RWorkheet_nandin#3b.Rmd

Mary coolen nandin

2024-10-05

1. Create a data frame using the table below.

a. Write the codes.

```r
data <- data.frame(
Respondents = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 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),
TypeofHouses = c(1, 2, 3, 1, 1, 3, 3, 1, 2, 3, 2, 3, 2, 2, 3, 3, 3, 3, 3, 2)
)
colnames(data) <- c("Respondents", "Sex", "Fathers Occupation", "Persons At Home", "Siblings At School"
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                  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
##    Type 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
```

```r
legend <- list(
Sex = c("Male" = 1, "Female" = 2),
FathersOccupation = c("Farmer" = 1, "Driver" = 2, "Others" = 3),
TypeofHouses = c("Wood" = 1, "Semi-concrete" = 2, "Concrete" = 3)
)
legend
```

```
## $Sex
##   Male Female
##      1      2
##
## $FathersOccupation
## Farmer Driver Others
##      1      2      3
##
## $TypeofHouses
##           Wood Semi-concrete       Concrete
##              1             2              3
```

b. Describe the data. Get the structure or the summary of the data

```r
str(data)
```

```
## 'data.frame':    20 obs. of  6 variables:
##  $ Respondents       : num  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 ...
##  $ Type of Houses    : num  1 2 3 1 1 3 3 1 2 3 ...
```

```r
summary(data)
```

```
##   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.: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
##  Siblings At School Type 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
```

Figure 1: R Chunk

c. Is the mean number of siblings attending is 5?

```
meansiblings <- mean(data$SiblingsAtSchool)
```

```
## Warning in mean.default(data$SiblingsAtSchool): argument is not numeric or
## logical: returning NA
```

d. Extract the 1st two rows and then all the columns using the subsetting functions. Write the codes and its output.

```
first2rows <- data[1:2, ]
first2rows
```

```
##   Respondents Sex Fathers Occupation Persons At Home Siblings At School
## 1           1   2                  1               5                  6
## 2           2   2                  3               7                  4
##   Type of Houses
## 1              1
## 2              2
```

e. Extract 3rd and 5th row with 2nd and 4th column. Write the codes and its result.

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

```
##   Sex Persons At Home
## 3   1               3
## 5   2               5
```

f. Select the variable types of houses then store the vector that results as types_houses. Write the codes.

```
types_houses <- data$TypeofHouses
types_houses
```

```
## NULL
```

g. Select only all Males respondent that their father occupation was farmer. Write the codes and its output.

```
malesfarmers <- data[data$Sex == 1 & data$FathersOccupation == 1, ]
malesfarmers
```

```
## [1] Respondents        Sex                 Fathers Occupation Persons At Home
## [5] Siblings At School Type of Houses
## <0 rows> (or 0-length row.names)
```

h. Select only all females respondent that have greater than or equal to 5 number of siblings attending school. Write the codes and its outputs.

```
femalessiblings <- data[data$Sex == 2 & data$SiblingsAtSchool >= 5, ]
femalessiblings
```

```
## [1] Respondents        Sex                 Fathers Occupation Persons At Home
## [5] Siblings At School Type of Houses
## <0 rows> (or 0-length row.names)
```

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)
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.

  3. Create a .csv file of this. Save it as HouseholdData.csv

   a. Import the csv file into the R environment. Write the codes.

```r
data1 = read.csv("/cloud/project/HouseholdData.csv")
data1
```

```
##    Respondents    Sex FathersOccupation PersonsAtHome SiblingsAtSchool
## 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
##     TypeofHouses  X
## 1           Wood NA
## 2       Congrete NA
## 3       Congrete NA
## 4           Wood NA
## 5  Semi-congrete NA
## 6  Semi-congrete NA
## 7           Wood NA
## 8  Semi-congrete NA
## 9  Semi-congrete NA
## 10          Wood NA
```

   b. Convert the Sex into factor using factor() function and change it into integer.[Legend: Male = 1 and Female = 2]. Write the R codes and its output.

```r
data1$Sex <- factor(data1$Sex, levels = c("Male", "Female"), labels = c(1, 2))
data1$Sex
```

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

c. Convert the Type of Houses into factor and change it into integer. [Legend: Wood = 1; Congrete = 2; Semi-Congrete = 3]. Write the R codes and its output.

```
data1$TypeofHouse <- factor(data1$TypeofHouse,
levels = c("Wood", "Congrete", "Semi-congrete"),
labels = c(1, 2, 3))
data1$TypeofHouse
```

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

d. On father's occupation, factor it as Farmer = 1; Driver = 2; and Others = 3. What is the R code and its output?

```
data1$FathersOccupation <- factor(data1$FathersOccupation,
levels = c(1, 2, 3),
labels = c("Farmer", "Driver", "Others"))
data1$FathersOccupation
```

```
## [1] Farmer Driver Others Others Farmer Driver Driver Others Farmer Others
## Levels: Farmer Driver Others
```

e. Select only all females respondent that has a father whose occupation is driver. Write the codes and its output.

```
femaledrivers <- data1[data1$Sex == 2 & data1$FathersOccupation == 2, ]
femaledrivers
```

```
## [1] Respondents       Sex                 FathersOccupation PersonsAtHome
## [5] SiblingsAtSchool  TypeofHouses        X                 TypeofHouse
## <0 rows> (or 0-length row.names)
```

f. Select the respondents that have greater than or equal to 5 number of siblings attending school. Write the codes and its output.

```
siblings_filter <- data1[data1$SiblingsAtSchool >= 5, ]
siblings_filter
```

```
##   Respondents Sex FathersOccupation PersonsAtHome SiblingsAtSchool
## 4           4   1             Others             8                5
## 9           9   2             Farmer            11                6
##     TypeofHouses  X TypeofHouse
## 4           Wood NA           1
## 9  Semi-congrete NA           3
```

4. Interpret the graph

The graph below plots the actual sentiment analysis of the tweets between July 14, 2020 to July 21, 2020, with them classified as negative, neutral, or positive. Most days were mainly dominated by negative sentiments, with the highest numbers of negative tweets found on July 15 and July 21, respectively, with a sheer peak of about 4000 on July 15. Meanwhile, the few days that actually had fewer tweets were July 18 and July 20, particularly the former with its low levels of negative tweets. Although positive tweets were relatively consistent across the period, they were always out-numbered by negative tweets, suggesting an overall pattern of negative slant to the week's experience.