

# Rworksheet\_Garcia#3b

Carol D. Garcia

2025-11-08

```
#1
#a
data <- data.frame(
  Respondents = 1:20,
  Sex = c(2, 2, 1, 2, 1, 2, 2, 2, 1, 2, 2, 1, 2, 2, 1, 3, 2, 1, 2, 1),
  Fathers_Occupation = c(1, 3, 2, 3, 1, 1, 2, 2, 1, 3, 2, 1, 1, 3, 1, 2, 3, 2, 2, 1),
  Persons_at_Home = c(5, 3, 7, 5, 7, 6, 3, 6, 8, 1, 2, 4, 5, 7, 7, 3, 5, 6, 3, 6),
  Siblings_at_School = c(6, 4, 1, 1, 6, 5, 5, 3, 4, 3, 5, 2, 2, 2, 4, 5, 2, 3, 5, 6),
  Types_of_Houses = c(1, 2, 2, 3, 1, 3, 1, 1, 2, 2, 3, 3, 2, 1, 2, 3, 2, 2, 3)
)
data

##   Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1             1   2                  1                 5                   6
## 2             2   2                  3                 3                   4
## 3             3   1                  2                 7                   1
## 4             4   2                  3                 5                   1
## 5             5   1                  1                 7                   6
## 6             6   2                  1                 6                   5
## 7             7   2                  2                 3                   5
## 8             8   2                  2                 6                   3
## 9             9   1                  1                 8                   4
## 10            10  2                  3                 1                   3
## 11            11  2                  2                 2                   5
## 12            12  1                  1                 4                   2
## 13            13  2                  1                 5                   2
## 14            14  2                  3                 7                   2
## 15            15  1                  1                 7                   4
## 16            16  3                  2                 3                   5
## 17            17  2                  3                 5                   2
## 18            18  1                  2                 6                   3
## 19            19  2                  2                 3                   5
## 20            20  1                  1                 6                   6
##   Types_of_Houses
## 1             1
## 2             2
## 3             2
## 4             3
## 5             1
## 6             3
## 7             1
## 8             1
## 9             1
```

```

## 10          2
## 11          2
## 12          3
## 13          3
## 14          2
## 15          1
## 16          2
## 17          3
## 18          2
## 19          2
## 20          3

#b
# Get the structure of the data
str(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 1 2 2 2 1 2 ...
## $ Fathers_Occupation: num 1 3 2 3 1 1 2 2 1 3 ...
## $ Persons_at_Home : num 5 3 7 5 7 6 3 6 8 1 ...
## $ Siblings_at_School: num 6 4 1 1 6 5 5 3 4 3 ...
## $ Types_of_Houses : num 1 2 2 3 1 3 1 1 1 2 ...

# Get the summary of the data
summary(data)

##   Respondents      Sex   Fathers_Occupation Persons_at_Home
##   Min.    : 1.00  Min.    :1.00    Min.    :1.00    Min.    :1.00
##   1st Qu.: 5.75  1st Qu.:1.00    1st Qu.:1.00    1st Qu.:3.00
##   Median  :10.50  Median  :2.00    Median  :2.00    Median  :5.00
##   Mean    :10.50  Mean    :1.7     Mean    :1.85    Mean    :4.95
##   3rd Qu.:15.25  3rd Qu.:2.00    3rd Qu.:2.25    3rd Qu.:6.25
##   Max.    :20.00  Max.    :3.0     Max.    :3.00    Max.    :8.00
##   Siblings_at_School Types_of_Houses
##   Min.    :1.00      Min.    :1
##   1st Qu.:2.00      1st Qu.:1
##   Median  :4.0       Median  :2
##   Mean    :3.7       Mean    :2
##   3rd Qu.:5.0       3rd Qu.:3
##   Max.    :6.0       Max.    :3

#c
# Calculate the mean of Siblings at School
mean_siblings <- mean(data$Siblings_at_School)

# Check if the mean is equal to 5
mean_siblings == 5

## [1] FALSE

#d
subset_1st_two_rows <- data[1:2, ]
print(subset_1st_two_rows)

##   Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1             1    2                      1                  5                     6

```

```

## 2          2   2           3           3           4
##   Types_of_Houses
## 1          1
## 2          2
#e
subset_3rd_5th_rows <- data[c(3, 5), c(2, 4)]
print(subset_3rd_5th_rows)

##   Sex Persons_at_Home
## 3   1          7
## 5   1          7
#f
types_houses <- data$Types_of_Houses
print(types_houses)

## [1] 1 2 2 3 1 3 1 1 1 2 2 3 3 2 1 2 3 2 2 3
#g
males_farmer <- subset(data, Sex == 1 & Fathers_Occupation == 1)
print(males_farmer)

##   Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 5            5   1                   1             7                 6
## 9            9   1                   1             8                 4
## 12           12   1                   1             4                 2
## 15           15   1                   1             7                 4
## 20           20   1                   1             6                 6
##   Types_of_Houses
## 5            1
## 9            1
## 12           3
## 15           1
## 20           3
#h
females_siblings_ge5 <- subset(data, Sex == 2 & Siblings_at_School >= 5)
print(females_siblings_ge5)

##   Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1            1   2                   1             5                 6
## 6            6   2                   1             6                 5
## 7            7   2                   2             3                 5
## 11           11   2                   2             2                 5
## 19           19   2                   2             3                 5
##   Types_of_Houses
## 1            1
## 6            3
## 7            1
## 11           2
## 19           2
#2
#a
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

#explanation
#The str() function outputs the structure of the empty data frame, showing that it has no data (0 rows)

#3
#a
household_data <- data.frame(
  Respondents = 1:10,
  Sex = c("Male", "Female", "Female", "Male", "Male", "Female", "Male", "Female", "Female", "Male"),
  Fathers_Occupation = c(1, 2, 3, 3, 1, 2, 2, 3, 1, 3), # Assuming these represent occupation codes
  Persons_at_Home = c(5, 7, 3, 8, 6, 4, 2, 2, 11, 6),
  Siblings_at_School = c(2, 3, 0, 5, 2, 3, 1, 2, 6, 2),
  Types_of_Houses = c("Wood", "Concrete", "Concrete", "Wood", "Wood", "Semi-concrete", "Semi-concrete",
                      "Wood", "Semi-concrete", "Concrete")
)
write.csv(household_data, file = "HouseholdData.csv", row.names = FALSE)
print(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     Male                 2                 2                   1
## 8             8   Female                 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           Wood
## 6  Semi-concrete
## 7  Semi-concrete

```

```

## 8          Wood
## 9      Semi-concrete
## 10     Concrete

imported_data <- read.csv("HouseholdData.csv")
print(imported_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 Male                  2                 2                      1
## 8             8 Female                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          Wood
## 6  Semi-congcrete
## 7  Semi-congcrete
## 8          Wood
## 9  Semi-concrete
## 10     Concrete

#b
imported_data$Sex <- factor(imported_data$Sex, levels = c("Male", "Female"), labels = c(1, 2))
print(imported_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   1                  2                 2                      1
## 8             8   2                  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          Wood
## 6  Semi-congcrete
## 7  Semi-congcrete
## 8          Wood
## 9  Semi-concrete

```

```

## 10      Concrete

#c
imported_data$Types_of_Houses <- factor(imported_data$Types_of_Houses,
                                         levels = c("Wood", "Concrete", "Semi-concrete"),
                                         labels = c(1, 2, 3))
print(imported_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   1                 2             2                  1
## 8           8   2                 3             2                  2
## 9           9   2                 1            11                  6
## 10          10  1                 3             6                  2

##   Types_of_Houses
## 1           1
## 2           <NA>
## 3           <NA>
## 4           1
## 5           1
## 6           <NA>
## 7           <NA>
## 8           1
## 9           3
## 10          2

#d
imported_data$Fathers_Occupation <- factor(imported_data$Fathers_Occupation,
                                              levels = c(1, 2, 3),
                                              labels = c("Farmer", "Driver", "Others"))
print(imported_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   1             Driver            2                  1
## 8           8   2             Others            2                  2
## 9           9   2             Farmer            11                 6
## 10          10  1             Others            6                  2

##   Types_of_Houses
## 1           1
## 2           <NA>
## 3           <NA>
## 4           1
## 5           1
## 6           <NA>
## 7           <NA>

```

```

## 8          1
## 9          3
## 10         2

#e
females_driver <- subset(imported_data, Sex == 2 & Fathers_Occupation == "Driver")
print(females_driver)

##   Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 2            2    2             Driver           7              3
## 6            6    2             Driver           4              3
##   Types_of_Houses
## 2            <NA>
## 6            <NA>

#f
siblings_ge5 <- subset(imported_data, Siblings_at_School >= 5)
print(siblings_ge5)

##   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
Date <- c(
"July 14", "July 14", "July 14",
"July 15", "July 15", "July 15",
"July 17", "July 17", "July 17",
"July 18", "July 18", "July 18",
"July 20", "July 20", "July 20",
"July 21", "July 21", "July 21"
)
Sentiment <- rep(c("Negative", "Neutral", "Positive"), times = 6)
Count <- c(
2400, 1600, 1700,
3800, 2900, 3200,
3300, 1700, 2500,
3300, 2000, 2600,
2200, 1400, 1600,
3700, 2800, 3400
)
sentiment_data <- data.frame(Date, Sentiment, Count)
sentiment_data

##      Date Sentiment Count
## 1 July 14 Negative  2400
## 2 July 14 Neutral   1600
## 3 July 14 Positive  1700
## 4 July 15 Negative  3800
## 5 July 15 Neutral   2900
## 6 July 15 Positive  3200
## 7 July 17 Negative  3300

```

```
## 8 July 17 Neutral 1700
## 9 July 17 Positive 2500
## 10 July 18 Negative 3300
## 11 July 18 Neutral 2000
## 12 July 18 Positive 2600
## 13 July 20 Negative 2200
## 14 July 20 Neutral 1400
## 15 July 20 Positive 1600
## 16 July 21 Negative 3700
## 17 July 21 Neutral 2800
## 18 July 21 Positive 3400
```

*#The graph shows the number of tweets classified into negative, neutral, and positive sentiments accross .*