

# ADsP 과목3 데이터 분석

인공지능소프트웨어학과

강환수 교수

1. R 기초와 데이터 마트
2. 통계 분석
3. 정형 데이터 마이닝

DONGYANG MIRAE UNIVERSITY  
Dept. of Artificial Intelligence



# ADsP

## (Advanced Data Analytics Semi-Professional)

과목명	주요항목	세부항목
데이터 이해	데이터의 이해	데이터와 정보
		데이터베이스의 정의와 특징
		데이터베이스 활용
	빅데이터의 이해	빅데이터의 이해
		빅데이터의 가치와 영향
		비즈니스 모델
		위기 요인과 통제 방안
		미래의 빅데이터
	가치 창조를 위한 데이터 사이언스와 전략 인사이트	빅데이터분석과 전략 인사이트
		전략 인사이트 도출을 위한 필요 역량
		빅데이터 그리고 데이터 사이언스의 미래
데이터분석 기획	데이터분석 기획의 이해	분석 기획 방향성 도출
		분석 방법론
		분석 과제 발굴
		분석 프로젝트 관리 방안
	분석 마스터 플랜	마스터 플랜 수립
		분석 거버넌스 체계 수립
데이터분석	R기초와 데이터 매트	R기초
		데이터 매트
		결측값 처리와 이상값 검색
	통계분석	통계학 개론
		기초 통계분석
		다변량 분석
		시계열 예측
	정형 데이터 마이닝	데이터 마이닝 개요
		분류분석(Classification)
		군집분석(Clustering)
		연관분석(Association Analysis)

- 인공지능소프트웨어학과
- 강환수 교수
- 2025.08.06(수)~07(목)
- 3교과목: 데이터분석

# R 기초와 데이터 마트

- 1. R기초
- 2. 데이터 마트
- 3. 결측값 처리와 이상값 검색

**ADsP**

**(Advanced Data Analytics Semi-Professional)**

**R과 R스튜디오의 설치 및 사용**

오픈소스 프로그램으로 통계, 데이터 마이닝과 그래프를 위한 언어

- 데이터 분석을 위한 자료 전처리, 통계 및 시각화를 지원하는 컴퓨터 언어 및 환경
- 1980년 AT&T 벨 연구소의 John Chambers가 개발한 S 언어를 기반으로 1993년 뉴질랜드 Auckland 대학의 통계학과 교수 Robert Gentleman과 Ross Ihaka 가 개발

R의 특징

- 다양한 최신 통계 분석 및 마이닝 기능을 R 플랫폼에서 제공
- 다양한 최신 알고리즘을 제공해 다양한 시도 가능
- 무료 소프트웨어
- R-Project와 CRAN (Comprehensive R Archive Network)에서 배포  
<https://www.r-project.org/>  
<https://cran.r-project.org/>

## 4. R과 R스튜디오의 설치 및 사용

### 1. R과 R스튜디오의 소개

- Python : 프로그래밍 언어로서의 특성이 강함
- R : 데이터 분석을 목적으로 개발, 통계 전문에서 확장되고 있는 상황
  - RStudio라는 훌륭한 작업환경 제공
  - 풍부한 패키지(함수와 관련 자료의 묶음) 제공
  - 미려한 데이터 시각화 패키지 제공



그림 1-10 R과 파이썬



그림 1-11 R을 쉽게 사용할 수 있는 환경을 제공하는 R스튜디오

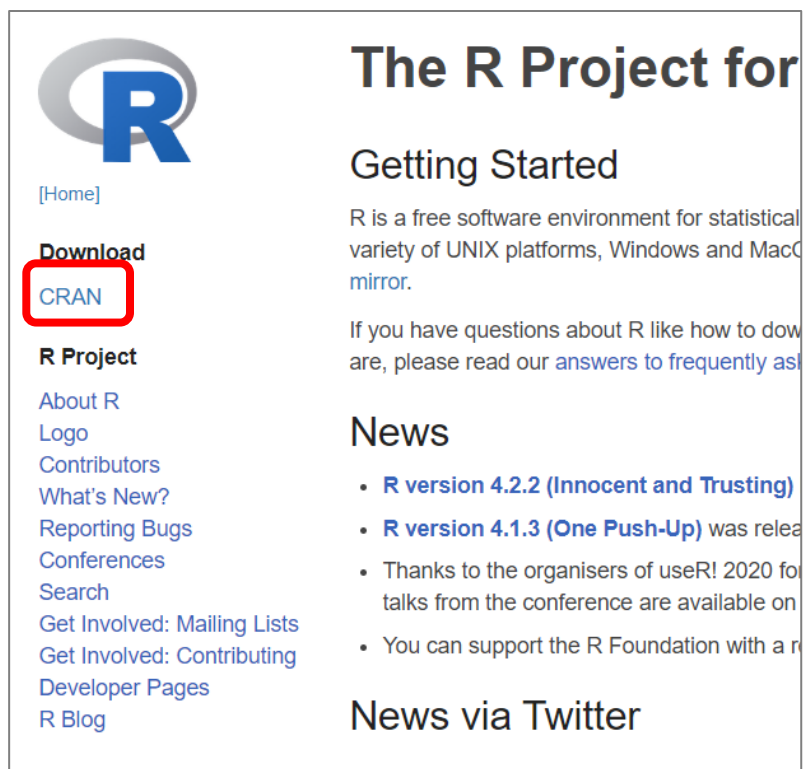
## 4. R과 R스튜디오의 설치 및 사용


### 2. R의 설치

01 <https://www.r-project.org/> 에 접속하여 설치 진행

2025년 8월 현재, 4.5.1

<https://cran.yu.ac.kr/>





[Home]

**Download**

**CRAN**

**R Project**

- About R
- Logo
- Contributors
- What's New?
- Reporting Bugs
- Conferences
- Search
- Get Involved: Mailing Lists
- Get Involved: Contributing
- Developer Pages
- R Blog

## The R Project for Statistical Computing

### Getting Started

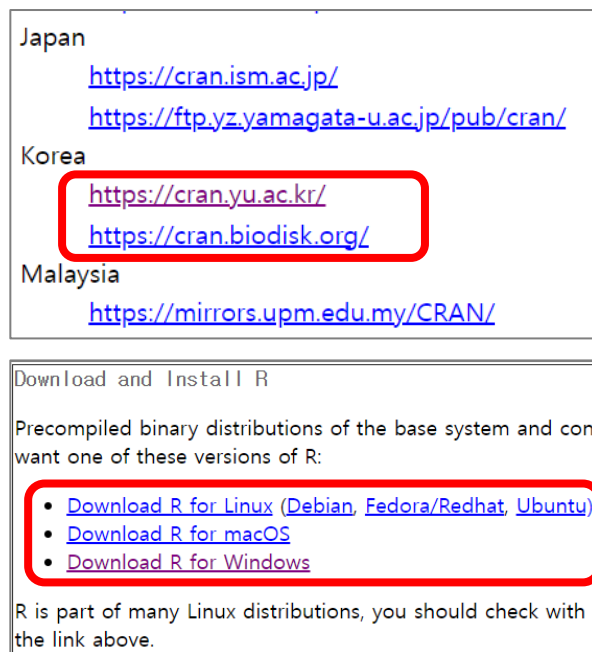
R is a free software environment for statistical computing and graphics. It is available for a variety of UNIX platforms, Windows and MacOS.

If you have questions about R like how to download or how to install, please read our [answers to frequently asked questions](#).

### News

- R version 4.2.2 (Innocent and Trusting)**
- R version 4.1.3 (One Push-Up)** was released
- Thanks to the organisers of useR! 2020 for the talks from the conference are available on [YouTube](#)
- You can support the R Foundation with a [donation](#)

### News via Twitter



Japan

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

Korea

- <https://cran.yu.ac.kr/>
- <https://cran.biodisk.org/>

Malaysia

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

### Download and Install R

Precompiled binary distributions of the base system and contributed packages are available for the following operating systems. Choose the one that matches your system.

- [Download R for Linux \(Debian, Fedora/Redhat, Ubuntu\)](#)
- [Download R for macOS](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your distribution for the link above.

자신의 운영체제에 맞는 버전을 설치한다

## 4. R과 R스튜디오의 설치 및 사용

02 [install R for the first time] 링크 클릭 → [Download R-4.5.1 for Windows] 클릭

### R for Windows

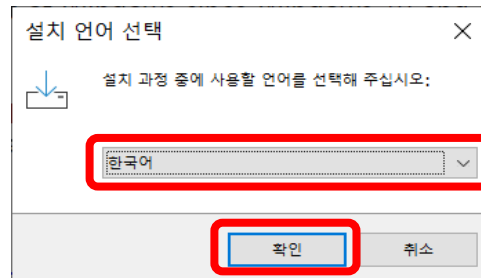
Subdirectories:

<a href="#">base</a>	Binaries for base distribution. This is what you want to <b>install R for the first time.</b>
<a href="#">contrib</a>	Binaries of contributed CRAN packages (for R >= 3.4.x).
<a href="#">old contrib</a>	Binaries of contributed CRAN packages for outdated versions of R (for R < 3.4.x).
<a href="#">Rtools</a>	Tools to build R and R packages. This is what you want to build your own packages or R itself.

### R-4.5.1 for Windows

[Download R-4.5.1 for Windows](#) (86 megabytes, 64 bit)  
[README on the Windows binary distribution](#)  
[New features in this version](#)

03 [한국어] 선택하고 [확인] 클릭

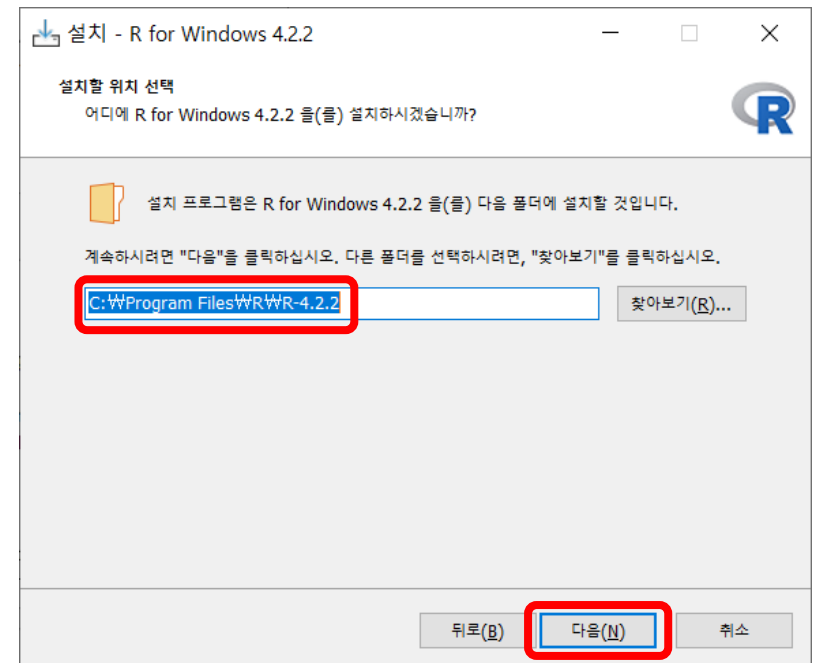
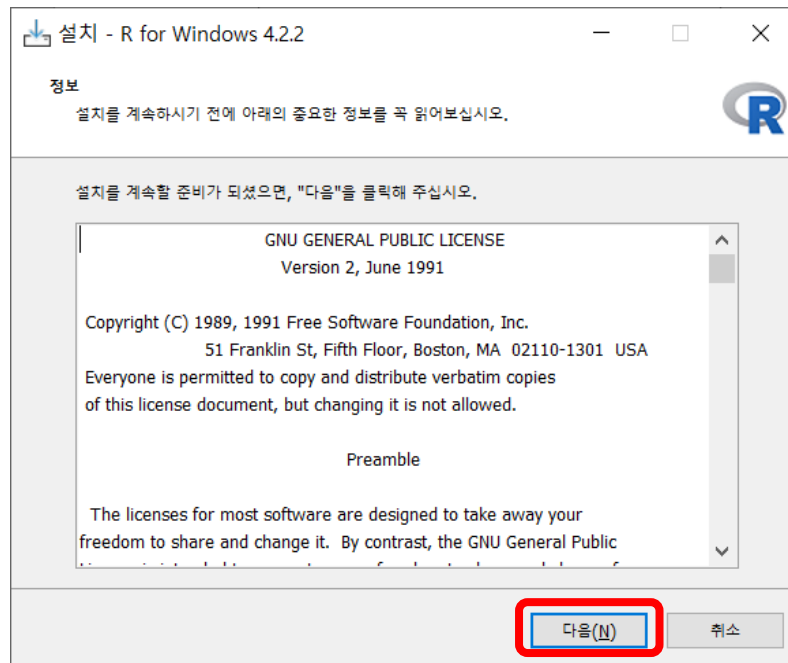




## 4. R과 R스튜디오의 설치 및 사용

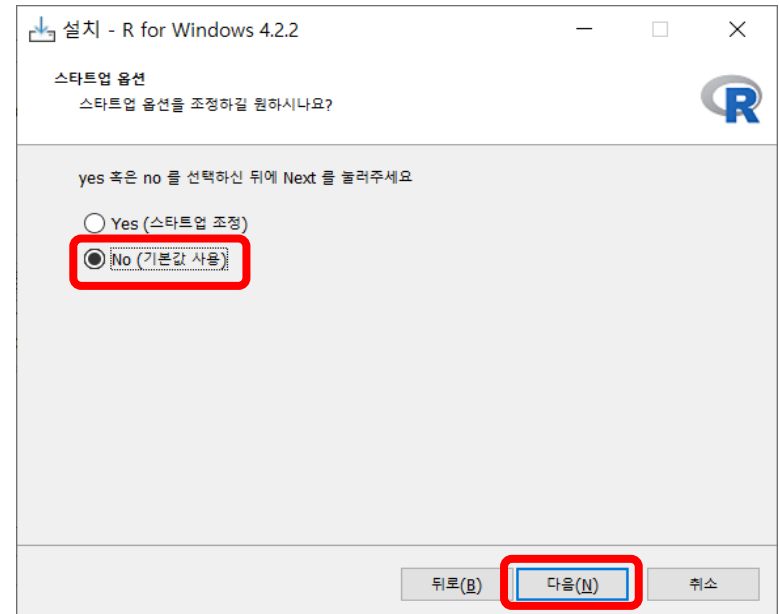
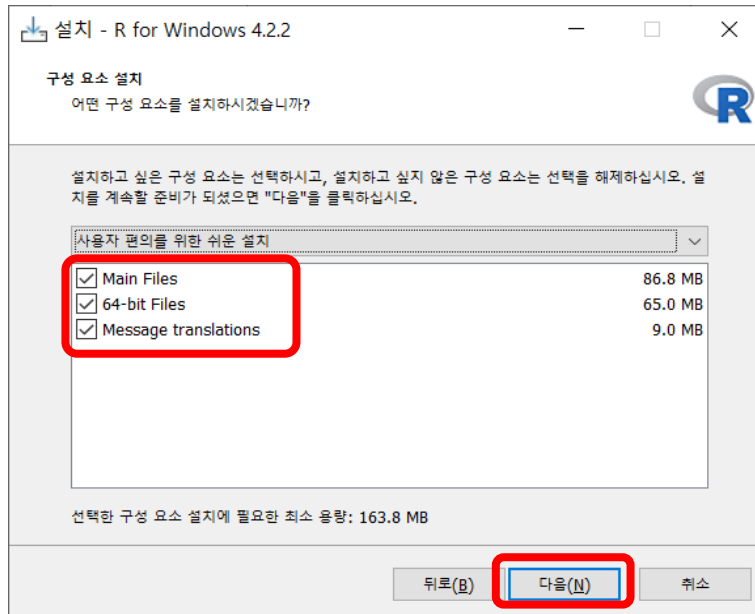
04 설치 정보가 나타나면 내용 확인하고 [다음] 버튼 클릭

→ 설치할 위치 선택에서 경로 변경하거나 유지한 채 [다음] 버튼 클릭



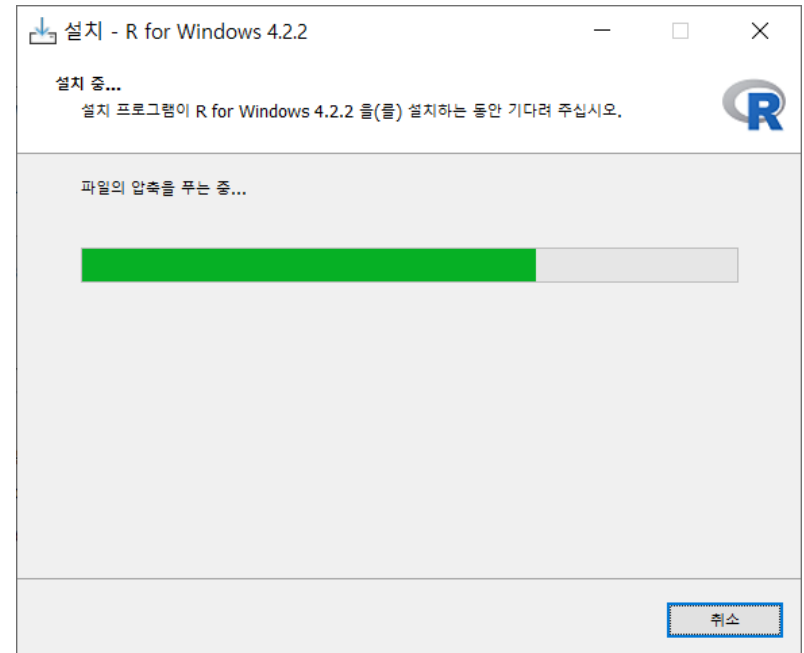
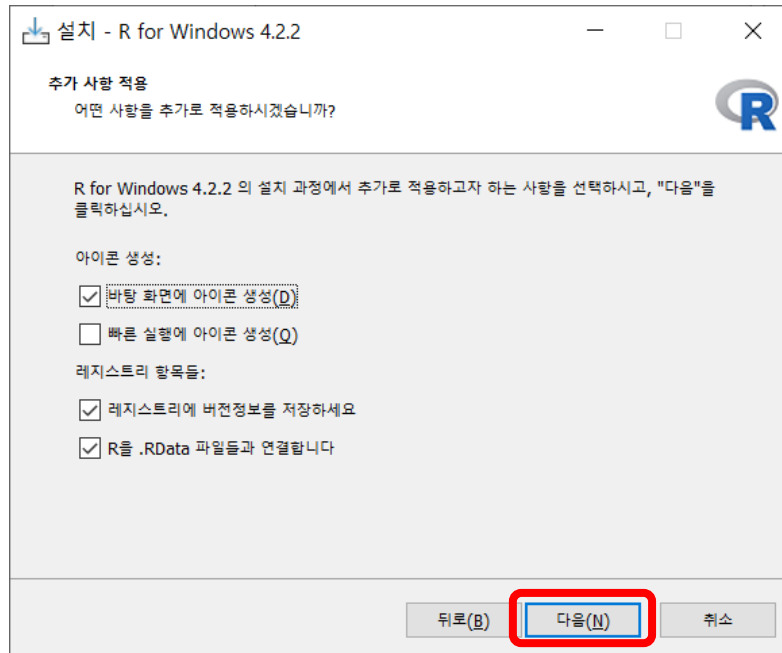
## 4. R과 R스튜디오의 설치 및 사용

05 구성 요소 설치에서 필요한 항목 체크하고 [다음] 버튼 클릭  
→ 스타트업 옵션에서 [No]를 선택 후, [다음] 버튼 클릭



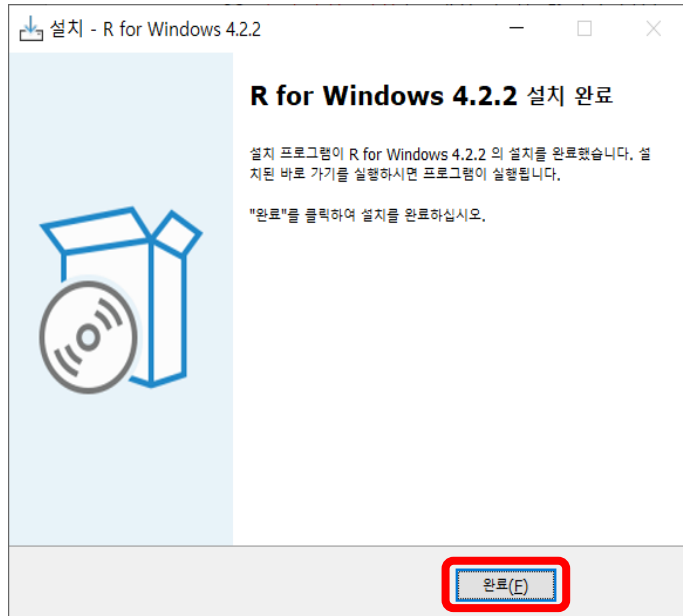
## 4. R과 R스튜디오의 설치 및 사용

06 추가 사항 적용에서 내용 변경 없이 [다음] 버튼 클릭 → 설치가 진행됨



## 4. R과 R스튜디오의 설치 및 사용

### 07 설치 완료 창 열리면 [완료] 버튼을 눌러 설치 완료



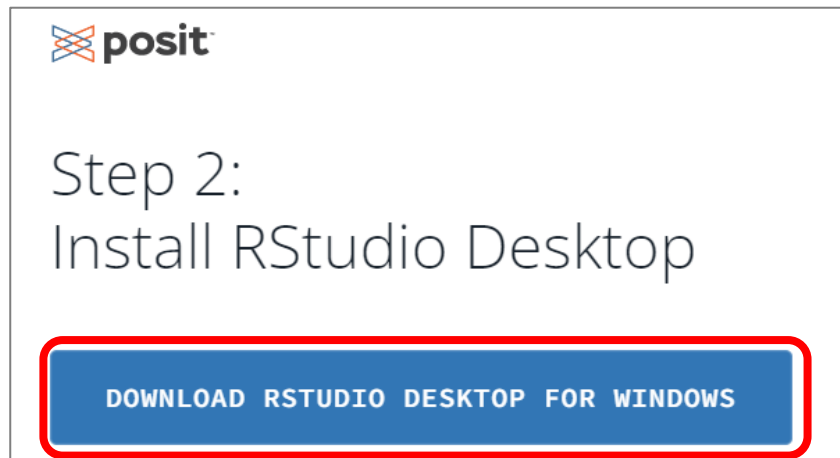
## 4. R과 R스튜디오의 설치 및 사용

### 3. R스튜디오의 설치

01 <https://posit.co/download/rstudio-desktop/>에 접속

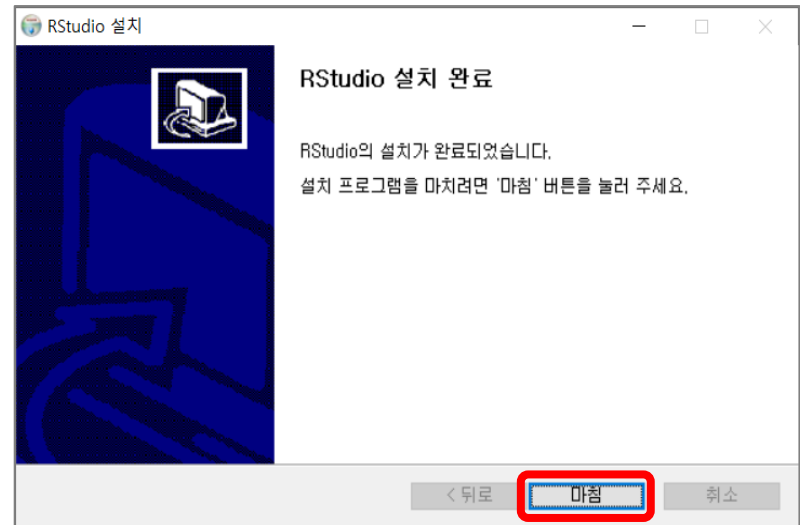
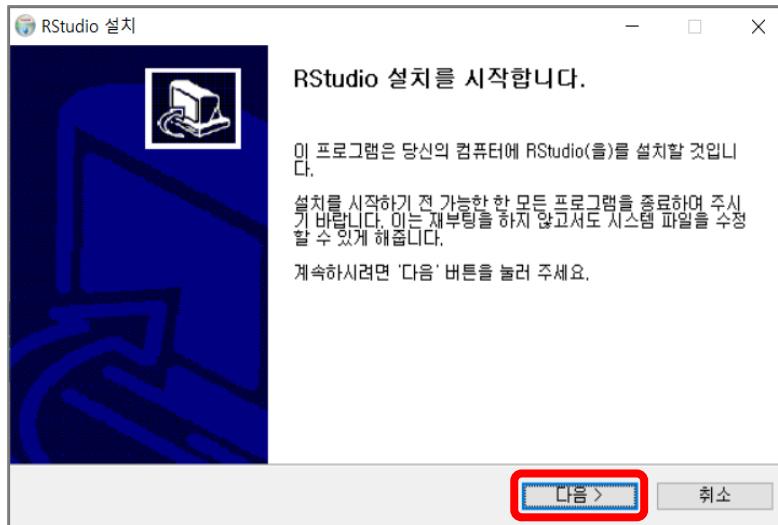
Step 1: Install R ► 이미 설치 했으므로 Skip

Step 2: Install RStudio Desktop ► 설치 파일을 다운로드



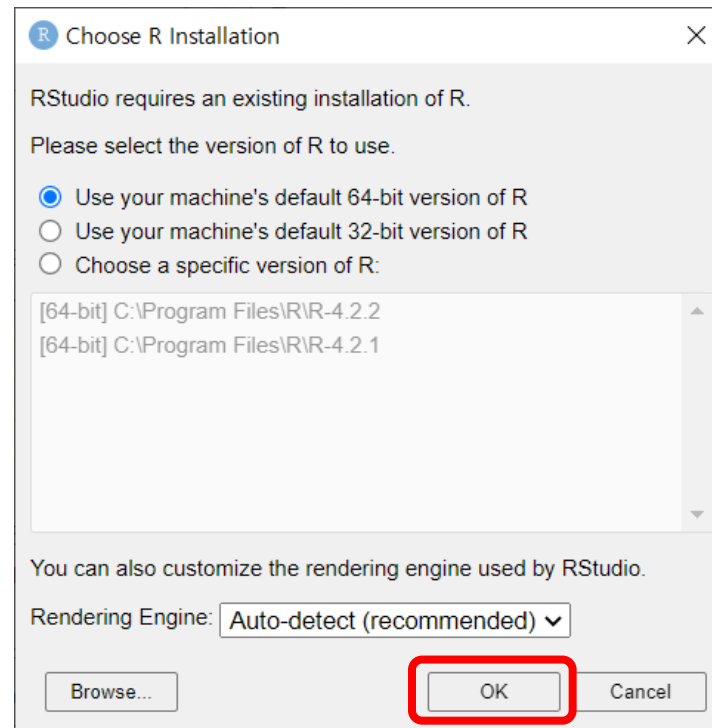
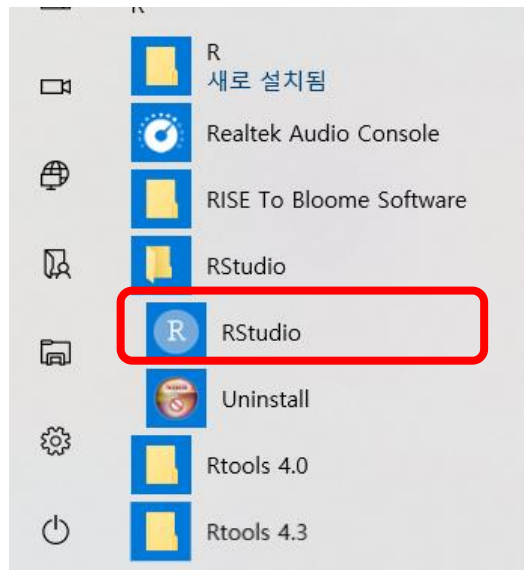
## 4. R과 R스튜디오의 설치 및 사용

02 설치 파일 더블클릭 → 계속해서 [다음] 버튼을 클릭 → R스튜디오 설치가 완료되면 [마침] 버튼 클릭



## 4. R과 R스튜디오의 설치 및 사용

04 설치가 완료되면 윈도우 시작 메뉴에서 [RStudio]-[RStudio] 클릭하여  
R스튜디오 실행



설치된 R 버전이 여러 개인 경우 어떤 버전을 사용할지 묻는다. 선택사항 변경 없이 [OK] 클릭

## 4. R과 R스튜디오의 설치 및 사용

### 4. R스튜디오의 화면 구성

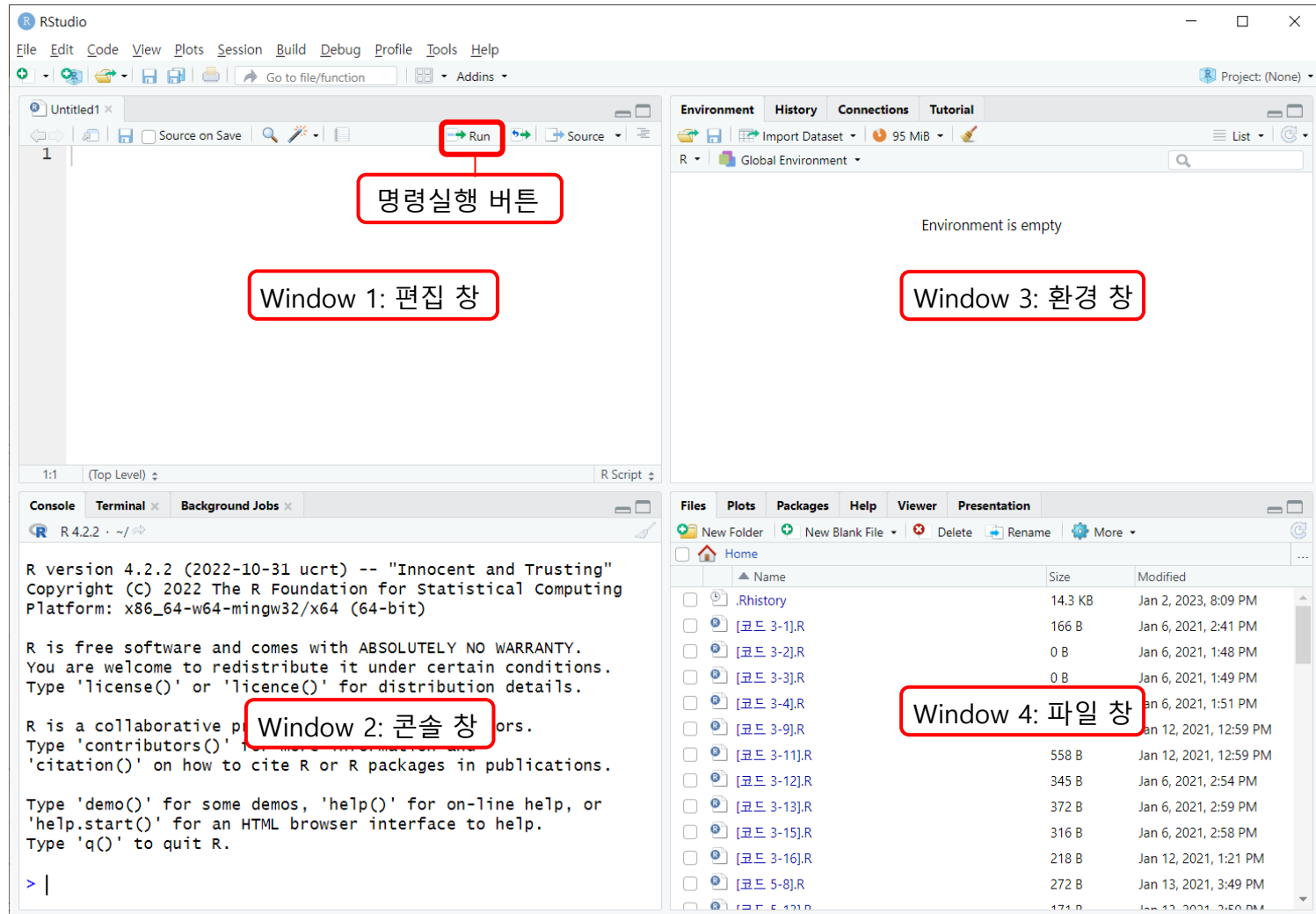


그림 1-12 R스튜디오 초기 화면



## 4. R과 R스튜디오의 설치 및 사용

### 4.1 편집(Script) 창

- R 명령문('R 스크립트' 라고도 한다.)들을 작성하고 실행하는 영역

### 4.2 콘솔(Console) 창

- 편집 창에서 R 명령문을 편집하고 실행 버튼을 클릭했을 때, 명령문의 실행 과정 및 결과를 표시하는 창

### 4.3 환경(Environment) 창

- R 명령문이 실행하는 동안 만들어지는 각종 변수나 자료구조의 내용을 보여주는 영역

### 4.4 파일(Files) 창

- 도움말, 패키지 설치 및 조회, 그래프 실행 내용 조회 등 유용한 기능을 제공하는 창

## 4. R과 R스튜디오의 설치 및 사용

### ❶ 파일(Files) 탭:

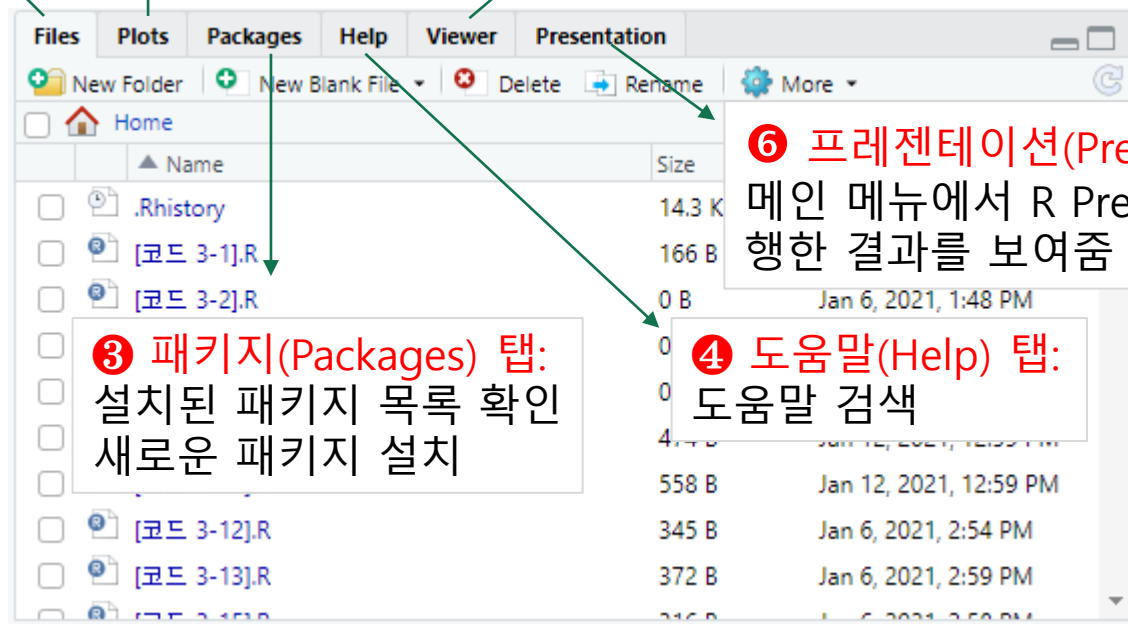
현재 작업 폴더의 내용을 탐색기처럼 보여준다

### ❷ 플롯(Plots) 탭:

그래프가 표시되는 영역

### ❸ 뷰어(Viewer) 탭:

결과가 웹브라우저에 나타나는 경우 여기에 표시



### ❻ 프레젠테이션(Presentation) 탭:

메인 메뉴에서 R Presentation을 실행한 결과를 보여줌

### ❹ 도움말(Help) 탭:

도움말 검색

### ❸ 패키지(Packages) 탭:

설치된 패키지 목록 확인  
새로운 패키지 설치

## 4. R과 R스튜디오의 설치 및 사용

### 5. R스튜디오 다루기

#### 5.1 R스튜디오 화면 재구성하기

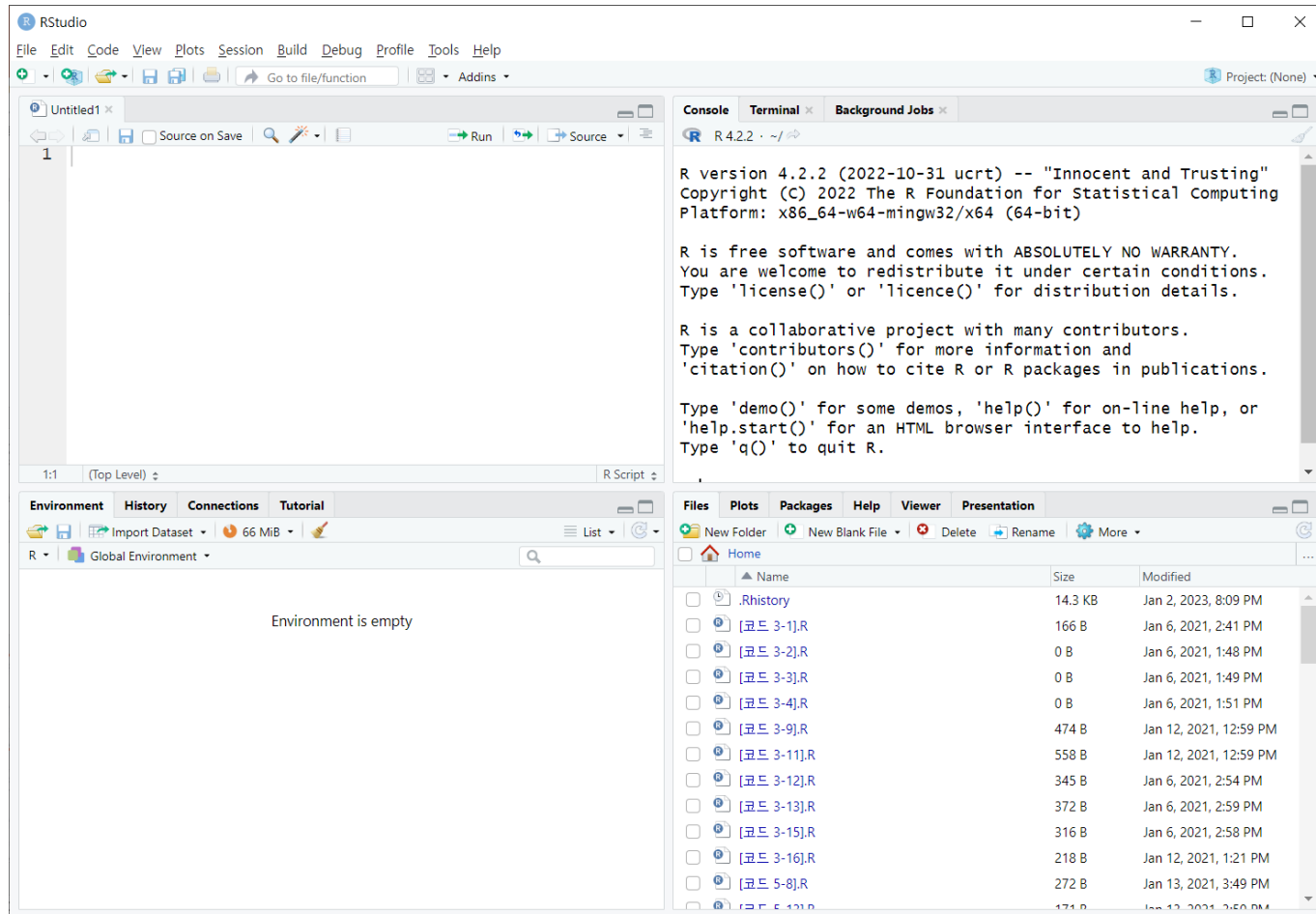


그림 1-13 콘솔 창 재배치 후의 R스튜디오

## 4. R과 R스튜디오의 설치 및 사용

### 5.2 R스튜디오에서 명령문의 실행

```
5+8  
3+(4*5)  
a <- 10  
print(a)
```

```
> 5+8  
[1] 13
```

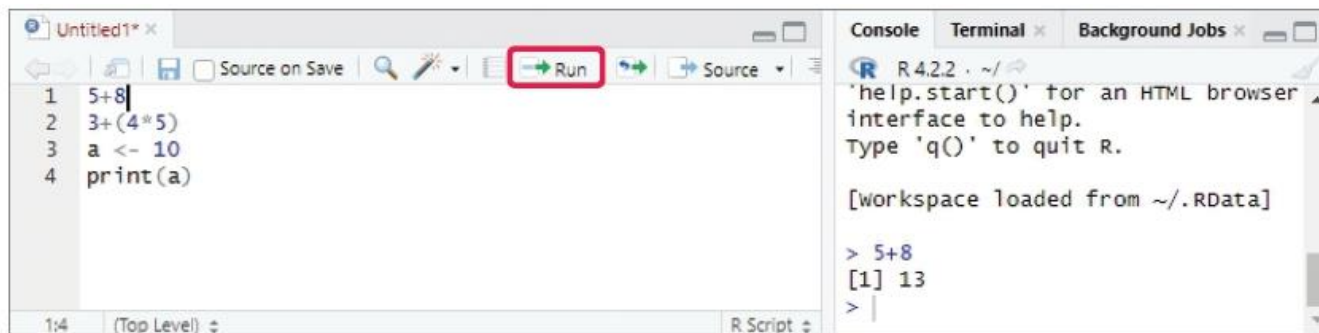


그림 1-14 편집 창 1행에 커서를 놓고 실행 아이콘을 클릭했을 때 콘솔 창에서의 실행 결과

## 4. R과 R스튜디오의 설치 및 사용

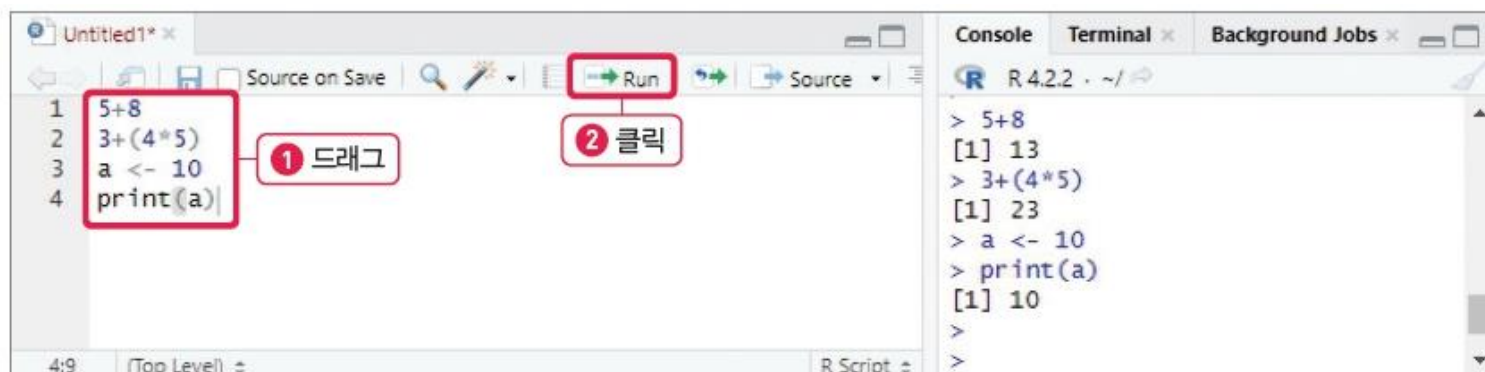


그림 1-15 편집 창 1~4행을 블록 선택하고 실행 아이콘을 클릭했을 때 콘솔 창에서의 실행 결과

표 1-1 명령어 실행과 관련된 단축키

명령어 실행	단축키
한 줄만 실행할 때	명령어가 있는 줄에서 <b>Ctrl</b> + <b>Enter</b>
여러 줄을 실행할 때	명령어들을 드래그하여 블록을 만든 후 <b>Ctrl</b> + <b>Enter</b>
편집된 모든 명령문들을 실행할 때	<b>Ctrl</b> + <b>Alt</b> + <b>R</b>
바로 직전에 실행한 명령을 다시 실행할 때	<b>Ctrl</b> + <b>Alt</b> + <b>P</b>

## 4. R과 R스튜디오의 설치 및 사용

### 5.3 R스튜디오에서의 저장과 종료

- 메뉴에서 [File]-[Save] 또는 [File]-[Save As]
- R 스크립트 파일의 확장자 이름은 일반적으로 'test.R'과 같이 '.R'을 붙임
- 아래와 같은 메시지가 출력되면 [Save] 클릭

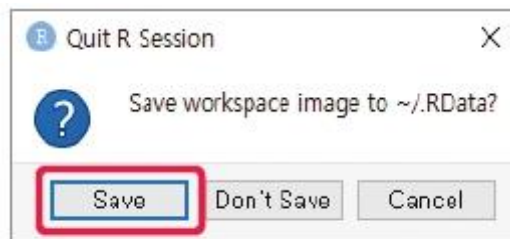


그림 1-16 R스튜디오 종료 대화상자

## 4. R과 R스튜디오의 설치 및 사용

### 5.4 패키지의 설치

- R에서는 데이터 분석을 위해서 매우 다양한 함수들을 제공
- 패키지(package) 는 이러한 함수들을 기능별로 묶어놓은 '꾸러미'
- 어떤 함수를 이용하기 위해서는 그 함수를 포함하고 있는 패키지를 사전에 설치해야 함



그림 1-17 현재 설치된 패키지 목록

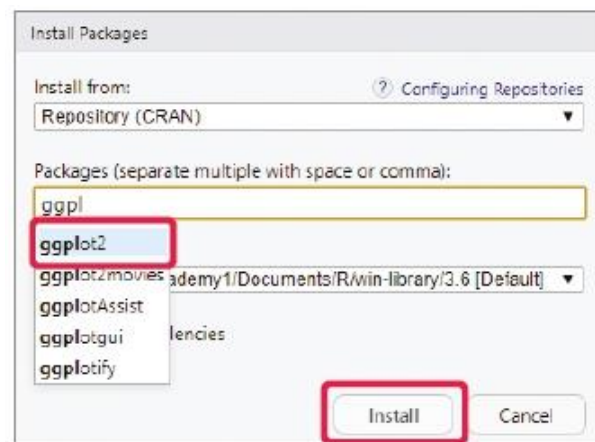
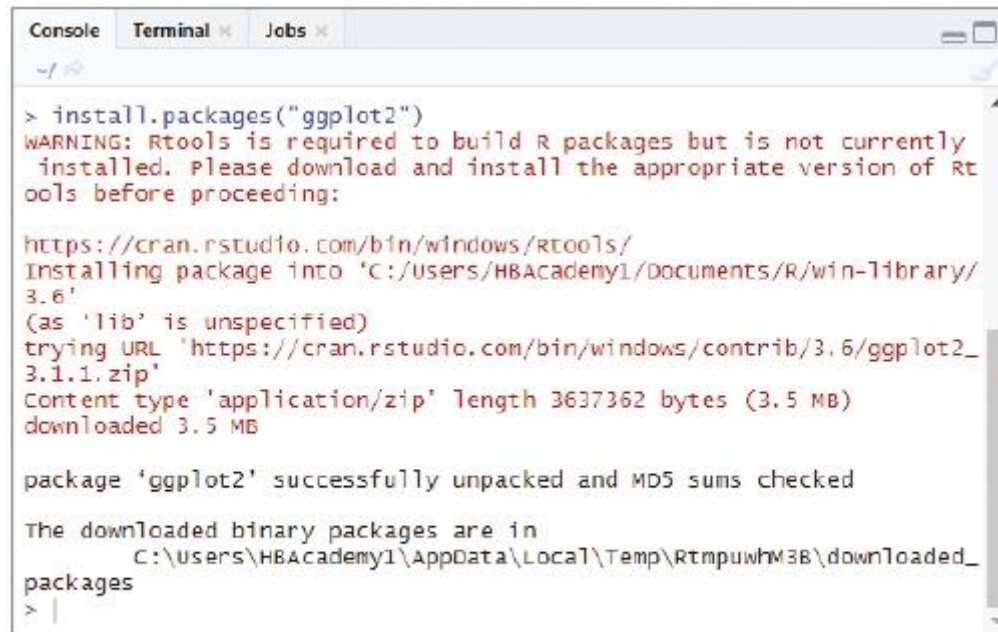


그림 1-18 패키지 설치 윈도우

## 4. R과 R스튜디오의 설치 및 사용



```
Console Terminal Jobs
~/
> install.packages("ggplot2")
WARNING: Rtools is required to build R packages but is not currently
installed. Please download and install the appropriate version of Rt
ools before proceeding:

https://cran.rstudio.com/bin/windows/rtools/
Installing package into 'C:/Users/HBAcademy1/Documents/R/win-library/
3.6'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/ggplot2_
3.1.1.zip'
Content type 'application/zip' length 3637362 bytes (3.5 MB)
downloaded 3.5 MB

package 'ggplot2' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:\Users\HBAcademy1\AppData\Local\Temp\RtmpuwhM3B\downloaded_
packages
> |
```

그림 1-19 패키지 설치가 성공한 경우의 일반적인 화면

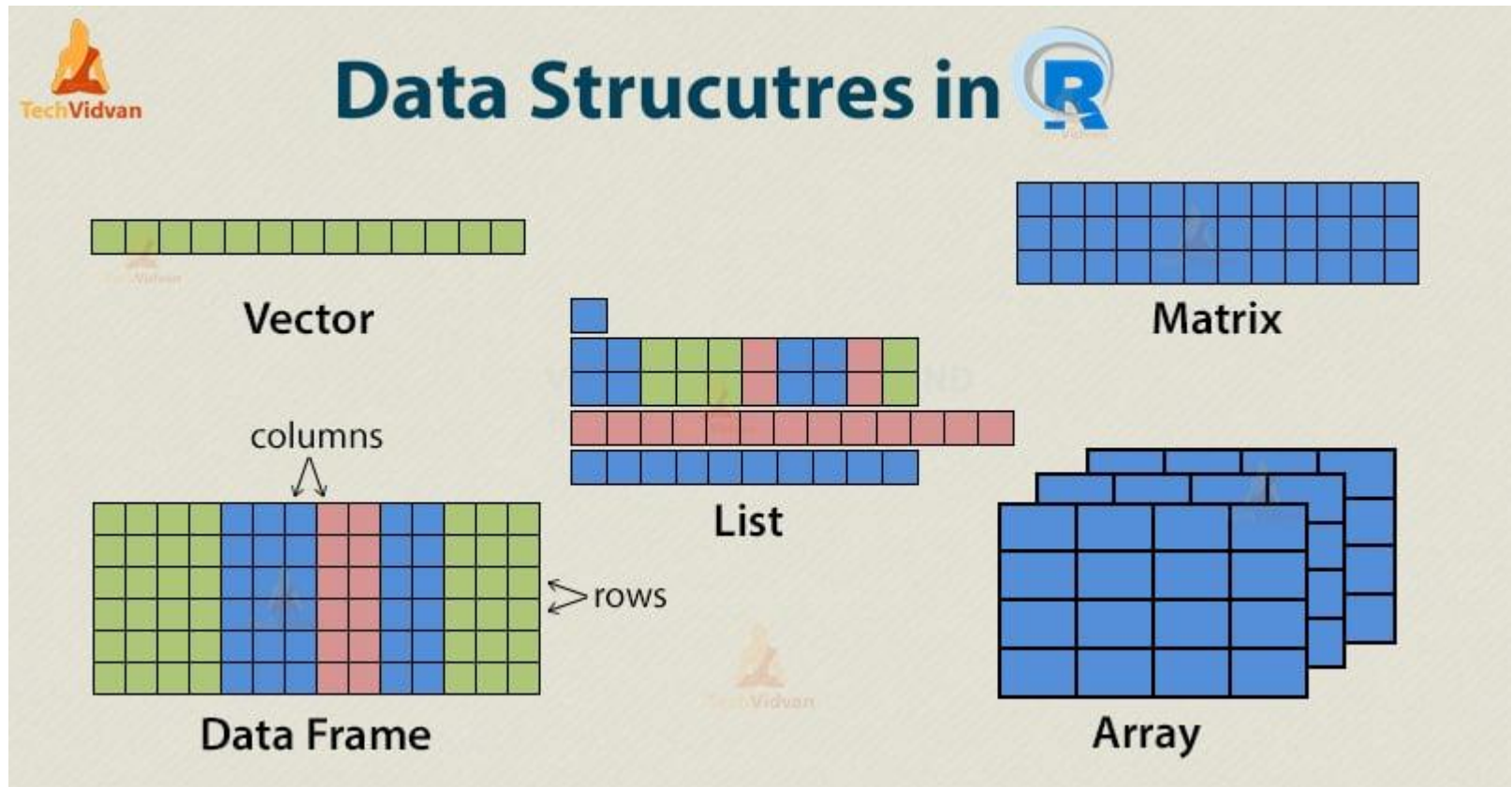
- 설치한 패키지 불러오기

```
library(ggplot2)
```



# R 자료 구조

스칼라, 벡터(1차원), 매트릭스(2차원), 배열(n차원), 리스트, 데이터프레임



## 다양한 자료형 확인 함수 기초

함수	설명	반환값 예시	사용 목적	예시 코드	출력 결과 (콘솔 출력 기준)
<b>str()</b>	객체의 구조를 요약해서 출력	"num [1:3] 1 2 3"	구조적 요약 (디버깅용)	str(c(1, 2, 3))	num [1:3] 1 2 3
<b>class()</b>	객체의 클래스 (S3 기준) 확인	"numeric"	객체 분류 및 메서드 선택용	class(c(1, 2, 3))	"numeric"
<b>typeof()</b>	내부 저장 타입 확인	"double"	메모리 표현 방식 분석	typeof(c(1, 2, 3))	"double"
<b>mode()</b>	일반적 연산 모드 확인	"numeric"	계산 가능한 유형 기준	mode(c(1, 2, 3))	"numeric"

# 다양한 자료형 확인 함수

## 다양한 객체 유형 비교표

객체 유형	예시 코드	str() 결과 요약	class()	typeof()	mode()
숫자 벡터	x <- c(1, 2, 3)	num [1:3] 1 2 3	"numeric"	"double"	"numeric"
정수 벡터	x <- c(1L, 2L)	int [1:2] 1 2	"integer"	"integer"	"numeric"
문자 벡터	x <- c("a", "b")	chr [1:2] "a" "b"	"character"	"character"	"character"
논리 벡터	x <- c(TRUE, FALSE)	logi [1:2] TRUE FALSE	"logical"	"logical"	"logical"
리스트	x <- list(a=1, b="x")	List of 2 \$ a: num 1 \$ b: chr "x"	"list"	"list"	"list"
행렬 (matrix)	x <- matrix(1:4, 2, 2)	int [1:2, 1:2] 1 3 2 4	"matrix"	"integer"	"numeric"
배열 (array)	x <- array(1:8, dim=c(2,2,2))	int [1:2, 1:2, 1:2] , , 1 1 3 2 4 ...	"array"	"integer"	"numeric"
데이터프레임	x <- data.frame(a=1:2, b=c("x", "y"))	data.frame: 2 obs. of 2 variables: \$ a: int 1 2 \$ b: chr "x" "y"	"data.frame"	"list"	"list"
팩터	x <- factor(c("A", "B"))	Factor w/ 2 levels "A","B": 1 2	"factor"	"integer"	"numeric"
함수	x <- function(a) a + 1	function (a)	"function"	"closure"	"function"
테이블 (table)	x <- table(c("A", "B", "A"))	'table' int [1:2(1d)] 2 1 - attr(*, "dimnames")=List of 1	"table"	"integer"	"numeric"

# 다양한 자료형 확인

함수 mode() class() typeof()

객체	예시	mode 함수 적용	class 함수 적용	typeof 함수
숫자	3	numeric	numeric	double
문자열	"hi"	character	character	character
숫자벡터	c(1,2,3)	numeric	numeric	double
문자열벡터	c("hi","hello")	character	character	character
Factor	factor(c("A","A","B"))	numeric	factor	integer
리스트	list(c(1,2),c("A"))	list	list	list
데이터프레임	data.frame(c(1,2))	list	data.frame	list
행렬	matrix(c(12,3,4),2,2)	numeric	matrix	double
배열	array(c(1,2,3,4),dim=c(2,2,3))	numeric	array	integer
테이블	table(c(1,1,2,3))	numeric	table	integer
함수	mean	function	function	closure

## 여기서 잠깐! R과 R스튜디오 설치 시 주의 사항

- R과 R스튜디오 설치 시 보통 문제없이 설치되지만 가끔 에러가 발생함.

이러한 에러를 방지하기 위해 다음 사항들을 사전에 점검.

- ① 현재 로그인한 윈도우의 계정 이름은 한글이 아니어야 함.  
한글 계정명을 사용하는 경우는 영문 계정을 새로 만들고,  
영문 계정으로 로그인 후 설치 작업을 해야 함.
- ② R이나 R스튜디오를 설치하는 폴더의 경로에 한글이 포함되지 않도록 함.
- ③ 원 드라이브나 구글 드라이브와 같이 클라우드와 연동된 폴더에 설치하는 것은 피하기.