

RWorksheet_Quebral#3b

Myles Andrei Quebral

2024-10-01

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

```
Respondents <- c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20)
Sex <- c(2,2,1,2,2,2,2,2,2,2,1,2,2,2,2,2,2,1,2)
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_frame <- data.frame(Respondents, Sex, Fathers_Occupation, Persons_at_Home, Siblings_at_School, Types_of_houses)
data_frame
```

##	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. Describe the data. Get the structure or the summary of the data

```
str(data_frame)
```

```
## 'data.frame':  20 obs. of  6 variables:
## $ Respondents      : num  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 ...
```

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

```
mean5 <- mean(data_frame[,5])
mean5
```

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

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

```
sub_set <- subset(data_frame[1:2,])
sub_set
```

```
##   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.

```
colrows <- subset(data_frame[c(3,5),c(2,4)])
colrows
```

```
##   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 <- c(Types_of_houses)
```

#g. Select only all Males respondent that their father occupation was farmer. Write the codes and its output.

```
male_farmers <- data_frame[data_frame$Sex == 1 & data_frame$Fathers_Occupation == 1, ]
male_farmers
```

```
## [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.

```
female_respondet <- data_frame[data_frame$Sex == 2 & data_frame$Siblings_at_School >= 5, ]
female_respondet
```

```
##      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.

```
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
```

#The result shows that the data frame has 0 observations (rows) and 5 variables (columns), where each column is specified with a different data type.

#3. Create a .csv file of this. Save it as HouseholdData.csv #a. Import the csv file into the R environment. Write the codes.

```
setwd("/cloud/project")
haus_data <- read.csv("Book1.csv")
haus_data
```

```
## Respondents Sex Father.s.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          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.

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

```
## Respondents Sex Father.s.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          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.

```
haus_data$Types.of.houses <- factor(haus_data$Types.of.houses, levels = c("Wood", "Concrete", "Semi-Concrete"), labels = c(1,2,3))
haus_data
```

```
##      Respondents Sex Father.s.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
```

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

```
haus_data$Father.s.Occupation <- factor(haus_data$Father.s.Occupation, levels = c(1,2,3), labels = c("Farmer", "Driver", "Others"))
haus_data
```

```
##      Respondents Sex Father.s.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
```

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

```
female_dad_driver <- subset(haus_data, Sex == 2 & Father.s.Occupation == "Driver")
female_dad_driver
```

```
##   Respondents Sex Father.s.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.

```
paybsiblings <- subset(haus_data, Respondents & Siblings.at.School >= 5)
paybsiblings
```

```
##   Respondents Sex Father.s.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. #The graphic presents a sentiment analysis of tweets (good, neutral, and negative) for a set of July 2020 dates. Most dates appear to be dominated by negative attitude, with the exception of July 15 and July 23, when sentiment is distributed more evenly or in a mixed manner.