

Rworksheet_sayson#3b

Adrian T. Sayson

2025-10-13

```
#1.  
#a  
data = data.frame(  
  respondents = c(1:20),  
  sex = c(2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, 1, 2),  
  fathers_occupation = c(1, 3, 3, 3, 1, 2, 3, 1, 1, 1, 3, 2, 1, 3, 3, 1, 3, 1, 2, 1),  
  persons_at_home = c(5, 3, 3, 6, 1, 9, 6, 7, 4, 1, 7, 5, 5, 7, 5, 8, 3, 6, 7, 6),  
  siblings_at_school = c(6, 4, 3, 5, 2, 1, 3, 3, 1, 2, 1, 3, 3, 5, 2, 2, 2, 5, 3, 3),  
  types_of_houses = c(1, 2, 3, 1, 1, 3, 1, 1, 2, 3, 3, 2, 3, 2, 2, 3, 2, 3, 3, 2))  
print(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	3	3	3
## 4	4	2	3	6	5
## 5	5	2	1	1	2
## 6	6	2	2	9	1
## 7	7	2	3	6	3
## 8	8	2	1	7	3
## 9	9	2	1	4	1
## 10	10	2	1	1	2
## 11	11	1	3	7	1
## 12	12	2	2	5	3
## 13	13	2	1	5	3
## 14	14	2	3	7	5
## 15	15	2	3	5	2
## 16	16	2	1	8	2
## 17	17	2	3	3	2
## 18	18	2	1	6	5
## 19	19	1	2	7	3
## 20	20	2	1	6	3
##	types_of_houses				
## 1	1				
## 2	2				
## 3	3				
## 4	1				
## 5	1				
## 6	3				
## 7	1				
## 8	1				
## 9	2				

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

```
#b
#The dataset contains information from 20 respondents, mostly female. Most fathers work in "other" occu
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 2 2 2 2 2 2 ...
## $ fathers_occupation: num 1 3 3 3 1 2 3 1 1 1 ...
## $ persons_at_home : num 5 3 3 6 1 9 6 7 4 1 ...
## $ siblings_at_school: num 6 4 3 5 2 1 3 3 1 2 ...
## $ types_of_houses : num 1 2 3 1 1 3 1 1 2 3 ...
```

```
#c
mean(data$siblings_at_school)
```

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

```
#No, the mean is 3 rounded up from 2.95.
```

```
#d
first_two_rows <- data[1:2, ]
first_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
extracted_data <- data[c(3,5), c(2,4)]
extracted_data
```

```
## sex persons_at_home
## 3      1      3
## 5      2      1
```

```
#f
types_houses <- data[, "types_of_houses"]
types_houses
```

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

```
#g
male_farmers <- subset(data, sex == 1 & fathers_occupation == 1)
print(male_farmers)
```

```
## [1] respondents      sex      fathers_occupation persons_at_home
## [5] siblings_at_school types_of_houses
## <0 rows> (or 0-length row.names)
```

```
#h
female_siblings_5plus <- subset(data, sex == 2 & siblings_at_school >= 5)
print(female_siblings_5plus)
```

```
##      respondents sex fathers_occupation persons_at_home siblings_at_school
## 1             1  2             1             5             6
## 4             4  2             3             6             5
## 14            14  2             3             7             5
## 18            18  2             1             6             5
##      types_of_houses
## 1             1
## 4             1
## 14            2
## 18            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
```

#a

#the results confirm that the program successfully created a well-structured, zero-row data frame with

#3.

```
respondents_data <- 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, 3, 1, 3),  
  Persons_at_Home = c(5, 7, 3, 8, 6, 4, 2, 2, 11, 6),  
  Siblings_at_School = c(2, 3, 0, 5, 2, 1, 1, 2, 6, 2),  
  Types_of_Houses = c("Wood", "Congrete", "Congrete", "Wood", "Semi-concrete", "Semi-concrete", "Wood",  
write.csv(respondents_data, "HouseholdData.csv", row.names = FALSE)  
respondents_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 1  
## 7 7 Female 2 2 1  
## 8 8 Male 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 Semi-concrete  
## 6 Semi-concrete  
## 7 Wood  
## 8 Semi-concrete  
## 9 Semi-concrete  
## 10 Congrete
```

#a

```
my_data <- read.csv("HouseholdData.csv")  
my_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 1  
## 7 7 Female 2 2 1  
## 8 8 Male 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      Semi-concrete
## 6      Semi-concrete
## 7           Wood
## 8      Semi-concrete
## 9      Semi-concrete
## 10          Congrete
```

```
#b
my_data1 <- read.csv("HouseholdData.csv")
my_data1$Sex <- factor(my_data1$Sex, levels = c("Male", "Female"))
my_data1$Sex <- as.integer(my_data1$Sex)
print(my_data1$Sex)
```

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

```
#c
my_data2 <- read.csv("HouseholdData.csv")
my_data2$Types_of_Houses <- factor(my_data2$Types_of_Houses, levels = c("Wood", "Congrete", "Semi-concrete"))
my_data2$Types_of_Houses <- as.integer(my_data2$Types_of_Houses)
print(my_data2$Types_of_Houses)
```

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

```
#d
my_data3 <- read.csv("HouseholdData.csv")
my_data3$Fathers_Occupation <- factor(my_data3$Fathers_Occupation, levels = c(1, 2, 3), labels = c("Farmer", "Teacher", "Other"))
my_data3$Fathers_Occupation <- as.integer(my_data3$Fathers_Occupation)
print(my_data3$Fathers_Occupation)
```

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

```
#e
data4 <- read.csv("HouseholdData.csv")
female_driver <- subset(data4, Sex == "Female" & Fathers_Occupation == 2)
female_driver
```

```
##      Respondents      Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 2              2 Female                2              7              3
## 6              6 Female                2              4              1
## 7              7 Female                2              2              1
##      Types_of_Houses
## 2              Congrete
## 6      Semi-concrete
## 7              Wood
```

```
#f
data5 <- read.csv("HouseholdData.csv")
greater_than_5_siblings <- subset(data5, Siblings_at_School >= 5)
greater_than_5_siblings

##   Respondents    Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 4           4   Male                3             8             5
## 9           9 Female                1             11             6
##   Types_of_Houses
## 4           Wood
## 9   Semi-concrete
```

```
#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
```