

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

f. Select the variable Types_of_Houses into a vector.

```
types_houses <- df$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 all male respondents whose father's occupation was Farmer.

```
subset(df, Sex == 1 & Fathers_Occupation == 1)
```

```
##   Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 8           8   1                  1               7                  3
##   Types_of_Houses
## 8               1
```

h. Select all female respondents with 5 siblings attending school.

```r
subset(df, Sex == 2 & Siblings_at_School >= 5)
```

```
##    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. Create an empty data frame

```r
df_empty <- data.frame(
  Ints = integer(),
  Doubles = double(),
  Characters = character(),
  Logicals = logical(),
  Factors = factor(),
  stringsAsFactors = FALSE
)

str(df_empty)
```

```
## 'data.frame':    0 obs. of  5 variables:
##  $ Ints      : int
##  $ Doubles   : num
##  $ Characters: chr
##  $ Logicals  : logi
##  $ Factors   : Factor w/ 0 levels:
```

---

3. Create a CSV file then import it

a. Import the CSV file

First create the object to save:

```r
household <- df
write.csv(household, "HouseholdData.csv", row.names = FALSE)

data <- read.csv("HouseholdData.csv")
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   1                  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                  2
## 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
```

b. Convert Sex to factor and then integer

(Male = 1, Female = 2)

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

c. Convert Type of House into factor (Wood=1; Concrete=2; Semi-Concrete=3)

```r
data$Types_of_Houses <- factor(
  data$Types_of_Houses,
```

```
  levels = c(1,2,3),
  labels = c("Wood", "Concrete", "Semi-Concrete")
)

data$Types_of_Houses <- as.integer(data$Types_of_Houses)
data$Types_of_Houses
```

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

d. Factor father's occupation (Farmer=1; Driver=2; Others=3)

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

data$Fathers_Occupation <- as.integer(data$Fathers_Occupation)
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   1                  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                  2
## 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
```

e. Select female respondents whose father is a driver.

```
female_driver <- subset(data, Sex == 2 & Fathers_Occupation == 2)
female_driver
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 6            6   2                  2               9                  1
## 12          12   2                  2               5                  2
##    Types_of_Houses
## 6                3
## 12               3
```

f. Select respondents with >= 5 siblings attending school.

```
siblings5 <- subset(data, Siblings_at_School >= 5)
siblings5
```

```
##    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
```

---

4. Interpretation of the Graph

The chart indicates that most respondents live in wooden houses, followed by concrete and semi-concrete types.

Those living in wooden houses tend to have more siblings attending school, suggesting larger families. In contrast, residents of concrete houses generally have fewer siblings, implying smaller family size and possibly higher economic stability.