

# R기반\_데이터 분석

# BIG 데이터 개요

## ➤ 빅데이터의 3V(가트너 그룹)

- 용량(Volume)

: 데이터 용량의 증가 → 대 단위 분석 이슈.

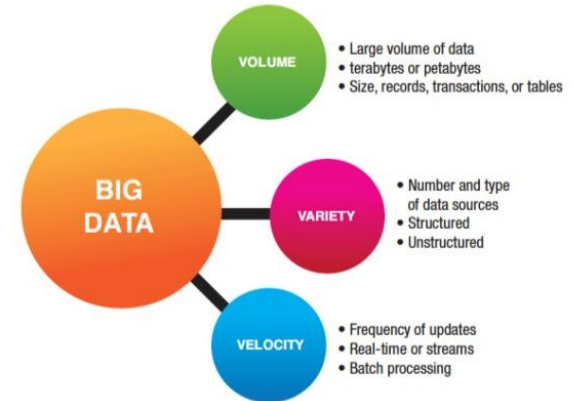
- 속도(Velocity)

: 빠른 데이터 생성 → 빠른 데이터 처리.

- 다양성(Variety)

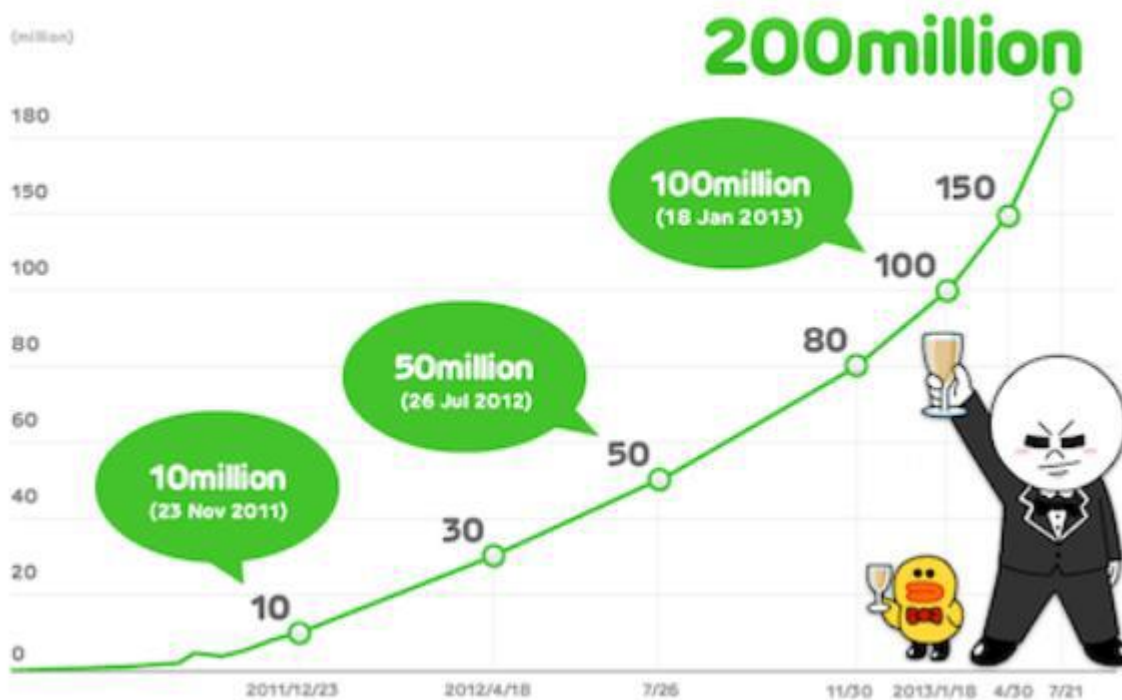
: 소셜미디어, 모바일, DB → 다차원 의사결정

### The Three Vs of Big Data



# BIG 데이터 등장 배경

- 스마트 폰 및 IT 디지털 기기 → 사용량 증가
- SNS 확산 → 비정형 데이터 증가
- 비정형 데이터 처리 기술 향상



# BIG 데이터 정의 및 특징

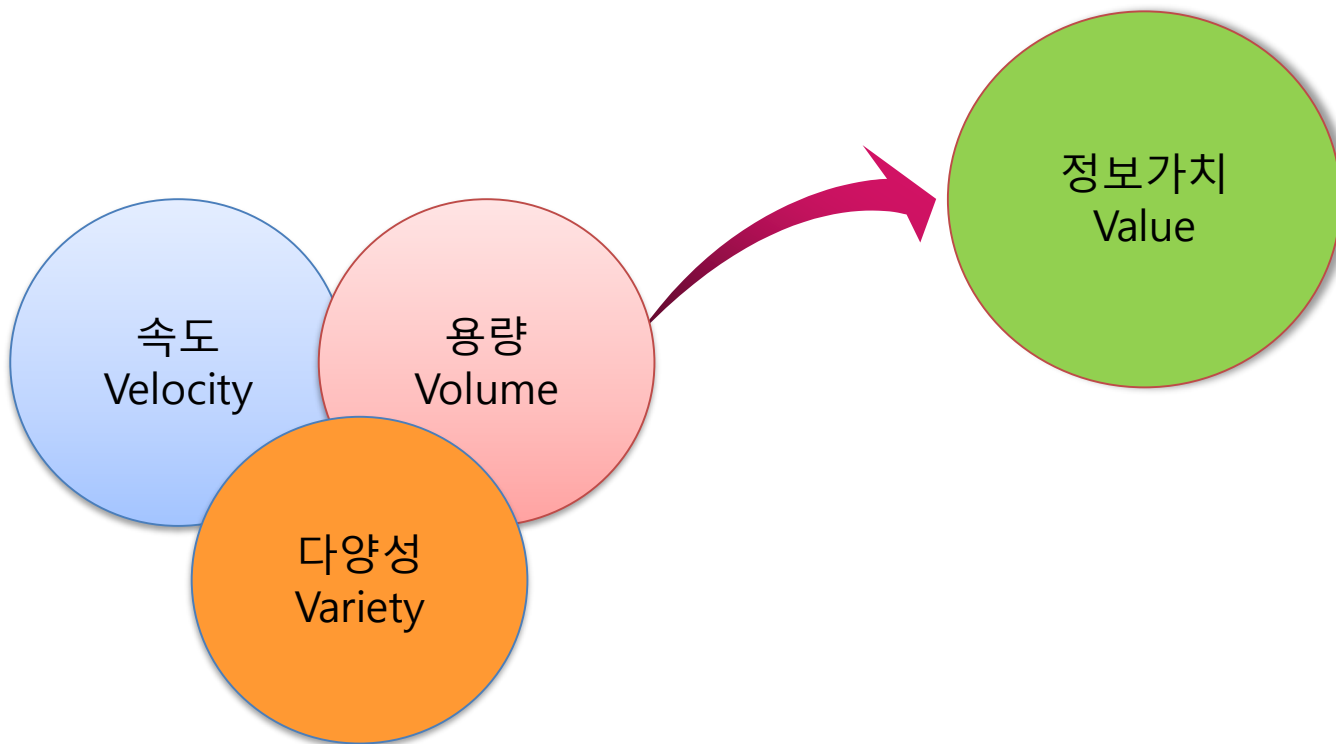
- 일반적으로 사용되는 데이터 수집, 관리, 처리 소프트웨어의 수용 한계를 초과한 데이터
- 기존 데이터베이스의 역량을 넘어서는 대량의 정형 또는 비정형 데이터의 집합
- 기존 데이터 규모에서 불가능했던 새로운 통찰이나 새로운 형태의 가치를 추출 및 예측 할 수 있는 데이터 집합체



# BIG 데이터 정의 및 특징

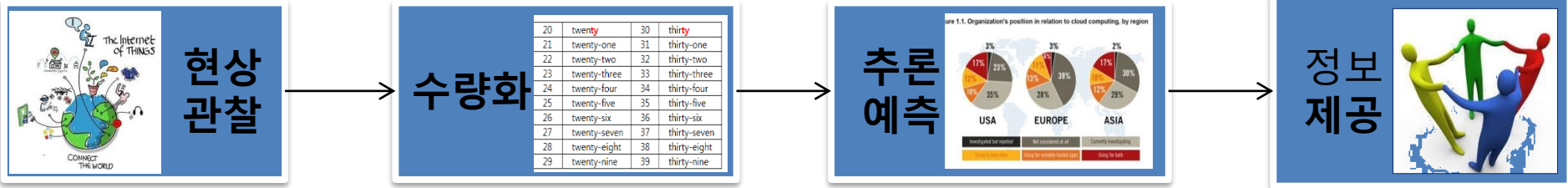
---

## ➤ 빅데이터 분석의 목적



# BIG 데이터 분석 과정

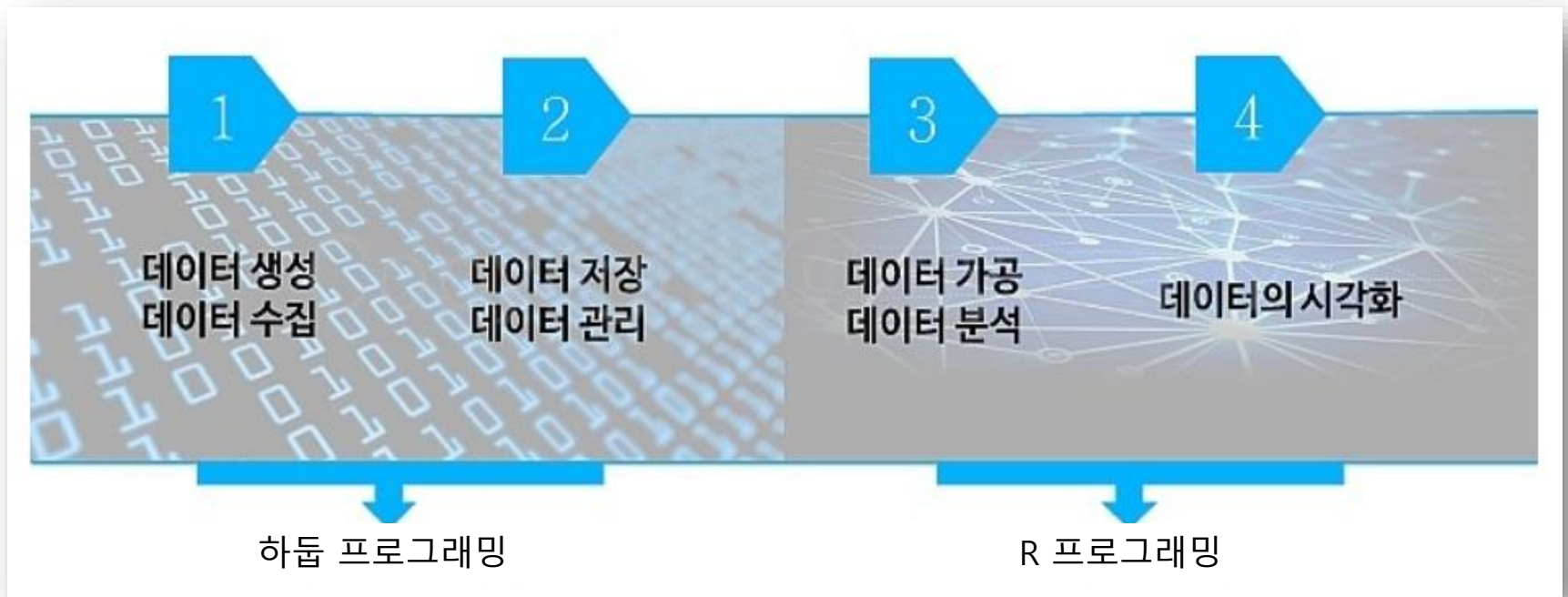
## ➤ 정형 / 비정형 데이터 분석 과정



- 실 세계의 현상을 관찰하여 수량화하고 이를 통해서 패턴을 분석하여 추론/예측하는 일련의 과정

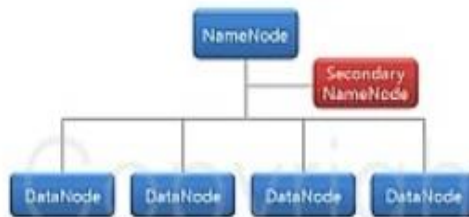
# BIG 데이터 분석 과정

## ➤ 하둡과 R 프로그래밍



# BIG 데이터 분석 과정

## ➤ 하둡과 R 프로그래밍



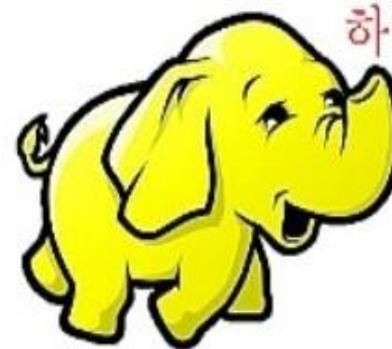
· HDFS의 기본 구성 ·

여러 대의 컴퓨터를 병렬구조로 하둡 클러스터라는 여러 대의 서버를 하나의 서버로 묶는 방식으로 방대한 양의 데이터를 분산처리하여 빠른 시간 내에 결과를 제공하는 데이터 관리 기술. 핵심기술은 분산파일시스템(HDFS)와 맵리듀스(MapReduce)가 있다.



· 맵리듀스의 Word Count Processing 예제 ·

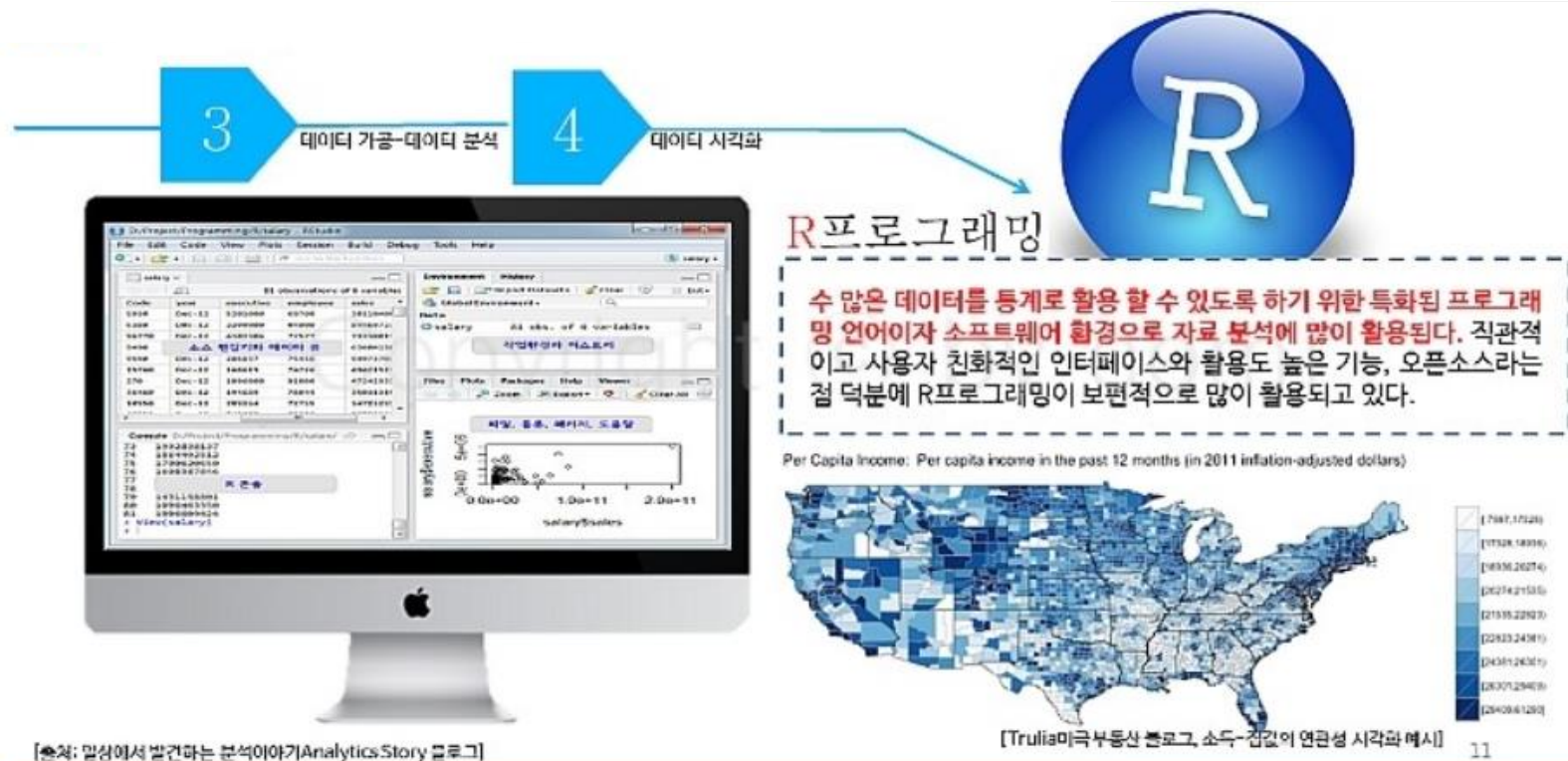
[출처: HBI 기술 연구소]



하둡 프로그래밍



# BIG 데이터 분석 과정



# BIG 데이터 분석 과정

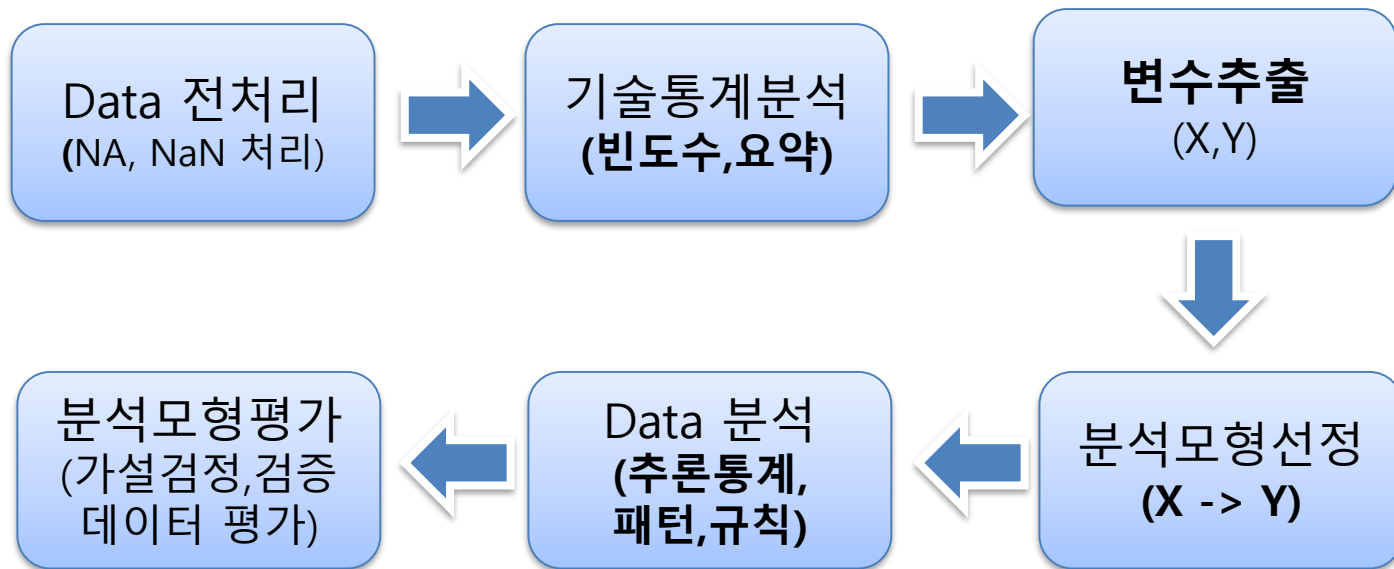
## ➤ 일반적인 분석 절차



# BIG 데이터 분석 과정

---

## ➤ 일반적인 분석 절차




# BIG 데이터 분석 방법

---

## ➤ 분석 유형

- 기술통계분석 : 분석 데이터의 특성 분석
- 추론통계분석 : 통계 및 수학 기초 분석 및 검정
- 텍스트 마이닝 : 키워드 및 연관어 분석
- 데이터 마이닝
  - ❖ 지도학습 : 인과관계 기반 예측분석
  - ❖ 비지도학습 : 패턴 기반 예측분석

# BIG 데이터 분석 방법

기술통계분석	추론통계분석	데이터마이닝																									
<ul style="list-style-type: none"><li>데이터 특성분석</li></ul>	<ul style="list-style-type: none"><li>모집단을 표본으로 분석하여 추론</li></ul>	<ul style="list-style-type: none"><li>변수의 관계와 패턴 분석으로 미래 예측</li></ul>																									
<table><tr><th colspan="2">학력수준</th><th>실패</th><th>진학</th></tr><tr><td rowspan="2">고졸</td><td>관찰빈도</td><td>40</td><td>49</td></tr><tr><td>기대빈도</td><td>36</td><td>54</td></tr><tr><td rowspan="2">대졸</td><td>관찰빈도</td><td>27</td><td>55</td></tr><tr><td>기대빈도</td><td>33</td><td>49</td></tr><tr><td rowspan="2">대학원졸</td><td>관찰빈도</td><td>23</td><td>31</td></tr><tr><td>기대빈도</td><td>21</td><td>32</td></tr></table>	학력수준		실패	진학	고졸	관찰빈도	40	49	기대빈도	36	54	대졸	관찰빈도	27	55	기대빈도	33	49	대학원졸	관찰빈도	23	31	기대빈도	21	32		
학력수준		실패	진학																								
고졸	관찰빈도	40	49																								
	기대빈도	36	54																								
대졸	관찰빈도	27	55																								
	기대빈도	33	49																								
대학원졸	관찰빈도	23	31																								
	기대빈도	21	32																								
<ul style="list-style-type: none"><li>합계, 평균, 빈도수, 비율, 표준편차, 분산, 교차분석 등</li><li>데이터의 특성 분석</li></ul>	<ul style="list-style-type: none"><li>카이제곱검정, 비율 검정, 평균차이검정, 상관분석, 분산분석</li><li>집단간 차이 분석</li></ul>	<ul style="list-style-type: none"><li>텍스트분석, 예측, 분류, 군집, 연관분석</li><li>패턴 및 규칙을 이용한 의사결정</li></ul>																									

# BIG 데이터 파급효과

---

- 기존 사실에 대한 체계적인 객관적 근거 제시
- 다변화된 현대 사회를 더욱 정확하게 예측 및 대응 가능
- 개인화된 현대 사회 구성원 마다 맞춤형 정보 제공, 관리, 분석 가능
- 정치, 사회, 경제, 문화, 과학 기술 등 전 영역에 걸쳐서 사회와 인류에게  
가치 있는 정보 제공(중요성 부각)

# BIG 데이터 활용 사례



▲ 서울시와 KT가 함께 추진한 심야버스 노선정책

# BIG 데이터 활용 사례

## 경제적 효과

	계수	2012	2015	2020
생산유발효과(억원)	1.32	95,573	119,736	171,296
부가가치 유발효과 (억원)	0.99	71,680	89,802	128,472
인력창출(명)	1.24	89,781	112,479	160,914

**\*유통업계 재고관리에 효율적인 빅데이터 사용 사례**  
[파리바게트의 날씨 마케팅]  
27도 이상의 맑은 날씨: 샌드위치 판매 상승  
20도 이하의 쌀쌀 날씨: 피자 빵

[출처: 한국통신학회 (정보통신) 제29권제11호, 48-54 (7 pages), 황기순 (빅데이터) 등장에 따른  
경제적 파급효과 및 법 (규제) 연구, 이규철, 김희선, 2012.10.]





# BIG 데이터 활용 사례

공공측면(재난대응)



싱가포르: 국가위협관리, 일본은 자연재해방지



# BIG 데이터 활용 사례

## ➤ 다양한 분야에서 빅데이터 활용 사례

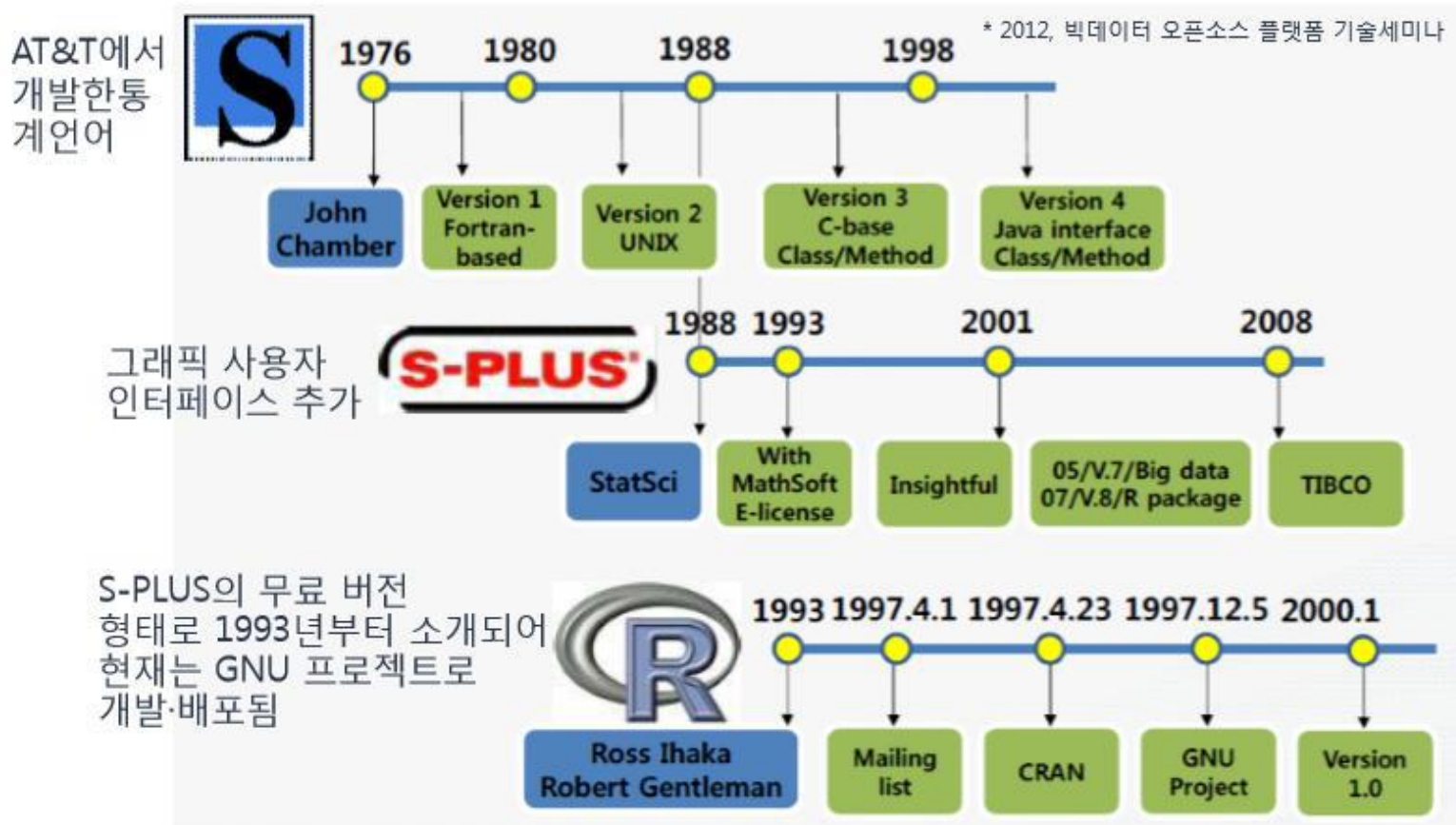
관련분야	빅 데이터 활용 사례
공공분야	▪ 교통, 부정부패, 세수 증감 데이터 분석
복지분야	▪ 자살 예보시스템, 실버 계층 의료 개선
기상분야	▪ 날씨 분석과 재난 예고
정치분야	▪ 소셜 데이터를 통한 선거 전략 수립, 맞춤형 캠페인
의료분야	▪ 인간 유전자 데이터 분석과 희귀병 치료
금융분야	▪ 주가지수 예측, 거시 변수 예측
의류분야	▪ 유행 디자인 사전 파악 및 시장 선호도 분석
스포츠분야	▪ 선수 부상예측, 상대팀 전술 분석 및 대응

# R 설치 및 데이터 분석 환경 만들기

# R 설치 및 개요

## ➤ R 프로그램 탄생과 발전 과정

- 1993년 뉴질랜드 오클랜드 대학의 통계학과 교수2명(Ross Ihaka, Robert Gentleman) 개발.



# R 설치 및 개요

---

## ➤ R이란?

- 데이터 분석을 위한 통계 및 그래픽스를 지원하는 무료 소프트웨어 환경.

## ➤ R 프로그램 특징

- In-Memory Computing
  - ✓ 빠른 처리 속도(H/W 메모리 크기에 영향 받음).
- Object-Oriented Programming
  - ✓ 데이터, 함수, 차트 등 모든 것이 object로 관리.
- Package
  - ✓ 최신의 알고리즘 및 방법론이 Package로 제공됨.
  - ✓ 통계, 기계 학습, 금융, 생물 정보학, 그래픽스에 이르는 다양한 통계 패키지를 갖추고 있으며, 이 모든 것이 무료로 제공.
- Visualization
  - ✓ 분석에 통찰을 부여할 수 있는 그래픽에 대한 강력한 지원.

# R 설치 및 개요

➤ 다운로드 : <https://www.r-project.org/>

- CRAN Site에 수많은 패키지 등록 : `dim(available.packages())`

The image shows a screenshot of a web browser displaying the R Project website. The main page, titled "The R Project for Statistical Computing", features a navigation menu on the left with links like "Download", "About R", and "Documentation". The "Download" link is highlighted with a red box, and the "CRAN" link below it is also highlighted. The main content area includes a "Getting Started" section and a "News" section. An inset window shows the "CRAN - Mirrors" page, which lists various mirrors for CRAN. The "Korea" section of the mirrors page is highlighted with a red box, showing links to mirrors at Pukyong National University, Yeungnam University, Bigdata Campus, Seoul Metropolitan Government, and The Genome Institute of UNIST.

**The R Project for Statistical Computing**

[Home]

**Download**

[CRAN](#)

**R Project**

[About R](#)

[Contributors](#)

[What's New?](#)

[Mailing Lists](#)

[Bug Tracking](#)

[Conferences](#)

[Search](#)

**R Foundation**

[Foundation](#)

[Board](#)

[Members](#)

[Donors](#)

[Donate](#)

**Documentation**

**Getting Started**

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred [CRAN mirror](#).

If you have questions about R like how the license terms are, please read our [FAQ](#) or you send an email.

**News**

- **R 3.2.1 (World-Famous Astronomer)** June 8. Final release is scheduled for June 8, 2015.
- **R version 3.2.0 (Full of Ingredients)** June 8, 2015.
- **R version 3.1.3 (Smooth Sidewalk)** June 8, 2015.
- **The R Journal Volume 6/2** is available.
- **useR! 2015**, will take place at the University of California, San Diego, July 3, 2015.
- **useR! 2014**, took place at the University of California, San Diego, July 3, 2014.

**CRAN - Mirrors**

<https://cran.ism.ac.jp/>  
<https://ftp.yz.yamagata-u.ac.jp/pub/cran/>

**Japan**

The Institute of Statistical Mathematics, Tokyo  
Yamagata University

**Korea**

<https://ftp.harukasan.org/CRAN/>  
<https://cran.yu.ac.kr/>  
<https://cran.seoul.go.kr/>  
<https://cran.biodisk.org/>

Information and Database Systems Laboratory, Pukyong National University  
Yeungnam University  
Bigdata Campus, Seoul Metropolitan Government  
The Genome Institute of UNIST (Ulsan National Institute of Science and Technology)

**Malaysia**

<https://mirrors.upm.edu.my/CRAN/>

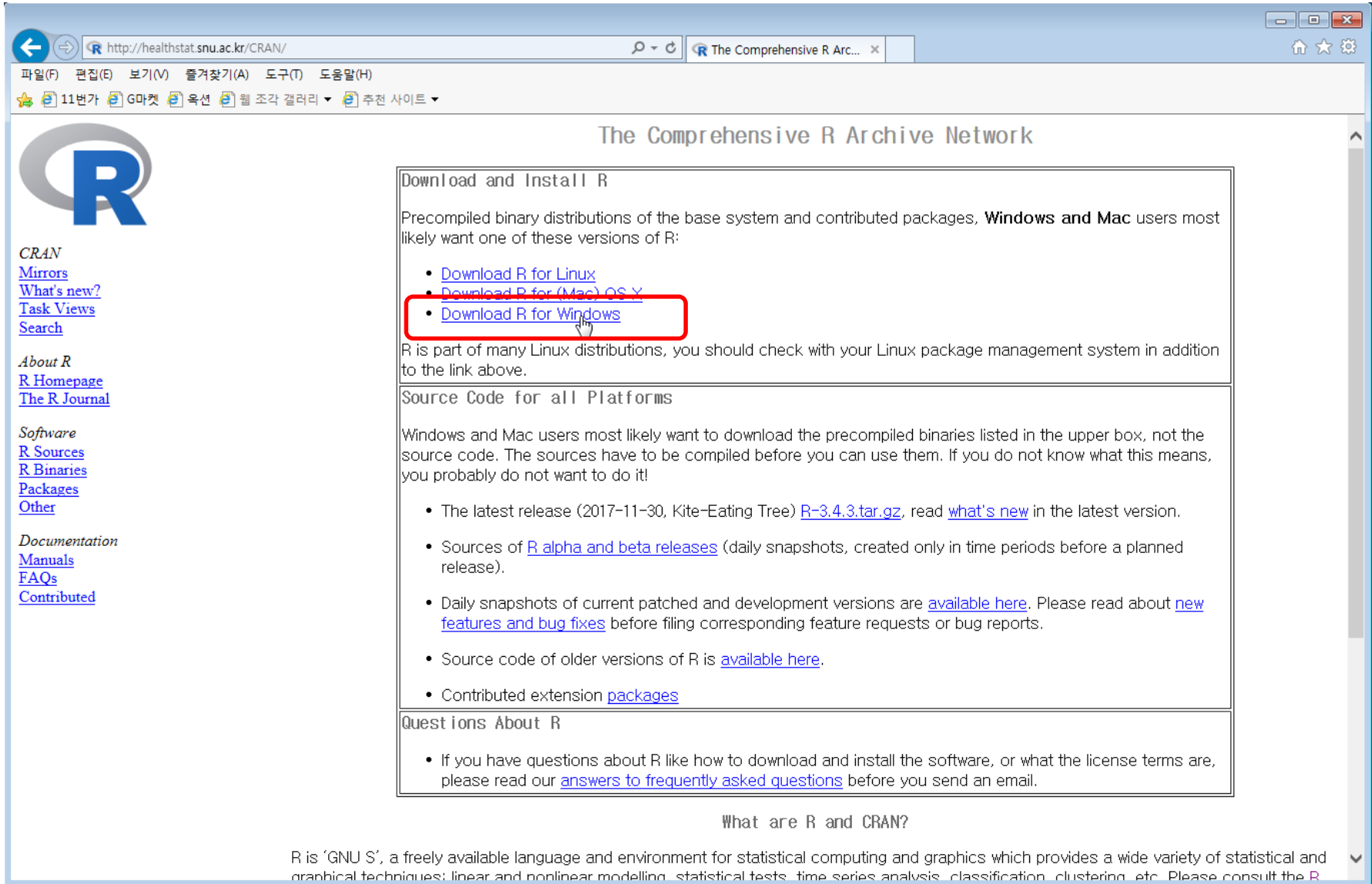
Universiti Putra Malaysia

**Mexico**

<https://cran.seoul.go.kr/itam.mx/>

Instituto Tecnológico Autónomo de México

# R 설치 및 개요



The screenshot shows a web browser window with the address bar displaying <http://healthstat.snu.ac.kr/CRAN/>. The page title is "The Comprehensive R Archive Network". The left sidebar contains links for CRAN (Mirrors, What's new?, Task Views, Search), About R (R Homepage, The R Journal), Software (R Sources, R Binaries, Packages, Other), and Documentation (Manuals, FAQs, Contributed). The main content area is titled "The Comprehensive R Archive Network" and contains three sections: "Download and Install R", "Source Code for all Platforms", and "Questions About R".

**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 \(Mac\) OS 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.

**Source Code for all Platforms**

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2017-11-30, Kite-Eating Tree) [R-3.4.3.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

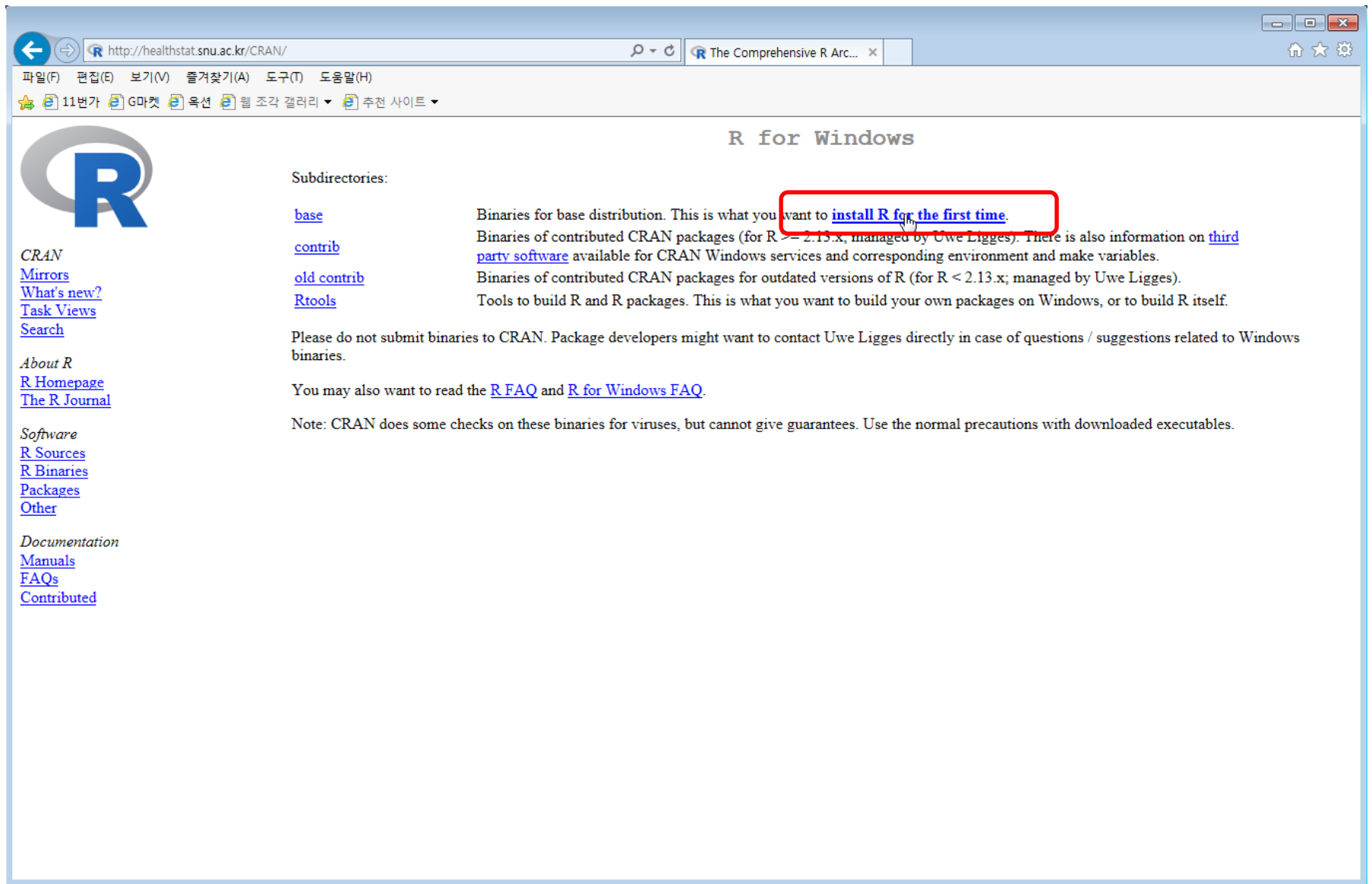
**Questions About R**

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

**What are R and CRAN?**

R is 'GNU S', a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc. Please consult the [R](#)

# R 설치 및 개요



The screenshot shows a web browser window with the address bar displaying <http://healthstat.snu.ac.kr/CRAN/>. The browser's menu bar includes options like '파일(F)', '편집(E)', '보기(V)', '즐거찾기(A)', '도구(T)', and '도움말(H)'. The toolbar shows icons for '11번가', 'G마켓', '옥션', '웹 조각 갤러리', and '추천 사이트'. The main content area is titled 'R for Windows' and features the R logo on the left. The logo consists of a large blue 'R' with a grey circle around it. To the right of the logo, under the heading 'Subdirectories:', there are four links: 'base', 'contrib', 'old contrib', and 'Rtools'. Each link has a corresponding description. The 'base' link is highlighted with a red box and a red circle around the text 'install R for the first time'. The descriptions for 'base', 'contrib', and 'old contrib' mention 'Uwe Ligges' and 'CRAN Windows services'. The 'Rtools' link mentions 'building R and R packages'. Below the subdirectories, there is a paragraph about not submitting binaries to CRAN and a note about virus checks. At the bottom, there are links for 'About R', 'Software', and 'Documentation'.

**R for Windows**

Subdirectories:

- [base](#): Binaries for base distribution. This is what you want to **install R for the first time**.
- [contrib](#): Binaries of contributed CRAN packages (for R >= 2.13.x, managed by Uwe Ligges). There is also information on [third party software](#) available for CRAN Windows services and corresponding environment and make variables.
- [old contrib](#): Binaries of contributed CRAN packages for outdated versions of R (for R < 2.13.x; managed by Uwe Ligges).
- [Rtools](#): Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.

Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

**CRAN**

- [Mirrors](#)
- [What's new?](#)
- [Task Views](#)
- [Search](#)

**About R**

- [R Homepage](#)
- [The R Journal](#)

**Software**


- [R Sources](#)
- [R Binaries](#)
- [Packages](#)
- [Other](#)

**Documentation**

- [Manuals](#)
- [FAQs](#)
- [Contributed](#)



# R 설치 및 개요



CRAN  
[Mirrors](#)  
[What's new?](#)  
[Task Views](#)  
[Search](#)

About R  
[R Homepage](#)  
[The R Journal](#)

Software  
[R Sources](#)  
[R Binaries](#)  
[Packages](#)  
[Other](#)

Documentation  
[Manuals](#)  
[FAQs](#)  
[Contributed](#)

## R-4.1.2 for Windows (32/64 bit)

[Download R 4.1.2 for Windows](#) (86 megabytes, 32/64 bit)  
[Installation and other instructions](#)  
[New features in this version](#)

If you want to double-check that the package you have downloaded matches the package distributed by CRAN, you can compare the [md5sum](#) of the .exe to the [fingerprint](#) on the master server. You will need a version of md5sum for windows: both [graphical](#) and [command line versions](#) are available.

### Frequently asked questions

- [Does R run under my version of Windows?](#)
- [How do I update packages in my previous version of R?](#)
- [Should I run 32-bit or 64-bit R?](#)

Please see the [R FAQ](#) for general information about R and the [R Windows FAQ](#) for Windows-specific information.

### Other builds

- Patches to this release are incorporated in the [r-patched snapshot build](#).
- A build of the development version (which will eventually become the next major release of R) is available in the [r-devel snapshot build](#).
- [Previous releases](#)

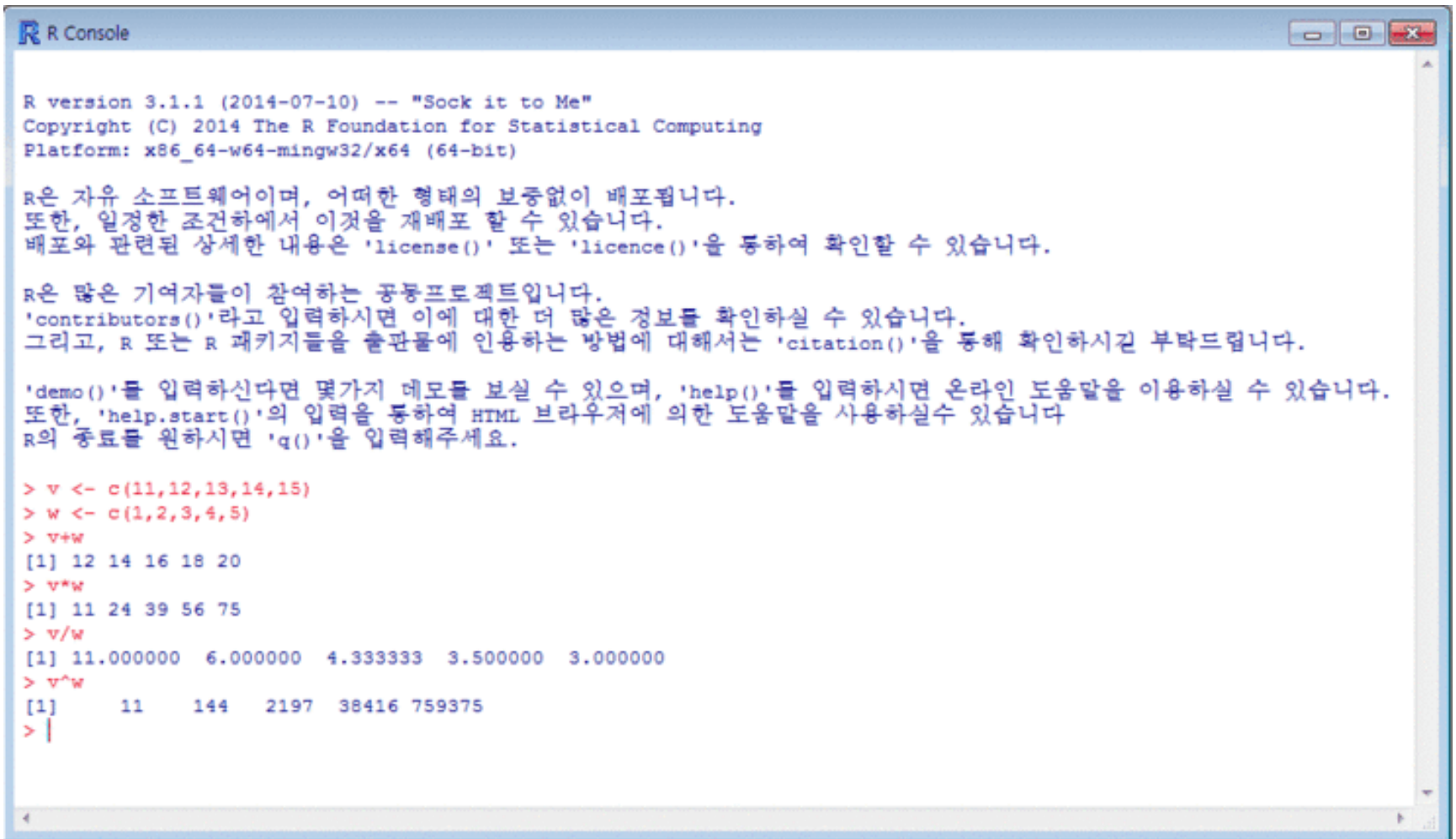
Note to webmasters: A stable link which will redirect to the current Windows binary release is [<CRAN MIRROR>/bin/windows/base/release.html](#).

---

Last change: 2021-11-01

# R 시작하기

## ➤ R console 화면



```
R Console

R version 3.1.1 (2014-07-10) -- "Sock it to Me"
Copyright (C) 2014 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()'을 통하여 확인할 수 있습니다.

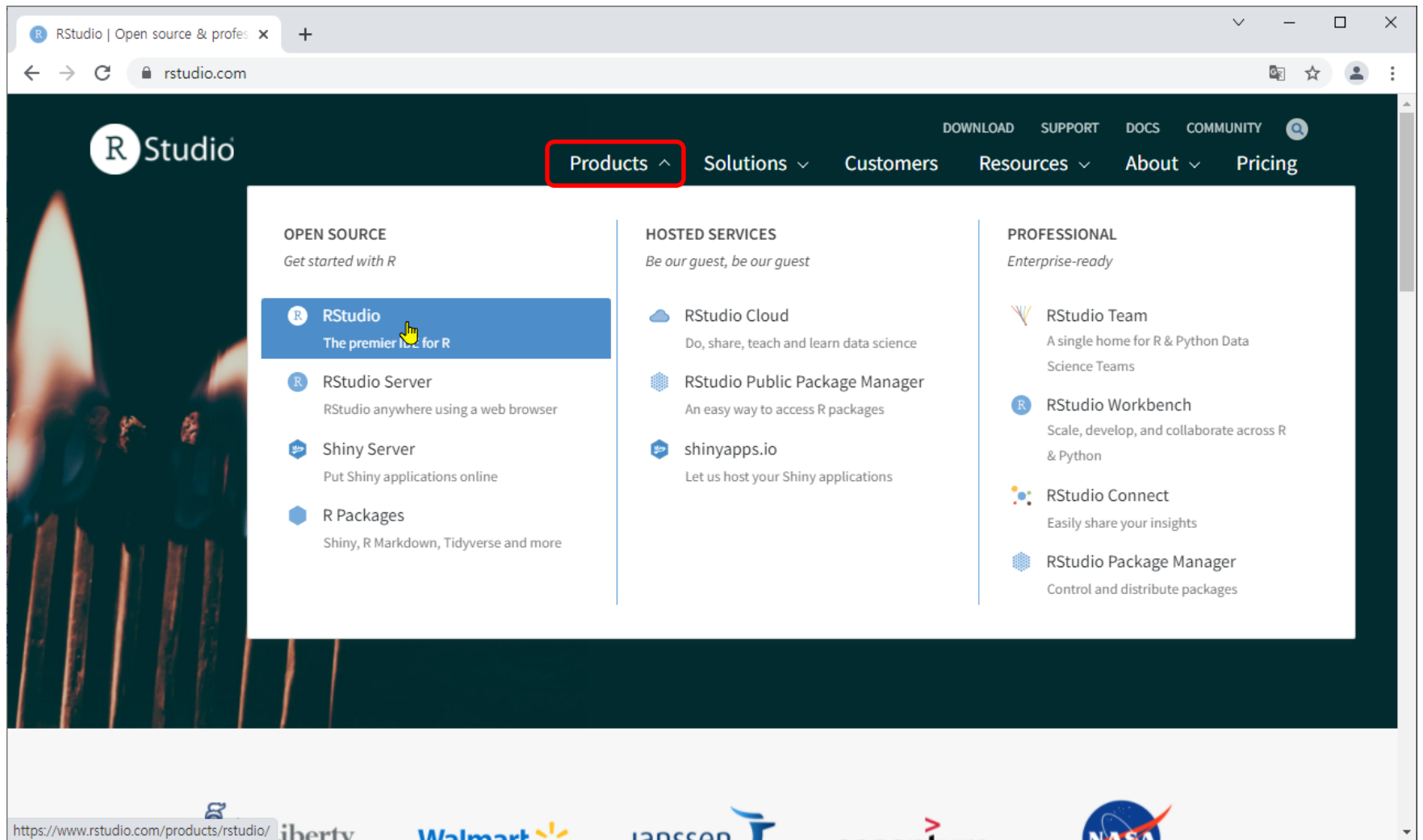
R은 많은 기여자들이 참여하는 공동프로젝트입니다.
'contributors()'라고 입력하시면 이에 대한 더 많은 정보를 확인하실 수 있습니다.
그리고, R 또는 R 패키지들을 출판물에 인용하는 방법에 대해서는 'citation()'을 통해 확인하시길 부탁드립니다.

'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.

> v <- c(11,12,13,14,15)
> w <- c(1,2,3,4,5)
> v+w
[1] 12 14 16 18 20
> v*w
[1] 11 24 39 56 75
> v/w
[1] 11.000000  6.000000  4.333333  3.500000  3.000000
> v^w
[1]      11      144     2197    38416   759375
> |
```

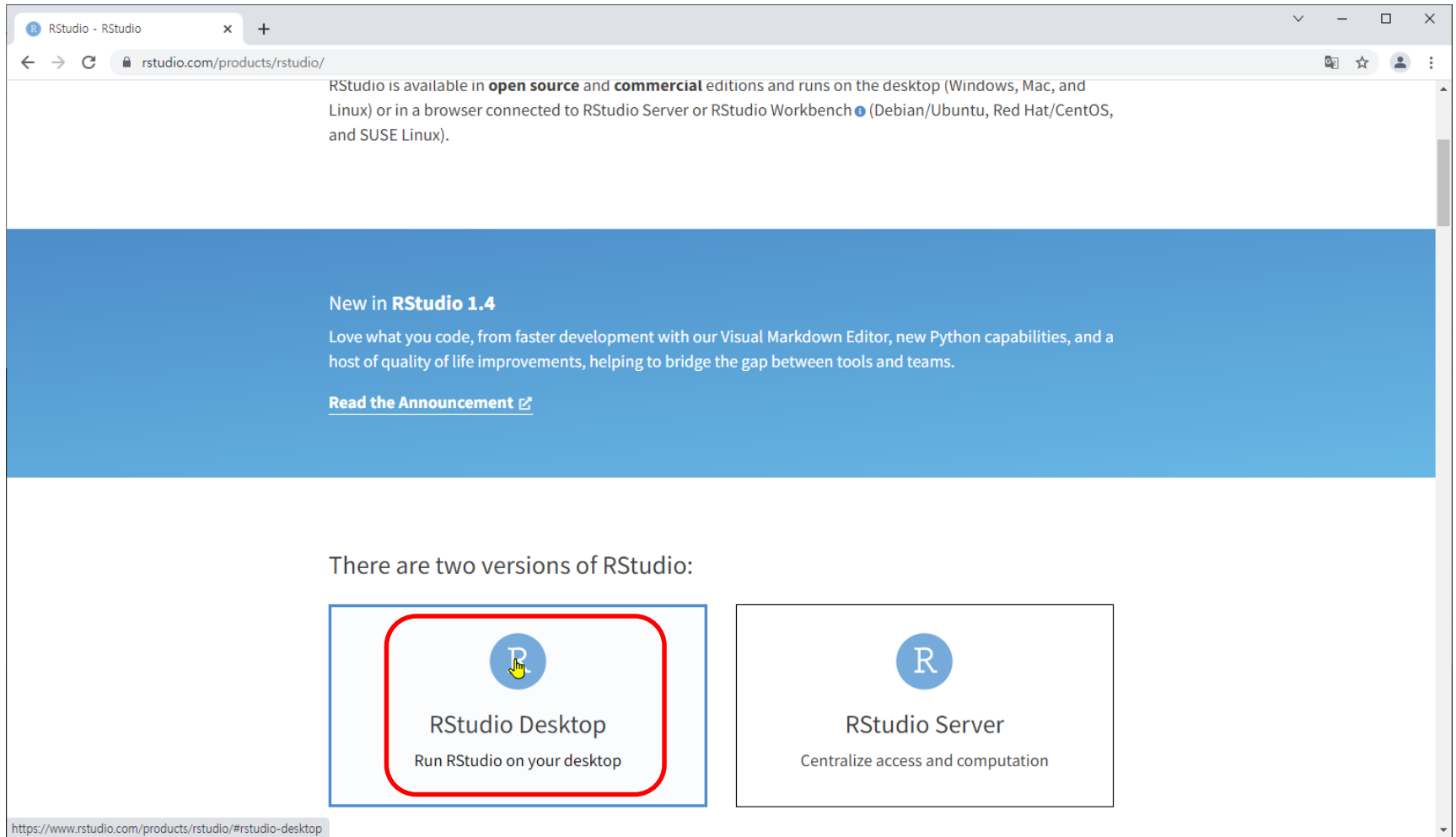
# R Studio 설치하기

➤ <https://www.rstudio.com/>



# R Studio 설치하기

## ➤ Desktop 선택



# R Studio 설치하기

## ➤ Desktop 선택

The screenshot shows the RStudio Desktop product page. The browser address bar displays 'rstudio.com/products/rstudio/'. The page features a comparison between the 'Open Source Edition' and 'RStudio Desktop Pro'. The 'Open Source Edition' is free and includes features like local access, syntax highlighting, and integrated R help. 'RStudio Desktop Pro' is a commercial license for \$995/year, offering priority support and remote access. A red box highlights the 'DOWNLOAD RSTUDIO DESKTOP' button, which is the recommended choice for installation.

	Open Source Edition	RStudio Desktop Pro
<b>Overview</b>	<ul style="list-style-type: none"><li>• Access RStudio locally</li><li>• Syntax highlighting, code completion, and smart indentation</li><li>• Execute R code directly from the source editor</li><li>• Quickly jump to function definitions</li><li>• View content changes in real-time with the Visual Markdown Editor</li><li>• Easily manage multiple working directories using projects</li><li>• Integrated R help and documentation</li><li>• Interactive debugger to diagnose and fix errors</li><li>• Extensive package development tools</li></ul>	<p>All of the features of open source; plus:</p> <ul style="list-style-type: none"><li>• A commercial license for organizations not able to use AGPL software</li><li>• Access to priority support</li><li>• <a href="#">RStudio Professional Drivers</a></li><li>• Connect directly to your RStudio Workbench instance remotely</li></ul>
<b>Support</b>	Community forums only	<ul style="list-style-type: none"><li>• Priority Email Support</li><li>• 8 hour response during business hours (ET)</li></ul>
<b>License</b>	AGPL v3	<a href="#">RStudio License Agreement</a>
<b>Pricing</b>	Free	\$995/year
<b>Download</b>	<a href="#">DOWNLOAD RSTUDIO DESKTOP</a>	<a href="#">DOWNLOAD FREE RSTUDIO DESKTOP PRO TRIAL</a>

<https://www.rstudio.com/products/rstudio/download/>

# R Studio 설치하기

## ➤ Desktop 선택

The screenshot shows the RStudio download page. The 'RStudio Desktop' option is highlighted with a red box around its 'DOWNLOAD' button. Below the product names is a table comparing features across the four options.

	RStudio Desktop	RStudio Desktop Pro	RStudio Server	RStudio Workbench
License	Open Source License	Commercial License	Open Source License	Commercial License
Price	Free	\$995 /year	Free	\$4,975 /year (5 Named Users)
Action	<b>DOWNLOAD</b> (highlighted)	BUY	DOWNLOAD	BUY
Learn more	Learn more	Learn more	Learn more	Evaluation   Learn more
Integrated Tools for R	✓	✓	✓	✓
Priority Support		✓		✓
Access via Web Browser			✓	✓
RStudio Professional Drivers		✓		✓
Connect to RStudio		✓		


# R Studio 설치하기

Download the RStudio IDE - RStudio


rstudio.com/products/rstudio/download/#download

## RStudio Desktop 2021.09.1+372 - Release Notes

1. Install R. RStudio requires R 3.0.1+.
2. Download RStudio Desktop. Recommended for your system:

**DOWNLOAD RSTUDIO FOR WINDOWS**  
2021.09.1+372 | 156.89MB




Requires Windows 10 (64-bit)



## All Installers

Linux users may need to [import RStudio's public code-signing key](#) prior to installation, depending on the operating system's security policy.

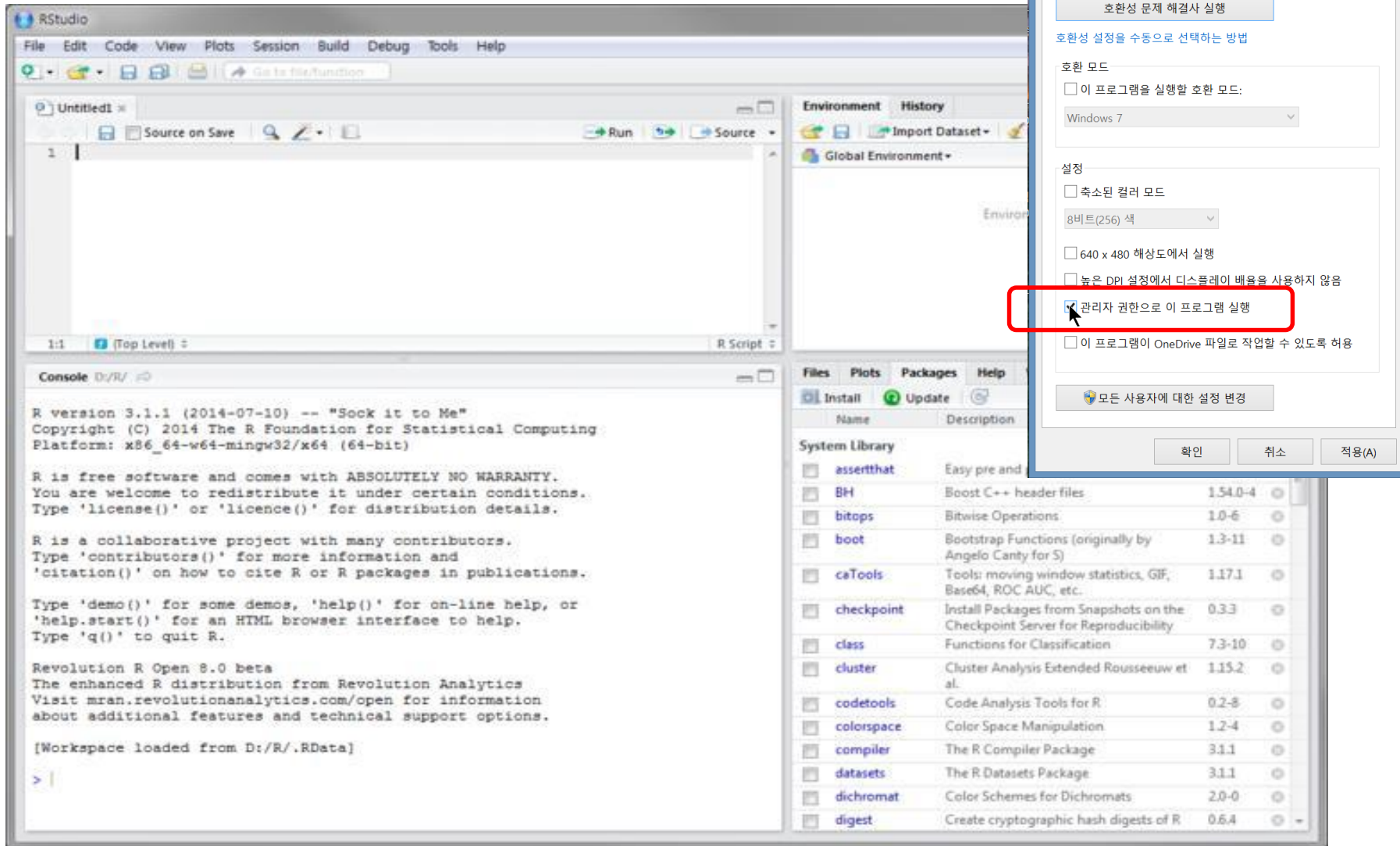
RStudio requires a 64-bit operating system. If you are on a 32 bit system, you can use an [older version of RStudio](#).

OS	Download	Size	SHA-256
Windows 10	 <a href="#">RStudio-2021.09.1-372.exe</a>	156.89 MB	1c3d27f5
macOS 10.14+	 <a href="#">RStudio-2021.09.1-372.dmg</a>	203.00 MB	daec6a40
Ubuntu 18/Debian 10	 <a href="#">rstudio-2021.09.1-372-amd64.deb</a>	117.89 MB	921b4f23

<https://download1.rstudio.org/desktop/windows/RStudio-2021.09.1-372.exe>

# R Studio 설정하기

## ➤ R Studio 실행 화면





# R Studio 설정하기

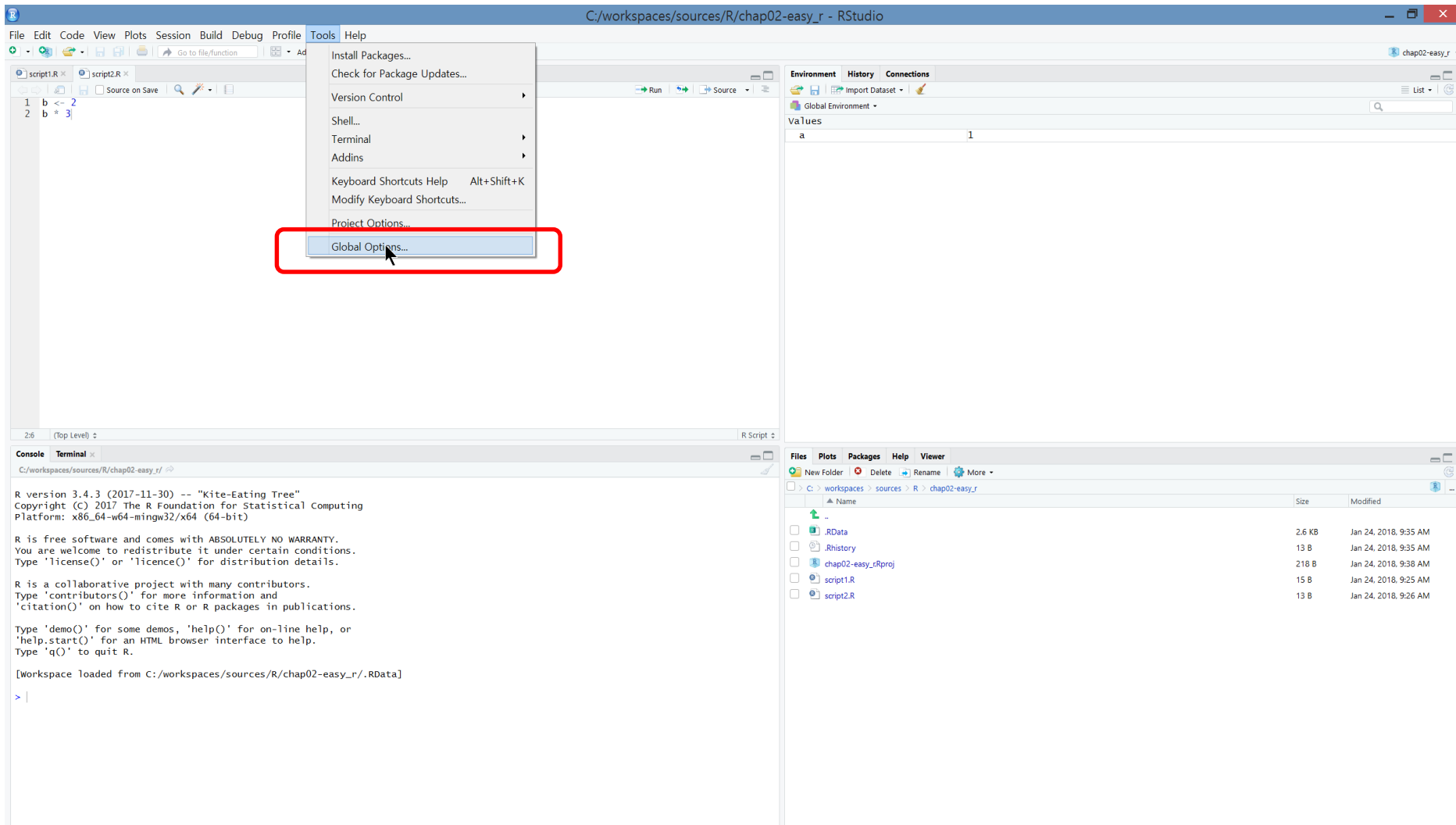
## ➤ R Studio 4가지 패널 창 기능

The screenshot shows the RStudio interface with four panels highlighted by blue boxes and labels:

- R Script**: Points to the top-left menu bar where the 'R Script' option is highlighted.
- 코딩 영역** (Coding Area): Points to the main editor window containing R code. The code includes comments in Korean and R commands like `order <- c("third", "second")`, `plot(order)`, `fa <- as.factor(order)`, and `plot(fa)`.
- 작업내역/환경** (Environment/History): Points to the top-right panel showing the 'Global Environment' with a list of objects (fa, fr, ga, hflig..., hong, i, iris, kang, kor, kor.m..., lat, lee) and their attributes.
- 콘솔(실행결과)** (Console/Execution Results): Points to the bottom-left panel showing the output of the R commands, including the message 'Type \'license()\' or \'licence()\' for distribution details.'
- 탐색기/그래프/도움말/패키지** (Browser/Plots/Help/Viewer): Points to the bottom-right panel showing a bar chart with three bars labeled 'first', 'second', and 'third'.

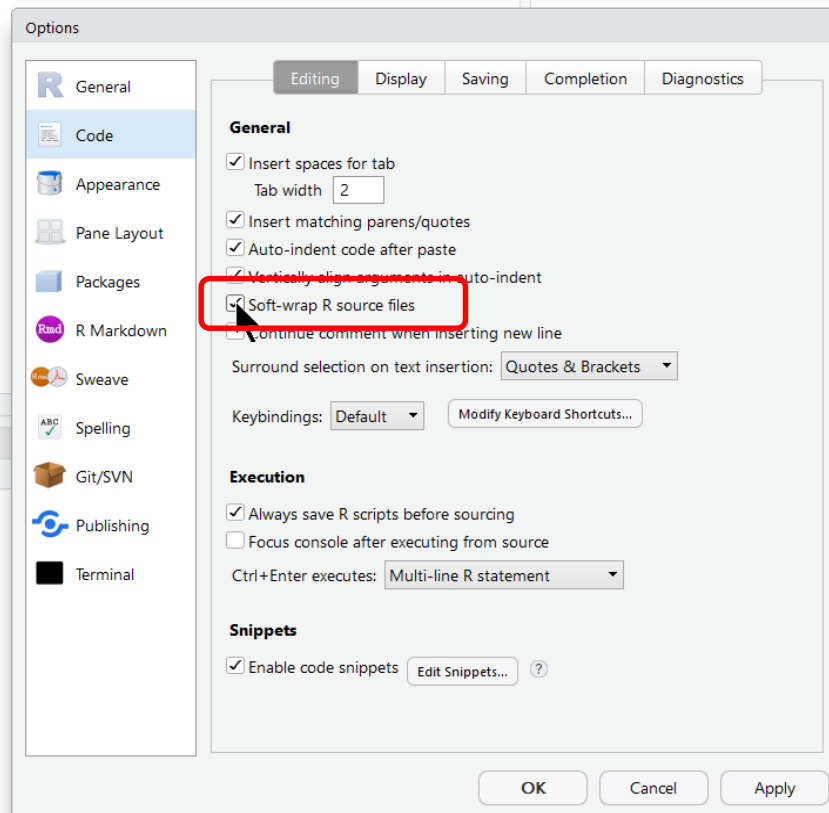
# R Studio 설정하기

## ➤ R 환경 설정



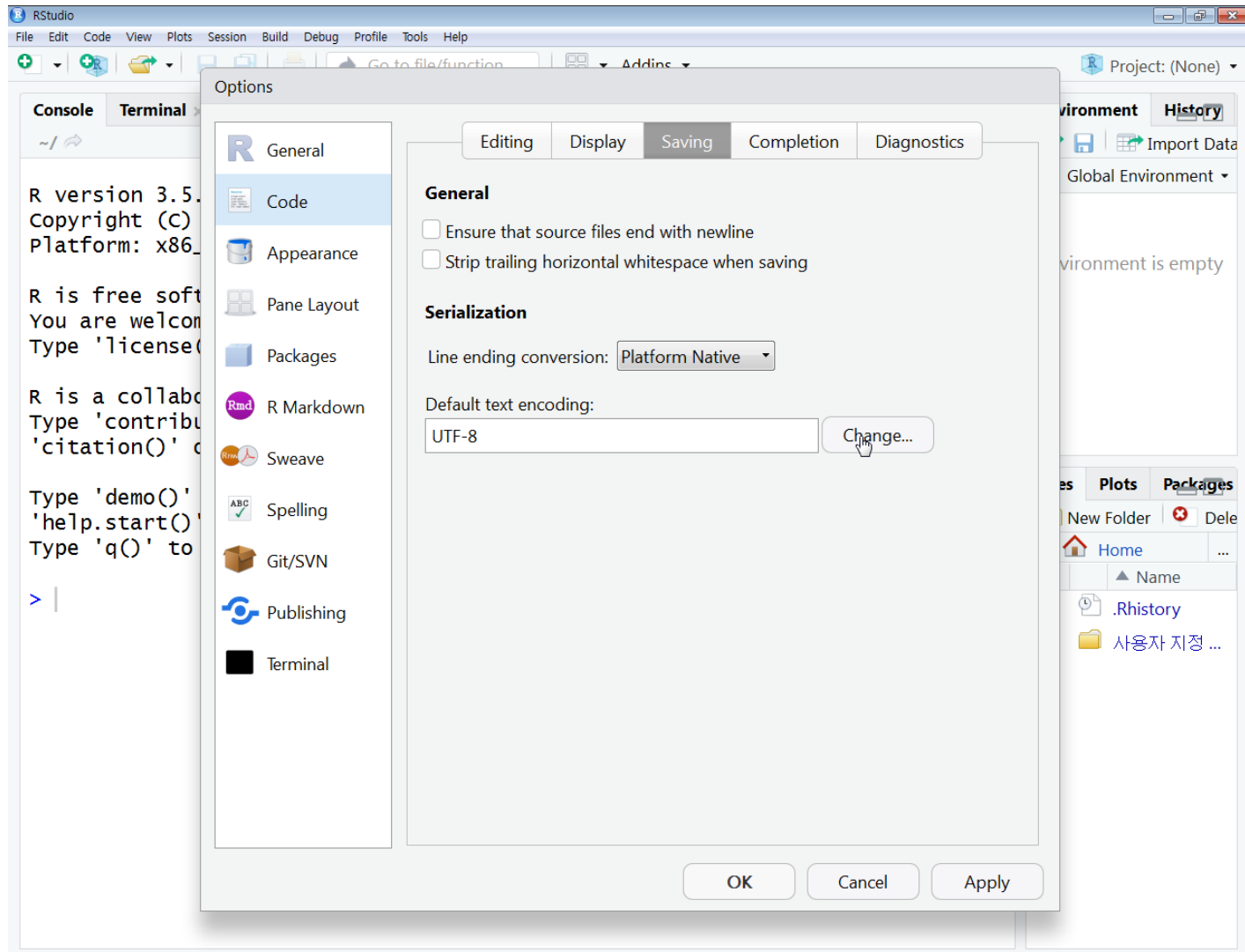
# R Studio 설정하기

## ➤ R 환경 설정 – 자동 줄바꿈 옵션 설정



# R Studio 설정하기

## ➤ R 환경 설정 – 인코딩 방식 설정



R 개 요

# 명령어 실행 방법

---

## ➤ Interaction 방식

- 줄 단위 실행
- 줄 선택 + Ctrl + R 또는 Ctrl + Enter

## ➤ CRAN Site에서 제공하는 패키지 보기.

```
> dim(available.packages()) # 17517  17
```

```
> available.packages()      # 패키지 상세보기
```

# R Session

---

## ➤ R session 정보

- Session : 사용자가 R프로그램을 기동한 이후 R 콘솔 시작 ~ 종료 전까지의 기간에 수행된 정보를 의미.

> sessionInfo() # 세션 정보 보기

R version 4.0.5 (2021-03-31)

Platform: x86\_64-w64-mingw32/x64 (64-bit)

Running under: Windows 10 x64 (build 19041)

Matrix products: default

locale:

[1] LC\_COLLATE=Korean\_Korea.949 LC\_CTYPE=Korean\_Korea.949 LC\_MONETARY=Korean\_Korea.949

[4] LC\_NUMERIC=C LC\_TIME=Korean\_Korea.949

attached base packages:

[1] stats graphics grDevices utils datasets methods base

loaded via a namespace (and not attached):

[1] compiler\_4.0.5 tools\_4.0.5

# 패키지 사용

---

- R에서 제공하는 패키지(Package)는 처리할 자료(data)와 기능(function) 그리고 알고리즘(Algorithm)등이 하나의 꾸러미 형태로 제공.
- 패키지는 CRAN Site에서 다운로드하여 사용자의 컴퓨터에 설치할 수 있음.
- R 프로그래밍의 실력은 얼마나 많은 패키지를 효과적으로 데이터 분석에 적용할 수 있는가 하는 능력에 따라 좌우된다고 할 수 있음.

- > `install.packages("stringr")` # stringr 패키지 설치

- > `installed.packages()` # 설치된 패키지 확인

- > `library(stringr)` # 패키지 메모리에 로드(load)

- > `search()` # 현재 로드된 패키지 확인

- > `remove.package("stringr")` # 패키지 제거

- > `data()` # R에서 제공하는 실습용 데이터 셋 보기



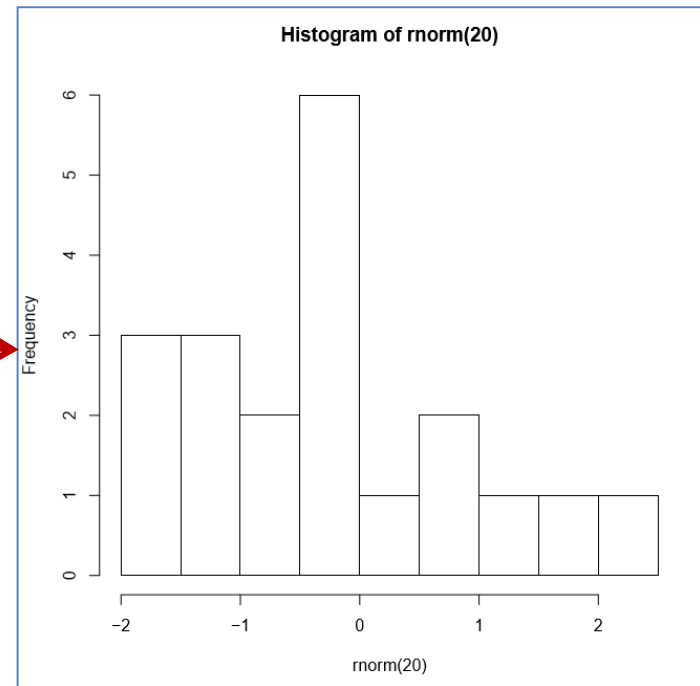
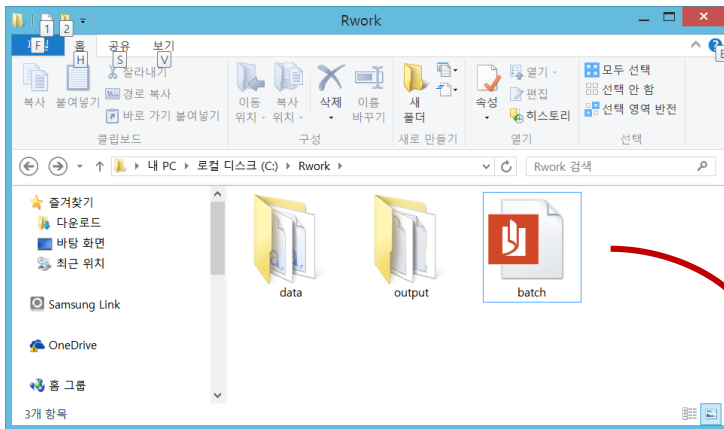
# 결과물 저장

---

## ➤ Batch 방식

- 여러 줄 소스코드 일괄처리
- Block 지정 + Ctrl + R
- Batch 방식 히스토그램 그리기 예
  - ✓ `pdf("c:/workspaces/Rwork/batch.pdf")` # 지정된 경로의 파일에 결과 출력
  - ✓ `hist(rnorm(20))` # 난수에 대한 히스토그램 그리기
  - ✓ `dev.off()` # 출력할 파일 닫음

# Batch 방식 히스토그램 결과



# 변수(Variable)

---

## ➤ 변수

- 값을 저장하는 메모리 이름
- R은 모두 변수가 객체(데이터+함수+차트 등)

## ➤ 변수명 작성 규칙

- 영문자 시작
- 두번째 단어 : 숫자, 언더바(\_), 점(member.id) 사용 가능
- 대소문자 구분
- 의미를 파악할 수 있는 이름으로 지정
- 정의된 변수는 재사용이 가능

# 데이터 유형(Data Type)

---

- R은 변수 선언 시 별도의 Type을 선언하지 않음

유형(Type)	값(value)	예
숫자형(Numeric)	정수, 실수	10, 25, 55.23
문자형(Character)	문자, 문자열	"홍길동"
논리형(Logical)	참, 거짓	TRUE 또는 T, FALSE 또는 F
결측 데이터	결측치, 비슷자	NA(Not Available), NaN(Not a Number)

# 데이터 유형(Data Type) 보기 함수

---

➤ mode()함수 : 데이터 유형 보기 함수

is.numeric(x)	수치형 여부	is.integer(x)	정수형 여부
is.logical(x)	논리형 여부	is.double(x)	실수형 여부
is.character(x)	문자형 여부	is.complex(x)	복소수형 여부
is.data.frame(x)	데이터프레임 여부	is.factor(x)	범주형 여부
is.na(x)	NA 여부	is.nan(x)	NaN 여부

# 데이터 타입 보기 함수

---

## ➤ 데이터 타입 보기 함수

- `is.character(string)`                      # TRUE
- `x <- is.numeric(int)`
- `x`    # TRUE
- `is.logical(boolean)`                      # TRUE
- `is.logical(x)`                              # TRUE -> TRUE
- `is.na(x)`                                      # NA 확인

## ➤ 현재 사용중인 변수객체 보기

- `ls()`

# 자료형과 자료구조

---

## ➤ 자료형과 자료구조(객체 타입) 보기

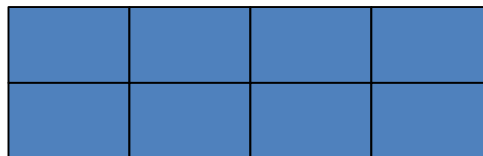
- `mode(변수)` : 자료형(Data Type) 보기
- `class(변수)` : 자료구조(객체) 보기

## ➤ `mode()` vs `class()`

- ex) 변수 <- 데이터
- `mode` : 데이터 성격(숫자형, 문자형, 논리형)
- `class` : 데이터 자료구조 성격(Array, List, Table 등)



Array 자료구조



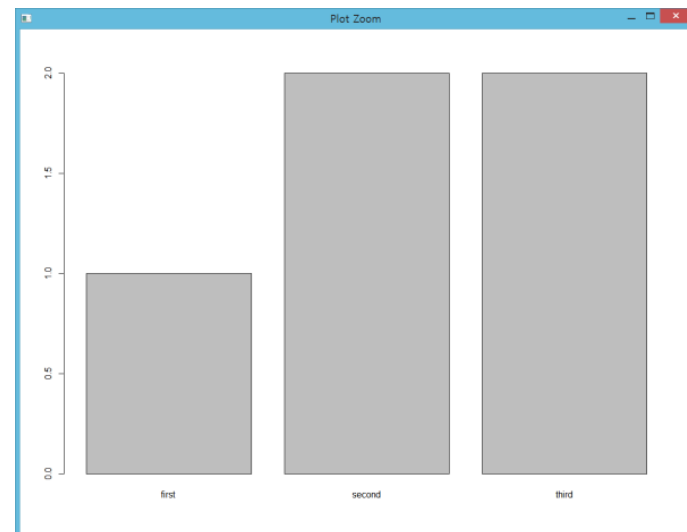
Matrix 자료구조

# 자료형과 자료구조

- Factor : 같은 성격인 값의 목록을 범주(category)로 갖는 벡터 자료.
  - ✓ 범주 : 변수가 가질 수 있는 값의 범위(ex: 성별 변수의 범주는 남자와 여자)
  - Nominal : level 순서의 값 무의미-알파벳 순서 정렬
  - Ordinal : level 순서의 값 의미 - 사용자가 직접 순서 지정

## # Factor Nominal 예

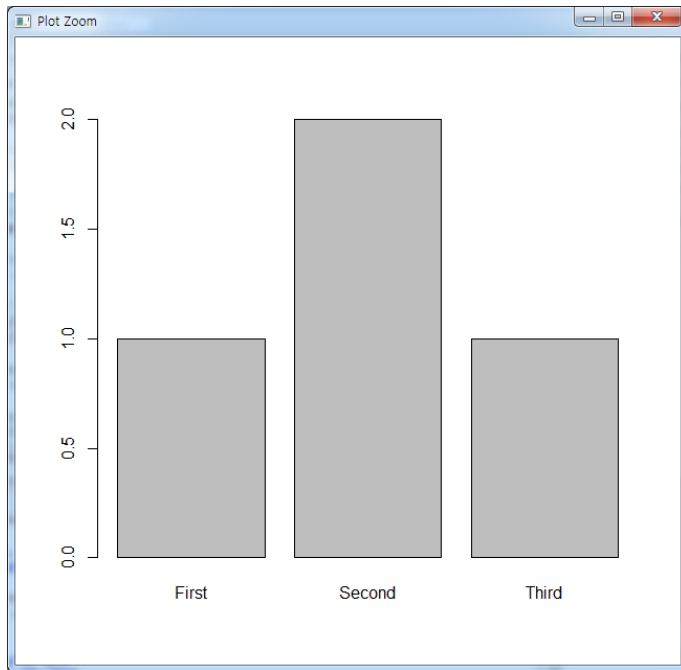
- `order <- c("third", "second", "first", "third", "second")`
- `plot(order)` # error(차트는 수치 데이터만 가능)
- `fa <- as.factor(order)` # factor형으로 변환
- `plot(fa)` # 막대차트 - 동일 문자열 수 카운터



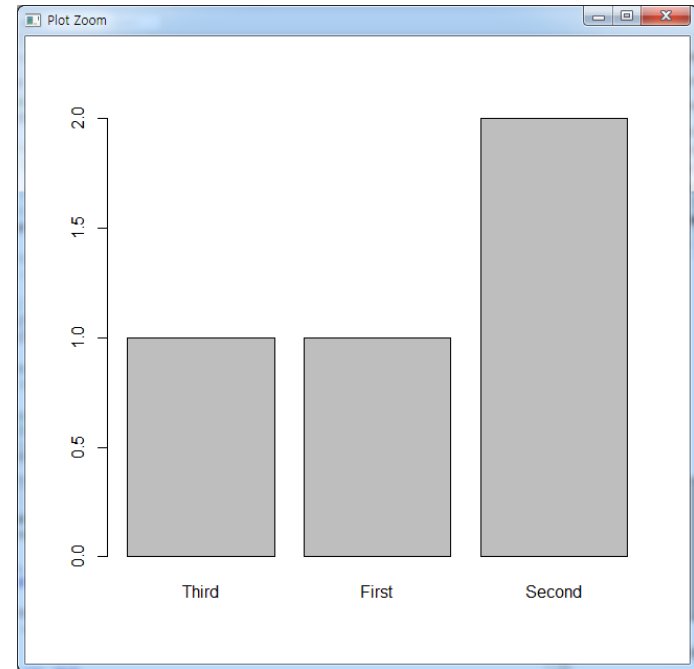


# 자료형과 자료구조

plot(level) # Nominal인 경우



plot(level2) # Ordinal인 경우



# 데이터 형 변환 함수

---

<code>as.numeric(x)</code>	수치형 변환	<code>as.integer(x)</code>	정수형 변환
<code>as.logical(x)</code>	논리형 변환	<code>as.double(x)</code>	실수형 변환
<code>as.character(x)</code>	문자형 변환	<code>as.complex(x)</code>	복소수형 변환
<code>as.data.frame(x)</code>	데이터프레임 변환	<code>as.factor(x)</code>	범주형 변환
<code>as.list(x)</code>	리스트형 변환	<code>as.vector(x)</code>	벡터형 변환

# 날짜와 시간 표현을 위한 제어문자

---

날짜	제어문자	시간	제어문자
년도 4자리	%Y	24 시간	%H
년도 2자리	%y	12 시간	%I
월	%m	분	%M
일	%d	초	%S

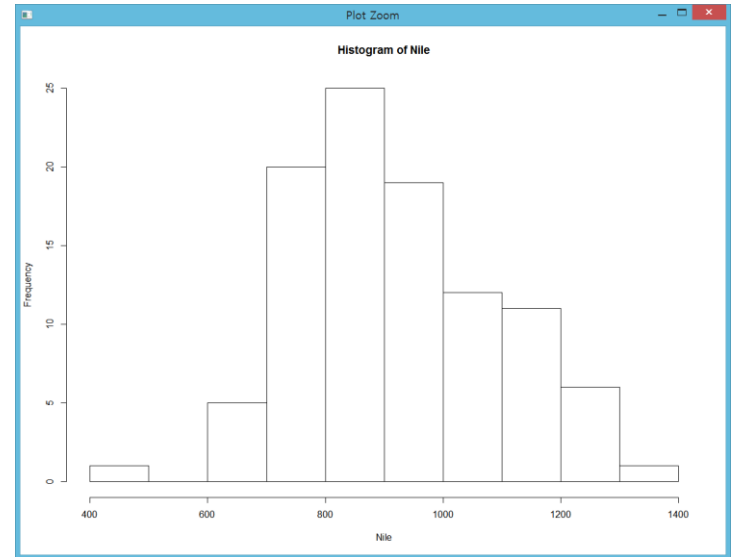
# 기본 함수

## ➤ 기본 함수 사용

- 함수 도움말 : `help(함수명)`, `?함수명`
- `help(mean)` # 브라우저 화면으로 제공
- `?sum`
- google에서 검색 : 함수() in r

## ➤ 데이터 세트와 리스트 보기

- `data()`
- `hist(Nile)` # 나일강 흐름에 대한 데이터 셋 히스토그램 그리기



# 기본 함수

data()

```
Data sets in package ;@arules;~:

Adult                Adult Data Set
AdultUCI              Adult Data Set
Epub                  Epub Data Set
Groceries             Groceries Data Set
Income                Income Data Set
IncomeESL             Income Data Set

Data sets in package ;@datasets;~:

AirPassengers         Monthly Airline Passenger Numbers 1949-1960
BJsales               Sales Data with Leading Indicator
BJsales.lead (BJsales) Sales Data with Leading Indicator
BOD                   Biochemical Oxygen Demand
CO2                   Carbon Dioxide Uptake in Grass Plants
ChickWeight           Weight versus age of chicks on different
                      diets
DNase                 Elisa assay of DNase
EuStockMarkets        Daily Closing Prices of Major European Stock
                      Indices, 1991-1998
Formaldehyde          Determination of Formaldehyde
HairEyeColor          Hair and Eye Color of Statistics Students
Harman23.cor           Harman Example 2.3
Harman74.cor           Harman Example 7.4
Indometh              Pharmacokinetics of Indomethacin
InsectSprays          Effectiveness of Insect Sprays
JohnsonJohnson       Quarterly Earnings per Johnson & Johnson
                      Share
LakeHuron             Level of Lake Huron 1875-1972
LifeCycleSavings      Intercountry Life-Cycle Savings Data
Loblolly              Growth of Loblolly pine trees
Nile                  Flow of the River Nile
Orange                Growth of Orange Trees
OrchardSprays         Potency of Orchard Sprays
PlantGrowth           Results from an Experiment on Plant Growth
Puromycin             Reaction Velocity of an Enzymatic Reaction
Seatbelts             Road Casualties in Great Britain 1969-84
Theoph                Pharmacokinetics of Theophylline
Titanic               Survival of passengers on the Titanic
ToothGrowth           The Effect of Vitamin C on Tooth Growth in
                      Guinea Pigs
UCBAdmissions         Student Admissions at UC Berkeley
UKDriverDeaths        Road Casualties in Great Britain 1969-84
```

R 제공  
데이터 셋

# 기본 함수

---

## ➤ 함수 파라미터 보기

- 형식) `args(함수명)`
- `args(max)` #
- # `max()` 함수 파라미터 : `(na.rm = FALSE)` <- NA 제거 안됨
- `max(10, 20, NA, 30)` # NA 출력

## ➤ 함수 예제 보기

- 형식) `example(함수명)` : 해당 함수의 사용 예 제공
- `example(seq)`
- `seq(10,20, by=2)` # 10 12 14 16 18 20
- `example(mean)`
- `mean(10:20)` # 15

# 작업공간

---

## ➤ 작업공간 보기

- `getwd()` # 기본함수

## ➤ 작업공간 지정

- `setwd("C:/workspaces/Rwork/Data")` # 구분자 : / or \
- `data <- read.csv("test.csv", header=T)` # 지정된 경로에서 파일 로딩
- `args(read.csv)` # (file, header = TRUE)
- `data`

```
> data
  A B C D E
1 2 4 4 2 2
2 1 2 2 2 2
3 2 3 4 3 3
4 3 5 5 3 3
5 3 2 4 4 4
6 4 3 3 4 2
```



test.csv  
파일 내용

# 작업공간

- 스크립트 파일 저장 및 깨짐 현상 처리 방법
  - 문자 인코딩 방식 지정 : UTF-8
  - 메뉴 : Tools > General Options > Code(Saving 탭) > Default text encoding-UTF-8

