

Worksheet-3b in R

Michael T. Simpron

2025-10-13

##1.

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

print(df)
```

##	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	1	1	6	5
## 8	8	2	1	7	3
## 9	9	2	1	8	1
## 10	10	1	1	8	2
## 11	11	1	3	7	3
## 12	12	2	1	3	2
## 13	13	2	1	4	2
## 14	14	2	3	7	5
## 15	15	2	3	8	2
## 16	16	2	1	8	3
## 17	17	2	3	3	2
## 18	18	1	1	11	7
## 19	19	1	1	7	3
## 20	20	2	1	7	3
##	Types_of_Houses				
## 1	1				
## 2	2				
## 3	3				
## 4	1				
## 5	1				

```
## 6      3
## 7      2
## 8      1
## 9      2
## 10     3
## 11     2
## 12     1
## 13     1
## 14     2
## 15     1
## 16     2
## 17     3
## 18     3
## 19     3
## 20     3
```

```
##b)
str(df)
```

```
## '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 1 2 2 1 ...
## $ Fathers_Occupation: num 1 3 3 3 1 2 1 1 1 1 ...
## $ Persons_at_Home : num 5 7 3 8 5 9 6 7 8 8 ...
## $ Siblings_at_School: num 6 4 4 1 2 1 5 3 1 2 ...
## $ Types_of_Houses : num 1 2 3 1 1 3 2 1 2 3 ...
```

```
summary(df)
```

```
## Respondents      Sex      Fathers_Occupation Persons_at_Home
## Min.   : 1.00   Min.   :1.0   Min.   :1.00      Min.   : 3.00
## 1st Qu.: 5.75   1st Qu.:1.0   1st Qu.:1.00      1st Qu.: 5.00
## Median :10.50   Median :2.0   Median :1.00      Median : 7.00
## Mean   :10.50   Mean   :1.7   Mean   :1.75      Mean   : 6.55
## 3rd Qu.:15.25   3rd Qu.:2.0   3rd Qu.:3.00      3rd Qu.: 8.00
## Max.   :20.00   Max.   :2.0   Max.   :3.00      Max.   :11.00
## Siblings_at_School Types_of_Houses
## Min.   :1.00     Min.   :1
## 1st Qu.:2.00     1st Qu.:1
## Median :3.00     Median :2
## Mean   :3.05     Mean   :2
## 3rd Qu.:4.00     3rd Qu.:3
## Max.   :7.00     Max.   :3
```

```
##c)
mean_siblings <- mean(df$Siblings_at_School)

mean_siblings
```

```
## [1] 3.05
```

```
mean_siblings == 5
```

```
## [1] FALSE
```

```
##d)
```

```
df[1:2, ]
```

```
## 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)
```

```
df[c(3,5), c(2,4)]
```

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

```
##f)
```

```
types_houses <- df$Types_of_Houses
types_houses
```

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

```
##g)
```

```
subset(df, Sex == 1 & Fathers_Occupation == 1)
```

```
## Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 7          7  1              1              6              5
## 10         10  1              1              8              2
## 18         18  1              1             11              7
## 19         19  1              1              7              3
## Types_of_Houses
## 7          2
## 10         3
## 18         3
## 19         3
```

```
##h)
```

```
subset(df, Sex == 2 & Siblings_at_School >= 5)
```

```
## Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1          1  2              1              5              6
## 14         14  2              3              7              5
## Types_of_Houses
## 1          1
## 14         2
```

```
##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
```

```
##a)
```

*#The data frame is defined but empty - it has 0 rows and 5 columns with various data types.
#This kind of structure is useful when you want to initialize a data frame first and then add rows to it*

```
##2.
```

```
HouseholdData <- data.frame(  
  Respondents = 1:10,  
  Sex = c("Male", "Female", "Female", "Male", "Male", "Female", "Female", "Male", "Female", "Male"),  
  Fathers_Occupation = c(1, 2, 3, 3, 1, 2, 2, 1, 1, 3),  
  Persons_at_Home = c(5, 7, 3, 8, 6, 4, 4, 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", "Semi-concrete",  
    "Semi-concrete", "Wood", "Semi-concrete", "Semi-concrete", "Concrete")  
)  
  
print(HouseholdData)
```

```
##      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 Female                2             4             1  
## 8             8   Male                1             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      Semi-concrete
## 6      Semi-concrete
## 7      Wood
## 8      Semi-concrete
## 9      Semi-concrete
## 10     Concrete
```

```
write.csv(HouseholdData, file = "HouseholdData.csv", row.names = FALSE)
```

```
##a)
data <- read.csv("HouseholdData.csv")

print(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 Female      2      4      1
## 8      8  Male      1      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      Semi-concrete
## 6      Semi-concrete
## 7      Wood
## 8      Semi-concrete
## 9      Semi-concrete
## 10     Concrete
```

```
##b)
data$Sex <- factor(data$Sex, levels = c("Male", "Female"), labels = c(1, 2))
data$Sex <- as.integer(as.character(data$Sex))
print(data$Sex)
```

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

```
##c)
colnames(data)[colnames(data) == "Types_of_Houses"] <- "Types_of_Houses"
data$Types_of_Houses <- factor(data$Types_of_Houses,
                               levels = c("Wood", "Concrete", "Semi-concrete"),
                               labels = c(1, 2, 3))
data$Types_of_Houses <- as.integer(as.character(data$Types_of_Houses))
print(data$Types_of_Houses)
```

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

```
##d)
data$Fathers_Occupation <- factor(data$Fathers_Occupation,
                                  levels = c(1, 2, 3),
                                  labels = c(1, 2, 3))
data$Fathers_Occupation <- as.integer(as.character(data$Fathers_Occupation))
print(data$Fathers_Occupation)
```

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

```
##e)
female_driver <- subset(data, Sex == 2 & Fathers_Occupation == 2)
print(female_driver)
```

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

```
##f)
siblings_5_or_more <- subset(data, Siblings_at_School >= 5)
print(siblings_5_or_more)
```

```
## Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 4          4 1          3          8          5
## 9          9 2          1         11          6
## Types_of_Houses
## 4          1
## 9          3
```

```
##4.
```

```
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

```

““