

## *Chapter 2. Getting Started with R:* 데이터 분석을 위한 준비

SEOUL WOMEN'S UNIVERSITY  
DEPT OF ECONOMICS

## 변수(Variables)

- R은 #로 시작하는 라인은 읽지 않는다. 필요한 코멘트를 남기고 싶을 때 #를 사용하면 된다.
- `a <- 1`  
a라는 변수에 1을 assign하라는 뜻이다. R Console에 `a <- 1`을 입력해보자. 그리고 Enter.

R Console

```
> a <- 1
>
```

- 이어서 a를 type해보자.

R Console

```
> a <- 1
> a
[1] 1
>
```

- 같은 방법으로 몇 개의 변수를 더 만들어본다.

## R Console

```
> b <- 2  
> b  
[1] 2  
>  
> c <- 3  
> c  
[1] 3  
>  
d <- 3.5  
> d  
[1] 3.5  
>
```

D:/MyR/CH00 - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
CH00

Source

Console

Terminal

Jobs

D:/MyR/CH00/
배표와 관련된 상세한 내용은 'license()' 또는 'licence()'을 통하여 확인할 수 있습니다.  
R은 많은 기여자들이 참여하는 공동프로젝트입니다.  
'contributors()'라고 입력하시면 이에 대한 더 많은 정보를 확인할 수 있습니다.  
그리고, R 또는 R 패키지들을 출판물에 인용하는 방법에 대해서는 'citation()'을 통해 확인하시길 부탁드립니다.  
'demo()'를 입력하신다면 몇가지 데모를 보실 수 있으며, 'help()'를 입력하시면 온라인 도움말을 이용하실 수 있습니다.  
또한, 'help.start()'의 입력을 통하여 HTML 브라우저에 의한 도움말을 이용하실 수 있습니다.  
R의 종료를 원하시면 'q()'을 입력해주세요.

```

> a<-1
> a
[1] 1
> b<-2
> b
[1] 2
>
> c<-3
> c
[1] 3
>
> d<-3.5
> d
[1] 3.5
>

```

Environment History Connections

Import Dataset
Global Environment

Values

a	1
b	2
c	3
d	3.5

Files Plots Packages Help Viewer

New Folder Delete Rename More
D: > MyR > CH00

	Name	Size	Modified
<input type="checkbox"/>	..		
<input type="checkbox"/>	.Rhistory	0 B	Mar 4, 2020, 11:42 PM
<input type="checkbox"/>	CH00.Rproj	218 B	Mar 4, 2020, 11:42 PM
<input type="checkbox"/>	Example1.R	11 B	Mar 4, 2020, 11:09 PM
<input type="checkbox"/>	Example2.R	11 B	Mar 4, 2020, 11:19 PM

- 간단한 계산

## R Console

```
> a+b  
[1] 3  
>  
> a+b+c  
[1] 6  
>  
> 4/b  
[1] 2  
>  
> 5*b  
[1] 10  
>
```

- R에서 벡터(vector) 형태의 변수도 만들 수 있다. 이 때 `c( )`를 사용한다.

## R Console

```
> v1 <- c(1,2,5,7,8)
> v1
[1] 1 2 5 7 8
>
```

- 1부터 5까지 1씩 증가하는 벡터(vector)

## R Console

```
> v2 <- c(1:5)
> v2
[1] 1 2 3 4 5
>
> v3 <- seq(1,5)
> v3
[1] 1 2 3 4 5
```

- seq( )의 응용: 1부터 10까지 2씩 증가하는 벡터(vector)

## R Console

```
> v4 <- seq(1, 10, by=2)
> v4
[1] 1 3 5 7 9
>
```

- seq( )의 응용: 1부터 10까지 3씩 증가하는 벡터(vector)

## R Console

```
> v5 <- seq(1, 10, by=3)
> v5
[1] 1 4 7 10
>
```



- 벡터(vector)의 연산: 벡터와 스칼라(scalar)의 덧셈

## R Console

```
> v1  
[1] 1 2 5 7 8  
>  
> v1 + 2  
[1] 3 4 7 9 10  
>
```

- 벡터(vector)의 연산: 벡터와 벡터의 덧셈

## R Console

```
> v1  
[1] 1 2 5 7 8  
>  
> v2  
[1] 1 2 3 4 5  
>  
> v1 + v2  
[1] 2 4 8 11 13  
>
```

- 문자로 된 변수(character variables): 변수에 문자를 넣을 때(assign)는 " "을 사용

## R Console

```
> c1 <- "a"
> c1
[1] "a"
>
> c2 <- "text"
> c2
[1] "text"
>
> c3 <- "Hello Seoul Women's University!"
> c3
[1] "Hello Seoul Women's University!"
>
```

- 문자로 된 벡터(character vectors): numeric 벡터를 만들 때 사용했던 `c( )`를 사용

## R Console

```
> c4 <- c("a", "b", "c")
> c4
[1] "a" "b" "c"
>
> c5 <- c("Hello", "my", "name", "is", "Weh-Sol", "!")
> c5
[1] "Hello" "my" "name" "is" "Weh-Sol" "!"
>
```

- 문자로 된 벡터(character vectors)들은 연산이 불가능하다. 특수한 함수를 사용해야 변형이 가능하다.

# 함수(Functions)

- 함수: 간단한 예를 통해서 함수를 이해할 수 있다. 벡터를 만들 때 사용했던 `c()`가 바로 여러 개의 숫자나 문자를 하나의 벡터로 변환해주는 함수
- `mean()`: 평균을 구하는 함수

## R Console

```
> x <- c(1, 2, 3)
> x
[1] 1 2 3
>
> mean(x)
[1] 2
>
> xbar <- mean(x)
> xbar
[1] 2
>
```

- `max( )`, `min( )`: 최대값과 최소값을 찾는 함수

## R Console

```
> x <- c(1, 2, 3)
> x
[1] 1 2 3
>
> max(x)
[1] 3
>
> min(x)
[1] 1
>
```

- `paste( , collapse="구분기호")`: 여러 문자(character)를 하나로 병합하는 함수. 문자와 문자 사이에 "구분기호"가 들어간다. input은 character vector.

## R Console

```
> c5 <- c("Hello", "my", "name", "is", "Weh-Sol", "!")
> c5
[1] "Hello" "my" "name" "is" "Weh-Sol" "!"
>
> paste(c5)
[1] "Hello" "my" "name" "is" "Weh-Sol" "!"
>
> paste(c5, collapse=",") # 구분기호로 콤마(,)를 사용
[1] "Hello,my,name,is,Weh-Sol,!"
>
> paste(c5, collapse=" ") # 구분기호로 space를 사용
[1] "Hello my name is Weh-Sol !"
>
> who <- paste(c5, collapse=" ")
> who
[1] "Hello my name is Weh-Sol !"
>
```

# Packages



- 어떤 함수들(mean, max, min)은 R에 이미 내장되어 있어서 별도로 설치할 필요가 없다.
- 모든 함수가 R에 내장되어있는 것은 아니다. 필요한 함수를 별도로 설치해야 하는데, 하나 혹은 2개 이상의 함수들을 묶어놓은 패키지(packages)를 다운로드해야 한다.
- Packages(패키지): 하나 혹은 그 이상의 함수들을 묶어놓은 것.
  - 사용하고 싶은 함수가 포함되어 있는 패키지를 찾아서 다운로드한다.
  - 스크립트에서 Code를 시작할 때 패키지를 로드(load)한다.
  - 그러면 패키지 안에 들어있는 모든 함수를 사용할 수 있다.
- 다양한 그래프를 그리기 위해서 `qplot()`이라는 함수를 사용하고 싶다. `qplot()`은 `ggplot2`라는 package에 포함되어 있다.
- 1단계: package `ggplot2`를 설치한다.

R Console

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

아래와 같이 자동으로 설치된다.

The screenshot shows the RStudio interface with the following components:

- Source Panel:** Contains the R script code being executed.
 

```
contributors()'라고 입력하시면 이에 대한 더 많은 정보를 확인하실 수 있습니다.
그리고, R 또는 R 패키지들을 출판물에 인용하는 방법에 대해서는 'citation()'을
통해 확인하시길 부탁드립니다.

'demo()'를 입력하신다면 몇가지 데모를 보실 수 있으며, 'help()'를 입력하시면
온라인 도움말을 이용하실 수 있습니다.
또한, 'help.start()'의 입력을 통하여 HTML 브라우저에 의한 도움말을 이용하실
수 있습니다
R의 종료를 원하시면 'q()'을 입력해주세요.

> install.packages("ggplot2")
WARNING: Rtools is required to build R packages but is not currently
installed. Please download and install the appropriate version of R
tools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/JOY/Documents/R/win-library/3.6'
(as 'lib' is unspecified)
URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/ggplot2_3.2.1.
zip'을 시도합니다
Content type 'application/zip' length 3975973 bytes (3.8 MB)
downloaded 3.8 MB

package 'ggplot2' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:/Users/JOY/AppData/Local/Temp/RtmpCCjpiK/downloaded_packag
es
> |
```
- Console Panel:** Displays the output of the R script execution, including the warning about Rtools and the successful installation of ggplot2.
- Environment Panel:** Shows the Global Environment, which is currently empty.
- Files Panel:** Shows the project files, including .Rhistory, CH00.Rproj, Example1.R, and Example2.R.

[Tools] → [Install Packages...]를 클릭해서 패키지를 설치할 수도 있다.

The screenshot shows the RStudio interface. The 'Tools' menu is open, and 'Install Packages...' is selected. The console displays the following text:

```
R version 3.6.3 (2020-02-29) -- "Holding the Fort"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R은 자유 소프트웨어이며, 어떠한 형태의 보증없이 배포됩니다.
또한, 일정한 조건하에서 이것을 재배포 할 수 있습니다.
배포와 관련된 상세한 내용은 'license()' 또는 'licence()'를
실행하십시오.

R은 많은 기여자들이 참여하는 공동프로젝트입니다.
'contributors()'라고 입력하시면 이에 대한 더 많은 정보를 확인할 수 있습니다.
그리고, R 또는 R 패키지들을 출판물에 인용하는 방법에 대해서는 'citation()'을 통해
확인하시길 부탁드립니다.

'demo()'를 입력하신다면 몇가지 데모를 보실 수 있으며, 'help()'를 입력하시면은
라인 도움말을 이용하실 수 있습니다.
또한, 'help.start()'의 입력을 통하여 HTML 브라우저에 의한 도움말을 이용하실 수
있습니다.
R의 종료를 원하시면 'q()'을 입력해주세요.

> |
```

The file explorer on the right shows the following files:

Name	Size	Modified
..		
.Rhistory	643 B	Mar 5, 2020, 1:53 AM
CH00.Rproj	218 B	Mar 5, 2020, 1:53 AM
Example1.R	11 B	Mar 4, 2020, 11:09 PM
Example2.R	11 B	Mar 4, 2020, 11:19 PM

ggplot2을 입력한 후 [Install]을 클릭한다.

The screenshot shows the RStudio interface with the 'Install Packages' dialog box open. The dialog box has a red border around the 'Packages' input field. The background shows the RStudio interface with the console and environment panes.

**Source**

**Console** **Terminal** **Jobs**

D:/MyR/CH00/

R version 3.6.3 (2020-02-29) -- "Holding the Windsock"  
Copyright (C) 2020 The R Foundation for Statistical Computing  
Platform: x86\_64-w64-mingw32/x64 (64-bit)

R은 자유 소프트웨어이며, 어떠한 형태의 보증없이  
또한, 일정한 조건하에서 이것을 재배포 할 수 있는  
배포와 관련된 상세한 내용은 'license()' 또는  
있습니다.

R은 많은 기여자들이 참여하는 공동프로젝트입니다.  
'contributors()'라고 입력하시면 이에 대한 더  
그리고, R 또는 R 패키지들을 출판물에 인용하는 방  
해 확인하시길 부탁드립니다.

'demo()'를 입력하신다면 몇가지 데모를 보실 수  
라인 도움말을 이용하실 수 있습니다.  
또한, 'help.start()'의 입력을 통하여 HTML 브  
있습니다.  
R의 종료를 원하시면 'q()'을 입력해주세요.

> |

**Environment** **History** **Connections**

Import Dataset

Global Environment

Environment is empty

**Help** **Viewer**

Rename More

	Size	Modified
	643 B	Mar 5, 2020, 1:53 AM
	218 B	Mar 5, 2020, 1:53 AM
	11 B	Mar 4, 2020, 11:09 PM
	11 B	Mar 4, 2020, 11:19 PM

**Install Packages**

Install from: [Configuring Repositories](#)

Repository (CRAN)

Packages (separate multiple with space or comma):

Install to Library: C:/Users/JOY/Documents/R/win-library/3.6 [Default]

☒ Install dependencies

Install Cancel

- 2단계: ggplot2 안에 들어있는 함수들을 사용하기 위해서 패키지를 로드(load)한다.

## R Console

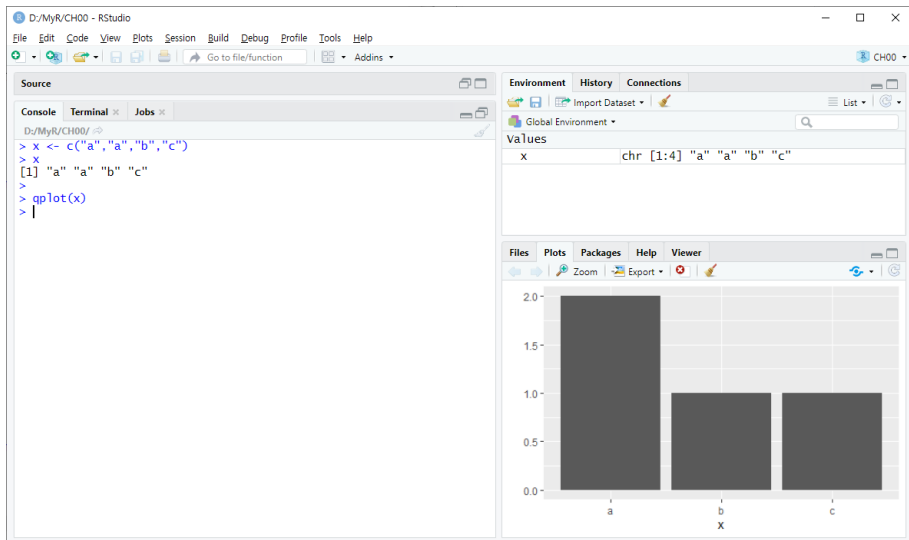
```
> library(ggplot2)
>
```

- qplot( )의 활용 1:

## R Console

```
> x <- c("a","a","b","c")
> x
[1] "a" "a" "b" "c"
>
> qplot(x)
>
```

y변수는 지정하지 않고 x변수만 입력하면 빈도(frequency)를 나타내는 bar chart를 얻는다.



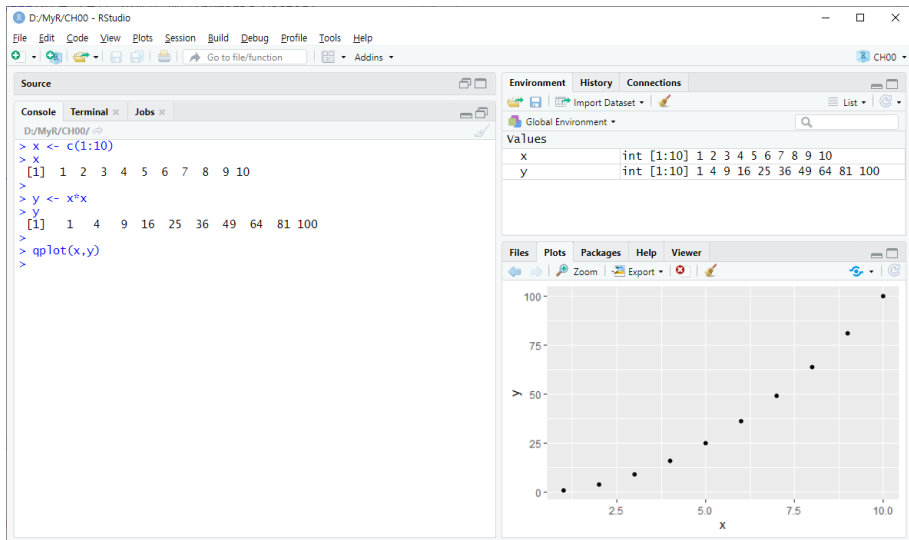
- `qplot()`의 활용 2:  $x$ 는 1부터 10까지 벡터(numeric vector),  $y$ 는  $x$ 의 제곱과 같다고 해보자.

$$y = x^2$$

## R Console

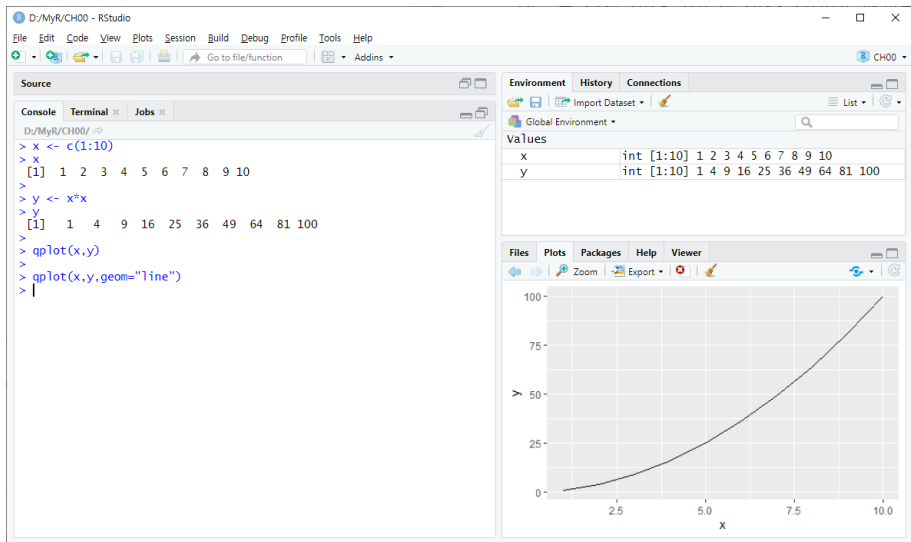
```
> x <- c(1:10)
> x
[1] 1 2 3 4 5 6 7 8 9 10
>
> y <- x*x
> y
[1] 1 4 9 16 25 36 49 64 81 100
>
> qplot(x,y)      # qplot(x,y,geom="point")라고 해도 결과는 같다.
>
```

$x$ 와  $y$ 를 모두 입력하면 두 변수 사이의 관계를 점(point)으로 나타낸 그래프를 얻는다.





- `qplot(x,y,geom="line")`이라고 입력하면 두 변수 사이의 관계를 직선(line)으로 나타낸 그래프를 얻는다.



- R에 설치된 함수 매뉴얼(manual)은 다음과 같이 확인할 수 있다.

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains the command `> ?qplot`, which is highlighted with a red box.
- Environment Panel:** Displays the current environment with variables `x` and `y`, both of type `int [1:10]`. The values for `x` are 1 through 10, and for `y` are 1 through 100.
- Help Panel:** Displays the documentation for the `qplot` function from the `ggplot2` package. The title is "Quick plot". The description states: "qplot is a shortcut designed to be familiar if you're used to base `plot()`. It's a convenient wrapper for creating a number of different types of plots using a consistent calling scheme. It's great for allowing you to produce plots quickly, but I highly recommend learning `ggplot()` as it makes it easier to create complex graphics." The usage section shows the function signature: `qplot(x, y, data, facets = NULL, margins = FALSE, ...)`.