RWorksheet_Sadsad#3b.Rmd

Missy Key Sadsad

2023-10-10

```r
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.

```r
#a.
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)
Father_Occupation <- 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)
TypeOfHouses <- c (1,2,3,1,1,3,3,1,2,3,2,3,2,2,3,3,3,3,3,2)

data_display <- data.frame(Respondents, Sex,Father_Occupation,PersonsAtHome,SiblingsAtSchool,TypeOfHouse
data_display
```

	Respondents	Sex	Father_Occupation	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

	Respondents	Sex	Father_Occupation	PersonsAtHome	SiblingsAtSchool
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

	TypeOfHouses
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
#in this dataset includes information from 20 respondents. It covers their gender, with 7 males and 13 
summary(data_display)
```

```
  Respondents            Sex        Father_Occupation 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  TypeOfHouses
 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. No
```

```
#d.
data1 <- subset (data_display)[1:2, 2:6, drop=FALSE]
data1
```

```
   Sex Father_Occupation PersonsAtHome SiblingsAtSchool TypeOfHouses
1    2               1             5               6              1
2    2               3             7               4              2
```

```
#e.
data2 <- data_display[c(3,5), c(2,4)]
data2
```

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

```
#f.
types_houses <- data_display[c(6)]
types_houses
```

```
   TypeOfHouses
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
selected_data <- data_display %>% select(1,2,3)
data3 <- selected_data[data_display$Sex == 1,]
data3
```

```
   Respondents Sex Father_Occupation
3            3   1                 3
11          11   1                 3
19          19   1                 2
```

```
#h
selected_data2 <- data_display %>% select(1,2,5)
female <- selected_data2[data_display$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
```

```r
colnames(data_display) <- c("Respondents", "Sex", "Fathers Occupation", "Persons At Home", "Siblings At
```

#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:"
```

```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
```

```r
#a.The output indicates that df is a dataframe with 0 observations (rows) and 5 variables (columns). It
```

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

```r
#a
new_Respondents <-  c(seq(1,10))
new_Sex <- c("Male","Female","Female","Male","Male","Female","Female","Male","Female","Male")
new_Fathers_Occupation <- c(1,2,3,3,1,2,2,3,1,3)
new_PersonsAtHome <- c(5,7,3,8,6,4,4,2,11,6)
new_SiblingsAtSchool <- c(2,3,0,5,2,3,1,2,6,2)
new_TypeOfHouses <- c ("Wood","Congrete","Congrete","Wood","Semi-Congrete","Semi-Congrete","Wood","Semi-

HouseholdData <- data.frame(new_Respondents, new_Sex,new_Fathers_Occupation,new_PersonsAtHome,new_Siblir


csv_file <- "HouseholdData.csv"
write.csv(HouseholdData, file = csv_file)
HouseholdData <- read.csv("HouseholdData.csv")
```

```
#b
data_display1 <- factor(HouseholdData$new_Sex, labels = c("Male", "Female"))
data_display1 <- as.integer(data_display1)
data_display1
```

```
 [1] 2 1 1 2 2 1 1 2 1 2
```

```
#c
data_display2 <- factor(HouseholdData$new_TypeOfHouses, labels = c("Wood" = 1, "Congrete" = 2,"Semi-Con
data_display2 <- as.integer(data_display2)
data_display2
```

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

```
#d
data_display3 <- factor(HouseholdData$new_Fathers_Occupation, labels = c("Farmer" = 1, "Driver" = 2,"Ot
data_display3 <- as.integer(data_display3)
data_display3
```

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

```
#e
selected_data3 <- HouseholdData %>% select(2, 3,4)
data4 <- selected_data3[HouseholdData$new_Fathers_Occupation == 2, ]
data4
```

```
  new_Respondents new_Sex new_Fathers_Occupation
2               2  Female                      2
6               6  Female                      2
7               7  Female                      2
```

```
#f
selected_data3 <- HouseholdData %>% select(2,6)
data4 <- selected_data3[HouseholdData$new_SiblingsAtSchool >= 5,]
data4
```

```
  new_Respondents new_SiblingsAtSchool
4               4                    5
9               9                    6
```

```
colnames(HouseholdData) <- c("Respondents", "Sex", "Fathers Occupation", "Persons At Home", "Siblings A
```

4. Interpret the Graph.

In the bar graph titled "Sentiment of Tweets per Day," it have three distinct sentiment categories: Negative, Neutral, and Positive. Each category represents the prevailing mood and tone of the tweets on specific dates in July 2020.

Negative Sentiment: This represents tweets expressing dissatisfaction, criticism, or negative emotions. On specific dates like July 15 and July 21, 2020, negative tweets surged, indicating intense discussions or concerns.

Neutral Sentiment: These tweets strike a balanced tone, presenting information objectively. Across various days in July 2020, including July 14, 15, 17, 18, and 21, neutral sentiments prevailed, reflecting varying levels of non-partisan discourse.

Positive Sentiment: This category showcases optimistic, enthusiastic, and positive tweets. Despite negative sentiments on certain days, like July 14, 15, 17, 18, and 21, positive tweets also shone, symbolizing resilience, hope, or a positive outlook amid diverse sentiments.

In summary, this graph succinctly captures Twitter's sentiment landscape in July 2020, highlighting fluctuations in Negative, Neutral, and Positive sentiments on specific dates.