

RWorksheet_Perez#3b

2024-10-03

1. Create a data frame using the table below.

a. Write the codes.

```
personal_info <- data.frame(Respondents = 1:20, Sex=c("Female", "Female", "Male", "Female","Female","Female", "Female", "Male", "Male", "Male", "Male", "Male", "Male", "Male", "Male", "Male", "Male", "Male", "Male", "Male", "Male"), Fathers_Occupation=c("Farmer", "Others", "Others", "Others", "Farmer", "Driver", "Others", "Farmer", "Farmer", "Farmer", "Others", "Driver", "Farmer", "Others", "Others", "Farmer", "Others", "Driver", "Farmer", "Farmer"), 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))
```

```
## Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1 1 Female Farmer 5 6
## 2 2 Female Others 7 4
## 3 3 Male Others 3 4
## 4 4 Female Others 8 1
## 5 5 Female Farmer 5 2
## 6 6 Female Driver 9 1
## 7 7 Female Others 6 5
## 8 8 Female Farmer 7 3
## 9 9 Female Farmer 8 1
## 10 10 Female Farmer 4 2
## 11 11 Male Others 7 3
## 12 12 Female Driver 5 2
## 13 13 Female Farmer 4 5
## 14 14 Female Others 7 5
## 15 15 Female Others 8 2
## 16 16 Female Farmer 8 1
## 17 17 Female Others 3 2
## 18 18 Female Farmer 11 5
## 19 19 Male Driver 7 3
## 20 20 Female Farmer 6 2
## Types_of_Houses
## 1 Wood
## 2 Semi-concrete
## 3 Concrete
## 4 Wood
## 5 Wood
## 6 Concrete
## 7 Concrete
## 8 Wood
## 9 Semi-concrete
## 10 Concrete
## 11 Semi-concrete
## 12 Concrete
## 13 Semi-concrete
## 14 Semi-concrete
## 15 Concrete
## 16 Concrete
## 17 Concrete
## 18 Concrete
```

```
## 19      Concrete
## 20      Semi-concrete
```

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

The data frame consists of information about a respondent's sex, their father's occupation, the number of persons at home, siblings at school, and the types of house they live in.

```
str(personal_info)
```

```
## 'data.frame':    20 obs. of  6 variables:
## $ Respondents      : int  1 2 3 4 5 6 7 8 9 10 ...
## $ Sex              : chr  "Female" "Female" "Male" "Female" ...
## $ Fathers_Occupation: chr  "Farmer" "Others" "Others" "Others" ...
## $ 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   : chr  "Wood" "Semi-concrete" "Concrete" "Wood" ...
```

```
summary(personal_info)
```

```
##   Respondents      Sex      Fathers_Occupation Persons_at_Home
##   Min.       : 1.00   Length:20      Length:20      Min.       : 3.0
##   1st Qu.: 5.75   Class :character   Class :character   1st Qu.: 5.0
##   Median :10.50   Mode  :character   Mode  :character   Median : 7.0
##   Mean    :10.50                                     Mean    : 6.4
##   3rd Qu.:15.25                                     3rd Qu.: 8.0
##   Max.     :20.00                                     Max.     :11.0
##   Siblings_at_School Types_of_Houses
##   Min.       :1.00      Length:20
##   1st Qu.:2.00      Class :character
##   Median :2.50      Mode  :character
##   Mean    :2.95
##   3rd Qu.:4.25
##   Max.     :6.00
```

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

```
mean(personal_info[,5])
```

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

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

```
subset(personal_info[1:2, ])
```

```
##   Respondents      Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1           1 Female      Farmer              5              6
## 2           2 Female      Others              7              4
##   Types_of_Houses
## 1           Wood
## 2   Semi-concrete
```

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

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

```
##      Sex Persons_at_Home
## 3   Male              3
## 5 Female              5
```

f. Select the variable types of houses then store the vector that results as `types_houses`. Write the codes.

```
types_houses <- personal_info$Types_of_Houses
types_houses
```

```
## [1] "Wood"          "Semi-concrete" "Concrete"      "Wood"
## [5] "Wood"          "Concrete"      "Concrete"      "Wood"
## [9] "Semi-concrete" "Concrete"      "Semi-concrete" "Concrete"
## [13] "Semi-concrete" "Semi-concrete" "Concrete"      "Concrete"
## [17] "Concrete"      "Concrete"      "Concrete"      "Semi-concrete"
```

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

```
subset(personal_info, Sex == "Male" & Fathers_Occupation == "Farmer")
```

```
## [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.

```
subset(personal_info, Sex == "Female" & Siblings_at_School >= 5)
```

```
##      Respondents      Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1              1 Female              Farmer              5              6
## 7              7 Female              Others              6              5
## 13             13 Female              Farmer              4              5
## 14             14 Female              Others              7              5
## 18             18 Female              Farmer             11              5
##      Types_of_Houses
## 1              Wood
## 7              Concrete
## 13 Semi-concrete
## 14 Semi-concrete
## 18              Concrete
```

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 output shows the structure of the empty data frame that was made from the code given. Since it is empty, the result is NULL. However, it includes different types. These types are integers, doubles, characters, logicals, and factors.

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("/cloud/project/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             Concrete
## 3             Concrete
## 4             Wood
## 5      Semi-concrete
## 6      Semi-concrete
## 7             Wood
## 8      Semi-concrete
## 9      Semi-concrete
## 10            Concrete
```

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.

```
sex_factor <- factor(household_data$Sex)
household_data$Sex <- factor(sex_factor, labels = c(2, 1))
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             Concrete
## 3             Concrete
## 4             Wood
```

```
## 5    Semi-concrete
## 6    Semi-concrete
## 7          Wood
## 8    Semi-concrete
## 9    Semi-concrete
## 10   Concrete
```

- c. Convert the Type of Houses into factor and change it into integer. [Legend: Wood = 1; Concrete = 2; Semi-Concrete = 3]. Write the R codes and its output.

```
house_factor <- factor(household_data$Types.of.Houses)
household_data$Types.of.Houses <- factor(house_factor, labels = c(2, 3, 1))
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?

```
father_occupation <- factor(household_data$Fathers.Occupation)
household_data$Fathers.Occupation <- factor(father_occupation, labels = c("Farmer", "Driver", "Others"))
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.

```
subset(household_data, Sex == 2 & Fathers.Occupation == "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.

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

```
## 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.

The bar graph displays the Sentiments of Tweets per Day over a series of dates wherein the bars represent the count of tweets with three different sentiments, namely: Negative (red), Neutral (yellow) and Positive (blue). The x-axis contains the dates July 14, 2020 to July 21, 2020 while the y-axis contains the number of tweets with the particular sentiment for the corresponding date.

The sentiment trends over time show that negative sentiment (red bars) dominates most dates, consistently having the highest count compared to other sentiments, with particularly large spikes on July 15 and 20, 2020. Neutral sentiment (yellow bars) remains relatively consistent throughout the period, with moderate tweet counts that do not fluctuate as dramatically as negative or positive sentiments. Positive sentiment (blue bars) shows more variability, generally lower than negative sentiment, but with a notable increase on July 21, 2020, indicating a shift toward a more positive outlook. The spikes in negative sentiment on July 15 and 20 suggest events or discussions triggering this response, while the rise in positive tweets on July 21 may indicate a change in conversation or a reaction to events in a more favorable direction.