```r
# Data structure: 4. Data frame
# A data frame is a two dimensional format of data structure that is useful to store
# data in tables. In general, each column is for a different item and
# each row is for a different unit.
# So, if you are interested in a specific item (for example, name or grade), you need to
# choose a column. If you are interested in information for a specific unit,
# you can choose a row.
# It is a collection of variables of the same length,
# but possibly of different type of variables.(numeric, factor, character)
# <-> Elements of a matrix should be of the same type.

rm(list=ls())
name <- c("Tom", "James", "Mary", "Paker")
score <- c(9, 7, 6, 10)
grade <- factor(c("A", "C", "D", "A"), ordered=TRUE, levels=c("D","C","B","A"))
Econ_dep <- data.frame(name, score, grade)

Econ_dep

str(Econ_dep) # gives us the number of observations (# of rows),
              # the number of variables (# of columns), full list of the variable names,

# Name is not a factor (categorical variable). Two persons with the same name are not
in the same category.
# A character vector is converted to a factor in data.frame if we don't specify as a
character.
Econ_dep1 <- data.frame(name, score, grade, stringsAsFactors = FALSE)
str(Econ_dep1)
Econ_dep2 <- data.frame(I(name), score, grade)
str(Econ_dep2)

Econ_dep$name
name

rm(name, score, grade)

# subset selection and subset elimination
Econ_dep[c(3,1),]
Econ_dep[c(1,3),]
Econ_dep[-c(2,3),]
Econ_dep[c(TRUE,TRUE,FALSE,FALSE),]
Econ_dep[Econ_dep$name=="James",]
Econ_dep[!(Econ_dep$name=="James"),]
subset(x=Econ_dep, subset=!(Econ_dep$name=="James"))  # subset()

Econ_dep[,2]
Econ_dep[,-2]
Econ_dep[,c("name","grade")]
Econ_dep[,c(TRUE, FALSE, TRUE)]
Econ_dep[,!c(TRUE, FALSE, TRUE)]

Econ_dep$grade <- NULL   # eliminate "grade" column
Econ_dep

# Add a vector to data.frame
attendance <- c("all", "some", "never", "all")
Econ_dep <- cbind(Econ_dep,attendance)
Econ_dep$gender <- c("Male", "Male", "Female", "Female")
str(Econ_dep)                          # Econ_dep$gender is a character vector not a
factor.

Econ_dep$year <- NA
Econ_dep$year[Econ_dep$name %in% c("Tom", "Mary")] <- 4
Econ_dep

# Sort midterm using order() function
rank <- order(Econ_dep$score, decreasing = TRUE)
Econ_dep[rank,]

```

```r
attach(Econ_dep)          # data.frame "midterm" is attached to the R search path.
                          # So, we don't need to put midterm$ to call a variable
                          # in midterm

name

detach(Econ_dep)
name

as.matrix(Econ_dep)       # Easy to convert data.frame to matrix

# Grouped data
ID <- 1:20
rand.number <- rnorm(20)
participant <- data.frame(ID, rand.number)

# Generate two data frames "group1" and "group2" from "participant"

group1 <- participant[participant$rand.number > 0,]   # Conditional selection
group2 <- subset(participant, subset = rand.number <= 0)  # You can use subset function.

# Create a new variable "level" in "participant" to group data.

participant$level[participant$rand.number > 0.5] <- "first"
participant$level[participant$rand.number > 0 & participant$rand.number <= 0.5] <-
"second"
participant$level[participant$rand.number > -0.5 & participant$rand.number <= 0] <-
"third"
participant$level[participant$rand.number <= -0.5] <- "fourth"

str(participant)
participant$level <- as.factor(participant$level)
levels(participant$level) = c("first", "second", "third", "fourth")

participant$group <- as.numeric(participant$level)

participant

# Exercise 1.
################################################################################
# Print out bulit-in R data frame
mtcars

# Try the function head() on mtcars.
head(mtcars)         # head() enables us to see the first few observations of a data
frame.

# Similarly tail() prints out the last few observation in the data set.
tail(mtcars)

str(mtcars)

# Regress mpg on cyl and hp. (Chekc lm() function)


# Select the rows for mercedes. Hint: They include "Merc" and check grep() function


# Exercise 2.
################################################################################
name <- c("Apple", "MS", "Google", "Honda", "GM", "Volks", "Hyundai", "Amazon")
type <- c("IT", "IT", "IT", "Auto", "Auto", "Auto", "Auto", "IT")
stock <- c(165.5, 55.48, 1119.20, 36.16, 41, 172.06, 162.5, 1429.95)
US <- c(TRUE, TRUE, TRUE, FALSE, TRUE, FALSE, FALSE, TRUE)

# Creat a data from from the vectors
portfolio <-

rm(name, type, stock, US)
```

```
131
132  # Check the structure of planet_df (Use str() function)
133
134  portfolio$name <- as.character(portfolio$name)
135
136  # Selection from data.frame.
137    # Print out stock price of google
138
139    # Print out data for Google (entire third row)
140
141    # Print out the first 5 values of stock column.
142
143    # Print out data for IT companies.
144
145    # Print out data for companies whose stock price is lower than Apple.
146
147
148  # Sort data.frame using order() function. From the most expensive one.
149  # Use rank to sort portforlio
150
151
```