

RWorksheet_Animas#3b

Marvin Anthony Y. Animas

2024-10-03

```
#1 Create the data frame
```

```
#a Write the codes.
```

```
df <- data.frame(  
  Respondents = 1:20,  
  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))  
df
```

##	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
#The data surveys a total of 20 respondents,it shows data like sex, Fathers occupation, Persons at home

#c. Is the mean number of siblings attending is 5?
mean_of_siblings<-mean(df[,5])
mean_of_siblings
```

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

```
#The mean number of siblings is not 5 the mean number is 2.95

#d. Extract the 1st two rows and then all the columns using the subsetting functions. Write the codes a
first_two<-df[1:2,]
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_Fifth<-df[c(3,5), c(2,4)]
Third_Fifth
```

```
##   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<-df[,6]
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
Males<-subset(df, df[,2] == 1 & df[,3] == 1 )
Males
```

```
## [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
Female<- subset(df, df[,2] == 2 & df[,5] >= 5)
Female
```

```
##      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())
```

```
#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 code shows an empty data frame named df in R with 5 variables (Ints, Doubles, Characters, Logicals)

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

```
respondents <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
sex <- c("Male", "Female", "Female", "Male", "Male", "Female", "Female", "Male", "Female", "Male")
father <- c(1, 2, 3, 3, 1, 2, 2, 3, 1, 3)
persons <- c(5, 7, 3, 8, 6, 4, 4, 2, 11, 6)
siblings <- c(2, 3, 0, 5, 2, 3, 1, 2, 6, 2)
houses <- c("Wood", "Congrete", "Congrete", "Wood", "Semi-concrete", "Semi-concrete", "Wood", "Semi-concrete", "Semi-concrete", "Congrete")

HOUSEdata <- data.frame(Respondents = respondents, Sex = sex, Fathers_Occupation = father, Persons_at_Home = persons, Siblings_at_School = siblings, Types_of_Houses = houses)
HOUSEdata
```

```
##      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(HOUSEdata, file = "HouseholdData.csv", FALSE)
```

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

```
Imported <- read.csv("HouseholdData.csv")
Imported
```

```
##      X Respondents      Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 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"), labels = c(1, 2))
Imported
```

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

```
#c. Convert the Type of Houses into factor and change it into integer. [Legend: Wood= 1; Concrete = 2; Semi-concrete = 3]
Imported$Types_of_Houses <- factor(Imported$Types_of_Houses, levels = c("Wood", "Congrete", "Semi-concrete"), labels = c(1, 2, 3))
Imported
```

```
##      X Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1      1      1      1      1      5      2
## 2      2      2      2      2      7      3
## 3      3      3      2      3      3      0
## 4      4      4      1      3      8      5
## 5      5      5      1      1      6      2
## 6      6      6      2      2      4      3
## 7      7      7      2      2      4      1
```

```
## 8      8      8      1      3      2      2
## 9      9      9      2      1      11      6
## 10    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

```
Imported$Fathers_Occupation <- factor(Imported$Fathers_Occupation, levels = c("Farmer", "Driver", "Other"))
Imported
```

```
##      X Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1      1      1      1      <NA>      5      2
## 2      2      2      2      <NA>      7      3
## 3      3      3      2      <NA>      3      0
## 4      4      4      1      <NA>      8      5
## 5      5      5      1      <NA>      6      2
## 6      6      6      2      <NA>      4      3
## 7      7      7      2      <NA>      4      1
## 8      8      8      1      <NA>      2      2
## 9      9      9      2      <NA>     11      6
## 10    10     10      1      <NA>      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

```
Female_Drivers <- subset(Imported, Sex == 2 & Fathers_Occupation == 2)
Female_Drivers
```

```
## [1] X Respondents Sex Fathers_Occupation
## [5] Persons_at_Home Siblings_at_School Types_of_Houses
## <0 rows> (or 0-length row.names)
```

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

```
Siblings_Up5 <- subset(Imported, Siblings_at_School >= 5)
Siblings_Up5
```

```
##      X Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 4 4          4      1          <NA>          8          5
## 9 9          9      2          <NA>         11          6
##      Types_of_Houses
## 4          1
## 9          3
```

#4 Interpret the graph.

#This graph displays the sentiment analysis of tweets over several days in July 2020.

#It shows the count of tweets categorized as negative, neutral, and positive for each day.

#The highest peak for negative sentiment occurs on July 15th. There are relative peaks for Positive comm