

RWorksheet_Junsay#3b

Syndric James Z. Junsay

2025-10-16

1. Create a data frame using the table below.

a. Write the codes.

```
Respondents <- 1:20
Sex <- c(2,2,1,2,2,2,2,2,2,1,1,2,2,2,2,2,2,1,1,2)
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(5,7,3,8,1,9,6,7,8,4,7,5,4,7,8,8,3,11,5,6)
Siblings_at_School <- c(6,4,4,1,4,1,5,3,1,2,3,2,5,3,2,1,2,5,3,2)
Types_of_Houses <- c(1,2,3,1,1,3,3,1,3,2,2,3,2,3,3,3,3,3,3,2)

# Combine into a data frame
survey_data <- data.frame(Respondents, Sex, Fathers_Occupation, Persons_at_Home,
                           Siblings_at_School, Types_of_Houses)

# Display the data
print(survey_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	1	4
## 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	3
## 15	15	2	3	8	2
## 16	16	2	1	8	1
## 17	17	2	3	3	2
## 18	18	1	1	11	5
## 19	19	1	2	5	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      3
## 10     2
## 11     2
## 12     3
## 13     2
## 14     3
## 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(survey_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 1 9 6 7 8 4 ...
## $ Siblings_at_School: num 6 4 4 1 4 1 5 3 1 2 ...
## $ Types_of_Houses : num 1 2 3 1 1 3 3 1 3 2 ...
```

```
summary(survey_data)
```

```
## Respondents      Sex      Fathers_Occupation Persons_at_Home
## Min.   : 1.00   Min.   :1.00   Min.   :1.00   Min.   : 1.00
## 1st Qu.: 5.75   1st Qu.:1.75   1st Qu.:1.00   1st Qu.: 4.75
## Median :10.50   Median :2.00   Median :2.00   Median : 6.50
## Mean   :10.50   Mean   :1.75   Mean   :1.95   Mean   : 6.10
## 3rd Qu.:15.25   3rd Qu.:2.00   3rd Qu.:3.00   3rd Qu.: 8.00
## Max.   :20.00   Max.   :2.00   Max.   :3.00   Max.   :11.00
## Siblings_at_School Types_of_Houses
## Min.   :1.00     Min.   :1.00
## 1st Qu.:2.00     1st Qu.:2.00
## Median :3.00     Median :3.00
## Mean   :2.95     Mean   :2.35
## 3rd Qu.:4.00     3rd Qu.:3.00
## Max.   :6.00     Max.   :3.00
```

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

```
mean(Siblings_at_School)
```

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

no, the mean number of siblings attending is not 5.

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

```
first_two <- survey_data[1:2, ]  
print(first_two)
```

```
## 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.

```
third_fourth <- survey_data[c(3, 5), c(2,4)]  
print(third_fourth)
```

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

- f. Select the variable types of houses then store the vector that results as types_houses. Write the codes.

```
types_houses <- survey_data$Types_of_Houses  
print(types_houses)
```

```
## [1] 1 2 3 1 1 3 3 1 3 2 2 3 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_farmers <- subset(survey_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  
## 18          18    1              1              11             5  
## Types_of_Houses  
## 10          2  
## 18          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_many_siblings <- subset(survey_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
##      Types_of_Houses
## 1             1
## 7             3
## 13            2
```

2. Write a R program to create an empty data frame. Using the following codes:

a. Describe the results.

```
df = data.frame(Ints=integer(),Doubles=double(), Characters=character(), Logicals=logical(), Factors=factors())
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

```
Respondents <- 1:10
Sex <- c("Male", "Female", "Female", "Male", "Male",
        "Female", "Female", "Male", "Female", "Male")
Fathers_Occupation <- c(1, 2, 3, 3, 1, 2, 2, 3, 1, 3)
Persons_at_Home <- c(5, 7, 3, 8, 6, 4, 4, 2, 11, 6)
Siblings_at_School <- c(2, 3, 0, 5, 2, 3, 1, 2, 6, 2)
Types_of_Houses <- c("Wood", "Congrete", "Congrete", "Wood",
                    "Semi-concrete", "Semi-concrete", "Wood",
                    "Semi-concrete", "Semi-concrete", "Congrete")

# Combine into a data frame
HouseholdData <- data.frame(Respondents, Sex, Fathers_Occupation,
                             Persons_at_Home, Siblings_at_School,
                             Types_of_Houses)

# View the data frame
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 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-concrete
## 6 Semi-concrete
## 7 Wood
## 8 Semi-concrete
## 9 Semi-concrete
## 10 Congrete
```

```
write.csv(HouseholdData, "HouseholdData.csv", row.names = FALSE)
```

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

```
HouseholdData <- read.csv("HouseholdData.csv")
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 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-concrete
## 6 Semi-concrete
## 7 Wood
## 8 Semi-concrete
## 9 Semi-concrete
## 10 Congrete
```

- 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.

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

```
##      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             6   2              2             4           3
## 7             7   2              2             4           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             Congrete
## 3             Congrete
## 4             Wood
## 5             Semi-concrete
## 6             Semi-concrete
## 7             Wood
## 8             Semi-concrete
## 9             Semi-concrete
## 10            Congrete
```

```
str(HouseholdData)
```

```
## 'data.frame':  10 obs. of  6 variables:
## $ Respondents      : int  1 2 3 4 5 6 7 8 9 10
## $ Sex              : Factor w/ 2 levels "1","2": 1 2 2 1 1 2 2 1 2 1
## $ Fathers_Occupation: int  1 2 3 3 1 2 2 3 1 3
## $ Persons_at_Home   : int  5 7 3 8 6 4 4 2 11 6
## $ Siblings_at_School: int  2 3 0 5 2 3 1 2 6 2
## $ Types_of_Houses   : chr  "Wood" "Congrete" "Congrete" "Wood" ...
```

- 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.

```
HouseholdData$Types_of_Houses <- factor(HouseholdData$Types_of_Houses,
                                          levels = c("Wood", "Congrete", "Semi-concrete"),
                                          labels = c(1, 2, 3))
print(HouseholdData)
```

```
##      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      6 2      2      4      3
## 7      7 2      2      4      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
```

```
str(HouseholdData)
```

```
## 'data.frame': 10 obs. of 6 variables:
## $ Respondents : int 1 2 3 4 5 6 7 8 9 10
## $ Sex : Factor w/ 2 levels "1","2": 1 2 2 1 1 2 2 1 2 1
## $ Fathers_Occupation: int 1 2 3 3 1 2 2 3 1 3
## $ Persons_at_Home : int 5 7 3 8 6 4 4 2 11 6
## $ Siblings_at_School: int 2 3 0 5 2 3 1 2 6 2
## $ Types_of_Houses : Factor w/ 3 levels "1","2","3": 1 2 2 1 3 3 1 3 3 2
```

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

```
HouseholdData$Fathers_Occupation <- factor(HouseholdData$Fathers_Occupation,
                                           levels = c(1, 2, 3),
                                           labels = c("Farmer", "Driver", "Others"))
print(HouseholdData)
```

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

```
str(HouseholdData)
```

```
## 'data.frame':  10 obs. of  6 variables:
## $ Respondents      : int  1 2 3 4 5 6 7 8 9 10
## $ Sex              : Factor w/ 2 levels "1","2": 1 2 2 1 1 2 2 1 2 1
## $ Fathers_Occupation: Factor w/ 3 levels "Farmer","Driver",...: 1 2 3 3 1 2 2 3 1 3
## $ Persons_at_Home   : int   5 7 3 8 6 4 4 2 11 6
## $ Siblings_at_School: int   2 3 0 5 2 3 1 2 6 2
## $ Types_of_Houses   : Factor w/ 3 levels "1","2","3": 1 2 2 1 3 3 1 3 3 2
```

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

```
female_driver <- subset(HouseholdData, Sex == 2 & Fathers_Occupation == "Driver")
print(female_driver)
```

```
##   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
```

- f. Select the respondents that have greater than or equal to 5 number of siblings attending school. Write the codes and its output.

```
five_or_more_siblings <- subset(HouseholdData, Siblings_at_School >= 5)
print(five_or_more_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
```

##4. Interpret the graph.


```
library(ggplot2)

sentiments <- data.frame(
  Date = rep(c("July 14, 2020", "July 15, 2020", "July 17, 2020",
              "July 18, 2020", "July 20, 2020", "July 21, 2020"), each = 3),
  Sentiment = rep(c("Negative", "Neutral", "Positive"), times = 6),
  Count = c(2500, 1500, 1800,
            4200, 2800, 3200,
            3300, 2100, 2500,
            3200, 2000, 2600,
            2300, 1500, 2000,
            4100, 2700, 3400)
)

ggplot(sentiments, aes(x = interaction(Date, Sentiment), y = Count, fill = Sentiment)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  labs(title = "Sentiments Of Tweets Per Day",
       x = "Day of Date / Sentiment",
       y = "Count of Tweets") +
  scale_fill_manual(values = c("red", "orange", "blue")) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

