

데이터과학을 위한 **R**프로그래밍

1주차. 데이터과학과 오픈소스



이혜선 교수

포항공과대학교 산업경영공학과



목차

1주차. 데이터과학과 오픈소스

1차시

오픈소스와 R프로그램

2차시

R프로그램 실행(RStudio)

3차시

기본스크립트와 R 추가패키지



1주차

2차시

R프로그램 실행 (RStudio)

RStudio의 레이아웃

The screenshot shows the RStudio IDE with the following components and annotations:

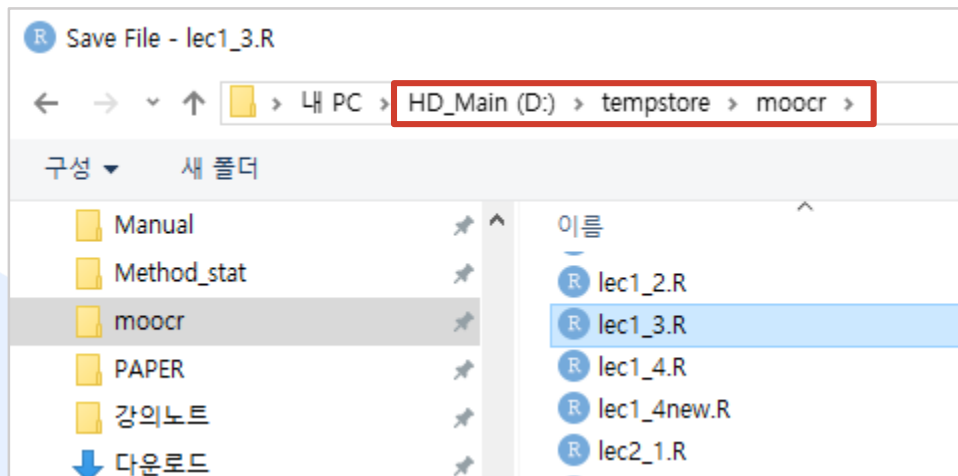
- Source Editor:** Contains R code for creating a vector `x1` and checking its class. An orange callout bubble points to the code with the text: "프로그램 입력창 (불러오기, 저장..)".
- Environment Pane:** Shows the current environment with a variable `x1` of type `num` containing values `[1:5] 1 3 5 7 9`. An orange callout bubble points to it with the text: "현재 생성된 데이터를 보여줌".
- Plots Pane:** Located at the bottom, it is currently empty. A red box highlights the `Plots` tab, with an orange callout bubble pointing to it saying "Plot(그래픽)".
- Help Pane:** Located at the bottom right, it shows a list of installed and available packages. A red box highlights the `Help` tab, with an orange callout bubble pointing to it saying "Help(도움말)".

The code in the Source Editor is as follows:

```
1 # lec1_2.r
2 # description
3
4 x1<-c(1,3,5,7,9)
5
6 # x is numeric or character?
7 class(x1)
8 is.numeric(x1)
9
10 # integer is numeric
11 is.integer(x1)
12
13 # need to define as.integer as.numeric as.character
14 x2<-as.integer(x1)
15
16 # length of x
17 length(x1)
18
19 # x is a vector?
20 is.vector(x1)
21
22 # class - character
23 "I like apple"
```

R프로그램 관리

☒ *.r (R프로그램)을 지정한 폴더에 저장

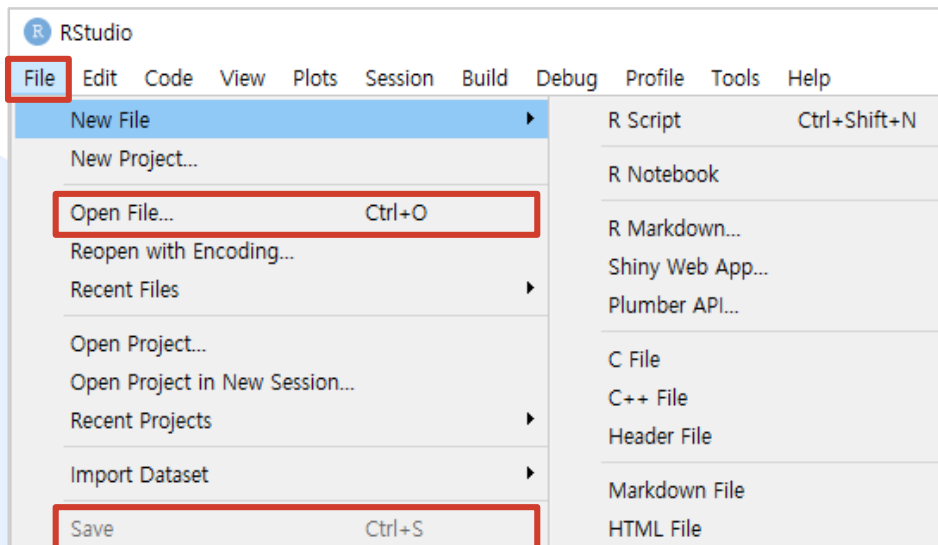


● R프로그램 열기/저장

☑ R프로그램 열기/저장

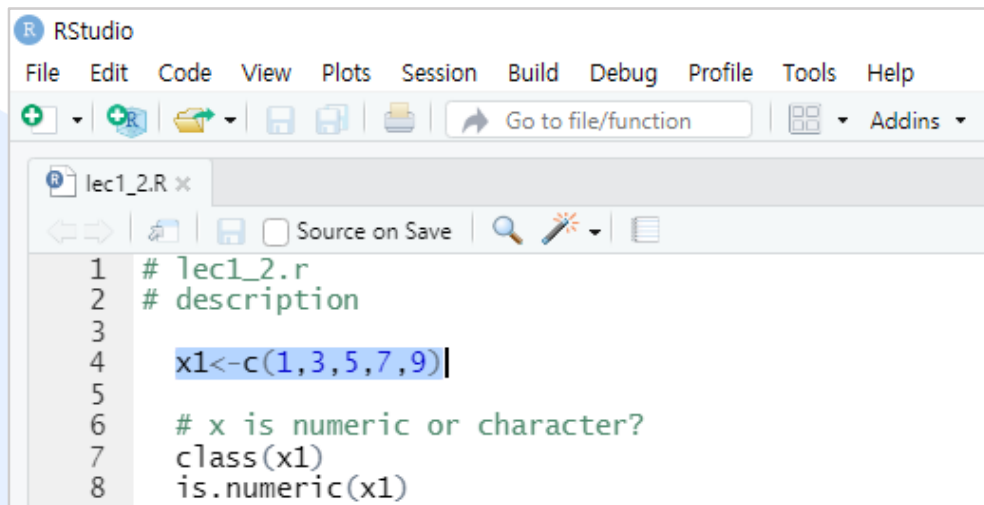
➤ File_Open File (lec1_2.r을 열기)

➤ File_Save



● R프로그램 수행

- ✓ R프로그램을 수행하는 단축키 : (Ctrl & Enter)
- ✓ 수행하고자 하는 프로그램 부분을 선택하고 수행
- ✓ 예제 : 첫 줄을 선택하고 수행 → (1,3,5,7,9)을 가진 벡터 ×1 생성



```
1 # lec1_2.r
2 # description
3
4 x1<-c(1,3,5,7,9)
5
6 # x is numeric or character?
7 class(x1)
8 is.numeric(x1)
```

R프로그램 수행

☑ lec1_2.r의 첫째 줄을 수행하면, 환경창에 x1 데이터가 생성됨을 보여줌

The screenshot shows the RStudio interface. The script editor on the left contains the following code:

```
1 # lec1_2.r
2 # description
3
4 x1<-c(1,3,5,7,9)
5
6 # x is numeric or character?
7 class(x1)
8 is.numeric(x1)
```

A red box highlights the line `x1<-c(1,3,5,7,9)`, and a red arrow points from it to the Environment pane on the right. The Environment pane shows the Global Environment with a variable `x1` of type `num`. The values are displayed as `[1:5] 1 3 5 7 9`.

Environment	History	Connections	Tutorial
Global Environment	Import Dataset		
Values			
x1			
num [1:5]			
1 3 5 7 9			

● R프로그램 수행

☑ 프로그램 코드설명 (lec1_2.r)

```
# lec1_2.r
# description

x1<-c(1,3,5,7,9)

# x is numeric or character?
class(x1)
is.numeric(x1)

# integer is numeric
is.integer(x1)

# need to define as.integer as.numeric as.character
x2<-as.integer(x1)

# length of x
length(x1)

# x is a vector?
is.vector(x1)

# class - character
"I like apple"
class("I like apple")
```

으로 시작하는 줄은 프로그램에 대한 설명
(자동적으로 초록색 폰트로 바뀜)

검정색 폰트의 줄은 R프로그램코드

R프로그램 수행

✓ 프로그램 코드설명 (lec1_2.r)

프로그램 편집 창

```
# lec1_2.r
# description

x1<-c(1,3,5,7,9)

# x1 is numeric or character?
class(x1)
is.numeric(x1)

# integer is numeric
is.integer(x1)

# need to define as.integer as.r
x2<-as.integer(x1)

# length of x1
length(x1)

# x1 is a vector?
is.vector(x1)

# class - character
"I like apple"
class("I like apple")
```

컨솔창

```
> x1<-c(1,3,5,7,9)
> x1
[1] 1 3 5 7 9
> class(x1)
[1] "numeric"
> is.numeric(x1)
[1] TRUE
> length(x1)
[1] 5
> is.vector(x1)
[1] TRUE
```

X의 범주는? (숫자 혹은 문자)

X는 숫자?

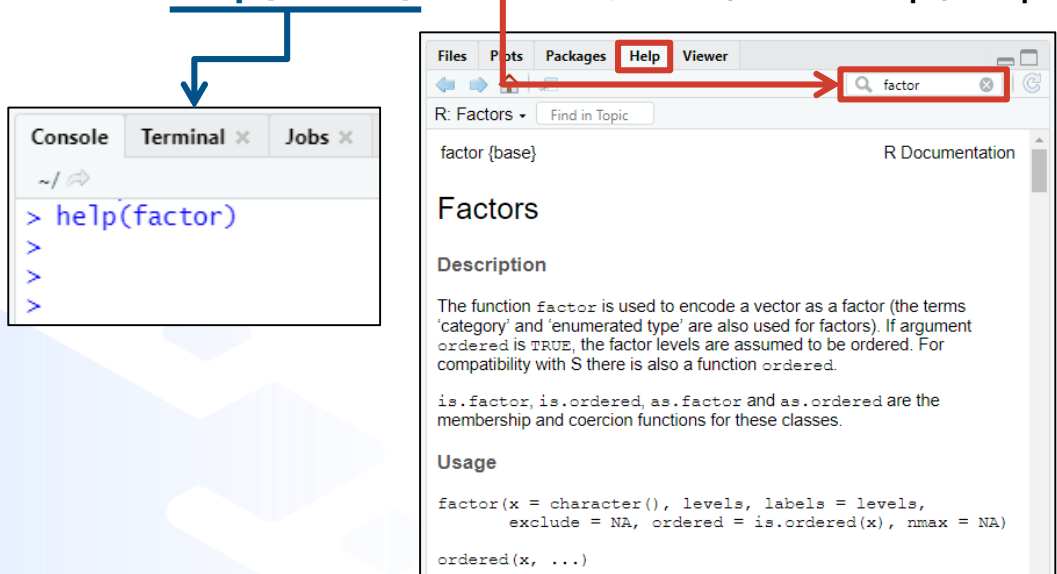
X의 길이는?

X는 벡터?

R 도움말의 활용

✓ help (도움말)의 활용

- ▶ Help창을 선택하고 검색란에 **'factor'**를 입력하면 매뉴얼로 자동 연결됨
- ▶ 혹은 콘솔창에 **help(factor)**라고 치면 동일함 (예제 : help(boxplot))



```
Console Terminal x Jobs x
~/
> help(factor)
>
>
>
```

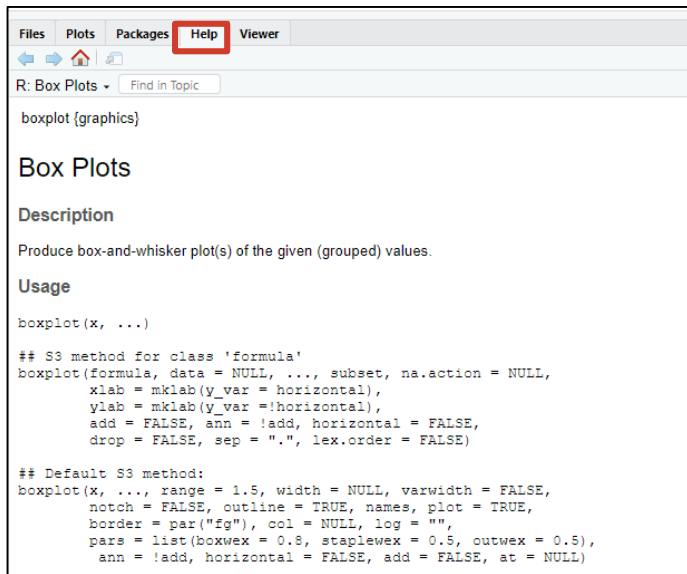


● R 도움말의 활용

☑ help (도움말)의 활용

- ▶ Help창을 선택하고 검색란에 **'boxplot'**를 입력하면 매뉴얼로 자동 연결됨
- ▶ 혹은 콘솔창에 **help(boxplot)**라고 치면 동일함 (예제 : help(boxplot))

help(boxplot)

A screenshot of the RStudio Help window. The 'Help' tab is selected in the top bar. The search bar at the top contains 'boxplot'. The main content area displays the help page for 'Box Plots', including a description, usage, and the underlying R code for the boxplot function.

```
boxplot (graphics)

Box Plots

Description
Produce box-and-whisker plot(s) of the given (grouped) values.

Usage
boxplot(x, ...)

## S3 method for class 'formula'
boxplot(formula, data = NULL, ..., subset, na.action = NULL,
  xlab = mklab(y_var = horizontal),
  ylab = mklab(y_var = !horizontal),
  add = FALSE, ann = !add, horizontal = FALSE,
  drop = FALSE, sep = ".", lex.order = FALSE)

## Default S3 method:
boxplot(x, ..., range = 1.5, width = NULL, varwidth = FALSE,
  notch = FALSE, outline = TRUE, names, plot = TRUE,
  border = par("fg"), col = NULL, log = "",
  pars = list(boxwex = 0.8, staplewex = 0.5, outwex = 0.5),
  ann = !add, horizontal = FALSE, add = FALSE, at = NULL)
```

R 함수 검색

✓ help.search(" ")를 사용

▶ 예제 : 선형모형(linear model)을 R에서 수행할 때 어떤 함수를 사용하는지 찾는 방법

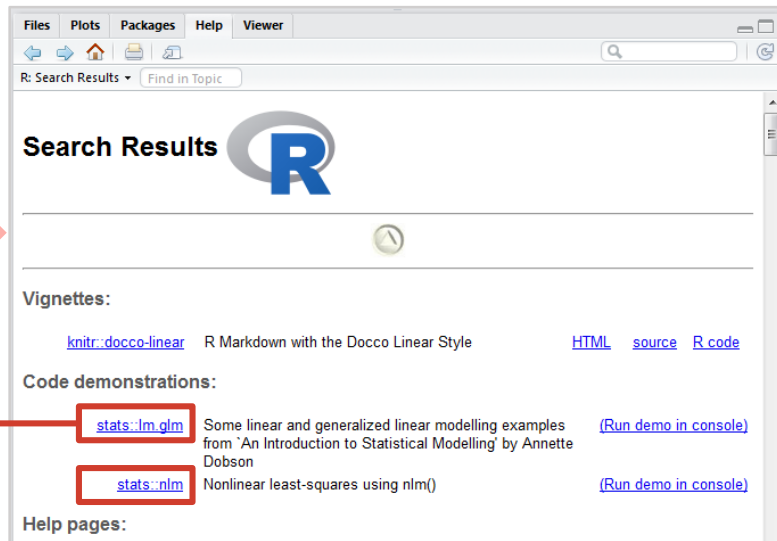
프로그램 편집 창

```
# to find out the package  
help.search("linear")
```

stats lm.glm

패키지 이름

함수이름



✧ stats 모듈은 R프로그램 베이스에 이미 탑재되어 있음!!! (추가패키지가 아님)