RWorksheet_Regacho#3b

STEP 1: Create the dataset manually

```
# a. Create a data frame
respondents <- c(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(1,2,3,1,1,3,3,1,2,3,2,3,2,2,3,3,3,3,3,2)

data <- data.frame(
  Respondents = respondents,
  Sex = sex,
  Fathers_Occupation = fathers_occupation,
  Persons_at_Home = persons_at_home,
  Siblings_at_School = siblings_at_school,
  Types_of_Houses = types_of_houses
)

# b. Describe the data and get structure/summary
cat("Structure of the data:\n")
```

```
## Structure of the data:
```

```
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 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 ...
```

```
cat("\nSummary of the data:\n")
```

```
##
## Summary of the data:
```

```
summary(data)
```

```
##   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.   :1.0
##  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. Check if mean number of siblings attending is 5
siblings_mean <- mean(data$Siblings_at_School)
cat("\nMean number of siblings attending school:", siblings_mean, "\n")
```

```
##
## Mean number of siblings attending school: 2.95
```

```
cat("Is the mean number of siblings attending 5?", siblings_mean == 5, "\n")
```

```
## Is the mean number of siblings attending 5? FALSE
```

```
# d. Extract first two rows and all columns
first_two_rows <- data[1:2, ]
cat("\nFirst two rows with all columns:\n")
```

```
##
## First two rows with all columns:
```

```
print(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. Extract 3rd and 5th row with 2nd and 4th column
subset_data <- data[c(3,5), c(2,4)]
cat("\n3rd and 5th row with 2nd and 4th column:\n")
```

```
##
## 3rd and 5th row with 2nd and 4th column:
```

```r
print(subset_data)
```

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

```r
# f. Select variable types of houses and store as types_houses
types_houses <- data$Types_of_Houses
cat("\nTypes of houses vector:\n")
```

```
##
## Types of houses vector:
```

```r
print(types_houses)
```

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

```r
# g. Select all Males respondent with father occupation as farmer
males_farmer <- subset(data, Sex == 1 & Fathers_Occupation == 1)
cat("\nMales with father occupation as farmer:\n")
```

```
##
## Males with father occupation as farmer:
```

```r
print(males_farmer)
```

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

```r
# h. Select all females respondent with >=5 siblings attending school
females_siblings <- subset(data, Sex == 2 & Siblings_at_School >= 5)
cat("\nFemales with >=5 siblings attending school:\n")
```

```
##
## Females with >=5 siblings attending school:
```

```r
print(females_siblings)
```

```
##    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
# Create empty data frame
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:"
```

```r
print(str(df))
```

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

```r
# a.
# The empty data frame has 0 observations(rows) and 5 variables(columns) with the specified data types.
```

```r
# Create the data frame from the table
household_data <- data.frame(
  Respondents = 1:10,
  Sex = c("Male", "Female", "Female", "Male", "Male", "Female", "Female", "Male", "Female", "Male"),
  Fathers_Occupation = c(1, 2, 3, 3, 1, 2, 2, 3, 1, 3),
  Persons_at_Home = c(5, 7, 3, 8, 6, 4, 4, 2, 11, 6),
  Siblings_at_School = c(2, 3, 0, 5, 2, 3, 1, 2, 6, 2),
  Types_of_Houses = c("Wood", "Congrete", "Congrete", "Wood", "Semi-congrete",
                      "Semi-congrete", "Wood", "Semi-congrete", "Semi-congrete", "Congrete")
)

# a. Create and import CSV file
write.csv(household_data, "HouseholdData.csv", row.names = FALSE)
imported_data <- read.csv("HouseholdData.csv")

cat("a. Imported data:\n")
```

```
## a. Imported data:
```

```r
print(imported_data)
```

```
##    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-congrete
## 6    Semi-congrete
## 7             Wood
## 8    Semi-congrete
## 9    Semi-congrete
## 10        Congrete
```

```r
# b. Convert Sex into factor and change to integer
imported_data$Sex <- factor(imported_data$Sex, levels = c("Male", "Female"))
imported_data$Sex <- as.integer(imported_data$Sex)

cat("\nb. Sex converted to factor and integer:\n")
```

```
##
## b. Sex converted to factor and integer:
```

```r
print(imported_data$Sex)
```

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

```r
# c. Convert Types of Houses into factor and change to integer
imported_data$Types_of_Houses <- factor(imported_data$Types_of_Houses,
                                        levels = c("Wood", "Congrete", "Semi-congrete"))
imported_data$Types_of_Houses <- as.integer(imported_data$Types_of_Houses)

cat("\nc. Types of Houses converted to factor and integer:\n")
```

```
##
## c. Types of Houses converted to factor and integer:
```

```r
print(imported_data$Types_of_Houses)
```

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

```r
# d. Convert Father's Occupation to factor
imported_data$Fathers_Occupation <- factor(imported_data$Fathers_Occupation,
                                           levels = c(1, 2, 3),
                                           labels = c("Farmer", "Driver", "Others"))

cat("\nd. Father's Occupation converted to factor:\n")
```

```
##
## d. Father's Occupation converted to factor:
```

```
print(imported_data$Fathers_Occupation)
```

```
##  [1] Farmer Driver Others Others Farmer Driver Driver Others Farmer Others
## Levels: Farmer Driver Others
```

```
# e. Select all females with father occupation as driver
females_driver <- subset(imported_data, Sex == 2 & Fathers_Occupation == "Driver")

cat("\ne. Females with father occupation as driver:\n")
```

```
##
## e. Females with father occupation as driver:
```

```
print(females_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. Select respondents with >=5 siblings attending school
high_siblings <- subset(imported_data, Siblings_at_School >= 5)

cat("\nf. Respondents with >=5 siblings attending school:\n")
```

```
##
## f. Respondents with >=5 siblings attending school:
```

```
print(high_siblings)
```

```
##   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
```

```
cat("\nFinal data frame after all conversions:\n")
```

```
##
## Final data frame after all conversions:
```

```r
print(imported_data)
```

```
##    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
```