RWorksheet_Cababasay#3b

Mike Jayson Cababasay

```r
# 1. Creating and Exploring a Data Frame

# a. Create the data frame
household <- data.frame(
  Sex = c("Male", "Female", "Male", "Female", "Female"),
  Types_of_Houses = c("Wood", "Concrete", "Semi-Concrete", "Concrete", "Wood"),
  Fathers_Occupation = c("Farmer", "Driver", "Farmer", "Others", "Farmer"),
  Number_of_Siblings_Attending = c(3, 5, 2, 6, 7)
)
household
```

```
##      Sex Types_of_Houses Fathers_Occupation Number_of_Siblings_Attending
## 1   Male            Wood             Farmer                           3
## 2 Female        Concrete             Driver                           5
## 3   Male   Semi-Concrete             Farmer                           2
## 4 Female        Concrete             Others                           6
## 5 Female            Wood             Farmer                           7
```

```r
# b. Describe the data (structure and summary)
str(household)
```

```
## 'data.frame':    5 obs. of  4 variables:
##  $ Sex                          : chr  "Male" "Female" "Male" "Female" ...
##  $ Types_of_Houses              : chr  "Wood" "Concrete" "Semi-Concrete" "Concrete" ...
##  $ Fathers_Occupation           : chr  "Farmer" "Driver" "Farmer" "Others" ...
##  $ Number_of_Siblings_Attending: num  3 5 2 6 7
```

```r
summary(household)
```

```
##      Sex            Types_of_Houses    Fathers_Occupation
##  Length:5           Length:5           Length:5
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##  Number_of_Siblings_Attending
##  Min.   :2.0
##  1st Qu.:3.0
##  Median :5.0
##  Mean   :4.6
##  3rd Qu.:6.0
##  Max.   :7.0
```

```r
# c. Is the mean number of siblings attending 5?
mean(household$Number_of_Siblings_Attending) == 5
```

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

```
mean(household$Number_of_Siblings_Attending)
```

```
## [1] 4.6
```

```
# d. Extract the 1st two rows and all columns
household[1:2, ]
```

```
##      Sex Types_of_Houses Fathers_Occupation Number_of_Siblings_Attending
## 1   Male            Wood             Farmer                           3
## 2 Female        Concrete             Driver                           5
```

```
# e. Extract 3rd and 5th rows with 2nd and 4th columns
household[c(3,5), c(2,4)]
```

```
##   Types_of_Houses Number_of_Siblings_Attending
## 3   Semi-Concrete                           2
## 5            Wood                           7
```

```
# f. Select only the variable "Types_of_Houses"
types_houses <- household$Types_of_Houses
types_houses
```

```
## [1] "Wood"          "Concrete"      "Semi-Concrete" "Concrete"
## [5] "Wood"
```

```
# g. Select all male respondents whose father's occupation was Farmer
subset(household, Sex == "Male" & Fathers_Occupation == "Farmer")
```

```
##    Sex Types_of_Houses Fathers_Occupation Number_of_Siblings_Attending
## 1 Male            Wood             Farmer                           3
## 3 Male   Semi-Concrete             Farmer                           2
```

```
# h. Select all female respondents with >=5 siblings attending school
subset(household, Sex == "Female" & Number_of_Siblings_Attending >= 5)
```

```
##      Sex Types_of_Houses Fathers_Occupation Number_of_Siblings_Attending
## 2 Female        Concrete             Driver                           5
## 4 Female        Concrete             Others                           6
## 5 Female            Wood             Farmer                           7
```

```
# 2. Creating an 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:"
```

```
str(df)
```

```
## 'data.frame':    0 obs. of  5 variables:
##  $ Ints      : int
```

```
##  $ Doubles   : num
##  $ Characters: chr
##  $ Logicals  : logi
##  $ Factors   : Factor w/ 0 levels:
```

```
# 3. Importing and Converting CSV Data

# a. Save and import the CSV file
write.csv(household, "HouseholdData.csv", row.names = FALSE)
imported_data <- read.csv("HouseholdData.csv")
imported_data
```

```
##      Sex Types_of_Houses Fathers_Occupation Number_of_Siblings_Attending
## 1   Male            Wood             Farmer                           3
## 2 Female        Concrete             Driver                           5
## 3   Male   Semi-Concrete             Farmer                           2
## 4 Female        Concrete             Others                           6
## 5 Female            Wood             Farmer                           7
```

```
# b. Convert Sex to factor, then to integer (Male = 1, Female = 2)
imported_data$Sex <- factor(imported_data$Sex, levels = c("Male", "Female"), labels = c(1, 2))
imported_data$Sex <- as.integer(imported_data$Sex)
imported_data
```

```
##   Sex Types_of_Houses Fathers_Occupation Number_of_Siblings_Attending
## 1   1            Wood             Farmer                           3
## 2   2         Concrete             Driver                           5
## 3   1    Semi-Concrete             Farmer                           2
## 4   2         Concrete             Others                           6
## 5   2            Wood             Farmer                           7
```

```
# c. Convert Type of Houses to factor, then to integer (Wood = 1, Concrete = 2, Semi-Concrete = 3)
imported_data$Types_of_Houses <- factor(imported_data$Types_of_Houses,
                                        levels = c("Wood", "Concrete", "Semi-Concrete"),
                                        labels = c(1, 2, 3))
imported_data$Types_of_Houses <- as.integer(imported_data$Types_of_Houses)
imported_data
```

```
##   Sex Types_of_Houses Fathers_Occupation Number_of_Siblings_Attending
## 1   1               1             Farmer                           3
## 2   2               2             Driver                           5
## 3   1               3             Farmer                           2
## 4   2               2             Others                           6
## 5   2               1             Farmer                           7
```

```
# d. Convert Fathers_Occupation to factor, then to integer (Farmer = 1, Driver = 2, Others = 3)
imported_data$Fathers_Occupation <- factor(imported_data$Fathers_Occupation,
                                           levels = c("Farmer", "Driver", "Others"),
                                           labels = c(1, 2, 3))
imported_data$Fathers_Occupation <- as.integer(imported_data$Fathers_Occupation)
imported_data
```

```
##   Sex Types_of_Houses Fathers_Occupation Number_of_Siblings_Attending
## 1   1               1                 1                           3
## 2   2               2                 2                           5
## 3   1               3                 1                           2
## 4   2               2                 3                           6
```

```
## 5   2              1                  1                            7
```

```
# e. Select all female respondents (Sex = 2) whose father's occupation is Driver (2)
subset(imported_data, Sex == 2 & Fathers_Occupation == 2)
```

```
##   Sex Types_of_Houses Fathers_Occupation Number_of_Siblings_Attending
## 2   2              2                  2                            5
```

```
# f. Select respondents with >=5 siblings attending school
subset(imported_data, Number_of_Siblings_Attending >= 5)
```

```
##   Sex Types_of_Houses Fathers_Occupation Number_of_Siblings_Attending
## 2   2              2                  2                            5
## 4   2              2                  3                            6
## 5   2              1                  1                            7
```

```
# 4. Interpretation
# The dataset represents household information including gender, housing type, father's occupation, and
# The numeric conversions allow for easier statistical analysis and plotting later.
```