

R 데이터 분석 입문

1주차

Introduction

오 세 종

 DANKOOK UNIVERSITY

Contents



1. The age of data
2. 빅데이터
3. Install R, Rstudio

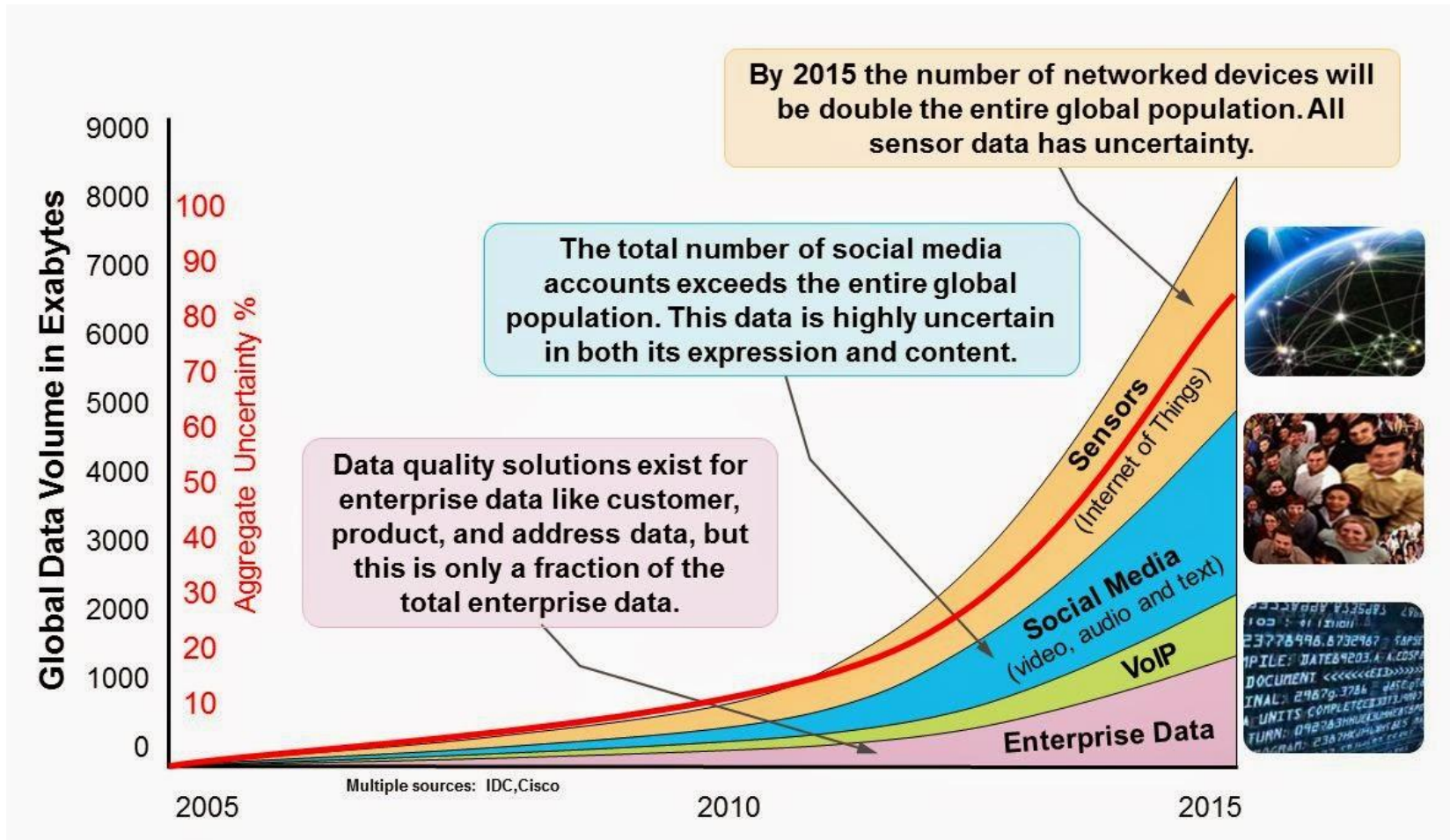
1. The age of data

- 우리는 데이터의 시대(the age of data)에 살고 있음
- 우리를 둘러싼 모든 것들이 데이터 소스와 연결되고, 우리 삶의 많은 부분들이 디지털 데이터화 되고 있음
 - 이메일, SNS, 전화사용 기록, 신용카드거래 기록, 병원 치료 기록, 성적
 - 인터넷, 주민정보, 등기정보, 판매정보, 주식거래 정보
- 데이터의 증가 속도는 점점 빨라지고 있으며, 전통적인 방법으로는 다루기 힘든 규모가 되고 있다
 - 빅데이터
- 정치,경제, 사회, 문화 등 모든 영역이 데이터와 연관



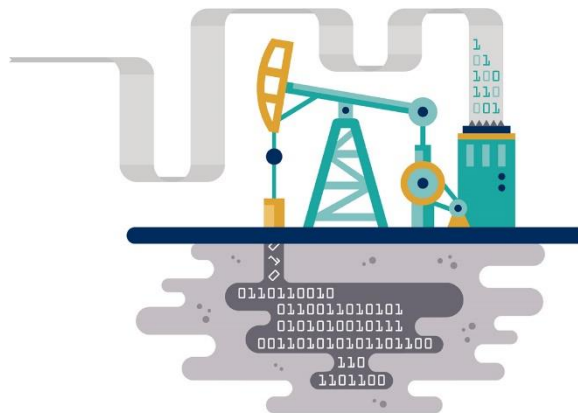
1. The age of data

- 데이터의 폭발적인 증가



1. The age of data

- “Data are becoming the new raw material of business” - Craig Mundie, Microsoft
- “Data tab has been turned on and will never be turned off” – Mike Hoskins, Actian CTO, 2013
- “You can have data without information, but you cannot have information without data” – Writer
- “It is a capital mistake to theorize before one has data” – Sherlock Homes
- “Data is the new oil” – Clive Humby, ANA Senior marketer’s submit



1. The age of data

데이터를 잘 다룰 줄 아는 기업, 개인이 성공하는 시대가 도래



1. The age of data

● 사례 1

한 남성이 화가 잔뜩 난 채로 대형마트에 들어서더니 “매니저 나와”라고 소리 질렀다. 손에는 이 대형마트가 딸에게 우편으로 보낸 아기 옷과 침대 등 유아용품 할인쿠폰이 들려 있었다. 남성은 “내 딸은 아직 고등학생인데 이런 쿠폰을 보내다니. 임신을 부추기는 거냐”고 따졌다. 영문을 몰랐던 매니저는 우선 남성을 진정시키고 돌려보냈다. 며칠 뒤 재차 사과를 하려고 전화를 건 매니저에게 뜻밖의 반응이 돌아왔다. 이 남성은 “딸이 임신을 한 걸 뒤늦게 알게 됐다”며 사과했다. 미성년 딸이 부모 모르게 아이를 가졌던 것이다. 대형마트 마케팅팀은 부모도 몰랐던 딸의 임신 사실을 어떻게 알아챘을까? 답은 ‘빅데이터 분석’에 있다. 이 얘기는 미국 대형마트 ‘타깃’의 미니애폴리스 점포에서 실제로 일어났던 일이다. <뉴욕 타임스>가 보도한 이 사례는 빅데이터 분석의 힘을 여실히 보여준다. 대형마트는 고객이 가족에게도 알리지 않은 비밀을 찾아내 앞으로 무엇이 필요할지 정확하게 예측해냈다.

http://www.hani.co.kr/arti/economy/economy_general/729868.html

1. The age of data

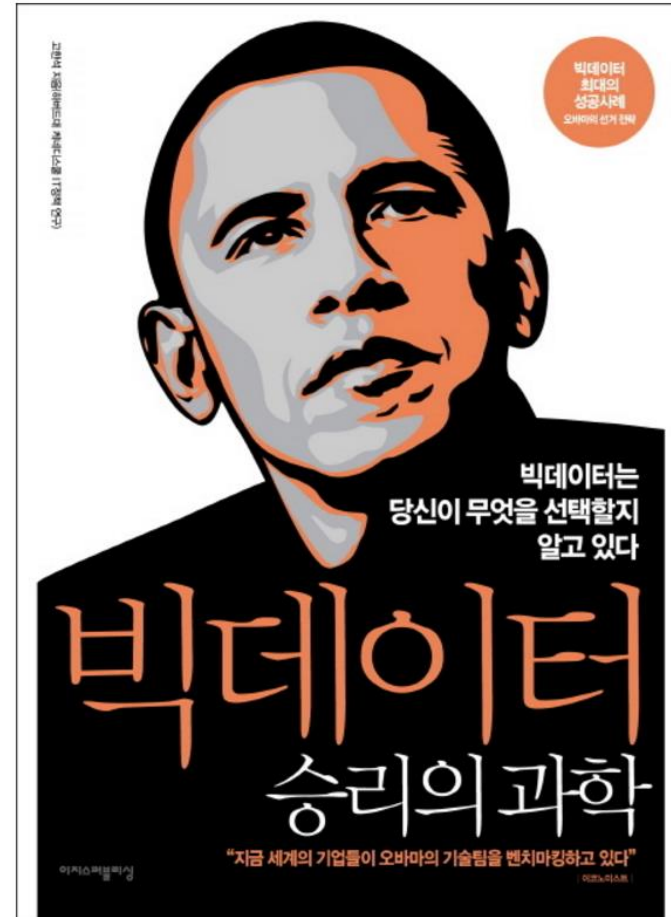
타깃의 빅데이터 전문가들은 고객의 25가지 구매 행태를 분석하면 여성의 임신과 출산을 상당히 정확하게 예측할 수 있다는 사실을 확인했다. 예를 들어 향이 나는 로션을 사던 여성이 무향의 로션으로 바꾸거나, 평소 사지 않던 미네랄 영양제를 갑자기 사들이는 경우다. 타깃은 고객 데이터 베이스에 이를 적용했고, 전국적으로 수만명의 임신 추정 여성들을 가려내 관련 할인쿠폰을 보냈다. 이를 통해 고객의 구매 행동을 이끌어낼 수 있기 때문이다.



<https://theamericangenius.com/tech-news/target-remodeling-stores/>

1. The age of data

- 사례 2. 오바마 : 데이터베이스에 기초한 개인 맞춤형 선거 홍보



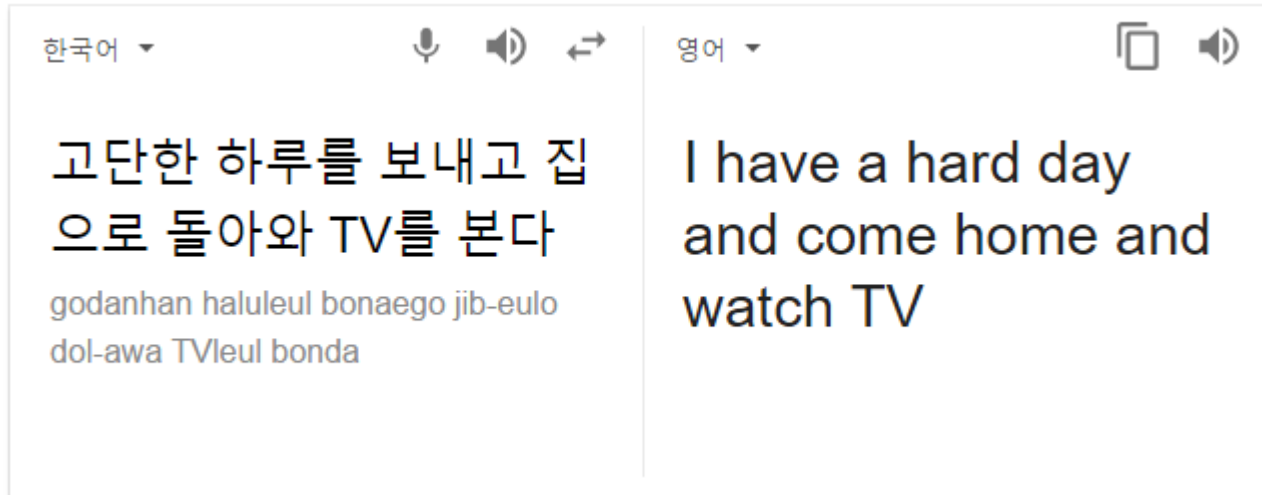
1. The age of data

- 후원금 모금액



1. The age of data

- 사례3. 구글 자동번역



과거 : 형태소 분석, 단어 by 단어로 번역

구글 : 번역된 예문 데이터를 기반으로 가장 적합한 번역 사례를 찾음
데이터가 쌓이면 쌓일수록 번역이 정교해짐

1. The age of data

- 사례 4. 빅데이터(유튜브) - 세계미래포럼



- <https://www.youtube.com/watch?v=X4hMFym0-uo>

2. 빅데이터*

- 데이터의 규모에 초점을 맞춘 정의

기존 데이터베이스 관리도구의 데이터 수집, 저장, 관리, 분석 역량을 넘어서는 데이터 (맥킨지 2011년 5월)

- 업무 수행 방식에 초점을 맞춘 정의

다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고, 데이터의 빠른 수집, 발굴, 분석을 지원하도록 고안된 차세대 기술 및 아키텍처 (IDC 2011년 6월)

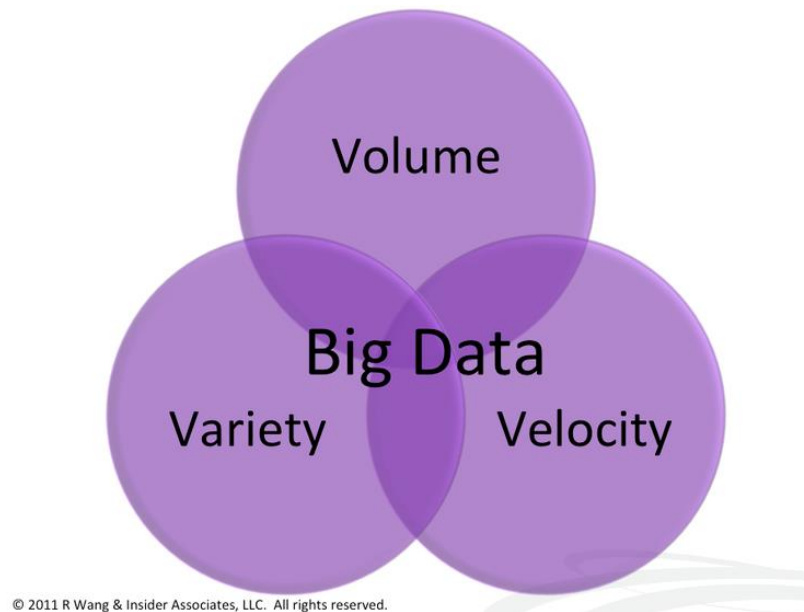
* Yonsei University 2nd Semester, 2014 Sanghyun Park

2. 빅데이터

- 빅데이터의 출현 배경
 - 스마트폰과 같은 디지털 기기의 보급과 소셜 네트워크서비스(SNS)의 부상으로 엄청난 양의 데이터가 생산되고 있음
 - 의료분야의 환자 데이터, 은행권의 거래 데이터, 교통 이용 데이터 등도 빅데이터임
 - 빅데이터는 기업, 정부, 의료, 교육 분야에서 그 가치가 입증되고 있음
 - 빅데이터를 적극적으로 활용하는 기업은 비즈니스 성과를 개선할 수 있음

2. 빅데이터

- BI/DW 리서치 기관인 TDWI 가 빅데이터의 3대 요소(3V)를 그림으로 표현



- 3대 요소 가운데 두 가지 이상의 요소만 충족된다면 빅데이터라고 볼 수 있음

2. 빅데이터



[출처:채승병 외, 빅데이터: 산업 지각변동의 진원, SERI CEO Information, 2012. 5]

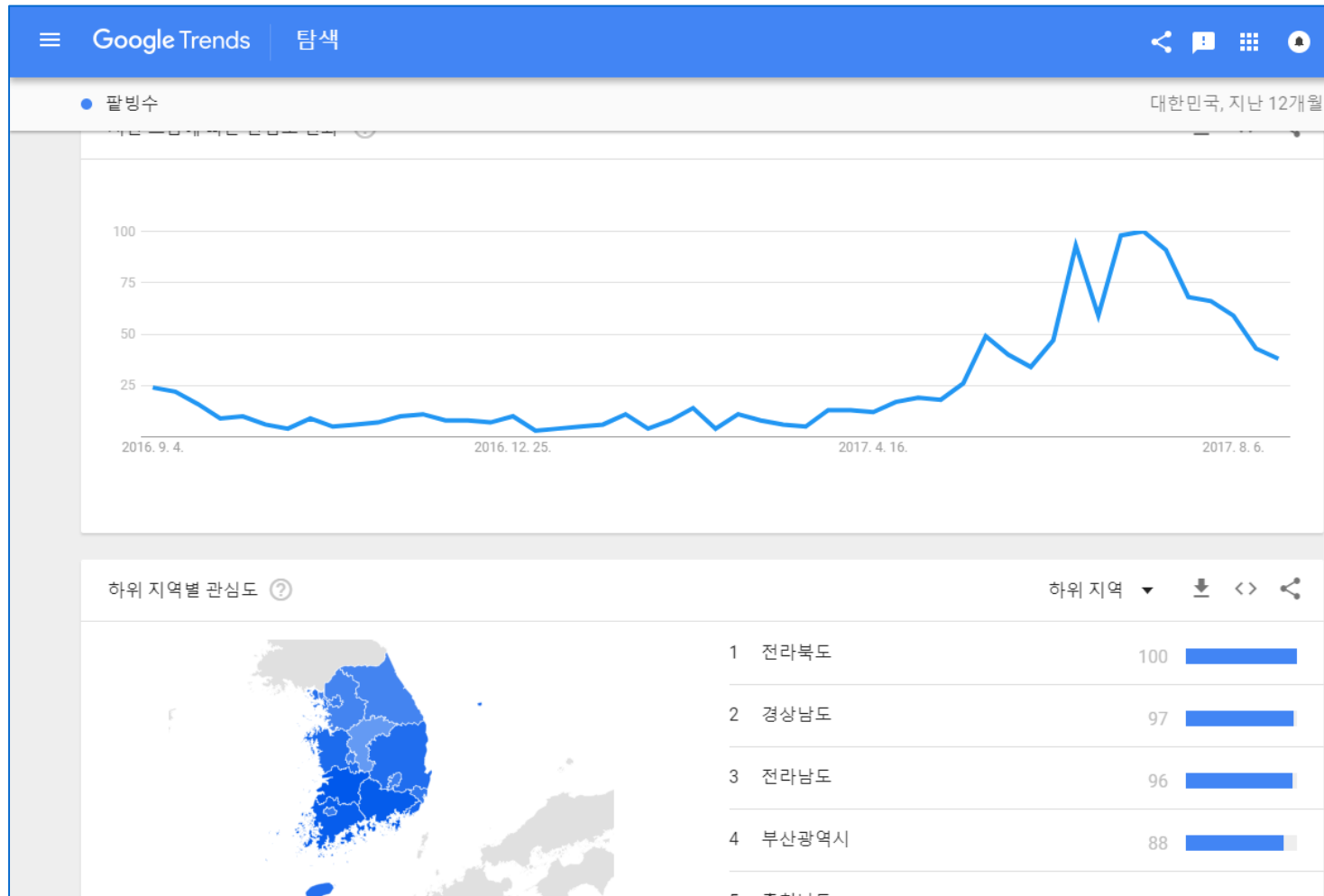
2. 빅데이터

- 공공데이터 포털
 - <https://www.data.go.kr/>

The screenshot shows the homepage of the Korea Public Data Portal. At the top, the logo "DATA . GO . KR" is displayed, with "공공데이터포털" (Public Data Portal) in smaller text. To the right of the logo are navigation links: "데이터셋" (Data Set), "활용사례" (Use Cases), "참여마당" (Participation Plaza), and "정보공유" (Information Sharing). Below these is a search bar with the placeholder text "검색어를 입력하세요." (Enter search terms). The main navigation bar is divided into three sections: "FILE DATA" (with a document icon), "OPEN API" (with a gear icon), and "STANDARD DATA" (with a bar chart icon). Below this bar is a grid of 16 icons representing various data categories: 교육 (Education), 국토관리 (Land Management), 공공행정 (Public Administration), 재정금융 (Finance), 산업고용 (Industry and Employment), 사회복지 (Social Welfare), 식품건강 (Food and Health), 문화관광 (Culture and Tourism), 보건의료 (Health and Medical Care), 재난안전 (Disaster and Safety), 교통물류 (Transportation and Logistics), 환경기상 (Environment and Weather), 과학기술 (Science and Technology), 농축수산 (Agriculture, Forestry, and Fisheries), 통일외교안보 (Unification, Foreign Affairs, and Security), and 법률 (Law). Below the grid are three featured sections: "국가중점개발데이터" (National Key Development Data) with a description and a right arrow icon; "공공데이터 활용사례" (Public Data Use Cases) featuring a "스마트 팔레트" (Smart Palette) and a plus icon; and "공공데이터 활용신청 TOP10" (Public Data Use Application TOP10) with icons for "FILE DATA" and "OPEN API".

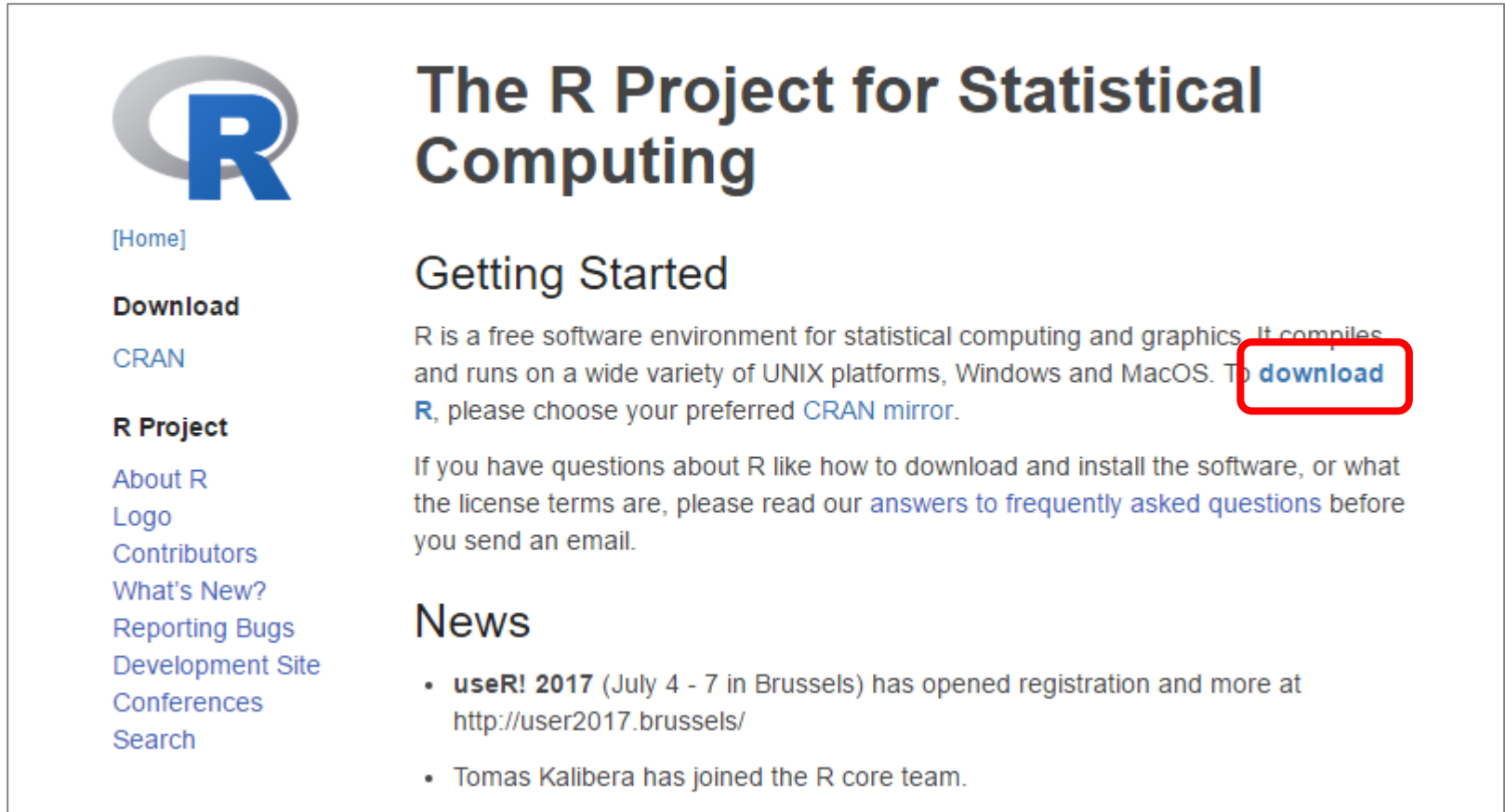
2. 빅데이터

- Google trend
 - <https://trends.google.co.kr/>




3. Install R, Rstudio

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



The screenshot shows the homepage of The R Project for Statistical Computing. On the left is a sidebar with navigation links. The main content area features the project title, a 'Getting Started' section with a 'download' button highlighted by a red box, and a 'News' section with two bullet points.



[\[Home\]](#)

Download

[CRAN](#)

R Project

[About R](#)

[Logo](#)

[Contributors](#)

[What's New?](#)

[Reporting Bugs](#)

[Development Site](#)

[Conferences](#)

[Search](#)

The R Project for Statistical Computing

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 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.

News

- **useR! 2017** (July 4 - 7 in Brussels) has opened registration and more at <http://user2017.brussels/>
- Tomas Kalibera has joined the R core team.

3. Install R, Rstudio

- 다운로드 사이트 선택

http://cran.stat.unipd.it/	University of Padua
http://dssm.unipa.it/CRAN/	Universita degli Studi di Palermo
Japan	
https://cran.ism.ac.jp/	The Institute of Statistical Mathematics, Tokyo
http://cran.ism.ac.jp/	The Institute of Statistical Mathematics, Tokyo
Korea	
http://cran.nexr.com/	NexR Corporation, Seoul
http://healthstat.snu.ac.kr/CRAN/	Graduate School of Public Health, Seoul National
http://cran.biodisk.org/	The Genome Institute of UNIST (Ulsan National In and Technology)
Lebanon	
http://rmirror.lau.edu.lb/	Lebanese American University, Byblos
Malaysia	
https://wbc.upm.edu.my/cran/	Univerisiti Putra Malaysia
http://wbc.upm.edu.my/cran/	Univerisiti Putra Malaysia

3. Install R, Rstudio

- 원도우용 버전 선택

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 (Monday 2016-10-31, Sincere Pumpkin Patch) [R-3.3.2.tar.gz](#), read

3. Install R, Rstudio

- base 버전 선택

R for Windows

Subdirectories:

base	Binaries for base distribution (managed by Duncan Murdoch). This is what you want to install R for the first time.
contrib	Binaries of contributed CRAN packages (for R >= 2.11.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 < 2.11.x; managed by Uwe Ligges).
Rtools	Tools to build R and R packages (managed by Duncan Murdoch). 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 Duncan Murdoch or Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

3. Install R, Rstudio

- 현재 최신 버전은 R3.4.1

R-3.4.1 for Windows (32/64 bit)

[Download R 3.4.1 for Windows](#) (62 megabytes, 32/64 bit)

[Installation and other instructions](#)

[New features in this version](#)

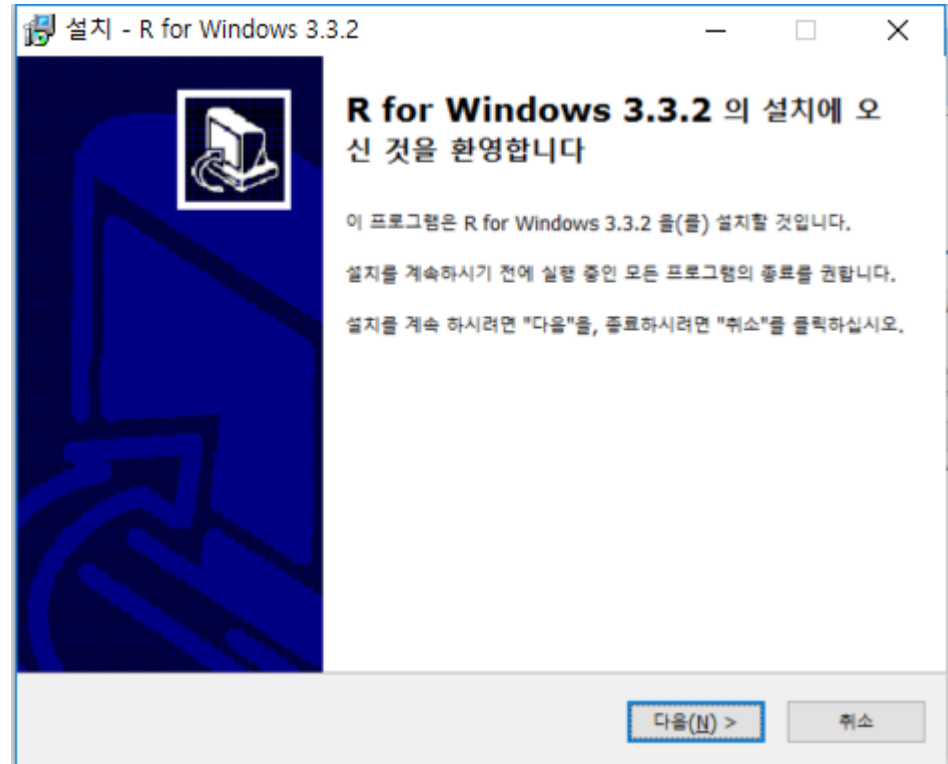
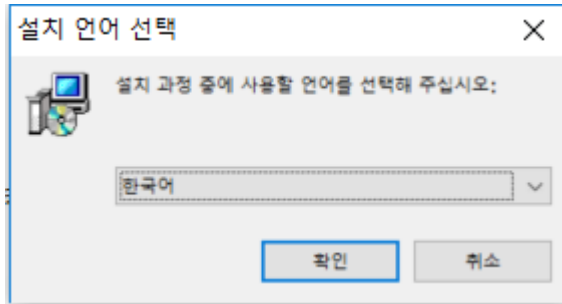
If you want to double-check that the package you have downloaded matches the package distributed by CRAN, you can check the [fingerprint](#) on the master server. You will need a version of md5sum for windows: both [graphical](#) and [command line version](#)

Frequently asked questions

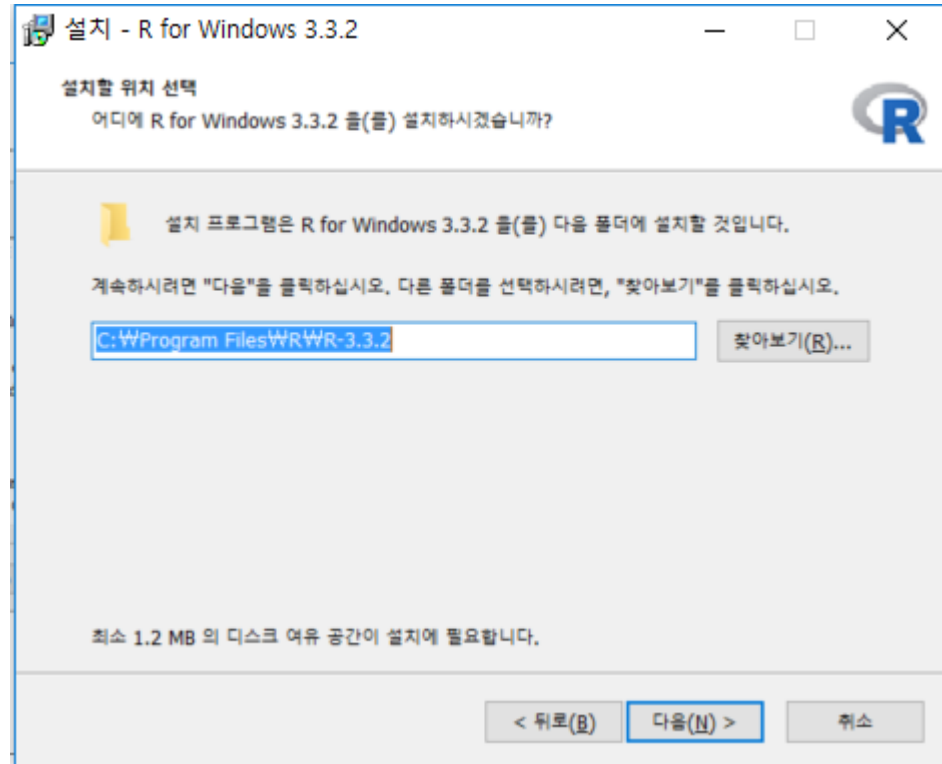
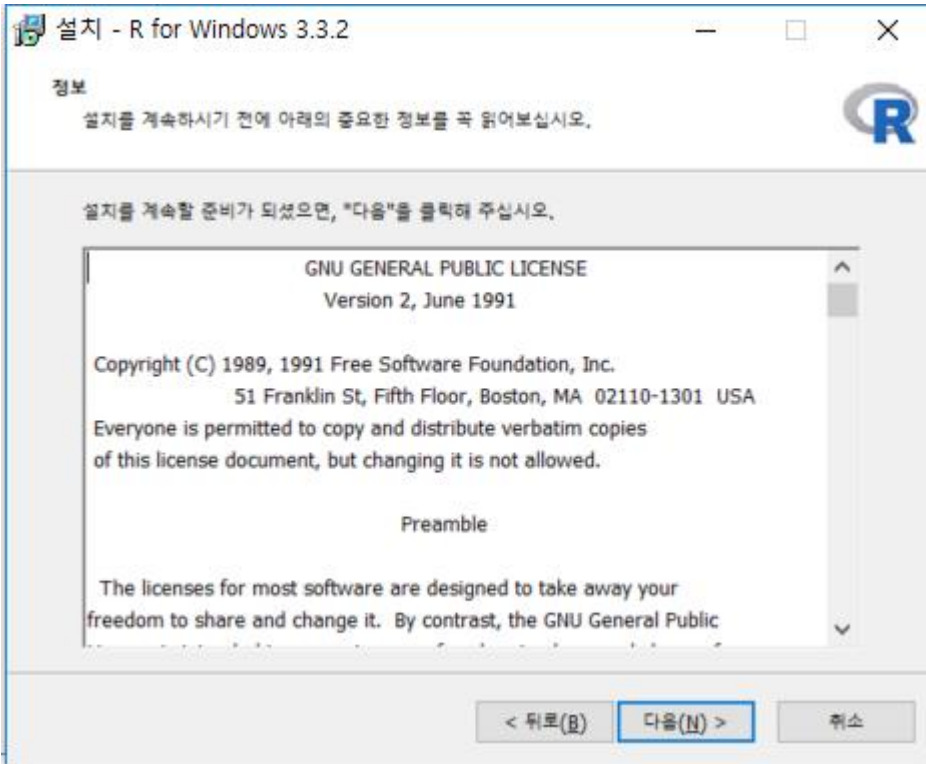
- [Does R run under my version of Windows?](#)
- [How do I update packages in my previous version of R?](#)
- [Should I run 32-bit or 64-bit R?](#)

3. Install R, Rstudio

- 다운로드 받은 설치 프로그램 실행

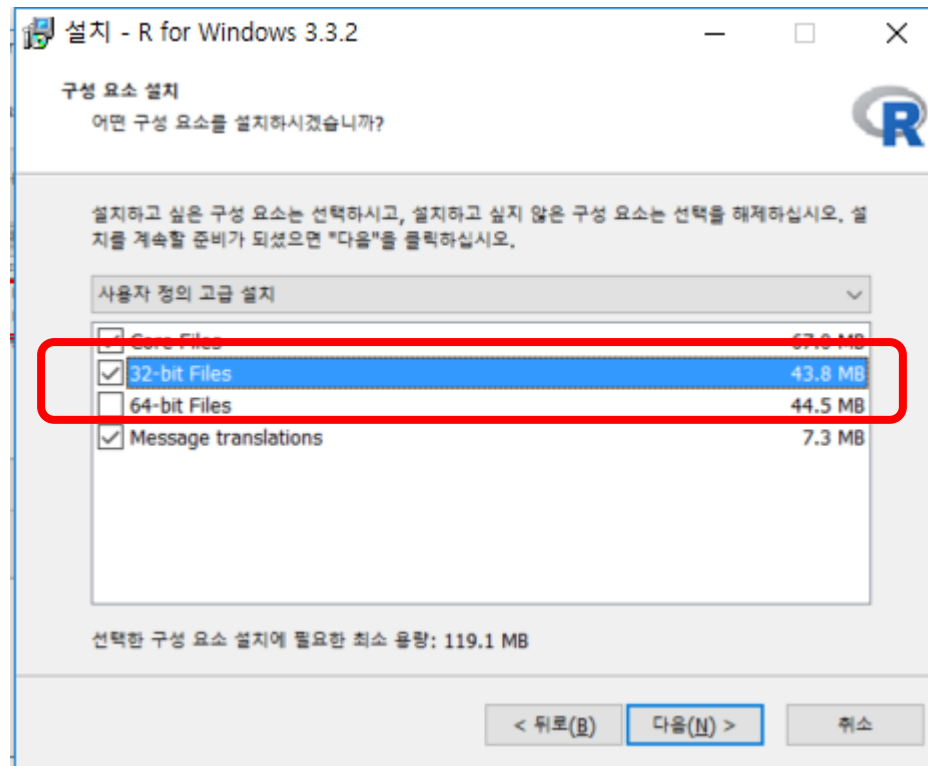


3. Install R, Rstudio



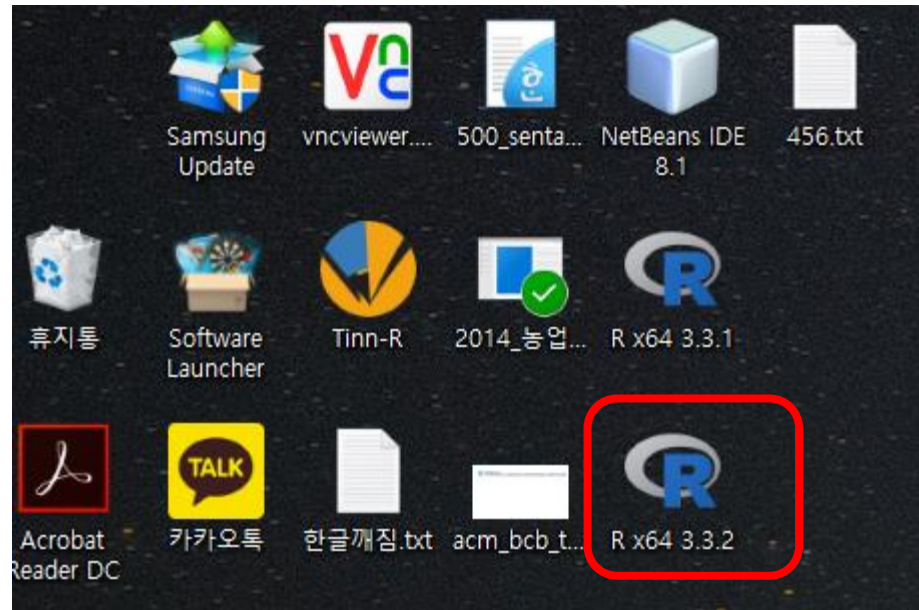
3. Install R, Rstudio

- 32 bit 또는 64 bit 버전 중 하나를 설치한다.
 - 운영체제의 버전에 따라 선택. 64 bit 버전이 더 큰 데이터를 더 효율적으로 처리할 수 있다



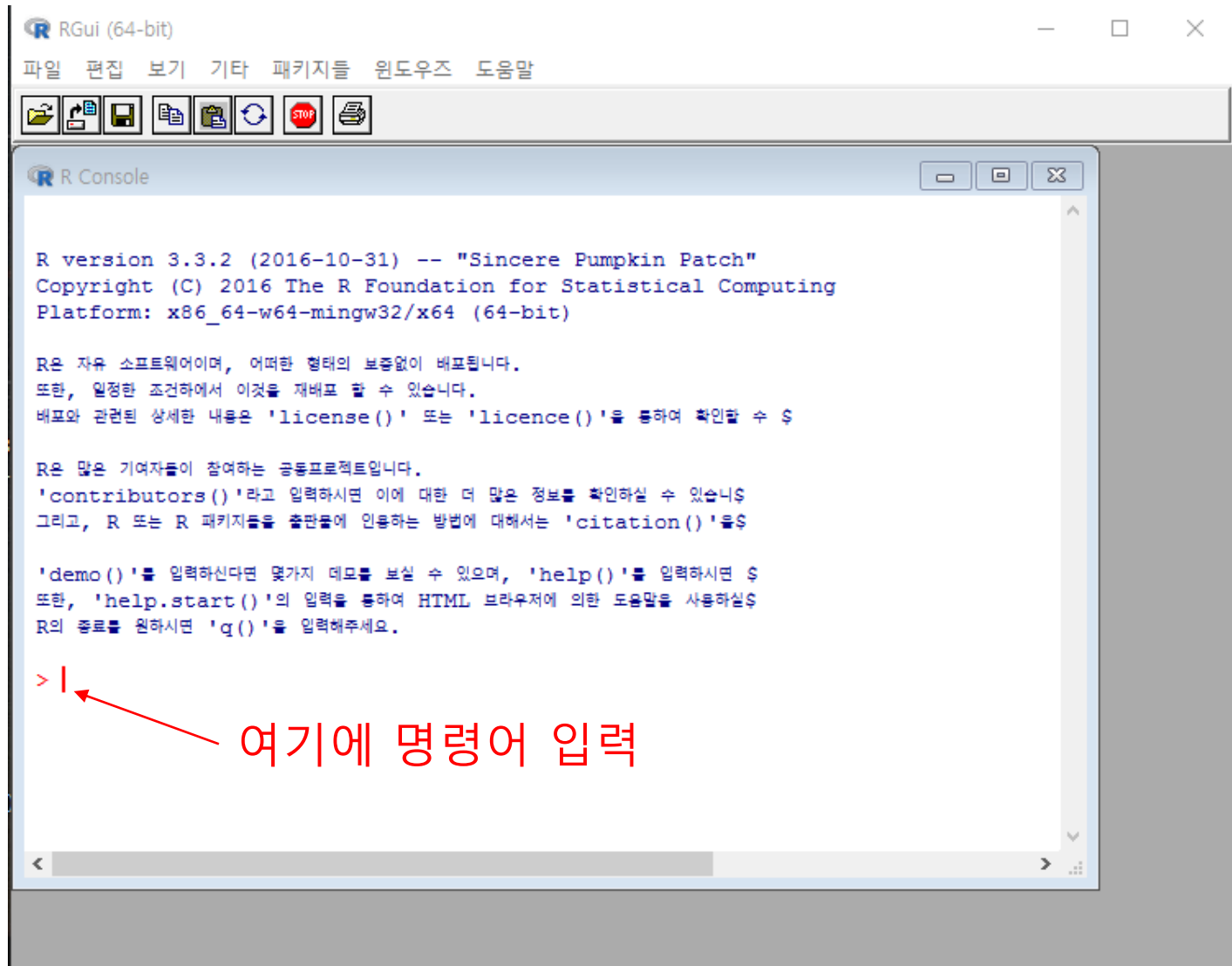
3. Install R, Rstudio

- 설치가 완료되면 바탕화면에 실행 아이콘 생성



3. Install R, Rstudio

- R 을 실행 화면



3. Install R, Rstudio

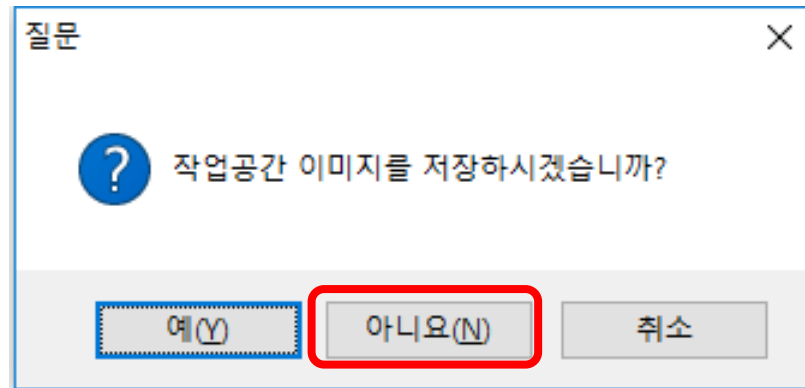
```
> 20+4  
[1] 24  
> a<-10  
> b<-20  
> a+b  
[1] 30  
>
```

```
> 20+4  
[1] 24  
> a<-10  
> b<-20  
> a+b  
[1] 30  
> |
```


3. Install R, Rstudio

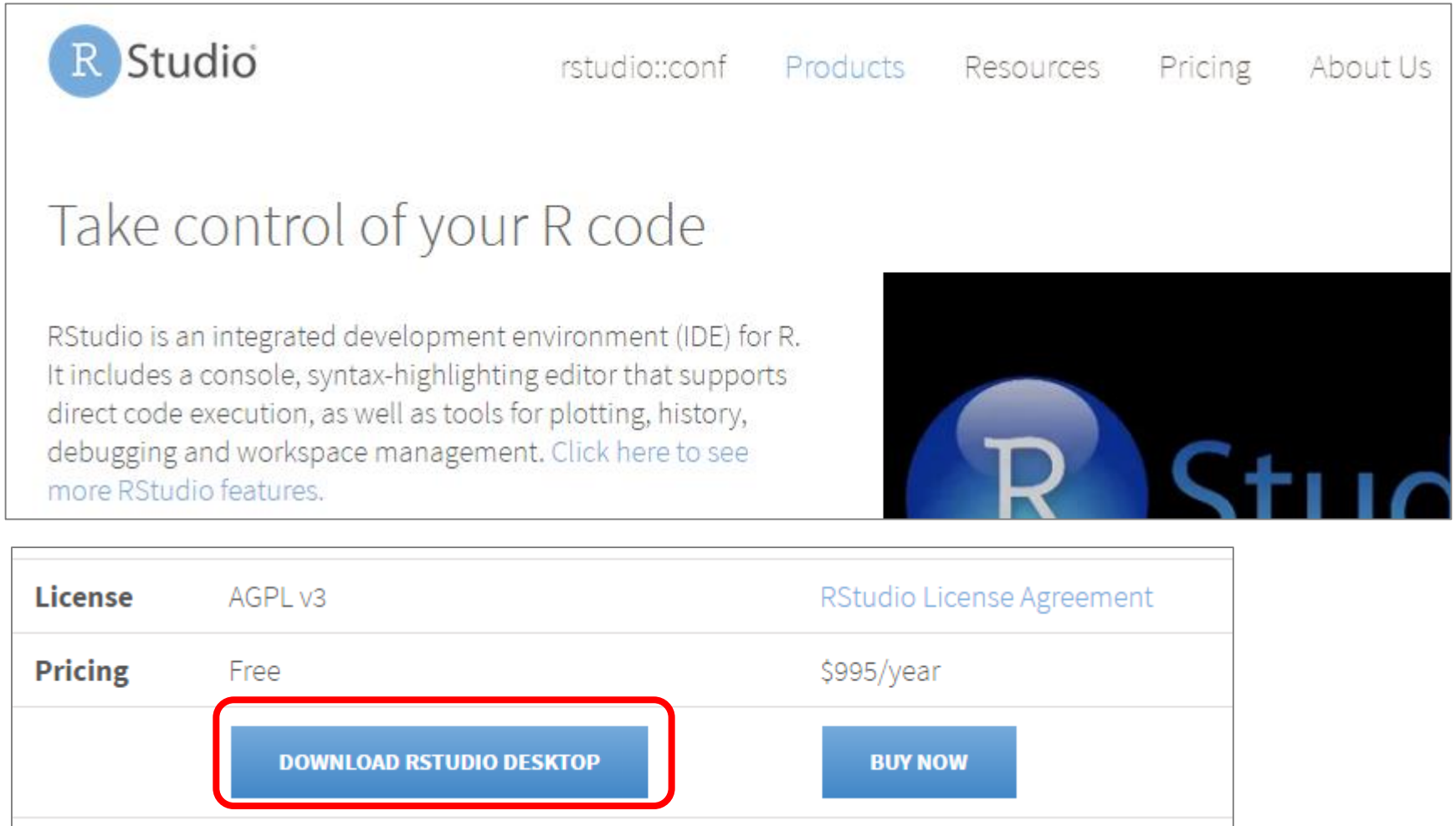
- R 실행 종료

- 종료버튼 클릭시 다음과 같은 화면이 나오는데, 현재까지 작업한 결과를 저장할 것인지를 묻는 것이다.
- 대부분의 경우는 실행한 명령어(R script)만 저장하고 작업공간 이미지(R script, 생성한 데이터 등)는 저장하지 않는다



3. Install R, Rstudio

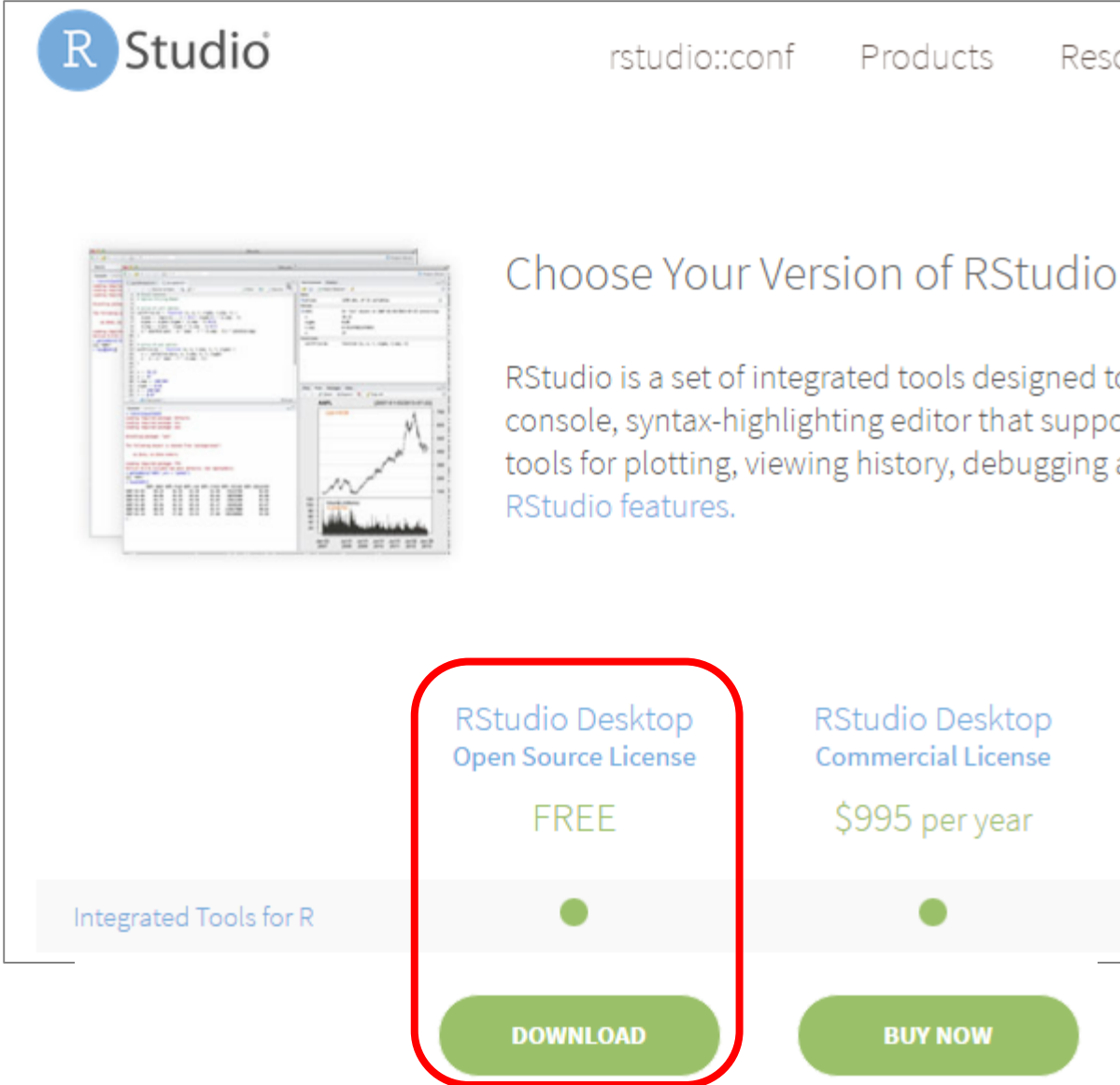
- <https://www.rstudio.com/products/rstudio/>
 - R 을 보다 편리하게 사용할 수 있는 환경 제공



The screenshot shows the RStudio website. At the top is the RStudio logo and navigation links: `rstudio::conf`, [Products](#), [Resources](#), [Pricing](#), and [About Us](#). The main heading is "Take control of your R code". Below it, a paragraph describes RStudio as an integrated development environment (IDE) for R, listing features like a console, syntax-highlighting editor, direct code execution, plotting, history, debugging, and workspace management. A link "Click here to see more RStudio features." is provided. To the right is a large image of the RStudio logo. Below the main content is a table with two rows: "License" (AGPL v3, with a link to the RStudio License Agreement) and "Pricing" (Free, with a price of \$995/year). At the bottom of the table are two buttons: "DOWNLOAD RSTUDIO DESKTOP" (highlighted with a red box) and "BUY NOW".

License	AGPL v3	RStudio License Agreement
Pricing	Free	\$995/year
	DOWNLOAD RSTUDIO DESKTOP	BUY NOW

3. Install R, Rstudio




The screenshot shows the RStudio website with the following elements:

- Header:** RStudio logo, navigation links for `rstudio::conf`, Products, and Resources.
- Image:** A collage of RStudio interface windows showing code, console, and plots.
- Section:** "Choose Your Version of RStudio"
- Description:** "RStudio is a set of integrated tools designed to... console, syntax-highlighting editor that supports... tools for plotting, viewing history, debugging and... RStudio features."
- License Comparison Table:**

License Type	Price	Download Button
RStudio Desktop Open Source License	FREE	DOWNLOAD
RStudio Desktop Commercial License	\$995 per year	BUY NOW

The "RStudio Desktop Open Source License" option is highlighted with a red rounded rectangle, and its "DOWNLOAD" button is visible.

3. Install R, Rstudio

 **RStudio**

[rstudio::conf](#) [Products](#) [Resources](#)

RStudio Desktop 1.0.136 — [Release Notes](#)

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

Installers for Supported Platforms

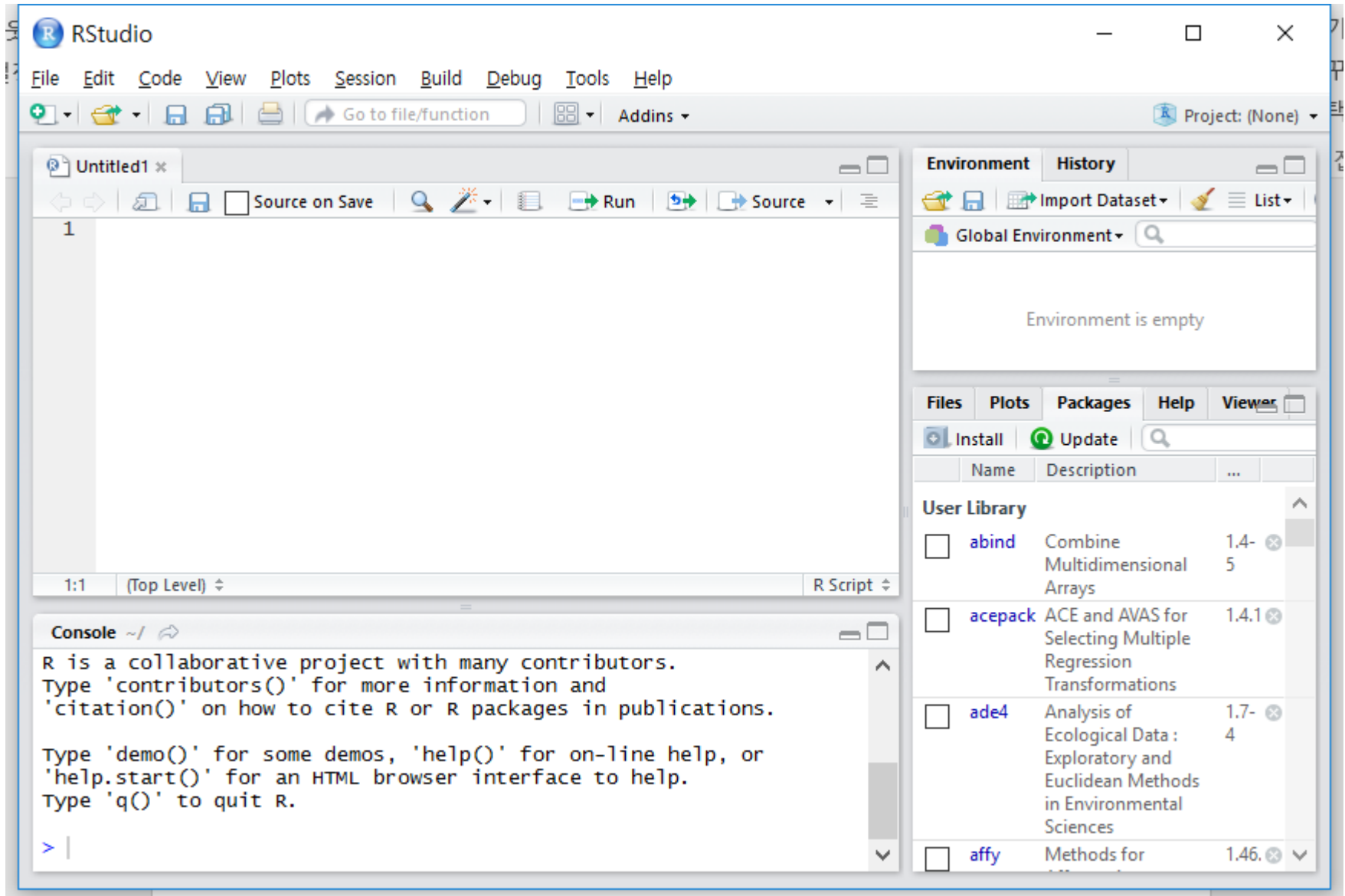
Installers	Size	Date
RStudio 1.0.136 - Windows Vista/7/8/10	81.9 MB	2016-12-21
RStudio 1.0.136 - Mac OS X 10.6+ (64-bit)	71.2 MB	2016-12-21
RStudio 1.0.136 - Ubuntu 12.04+/Debian 8+ (32-bit)	85.5 MB	2016-12-21
RStudio 1.0.136 - Ubuntu 12.04+/Debian 8+ (64-bit)	92.1 MB	2016-12-21
RStudio 1.0.136 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	84.7 MB	2016-12-21
RStudio 1.0.136 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	85.7 MB	2016-12-21

Zip/Tarballs

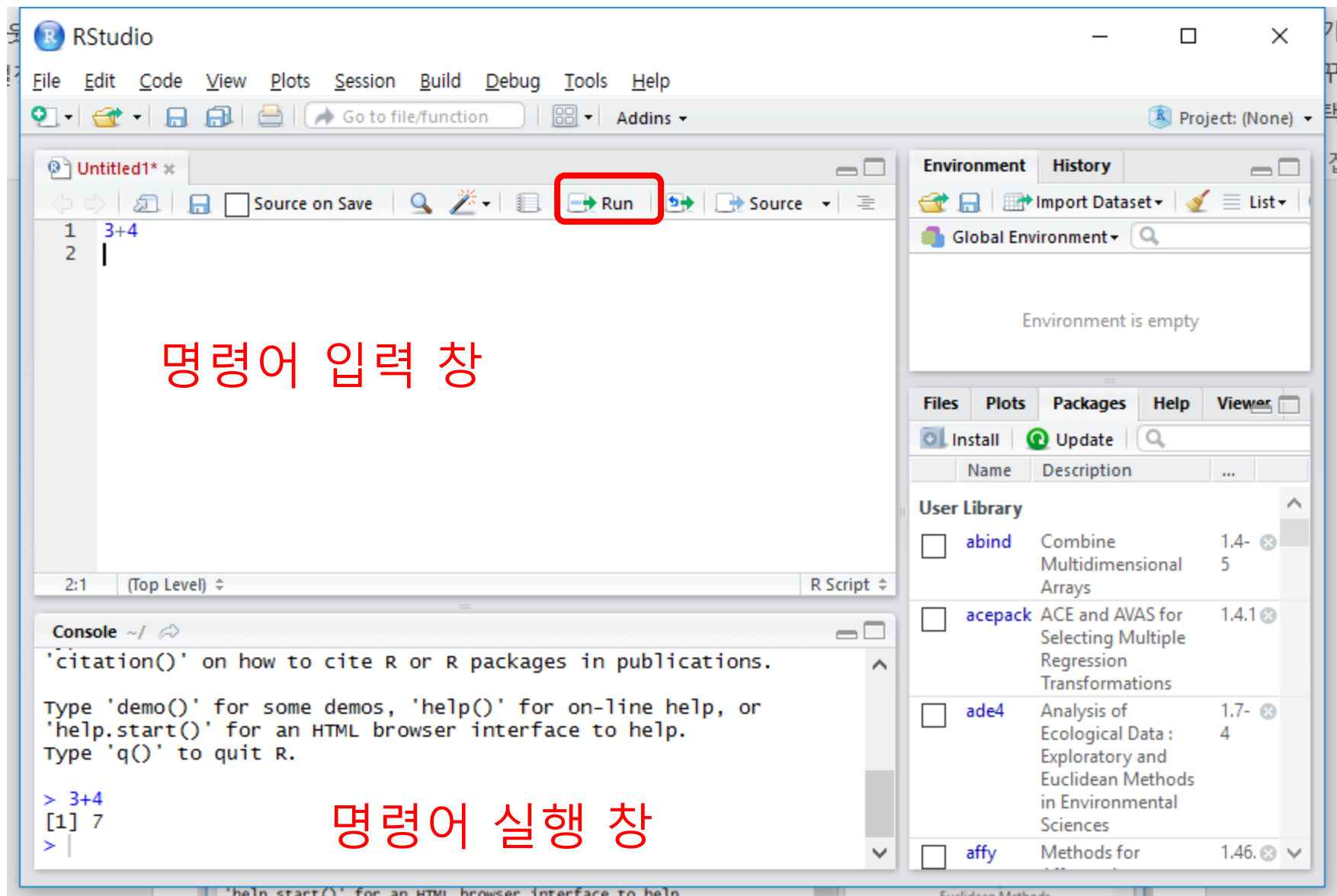
Zip/tar archives	Size	Date
RStudio 1.0.136 - Windows Vista/7/8/10	117.5 MB	2016-12-21
RStudio 1.0.136 - Ubuntu 12.04+/Debian 8+ (32-bit)	86.2 MB	2016-12-21
RStudio 1.0.136 - Ubuntu 12.04+/Debian 8+ (64-bit)	93.2 MB	2016-12-21
RStudio 1.0.136 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	85.4 MB	2016-12-21
RStudio 1.0.136 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	86.6 MB	2016-12-21

3. Install R, Rstudio


- R Studio 초기화면



3. Install R, Rstudio

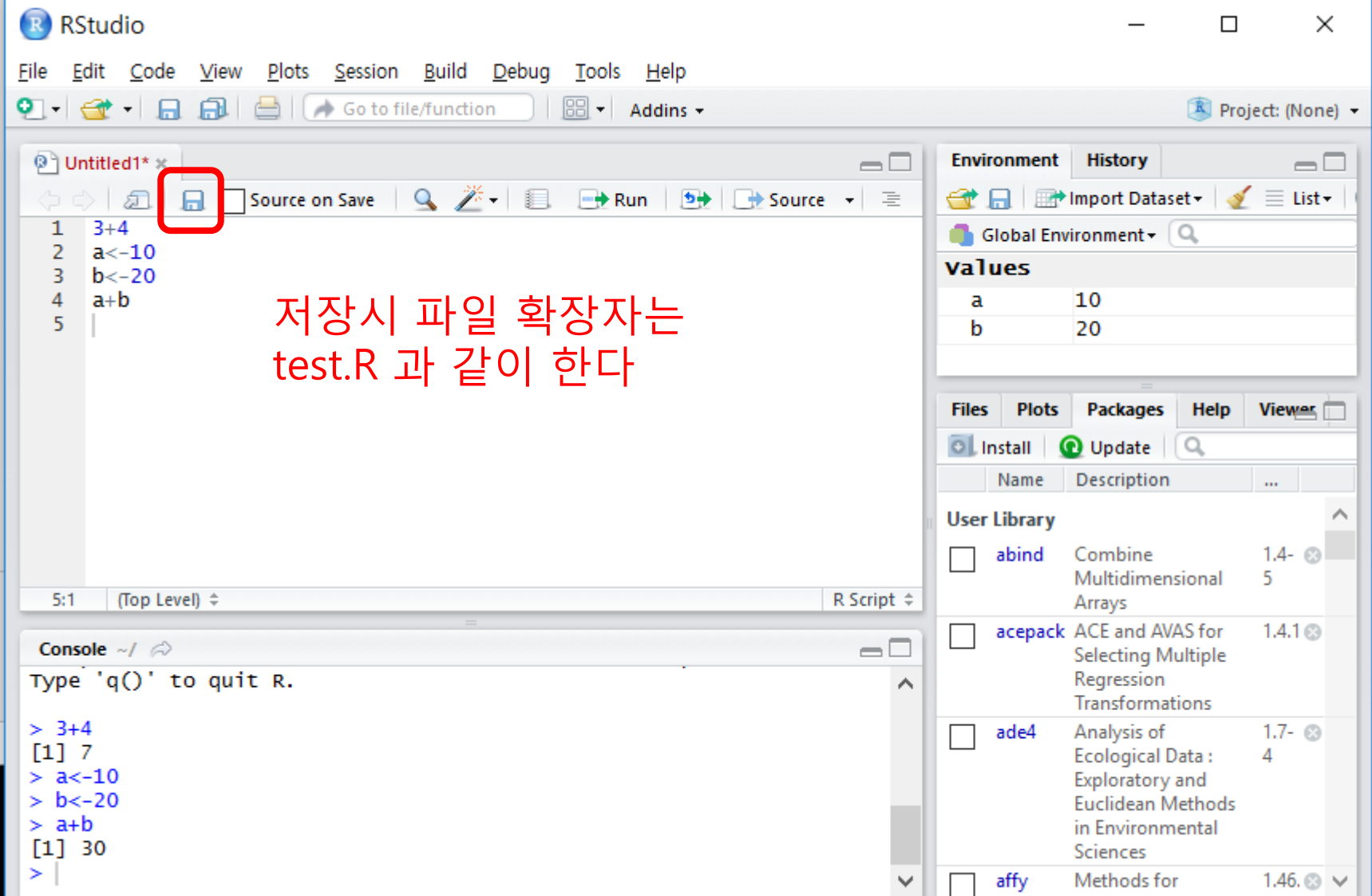


3. Install R, Rstudio

-  : 명령어 실행 버튼
 - ◉ 현재 커서가 위치한 라인의 명령어 실행
 - ◉ 여러 명령어 라인을 선택한 상태이면 선택한 명령어들을 실행
 - ◉ 명령어 실행 버튼을 클릭하는 대신 <Ctrl> + <Enter> 를 해도 된다

3. Install R, Rstudio

- 현재 작업 내용 (명령어들) 파일에 저장하기



The screenshot shows the RStudio interface. In the top-left pane, a script named 'Untitled1*' is open, containing the following R code:

```
1 3+4
2 a<-10
3 b<-20
4 a+b
5
```

The 'Save' icon (a floppy disk) in the toolbar is highlighted with a red square. Below the script editor, the console shows the execution of the code:

```
> 3+4
[1] 7
> a<-10
> b<-20
> a+b
[1] 30
>
```

On the right side, the 'Environment' pane shows the current values of variables:

values	
a	10
b	20

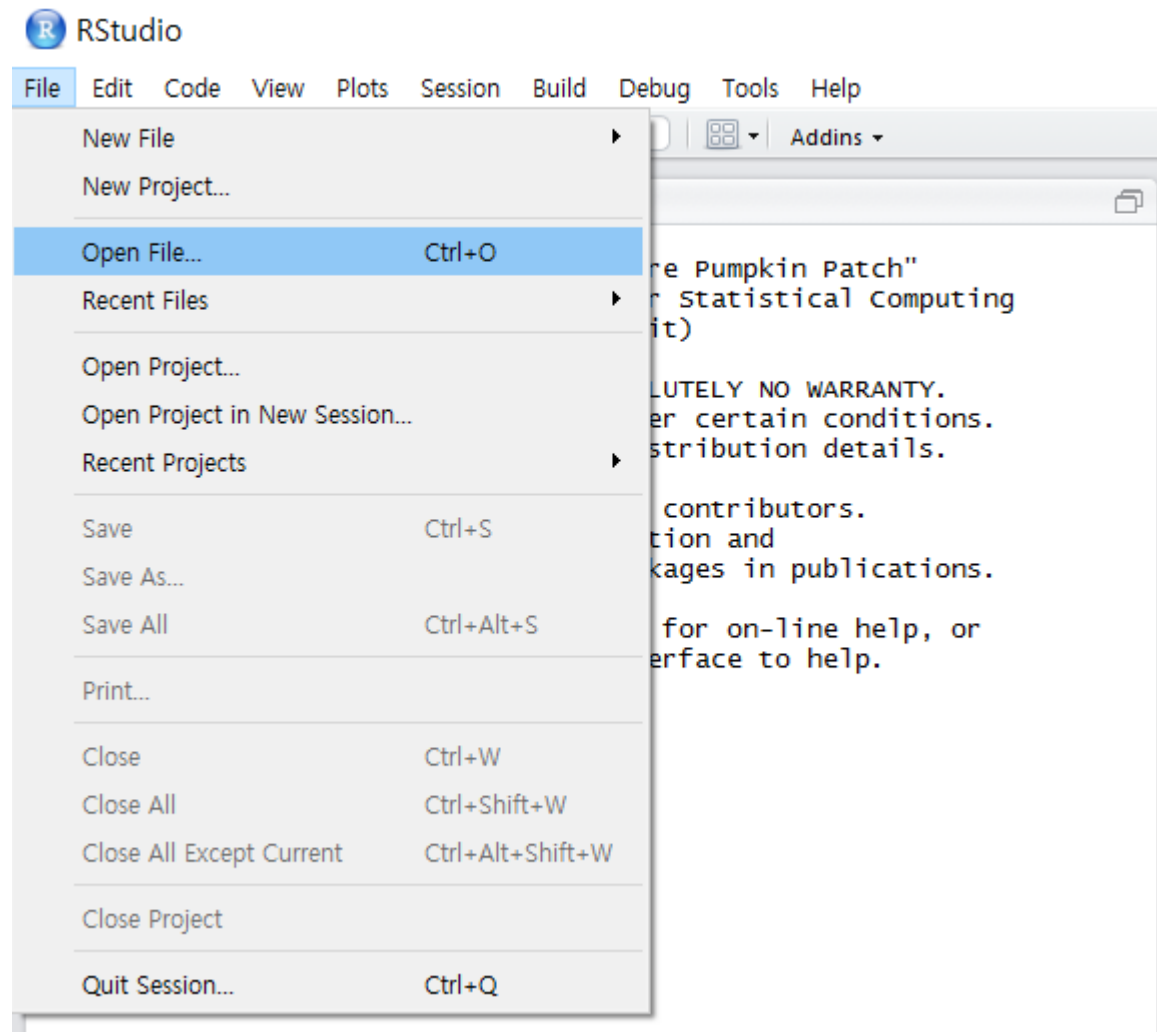
Below the Environment pane, the 'Packages' pane lists installed and available packages:

Name	Description	Version
abind	Combine Multidimensional Arrays	1.4-5
acepack	ACE and AVAS for Selecting Multiple Regression Transformations	1.4.1
ade4	Analysis of Ecological Data : Exploratory and Euclidean Methods in Environmental Sciences	1.7-4
affy	Methods for	1.46.

Red text overlaid on the script editor reads: 저장시 파일 확장자는 test.R 과 같이 한다 (When saving, the file extension should be like test.R).

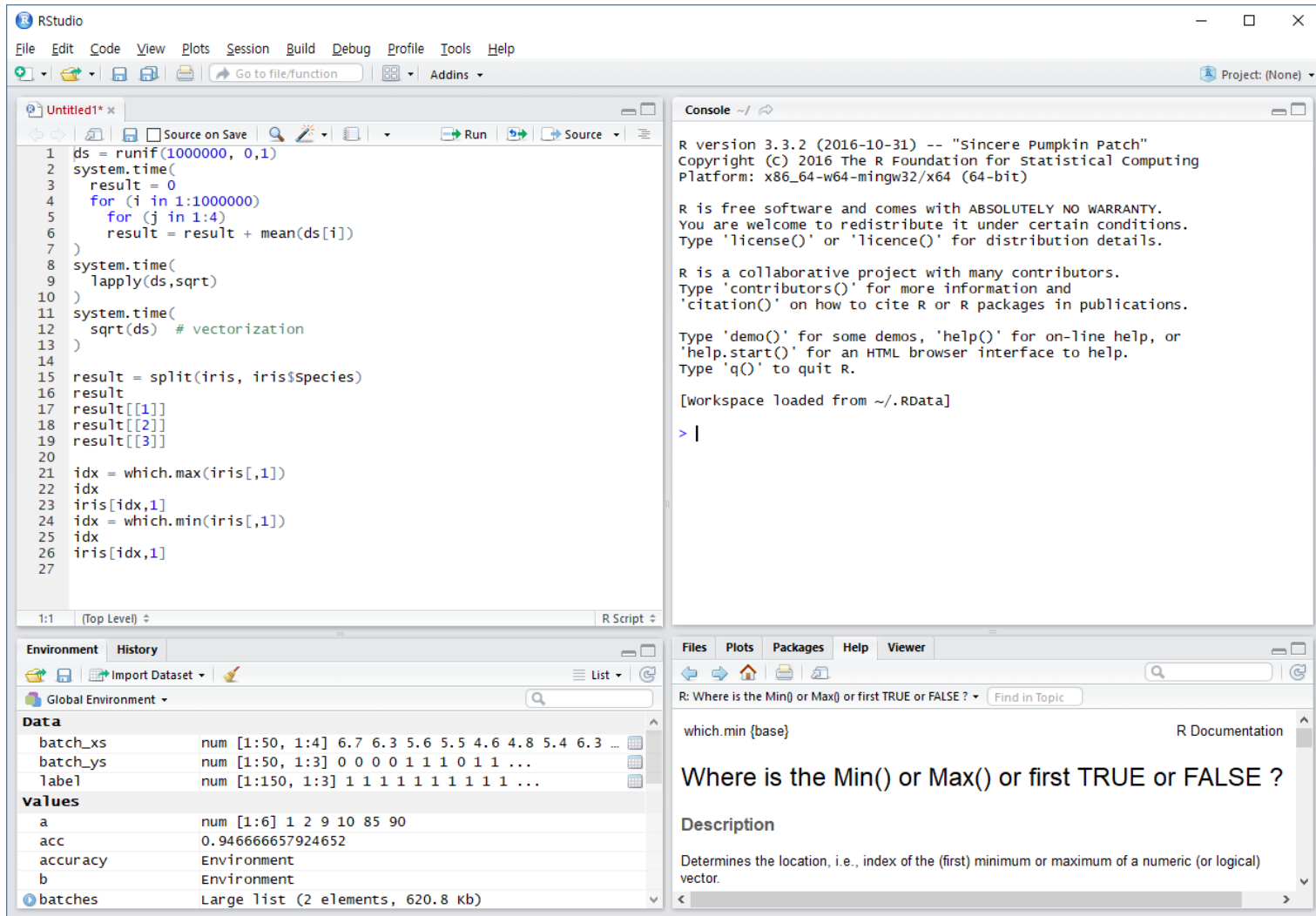
3. Install R, Rstudio

- R script 파일 불러오기



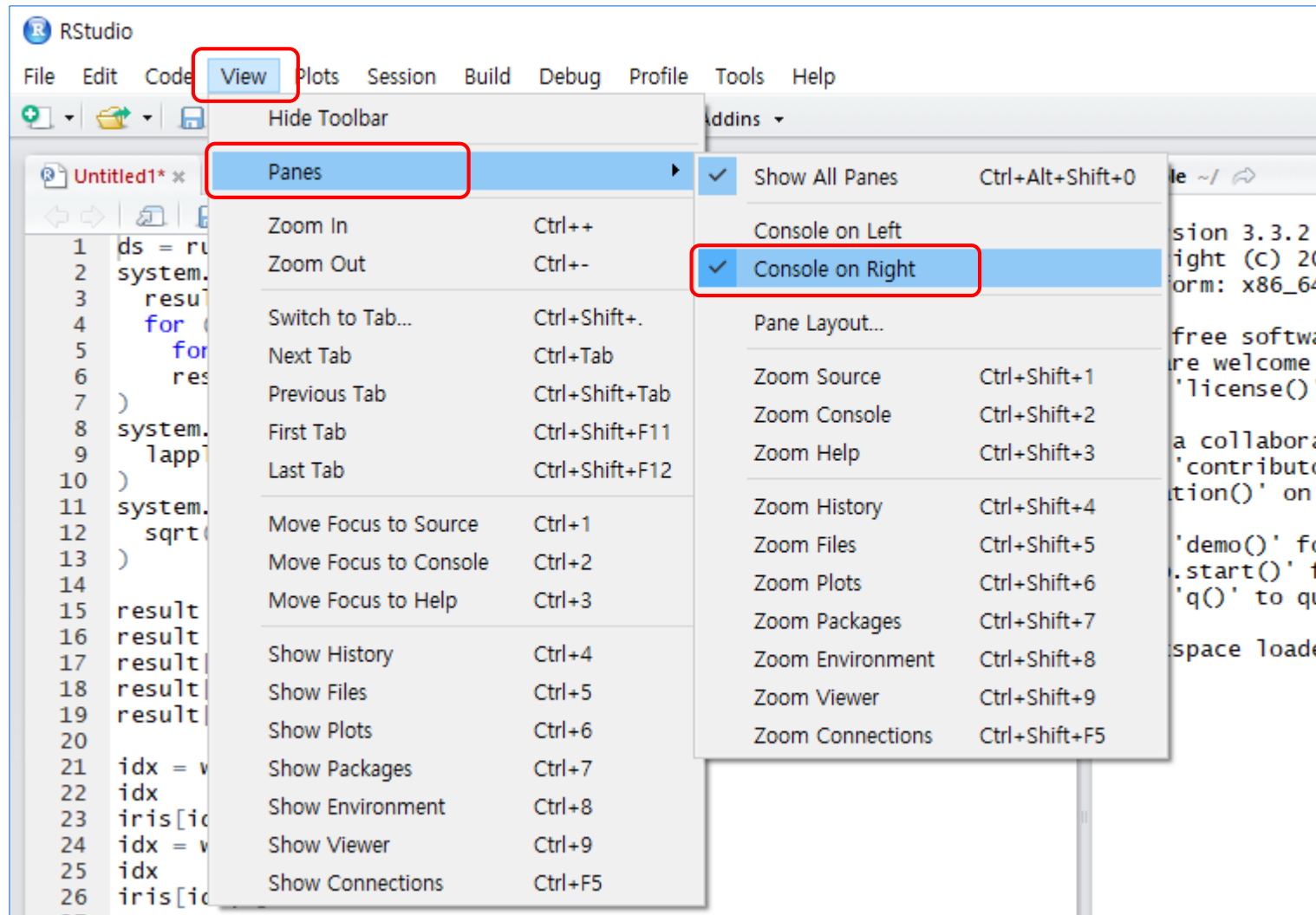
3. Install R, Rstudio

- 콘솔 화면이 오른쪽으로 가도록 조정하기



3. Install R, Rstudio

- 콘솔 화면이 오른쪽으로 가도록 조정하기



[R Tip]

- Rstudio 단축키

- ✓ 코드실행 :
- ✓ 주석처리/제거:
- ✓ 콘솔창이동 :
- ✓ 스크립트창 이동 :

[Ctrl] + [Enter]

[Ctrl] + [Shift] + [c]

[Ctrl] + [1]

[Ctrl] + [2]

