

RWorksheet_Liza#3b

Jowelyn Liza

2024-12-18

#1a

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

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

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

#1b

#The data contains information from 20 individuals regarding their families and residences. The majority

```
str(respondents_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 1 2 1 2 1 2 ...
## $ Fathers_Occupation: num 1 3 1 3 3 1 3 2 3 1 ...
## $ Person_at_Home : num 5 7 3 8 9 6 9 6 4 3 ...
## $ Siblings_at_School: num 6 4 4 1 1 3 3 5 3 2 ...
## $ Types_of_Houses : num 1 2 1 1 3 3 3 2 1 3 ...
```

```
summary(respondents_data)
```

```
## Respondents      Sex      Fathers_Occupation Person_at_Home
## Min.   : 1.00    Min.   :1.0    Min.   :1      Min.   : 3.00
## 1st Qu.: 5.75    1st Qu.:1.0    1st Qu.:1      1st Qu.: 4.75
## Median :10.50    Median :1.5    Median :2      Median : 6.00
## Mean   :10.50    Mean   :1.5    Mean   :2      Mean   : 6.20
## 3rd Qu.:15.25    3rd Qu.:2.0    3rd Qu.:3      3rd Qu.: 7.25
## Max.   :20.00    Max.   :2.0    Max.   :3      Max.   :11.00
## Siblings_at_School Types_of_Houses
## Min.   :1.00      Min.   :1.00
## 1st Qu.:2.00      1st Qu.:1.00
## Median :3.00      Median :2.00
## Mean   :3.15      Mean   :1.95
## 3rd Qu.:4.00      3rd Qu.:3.00
## Max.   :6.00      Max.   :3.00
```

#1c

#no, the mean of the number of siblings is 3.15

```
mean_siblings <- mean(respondents_data$Siblings_at_School)
mean_siblings == 5
```

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

```
mean_siblings
```

```
## [1] 3.15
```

```
#1d
```

```
subset_data <- respondents_data[1:2,]  
subset_data
```

```
## Respondents Sex Fathers_Occupation Person_at_Home Siblings_at_School  
## 1          1 2              1              5              6  
## 2          2 2              3              7              4  
## Types_of_Houses  
## 1          1  
## 2          2
```

```
#1e
```

```
subset_data2 <- respondents_data[c(3, 5), c(2, 4)]  
subset_data2
```

```
## Sex Person_at_Home  
## 3 1              3  
## 5 1              9
```

```
#1f
```

```
types_houses <- respondents_data$Types_of_Houses  
types_houses
```

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

```
#1g
```

```
male_farmers <- subset(respondents_data, Sex == 1 & Fathers_Occupation == 1)  
male_farmers
```

```
## Respondents Sex Fathers_Occupation Person_at_Home Siblings_at_School  
## 3          3 1              1              3              4  
## 12         12 1              1              5              2  
## 16         16 1              1              7              3  
## 18         18 1              1              7              3  
## 20         20 1              1              6              2  
## Types_of_Houses  
## 3          1  
## 12         2  
## 16         3  
## 18         3  
## 20         2
```

```
#1h
```

```
femalesibs <- subset(respondents_data, Sex == 2 & Siblings_at_School >= 5)  
femalesibs
```

```
## Respondents Sex Fathers_Occupation Person_at_Home Siblings_at_School  
## 1          1 2              1              5              6  
## 8          8 2              2              6              5  
## 17         17 2              3             11              5  
## Types_of_Houses  
## 1          1  
## 8          2  
## 17         1
```

```
#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
```

#It is an empty data frame with 5 columns designated for integers, decimals, text, and boolean values (

```
#3
```

```

household_data <- data.frame(
  Respondents = 1:10,
  Sex = c("Male", "Female", "Female", "Male", "Male",
          "Female", "Female", "Male", "Female", "Male"),

  Fathers_Occupation = c("Farmer", "Farmer", "Farmer", "Farmer", "Driver", "Driver", "Driver", "Driver",
                          "Driver", "Driver"),

  Persons_at_Home = c(5, 7, 3, 8, 1, 2, 4, 3, 1, 6),

  Siblings_at_School = c(5, 7, 3, 8, 1, 4, 2, 6, 11, 6),

  Types_of_Houses = c("Wood", "Concrete", "Concrete", "Wood", "Semi-concrete", "Semi-concrete", "Concrete",
                      "Concrete", "Wood", "Semi-concrete")
household_data

```

```

##   Respondents    Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1           1   Male           Farmer              5           5
## 2           2 Female           Farmer              7           7
## 3           3 Female           Farmer              3           3
## 4           4   Male           Farmer              8           8
## 5           5   Male           Driver              1           1
## 6           6 Female           Driver              2           4
## 7           7 Female           Driver              4           2
## 8           8   Male           Driver              3           6
## 9           9 Female           Others              1          11
## 10          10   Male           Others              6           6
##   Types_of_Houses
## 1             Wood
## 2           Concrete
## 3           Concrete
## 4             Wood
## 5   Semi-concrete
## 6   Semi-concrete

```

```
## 7      Concrete
## 8      Wood
## 9      Semi-concrete
## 10     Concrete
```

#3a

```
write.csv(household_data, "HouseHoldData.csv", row.names = FALSE)
library(readr)
household_data <- read.csv("HouseHoldData.csv")
household_data
```

```
##      Respondents      Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1             1    Male           Farmer             5           5
## 2             2  Female           Farmer             7           7
## 3             3  Female           Farmer             3           3
## 4             4    Male           Farmer             8           8
## 5             5    Male           Driver             1           1
## 6             6  Female           Driver             2           4
## 7             7  Female           Driver             4           2
## 8             8    Male           Driver             3           6
## 9             9  Female           Others             1          11
## 10           10    Male           Others             6           6
```

```
##      Types_of_Houses
## 1             Wood
## 2             Concrete
## 3             Concrete
## 4             Wood
## 5      Semi-concrete
## 6      Semi-concrete
## 7             Concrete
## 8             Wood
## 9      Semi-concrete
## 10            Concrete
```

#3b

```
household_data$Sex <- factor(household_data$Sex, levels = c("Male", "Female"), labels = c(1,2))
household_data$Sex
```

```
## [1] 1 2 2 1 1 2 2 1 2 1
## Levels: 1 2
```

#3c

```
household_data$Type_of_Houses <- factor(household_data$Types_of_Houses, levels = c("Wood", "Concrete", "Semi-concrete"), labels = c(1,2,3))
household_data$Types_of_Houses
```

```
## [1] "Wood"      "Concrete"   "Concrete"   "Wood"
## [5] "Semi-concrete" "Semi-concrete" "Concrete"   "Wood"
## [9] "Semi-concrete" "Concrete"
```

#3d

```
household_data$Fathers_Occupation <- factor(household_data$Fathers_Occupation, levels = c(1, 2, 3), labels = c("Farmer", "Driver", "Others"))
household_data$Fathers_Occupation
```

```
## [1] <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## Levels: Farmer Driver Others
```

#3e

```
female_driver <- subset(household_data, Sex == 2 & Fathers_Occupation == "Driver")
```

```
female_driver
```

```
## [1] Respondents      Sex      Fathers_Occupation Persons_at_Home  
## [5] Siblings_at_School Types_of_Houses  Type_of_Houses  
## <0 rows> (or 0-length row.names)
```

```
#3f
```

```
siblings_5_or_more <- subset(household_data, Siblings_at_School >= 5)  
siblings_5_or_more
```

```
##      Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School  
## 1             1   1             <NA>             5             5  
## 2             2   2             <NA>             7             7  
## 4             4   1             <NA>             8             8  
## 8             8   1             <NA>             3             6  
## 9             9   2             <NA>             1            11  
## 10            10   1             <NA>             6             6  
##      Types_of_Houses Type_of_Houses  
## 1             Wood      1  
## 2             Concrete  2  
## 4             Wood      1  
## 8             Wood      1  
## 9      Semi-concrete  3  
## 10            Concrete  2
```

```
#4
```

```
#The graph illustrates the sentiments expressed in tweets: negative sentiments are represented in red, .
```