



2주차: 데이터 사이언스 세계로

ChulSoo Park

School of Computer Engineering & Information Technology

Korea National University of Transportation



학습목표 (2주차)

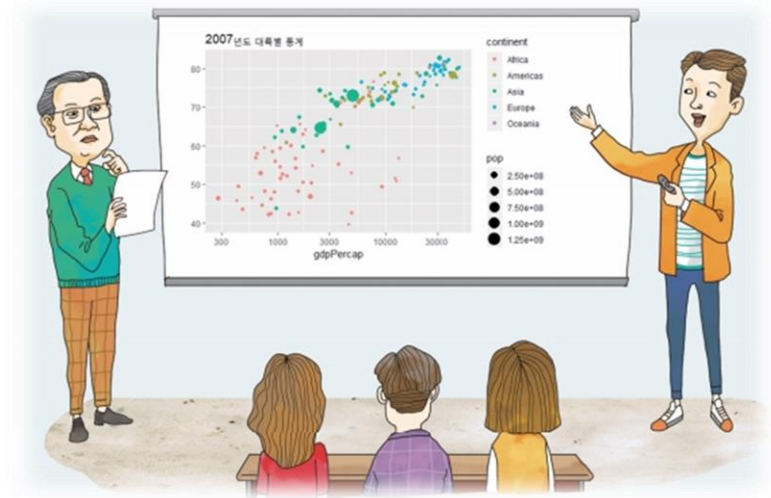
- ❖ 분석 도구 챙기기(R, Rstudio 설치)
- ❖ 분석 도구(R) 익히기
- ❖ R 분석도구로 데이터 핸들링
- ❖ 데이터 시각화 맛보기
- ❖ 데이터 과학 Process 이해



02

CHAPTER

데이터 과학으로 풍덩



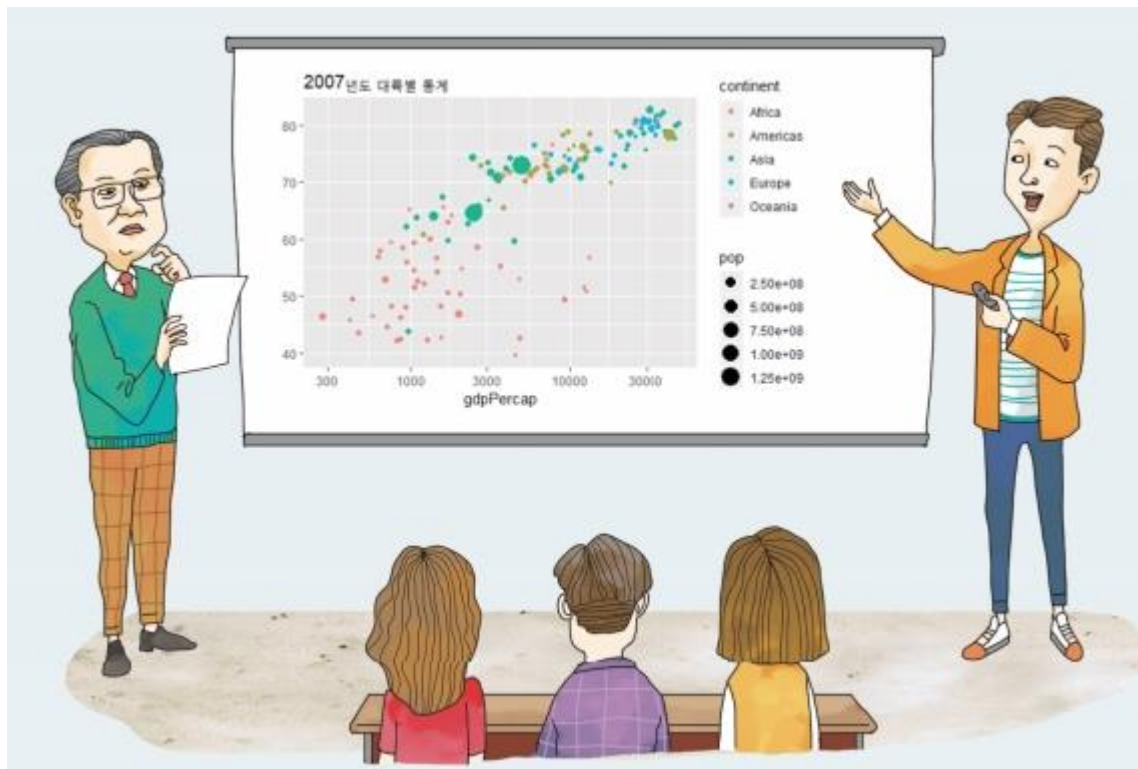
CONTENTS

- 2.1 도구 챙기기
- 2.2 데이터와 친해지기
- 2.3 데이터 시각화 맛보기
- 2.4 데이터 과학을 위한 좋은 습관 알아보기
- 2.5 좋은 도구 익히기
- 2.6 데이터와 더 친해지기
- 요약



■ 데이터 과학

- 데이터로부터 유용한 정보와 통찰을 끄집어내고 합리적인 의사결정을 돕는 현대적인 학문 분야



4차산업혁명 : 사물인터넷(IoT),빅데이터(Big Data),인공지능(AI) → 의사결정,현업활용

출처: 제4차 산업혁명, 김진호,최용주, p25



4차산업혁명 환경하의 디지털 경영혁신

이성열 · 강성근 · 김순신 지음



기업은 변화의 혁신을 통하여 성장한다. 4차 산업혁명과 디지털 혁신으로 바뀌고 있는 경영환경의 급격한 변화에서 전통기업들이 일어나 신속하게 디지털 경영혁신을 통하여 변화에 나갈 수 있는 가는 향후 기업생존의 핵심이 될 것이다.



	제1차 산업혁명	제2차 산업혁명	제3차 산업혁명	제4차 산업혁명
시기	18세기	19~20세기 초	20세기 후반	20세기
특징	증기기관 기반의 '기계화 혁명'	전기 에너지 기반의 '대량생산 혁명'	컴퓨터와 인터넷 기반의 '디지털 혁명'	사물인터넷(IoT)과 빅데이터, 인공지능(AI) 기반의 '만물 초지능 혁명'
영향	수공업 시대에서 증기기관을 활용한 기계가 물건을 생산하는 기계화 시대로 변화	전기와 생산조립 라인의 출현으로 대량생산 체계 구축	반도체와 컴퓨터, 인터넷 혁명으로 정보의 생성·가공·공유를 가능케 하는 정보기술시대의 개막	사람, 사물, 공간을 연결하고 자동화 지능화되어 디지털·물리적·생물학적 영역의 경계가 사라지면서 기술이 융합되는 새로운 시대

자료 : 미래에셋대우 글로벌투자전략부



■ 데이터 과학 관련 학문 분야

Data Science Is Multidisciplinary
By Brenden Tierney, 2012

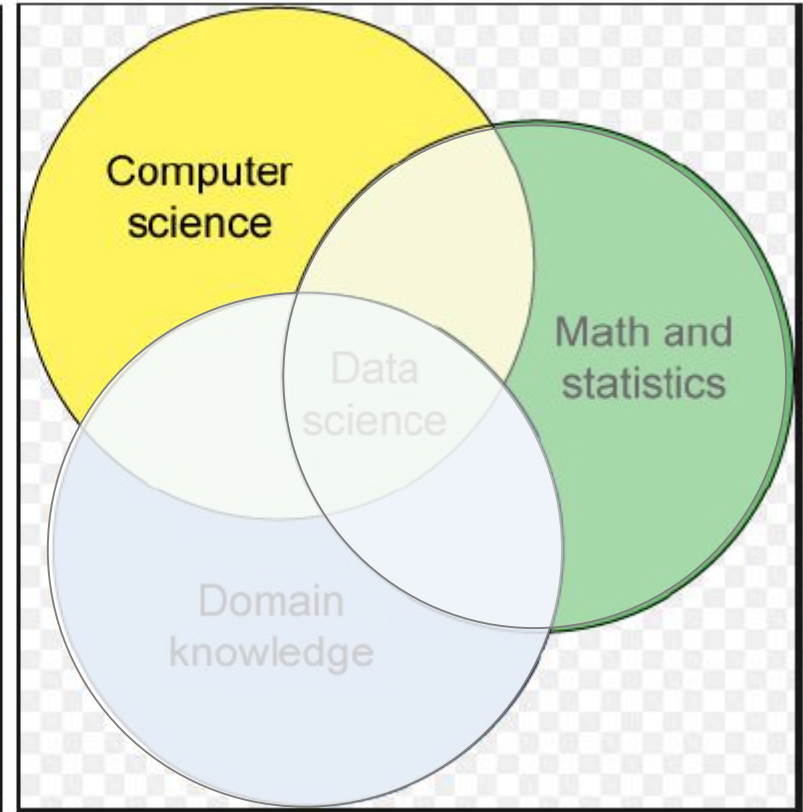
이수구분	교과목명	이수구분	교과목명	이수구분	교과목명
학문의기초	통계계산프로그래밍	학문의기초	컴퓨터프로그래밍I	전공선택	빅데이터분석
학문의기초	전산수학I	학문의기초	컴퓨터프로그래밍II	전공선택	데이터과학응용 I
학문의기초	행렬이론	전공선택	자료구조	전공선택	데이터과학응용 II
학문의기초	전산수학II	전공선택	알고리즘	전공선택	데이터과학응용 III
학문의기초	기초통계학	전공선택	컴퓨터구조	전공선택	데이터과학응용 IV
학문의기초	미적분학및연습 I	전공선택	운영체제	전공선택	졸업프로젝트(캡스톤디자인) I
학문의기초	통계수학	전공선택	컴퓨터네트워크	전공선택	졸업프로젝트(캡스톤디자인) II
전공선택	확률및랜덤과정	전공선택	인공지능	전공선택	기업현장학습 I
전공선택	데이터실험설계및분석	전공선택	정보보호	전공선택	기업현장학습 II
전공선택	선형대수 I 및연습	전공선택	기계학습		
전공선택	확률과정개론	전공선택	데이터베이스		
전공선택	확률론 입문	전공선택	블록최적화입문		
전공선택	수리통계학	전공선택	인간컴퓨터상호작용		
전공선택	시계열분석	전공선택	인터랙티브시각화		
전공선택	통계계산방법	전공선택	자연어처리		
전공선택	회귀분석	전공선택	정보검색		
전공선택	범주형자료분석	전공선택	딥러닝		
전공선택	통계적데이터과학 I	전공선택	기초컴퓨터비전이론 및 응용		
전공선택	다변량통계분석	전공선택	고급기계학습		
전공선택	베이지스통계입문	전공선택	디자인사고		
전공선택	통계적데이터과학 II	전공선택	데이터시각화		
전공선택	통계적머신러닝	전공선택	강화학습		
전공선택	선형방정식	전공선택	데이터프리젠테이션		

네이버와 넥슨의 신입 개발자 초임은 지난해 고용노동부가 공개한 직원 500명 이상 대기업의 2020년 대졸 신입 사무직 근로자의 평균 연봉 3,347만 원에 비하면 1.5배에 달하는 수준이다. 2021. 2. 1.

주요 ICT기업 초임 연봉 수준

네이버	5,000만원 (개발직군) 4,500만원 (비개발직군)
넥슨	5,000만원 (개발직군) 4,500만원 (비개발직군)
삼성전자	4,500만원 (대졸 공채 평균)
SK텔레콤	4,900만원 (대졸 공채 평균)

자료 : 각 사 취업

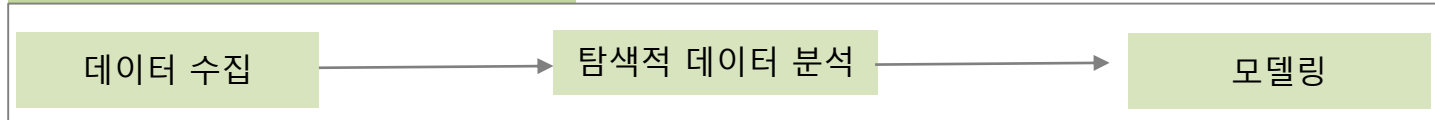


출처 : www.oralytics.com

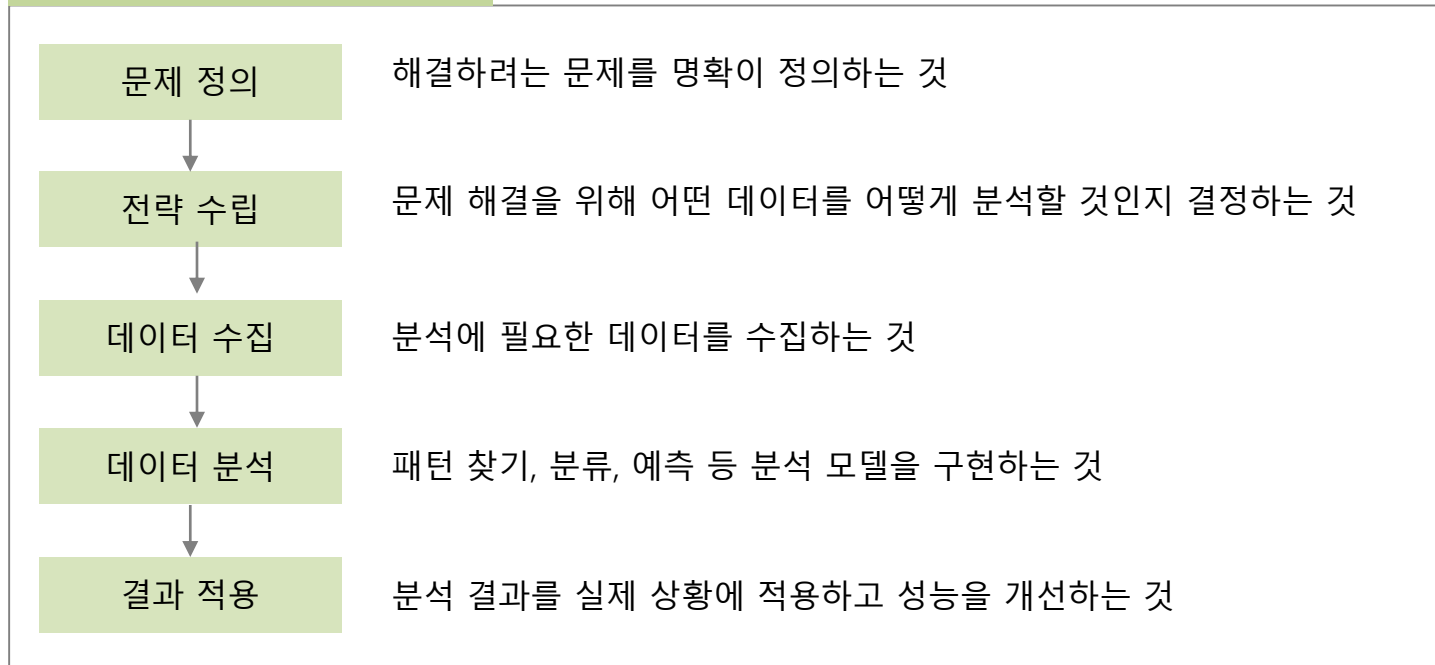
데이터 사이언스 학과(대학원)



데이터 과학의 절차(그림 1- 5)

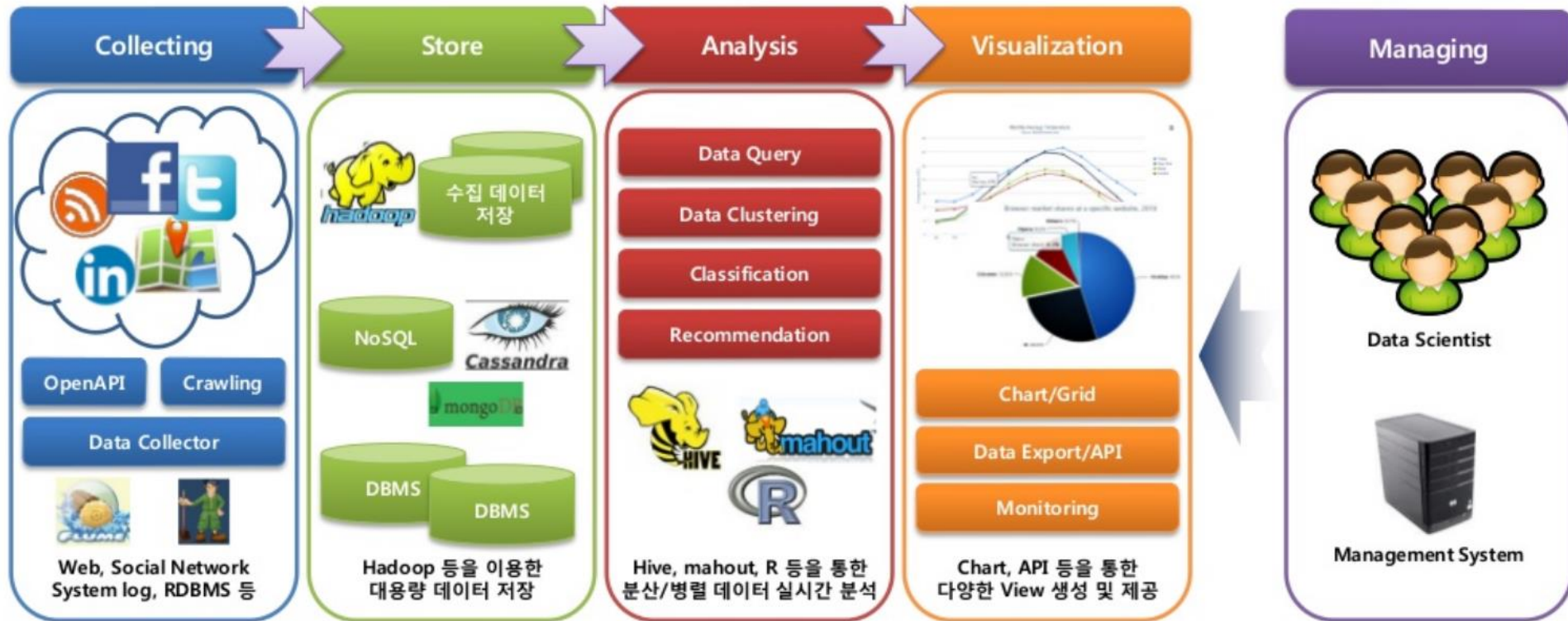


데이터 사이언스 프로세스

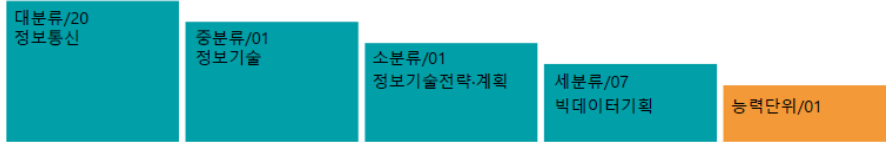


솔루션 기능 요건 : Big Data Solution은 방대하고 복잡한 정형/비정형 Data의 수집, 정방(**modify**), 분석과 다양한 뷰를 통한 정보 제공의 4가지 핵심기능을 기본으로 각 기능을 담당할 인력(Data Scientist) 및 관리 시스템을 필요로 함.

Function of Big Data Solution



■ 국가직무능력표준(NCS)



- ◆ 1능력단위: 3~5개 능력단위요소로 구성
- ◆ 1능력단위 : 80 Page 내외

목적 : 빅데이터 기획은 **가치 창출**을 위하여 빅데이터 서비스, 환경분석, 분석모델 및 플랫폼을 정의하고 이에 대한 활용과 성과 및 운영관리를 수립하는 일이다.

NCS학습모듈 빅데이터 서비스 기획

능력단위	수준	능력단위요소
빅데이터 서비스기획	7	빅데이터서비스목표수립하기
		빅데이터 서비스 모델 수립하기
		빅데이터 서비스 전략 수립하기
		빅데이터 서비스 정책 수립하기
빅데이터 환경분석	5	빅데이터 서비스 환경 분석하기
		빅데이터 데이터 환경 분석하기
		빅데이터 기술 환경 분석하기
		빅데이터 인력 환경 분석하기
빅데이터 분석 기획	5	빅데이터 분석 요건 정의하기
		빅데이터 분석 데이터 확보 기획하기
		빅데이터 분석 데이터 탐색 기획하기
		빅데이터 분석 모델링 기획하기
빅데이터 기술 플랫폼 기획	5	빅데이터 분석 결과 적용 계획하기
		빅데이터 아키텍처 기획하기
		빅데이터 수집 기술 기획하기
		빅데이터 저장 기술 기획하기
빅데이터 성과 관리 기획	6	빅데이터 처리 기술 기획하기
		빅데이터 성과기준 기획하기
		빅데이터 성과측정 방법 기획하기
		빅데이터 성과평가 실행 기획하기
빅데이터 활용 기획	7	빅데이터 성과평가 피드백 기획하기
		빅데이터 분석 결과 활용 기획하기
		빅데이터 비즈니스 모델 활용 기획하기
		빅데이터 시장 활용 기획하기
빅데이터 운영 기획	7	빅데이터 상품화 기획하기
		빅데이터 품질관리 기획하기
		빅데이터 보안관리 기획하기
		빅데이터 조직 수립하기
	7	빅데이터 운영관리 기획하기



차이코프스키 국제 바이올린 경연대회 바이올린 제작 부문 대상 수상



시상 기준은 무언인가요?

많은 콩쿠르가 있지만 대부분이 100점 만점

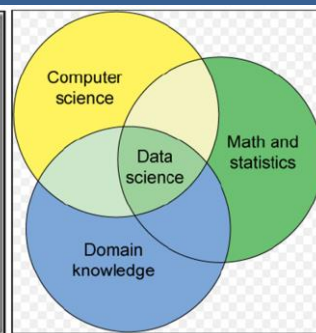
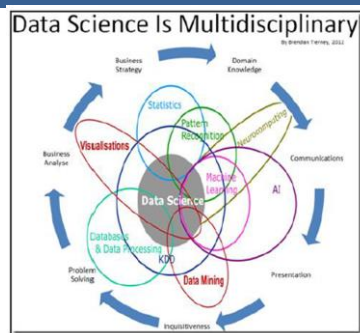
제작 부문 50점 : 소재, 디자인, 완성도 등

소리 부분 50점 : 실제로 연주해 보았을 때 어떤 **소리**의 울림



2. 심의 일정 및 대상 종목

- (회의 일시) '18. 11. 8(목), 13:30~
- (회의 장소) 서울 LW컨벤션 회의실 3층
- (심의 대상) 빅데이터분석기사, 신발개발산업기사, 타워크레인설치·해체기능사



연번	세부직무분야	유형	종목명	소관부처(담당과)
1	정보관리 (추가)	신설	빅데이터분석기사	과학기술정보통신부(융합신산업과), 통계청(통계정책과)
2	신발	신설	신발개발산업기사	산업통상자원부(섬유세라믹과)
3	건설기계 (추가)	신설	타워크레인설치·해체기능사	고용노동부(산업안전과)

* (국가기술자격 정책심의위원회 운영규정 제6조 제2항) 종목의 신설 등으로 해당하는 세부직무 분야가 없는 경우에는 관련성이 큰 세부직무분야별전문위원회에 「국가기술자격법 시행규칙」 제7조에 따른 주무부장관의 추천을 받은 5명 이내의 해당 분야 전문가를 추가하여 심의할 수 있음



■ 데이터 과학을 배우는 가장 좋은 방법은 예제에 의한 학습

- 수영: 물에 뛰어들어 걸어다니고 물장구 치고...
- 목공: 참나무, 소나무 등에 드릴링해보며 나무 강도를 익히고...
- 데이터 과학: 데이터를 가지고 그려보고 통계를 내보고...

■ 데이터는 천의 얼굴

- 거리나 체중을 나타내는 실수, 혈액형을 나타내는 범주값, 주소나 이름의 문자열 등

■ 이 책의 주요 목표

- 독자가 다양한 데이터를 접하도록 유도하고 배려
- 이 장에서는 단순한 몇 가지 데이터를 다룸 (cars, women, iris, tips)



[illegible]

2.1 도구 챙기기

■ 가장 먼저 좋은 도구를 챙겨야 함

- 목공에서는 드릴, 톱, 망치, 테니스를 배운다면 좋은 라켓과 공 → 비용 발생
- 데이터 과학 → 무료 설치, 무료 업그레이드, 10분 정도면 설치 가능

■ R 설치

1. <https://www.r-project.org> 접속
2. CRAN 클릭 후, Korea 서버의 URL 클릭
3. 운영체제 선택
4. Subdirectories에서 base 선택
5. 64비트와 32비트 중에 선택(지금은 대부분 컴퓨터가 64비트임)
6. 설치 파일을 다운로드한 후, 안내에 따라 설치



■ 먼저 설치된 R 툴 버전 확인

제어판 > 프로그램 > 프로그램 및 기능

프로그램 제거 또는 변경

프로그램을 제거하려면 목록에서 선택한 후 [제거], [변경] 또는 [복구]를 클릭하십시오.

구성 ▾				
이름	게시자	설치 날짜	크기	버전
Pixillion Image Converter	NCH Software	2020-06-02	3.58MB	7.12
Prism 동영상 파일 변환기	NCH Software	2020-06-02	3.49MB	6.39
R for Windows 4.0.2	R Core Team	2020-07-04	204MB	4.0.2
RansomDefender	Clonix Co., Ltd.	2020-05-04		1.0
Realtek Card Reader	Realtek Semiconductor Corp.	2020-05-04	14.6MB	10.0.18362.31255
Realtek Ethernet Controller Driver	Realtek	2020-05-04	5.09MB	10.36.701.2019
Realtek High Definition Audio Driver	Realtek Semiconductor Corp.	2020-05-04	83.8MB	6.0.8858.1
REPORTX	DigitalZone Co., Ltd.	2021-01-14	9.23MB	
RStudio	RStudio	2020-07-04		1.3.959
Samsung DPI Configuration	Samsung Electronics Co., Ltd.	2020-05-04	11.3MB	1.0.17
Samsung USB Driver for Mobile Phones	Samsung Electronics Co., Ltd.	2020-09-07	58.1MB	1.7.31.0
Smart Switch	Samsung Electronics Co., Ltd.	2020-09-07	68.2MB	4.2.20072.4
Switch 사운드 파일 변환기	NCH Software	2020-06-02	5.37MB	8.10
TouchEn nxKey with E2E for 32bit	RaonSecure Co., Ltd.	2021-01-05	11.1MB	1.0.0.75
Veraport(보안모듈 관리 프로그램) G3 - 3,8,5,1	Wizvera	2021-01-05	20.4MB	3,8,5,1
VLC media player	VideoLAN	2021-01-08		3.0.11

■ 먼저 설치된 R 툴 삭제 확인

▸ 프로그램 ▸ 프로그램 및 기능



프로그램 제거 또는 변경

프로그램을 제거하려면 목록에서 선택한 후 [제거], [변경] 또는 [복구]를 클릭하십시오.

구성 ▾

이름	게시자	설치 날짜	크기	버전
Pixillion Image Converter	NCH Software	2020-06-02	3.58MB	7.12
Prism 동영상 파일 변환기	NCH Software	2020-06-02	3.49MB	6.39
RansomDefender	Clonix Co., Ltd.	2020-05-04		1.0
Realtek Card Reader	Realtek Semiconductor Corp.	2020-05-04	14.6MB	10.0.18362.31255
Realtek Ethernet Controller Driver	Realtek	2020-05-04	5.09MB	10.36.701.2019
Realtek High Definition Audio Driver	Realtek Semiconductor Corp.	2020-05-04	83.8MB	6.0.8858.1
REPORTX	DigitalZone Co., Ltd	2021-01-14	9.23MB	
Samsung DPI Configuration	Samsung Electronics Co., Ltd.	2020-05-04	11.3MB	1.0.17
Samsung USB Driver for Mobile Phones	Samsung Electronics Co., Ltd.	2020-09-07	58.1MB	1.7.31.0
Smart Switch	Samsung Electronics Co., Ltd.	2020-09-07	68.2MB	4.2.20072.4
Switch 사운드 파일 변환기	NCH Software	2020-06-02	5.37MB	8.10
TouchEn nxKey with E2E for 32bit	RaonSecure Co., Ltd.	2021-01-05	11.1MB	1.0.0.75
Veraport(보안모듈 관리 프로그램) G3 - 3,8,5,1	Wizvera	2021-01-05	20.4MB	3,8,5,1
VLC media player	VideoLAN	2021-01-08		3.0.11
WeGuardia™ SSLplus V2.0	Future Systems Inc.	2021-01-22	21.6MB	1.0.10
Windows 드라이버 패키지 - Arduino LLC (www.arduino.cc) A...	Arduino LLC (www.arduino.cc)	2020-07-10		01/04/2013 1.0.0.0
WIZVERA Process Manager 1,0,5,4	WIZVERA	2021-01-05	2.29MB	1,0,5,4



2.1 도구 챙기기

제어판 > 시스템 및 보안 > 시스템

컴퓨터에 대한 기본 정보 보기

Windows 버전

Windows 10 Enterprise

© 2019 Microsoft Corporation. All rights reserved.



시스템

프로세서: Intel(R) Core(TM) i5-10210U CPU @ 1.60GHz 2.11 GHz

설치된 메모리(RAM): 8.00GB(7.82GB 사용 가능)

시스템 종류: 64비트 운영 체제, x64 기반 프로세서

펜 및 터치: 이 디스플레이에 사용할 수 있는 펜 또는 터치식 입력이 없습니다.

컴퓨터 이름, 도메인 및 작업 그룹 설정

컴퓨터 이름: 박철수노트북1

전체 컴퓨터 이름: 박철수노트북1.박철수노트북1

컴퓨터 설명: 박철수 노트북1

작업 그룹: 박철수 가족

설정 변경

Windows 정품 인증

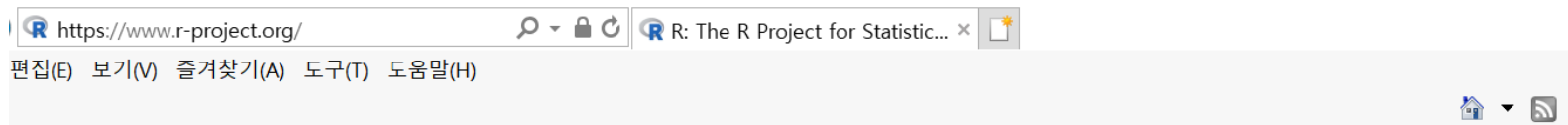
Windows 정품 인증을 받았습니다. [Microsoft 소프트웨어 사용 조건 읽기](#)

제품 ID: 00329-10330-60585-AA115

제품 키 변경



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



[Home]

Download

CRAN

R Project

About R

Logo

Contributors

What's New?

Reporting Bugs

Conferences

Search

Get Involved: Mailing Lists

Developer Pages

R Blog

R Foundation

Foundation

Board

Members

Donors

Donate

Help With R

Getting Help

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

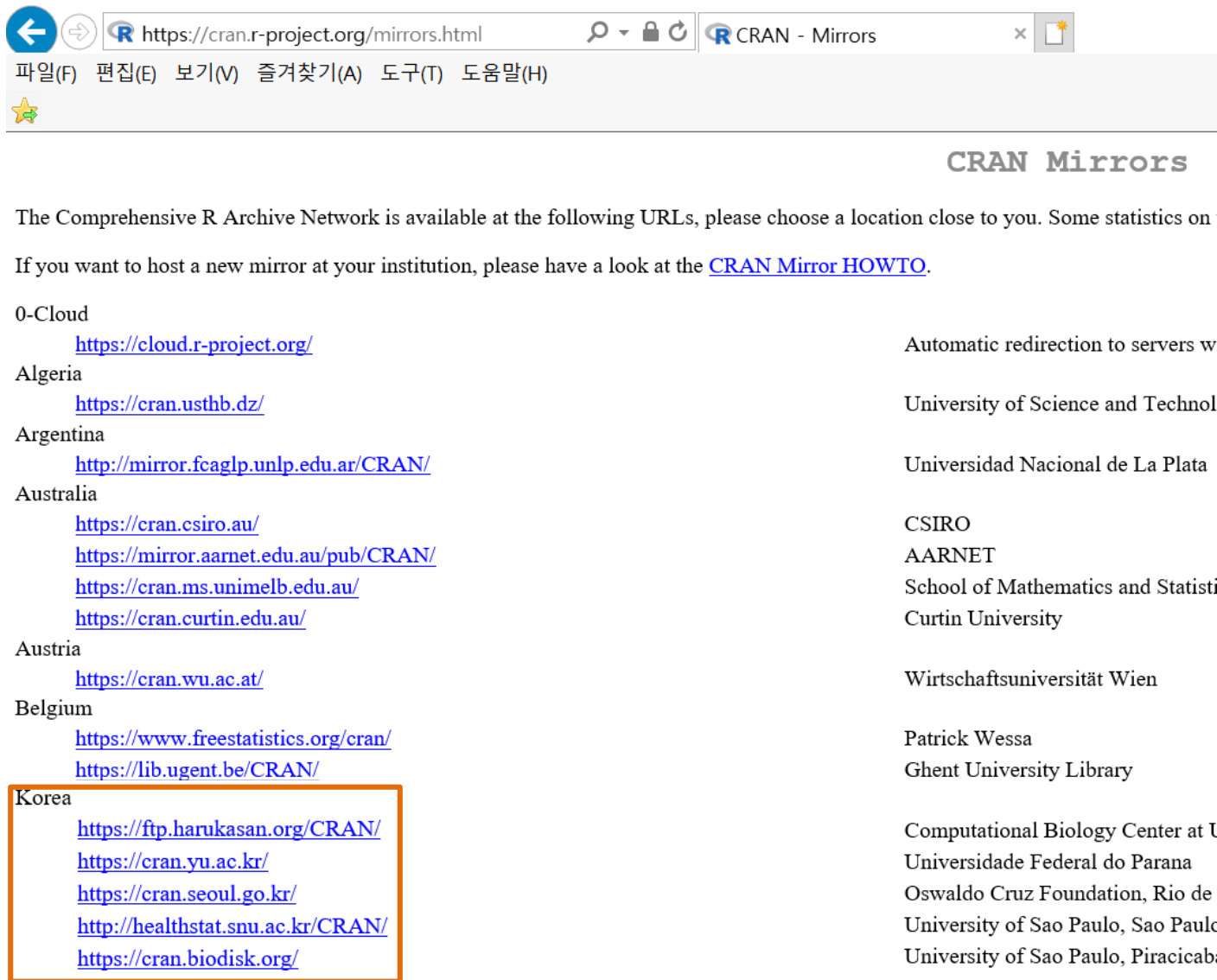
- **R version 4.0.4 (Lost Library Book) prerelease versions** will appear starting Friday 2021-02-05. Final release is scheduled for Monday 2021-02-15.
- **R version 4.0.3 (Bunny-Wunnies Freak Out)** has been released on 2020-10-10.
- Thanks to the organisers of useR! 2020 for a successful online conference. Recorded tutorials and talks from the conference are available on the [R Consortium YouTube channel](#).
- **R version 3.6.3 (Holding the Windsock)** was released on 2020-02-29.
- You can support the R Foundation with a renewable subscription as a [supporting member](#)

News via Twitter

[News from the R Foundation](#)



2.1 도구 챙기기



CRAN Mirrors

The Comprehensive R Archive Network is available at the following URLs, please choose a location close to you. Some statistics on

If you want to host a new mirror at your institution, please have a look at the [CRAN Mirror HOWTO](#).

0-Cloud	https://cloud.r-project.org/	Automatic redirection to servers w
Algeria	https://cran.usthb.dz/	University of Science and Technol
Argentina	http://mirror.fcaglp.unlp.edu.ar/CRAN/	Universidad Nacional de La Plata
Australia	https://cran.csiro.au/	CSIRO
	https://mirror.aarnet.edu.au/pub/CRAN/	AARNET
	https://cran.ms.unimelb.edu.au/	School of Mathematics and Statisti
	https://cran.curtin.edu.au/	Curtin University
Austria	https://cran.wu.ac.at/	Wirtschaftsuniversität Wien
Belgium	https://www.freeststatistics.org/cran/	Patrick Wessa
	https://lib.ugent.be/CRAN/	Ghent University Library
Korea	https://ftp.harukasan.org/CRAN/	Computational Biology Center at U
	https://cran.yu.ac.kr/	Universidade Federal do Parana
	https://cran.seoul.go.kr/	Oswaldo Cruz Foundation, Rio de
	http://healthstat.snu.ac.kr/CRAN/	University of Sao Paulo, Sao Paul
	https://cran.biodisk.org/	University of Sao Paulo, Piracicab



2.1 도구 챙기기

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

Source Code

Windows and Mac
be compiled be

- The latest
- Sources of
- Daily snapshots
corresponding
- Source code
- Contributed

Questions

- If you have questions about R, see the [frequently asked questions](#) below.

R for Windows

Subdirectories:

[base](#)

[contrib](#)

[old contrib](#)

[Rtools](#)

Please do not submit

You may also want

Note: CRAN does

Binaries for base distribution. This is what you want to [install R for the first time](#).

Binaries of contributed CRAN packages (for R \geq 2.13.x; managed by Uwe Ligges). There are also binaries available for CRAN Windows services and corresponding environment and make variables.

R-4.0.3 for Windows (32/64 bit)

[Download R 4.0.3 for Windows](#) (85 megabytes, 32/64 bit)

[Installation and other instructions](#)

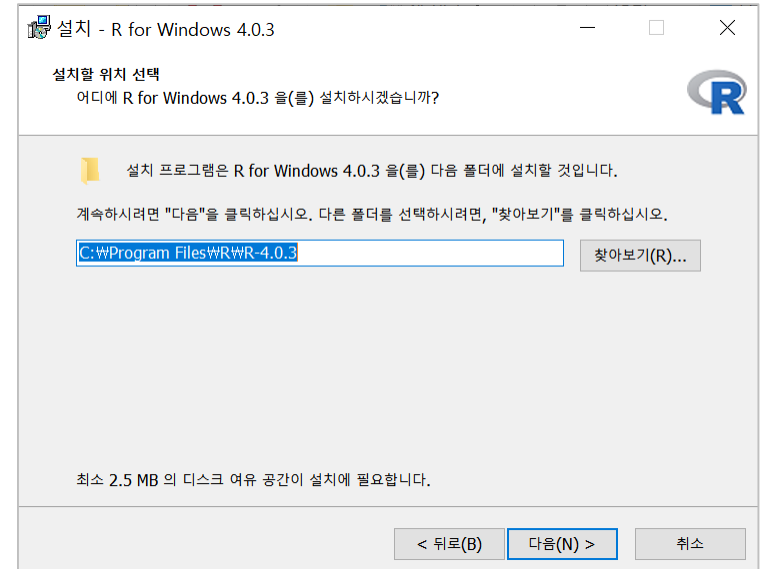
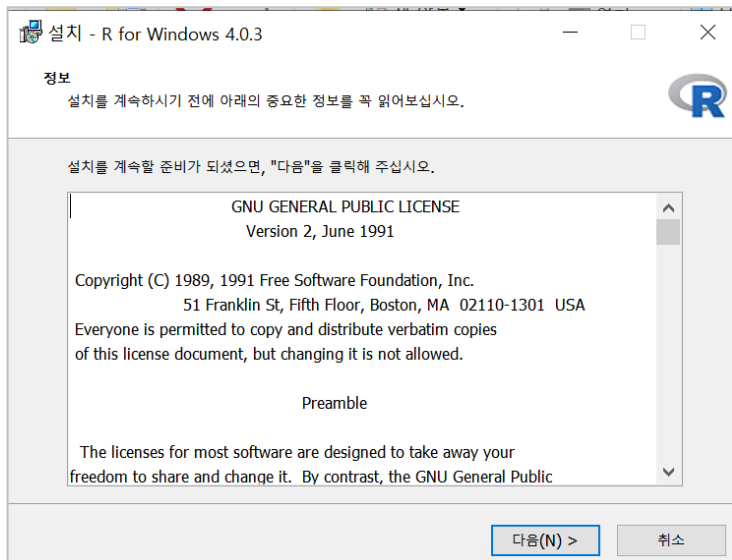
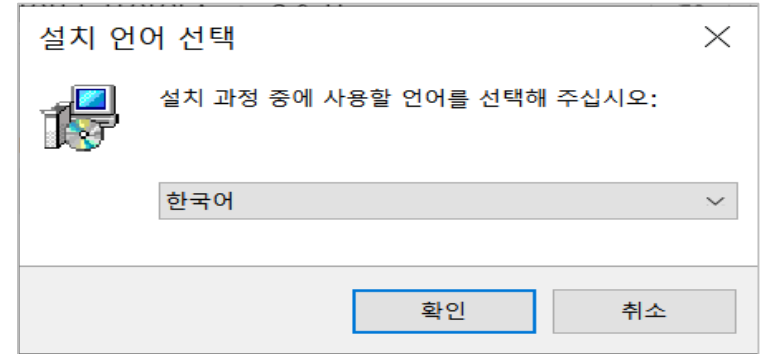
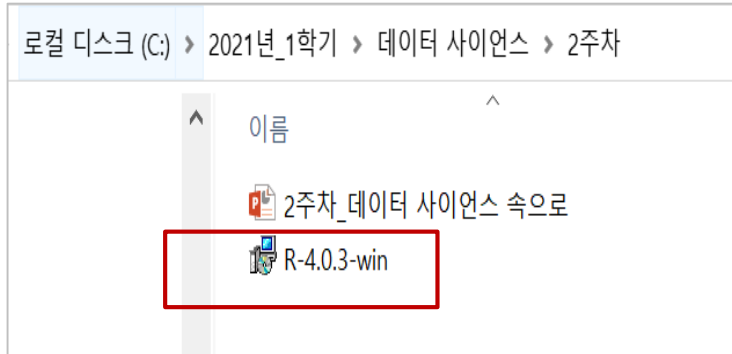
[New features in this version](#)

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

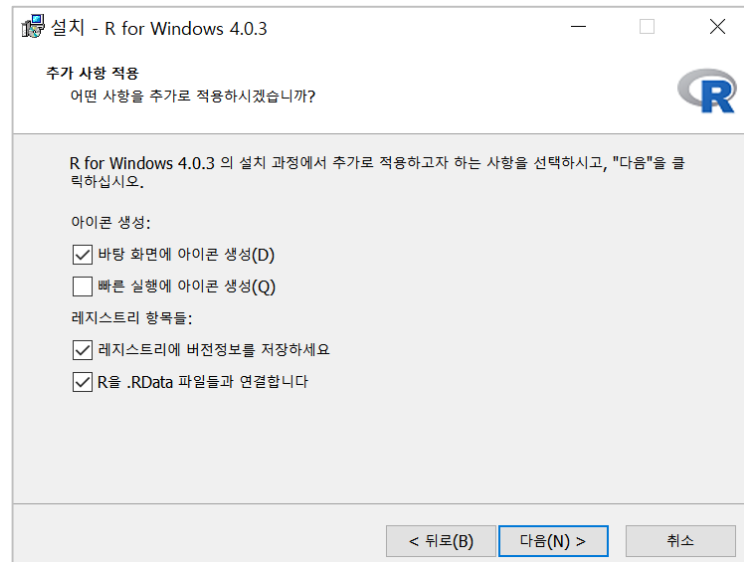
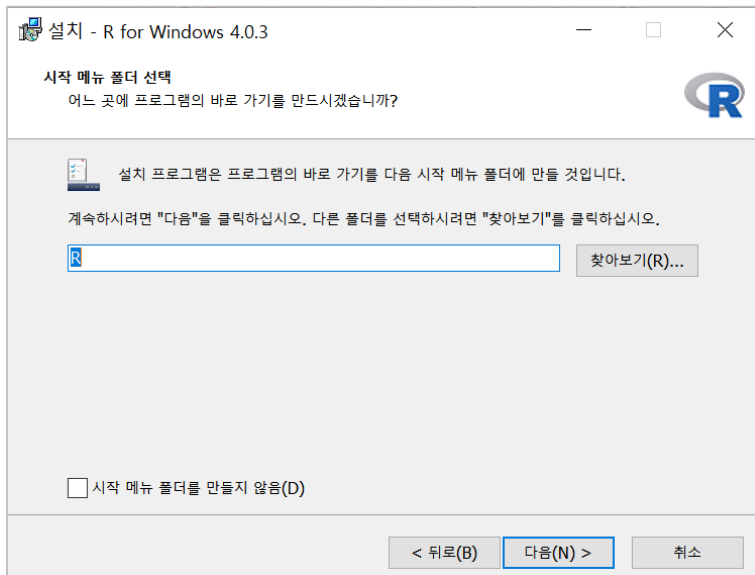
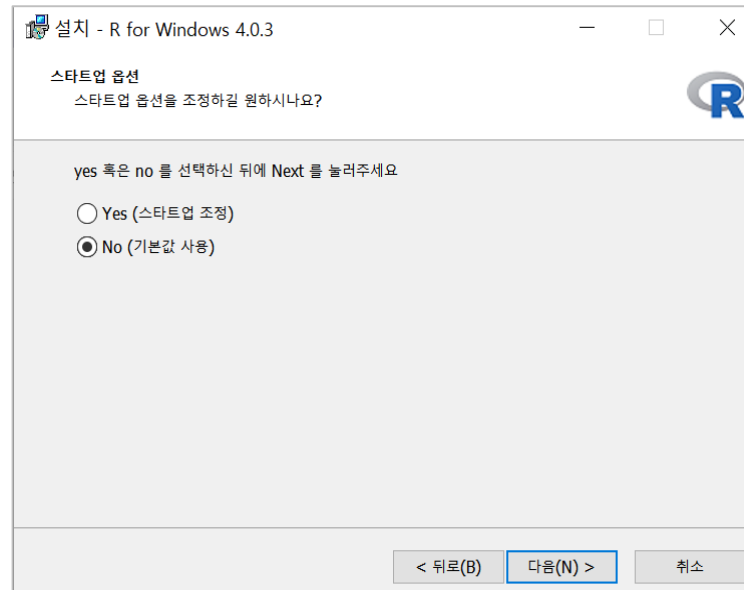
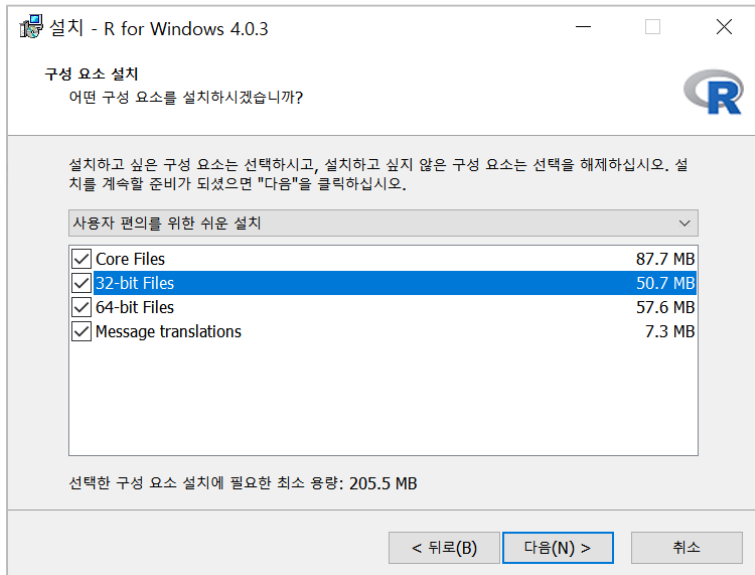
Frequently asked questions

- [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?](#)

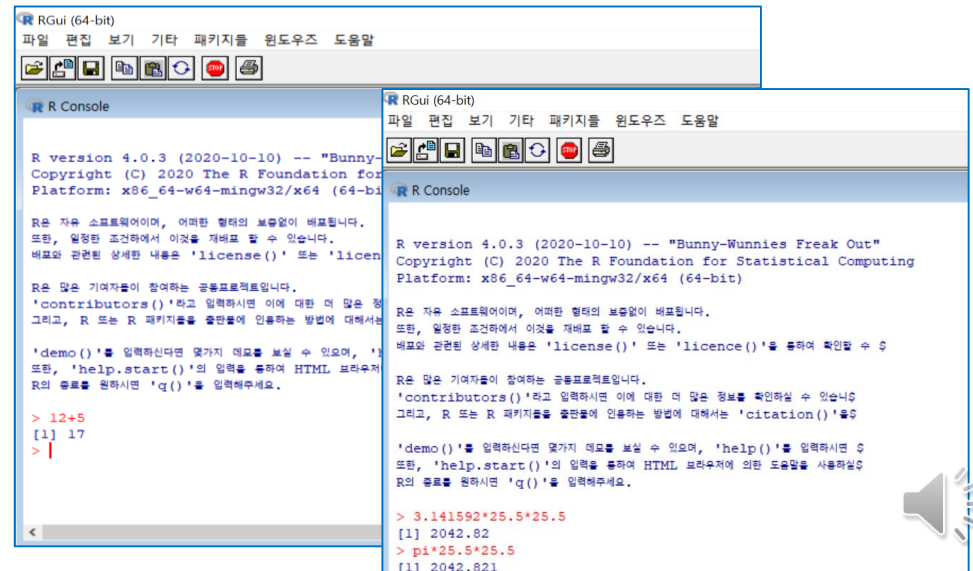
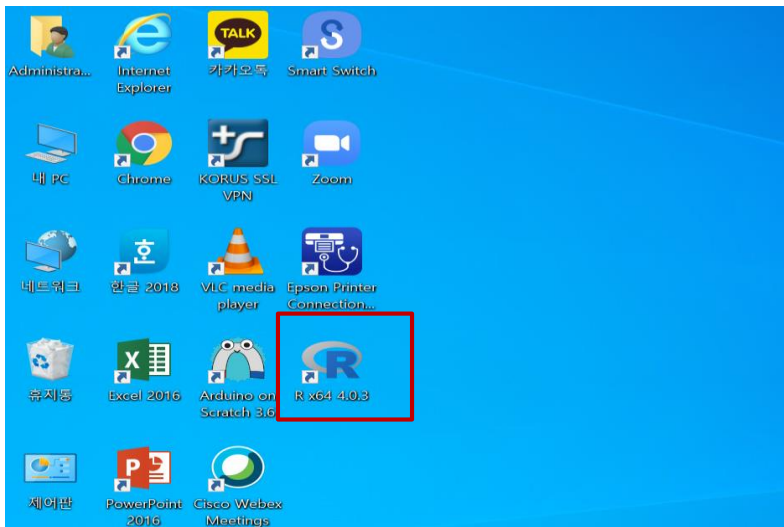
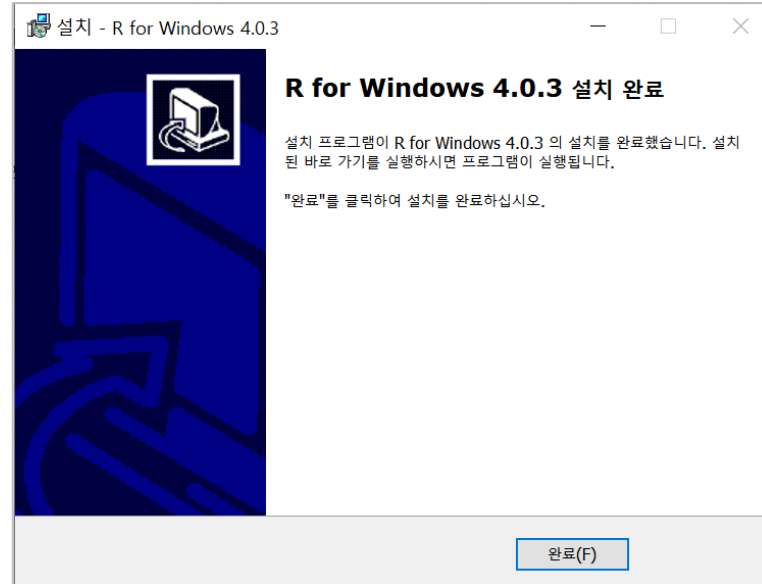
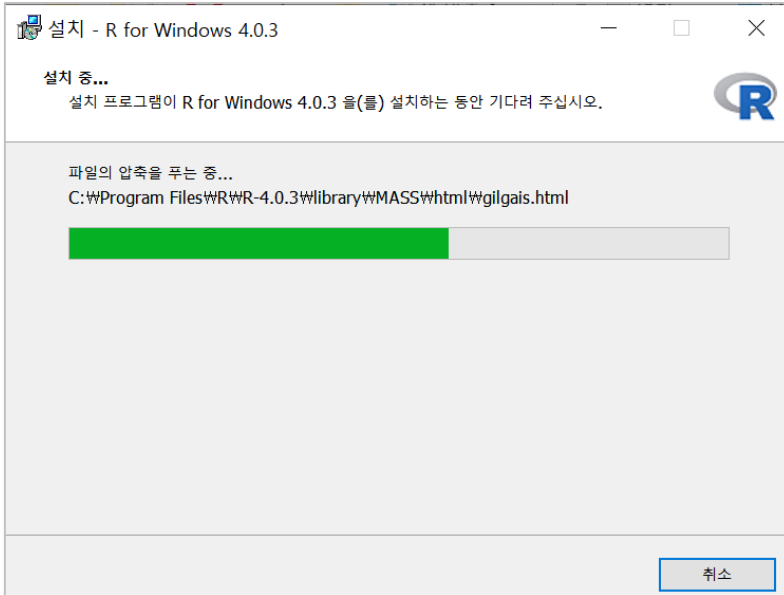
2.1 도구 챙기기



2.1 도구 챙기기



2.1 도구 챙기기



Thank you

