Rworksheet.Sabando#3b.Rmd

Samantha L. Sabando

2025-10-13

```
# Create the dataset
# Creating the dataset based on the given table
#A
data <- data.frame(
  Respondents = 1:20,
  Sex = c(2,2,1,2,2,2,2,2,2,1,1,2,2,2,2,1,2,2,1,2),
  Fathers_Occupation = c(1,3,3,3,1,2,3,1,1,1,3,2,1,3,1,1,3,1,2,1),
  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(6,4,4,1,2,1,5,3,1,2,3,2,5,5,2,1,2,5,3,2),
  Types_of_Houses = c(1,2,3,1,1,3,3,1,2,3,2,3,2,2,2,3,3,3,3,2)
)

# Display the dataset
data
```

```
##    Respondents Sex Fathers_Occupation Persons_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   1                  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                  1               8                  2
## 16          16   1                  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
##    Types_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               2
## 16               3
## 17               3
## 18               3
## 19               3
## 20               2
```

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

```
# Get the summary of the data
#B
summary(data)
```

```
##   Respondents         Sex       Fathers_Occupation Persons_at_Home
##  Min.   : 1.00   Min.   :1.00   Min.   :1.00       Min.   : 3.0
##  1st Qu.: 5.75   1st Qu.:1.75   1st Qu.:1.00       1st Qu.: 5.0
##  Median :10.50   Median :2.00   Median :1.50       Median : 7.0
##  Mean   :10.50   Mean   :1.75   Mean   :1.85       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 Types_of_Houses
##  Min.   :1.00       Min.   :1.00
##  1st Qu.:2.00       1st Qu.:2.00
##  Median :2.50       Median :2.00
##  Mean   :2.95       Mean   :2.25
##  3rd Qu.:4.25       3rd Qu.:3.00
##  Max.   :6.00       Max.   :3.00
```

```
#C mean number
mean_siblings <- mean(data$Siblings_at_School)
is_mean_5 <- mean_siblings == 5
print(is_mean_5)
```

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

```
print(mean_siblings)
```

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

```
#D The 1st two rows and then all the columns using the subsetting functions
subset_data <- data[1:2, ]
```

```r
print(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
```

```r
#E 3rd and 5th row with 2nd and 4th column
subset_data_2 <- data[c(3, 5), c(2, 4)]
print(subset_data_2)
```

```
##   Sex Persons_at_Home
## 3   1               3
## 5   2               5
```

```r
#F the variable types of houses then store the vector that results as types_houses
types_houses <- data$Types_of_Houses
print(types_houses)
```

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

```r
#G only all Males respondent that their father occupation was farmer
male_farmers <- subset(data, Sex == 1 & Fathers_Occupation == 1)
print(male_farmers)
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 10          10   1                  1               4                  2
## 16          16   1                  1               8                  1
##    Types_of_Houses
## 10               3
## 16               3
```

```r
#H only all females respondent that have greater than or equal to 5 number of siblings attending school
female_many_siblings <- subset(data, Sex == 2 & Siblings_at_School >= 5)
print(female_many_siblings)
```

```
##    Respondents Sex Fathers_Occupation Persons_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
##    Types_of_Houses
## 1                1
## 7                3
## 13               2
## 14               2
## 18               3
```

```r
#2 R program to create an empty data frame.
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
```

```
#3 Create a .csv file of this. Save it as HouseholdData.csv
data <- read.csv("HouseholdData.csv")
print(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.
data$Sex <- factor(data$Sex, levels = c("Male", "Female"), labels = c(1, 2))
print(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
data$Types.of.Houses <- factor(data$Types.of.Houses, levels = c("Wood", "Congrete", "Semi-congrete"), la
print(data$Types.of.Houses)
```

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

```r
#D Fathers occupation
data$Fathers.Occupation <- factor(data$Fathers.Occupation, levels = c(1, 2, 3), labels = c("Farmer", "D
print(data$Fathers.Occupation)
```

```
##  [1] Farmer Driver Others Others Farmer Driver Driver Others Farmer Others
## Levels: Farmer Driver Others
```

```r
#E  only all females respondent that has a father whose occupation is driver.
female_drivers <- subset(data, Sex == "2" & Fathers.Occupation == "Driver")
print(female_drivers)
```

```
##   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
```

```r
#F f. Select the respondents that have greater than or equal to 5 number of siblings attending school.
respondents_5_siblings <- subset(data, Siblings.at.School >= 5)
print(respondents_5_siblings)
```

```
##   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
```

```r
# Figure 3:  Interpretation
# The bar chart illustrates the distribution of tweet sentiments per day
# from July 14 to July 21, 2020. Negative sentiments consistently appear
# as the most dominant category, peaking notably on July 15 and July 21.
# Neutral tweets show relatively stable counts throughout the observed period,
# while positive tweets remain lower but display minor increases on certain days
# such as July 17 and July 20. Overall, the figure indicates that Twitter users
# expressed more negative sentiments during this timeframe compared to neutral
# and positive reactions.
```