RWorksheet_Delgado#3b.Rmd

Clyde Marcelo Delgado

2024-09-30

```
# 1. Create a data frame using the table provided
respondent_data <- data.frame(
  Respondent = 1:10,
  Sex = c('Male', 'Female', 'Female', 'Male', 'Female', 'Male', 'Male', 'Female', 'Male', 'Female'),
  Father_Occupation = c('Farmer', 'Driver', 'Others', 'Farmer', 'Driver', 'Others', 'Farmer', 'Driver',
  Num_Siblings_Attending = c(5, 2, 7, 3, 5, 2, 4, 6, 1, 3),
  Type_of_House = c('Wood', 'Concrete', 'Semi-Concrete', 'Wood', 'Concrete', 'Semi-Concrete', 'Concrete
)

# a. Describe the data. Get the structure or the summary of the data.
summary(respondent_data)
```

```
##    Respondent        Sex            Father_Occupation  Num_Siblings_Attending
##  Min.   : 1.00   Length:10          Length:10          Min.   :1.00
##  1st Qu.: 3.25   Class :character   Class :character   1st Qu.:2.25
##  Median : 5.50   Mode  :character   Mode  :character   Median :3.50
##  Mean   : 5.50                                         Mean   :3.80
##  3rd Qu.: 7.75                                         3rd Qu.:5.00
##  Max.   :10.00                                         Max.   :7.00
##  Type_of_House
##  Length:10
##  Class :character
##  Mode  :character
##
##
##
```

```
str(respondent_data)
```

```
## 'data.frame':    10 obs. of  5 variables:
##  $ Respondent            : int  1 2 3 4 5 6 7 8 9 10
##  $ Sex                   : chr  "Male" "Female" "Female" "Male" ...
##  $ Father_Occupation     : chr  "Farmer" "Driver" "Others" "Farmer" ...
##  $ Num_Siblings_Attending: num  5 2 7 3 5 2 4 6 1 3
##  $ Type_of_House         : chr  "Wood" "Concrete" "Semi-Concrete" "Wood" ...
```

```
# b. Is the mean number of siblings attending school 5?
mean_siblings <- mean(respondent_data$Num_Siblings_Attending)
mean_siblings == 5 # Output will be TRUE if mean is 5, otherwise FALSE.
```

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

```
# c. Extract the 1st two rows and all columns using subsetting functions.
first_two_rows <- respondent_data[1:2, ]
first_two_rows
```

```
##   Respondent    Sex Father_Occupation Num_Siblings_Attending Type_of_House
## 1          1   Male            Farmer                      5          Wood
## 2          2 Female            Driver                      2      Concrete
```

```
# d. Extract 3rd and 5th row with 2nd and 4th column.
rows_3_and_5_cols_2_and_4 <- respondent_data[c(3, 5), c(2, 4)]
rows_3_and_5_cols_2_and_4
```

```
##      Sex Num_Siblings_Attending
## 3 Female                      7
## 5 Female                      5
```

```
# e. Select the variable "Type_of_House" and store it as types_houses.
types_houses <- respondent_data$Type_of_House
types_houses
```

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

```
# f. Select all male respondents whose father occupation was "Farmer".
male_farmers <- respondent_data[respondent_data$Sex == "Male" & respondent_data$Father_Occupation == "Fa
male_farmers
```

```
##   Respondent  Sex Father_Occupation Num_Siblings_Attending Type_of_House
## 1          1 Male            Farmer                      5          Wood
## 4          4 Male            Farmer                      3          Wood
## 7          7 Male            Farmer                      4      Concrete
## 9          9 Male            Farmer                      1      Concrete
```

```
# g. Select all female respondents with greater than or equal to 5 siblings attending school.
female_five_siblings <- respondent_data[respondent_data$Sex == "Female" & respondent_data$Num_Siblings_A
female_five_siblings
```

```
##   Respondent    Sex Father_Occupation Num_Siblings_Attending Type_of_House
## 3          3 Female            Others                      7 Semi-Concrete
## 5          5 Female            Driver                      5      Concrete
## 8          8 Female            Driver                      6          Wood
```

```
# 2. Write an R program to create an empty data frame
df <- data.frame(
  Ints = integer(),
  Doubles = double(),
  Characters = character(),
  Logicals = logical(),
  Factors = factor(),
  stringsAsFactors = FALSE
)

# Print the structure of the empty data frame
print("Structure of the empty dataframe:")
```

```
## [1] "Structure of the empty dataframe:"
```

```
str(df)
```

```
## 'data.frame':    0 obs. of  5 variables:
##  $ Ints      : int
##  $ Doubles   : num
```

```
##  $ Characters: chr
##  $ Logicals  : logi
##  $ Factors   : Factor w/ 0 levels:
```

```
# a. Describe the results:
# The data frame is empty but predefined to have columns with specific data types: integers, doubles, c


# 3. Create a .csv file of the respondent data frame and save it as "HouseholdData.csv".
write.csv(respondent_data, "HouseholdData.csv", row.names = FALSE)

# a. Import the csv file into the R environment.
imported_data <- read.csv("HouseholdData.csv")
print(imported_data)
```

```
##    Respondent    Sex Father_Occupation Num_Siblings_Attending Type_of_House
## 1           1   Male            Farmer                      5          Wood
## 2           2 Female            Driver                      2      Concrete
## 3           3 Female            Others                      7 Semi-Concrete
## 4           4   Male            Farmer                      3          Wood
## 5           5 Female            Driver                      5      Concrete
## 6           6   Male            Others                      2 Semi-Concrete
## 7           7   Male            Farmer                      4      Concrete
## 8           8 Female            Driver                      6          Wood
## 9           9   Male            Farmer                      1      Concrete
## 10         10 Female            Others                      3          Wood
```

```
# b. Convert the Sex column into factor and change it to integer (Male = 1, Female = 2).
imported_data$Sex <- factor(imported_data$Sex, levels = c("Male", "Female"), labels = c(1, 2))
imported_data$Sex
```

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

```
# c. Convert the Type_of_House into factor and change it into integer (Wood = 1, Concrete = 2, Semi-Con
imported_data$Type_of_House <- factor(imported_data$Type_of_House, levels = c("Wood", "Concrete", "Semi
imported_data$Type_of_House
```

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

```
# d. Factor the Father's occupation (Farmer = 1, Driver = 2, Others = 3).
imported_data$Father_Occupation <- factor(imported_data$Father_Occupation, levels = c("Farmer", "Driver
imported_data$Father_Occupation
```

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

```
# e. Select all female respondents whose father is a driver.
female_driver <- imported_data[imported_data$Sex == 2 & imported_data$Father_Occupation == 2, ]
female_driver
```

```
##   Respondent Sex Father_Occupation Num_Siblings_Attending Type_of_House
## 2          2   2                 2                      2             2
## 5          5   2                 2                      5             2
## 8          8   2                 2                      6             1
```

```r
# f. Select respondents who have greater than or equal to 5 siblings attending school.
siblings_five_or_more <- imported_data[imported_data$Num_Siblings_Attending >= 5, ]
siblings_five_or_more
```

```
##   Respondent Sex Father_Occupation Num_Siblings_Attending Type_of_House
## 1          1   1                 1                      5             1
## 3          3   2                 3                      7             3
## 5          5   2                 2                      5             2
## 8          8   2                 2                      6             1
```

```r
# 4. The interpretation of the graph can be done when the graph is provided in the exercise.


#The graph titled "Sentiments of Tweets Per Day" presents the distribution of tweet sentiments from Jul

# Key Points:
#*July 14, 2020:* The majority of tweets were negative, with around 2500, while neutral and positive tw
#*July 15, 2020:* There was a significant increase in negative tweets, exceeding 4000, while neutral an
#*July 18, 2020:* Negative tweets decreased to approximately 2500, and the neutral and positive tweets
#*July 20, 2020:* The counts of negative and neutral tweets were nearly the same, around 2500, but the
#*July 21, 2020:* Negative tweets remained the highest (about 3500), followed by positive (~3000) and n

# Conclusion:
#The chart shows that negative sentiment tweets were generally more common than neutral and positive on
```