

RWorksheet_Delatina#3B

Angel

2024-10-03

#1A

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

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

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

#1B

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

#1C

```
mean_siblings <- mean(data$Siblings_at_School)
mean_siblings
```

```
## [1] 3.1
```

#1D

```
subset_rows_1_2 <- data[1:2,]
subset_rows_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
```

#1E

```
subset_3_5_rows_2_4_col <- data[c(3,5),c(2,4)]
subset_3_5_rows_2_4_col
```

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

```
#1F
```

```
types_houses <- data$Types_of_Houses
types_houses
```

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

```
#1G
```

```
male_farmers <- data[data$Sex == 1 & data$Fathers_Occupation == 1, ]
male_farmers
```

```
## Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 6 6 1 1 3 3
## 10 10 1 1 4 2
## 15 15 1 1 8 1
## 18 18 1 1 11 3
## 19 19 1 1 7 2
## Types_of_Houses
## 6 3
## 10 3
## 15 3
## 18 3
## 19 3
```

```
#1H
```

```
female_siblings_5_or_more <- data[data$Sex == 2 & data$Siblings_at_School >= 5,]
female_siblings_5_or_more
```

```
## Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1 1 2 1 5 6
## 13 13 2 3 7 5
## 14 14 2 2 3 7
## Types_of_Houses
## 1 1
## 13 2
## 14 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
```

```
#3A
```

```
household_data <- read.csv("~/DataScience/CS101/worksheet3B/HouseholdsData.csv")
head(household_data)
```

```
## Respondents Sex Fathers.Occupation Person.at.Home Siblings.at.Home
## 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
## Types.of.Houses
## 1 Wood
## 2 Congrete
## 3 Congrete
## 4 Wood
## 5 Semi-Congrete
## 6 Semi-Congrete
```

```
#3B
```

```
household_data$Sex <- factor(household_data$Sex, levels = c("Male", "Female"))
household_data$Sex <- as.integer(household_data$Sex)
household_data
```

```
## Respondents Sex Fathers.Occupation Person.at.Home Siblings.at.Home
## 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 2 3 6 2
## Types.of.Houses
## 1 Wood
```

```
## 2      Congrete
## 3      Congrete
## 4      Wood
## 5      Semi-Congrete
## 6      Semi-Congrete
## 7      Wood
## 8      Semi-Congrete
## 9      Semi-Congrete
## 10     Congrete
```

```
#3C
```

```
str(household_data)
```

```
## 'data.frame':  10 obs. of  6 variables:
## $ Respondents      : int  1 2 3 4 5 6 7 8 9 10
## $ Sex              : int  1 2 2 1 1 2 2 1 2 2
## $ Fathers.Occupation: int  1 2 3 3 1 2 2 3 1 3
## $ Person.at.Home   : int  5 7 3 8 6 4 4 2 11 6
## $ Siblings.at.Home : int  2 3 0 5 2 3 1 2 6 2
## $ Types.of.Houses  : chr  "Wood" "Congrete" "Congrete" "Wood" ...
```

```
head(household_data)
```

```
##   Respondents Sex Fathers.Occupation Person.at.Home Siblings.at.Home
## 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
##   Types.of.Houses
## 1      Wood
## 2    Congrete
## 3    Congrete
## 4      Wood
## 5 Semi-Congrete
## 6 Semi-Congrete
```

```
unique(household_data$Types_of_Houses)
```

```
## NULL
```

```
if("Types_of_Houses"%in% names(household_data)){
  household_data$Types_of_Houses<- factor(household_data$Types_of_Houses,
                                           levels = c("Wood", "Concrete", "Semi-Concret"),
  household_data$Types_of_Houses <- as.integer(household_data$Types_of_Houses)

  print(head(household_data))
}else{
  print("The column 'Types_of_Houses' does not exist in the data frame.")
}
```

```
## [1] "The column 'Types_of_Houses' does not exist in the data frame."
```

```
#3D
```

```
str(household_data)
```

```
## 'data.frame': 10 obs. of 6 variables:
## $ Respondents : int 1 2 3 4 5 6 7 8 9 10
## $ Sex : int 1 2 2 1 1 2 2 1 2 2
## $ Fathers.Occupation: int 1 2 3 3 1 2 2 3 1 3
## $ Person.at.Home : int 5 7 3 8 6 4 4 2 11 6
## $ Siblings.at.Home : int 2 3 0 5 2 3 1 2 6 2
## $ Types.of.Houses : chr "Wood" "Congrete" "Congrete" "Wood" ...
```

```
head(household_data)
```

```
## Respondents Sex Fathers.Occupation Person.at.Home Siblings.at.Home
## 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
## Types.of.Houses
## 1 Wood
## 2 Congrete
## 3 Congrete
## 4 Wood
## 5 Semi-Congrete
## 6 Semi-Congrete
```

```
names(household_data)[names(household_data) == "Fathers Occupation"]<- "Fathers-Occupation"
names(household_data)
```

```
## [1] "Respondents" "Sex" "Fathers.Occupation"
## [4] "Person.at.Home" "Siblings.at.Home" "Types.of.Houses"
```

```
if("Fathers_Occupation" %in% names(household_data)){
  household_data$Fathers_Occupation <- factor(household_data$Fathers_Occupation,
                                              levels= c("Farmer", "Driver", "Others"),
                                              labels= c(1, 2, 3))
  household_data$Fathers-Occupation <- as.integer(household_data$Fathers_Occupation)
  print(head(household_data))
} else{
  print("The column 'Fathers_Occupation' still does not exist in the data frame")
}
```

```
## [1] "The column 'Fathers_Occupation' still does not exist in the data frame"
```

```
#3E
```

```
str(household_data)
```

```
## 'data.frame':  10 obs. of  6 variables:
## $ Respondents      : int  1 2 3 4 5 6 7 8 9 10
## $ Sex              : int  1 2 2 1 1 2 2 1 2 2
## $ Fathers.Occupation: int  1 2 3 3 1 2 2 3 1 3
## $ Person.at.Home   : int  5 7 3 8 6 4 4 2 11 6
## $ Siblings.at.Home : int  2 3 0 5 2 3 1 2 6 2
## $ Types.of.Houses  : chr  "Wood" "Congrete" "Congrete" "Wood" ...
```

```
names(household_data)
```

```
## [1] "Respondents"      "Sex"              "Fathers.Occupation"
## [4] "Person.at.Home"   "Siblings.at.Home" "Types.of.Houses"
```

```
head(household_data)
```

```
##   Respondents Sex Fathers.Occupation Person.at.Home Siblings.at.Home
## 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
##   Types.of.Houses
## 1           Wood
## 2           Congrete
## 3           Congrete
## 4           Wood
## 5   Semi-Congrete
## 6   Semi-Congrete
```

```
females_with_driver_father <- subset(household_data, Sex == "Female"& 'Fathers_Occupation' == "Driver")
```

```
#3F
```

```
str(household_data)
```

```
## 'data.frame':  10 obs. of  6 variables:
## $ Respondents      : int  1 2 3 4 5 6 7 8 9 10
## $ Sex              : int  1 2 2 1 1 2 2 1 2 2
## $ Fathers.Occupation: int  1 2 3 3 1 2 2 3 1 3
## $ Person.at.Home   : int  5 7 3 8 6 4 4 2 11 6
## $ Siblings.at.Home : int  2 3 0 5 2 3 1 2 6 2
## $ Types.of.Houses  : chr  "Wood" "Congrete" "Congrete" "Wood" ...
```

```
names(household_data)
```

```
## [1] "Respondents"      "Sex"              "Fathers.Occupation"
## [4] "Person.at.Home"   "Siblings.at.Home" "Types.of.Houses"
```

```
head(household_data)
```

```
## Respondents Sex Fathers.Occupation Person.at.Home Siblings.at.Home
## 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
## Types.of.Houses
## 1          Wood
## 2          Congrete
## 3          Congrete
## 4          Wood
## 5      Semi-Congrete
## 6      Semi-Congrete
```

```
females_with_driver_father <- subset(household_data, Sex == "Female" & `Fathers.Occupation` == "Driver")
```