RWorksheet_asenjo#3b

Samuel Asenjo

2024-09-30

1.

a.

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

```
##    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. The data has 20 rows and 6 columns

```
summary(data)
```

```
##   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. No, the mean is 2.95

```
SiblingsatSchool =  c(6, 4, 4, 1, 2, 1, 5, 3, 1, 2, 3, 2, 5, 5, 2, 1, 2, 5, 3, 2)
mean(SiblingsatSchool)
```

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

d.

```
sd <- data[1:2, ]
sd
```

```
##   Respondents Sex FathersOccupation PersonsatHome SiblingsatSchool
## 1           1   2                 1             5                6
## 2           2   2                 3             7                4
##   TypesofHouses
## 1             1
## 2             2
```

e.

```
ed <- data[c(3, 5), c(2,4)]
ed
```

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

f.

```
types_houses <- data[, "TypesofHouses"]

types_houses
```

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

g.

```
mf <- subset(data, Sex == 1 & FathersOccupation == 1)
mf
```

```
## [1] Respondents       Sex               FathersOccupation PersonsatHome
## [5] SiblingsatSchool  TypesofHouses
## <0 rows> (or 0-length row.names)
```

h.

```
fs <- subset(data, Sex ==2 & SiblingsatSchool >=5)
fs
```

```
##    Respondents Sex FathersOccupation PersonsatHome SiblingsatSchool
## 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
##    TypesofHouses
## 1              1
## 7              3
## 13             2
## 14             2
## 18             3
```

2.

a. it displays the data types of each objects.

```
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
```

3. a.

```
HouseholdData <- read.csv("/cloud/project/Worksheet 3/HouseholdData.csv")
```

b.

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

```
##    Respondents Sex FathersOccupation PersonsAtHome SiblingsAtSchool
## 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
##    TypesOfHouses
## 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.

```
HouseholdData$TypesOfHouses <- factor(HouseholdData$TypesOfHouses, levels = c("Wood", "Congrete", "Semi
HouseholdData
```

```
##    Respondents Sex FathersOccupation PersonsAtHome SiblingsAtSchool
## 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
##    TypesOfHouses
## 1              1
## 2              2
## 3              2
## 4           <NA>
## 5              3
## 6              3
## 7              1
## 8              3
## 9              3
```

```
## 10             2
```

d.

```
HouseholdData$FathersOccupation <- factor(HouseholdData$FathersOccupation, levels = c(1, 2, 3), labels =
HouseholdData
```

```
##    Respondents Sex FathersOccupation PersonsAtHome SiblingsAtSchool
## 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
##    TypesOfHouses
## 1              1
## 2              2
## 3              2
## 4           <NA>
## 5              3
## 6              3
## 7              1
## 8              3
## 9              3
## 10             2
```

e.

```
fr <- subset(HouseholdData, Sex==2 & FathersOccupation =="Driver")
fr
```

```
##   Respondents Sex FathersOccupation PersonsAtHome SiblingsAtSchool
## 2           2   2            Driver             7                3
## 6           6   2            Driver             4                3
## 7           7   2            Driver             4                1
##   TypesOfHouses
## 2             2
## 6             3
## 7             1
```

f.

```
ns <- subset(HouseholdData, Respondents & SiblingsAtSchool >=  5)
ns
```

```
##   Respondents Sex FathersOccupation PersonsAtHome SiblingsAtSchool
## 4           4   1            Others             8                5
## 9           9   2            Farmer            11                6
##   TypesOfHouses
## 4          <NA>
## 9             3
```

The graph illustrates the daily sentiment of tweets from July 14 to July 21, 2020, revealing that negative sentiments (red) were the most prominent, particularly on July 15 and July 21, while positive sentiments (blue) peaked on July 21. The neutral sentiments (yellow) also peaked on July 15. In general, the data indicates that

public discussion during this time was mainly critical, with significantly fewer positive expressions compared to negative ones.