RWorksheet_Andica#3b

John Laurence Andica

2024-10-02

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

a. Write the codes.

```
library(readr)
myData <- 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),
  Fathers_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)
)
myData
```

```
##    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

```
#The dataset consists of 20 respondents with a slight female majority, varying household sizes (3 to 11
str(myData)
```

```
## '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 ...
##  $ 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_siblings <- mean(myData$Siblings_at_School)
mean_siblings
```

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

```
#No, the mean is 2.95
```

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

```
subset_data <- myData[1:2, ]
subset_data
```

```
##   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.

```
subset_data <- myData[c(3, 5), c(2, 4)]
subset_data
```

```
##   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 <- myData$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
```

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

```
male_farmers_son <- subset(myData, Sex == 1 & Fathers_Occupation == 1)

male_farmers_son
```

```
## [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_with_siblings = subset(myData, Sex == 2 & Siblings_at_School >= 5)
female_with_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:")
```

```
## [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
```

```
#The code outputs 0 observations or what we call row and 5 variables or what we call column.
```

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

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

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

```
## Rows: 10 Columns: 6
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): Sex, Types_of_Houses
## 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.
```

```
my_Data
```

```
## # A tibble: 10 x 6
##    Respondents Sex    Fathers_Occupation Persons_at_Home Siblings_at_School
##          <dbl> <chr>               <dbl>           <dbl>              <dbl>
##  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
## # i 1 more variable: Types_of_Houses <chr>
```

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.

```
my_Data$Sex <- factor(my_Data$Sex, levels = c("Male", "Female"), labels = c(1, 2))
my_Data$Sex <- as.integer(my_Data$Sex)
my_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.

```
my_Data$Types_of_Houses <- factor(my_Data$Types_of_Houses,
                                  levels = c("Wood", "Congrete", "Semi-congrete"),
                                  labels = c(1, 2, 3))
my_Data$Types_of_Houses <- as.integer(my_Data$Types_of_Houses)
my_Data$Types_of_Houses
```

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

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

```
fathers_occupation_vector <- my_Data$Fathers_Occupation
fathers_occupation_factor <- factor(fathers_occupation_vector,
                                    levels = c(1, 2, 3),
                                    labels = c("Farmer", "Driver", "Others"),
                                    exclude = NULL)
fathers_occupation_character <- as.character(fathers_occupation_factor)
fathers_occupation_character
```

```
##  [1] "Farmer" "Driver" "Others" "Others" "Farmer" "Driver" "Driver" "Others"
##  [9] "Farmer" "Others"
```

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

```
female_drivers <- my_Data[my_Data$Sex == 2 & fathers_occupation_character == "Driver", ]
female_drivers
```

```
## # A tibble: 3 x 6
##   Respondents   Sex Fathers_Occupation Persons_at_Home Siblings_at_School
##         <dbl> <int>              <dbl>           <dbl>              <dbl>
## 1           2     2                  2               7                  3
## 2           6     2                  2               4                  3
## 3           7     2                  2               4                  1
## # i 1 more variable: Types_of_Houses <int>
```

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

```
more_siblings <- my_Data[my_Data$Siblings_at_School >= 5, ]
more_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: Types_of_Houses <int>
```