

RWorksheet_Benliro#3b

Rica Marie Benliro

2023-10-15

1. Create a data frame using the table below.

a. Write the codes.

```
household_data <- data.frame(  
  respondents = c(1:20),  
  sex = c(2,2,1,2,2,2,2,2,2,2,1,2,2,2,2,2,2,1,2),  
  father_occupation = c(1,3,3,3,1,2,3,1,1,1,3,2,1,3,3,1,3,1,2,1),  
  personAtHome = c(5,7,3,8,5,9,6,7,8,4,7,5,4,7,8,8,3,11,7,6),  
  siblingAtSchool = c(6,4,4,1,2,1,5,3,1,2,3,2,5,5,2,1,2,5,3,2),  
  typeOfHouses = c(1,2,3,1,1,3,3,1,2,3,2,3,2,2,3,3,3,3,3,2)  
)
```

household_data

##	respondents	sex	father_occupation	personAtHome	siblingAtSchool	typeOfHouses
## 1	1	2	1	5	6	1
## 2	2	2	3	7	4	2
## 3	3	1	3	3	4	3
## 4	4	2	3	8	1	1
## 5	5	2	1	5	2	1
## 6	6	2	2	9	1	3
## 7	7	2	3	6	5	3
## 8	8	2	1	7	3	1
## 9	9	2	1	8	1	2
## 10	10	2	1	4	2	3
## 11	11	1	3	7	3	2
## 12	12	2	2	5	2	3
## 13	13	2	1	4	5	2
## 14	14	2	3	7	5	2
## 15	15	2	3	8	2	3
## 16	16	2	1	8	1	3
## 17	17	2	3	3	2	3
## 18	18	2	1	11	5	3
## 19	19	1	2	7	3	3
## 20	20	2	1	6	2	2

b. Describe the data. Get the structure or the summary of the data

```
summary(household_data)
```

##	respondents	sex	father_occupation	personAtHome
## 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
## siblingAtSchool typeOfHouses
## 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
```

c. Is the mean number of siblings attending is 5?

```
mean(household_data$siblingAtSchool)
```

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

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

```
household_data[1:2, ]
```

```
## respondents sex father_occupation personAtHome siblingAtSchool typeOfHouses
## 1          1 2              1          5          6          1
## 2          2 2              3          7          4          2
```

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

```
household_data[c(3,5), c(2,4)]
```

```
## sex personAtHome
## 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_data$typeOfHouses
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.

```
selection1 <- subset(household_data, sex == 1 & father_occupation == 1)
selection1
```

```
## [1] respondents sex father_occupation personAtHome
## [5] siblingAtSchool typeOfHouses
## <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.

```
selection2 <- subset(household_data, sex == 2 & siblingAtSchool >=5)
selection2
```

```
## respondents sex father_occupation personAtHome siblingAtSchool typeOfHouses
## 1          1 2              1          5          6          1
## 7          7 2              3          6          5          3
## 13         13 2              1          4          5          2
## 14         14 2              3          7          5          2
```

```
## 18          18    2          1          11          5          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 results.

The results shows an empty data frame with 5 different data types but no observations since we have not entered any values in the data frame.

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

```
householdDataCSV <- read.csv(file = "HouseholdData.csv", header = T, stringsAsFactors = F, sep = ",")
householdDataCSV
```

```
## Respondents Sex Fathers.Occupation Person.s.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-concrete
## 6          Semi-concrete
## 7          Wood
## 8          Semi-concrete
## 9          Semi-concrete
## 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.

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

```
householdDataCSV
```

```
##      Respondents Sex Fathers.Occupation Person.s.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-concrete
## 6             Semi-concrete
## 7             Wood
## 8             Semi-concrete
## 9             Semi-concrete
## 10            Congrete
```

- 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.

```
householdDataCSV$Types.of.Houses <- factor(householdDataCSV$Types.of.Houses, levels = c("Wood", "Congrete", "Semi-Congrete"))
householdDataCSV$Types.of.Houses <- as.integer(householdDataCSV$Types.of.Houses)
```

```
householdDataCSV
```

```
##      Respondents Sex Fathers.Occupation Person.s.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?

```
householdDataCSV$Fathers.Occupation <- factor(householdDataCSV$Fathers.Occupation, levels = c("1", "2", "3"))
```

```
householdDataCSV
```

```
## Respondents Sex Fathers.Occupation Person.s.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_selections <- subset(householdDataCSV, Sex == 2 & Fathers.Occupation == "Driver")
female_selections
```

```
## Respondents Sex Fathers.Occupation Person.s.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.

```
withSiblings_selection <- subset(householdDataCSV, Siblings.at.School >= 5)
withSiblings_selection
```

```
## Respondents Sex Fathers.Occupation Person.s.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.

#In the graph entitled “Sentiments of Twitter Per Day”, shows a summary of sentiments scraped from Twitter from July 14, 2020 to July 21, 2020. There are 3 categories of sentiments: Negative, Neutral, and Positive. We can see in the graph that negative sentiments are significantly high in all of the dates. This means that Twitter produces more negative sentiments more than positive which indicates a very toxic and harmful social media environment during the span of the said dates.