

강의교안 이용 안내

- 본 강의교안의 저작권은 이윤환과 한빛아카데미(주)에 있습니다.
- 이 자료를 무단으로 전제하거나 배포할 경우 저작권법 136조에 의거하여 벌금에 처할 수 있고 이를 병과(併科)할 수도 있습니다.



제작자의 이야기

처음으로 교재를 제작하고 강의교안을 제작하다 보니 많은 오타자와 부족한 점이 눈에 계속 들어옵니다.

이런 부족함을 조금이라도 만회하고자 강의교안에 조금 더 보충을 하였습니다. 교재에 없는 내용이 있더라도 이해해 주세요. (또한 슬라이드 노트에 몇가지 전해드리는 글이 있는 슬라이드도 있습니다.)

그리고, R과 R Studio는 계속해서 업데이트가 이뤄져 교재와 다른 곳이 있을 수 있음을 이해해 주세요.

마지막으로 블로그 randstat.tistory.com 도 종종 방문해주셨으면 합니다.

늘 건강과 행복이 함께하시길 바라오며, 부족하고 잘못된 것은 openx3@tistory.com 으로 가르침 주세요.

지은이 이윤환 드림



제대로 알고 쓰는

R 통계분석

CHAPTER 01

통계학과 R의 시작

Contents

1.1

통계학

- 통계학이란
- 모집단과 표본, 그리고 기본원리
- 통계에서의 자료

1.2

R과 R Studio

- R이란
- R 기초

A

R과 R Studio 설치

- R 설치하기
- R Studio 설치하기



01. 통계학

: 자료를 통해 각종 현상을 밝히는 학문

1. 통계학의 배경지식과 통계학의 기본원리에 대해 학습한다.
2. 통계학에서 다루는 자료에 대해 알아본다.

통계학이란

• 크림전쟁



오스만제국의 쇠락과
러시아 제국의 남하정책

러시아 제국 견제를 위해
영국, 프랑스 등의 충돌
-> 사실상의 1차 세계대전

[그림 1-1] 흑해 주변¹, 붉은 타원이 크림반도이다.

그림 출처 https://commons.wikimedia.org/wiki/File:Black_Sea_map.png by NormanEinstein (CC BY-SA 3.0)

통계학이란



플로렌스 나이팅게일
(Florence Nightingale, 1820~1910)

- 백의의 천사

“전투에 의한 사망자보다,
병원 환경에 의한 사망자가 더 많다”

크림전쟁에서 나이팅게일

- 정부의 요청으로 야전병원으로 파견
- 병원 행정의 정상화
- 환자에 대한 정확한 기록과 관리
 - 위생의 중요성 강조와 설득(어떻게?)
 - 병원 내 사망률을 획기적으로 줄임
 - 위생의 중요성을 사회에 알림

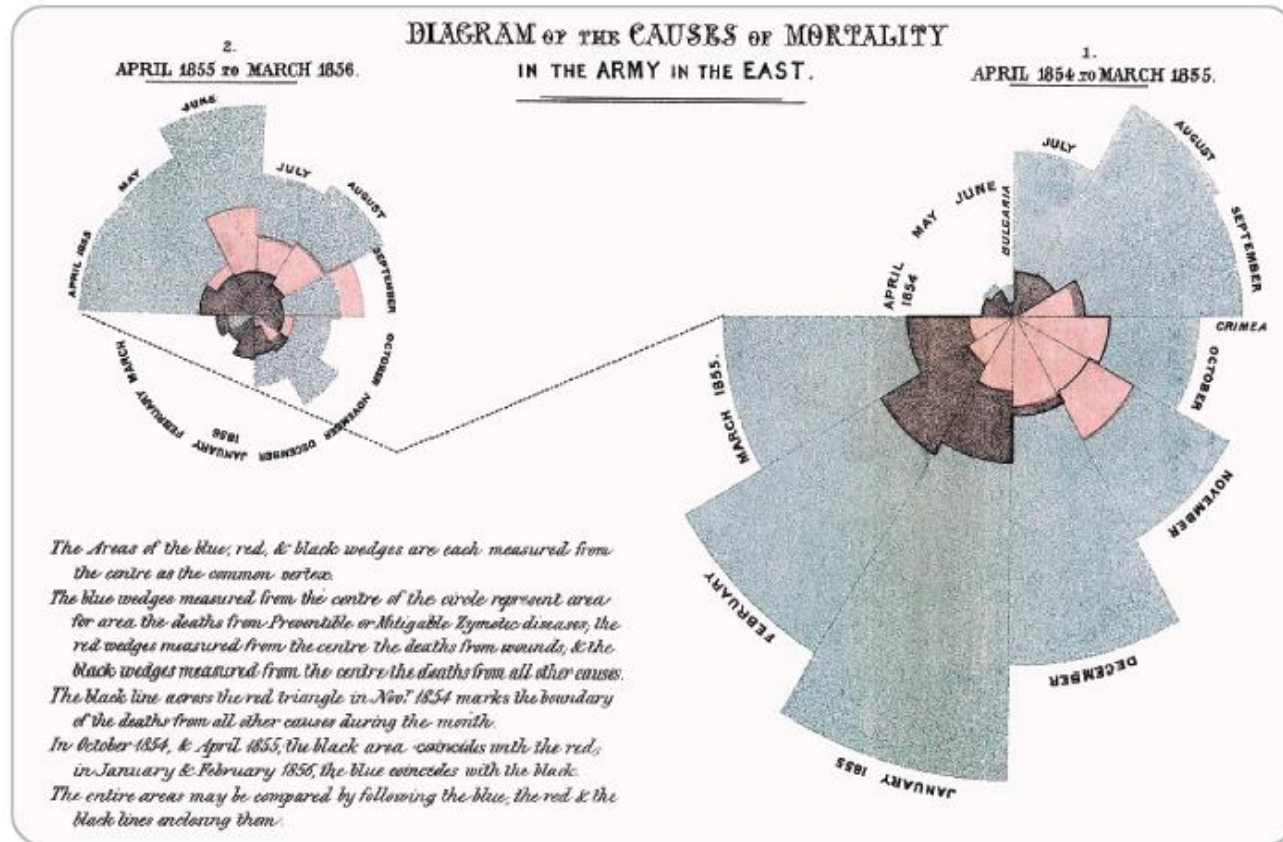
통계학이란

• 통계학자로서의 나이팅게일

- ▣ 자료를 수집하고 요약하여 설득의 도구로 자료와 통계를 사용
- ▣ 자료를 숫자로만 제시하는 것이 아니라 도표를 활용하여 변화를 누구나 쉽게 알 수 있도록 함 : 시각화의 선구자.
- ▣ 나이팅게일의 **장미도표**(Rose diagram, coxombs)
 - ① 원을 중심각 30도씩 12조각으로 나눈 후 각 조각을 월로 나타냅니다.
 - ② 각 조각별로 사망원인을 나타내는 세 개의 썰기(wedge)를 겹쳐놓습니다.
사망원인에 따라 파란색은 질병, 빨간색은 부상, 검은색은 기타 이유를 나타냅니다.
 - ③ 해당 월별 사망자 수를 각 썰기의 넓이로 합니다. 면적이 넓은 썰기를 뒤에 배치하여 작은 면적을 갖는 썰기를 가리지 않도록 합니다.
 - ④ 나이팅게일의 활동(위생환경 개선) 이전 1년과 이후 1년으로 두 장을 그린 후, 함께 배치하여 비교할 수 있도록 합니다.

통계학이란

- 나이팅게일의 장미도표(Rose diagram, coxombs)



[그림 1-2] 나이팅게일의 장미 도표

그림 출처 <https://commons.wikimedia.org/wiki/File:Nightingale-mortality.jpg> (Public domain)

통계학이란

- **통계학(Statistics, 統計學)**

- 사전적 정의(위키피디아)

- 수량적 비교를 기초로 하여, 많은 사실을 통계적으로 관찰하고 처리하는 방법을 연구하는 학문

- R. A. Fisher의 통계학

- 관찰자료에 수학적 원리를 적용하는 응용수학의 한 분야

- 다양한 사회 현상에 대해 자료를 바탕으로 신뢰할 만한 정보를 제공하여 사회 현상을 파악하게 하는 학문

- 보다 효율적인 정보 전달 방법을 연구하는 분야 또한 중요한 통계학의 분야

- 연구대상 : R. A. Fisher

- 모집단, 변동량, 자료축약방법

모집단과 표본, 그리고 기본원리

- 모집단과 표본

- 모집단

- ‘우리가 알고자 하는 대상 전체’
 - 조사 대상의 범위
 - 전수조사 : 모집단 전체를 조사하는 방법

- 표본

- ‘모집단으로부터 조사하기 위해 선택된 조사대상’
 - 모집단 전체를 조사하는 것이 불가능하거나, 수류탄과 같이 조사하면 사라지는 특성을 가진 조사대상, 시간적/공간적 제약이 있을시 모집단을 잘 대표할 수 있는 조사대상으로 실제 조사대상이 됨
 - 표본조사 : 표본을 조사대상으로 조사하는 방법

모집단과 표본, 그리고 기본원리

- 1936년 리터러리 다이제스트의 미국대선 여론조사에서 다음에 해당하는 대상을 이야기해 봅시다.
 - 모집단
 - 표본

모집단과 표본, 그리고 기본원리

- 통계학의 기본원리

- “표본추출을

공평하게 하면

(표본의 특성이) 모집단의 특성과
잘 닮는다”

통계에서의 자료

- 통계에서의 자료

- 관찰대상

- 조사를 위해 자료를 수집하는 대상으로 표로 자료들을 나타낼 때 행요소에 해당

- 속성(혹은 변수)

- 조사에서 값을 측정하는 각 관찰대상이 갖고 있는 개별 특성
 - 일반적으로 하나의 속성(변수)가 갖는 값들을 자료라고 함
 - 다양한 정의가 있으며, 개별 특성에 대해 하나의 속성이 아닌 여러 개의 속성이 필요할 경우도 있음(예를 들어 건강상태를 하나의 속성으로 볼 때 이를 위해 체중, 혈당, 혈압 등 여러 개의 측정대상이 존재)

- Dataset

- 관찰대상과 각 관찰대상으로 관찰한 속성들의 값을 표의 형태로 저장 혹은 나타낸 자료들의 모임

통계에서의 자료

Dataset : 관찰대상으로 부터 속성들을 측정된 자료들의 모임

[표 1-1] 2010년 인구주택총조사 중 일부

ID	성별	나이	가구주와의 관계	학력	출생아 수
1	여자	68	가구주의 배우자	초등학교	3
2	여자	29	가구주의 배우자	초등학교	0
3	여자	7	자녀	초등학교	결측
4	여자	3	자녀	안 받았음	결측
5	남자	26	자녀	중학교	결측
6	여자	52	가구주의 배우자	초등학교	2
7	여자	62	가구주의 배우자	고등학교	1
8	여자	10	자녀	초등학교	결측
9	남자	58	가구주	중학교	결측

행 : 관찰대상(개체)

열 : 속성, 변수 (측정, 자료)

통계에서의 자료

- 통계학의 자료

- 질적변수 : 자료의 값이 아니라 각 값이 나타내는(가리키는) 의미를 나타내는 값으로 이뤄진 변수
 - 명목형 자료 : 분류를 목적으로 사용하는 자료
 - 순위형 자료 : 순서로 분류할 때 사용하는 자료
- 양적변수 : 측정 대상의 크기를 나타내는 값 자체를 사용하는 값으로 이뤄진 변수
 - 이산형 자료 : 셀 수 있는 값을 나타낼 때 사용하는 자료
 - 연속형 자료 : 측정 대상의 크기 변화가 연속적일 때 사용하는 자료

통계에서의 자료

• 통계학의 자료 : 예시

[표 1-2] 통계에서의 자료

변수 유형	자료 유형	인구주택총조사 자료	예
질적변수	명목형 자료	성별, 배우자와의 관계	거주지역, 혈액형 등
	순서형 자료	학력	학점, 설문문항 등
양적변수	이산형 자료	출생아 수	형제 수, 수강과목 수 등
	연속형 자료	연령	키, 몸무게 등

❖ 척도 : 자료들을 측정하기 위한 측정 도구

- 명목척도와 서열척도
 - 각각 명목형 자료, 순서형 자료에 사용하는 척도입니다.
- 등간척도와 비율척도 : 양적변수에 사용되는 척도로 사칙연산이 가능합니다.
 - 자료의 각 값이 동일한 간격으로 이뤄진 자료를 등간척도라 하고 비율척도와의 차이는 절대 영(o)점이 존재하면 비율척도라 합니다. 절대 영점은 값의 크기가 0으로 결측과 구별해 주시기 바랍니다.
 - 기온에서 0도는 절대영점일까 생각해 봅시다.



부록 A. R과 R Studio 설치하기

: 수업을 진행하는데 필수적인 프로그램을
을 설치합니다.

1. R을 설치한다.
2. R Studio를 설치하고 작업환경을 구축한다.

R 설치하기

- **R**

- 통계 계산과 그래프 작성을 위한 프로그래밍 언어이자 소프트웨어 환경
- 1993년 뉴질랜드 오클랜드 대학의 로스 이하카(Ross Ihaka)와 로버트 젠틀맨(Robert Gentleman)에 의해 공개
- R 코어팀에 의해 지속적으로 개발
- GPL하에 배포되어 비용 부담 없이 자유롭게 사용
- 통계소프트웨어 개발과 자료 분석에 널리 사용되고 있으며, 통계학자들뿐만 아니라 계량 연구를 하는 분야에서 폭넓게 사용

- **홈페이지**

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

R 설치하기

- **R**

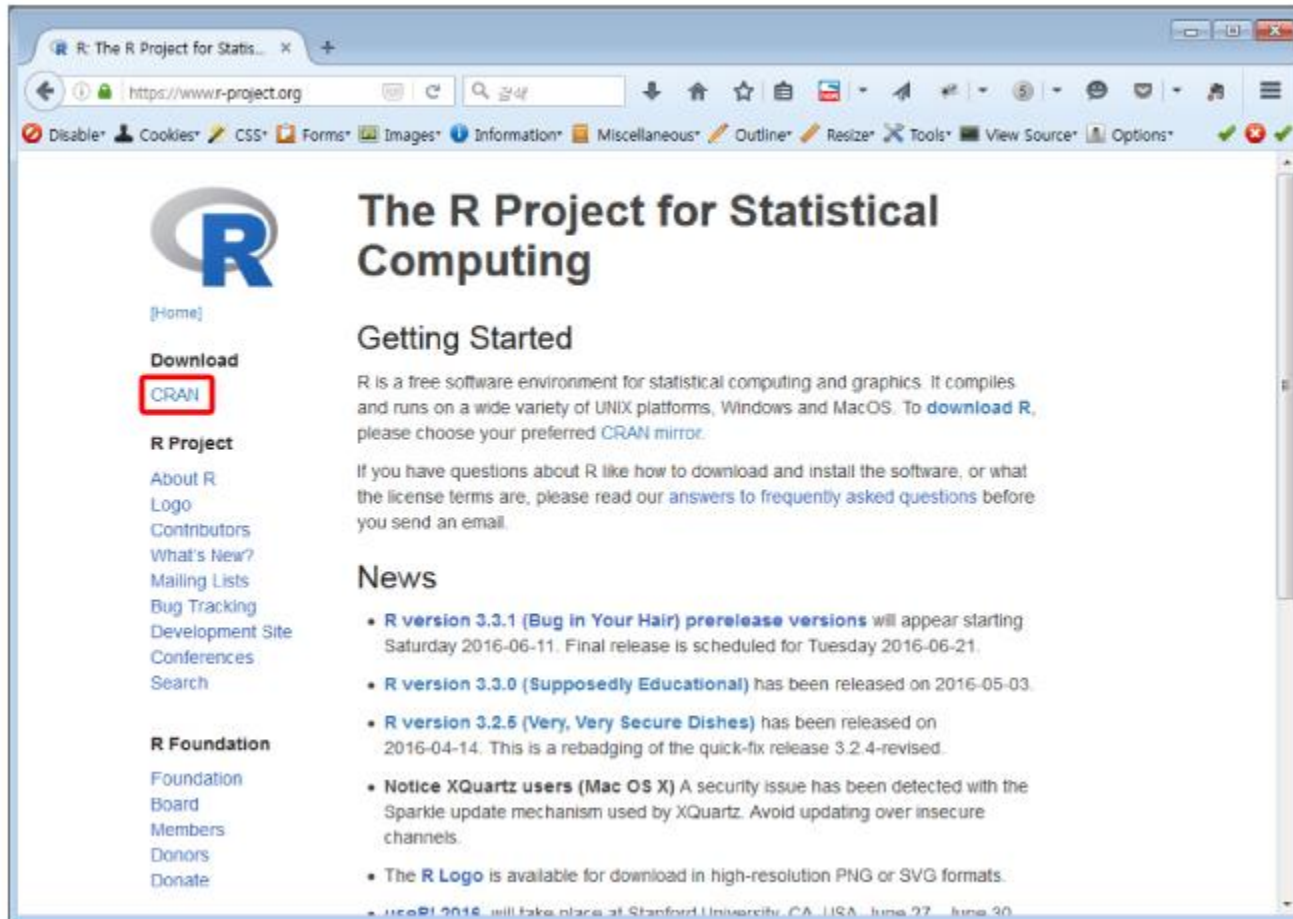
- 통계 계산과 그래프 작성을 위한 프로그래밍 언어이자 소프트웨어 환경
- 1993년 뉴질랜드 오클랜드 대학의 로스 이하카(Ross Ihaka)와 로버트 젠틀맨(Robert Gentleman)에 의해 공개
- R 코어팀에 의해 지속적으로 개발
- GPL하에 배포되어 비용 부담 없이 자유롭게 사용
- 통계소프트웨어 개발과 자료 분석에 널리 사용되고 있으며, 통계학자들뿐만 아니라 계량 연구를 하는 분야에서 폭넓게 사용

- **홈페이지**

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

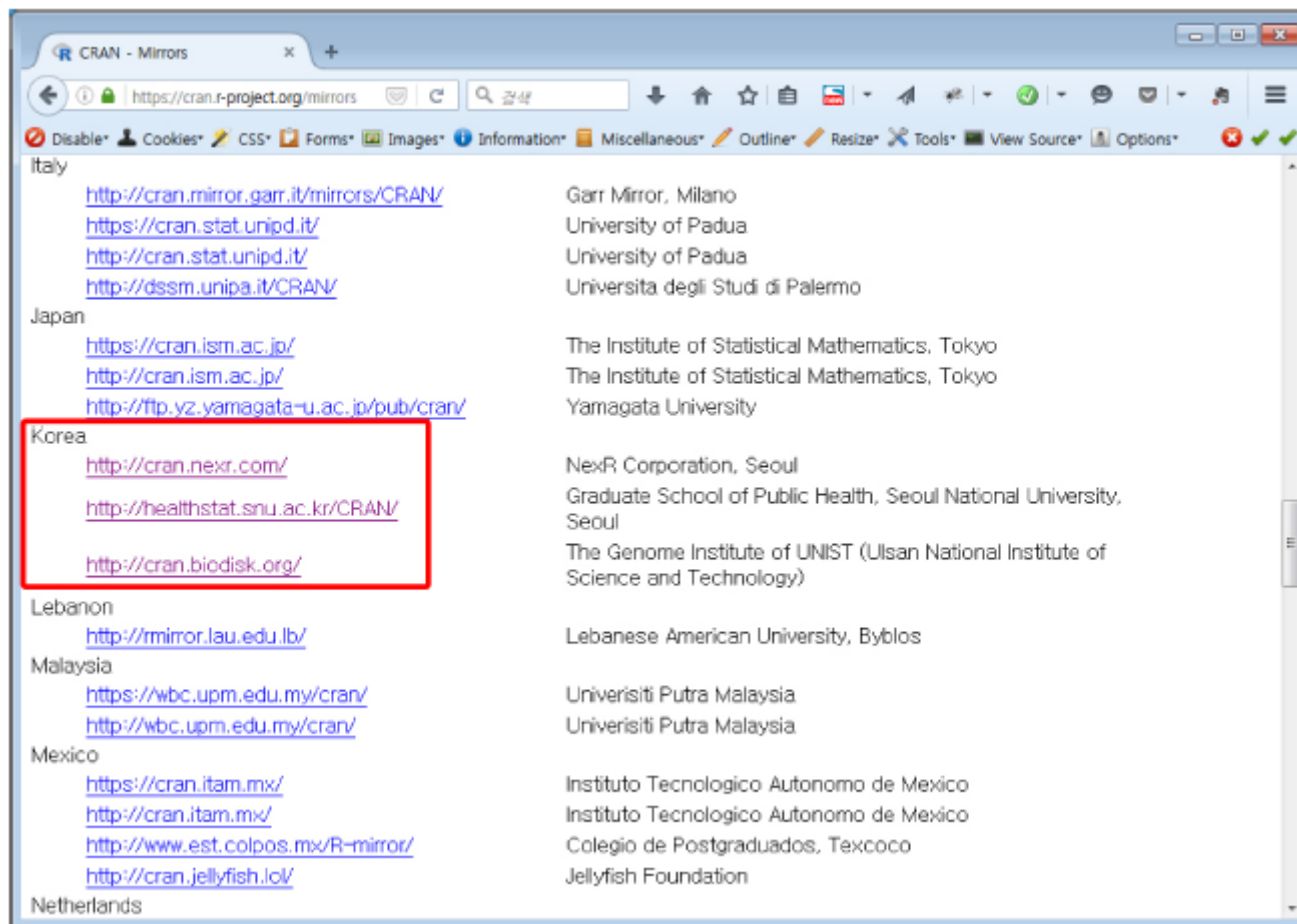
R 설치하기

- 홈페이지에서 좌측 항목 중 'Download'의 'CRAN'을 클릭합니다



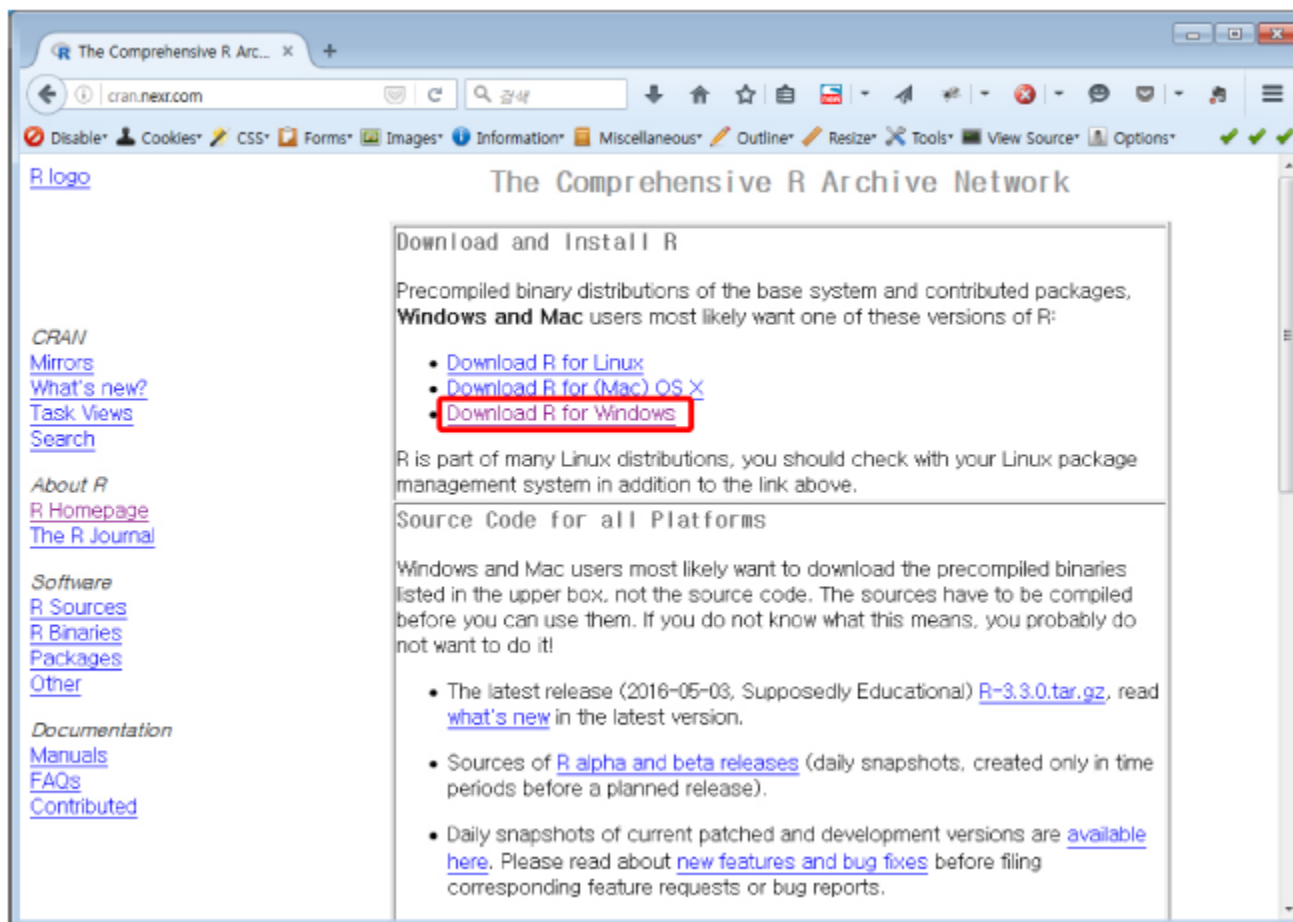
R 설치하기

- 국가별로 다운로드 받을 수 있는 곳의 링크가 나옵니다. 대한민국(Korea)의 링크 사이트 들을 찾아 세 군데 중 한 곳을 선택합니다.



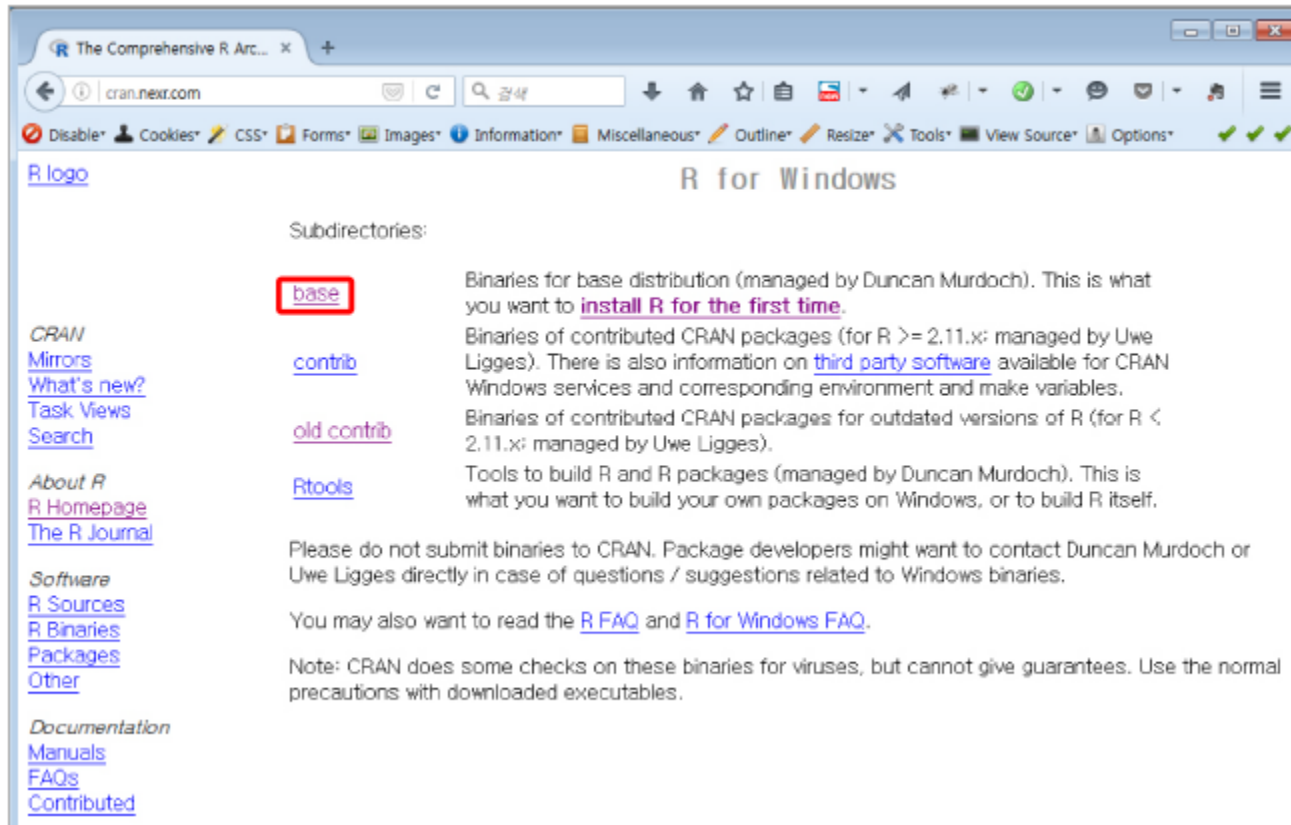
R 설치하기

- 윈도우를 기준으로 설명하므로, ‘Download R for Windows’를 클릭합니다. (Max OS X, Linux 사용자들도 크게 다르지 않습니다.)



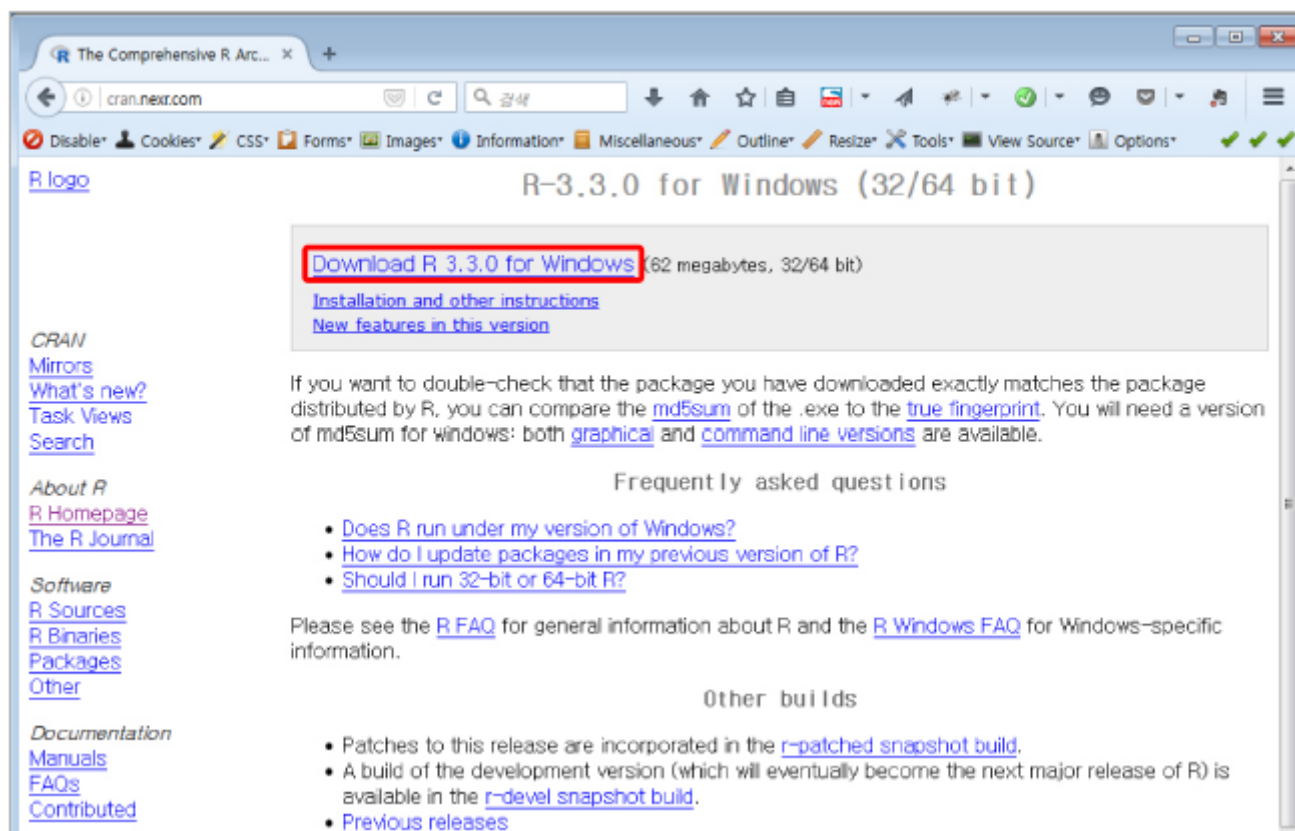
R 설치하기

- R을 설치하기 위해 여기서 'base'를 선택합니다. 나머지 링크들은 패키지 제작 및 R 환경을 위한 링크입니다. ('install R for the first time'을 클릭해도 됩니다.)



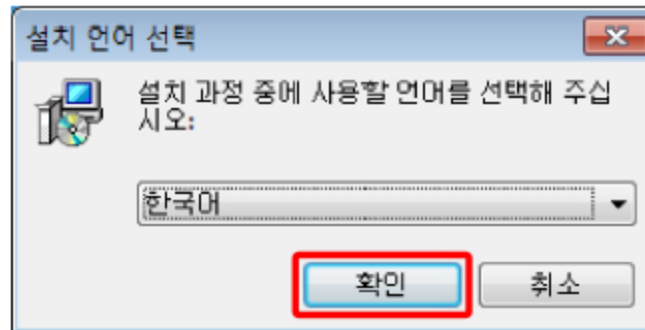
R 설치하기

- 'Download R.x.x.x for Windows'를 클릭하여 설치 파일을 다운로드합니다.



R 설치하기

- 설치과정
 - 다운로드 받은 파일을 실행하여 윈도우 설치과정을 진행합니다.

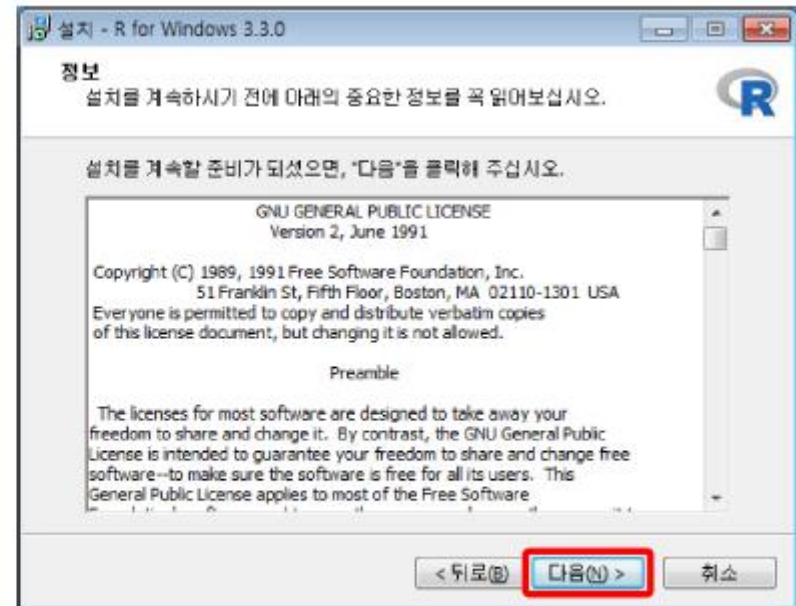


[그림 A-6] 설치 시 사용할 언어 선택

R 설치하기

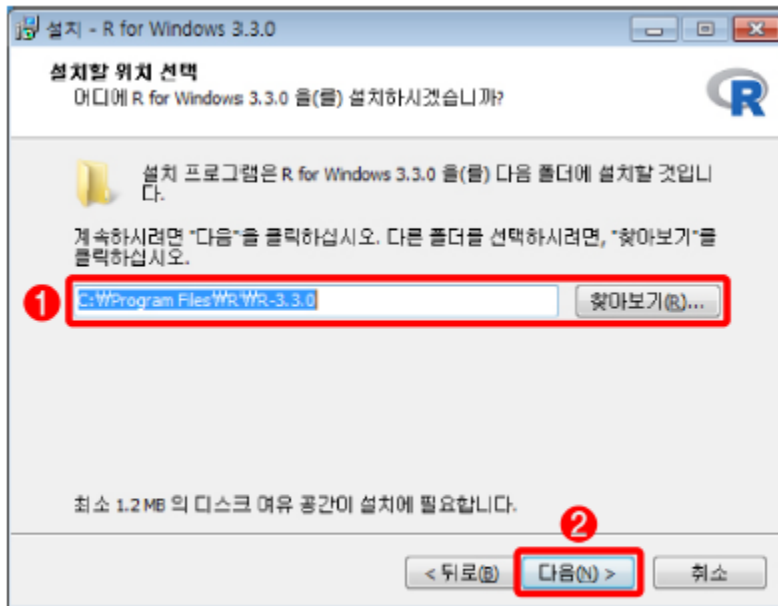


[그림 A-7] 설치 시작

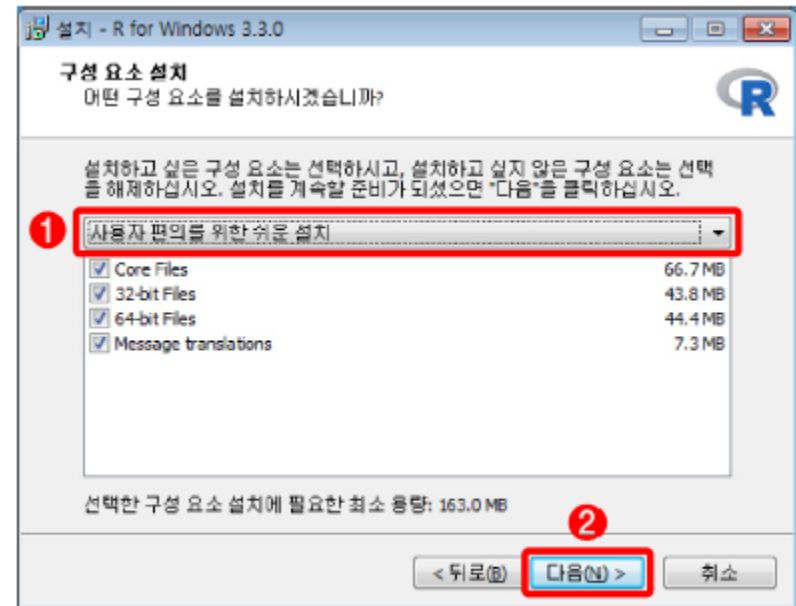


[그림 A-8] 라이선스 확인

R 설치하기

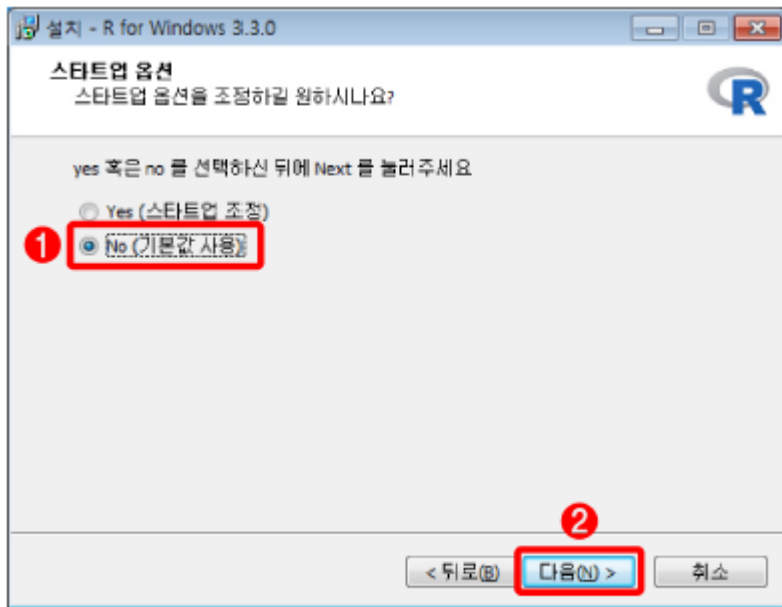


[그림 A-9] 설치 위치 확인

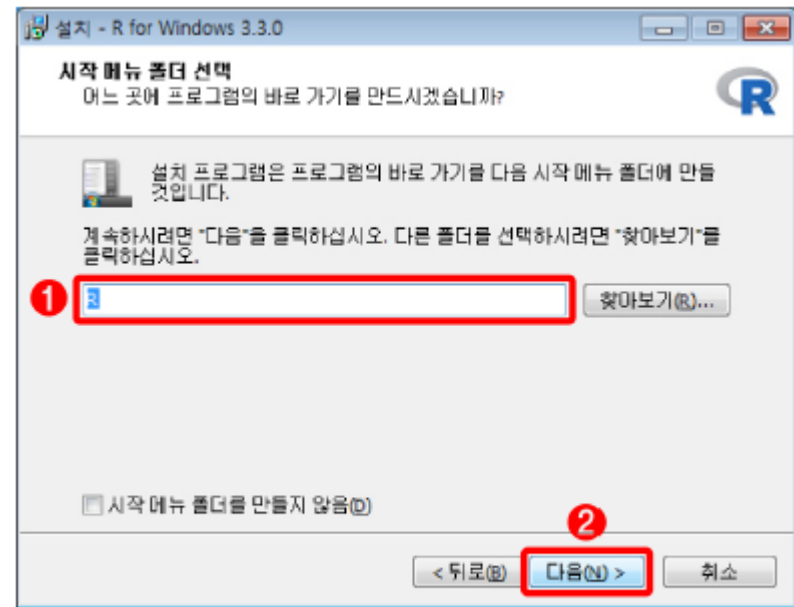


[그림 A-10] 설치 내용 확인

R 설치하기

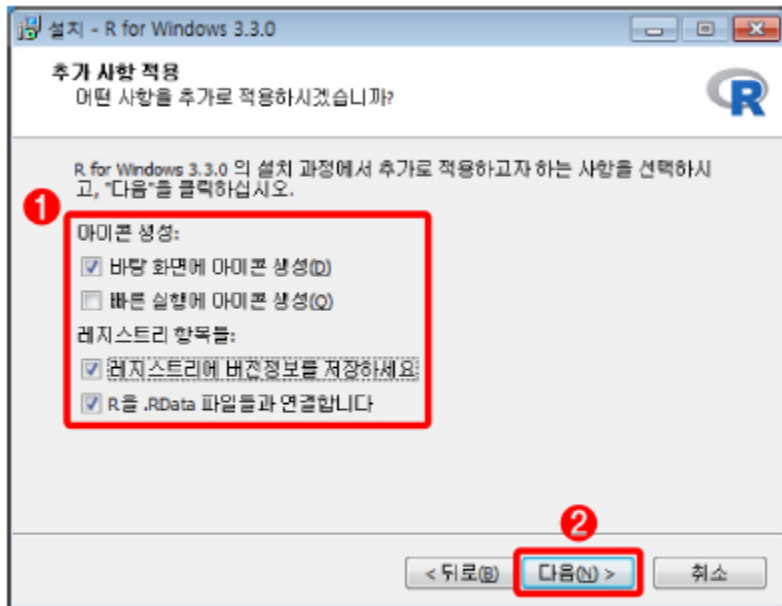


[그림 A-11] 기본 시작 옵션으로 설치

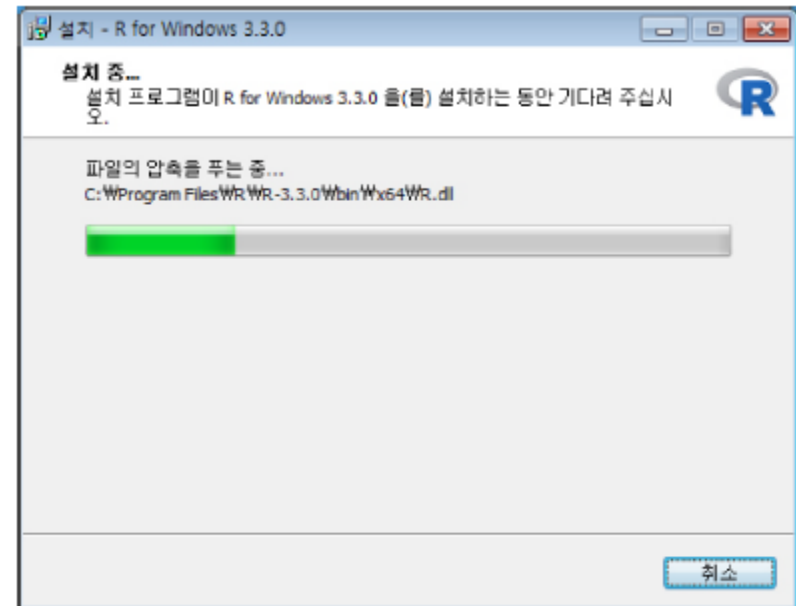


[그림 A-12] Windows의 프로그램 그룹 등록

R 설치하기

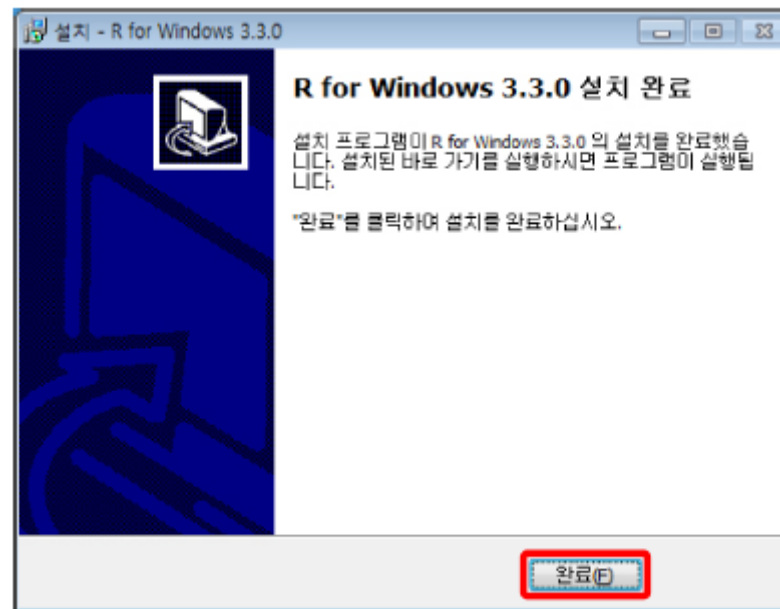


[그림 A-13] 추가 적용사항 확인



[그림 A-14] 설치 진행

R 설치하기



[그림 A-15] 설치 완료

R Studio 설치하기

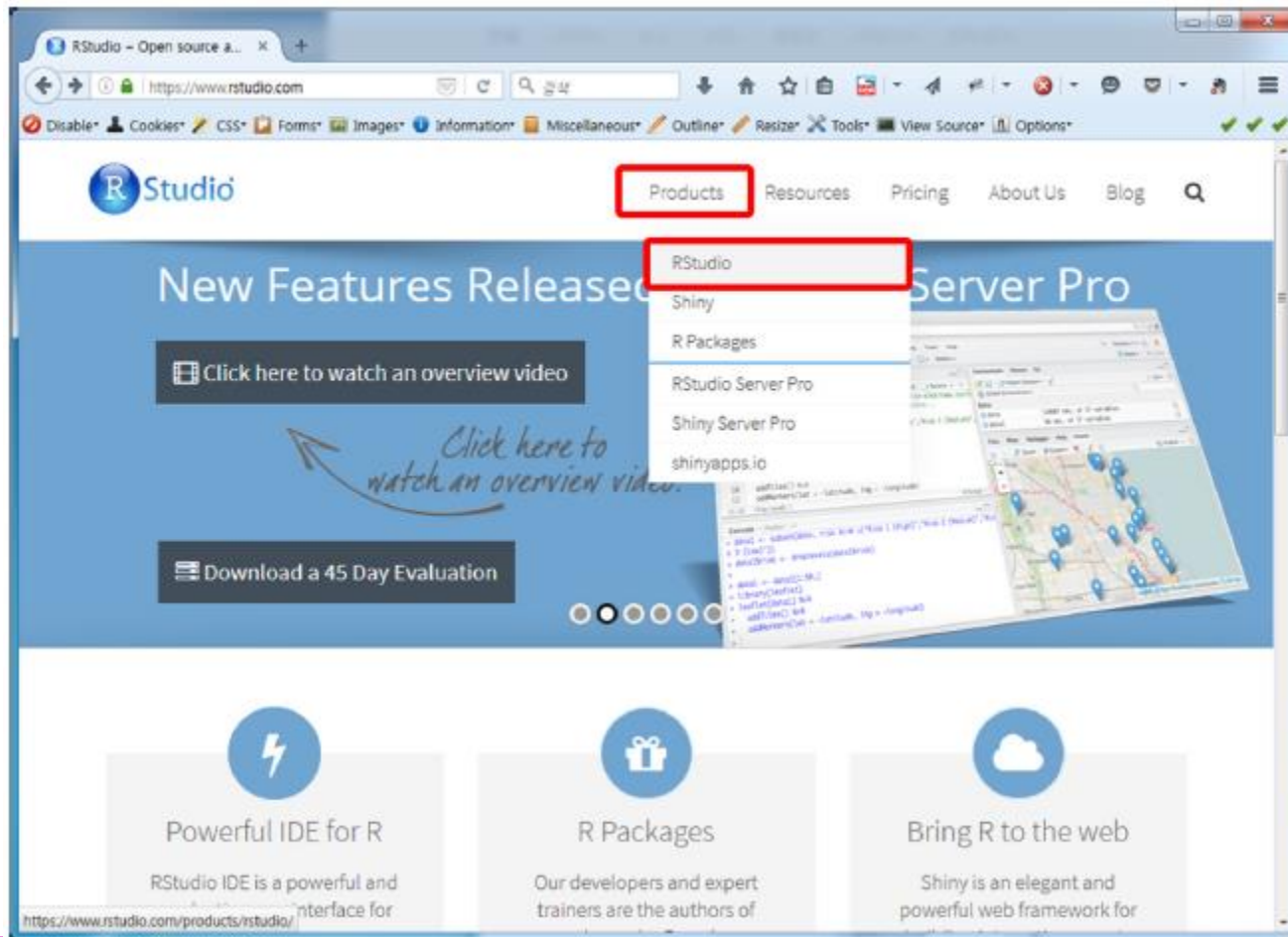
- **R Studio**

- R Studio는 R을 위한 통합개발환경(IDE)입니다.
 - 통합개발환경(IDE, Integrated Development Environment)
 - 코딩, 디버그, 컴파일, 배포 등 프로그램 개발에 관련된 모든 작업을 하나의 프로그램 안에서 처리하는 환경을 제공하는 소프트웨어이다(위키피디아, 통합개발환경)
 - R 통합개발환경을 위해 R Studio는 콘솔, 직접 코드를 실행시킬 수 있는 구문강조 (Syntax Highlighting) 기능이 있는 편집기와 그림을 그리고 코드 이력을 기록하고 코드 내 버그 찾기 기능과 작업공간 관리 기능을 수행합니다.
 - 오픈소스로 개발되어 있어 무료로 사용할 수 있으며, 일정 비용을 지불하고 각종 지원을 받을 수 있는 유료버전이 있습니다.
 - 홈페이지 : <https://www.R-Studio.com>

R Studio 설치하기

- 설치 파일을 찾아갑시다

- 먼저 홈페이지를 방문하여 상단 메뉴의 Products의 R Studio를 선택합니다.







R Studio 설치하기

- ▣ 개인 사용자의, ‘Desktop’과 서버 설치용 ‘Server’ 두 가지가 있습니다.
- ▣ ‘R Studio Desktop(Free License)’을 선택합니다.

Choose your version of RStudio

Products Resources Pricing About Us Blog Q

	RStudio Desktop (Free License)	RStudio Desktop (Commercial License)	RStudio Server (Free License)	RStudio Server Pro (Commercial License)
Integrated Development Environment for R	✓	✓	✓	✓
Priority support		✓		✓
Access via Web Browser			✓	✓
Enterprise Security and Access Controls				✓
Project Sharing				✓
Access to Multiple Versions of R				✓
Multiple Concurrent Sessions				✓
Administrative Dashboard				✓
Load Balancing and Resource Management				✓
License	AGPL	Commercial	AGPL	Commercial
	 DOWNLOAD	 DOWNLOAD	 DOWNLOAD	 DOWNLOAD

R Studio 설치하기

- 윈도우 하에서 설치를 진행합니다. (다른 운영체제도 크게 다를 바 없습니다.)
- Installers 목록 중 'R Studio-x.yy.zzz-Windows Vista/7/8/10'을 클릭합니다.
 - 다운로드가 진행됩니다.

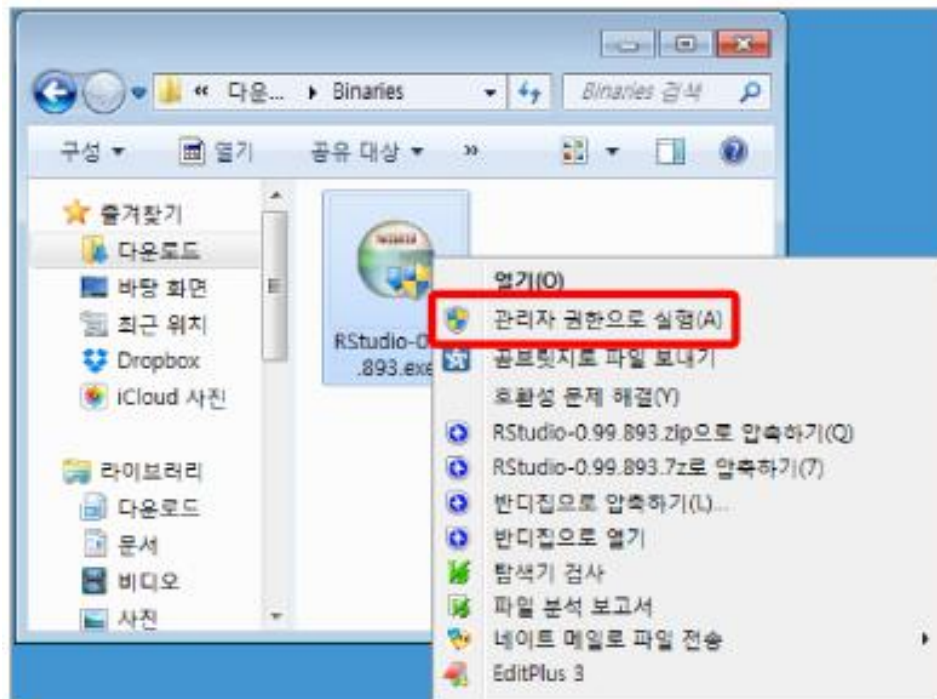
RStudio requires R 2.11.1+. If you don't already have R, download it [here](#).

Installers for Supported Platforms

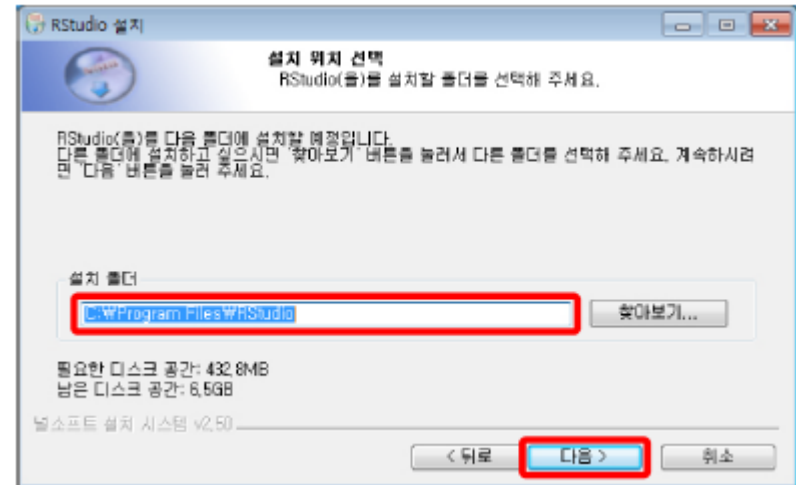
Installers	Size	Date	MD5
RStudio 0.99.903 - Windows Vista/7/8/10	77.1 MB	2016-07-18	716f28f2143c5e21f4acea5752e284f8
RStudio 0.99.903 - Mac OS X 10.6+ (64-bit)	60 MB	2016-07-18	d14a1585b5a5ac0839507b9c04d460d6
RStudio 0.99.903 - Ubuntu 12.04+/Debian 8+ (32-bit)	81.6 MB	2016-07-18	761eae80b0ba4d4cd9051a802a2c44e2
RStudio 0.99.903 - Ubuntu 12.04+/Debian 8+ (64-bit)	88.3 MB	2016-07-18	98ea59d3db00e0083d3e4053514f764d
RStudio 0.99.903 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	81 MB	2016-07-18	ce2ea1023d99175cb909def0fe66eba7
RStudio 0.99.903 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	81.9 MB	2016-07-18	152f247255e86904cf3354afbc7b3b99

R Studio 설치하기

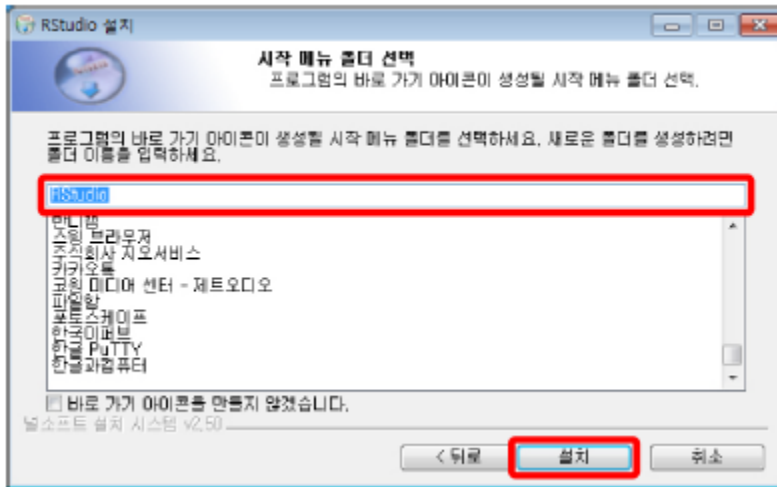
- 설치과정
 - 다운로드 받은 파일을 실행하여 윈도우 설치과정을 진행합니다.
 - 마우스 오른쪽 버튼을 클릭하여 관리자 권한으로 실행합니다.



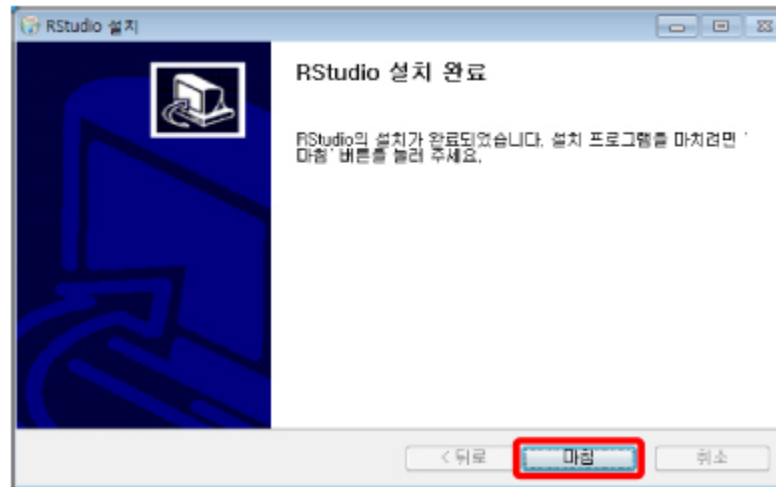
R Studio 설치하기



R Studio 설치하기



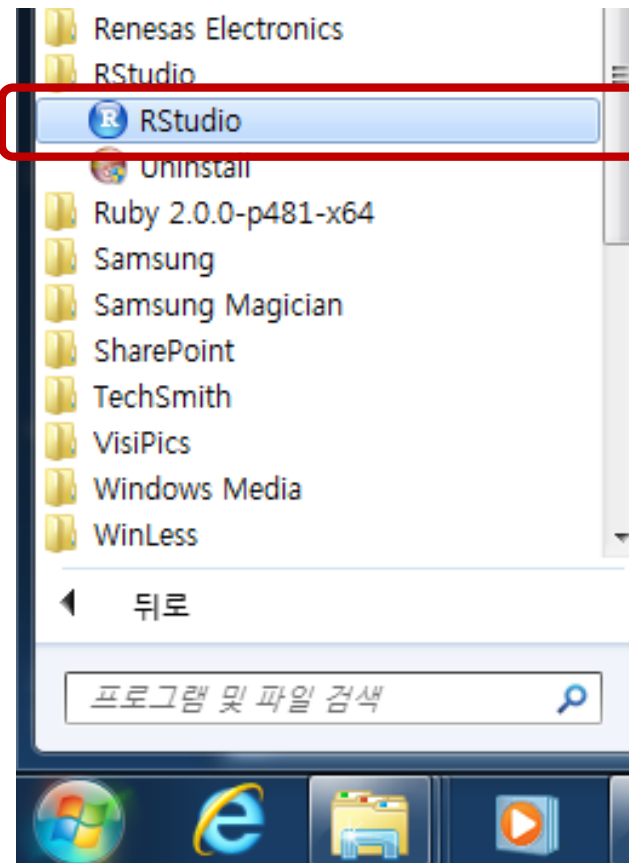
R Studio 설치하기



R Studio 실행과 기본설정 : 작업환경 구축

- **R Studio 실행**

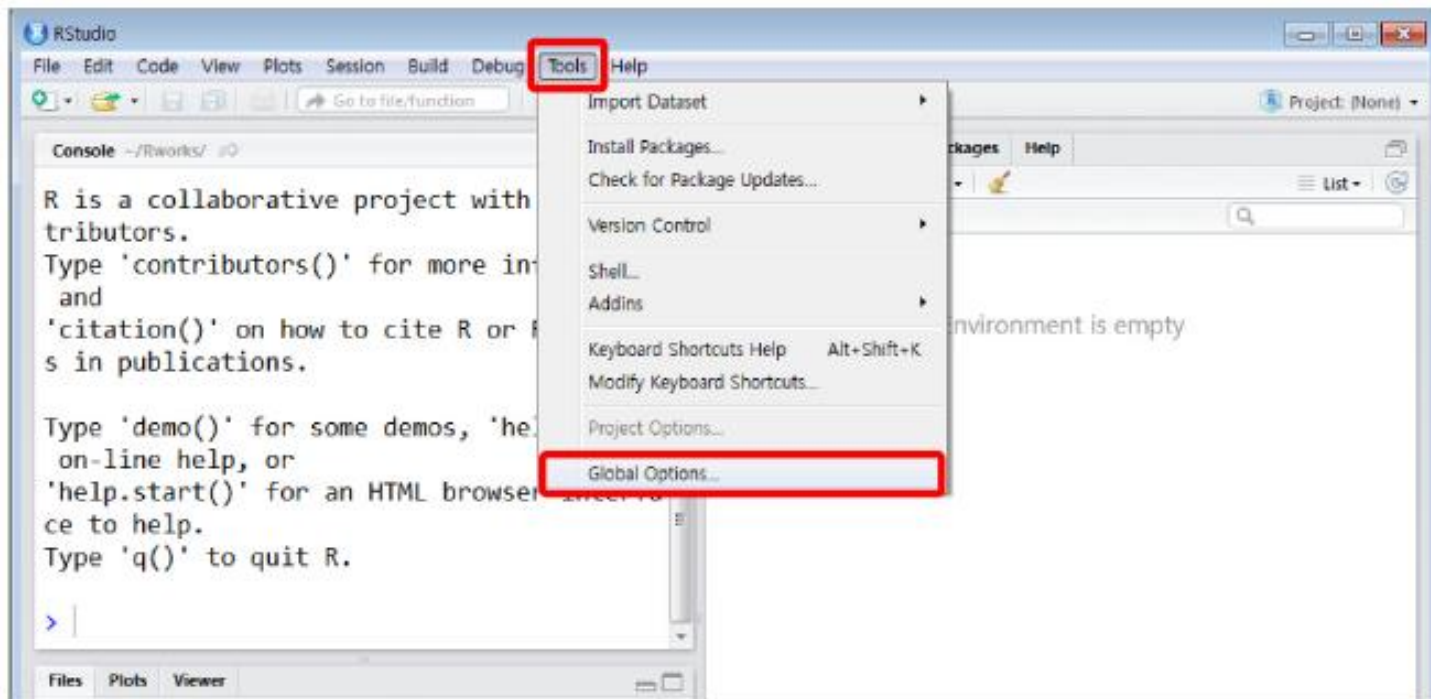
- ▣ 프로그램 그룹에서 “R Studio” 항목을 찾습니다.
 - 항목안에 “R Studio”를 클릭하여 실행합니다.



R Studio 실행과 기본설정 : 작업환경 구축

• R Studio 설정하기

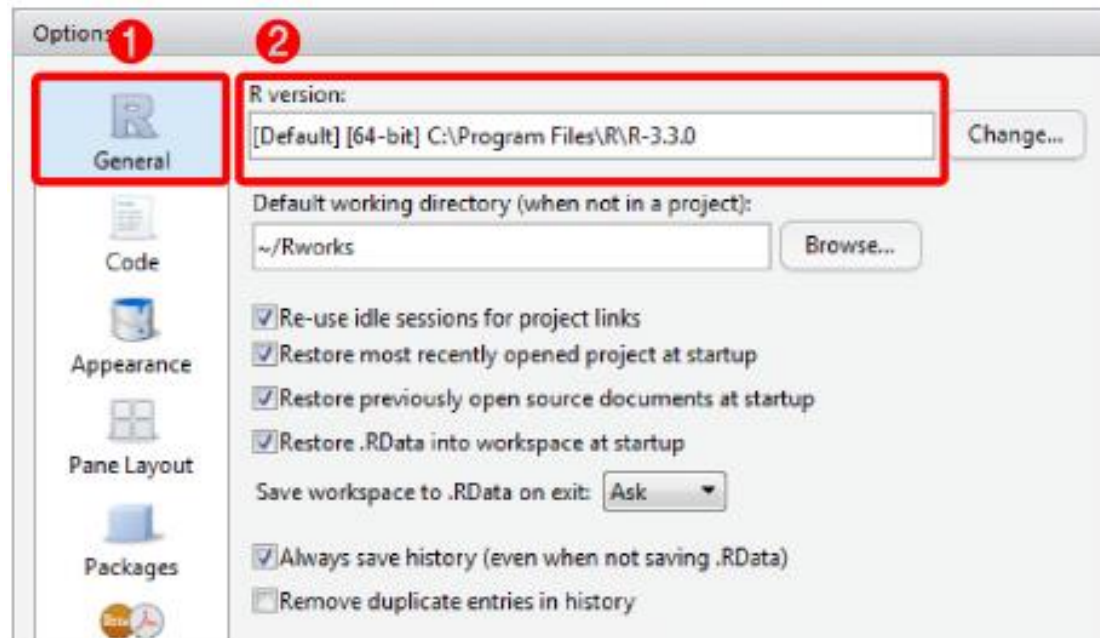
- 작업화면을 작업에 용이하게 변경해 봅시다.
- 상단의 'Tools' 메뉴에서 'Global Options'를 클릭합니다.



R Studio 실행과 기본설정 : 작업환경 구축

• R Studio 설정하기 : General

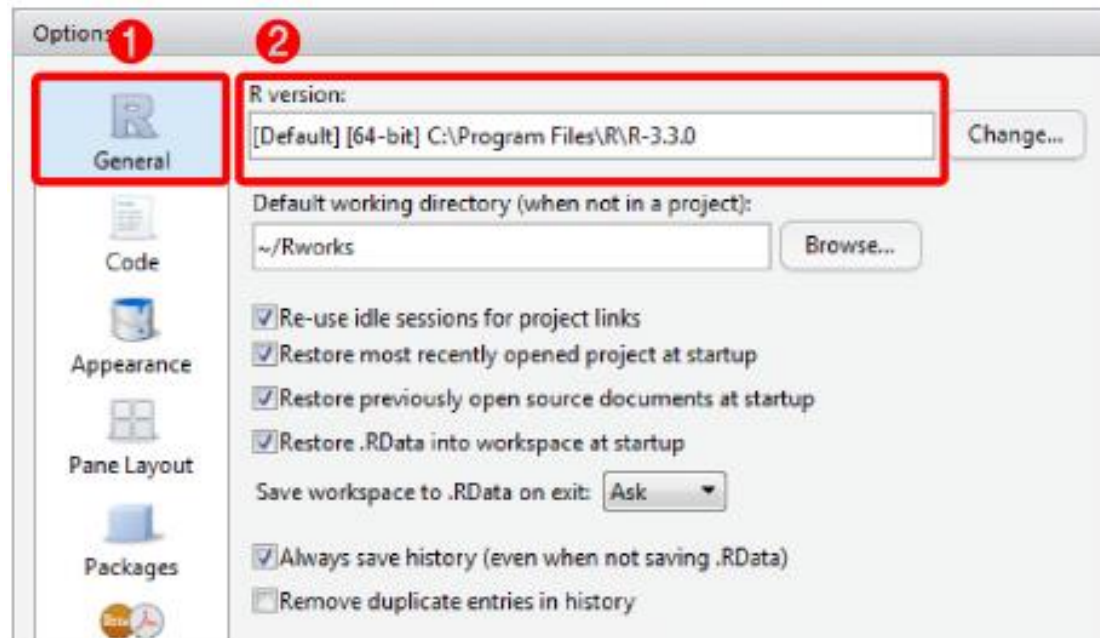
- ▣ 실행할 R 버전 확인
- ▣ 기본 시작 위치 지정 : R 버전 하단에서 “~”는 사용자의 Windows 상에서의 홈디렉토리 입니다.



R Studio 실행과 기본설정 : 작업환경 구축

• R Studio 설정하기 : Code

- 설치되는 언어 저장 방식 설정
- 상단의 “Saving” 탭을 선택하고, “Default Encoding”을 전세계에서 표준의 역할을 하는 유니코드 저장을 위해 사용하는 “**UTF-8**”로 합니다.

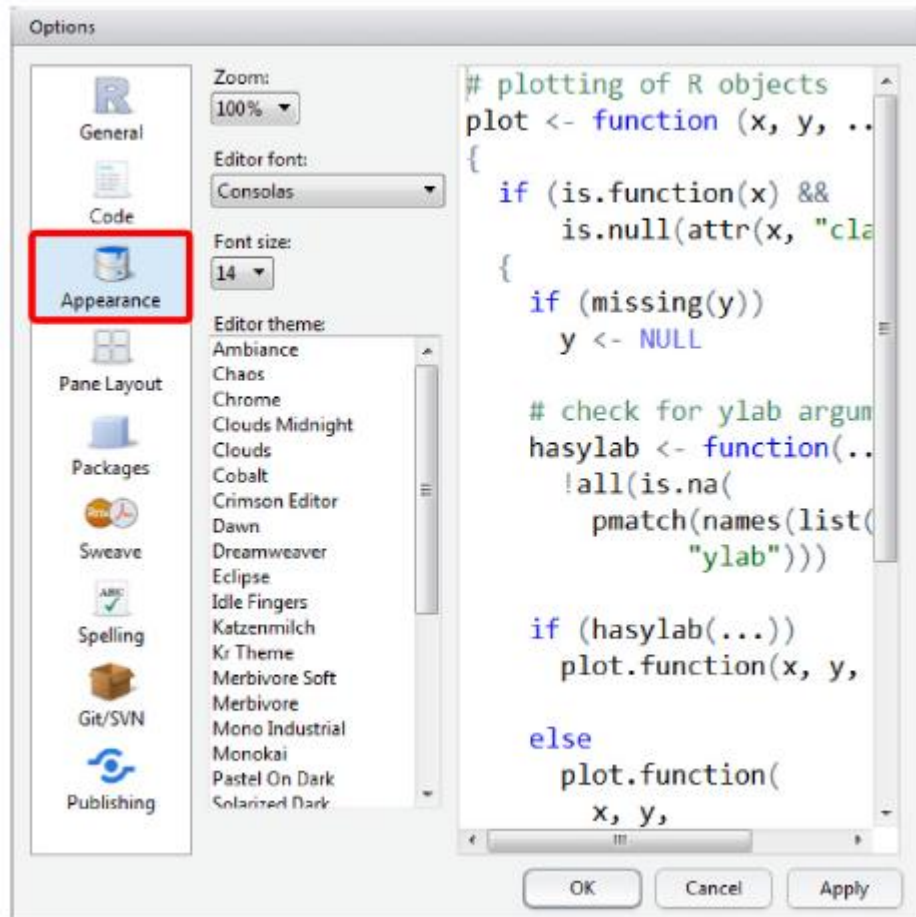


R Studio 실행과 기본설정 : 작업환경 구축

• R Studio 설정하기 :

Appearance

- 화면상의 글꼴 설정
- 개인의 취향입니다만 저는 개인적으로 “Consolas” 폰트에 크기 14 정도로 하였습니다. 여러분들이 가장 눈에 잘 들어오는 글꼴과 크기를 결정해 주세요.
(우측에 미리보기를 통해 확인할 수 있습니다.)

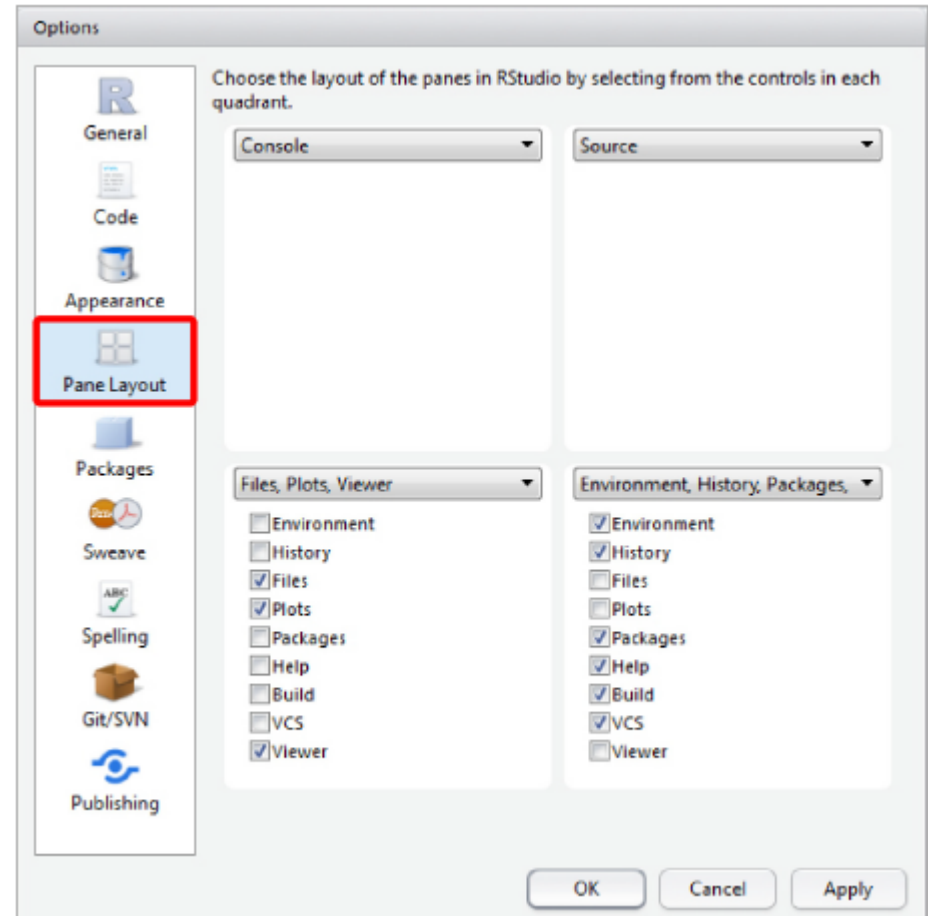


R Studio 실행과 기본설정 : 작업환경 구축

• R Studio 설정하기 :

Pane Layout

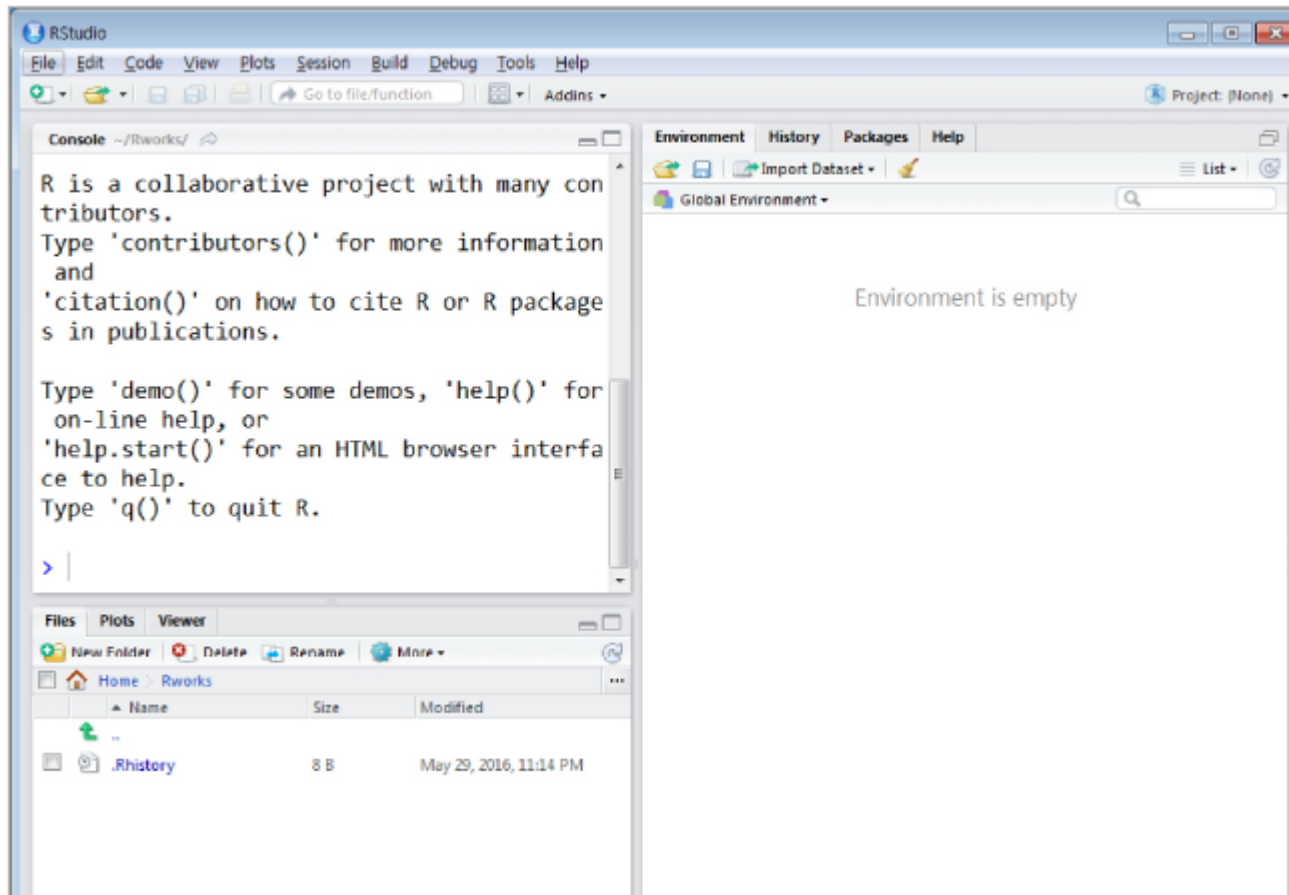
- ▣ 작업창 설정
- ▣ 글꼴과 마찬가지로 각자 맞춰 변경합니다.
- ▣ 본 교재에서는 출력은 왼쪽, 입력 등의 작업은 오른쪽을 기본으로
 - 좌측 상단 : Console
 - 좌측 하단 : Files, Plots, Viewer
 - 우측 상단 : Source
 - 우측 하단 : Help, Environment, History, Packages 등으로 지정했습니다.



R Studio 실행과 기본설정 : 작업환경 구축

- **R Studio 설정하기 : 확인**

- 앞서 기술한 바와 같이 설정하시면 눈에 띄는 변화는 Pane(작업창)의 변화와 Console 창의 글꼴 변화등이 눈에 들어옵니다.





02. R과 R Studio

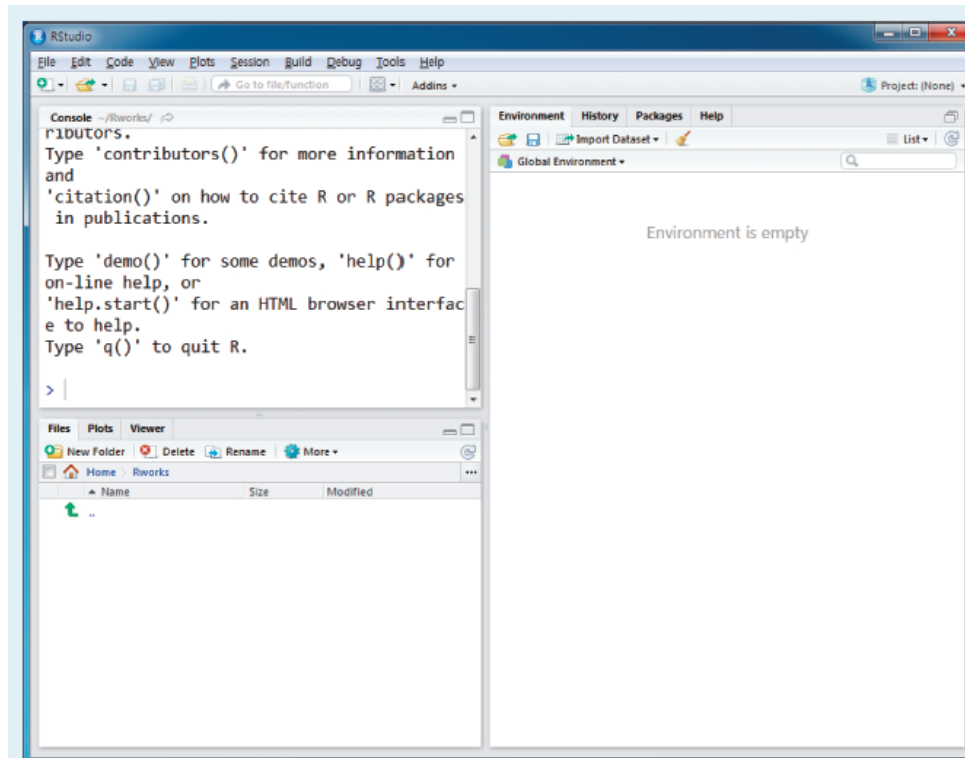
: 통계에 필요한 계산과 그림을 멋지게
처리하는 도구

1. R Studio를 통해 작업 환경을 구축한다.
2. 계산기로서의 R을 사용한다.
3. R에서 자료를 저장하는 방식에 대해 학습한다.

R Studio와 함께 R 시작하기

예제 1-1 프로젝트(Project) 생성

1장에서 사용할 내용들을 효율적으로 관리하기 위해 'Chapter01'이라는 이름으로 프로젝트를 생성합니다.



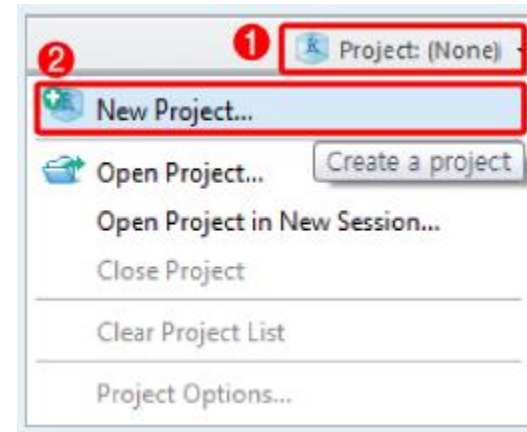
[그림 1-4] R Studio 실행화면

R Studio와 함께 R 시작하기

예제 1-1 프로젝트(Project) 생성

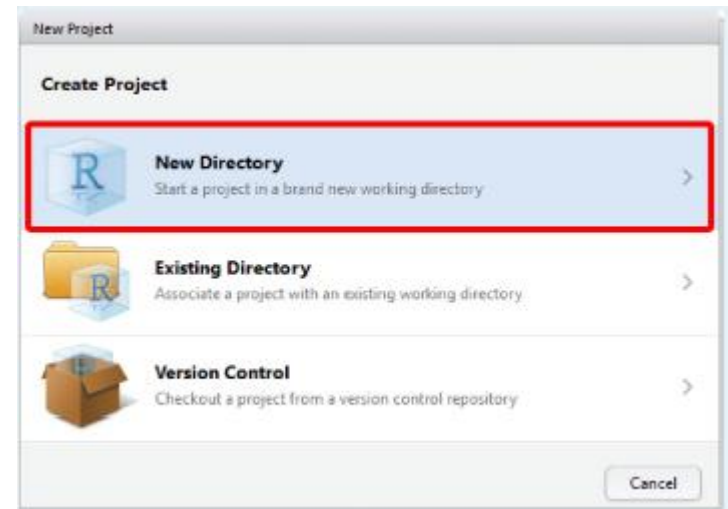
프로젝트 생성 메뉴

- ① R Studio 우측 상단의 “Project” 클릭
- ② “New Project” 클릭



프로젝트 경로 설정

- ① “New Directory”를 눌러 새로운 위치에 프로젝트를 생성합니다.

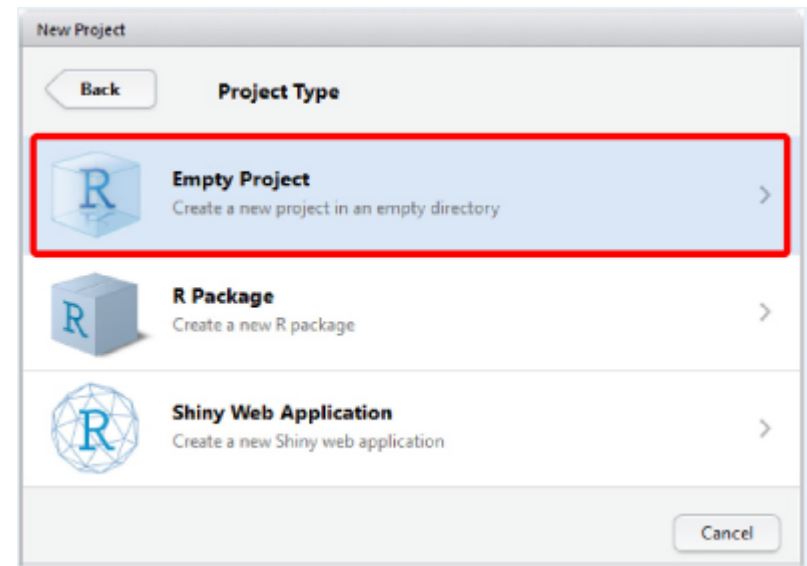


R Studio와 함께 R 시작하기

예제 1-1 프로젝트(Project) 생성

프로젝트 유형 설정

① “Empty Project”를 선택합니다.

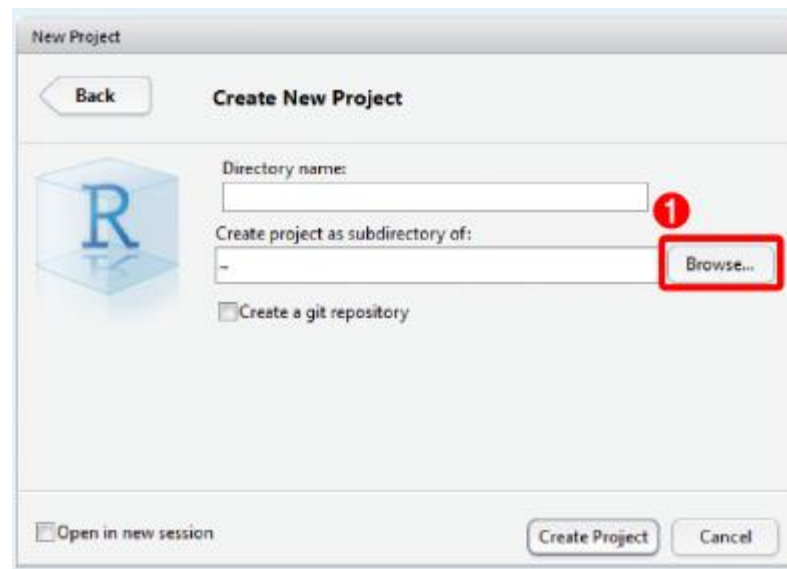


R Studio와 함께 R 시작하기

예제 1-1 프로젝트(Project) 생성

(내)문서 폴더 아래에 “StatwithR”이라는 폴더를 만듭니다.
이 폴더는 앞으로 사용할 프로젝트들의 상위 폴더가 될 것입니다.

- ① 기본 저장위치는 “~”로 이는 윈도우에서는 사용자의 (내)문서입니다.
“Browse”를 눌러 변경해 봅시다.

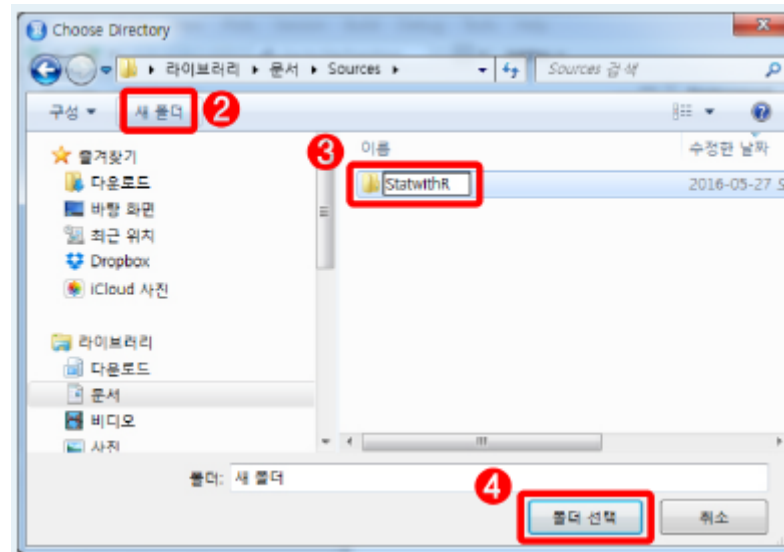


R Studio와 함께 R 시작하기

예제 1-1 프로젝트(Project) 생성

프로젝트 저장 위치 설정

- ② 디렉토리 선택창이 나타나면 “새폴더”를 누릅니다.
- ③ 생성된 “새폴더”의 이름을 “StatwithR”로 변경합니다.(F2 키 등 이용)
- ④ 이름을 변경한 “StatwithR” 디렉토리를 선택한 후 “폴더선택”을 누릅니다.



R Studio와 함께 R 시작하기

예제 1-1 프로젝트(Project) 생성

프로젝트 저장 위치 설정

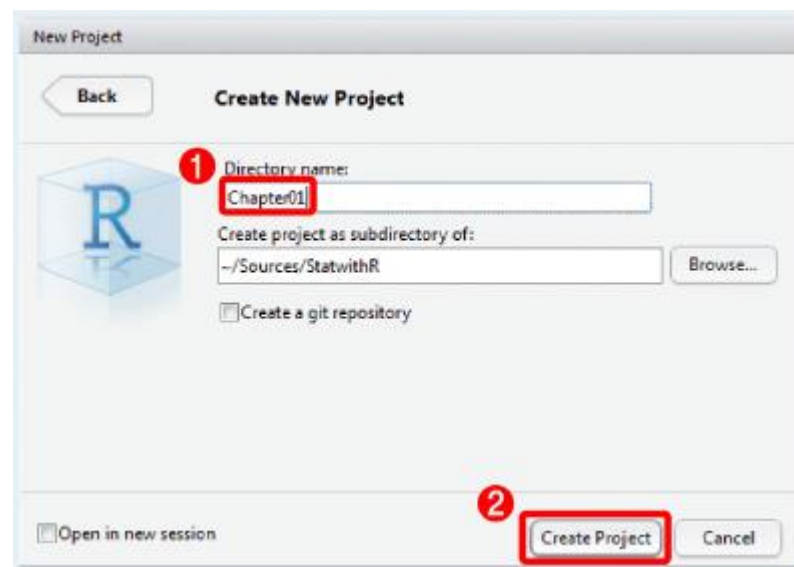
각 장별로 프로젝트를 만들고자 합니다.

먼저 1장의 프로젝트명을 “Chapter01”으로 하고 이를 생성합니다.

① Directory name을 “Chapter01”으로 합니다.

② 하단의 “Create Project”를 클릭합니다.

※ 앞서 지정한 ~/StatwithR 아래에
저장되는지 확인합니다.



R Studio와 함께 R 시작하기

예제 1-1 프로젝트(Project) 생성

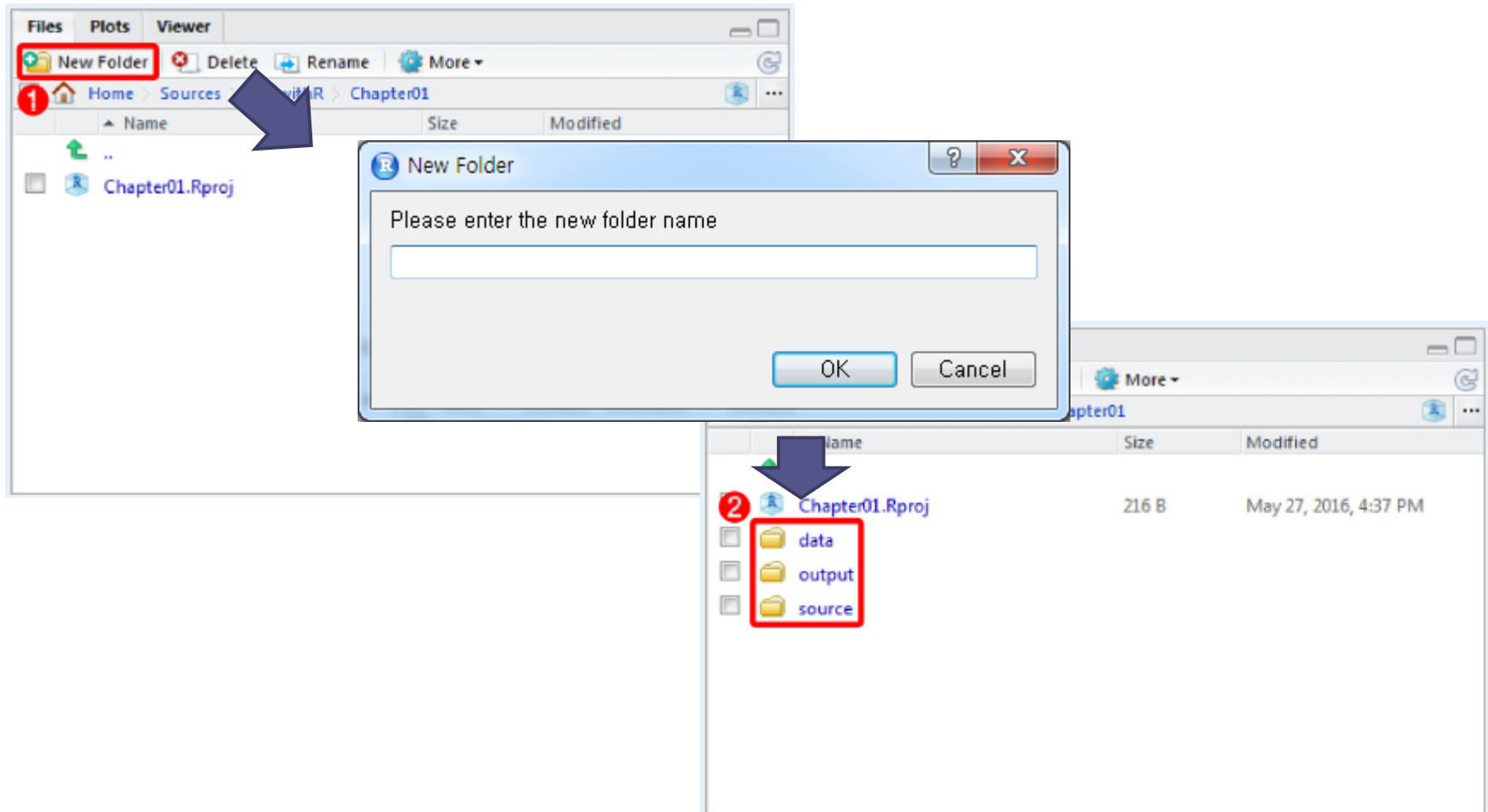
프로젝트에서 사용하는 각종 자원들을 위한 하위 폴더 생성

프로젝트 내에서 R 코드는 “source” 폴더, 데이터 셋은 “data” 폴더, 프로젝트에서 생성한 각종 파일은 “output” 폴더에 저장하고 관리합니다.

- ① R Studio 좌측 하단(부록 A의 R Studio 설치 후 설정을 동일하게 한 경우)의 “File” 탭에서 “New Folder”를 누릅니다.
- ② 생성할 폴더에 이름을 묻는 창에 “source”를 입력하고 “Ok”를 누릅니다.
(이와 마찬가지로 “data”, “output” 폴더를 생성합니다. 순서는 무관합니다.)

R Studio와 함께 R 시작하기

예제 1-1 프로젝트(Project) 생성



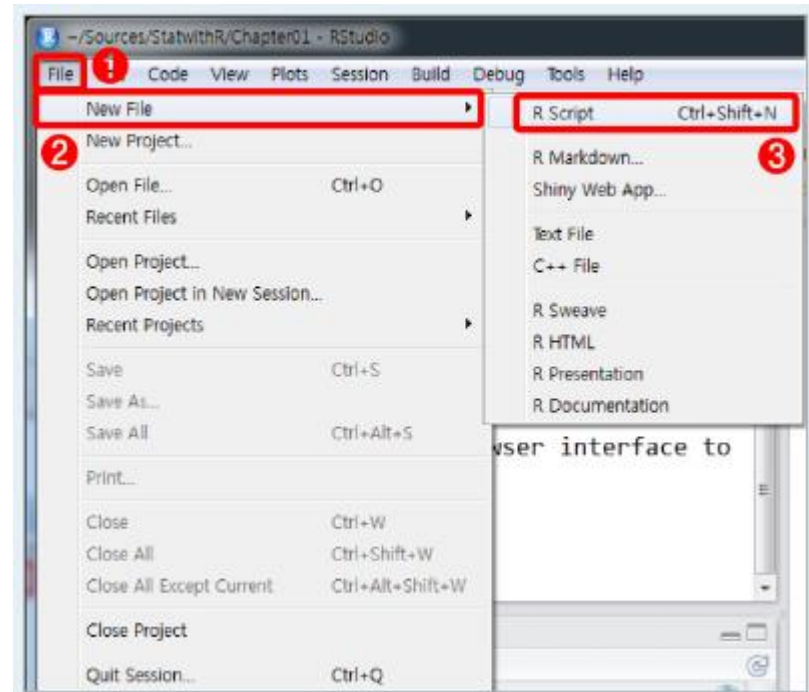
R Studio와 함께 R 시작하기

예제 1-2 R Script 파일 생성, 명령 실행하기와 종료

R Studio에서 R 파일을 생성하고 실행해 봅시다.

R 에 명령을 내릴 R 명령들로 구성된 R 파일(R Studio에서는 R script라고 합니다) 을 생성합니다.

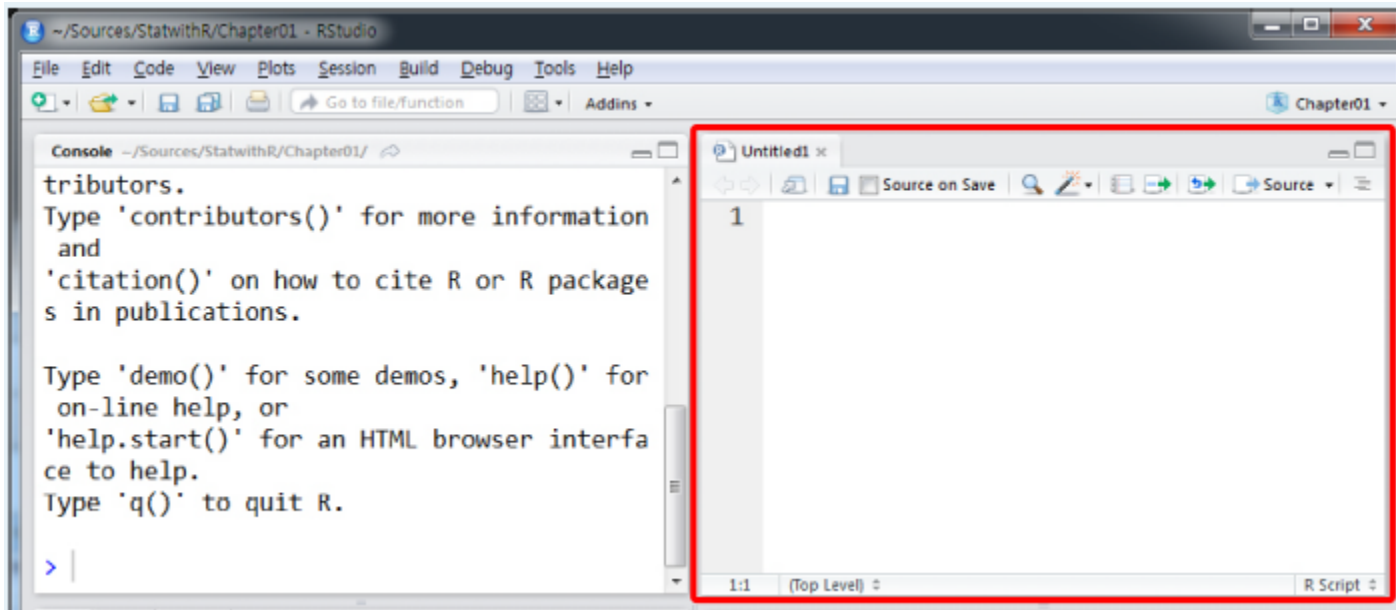
- ① File 메뉴를 클릭합니다.
- ② “New File”을 클릭합니다.
- ③ “R Script”를 클릭합니다.
(단축키로 Ctrl키, Shift키, N키를
동시에 누르셔도 됩니다.)



R Studio와 함께 R 시작하기

예제 1-2 R Script 파일 생성, 명령 실행하기와 종료

R Studio 우측에 R 명령들을 편집하고 R 스크립트 파일로 저장하기 위한 “Untitled01” 이라는 이름을 갖는 스크립트 탭이 생성됩니다.



R Studio와 함께 R 시작하기


예제 1-2 R Script 파일 생성, 명령 실행하기와 종료

R 스크립트 파일에 R 명령을 입력하고 이를 실행해 봅시다.

열려있는 R 스크립트 탭의 1번 줄에 `help(q)`를 입력합니다.

- 대소문자를 구별하므로 모두 소문자로 입력합니다.

- 화면에 커서가 2번줄에 있으나 실행 후 화면으로 엔터를 입력하지 않습니다.

① 열려있는 스크립트 탭 우측 상단의  을 클릭합니다.

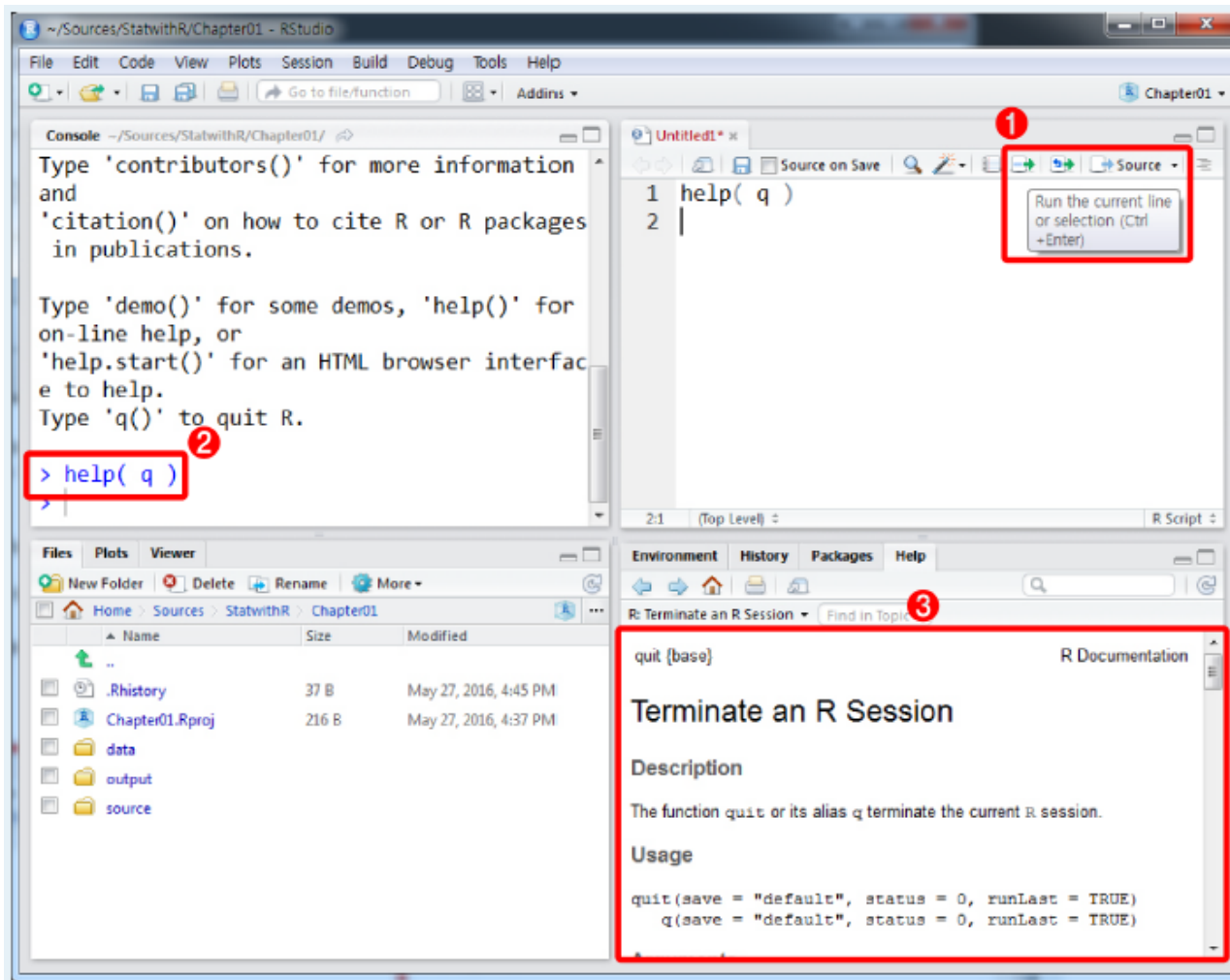
② R Studio 좌측 상단의 Console 창으로 스크립트 창에 입력한 R 명령어인 `help(q)`가 나타나고 이를 실행합니다.

③ `help()` 명령은 도움말을 구하는 R의 명령어(실재로는 함수)로 `q` 명령에 대한 도움말을 보여줍니다. R Studio에서는 우측 하단의 “Help” 탭에 그 내용이 나옵니다.

- `help()` 명령은 새롭게 나오는 R 명령마다 여러분들이 실행하셔서 그 내용을 확인해 주시기 바랍니다.

R Studio와 함께 R 시작하기


예제 1-2 R Script 파일 생성, 명령 실행하기와 종료



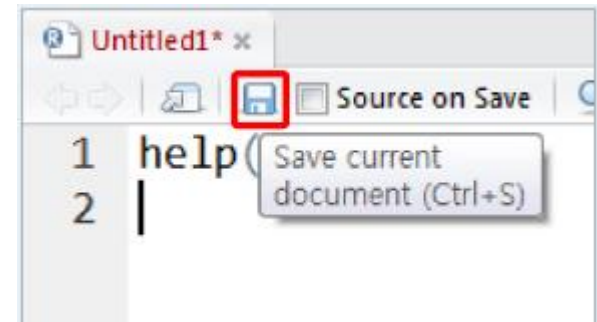
R Studio와 함께 R 시작하기

예제 1-2 R Script 파일 생성, 명령 실행하기와 종료

R 스크립트 파일을 저장해 봅시다.

스크립트 탭 상단의 저장을 위한 디스크 아이콘 ()을 클릭합니다.

- 단축키로 ctrl키와 s 키를 동시에 누르셔도 됩니다.



R Studio와 함께 R 시작하기

예제 1-2 R Script 파일 생성, 명령 실행하기와 종료

파일 저장을 위한 창이 생성됩니다.

① 앞서 R 스크립트 파일들은 source 폴더에 넣고 관리하기로 하였으니, 먼저 “source” 폴더를 더블클릭하여 엽니다.

② 파일이름으로 “code01.R”을 입력합니다.

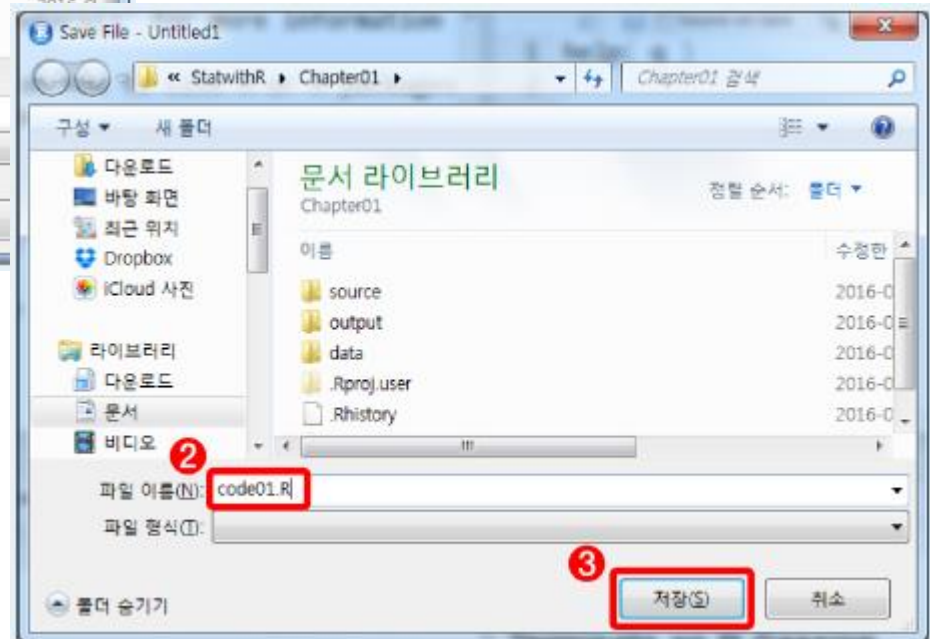
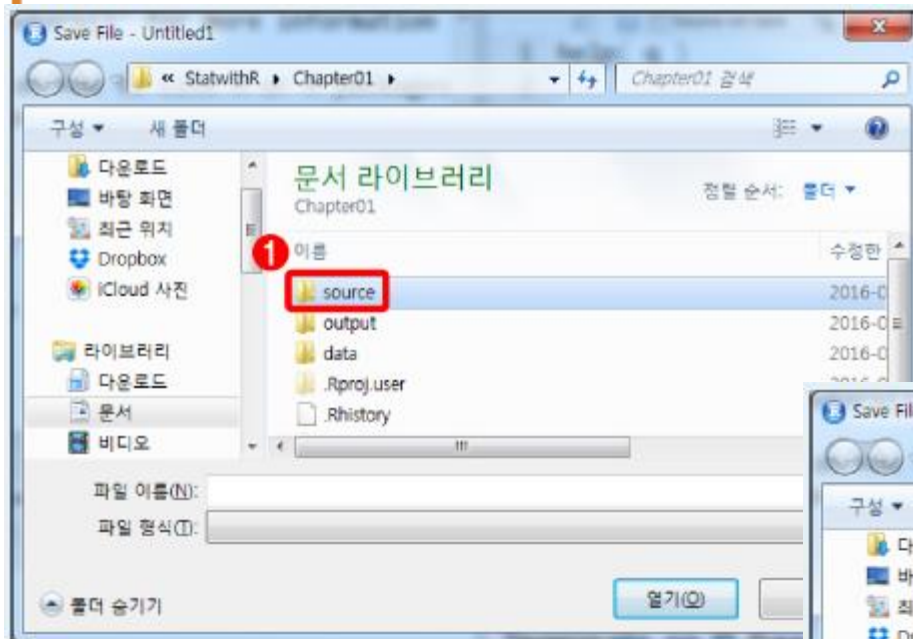
- R 스크립트 파일의 확장자는 .R(대문자)입니다. R Studio 버전에 따라 확장자를 자동으로 붙여주기도 하지만, 상황에 따라 다르므로 확장자를 같이 붙혀 저장합니다.

③ “저장”을 누릅니다.

- 스크립트 탭의 상단에 code01.R이라는 이름으로 변경이 될 것 입니다.

R Studio와 함께 R 시작하기

예제 1-2 R Script 파일 생성, 명령 실행하기와 종료



R Studio와 함께 R 시작하기

예제 1-2 R Script 파일 생성, 명령 실행하기와 종료

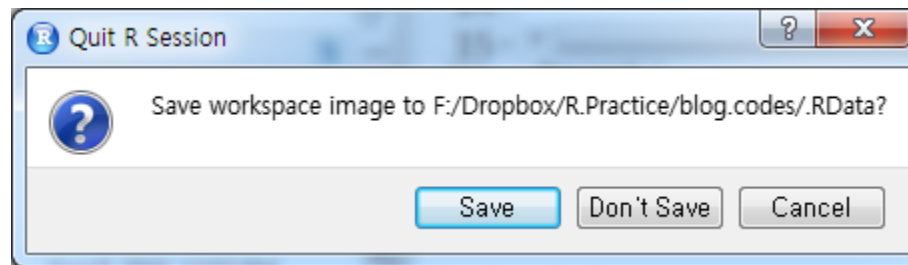
R Studio를 종료해 봅시다.

- ① File 메뉴를 클릭하면, 맨 아래에 “Quit Session”이 나옵니다. 클릭하면 R이 종료 됩니다. (교재 내용)
 - ② “Quit Session” 을 클릭하는 역할의 단축키는 Ctrl-Q 입니다.
 - ③ Console로 이동하여 “q()”라고 입력하고 엔터키를 입력합니다.
- 위의 세가지 방법 중 편한 방법을 사용하시면 됩니다.

R Studio와 함께 R 시작하기

예제 1-2 R Script 파일 생성, 명령 실행하기와 종료

R Studio가 종료하면서 변경사항이 있을 시 현재 생성한 각종 내용들을 저장할 것인지 물어봅니다. 스크립트 파일 등은 저장하고, R을 수행하면서 생성한 각종 자원들을 작업공간의 이미지로 저장하여 다음 사용시에 사용할 것인지를 물어옵니다. 우리는 모든 명령을 R 스크립트로 저장할 것이므로 굳이 저장하지 않아도 되지만, 여러분들이 원하는 대로 선택하시면 됩니다.



R 기초 : 계산기로서의 R

• 연산자

- 사전 : 벡터 공간, 함수 공간의 원소를 다른 원소에 대응시키는 변환 (네이버 어학 사전 “연산자”)
 - 익숙한 연산자로 사칙연산자가 있습니다.
- 컴퓨터 분야에서의 연산자
 - 프로그래밍은 주어진 자료를 바탕으로 각종 연산을 수행하여 결과를 도출하는 과정입니다. 프로그래밍에서 연산자 가장 기본이 되는 연산을 수행하기 위한 기호로 프로그래밍에서 각종 연산의 바탕이 됩니다.
 - 연산의 결과가 수치값을 갖는 산술연산자, 연산의 결과가 논리값을 갖는 논리연산자가 있으며 이외에도 다양한 연산자가 있습니다.

R 기초 : 계산기로서의 R

• R에서의 연산자


▣ 산술연산자

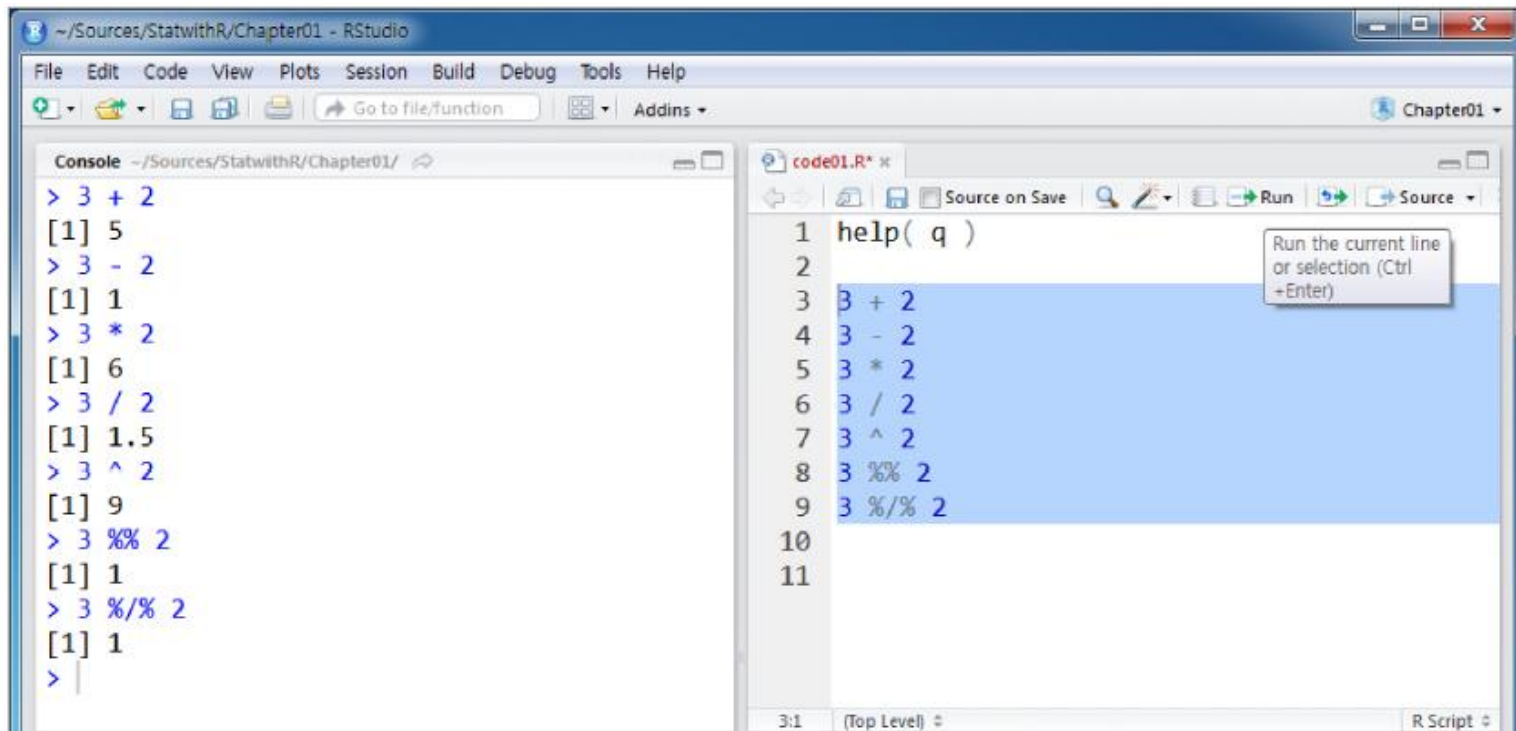
[표 1-3] R의 산술연산자

연산자	설명	예	결과
+	더하기	$3 + 2$	5
-	빼기	$3 - 2$	1
*	곱하기	$3 * 2$	6
/	나누기	$3 / 2$	1.5
$^$ 혹은 $**$	승수	$3 ^ 2$	9
$x \% y$	X를 y로 나눈 나머지 값 반환	$3 \% 2$	1
$x \%\% y$	나누기의 결과를 정수로	$3 \%\% 2$	1

R 기초 : 계산기로서의 R

• R code : code01.R

- 스크립트 창에 다음과 같이 코드를 입력하고 실행할 코드들을 마우스 드래그로 선택한 후  버튼을 누르거나 “Ctrl + Enter”를 동시에 키보드로 입력하면 R 코드의 결과가 좌측의 콘솔 창에 나옵니다. (드래그한 코드 모두 실행)



R 기초 : 계산기로서의 R

• R에서의 연산자

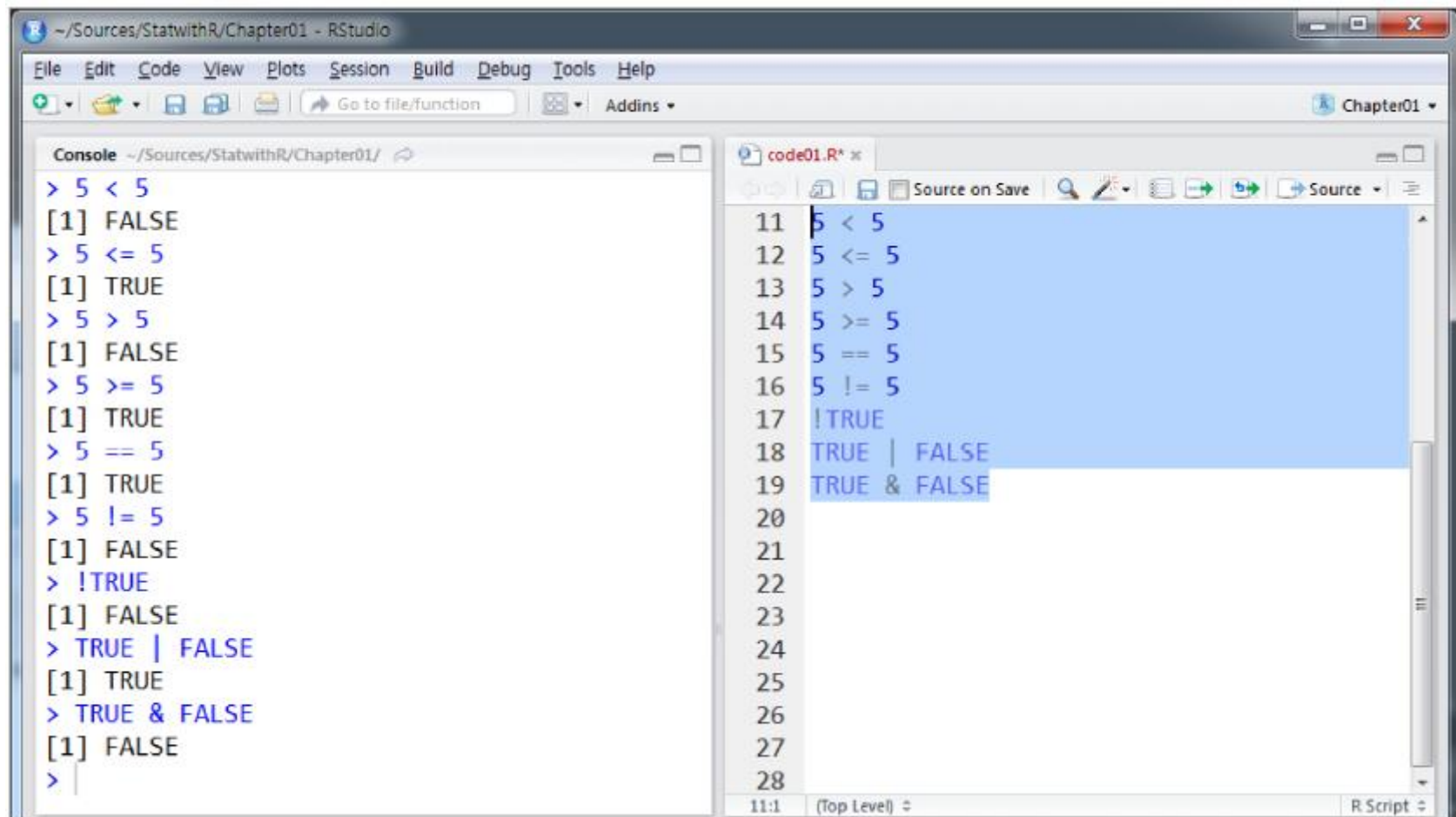
- 논리연산자 : 연산의 결과가 논리값(TRUE 혹은 FALSE)

[표 1-4] R의 논리 연산자

연산자	설명	예	결과
<	좌변이 보다 작은	5 < 5	FALSE
<=	좌변 이하	5 <= 5	TRUE
>	좌변이 보다 큰	5 > 5	FALSE
>=	좌변 이상	5 >= 5	TRUE
==	값이 같은	5 == 5	TRUE
!=	값이 다른	5 != 5	FALSE
!x	부정형 연산	!TRUE	FALSE
x y	x OR y (논리합)	TRUE FALSE	TRUE
x & y	x AND y (논리곱)	TRUE & FALSE	FALSE

R 기초 : 계산기로서의 R

- R code : code01.R**



The screenshot shows the RStudio interface with a script editor on the right and a console on the left. The script editor contains the following R code:

```
11 5 < 5
12 5 <= 5
13 5 > 5
14 5 >= 5
15 5 == 5
16 5 != 5
17 !TRUE
18 TRUE | FALSE
19 TRUE & FALSE
20
21
22
23
24
25
26
27
28
```

The console on the left shows the output of these operations:

```
> 5 < 5
[1] FALSE
> 5 <= 5
[1] TRUE
> 5 > 5
[1] FALSE
> 5 >= 5
[1] TRUE
> 5 == 5
[1] TRUE
> 5 != 5
[1] FALSE
> !TRUE
[1] FALSE
> TRUE | FALSE
[1] TRUE
> TRUE & FALSE
[1] FALSE
>
```

R 기초 : 변수 - 사용할 자료 저장

- **TRUE, FALSE는 문자일까요? 숫자일까요?**

- TRUE와 FALSE 은 R에서 무언가를 나타내는 이름으로
 - TRUE(단축형으로 T) : R에서 참인 상태를 나타냄
 - FALSE(단축형으로 F) : R에서 거짓인 상태를 나타냄
- TRUE와 FALSE가 상황에 따라 각각 나타내는 바가 다르면 어떨까요?
- 상수 : 나타내는 값을 바꿀 수 없는 자료

- **변수**

- 상수는 프로그래밍을 좀 더 유연하고 편하게 하기 위해 미리 정의되어 있는 자료들이며, 실제 사용되는 자료들은 상황에 따라 달라집니다.
- 상황에 따라 값을 바꿀 수 있는 자료들을 변수라고 합니다.

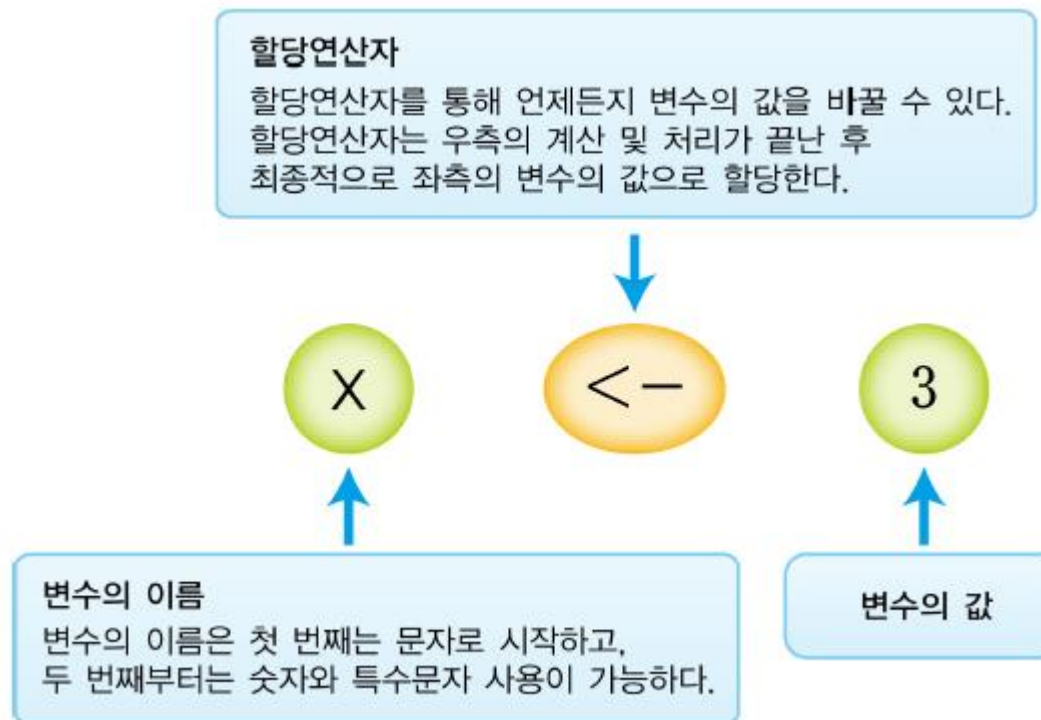
R 기초 : 변수 - 사용할 자료 저장

• 변수 사용하기

- 이름 : 저장하고자 하는 값을 가장 잘 나타낼 이름을 짓습니다.
 - 문자, 숫자, 특수문자(점(.), 밑줄(_))을 사용할 수 있습니다.
 - 변수의 이름은 숫자로 시작할 수 없습니다.
 - 점(.)로도 시작할 수 있으나 바로 뒤에 숫자가 나올 수 없습니다.
 - R에서 사용하는 예약어들은 변수명으로 사용할 수 없습니다. (for, function 등)
 - 변수명은 **대소문자를 구분**합니다.
- 대입연산자(혹은 할당연산자) : “<- “
 - 할당연산자를 이용하여 변수에 원하는 값을 저장합니다.
- 값
 - 대입연산자를 이용하여 변수에 값을 저장합니다.
 - 변수의 초기화
 - 변수를 지금 당장 사용하지 않지만, 변수를 미리 만들어 놓고 값으로 결측상태를 나타내는 NA, 혹은 값이 정해지지 않은 상태를 나타내는 NULL을 이용하거나 연산에 따라 항등원을 이용하여 값을 초기화하고 필요한 시점에 원하는 값을 저장하여 사용합니다.

R 기초 : 변수 - 사용할 자료 저장

- 변수에 값 할당하기



R 기초 : 변수 - 사용할 자료 저장

- 변수 사용하기 : 01.variable.R

```
# 변수를 만들고 값을 대입합니다.
```

```
> x <- 3
```

```
# 변수 x의 값을 사용합니다.
```

```
> x
```

```
[1] 3
```

```
# 변수 x의 이름을 입력합니다.
```

```
# 변수 x가 갖고 있는 값을 돌려줍니다.
```

```
# 변수의 값 바꾸기
```

```
> x <- 5
```

```
> x
```

```
[1] 5
```

```
# 대입연산자를 이용하여 변수의 값을 변경합니다.
```

```
# 초기화 연습
```

```
> x <- NULL
```

```
# NULL로 변수를 초기화하였습니다.
```

```
# 초기화는 변수를 미리 마련하고
```

```
# 추후 필요시 원하는 값으로 대입합니다.
```

R 기초 : 자료형

• 변수 사용하기 : 01.variable.R

```
# 두 변수 x와 y의 값을 서로 교환합니다.
> y <- 3                # 변수 y에 3을 대입합니다.
> temp <- y             # 새로운 변수 temp에 y의 값을 대입합니다.
> temp                 # temp는 y의 값인 3을 갖고 있습니다.
[1] 3
> y                    # y는 값을 대입한적이 없으므로 그대로 3입니다.
[1] 3
> y <- x               # 변수 y에 x가 갖고 있는 값을 대입합니다.
> y                   # 변수 y는 이제 x의 값인 5를 갖습니다.
[1] 5
> x                   # x는 여전히 5입니다.
[1] 5
> x <- temp            # 변수 x에 temp가 갖고 있던 y의 값을 대입합니다.
> x                   # 변수 x와 y가 서로 값을 바꾸었습니다.
[1] 3
> y
[1] 5
```

R 기초 : 자료형 - 컴퓨터의 자료저장 방법

- 컴퓨터의 자료 저장은 0과 1만
- 자료형 : 저장하고자 하는 자료에 맞는 저장방식
 - 정수형 : 2진수의 형태로 저장합니다.
 - 실수형 : 부동소수점으로 지수부와 가수부를 2진수로 저장합니다.
 - 부동소수점 : 소수점의 위치를 고정하지 않고 그 위치를 나타내는 수를 따로 적는 것 (위키피디아, 부동소수점)
 - 예를들어 1.2345 , $12.345 * 10^{-1}$, $123.45 * 10^{-2}$ 은 모두 같은 숫자
 - 여기서 12345는 가수부, 10은 밑, 각 승수 -1, -2는 지수부에 해당
 - 컴퓨터에서 부동소수점 수는 근사값이며, 유효숫자 범위에 따라 정밀도 구분
 - 문자형 : 저장하고자 하는 문자에 해당하는 숫자로 구성된 문자표에 의해 인코딩합니다.
 - 저장시 문자표에 따라 숫자로 저장하고 읽어올 시 저장된 숫자에 맞는 문자 출력
 - 문자저장시 큰따옴표 혹은 작은따옴표로 저장할 값을 감쌉니다. (“123”, “abc”)
 - 큰 따옴표 혹은 작은 따옴표로 감싸지 않은 문자열은 R에서 이름으로 사용합니다.

R 기초 : 자료형 - 컴퓨터의 자료저장 방법

• 기본 자료형 : o2.datatype.R

```

> x <- 3
> y <- 2
> x / y
[1] 1.5
> xi <- 1 + 2i
> yi <- 1 - 2i
> xi + yi
[1] 2+0i
> str <- "string"
> str
[1] "string"
>
> TRUE
[1] TRUE
> FALSE
[1] FALSE

```

수치형 자료
R에서는 기본적으로 실수 사용

실수/실수의 형태로 결과는 실수
복소수형의 경우 허수부에
소문자 i를 붙혀 복소수임을 알림

복소수간의 연산결과
문자열의 경우 큰/작은따옴표 사용

문자열임을 알리는 큰 따옴표

TRUE/FALSE는 논리형



Q & A



수고하셨습니다.