RWorksheet_Vicinte#3b

Anthony Dave Vicinte

2024-09-30

#1. Create a data frame using the table below. #a.

```
data <- data.frame(
  Respondents = 1:20,
  Sex = c(2, 1, 2, 2, 1, 1, 2, 2, 2, 1, 1, 2, 2, 1, 2, 2, 1, 2, 1, 2),
  FatherOccupation = c(1, 2, 3, 1, 2, 1, 3, 2, 3, 3, 1, 3, 2, 1, 3, 1, 3, 3, 1, 1),
  Persons_at_Home = c(5, 7, 3, 5, 5, 3, 6, 6, 7, 7, 3, 7, 4, 7, 8, 8, 3, 11, 8, 6),
  Siblings_at_School = c(6, 4, 3, 2, 3, 3, 5, 5, 4, 5, 3, 7, 5, 2, 1, 3, 1, 5, 3, 2),
  Types_of_Houses = c(1, 2, 3, 1, 3, 1, 3, 3, 3, 1, 3, 3, 3, 1, 3, 3, 3, 3, 3, 2)
)
data
```

```
##    Respondents Sex FatherOccupation Persons_at_Home Siblings_at_School
## 1            1   2                1               5                  6
## 2            2   1                2               7                  4
## 3            3   2                3               3                  3
## 4            4   2                1               5                  2
## 5            5   1                2               5                  3
## 6            6   1                1               3                  3
## 7            7   2                3               6                  5
## 8            8   2                2               6                  5
## 9            9   2                3               7                  4
## 10          10   1                3               7                  5
## 11          11   1                1               3                  3
## 12          12   2                3               7                  7
## 13          13   2                2               4                  5
## 14          14   1                1               7                  2
## 15          15   2                3               8                  1
## 16          16   2                1               8                  3
## 17          17   1                3               3                  1
## 18          18   2                3              11                  5
## 19          19   1                1               8                  3
## 20          20   2                1               6                  2
##    Types_of_Houses
## 1                1
## 2                2
## 3                3
## 4                1
## 5                3
## 6                1
## 7                3
## 8                3
## 9                3
```

```
## 10               1
## 11               3
## 12               3
## 13               3
## 14               1
## 15               3
## 16               3
## 17               3
## 18               3
## 19               3
## 20               2
```

#b.

```
str(data)
```

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

```
summary(data)
```

```
##   Respondents         Sex      FatherOccupation Persons_at_Home
##  Min.   : 1.00   Min.   :1.0   Min.   :1        Min.   : 3.00
##  1st Qu.: 5.75   1st Qu.:1.0   1st Qu.:1        1st Qu.: 4.75
##  Median :10.50   Median :2.0   Median :2        Median : 6.00
##  Mean   :10.50   Mean   :1.6   Mean   :2        Mean   : 5.95
##  3rd Qu.:15.25   3rd Qu.:2.0   3rd Qu.:3        3rd Qu.: 7.00
##  Max.   :20.00   Max.   :2.0   Max.   :3        Max.   :11.00
##  Siblings_at_School Types_of_Houses
##  Min.   :1.00       Min.   :1.00
##  1st Qu.:2.75       1st Qu.:1.75
##  Median :3.00       Median :3.00
##  Mean   :3.60       Mean   :2.40
##  3rd Qu.:5.00       3rd Qu.:3.00
##  Max.   :7.00       Max.   :3.00
```

#c.

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

```
## [1] 3.6
```

#d.

```
data[1:2, ]
```

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

#e.

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

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

#f.

```r
types_houses <- data$Types_of_Houses
types_houses
```

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

#g.

```r
MaleFarmers <- subset(data, Sex == 1 & FatherOccupation == 1)
MaleFarmers
```

```
##    Respondents Sex FatherOccupation Persons_at_Home Siblings_at_School
## 6            6   1                1               3                  3
## 11          11   1                1               3                  3
## 14          14   1                1               7                  2
## 19          19   1                1               8                  3
##    Types_of_Houses
## 6                1
## 11               3
## 14               1
## 19               3
```

#h.

```r
Female_Siblings <- subset(data, Sex == 2 & Siblings_at_School >= 5)
Female_Siblings
```

```
##    Respondents Sex FatherOccupation Persons_at_Home Siblings_at_School
## 1            1   2                1               5                  6
## 7            7   2                3               6                  5
## 8            8   2                2               6                  5
## 12          12   2                3               7                  7
## 13          13   2                2               4                  5
## 18          18   2                3              11                  5
##    Types_of_Houses
## 1                1
## 7                3
## 8                3
## 12               3
## 13               3
## 18               3
```

#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)
```

```r
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
```

#a. Describe the results: The empty data frame has 5 columns: integers, doubles, characters, logicals, and factors, but no data is stored yet. #3. Create a .csv file of this. Save it as HouseholdData.csv #a

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

#b.

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

```
##    Respondents Sex Father.s.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.

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

```
##    Respondents Sex Father.s.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.

```
datac$Father.s.Occupation <- factor(datac$Father.s.Occupation, levels = c(1,2,3), labels = c("Farmer",
datac
```

```
##    Respondents Sex Father.s.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.

```
FemaleDriverDad <- subset(datac, Sex == 2 & Father.s.Occupation =="Driver")
FemaleDriverDad
```

```
##   Respondents Sex Father.s.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.

```
manysiblings <- subset(datac, Respondents & Siblings.at.School >= 5)
manysiblings
```

```
##   Respondents Sex Father.s.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. Interpret the graph. #The graph shows the number of tweets per day from July 14 to July 21, 2020, categorized by sentiment (negative, neutral, and positive). On most days, negative tweets are the highest, followed by neutral and positive tweets.