

# 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

3차시

# 기본스크립트와 R 추가패키지

## ● 기본스크립트와 함수

☑ 기본연산 (더하기, 곱하기, 제곱근, 최소값, 최대값, 평균)

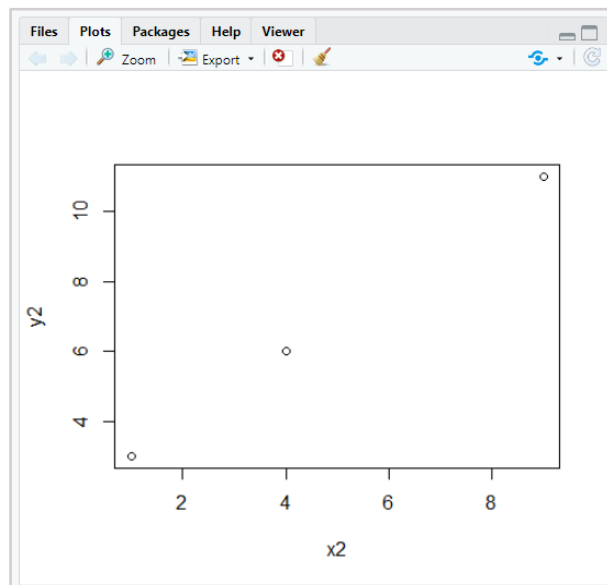
### 프로그램 편집 창

```
# lec1_3.r  
# Basic operations  
  
x2<-c(1,4,9)  
sqrt(x2)  
min(x2)  
max(x2)  
mean(x2)  
  
y2=2+x2  
  
plot(x2, y2)
```

### 콘솔창

```
> x2<-c(1,4,9)  
> sqrt(x2)  
[1] 1 2 3  
> min(x2)  
[1] 1  
> max(x2)  
[1] 9  
> mean(x2)  
[1] 4.666667  
>  
> y2=2+x2  
> y2  
[1] 3 6 11
```

### Plot



## ● 기본스크립트와 함수

### ☑ 함수 (로그, 지수, 사인, 코사인)

```
# functions
```

```
log10(10)
```

```
log(10)
```

```
exp(10)
```

```
sin(pi/2)
```

```
> log10(10)
```

```
[1] 1
```

```
> log(10)
```

```
[1] 2.302585
```

```
> exp(10)
```

```
[1] 22026.47
```

### ☑ 소문자, 대문자 변경, ls(), rm()

```
# list
```

```
ls()
```

```
# remove object
```

```
rm(x1)
```

```
# change to lower or upper case
```

```
c1 <- "MiXeD cAsE 123"
```

```
tolower(c1)
```

```
toupper(c1)
```

소문자로 변경

대문자로 변경

```
> c1 <- "MiXeD cAsE 123"
```

```
> tolower(c1)
```

```
[1] "mixed case 123"
```

```
> toupper(c1)
```

```
[1] "MIXED CASE 123"
```

## ● 기본스크립트와 함수

### ☑ 정규분포로부터 데이터생성 함수 : rnorm

random normal distribution

```
# generating data from distributions
# 100 values from normal distribution with mean=0, sd=1
x3<-rnorm(100)
head(x3)
mean(x3)
sd(x3)
min(x3)
max(x3)

# 10000 values from normal distribution with mean=0, sd=1
x4<-rnorm(10000)
mean(x4)
sd(x4)|
```

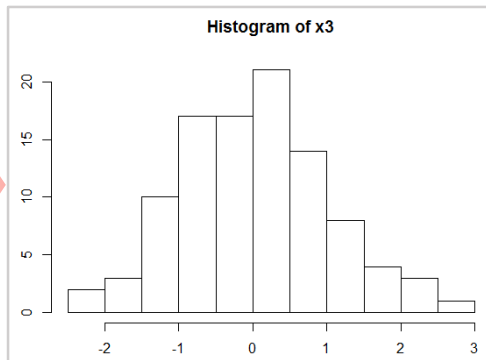
X<-rnorm(n, mean, std)  
X<-rnorm(100,0,5)  
X<-rnorm(100,5,5)

```
> x3<-rnorm(100)
> head(x3)
[1]  1.3759941 -0.1978637  0.9209631  1.1059498 -0.8540965
[6]  1.2116593
> mean(x3)
[1] -0.1774837
> sd(x3)
[1] 1.076244
> min(x3)
[1] -3.345342
> max(x3)
[1] 1.880073
>
> # 10000 values from normal distribution with mean=0, sd=1
> x4<-rnorm(10000)
> mean(x4)
[1] 0.009022544
> sd(x4)
[1] 1.00527
```

## ● 기본스크립트와 함수

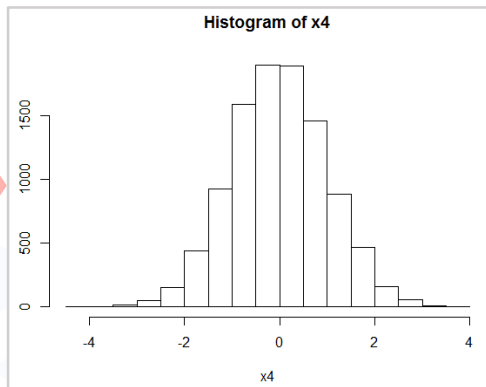
```
x3<-rnorm(100)
# head(x3)
mean(x3)
sd(x3)
min(x3)
max(x3)

hist(x3)
```



```
# 10000 values from normal
x4<-rnorm(10000)
mean(x4)
sd(x4)

hist(x4)
```



# Q1

X3과 X4의  
히스토그램을 보면  
어떤 데이터가 더  
정규분포 형태와 유사한가?

## ● R 추가패키지 설치

☑ R프로그램은 기본 program에 추가적으로 필요한 분석 툴을 설치

그래픽에  
필요한 툴

ggplot2

Grammar of Graphics

텍스트마이닝에  
필요한 패키지

NLP, tm, wordcloud

Natural Language Processing  
text mining  
wordcloud

Base program

ggplot2-package (ggplot2)

R Documentation

ggplot2: Create Elegant Data Visualisations  
Using the Grammar of Graphics

### Description

A system for 'declaratively' creating graphics, based on "The Grammar of Graphics". You provide the data, tell 'ggplot2' how to map variables to aesthetics, what graphical primitives to use, and it takes care of the details.

### Author(s)

**Maintainer:** Hadley Wickham [hadley@rstudio.com](mailto:hadley@rstudio.com)

### Authors:

- Winston Chang [winston@rstudio.com](mailto:winston@rstudio.com)



## ● R 추가패키지 설치

✓ 추가패키지 설치 (install.packages)

프로그램 편집 창

```
# install package
```

1 ggplot2라는 패키지를 설치

```
install.packages("ggplot2")
```

```
library(ggplot2)
```

```
help(ggplot2)
```

2

scatterplot3d 패키지 설치

```
# install scatterp.
```

```
install.packages("scatterplot3d")
```

```
library(scatterplot3d)
```

scatterplot3d

ggplot2



R 기본프로그램

✳ 해당패키지를 사용할 때 library(scatterplot3d)는 사용할 때마다 필요함 !!! (설치는 한번만 함)

## ● R 추가패키지 사용 예제

☑ 3d plot을 그리고자 할 때 : "scatterplot3d" 패키지

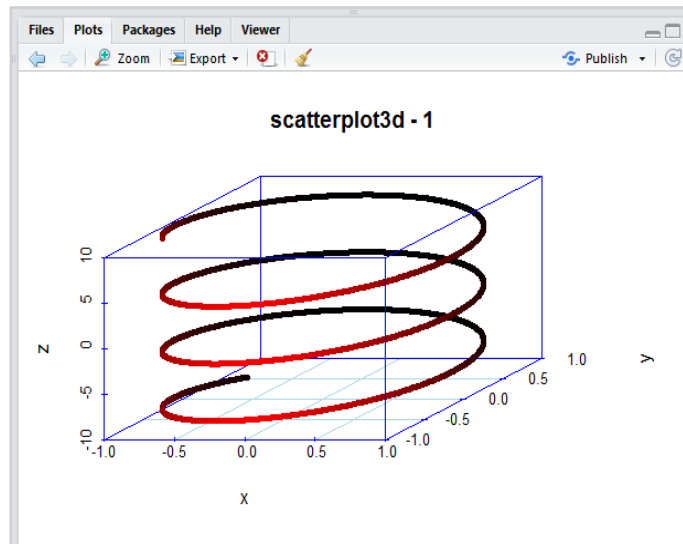
```
# install package

install.packages("ggplot2")
library(ggplot2)
help(ggplot2)

# install scatterplot3d

install.packages("scatterplot3d")
library(scatterplot3d)
help(scatterplot3d)

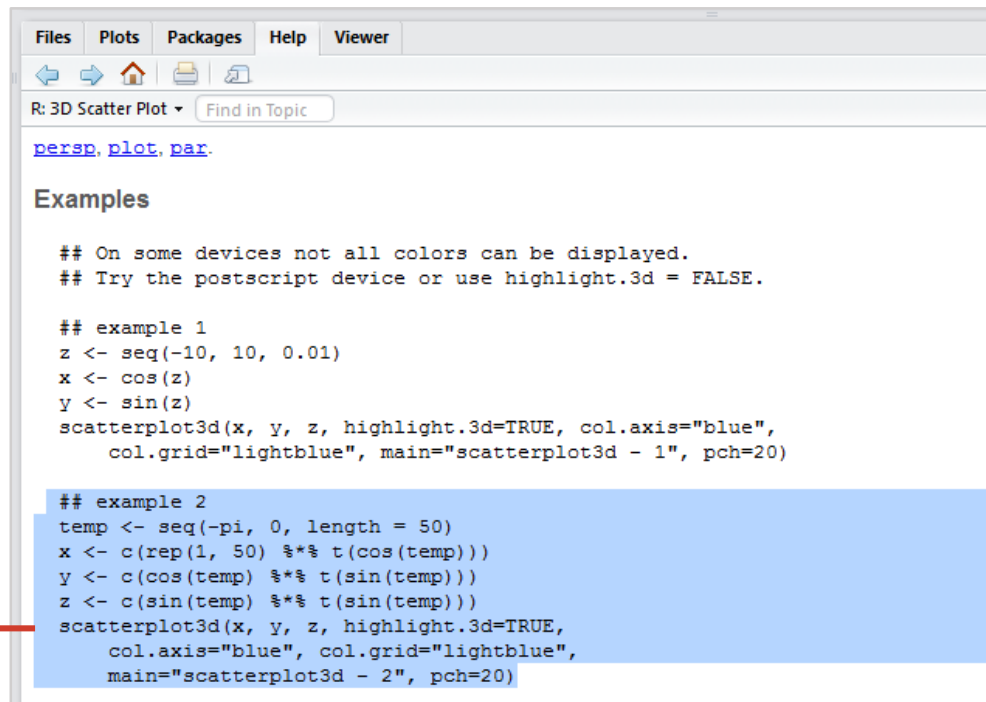
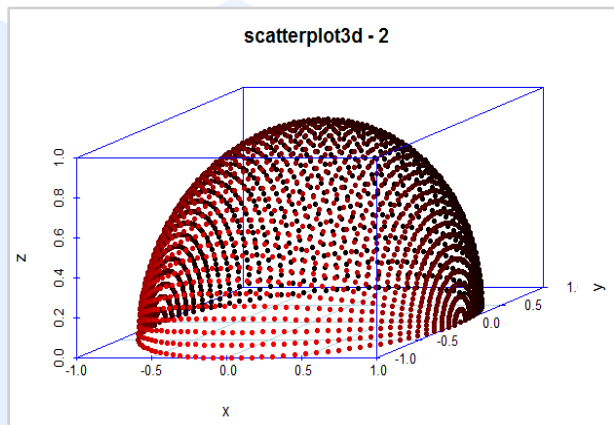
# example program using scatterplot3d
z <- seq(-10, 10, 0.01)
x <- cos(z)
y <- sin(z)
scatterplot3d(x, y, z, highlight.3d=TRUE, col.axis="blue",
              col.grid="lightblue", main="scatterplot3d - 1", pch=20)
```



## R 추가패키지 사용 예제

✓ help 매뉴얼의 예제 프로그램 사용

`help(scatterplot3d)`



```
## On some devices not all colors can be displayed.
## Try the postscript device or use highlight.3d = FALSE.

## example 1
z <- seq(-10, 10, 0.01)
x <- cos(z)
y <- sin(z)
scatterplot3d(x, y, z, highlight.3d=TRUE, col.axis="blue",
  col.grid="lightblue", main="scatterplot3d - 1", pch=20)

## example 2
temp <- seq(-pi, 0, length = 50)
x <- c(rep(1, 50) %*% t(cos(temp)))
y <- c(cos(temp) %*% t(sin(temp)))
z <- c(sin(temp) %*% t(sin(temp)))
scatterplot3d(x, y, z, highlight.3d=TRUE,
  col.axis="blue", col.grid="lightblue",
  main="scatterplot3d - 2", pch=20)
```

## ● R 추가패키지 설치

