Untitled

Jeron J. Aurelio

2025-11-24

###1. Create a data frame using the table below ###a. Write the codes. ### 1. Create a data frame using the table below ### a. Write the codes

```
respondents_data <- data.frame(
  Respondents = 1:20,
  Sex = c(2,2,1,2,2,2,2,2,1,2,1,2,2,2,2,2,2,1,2,2),
  Fathers_Occupation = c(1,3,3,2,1,1,3,1,1,2,3,1,2,1,3,3,3,2,1,2),
  Persons_at_Home = c(5,7,3,8,5,9,6,4,8,7,7,4,4,7,8,8,1,11,7,6),
  Siblings_at_School = c(6,4,4,1,2,3,5,1,1,2,3,5,5,2,5,1,2,5,3,2),
  Types_of_Houses = c(1,2,3,1,1,3,3,3,2,2,3,2,2,2,3,3,3,3,3,2)
)
# View the data
print(respondents_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                  2               8                  1
## 5            5   2                  1               5                  2
## 6            6   2                  1               9                  3
## 7            7   2                  3               6                  5
## 8            8   2                  1               4                  1
## 9            9   1                  1               8                  1
## 10          10   2                  2               7                  2
## 11          11   1                  3               7                  3
## 12          12   2                  1               4                  5
## 13          13   2                  2               4                  5
## 14          14   2                  1               7                  2
## 15          15   2                  3               8                  5
## 16          16   2                  3               8                  1
## 17          17   2                  3               1                  2
## 18          18   2                  2              11                  5
## 19          19   1                  1               7                  3
## 20          20   2                  2               6                  2
##    Types_of_Houses
## 1                1
## 2                2
## 3                3
## 4                1
## 5                1
## 6                3
## 7                3
```

```
## 8                3
## 9                2
## 10               2
## 11               3
## 12               2
## 13               2
## 14               2
## 15               3
## 16               3
## 17               3
## 18               3
## 19               3
## 20               2
```

```
# Get summary of data
summary(respondents_data)
```

```
##   Respondents         Sex      Fathers_Occupation Persons_at_Home
##  Min.   : 1.00   Min.   :1.0   Min.   :1.00       Min.   : 1.00
##  1st Qu.: 5.75   1st Qu.:2.0   1st Qu.:1.00       1st Qu.: 4.75
##  Median :10.50   Median :2.0   Median :2.00       Median : 7.00
##  Mean   :10.50   Mean   :1.8   Mean   :1.95       Mean   : 6.25
##  3rd Qu.:15.25   3rd Qu.:2.0   3rd Qu.:3.00       3rd Qu.: 8.00
##  Max.   :20.00   Max.   :2.0   Max.   :3.00       Max.   :11.00
##  Siblings_at_School Types_of_Houses
##  Min.   :1.0        Min.   :1.00
##  1st Qu.:2.0        1st Qu.:2.00
##  Median :3.0        Median :2.50
##  Mean   :3.1        Mean   :2.35
##  3rd Qu.:5.0        3rd Qu.:3.00
##  Max.   :6.0        Max.   :3.00
```

b. Description of the Data

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

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

```
mean_siblings <- mean(respondents_data$Siblings_at_School)
mean_siblings
```

```
## [1] 3.1
```

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

```
respondents_data[1:2, ]
```

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

```
respondents_data[c(3,5),c(2,4) ]
```

```
##   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 <- (respondents_data$Types_of_Houses)
types_houses
```

```
##  [1] 1 2 3 1 1 3 3 3 2 2 3 2 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(respondents_data, Sex == 1 & Fathers_Occupation == 1)
male_farmer
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 9            9   1                  1               8                  1
## 19          19   1                  1               7                  3
##    Types_of_Houses
## 9                2
## 19               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.

```
females_respondent <- subset(respondents_data, Sex == 2 & Siblings_at_School >= 5)
females_respondent
```

```
##    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                  1               4                  5
## 13          13   2                  2               4                  5
## 15          15   2                  3               8                  5
## 18          18   2                  2              11                  5
##    Types_of_Houses
## 1                1
## 7                3
## 12               2
## 13               2
## 15               3
```

```
## 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:")
```

```
## [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.

An empty data frame was created — meaning it has no rows (0 observations) but has 5 columns (variables).

Each column has a specific data type:

Ints → stores integer values (int)

Doubles → stores numeric / double values (num)

Characters → stores text / string values (chr)

Logicals → stores TRUE/FALSE values (logi)

Factors → stores categorical data (with 0 levels because it's still empty)

stringsAsFactors = FALSE

This tells R not to automatically convert character strings into factors (a common behavior in older R versions).

Because the data frame is empty, all columns exist but contain no data yet.

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

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

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

print(household_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              3                  5
```

```
## 5            5   Male                  1               6                  2
## 6            6 Female                  2               4                  3
## 7            7 Female                  2               8                  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         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.

```r
household_data <- read.csv("HouseholdData.csv")

# Convert 'Sex' to numeric (Male = 1, Female = 2)

household_data$Sex <- factor(household_data$Sex,
                             levels = c("Male", "Female"),
                             labels = c(1, 2))
household_data$Sex <- as.integer(as.character(household_data$Sex))
print(household_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               3                  5
## 5            5   1                  1               6                  2
## 6            6   2                  2               4                  3
## 7            7   2                  2               8                  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         Concrete
## 3         Concrete
## 4             Wood
## 5    Semi-concrete
## 6    Semi-concrete
## 7             Wood
## 8    Semi-concrete
## 9    Semi-concrete
## 10        Concrete
```

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.

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

# Convert the factor into integer
household_data$Types_of_Houses <- as.integer(as.character(household_data$Types_of_Houses))

# Display updated data
print(household_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               3                  5
## 5            5   1                  1               6                  2
## 6            6   2                  2               4                  3
## 7            7   2                  2               8                  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?

```
household_data$Fathers_Occupation <- as.numeric(household_data$Fathers_Occupation)
# Display the updated data

print(household_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               3                  5
## 5            5   1                  1               6                  2
## 6            6   2                  2               4                  3
## 7            7   2                  2               8                  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
```

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

```
# Convert Fathers_Occupation into factor with numeric labels
female_father <- subset(household_data, Sex == 2 & Fathers_Occupation == 2)
female_father
```

```
##   Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 2           2   2                  2              7                  3
## 6           6   2                  2              4                  3
## 7           7   2                  2              8                  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.

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

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

4. Interpret the graph

Title: "Sentiments of Tweets Per Day"

The chart shows how many tweets were negative, neutral, or positive on each date from July 14–21, 2020.

Each group of three bars represents a single day:

Red bar → Negative tweets

Yellow bar → Neutral tweets

Blue bar → Positive tweets

The y-axis (vertical) shows the Count of Tweets, while the x-axis (horizontal) shows the Dates.

Step-by-Step Interpretation

General pattern

Every day, there are tweets of all three sentiment types (negative, neutral, positive).

The height of each bar shows how many tweets fall into that sentiment.

Which sentiment dominates?

The red (Negative) bars are often the tallest — especially on July 15 and July 21, showing more negative tweets on those days.

The yellow (Neutral) and blue (Positive) bars vary, but they are generally lower.

Changes across time

On July 14, the number of tweets is moderate, but mostly negative.

On July 15, negative tweets peaked — the highest across all days.

Between July 17–20, tweet counts drop a bit.

On July 21, negative tweets rise again sharply.

Possible interpretation

Sentiment on social media fluctuated daily.

Negative sentiments dominate throughout the observed period.

Positive tweets remain fairly stable.

Neutral tweets fluctuate between 1,500–3,000 counts.