Worksheet-3b in R

Drake Francis M. Jaculina

2022-11-23

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

1. Create a data frame using the table below.

a. Write the codes.

```
statistics1 <- data.frame(
        Respondents = c(seq(1,20)),
        Sex = c(rep( c(2,1,2,1,2,1,2), c(2,1,7,1,7,1,1))),
        FathersOccupation = 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),
        Typesofhouses = c(1,2,3,1,1,3,3,1,2,3,2,3,2,2,3,3,3,3,3,2)
        )
statistics1
```

```
##    Respondents Sex FathersOccupation PersonsatHome Siblingsatschool
## 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
##    Typesofhouses
## 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.

The provided data displays the Respondent's responses in Sex, Fathers Occupation, Persons at home, Siblings at school, and Types of houses. Females outnumber males in the column sex. This is a survey data. In the fathers occupation, the highest number is the farmer followed by others and last is the Driver. Respondent number 18 had the highest number of people at home in the "Persons at home" column. The most siblings in school are 6, while the fewest are 1. In the Types of houses column, the concrete is the highest with 10 respondents followed by semi-concrete with 6 respondents, while wood is the lowest with 4 respondents.

Get the structure or the summary of the data.

```
summary(statistics1)
```

```
##   Respondents          Sex       FathersOccupation PersonsatHome
##  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
##  Siblingsatschool Typesofhouses
##  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, because siblings at school has 2.95 mean.

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

```
Extract_SubsetD <- subset(statistics1[1:2,  ])
Extract_SubsetD
```

```
##   Respondents Sex FathersOccupation PersonsatHome Siblingsatschool
## 1           1   2                 1             5                6
## 2           2   2                 3             7                4
##   Typesofhouses
## 1             1
## 2             2
```

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

```
Extract_SubsetE <- subset(statistics1[c(3,5),c(2,4)])
Extract_SubsetE
```

```
##   Sex PersonsatHome
## 3   1             3
## 5   2             5
```

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

```
types_houses <- select(statistics1, Typesofhouses)
types_houses
```

```
##    Typesofhouses
## 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
```

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

```
Father_farmer <- subset(statistics1[c(1:20), c(2,3)])
Father_farmer
```

```
##    Sex FathersOccupation
## 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
```

```
Male_respo <- subset(statistics1,Sex == '1' & FathersOccupation == '1')
Male_respo
```

```
## [1] Respondents       Sex               FathersOccupation PersonsatHome
## [5] Siblingsatschool  Typesofhouses
## <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.

```
siblings <- subset(statistics1[c(1:20), c(1,2,5)])
siblings
```

```
##    Respondents Sex Siblingsatschool
## 1            1   2                6
## 2            2   2                4
## 3            3   1                4
## 4            4   2                1
## 5            5   2                2
## 6            6   2                1
## 7            7   2                5
## 8            8   2                3
## 9            9   2                1
## 10          10   2                2
## 11          11   1                3
## 12          12   2                2
## 13          13   2                5
```

```
## 14          14   2                5
## 15          15   2                2
## 16          16   2                1
## 17          17   2                2
## 18          18   2                5
## 19          19   1                3
## 20          20   2                2
```

```
female <- siblings[statistics1$Siblingsatschool >= '5',]
female
```

```
##    Respondents Sex Siblingsatschool
## 1            1   2                6
## 7            7   2                5
## 13          13   2                5
## 14          14   2                5
## 18          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)
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 creates an empty data frame.It resulted to 0 observations, 5 variables, and NULL due to Initializing a data frame with an empty column.

3. Interpret the graph.

Figure 1: Sentiments of Tweets per day - Donald Trump

The graph presents the sentiments of tweets per day - Donald Trump. It has a legend at the right upper side, red for negative, yellow for neutral, and blue for positive. As you can see in the graph, there are more negative sentiments than neutral and positive sentiments. The negative sentiment (shown by the red color) is the highest among the three for six consecutive days from July 14, 2020, to July 21, 2020. And then Positive and Neutral sentiments as follows.