RWorksheet_Freires#3b

2024-10-02

1. Create a data frame using the table below.

   a. Write the codes.

```r
respondents <- c(1:20)
print(respondents)
```

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

```r
sex <- factor(c(2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 1, 2))
print(sex)
```

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

```r
fathers_occupation <- factor(c(1, 2, 2, 2, 1, 2, 3, 1, 1, 1, 3, 2, 1, 3, 3, 1, 3, 1, 2, 1))
print(fathers_occupation)
```

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

```r
persons <- c(5, 7, 3, 8, 5, 9, 6, 7, 8, 4, 7, 5, 4, 7, 8, 8, 3, 11, 7, 6)
print(persons)
```

```
##  [1]  5  7  3  8  5  9  6  7  8  4  7  5  4  7  8  8  3 11  7  6
```

```r
siblings <- c(6, 4, 4, 1, 2, 1, 5, 3, 1, 2, 3, 2, 5, 5, 2, 1, 2, 5, 3, 2)
print(siblings)
```

```
##  [1] 6 4 4 1 2 1 5 3 1 2 3 2 5 5 2 1 2 5 3 2
```

```r
houses <- c(1, 2, 3, 1, 1, 3, 3, 1, 2, 3, 2, 3, 2, 2, 3, 3, 3, 3, 3, 2)
print(houses)
```

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

```r
data <- data.frame(respondents, sex, fathers_occupation, persons, siblings, houses)
print(data)
```

```
##    respondents sex fathers_occupation persons siblings houses
## 1            1   2                  1       5        6      1
## 2            2   2                  2       7        4      2
## 3            3   1                  2       3        4      3
## 4            4   2                  2       8        1      1
## 5            5   2                  1       5        2      1
## 6            6   2                  2       9        1      3
## 7            7   2                  3       6        5      3
## 8            8   2                  1       7        3      1
## 9            9   2                  1       8        1      2
## 10          10   2                  1       4        2      3
## 11          11   1                  3       7        3      2
```

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

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

```r
str(data)
```

```
## 'data.frame':    20 obs. of  6 variables:
##  $ respondents      : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ sex              : Factor w/ 2 levels "1","2": 2 2 1 2 2 2 2 2 2 2 ...
##  $ fathers_occupation: Factor w/ 3 levels "1","2","3": 1 2 2 2 1 2 3 1 1 1 ...
##  $ persons          : num  5 7 3 8 5 9 6 7 8 4 ...
##  $ siblings         : num  6 4 4 1 2 1 5 3 1 2 ...
##  $ houses           : num  1 2 3 1 1 3 3 1 2 3 ...
```

- The data has six objects such as respondents, sex, fathers occupation, persons at home, siblings at school, types of houses. It has the data type of character, factors and levels, and its values.

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

```r
num <- c(siblings)
mean(num)
```

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

- The mean number of siblings attending is not 5.

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

```r
subdata <- data[1:2, ]
print(subdata)
```

```
##   respondents sex fathers_occupation persons siblings houses
## 1           1   2                  1       5        6      1
## 2           2   2                  2       7        4      2
```

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

```r
extracted <- data[c(3, 5), c(2, 3)]
print(extracted)
```

```
##   sex fathers_occupation
## 3   1                  2
## 5   2                  1
```

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

```r
types_houses <- class(houses)
print(types_houses)
```

```
## [1] "numeric"
```

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

```r
male_respondents <- subset(data, sex == 1 & fathers_occupation == 1)
print(male_respondents)
```

```
## [1] respondents       sex              fathers_occupation persons
## [5] siblings          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 and its outputs.

```r
female_respondents <- subset(data, sex == 2 & siblings <= 5)
print(female_respondents)
```

```
##    respondents sex fathers_occupation persons siblings houses
## 2            2   2                  2       7        4      2
## 4            4   2                  2       8        1      1
## 5            5   2                  1       5        2      1
## 6            6   2                  2       9        1      3
## 7            7   2                  3       6        5      3
## 8            8   2                  1       7        3      1
## 9            9   2                  1       8        1      2
## 10          10   2                  1       4        2      3
## 12          12   2                  2       5        2      3
## 13          13   2                  1       4        5      2
## 14          14   2                  3       7        5      2
## 15          15   2                  3       8        2      3
## 16          16   2                  1       8        1      3
## 17          17   2                  3       3        2      3
## 18          18   2                  1      11        5      3
## 20          20   2                  1       6        2      2
```

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:") print(str(df))

```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.

- The structure of the empty frame does not have rows but has 5 data types with Int, Double, Character, Logical, and Factors.

3. Create a .csv file of this. Save it as HouseholdData.csv

```r
household_data <- data.frame(
  Respondents = 1:10,
  Sex = c("Male", "Female", "Female", "Male", "Male", "Female", "Female", "Male", "Female", "Male"),
  Fathers_Occupation = c(1, 2, 3, 3, 1, 2, 2, 1, 1, 3),
  Persons_at_Home = c(5, 7, 3, 8, 6, 4, 2, 4, 11, 6),
  Siblings_at_School = c(5, 3, 3, 5, 6, 3, 1, 2, 6, 6),
  Types_of_Houses = c("Wood", "Concrete", "Concrete", "Wood", "Semi-concrete", "Semi-concrete", "Wood",
)
print(household_data)
```

```
##     Respondents    Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1             1   Male                  1               5                  5
## 2             2 Female                  2               7                  3
## 3             3 Female                  3               3                  3
## 4             4   Male                  3               8                  5
## 5             5   Male                  1               6                  6
## 6             6 Female                  2               4                  3
## 7             7 Female                  2               2                  1
## 8             8   Male                  1               4                  2
## 9             9 Female                  1              11                  6
## 10           10   Male                  3               6                  6
##     Types_of_Houses
## 1              Wood
## 2          Concrete
## 3          Concrete
## 4              Wood
## 5     Semi-concrete
## 6     Semi-concrete
## 7              Wood
## 8     Semi-concrete
## 9     Semi-concrete
## 10         Concrete
```

```r
write.csv(household_data, file = "HouseholdData.csv", row.names = FALSE)
```

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

```r
household_data <- read.csv("HouseholdData.csv")

print(household_data)
```

```
##     Respondents    Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1             1   Male                  1               5                  5
## 2             2 Female                  2               7                  3
## 3             3 Female                  3               3                  3
## 4             4   Male                  3               8                  5
## 5             5   Male                  1               6                  6
```

```
## 6             6 Female              2              4              3
## 7             7 Female              2              2              1
## 8             8   Male              1              4              2
## 9             9 Female              1             11              6
## 10           10   Male              3              6              6
##    Types_of_Houses
## 1             Wood
## 2         Concrete
## 3         Concrete
## 4             Wood
## 5    Semi-concrete
## 6    Semi-concrete
## 7             Wood
## 8    Semi-concrete
## 9    Semi-concrete
## 10        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.

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

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1            1   1                  1               5                  5
## 2            2   2                  2               7                  3
## 3            3   2                  3               3                  3
## 4            4   1                  3               8                  5
## 5            5   1                  1               6                  6
## 6            6   2                  2               4                  3
## 7            7   2                  2               2                  1
## 8            8   1                  1               4                  2
## 9            9   2                  1              11                  6
## 10          10   1                  3               6                  6
##    Types_of_Houses
## 1             Wood
## 2         Concrete
## 3         Concrete
## 4             Wood
## 5    Semi-concrete
## 6    Semi-concrete
## 7             Wood
## 8    Semi-concrete
## 9    Semi-concrete
## 10        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.

```r
household_data$Types_of_Houses <- factor(household_data$Types_of_Houses, levels = c("Wood", "Concrete",

household_data$Types_of_Houses <- as.integer(household_data$Types_of_Houses)

print(household_data)
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1            1   1                  1               5                  5
## 2            2   2                  2               7                  3
## 3            3   2                  3               3                  3
## 4            4   1                  3               8                  5
## 5            5   1                  1               6                  6
## 6            6   2                  2               4                  3
## 7            7   2                  2               2                  1
## 8            8   1                  1               4                  2
## 9            9   2                  1              11                  6
## 10          10   1                  3               6                  6
##    Types_of_Houses
## 1                1
## 2                2
## 3                2
## 4                1
## 5                3
## 6                3
## 7                1
## 8                3
## 9                3
## 10               2
```

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

```
household_data$Fathers_Occupation <- factor(household_data$Fathers_Occupation, levels = c(1, 2, 3), lab
household_data$Fathers_Occupation <- as.integer(household_data$Fathers_Occupation)

print(household_data)
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1            1   1                  1               5                  5
## 2            2   2                  2               7                  3
## 3            3   2                  3               3                  3
## 4            4   1                  3               8                  5
## 5            5   1                  1               6                  6
## 6            6   2                  2               4                  3
## 7            7   2                  2               2                  1
## 8            8   1                  1               4                  2
## 9            9   2                  1              11                  6
## 10          10   1                  3               6                  6
##    Types_of_Houses
## 1                1
## 2                2
## 3                2
## 4                1
## 5                3
## 6                3
## 7                1
## 8                3
## 9                3
## 10               2
```

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

```
female_respondents <- subset(household_data, Sex == 2 & Fathers_Occupation == 2)
print(female_respondents)
```

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

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

```
siblings <- subset(household_data, Siblings_at_School >= 5)
print(siblings)
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1            1   1                  1               5                  5
## 4            4   1                  3               8                  5
## 5            5   1                  1               6                  6
## 9            9   2                  1              11                  6
## 10          10   1                  3               6                  6
##    Types_of_Houses
## 1                1
## 4                1
## 5                3
## 9                3
## 10               2
```

4. Interpret the graph.

- This graph illustrates the analysis of Sentiments of Tweets Per Day. It is categorized by the following sentiments such as negative, neutral, and positive.

  The negative sentiment which are the red bars are the most prominent in the graph. It reached most tweets with the number of 4000 on July 15, 2020 and July 21, 2020.

  The positive sentiment which is symbolized by the blue bars are comparatively lower through out the week with only 1500-3000 tweets.

  The neutral sentiment is more in between the negative and positive sentiments with the amount of 1500- below 3000 tweets per day.

  My interpretation of the graph shows how negative sentiments dominates the sentiments of tweets per day while neutral being the lesser topic.