Rworksheet#3b

Lance Y. Sarabia

2022-11-23

1. Create a data frame using the table below. a. Write the codes.

```
data1 <- data.frame(Respondents = c(seq(1:20)),
                    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)
)
data1
```

```
##    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

```
sum_data1 <- summary(data1)
sum_data1
```

```
##   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 Types_of_houses
##  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? - No, the mean of siblings at school is 2.95

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

```
sub_data1 <- subset(data1[1:2, 1:6])
sub_data1
```

```
##   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.

```
sub_data2 <- subset(data1[c(3,5),c(2,4)])
sub_data2
```

```
##   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.

```
sub_data3 <- c(data1$Types_of_houses)
sub_data3
```

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

```
types_houses <- sub_data3
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.

```
sub_data4 <- subset(data1[c(1:20), c(2,3)])
sub_data4
```

```
##    Sex Fathers_Occupation
## 1    2                  1
## 2    2                  3
## 3    1                  3
## 4    2                  3
## 5    2                  1
## 6    2                  2
## 7    2                  3
## 8    2                  1
## 9    2                  1
## 10   2                  1
## 11   1                  3
## 12   2                  2
## 13   2                  1
## 14   2                  3
## 15   2                  3
## 16   2                  1
## 17   2                  3
## 18   2                  1
## 19   1                  2
## 20   2                  1
```

```
data_male <- sub_data4[data1$Fathers_Occupation == '1',]
data_male
```

```
##    Sex Fathers_Occupation
## 1    2                  1
```

```
## 5    2                  1
## 8    2                  1
## 9    2                  1
## 10   2                  1
## 13   2                  1
## 16   2                  1
## 18   2                  1
## 20   2                  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.

```
sub_data5 <- subset(data1[c(1:20), c(2,5)])
sub_data5
```

```
##    Sex Siblings_at_school
## 1    2                  6
## 2    2                  4
## 3    1                  4
## 4    2                  1
## 5    2                  2
## 6    2                  1
## 7    2                  5
## 8    2                  3
## 9    2                  1
## 10   2                  2
## 11   1                  3
## 12   2                  2
## 13   2                  5
## 14   2                  5
## 15   2                  2
## 16   2                  1
## 17   2                  2
## 18   2                  5
## 19   1                  3
## 20   2                  2
```

```
data_female <- sub_data5[data1$Siblings_at_school >= '5',]
data_female
```

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

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)
paste("Structure of the empty dataframe:")
```

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

```
paste(str(df))
```

```
## 'data.frame':    0 obs. of  5 variables:
##  $ Ints      : int
##  $ Doubles   : num
##  $ Characters: chr
##  $ Logicals  : logi
##  $ Factors   : Factor w/ 0 levels:
```

```
## character(0)
```

a. Describe the results. - The data shows that there is no rows but has 5 columns with zero(0) factor levels.

3. Interpret the graph. -From the data it shows that there is more negative tweets than the neutral and positive, and it skewed from left to right within the timeframe of july 15, 2020 - july 21, 2020.