RWorksheet_Pineda#3b

Epiphany Louise O. Pineda

2025-10-13

```r
#(1.) Create a data frame using the table below.

#a. Write the codes.
data <- data.frame(
  Respondents = 1:20,
  Sex = c(2, 2, 1, 2, 2, 2, 1, 2, 1, 2,
          1, 2, 2, 2, 2, 2, 2, 1, 1, 2),
  Fathers_Occupation = c(1, 3, 1, 3, 2, 2, 3, 1, 1, 1,
                         1, 2, 1, 1, 3, 3, 3, 3, 2, 1),
  Persons_at_Home = c(5, 7, 3, 8, 5, 9, 6, 7, 8, 4,
                      7, 5, 4, 5, 8, 2, 3, 11, 7, 6),
  Siblings_at_School = c(6, 4, 1, 1, 2, 1, 5, 3, 1, 2,
                         3, 2, 5, 1, 2, 2, 2, 5, 3, 2),
  Types_of_Houses = c(1, 2, 3, 1, 1, 3, 3, 1, 2, 3,
                      3, 3, 2, 3, 3, 3, 3, 3, 3, 2)
)

#b. Describe the data. Get the structure or the summary of the data
str(data)
#There are 20 respondents (rows).
#Each variable is numeric (integer type).

#c. Is the mean number of siblings attending is 5?
mean_siblings <- mean(data$Siblings_at_School)
mean_siblings
mean_siblings == 5
#Output: [1] 2.8
#        [1] FALSE
#No, the mean number of siblings attending school is 2.8, not 5.

#d. Extract the 1st two rows and then all the columns using the subsetting functions.
#Write the codes and its output.
first_two_rows <- data[1:2, ]
first_two_rows
#Output:
#   Respondents  Sex   Fathers_Occupation Persons_at_Home  Siblings_at_School  Types_of_Houses
#1           1    2                   1               5                   6                1
#2           2    2                   3               7                   4                2

#e. Extract 3rd and 5th row with 2nd and 4th column. Write the codes and its result.
subset_rows_cols <- data[c(3,5), c(2,4)]
subset_rows_cols
```

```
#Output:
#   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.
#Write the codes.
types_houses <- data$Types_of_Houses
types_houses
#Output:[1] 1 2 3 1 1 3 3 1 2 3 3 3 2 3 3 3 3 3 3 2

#g. Select only all Males respondent that their father occupation was farmer. Write the codes and its o
male_farmer <- subset(data, Sex == 1 & Fathers_Occupation == 1)
male_farmer
#OutPut:
#    Respondents Sex  Fathers_Occupation  Persons_at_Home  Siblings_at_School  Types_of_Houses
#3           3    1                 1              3                  1                3
#9           9    1                 1              8                  1                2
#11         11    1                 1              7                  3                3

#h. Select only all females respondent that have greater than or equal to 5 number of siblings attendin
female_5siblings <- subset(data, Sex == 2 & Siblings_at_School >= 5)
female_5siblings
#Output:
#     Respondents Sex  Fathers_Occupation  Persons_at_Home  Siblings_at_School Types_of_Houses
#1           1     2                 1              5                  6                1
#13         13     2                 1              4                  5                2



#(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:")
#print(str(df))

#a. Describe the results.
#[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:



#(3.) Create a .csv file of this. Save it as HouseholdData.csv
```

```r
#a. Import the csv file into the R environment. Write the codes.
HouseholdData <- read.csv("HouseholdData.csv")

print(HouseholdData)

#b. Convert the Sex into factor using factor() function and change it into integer.[Legend: Male = 1 an
HouseholdData$Sex <- factor(HouseholdData$Sex,
                            levels = c("Male", "Female"),
                            labels = c(1, 2))

HouseholdData$Sex
#Output:[1] 1 2 2 1 1 2 2 1 2 1
#Levels: 1 2

#c. Convert the Type of Houses into factor and change it into integer. [Legend: Wood = 1; Congrete = 2;
HouseholdData$Types.of.Houses <- factor(HouseholdData$Types.of.Houses,
                                        levels = c("Wood", "Concrete", "Semi-concrete"),
                                        labels = c(1, 2, 3))

HouseholdData$Types.of.Houses
#Output: [1] 1 2 2 1 3 3 1 3 3 2
#Levels: 1 2 3

#d. On father's occupation, factor it as Farmer = 1; Driver = 2; and Others = 3. What is the R code and
HouseholdData$Fathers.Occupation <- factor(HouseholdData$Fathers.Occupation,
                                           levels = c(1, 2, 3),
                                           labels = c("Farmer", "Driver", "Others"))

HouseholdData$Fathers.Occupation
#Output: [1] Farmer Driver Others Others Farmer Driver Farmer Others Farmer Others
#Levels: Farmer Driver Others

#e. Select only all females respondent that has a father whose occupation is driver. Write the codes an
female_driver <- subset(HouseholdData, Sex == 2 & Fathers.Occupation == "Driver")
female_driver
#Output:
#  Respondents Sex  Fathers.Occupation  Persons.at.Home  Siblings.at.School  Types.of.Houses
#2           2     2             Driver                7                   3                2
#6           6     2             Driver                4                   3                3

#f. Select the respondents that have greater than or equal to 5 number of siblings attending school. Wr
siblings_5up <- subset(HouseholdData, Siblings.at.School >= 5)
siblings_5up
#Output:
#   Respondents Sex Fathers.Occupation  Persons.at.Home  Siblings.at.School  Types.of.Houses
#4           4     1            Others                8                   5                1
#9           9     2            Farmer               11                   6                3



#(4.) Interpret the graph.
#The graph shows how people felt on Twitter each day from July 14 to July 21, 2020.
#The red bars (negative) are the tallest most of the time, meaning most tweets had negative feelings.
```

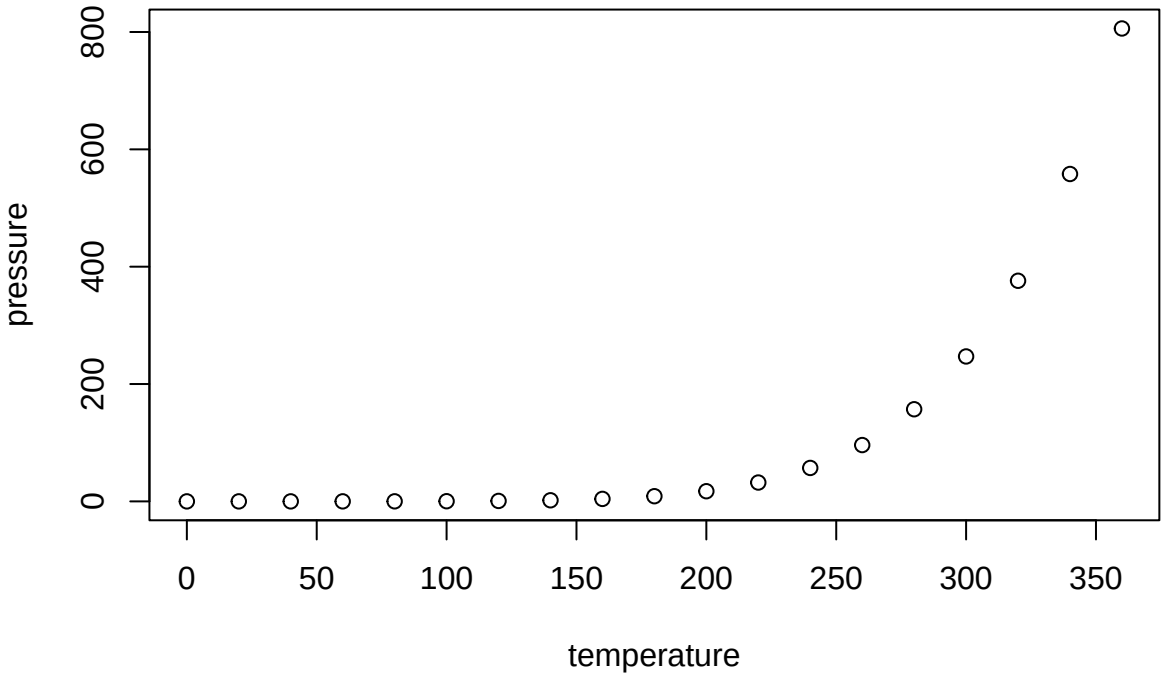
```
#The orange bars (neutral) are in the middle, showing some tweets were neither happy nor sad.
#The blue bars (positive) are usually the shortest, meaning few tweets were happy or positive.
#The highest negative tweets happened on July 15 and July 21, which means people were more upset or unh

#In short: Most people tweeted negative thoughts, fewer were neutral, and the least were positive durin
```

```
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.