RWorksheet_Gallenero#3b

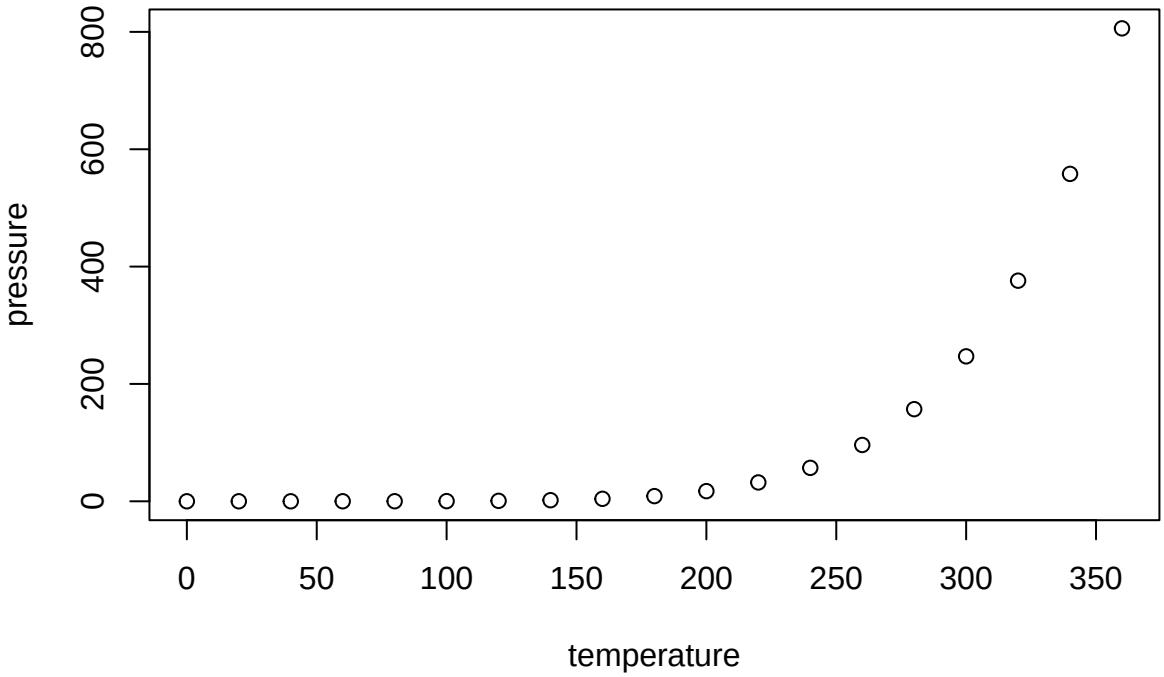
2023-10-11

```
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:




```
#1.a Create a data frame
respondents_number <- 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)
father_occ <- c(1,3,3,3,1,2,3,1,1,1,3,2,1,3,3,1,3,1,2,1)
person_home <- c(5,7,3,8,5,9,6,7,8,4,7,5,4,7,8,8,3,11,7,6)
sibs_schl <- c(6,4,4,1,2,1,5,3,1,2,3,2,5,5,2,1,2,5,3,2)
house_type<- c(1,2,3,1,1,3,3,1,2,3,2,3,2,2,3,3,3,3,3,2)


Household_data <- data.frame(
```

```
  Respondents = respondents_number,
  Sex = sex,
  FatherOccupation = father_occ,
  PersonAtHome = person_home,
  SiblingsAtSchool = sibs_schl,
  TypesOfHouses = house_type
)
Household_data
```

```
##    Respondents Sex FatherOccupation PersonAtHome SiblingsAtSchool TypesOfHouses
## 1            1   2                1            5                6             1
## 2            2   2                3            7                4             2
## 3            3   1                3            3                4             3
## 4            4   2                3            8                1             1
## 5            5   2                1            5                2             1
## 6            6   2                2            9                1             3
## 7            7   2                3            6                5             3
## 8            8   2                1            7                3             1
## 9            9   2                1            8                1             2
## 10          10   2                1            4                2             3
## 11          11   1                3            7                3             2
## 12          12   2                2            5                2             3
## 13          13   2                1            4                5             2
## 14          14   2                3            7                5             2
## 15          15   2                3            8                2             3
## 16          16   2                1            8                1             3
## 17          17   2                3            3                2             3
## 18          18   2                1           11                5             3
## 19          19   1                2            7                3             3
## 20          20   2                1            6                2             2
```

```
#1.b

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 ...
##  $ FatherOccupation: num  1 3 3 3 1 2 3 1 1 1 ...
##  $ PersonAtHome    : num  5 7 3 8 5 9 6 7 8 4 ...
##  $ SiblingsAtSchool: num  6 4 4 1 2 1 5 3 1 2 ...
##  $ TypesOfHouses   : num  1 2 3 1 1 3 3 1 2 3 ...
```

```
summary(Household_data)
```

```
##   Respondents         Sex       FatherOccupation  PersonAtHome
##  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
##  SiblingsAtSchool TypesOfHouses
##  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
```

```
# There are 20 observations (rows) and 6 variables (columns) in the data frame.
# the data frame has 6 variables (columns) and 20 observations (rows).


# respondents - a numeric identifier for each respondent
# sex - (1 for male, 2 for female) indicates the respondent's gender.
# occupation of father - shows the occupation of the father (1 for farmer, 2 for driver, 3 for others).
# persons at home - displays the number of people at home.
# siblings at school - the number of siblings at the school.
# type of house - defines the type of house (1 for wood, 2 for semi-concrete, 3 for concrete).

#1.c
sibs_schl_mean <- mean(Household_data$SiblingsAtSchool)
sibs_schl_mean
```

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

```
# no, it is not 5 because the mean of siblings at school is 2.95

#1.d
first_Two_rows <- Household_data[1:2,]
first_Two_rows
```

```
##   Respondents Sex FatherOccupation PersonAtHome SiblingsAtSchool TypesOfHouses
## 1           1   2                1            5                6             1
## 2           2   2                3            7                4             2
```

```
#1.e
extract_rows <- Household_data[c(3,5),c(2,4)]
extract_rows
```

```
##   Sex PersonAtHome
## 3   1            3
## 5   2            5
```

```
#1.f
types_Houses <- Household_data$TypesOfHouses
types_Houses
```

```
##  [1] 1 2 3 1 1 3 3 1 2 3 2 3 2 2 3 3 3 3 3 2
```

```
#1.g
male_FatherFarmer <- Household_data[Household_data$Sex == 1 & Household_data$FatherOccupation == 1,]

male_FatherFarmer
```

```
## [1] Respondents      Sex              FatherOccupation PersonAtHome
## [5] SiblingsAtSchool TypesOfHouses
## <0 rows> (or 0-length row.names)
```

```
#1.h

female_Respondent <- Household_data[Household_data$SiblingsAtSchool >= 5,]
female_Respondent
```

```
##    Respondents Sex FatherOccupation PersonAtHome SiblingsAtSchool TypesOfHouses
## 1            1   2                1            5                6             1
## 7            7   2                3            6                5             3
## 13          13   2                1            4                5             2
## 14          14   2                3            7                5             2
## 18          18   2                1           11                5             3
```

```
# 2
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
```

```
# df have a empty data frame with 0 rows and 5 columns
# the columns has the following data type:
# ints = integer
# doubles = double
# characters = character
# logicals = logical
# factors  = factor
#(0 levels which means empty)

# can be serve as a template that can be populated with data


# 3

new_respondents <- c(1,2,3,4,5,6,7,8,9,10)
new_sex <- c("Male", "Female", "Female", "Male", "Male", "Female", "Female", "Male", "Female", "Male")
new_fatherocc <- c(1,2,3,3,1,2,2,3,1,3)
new_persons <- c(5,7,3,8,6,4,4,2,11,6)
new_sibs <- c(2,3,0,5,2,3,1,2,6,2)
new_type <- c("Wood", "Congrete", "Congrete", "Wood", "Semi-congrete", "Semi-congrete", "Wood", "Semi-c

Household_Data <- data.frame(
  Respondents = new_respondents,
  Sex = new_sex,
  FatherOccupation = new_fatherocc,
```

```
  PersonsAtHome = new_persons,
  SiblingsAtSchool = new_sibs,
  HouseType = new_type
)
Household_Data
```

```
##    Respondents    Sex FatherOccupation PersonsAtHome SiblingsAtSchool
## 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
##        HouseType
## 1           Wood
## 2       Congrete
## 3       Congrete
## 4           Wood
## 5  Semi-congrete
## 6  Semi-congrete
## 7           Wood
## 8  Semi-congrete
## 9  Semi-congrete
## 10      Congrete
```

```
write.csv(Household_Data, file = "HouseholdData.csv")

#3a

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

```
##     X Respondents    Sex FatherOccupation PersonsAtHome SiblingsAtSchool
## 1   1           1   Male                1             5                2
## 2   2           2 Female                2             7                3
## 3   3           3 Female                3             3                0
## 4   4           4   Male                3             8                5
## 5   5           5   Male                1             6                2
## 6   6           6 Female                2             4                3
## 7   7           7 Female                2             4                1
## 8   8           8   Male                3             2                2
## 9   9           9 Female                1            11                6
## 10 10          10   Male                3             6                2
##        HouseType
## 1           Wood
## 2       Congrete
## 3       Congrete
## 4           Wood
## 5  Semi-congrete
## 6  Semi-congrete
```

```
## 7           Wood
## 8  Semi-congrete
## 9  Semi-congrete
## 10      Congrete
```

```r
#3b

imported$Sex <- factor(imported$Sex, levels = c("Male", "Female"))
imported$Sex <- as.integer(imported$Sex)

imported$Sex
```

```
##  [1] 1 2 2 1 1 2 2 1 2 1
```

```r
#output: 1 2 2 1 1 2 2 1 2 1

#3c

imported$HouseType <- factor(imported$HouseType, levels = c("Wood", "Congrete", "Semi-congrete"))
imported$HouseType <- as.integer(imported$HouseType)

imported$HouseType
```

```
##  [1] 1 2 2 1 3 3 1 3 3 2
```

```r
#Output: 1 2 2 1 3 3 1 3 3 2

#3d

imported$FatherOccupation <- factor(imported$FatherOccupation, levels = c(1,2,3), labels = c("Farmer",

imported$FatherOccupation
```

```
##  [1] Farmer Driver Others Others Farmer Driver Driver Others Farmer Others
## Levels: Farmer Driver Others
```

```r
#3e

femaleDriver <- imported[imported$Sex == 2 & imported$FatherOccupation == "Driver", c (3,4)]
femaleDriver
```

```
##   Sex FatherOccupation
## 2   2           Driver
## 6   2           Driver
## 7   2           Driver
```

```r
#3f

greater_five <- imported[imported$SiblingsAtSchool >= 5,]
greater_five
```

```
##   X Respondents Sex FatherOccupation PersonsAtHome SiblingsAtSchool HouseType
## 4 4           4   1           Others             8                5         1
## 9 9           9   2           Farmer            11                6         3
```

```r
#4
# On this day, July 14, negative sentiments surpass all other sentiments. This signifies that some topi

# On this day, July 15 the negative sentiment is still at the highest even if all the sentiments increa
```

```
# On these days, July 17 and 18, negative sentiments are still high, whereas neutral and positive senti
# On July 20, all sentiments reached their lowest point, despite the fact that there were still more ne
# On July 21, all sentiments increase, with the negative remaining at the top. This could imply that so
# Based on this data, we may conclude that public sentiment is prone to external forces and varies over
```