RWorksheet_Caneso#3b

Rovel Jan Caneso

2024-10-02

Number 1

```
#[1.] Create a data frame using the table below.
#     a. Write the codes.
akongData <- data.frame(
  Respondents = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20),
  Sex = c(2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 1, 2, 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)
)
#     b. Describe the data. Get the structure or the summary of the data
str(akongData)
```

```
## 'data.frame':    20 obs. of  6 variables:
##  $ Respondents       : num  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 ...
```

```
#     c. Is the mean number of siblings attending is 5?
sm <- mean(akongData$Siblings_at_School)
sm == 5
```

```
## [1] FALSE
```

```
#     d. Extract the 1st two rows and then all the columns using the subsetting functions.
#        Write the codes and its output.
akongData[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_Houses
## 1               1
## 2               2
```

```
#     e. Extract 3rd and 5th row with 2nd and 4th column. Write the codes and its result.
akongData[c(3,5),c(2,4)]
```

```
##   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 <- akongData$Types_of_Houses
#     g. Select only all Males respondent that their father occupation was farmer. Write the
#        codes and its output.
subset(akongData, Sex == 1 & Fathers_Occupation == 1)
```

```
## [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. Write the codes and its outputs.
subset(akongData[,c(2,5)], Sex == 2 & Siblings_at_School >= 5)
```

```
##    Sex Siblings_at_School
## 1    2                  6
## 7    2                  5
## 13   2                  5
## 14   2                  5
## 18   2                  5
```

Number 2

```
#[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.
# [It had the output of the data frames structure or summary.]
```

Number 3

```
#[3.] Create a .csv file of this. Save it as HouseholdData.csv
#     a. Import the csv file into the R environment. Write the codes.
houseData <- read.csv("HouseholdData.csv")
#     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.
houseData$Sex <- as.numeric(factor(houseData$Sex, levels = c("Male", "Female")))
houseData
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 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
##    Types_of_Houses
## 1             Wood
## 2         Congrete
## 3         Congrete
## 4             Wood
## 5    Semi-Congrete
## 6    Semi-Congrete
## 7             Wood
## 8    Semi-Congrete
## 9    Semi-Congrete
## 10        Congrete
```

```
#     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.
houseData$Types_of_Houses <- as.numeric(factor(houseData$Types_of_Houses, levels = c("Wood", "Congrete"
houseData$Types_of_Houses
```

```
##  [1] 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 its output?
houseData$Fathers_Occupation <- as.character(factor(houseData$Fathers_Occupation, labels = c("Farmer",
houseData$Fathers_Occupation
```

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

```r
#     e. Select only all females respondent that has a father whose occupation is driver. Write
#        the codes and its output.
subset(houseData[,c(2:3)], Sex == 2 & Fathers_Occupation == "Driver")
```

```
##   Sex Fathers_Occupation
## 2   2             Driver
## 6   2             Driver
## 7   2             Driver
```

```r
#     f. Select the respondents that have greater than or equal to 5 number of siblings attending
#        school. Write the codes and its output.
subset(houseData[,c(1,5)], Siblings_at_School >= 5)
```

```
##   Respondents Siblings_at_School
## 4           4                 5
## 9           9                 6
```

Number 4

```
## [1] "A graph of people's mood by their tweets between July 14 and July 21, 2020, can be depicted"
## [1] "as three types of emotions: negative, neutral, and positive. From the graph shown above, a"
## [1] "high quantity of negative tweets can be observed on these days with an unevenly larger number"
## [1] "on July 15 and 21. There is a smaller amount of positive tweets, which can be noticed on some"
## [1] "days, with a constant neutral number throughout the week. By utilizing the ggplot2 of the R"
## [1] "programming, this graph can be employed for text data classification, determination of"
## [1] "frequency of positive, negative, and neutral sentiment, and identification of trends in"
## [1] "large datasets."
```