

RWorksheet3b in R

Keir G. Sumayo

2025-10-19

###1.

```
#A.
respondents <- c(1:20)
sex <- c(2,2,1,2,2,2,2,2,2,1,2,2,2,2,2,2,1,2)
f_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(5,7,3,8,5,9,6,7,8,4,7,5,4,7,8,8,3,11,7,6)
siblings_at_school <- c(6,4,4,1,2,1,5,3,1,2,3,2,5,5,2,1,2,5,3,2)
types_of_houses <- c(1,2,3,1,1,3,3,1,2,3,2,3,2,2,3,3,3,3,3,2)

table <- data.frame (
  Respondents = respondents,
  Sex = sex,
  Fathers_Occupation = f_occupation,
  Persons_at_Home = persons_at_home,
  Siblings_at_School = siblings_at_school,
  Types_of_houses = types_of_houses
)

table
```

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

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

```
#B.
#These are datas of 20 people, all 6 columns has no missing data.
str(table)
```

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

```
#C. Is the mean number of siblings attending is 5?
mean_siblings <- mean(siblings_at_school)
mean_siblings
```

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

```
#D.
#Using head()
head(table, n = 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.
target_cols <- c("Sex", "Persons_at_Home")
```

```
#Using head()
head_table <- head(table, n = 5)
head_table[c(3, 5), target_cols]
```

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

```
#F.
types_houses <- table$Types_of_houses
types_houses
```

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

```
#G.
#Using subset()
selected_data_male <- subset(table, subset = Sex == 1 & Fathers_Occupation == 1)
selected_data_male
```

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

#There is none, because there is no male that has a father who has a occupation of farmer.

```
#H.
selected_data_female <- subset(table, subset = Sex == 2 & Siblings_at_School >= 5)
selected_data_female
```

```
##      Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1              1  2              1              5              6
## 7              7  2              3              6              5
## 13             13  2              1              4              5
## 14             14  2              3              7              5
## 18             18  2              1             11              5
##      Types_of_houses
## 1              1
## 7              3
## 13             2
## 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.
str(df)
```

```
## 'data.frame': 0 obs. of 5 variables:
## $ Ints : int
## $ Doubles : num
## $ Characters: chr
## $ Logicals : logi
## $ Factors : Factor w/ 0 levels:
```

```
###3.
```

```
#A.
householddata_imported <- read.csv("HouseholdData.csv",
                                   header = TRUE)
head(householddata_imported, n = 10)
```

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

```
##      Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1          1    1              1              5              2
## 2          2    2              2              7              3
## 3          3    2              3              3              0
## 4          4    1              3              8              5
## 5          5    1              1              6              2
## 6          6    2              2              4              3
## 7          7    2              2              4              1
## 8          8    1              3              2              2
## 9          9    2              1             11              6
## 10         10    1              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
```

```
#C.
householddata_imported$Types_of_Houses <- factor(householddata_imported$Types_of_Houses,
                                                  levels = c("Wood", "Congrete", "Semi-concrete"),
                                                  labels = c(1,2,3))
householddata_imported
```

```
##      Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1          1    1              1              5              2
## 2          2    2              2              7              3
## 3          3    2              3              3              0
## 4          4    1              3              8              5
## 5          5    1              1              6              2
## 6          6    2              2              4              3
## 7          7    2              2              4              1
## 8          8    1              3              2              2
```

```
## 9          9  2          1          11          6
## 10         10  1          3          6          2
##   Types_of_Houses
## 1          1
## 2          2
## 3          2
## 4          1
## 5          3
## 6          3
## 7          1
## 8          3
## 9          3
## 10         2
```

#D.

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

```
##   Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1          1  1          Farmer          5          2
## 2          2  2          Driver          7          3
## 3          3  2          Others          3          0
## 4          4  1          Others          8          5
## 5          5  1          Farmer          6          2
## 6          6  2          Driver          4          3
## 7          7  2          Driver          4          1
## 8          8  1          Others          2          2
## 9          9  2          Farmer         11          6
## 10         10  1          Others          6          2
##   Types_of_Houses
## 1          1
## 2          2
## 3          2
## 4          1
## 5          3
## 6          3
## 7          1
## 8          3
## 9          3
## 10         2
```

#E.

```
select_f_driver <- subset(householddata_imported, subset = Sex == 2 & Fathers_Occupation == "Driver")
select_f_driver
```

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

```
## 2          2
## 6          3
## 7          1
```

#F.

```
selected_data_both <- subset(householddata_imported, subset = Siblings_at_School >= 5)
```

```
selected_data_both
```

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

####4.

#This graph is about sentiment tweets per day, and there is a spike of negative tweets from july 15 t