

## RWorksheet\_Talaban#3b

```
#1a
Respondent <- 1:20
Sex <- c(2,2,1,2,2,2,2,2,2,1,2,2,2,2,2,2,1,2)
Fathers_Occ <- c(1,3,3,3,1,2,3,1,1,1,3,2,1,3,3,1,3,1,2,1)
PersonsAtHome <- c(5,7,3,8,5,9,6,7,8,4,7,5,4,7,8,8,3,11,7,6)
SiblingsAtSchool <- c(6,4,4,1,2,1,5,3,1,2,3,2,5,5,2,1,2,5,3,2)
TypesOfHouses <- c(1,2,3,1,1,3,3,1,2,3,2,3,2,2,3,3,3,3,3,2)

df <- data.frame(Respondent, Sex, Fathers_Occ, PersonsAtHome, SiblingsAtSchool, TypesOfHouses)

##   Respondent Sex Fathers_Occ PersonsAtHome 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

#1b
str(df)

## 'data.frame': 20 obs. of  6 variables:
## $ Respondent : int  1 2 3 4 5 6 7 8 9 10 ...
## $ Sex        : num  2 2 1 2 2 2 2 2 2 2 ...
## $ Fathers_Occ : num  1 3 3 3 1 2 3 1 1 1 ...
## $ PersonsAtHome : 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 ...

#1c
mean(SiblingsAtSchool)

## [1] 2.95
```

```

#No because the mean is 2.95 only

#1d
df_subset <- df[1:2,]

df_subset

##   Respondent Sex Fathers_Occ PersonsAtHome SiblingsAtSchool TypesOfHouses
## 1           1   2           1           5           6           1
## 2           2   2           3           7           4           2

#1e
df_subset2 <- df[c(3,5), c(2,4)]

df_subset2

##   Sex PersonsAtHome
## 3   1           3
## 5   2           5

#1f
types_houses <- df$Types.of.houses

#1g
m_farmer <- df[df$Sex == 1 & df$Fathers_Occ == 1,]
m_farmer

## [1] Respondent      Sex          Fathers_Occ    PersonsAtHome
## [5] SiblingsAtSchool TypesOfHouses
## <0 rows> (or 0-length row.names)

#1h
f_siblings <- df[df$Sex == 2 & df$SiblingsAtSchool >= 5,]
f_siblings

##   Respondent Sex Fathers_Occ PersonsAtHome 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

#2a
#There are 0 observations of 5 variables. The result is NULL.

#3a
household <- read.csv("HouseholdData.csv")
print(household)

##      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            4                 8                  5
## 5           5 Male            5                 6                  2
## 6           6 Female          6                 2                 3
## 7           7 Female          7                 2                 1
## 8           8 Male            8                 3                 2
## 9           9 Female          9                 1                11                 6
## 10          10 Male           10                3                 6                  2
##      Types.of.Houses
## 1           Wood
## 2           Concrete
## 3           Concrete
## 4           Wood
## 5           Semi-concrete
## 6           Semi-concrete
## 7           Wood
## 8           Semi-concrete
## 9           Semi-concrete
## 10          Concrete

#3b
household$Sex <- factor(household$Sex, levels = c("Male", "Female"))
household$SexInt <- as.integer(household$Sex)
print(household[, c("Sex", "SexInt")])

##      Sex SexInt
## 1   Male     1
## 2 Female     2
## 3 Female     2
## 4   Male     1
## 5   Male     1
## 6 Female     2
## 7 Female     2
## 8   Male     1
## 9 Female     2
## 10  Male     1

#3c
household$Types.of.Houses <- factor(household$Types.of.Houses, levels = c("Wood", "Concrete", "Semi-con"))
household$Types.of.Houses.Int <- as.integer(household$Types.of.Houses)
print(household[, c("Types.of.Houses", "Types.of.Houses.Int")])

```

```

##      Types.of.Houses Types.of.Houses.Int
## 1          Wood           1
## 2      Concrete          2
## 3      Concrete          2
## 4          Wood           1
## 5  Semi-concrete         3
## 6  Semi-concrete         3
## 7          Wood           1
## 8  Semi-concrete         3
## 9  Semi-concrete         3
## 10     Concrete          2

#3d
household$Fathers.Occupation <- factor(household$Fathers.Occupation, levels = c(1, 2, 3), labels = c("Farmer", "Driver", "Other"))
household$Fathers.Occupation.Int <- as.integer(household$Fathers.Occupation)
print(household[, c("Fathers.Occupation", "Fathers.Occupation.Int")])

##      Fathers.Occupation Fathers.Occupation.Int
## 1          Farmer            1
## 2          Driver            2
## 3          Other             3
## 4          Other             3
## 5          Farmer            1
## 6          Driver            2
## 7          Driver            2
## 8          Other             3
## 9          Farmer            1
## 10         Other             3

#3e
filtered_a <- household[
  household$Sex == "Female" & household$Fathers_Occ == "Driver",]
filtered_a

## [1] Respondents      Sex          Fathers.Occupation
## [4] Persons.At.Home Siblings.At.School Types.of.Houses
## [7] SexInt          Types.of.Houses.Int  Fathers.Occupation.Int
## <0 rows> (or 0-length row.names)

#3f
filtered_b <- household[household$SiblingsAtSchool >= 5,]
filtered_b

## [1] Respondents      Sex          Fathers.Occupation
## [4] Persons.At.Home Siblings.At.School Types.of.Houses
## [7] SexInt          Types.of.Houses.Int  Fathers.Occupation.Int
## <0 rows> (or 0-length row.names)

#4
date <- c("July 14, 2020", "July 15, 2020", "July 17, 2020", "July 18, 2020", "July 20, 2020", "July 21, 2020")

negative <- c(2450, 4200, 3250, 3250, 2350, 4050)
neutral <- c(1575, 2800, 1900, 2050, 1450, 2700)
positive <- c(1725, 3200, 2400, 2525, 1700, 3375)

sentimentsoftweets_df <- data.frame(
  Date = as.Date(date, format = "%B %d, %Y"),
  Sentiment = c("negative", "neutral", "positive"))

```

```

Negative = negative,
Neutral = neutral,
Positive = positive)

print(sentimentsoftweets_df)

##           Date Negative Neutral Positive
## 1 2020-07-14      2450     1575    1725
## 2 2020-07-15      4200     2800    3200
## 3 2020-07-17      3250     1900    2400
## 4 2020-07-18      3250     2050    2525
## 5 2020-07-20      2350     1450    1700
## 6 2020-07-21      4050     2700    3375

```

*#The chart shows that tweet volumes vary by day, but negative and positive sentiments generally outnumber neutral ones.*