

RWorksheet_Nava#3b

Samantha Nicole Nava

2024-09-31

1.

a.

```
library(readxl)
resdata <- read_excel("respondents_data.xlsx")
head(resdata)
```

```
## # A tibble: 6 x 6
##   Respondents Sex `Fathers Occupation` `Persons at Home` `Siblings at school`
##   <dbl> <dbl>          <dbl>          <dbl>          <dbl>
## 1         1     1             1             5             6
## 2         2     2             3             7             4
## 3         3     2             3             3             4
## 4         4     1             3             8             1
## 5         5     2             1             5             2
## 6         6     2             2             9             1
## # i 1 more variable: `Types of houses` <dbl>
```

b.

```
str(resdata)
```

```
## tibble [20 x 6] (S3: tbl_df/tbl/data.frame)
##  $ Respondents      : num [1:20] 1 2 3 4 5 6 7 8 9 10 ...
##  $ Sex              : num [1:20] 1 2 2 1 2 2 2 2 1 2 ...
##  $ Fathers Occupation: num [1:20] 1 3 3 3 1 2 3 1 1 1 ...
##  $ Persons at Home   : num [1:20] 5 7 3 8 5 9 6 7 8 4 ...
##  $ Siblings at school: num [1:20] 6 4 4 1 2 1 5 3 1 2 ...
##  $ Types of houses   : num [1:20] 1 2 3 3 1 3 3 1 2 3 ...
```

```
summary(resdata)
```

```
##   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.:1.75   1st Qu.:1.00   1st Qu.: 5.0
##   Median :10.50   Median :2.00   Median :2.00   Median : 7.0
##   Mean   :10.50   Mean   :1.75   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 :3.00     Median :3.0
##   Mean   :3.00     Mean   :2.4
##   3rd Qu.:4.25     3rd Qu.:3.0
```

```
## Max. :6.00 Max. :3.0
```

c. yes
d.

```
firsttwo <- resdata[1:2, ]  
firsttwo
```

```
## # A tibble: 2 x 6  
## Respondents Sex `Fathers Occupation` `Persons at Home` `Siblings at school`  
## <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 1 1 5 6  
## 2 2 2 3 7 4  
## # i 1 more variable: `Types of houses` <dbl>
```

e.

```
selected <- resdata[c(3, 5), c(2, 4)]  
selected
```

```
## # A tibble: 2 x 2  
## Sex `Persons at Home`  
## <dbl> <dbl>  
## 1 2 3  
## 2 2 5
```

f.

```
types_houses <- resdata$`Types of houses`  
types_houses
```

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

g.

```
male_farmers <- resdata[resdata$Sex == 1 & resdata$`Fathers Occupation` == 1, ]  
male_farmers
```

```
## # A tibble: 2 x 6  
## Respondents Sex `Fathers Occupation` `Persons at Home` `Siblings at school`  
## <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 1 1 5 6  
## 2 9 1 1 8 1  
## # i 1 more variable: `Types of houses` <dbl>
```

h.

```
females_with_siblings <- resdata[resdata$Sex == 2 & resdata$`Siblings at school` >= 5, ]  
females_with_siblings
```

```
## # A tibble: 4 x 6  
## Respondents Sex `Fathers Occupation` `Persons at Home` `Siblings at school`  
## <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 7 2 3 6 5  
## 2 13 2 1 4 5  
## 3 14 2 3 7 5  
## 4 18 2 1 11 5  
## # i 1 more variable: `Types of houses` <dbl>
```

2.

```
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. The dataframe is empty but it has 5 columns of different types.

3.

a.

```
respondents_data <- read.csv("respondents_data.csv")
respondents_data
```

```
## 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 2 4
## 8 8 Male 3 2 2
## 9 9 Female 1 11 1
## 10 10 Male 3 6 2
## Types.of.Houses
## 1 Wood
## 2 Conrete
## 3 Conrete
## 4 Wood
## 5 Semi-concrete
## 6 Semi-concrete
## 7 Wood
## 8 Semi-concrete
## 9 Semi-concrete
## 10 Conrete
```

b.

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

```
## [1] 1 2 2 1 1 2 2 1 2 1
## Levels: 1 2
```

c.

```
respondents_data$Types.of.Houses <- factor(respondents_data$Types.of.Houses, levels = c("Wood", "Concrete", "Brick"))
respondents_data$Types.of.Houses
```

```
## [1] 1 <NA> <NA> 1 3 3 1 3 3 <NA>
## Levels: 1 2 3
```

d.

```
respondents_data$Fathers.Occupation <- factor(respondents_data$Fathers.Occupation, levels = c(1, 2, 3),
respondents_data$Fathers.Occupation <- as.integer(respondents_data$Fathers.Occupation)
respondents_data$Fathers
```

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

e.

```
respondents_data$Fathers.Occupation <- factor(respondents_data$Fathers.Occupation, levels = c(1, 2, 3),
respondents_data$Fathers.Occupation <- as.integer(respondents_data$Fathers.Occupation)
respondents_data$Fathers
```

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

f.

```
siblings5 <- respondents_data[respondents_data$Siblings.at.School >= 5, ]
siblings5
```

```
## Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 4 4 1 3 8 5
## Types.of.Houses
## 4 1
```

4. The graph plots the daily sentiment of the tweets, categorized by color to represent different tweet types. It can be seen that there is a persistent pattern where negative tweets are always greater than neutral or positive tweets. That means Twitter is primarily used to voice critical opinions or negativity.