RWORKSHEET_GENTAPAO#3B

BEA JULIETTE L. GENTAPAO

2025-10-13

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

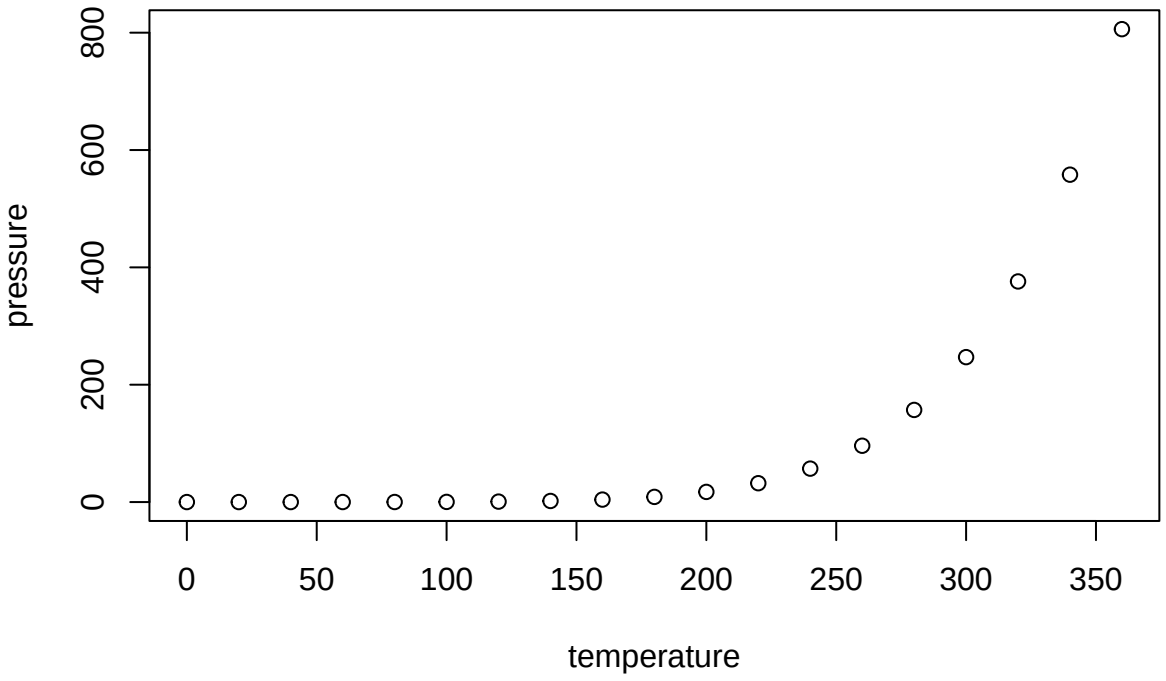
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Data Frames and Data Manipulation

1. Create a data frame from the household table

a. Write the codes

```
# Create the household data frame
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, 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)
)

household_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   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 summary

```
# Get the structure of the data
cat("Structure of the data:\n")
```

```
## Structure of the data:
```

```
str(household_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 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 ...
```

```
cat("\n\nSummary of the data:\n")
```

```
##
##
## Summary of the data:
```

```
summary(household_data)
```

```
##   Respondents         Sex       Fathers_Occupation Persons_at_Home
##  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
##  Siblings_at_School Types_of_Houses
##  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
```

Description: The data frame contains 20 observations (respondents) and 6 variables. The variables include respondent number, sex (coded as integers), father's occupation (coded as integers), number of persons at home, number of siblings attending school, and type of house (coded as integers). All variables are numeric/integer type. The data represents household information from a survey.

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

```
mean_siblings <- mean(household_data$Siblings_at_School)
cat("Mean number of siblings attending school:", mean_siblings, "\n")
```

```
## Mean number of siblings attending school: 2.95
```

```
if (mean_siblings == 5) {
  cat("Yes, the mean is 5.")
} else {
  cat("No, the mean is not 5. It is", mean_siblings)
}
```

```
## No, the mean is not 5. It is 2.95
```

d. Extract the 1st two rows and all columns

```
first_two_rows <- household_data[1:2, ]
first_two_rows
```

```
##   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

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

```
##   Sex Persons_at_Home
## 3   1               3
## 5   2               5
```

f. Select the variable types of houses and store as types_houses

```
types_houses <- household_data$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 respondents whose father occupation was farmer

```
# Male = 1, Farmer = 1
male_farmers <- household_data[household_data$Sex == 1 &
                               household_data$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. Select only all females with >= 5 siblings attending school

```
# Female = 2, Siblings >= 5
female_five_siblings <- household_data[household_data$Sex == 2 &
                                       household_data$Siblings_at_School >= 5, ]
female_five_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. 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
```

Description of results: The output shows an empty data frame with 0 observations (rows) but with 5 defined variables (columns). Each column has a specific data type: `Ints` for integers, `Doubles` for numeric decimal values, `Characters` for text strings, `Logicals` for TRUE/FALSE values, and `Factors` for categorical data. Although the structure is defined, there is no actual data stored in the data frame yet.

3. Working with CSV files

Create and save HouseholdData.csv

```
# Create the household data
household_csv_data <- data.frame(
  Respondents = c(1:20),
  Sex = c("Male", "Female", "Female", "Male", "Male", "Male", "Female",
          "Female", "Male", "Female", "Male", "Female", "Male", "Female",
          "Male", "Male", "Male", "Female", "Male", "Female"),
  Fathers_Occupation = c("Farmer", "Others", "Others", "Others", "Farmer",
                         "Driver", "Others", "Farmer", "Farmer", "Farmer",
                         "Others", "Driver", "Farmer", "Others", "Others",
                         "Farmer", "Others", "Farmer", "Driver", "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),
  Types_of_Houses = c("Wood", "Congrete", "Semi-Congrete", "Wood", "Wood",
                      "Semi-Congrete", "Semi-Congrete", "Wood", "Congrete",
                      "Semi-Congrete", "Congrete", "Semi-Congrete", "Congrete",
                      "Congrete", "Semi-Congrete", "Semi-Congrete", "Semi-Congrete",
                      "Semi-Congrete", "Semi-Congrete", "Congrete")
)

# Save to CSV
write.csv(household_csv_data, file = "HouseholdData.csv", row.names = FALSE)
cat("CSV file created successfully!\n")
```

```
## CSV file created successfully!
```

a. Import the CSV file into R environment

```
household_imported <- read.csv("HouseholdData.csv")
head(household_imported)
```

```
##   Respondents    Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1           1   Male             Farmer               5                  6
## 2           2 Female             Others               7                  4
## 3           3 Female             Others               3                  4
## 4           4   Male             Others               8                  1
## 5           5   Male             Farmer               5                  2
## 6           6   Male             Driver               9                  1
##   Types_of_Houses
## 1            Wood
## 2        Congrete
## 3   Semi-Congrete
## 4            Wood
## 5            Wood
## 6   Semi-Congrete
```

b. Convert Sex into factor and change to integer

```
household_imported$Sex <- factor(household_imported$Sex,
                                 levels = c("Male", "Female"),
                                 labels = c(1, 2))
```

```
household_imported$Sex <- as.integer(household_imported$Sex)

cat("Sex converted to integer:\n")
```

```
## Sex converted to integer:
```

```
head(household_imported)
```

```
##   Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1           1   1             Farmer               5                  6
## 2           2   2             Others               7                  4
## 3           3   2             Others               3                  4
## 4           4   1             Others               8                  1
## 5           5   1             Farmer               5                  2
## 6           6   1             Driver               9                  1
##   Types_of_Houses
## 1            Wood
## 2        Congrete
## 3   Semi-Congrete
## 4            Wood
## 5            Wood
## 6   Semi-Congrete
```

c. Convert Type of Houses into factor and change to integer

```
household_imported$Types_of_Houses <- factor(household_imported$Types_of_Houses,
                                             levels = c("Wood", "Congrete",
                                                        "Semi-Congrete"),
                                             labels = c(1, 2, 3))
household_imported$Types_of_Houses <- as.integer(household_imported$Types_of_Houses)

cat("Types of Houses converted to integer:\n")
```

```
## Types of Houses converted to integer:
```

```
head(household_imported)
```

```
##   Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1           1   1             Farmer               5                  6
## 2           2   2             Others               7                  4
## 3           3   2             Others               3                  4
## 4           4   1             Others               8                  1
## 5           5   1             Farmer               5                  2
## 6           6   1             Driver               9                  1
##   Types_of_Houses
## 1               1
## 2               2
## 3               3
## 4               1
## 5               1
## 6               3
```

d. Convert Father's Occupation into factor

```
household_imported$Fathers_Occupation <- factor(household_imported$Fathers_Occupation,
                                                levels = c("Farmer", "Driver", "Others"),
                                                labels = c(1, 2, 3))
household_imported$Fathers_Occupation <- as.integer(household_imported$Fathers_Occupation)

cat("Father's Occupation converted to integer:\n")
```

```
## Father's Occupation converted to integer:
```

```
head(household_imported)
```

```
##   Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1           1   1                  1               5                  6
## 2           2   2                  3               7                  4
## 3           3   2                  3               3                  4
## 4           4   1                  3               8                  1
## 5           5   1                  1               5                  2
## 6           6   1                  2               9                  1
##   Types_of_Houses
## 1               1
## 2               2
## 3               3
## 4               1
## 5               1
## 6               3
```

```
str(household_imported)
```

```
## 'data.frame':    20 obs. of  6 variables:
##  $ Respondents       : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Sex               : int  1 2 2 1 1 1 2 2 1 2 ...
##  $ Fathers_Occupation: int  1 3 3 3 1 2 3 1 1 1 ...
##  $ Persons_at_Home   : int  5 7 3 8 5 9 6 7 8 4 ...
##  $ Siblings_at_School: int  6 4 4 1 2 1 5 3 1 2 ...
##  $ Types_of_Houses   : int  1 2 3 1 1 3 3 1 2 3 ...
```

e. Select only females whose father's occupation is driver

```
# Female = 2, Driver = 2
female_driver <- household_imported[household_imported$Sex == 2 &
                                    household_imported$Fathers_Occupation == 2, ]
cat("Female respondents with father as driver:\n")
```

```
## Female respondents with father as driver:
```

```
female_driver
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 12          12   2                  2               5                  2
##    Types_of_Houses
## 12               3
```

f. Select respondents with >= 5 siblings attending school

```
five_or_more_siblings <- household_imported[household_imported$Siblings_at_School >= 5, ]
cat("Respondents with 5 or more siblings attending school:\n")
```

```
## Respondents with 5 or more siblings attending school:
```

```
five_or_more_siblings
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1            1   1                  1               5                  6
## 7            7   2                  3               6                  5
## 13          13   1                  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
```

4. Interpret the Graph

```
## Graph Interpretation:
```

Interpretation of the Sentiment Analysis Graph:

The bar graph shows the distribution of sentiments (Negative, Neutral, and Positive) across different days of the week from July 14, 2020 to July 21, 2020.

Key Observations:

1. **Negative Sentiment (Red bars):** Remains relatively low and consistent throughout the week, with slight variations. The negative sentiment appears to be lowest on July 15 and slightly higher on July 21.
2. **Neutral Sentiment (Yellow/Orange bars):** Shows moderate levels throughout the week. There's a noticeable spike on July 17 and July 18, suggesting more neutral or ambivalent responses on those days.
3. **Positive Sentiment (Blue/Teal bars):** This is the dominant sentiment throughout the entire week. Positive sentiment is highest on July 15 and July 21, and relatively lower (but still dominant) on July 17 and July 18.
4. **Overall Trend:** The data suggests predominantly positive sentiment throughout the week, with neutral sentiment being the second most common. Negative sentiment remains consistently low across all days.
5. **Notable Patterns:** July 15 shows the strongest positive sentiment with very low neutral sentiment, while July 17-18 show a shift toward more neutral sentiments, possibly indicating a mid-week change in mood or topics being discussed.