Work sheet #3b

Kenan Jake Jimenez

2025-10-13

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

a. Write the codes.

```
household_df <- data.frame(
  sex = c("Male", "Female", "Female", "Male", "Male",
          "Female", "Female", "Male", "Female", "Male",
          "Male", "Female", "Male", "Female", "Male",
          "Female", "Male", "Female", "Female", "Male"),

  fathers_occupation = c("Farmer", "Driver", "Others", "Farmer", "Driver",
                         "Farmer", "Others", "Driver", "Farmer", "Others",
                         "Driver", "Farmer", "Driver", "Others", "Farmer",
                         "Driver", "Others", "Farmer", "Driver", "Others"),

  persons_at_home = c(5, 6, 4, 7, 5,
                      8, 3, 6, 5, 4,
                      6, 5, 4, 7, 5,
                      6, 5, 4, 3, 4),

  siblings_at_school = c(3, 5, 2, 6, 4,
                     7, 3, 5, 4, 2,
                     6, 5, 3, 7, 4,
                     6, 5, 4, 2, 3),

  type_of_house = c("Wood", "Concrete", "Semi-Concrete", "Wood", "Concrete",
                    "Wood", "Semi-Concrete", "Concrete", "Wood", "Semi-Concrete",
                    "Concrete", "Wood", "Semi-Concrete", "Concrete", "Wood",
                    "Semi-Concrete", "Concrete", "Wood", "Concrete", "Semi-Concrete")

)

household_df
```

```
##       sex fathers_occupation persons_at_home siblings_at_school type_of_house
## 1    Male             Farmer               5                  3          Wood
## 2  Female             Driver               6                  5      Concrete
## 3  Female             Others               4                  2 Semi-Concrete
## 4    Male             Farmer               7                  6          Wood
## 5    Male             Driver               5                  4      Concrete
## 6  Female             Farmer               8                  7          Wood
```

```
## 7  Female            Others              3                 3 Semi-Concrete
## 8    Male            Driver              6                 5      Concrete
## 9  Female            Farmer              5                 4          Wood
## 10   Male            Others              4                 2 Semi-Concrete
## 11   Male            Driver              6                 6      Concrete
## 12 Female            Farmer              5                 5          Wood
## 13   Male            Driver              4                 3 Semi-Concrete
## 14 Female            Others              7                 7      Concrete
## 15   Male            Farmer              5                 4          Wood
## 16 Female            Driver              6                 6 Semi-Concrete
## 17   Male            Others              5                 5      Concrete
## 18 Female            Farmer              4                 4          Wood
## 19 Female            Driver              3                 2      Concrete
## 20   Male            Others              4                 3 Semi-Concrete
```

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

```
str(household_df)
```

```
## 'data.frame':    20 obs. of  5 variables:
##  $ sex               : chr  "Male" "Female" "Female" "Male" ...
##  $ fathers_occupation: chr  "Farmer" "Driver" "Others" "Farmer" ...
##  $ persons_at_home   : num  5 6 4 7 5 8 3 6 5 4 ...
##  $ siblings_at_school: num  3 5 2 6 4 7 3 5 4 2 ...
##  $ type_of_house     : chr  "Wood" "Concrete" "Semi-Concrete" "Wood" ...
```

```
summary(household_df)
```

```
##      sex            fathers_occupation persons_at_home siblings_at_school
##  Length:20          Length:20          Min.   :3.0     Min.   :2.00
##  Class :character   Class :character   1st Qu.:4.0     1st Qu.:3.00
##  Mode  :character   Mode  :character   Median :5.0     Median :4.00
##                                        Mean   :5.1     Mean   :4.30
##                                        3rd Qu.:6.0     3rd Qu.:5.25
##                                        Max.   :8.0     Max.   :7.00
##  type_of_house
##  Length:20
##  Class :character
##  Mode  :character
##
##
##
```

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

```
mean(household_df$siblings_at_school)
```

```
## [1] 4.3
```

d. Extract the 1st two rows and then all the columns using the subsetting functions. Write the codes and its output.

```
household_df[1:2, ]
```

```
##      sex fathers_occupation persons_at_home siblings_at_school type_of_house
## 1   Male             Farmer               5                  3          Wood
## 2 Female             Driver               6                  5      Concrete
```

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

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

```
##   fathers_occupation siblings_at_school
## 3             Others                 2
## 5             Driver                 4
```

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

```
type_houses <- household_df$type_of_house
type_houses
```

```
##  [1] "Wood"          "Concrete"      "Semi-Concrete" "Wood"
##  [5] "Concrete"      "Wood"          "Semi-Concrete" "Concrete"
##  [9] "Wood"          "Semi-Concrete" "Concrete"      "Wood"
## [13] "Semi-Concrete" "Concrete"      "Wood"          "Semi-Concrete"
## [17] "Concrete"      "Wood"          "Concrete"      "Semi-Concrete"
```

g. Select only all Males respondent that their father occupation was farmer. Write the codes and its output.

```
subset_males_farmer <- subset(household_df, sex == "Male" & fathers_occupation == "Farmer")
subset_males_farmer
```

```
##     sex fathers_occupation persons_at_home siblings_at_school type_of_house
## 1  Male             Farmer               5                  3          Wood
## 4  Male             Farmer               7                  6          Wood
## 15 Male             Farmer               5                  4          Wood
```

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.

```
subset(household_df, sex == "Female" & siblings_at_school >= 5)
```

```
##       sex fathers_occupation persons_at_home siblings_at_school type_of_house
## 2  Female             Driver               6                  5      Concrete
## 6  Female             Farmer               8                  7          Wood
## 12 Female             Farmer               5                  5          Wood
## 14 Female             Others               7                  7      Concrete
## 16 Female             Driver               6                  6 Semi-Concrete
```

#2. Write a R program to create an empty data frame. Using the following codes: a. Describe the results.

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

str(df)
```

```
## '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 of this. Save it as HouseholdData.csv

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

```
write.csv(household_df, "HouseholdData.csv", row.names = FALSE)

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

```
##       sex fathers_occupation persons_at_home siblings_at_school type_of_house
## 1    Male             Farmer               5                  3          Wood
## 2  Female             Driver               6                  5      Concrete
## 3  Female             Others               4                  2 Semi-Concrete
## 4    Male             Farmer               7                  6          Wood
## 5    Male             Driver               5                  4      Concrete
## 6  Female             Farmer               8                  7          Wood
## 7  Female             Others               3                  3 Semi-Concrete
## 8    Male             Driver               6                  5      Concrete
## 9  Female             Farmer               5                  4          Wood
## 10   Male             Others               4                  2 Semi-Concrete
## 11   Male             Driver               6                  6      Concrete
## 12 Female             Farmer               5                  5          Wood
## 13   Male             Driver               4                  3 Semi-Concrete
## 14 Female             Others               7                  7      Concrete
## 15   Male             Farmer               5                  4          Wood
## 16 Female             Driver               6                  6 Semi-Concrete
## 17   Male             Others               5                  5      Concrete
## 18 Female             Farmer               4                  4          Wood
## 19 Female             Driver               3                  2      Concrete
## 20   Male             Others               4                  3 Semi-Concrete
```

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.

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

```
##    sex fathers_occupation persons_at_home siblings_at_school type_of_house
## 1    1             Farmer               5                  3          Wood
## 2    2             Driver               6                  5      Concrete
## 3    2             Others               4                  2 Semi-Concrete
## 4    1             Farmer               7                  6          Wood
## 5    1             Driver               5                  4      Concrete
## 6    2             Farmer               8                  7          Wood
## 7    2             Others               3                  3 Semi-Concrete
## 8    1             Driver               6                  5      Concrete
## 9    2             Farmer               5                  4          Wood
## 10   1             Others               4                  2 Semi-Concrete
## 11   1             Driver               6                  6      Concrete
## 12   2             Farmer               5                  5          Wood
## 13   1             Driver               4                  3 Semi-Concrete
## 14   2             Others               7                  7      Concrete
## 15   1             Farmer               5                  4          Wood
## 16   2             Driver               6                  6 Semi-Concrete
## 17   1             Others               5                  5      Concrete
## 18   2             Farmer               4                  4          Wood
## 19   2             Driver               3                  2      Concrete
## 20   1             Others               4                  3 Semi-Concrete
```

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.

```
data$type_of_house <- factor(data$type_of_house,
levels = c("Wood", "Concrete", "Semi-Concrete"),
labels = c(1, 2, 3))
data$type_of_house <- as.integer(data$type_of_house)
data$type_of_house
```

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

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

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

```
##    sex fathers_occupation persons_at_home siblings_at_school type_of_house
## 1    1                  1               5                  3             1
## 2    2                  2               6                  5             2
## 3    2                  3               4                  2             3
## 4    1                  1               7                  6             1
```

```
## 5    1                 2               5                 4              2
## 6    2                 1               8                 7              1
## 7    2                 3               3                 3              3
## 8    1                 2               6                 5              2
## 9    2                 1               5                 4              1
## 10   1                 3               4                 2              3
## 11   1                 2               6                 6              2
## 12   2                 1               5                 5              1
## 13   1                 2               4                 3              3
## 14   2                 3               7                 7              2
## 15   1                 1               5                 4              1
## 16   2                 2               6                 6              3
## 17   1                 3               5                 5              2
## 18   2                 1               4                 4              1
## 19   2                 2               3                 2              2
## 20   1                 3               4                 3              3
```

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

```
subset(data, sex == 2 & fathers_occupation == 2)
```

```
##    sex fathers_occupation persons_at_home siblings_at_school type_of_house
## 2    2                 2               6                 5              2
## 16   2                 2               6                 6              3
## 19   2                 2               3                 2              2
```

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

```
subset(data, siblings_at_school >= 5)
```

```
##    sex fathers_occupation persons_at_home siblings_at_school type_of_house
## 2    2                 2               6                 5              2
## 4    1                 1               7                 6              1
## 6    2                 1               8                 7              1
## 8    1                 2               6                 5              2
## 11   1                 2               6                 6              2
## 12   2                 1               5                 5              1
## 14   2                 3               7                 7              2
## 16   2                 2               6                 6              3
## 17   1                 3               5                 5              2
```

#4. Interpret the graph. The graph in general indicates that the majority of respondents reside in wooden houses, then concrete and semi-concrete homes. Wooden house residents tend to have more school-going siblings, implying that big families are prevalent in such dwellings.