

## 제 1 장. R Studio의 소개

### 1.1 R의 소개

- R은 데이터 분석을 위한 통계 및 그래픽스를 지원하는 자유 소프트웨어 환경임
- R은 하나의 컴퓨터 언어이자 다양한 패키지(또는 라이브러리)의 집합으로써 자유롭게 데이터 분석을 수행할 수 있는 장점을 갖고 있음
- R은 통계, 금융, 바이오인포머틱스, 그래픽스에 이르는 다양한 통계 패키지를 갖고 있으며 이 모든 것이 무료로 제공됨

### 1.2 R Studio의 소개

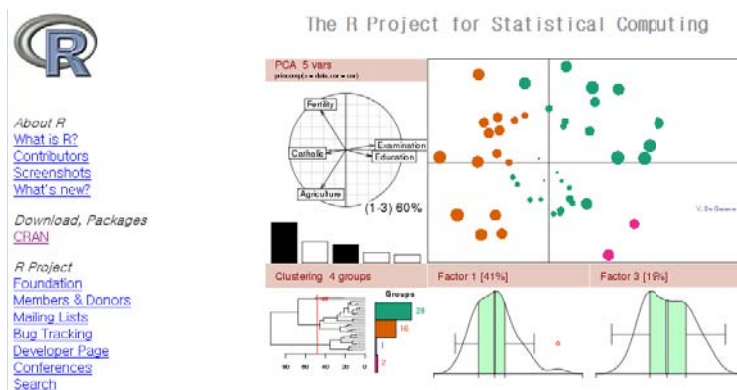
- R을 위한 통합 환경 (Integrated Development Environment)으로 R 사용에 필요한 부가 정보들을 한눈에 볼 수 있게 해줌

### 1.3 R Studio의 설치


: R Studio를 설치하기 위해서는 먼저 R을 설치해야 한다

#### 1.3.1 R의 설치

- 1) R 웹사이트 : <http://www.r-project.org>



2) [CRAN] 링크를 선택하고 대한민국의 R Mirror 사이트를 선택한 후, 해당되는 운영체계에 적절한 링크를 선택한다.



About R  
What is R?  
Contributors  
Screenshots  
What's new?

Download, Packages  
CRAN

R Project  
Foundation  
Members & Donors  
Mailing Lists  
Bug Tracking  
Developer Page  
Conferences  
Search

Indonesia	<a href="http://ftp.itim.ac.id/cran/">http://ftp.itim.ac.id/cran/</a>	Indian Institute of Technol
	<a href="http://cran.repo.bppt.go.id/">http://cran.repo.bppt.go.id/</a>	Agency for The Applicatio Technology
	<a href="http://cran.unsj.ac.id/">http://cran.unsj.ac.id/</a>	The University of Jember
Iran	<a href="http://cran.um.ac.ir/">http://cran.um.ac.ir/</a>	Ferdowsi University of Ma
Ireland	<a href="http://ftp.heanet.ie/mirrors/cran/r-project.org/">http://ftp.heanet.ie/mirrors/cran/r-project.org/</a>	HEAnet, Dublin
Italy	<a href="http://cran.mirror.sar.it/mirrors/CRAN/">http://cran.mirror.sar.it/mirrors/CRAN/</a>	Garr Mirror, Milano
	<a href="http://cran.stat.unipd.it/">http://cran.stat.unipd.it/</a>	University of Padua
	<a href="http://dsrm.unipa.it/CRAN/">http://dsrm.unipa.it/CRAN/</a>	Università degli Studi di P
Japan	<a href="http://easrc.hyogo-u.ac.jp/cran/">http://easrc.hyogo-u.ac.jp/cran/</a>	Hyogo University of Tead
	<a href="http://cran.ism.ac.jp/">http://cran.ism.ac.jp/</a>	Institute of Statistical Mat
	<a href="http://cran.md.tsukuba.ac.jp/">http://cran.md.tsukuba.ac.jp/</a>	University of Tsukuba
Korea	<a href="http://cran.nexr.com/">http://cran.nexr.com/</a>	NexR Corporation, Seoul
	<a href="http://biostat.cau.ac.kr/CRAN/">http://biostat.cau.ac.kr/CRAN/</a>	Chung-Ang University, Se
Lebanon	<a href="http://mirror.lau.edu.lb/">http://mirror.lau.edu.lb/</a>	Lebanese American Unive

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for MacOS X](#)
- [Download R for Windows](#)

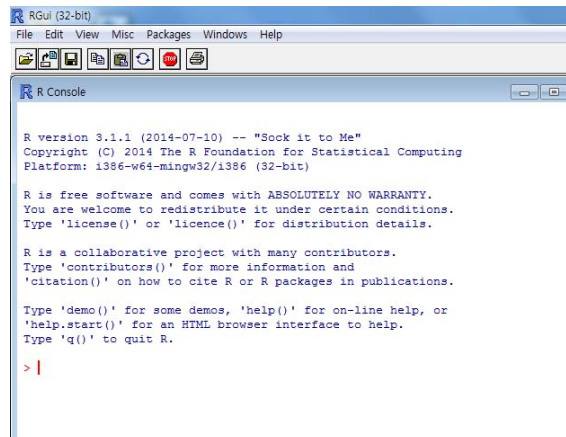
R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

3) [base] 링크를 선택하여 기본적인 프로그램 설치 파일의 링크를 선택한다.



[contrib] 링크에는 부가적인 패키지들이 있고, 이것들은 필요할 때 R 프로그램 내에서 설치할 수 있으므로 초기에 반드시 설치할 필요는 없다.

4) 다운로드 링크를 선택 한 후, 바로 실행 버튼을 눌러 프로그램을 설치한다. 설치 후 환경 설정을 수정할 수 있으므로 제시하는 디폴트 옵션으로 설치할 것을 권한다. 설치 후 바탕화면에 나타난 아이콘을 선택하면 다음과 같은 첫 화면이 나타난다.



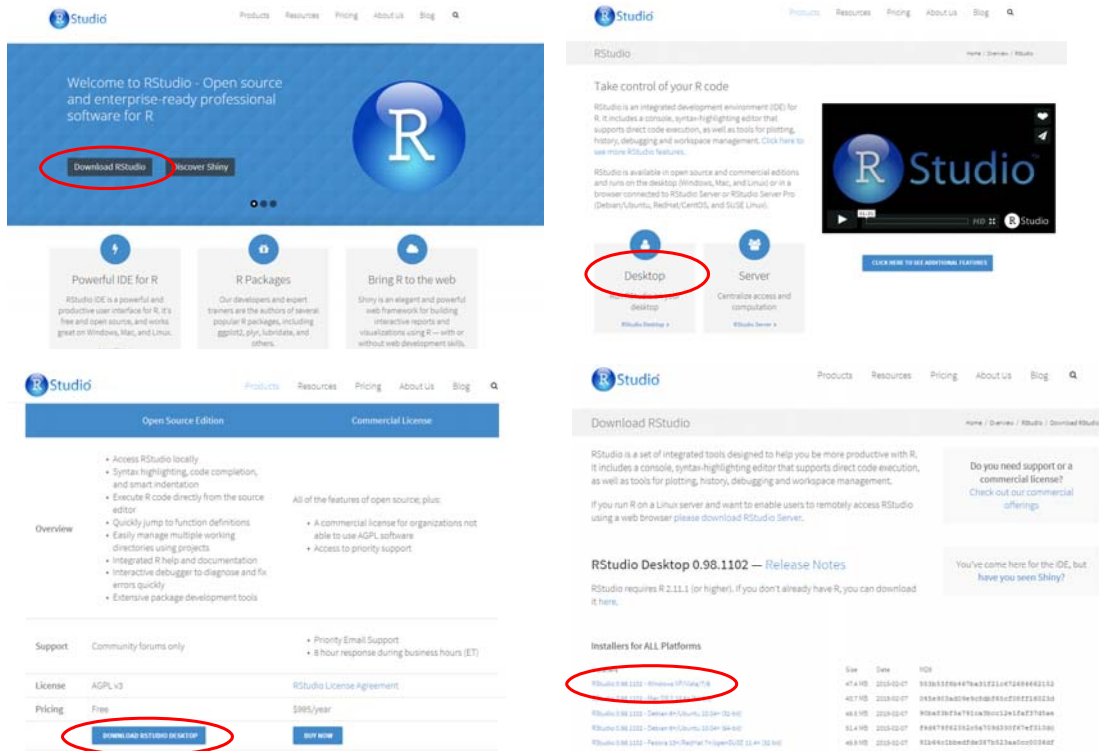
R console 창에서는 라인 실행이 가능하며 프로그램(R에서는 Script 라고 함)의 입력과 실행, 결과 출력, 오류 메시지 등이 출력된다. 그래프는 필요 시 그래프 창이 새로 열리며 그려지고 프로그램은 별도의 스크립트 창에서도 작성이 가능하다.

### 1.3.2 R Studio의 설치

1) R studio 웹사이트 : <http://www.rstudio.com/>

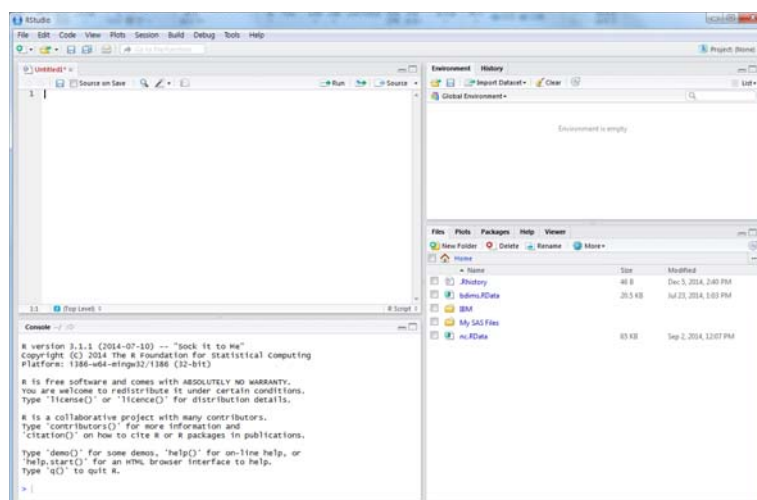


2) 첫 화면에서 [Download RStudio]를 선택한 후 [Desktop]을 선택하여 설치 페이지로 이동하고 해당되는 운영체제를 선택하여 설치한다.



## 1.4 R Studio의 실행

R Studio는 기본적으로 다음과 같이 네 개의 창으로 구성되어 있다.



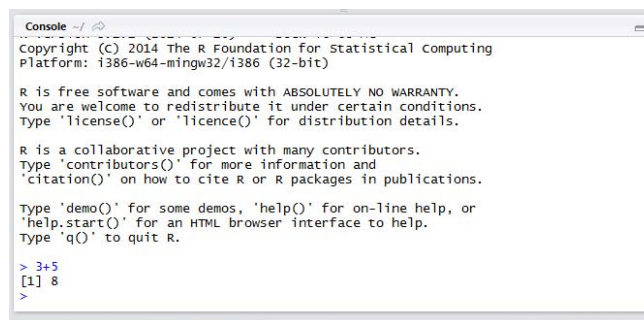
1) R script and data view : R 스크립트(명령어)를 작성하고 데이터를 볼 수 있다

- 2) Console : R 스크립트를 실행하고 실행 결과를 볼 수 있다
- 3) Workspace and History : R session이 실행되는 동안의 모든 작업 과정들과 생성물들이 기록되어 있다
- 4) Files, plots, packages and help : 디렉토리 내의 파일들을 확인하거나 그래프를 출력한다. 또한 추가적으로 패키지를 설치할 수 있고 도움말을 통해 R studio에 관한 추가적인 정보를 얻을 수 있다.

## 1.5 R의 기본 사용법

R console 창에 직접 명령어(스크립트)를 입력함으로써 간단한 계산이나 통계 분석, 그래프 등의 분석을 할 수 있다.

예를 들어, '3+5'의 계산을 하기 위한 경우, 다음과 같이 R console 창에 '3+5'를 입력하고 엔터를 누르면 바로 출력 결과를 확인할 수 있다.



```
Console ~/
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-bit)

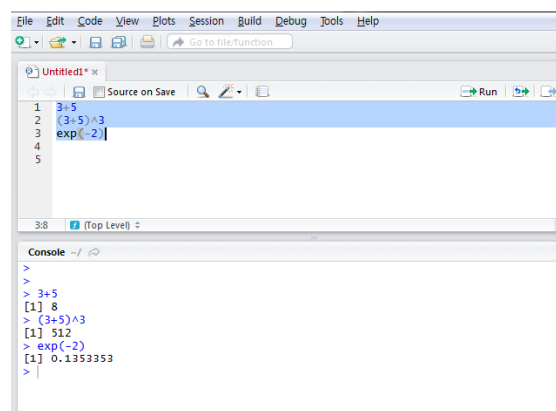
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

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.

> 3+5
[1] 8
>
```

다만 R console 창에 직접 명령어를 입력하게 되면 명령어의 재사용이나 오타의 수정 등이 어려워질 수 있으므로 R script 창에 명령어를 작성 한 후 이를 console 창에서 실행하는 방법을 사용한다. R script 창은 명령어 편집기로써 일종의 메모장이라고 생각하면 될 것이다. R script 창에 명령어를 입력 한 후 문장의 끝에 커서를 위치하게 한 후, 'Ctrl + Enter'를 실행하면 해당되는 라인의 명령어가 console 창에서 실행된다. 만약 여러 줄의 명령어를 한 번에 모두 실행하고 싶다면 실행을 원하는 스크립트를 동시에 선택한 후 'Ctrl + Enter'를 실행한다.



```
File Edit Code View Plots Session Build Debug Tools Help
[Icons] Go to file/function

Untitled1 *
1 3+5
2 (3+5)^3
3 exp(-2)
4
5

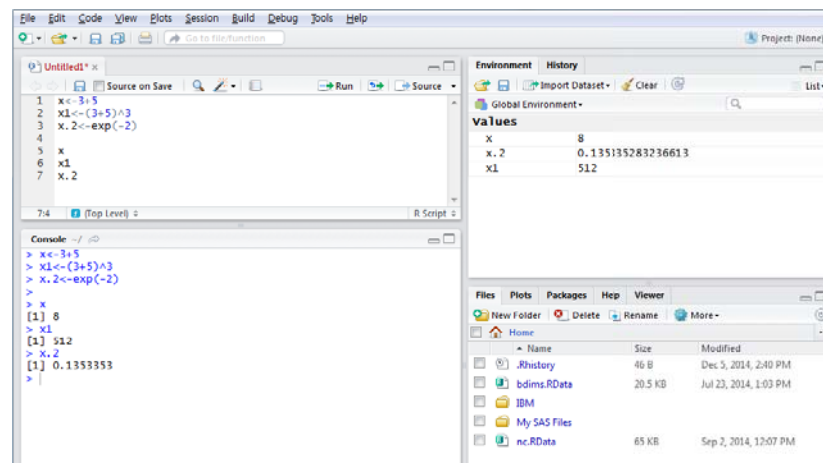
3:8 (Top Level)

Console ~/
>
> 3+5
[1] 8
> (3+5)^3
[1] 512
> exp(-2)
[1] 0.1353353
>
```

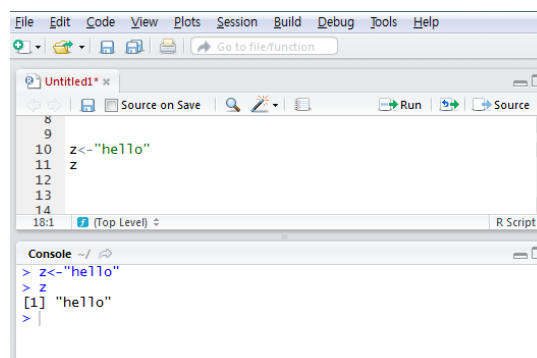
이렇게 실행된 계산 결과는 결과물의 출력만 이루어 질 뿐 변수로써 저장되지 않는다. 따라서 자료를 변수에 입력하여 저장하는 작업이 필요하다.

R의 변수명은 알파벳, 숫자, ‘\_’, ‘.’ 으로 구성되며 첫 글자는 문자 또는 ‘.’으로 시작해야 한다. 또한 이름 중간에는 공백을 가질 수 없으며 대소문자를 구별한다.

변수에 값을 할당할 때는 ‘<-’ 또는 ‘=’, ‘->’를 사용한다. R에서 자료가 저장되는 변수를 오브젝트(object)라고 하며 오브젝트의 출력은 오브젝트 명을 입력하여 실행하면 된다. 다음은 앞서 실행된 계산 결과들을 각각의 오브젝트로 저장한 결과이다.



오브젝트가 생성되면 Workspace에서 생성된 오브젝트 리스트를 확인할 수 있다. 문자열을 변수로 저장하기 위해서는 문자열을 따옴표로 묶으면 된다.

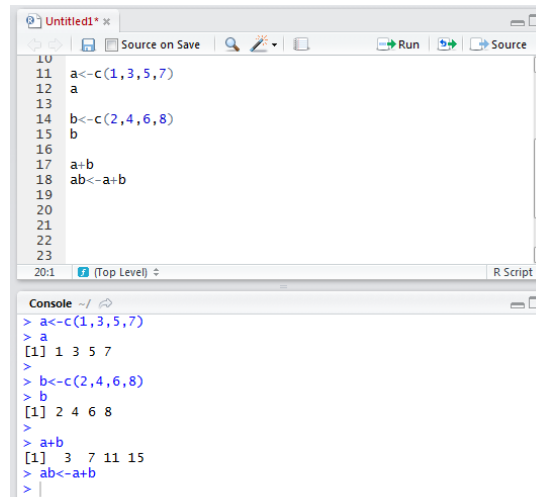


오브젝트에 저장될 수 있는 내용은 스칼라, 벡터, 행렬, 함수 등이 있으며 자료와 관련된 오브젝트의 특징은 다음과 같다.

- 스칼라 (scalars) : 하나의 숫자 또는 문자로 이루어 짐.
- 벡터 (vectors) : 스칼라의 집합. 여러 개의 숫자 또는 문자로 이루어 짐.
- 행렬 (matrices) : 벡터의 집합. 행과 열로 이루어져 있음.

### 1.5.1 벡터 (vectors)

R에서 벡터를 생성하기 위해서는 `c()` 함수를 사용한다. 예를 들어 (1,3,5,7)로 구성된 벡터를 생성하기 위해서는 다음과 같이 실행한다. 숫자로 이루어진 벡터의 연산은 같은 위치에 있는 스칼라들끼리 이루어진다. 생성된 두 벡터 오브젝트의 연산 결과 역시 또 다른 오브젝트로 저장 가능하다.



```
10
11 a<-c(1,3,5,7)
12 a
13
14 b<-c(2,4,6,8)
15 b
16
17 a+b
18 ab<-a+b
19
20
21
22
23
```

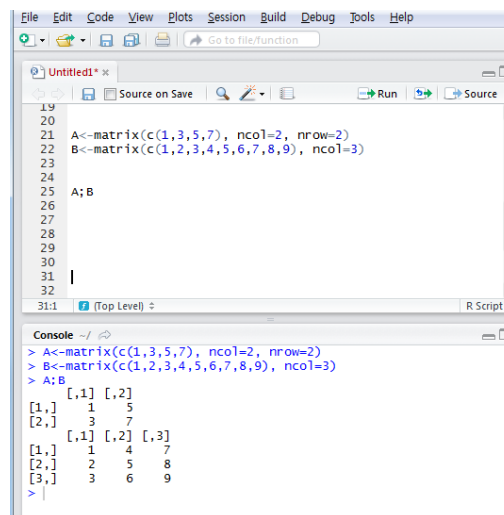
Console

```
> a<-c(1,3,5,7)
> a
[1] 1 3 5 7
> b<-c(2,4,6,8)
> b
[1] 2 4 6 8
> a+b
[1] 3 7 11 15
> ab<-a+b
> |
```

벡터는 동일한 타입의 스칼라로 구성되기 때문에 하나의 벡터 안에 문자와 숫자가 동시에 입력 되는 경우에는 숫자형 스칼라가 문자형으로 바뀌게 된다.

### 1.5.2 행렬(Matrices)

행렬은 열(columns)과 행(rows)으로 이루어진 2차원 벡터이다. 행렬의 생성은 `matrix()` 함수를 이용한다. 행렬을 구성하는 원소들을 벡터의 형태로 입력 한 후, 행과 열의 수를 지정한다.



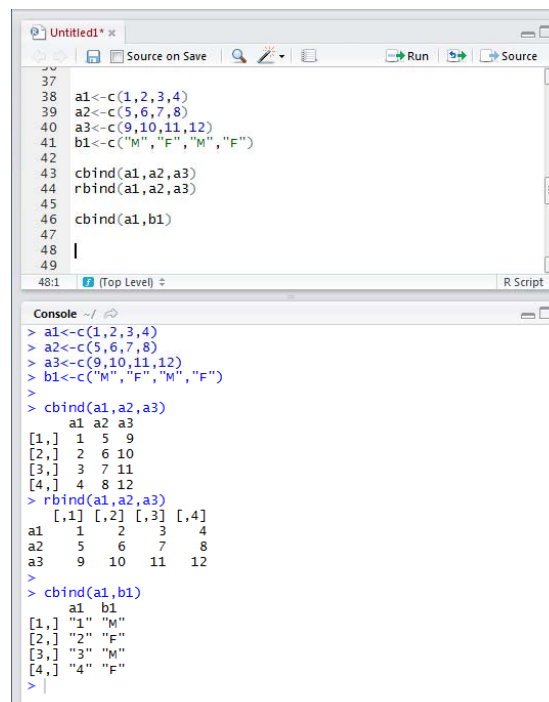
```
19
20
21 A<-matrix(c(1,3,5,7), ncol=2, nrow=2)
22 B<-matrix(c(1,2,3,4,5,6,7,8,9), ncol=3)
23
24 A;B
25
26
27
28
29
30
31
32
```

Console

```
> A<-matrix(c(1,3,5,7), ncol=2, nrow=2)
> B<-matrix(c(1,2,3,4,5,6,7,8,9), ncol=3)
> A;B
      [,1] [,2]
[1,] 1     5
[2,] 3     7
      [,1] [,2] [,3]
[1,] 1     4     7
[2,] 2     5     8
[3,] 3     6     9
> |
```

R 스크립트는 하나의 명령어가 한 줄에 입력되는 것이 기본이지만 한 줄(line)에서 명령어를 나누기 위해서는 세미콜론(:)을 사용한다.

길이가 동일한 벡터를 이용해서 행렬을 생성하기 위해서는 cbind() 또는 rbind()의 함수를 이용한다. 벡터와 마찬가지로 행렬에는 한 가지 유형의 스칼라만 저장할 수 있다.



```
37
38 a1<-c(1,2,3,4)
39 a2<-c(5,6,7,8)
40 a3<-c(9,10,11,12)
41 b1<-c("M","F","M","F")
42
43 cbind(a1,a2,a3)
44 rbind(a1,a2,a3)
45
46 cbind(a1,b1)
47
48 |
49
```

```
> a1<-c(1,2,3,4)
> a2<-c(5,6,7,8)
> a3<-c(9,10,11,12)
> b1<-c("M","F","M","F")
>
> cbind(a1,a2,a3)
  a1 a2 a3
[1,] 1  5  9
[2,] 2  6 10
[3,] 3  7 11
[4,] 4  8 12
> rbind(a1,a2,a3)
  a1 [,2] [,3] [,4]
a1  1    2    3    4
a2  5    6    7    8
a3  9   10   11   12
>
> cbind(a1,b1)
  a1 b1
[1,] "1" "M"
[2,] "2" "F"
[3,] "3" "M"
[4,] "4" "F"
> |
```

벡터와 행렬에 관해서는 제 3장에서 더 자세히 다루기로 한다.

## 1.6 자료의 입력과 저장

R에서 자료를 입력하는 방법은 R console창에서 직접 입력하거나 외부에서 불러오는 방법이 있다.

### 1.6.1 데이터 프레임(Data frame) 정의

데이터 프레임은 R에서 가장 중요한 자료형이다. 데이터 프레임은 행렬과 마찬가지로의 모습을 하고 있지만 행렬과 달리 다양한 변수, 관측치, 범주 등을 표현하기 위해 특화되어 있다. 즉, 행렬은 한 가지 유형의 스칼라만 저장이 가능하지만 데이터 프레임의 경우에는 문자와 숫자 등 여러 가지 자료 유형을 혼용해서 사용할 수 있다.

예를 들어 다음과 같은 5명의 학생들에 대한 시험 성적 자료가 있다고 하자.

번호	성별	수학	영어
1	M	66	70
2	M	64	68
3	F	48	46
4	F	46	48
5	M	78	84

이 자료는 5개의 행과 4개의 열로 이루어진 행렬의 형태로 생각할 수 있다. 그러나 번호와 수학점수, 영어점수는 숫자형 벡터이고 성별은 문자형 벡터이다. 따라서 이 자료는 행렬이 아닌 데이터 프레임으로 정의해야 한다. 아래의 결과를 확인해 보자.

```

48
49
50 id<-c(1,2,3,4,5)
51 gender<-c("M","M","F","F","M")
52 math<-c(66,64,48,46,78)
53 eng<-c(70,68,46,48,84)
54
55
56 data.frame(id,gender,math,eng)
57
58 cbind(id,gender,math,eng)
59
60

```

```

> id<-c(1,2,3,4,5)
> gender<-c("M","M","F","F","M")
> math<-c(66,64,48,46,78)
> eng<-c(70,68,46,48,84)
>
>
> data.frame(id,gender,math,eng)
  id gender math eng
1  1     M   66  70
2  2     M   64  68
3  3     F   48  46
4  4     F   46  48
5  5     M   78  84
>
> cbind(id,gender,math,eng)
  id gender math eng
[1,] "1"  "M"   "66" "70"
[2,] "2"  "M"   "64" "68"
[3,] "3"  "F"   "48" "46"
[4,] "4"  "F"   "46" "48"
[5,] "5"  "M"   "78" "84"
>

```

위에서 생성된 데이터 프레임을 오브젝트로 저장하게 되면 work space의 리스트에 새로운 오브젝트가 추가된 것을 확인할 수 있다. 또한 work space에서 오브젝트 명을 클릭하면 스크립트창에서 자료를 볼 수 있다.

```

48
49
50 id<-c(1,2,3,4,5)
51 gender<-c("M","M","F","F","M")
52 math<-c(66,64,48,46,78)
53 eng<-c(70,68,46,48,84)
54
55
56 score<-data.frame(id,gender,math,eng)
57
58
59
60

```

```

Environment History
Global Environment
Data
score 5 obs. of 4 variables
values
eng      num [1:5] 70 68 46 48 84
gender   chr [1:5] "M" "M" "F" "F" "M"
id        num [1:5] 1 2 3 4 5
math     num [1:5] 66 64 48 46 78

```



Environment History

Global Environment

Data

score 5 obs. of 4 variables

values

variable	type	values
eng	num [1:5]	70 68 46 48 84
gender	chr [1:5]	"M" "M" "F" "F" "M"
id	num [1:5]	1 2 3 4 5
math	num [1:5]	66 64 48 46 78

	id	gender	math	eng
1	1	M	66	70
2	2	M	64	68
3	3	F	48	46
4	4	F	46	48
5	5	M	78	84

## 1.6.2 외부 자료 불러오기

R에서는 다양한 형태의 자료들을 불러와서 사용할 수 있지만 가장 많이 사용되는 형태는 텍스트 형태의 파일이다. 텍스트 형태의 자료를 불러오기 위해서는 `read.table()`의 함수를 사용하고 csv 형태의 자료를 불러오기 위해서는 `read.csv()`를 사용할 수 있다. 읽어 들인 파일은 데이터 프레임으로 반환되어 저장된다.

이 부분에 관해서는 2장 이후의 자료 분석 예제를 통해 알아보기로 한다.