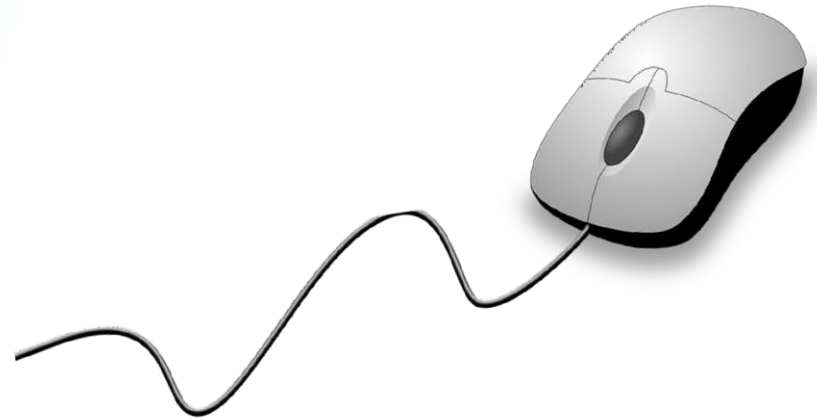


공개SW 솔루션 설치 & 활용 가이드

시스템SW > SW공학도구



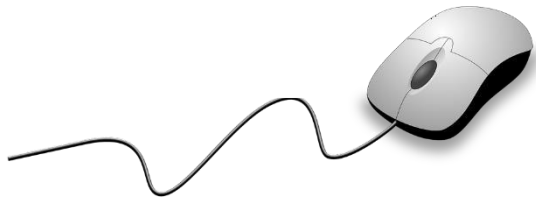
제대로 배워보자

How to Use Open Source Software

Open Source Software Installation & Application Guide



오픈소스 소프트웨어 통합지원센터
Open Source Software Support Center



CONTENTS

1. 개요
2. 기능요약
3. 실행환경
4. 설치 및 실행
5. 기능소개
6. 활용예제
7. FAQ
8. 용어정리

1. 개요



소개	<ul style="list-style-type: none"> • 통계 계산 및 그래픽 시각화, 데이터 마이닝에 많이 사용되는 프로그래밍 언어 • 통계 소프트웨어 개발과 자료 분석에 많이 사용되는 오픈소스 기반의 언어 • CRAN을 통해서 다양한 패키지들을 사용 가능 		
주요기능	<ul style="list-style-type: none"> • 각종 통계 알고리즘 제공 • 머신러닝 및 데이터 분석에 사용하기 용이 • 필요한 기능은 CRAN에서 내려받아 사용 		
대분류	• 시스템 SW	소분류	• SW공학도구
라이선스 형태	• GNU General Public License v2	사전설치 솔루션	• N/A
운영체제	<ul style="list-style-type: none"> • Windows • macOS • Linux 	버전	• 3.5.1(2018년 10월 기준)
특징	<ul style="list-style-type: none"> • 프로그래밍이 쉽고 많은 기능들이 CRAN에 구현 • 각종 기계학습 알고리즘 사용 가능 • 통계 및 시각화에 강한 장점 • SPSS, SAS, STATA 등 통계소프트웨어를 대체 가능 		
보안취약점	<ul style="list-style-type: none"> • 취약점 ID : CVE-2016-8714 • 심각도 : 7.5 HIGH(V3) • 취약점 설명 : R 프로그래밍 언어 버전 3.3.0의 LoadEncoding 기능에 악용 가능한 버퍼 오버플로우 취약점이 존재 • 대응방안 : 3.3.3-1 이상 업그레이드 • 참고 경로 : https://lists.debian.org/debian-security-announce/2017/msg00068.html 		
개발회사/커뮤니티	• R Core Team		
공식 홈페이지	• https://www.r-project.org/		



2. 기능요약



- R은 데이터 분석을 위한 통계 계산 및 그래픽을 위한 프로그래밍 언어
- 벨 연구소에서 만들어진 통계분석 언어인 S를 기반으로 뉴질랜드 오클랜드 대학의 로버트 젠틀맨과 로스 이하카가 개발
- 현재 데이터 분석을 위한 도구로 많은 인기를 얻고 있음

Language Rank	Types	Spectrum Ranking
1. Python	🌐 🖥️	100.0
2. C	📱 🖥️ 🖨️	100.0
3. Java	🌐 📱 🖥️	99.4
4. C++	📱 🖥️ 🖨️	96.9
5. C#	🌐 📱 🖥️	88.6
6. R	🖥️	88.1
7. JavaScript	🌐 📱	85.3
8. PHP	🌐	81.1
9. Go	🌐 🖥️	75.7
10. Swift	📱 🖥️	74.3

IEEE SPECTRUM의 2017년 The Top Programming Languages

- R언어는 오픈소스이며 데이터 분석을 위한 데이터 마이닝이나 기계학습, 통계 분석, 시각 화등을 위한 패키지들을 CRAN을 통해 무료로 공개



3. 실행환경



- 지원 OS
 - Windows 32bit / 64bit
 - MacOS
 - Linux (RedHat, CentOS, Debian, Ubuntu, SuSE 등)



4. 설치 및 실행



세부 목차

1. 설치 파일 다운로드
2. 설치 진행 – Windows 및 Linux
3. 설치 완료 – Windows 및 Linux
4. R Studio 다운로드
5. R Studio 설치 및 완료



4. 설치 및 실행



4.1 설치 파일 준비

- R을 사용하기 위해서는 아래 사이트에서 R을 다운로드
- <https://cloud.r-project.org/>

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 \(Mac\) 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 (2018-04-23, Joy in Playing) [R-3.5.0.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.

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 \geq 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.

R-3.5.0 for Windows (32/64 bit)

[Download R 3.5.0 for Windows](#) (62 megabytes, 32/64 bit)
[Installation and other instructions](#)
[New features in this version](#)

현재 사용중인 운영체제가 윈도우라면 R for Windows를 클릭



4. 설치 및 실행



4.2 설치 진행 - Windows(1/2)

- 1 단계 - 언어 선택

한글로 구성된 메뉴를 원하면 아래 선택에서 한글을 선택하고 확인 클릭

- 2 단계 - 라이선스 선택

라이선스를 확인하고 다음 버튼을 클릭하여 진행

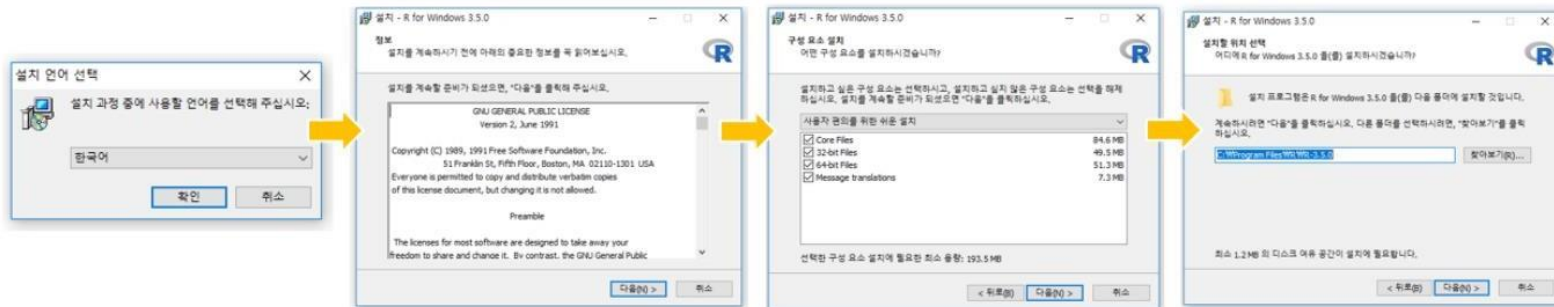
- 3 단계 - 설치할 위치 선택

설치될 경로를 설정하며, 설치 경로를 기본값으로 사용할 경우 다음 버튼을 눌러 진행

* 설치될 경로에 한글이 있으면 안됨

- 4 단계 - 구성요소 선택

설치할 컴퓨터의 OS의 버전(32비트 / 64비트)에 따라 선택하여 설치할 수 있으나, 전체 선택을 하여도 크게 사용하는데 지장은 없으며, 전체 선택을 하여 설치 진행



4. 설치 및 실행



4.2 설치 진행 - Windows(2/2)

- 5 단계 - 스타트업 옵션 설정

사용자 환경을 변경하고 싶다면 Yes를 선택하고 진행하지만, 기본값으로 사용할 경우 No를 선택하고 진행 후 Yes를 선택하면 디스플레이 모드를 선택 가능

* 기본값 No 권장

디스플레이 모드

MDI(multiple-document interface)와 SDI(single-document interface) 선택

MDI를 선택하면 R 코드로 그래프를 생성하면 내부 콘솔에 그래프 생성 S

DI는 선택하면 R 코드로 그래프를 출력하면 다른 작업 창 생성

- 6 단계 - 시작 메뉴 폴더 선택 및 추가 사항 적용 후 기본값을 사용하여 설치 완료



4. 설치 및 실행



4.2 설치 진행 및 확인 -Linux(4/4)

- R을 리눅스 계열인 우분투에서 설치할 경우 터미널에서 명령문으로 설치 가능하며, 터미널을 실행한 후에 아래 명령어 실행

```
sudo apt-get update
sudo apt-get install r-base
sudo apt-get install r-base-dev
```

- 설치 후 터미널에서 R이라는 명령어를 실행하면 R 실행

```
gen2@DESKTOP-554ALNO:~$ R

R version 3.2.3 (2015-12-10) -- "Wooden Christmas-Tree"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-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.
```



4. 설치 및 실행



4.3 설치 완료 - Windows

- 설치 완료 확인하기

R이 정상적으로 설치가 되었다면 시작 메뉴에 R 메뉴 생성

R x64 3.5.0를 선택하여 R 실행

- * x64는 운영체제의 버전이 64비트 의미
- * i386은 운영체제의 버전이 32비트 의미
- * 32비트 운영체제에는 64비트 R이 설치되지 않음

정상적으로 설치가 완료되었다면 아래 화면 출력

```
R GUI (64-bit)
파일 편집 보기 기타 패키지를 윈도우즈 도움말

R Console

R version 3.5.0 (2018-04-23) -- "Joy in Playing"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

같은 이름으로 존재하며, 어떤 형태로든 사용될 수 없습니다.
또한, 동일한 조건에서 이걸 가져오지 못합니다.
이러한 권한은 사용자 'license()' 또는 'licence()'를 통해 확인될 수 있습니다.

같은 이름으로 존재하는 공유 라이브러리입니다.
'contributors()'도 알려주지만 어떤 어떤 이 같은 정보를 확인하실 수 있습니다.
그리고, 이 또는 이 패키지를 출판할 때 사용하는 방법에 대해서는 'citation()'을 보십시오.

'demo()'를 실행하려면 몇 가지 파일을 보십시오. 'help()'를 실행하십시오.
또한, 'help.start()'의 URL을 통하여 HTML 웹-페이지의 도움말을 보십시오.
이의 도움을 받으려면 'q()'를 실행하십시오.

> |
```



4. 설치 및 실행



4.4 R Studio 다운로드(1/2)

- R Console이라는 부분에서 R Code를 작성하여 개발 및 분석 진행 가능
- R Console만 사용하여 분석을 진행 하다 보면 많은 불편한 점 발생
- 이 때 R을 좀 더 편하게 사용하기 위해서 통합개발환경 프로그램 설치함
- 통합개발환경(IDE)는 사용자가 개발 및 분석을 편하게 진행할 수 있도록 도와 주는 프로그램
- R의 대표적인 IDE 프로그램으로는 R-Studio 사용
- 하지만 R의 IDE의 프로그램이 R-Studio만 존재하는 것은 아니며, 다수의 프로그램이 존재 하지만 R-Studio를 가장 많이 사용
- Tinn-R이나 MS에서 제공하는 Visual Studio에서도 R 지원



4. 설치 및 실행



4.4 R Studio 다운로드(2/2)

- R만 설치하여 콘솔 창에서 작업을 진행하다 보면 사용하는데 많은 불편함 발생
- 편리하게 사용할 수 있게 도와주는 IDE(Integrated Development Environment, 통합 개발 환경) 프로그램인 R-Studio 설치
- 아래 url에 접속하여 R-Studio 다운로드
- <https://www.rstudio.com/>

Download 링크를 클릭합니다. 클릭하면 RStudio를 다운로드 받을 수 있는 페이지로 이동합니다.

RStudio
RStudio makes R easier to use. It includes a code editor, debugging & visualization tools.
[Download](#) [Learn More](#)

Shiny
Shiny helps you make interactive web applications for visualizing data. Bring R data analysis to life.
[Learn More](#)

R Packages
Our developers create popular packages to expand the features of R. Includes ggplot2, dplyr, R Markdown & more.
[Learn More](#)



4. 설치 및 실행



4.5 R Studio 설치 및 완료(1/6)

- RStudio Desktop Open Source License 버전을 클릭하여 다운로드
- Desktop 버전은 AGPL v3 라이선스
- 상업용 버전은 비용을 지불하고 사용

	RStudio Desktop Open Source License	RStudio Desktop Commercial License	RStudio Server Open Source License	RStudio Server Pro Commercial License	RStudio Server Pro + RStudio Connect Commercial License
	FREE	\$995 per year	FREE	\$9,995 per year	\$29,995 per year
	DOWNLOAD Learn More	BUY Learn More	DOWNLOAD Learn More	DOWNLOAD Learn More	TALK Learn More
Integrated Tools for R	●	●	●	●	●
Priority Support		●		●	●
Access via Web Browser			●	●	●
Enterprise Security				●	●
Project Sharing				●	●
Manage Multiple R Sessions & Versions				●	●
Admin Dashboard				●	●
Load Balancing				●	●
One-Click Publishing					●
Managed R					●
Scheduled Reports					●
License	AGPL	Commercial	AGPL	Commercial	Commercial
Pricing	FREE	\$995/yr	FREE	\$9,995/yr	\$29,995/yr
	RStudio Desktop Open Source DOWNLOAD NOW Learn More	RStudio Desktop Commercial BUY NOW Learn More	RStudio Server Open Source DOWNLOAD NOW Learn More	RStudio Server Pro DOWNLOAD NOW Learn More	RStudio Server Pro + RStudio Connect CONTACT SALES Learn More about RStudio Connect

Installers for Supported Platforms

Installers
RStudio 1.1.447 - Windows Vista/7/8/10
RStudio 1.1.447 - Mac OS X 10.8+ (64-bit)
RStudio 1.1.447 - Ubuntu 12.04-15.10/Debian 8 (32-bit)
RStudio 1.1.447 - Ubuntu 12.04-15.10/Debian 8 (64-bit)
RStudio 1.1.447 - Ubuntu 16.04+/Debian 9+ (64-bit)
RStudio 1.1.447 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)
RStudio 1.1.447 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)

Size	Date	MD5
85.8 MB	2018-04-18	359df07f279db25c99d0f91449b0fc33
74.5 MB	2018-04-18	13d679b7ec208bd31f22550c0dd6c99
89.3 MB	2018-04-18	8186c1b10793a6ff3138cd16af3ea433
97.4 MB	2018-04-18	ad9410a0c74eb68d4c34bd880fd103a6
64.5 MB	2018-04-18	1eecbf2c80fb329a8a1bcb92a68e45e
88.1 MB	2018-04-18	aac10abbc88315ddd3bf2b25ce81a4
90.6 MB	2018-04-18	0dd507d4eccc283a64b63f7599

Zip/Tarballs

Zip/tar archives
RStudio 1.1.447 - Windows Vista/7/8/10
RStudio 1.1.447 - Ubuntu 12.04-15.10/Debian 8 (32-bit)
RStudio 1.1.447 - Ubuntu 12.04-15.10/Debian 8 (64-bit)
RStudio 1.1.447 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)
RStudio 1.1.447 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)

Size	Date	MD5
122.9 MB	2018-04-18	6dc0ecccfb2a0a3a978967b49ce550c1b
90 MB	2018-04-18	cf87670c1f0cfed7e5b5d55a2ec8072
98.3 MB	2018-04-18	b05b723afc433817d9fcaef09f7ed012d
88.8 MB	2018-04-18	9db18a7481ff94203b60ebc78d6b05ea
91.4 MB	2018-04-18	ac74ff99c73d3623e57f6b38f0f9901

Source Code

A tarball containing source code for RStudio v1.1.447 can be downloaded from [here](#)

4. 설치 및 실행

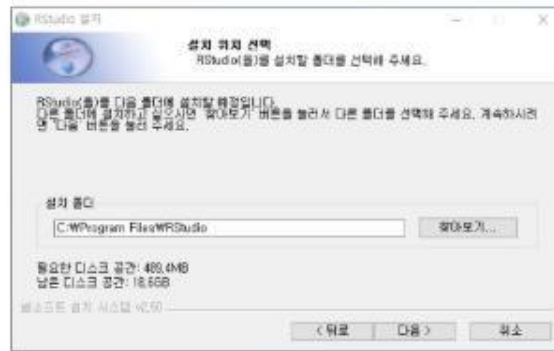


4.5 R Studio 설치 및 완료(2/6)

- 다운로드 받은 설치 파일을 실행하여 내 컴퓨터에 RStudio 설치



다음 버튼을 클릭하여 진행한다.

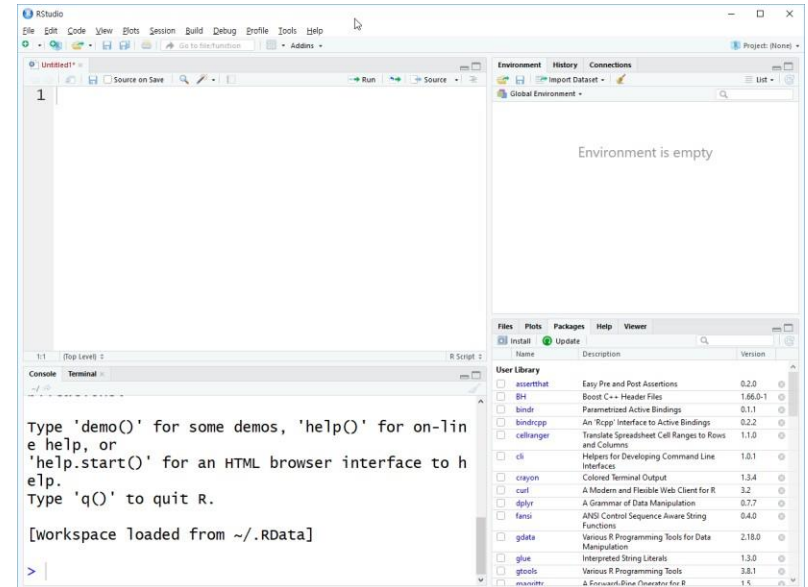


설치 폴더 설정은 특별한 경우가 아니라면 기본 설정 폴더에 설치합니다.



마침을 클릭하여 설치를 끝냅니다.

- 윈도우 시작 버튼을 클릭하여 RStudio 메뉴에서 RStudio 실행



4. 설치 및 실행



4.5 R Studio 설치 및 완료(3/6)

- 이번에는 우분투에서 설치하는 방법
- Ubuntu용으로 다운로드
- 사용 중인 운영체제 버전 해당하는 설치 파일 다운로드

Installers for Supported Platforms

Installers	Size	Date	MD5
RStudio 1.1.447 - Windows Vista/7/8/10	85.8 MB	2018-04-18	359df07f279db25c99d0f91449b0fc33
RStudio 1.1.447 - Mac OS X 10.6+ (64-bit)	74.5 MB	2018-04-18	13d679b7ec208bd31f22550c0cdd6c99
RStudio 1.1.447 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	89.3 MB	2018-04-18	8186c1b10793a6ff3138cd16af3ea433
RStudio 1.1.447 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	97.4 MB	2018-04-18	ad9410a0c74eb68d4c34bd880fd103a6
RStudio 1.1.447 - Ubuntu 16.04+/Debian 9+ (64-bit)	64.5 MB	2018-04-18	1eecbf2c80fb329a8a1bcb92a68e459e
RStudio 1.1.447 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	88.1 MB	2018-04-18	aac10abbc88315ddd3bf2b25ce8d9814
RStudio 1.1.447 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	90.6 MB	2018-04-18	0dd507d4ece283a64b63f759dc9e4fad

Zip/Tarballs

Zip/tar archives	Size	Date	MD5
RStudio 1.1.447 - Windows Vista/7/8/10	122.9 MB	2018-04-18	6dc0ecef2a0a3a978967b49ce550c1b
RStudio 1.1.447 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	90 MB	2018-04-18	efd87670c1f0cfed7e5b5d55a2ec8072
RStudio 1.1.447 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	98.3 MB	2018-04-18	b05b723afc433817d9fcaf09f7ed012d
RStudio 1.1.447 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	88.8 MB	2018-04-18	9db18a7481ff94203b60ebc78d6b05ea
RStudio 1.1.447 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	91.4 MB	2018-04-18	ac74fff99c73d3623e57f6b38f0f9901

Source Code

A tarball containing source code for RStudio v1.1.447 can be downloaded from [here](#)



4. 설치 및 실행



4.5 R Studio 설치 및 완료(4/6)

- 다운로드 받은 deb 파일을 클릭해서 설치를 하는 방법과 터미널에서 명령어를 이용해서 설치하는 방법이 있으며, deb파일을 클릭하면 아래와 같이 설치를 진행 가능(관리자권한이 필요하여 비밀번호 필요)



- 터미널에서 dpkg 명령어를 사용하여 설치 진행 가능

```
nipa@ubuntu:~/Downloads$ sudo dpkg -i ./rstudio-xenial-1.1.447-amd64.deb
```



4. 설치 및 실행



4.5 R Studio 설치 및 완료(5/6)

- 우분투에서 RStudio 실행하기

RStudio 검색 후 실행



프로그램 표시 클릭

- 터미널 화면에서 `rstudio`라고 명령을 실행하면 위와 동일하게 RStudio가 실행
`nipa@ubuntu:~/Downloads$ rstudio`



4. 설치 및 실행



4.5 R Studio 설치 및 완료(6/6)

- RStudio 구성

- Source

RScript 소스 내용 표시

Ctrl + Enter를 해당 라인에서 실행하면

Console창에 실행

- Console

R 명령어를 실행할 수 있는 화면

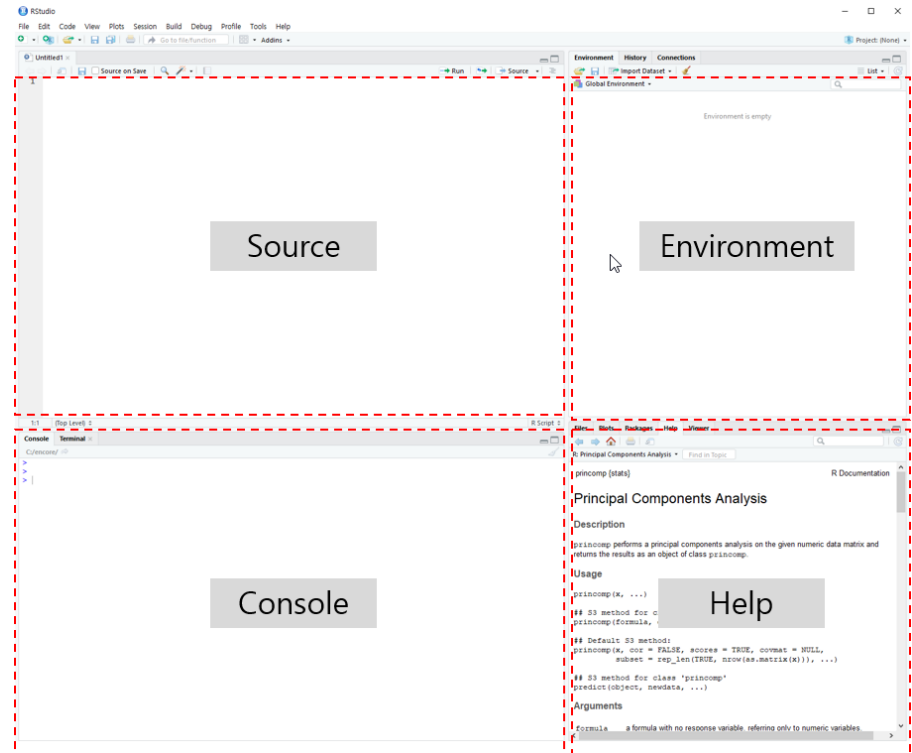
Ctrl + I을 실행하면 화면 정리 가능

- Environment

R에서 사용하는 데이터들이 메모리에 로딩되면 이 창에 표시

- Help

도움말 화면을 표시



5. 기능소개



세부 목차

1. 변수
2. Vector
3. List
4. Matrix
5. Data.Frame
6. Array



5. 기능소개



5.1 변수

- 변수란 데이터를 저장하는 공간이며, R에서는 변수의 이름을 선언하고 값을 할당하면 메모리에 로딩되어 관리
- 할당된 값에 수정은 지원하지 않고, 사용자가 R의 변수 값을 수정할 경우가 생기면 해당 변수의 값이 직접 수정되는 것이 아니라 새로운 값을 메모리에 할당하고 기존의 변수가 이 값을 사용할 수 있도록 연결
- R의 변수 선언 규칙
 - 변수의 이름을 선언할 때는 알파벳, 숫자, '_', '.' 등으로 구성되며, '.'는 사용 불가
 - 변수명의 첫 글자는 숫자로 시작 불가
- 변수에 값을 할당
 - '<->'와 '='를 사용하여 값을 할당할 수 있지만, 일반적으로 R에서는 '<->'를 사용

```
# 변수를 선언
x <- 1
y <- 'Hello'
```

5. 기능소개



5.2 Vector

- R의 기본 자료형은 숫자를 나타내는 Numeric, 참(TRUE)과 거짓(FALSE)를 표현하는 logical, 문자 및 문자열을 저장할 수 있는 character형 있음
- Vector는 동일한 데이터 형식의 값으로 구성된 1차 행렬 객체
- Vector 생성에 사용되는 함수는 Combine의 앞 글자를 사용한 c() 함수 사용
- c() 함수는 일련의 입력된 데이터를 병합하는 함수
- R은 모든 데이터를 기본적으로 Vector 형식으로 저장

```
# oss라는 변수에 숫자 1,2,3,4,5를 저장  
oss <- c(1,2,3,4,5)
```

- Vector는 1차 행렬 구조이며, 각각의 원소라고 불리는 데이터는 Index라고 불리는 번호표를 가지고 있으며, 이를 사용하여 해당 번호의 데이터를 접근할 수 있음

```
king <- c("태조", "정종", "태종", "세종", "문종", "단종", "세조", "예종",  
"성종", "연산군")  
king[3]
```

결과

```
[1] "태종"
```



5. 기능소개



5.3 List

- 통계 관련 함수 대부분이 리스트 형식으로 결과 출력
- 선형 회귀 함수인 'lm' 함수 출력은 벡터, 데이터프레임, 리스트 등을 포함한 12개의 복합 리스트

- 리스트의 생성

리스트를 생성하려면 list() 함수를 사용하여 생성

list(데이터1, 데이터2, 데이터3 ...) 형식으로 서로 다른 데이터 구조 저장

```
list.data <- list(1:10,  
                data.frame(a=1:10, b = LETTERS[1:10]),  
                matrix(1:9, ncol=3))
```

- 리스트 원소에서 Index 번호를 가지고 데이터 추출 가능 리스트
는 [[번호]] 혹은 []를 사용하여 해당 리스트 원소 추출 가능 []를
사용하여 데이터를 추출하면 해당 결과는 list 구조 출력

```
list.data[1]
```



5. 기능소개



5.4 Matrix

- Matrix는 수학에서 배운 행렬에서 사용하는 행과 열로 구성된 데이터 구조
- 벡터 데이터 형식처럼 한 가지의 유형으로만 저장되어야 하며, 첫번째 행은 숫자로 구성되고, 두번째 행은 문자로 구성되는 형태는 허용 안함
- 수학적 행렬 연산이 가능하며, 수학연산을 실행할 시에는 행렬 연산에 대한 수학적 지식 필요
- 행렬 생성

행렬은 `matrix()` 함수를 사용하여 생성 가능

`matrix()` 함수는 데이터 항목을 채워 나가는 방식으로 이루어져 있으며, 위에서 아래 방향으로 데이터를 채워감

`matrix(data, row, col, byrow=FALSE, dimnames=NULL)`

- 1~6 숫자 데이터를 사용하여 `matrix()` 함수를 사용하여 행렬 생성

```
matrix.1 <- matrix(1:6)
```



5. 기능소개



5.5 data.frame

- 데이터프레임은 데이터 분석 시에 가장 많이 사용되며, 가장 기본이 되는 데이터 형식
- 엑셀의 스프레드시트처럼 데이터가 행과 열로 구성되어 표 형태로 정리된 모습
- 엑셀에서 행(1,2,3.....)과 열(A, B, C.....)의 이름이 있듯이 데이터프레임도 각각의 행과 열에 이름을 부여할 수 있고, 열의 이름에는 관측값의 이름을 저장하며, 행의 이름에는 관측값이 가지고 있는 시간, 번호 등 부여 가능
- 행렬(Matrix)는 저장된 데이터의 타입이 모두 동일해야 하지만, 데이터프레임은 다양한 데이터 타입 저장 가능
- 데이터 프레임 생성
데이터프레임은 `data.frame()` 함수를 사용하여 생성할 수 도 있고, csv 파일 형태의 파일을 읽어 들일 때 사용하는 `read.csv()` 함수는 자동으로 데이터 프레임 형식으로 데이터 읽음

```
data.frame(..., row.names = NULL, check.rows = FALSE, check.names = TRUE,  
fix.empty.names = TRUE, stringsAsFactors = default.stringsAsFactors())
```



5. 기능소개



5.5 data.frame

- x변수에 1~5까지 데이터를 넣고 y변수에는 a~e까지 데이터를 넣고 data.frame()함수에 값을 전달하여 새로운 데이터프레임 생성

```
df <- data.frame(x = c(1,2,3,4,5), y=c('a', 'b', 'c', 'd', 'e'))
```

```
> df
  x y
1 1 a
2 2 b
3 3 c
4 4 d
5 5 e
```



5. 기능소개



5.6 Array

- 배열은 다차원 벡터이며, 행렬은 2차원이라면 배열은 2차원을 넘어서 다차원 데이터
- 배열의 값은 모두 같은 데이터 유형을 가져야 함
- 대괄호를 사용해 개별 값에 접근 가능
- 배열은 `array()` 함수를 사용해서 배열 생성

`array(data, dim, dimnames)`

- 첫번째 매개변수에는 `data`를 입력하고, 두번째 매개변수에는 배열의 차원을 설정하고 이 값을 지정하지 않는다면 기본적으로 1차원 배열 생성
- 세 번째 매개변수는 차원의 이름 부여
- 아래 예제는 3x4 차원의 배열 생성

```
array.1 <- array(1:12, dim=c(3,4))
```



6. 활용예제



세부 목차

- 1. 기계학습
- 2. 시각화



6. 활용예제



6.1 기계학습(1/2)

- R은 많은 기계학습 알고리즘을 제공하고 있고 그 중에서 회귀분석에 대해서 알아봄
- 모델 생성

lm() 함수를 사용하여 선형 회귀 모형을 만들 수 있음

인자로 $y \sim x$ 형태의 formula를 만들고 데이터 입력

R에서는 기본적으로 연습 데이터를 제공하는데 아래는 그중 cars라는 데이터임

```
model <- lm(dist ~ speed, cars)
```

cars의 데이터에서 speed를 설명변수로 dist를 종속변수로 하여 선형회귀 모델 생성

- 선형회귀 결과 출력

선형회귀의 절편과 기울기를 아래와 같이 확인 가능

```
coef(model)
```

```
> coef(model)
```

```
(Intercept)      speed
-17.579095      3.932409
```



6. 활용예제



6.1 기계학습(2/2)

- 생성된 모델을 사용하여 예측값 출력

각각의 speed 값에 대한 dist의 값을 예측한 예제값을 fitted() 함수로 구할 수 있음

```
fitted(model)
```

```
      1      2      3      4      5      6      7      8      9     10     11     12     13     14     15
-1.849460 -1.849460  9.947766  9.947766 13.880175 17.812584 21.744993 21.744993 21.744993 25.677401 25.677401 29.609810 29.609810 29.609810
      16      17      18      19      20      21      22      23      24      25      26      27      28      29      30
33.542219 33.542219 33.542219 33.542219 37.474628 37.474628 37.474628 37.474628 41.407036 41.407036 41.407036 45.339445 45.339445 49.271854 49.271854
      31      32      33      34      35      36      37      38      39      40      41      42      43      44      45
49.271854 53.204263 53.204263 53.204263 53.204263 57.136672 57.136672 57.136672 61.069080 61.069080 61.069080 61.069080 61.069080 68.933898 72.866307
      46      47      48      49      50
76.798715 76.798715 76.798715 76.798715 80.731124
```

- predict() 함수는 모델에 대한 새로운 데이터(3,7,12)의 dist 예측값 출력 가능

```
predict(model, newdata = data.frame(speed = c(3, 7, 12)))
```

```
      1      2      3
-5.781869  9.947766 29.609810
```

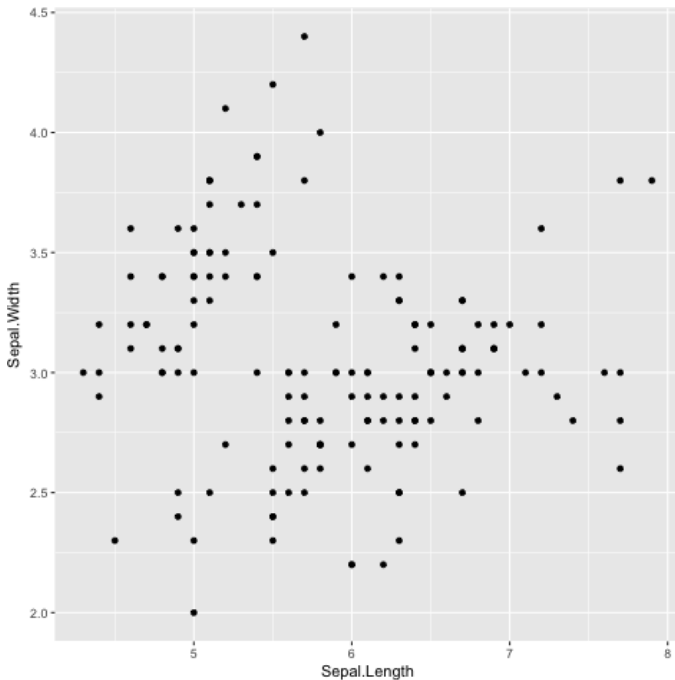


6. 활용예제



6.2 시각화

- R은 시각화에 매우 뛰어난 장점을 가지고 있음
- 그 중에서 ggplot2라는 패키지는 많은 R의 사용자들이 자주 사용하는 시각화 패키지이며, 많은 기능들을 다루고 있음
- 아래 예제는 ggplot2의 기능을 사용하여 scatter plot를 그린 예제



```
# library
```

```
library(ggplot2)
```

```
head(iris)
```

```
ggplot(iris, aes(x=Sepal.Length, y=Sepal.Width)) +  
  geom_point()
```





Q R을 공부를 시작하는데 통계가 필요한가요?

A R은 통계 프로그램에서 시작을 했기 때문에 통계에 대한 강력한 기능을 제공하고 있습니다. R을 제대로 사용을 하기 위해서는 통계에 대한 지식을 반드시 필요하진 않지만, 좀 더 고급 기술을 사용하기 위해서는 필요합니다.

Q R을 설치하고 R-Studio를 왜 설치해야 하나요?

A R을 설치하고 기본 R을 가지고만 분석 및 진행을 할 수 있습니다. 하지만 사용해보면 느낄 수 있겠지만, 기본적으로 제공하는 R은 사용자에게 많은 정보를 제공해 주지 못하고 사용 또한 불편합니다. 하지만 익숙해진다면 R-Studio같은 IDE 프로그램을 사용하지 않으셔도 됩니다.





Q R을 통계에서 사용하지 않고 다른 곳에 어디에 사용하나요?

A R은 데이터 전처리를 위한 dplyr라는 패키지를 제공하며, 해당 패키지를 사용하면 데이터 전처리를 쉽게 할 수 있습니다. R을 통계에 사용하지 않더라도 dplyr 같은 패키지만을 사용해도 데이터 전처리를 간단하게 진행할 수 있습니다.

Q R Shiny는 무엇인가요?

A R의 가장 큰 장점인 시각화를 웹하고 연동하여 인터랙티브하게 시각화를 구성할 수 있습니다. Shiny라는 패키지를 따로 설치가 필요하며, 웹에 대한 지식을 가지고 있다면 조금 더 화려한 시각화를 구성할 수 있습니다.



8. 용어정리



용어	설명
IDE	Integrated Development Environment(통합 개발 환경)
선형회귀	종속 변수 y 와 한 개 이상의 독립 변수 X 와의 선형 상관 관계를 모델링하는 회귀 분석 기법
GPL	GNU General Public License의 약자. 라이선스의 종류 중 하나. 소스의 취득, 수정, 배포, 공개가 자유로우나 최종 배포 시 GPL라이선스를 따라야 함. 수정 배포 시 자체 개발한 소스도 공개해야 함
Scatter Plot	산점도라고 불리며 직교 좌표계를 이용해 두 개 변수간의 관계를 그림으로 나타내는 방법
S 언어	벨 연구소에서 개발된 통계적인 프로그래밍 언어
GUI	Graphical User Interface
CLI	Command Line Interface



Open Source Software Installation & Application Guide



이 저작물은 크리에이티브 커먼즈 [저작자표시-비영리-동일조건 변경허락 2.0 대한민국 라이선스]에 따라 이용하실 수 있습니다.