Experiment 4: Data Frames and Factors

# Data Frames

## Creating a Data Frame

n <- 1:5  
s <- rep('Name', 5)  
x <- c(rep(TRUE, 3), rep(FALSE, 2))  
  
df <- data.frame(n, s, x)  
df

## n s x  
## 1 1 Name TRUE  
## 2 2 Name TRUE  
## 3 3 Name TRUE  
## 4 4 Name FALSE  
## 5 5 Name FALSE

class(df)

## [1] "data.frame"

class(df[1])

## [1] "data.frame"

class(df[[1]])

## [1] "integer"

is.vector(df[[1]])

## [1] TRUE

class(df$n)

## [1] "integer"

## Naming columns

names(df) = c("Number", "Name", "Student")  
df

## Number Name Student  
## 1 1 Name TRUE  
## 2 2 Name TRUE  
## 3 3 Name TRUE  
## 4 4 Name FALSE  
## 5 5 Name FALSE

## Strings As Factors

class(df$Name)

## [1] "factor"

df2 <- data.frame(n, s, x, stringsAsFactors = F)  
class(df2$s)

## [1] "character"

## Create a sample data frame

serialNum <- 1:10  
names <- paste("Name", as.character(serialNum))  
salary <- 33000 \* serialNum  
joiningDates <- rep(Sys.Date(), length(serialNum))  
  
employees <- data.frame(  
 serialNum, names, salary, joiningDates,  
 stringsAsFactors = F  
)  
print(employees)

## serialNum names salary joiningDates  
## 1 1 Name 1 33000 2019-01-29  
## 2 2 Name 2 66000 2019-01-29  
## 3 3 Name 3 99000 2019-01-29  
## 4 4 Name 4 132000 2019-01-29  
## 5 5 Name 5 165000 2019-01-29  
## 6 6 Name 6 198000 2019-01-29  
## 7 7 Name 7 231000 2019-01-29  
## 8 8 Name 8 264000 2019-01-29  
## 9 9 Name 9 297000 2019-01-29  
## 10 10 Name 10 330000 2019-01-29

# Factors

## Play with Factors

gender <- factor(c("Male", "Female", "Others"))  
class(gender)

## [1] "factor"

directions <- c("East", "West", "North", "South")  
print(directions)

## [1] "East" "West" "North" "South"

print(is.factor(directions))

## [1] FALSE

gender[[1]]

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

levels(gender) <- c("male", "female", "others")  
gender

## [1] female male others  
## Levels: male female others

print(as.factor(directions))

## [1] East West North South  
## Levels: East North South West

## Create a Data Frame and Check for factors

height <- seq(30, 50, by = 1.8)  
weight <- seq(40, 60, by = 1.8)  
halfLength <- length(height) / 2  
gender <- c(rep('Male', halfLength), rep('Female', halfLength))  
  
bmi <- data.frame(height, weight, gender)  
bmi

## height weight gender  
## 1 30.0 40.0 Male  
## 2 31.8 41.8 Male  
## 3 33.6 43.6 Male  
## 4 35.4 45.4 Male  
## 5 37.2 47.2 Male  
## 6 39.0 49.0 Male  
## 7 40.8 50.8 Female  
## 8 42.6 52.6 Female  
## 9 44.4 54.4 Female  
## 10 46.2 56.2 Female  
## 11 48.0 58.0 Female  
## 12 49.8 59.8 Female

is.factor(bmi$gender)

## [1] TRUE