

Big Data



About..

컴퓨터소프트웨어공학과
김 원 일



목차



- Big Data
- R installation
- R Execution
- R 산술 연산
- R 기본 지원 함수
- R 변수 선언과 사용



Big Data – 1



• 기본 정의

- Data Vs. Information
- 기존 데이터베이스 관리도구의 능력을 넘어서는 용량 처리
 - DBMS 능력을 넘어서는 용량. DBMS는 결국 파일로 구성된 데이터
 - 윈도우 파일 시스템 NTFS의 최대 지원 파일 크기는 16 TB
- 정형 또는 비정형 데이터 집합에 대한 처리
 - DBMS는 스키마(Schema) 정의가 반드시 동반되어야 처리가 가능
 - RDB 와 NoSQL(OODB)은 결국 비슷한 구조
- 특별히 지정된 데이터의 종류가 아닌 데이터에 대한 처리
 - 인식과 분류가 가능한 데이터 종류에 대한 처리
 - 아무 데이터나 모두 데이터가 될 수는 없음. 모든 데이터는 A.I의 영역
- 유의미한 정보를 추출하는 데이터 분석 기술



Big Data – 2



- **최종 목표**

- 다중 복합 대량 데이터 관리를 통해 대량의 Data에서 Information 추출
- 지속적인 데이터 관리를 통해 최신 정보를 항상 유지하여 다양한 분석 결과를 제공

- **Big Data 3V**

- Volume(볼륨) : 데이터의 규모. 테라부터 페타 바이트 이상
- Varsity(다양성) : 다양한 데이터 및 형식의 데이터를 포함
- velocity(속도) : 하루 단위의 데이터를 수집, 저장, 처리 및 분석하는 속도

- **분석 결과 활용**

- 기술적 분석 : 어떤 일이 발생하였고, 그 이유는 무엇인지 분석
- 예측 분석 : 미래에 발생할 일에 대한 예측으로 알림, 탐지를 통한 관리
- 규범 분석 : 권장 사항이나 임계 값에 대한 분석 정보 제공



- Big Data 처리

- 수집 : 원시 데이터의 수집 단계
- 저장 : 원시 및 분석 데이터를 안전하고, 확장 가능하게 보관하는 단계
- 처리 및 분석 : 정렬, 집계 및 조인 등으로 정보를 추출하는 단계
- 사용 및 시각화 : 유의미한 정보 추출이나 예측을 수행하는 단계

- Big Data 동작

- 기본적으로 배치(Batch) 작업을 통해 수행하는 것이 기본
- 최근 수집, 저장, 분석 단계를 분리하여 수행하는 형태로 진화


- Big Data 주의점

- 실제 사용되는 Big Data는 초 대용량의 메모리와 디스크를 필요로 함
- Big Data가 필요하지 않은 경우는 굳이 적용할 필요가 없음
- 테스트를 위한 데이터 셋을 구할 때, 사용 가능한지 여부를 확인해야 함



R installation – 1

- 다운로드 URL: <https://cloud.r-project.org/>



[CRAN](#)
[Mirrors](#)
[What's new?](#)
[Task Views](#)
[Search](#)

[About R](#)
[R Homepage](#)
[The R Journal](#)

[Software](#)
[R Sources](#)
[R Binaries](#)
[Packages](#)
[Other](#)

[Documentation](#)
[Manuals](#)
[FAQs](#)
[Contributed](#)

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](#) (Debian, Fedora/Redhat, Ubuntu)
- [Download R for macOS](#)
- [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 (2021-08-10, Kick Things) [R-4.1.1.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



R installation – 2



- base → install R for the first time 선택

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 >= 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.

Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.



R installation – 3

- Download R 4.X for Windows
 - linux와 MAC 버전도 존재

R-4.1.1 for Windows (32/64 bit)

[Download R 4.1.1 for Windows](#) (86 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 [md5sum](#) of the .exe to the [fingerprint](#) on the master server. You will need a version of md5sum for windows: both [graphical](#) and [command line versions](#) are available.

Frequently asked questions

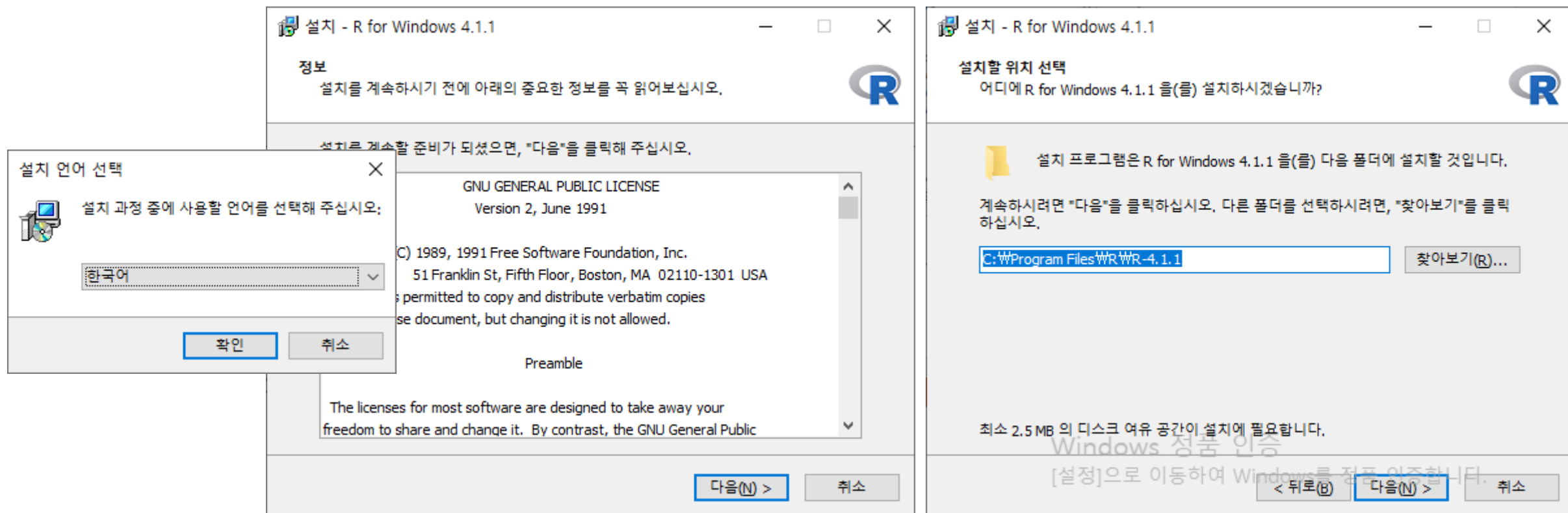
- [Does R run under my version of Windows?](#)
- [How do I update packages in my previous version of R?](#)
- [Should I run 32-bit or 64-bit R?](#)

Please see the [R FAQ](#) for general information about R and the [R Windows FAQ](#) for Windows-specific information.



R installation – 4

- 언어 선택
- 라이선스 선택 후, 설치 경로 설정

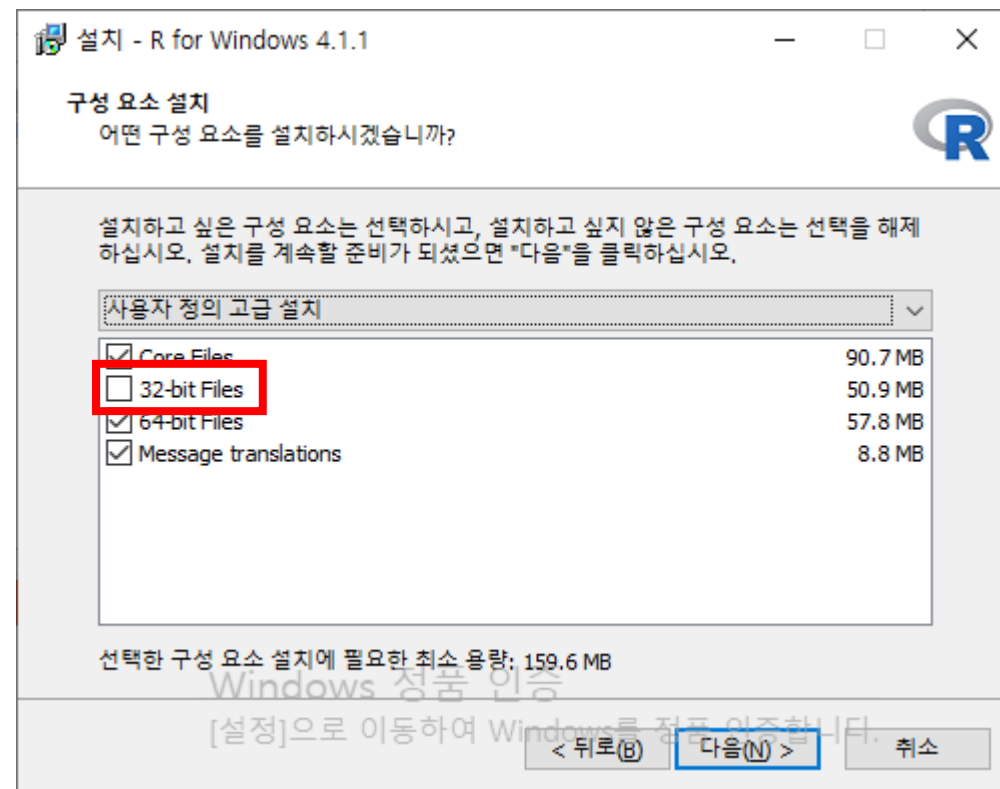
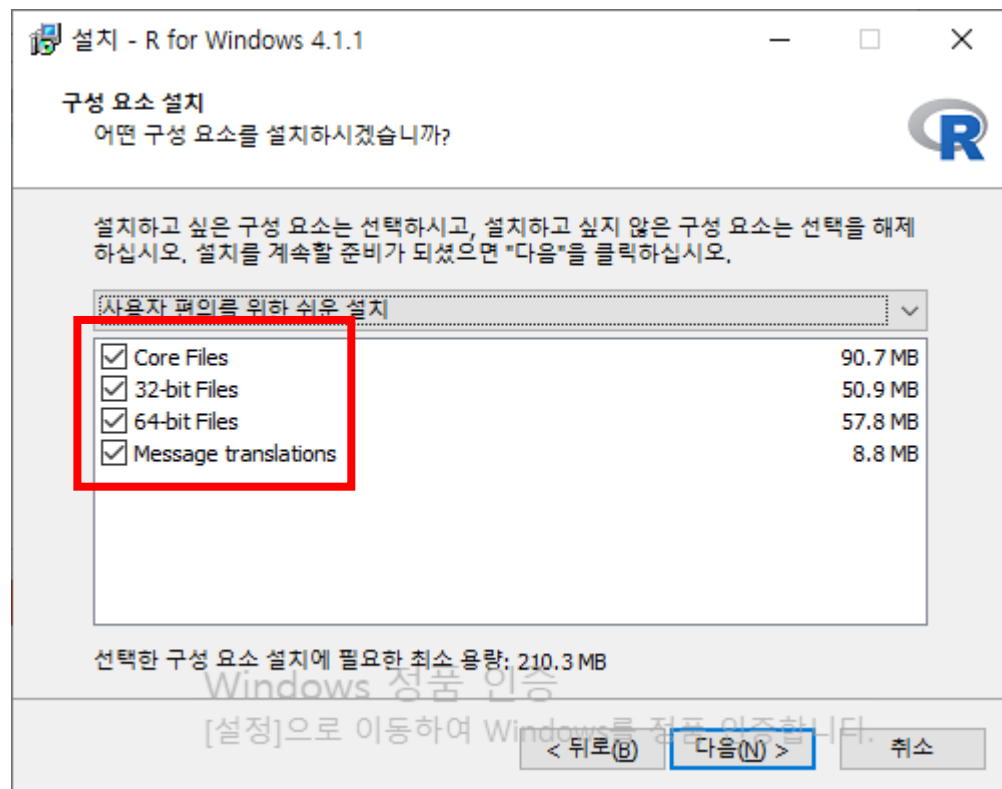




R installation – 5

• 아키텍처 선택

- 설치 시스템에 맞게 선택하는 것이 좋음

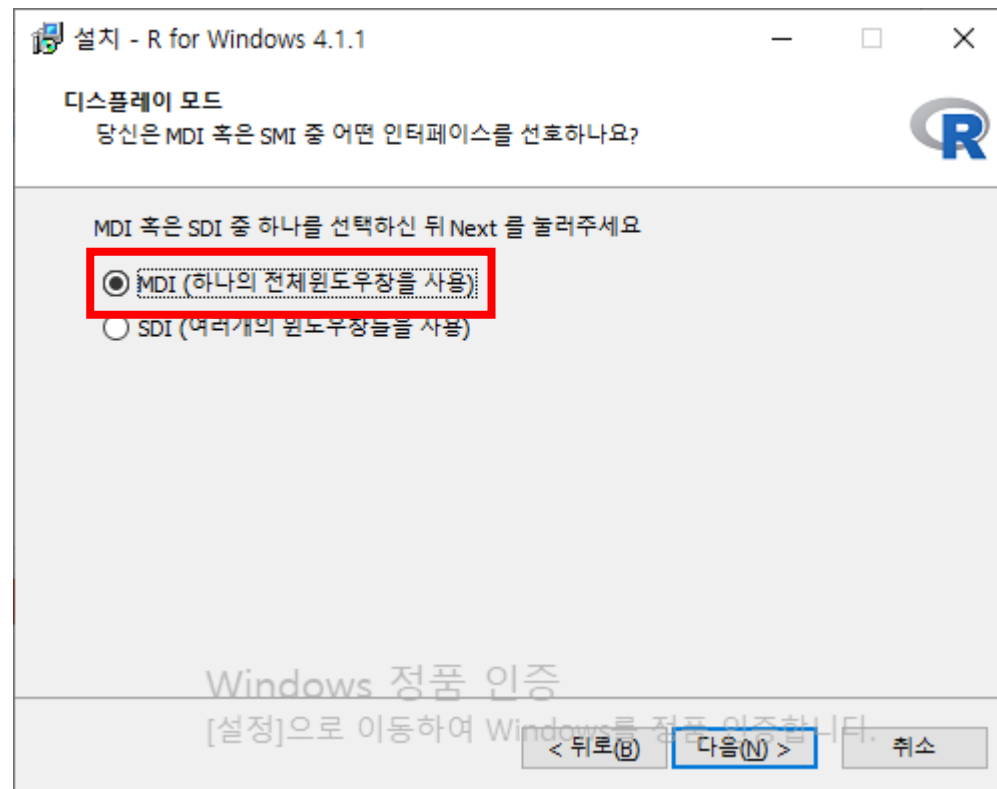
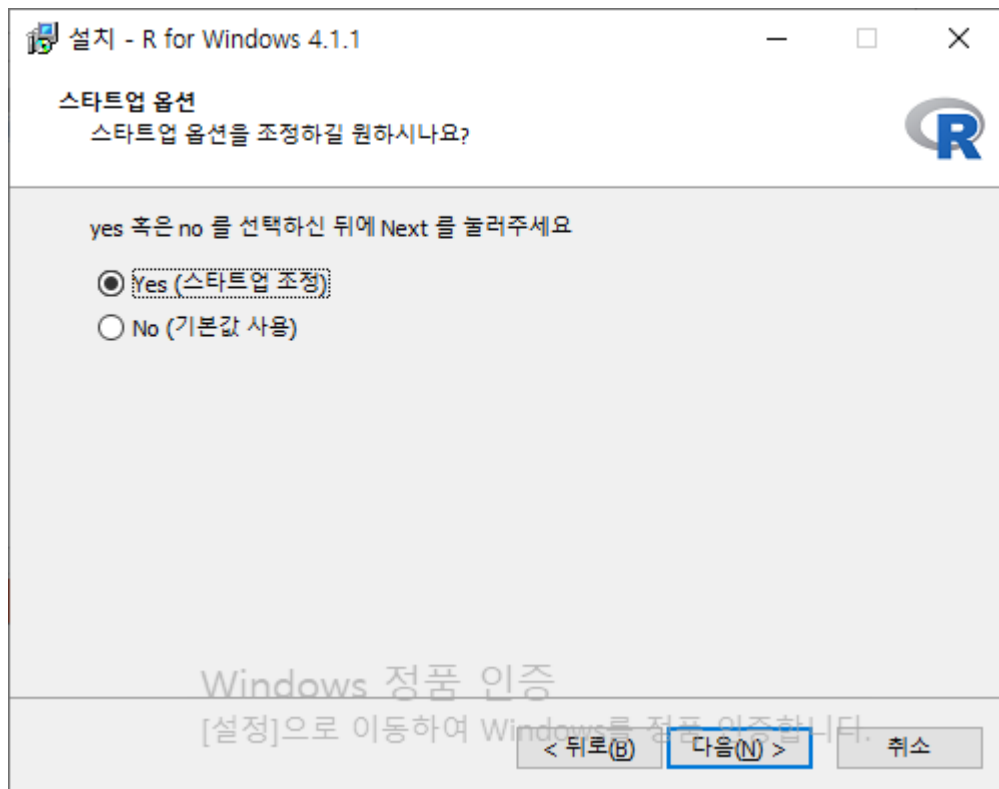




R installation – 6

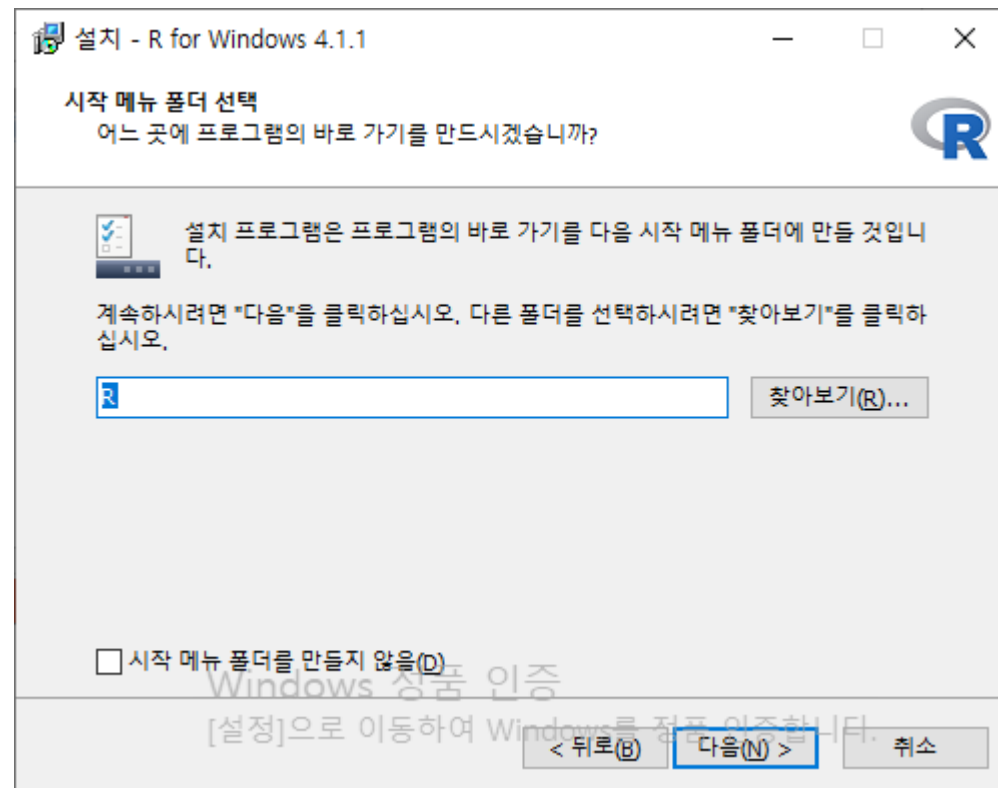
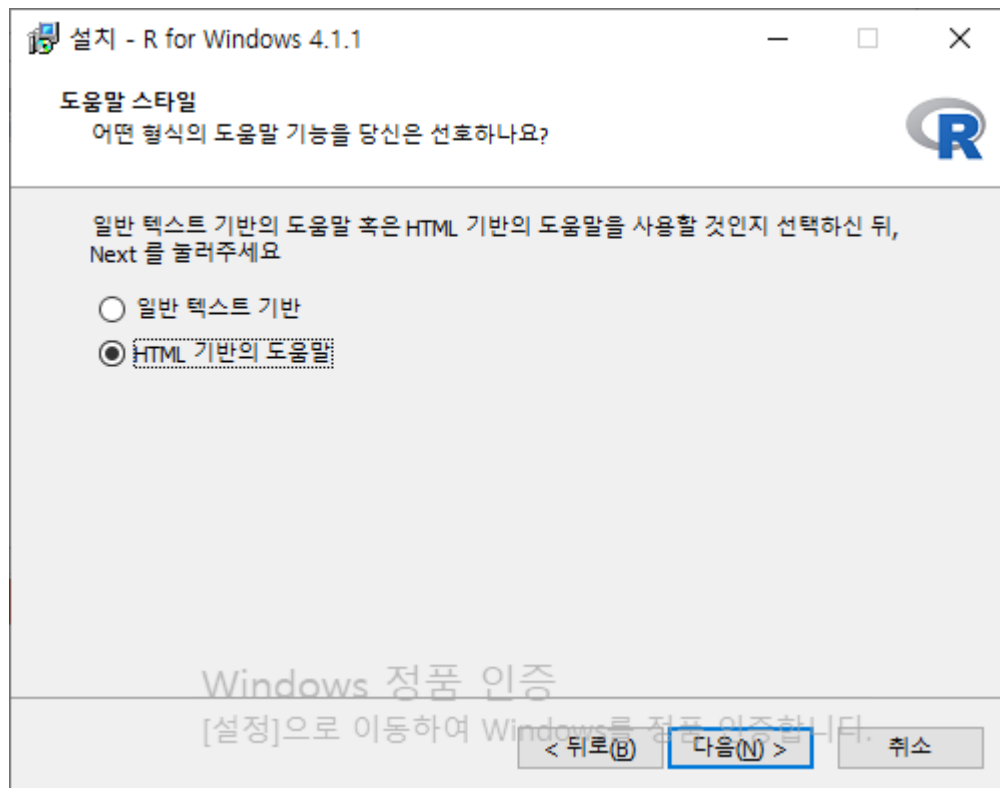
• 시작 옵션과 실행 모드 설정

- MDI (Multiple Document Interface) 또는 SDI (Single Document Interface)





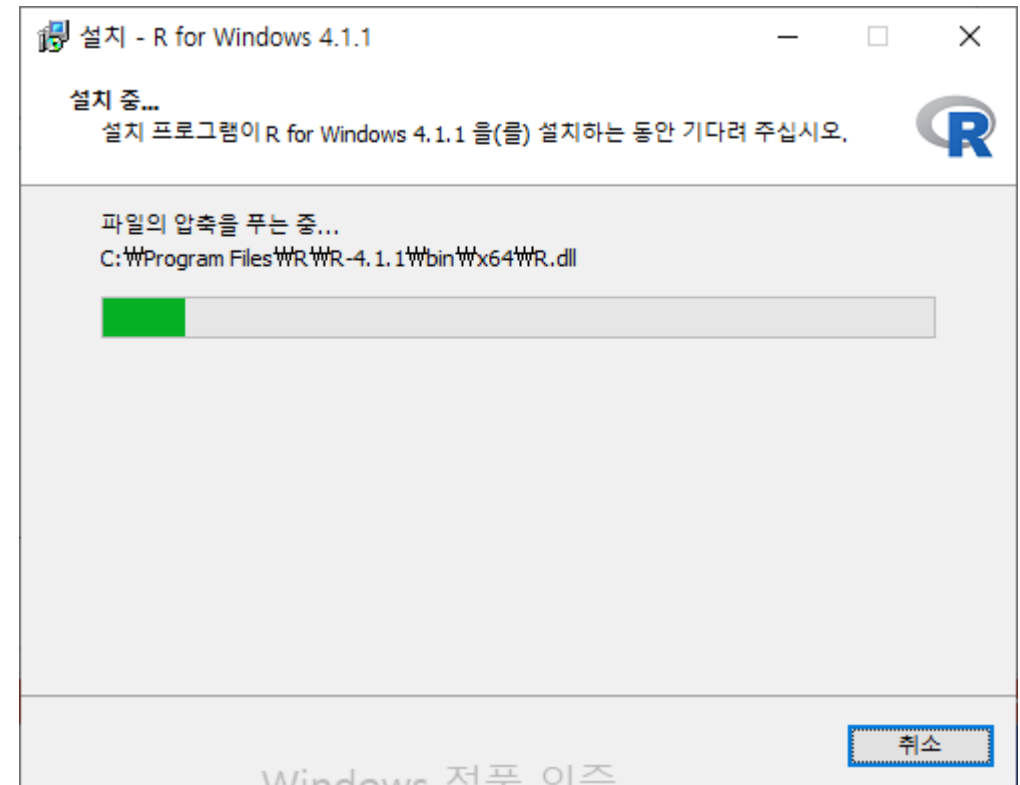
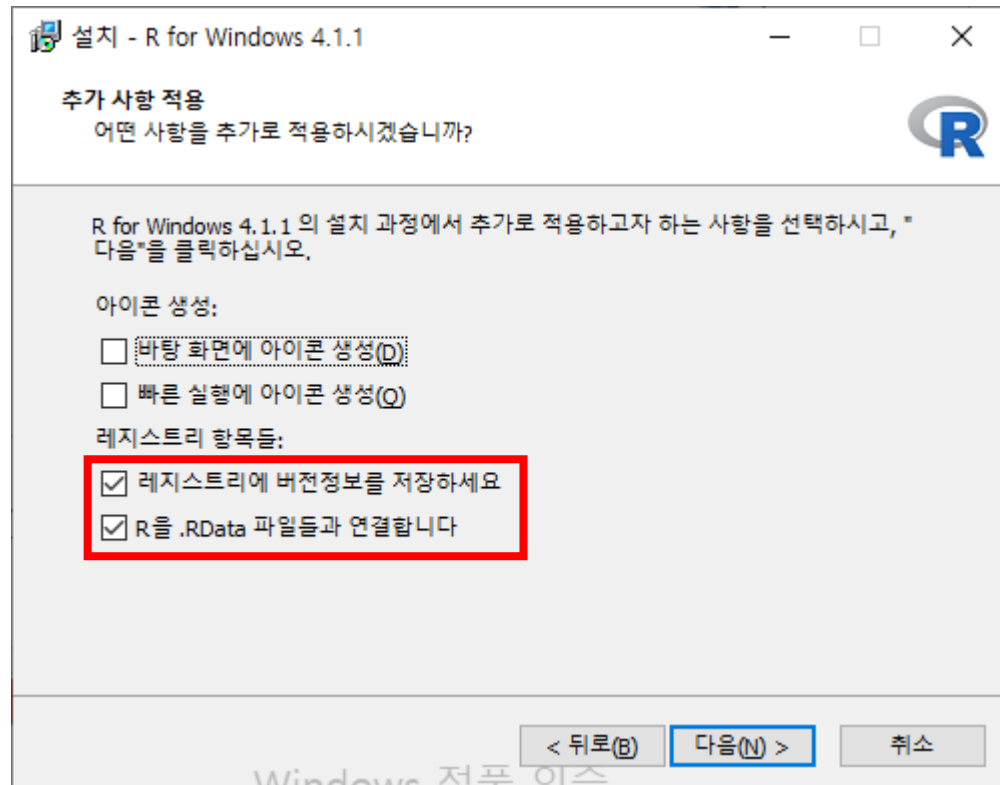
- 도움말 스타일과 시작 메뉴 폴더 선택





R installation – 8

- 필요한 경우 아이콘 생성





R installation – 9

- 설치 완료 및 아이콘 확인

