RWorksheet_Benedicto#3b.R

Gabriel R. Benedicto

2024-10-02

1. Create a data frame using the table below. A. Write the codes.

```
data <- data.frame(
  Respondents = 1:20,
  Sex = c(2, 2, 1, 2, 2, 2, 2, 2, 2, 1, 1, 2, 1, 2, 3, 2, 3, 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)
)
data
```

```
##    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   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   3                  3               8                  2
## 16          16   2                  1               8                  1
## 17          17   3                  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. (The data consist of the following Respondents, Sex , Farmers Occupation, Person at Home, Siblings at School and Type of Houses) Get the structure or the summary of the data

```
str(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 ...
##  $ 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?(NO it is 2.95)

```
mean_siblings_at_school <- mean(data$Siblings_at_School)
mean_siblings_at_school
```

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

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

```
sub_data <- data[1:2, ]
sub_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.

```
sub_data2 <- data[c(3,5), c(2,4)]
sub_data2
```

```
##   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.

```
type_houses <- data$Types_of_Houses
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 output.

```
male_farmers <- data[data$Sex == 1 & data$Fathers_Occupation == 1, ]
male_farmers
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 10          10   1                  1               4                  2
## 13          13   1                  1               4                  5
##    Types_of_Houses
## 10               3
## 13               2
```

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_respondents <- data[data$Sex == 2 & data$Siblings_at_School >= 5, ]
female_respondents
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1            1   2                  1               5                  6
## 7            7   2                  3               6                  5
## 14          14   2                  3               7                  5
## 18          18   2                  1              11                  5
##    Types_of_Houses
## 1                1
## 7                3
## 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
```

A. Describe the output: The structure describes an empty data frame with 5 variables but no observations. The variables are:

Ints: Integer type Doubles: Numeric type Characters: Character type Logicals: Logical type Factors: Factor type with no levels defined

3. Create a .csv file of this. Save it as HouseholdData.csv A. Import the csv file into the R environment. Write the codes.

```
household_data <- read.csv("HouseholdData.csv")
household_data
```

```
##    Respondents    Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 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
##    Types.of.Houses
## 1             Wood
## 2         Congrete
## 3         Congrete
## 4             Wood
## 5    Semi-congrete
## 6    Semi-congrete
## 7             Wood
## 8    Semi-congrete
## 9    Semi-congrete
## 10        Congrete
```

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

```
##    Respondents Sex Fathers.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
##    Types.of.Houses
## 1             Wood
## 2         Congrete
## 3         Congrete
## 4             Wood
## 5    Semi-congrete
## 6    Semi-congrete
## 7             Wood
## 8    Semi-congrete
## 9    Semi-congrete
## 10        Congrete
```

C. 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.

```r
head(household_data)
```

```
##   Respondents Sex Fathers.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
##   Types.of.Houses
## 1            Wood
## 2        Congrete
## 3        Congrete
## 4            Wood
## 5   Semi-congrete
## 6   Semi-congrete
```

```r
household_data$Types.of.Houses <- as.integer(factor(household_data$Types.of.Houses, levels = c("Wood", "
household_data
```

```
##    Respondents Sex Fathers.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
##    Types.of.Houses
## 1                1
## 2                2
## 3                2
## 4                1
## 5                3
## 6                3
## 7                1
## 8                3
## 9                3
## 10               2
```

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

```r
household_data$Fathers.Occupation <- as.character(factor(household_data$Fathers.Occupation, levels = c(
household_data
```

```
##    Respondents Sex Fathers.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
##    Types.of.Houses
## 1                1
## 2                2
## 3                2
## 4                1
## 5                3
## 6                3
## 7                1
## 8                3
## 9                3
## 10               2
```

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

```
female_driver <- subset(household_data, Sex == 2 & Fathers.Occupation == "Driver")
female_driver
```

```
##   Respondents Sex Fathers.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
##   Types.of.Houses
## 2               2
## 6               3
## 7               1
```

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

```
siblings_5_or_more <- subset(household_data, Siblings.at.School >= 5)
siblings_5_or_more
```

```
##   Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 4           4   1            Others               8                 5
## 9           9   2            Farmer              11                 6
##   Types.of.Houses
## 4               1
## 9               3
```

4. Interpret the Graph *July 15, 2020: There is a very high count of negative tweets (over 4000), the highest among all categories and dates.* July 14, 2020: There is a balanced distribution of tweets with a slight tilt towards negative and positive tweets. Neutral tweets are relatively fewer. *July 17, 2020: The number of negative tweets is higher than positive tweets, while neutral tweets are lower.* July 18, 2020: The number of negative tweets is higher than positive tweets, while neutral tweets are lower. *July 20, 2020: Negative tweets outnumber positive ones, while neutral tweets remain low.* July 21, 2020: Negative tweets have a significant rise again, close to 4000, while positive and neutral tweets follow behind.