

RWorksheet_rocillo#3b

Cassandra Jielin Rocillo

2024-10-02

1. Create a data frame using the table below.

a. Write the codes. Legend: Male-1
Female-2

Farmer-1 Driver- Others-3

Wood-1 Semi-concrete-2 Concete-3

```
respondents <- c(1:20)
sex <- c(2, 2, 1, 2, 2, 2, 2, 2, 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)
person_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)
type_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 = respondents, Sex = sex, Fathers_Occupation = fathers_occupation, Person_at_Home = person_at_home, Siblings_at_School = siblings_at_school, Type_of_Houses = type_of_houses)
print(data)
```

##	Respondents	Sex	Fathers_Occupation	Person_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
##	Type_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. Get the structure or the summary of the data

```
str(data)
```

```
## '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 ...
## $ Person_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 ...
## $ Type_of_Houses : num 1 2 3 1 1 3 3 1 2 3 ...
```

```
summary(data)
```

```
## Respondents Sex Fathers_Occupation Person_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 Type_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

```
mean_num <- mean(siblings_at_school)
print(mean_num)
```

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

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

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

```
## Respondents Sex Fathers_Occupation Person_at_Home Siblings_at_School
## 1          1 2              1              5              6
## 2          2 2              3              7              4
## Type_of_Houses
## 1          1
## 2          2
```

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

```
extracted_data <- data[c(3, 5), c(2, 4)]
print(extracted_data)
```

```
## Sex Person_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 codes.

```
types_houses <- class(type_of_houses)
print(types_houses)
```

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

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

```
male_farmers <- data[data$Sex == 1 & data$Fathers_Occupation == 1, ]
print(male_farmers)
```

```
## [1] Respondents      Sex      Fathers_Occupation Person_at_Home
## [5] Siblings_at_School Type_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 and its outputs.

```
fem <- data[data$Sex == 2 & data$Siblings_at_School >= 5, ]
print(fem)
```

```
## Respondents Sex Fathers_Occupation Person_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
## Type_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:") print(str(df))

```
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.

- The results shows a structure of empty data frame.

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

```
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", "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
```

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

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

```
import_data <- read.csv("HouseholdData.csv")
print(import_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.

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

```
## Respondents Sex Fathers_Occupation Person_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
##      Type_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. 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$Type_of_Houses <- factor(data$Type_of_Houses, levels = c("Wood", "Concrete", "Semi-concrete"), lab
data$Type_of_Houses <- as.integer(data$Type_of_Houses)
print(data)
```

```
##      Respondents Sex Fathers_Occupation Person_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
##      Type_of_Houses
## 1      NA
## 2      NA
## 3      NA
## 4      NA
## 5      NA
## 6      NA
## 7      NA
## 8      NA
## 9      NA
## 10     NA
## 11     NA
## 12     NA
## 13     NA
## 14     NA
## 15     NA
## 16     NA
## 17     NA
## 18     NA
## 19     NA
## 20     NA
```

- 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(1, 2, 3), labels = c("Farmer", "D
data$Fathers_Occupation <- as.integer(data$Fathers_Occupation)

print(data)
```

```
##      Respondents Sex Fathers_Occupation Person_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
##      Type_of_Houses
## 1      NA
```

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

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

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

```
##      Respondents Sex Fathers_Occupation Person_at_Home Siblings_at_School
## 6              6  2              2              9              1
## 12             12  2              2              5              2
##      Type_of_Houses
## 6              NA
## 12             NA
```

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

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

```
##      Respondents Sex Fathers_Occupation Person_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
##      Type_of_Houses
## 1              NA
## 7              NA
## 13             NA
## 14             NA
## 18             NA
```

4. Interpret the graph.

-The “Sentiments of Tweets Per Day” bar graph offers a visual breakdown of the sentiment distribution of tweets made between July 14, 2020, and July 21, 2020. Red, orange, and blue bars, correspondingly, indicate the “negative”, “neutral”, and “positive” emotion categories.

- The graph shows that “negative sentiments” lead across all of the days, peaking on July 15 and July 20 when there are more than 4,000 unfavorable tweets. This implies that there were more negative conversations or responses on social media over these days. Though they trail closely behind negative feelings, “neutral sentiments” show a more measured tone in some tweets. However, “positive sentiments” are usually the least common; at this time, positive responses were less common.