RWorksheet_Redosendo#3b

Kyle Edward B. Redosendo

2025-11-16

```
# 1. Create a data frame using the table below
#a
df <- 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(5,7,3,8,5,9,6,7,8,4,7,5,4,7,8,8,3,11,7,6),
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(1,2,3,1,1,3,3,1,2,3,2,3,2,2,3,3,2,3,3,2)
)
```

```
#b
print("Data Frame:")
```

```
## [1] "Data Frame:"
```

```
df
```

```
##    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               2
## 18               3
## 19               3
## 20               2
```

```
print("Structure of the Data:")
```

```
## [1] "Structure of the Data:"
```

```
str(df)
```

```
## '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 ...
```

```
print("Summary of the Data:")
```

```
## [1] "Summary of the Data:"
```

```
summary(df)
```

```
##   Respondents          Sex       Fathers_Occupation Persons_at_Home
##  Min.   : 1.00   Min.   :1.00   Min.   :1.00       Min.   : 3.0
##  1st Qu.: 5.75   1st Qu.:2.00   1st Qu.:1.00       1st Qu.: 5.0
##  Median :10.50   Median :2.00   Median :2.00       Median : 7.0
##  Mean   :10.50   Mean   :1.85   Mean   :1.95       Mean   : 6.4
##  3rd Qu.:15.25   3rd Qu.:2.00   3rd Qu.:3.00       3rd Qu.: 8.0
##  Max.   :20.00   Max.   :2.00   Max.   :3.00       Max.   :11.0
##  Siblings_at_School Types_of_Houses
##  Min.   :1.00       Min.   :1.00
##  1st Qu.:2.00       1st Qu.:2.00
##  Median :2.50       Median :2.00
##  Mean   :2.95       Mean   :2.25
##  3rd Qu.:4.25       3rd Qu.:3.00
##  Max.   :6.00       Max.   :3.00
```

```
# c. Mean siblings

sibmean <- mean(df$Siblings_at_School)
print(paste("Actual mean is:", sibmean))
```

```
## [1] "Actual mean is: 2.95"
```

```r
sibmean == 5
```

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

```r
# d. Extract 1st two rows

subset_d <- df[1:2, ]
subset_d
```

```
##   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
```

```r
# e. Extract 3rd & 5th rows, 2nd & 4th cols

subset_e <- df[c(3, 5), c(2, 4)]
subset_e
```

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

```r
# f. Vector of Types_of_Houses

types_houses <- df$Types_of_Houses
types_houses
```

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

```r
# g. Male respondents (Sex = 1) with farmer father (1)

subset_g <- df[df$Sex == 1 & df$Fathers_Occupation == 1, ]
subset_g
```

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

```r
# h. Female respondents (2) with >=5 siblings in school

subset_h <- df[df$Sex == 2 & df$Siblings_at_School >= 5, ]
subset_h
```

```
##    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
```

```r
# 2. Empty dataframe

df_empty <- 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:"
```

```r
str(df_empty)
```

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

```r
# 3. Import CSV (must be uploaded to Posit Cloud)

house <- read.csv("HouseholdData.csv", header = TRUE)
print("Imported CSV Data:")
```

```
## [1] "Imported CSV Data:"
```

```r
house
```

```
##    Respondents    Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 1            1   Male                  1               5                  2
## 2            2 Female                  2               2                  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               2                  1
## 8            8   Male                  3               2                  2
## 9            9 Female                  1               1                  6
## 10          10   Male                  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
```

```r
# 3b. Convert Sex

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

```
##  [1] <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## Levels: Male Female
```

```r
# 3c. Convert Types_of_Houses

if ("Types_of_Houses" %in% names(house) && length(house$Types_of_Houses) > 0) {
  house$Types_of_Houses <- factor(
    house$Types_of_Houses,
    levels = c(1, 2, 3),
    labels = c("Wood", "Concrete", "Semi-Congrete")
  )
} else {
  warning("Column 'Types_of_Houses' is missing or empty in the dataset.")
}
```

```
## Warning: Column 'Types_of_Houses' is missing or empty in the dataset.
```

```r
house$Types_of_Houses
```

```
## NULL
```

```r
# 3d. Convert Fathers_Occupation
if ("Fathers_Occupation" %in% names(house) && length(house$Fathers_Occupation) > 0) {
  house$Fathers_Occupation <- factor(
    house$Fathers_Occupation,
    levels = c(1, 2, 3),
    labels = c("Farmer", "Driver", "Others")
  )
} else {
  # If column missing or empty, create it with NAs of correct length
  warning("Column 'Fathers_Occupation' is missing or empty. Filling with NA.")
  house$Fathers_Occupation <- factor(rep(NA, nrow(house)))
}
```

```
## Warning: Column 'Fathers_Occupation' is missing or empty. Filling with NA.
```

```r
house$Fathers_Occupation
```

```
##  [1] <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## Levels:
```

```r
# 3e. Female respondents with driver father

subset_3e <- subset(house, Sex == "Female" & Fathers_Occupation == "Driver")
subset_3e
```

```
## [1] Respondents        Sex                Fathers.Occupation Persons.at.Home
## [5] Siblings.at.School Types.of.Houses    Fathers_Occupation
## <0 rows> (or 0-length row.names)
```

```r
# 3f. Respondents with >=5 siblings in school

subset_3f <- subset(df, Siblings_at_School >= 5)
subset_3f
```

```
##    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
```