RWorksheet_Celestra#3b

Kenneth Celestra

2024-10-02

1.

a.

```r
# Create a data frame
data <- data.frame(
  Respondents = 1:20,
  Sex = c(2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 1, 2),
  Fathers_Occupation = c(1, 3, 3, 3, 1, 2, 3, 1, 1, 1, 3, 2, 1, 3, 3, 1, 3, 1, 2, 1),
  Persons_at_Home = c(5, 7, 3, 8, 5, 9, 6, 7, 8, 4, 7, 5, 4, 7, 8, 8, 3, 11, 7, 6),
  Siblings_at_School = c(6, 4, 4, 1, 2, 1, 5, 3, 1, 2, 3, 2, 5, 5, 2, 3, 1, 5, 3, 2),
  Types_of_Houses = c(1, 2, 3, 1, 1, 3, 3, 1, 2, 3, 2, 3, 2, 2, 3, 3, 3, 3, 3, 2)
)

print(data)
```

```
##    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   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                  3
## 17          17   2                  3               3                  1
## 18          18   2                  1              11                  5
## 19          19   1                  2               7                  3
## 20          20   2                  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. The dataset contains information on 20 respondents, each described by several categorical and numerical variables. The columns include the respondents' sex, their father's occupation, the number of persons at home, the number of siblings attending school, and the type of house they live in.

```r
str(data)
```

```
## '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 2 ...
##  $ 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 ...
```

    c. No it is 3

```r
mean(data$Siblings_at_School)
```

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

    d.

```r
subset <- data[1:2, ]

print(subset)
```

```
##   Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1           1   2                  1               5                  6
## 2           2   2                  2               3               7                  4
##   Types_of_Houses
## 1               1
## 2               2
```

    e.

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

print(subsetdata)
```

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

    f.

```
types_houses <- data$Types_of_Houses

print(types_houses)

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

g.

```
male_farmers <- data[data$Sex == 1 & data$Fathers_Occupation == 1, ]
print(male_farmers)

## [1] Respondents       Sex              Fathers_Occupation Persons_at_Home
## [5] Siblings_at_School Types_of_Houses
## <0 rows> (or 0-length row.names)
```

h.

```
female_with_5_siblings <- data[data$Sex == 2 & data$Siblings_at_School >= 5, ]

print(female_with_5_siblings)

##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 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
##    Types_of_Houses
## 1                1
## 7                3
## 13               2
## 14               2
## 18               3
```

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. It show the structure of an empty data frame that has 0 observations and 5 variables which are Int, Double, Char, Logical, and Factors

3.

a.

```r
householdData <-read.csv("HouseholdData.csv")
householdData
```

```
##    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               4                  1
## 8            8   Male                  3               2                  2
## 9            9 Female                  1              11                  6
## 10          10   Male                  3               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.

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

householdData$Sex <- as.integer(householdData$Sex)

householdData
```

```
##    Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 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
##     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
```

c.

```
householdData$Types.of.Houses <- factor(householdData$Types.of.Houses, levels = c("Wood", "Congrete", "S

householdData$Types.of.Houses <- as.integer(householdData$Types.of.Houses)

householdData
```

```
##    Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 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
##    Types.of.Houses
## 1                1
## 2                2
## 3                2
## 4                1
## 5                3
## 6                3
## 7                1
## 8                3
## 9                3
## 10               2
```

d.

```
householdData$Fathers.Occupation <- factor(householdData$Fathers.Occupation, levels = c(1,2,3), labels

householdData
```

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

```
## 5                 3
## 6                 3
## 7                 1
## 8                 3
## 9                 3
## 10                2
```

   e.

```
female_drivers <- householdData[householdData$Sex == 2 & householdData$Fathers.Occupation == "Driver", ]
```

```
female_drivers
```

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

   f.

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

```
siblings5
```

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

4. The bar graph titled "Sentiments of Tweets Per Day" shows how tweets with different feelings (Negative, Neutral, and Positive) were shared between July 14 and July 21, 2020. The number of tweets is shown on the y-axis, and each day is divided into the three types of sentiment. The graph highlights that Negative sentiment was the most common throughout these days, especially on July 15 and July 21, when negative tweets went over 4,000. Neutral sentiment stayed fairly steady with some small changes, while Positive sentiment was usually lower but increased a bit on July 17 and July 18. This suggests that during this time, most tweets had a negative tone, likely in reaction to events happening at the time. In summary, negative tweets were the most frequent, with neutral and positive feelings being shared less often.