RWorksheet_Cahutay#3b

Mark Anton Cahutay

2024-10-02

1. Data Frame from the Table

```
#A. R codes for data frame from the table
survey_df <- 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)
)
survey_df
```

```
##    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 data consist responses from 20 individuals, detailing their gender,
#fathers' occupations, household sizes, number of siblings in school, and types of housings.
str(survey_df)
```

```
## '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. Mean of siblings attending
sibling_attendance <- mean(survey_df$Siblings_at_School)
sibling_attendance
```

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

```
sibling_attendance == 5
```

```
## [1] FALSE
```

```
#D. Extracted the 1st two rows and all the columns using the subsetting functions.
extracted_df <- survey_df[1:2,]
extracted_df
```

```
##   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. Extracted the 3rd and 5th row with 2nd and 4th column.
extracted_df2 <- survey_df[c(3,5), c(2,4)]
extracted_df2
```

```
##   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.
types_houses <- c(survey_df$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.
male_farmers <- survey_df[survey_df$Sex == 1 & survey_df$Fathers_Occupation == 1, ]
male_farmers
```

```
## [1] Respondents        Sex                Fathers_Occupation Persons_at_Home
## [5] Siblings_at_School Types_of_Houses
## <0 rows> (or 0-length row.names)
```

```
#H. Female respondents that have greater than or equal to 5 number of siblings attending school.
female_siblings <- survey_df[survey_df$Sex == 2 & survey_df$Siblings_at_School >= 5, ]
female_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 an R program to create an empty data frame.

```
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. Results Description The output indicates that an empty data frame has been created with 0 observations and 5 columns which are the Ints, Doubles, Characters, Logicals, and Factors, each defined by its respective data type but without any values.

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

```
#A. Imported the csv file into the R environment
household <- read.csv("HouseholdData.csv")
household
```

```
##    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. Converted the Sex into factor using factor() function
household$Sex <- as.integer(
                 factor(household$Sex,
                 levels = c("Male", "Female"),
                 labels = c(1, 2))
                 )
household
```

```
##    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. Converted the Type of Houses into factor and changed it into integer.
household$Types_of_Houses <- as.integer(
                             factor(household$Types_of_Houses,
                             levels = c("Wood", "Congrete", "Semi-Congrete"),
                             labels = c(1, 2, 3))
                             )
household
```

```
##    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.
household$Fathers_Occupation <- as.character(
```

```
                                  factor(household$Fathers_Occupation,
                                  levels = c(1,2,3),
                                  labels =  c("Farmer", "Driver", "Others"))
                                  )
household
```

```
##    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.
subset(household[,c(2:3)], Sex == 2 & Fathers_Occupation == "Driver")
```

```
##   Sex Fathers_Occupation
## 2   2             Driver
## 6   2             Driver
## 7   2             Driver
```

```
#F. Respondents that have greater than or equal to 5 number of siblings attending school.
subset(household[,c(1,5)], Siblings_at_School >= 5)
```

```
##   Respondents Siblings_at_School
## 4           4                  5
## 9           9                  6
```

4. Interpret the graph

- The bar graph illustrates sentiments of tweets per day from July 14 to July 21, 2020, which were categorized into negative, positive, and neutral. Throughout the period, Negative tweets were more prevalent especially on July 15 and 21 where both reached over 4000. On the other hand, the Positive and Negative category remained steady throughout the period. Overall, the data highlights that negative tweets were more dominant while positive and neutral maintained balance.