Rworksheet_Montealto#3b

Pierre Vincent S. Montealto Jr.

2024-10-11

1. Create a data frame using the table below.

a. Write the codes.

```r
Data <- data.frame(
  Respondents        = 1:20,
  Sex                = c(2, 1, 2, 2, 1, 1, 2, 2, 2, 1, 1, 2, 2, 1, 2, 2, 1, 2, 1, 2),
  FatherOccupation   = c(1, 2, 3, 1, 2, 1, 3, 2, 3, 3, 1, 3, 2, 1, 3, 1, 3, 3, 1, 1),
  PersonsAtHome      = c(5, 7, 3, 5, 5, 3, 6, 6, 7, 7, 3, 7, 4, 7, 8, 8, 3, 11, 8, 6),
  SiblingsAtSchool   = c(6, 4, 3, 2, 3, 3, 5, 5, 4, 5, 3, 7, 5, 2, 1, 3, 1, 5, 3, 2),
  TypesOfHouses      = c(1, 2, 3, 1, 3, 1, 3, 3, 3, 1, 3, 3, 3, 1, 3, 3, 3, 3, 3, 2)
)

Data
```

```
##    Respondents Sex FatherOccupation PersonsAtHome SiblingsAtSchool
## 1            1   2                1             5                6
## 2            2   1                2             7                4
## 3            3   2                3             3                3
## 4            4   2                1             5                2
## 5            5   1                2             5                3
## 6            6   1                1             3                3
## 7            7   2                3             6                5
## 8            8   2                2             6                5
## 9            9   2                3             7                4
## 10          10   1                3             7                5
## 11          11   1                1             3                3
## 12          12   2                3             7                7
## 13          13   2                2             4                5
## 14          14   1                1             7                2
## 15          15   2                3             8                1
## 16          16   2                1             8                3
## 17          17   1                3             3                1
## 18          18   2                3            11                5
## 19          19   1                1             8                3
## 20          20   2                1             6                2
##    TypesOfHouses
## 1              1
## 2              2
## 3              3
## 4              1
## 5              3
## 6              1
## 7              3
```

```
## 8               3
## 9               3
## 10              1
## 11              3
## 12              3
## 13              3
## 14              1
## 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

```r
str(Data)
```

```
## 'data.frame':    20 obs. of  6 variables:
##  $ Respondents    : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Sex            : num  2 1 2 2 1 1 2 2 2 1 ...
##  $ FatherOccupation: num  1 2 3 1 2 1 3 2 3 3 ...
##  $ PersonsAtHome  : num  5 7 3 5 5 3 6 6 7 7 ...
##  $ SiblingsAtSchool: num  6 4 3 2 3 3 5 5 4 5 ...
##  $ TypesOfHouses  : num  1 2 3 1 3 1 3 3 3 1 ...
```

```r
summary(Data)
```

```
##   Respondents         Sex       FatherOccupation PersonsAtHome
##  Min.   : 1.00   Min.   :1.0   Min.   :1         Min.   : 3.00
##  1st Qu.: 5.75   1st Qu.:1.0   1st Qu.:1         1st Qu.: 4.75
##  Median :10.50   Median :2.0   Median :2         Median : 6.00
##  Mean   :10.50   Mean   :1.6   Mean   :2         Mean   : 5.95
##  3rd Qu.:15.25   3rd Qu.:2.0   3rd Qu.:3         3rd Qu.: 7.00
##  Max.   :20.00   Max.   :2.0   Max.   :3         Max.   :11.00
##  SiblingsAtSchool TypesOfHouses
##  Min.   :1.00     Min.   :1.00
##  1st Qu.:2.75     1st Qu.:1.75
##  Median :3.00     Median :3.00
##  Mean   :3.60     Mean   :2.40
##  3rd Qu.:5.00     3rd Qu.:3.00
##  Max.   :7.00     Max.   :3.00
```

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

```r
mean(Data$SiblingsAtSchool)
```

```
## [1] 3.6
```

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

```r
Data[1:2, ]
```

```
##   Respondents Sex FatherOccupation PersonsAtHome SiblingsAtSchool TypesOfHouses
## 1           1   2                1             5                6             1
## 2           2   1                2             7                4             2
```

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

```r
Data[c(3, 5), c(2, 4)]
```

```
##   Sex PersonsAtHome
## 3   2             3
## 5   1             5
```

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

```r
types_houses <- Data$TypesOfHouses
types_houses
```

```
##  [1] 1 2 3 1 3 1 3 3 3 1 3 3 3 1 3 3 3 3 3 2
```

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

```r
MalesRespo <- subset(Data, Sex == 1 & FatherOccupation == 1)
MalesRespo
```

```
##     Respondents Sex FatherOccupation PersonsAtHome SiblingsAtSchool
## 6             6   1                1             3                3
## 11           11   1                1             3                3
## 14           14   1                1             7                2
## 19           19   1                1             8                3
##     TypesOfHouses
## 6               1
## 11              3
## 14              1
## 19              3
```

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.

```r
FemalesRespo <- subset(Data, Sex == 2 & SiblingsAtSchool >= 5)
FemalesRespo
```

```
##     Respondents Sex FatherOccupation PersonsAtHome SiblingsAtSchool
## 1             1   2                1             5                6
## 7             7   2                3             6                5
## 8             8   2                2             6                5
## 12           12   2                3             7                7
## 13           13   2                2             4                5
## 18           18   2                3            11                5
##     TypesOfHouses
## 1               1
## 7               3
## 8               3
## 12              3
## 13              3
## 18              3
```

2. Write a R program to create an empty data frame. Using the following codes:

```r
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.

-The output clearly indicates that It have successfully created an empty dataframe with specified column types but no data.

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

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

```
filedata <- 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.

```
filedata$Sex <- factor(filedata$Sex, levels = c("Male", "Female"), labels = c(1,2))
filedata
```

```
##    Respondents Sex FathersOccupation PersonsAtHome SiblingsAtSchool
## 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
##     TypesOfHouses
## 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.

```r
filedata$TypesOfHouses <- factor(filedata$TypesOfHouses, levels = c("Wood", "Congrete", "Semi-congrete")
filedata
```

```
##    Respondents Sex FathersOccupation PersonsAtHome SiblingsAtSchool
## 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
##    TypesOfHouses
## 1              1
## 2              2
## 3              2
## 4           <NA>
## 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 its output?

```r
filedata$FathersOccupation <- factor(filedata$FathersOccupation, levels = c(1,2,3), labels = c("Farmer"
filedata
```

```
##    Respondents Sex FathersOccupation PersonsAtHome SiblingsAtSchool
## 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            Driver             4                1
## 8            8   1            Others             2                2
## 9            9   2            Farmer            11                6
## 10          10   1            Others             6                2
##    TypesOfHouses
## 1              1
## 2              2
## 3              2
## 4           <NA>
## 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 its

output.

```
DriverDad <- subset(filedata, Sex == 2 & FathersOccupation == "Driver")
DriverDad
```

```
##   Respondents Sex FathersOccupation PersonsAtHome SiblingsAtSchool
## 2           2   2            Driver             7                3
## 6           6   2            Driver             4                3
## 7           7   2            Driver             4                1
##   TypesOfHouses
## 2             2
## 6             3
## 7             1
```

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

```
FiveSiblings <- subset(filedata, Respondents & SiblingsAtSchool >= 5)
FiveSiblings
```

```
##   Respondents Sex FathersOccupation PersonsAtHome SiblingsAtSchool
## 4           4   1            Others             8                5
## 9           9   2            Farmer            11                6
##   TypesOfHouses
## 4          <NA>
## 9             3
```

4. Interpret the graph.

-The graph reflects a general mood of concern or dissatisfaction on social media during the period, with negative sentiments dominating the conversation. It shows the number of tweets per day from July 14 to July 21, 2020, with sentiment categories reveals clear patterns in social media behavior.