RWorksheet_Subosa#3b.Rmd

Gian Adree Subosa

2024-10-02

#1. Create a data frame using the table below

```
#a.) Write the codes
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, 1, 2, 2, 1, 2)
father <- c(1, 3, 3, 3, 1, 2, 3, 1, 1, 2, 3, 2, 1, 3, 3, 1, 3, 1, 2, 1)
persons <- c(5, 7, 3, 8, 5, 9, 6, 7, 8, 4, 7, 5, 4, 7, 8, 8, 3, 11, 7, 6)
siblings <- c(6, 4, 4, 1, 2, 1, 5, 3, 1, 2, 3, 2, 5, 5, 2, 1, 2, 5, 3, 2)
houses <- c(1, 2, 3, 1, 1, 3, 3, 1, 2, 3, 2, 3, 2, 2, 3, 3, 3, 3, 3, 2)

data_table <- data.frame(Respondents = respondents, Sex = sex, Fathers_Occupation = father, Persons_at_
print(data_table)
```

```
##    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                  2              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   1                  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
##    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.) Describe the data. Get the structure or the summary of the data

#The data was organized by using data frame, enable to generate a table output

#Structure and summary of the data
summary(data_table)
```

```
##   Respondents          Sex      Fathers_Occupation Persons_at_Home
##  Min.   : 1.00   Min.   :1.0   Min.   :1          Min.   : 3.0
##  1st Qu.: 5.75   1st Qu.:2.0   1st Qu.:1          1st Qu.: 5.0
##  Median :10.50   Median :2.0   Median :2          Median : 7.0
##  Mean   :10.50   Mean   :1.8   Mean   :2          Mean   : 6.4
##  3rd Qu.:15.25   3rd Qu.:2.0   3rd Qu.:3          3rd Qu.: 8.0
##  Max.   :20.00   Max.   :2.0   Max.   :3          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
```

```
str(data_table)
```

```
## '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 2 ...
##  $ 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.) Is the mean number of siblings attending is 5?
mean_sibling <- mean(data_table$Siblings_at_school)
mean_sibling == 5
```

```
## [1] FALSE
```

```
print(mean_sibling)
```

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

```
#The answer is NO. The mean number of siblings attending is not 5 but 2.95
```

```r
#d.) Extract the 1st two rows and then all the columns using the subsetting functions
first_2_rows <- data_table[1:2, ]
print(first_2_rows)
```

```
##   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
```

```r
#e.) Extract 3rd and 5th row with 2nd and 4th column
extract_df <- data_table[c(3, 5), c(2, 4)]
print(extract_df)
```

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

```r
#f.) Select the variable types of houses then store the vector that results as types_houses
types_houses <- data_table$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
```

```r
#g.) Select only all Males respondent that their father occupation was farmer
m_farmers <- data_table[data_table$Sex == 1 & data_table$Fathers_Occupation == 1, ]
print(m_farmers)
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_school
## 16          16   1                  1               8                  1
##    Types_of_houses
## 16               3
```

```r
#h.) Select only all females respondent that have greater than or equal to 5 number of siblings attendi
f_siblings <- data_table[data_table$Sex == 2 & data_table$Siblings_at_school >= 5, ]
print(f_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. Write a R program to create an empty data frame

```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:"
```

```
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 result of this program is NULL and it has no observations.
```

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

```
respondents <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
sex <- c("Male", "Female", "Female", "Male", "Male", "Female", "Female", "Male", "Female", "Male")
father <- c(1, 2, 3, 3, 1, 2, 2, 3, 1, 3)
persons <- c(5, 7, 3, 8, 6, 4, 4, 2, 11, 6)
siblings <- c(2, 3, 0, 5, 2, 3, 1, 2, 6, 2)
houses <- c("Wood", "Congrete", "Congrete", "Wood", "Semi-congrete", "Semi-congrete", "Wood", "Semi-cong

houseData <- data.frame(Respondents = respondents, Sex = sex, Fathers_Occupation = father, Persons_at_H
print(houseData)
```

```
##    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
```

```
#a.) Import the csv file into the R environment
write.csv(houseData, file = "HouseholdData.csv", row.names = FALSE)
importedData <- read.csv("HouseholdData.csv")
```

```r
print(importedData)
```

```
##    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
```

```r
#b.) Convert the Sex into factor using factor() function and change it into integer
importedData$Sex <- factor(importedData$Sex, levels = c("Male", "Female"), labels = c(1, 2))
print(importedData)
```

```
##    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
```

```r
#c.) Convert Type of Houses into a factor and change to integer
importedData$Types_of_Houses <- factor(importedData$Types_of_Houses, levels = c("Wood", "Congrete", "Se
print(importedData)
```

```
##    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.) On father's occupation, factor it as Farmer = 1; Driver = 2; and Others = 3
importedData$Fathers_Occupation <- factor(importedData$Fathers_Occupation, levels = c("Farmer", "Driver
print(importedData)
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1            1   1               <NA>               5                  2
## 2            2   2               <NA>               7                  3
## 3            3   2               <NA>               3                  0
## 4            4   1               <NA>               8                  5
## 5            5   1               <NA>               6                  2
## 6            6   2               <NA>               4                  3
## 7            7   2               <NA>               4                  1
## 8            8   1               <NA>               2                  2
## 9            9   2               <NA>              11                  6
## 10          10   1               <NA>               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.) Select only all females respondent that has a father whose occupation is driver
femresp_drivers <- subset(importedData, Sex == 2 & Fathers_Occupation == 2)
print(femresp_drivers)
```

```
## [1] Respondents        Sex                Fathers_Occupation Persons_at_Home
```

```
## [5] Siblings_at_School Types_of_Houses
## <0 rows> (or 0-length row.names)
```

```
#f.) Select respondents with greater than or equal to 5 siblings attending school
siblings_grth_5 <- subset(importedData, Siblings_at_School >= 5)
print(siblings_grth_5)
```

```
##   Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 4           4   1               <NA>               8                  5
## 9           9   2               <NA>              11                  6
##   Types_of_Houses
## 4               1
## 9               3
```

#4. Interpret the graph

```
#This graph shows the sentiment analysis of tweets from July 14, 2020 to July 21, 2020, grouped into ne
```