RWorksheet_Gener#3b.Rmd

Kent Gener

2023-10-25

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:

```r
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Worksheet-3b in R

Instructions: • Use RStudio or the RStudio Cloud accomplish this worksheet. • Inside the folder worksheet#3, create an .Rmd (R Markdown) for this worksheet and saved it as RWorksheet_lastname#3b.Rmd • Knit to pdf to render a pdf file. • On your own GitHub repository, push the .Rmd file, as well as the pdf worksheet knitted to the repo you have created before. • Do not forget to comment your Git repo on our VLE • Accomplish this worksheet by answering the questions being asked and writing the code manually.

1. Create a data frame using the table below.

a. Write the codes.

```r
Household <- data.frame(
  Respondents = 1:20,
  Sex = c(2,2,1,rep(2,7),1,rep(2,7),1,2),
  Fathers_Occupation = c(1, rep(3,3),1:3,rep(1,3), (3:1),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(6,4,4,1,2,1,5,3,1,2,3,2,5,5,2,1,2,5,3,2),
  Types_of_Houses = c(1:3,1,1,3,3,1:3,2,3,2,2,rep(3,5),2)
)
Household
```

```
##    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
##     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

```
summary(Household)
```

```
##    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 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
```

```r
str(Household)
```

```
## '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. Is the mean number of siblings attending is 5?

```r
mean_siblings_attending <- mean(Household$Siblings_at_School)
mean_siblings_attending == 5
```

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

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

```r
subset1 <- Household[1:2, ]
subset1
```

```
##    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
```

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

```r
subset2 <- Household[c(3, 5), c(2, 4)]
subset2
```

```
##    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 <- Household$Types_of_Houses
types_houses
```

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

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

```
male_farmers <- Household[Household$Sex == 1 & Household$Fathers_Occupation == 1, ]
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. Select only all females respondent that have greater than or equal to 5 number of siblings attending school. Write the codes and its outputs.

```
female_greater_than_5_siblings <- Household[Household$Sex == 2 & Household$Siblings_at_School >= 5, ]
female_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                 3
## 13                2
## 14                2
## 18                3
```

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:") print(str(df))

```
df = data.frame(
  Ints = integer(),
  Doubles = double(),
  Characters = character(),
  Logicals = logical(),
  Factors = factor(),
  stringsAsFactors = FALSE
)
```

    a. Describe the results.

```
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
```

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

```
library(readr)
your_data <- read_csv("HouseholdData.csv")
```

```
## Rows: 10 Columns: 6
## -- Column specification --------------------------------------------------
## Delimiter: ","
## chr (2): Sex, TypeofHouses
## dbl (4): Respondents, Fathers Occupation, Persons at Home, Siblings at School
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
View(your_data)
```

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

```
your_data <- read_csv("HouseholdData.csv")
```

```
## Rows: 10 Columns: 6
## -- Column specification --------------------------------------------------
## Delimiter: ","
## chr (2): Sex, TypeofHouses
## dbl (4): Respondents, Fathers Occupation, Persons at Home, Siblings at School
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

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.

```
your_data$Sex <- factor(your_data$Sex, levels = c("Male", "Female"))
your_data$Sex <- as.integer(your_data$Sex)
your_data$Sex
```

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

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.

```
your_dataTypeofHouses <- factor(your_data$TypeofHouses, levels = c("Wood", "Concrete", "Semi-concrete")

your_dataTypeofHouses
```

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

```
legend_fathers_occupation <- c("Farmer", "Driver", "Others")

your_dataFathers_Occupation <- factor(your_data$'Fathers Occupation', levels= c("1","2","3"), c("Farmer

your_dataFathers_Occupation
```

```
## [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.

```
female_drivers <- your_data[your_data$Sex == 2 & your_dataFathers_Occupation == 2, ]
female_drivers
```

```
## # A tibble: 0 x 6
## # i 6 variables: Respondents <dbl>, Sex <int>, Fathers Occupation <dbl>,
## #   Persons at Home <dbl>, Siblings at School <dbl>, TypeofHouses <chr>
```

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

```
more_than_5_siblings <- your_data[your_data$'Siblings at School' >= 5, ]
more_than_5_siblings
```

```
## # A tibble: 2 x 6
##   Respondents   Sex 'Fathers Occupation' 'Persons at Home' 'Siblings at School'
##         <dbl> <int>                <dbl>            <dbl>                <dbl>
## 1           4     1                    3                8                    5
## 2           9     2                    1               11                    6
## # i 1 more variable: TypeofHouses <chr>
```

4. Interpret the graph.

```
#NUMBER 3 INTERPRET PARAGRAPH
#The graph consistently shows that the negative sentiments are the \n greatest in number, The neutral b
```