RWorkSheet_Barrientos#3b

Barrientos, Milfrance D

2024-10-02

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

#a. Write the codes.
respondents_data <- data.frame (
respondents = 1:20,
sex = c(2, 2, 1, 2, 2, 2, 2, 2, 2, 1, 1, 2, 1, 2, 2, 1, 2, 1,1, 2),
fathers_occupation = c(1, 3, 3, 3, 1, 2, 3, 1, 1, 1, 3, 2, 1, 3,3, 1, 3, 1, 2, 1),
person_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(
)
respondents_data
```

```
##    respondents sex fathers_occupation person_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   1                  1              4                  2
## 11          11   1                  3              7                  3
## 12          12   2                  2              5                  2
## 13          13   1                  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   1                  1             11                  5
## 19          19   1                  2              7                  3
## 20          20   2                  1              6                  2
##    typeOfHouses
## 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
str(respondents_data)
```

```
## '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 1 ...
##  $ fathers_occupation: num  1 3 3 3 1 2 3 1 1 1 ...
##  $ person_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 ...
##  $ typeOfHouses      : num  1 2 3 1 1 3 3 1 2 3 ...
```

```
respondents_data
```

```
##    respondents sex fathers_occupation person_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   1                  1              4                   2
## 11          11   1                  3              7                   3
## 12          12   2                  2              5                   2
## 13          13   1                  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   1                  1             11                   5
## 19          19   1                  2              7                   3
## 20          20   2                  1              6                   2
##    typeOfHouses
## 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
```

```
# c.Is the mean number of siblings attending is 5?
mean(respondents_data$siblings_at_school)
```

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

```
# no it is 2.95

# d. Extract the 1st two rows and then all the columns using the subsetting functions.
# Write the codes and its output.
firsttworows <- respondents_data [1:2, ]
firsttworows
```

```
##   respondents sex fathers_occupation person_at_home siblings_at_school
## 1           1   2                  1              5                  6
## 2           2   2                  3              7                  4
##   typeOfHouses
## 1            1
## 2            2
```

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

```
##   sex person_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
type_houses <- respondents_data$typeOfHouses
type_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
maleFarmers <- respondents_data[respondents_data$sex == 1 & respondents_data$fathers_occupation == 1, ]
maleFarmers
```

```
##    respondents sex fathers_occupation person_at_home siblings_at_school
## 10          10   1                  1              4                  2
## 13          13   1                  1              4                  5
## 16          16   1                  1              8                  1
## 18          18   1                  1             11                  5
##    typeOfHouses
## 10            3
## 13            2
## 16            3
## 18            3
```

```
# h. Select only all females respondent that have greater than or equal to 5 number of siblings attendi
# Write the codes and its outputs.
fem <- respondents_data[respondents_data$sex == 2 & respondents_data$siblings_at_school >=5, ]
fem
```

```
##    respondents sex fathers_occupation person_at_home siblings_at_school
## 1            1   2                  1              5                  6
## 7            7   2                  3              6                  5
## 14          14   2                  3              7                  5
##    typeOfHouses
## 1             1
## 7             3
## 14            2
```

```
#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
```

```
# a. describe the results - the result is an empty data frame with 0 observations of 5 variables
```

```
# 3.Create a .csv file of this. Save it as HouseholdData.csv
# a. Import the csv file into the R environment. Write the codes.
library(readxl)
Household <- read_excel("C:/PROJ/HouseholdData.xlsx")
Household
```

```
## # 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                    1               8                  5
##  5           5 Male                    1               6                  2
##  6           6 Female                  2               2                  3
##  7           7 Female                  2               4                  1
##  8           8 Male                    2               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.
Household$Sex <- factor(Household$Sex)
Household$Sex <- as.integer(Household$Sex)
Household
```

```
## # A tibble: 10 x 6
##    Respondents   Sex Fathers_Occupation Persons_at_Home Siblings_at_School
##          <dbl> <int>               <dbl>           <dbl>              <dbl>
##  1           1     2                   1               5                  2
##  2           2     1                   2               7                  3
##  3           3     1                   3               3                  0
##  4           4     2                   1               8                  5
##  5           5     2                   1               6                  2
##  6           6     1                   2               2                  3
##  7           7     1                   2               4                  1
##  8           8     2                   2               2                  2
##  9           9     1                   1              11                  6
## 10          10     2                   3               6                  2
## # i 1 more variable: Types_of_Houses <chr>
```

```
# c. Convert the Type of Houses into factor and change it into integer. [Legend: Wood = 1; Congrete = 2
Household$Types_of_Houses <- factor(Household$Types_of_Houses, levels = c("Wood", "Congrete", "Semi-con
Household$Types_of_Houses <- as.integer(as.character(Household$Types_of_Houses))
Household
```

```
## # A tibble: 10 x 6
##    Respondents   Sex Fathers_Occupation Persons_at_Home Siblings_at_School
##          <dbl> <int>               <dbl>           <dbl>              <dbl>
##  1           1     2                   1               5                  2
##  2           2     1                   2               7                  3
##  3           3     1                   3               3                  0
##  4           4     2                   1               8                  5
##  5           5     2                   1               6                  2
##  6           6     1                   2               2                  3
##  7           7     1                   2               4                  1
##  8           8     2                   2               2                  2
##  9           9     1                   1              11                  6
```

```
## 10          10     2                   3               6                   2
## # i 1 more variable: Types_of_Houses <int>
```

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

#e. Select only all females respondent that has a father whose occupation is driver.
# Write the codes and its output.
fem <- Household[Household$Sex == 2 & Household$Fathers_Occupation == 2, ]
fem
```

```
## # A tibble: 5 x 6
##   Respondents   Sex Fathers_Occupation Persons_at_Home Siblings_at_School
##         <dbl> <int>              <int>           <dbl>               <dbl>
## 1          NA    NA                 NA              NA                  NA
## 2          NA    NA                 NA              NA                  NA
## 3          NA    NA                 NA              NA                  NA
## 4          NA    NA                 NA              NA                  NA
## 5          NA    NA                 NA              NA                  NA
## # 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.
sib <- Household[Household$Siblings_at_School >= 5, ]
sib
```

```
## # A tibble: 2 x 6
##   Respondents   Sex Fathers_Occupation Persons_at_Home Siblings_at_School
##         <dbl> <int>              <int>           <dbl>               <dbl>
## 1           4     2                 NA               8                   5
## 2           9     1                 NA              11                   6
## # i 1 more variable: Types_of_Houses <int>
```

```
# 4. interpret the graph
# The graph shows the sentiment of tweets collected on different days.
# The sentiment is classified into three categories:
# the positive represented by Blue, the negative represented by Red,
# and the neutral represented by Yellow.
# The data is shown from July 14, 2020, to July 20, 2020.
# The Y-axis represents the count of tweets, while the X-axis
# shows the different sentiment categories for each day.
# On almost every day, the Red or Negative sentiment is the highest,
# followed by the Blue or Positive sentiment, and the Yellow
# or Neutral sentiment is the lowest.
# The highest count of tweets is on July 15, 2020,
# while the lowest count of tweets is on July 20, 2020.
# The sentiment of the tweets is mostly negative, followed by positive,
# and the least is neutral.
```