

RWorksheet_Canonicato#3b

Dianah Marie Canonicato

2023-10-11

#Worksheet-3b in R

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

#a. Write the codes.

```
respondents <- c(1:20)
sex <- c(2,2,1,2,2,2,2,2,2,1,2,2,2,2,2,1,2)
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)
```

```
Household_Data <- data.frame(respondents,sex,fathers_occupation,persons_at_home,siblings_at_school,types_of_houses)
Household_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. Describe the data. Get the structure or the summary of the data

```
str(Household_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(Household_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?

```
siblings_mean <- mean(siblings_at_school)
siblings_mean
```

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

the mean number of siblings attending school is not 5 but 2.95

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

```
firstTwo_rows <- head(Household_Data, 2)
firstTwo_rows
```

```
## 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
```

#Output:

```
# respondents sex fathers_occupation persons_at_home siblings_at_school
#          1  2              1              5              6
#          2  2              3              7              4
# types_of_houses
#          1
#          2
```

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

```
extract_rows_column <- Household_Data[c(3,5),c(2,4)]
extract_rows_column
```

```
## 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 c

```
types_houses <- Household_Data$types_of_houses
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 o

```
male_farmer= Household_Data[Household_Data$sex == 1 & Household_Data$fathers_occupation == 1,]
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

```
females_siblings_school <- Household_Data[Household_Data$siblings_at_school >= 5,]
females_siblings_school
```

```
## 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:")
```

```
## [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.

#This output shows that the dataframe `df` has 0 observations (rows) and 6 variables (columns). The variables are empty. Since the dataframe is empty, all the variables are also empty. This means that the `Ints`, `Doubles`,

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

```
New_Respondents <- c(1:10)
New_Sex <- c("Male", "Female", "Female", "Male", "Male", "Female", "Female", "Male", "Female", "Male")
New_Fathers_Occupation <- c(1,2,3,3,1,2,2,3,1,3)
New_Persons_At_Home <- c(5,7,3,8,6,4,4,2,11,6)
New_SiblingsAt_School <- c(2,3,0,5,2,3,1,2,6,2)
New_Types_Houses <- c("Wood", "Congrete", "Congrete", "Wood", "Semi-concrete", "Semi-concrete", "Wood", "Congrete", "Congrete", "Wood")

Household_Data <- data.frame( Respondents=New_Respondents, Sex=New_Sex, FatherOccupation=New_Fathers_Occupation,
                             Persons_At_Home=New_Persons_At_Home, Siblings_At_School=New_Siblings_At_School,
                             Types_Of_Houses=New_Types_Houses)

write.csv(Household_Data, "HouseholdData.csv")
```

```
##      Respondents    Sex FatherOccupation Persons_At_Home Siblings_AtSchool
## 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(Household_Data, file ="Household_Data.csv")
```

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

```
imported <- read.csv("Household_Data.csv")
imported
```

```
##      X Respondents      Sex FatherOccupation Persons_At_Home Siblings_AtSchool
## 1    1              1  Male                1                5                2
## 2    2              2 Female                2                7                3
## 3    3              3 Female                3                3                0
## 4    4              4  Male                3                8                5
## 5    5              5  Male                1                6                2
## 6    6              6 Female                2                4                3
## 7    7              7 Female                2                4                1
## 8    8              8  Male                3                2                2
## 9    9              9 Female                1               11                6
## 10  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]

```
imported$Sex <- factor(imported$Sex, levels = c("Male", "Female"))
imported$Sex<- as.integer(imported$Sex)
imported$Sex
```

```
##      [1] 1 2 2 1 1 2 2 1 2 1
```

#Output: 1 2 2 1 1 2 2 1 2 1

#c. Convert the Type of Houses into factor and change it into integer. [Legend: Wood = 1; Congrete = 2; Semi-concrete = 3]

```
imported$New_Types_Houses <- factor(imported$Types_Of_Houses, levels = c("Wood", "Congrete", "Semi-concrete"))
imported$Types_Of_Houses <- as.integer(imported$Types_Of_Houses)
```

```
## Warning: NAs introduced by coercion
```

```
imported$Types_Of_Houses
```

```
## [1] NA NA NA NA NA NA NA NA NA NA
```

```
#Output: 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
```

```
imported$FatherOccupation <- factor(imported$FatherOccupation, levels = c(1,2,3), labels = c("Farmer",  
imported$FatherOccupation
```

```
## [1] Farmer Driver Others Others Farmer Driver Driver Others Farmer Others
```

```
## Levels: Farmer Driver Others
```

```
#Output: Farmer Driver Others Others Farmer Driver Driver Others Farmer Others
```

```
# e. Select only all females respondent that has a father whose occupation is driver. Write the codes a  
imported
```

```
##      X Respondents Sex FatherOccupation Persons_At_Home Siblings_AtSchool  
## 1    1            1   1           Farmer              5              2  
## 2    2            2   2           Driver              7              3  
## 3    3            3   2           Others              3              0  
## 4    4            4   1           Others              8              5  
## 5    5            5   1           Farmer              6              2  
## 6    6            6   2           Driver              4              3  
## 7    7            7   2           Driver              4              1  
## 8    8            8   1           Others              2              2  
## 9    9            9   2           Farmer             11              6  
## 10  10           10   1           Others              6              2
```

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

```
femaleDriver <- imported[imported$Sex == 2 & imported$FatherOccupation == "Driver" , c(3,4)]  
femaleDriver
```

```
##      Sex FatherOccupation  
## 2    2           Driver  
## 6    2           Driver  
## 7    2           Driver
```

```
# f. Select the respondents that have greater than or equal to 5 number of siblings attending school. W  
upper_five <- imported[imported$SiblingsAtSchool >= 5,]  
upper_five
```

```
## [1] X Respondents Sex FatherOccupation  
## [5] Persons_At_Home Siblings_AtSchool Types_Of_Houses New_Types_Houses  
## <0 rows> (or 0-length row.names)
```

#4. Interpret the graph.

#The graph shows the sentiment of tweets per day for the period of July 14, 2020 to July 21, 2020.

#The sentiment is divided into three categories: negative, neutral, and positive.

#On this day, July 14, the negative sentiment was the most prevalent among the other sentiments. This s

#On the following day, July 15, the negative sentiment remained at its peak, even as all the sentiments

#Over the next few days, from July 17 to July 18, the negative sentiment remained consistently high, wh

#On the day, July 20, all sentiments reached their lowest point, even though there were still more nega

#On this day, July 21, all sentiments increased, with the negative sentiment still being the most promi

#In conclusion, the graphs provide valuable insights into the sentiment analysis of tweets per day, all