Rworksheet_Aguirre#3

Ryza Faith

2024-10-01

1.

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

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

```
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 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 8 3 8 5 9 6 7 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 3 1 2 3 ...
```

```
summary(respondents_data)
```

```
##   Respondents         Sex       Fathers_Occupation Persons_at_Home
##  Min.   : 1.00   Min.   :1.00   Min.   :1.00       Min.   :3.00
##  1st Qu.: 5.75   1st Qu.:2.00   1st Qu.:1.00       1st Qu.:4.75
##  Median :10.50   Median :2.00   Median :2.00       Median :6.50
##  Mean   :10.50   Mean   :1.85   Mean   :1.95       Mean   :6.05
##  3rd Qu.:15.25   3rd Qu.:2.00   3rd Qu.:3.00       3rd Qu.:8.00
##  Max.   :20.00   Max.   :2.00   Max.   :3.00       Max.   :9.00
##  Siblings_at_School Types_of_Houses
##  Min.   :1.00       Min.   :1.0
##  1st Qu.:2.00       1st Qu.:2.0
##  Median :2.50       Median :2.5
##  Mean   :2.95       Mean   :2.3
##  3rd Qu.:4.25       3rd Qu.:3.0
##  Max.   :6.00       Max.   :3.0
```

C.

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

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

D.

```r
# Extract the first two rows and all columns
first_two_rows <- respondents_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               7                  4
##   Types_of_Houses
## 1               1
## 2               2
```

E.

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

```
##   Sex Persons_at_Home
## 3   1               8
## 5   2               8
```

F.

```r
types_houses <- respondents_data$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.

```r
males_farmers <- respondents_data[respondents_data$Sex == 1 & respondents_data$Fathers_Occupation == 1,
males_farmers
```

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

H.

```r
females_with_siblings <- respondents_data[respondents_data$Sex == 2 & respondents_data$Siblings_at_Scho
females_with_siblings
```

```
##    Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1            1   2                  1               5                  6
## 7            7   2                  3               9                  5
## 13          13   2                  1               5                  5
## 14          14   2                  3               4                  5
## 18          18   2                  1               3                  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
```

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.

```
unique(household_data$Sex)
```

```
## [1] "Male"   "Female"
```

```
household_data$Sex <- ifelse(tolower(household_data$Sex) == "male", 1,
                             ifelse(tolower(household_data$Sex) == "female", 2, NA))
household_data$Sex <- factor(household_data$Sex, levels = c(1, 2), labels = c("Male", "Female"))
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
```

```
unique(household_data$Sex)
```

```
## [1] Male   Female
## Levels: Male Female
```

2C.

```
str(household_data)
```

```
## 'data.frame':    10 obs. of  6 variables:
##  $ Respondents       : int  1 2 3 4 5 6 7 8 9 10
##  $ Sex               : Factor w/ 2 levels "Male","Female": 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   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
```

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

```
## NULL
```

```r
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"),
                                             labels = c(1, 2, 3))

    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.

```r
str(household_data)
```

```
## 'data.frame':    10 obs. of  6 variables:
##  $ Respondents       : int  1 2 3 4 5 6 7 8 9 10
##  $ Sex               : Factor w/ 2 levels "Male","Female": 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" ...
```

```r
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
```

```r
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"
```

```r
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.

```r
# Check the structure of the data frame
str(household_data)
```

```
## 'data.frame':    10 obs. of  6 variables:
##  $ Respondents       : int  1 2 3 4 5 6 7 8 9 10
##  $ Sex               : Factor w/ 2 levels "Male","Female": 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" ...
```

```r
# Check the column names in the data frame
names(household_data)
```

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

```r
# View the first few rows of the data frame
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
```

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

3F.

```r
# Check the structure of the data frame
str(household_data)
```

```
## 'data.frame':    10 obs. of  6 variables:
##  $ Respondents       : int  1 2 3 4 5 6 7 8 9 10
##  $ Sex               : Factor w/ 2 levels "Male","Female": 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" ...
```

```r
# Check the column names in the data frame
names(household_data)
```

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

```r
# Adjusting to the correct column name
respondents_with_5_siblings <- subset(household_data, Siblings.at.Home >= 5)
print(respondents_with_5_siblings)
```

```
##   Respondents    Sex Fathers.Occupation Person.at.Home Siblings.at.Home
## 4           4   Male                  3              8                5
## 9           9 Female                  1             11                6
##   Types.of.Houses
## 4            Wood
## 9   Semi-Congrete
```