RWorksheet_Leysa#3b

Camilo Leysa

2024-10-01

#Problem 1# a.

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

```
##    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   1                  1              4                  2
## 11          11   1                  3              7                  3
## 12          12   2                  2              5                  2
## 13          13   1                  1              4                  5
## 14          14   2                  3              7                  5
## 15          15   2                  3              8                  2
## 16          16   2                  1              8                  1
## 17          17   1                  3              3                  2
## 18          18   2                  1             11                  5
## 19          19   2                  2              7                  3
## 20          20   1                  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.

```
str(tabdata)
```

```
## 'data.frame':    20 obs. of  6 variables:
##  $ Respondents       : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Sex               : num  2 2 1 2 2 2 2 2 2 1 ...
##  $ Fathers_Occupation: num  1 3 3 3 1 2 3 1 1 1 ...
##  $ Persons_at_Home   : num  5 7 3 8 5 9 6 7 8 4 ...
##  $ Siblings_at_School: num  6 4 4 1 2 1 5 3 1 2 ...
##  $ Types_of_Houses   : num  1 2 3 1 1 3 3 1 2 3 ...
```

```
summary(tabdata)
```

```
##   Respondents         Sex      Fathers_Occupation Persons_at_Home
##  Min.   : 1.00   Min.   :1.0   Min.   :1.00       Min.   : 3.0
##  1st Qu.: 5.75   1st Qu.:1.0   1st Qu.:1.00       1st Qu.: 5.0
##  Median :10.50   Median :2.0   Median :2.00       Median : 7.0
##  Mean   :10.50   Mean   :1.7   Mean   :1.95       Mean   : 6.4
##  3rd Qu.:15.25   3rd Qu.:2.0   3rd Qu.:3.00       3rd Qu.: 8.0
##  Max.   :20.00   Max.   :2.0   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.

```
mean_sibling <- mean(c(6,4,4,1,2,1,5,3,1,2,3,2,5,5,2,1,2,5,3,2))
mean_sibling
```

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

d.

```
first_two <- tabdata[1:2,]
first_two
```

```
##   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.

```
sub_table <- tabdata[c(3,5), c(2,4)]
sub_table
```

```
##   Sex Persons_at_Home
## 3   1               3
## 5   2               5
```

f.

```
types_houses <- tabdata$Types_of_Houses
types_houses
```

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

g.

```
female_students <- tabdata[tabdata$Sex == 2 & tabdata$Siblings_at_School >=5,]
female_students
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1            1   2                  1               5                  6
## 7            7   2                  3               6                  5
## 14          14   2                  3               7                  5
## 18          18   2                  1              11                  5
##    Types_of_Houses
## 1                1
## 7                3
## 14               2
## 18               3
```

#Problem 2# a.

```
df = data.frame(Ints=integer(), Doubles=double(),Characters=character(), Logicals=logical(), Factors=fa
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
```

#Problem 3# a.

```
hdata <- read.csv("HouseholdData.csv")
hdata
```

```
##    Respondents    Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 1            1   Male             Farmer               5                  2
## 2            2 Female             Driver               7                  3
## 3            3 Female             Others               3                  0
## 4            4   Male             Others               8                  5
## 5            5   Male             Farmer               6                  2
## 6            6 Female             Driver               4                  3
## 7            7 Female             Driver               4                  1
## 8            8   Male             Others               2                  2
## 9            9 Female             Farmer              11                  6
## 10          10   Male             Others               6                  2
##    Types.of.Houses
## 1             Wood
## 2         Congrete
## 3         Congrete
## 4             Wood
## 5    Semi-congrete
## 6    Semi-congrete
## 7             Wood
## 8    Semi-congrete
## 9    Semi-congrete
## 10        Congrete
```

b.

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

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

c.

```
hdata$Types.of.Houses <- factor(hdata$Types.of.Houses, levels = c("Wood","Congrete","Semi-congrete"), la
hdata$Types.of.Houses
```

```
##  [1] 1 2 2 1 3 3 1 3 3 2
## Levels: 1 2 3
```

d.

```
hdata$Fathers.Occupation <- factor(hdata$Fathers.Occupation, levels = c("Farmer","Driver","Others"), lab
hdata$Fathers.Occupation
```

```
##  [1] 1 2 3 3 1 2 2 3 1 3
## Levels: 1 2 3
```

e.

```
rfemale <- subset(hdata, Sex ==2 & Fathers.Occupation==2)
rfemale
```

```
##   Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 2           2   2                  2               7                  3
## 6           6   2                  2               4                  3
## 7           7   2                  2               4                  1
##   Types.of.Houses
## 2               2
## 6               3
## 7               1
```

f.

```
fildata <- subset(hdata,Siblings.at.School >=5)
fildata
```

```
##   Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 4           4   1                  3               8                  5
## 9           9   2                  1              11                  6
##   Types.of.Houses
## 4               1
## 9               3
```

4.

```
#This bar chart shows the sentiment of tweets per day within the range dates of
#July 14, 2020 to July 21,2020. The sentiment categories are represented by
#"Negative", "Neutral", and "Positive", where each sentiment is represented by
#different colored bars. As we can see in the graph, there is a huge spike of
#negative tweets. In addition, on most days, negative sentiments appear to
#dominate over positive and neutral ones.
```