RWorksheet_Songaling#3b.Rmd

Charles Daniel Songaling

2023-10-11

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
#code here

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

#a. Write the codes.
household_data <- data.frame(
  Respondents =c(1:20),
  Sex = c("Female","Female","Male","Female","Female","Female","Female","Female","Female","Female","Male

  Fathers_Occupation = c("Farmer","Others","Others","Others","Farmer","Driver","Others","Farmer","Farme

  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(6,4,4,1,2,1,5,3,1,2,3,2,5,5,2,1,2,5,3,2),

  Types_of_houses = c("Wood","Semi-Concrete","Concrete","Wood","Wood","Concrete","Concrete","Wood","Sem
)
household_data
```

```
##    Respondents    Sex Fathers_Occupation Persons_at_home Siblings_at_school
## 1            1 Female             Farmer               5                  6
## 2            2 Female             Others               7                  4
## 3            3   Male             Others               3                  4
## 4            4 Female             Others               8                  1
## 5            5 Female             Farmer               5                  2
## 6            6 Female             Driver               9                  1
## 7            7 Female             Others               6                  5
## 8            8 Female             Farmer               7                  3
## 9            9 Female             Farmer               8                  1
## 10          10 Female             Farmer               4                  2
## 11          11   Male             Others               7                  3
## 12          12 Female             Driver               5                  2
## 13          13 Female             Farmer               4                  5
## 14          14 Female             Others               7                  5
## 15          15 Female             Others               8                  2
## 16          16 Female             Farmer               8                  1
```

```
## 17          17 Female             Others               3                 2
## 18          18 Female             Farmer              11                 5
## 19          19   Male             Driver               7                 3
## 20          20 Female             Farmer               6                 2
##    Types_of_houses
## 1             Wood
## 2    Semi-Concrete
## 3         Concrete
## 4             Wood
## 5             Wood
## 6         Concrete
## 7         Concrete
## 8             Wood
## 9    Semi-Concrete
## 10        Concrete
## 11   Semi-Concrete
## 12        Concrete
## 13   Semi-Concrete
## 14   Semi-Concrete
## 15        Concrete
## 16        Concrete
## 17        Concrete
## 18        Concrete
## 19        Concrete
## 20   Semi-Concrete
```

```
#b. Describe the data. Get the structure or the summary of the data
summary(household_data)
```

```
##   Respondents        Sex            Fathers_Occupation Persons_at_home
##  Min.   : 1.00   Length:20          Length:20          Min.   : 3.0
##  1st Qu.: 5.75   Class :character   Class :character   1st Qu.: 5.0
##  Median :10.50   Mode  :character   Mode  :character   Median : 7.0
##  Mean   :10.50                                         Mean   : 6.4
##  3rd Qu.:15.25                                         3rd Qu.: 8.0
##  Max.   :20.00                                         Max.   :11.0
##  Siblings_at_school Types_of_houses
##  Min.   :1.00       Length:20
##  1st Qu.:2.00       Class :character
##  Median :2.50       Mode  :character
##  Mean   :2.95
##  3rd Qu.:4.25
##  Max.   :6.00
```

```
#c. Is the mean number of siblings attending is 5?
mean_siblings <- mean(household_data$Siblings_at_school)
is_mean_5 <- mean_siblings == 5
print(is_mean_5)
```

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

```
#The answer is no because the mean number is 2.95.

#d. Extract the 1st two rows and then all the columns using the subsetting functions. Write the codes a
first_two_rows_all_columns <- household_data[1:2, ]
print(first_two_rows_all_columns)
```

```
##   Respondents    Sex Fathers_Occupation Persons_at_home Siblings_at_school
## 1           1 Female             Farmer               5                  6
## 2           2 Female             Others               7                  4
##   Types_of_houses
## 1            Wood
## 2   Semi-Concrete
```

```
#It shows the data in 1st two rows and the columns in those rows.

#e. Extract 3rd and 5th row with 2nd and 4th column. Write the codes and its result.
selected_rows_columns <- household_data[c(3, 5), c(2, 4)]
print(selected_rows_columns)
```

```
##      Sex Persons_at_home
## 3   Male               3
## 5 Female               5
```

```
#The output is the data in the 3rd and 5th row with the 2nd and 4th column.

#f. Select the variable types of houses then store the vector that results as types_houses. Write the c
types_houses <- household_data$Types_of_houses
print(types_houses)
```

```
##  [1] "Wood"          "Semi-Concrete" "Concrete"      "Wood"
##  [5] "Wood"          "Concrete"      "Concrete"      "Wood"
##  [9] "Semi-Concrete" "Concrete"      "Semi-Concrete" "Concrete"
## [13] "Semi-Concrete" "Semi-Concrete" "Concrete"      "Concrete"
## [17] "Concrete"      "Concrete"      "Concrete"      "Semi-Concrete"
```

```
#g. Select only all Males respondent that their father occupation was farmer. Write the codes and its o
males_farmers <- household_data[household_data$Sex == "Male" & household_data$Fathers_Occupation == "Fa
print(males_farmers)
```

```
## [1] Respondents        Sex                Fathers_Occupation Persons_at_home
## [5] Siblings_at_school Types_of_houses
## <0 rows> (or 0-length row.names)
```

```
#The output is empty because there is no male with a father whose occupation is farmer.

#h. Select only all females respondent that have greater than or equal to 5 number of siblings attendin
females_greater_than_5_siblings <- household_data[household_data$Sex == "Female" & household_data$Sibli
print(females_greater_than_5_siblings)
```

```
##    Respondents    Sex Fathers_Occupation Persons_at_home Siblings_at_school
## 1            1 Female             Farmer               5                  6
## 7            7 Female             Others               6                  5
## 13          13 Female             Farmer               4                  5
## 14          14 Female             Others               7                  5
## 18          18 Female             Farmer              11                  5
##    Types_of_houses
## 1             Wood
## 7          Concrete
## 13   Semi-Concrete
## 14   Semi-Concrete
## 18        Concrete
```

```
#The output is the data of all female respondents that have 5 or more siblings attending school.
```

```r
#2. Write a R program to create an empty data frame. Using the following codes:
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
```

```r
#a. Describe the results.
#The data frame is empty because no data has been added yet.

#3. Create a .csv file of this. Save it as HouseholdData.csv
write.csv(household_data, file = "HouseholdData.csv", row.names = FALSE)

#a. Import the csv file into the R environment. Write the codes.
imported_data <- read.csv("HouseholdData.csv")

#b. Convert the Sex into factor using factor()function and change it into integer.[Legend: Male = 1 and
imported_data$Sex <- factor(imported_data$Sex, levels = c("Male", "Female"))
imported_data$Sex <- as.integer(imported_data$Sex)

#c. Convert the Type of Houses into factor and change it into integer. [Legend: Wood = 1; Concrete = 2;
imported_data$Types_of_houses <- factor(imported_data$Types_of_houses, levels = c("Wood", "Concrete", "S
imported_data$Types_of_houses <- as.integer(imported_data$Types_of_houses)

#d.On father's occupation, factor it as Farmer = 1; Driver = 2; and Others = 3. What is the R code and
imported_data$Fathers_Occupation <- factor(imported_data$Fathers_Occupation, levels = c("Farmer", "Drive
imported_data$Fathers_Occupation <- as.integer(imported_data$Fathers_Occupation)

#e. Select only all females respondent that has a father whose occupation is driver. Write the codes an
female_drivers <- imported_data[imported_data$Sex == 2 & imported_data$Fathers_Occupation == 2, ]
print(female_drivers)
```

```
##    Respondents Sex Fathers_Occupation Persons_at_home Siblings_at_school
## 6            6   2                  2              9                  1
## 12          12   2                  2              5                  2
##    Types_of_houses
## 6                2
## 12               2
```

```r
#The output is the data of all the female respondent that has a father whose occupation is a driver but

#f. Select the respondents that have greater than or equal to 5 number of siblings attending school. Wr
greater_than_5_siblings <- imported_data[imported_data$Siblings_at_school >= 5, ]
```

```r
print(greater_than_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                2
## 13               3
## 14               3
## 18               2
```

```r
#The output is the data of all the respondents that have greater than or equal to 5 number of siblings

#4. Interpret the graph.
# The graph in figure 3 shows the sentiments of people when tweeting during July 14 until July 21, 2020
```