RWorksheet_Alpanghe#3B

Jersey Gabriel E. Alpanghe

2024-10-02

1. Create a data frame using the table below.

a. Write the codes.

```
df <- data.frame(
    Respondent = c(1, 2, 3, 4, 5),
    Sex = c("Male", "Female", "Male", "Female", "Female"),
    Father_Occupation = c("Farmer", "Driver", "Others", "Farmer", "Others"),
    Siblings_Attending = c(5, 2, 3, 6, 4),
    House_Type = c("Wood", "Concrete", "Wood", "Concrete", "Semi-Concrete")
)
```

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

```
str(df)
```

```
## 'data.frame':    5 obs. of  5 variables:
##  $ Respondent        : num  1 2 3 4 5
##  $ Sex               : chr  "Male" "Female" "Male" "Female" ...
##  $ Father_Occupation : chr  "Farmer" "Driver" "Others" "Farmer" ...
##  $ Siblings_Attending: num  5 2 3 6 4
##  $ House_Type        : chr  "Wood" "Concrete" "Wood" "Concrete" ...
```

```
summary(df)
```

```
##    Respondent     Sex            Father_Occupation  Siblings_Attending
##  Min.   :1    Length:5           Length:5           Min.   :2
##  1st Qu.:2    Class :character   Class :character   1st Qu.:3
##  Median :3    Mode  :character   Mode  :character   Median :4
##  Mean   :3                                          Mean   :4
##  3rd Qu.:4                                          3rd Qu.:5
##  Max.   :5                                          Max.   :6
##   House_Type
##  Length:5
##  Class :character
##  Mode  :character
##
##
##
```

Figure 1: R Chunk

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

```
mean(df$Siblings_Attending) == 5
```

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

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

```
df[1:2, ]
```

```
##   Respondent    Sex Father_Occupation Siblings_Attending House_Type
## 1          1   Male            Farmer                  5       Wood
## 2          2 Female            Driver                  2   Concrete
```

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

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

```
##      Sex Siblings_Attending
## 3   Male                  3
## 5 Female                  4
```

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

```
types_houses <- df$House_Type
types_houses
```

```
## [1] "Wood"          "Concrete"      "Wood"          "Concrete"
## [5] "Semi-Concrete"
```

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

```
subset(df, Sex == "Male" & Father_Occupation == "Farmer")
```

```
##   Respondent  Sex Father_Occupation Siblings_Attending House_Type
## 1          1 Male            Farmer                  5       Wood
```

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(df, Sex == "Female" & Siblings_Attending >= 5)
```

```
##   Respondent    Sex Father_Occupation Siblings_Attending House_Type
## 4          4 Female            Farmer                  6   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. Output: [1] "Structure of the empty dataframe:" 'data.frame': 0 obs. of 5 variables: $ Ints : int $ Doubles : num $ Characters: chr $ Logicals : logi $ Factors : Factor w/ 0 levels: NULL

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

Figure 2: Figure 2: Sentiment Analysis

a. Import the csv file into the R environment. Write the codes.

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

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$Sex <- factor(household_data$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. 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.

```
household_data$House_Type <- factor(household_data$Types.of.Houses, levels = c("Wood", "Concrete", "Sem
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 House_Type
```

```
## 1             Wood          1
## 2         Congrete       <NA>
## 3         Congrete       <NA>
## 4             Wood          1
## 5    Semi-Congrete       <NA>
## 6    Semi-Congrete       <NA>
## 7             Wood          1
## 8    Semi-Congrete       <NA>
## 9    Semi-Congrete       <NA>
## 10        Congrete       <NA>
```

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

```
household_data$Father_Occupation <- factor(household_data$Fathers.Occupation, levels = c("Farmer", "Dri
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 House_Type Father_Occupation
## 1             Wood          1              <NA>
## 2         Congrete       <NA>              <NA>
## 3         Congrete       <NA>              <NA>
## 4             Wood          1              <NA>
## 5    Semi-Congrete       <NA>              <NA>
## 6    Semi-Congrete       <NA>              <NA>
## 7             Wood          1              <NA>
## 8    Semi-Congrete       <NA>              <NA>
## 9    Semi-Congrete       <NA>              <NA>
## 10        Congrete       <NA>              <NA>
```

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 == 2)
```

```
##   Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 2           2   2                  2              7                  3
## 6           6   2                  2              4                  3
## 7           7   2                  2              4                  1
##   Types.of.Houses House_Type Father_Occupation
## 2        Congrete       <NA>              <NA>
## 6   Semi-Congrete       <NA>              <NA>
## 7            Wood          1              <NA>
```

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

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

```
##   Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 4           4   1                  3              8                  5
## 9           9   2                  1             11                  6
##   Types.of.Houses House_Type Father_Occupation
## 4            Wood          1              <NA>
## 9   Semi-Congrete       <NA>              <NA>
```

4. Interpret the graph.

Graph Title: "Sentiments Of Tweets Per Day" Sentiment Analysis on the basis of tweets for a few days (July 14, 2020-July 21, 2020). The sentiment has been classified into three types: Negative (red), Neutral (blue), and Positive (orange).

July 14, 2020: Majority of the tweets are negative in nature. Neutral sentiments occupy a slightly higher percentage The minimum percentage of positive sentiments

July 15, 2020: There is more than 4000 of negativity with tweets, that is the highest number during the picked period. Next comes the neutral ones, which represent the least of positivity.

July 17, 2020: Positive and negative have a corresponding number with themselves where both equal each other Neutral exceeds the number of negative ones, whereas the positivity ranks as the least.

July 18, 2020: Negative corresponds to the highest number The count of neutral is lower than that of negative and also positivity is the least.

July 20, 2020: Bad ones again lead the number, this time by neutral ones. Positive ones stay within low range.

July 21, 2020: Bad ones are again out of the pack and this time nigh to 4000. Neutral and good ones have nearly the same number with neutral being slightly ahead.

Overall: In the entire observed period, bad ones always lead the pack and have the highest number, thus indicating that negativity leads the ways in conversations on these dates.Neutral tweets always appear as the second-highest and positive as the lowest frequency across all dates. The spikes of negative sentiment are especially very high on July 15 and July 21.

This pattern may reflect some events or subjects, which provoked the strongest felt emotion-bias toward negative sentiments, specially during this period.