

# 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 – 3



- 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:

<a href="#">base</a>	Binaries for base distribution. This is what you want to <a href="#">install R for the first time</a> .
<a href="#">contrib</a>	Binaries of contributed CRAN packages (for R >= 2.13.x; managed by Uwe Ligges). There is also information on <a href="#">third party software</a> available for CRAN Windows services and corresponding environment and make variables.
<a href="#">old_contrib</a>	Binaries of contributed CRAN packages for outdated versions of R (for R < 2.13.x; managed by Uwe Ligges).
<a href="#">Rtools</a>	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?](#)

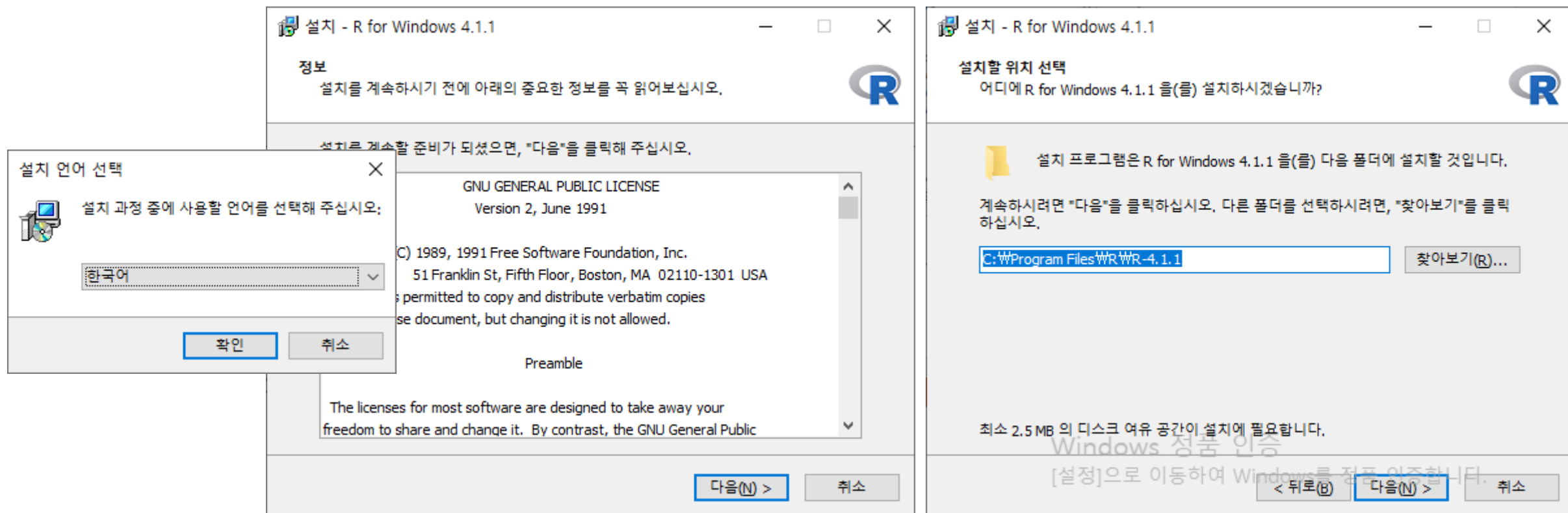
Please see the [R FAQ](#) for general information about R and the [R Windows FAQ](#) for Windows-specific information.





## R installation – 4

- 언어 선택
- 라이선스 선택 후, 설치 경로 설정

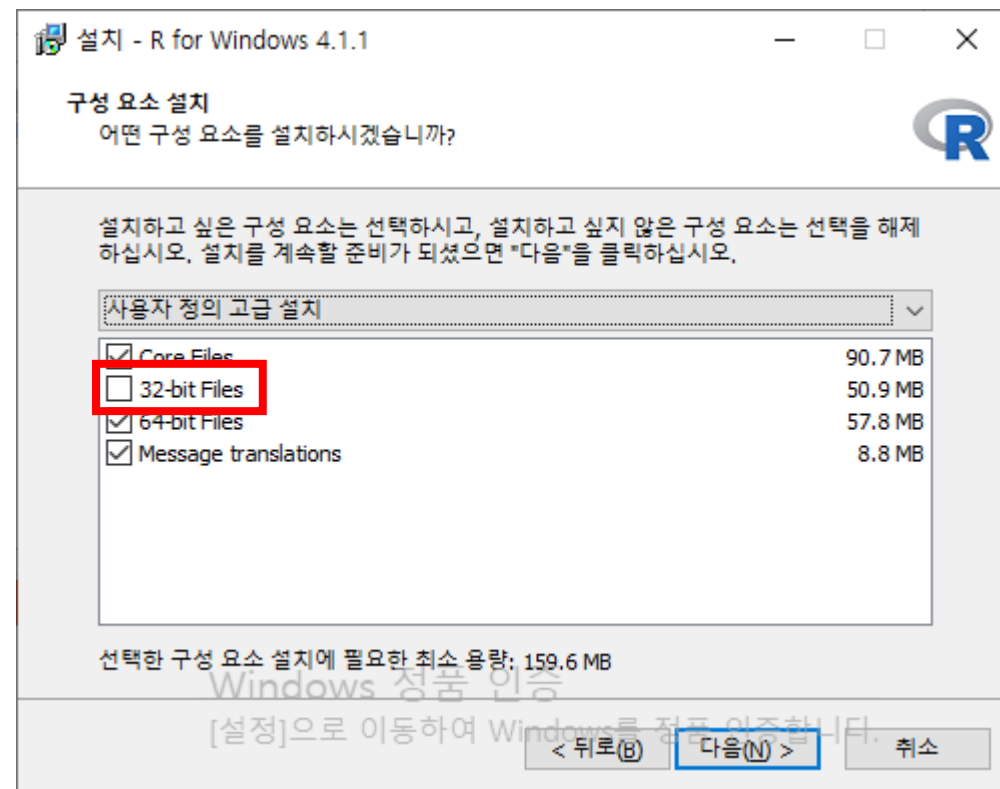
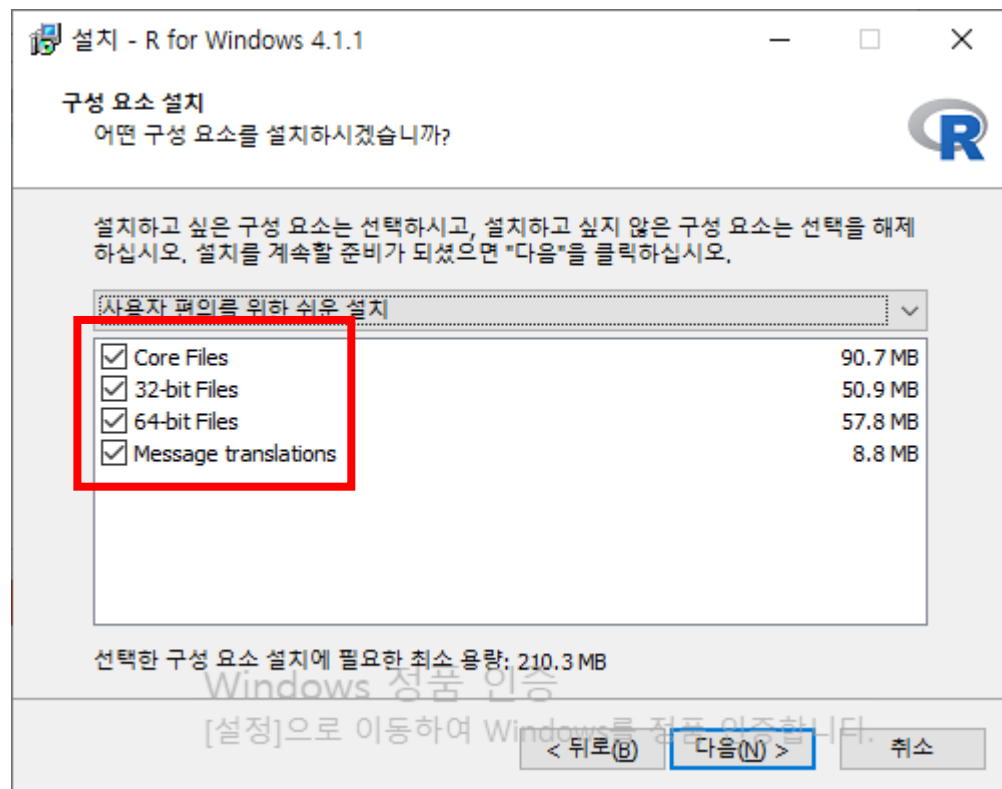




## R installation – 5

### • 아키텍처 선택

- 설치 시스템에 맞게 선택하는 것이 좋음

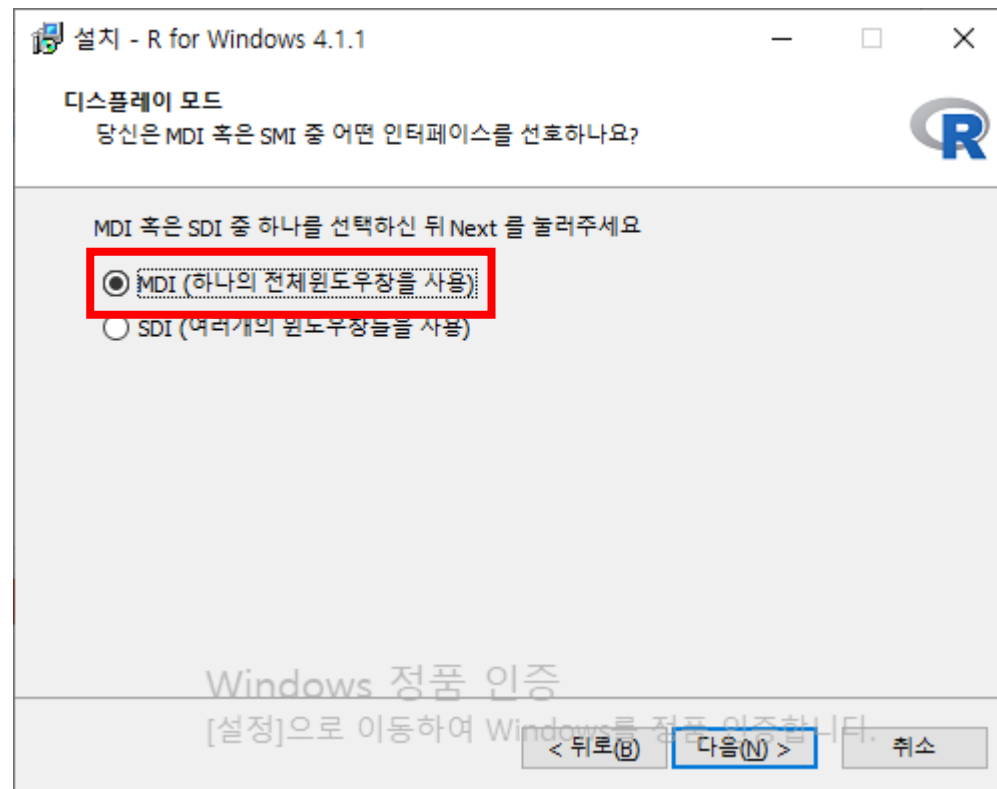
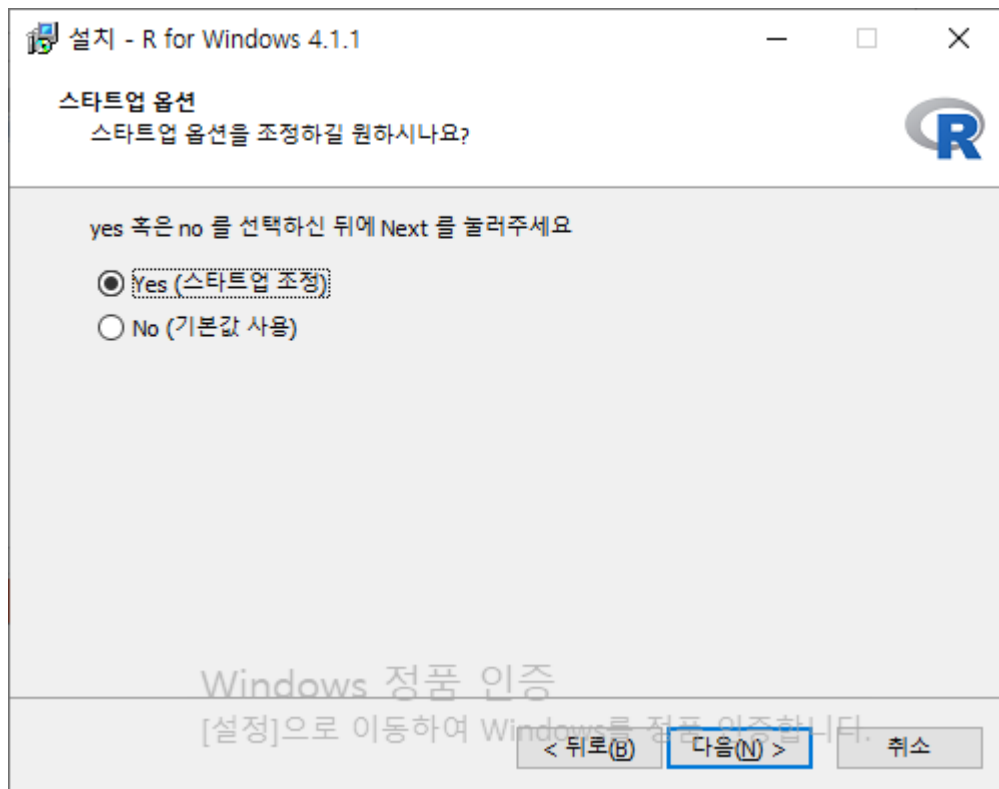




## R installation – 6

### • 시작 옵션과 실행 모드 설정

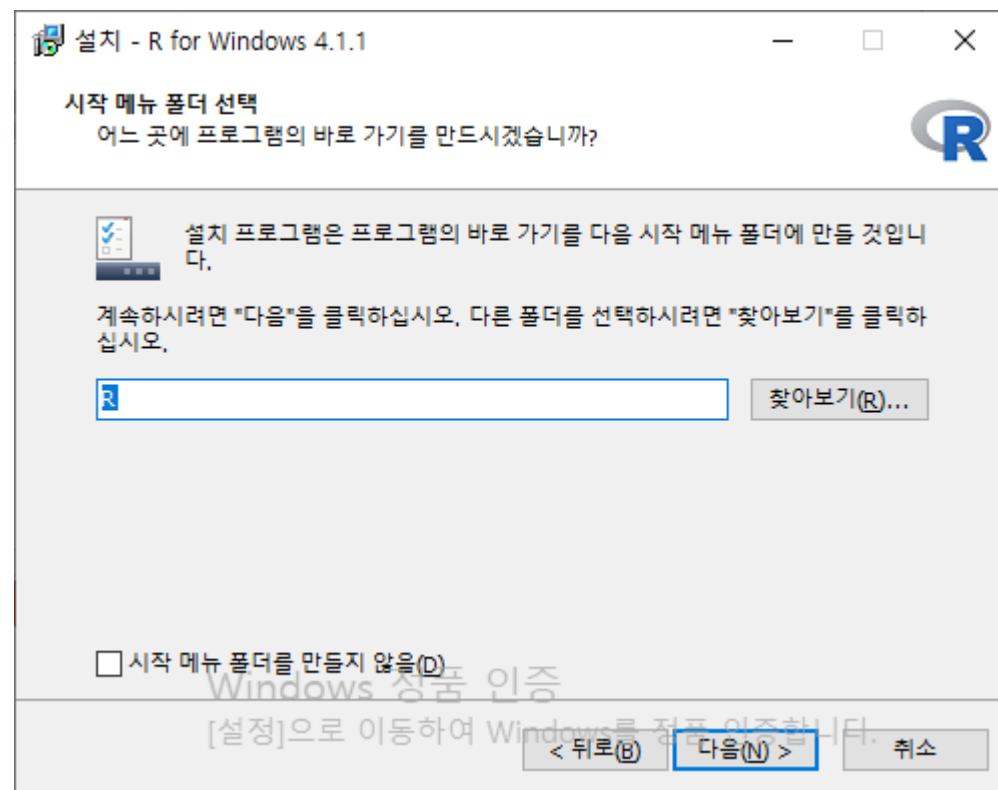
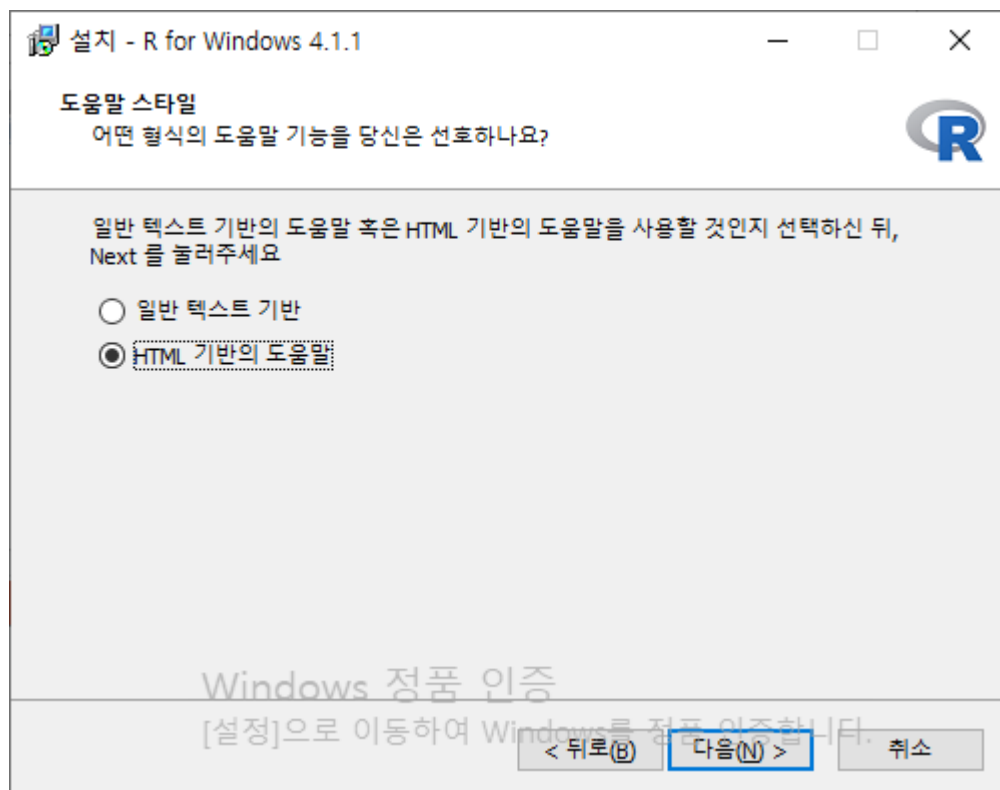
- MDI (Multiple Document Interface) 또는 SDI (Single Document Interface)





## R installation – 7

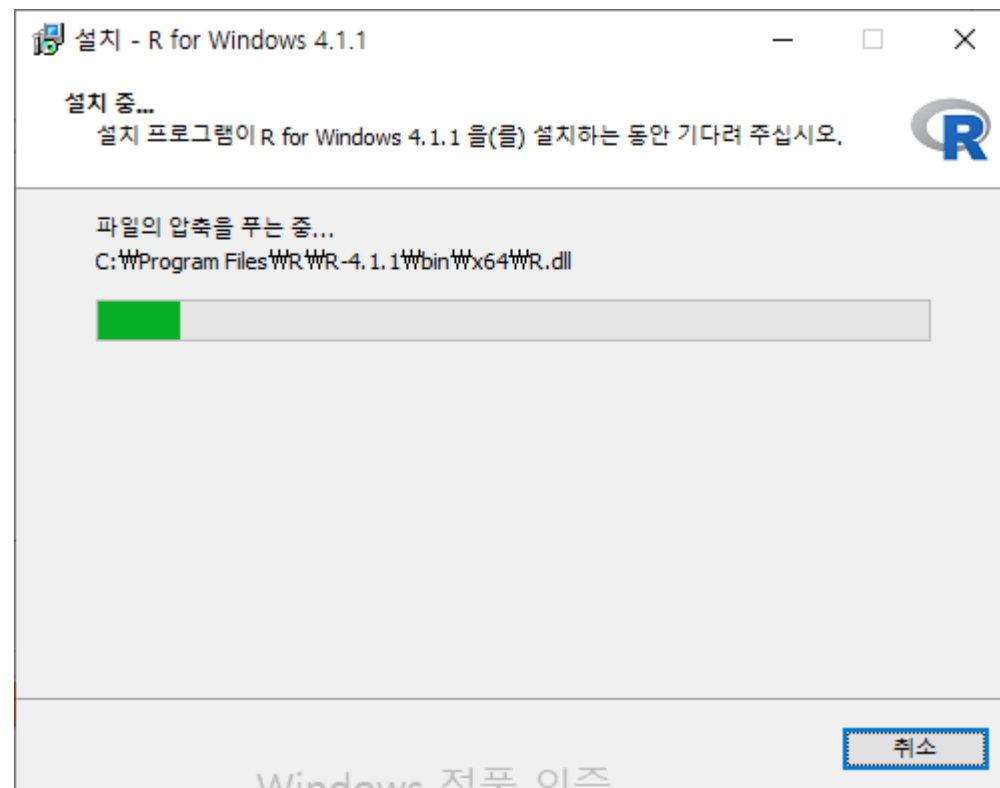
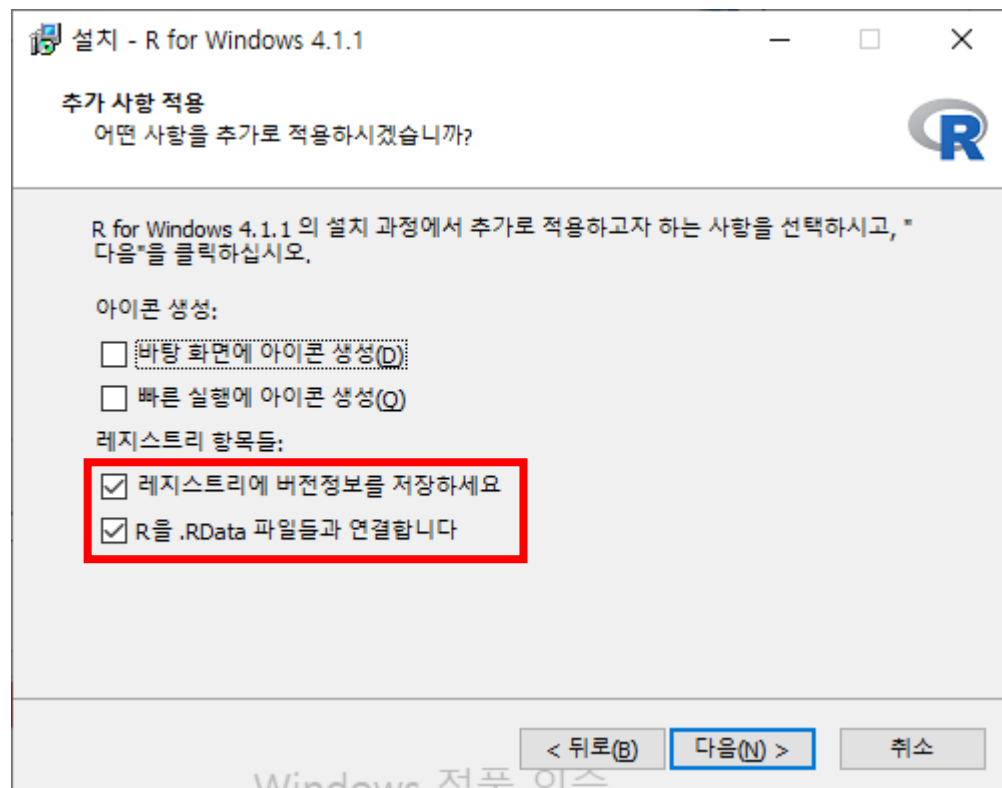
- 도움말 스타일과 시작 메뉴 폴더 선택





## R installation – 8

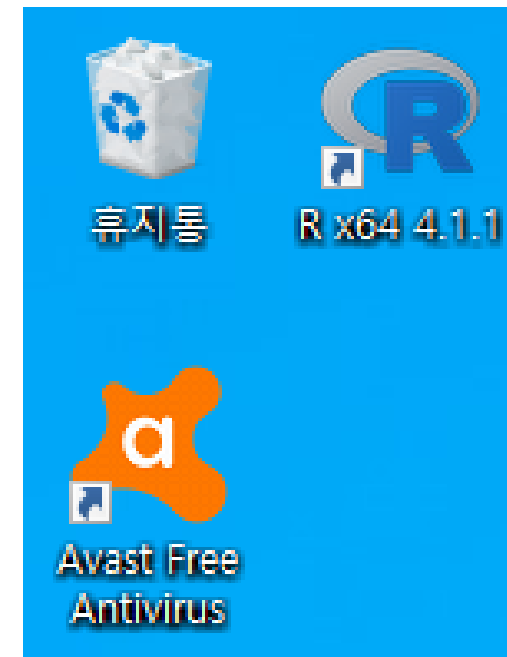
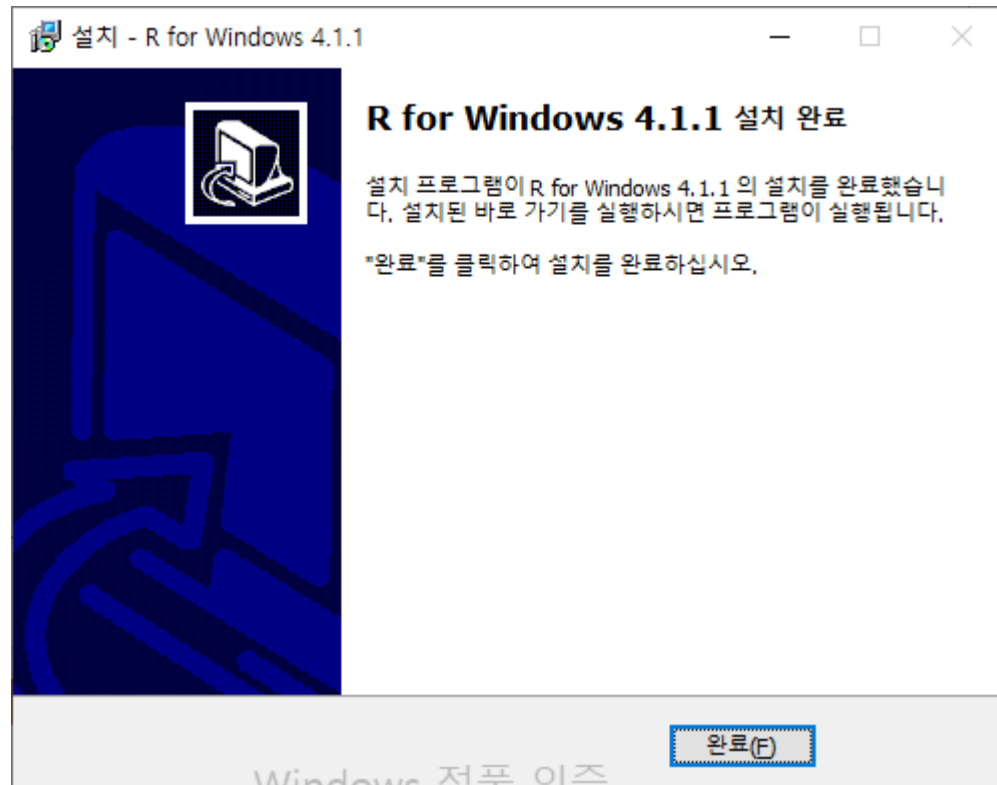
- 필요한 경우 아이콘 생성





## R installation – 9

- 설치 완료 및 아이콘 확인





# R Execution – 1

## • R 실행과 MDI 실행 모드

```
RGui (64-bit)
파일 편집 보기 기타 패키지를 윈도우즈 도움말

R Console

R version 4.1.1 (2021-08-10) -- "Kick Things"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R은 자유 소프트웨어이며, 어떠한 형태의 보증없이 배포됩니다.
또한, 일정한 조건하에서 이것을 재배포 할 수 있습니다.
배포와 관련된 상세한 내용은 'license()' 또는 'licence()'를 통하여 확인할 수 있습니다.

R은 많은 기여자들이 참여하는 공동프로젝트입니다.
'contributors()'라고 입력하시면 이에 대한 더 많은 정보를 확인할 수 있습니다.
그리고, R 또는 R 패키지를 출판물에 인용하는 방법에 대해서는 'citation()'을 입력하십시오.

'demo()'를 입력하신다면 몇가지 데모를 보실 수 있으며, 'help()'를 입력하시면 R의
또한, 'help.start()'의 입력을 통하여 HTML 브라우저에 의한 도움말을 사용하실 수
R의 종료를 원하시면 'q()'를 입력해주세요.

> |
```

```
RGui (64-bit)
파일 편집 패키지를 윈도우즈 도움말

R Console

R version 4.1.1 (2021-08-10) -- "Kick Things"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R은 자유 소프트웨어이며, 어떠한 형태의 보증없이 배포됩니다.
또한, 일정한 조건하에서 이것을 재배포 할 수 있습니다.
배포와 관련된 상세한 내용은 'license()' 또는 'licence()'를 통하여 확인할 수 있습니다.

R은 많은 기여자들이 참여하는 공동프로젝트입니다.
'contributors()'라고 입력하시면 이에 대한 더 많은 정보를 확인할 수 있습니다.
그리고, R 또는 R 패키지를 출판물에 인용하는 방법에 대해서는 'citation()'을 입력하십시오.

'demo()'를 입력하신다면 몇가지 데모를 보실 수 있으며, 'help()'를 입력하시면 R의
또한, 'help.start()'의 입력을 통하여 HTML 브라우저에 의한 도움말을 사용하실 수
R의 종료를 원하시면 'q()'를 입력해주세요.

> |
```

```
제목없음 - R 편집기

> v=3
> print v
[1] 3
> print
[1] 3
> function(x) {
  UseMethod("print")
}
> v
[1] 3
> function(x) {
  UseMethod("print")
}
> v
[1] 3
> asdf
[1] 3
> asdf = 3
> asdf
[1] 3
> asdf+v
[1] 6
> asdf*v
[1] 9
> |
```



- SDI 모드로 실행

```
R Console (64-bit)
파일 편집 기타 패키지들 윈도우즈 도움말

R version 4.1.1 (2021-08-10) -- "Kick Things"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R은 자유 소프트웨어이며, 어떠한 형태의 보증없이 배포됩니다.
또한, 일정한 조건하에서 이것을 재배포 할 수 있습니다.
배포와 관련된 상세한 내용은 'license()' 또는 'licence()'를 통하여 확인할 수 있습니다.

R은 많은 기여자들이 참여하는 공동프로젝트입니다.
'contributors()'라고 입력하시면 이에 대한 더 많은 정보를 확인할 수 있습니다.
그리고, R 또는 R 패키지들을 출판물에 인용하는 방법에 대해서는 'citation()'을 입력하십시오.

'demo()'를 입력하신다면 몇가지 데모를 보실 수 있으며, 'help()'를 입력하십시오.
또한, 'help.start()'의 입력을 통하여 HTML 브라우저에 의한 도움말을 사용할 수 있습니다.
R의 종료를 원하시면 'q()'를 입력하십시오.

> |
```





## R Execution – 3

### • 스크립트 코드의 이해

- 간단한 명령문의 구조
- 일반적인 상수 연산은 즉시 결과 확인이 가능
- 연산자는 일반적인 산술연산자만 허용

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지를 윈도우즈 도움말

'contributors()'라고 입력하시면 이에 대한 더 많은 정보를 확인하실 수 $
그리고, R 또는 R 패키지들을 출판물에 인용하는 방법에 대해서는 'citation$

'demo()'를 입력하신다면 몇가지 데모를 보실 수 있으며, 'help()'를 입력$
또한, 'help.start()'의 입력을 통하여 HTML 브라우저에 의한 도움말을 사$
R의 종료를 원하시면 'q()'를 입력해주세요.

[이전에 저장한 작업공간을 복구하였습니다]

> 26/38*100
[1] 68.42105
> |
```

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지를 윈도우즈 도움말

'demo()'를 입력하신다면 몇가지 데모를 보실 수 있으며, 'help()'를 입력$
또한, 'help.start()'의 입력을 통하여 HTML 브라우저에 의한 도움말을 사$
R의 종료를 원하시면 'q()'를 입력해주세요.

[이전에 저장한 작업공간을 복구하였습니다]

> 26/38*100
[1] 68.42105
> 36%90
에러: 예상하지 못한 입력입니다. in "36%90"
> |
```



- 일반적인 산술 연산 지원

- 2항 연산이 기본

- + : 더하기
- - : 빼기
- \* : 곱하기
- / : 나누기

- 매우 큰 수도 지원

- 통계 프로그램 기본
- 기본 소수점 7자리

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지를 윈도우즈 도움말
[Icons: File Explorer, R Script, Save, Print, Run, Stop, Copy]

> 3+5
[1] 8
> 3*90
[1] 270
> 65535 * 65535
[1] 4294836225
> 65535 * 655359900
[1] 4.294901e+13
> 338982 / 67
[1] 5059.433
> 9898456789675467 - 4567876456782376
[1] 5.33058e+15
> 3.14 * 78
[1] 244.92
> 2 / 3
[1] 0.6666667
> |
```



## • 기본 산술 연산

- $\wedge$  : 제곱
- $\%\%$  : 나눗셈의 나머지
- 0으로 나누는 문제
- ( , )와 { , } 괄호 지원
- [ , ]는 지원하지 않음

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지들 윈도우즈 도움말

> 3.14 * 78
[1] 244.92
> 2 / 3
[1] 0.6666667
> 3 / 0
[1] Inf
> -3 / 0
[1] -Inf
> 65535 %% 16
[1] 15
> 2 ^ 23
[1] 8388608
> ( 2 ^ 23 ) - ( 3 ^ 9 )
[1] 8368925
> { ( 2 ^ 23 ) - ( 3 ^ 9 ) } * 3
[1] 25106775
> [ { ( 2 * 3 ) + 90 } * 3 ]
에러: 예기치 않은 '['입니다 in "[
> |
```



- **기본 지원 함수들을 제공**
  - 기본적으로 필요한 함수들은 내장 함수로 제공

함수	의미	사용 예
log( )	로그 함수	log(10), log(10, base=2)
sqrt( )	제곱근	sqrt(36)
max( )	최대 값	max(3, 9, 5)
min( )	최소 값	min(3, 9, 5)
abs( )	절대 값	abs(-10)
factorial( )	팩토리얼	factorial(5)
sin( ), cos( ), tan( )	삼각함수	sin(pi/2)



- **내장 함수 즉시 사용 가능**
  - pi는 내장 변수로 선언
  - 내장 변수 값도 수정 가능
    - 수정하지 않아야 혼동 없음
- **함수명도 변수로 사용 가능**
  - 혼동 발생 가능성이 높음

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지들 윈도우즈 도움말

> log(10, base=2)
[1] 3.321928
> sqrt(36)
[1] 6
> max(3, 9, 5)
[1] 9
> min(3, 9, 5)
[1] 3
> abs(-10)
[1] 10
> factorial(5)
[1] 120
> sin(pi/2)
[1] 1
> cos(pi/2)
[1] 6.123032e-17
> tan(pi/2)
[1] 1.633124e+16
> |
```



## • 일반적인 변수 명명법과 동일

- 숫자는 첫 글자로 올 수 없음, 각종 특수 기호는 첫 글자로 올 수 없음
- 첫 글자 공백은 특수 기호로 인식되지 않음
- 첫 글자 이후로 "\_", "." 과 같은 특수 기호 사용이 가능

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지들 윈도우즈 도움말

> 12a <- 78
에러: 예상하지 못한 기호(symbol)입니다. in "12a"
> _a <- 78
에러: 예상하지 못한 입력입니다. in "_"
> @a <- 78
에러: 예기치 않은 '@'입니다 in "@"
> a <- 78
> a
[1] 78
> |
```

Windows 정품 인증  
[설정]으로 이동하여 Windows를 정품 인증합니다.



- 변수에 값 대입/치환 및 확인

- "변수 <- 값" [ENTER] : 값을 변수에 대입
- "변수" [ENTER] : 변수의 값을 화면으로 출력

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지를 윈도우즈 도움말

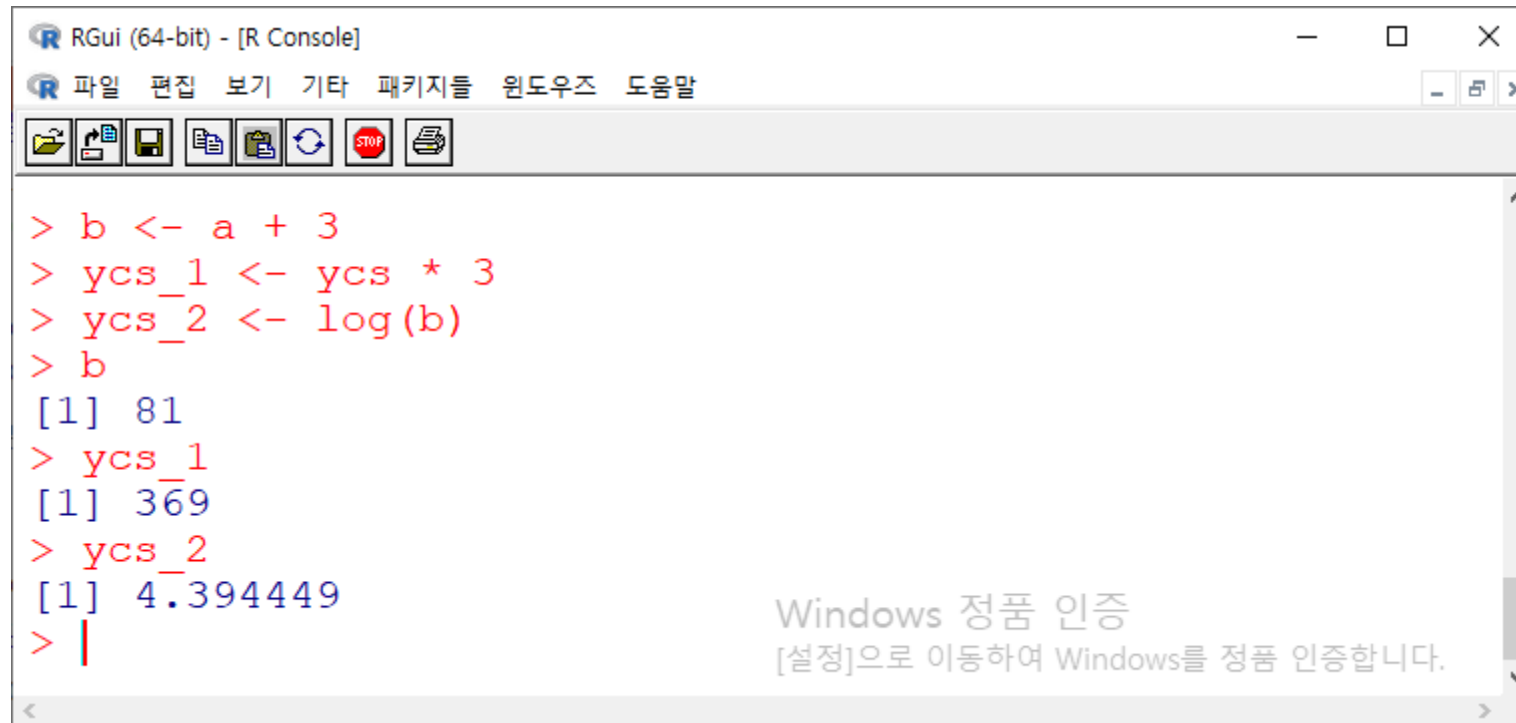
> a <- 78
> ycs <- 123
> ycs.name <- 234
> a
[1] 78
> ycs
[1] 123
> ycs.name
[1] 234
> |
```

Windows 정품 인증  
[설정]으로 이동하여 Windows를 정품 인증합니다.



- 다양한 초기화 형태를 지원

- 기존 변수를 이용하여 초기화 가능
- 별도의 연산과 변수들을 이용하여 초기화 가능



```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지를 윈도우즈 도움말

> b <- a + 3
> ycs_1 <- ycs * 3
> ycs_2 <- log(b)
> b
[1] 81
> ycs_1
[1] 369
> ycs_2
[1] 4.394449
> |
```

Windows 정품 인증  
[설정]으로 이동하여 Windows를 정품 인증합니다.





### • 변수 목록 확인

- ls( ) 명령으로 현재 변수 목록 확인 가능
- 스크립트 언어이므로, 변수 명 이외의 정보는 출력되지 않음
- 각 변수 값을 출력해야만 값 확인이 가능하므로, 변수 명을 잘 지어야 함

```
RGGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지들 윈도우즈 도움말
[Icons: File Explorer, Run, Save, Copy, Paste, Undo, Redo, Stop, Print]

> ls( )
[1] "a"          "b"          "c"
[4] "d"          "userInfo.addr" "userInfo.name"
[7] "userInfo.phone" "ycs"        "ycs.name"
[10] "ycs_1"      "ycs_2"

> |
```



## • 변수 이름 활용

- 구조체 형식으로 이름 사용 가능
- 형식만 사용한 별도의 변수 취급
- 문자열은 연산으로 동시 출력 불가

```
RGU (64-bit) - [R Console]
파일 편집 보기 기타 패키지들 윈도우즈 도움말

> userInfo.name <- 'wikim'
> userInfo.addr <- 'kimpo'
> userInfo.phone <- '02-2610-0764'
> userInfo.name
[1] "wikim"
> userInfo.addr
[1] "kimpo"
> userInfo.phone
[1] "02-2610-0764"
> print( userInfo.name & userInfo.addr )
userInfo.name & userInfo.addr에서 다음과 같은 에러$
오로지 숫자, 논리, 또는 복소수 유형에 대한 연산들$
> print( userInfo.name && userInfo.addr )
userInfo.name && userInfo.addr에서 다음과 같은 에러$
'x' 유형은 'x && y'에 유효하지 않습니다
> print( userInfo.name + userInfo.addr )
userInfo.name + userInfo.addr에서 다음과 같은 에러$
이항연산자에 수치가 아닌 인수입니다
> print( userInfo.name | userInfo.addr )
userInfo.name | userInfo.addr에서 다음과 같은 에러$
```