RWorksheet-3_Amuan#3B

Marivic Amuan

2023-10-16

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

```
respondent <- c(1:20)
sex <- c(2,2,1,2,2,2,2,2,2,2,1,2,2,2,2,2,2,2,1,2)
fathersOccupution <- c(1,3,3,3,1,2,3,1,1,1,3,2,1,3,3,1,3,1,2,1)
personsAtHome <- c(5,7,3,8,5,9,6,7,8,4,7,5,4,7,8,8,3,11,7,6)
siblingsAtSchool <- c(6,4,4,1,2,1,5,3,1,2,3,2,5,5,2,1,2,5,3,2)
typeOfHouse <- c(1,2,3,1,1,3,3,1,2,3,2,3,2,2,3,3,3,3,3,2)


householdData <- data.frame(

  Respondents = respondent,
  Sex = sex,
  FathersOccupation = fathersOccupution,
  PersonAtHome = personsAtHome,
  SiblingsAtSchool = siblingsAtSchool,
  HouseType = typeOfHouse

)

householdData
```

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

```
str(householdData)
summary(householdData)
```

#The data frame consists of 20 observations and 6 variables. #EXPLAINATIONS:

#Respondents: A variable is type of integer and there are 20 respondents in dataset. #Sex: A variable is type of numeric with values 1 and 2 represent the gender of male and female. #FathersOccupation: A variable is type of numeric with values 1, 2, and 3 , maybe it represent the level of occupation. #PersonAtHome: A variable is type of numeric, and it represent the number of people at home for each respondent. #SiblingsAtSchool: A variable is type of numeric and represent the number of siblings that the respondent has at school. #HouseType: A variable is numeric with values 1, 2, and 3, it represent kind of houses.

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

```
siblingsMean <- mean(householdData$SiblingsAtSchool)

siblingsMean
```

#EXPLANATION: No, The mean number of siblings attending is 2.95.

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

```
firstTwoRows <- householdData[1:2,]
firstTwoRows
```

#OUTPUT:

#Respondents Sex FathersOccupation PersonAtHome SiblingsAtSchool HouseType #1 1 2 1 5 6 1 #2 2 2 3 7 4 2

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

```
thirdAndFifthRows <- householdData[c(3,5),c(2,4)]
thirdAndFifthRows
```

#OUTPUT #3 1 3 #5 2 5

#1 F.Select the variable types of houses then store the vector that results as types_houses. #Write the codes.

```
typesHouses <- householdData$HouseType
typesHouses
```

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

```
householdData[householdData$Sex == 1 & householdData$FathersOccupation == "farmer", ]
householdData$FathersOccupation
```

#OUPUT <0 rows> (or 0-length row.names)

#1 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 <- householdData[householdData$SiblingsAtSchool >= 5,]
female
```

#OUTPUT: There are five observations #Respondents Sex FathersOccupation PersonAtHome SiblingsAtSchool HouseType #1 1 2 1 5 6 1 #7 7 2 3 6 5 3 #13 13 2 1 4 5 2 #14 14 2 3 7 5 2 #18 18 2 1 11 5 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))
```

#2.A EXPLANATION:

#Data frame is empty, this part of the output indicates that you have a data frame with 0 observations (rows) and 5 variables (columns). # Ints,column integer data type. # Doubles, column is numeric (double) data type. # Characters, column is character data type. # Logicals, column is logical (boolean) data type. # Factors, column a factor variable with 0 levels, currently has no unique levels.

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

```
otherRespondent <- c(1:10)
otherSex <- c("Male", "Female", "Female", "Male", "Male", "Female", "Female", "Male", "Female", "Male
otherFathersOccupution <- c(1,2,3,3,1,2,2,3,1,3)
```

```
  otherPersonsAtHome <- c(5,7,3,8,6,4,4,2,11,6)
  otherSiblingsAtSchool <- c(2,3,0,5,2,3,1,2,6,2)
  otherTypeOfHouse <- c("Wood", "Congrete", "Congrete", "Wood", "Semi-congrete", "Semi-congrete", "Wood

  householdData <- data.frame(
      Respondents = otherRespondent,
      Sex = otherSex,
      FatherOccupation = otherFathersOccupution,
      PersonAtHome =otherPersonsAtHome,
      SiblingsAtSchool = otherSiblingsAtSchool,
      HouseType = otherTypeOfHouse
)

  write.csv(householdData, file = "householdData.csv")
```

#3 A. Import the csv file into the R environment. Write the codes.

```
  householdData <- read.csv("householdData.csv")
  head(householdData)
```

#3 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"))
  householdData$Sex <- as.integer(householdData$Sex)

  householdData$Sex
```

#OUTPUT 1 2 2 1 1 2 2 1 2 1

#3 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$HouseType <- factor(householdData$HouseType, levels = c("Wood", "Congrete", "Semi-congr
  householdData$HouseType <- as.integer(householdData$HouseType)

  householdData$HouseType
```

#OUTPUT 1 2 2 1 3 3 1 3 3 2

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

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

#OUTPUT:

#Farmer, Driver, Others, Others, Farmer, Driver, Driver, Others, Farmer, Others, #Levels: Farmer, Driver, Others,

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

```
  female <- householdData[householdData$Sex == 2 & householdData$FatherOccupation == "Driver",]
  female
```

#OUTPUT:

#[1] X Respondents Sex FatherOccupation PersonAtHome
#[6] SiblingsAtSchool HouseType
#<0 rows> (or 0-length row.names)

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

```
five <- householdData[householdData$SiblingsAtSchool >= 5,]
five
```

#OUTPUT:

#X Respondents Sex FatherOccupation PersonAtHome SiblingsAtSchool HouseType #4 4 4 1 Others, 8 5 1 #9 9 9 2 Farmer, 11 6 3

#4. Interpret the graph.

#2022 #On July 14, the sentiments stated in tweets were as follows: 2500 unfavorable, 1500 neutral, and a range of 1500 to 2000 positive. #On July 15, the number of negative tweets climbed from 2500 to 4000 compared to the previous day. The number of neutral tweets increased from 2500 to 3000, while the number of positive tweets increased from 3000 to 3500. #On July 17, the number of negative sentiments fell to between 3000 and 3500, while neutral sentiments fell to approximately 2000 and good sentiments fell to nearly 2500. #On July 18, the number of unfavorable attitudes remained unchanged from the previous day, with a range of 3000 to 3500. The amount of neutral sentiments climbed to 2000, while positive sentiments increased to 2500. #On July 20, the quantity of negative attitudes in tweets fell from the previous day to over 2500. Positive sentiment declined to a range between 1500 and 2000, while neutral sentiment decreased to about 1500. #On July 21, the number of tweets with negative feelings increased from the previous day, reaching over 4100. Positive sentiment climbed from the previous day, almost reaching 3500, while neutral sentiment increased from the previous day, ranging between 2500 and 3000.

#The particular reasons for these shifts in mood could only be explained with additional context and study of the tweets.