RWorksheet_Basa#3b

Jared J Basa

2024-12-18

1. Create a data frame using the table below.

a. Write the codes.

```
Table <- data.frame(
  Respondents = c(1:20),
  Sex = c(2,2,1,2,2,2,2,2,2,1,2,2,2,2,2,2, 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),
  Type_of_House = c(1, 2, 3, 1, 1, 3, 3, 1, 2, 3, 2, 3, 2, 2, 3, 3, 3, 3, 3, 2)
)

Table
```

```
##    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
##    Type_of_House
## 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(Table)
```

```
## '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 ...
##  $ Type_of_House     : num  1 2 3 1 1 3 3 1 2 3 ...
```

```
summary(Table)
```

```
##   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 Type_of_House
##  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?

```
meanSiblings_at_School <- mean(Table$Siblings_at_School)
meanSiblings_at_School
```

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

no

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

```
subsetRespondents <- subset(Table, Respondents <= 2)
subsetRespondents
```

```
##   Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
```

```
## 1           1   2                  1               5                  6
## 2           2   2                  3               7                  4
##   Type_of_House
## 1             1
## 2             2
```

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

```
rowCols <- Table[c(3, 5), c("Sex", "Persons_at_Home")]
rowCols
```

```
##   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.

```
Type_of_House <- Table$Type_of_House
Type_of_House
```

```
##  [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.

```
maleFarmers <- Table[Table$Sex == 1 & Table$Fathers_Occupation == 1, ]
maleFarmers
```

```
## [1] Respondents        Sex                Fathers_Occupation Persons_at_Home
## [5] Siblings_at_School Type_of_House
## <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.

```
femaleMoreSiblings <- Table[Table$Sex == 2 & Table$Siblings_at_School >= 5, ]
femaleMoreSiblings
```

```
##    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
##    Type_of_House
## 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)
```

```r
print("Structure of the empty dataframe:")
```

```
## [1] "Structure of the empty dataframe:"
```

```r
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 shows data frame is empty and shows structure and types of columns that was defined.

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

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

```r
houseHold <- read.csv("HouseholdData.csv")
houseHold
```

```
##    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-congrete
## 6    Semi-congrete
## 7             Wood
## 8    Semi-congrete
## 9    Semi-congrete
## 10        Congrete
```

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.

```r
houseHold$Sex <- factor(houseHold$Sex, levels = c("Male", "Female"), labels = c(1, 2))
houseHold$Sex <- as.integer(houseHold$Sex)
houseHold
```

```
##    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.

```
houseHold$Types.of.Houses <- factor(houseHold$Types.of.Houses,
                                    levels = c("Wood", "Concrete", "Semi-concrete"),
                                    labels = c(1, 2, 3))
houseHold$Types.of.Houses <- as.integer(houseHold$Types.of.Houses)
houseHold
```

```
##    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                1
## 2               NA
## 3               NA
## 4                1
## 5               NA
## 6               NA
## 7                1
## 8               NA
## 9               NA
## 10              NA
```

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

```
houseHold$Fathers.Occupation <- factor(houseHold$Fathers.Occupation, levels = c(1, 2, 3), labels = c("F
houseHold
```

```
##    Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 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
##    Types.of.Houses
## 1                1
## 2               NA
## 3               NA
## 4                1
## 5               NA
## 6               NA
## 7                1
## 8               NA
## 9               NA
## 10              NA
```

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

```
fatherDriver <- subset(houseHold, Sex == 2 & Fathers.Occupation == "Driver")
fatherDriver
```

```
##   Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 2           2   2             Driver               7                  3
## 6           6   2             Driver               4                  3
## 7           7   2             Driver               4                  1
##   Types.of.Houses
## 2              NA
## 6              NA
## 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.

```
SiblingSchool <- subset(houseHold, Siblings.at.School >= 5)
SiblingSchool
```

```
##   Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 4           4   1             Others               8                  5
## 9           9   2             Farmer              11                  6
##   Types.of.Houses
## 4               1
## 9              NA
```

4. Interpret the graph. The graph shows from July 14 to July 21, 2020, the daily count of negative tweets is higher than any other category.Negative tweets reached its peak in July 15. Positive and neutral tweets is much lesser than negative tweets each day.