

Untitled

Zydrick Ceniza

2023-10-04

RWorksheet__Ceniza#3b

Zydrick Ceniza

1. Create a data frame using the table below.

a. Write the codes.

```
res<-c(1:20)
Male<-1
Female<-2
sex<-c(Female,Female,Male,Female,Female,Female,Female,Female,Female,Female,Male,Female,Female,Female,Female,
Farmer<-1
Driver<-2
Others<-3
father<-c(Farmer,Others,Others,Others,Farmer,Driver,Others,Farmer,Farmer,Farmer,Others,Driver,Farmer,
per<-c(5,7,3,8,5,9,6,7,8,4,7,5,4,7,8,8,3,11,7,6)
sib<-c(6,4,4,1,2,1,5,3,1,2,3,2,5,5,2,1,2,5,3,2)
Wood<-1
SemiConcrete<-2
Concrete<-3
th<-c(Wood,SemiConcrete,Concrete,Wood,Wood,Concrete,Concrete,Wood,SemiConcrete,Concrete,SemiConcrete,
num1a<-{
  data.frame(
    q=c(res),
    w=c(sex),
    e=c(father),
    r=c(per),
    t=c(sib),
    y=c(th)
  ) }
colnames(num1a)<-c("Respondents","Sex","Fathers Occupation","Persons at Home","Siblings at School","Type of House")
num1a
```

##	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 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

```
str(num1a)

## '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 house    : num  1 2 3 1 1 3 3 1 2 3 ...
```

The Structure contains 6 variables that contains 20 objects per variables

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

```
mean(num1a$`Siblings at School`)
```

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

No the mean is 2.95

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

its output.

```
num1a[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 house  
## 1 1  
## 2 2
```

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

```
e35<-c(3,5)  
e24<-c(2,4)  
num1a[e35,e24]
```

```
## 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("Wood","Semi-Concrete","Concrete")  
types_houses
```

```
## [1] "Wood" "Semi-Concrete" "Concrete"
```

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

```
g<-subset(num1a, sex==Male & `Fathers Occupation`==Farmer)  
g
```

```
## [1] Respondents Sex Fathers Occupation Persons at Home  
## [5] Siblings 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.

```
subset_result <- subset(num1a, Sex == 2 & `Siblings at School` >= 5, )  
subset_result
```

```
## 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 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))
```

```
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
```

- a. Describe the results. there are no objects in 5 variables

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 Home
##
## 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 Home`
##   <dbl> <chr>          <dbl>          <dbl>          <dbl>
## 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            3              4              1
## 8         8 Male              3              2              2
## 9         9 Female            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 Home`
##   <dbl> <fct>      <dbl>      <dbl>      <dbl>
## 1      1 1 1      1      5      2
## 2      2 2 2      2      7      3
## 3      3 2 2      3      3      0
## 4      4 1 3      3      8      5
## 5      5 1 1      1      6      2
## 6      6 2 2      2      4      3
## 7      7 2 3      3      4      1
## 8      8 1 3      2      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-Congrete = 3]. Write the R codes and its output.

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

```
HouseholdData
```

```
## # A tibble: 10 x 6
##   Respondents Sex `Fathers Occupation` `Persons at Home` `Siblings at Home`
##   <dbl> <fct>      <dbl>      <dbl>      <dbl>
## 1      1 1 1      1      5      2
## 2      2 2 2      2      7      3
## 3      3 2 3      3      3      0
## 4      4 1 3      3      8      5
## 5      5 1 1      1      6      2
## 6      6 2 2      2      4      3
## 7      7 2 3      3      4      1
## 8      8 1 3      2      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 Home`
##   <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      Others      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.

```
num3e<-subset(HouseholdData, Sex == 2 & `Fathers Occupation` == "Driver")
num3e
```

```
## # A tibble: 2 x 6
##   Respondents Sex `Fathers Occupation` `Persons at Home` `Siblings at Home`
##         <dbl> <fct> <fct>           <dbl>           <dbl>
## 1         2 2      Driver             7             3
## 2         6 2      Driver             4             3
## # i 1 more variable: `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.

```
num3f<-subset(HouseholdData, `Siblings at Home`>=5 )
num3f
```

```
## # A tibble: 2 x 6
##   Respondents Sex `Fathers Occupation` `Persons at Home` `Siblings at Home`
##         <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.

In the Sentiments of Tweets Per Day the Negative Tweets is always higher from July 14,15,17,18,20,21 in year 2020 and the Positive Tweets is in the Second place while the last place is the Neutral.

In July 14, 2020: There were a total of 5,800 tweets. Among these, 2,450 were negative, 1,500 were neutral, and 1,750 were positive.

In July 15, 2020: On this day, there were a total of 10,100 tweets. Out of these, 4,200 were negative, 2,750 were neutral, and 3,200 were positive.

In July 17, 2020: There were 8,100 tweets in total. Of these, 3,250 were negative, 1,900 were neutral, and 2,400 were positive.

In July 18, 2020: On this day, there were 8,100 tweets as well. Out of these, 3,250 were negative, 2,050 were neutral, and 2,500 were positive.

In July 20, 2020: There were 5,100 tweets in total. Among these, 2,400 were negative, 1,450 were neutral, and 1,600 were positive.

In July 21, 2020: The total number of tweets was 10,300. Out of these, 4,000 were negative, 2,700 were neutral, and 3,400 were positive.

Therefore, July 15, 2020 is the highest negative tweets, In July 21, 2020 is the highest positive tweets, while the highest neutral tweets is in July 25, 2020