



## [실습1] R 기초

- ✓ R 개요
- ✓ R 설치
- ✓ RStudio
- ✓ R 라이브러리
- ✓ R 기초





# R ?



- 데이터 분석을 위한 통계 및 그래픽스를 지원하는 프로그래밍 언어  
: <https://www.r-project.org/>
- CRAN(the Comprehensive R Archive Network)에서 제공  
: <http://cran.r-project.org>
- 사용자 제작 패키지(R 라이브러리)를 통하여 확장 가능
  - R 제공 패키지 목록 : <http://cran.nexr.com/web/packages/>
  - 다양한 분야(통계, 머신러닝, 금융, 바이오인포메틱스, 그래픽스 등 )의 패키지를 제공 : <https://cran.r-project.org/web/packages/index.html>

# R 예제



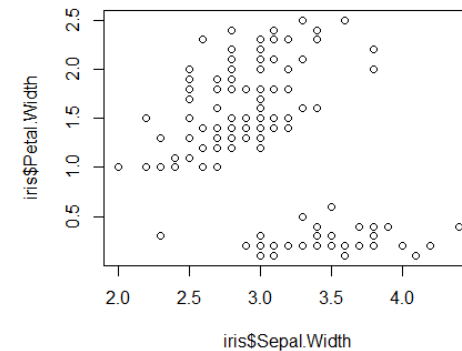
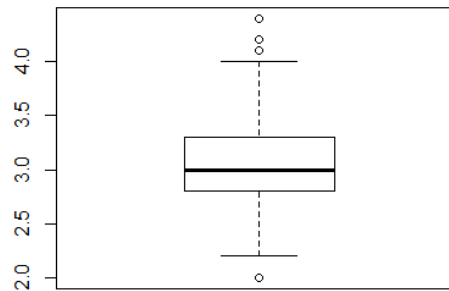
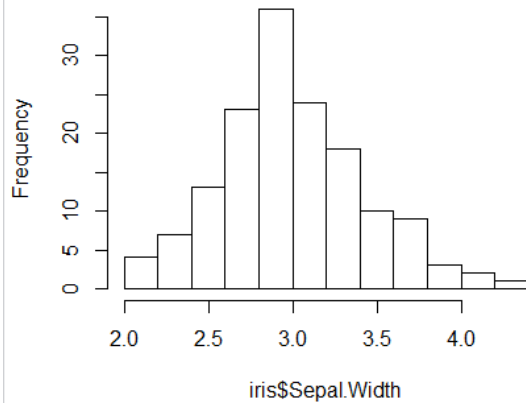
```
> data(iris)      #iris 내장데이터 가져오기
> str(iris)       #iris 구조 확인
'data.frame':   150 obs. of  5 variables:
 $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.
4 4.9 ...
 $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4
2.9 3.1 ...
 $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.
5 1.4 1.5 ...
 $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.
2 0.2 0.1 ...
 $ Species      : Factor w/ 3 levels "setosa","versic
olor",...: 1 1 1 1 1 1 1 1 1 ...
> head(iris)      #상위 6개 확인
  Sepal.Length Sepal.Width Petal.Length
1           5.1           3.5           1.4
2           4.9           3.0           1.4
3           4.7           3.2           1.3
4           4.6           3.1           1.5
5           5.0           3.6           1.4
6           5.4           3.9           1.7
  Petal.Width Species
1          0.2  setosa
2          0.2  setosa
3          0.2  setosa
4          0.2  setosa
5          0.2  setosa
6          0.4  setosa
> summary(iris)   #iris 데이터의 통계적요약값
  Sepal.Length      Sepal.Width      Petal.Length
Min.   :4.300    Min.   :2.000    Min.   :1.000
1st Qu.:5.100    1st Qu.:2.800    1st Qu.:1.600
Median :5.800    Median :3.000    Median :4.350
Mean   :5.843    Mean   :3.057    Mean   :3.758
3rd Qu.:6.400    3rd Qu.:3.300    3rd Qu.:5.100
Max.   :7.900    Max.   :4.400    Max.   :6.900
  Petal.Width      Species
Min.   :0.100    setosa   :50
1st Qu.:0.300    versicolor:50
Median :1.300    virginica :50
Mean   :1.199
3rd Qu.:1.800
Max.   :2.500
```

# R 예제



```
hist(iris$Sepal.Width) #꽃받침의 폭에대한 히스토그램  
boxplot(iris$Sepal.Width)#꽃받침의 폭에대한 박스플롯  
plot(iris$Sepal.Width, iris$Petal.Width) #꽃받침폭과 꽃잎폭에 대한 plot
```

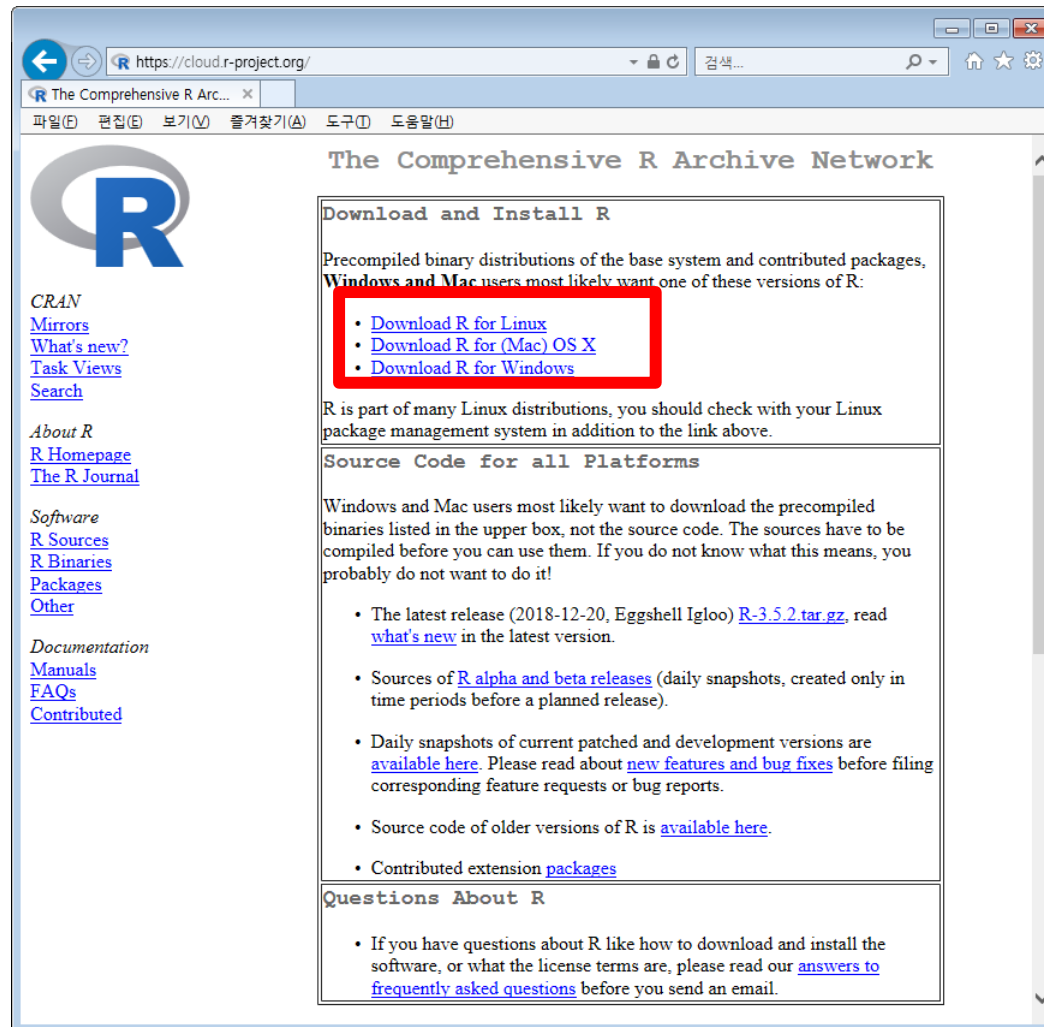
**Histogram of iris\$Sepal.Width**





# R 설치

(1) <https://cloud.r-project.org/>



# R Documentation



**The R Manuals**  
*edited by the R Development Core Team.*

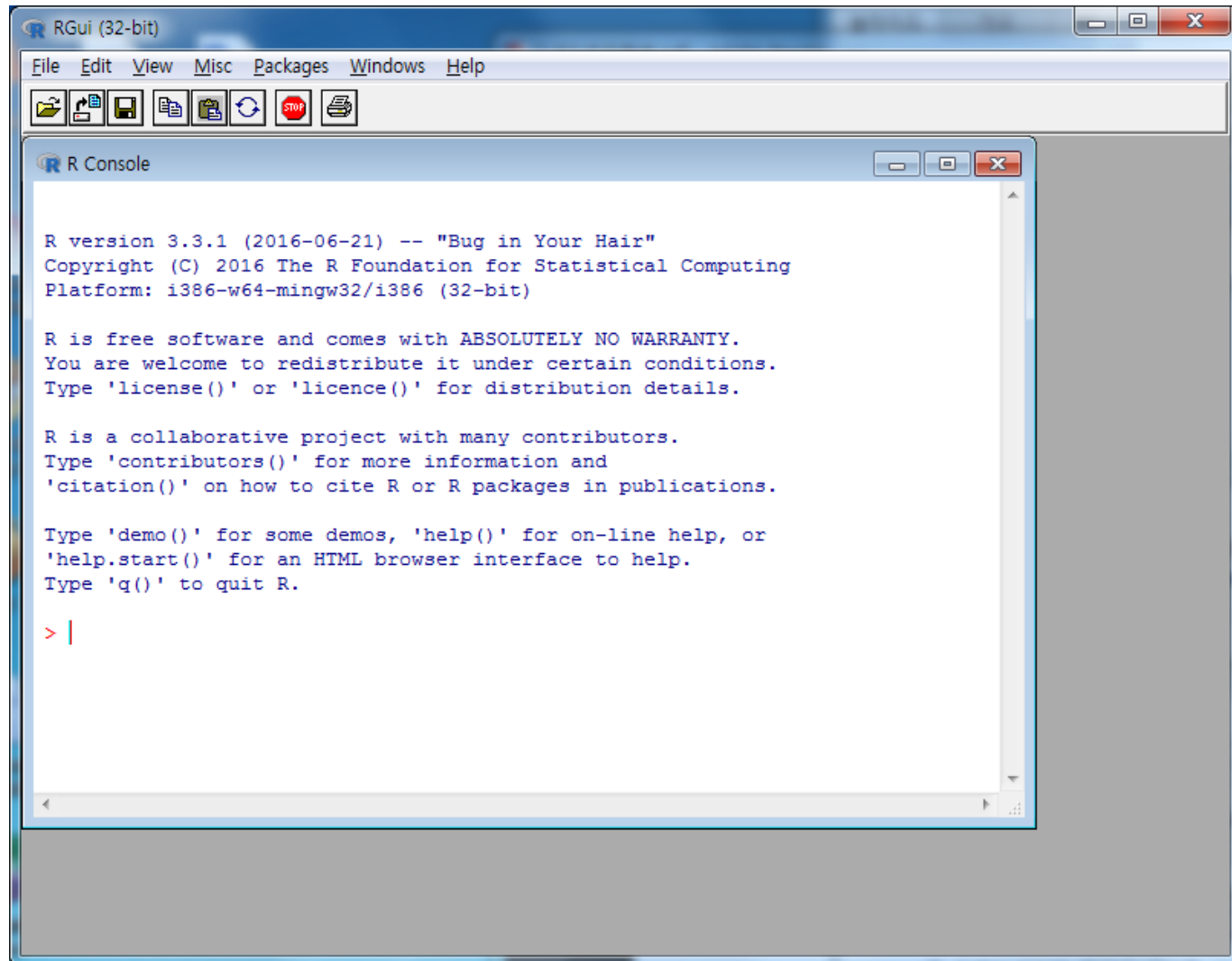
The following manuals for R were created on Debian Linux and may differ from the manuals for Mac or Windows on platform-specific pages, but most parts will be identical for all platforms. The correct version of the manuals for each platform are part of the respective R installations. The manuals change with R, hence we provide versions for the most recent released R version (R-release), a very current version for the patched release version (R-patched) and finally a version for the forthcoming R version that is still in development (R-devel).

Here they can be downloaded as PDF files, EPUB files, or directly browsed as HTML:

Manual	R-release	R-patched	R-devel
<b>An Introduction to R</b> is based on the former "Notes on R", gives an introduction to the language and how to use R for doing statistical analysis and graphics.	<a href="#">HTML</a>   <a href="#">PDF</a>   <a href="#">EPUB</a>	<a href="#">HTML</a>   <a href="#">PDF</a>   <a href="#">EPUB</a>	<a href="#">HTML</a>   <a href="#">PDF</a>   <a href="#">EPUB</a>
<b>R Data Import/Export</b> describes the import and export facilities available either in R itself or via packages which are available from CRAN.	<a href="#">HTML</a>   <a href="#">PDF</a>   <a href="#">EPUB</a>	<a href="#">HTML</a>   <a href="#">PDF</a>   <a href="#">EPUB</a>	<a href="#">HTML</a>   <a href="#">PDF</a>   <a href="#">EPUB</a>
<b>R Installation and Administration</b>	<a href="#">HTML</a>   <a href="#">PDF</a>   <a href="#">EPUB</a>	<a href="#">HTML</a>   <a href="#">PDF</a>   <a href="#">EPUB</a>	<a href="#">HTML</a>   <a href="#">PDF</a>   <a href="#">EPUB</a>
<b>Writing R Extensions</b> covers how to create your own packages, write R help files, and the foreign language (C, C++, Fortran, ...) interfaces.	<a href="#">HTML</a>   <a href="#">PDF</a>   <a href="#">EPUB</a>	<a href="#">HTML</a>   <a href="#">PDF</a>   <a href="#">EPUB</a>	<a href="#">HTML</a>   <a href="#">PDF</a>   <a href="#">EPUB</a>
A draft of <b>The R language definition</b> documents the language <i>per se</i> . That is, the objects that it works on, and the details of the expression evaluation process, which are useful to know when programming R functions.	<a href="#">HTML</a>   <a href="#">PDF</a>   <a href="#">EPUB</a>	<a href="#">HTML</a>   <a href="#">PDF</a>   <a href="#">EPUB</a>	<a href="#">HTML</a>   <a href="#">PDF</a>   <a href="#">EPUB</a>
<b>R Internals</b> : a guide to the internal structures of R and coding standards for the core team working on R itself.	<a href="#">HTML</a>   <a href="#">PDF</a>   <a href="#">EPUB</a>	<a href="#">HTML</a>   <a href="#">PDF</a>   <a href="#">EPUB</a>	<a href="#">HTML</a>   <a href="#">PDF</a>   <a href="#">EPUB</a>
<b>The R Reference Index</b> : contains all help files of the R standard and recommended packages in printable form. (9MB, approx. 3500 pages)	<a href="#">PDF</a>	<a href="#">PDF</a>	<a href="#">PDF</a>



# R 실행



The screenshot shows the RGui (32-bit) window. The menu bar includes File, Edit, View, Misc, Packages, Windows, and Help. Below the menu bar is a toolbar with icons for file operations and execution. The R Console window is open, displaying the following text:

```
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 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.

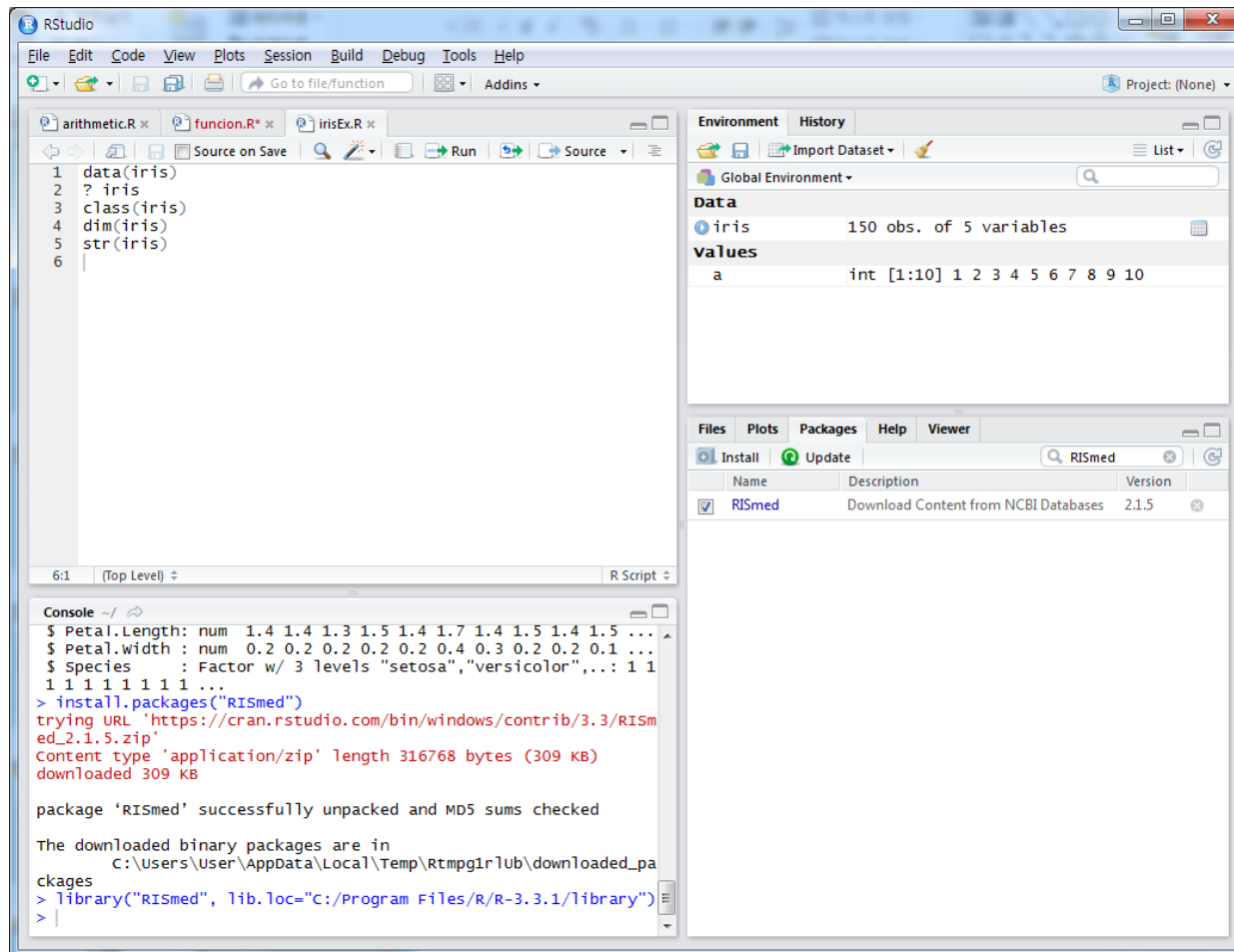
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```

# RStudio



- 명령 콘솔, 파일 편집, 데이터 보기, History, 차트 등을 접근할 수 있는 통합개발환경 제공

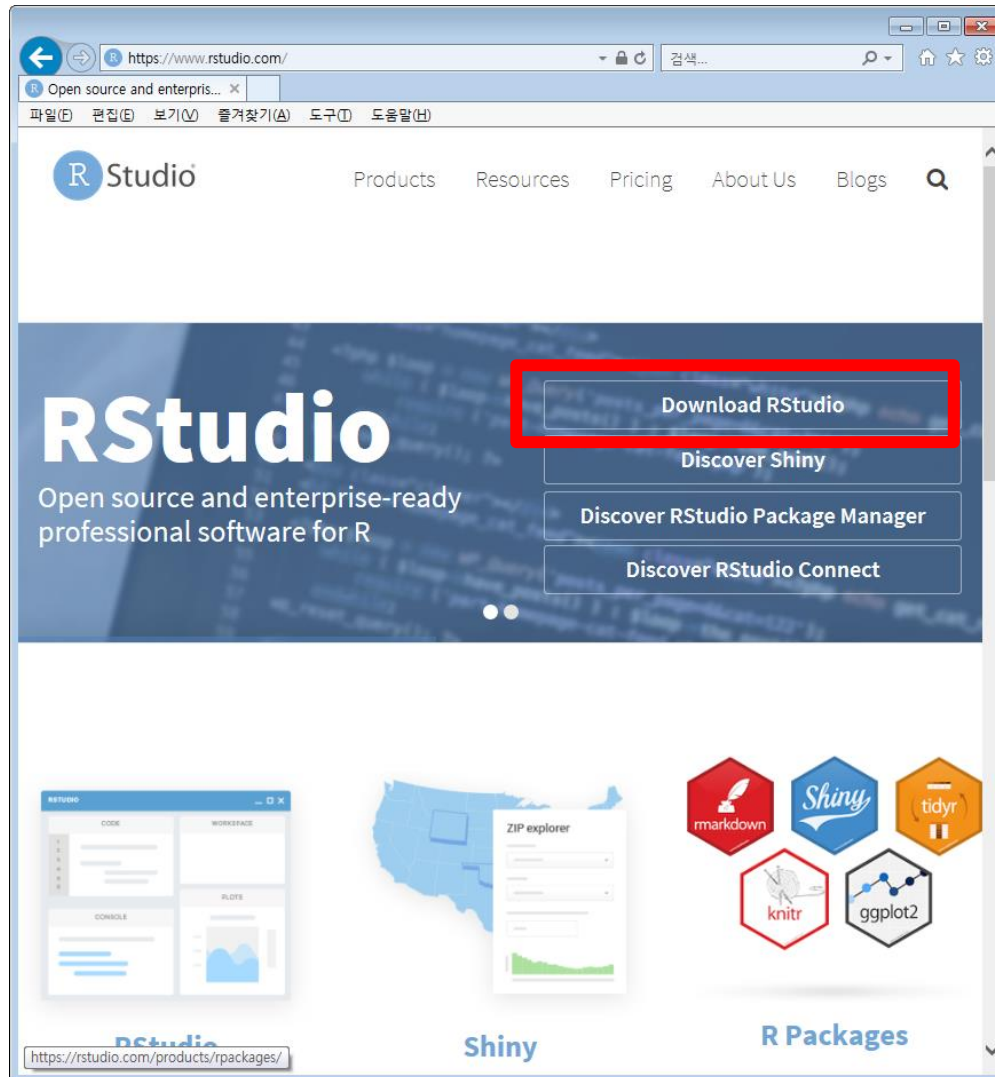




# RStudio



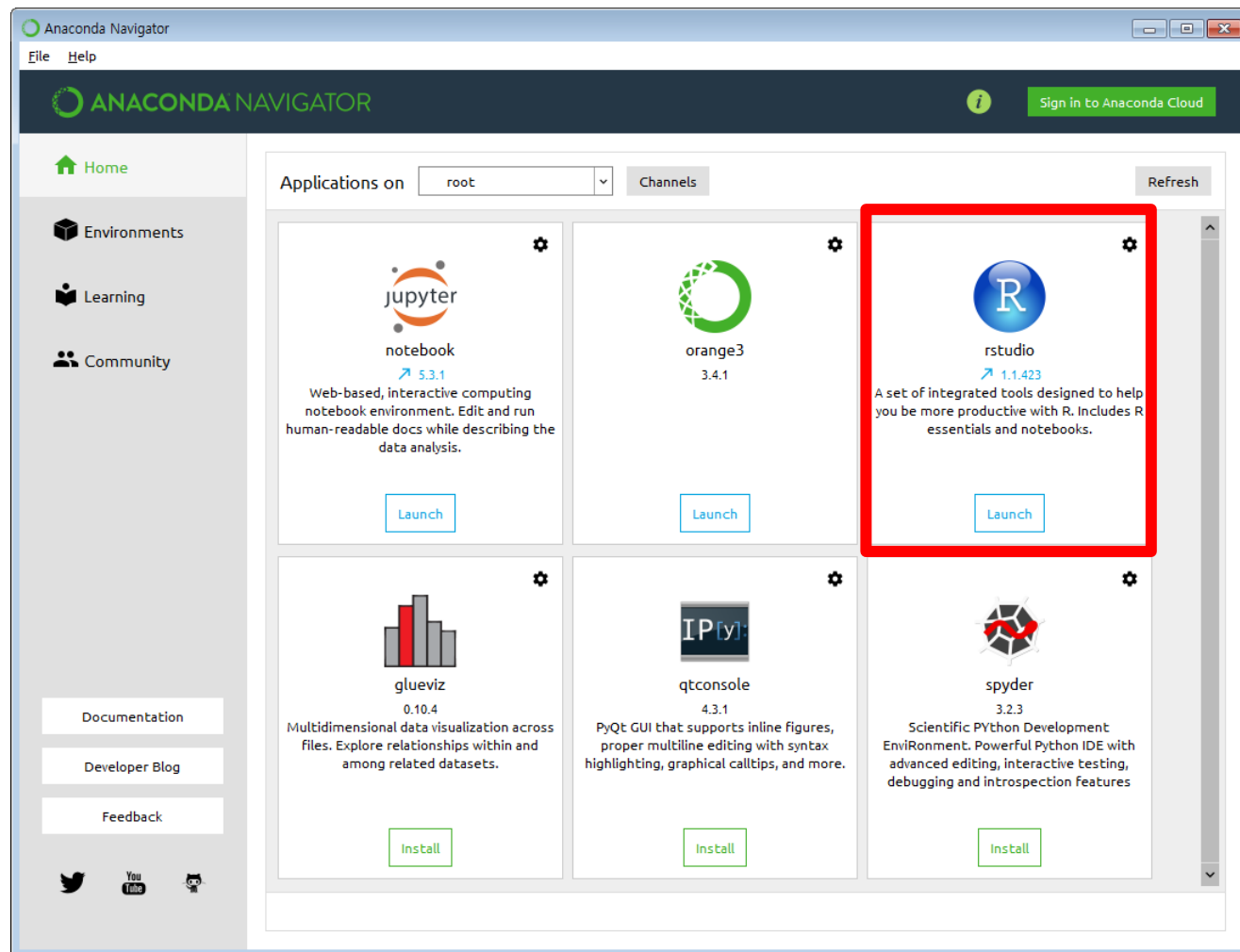
Rstudio 설치 <https://www.rstudio.com/>



# RStudio



- Anaconda Navigator에서 설치 및 시작



# RStudio 화면구성



The screenshot displays the RStudio environment with several components labeled in Korean:

- Source Script 창**: Points to the script editor on the left containing R code.
- Environment/History 창**: Points to the top-right pane showing the current environment and execution history.
- Console 창**: Points to the bottom-left pane showing the output of the executed R code.
- Files**: Points to the bottom-middle pane showing the file explorer.
- Plot**: Points to the bottom-right pane showing a scatter plot of the variable 'a'.
- Help**: Points to the bottom-right pane showing the help documentation.

The R code in the Source Script window is as follows:

```
1 10+20
2 10-20
3 2^5
4 a <- c(1:10)
5 a[1]
6 a[10]
7 plot(a)
8
9
```

The Console window output is:

```
> a <- c(1:10)
> a[1]
[1] 1
> a[10]
[1] 10
> 10+20
[1] 30
> 10-20
[1] -10
> 2^5
[1] 32
> a <- c(1:10)
> a[1]
[1] 1
> a[10]
[1] 10
```

The Environment/History window shows the variable 'a' as an integer vector of length 10.

The Files window shows the project structure with a file named 'test.R'.

The Plot window shows a scatter plot of the variable 'a' with values ranging from 1 to 10 on the x-axis and 1 to 10 on the y-axis.

The Help window provides links to R Resources, RStudio IDE Support, RStudio Cheat Sheets, RStudio Tip of the Day, RStudio Packages, RStudio Products, Manuals, and Reference.

# RStudio 기본조작



- R 콘솔창
  - 프롬프트('>')에 수식을 넣고 엔터키를 눌러 한 줄 계산식 처리
  - 이전 명령은 '↑'키를 눌러 재호출
  - 내용지우기 : "ctrl+L"
- R Script 창
  - 수식을 여러 줄에 넣어 .R 파일로 저장하여 한꺼번에 배치처리
  - 'Run'버튼을 통해 한 줄 또는 선택 줄 계산식 처리(실행)
- 도움말(help)
  - ?명령
- 주석 : '#'뒤에 설명입력
- 산술연산자
  - +(덧셈),-(뺄셈),\*(곱셈),/(나눗셈), %/(정수 나누기, 몫), %% (나머지), ^ (거듭제곱)
  - 우선순위를 위해 괄호() 사용

# RStudio 기본조작 예



The screenshot displays the RStudio interface with the following components:

- Script Editor:** Contains an R script with various arithmetic and function calls. The code is as follows:

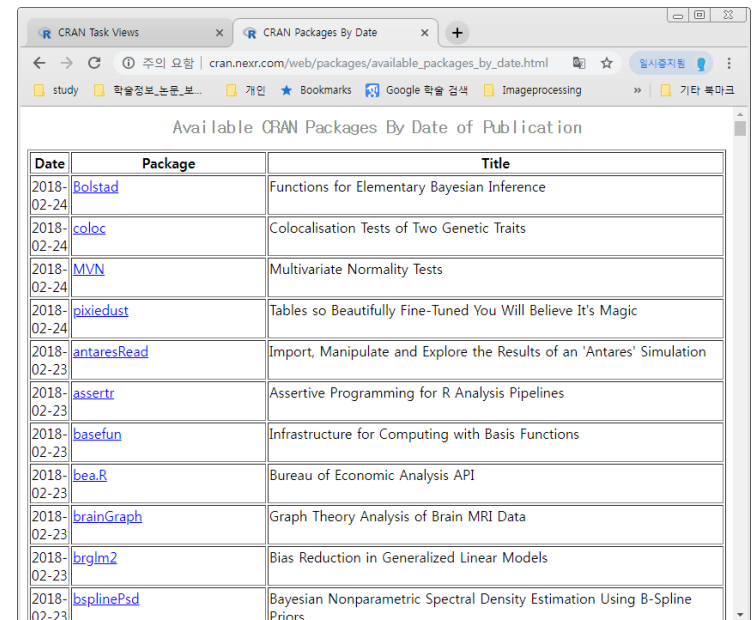
```
1 #사칙연산 +,-,*,/  
2 10+20  
3 10-20  
4 10*20  
5 10/20  
6 #정수나누기%/, 나머지%%  
7 10 %/% 3  
8 10 %% 3  
9 #제곱 ^  
10 2^5  
11 #우선순위  
12 (10+20)*3 + 2^5 - (40 -3)  
13  
14 #함수 round(), sqrt()  
15 sqrt(4 ^ 2)  
16 round(34.2)  
17 round(34.7)  
18  
19 #help  
20 ?round
```
- Console:** Shows the output of the executed code:

```
> #우선순위  
> (10+20)*3 + 2^5 - (40 -3)  
[1] 85  
>  
> #함수 round(), sqrt()  
> sqrt(4 ^ 2)  
[1] 4  
> round(34.2)  
[1] 34  
> round(34.7)  
[1] 35  
>  
> #help  
> ?round  
> 10 %/% 3
```
- Environment:** Displays "Global Environment" and "Environment is empty".
- Help Viewer:** Shows the documentation for the `round` function, titled "Rounding of Numbers". The description includes:
  - Round {base}** (R Documentation)
  - Description**
    - `ceiling` takes a single numeric argument `x` and returns a numeric vector containing the smallest integers not less than the corresponding elements of `x`.
    - `floor` takes a single numeric argument `x` and returns a numeric vector containing the largest integers not greater than the corresponding elements of `x`.
    - `trunc` takes a single numeric argument `x` and returns a numeric vector containing the integers formed by truncating the values in `x` toward 0.
    - `round` rounds the values in its first argument to the specified number of decimal places (default 0).

# R 라이브러리



- R의 가장 큰 장점중 하나는 개발자들이 구축해 다양한 라이브러리 (package) 사용
- <http://cran.nexr.com/web/packages/>에서 확인
- 설치 : `install.packages("패키지명")`
- 업데이트 : `update.packages()` : 설치된 패키지들을 확인해 최신 버전을 설치.
- 도움말 : `help(package = "패키지명")`
- 패키지 사용 : `library ("패키지명")`

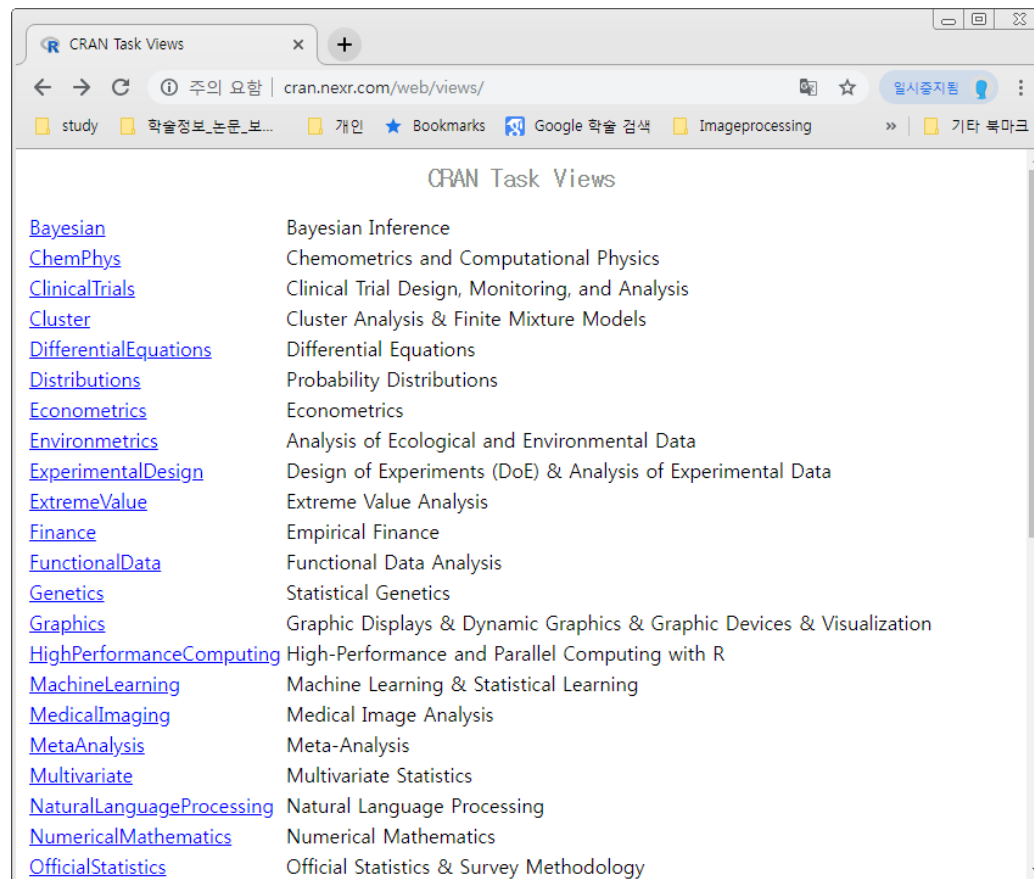


Available CRAN Packages By Date of Publication

Date	Package	Title
2018-02-24	<a href="#">Bolstad</a>	Functions for Elementary Bayesian Inference
2018-02-24	<a href="#">coloc</a>	Colocalisation Tests of Two Genetic Traits
2018-02-24	<a href="#">MVN</a>	Multivariate Normality Tests
2018-02-24	<a href="#">pixiedust</a>	Tables so Beautifully Fine-Tuned You Will Believe It's Magic
2018-02-23	<a href="#">antaresRead</a>	Import, Manipulate and Explore the Results of an 'Antares' Simulation
2018-02-23	<a href="#">assertr</a>	Assertive Programming for R Analysis Pipelines
2018-02-23	<a href="#">basefun</a>	Infrastructure for Computing with Basis Functions
2018-02-23	<a href="#">bea.R</a>	Bureau of Economic Analysis API
2018-02-23	<a href="#">brainGraph</a>	Graph Theory Analysis of Brain MRI Data
2018-02-23	<a href="#">brglm2</a>	Bias Reduction in Generalized Linear Models
2018-02-23	<a href="#">bsplinePsd</a>	Bayesian Nonparametric Spectral Density Estimation Using B-Spline Priors

# R 라이브러리

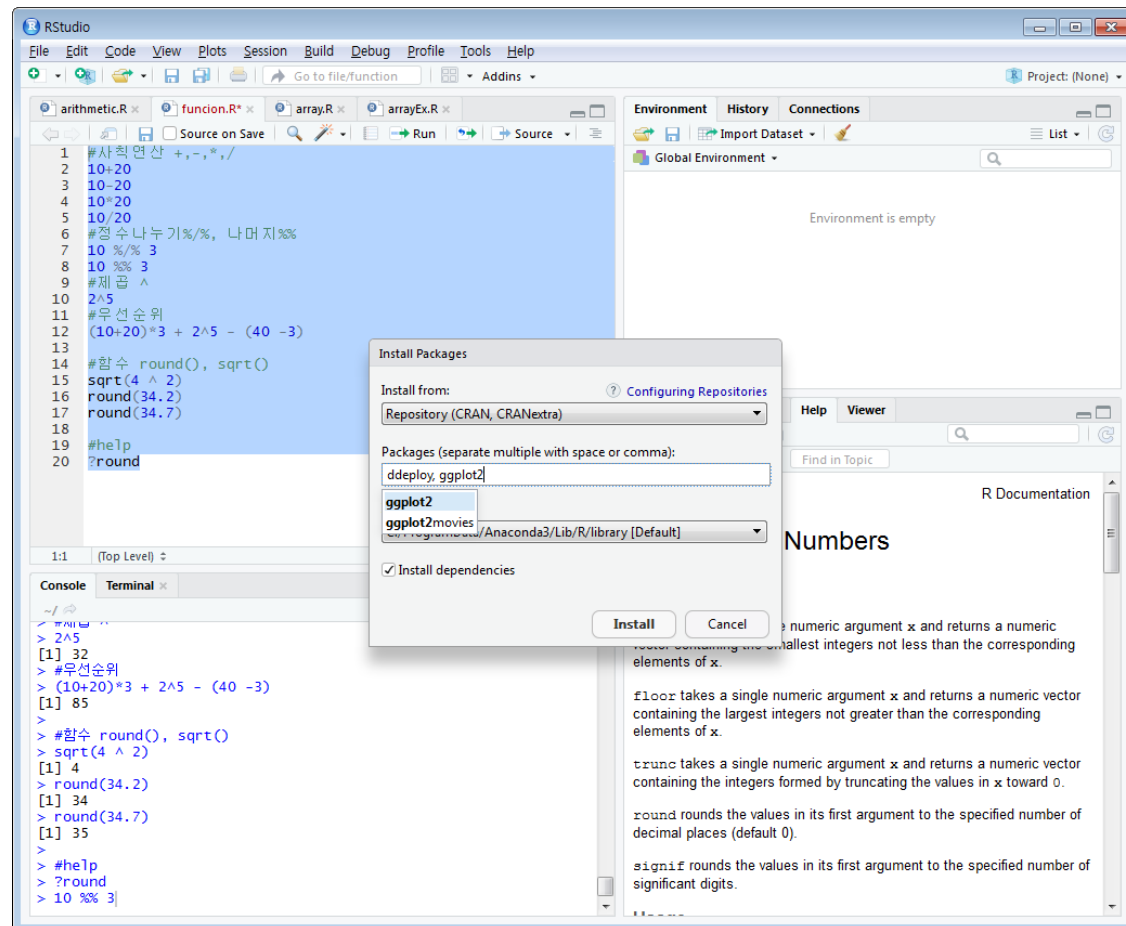
- CRAN Task Views
  - 적용 분야에 대한 문서 및 패키지





# R 라이브러리 설치

- 데이터가공, 시각화, 모형화를 위한 dplyr, ggplot2 라이브러리 설치
- RStudio 메뉴 [Tools]-[Install Package]-dplyr, ggplot2





# R 기초\_Data



- 숫자형
  - 정수(1, -5, 0..), 부동 소수(34.78, -0.05..), 복소수(  $2i + 4$  )
- 문자열
  - 작은 따옴표나 큰 따옴표로 묶여진 자료
  - 'this is string' , "this is string"
- 진리값
  - TRUE, T : 참값
  - FALSE, F : 거짓
  - 비교 연산자 : ==, !=, <, <=, >, >=
  - 논리연산자 : &, && (AND), |, || (OR), ! (NOT)
- NA
  - 데이터에 값이 존재하지 않는 결측치 표현
- NULL
  - 변수가 초기화 되지 않은 경우 사용
  - is.null()을 사용하여 판단

# R기초\_변수(variable)



- 값이나 계산 결과를 대입하여 저장
- 변수명 규칙
  - 알파벳, 숫자, '\_', '.'로 구성
  - 대소문자 구분
  - 영문자와 .으로만 첫글자 사용 가능
  - R 예약어 사용불가 : NA, TRUE, function, if, else, for, next, in
- 변수명 예
  - a, a1, name, a.name, a123
- 변수에 값 할당 : <-, <<-, =

```
변수 <- 값  
변수 <- 수식
```

```
#변수 사용 예  
#두수의 사칙연산  
a=100  
b=200  
sum= a+b  
sub= a-b  
mul = a*b  
div = a/b  
mod = a %% b
```

```
#원넓이, 원둘레 계산  
r=50.6  
area = pi * r ^ 2  
per = 2 * pi * r
```

```
#문자열  
name = "홍길동"
```

```
#진리값  
male = F  
female = TRUE
```

# R기초\_ 비교, 논리 연산



- 연산결과가 진리 값이 나오거나 진리 값을 대상으로 연산

#비교 연산

5 == 6

5 != 6

5 < 6

5 <= 6

5 > 6

5 >= 6

!(5 == 6)

#논리연산

! TRUE

TRUE & TRUE

TRUE & FALSE

TRUE | FALSE

#비교, 논리 연산으로 변수의 값 범위 확인

x = 100

x >= 10

#x가 10이상인가 확인

(x >= 30) & (x <= 50)

#x가 30-50사이의 범위 인가 확인

(x < 30) || (x > 50)

#x가 30-50 범위를 벗어나는가 확인

# R기초\_출력 함수



- `print()` : 객체를 출력
- `cat()` : 문자열을 콤마로 나열하여 출력

```
x=10
y=23.6
z="hallym"
print(x)
print(y)
print(z)
cat("x=" , x , "y=" ,y , "\n", "z=" , z)
```

# 출력 함수



- `sprintf("형식", 변수)` : 여러 타입의 값을 하나의 문자열로 묶어서 출력  
형식 : %d (십진수), %f(실수), %s(문자열)

```
> sprintf("x=%5d y=%10.2f, x=%s", x,y,z)
[1] "x=  10 y=  23.60, x=hallym"
> sprintf("%f", pi)
[1] "3.141593"
> sprintf("%.3f", pi)
[1] "3.142"
> sprintf("%1.0f", pi)
[1] "3"
> sprintf("%5.1f", pi)
[1] " 3.1"
> sprintf("%05.1f", pi)
[1] "003.1"
> sprintf("%+f", pi)
[1] "+3.141593"
> sprintf("% f", pi)
[1] " 3.141593"
> sprintf("%-10f", pi) # left justified
[1] "3.141593 "
> sprintf("%e", pi) # exponential
[1] "3.141593e+00"
> sprintf("%E", pi)
[1] "3.141593E+00"
```

# 연습문제



(1) 본인의 성명, 키와 몸무게를 이용하여 신체질량지수 (body mass index, BMI) 계산

$$\text{BMI} = \text{체중(kg)} \div \{\text{신장(m)} \times \text{신장(m)}\}$$

(2)미터(m) 값을 변수에 대입하여 인치(inch), 야드(yd), 마일(mile), 피트(ft)로 변환하는 R스크립트를 작성하여 실행

(3)나이를 변수에 대입하여 50이하이면 TRUE, 아니면 FALSE 결과 확인

(4)나이가 20대이면 TRUE, 아니면 FALSE 결과 확인