RWorksheet_#3b

Kathrina Casandra Sison

2024-10-14

1. Create a data frame using the table below.

a. Write the codes.

```
respo_data <- data.frame(
Respondents = 1:20,
Sex = c(2, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 1, 2, 2, 1, 1, 2, 1, 2, 1),
Fathers_Occupation = c(1, 3, 1, 3, 3, 1, 3, 2, 3, 1, 2, 1, 3, 1, 3, 1, 3, 1, 3, 1),
Persons_at_Home = c(5, 7, 3, 8, 9, 6, 9, 6, 4, 3, 4, 5, 7, 8, 3, 7, 11, 7, 6, 6),
Siblings_at_School = c(6, 4, 4, 1, 1, 3, 3, 5, 3, 2, 4, 2, 3, 4, 3, 3, 5, 3, 2, 2),
Types_of_Houses = c(1, 2, 1, 1, 3, 3, 3, 2, 1, 3, 1, 2, 1, 3, 1, 3, 1, 3, 2, 2)
)
```

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

```
str(respo_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 1 2 1 2 1 2 ...
##  $ Fathers_Occupation: num  1 3 1 3 3 1 3 2 3 1 ...
##  $ Persons_at_Home   : num  5 7 3 8 9 6 9 6 4 3 ...
##  $ Siblings_at_School: num  6 4 4 1 1 3 3 5 3 2 ...
##  $ Types_of_Houses   : num  1 2 1 1 3 3 3 2 1 3 ...
```

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

```
mean_siblings <- mean(respo_data$Siblings_at_School)
mean_siblings == 5
```

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

```
mean_siblings
```

```
## [1] 3.15
```

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

```
subset_data <- respo_data[1:2, ]
subset_data
```

```
##   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. Write the codes and its result.

```
subset_data_2 <- respo_data[c(3, 5), c(2, 4)]
subset_data_2
```

```
##   Sex Persons_at_Home
## 3   1               3
## 5   1               9
```

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

```
types_houses <- respo_data$Types_of_Houses
types_houses
```

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

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

```
male_farmers <- subset(respo_data, Sex == 1 & Fathers_Occupation == 1)
male_farmers
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 3            3   1                  1               3                  4
## 12          12   1                  1               5                  2
## 16          16   1                  1               7                  3
## 18          18   1                  1               7                  3
## 20          20   1                  1               6                  2
##    Types_of_Houses
## 3                1
## 12               2
## 16               3
## 18               3
## 20               2
```

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.

```
female_siblings <- subset(respo_data, Sex == 2 & Siblings_at_School >= 5)
female_siblings
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1            1   2                  1               5                  6
## 8            8   2                  2               6                  5
## 17          17   2                  3              11                  5
##    Types_of_Houses
## 1                1
## 8                2
## 17               1
```

2. Write a R program to create an empty data frame. Using the following codes: df = data.frame(Ints=integer(),

```
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:"
```

3.

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.

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

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

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

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

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.

```
household_data$Types_of_Houses <- factor(household_data$Types_of_Houses,
levels = c("Wood", "Concrete", "Semi-concrete"), labels = c(1, 2, 3))
household_data$Types_of_Houses
```

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

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

```
##  [1] <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## Levels: Farmer Driver Others
```

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

```
female_driver <- subset(household_data, Sex == 2 & Fathers_Occupation == "Driver")
female_driver
```

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

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

```
siblings_5_or_more <- subset(household_data, Siblings_at_School >= 5)
print(siblings_5_or_more)
```

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

4. Interpret the graph. The graph shows the different sentiment based on its color. The highest sentiment of tweets was negative and the lowest is neutral.