

RWorksheet_Malayas#3b

Andrew Miguel M. Malayas BSIT2A

2024-10-02

1. Create a data frame using the table below.

a. Write the codes.

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

##	Respondents	Sex	Fathers_Occupaton	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	2	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	3	8	2
## 16	16	2	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         3
## 16         3
## 17         3
## 18         3
## 19         3
## 20         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 2 2 ...
## $ Fathers_Occupaton : 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 ...
```

```
summary(data)
```

```
## Respondents Sex Fathers_Occupaton Persons_at_Home
## Min. : 1.00 Min. :1.00 Min. :1.00 Min. : 3.0
## 1st Qu.: 5.75 1st Qu.:2.00 1st Qu.:1.00 1st Qu.: 5.0
## Median :10.50 Median :2.00 Median :2.00 Median : 7.0
## Mean :10.50 Mean :1.85 Mean :1.95 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. :1.0
## 1st Qu.:2.00 1st Qu.:2.0
## Median :2.50 Median :2.5
## Mean :2.95 Mean :2.3
## 3rd Qu.:4.25 3rd Qu.:3.0
## Max. :6.00 Max. :3.0
```

```
data
```

```
## Respondents Sex Fathers_Occupaton 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 2     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     3      8      2
## 16     16 2     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     3
## 16     3
## 17     3
## 18     3
## 19     3
## 20     2
```

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

```
meanSiblings <- mean(data$Siblings_at_School)
meanSiblings
```

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

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

```
subsetdata <- subset(data, Respondents <=2)
subsetdata
```

```
## Respondents Sex Fathers_Occupaton 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.

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

```
##  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.

```
males_farmers <- data[data$Sex == 1 & data$Fathers_Occupaton,]
males_farmers
```

```
##  Respondents Sex Fathers_Occupaton Persons_at_Home Siblings_at_School
## 3           3   1             3             3             4
## 11          11   1             3             7             3
## 19          19   1             2             7             3
##  Types_of_Houses
## 3           3
## 11          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.

```
female_siblings <- data[data$Sex == 2 & data$Siblings_at_School >= 5,]
female_siblings
```

```
##  Respondents Sex Fathers_Occupaton 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
```

2. Write a R program to create an empty data frame. Using the following codes: `df = data.frame(Ids=integer(),Doubles=double(), Characters=character(), Logicals=logical(), Factors=factor(), stringsAsFactors=FALSE)`

`print("Structure of the empty dataframe:")` `print(str(df))` a. Describe the results. - The output shows a data frame with no rows and five columns, where each column has a defined data type.

```
df <- data.frame(
  Ids=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:
## $ Ids      : int
## $ Doubles   : num
## $ Characters: chr
## $ Logicals  : logi
## $ Factors   : Factor w/ 0 levels:
## NULL
```

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

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

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

```
## Types.of.Houses X
## 1      Wood NA
## 2      Congrete NA
## 3      Congrete NA
## 4      Wood NA
## 5      Semi-concrete NA
## 6      Wood NA
## 7      Semi-concrete NA
## 8      Semi-concrete NA
## 9      Congrete NA
```

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

household$Sex <-as.integer(household$Sex)

household
```

```
## 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      7      2      2      4      1
## 7      8      1      3      2      2
## 8      9      2      1     11      6
## 9     10      1      3      6      2
## Types.of.Houses X
## 1      Wood NA
## 2      Congrete NA
## 3      Congrete NA
## 4      Wood NA
## 5      Semi-concrete NA
## 6      Wood NA
## 7      Semi-concrete NA
## 8      Semi-concrete NA
## 9      Congrete NA
```

- 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$Types.of.Houses <- factor(household$Types.of.Houses, levels = c("Wood", "Congrete", "Semi-Congrete"), labels = c(1,2,3))

household$Types.of.Houses <- as.integer(household$Types.of.Houses)

household
```

```
## 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      7  2      2      4      1
## 7      8  1      3      2      2
## 8      9  2      1     11      6
## 9     10  1      3      6      2
##  Types.of.Houses  X
## 1      1 NA
## 2      2 NA
## 3      2 NA
## 4      1 NA
## 5     NA NA
## 6      1 NA
## 7     NA NA
## 8     NA NA
## 9      2 NA
```

- d. On father's occupation, factor it as Farmer = 1; Driver = 2; and Others = 3. What is the R code and its output?

```
household$Fathers.Occupation <- factor(household$Fathers.Occupation, levels = c(1, 2, 3), labels = c("F", "D", "O"))
```

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

```
FemaleFatherDRIVER <- household[household$Sex == 2 & household$Fathers_Occupation == 2,]
FemaleFatherDRIVER
```

```
## [1] Respondents      Sex      Fathers.Occupation Persons.at.Home
## [5] Siblings.at.School Types.of.Houses  X
## <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.

```
FMhousehold <- household[household$Siblings_at_School >= 5,]
FMhousehold
```

```
## [1] Respondents      Sex      Fathers.Occupation Persons.at.Home
## [5] Siblings.at.School Types.of.Houses  X
## <0 rows> (or 0-length row.names)
```

4. Interpret the graph. The graph displays the daily count of tweets categorized by sentiment (negative, neutral, and positive) between July 14 and July 21, 2020. In general, it indicates that negative sentiment was more prevalent than neutral and positive sentiments during this time frame.