

RWorksheet_Arlante#3.Rmd

Stephanie Lois Amber Arlante

2024-10-02

#1. Data Frame

#a: Create a Data Frame

```
df <- data.frame(
  Respondent = 1:5,
  Sex = c("Male", "Female", "Female", "Male", "Male"),
  FathersOccupation = c("Farmer", "Driver", "Others", "Farmer", "Others"),
  SiblingsAttending = c(4, 5, 6, 3, 5),
  TypeOfHouse = c("Wood", "Concrete", "Semi-Concrete", "Wood", "Concrete")
)
(df)
```

	Respondent	Sex	FathersOccupation	SiblingsAttending	TypeOfHouse
## 1	1	Male	Farmer	4	Wood
## 2	2	Female	Driver	5	Concrete
## 3	3	Female	Others	6	Semi-Concrete
## 4	4	Male	Farmer	3	Wood
## 5	5	Male	Others	5	Concrete

#b: Structure and Summary of the Data

```
str(df)
```

```
## 'data.frame':    5 obs. of  5 variables:
## $ Respondent      : int  1 2 3 4 5
## $ Sex              : chr  "Male" "Female" "Female" "Male" ...
## $ FathersOccupation: chr  "Farmer" "Driver" "Others" "Farmer" ...
## $ SiblingsAttending: num  4 5 6 3 5
## $ TypeOfHouse      : chr  "Wood" "Concrete" "Semi-Concrete" "Wood" ...
```

```
summary(df)
```

	Respondent	Sex	FathersOccupation	SiblingsAttending
## Min. :1	Length:5		Length:5	Min. :3.0
## 1st Qu.:2	Class :character		Class :character	1st Qu.:4.0
## Median :3	Mode :character		Mode :character	Median :5.0
## Mean :3				Mean :4.6
## 3rd Qu.:4				3rd Qu.:5.0
## Max. :5				Max. :6.0
## TypeOfHouse				
## Length:5				
## Class :character				
## Mode :character				
##				
##				

```
##
#c: Checks if the mean number of siblings attending is 5
meansiblings <- mean(df$SiblingsAttending)
(meansiblings)

## [1] 4.6
meansiblings == 5 # Check if the mean is equal to 5

## [1] FALSE
#d: Extract first two rows and all columns
subsetdf <- df[1:2, ]
(subsetdf)

##   Respondent   Sex FathersOccupation SiblingsAttending TypeOfHouse
## 1          1  Male           Farmer                4         Wood
## 2          2 Female           Driver                5         Concrete

#e: Extract 3rd and 5th rows with 2nd and 4th columns
subsetdf <- df[c(3, 5), c(2, 4)]
(subsetdf)

##      Sex SiblingsAttending
## 3 Female                6
## 5  Male                  5

#f: Select the variable "Type_of_House" and store it in types_houses
typeshouses <- df$TypeOfHouse
(typeshouses)

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

#g: Select male respondents with father's occupation as "Farmer"
malefarmers <- subset(df, Sex == "Male" & FathersOccupation == "Farmer")
(malefarmers)

##   Respondent   Sex FathersOccupation SiblingsAttending TypeOfHouse
## 1          1  Male           Farmer                4         Wood
## 4          4  Male           Farmer                3         Wood

#h: Select female respondents with 5 or more siblings attending school
femalesiblings <- subset(df, Sex == "Female" & SiblingsAttending >= 5)
(femalesiblings)

##   Respondent   Sex FathersOccupation SiblingsAttending TypeOfHouse
## 2          2 Female           Driver                5         Concrete
## 3          3 Female           Others                6 Semi-Concrete

#2: Empty Data Frame

dfempty <- data.frame(
  Ints = integer(),
  Doubles = double(),
  Characters = character(),
  Logicals = logical(),
  Factors = factor(),
  stringsAsFactors = FALSE
)
```

```

)

("Structure of the empty dataframe:")

## [1] "Structure of the empty dataframe:"
str(dfempty)

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

#3. HouseholdData.csv

#a: Save the Data Frame as a CSV File
write.csv(df, "HouseholdData.csv", row.names = FALSE)

#b: Import the CSV File into the R Environment
dfimported <- read.csv("HouseholdData.csv")
(dfimported)

##   Respondent   Sex FathersOccupation SiblingsAttending  TypeOfHouse
## 1          1  Male           Farmer              4         Wood
## 2          2 Female           Driver              5         Concrete
## 3          3 Female           Others              6 Semi-Concrete
## 4          4  Male           Farmer              3         Wood
## 5          5  Male           Others              5         Concrete

#c: Convert "Sex" into a factor and change it into integers (Male = 1, Female = 2)
dfimported$Sex <- factor(dfimported$Sex, levels = c("Male", "Female"), labels = c(1, 2))
(dfimported$Sex)

## [1] 1 2 2 1 1
## Levels: 1 2

#d: Convert "Type_of_House" into a factor (Wood = 1, Concrete = 2, Semi-Concrete = 3)
dfimported$TypeOfHouse <- factor(dfimported$TypeOfHouse, levels = c("Wood", "Concrete", "Semi-Concrete"), labels = c(1, 2, 3))
(dfimported$TypeOfHouse)

## [1] 1 2 3 1 2
## Levels: 1 2 3

#e: Factor father's occupation (Farmer = 1, Driver = 2, Others = 3)
dfimported$FathersOccupation <- factor(dfimported$FathersOccupation, levels = c("Farmer", "Driver", "Others"), labels = c(1, 2, 3))
(dfimported$FathersOccupation)

## [1] 1 2 3 1 3
## Levels: 1 2 3

#f: Select female respondents whose father is a driver
femaledriver <- subset(dfimported, Sex == 2 & FathersOccupation == 2)
(femaledriver)

##   Respondent Sex FathersOccupation SiblingsAttending TypeOfHouse
## 2          2  2              2              5         2

```

```
#g: Select respondents with 5 or more siblings attending school
siblings5plus <- subset(dfimported, SiblingsAttending >= 5)
(siblings5plus)
```

```
##   Respondent Sex FathersOccupation SiblingsAttending TypeOfHouse
## 2          2   2                 2                 5             2
## 3          3   2                 3                 6             3
## 5          5   1                 3                 5             2
```