

---

# 역사 데이터 시각화 분석

---

# What is R?

Big data Analysis Tool : R

# Where to get R?

---

<http://www.r-project.org/>

## 1) CRAN



*About R*  
[What is R?](#)  
[Contributors](#)  
[Screenshots](#)  
[What's new?](#)

*Download, Packages*  
[CRAN](#)

## 2) 국가선택

Korea

<http://cran.nexr.com/>  
<http://healthstat.snu.ac.kr/CRAN/>  
<http://cran.biodisk.org/>

## 3) 운영체제 선택

### Download and Install R

Precompiled binary distributions of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

## 4) 모드 선택

Subdirectories:

[base](#)  
[contrib](#)  
[Rtools](#)

Please do not submit to Windows binaries.

## 5) 다운로드

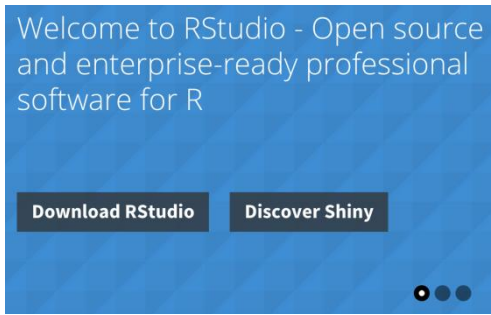
[Download R 3.1.2 for Windows](#) (54 megabytes, 32/64 bit)  
[Installation and other instructions](#)  
[New features in this version](#)

# Where to get R?

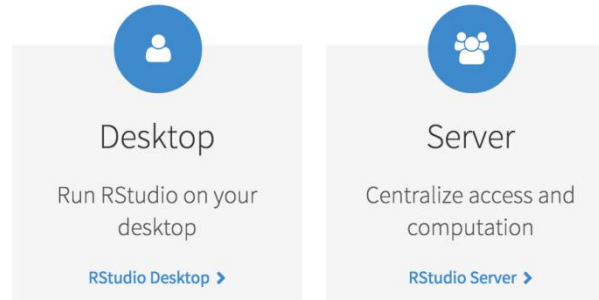


<http://www.rstudio.com/>

## 1) Download Main



## 2) 설치 위치 선택



## 3) 상품 선택

**Support** Community forums only

**License** AGPL v3

**Pricing** Free

[DOWNLOAD RSTUDIO DESKTOP](#)

## 4) 다운로드

### Installers for ALL Platforms

Installers	Size	Date	MD5
<a href="#">RStudio 0.98.1102 - Windows XP/Vista/7/8</a>	47.4 MB	2015-02-07	553b53f8b467ba31f21c672686662152
<a href="#">RStudio 0.98.1102 - Mac OS X 10.6+ (64-bit)</a>	43.7 MB	2015-02-07	045e903ad09e9c8dbf65cf08ff16023d
<a href="#">RStudio 0.98.1102 - Debian 6+/Ubuntu 10.04+ (32-bit)</a>	49.5 MB	2015-02-07	90ba83bf5a791ca3bcc12e1faf37d5ae
<a href="#">RStudio 0.98.1102 - Debian 6+/Ubuntu 10.04+ (64-bit)</a>	51.4 MB	2015-02-07	f4d479f62352c5a709d330f67ef310dc
<a href="#">RStudio 0.98.1102 - Fedora 13+/RedHat 7+/openSUSE 11.4+ (32-bit)</a>	49.9 MB	2015-02-07	91b64c1bbedfde387b523aa0cc0036df
<a href="#">RStudio 0.98.1102 - Fedora 13+/RedHat 7+/openSUSE 11.4+ (64-bit)</a>	51.5 MB	2015-02-07	dac3eb2127d82fa0ef35e8c4773c1f6a

# What is R?

---

R is a free software programming language and software environment for statistical computing and graphics

## ● 개발(Development)

- 뉴질랜드 오클랜드 대학 로스 이하카, 로버트 젠틀만이 최초 개발
- R-Core Team 1997

## ● 환경(environment)

- 대화식 프로그램 수행
- 대용량 데이터 관리 및 처리
- 행렬연산
- 그래픽환경

## ● 확장성 및 범용성

- Linux, Mac, Windows 운영체제에서 사용 가능
- Java, C, Fortran 프로그래밍 언어에 인터페이스 제공
- DBMS 데이터 접근 용이
- Embedded R in Excel

## ● Free software and Open source

- GPL(General Public License) 개념으로  
CRAN(Comprehensive R Archive Network)에서 배포

---

# **NA(Missing Value) Handling**

# NA Handling

- NA(Not available)

- 값이 누락되거나 값이 없는 값을 나타내는 문자

- 예제1

- 변수 생성

```
> X<-c(1,2,3,4,5,6,7,8,NA)
> X
[1] 1 2 3 4 5 6 7 8 NA
```

- NA값으로 변환

```
> X[X==2]<-NA
> X
[1] 1 NA 3 4 5 6 7 8 NA
```

- 변수 요약

```
> summary(X)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
  1.00   2.75   4.50   4.50   6.25   8.00     1
```

- 변수 연산하기

```
> sum(X)
[1] NA
> mean(X)
[1] NA
> sum(X,na.rm=T)
[1] 34
> mean(X,na.rm=T)
[1] 4.857143
```

# NA Handling

---

- 예제2

- 남녀간의 영어,수학 점수를 나타내는 데이터셋 생성

```
> Eng<-c(34,45,56,67,78,89,NA)
> Math<-c(98,NA,87,76,65,54,43)
> Gender<-c("M", "F", "M", "F", "M", "M", "M")
> Test<-data.frame(Eng=Eng,Math=Math,Gender,Gender)
> Test
```

	Eng	Math	Gender	Gender.1
1	34	98	M	M
2	45	NA	F	F
3	56	87	M	M
4	67	76	F	F
5	78	65	M	M
6	89	54	M	M
7	NA	43	M	M

- 데이터 확인

```
> str(Test)
'data.frame': 7 obs. of 4 variables:
 $ Eng      : num  34 45 56 67 78 89 NA
 $ Math     : num  98 NA 87 76 65 54 43
 $ Gender   : Factor w/ 2 levels "F","M": 2 1 2 1 2 2 2
 $ Gender.1: Factor w/ 2 levels "F","M": 2 1 2 1 2 2 2
```



# NA Handling

---

- NA를 포함한 행을 제거한 데이터 세트 생성

```
> na.omit(Test)
  Eng Math Gender Gender.1
1  34   98      M         M
3  56   87      M         M
4  67   76      F         F
5  78   65      M         M
6  89   54      M         M
```

- 영어와 수학 점수만으로 구성된 데이터 세트 생성

```
> Test2<-Test[,c("Eng", "Math")]
```

```
> Test2
```

```
  Eng Math
1  34   98
2  45   NA
3  56   87
4  67   76
5  78   65
6  89   54
7  NA   43
```

```
> apply(Test2,2,mean)
```

```
Eng Math
NA  NA
```

```
> apply(Test2,2,mean,na.rm=TRUE)
```

```
Eng Math
61.5 70.5
```

# NA Handling

---

- 데이터 세트 연산하기

```
> apply(Test2,2,mean)
Eng Math
NA    NA
> apply(Test2,2,mean,na.rm=TRUE)
Eng Math
61.5 70.5
|
```

# NA Handling

- 실습 문제
- 다음은 어느 한 반의 기말고사 성적이다.

	A	B	C	D	E	F	G	H
1	Num	Name	Eng	Math	Korean	Society	Science	Gender
2	1	Amy	97	74	93	96	76	F
3	2	Alexis	75	90	80	84	88	M
4	3	Lexi	86	76	88	90	71	F
5	4	Katie	88	77	85	89	74	F
6	5	Ivy	87	84	90	88	83	F
7	6	Teddy	83	97	75	86	89	M
8	7	Rocky	82	93	79	83	91	M
9	8	Becca	89	73	95	83	80	F
10	9	Taylor	80	94	73	81	94	M
11	10	Sam	79	95	72	78	93	M
12	11	Lauren	90	77	88	86	87	F
13	12	Iris	93	79	86	87	77	F
14	13	Simon	84	89	78	77	90	M
15	14	Tom	83	94	77	74	89	M
16	15	Becca	96	69	89	90	78	F
17	16	Cece	94	74	95	93	74	F
18	17	Kabin	77	95	74	85	95	M
19	18	Robert	86	90	78	82	94	M
20	19	Rebecca	96	79	90	86	67	F
21	20	Garin	85	95	83	80	90	M

## NA Handling

---

- 실습 문제
  - (1) 각 과목의 총합을 구하고, 분석에 사용한 R 코드를 적으시오.
  - (2) 각 과목의 평균을 구하고, 분석에 사용한 R 코드를 적으시오.
  - (3) 각 과목별로 성적이 가장 높은 학생은 누구인지 찾고, 분석에 사용한 R 코드를 적으시오. (결과 값에 각 과목의 성적과 학생의 이름이 모두 출력되게 하시오.)