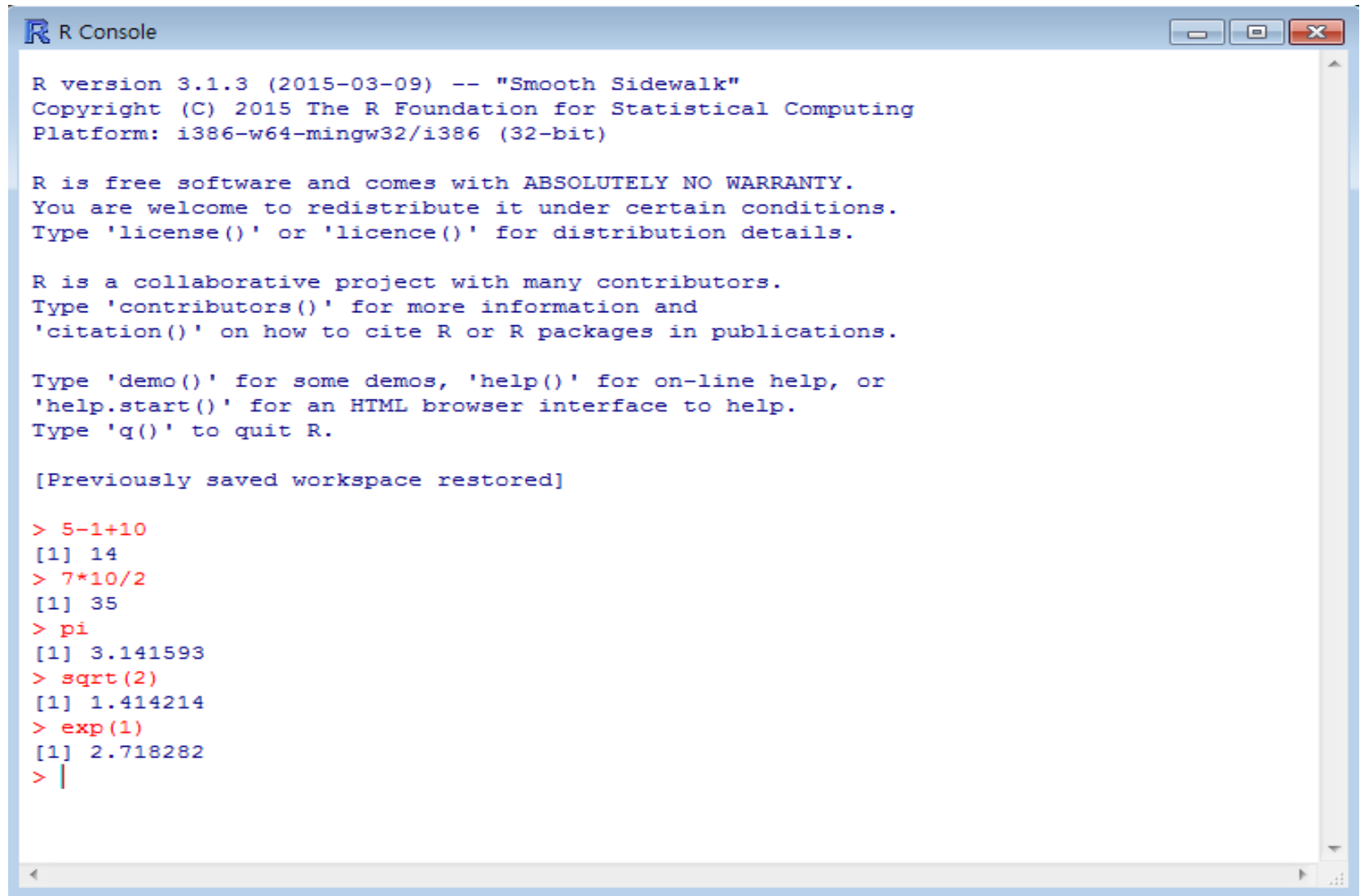


R 기초통계

2017. 11. 2

강창완

1. R as a calculator



```
R R Console

R version 3.1.3 (2015-03-09) -- "Smooth Sidewalk"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> 5-1+10
[1] 14
> 7*10/2
[1] 35
> pi
[1] 3.141593
> sqrt(2)
[1] 1.414214
> exp(1)
[1] 2.718282
> |
```

2. 데이터 입력하기

small dataset, 데이터 직접입력

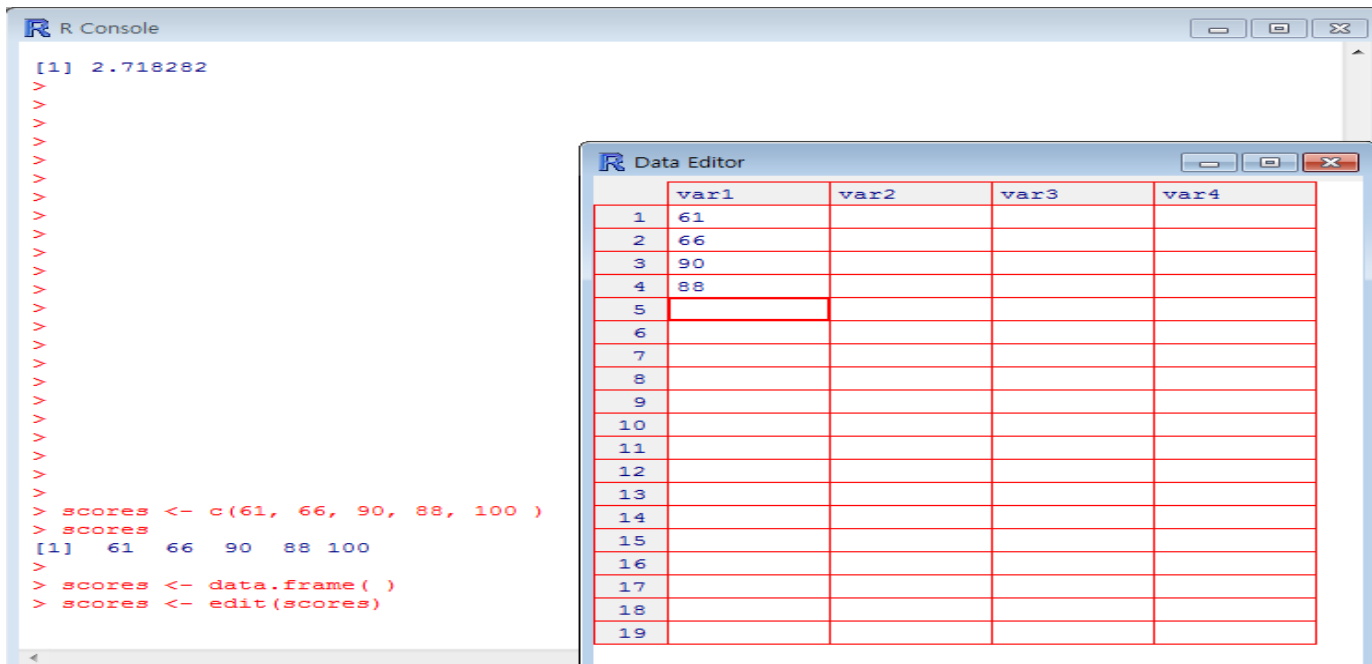
방법 1

```
> scores <- c( 61, 66, 90, 88, 100 )
```

방법 2 : spreadsheet-like editor

```
> scores <- data.frame()
```

```
> scores <- edit(scores)
```



문자형 데이터 입력하기

```
blood = c("O","O","A","B","A","O","A","A","B","O",  
  "B","O","B","O", "O","A","O","O","B","A",  
  "B","A","AB","A","B","A","A","O","O","A",  
  "O","O","A","A","A","O","B","O","O","AB")  
cnt = table(blood)  
prop = prop.table(cnt)  
cbind(cnt,prop)  
barplot(cnt)  
pie(cnt)
```

3. 데이터 불러오기

- 텍스트데이터 불러오기
- `deu <- read.table("c://bigdata/deu_ks.txt", header=TRUE)`

```
R Console
> deu <- read.table("c://bigdata/deu_ks.txt", header=TRUE)
> deu
```

	univ	id	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	p1	p2	p3
1	1	1	2	4	4	2	2	2	3	2	3	2	1	4	2
2	1	2	1	2	2	2	2	3	5	2	3	2	1	1	2
3	1	3	2	3	4	3	3	2	5	4	4	4	1	4	3
4	1	4	3	5	5	3	3	4	5	2	4	4	1	4	2
5	1	5	3	2	4	4	4	3	5	2	3	4	1	1	1
6	1	6	4	3	3	4	2	1	3	4	4	4	1	2	1
7	1	7	3	4	4	2	2	3	5	2	3	3	1	1	3
8	1	8	4	2	4	4	4	4	5	3	4	4	1	1	5
9	1	9	4	4	5	3	4	4	4	3	4	4	1	4	5
10	1	10	4	2	4	2	3	2	5	4	3	4	1	1	5
11	1	11	4	4	4	3	3	5	2	2	4	3	1	3	2
12	1	12	3	1	4	3	4	4	5	1	4	3	1	1	2
13	1	13	3	2	4	3	4	4	4	3	4	4	1	1	2
14	1	14	3	4	3	4	5	5	5	2	3	2	1	1	5
15	1	15	1	3	2	2	3	2	4	2	2	2	1	1	5
16	1	16	3	4	2	3	4	4	5	3	4	4	1	1	5
17	1	17	3	2	1	1	3	2	5	2	2	1	1	1	5
18	1	18	3	3	3	4	3	4	5	4	4	3	1	1	5
19	1	19	2	2	4	3	2	2	5	3	4	4	1	1	5
20	1	20	3	4	4	3	4	3	5	2	4	3	1	1	2
21	1	21	3	4	5	2	4	2	5	2	3	4	1	4	5
22	1	22	4	3	4	2	3	2	5	4	3	2	1	1	1
23	1	23	2	2	2	1	4	3	5	4	1	1	1	1	1
24	1	24	2	3	4	3	5	3	4	3	4	4	2	1	5
25	1	25	3	4	4	2	4	2	5	2	3	3	2	1	1
26	1	26	4	2	3	3	2	4	5	2	4	4	1	1	5
27	1	27	4	4	4	4	3	3	5	3	4	4	2	1	5
28	1	28	4	4	5	3	3	3	5	2	4	3	2	1	5

```
> f<-read.csv("c://bigdata/buytest.csv", header=TRUE)
> summary(f)
```

ID	RESPOND	AGE	INCOME	\$
000054889:	1 Min. :0.0000	Min. :18.00	Min. : 15.0\$	
000219612:	1 1st Qu.:0.0000	1st Qu.:38.00	1st Qu.: 35.0\$	
001044039:	1 Median :0.0000	Median :44.00	Median : 50.0\$	
001079946:	1 Mean :0.0767	Mean :44.56	Mean : 47.9\$	
001108462:	1 3rd Qu.:0.0000	3rd Qu.:51.00	3rd Qu.: 61.0\$	
001109024:	1 Max. :1.0000	Max. :75.00	Max. :114.0\$	
(Other) :9994		NA's :234	NA's :234	\$
MARRIED	FICO	OWNHOME	LOC	\$
Min. :0.0000	Min. :577.0	Min. :0.0000	E :2261	\$
1st Qu.:0.0000	1st Qu.:676.0	1st Qu.:0.0000	F :2168	\$
Median :1.0000	Median :695.0	Median :0.0000	B :1828	\$
Mean :0.5842	Mean :694.3	Mean :0.3341	H :1093	\$
3rd Qu.:1.0000	3rd Qu.:714.0	3rd Qu.:1.0000	G : 950	\$
Max. :1.0000	Max. :800.0	Max. :1.0000	A : 585	\$
NA's :234	NA's :39	NA's :234	(Other):1115	\$
BUY6	BUY12	BUY18	VALUE24	\$
Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. : 60\$	
1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.: 149\$	
Median :0.0000	Median :0.0000	Median :0.0000	Median : 214\$	
Mean :0.1283	Mean :0.2099	Mean :0.3471	Mean : 254\$	
3rd Qu.:0.0000	3rd Qu.:0.0000	3rd Qu.:1.0000	3rd Qu.: 312\$	
Max. :2.0000	Max. :3.0000	Max. :3.0000	Max. :1253\$	

엑셀데이터 불러오기

```
library(readxl)
deu_ks <- read_excel("C:/data/deu_ks.xlsx")
View(deu_ks)
deu_ks
```

	univ	id	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	p1	p2	p3
1	1	1	2	4	4	2	2	2	3	2	3	2	1	4	2
2	1	2	1	2	2	2	2	3	5	2	3	2	1	1	2
3	1	3	2	3	4	3	3	2	5	4	4	4	1	4	3
4	1	4	3	5	5	3	3	4	5	2	4	4	1	4	2
5	1	5	3	2	4	4	4	3	5	2	3	4	1	1	1
6	1	6	4	3	3	4	2	1	3	4	4	4	1	2	1
7	1	7	3	4	4	2	2	3	5	2	3	3	1	1	3
8	-----														

- SPSS데이터 불러오기

```
library(haven)
```

```
deusav<-read_sav("C:/data/deu_ks.sav")
```

```
deusav
```

```
write.table(deusav, "c://data/d2.txt",  
sep="wt", row.names=F, quote=F)
```


문자 데이터 불러오기 (website, HTML, pdf 등)

```
oz<-file("http://www.gutenberg.org/cache/epub/55/pg55.txt")
oz1<-readLines(oz)
oz1
write.table(oz1,"c://data/oz1.txt", sep="Wt",row.names=F,
quote=F)
```

패키지 내 데이터 불러오기

```
library(treemap)
data(GNI2014)
write.table(GNI2014, "c://cwk/a.txt")
a<-read.table("c://cwk/a.txt")
a
```

Package foreign provides function read.ssd() for importing SAS datasets (.sas7bdat les) into R.

```
library(foreign) # for importing SAS data
```

```
# the path of SAS on your computer
```

```
sashome <- "C:/Program Files/SAS/SASFoundation/9.2"
```

```
filepath <- "c://data/"
```

```
# filename should be no more than 8 characters, without extension
```

```
fileName <- "buytot"
```

```
# read data from a SAS dataset
```

```
a <- read.ssd(file.path(filepath), fileName,  
              sascmd=file.path(sashome, "sas.exe"))
```

기초통계 계산하기

Computing Basic Statistics

- `mean()`
- `median()`
- `sd()`
- `var()`
- `cor(x, y)`
- `cov(x, y)`
- `fivenum()`
- `stem()`

R Console

```
> mean(king$age)
[1] 46.03704
> median(king$age)
[1] 45
> sd(king$age)
[1] 16.1733
> var(king$age)
[1] 261.5755
> fivenum(king$age)
[1] 16 33 45 56 82
> stem(king$age)
```

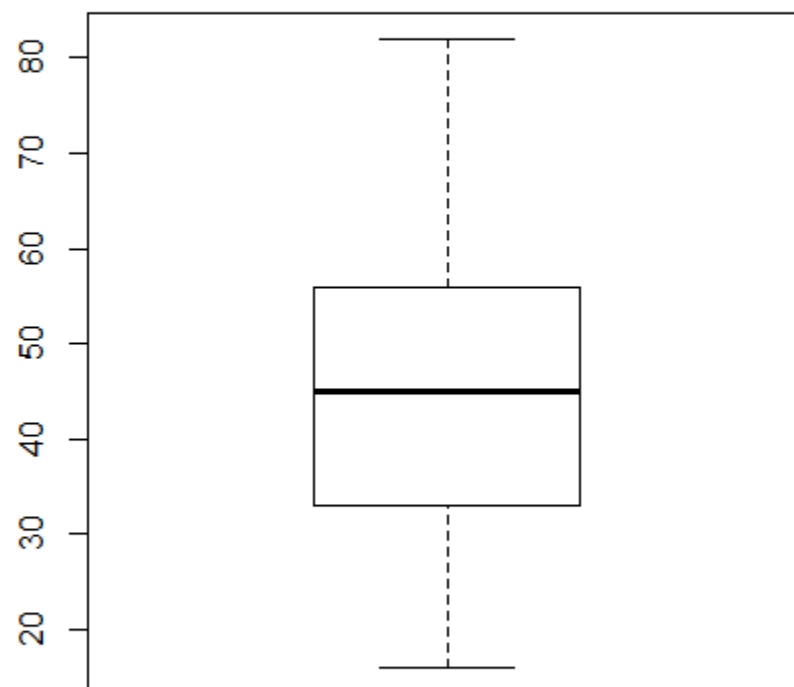
The decimal point is 1 digit(s) to the right of the |

```
1 | 6
2 | 28
3 | 00233678
4 | 0458
5 | 1234669
6 | 267
7 | 3
```

R Console

```
9   sungjong 37
10  yeoun    30
11  jungjong 56
12  injong   30
13  myungjong 33
14  sunjo    56
15  gwang    66
16  injo     54
17  hyojong  40
18  hyunjong 33
19  sukjong  59
20  gyeongjong 36
21  yeongjo  82
22  jeongjo  48
23  sunjo    44
24  hyeonjong 22
25  chuljong 32
26  gojong   67
27  sunjong  52
> boxplot(king$age)
> |
```

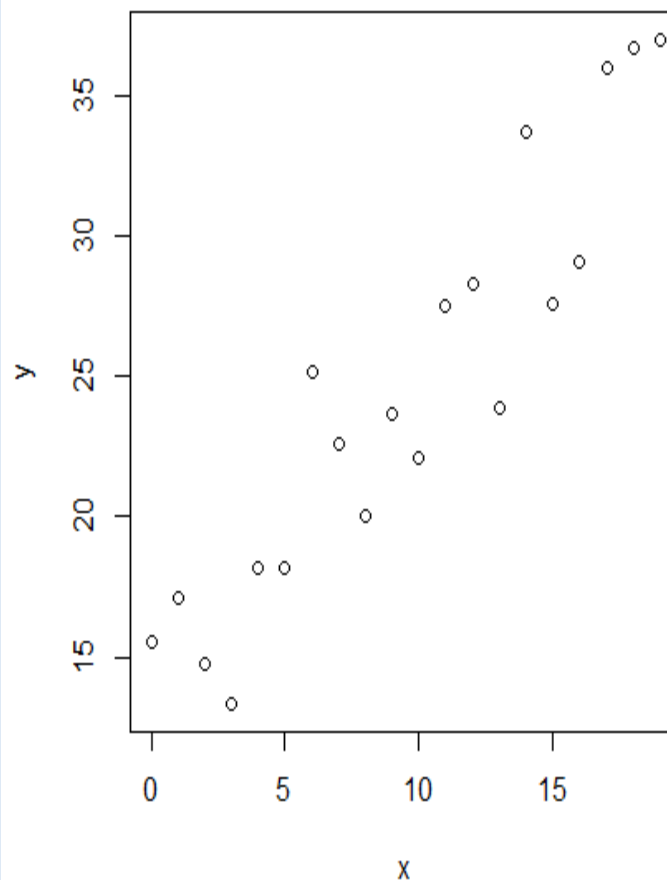
R Graphics: Device 2 (ACTIVE)



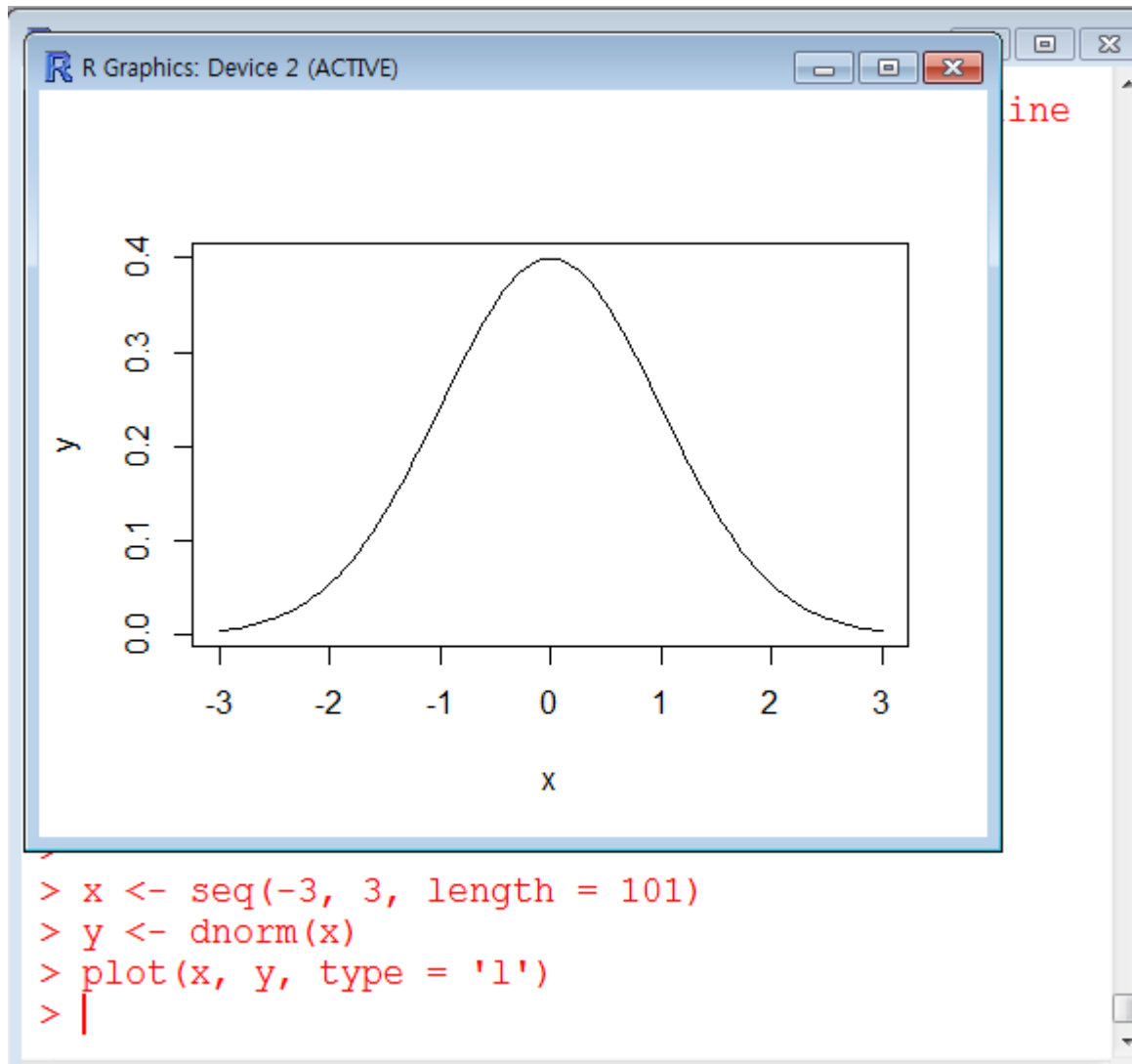
R Console

```
20 gyeongjong 36
21 yeongjo 82
22 jeongjo 48
23 sunjo 44
24 hyeonjong 22
25 chuljong 32
26 gojong 67
27 sunjong 52
> boxplot(king$sage)
>
>
>
> x<- 0:9
> x
[1] 0 1 2 3 4 5 6 7 8 9
> x<- 0:19
> x
[1] 0 1 2 3 4 5 6 7 8 9 10 11 12 13$
> x[5]
[1] 4
> y<- x+runif(20, min=10, max=20)
> plot(x, y)
> |
```

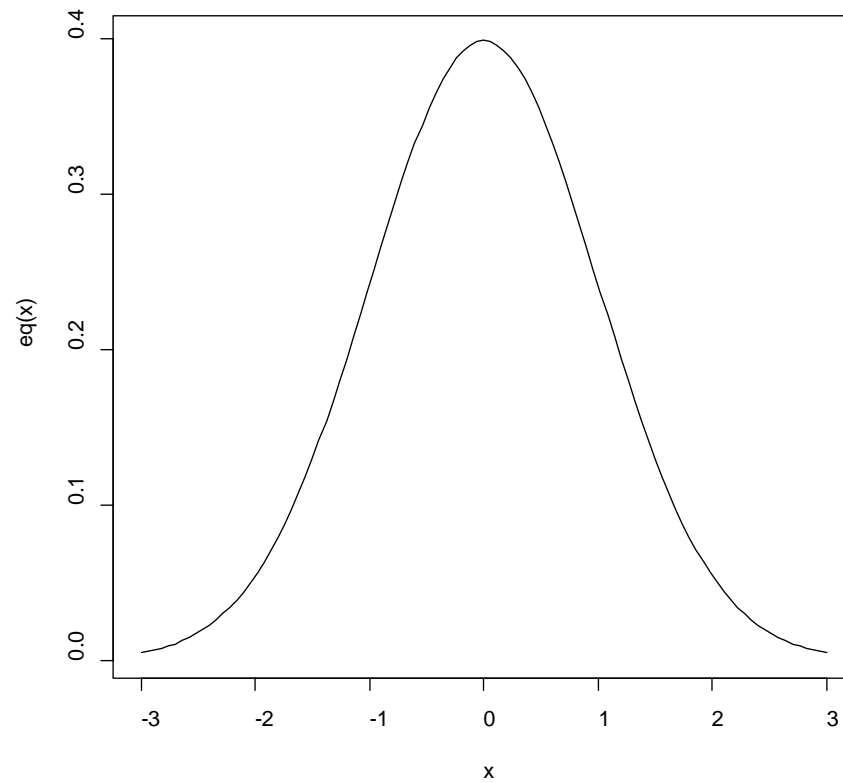
R Graphics: Device 2 (ACTIVE)



- `x <- seq(-3, 3, length = 101)`
- `y <- dnorm(x)`
- `plot(x, y, type = 'l')`



```
f=function(x){1/sqrt(2*pi)*exp(-x^2/2)}  
> curve(f, -3,3)
```



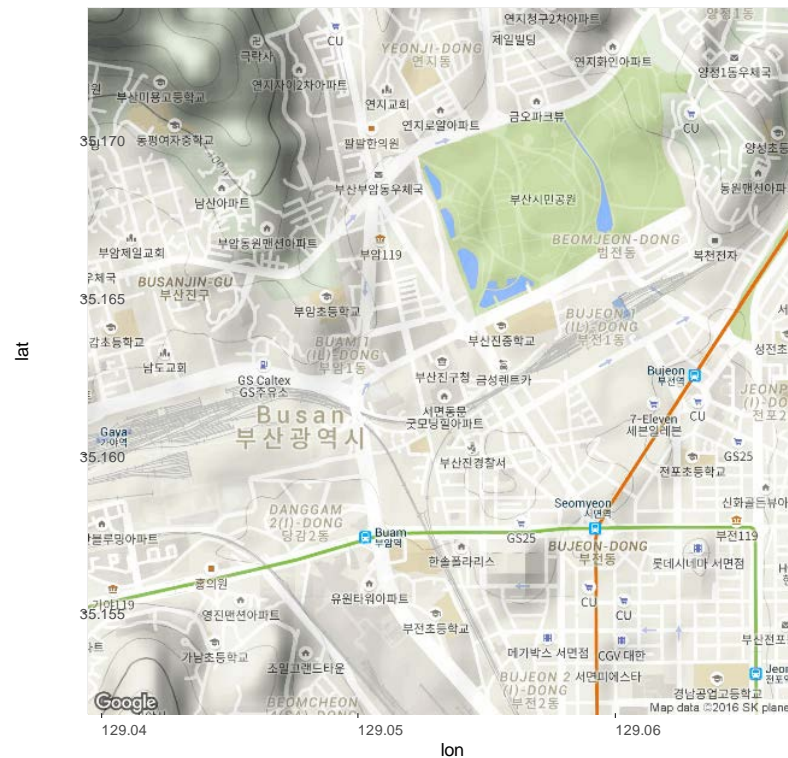
Another example is to find out

$$\int_{-1.96}^{1.96} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

```
>f <- function(x) {1/sqrt(2*pi)*exp(-x^2/2)}  
>integrate(f, lower = -1.96, upper = 1.96)  
0.9500042 with absolute error < 1.0e-11
```

R을 이용한 지도그리기

- 1) googlemap 이용하기
- Library(ggmap)
- Library(ggplot2)
- A<-get_map(location ="busanjin-gu" ,zoom=15)
- ggmap(A)



```
> a <- read.table("d://spatial/earth.txt", header=T)
```

```
> a
```

	area	case	long	lat
1	강원도	65	128.79	37.68
2	경기도	17	126.20	37.39
3	경남	44	128.89	35.24
4	경북	123	129.05	36.33
5	남해	2	127.10	34.25
6	서해	28	125.27	37.25
7	인천	17	124.80	37.62
8	전남	57	125.87	34.91
9	전북	36	126.48	35.81
10	제주도	27	126.40	33.53
11	충남	70	126.41	36.47
12	충북	18	127.73	36.49
13	평남	56	125.78	39.11
14	평북	11	125.75	40.13
15	함남	8	127.11	39.49
16	황해도	58	125.82	38.33

```
> korea<-get_map("south korea", zoom=6, maptype='roadmap'))  
> ggmap(korea)+geom_point(data=a, aes(x=long, y=lat, size=case),  
colour="red")
```

