RWorksheet#3b

Tamonan

2024-10-07

```r
# 1. Create a data frame using the table below.
# 1.a Write the codes
respondent_data <- data.frame(
  Respondent = 1:20,
  Sex = c(2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 1, 2),
  Father_Occupation = c(1, 3, 3, 3, 1, 2, 3, 1, 1, 1, 3, 2, 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, 2, 3, 1, 1, 3, 3, 1, 2, 3, 2, 3, 2, 2, 3, 3, 3, 3, 3, 2))

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

```
##    Respondent         Sex        Father_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(respondent_data)
```

```
## 'data.frame':    20 obs. of  6 variables:
##  $ Respondent        : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Sex               : num  2 2 1 2 2 2 2 2 2 2 ...
##  $ Father_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 ...
```

```r
# 1.c Is the mean number of siblings attending is 5?
mean_siblings <- mean(respondent_data$Siblings_at_school)
mean_siblings == 5
```

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

```
mean_siblings
```

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

```
# 1.d Extract the 1st two rows and all columns using subsetting functions.
subset_data <- respondent_data[1:2, ]
subset_data
```

```
##   Respondent Sex Father_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
```

```
# 1.e Extract 3rd and 5th row with 2nd and 4th column.
subset_data_3rd_5th <- respondent_data[c(3, 5), c(2, 4)]
subset_data_3rd_5th
```

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

```
# 1.f Select the variable "Type_of_House" and store it as types_houses.
types_houses <- respondent_data$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 2
```

```
# 1.g Select all male respondents whose father occupation was "Farmer".
male_farmers <- respondent_data[respondent_data$Sex == 1 & respondent_data$Father_Occupation == 1, ]
male_farmers
```

```
## [1] Respondent         Sex                Father_Occupation  Persons_at_Home
## [5] Siblings_at_school Types_of_houses
## <0 rows> (or 0-length row.names)
```

```
# 1.h Select all female respondents with greater than or equal to 5 siblings attending school.
female_siblings <- respondent_data[respondent_data$Sex == 2 & respondent_data$Siblings_at_school >= 5, ]
female_siblings
```

```
##    Respondent Sex Father_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
```

```r
# 2. Write an R program to create an empty data frame
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
# 2.a Describe the results:
# The DataFrame is empty right now, but it is prepared with columns that are meant to store specific ty
```

```r
# 3. Create a .csv file of the respondent data frame and save it as "HouseholdData.csv".
# 3.a Import the csv file into the R environment.
respondent_data <- read.csv("D:/RStudio/RWorksheet3/3b/HouseholdData.csv")

# 3.b Convert the Sex column into factor and change it to integer (Male = 1, Female = 2).
respondent_data$Sex <- factor(respondent_data$Sex, levels = c("Male", "Female"), labels = c(1,2))
respondent_data
```

```
##    Respondent Sex Father_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
##    Type_of_House
## 1          wood
## 2       congrete
```

```
## 3          congrete
## 4             wood
## 5   semi-congrete
## 6   semi-congrete
## 7             wood
## 8   semi-congrete
## 9   semi-congrete
## 10         congrete
```

```
# 3.c Convert the Type_of_House into factor and change it into integer (Wood = 1, Concrete = 2, Semi-Co
respondent_data$Type_of_House <- factor(respondent_data$Type_of_House, levels = c("Wood", "Congrete", "S
respondent_data
```

```
##     Respondent Sex Father_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
##     Type_of_House
## 1            <NA>
## 2            <NA>
## 3            <NA>
## 4            <NA>
## 5            <NA>
## 6            <NA>
## 7            <NA>
## 8            <NA>
## 9            <NA>
## 10           <NA>
```

```
# 3.d Factor the Father's occupation (Farmer = 1, Driver = 2, Others = 3).
respondent_data$Father_Occupation <- factor(respondent_data$Father_Occupation, levels = c(1, 2, 3), labe
respondent_data
```

```
##     Respondent Sex Father_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
##     Type_of_House
## 1            <NA>
```

```
## 2               <NA>
## 3               <NA>
## 4               <NA>
## 5               <NA>
## 6               <NA>
## 7               <NA>
## 8               <NA>
## 9               <NA>
## 10              <NA>
```

```r
# 3.e Select all female respondents whose father is a driver.
female_fd <- subset(respondent_data, Sex==2 & Father_Occupation == "Driver")
female_fd
```

```
##   Respondent Sex Father_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
##   Type_of_House
## 2          <NA>
## 6          <NA>
## 7          <NA>
```

```r
# 3.f Select respondents who have greater than or equal to 5 siblings attending school.
siblings_five_or_more <- subset(respondent_data, Respondent & Siblings_at_School >=  5)
siblings_five_or_more
```

```
##   Respondent Sex Father_Occupation Persons_at_Home Siblings_at_School
## 4          4   1            Others               8                  5
## 9          9   2            Farmer              11                  6
##   Type_of_House
## 4          <NA>
## 9          <NA>
```

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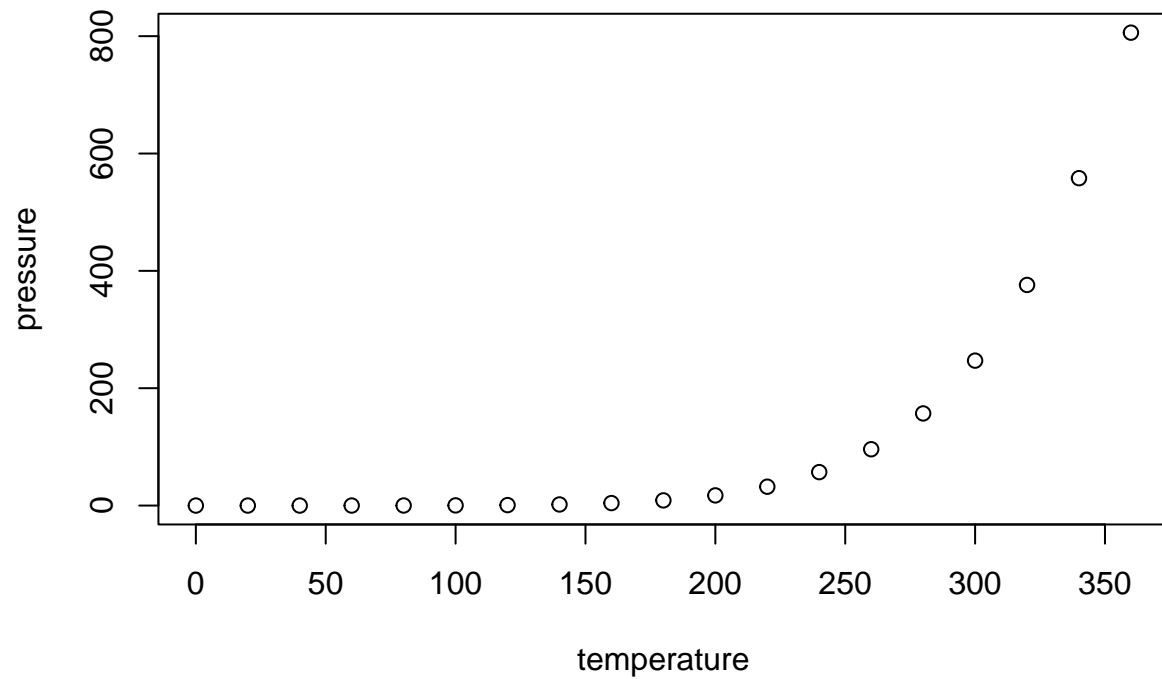
Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.