

RMarkdown_Lego3b

2023-10-04

R Markdown

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

```
respo<- c(1:20)
male<- 1
female<- 2
sex<- c(female, female, male, female, female, female, female, female, female, female, male, female, female, female, female, female, female, female, female, female)
farmer<-1
driver<-2
others<-3
occupation<- c (farmer, others, others, others, farmer, driver, others, farmer, farmer, farmer, farmer, others, farmer, others, farmer, farmer, farmer, farmer, farmer, farmer)
person<-c(5,7,3,8,5,9,6,7,8,4,7,5,4,7,8,8,3,11,7,6)
  sibling<-c(6,4,4,1,2,1,5,3,1,2,3,2,5,5,2,1,2,5,3,2)
  wood<-1
  semi_concrete<-2
  concrete<-3
  house<- c(wood,semi_concrete,concrete,wood,wood,concrete,concrete,wood,semi_concrete,concrete,semi_concrete,concrete,semi_concrete,semi_concrete,semi_concrete,semi_concrete,semi_concrete,semi_concrete,semi_concrete,semi_concrete)

num1<-data.frame(
  respondents<- c(respo),
  sex<- c(sex),
  occupations<- c(occupation),
  person<- c(person),
  siblings<- c(sibling),
  type<-(house)
)
colnames(num1)<- c("Respondents", "Sex", "Father Occupation", "Person at Home", "Sibling at School", "Type of House")
num1
```

##	Respondents	Sex	Father	Occupation	Person at Home	Sibling 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 house
## 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.

##c. Is the mean number of siblings attending is 5? No!!

```
mean(siblings)
```

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

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

```
num1[1:2, ]
```

```
##      Respondents Sex Father Occupation Person at Home Sibling at School
## 1      1      2      1      5      6
## 2      2      2      3      7      4
##      Types of house
## 1      1
## 2      2
```

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

```
num1[c(3, 5), c(2, 4)]
```

```
##      Sex Person 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(house)
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.

```
num1_g<- subset(num1, sex == 1 & occupation == 1)
num1_g
```

```
## [1] Respondents      Sex      Father Occupation Person at Home
## [5] Sibling at School Types of house
## <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.

```
num1_h<- subset(num1, sex == 2 & sibling >= 5)
num1_h
```

```
##      Respondents Sex Father Occupation Person at Home Sibling 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 house
## 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
```

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

```
library(readr)
HouseholdData <- read_csv("HouseholdData.csv")
```

```
## Rows: 10 Columns: 6
## -- Column specification -----
## Delimiter: ","
## chr (2): Sex, Types of Houses
## dbl (4): Respondents, Fathers Occupation, Persons at Home, Siblings at School
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
HouseholdData
```

```
## # A tibble: 10 x 6
##   Respondents Sex `Fathers Occupation` `Persons at Home` `Siblings at School`
##   <dbl> <chr> <dbl> <dbl> <dbl>
## 1 1 Male 1 5 2
## 2 2 Fema~ 2 7 3
## 3 3 Fema~ 3 3 0
## 4 4 Male 3 8 5
## 5 5 Male 1 6 2
## 6 6 Fema~ 2 4 3
## 7 7 Fema~ 2 4 1
## 8 8 Male 3 2 2
## 9 9 Fema~ 1 11 6
## 10 10 Male 3 6 2
## # i 1 more variable: `Types of Houses` <chr>
```

##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))
HouseholdData
```

```
## # A tibble: 10 x 6
##   Respondents Sex `Fathers Occupation` `Persons at Home` `Siblings at School`
##   <dbl> <fct> <dbl> <dbl> <dbl>
## 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
## # i 1 more variable: `Types of Houses` <chr>
```

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

Semi-Concrete = 3]. ##Write the R codes and its output.

```
HouseholdData$`Types of Houses` <- factor(HouseholdData$`Types of Houses`, levels = c("Wood", "Semi-concrete", "Semi-Concrete", "Concrete"))
HouseholdData
```

```
## # A tibble: 10 x 6
##   Respondents Sex `Fathers Occupation` `Persons at Home` `Siblings at School`
##   <dbl> <fct>          <dbl>          <dbl>          <dbl>
## 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
## # i 1 more variable: `Types of Houses` <fct>
```

##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"))
HouseholdData
```

```
## # A tibble: 10 x 6
##   Respondents Sex `Fathers Occupation` `Persons at Home` `Siblings at School`
##   <dbl> <fct> <fct>          <dbl>          <dbl>
## 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
## # i 1 more variable: `Types of Houses` <fct>
```

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

```
num3_e <- subset(HouseholdData, Sex == "Female" & `Fathers Occupation` == "Driver")
num3_e
```

```
## # A tibble: 0 x 6
## # i 6 variables: Respondents <dbl>, Sex <fct>, Fathers Occupation <fct>,
## #   Persons at Home <dbl>, Siblings at School <dbl>, Types of Houses <fct>
```

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

```
num3_f <- subset(HouseholdData, `Siblings at School` >= 5)
num3_f
```

```
## # A tibble: 2 x 6
##   Respondents Sex `Fathers Occupation` `Persons at Home` `Siblings at School`
## 1         9 2 Farmer         11          6
```

```
##          <dbl> <fct> <fct>          <dbl>          <dbl>
## 1          4 1      Others          8              5
## 2          9 2      Farmer         11              6
## # i 1 more variable: `Types of Houses` <fct>
```

##4. Interpret the graph. ##Sentiments of Tweets Per Day ##1.July 14, 2020: There were a total of 5,800 tweets. Among these, 2,500 were negative, 1,500 were neutral, and 1,800 were positive.

##2.July 15, 2020: On this day, there were a total of 10,100 tweets. Out of these, 4,200 were negative, 2,800 were neutral, and 3,100 were positive.

##3. July 17, 2020: There were 8,100 tweets in total. Of these, 3,300 were negative, 2,000 were neutral, and 2,800 were positive.

##4.July 18, 2020: On this day, there were 8,100 tweets as well. Out of these, 3,300 were negative, 2,200 were neutral, and 2,600 were positive.

##5.July 20, 2020: There were 5,100 tweets in total. Among these, 2,300 were negative, 1,300 were neutral, and 1,500 were positive.

##6.July 21, 2020: The total number of tweets was 10,300. Out of these, 4,000 were negative, 2,800 were neutral, and 3,500 were positive.

##According to the graph the negative sentiments are most tweets per day, while positive sentiments are in second place and neutral sentiments are always been last.