RWorksheet_lastname#3b

Angel Blase

2024-10-01

1. Create a data frame using the table below.

a. Write the codes.

```
data <- data.frame(
  Respondents = 1:20,
  Sex = c(2, 2, 1, 2, 2, 2, 2, 2, 1, 2, 3, 2, 2, 1, 2, 2, 2, 1, 1, 2),
  Fathers_Occupation = c(1, 3, 1, 3, 1, 2, 3, 1, 1, 1, 3, 2, 1, 3, 1, 2, 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, 3, 5, 3, 2, 4, 2, 5, 5, 5, 2, 1, 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)
)
```

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 1 2 ...
##  $ Fathers_Occupation: num  1 3 1 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 3 5 3 2 4 ...
##  $ Types_of_Houses   : num  1 2 3 1 1 3 3 1 2 3 ...
```

```
summary(data)
```

```
##   Respondents         Sex       Fathers_Occupation Persons_at_Home
##  Min.   : 1.00   Min.   :1.00   Min.   :1.0        Min.   : 3.0
##  1st Qu.: 5.75   1st Qu.:1.75   1st Qu.:1.0        1st Qu.: 5.0
##  Median :10.50   Median :2.00   Median :1.5        Median : 7.0
##  Mean   :10.50   Mean   :1.80   Mean   :1.8        Mean   : 6.4
##  3rd Qu.:15.25   3rd Qu.:2.00   3rd Qu.:3.0        3rd Qu.: 8.0
##  Max.   :20.00   Max.   :3.00   Max.   :3.0        Max.   :11.0
##  Siblings_at_School Types_of_Houses
##  Min.   :1.0        Min.   :1.0
##  1st Qu.:2.0        1st Qu.:2.0
##  Median :3.0        Median :2.5
##  Mean   :3.3        Mean   :2.3
##  3rd Qu.:5.0        3rd Qu.:3.0
##  Max.   :6.0        Max.   :3.0
```

The data consists of 20 respondents, with variables including their sex, father's occupation, number of persons at home, siblings attending school, and types of houses. The dataset includes both male and female respondents, and the father's occupations range from farmers to drivers and others, with varying family sizes and housing types.

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

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

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

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

```
first_two_rows <- data[1:2, ]
first_two_rows
```

```
##   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.

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

```
##   Sex Persons_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 <- data$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 only all Males respondent that their father occupation was farmer. Write the codes and its output.

```
male_farmer <- subset(data, Sex == 1 & Fathers_Occupation == 1)
male_farmer
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 3            3   1                  1               3                  4
## 9            9   1                  1               8                  2
## 18          18   1                  1              11                  5
##    Types_of_Houses
## 3                3
## 9                2
## 18               3
```

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(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
## 7            7   2                  3               6                  5
## 12          12   2                  2               5                  5
```

```
## 13          13   2                   1               4                   5
##    Types_of_Houses
## 1                1
## 7                3
## 12               3
## 13               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:")
```

```
## [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
```

Description: The R code creates an empty data frame with five columns: integers, doubles, characters, logicals, and factors. Each column is initialized without any values. The `stringsAsFactors=FALSE` option ensures that character columns remain as text and aren't converted to factors. When the structure of the data frame is printed using the "str()" function, it shows that there are zero observations and five variables. Each column is listed with its data type, but since no data has been added, all columns are empty. The factor column also has zero 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.

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

```
##    Respondents    Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1            1   Male                  1               5                  2
## 2            2 Female                  2               7                  3
## 3            3 Female                  3               3                  0
## 4            4   Male                  3               8                  5
## 5            5   Male                  1               6                  2
## 6            6 Female                  2               4                  3
## 7            7 Female                  2               4                  1
## 8            8   Male                  3               2                  2
## 9            9 Female                  1              11                  6
## 10          10   Male                  3               6                  2
##    Types_of_Houses
## 1             Wood
## 2         Congrete
```

```
## 3         Congrete
## 4             Wood
## 5    Semi-congrete
## 6    Semi-congrete
## 7             Wood
## 8    Semi-congrete
## 9    Semi-congrete
## 10        Congrete
```

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
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1            1   1                  1               5                  2
## 2            2   2                  2               7                  3
## 3            3   2                  3               3                  0
## 4            4   1                  3               8                  5
## 5            5   1                  1               6                  2
## 6            6   2                  2               4                  3
## 7            7   2                  2               4                  1
## 8            8   1                  3               2                  2
## 9            9   2                  1              11                  6
## 10          10   1                  3               6                  2
##    Types_of_Houses
## 1             Wood
## 2         Congrete
## 3         Congrete
## 4             Wood
## 5    Semi-congrete
## 6    Semi-congrete
## 7             Wood
## 8    Semi-congrete
## 9    Semi-congrete
## 10        Congrete
```

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

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1            1   1                  1               5                  2
## 2            2   2                  2               7                  3
## 3            3   2                  3               3                  0
## 4            4   1                  3               8                  5
## 5            5   1                  1               6                  2
## 6            6   2                  2               4                  3
```

```
## 7            7   2                  2               4                  1
## 8            8   1                  3               2                  2
## 9            9   2                  1              11                  6
## 10          10   1                  3               6                  2
##    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?

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

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1            1   1             Farmer               5                  2
## 2            2   2             Driver               7                  3
## 3            3   2             Others               3                  0
## 4            4   1             Others               8                  5
## 5            5   1             Farmer               6                  2
## 6            6   2             Driver               4                  3
## 7            7   2             Driver               4                  1
## 8            8   1             Others               2                  2
## 9            9   2             Farmer              11                  6
## 10          10   1             Others               6                  2
##    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_driver <- subset(data, Sex == 2 & Fathers_Occupation == "Driver")
female_driver
```

```
##   Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 2           2   2             Driver               7                  3
## 6           6   2             Driver               4                  3
## 7           7   2             Driver               4                  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_5_or_more <- subset(data, Siblings_at_School >= 5)
siblings_5_or_more
```

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

4. Interpret the graph:

The sentiment analysis suggests that across this time period, Negative and Positive tweets alternate in dominance, with spikes in both categories on specific days. Neutral sentiments are consistent but less frequent compared to Negative and Positive ones. The sentiment trends could reflect external events influencing public opinion on these days, leading to shifts in tweet sentiments.