RWorksheet_ESTOCE-3b

John Rick Estoce

2025-10-15

1. Create a data frame using the table below.

a. Write the codes.

```
# Create a data frame using the table

Respondents <- 1:20
Sex <- c(2,2,1,2,2,2,2,2,2,2,1,2,2,2,2,2,2,2,1,2)
Fathers_Occupation <- 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 <- data.frame(Respondents, Sex, Fathers_Occupation,
                   Persons_at_Home, Siblings_at_School, Types_of_Houses)

# View dat
print(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   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. Get Structure & Summary

```
# Check structure
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_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 ...
```

```
# Summary of the data
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.: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
```

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

```
mean(data$Siblings_at_School)
```

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

```
#No, the mean number of siblings attending school is not 5.
#The actual average is significantly lower (around 2.95).
```

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

```
# Method 1: Using row numbers
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
```

```
# Method 2: Using head() function
head(data, 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.

```
# The extraction of 3rd and 5th row with 2nd and 4th column
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.

```
# Select the variable Types_of_Houses and store it as a vector
types_houses <- data$Types_of_Houses

# Display the vector
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.

```
# Select only male respondents with father occupation as farmer
male_farmer <- subset(data, Sex == 1 & Fathers_Occupation == 1)

# Display the result
male_farmer
```

```
## [1] Respondents        Sex                Fathers_Occupation Persons_at_Home
## [5] Siblings_at_School Types_of_Houses
## <0 rows> (or 0-length row.names)
```

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.

```
# Select only female respondents with 5 or more siblings attending school
female_5siblings <- subset(data, Sex == 2 & Siblings_at_School >= 5)
```

```r
# Display the result
female_5siblings
```

```
##    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
```

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:") print(str(df)) a. Describe the results.

```r
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:"
```

```r
print(str(df))
```

```
## 'data.frame':    0 obs. of  5 variables:
##  $ Ints      : int 
##  $ Doubles   : num 
##  $ Characters: chr 
##  $ Logicals  : logi 
##  $ Factors   : Factor w/ 0 levels: 
## NULL
```

```r
#The result shows an empty data frame with 5 columns but no data. Each column is already assigned a spe
```

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

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

```r
# 3(a) Import the CSV file into R
HouseholdData <- read.csv("HouseholdData.csv", header = TRUE)

# View the imported data
print(HouseholdData)
```

```
##       Sex Type_of_House Father_Occupation Siblings
## 1   Male           Wood            Farmer        4
## 2 Female      Concrete            Driver       3
## 3   Male          Wood            Others       5
## 4 Female Semi-Concrete            Farmer       6
## 5 Female      Concrete            Driver       2
```

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.

```
# Manually create HouseholdData (example based on image)
HouseholdData <- data.frame(
  Respondents = 1:10,
  Sex = c("Male","Female","Female","Male","Male","Female","Female","Male","Female","Male"),
  Fathers_Occupation = c("Farmer","Driver","Farmer","Farmer","Farmer","Driver","Others","Farmer","Farme
  Persons_at_Home = c(5,7,3,8,6,5,4,3,11,6),
  Siblings_at_School = c(2,3,0,5,2,3,1,2,6,2),
  Types_of_Houses = c("Wood","Concrete","Concrete","Wood","Semi-concrete","Semi-concrete","Wood","Semi-
)

print(HouseholdData)
```

```
##    Respondents    Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1            1   Male             Farmer               5                  2
## 2            2 Female             Driver               7                  3
## 3            3 Female             Farmer               3                  0
## 4            4   Male             Farmer               8                  5
## 5            5   Male             Farmer               6                  2
## 6            6 Female             Driver               5                  3
## 7            7 Female             Others               4                  1
## 8            8   Male             Farmer               3                  2
## 9            9 Female             Farmer              11                  6
## 10          10   Male             Farmer               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
```

Convert Sex column to factor

HouseholdData$Sex < -factor(HouseholdData$Sex, levels = c("Male", "Female"))

Convert factor to integer values (Male = 1, Female = 2)

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

Display the updated dataset

print(HouseholdData)

~~~
c. Convert the Type of Houses into factor and change it into integer. [Legend: Wood = 1; Congrete = 2; S

``` r
# Convert Types of Houses to factor with specific order
HouseholdData$Types_of_Houses <- factor(HouseholdData$Types_of_Houses,
                                        levels = c("Wood", "Concrete", "Semi-concrete"))

# Convert to integer
HouseholdData$Types_of_Houses <- as.integer(HouseholdData$Types_of_Houses)

print(HouseholdData)
~~~

```
##    Respondents    Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1            1   Male             Farmer               5                  2
## 2            2 Female             Driver               7                  3
## 3            3 Female             Farmer               3                  0
## 4            4   Male             Farmer               8                  5
## 5            5   Male             Farmer               6                  2
## 6            6 Female             Driver               5                  3
## 7            7 Female             Others               4                  1
## 8            8   Male             Farmer               3                  2
## 9            9 Female             Farmer              11                  6
## 10          10   Male             Farmer               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?

```r
# Convert Father's Occupation to factor and then integer
HouseholdData$Fathers_Occupation <- factor(HouseholdData$Fathers_Occupation,
                                           levels = c("Farmer", "Driver", "Others"))

HouseholdData$Fathers_Occupation <- as.integer(HouseholdData$Fathers_Occupation)


print(HouseholdData)
```

```
##    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                  1               3                  0
## 4            4   Male                  1               8                  5
## 5            5   Male                  1               6                  2
## 6            6 Female                  2               5                  3
## 7            7 Female                  3               4                  1
```

```
## 8            8   Male                  1               3                  2
## 9            9 Female                  1              11                  6
## 10          10   Male                  1               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 Female respondents with Driver fathers.

```r
# Select females with father as driver
female_driver <- subset(HouseholdData, Sex == 2 & Fathers_Occupation == 2)

# View result
print(female_driver)
```

```
## [1] Respondents        Sex                Fathers_Occupation Persons_at_Home
## [5] Siblings_at_School Types_of_Houses
## <0 rows> (or 0-length row.names)
```

f. Select respondents with 5 siblings attending school.

```r
# Select respondents with 5 or more siblings at school
siblings_5plus <- subset(HouseholdData, Siblings_at_School >= 5)
# View result
print(siblings_5plus)
```

```
##   Respondents    Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 4           4   Male                  1               8                  5
## 9           9 Female                  1              11                  6
##   Types_of_Houses
## 4               1
## 9               3
```

4. Interpret the graph.

This graph shows how people felt on Twitter from July 14 to July 21, 2020. There are three kinds of feelings in the tweets:

Most tweets were Negative. Almost every day, the red bars are the tallest. This means many people were upset, angry, or sad during this time.

Neutral tweets are in the middle. Some people were just sharing facts or opinions without strong emotions.

Positive tweets are the lowest. There were happy or encouraging tweets, but not as many compared to negative ones.

July 15 and July 21 had the highest number of tweets, especially negative ones.

People on Twitter were mostly feeling negative during these days. There were some neutral thoughts and only a few positive messages.