

R을 이용한 자료처리 및 시각화

기본 환경 설정부터 재현가능한 연구 방법 입문

한국한의학연구원 한의기반연구부

Korea Institute of Oriental Medicine, KM Fundamental Research Division

구본초, *Ph.D. in Statistics*

2017-11-23

Contents

| | |
|--|-----------|
| 1 R을 사용하기 위한 환경 설정 | 5 |
| 1.1 Overview | 5 |
| 1.2 R 설치하기(Windows 용) | 7 |
| 1.3 R 시작 및 작동 체크 | 14 |
| 1.4 RStudio 설치하기 | 18 |
| 1.5 RStudio의 구성 | 21 |
| 1.6 RStudio에서 배치 파일 생성 및 실행 | 25 |
| 1.7 R 패키지 설치하기 | 25 |
| 1.8 rJava 설치하기 | 26 |
| 1.9 RStudio 프로젝트 생성 및 ProjectTemplate 패키지 연동 | 29 |
| 1.10 R + RStudio를 잘 활용 하려면? | 29 |
| 2 R의 기본 사용 | 31 |
| 2.1 R 객체 입력 방법 및 변수 설정 규칙 | 32 |
| 2.2 스칼라 | 33 |
| 2.3 벡터 | 33 |
| 2.4 리스트 | 33 |
| 2.5 행렬 | 33 |
| 2.6 배열 | 33 |
| 2.7 데이터 프레임 | 33 |
| 2.8 데이터 연산 | 33 |
| 3 데이터 조작 | 35 |

| | | |
|----------|-----------------------------------|-----------|
| 3.1 | 실습 데이터 설명 | 35 |
| 3.2 | 파일 입출력 | 35 |
| 3.3 | 벡터화 연산(apply 계열 함) | 35 |
| 3.4 | dplyr + tidyr을 이용한 data wrangling | 35 |
| 4 | ggplot2를 이용한 자료 시각화 | 37 |
| 4.1 | ggplot() vs. plot() | 37 |
| 4.2 | Grammar of Graphics | 37 |
| 5 | 재현가능한 데이터 연동 문서 작성 | 39 |

Chapter 1

R을 사용하기 위한 환경 설정

1.1 Overview

1.1.1 What is R?

- R 언어: 통계 및 자료 시각화를 지원하는 언어 및 환경
- 1980년 AT&T 연구소의 John Chambers가 개발한 S 언어를 기반으로 1995년 뉴질랜드 Auckland Univ.의 Robert Gentleman & Ross Ihaka 개발이 그 기원임
- GNU 기반 오픈소스

1.1.2 Why R?

1. 장점

- **Free software**
- 주요 운영체제(Unix, Windows, Macintosh)에서 사용 가능
- 현존하는 거의 대부분의 통계 방법론들이 package로 구현
- 강력한 그래픽 기능
- 빠른 update 및 연산속도(?)
- 방대한 community 및 R package에 공개 및 공유된 자료

2. Do we have to learn R language?

- 물론 꼭 배울 필요는 없다!! (요리를 하는데 꼭 좋은 칼이 필요 없듯이...)
- 다양한 통계 소프트웨어 및 분석 언어(SPSS, SAS, WEKA, MATLAB, PYTHON, ...) 존재
- 프로그래밍에 익숙하지 않은 사용자들의 접근성이 떨어짐
- 하지만 배워두면 좋은 이유
 - 통계분석과 보고서 작성을 위한 최적 환경
 - * R + Rstudio + Rmarkdown을 통해 분석에서부터 보고서 작성까지 한번에 가능
 - SPSS, SAS에 비해 확장성이 높기 때문에 다양한 문제에 적용 가능
 - 방대한 양의 메뉴얼 및 서적들
 - 그리고 이 모든 것을 거의 대부분 무료 사용 가능

Tips

- R에서 구현한 모든 package들은 CRAN (The Comprehensive R Archive Network, <http://cran.r-project.org/web/view>)에서 살펴볼 수 있음
- R과 관련한 대부분의 문제는 Google 검색과 스택 오버플로우 (<http://stackoverflow.com>)에서 해결할 수 있음
- 해당 문서는 서민구 선생님의 “R을 이용한 데이터 처리 & 실무” [서민구, 2014], 고석범 선생님의 “R과 Knitr을 활용한 데이터 연동형 문서 만들기” [고석범, 2014], R Friend 님의 블로그 “R, Python 분석과 프로그래밍” [?] 의 내용을 주로 참조함.
- 해당 문서는 RMarkdown + L^AT_EX + knitr로 작성된 문서이며, 이를 위해 작성한 모든 소스 코드 및 자료는 <https://github.com/Secondmoon-Kiom/CNUH-R-Lecture-2017>에서 확인할 수 있음.

1.2 R 설치하기(Windows 용)

R은 공개 소프트웨어로 <http://www.r-project.org/>에서 다운로드 및 설치 가능

1. 웹브라우저(i.e. explore, chrome, firefox 등)에서 <http://www.r-project.org> 이동
2. 좌측 R Logo 하단 Download 아래 ‘CRAN’ 클릭

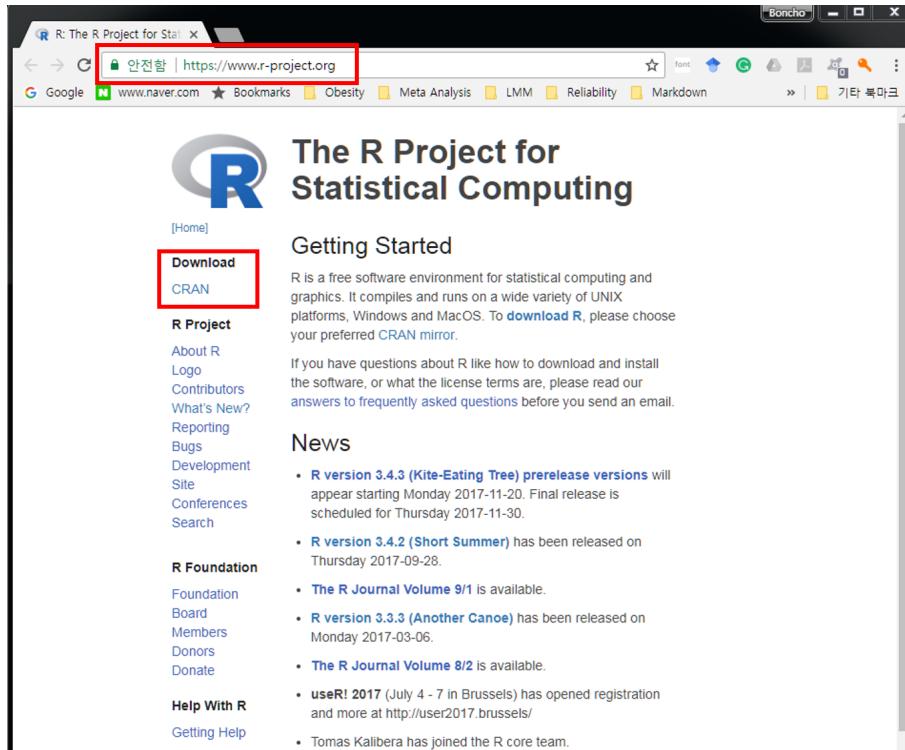


Figure 1.1: www.r-project.org 메인화면

3. 클릭 후 연결 창에서 스크롤 후 “Korea” 아래 링크 클릭 (그림 1.2 참조)

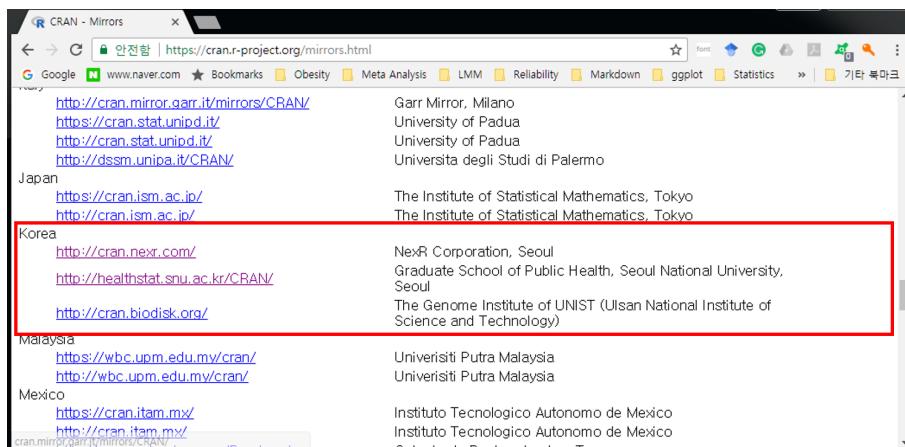


Figure 1.2: CRAN 국가별 mirrors

4. 클릭 후 세 가지 운영체제(Linux, Mac OS X, Windows)에 따른 R 버전 선택 가능

- 본 문서에서는 Windows 버전 설치만 다룸

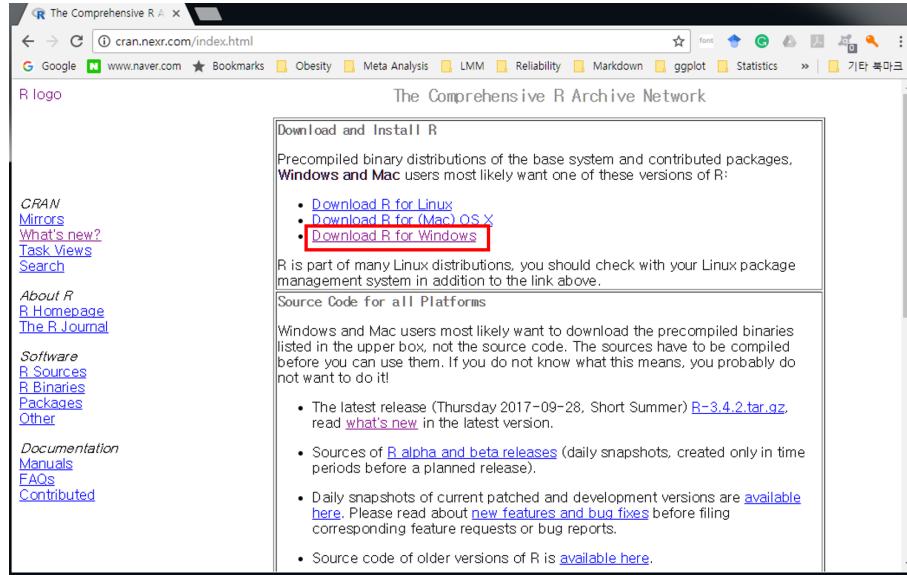


Figure 1.3: 운영체제 별 R 버전 선택

5. “Downloads R for Windows” 링크 클릭하면 다음과 같은 화면으로 이동

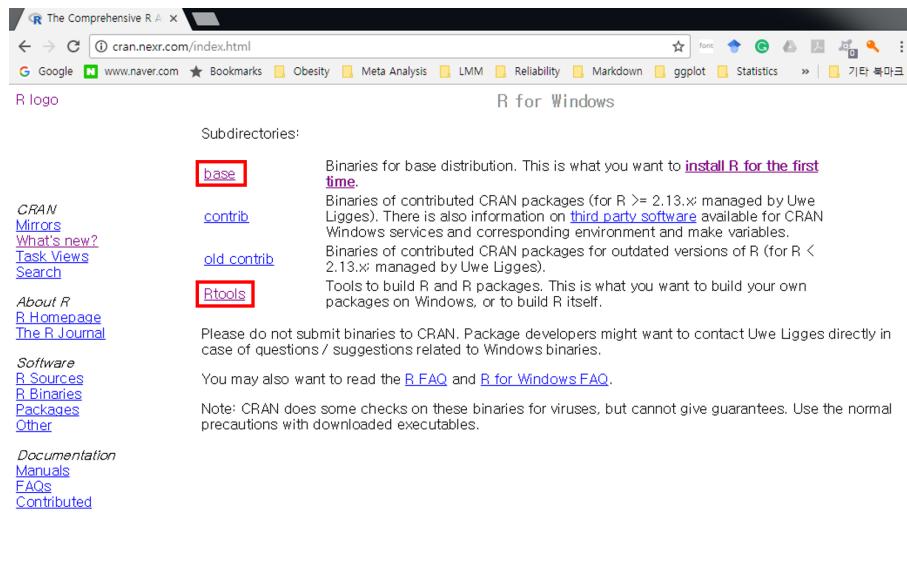


Figure 1.4: Windows 용 R base 및 구성요소 다운로드

6. R을 구성하는 하위구조 중 “base” 링크 클릭 후 다음 화면에서 “Downloads R 3.4.2 for Windows”를 클릭 후 설치 파일을 임의의 디렉토리에 저장 후 실행(그림 1.5 참조)

7. 참고로 3개 subdirectories에 대한 간략한 설명은 아래와 같음

- **base:** R 실행 프로그램

- contrib: R package의 바이너리 파일
- Rtools: R package 개발 및 배포를 위한 프로그램

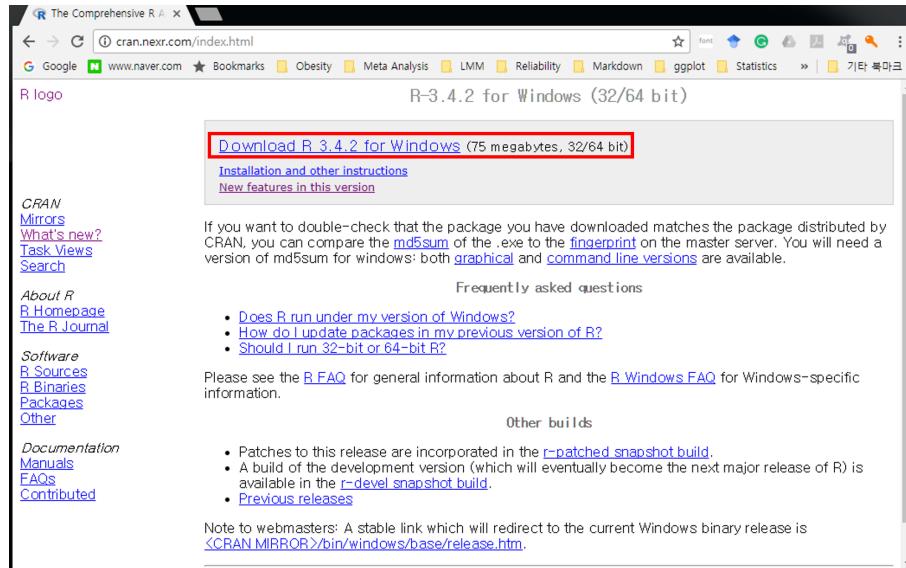


Figure 1.5: Windows 용 R 설치 파일 다운로드 페이지

8. 다운로드한 파일을 실행하면 아래와 같은 대화창이 나타남

- 한국어 선택 → 환영 화면에서 [다음(N)>] 클릭

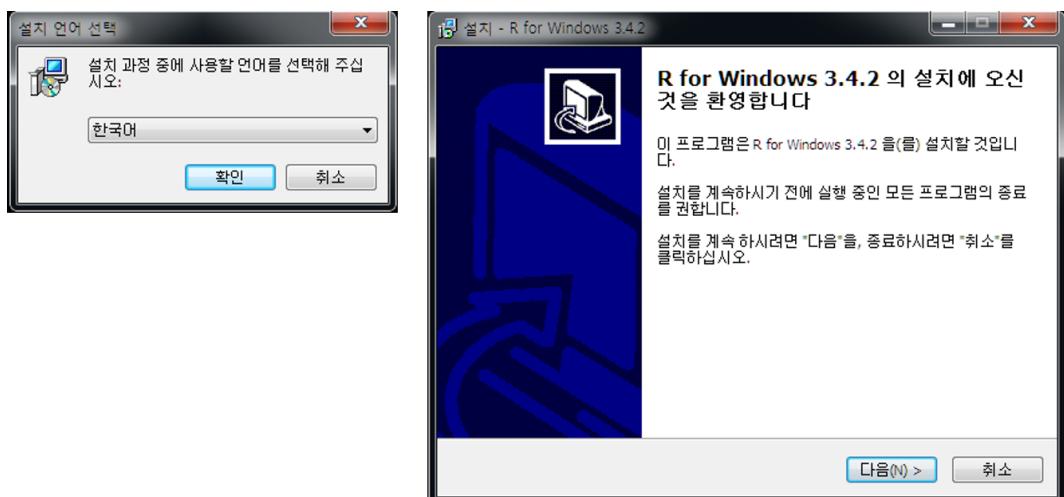


Figure 1.6: R 설치과정 01

9. GNU 라이센스에 대한 설명 및 동의 여부([다음(N)>]) 클릭 (그림 1.7)

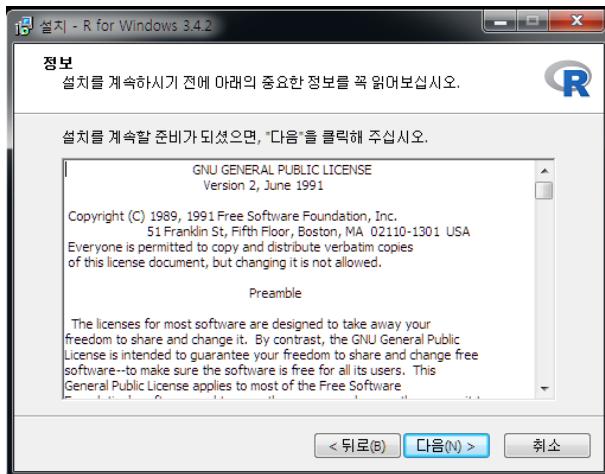


Figure 1.7: R GNU general license

10. 설치 디렉토리 설정 및 구성요소 설치 여부

- 원하는 디렉토리 설정(예: C:\R\R-3.4.2) (그림 1.8)
- 기본 프로그램("Core Files"), 32 또는 64 bit 용 설치 파일, R console 한글 번역 모두 체크 뒤 [다음(N)>] 클릭 (그림 1.9)

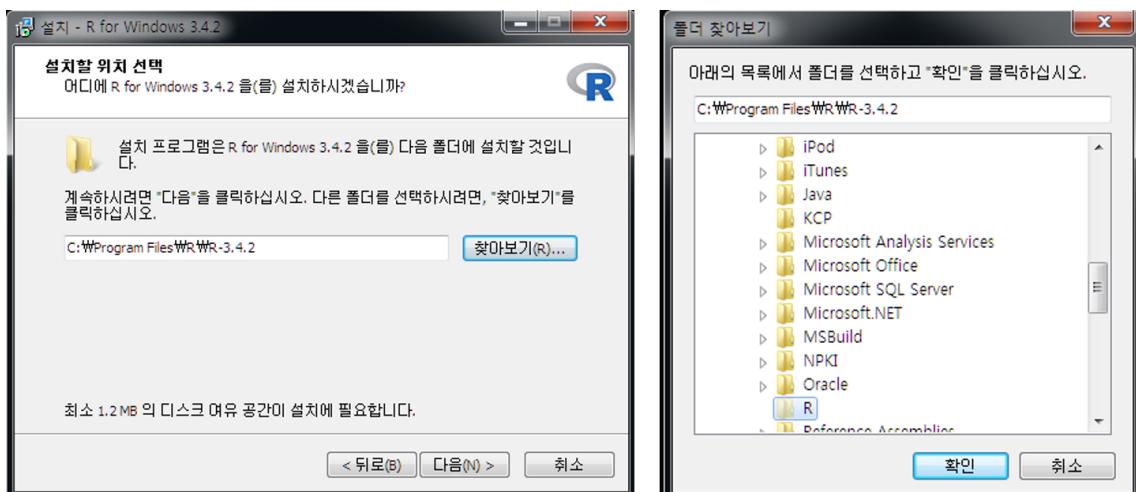


Figure 1.8: R 설치 디렉토리 설정

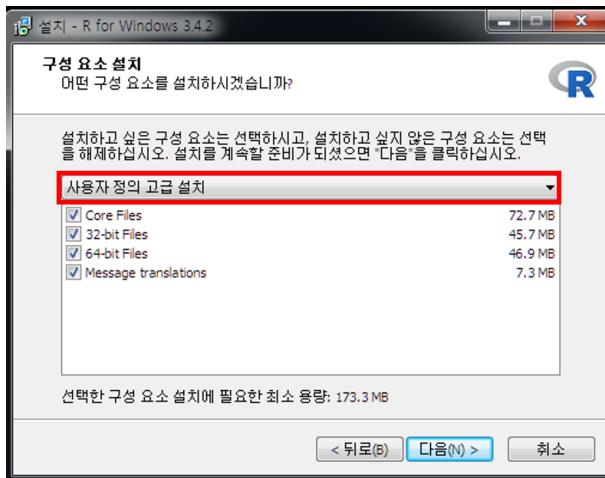


Figure 1.9: R 구성요소 설치

11. R 스타트업 옵션 지정

- 기본값("No" check-button)으로도 설치 진행 가능
- 본 문서에서는 스타트업 옵션 변경으로 진행

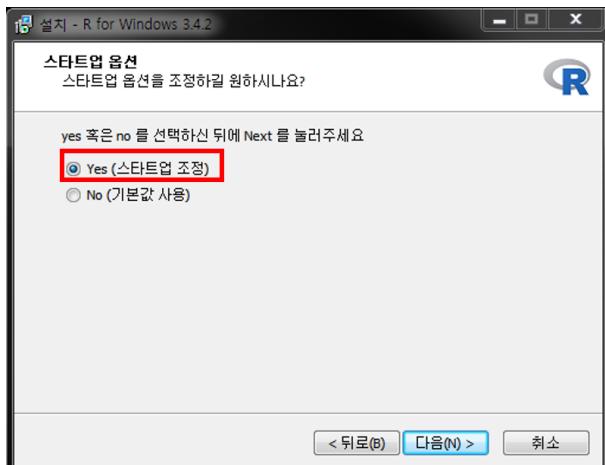


Figure 1.10: R 스타트업 옵션 변경

12. 화면표시방식(디스플레이) 모드 설정 변경

- MDI: 한 윈도우 내에서 script 편집창, 출력, 도움말 창 사용
- SDI: 다중 창에서 각각 script 편집창, 출력, 도움말 등을 독립적으로 열기

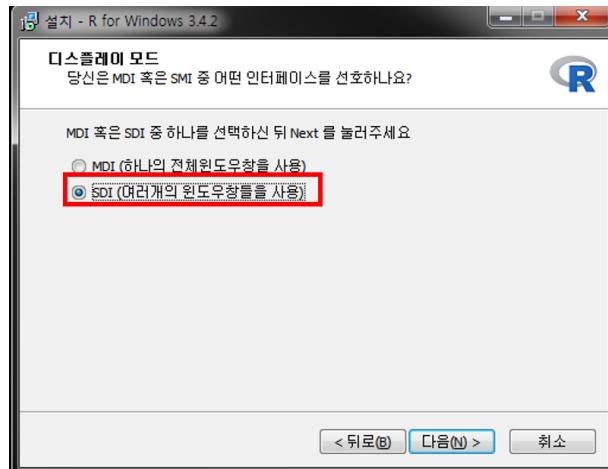


Figure 1.11: R 화면표시방식 설정 변경

13. 도움말 형식에서 HTML 도움말 기반 선택

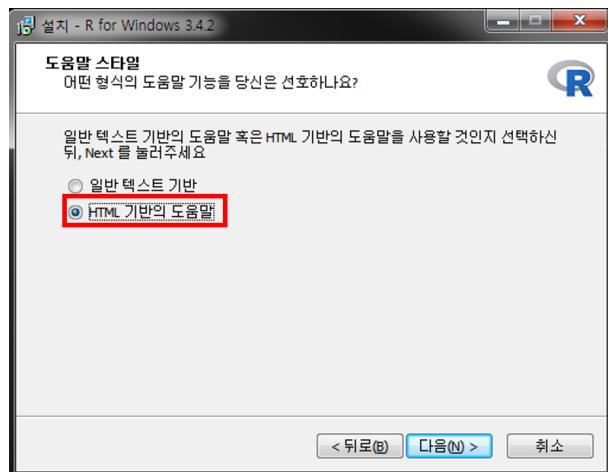


Figure 1.12: R 도움말 형식 변경

14. 시작메뉴 폴더 선택(그림 1.13)

- “바로가기”를 생성할 시작 메뉴 폴더 지정 후 [다음(N)>] 클릭 후 설치 진행
- 하단 “시작메뉴 폴더 만들지 않음” 체크박스 표시 시 시작메뉴에 “바로가기” 생성되지 않음(실행에 전혀 지장 없음)



Figure 1.13: 시작메뉴 폴더 선택

15. 추가 옵션 지정: 바탕화면 아이콘 생성 등 추가적 작업 옵션 체크 후 [다음(N)>] 클릭

→ 설치 진행

- 설치된 R 버전 정보 레지스트리 저장 여부
- .Rdata 확장자를 R 실행파일과 자동 연계

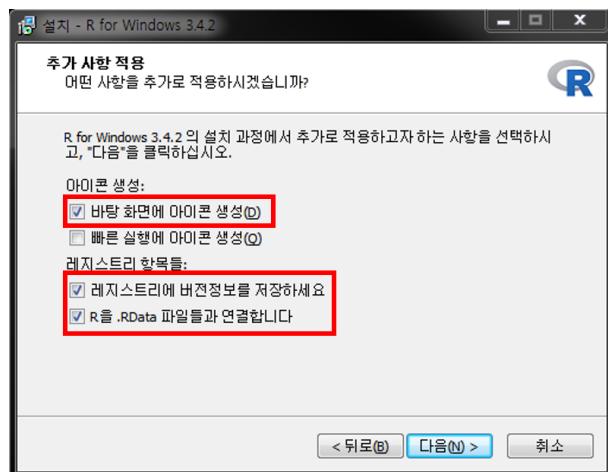


Figure 1.14: R 추가옵션 사항 선택

16. 설치 완료 후 바탕화면의 R 아이콘을 더블클릭하면 Rgui가 실행

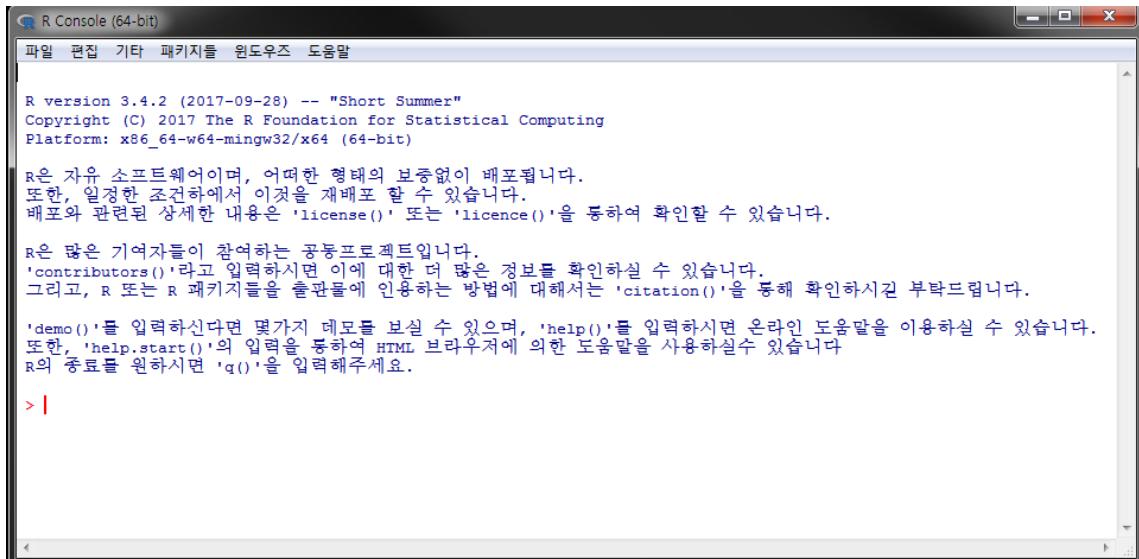


Figure 1.15: Windows에서 R 실행화면(콘솔 창, SDI 모드)

1.3 R 시작 및 작동 체크

위 R 시작화면에서 간단한 명령어들을 체크

- 그림 1.15에서 > 기호는 R의 명령 프롬프트임.
- Checklist
 - 1) “Hello R” 출력
 - 2) 1부터 100까지 정수 출력
 - 3) 간단한 histogram 출력

```
> # 문자열 출력
> print("Hello R")
```

```
[1] "Hello R"
```

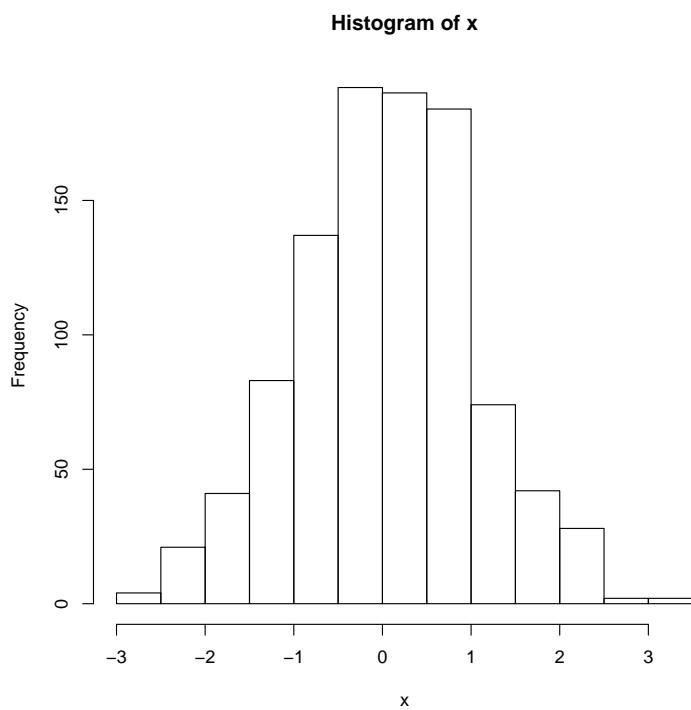
여기서 # 기호는 주석의 시작을 의미하며 같은 행에서 # 뒤 내용의 코드는 실행되지 않음

```
> # 1부터 100까지 수열 출력
> seq(1:100) # print('Hello R')
```

```
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
```

```
[18] 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34
[35] 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51
[52] 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68
[69] 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85
[86] 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
```

```
> # 간단한 히스토그램
> set.seed(12345) # random seed 지정
> x <- rnorm(1000) # 평균 0, 분산 1인 정규분포에서 난수 1000개 생성
> hist(x) # 히스토그램
```



Tips

R 명령어 실행 시 간혹 아니 매우 빈번히 오류가 나타나는데, 이를 해결할 수 있는 가장 손쉬운 방법은 Googling과 R의 도움말을 이용하는 것이 가장 효율적임.

도움말을 볼 수 있는 R 함수 리스트

Table 1.1: R help 관련 명령어 리스트

| 도움말 보기 명령어 | 설명 | 사용법 |
|-------------------|--|---|
| help 또는 ? | 도움말 시스템 호출 | help(topic # 도움말을 찾고자 하는 대상 또는 함수) |
| help.search 또는 ?? | 주어진 문자열을 포함한 문서 검색 | help.search(pattern # 찾고자 하는 문자열) |
| example | topic의 도움말 페이지에 있는 examples section 실행 | example(topic # 예제를 실행하고자 하는 topic 또는 함수) |
| vignette | topic의 pdf 또는 html 레퍼런스 메뉴얼 불러오기 | vignette(topic # topic에 저장된 reference manual) |

Tips

- vignette에서 제공하는 문서는 실제 분석을 기반으로 작성한 문서이기 때문에 초보자들이 R 패키지의 접근성을 높혀줌. browseVignettes()으로 현존하는 모든 R 패키지들의 vignette을 볼 수 있기 때문에 매우 유용함.
- R 세션(R 콘솔이 시작해서 종료까지)에 설정된 옵션은 options() 실행으로 확인 가능
- 현재 R 세션에 대한 정보는 sessionInfo() 함수로 확인 가능

참고

```
> # 현재 설정된 출력 자리수 확인
> options("digits")
```

```
$digits
```

```
[1] 7
```

```
> pi
```

```
[1] 3.141593
```

```
> # 출력 자리수를 7에서 3으로 변경
> options(digits = 3)
> pi
```

```
[1] 3.14
```

1.3.1 이것으로 R 설치 완료??

- 기본적 R 사용방식은 입력한 명령어와 실행결과를 확인하는 대화형(interpreter) 방식
- R 기본 콘솔창 안에서도 TAB을 누르면 자동완성 기능이라던가 ↑, ↓를 누르면 이전/ 이후 명령 기록을 볼 수 있음.
- 여러 줄 이상의 R 명령어라든가 반복적, 장기간 작업을 수행해야 하는 경우라면 R 명령어로 구성된 스크립트 작성 후 일괄 실행하는 것이 일반적임.
- 이러한 다중 명령 코딩 시 콘솔창에 직접 입력하는 것은 비효율적 → 스크립트 에디터를 주로 사용
- R 자체적으로 기본적인 스크립트 에디터 제공(R editor) → 가독성 및 코딩 효율이 떨어짐
- 대표적 R 에디터: WinEdt (<http://www.winedt.com>), Tinn-R (<https://sourceforge.net/projects/tinn-r/>), Vim (http://www.vim.org/scripts/script.php?script_id=2628)
- 대부분의 분석 및 개발 환경이 R GUI 만으로 구성되어 있지 않음 → **RStudio**를 이용한 통합 분석 환경 설정

1.4 RStudio 설치하기

- Rstudio: R 통합 분석/개발 환경(integrated development environment, IDE)으로 현재 가장 대중적으로 사용되고 있는 R 사용 환경
- 명령 콘솔 외 파일 편집, 데이터 보기, 명령 기록(.history), 그래프 등에 쉽게 접근 가능
- R과 마찬가지로 무료 소프트웨어임

1. Rstudio 사이트 접속

- 웹 브라우저를 통해 <https://www.rstudio.com> 연결 후 메인 화면에서 “Download RStudio” 클릭
- 혹은 상단 Pop-up menu 중 Products → RStudio 클릭 후 연결된 화면에서 다운로드 진행

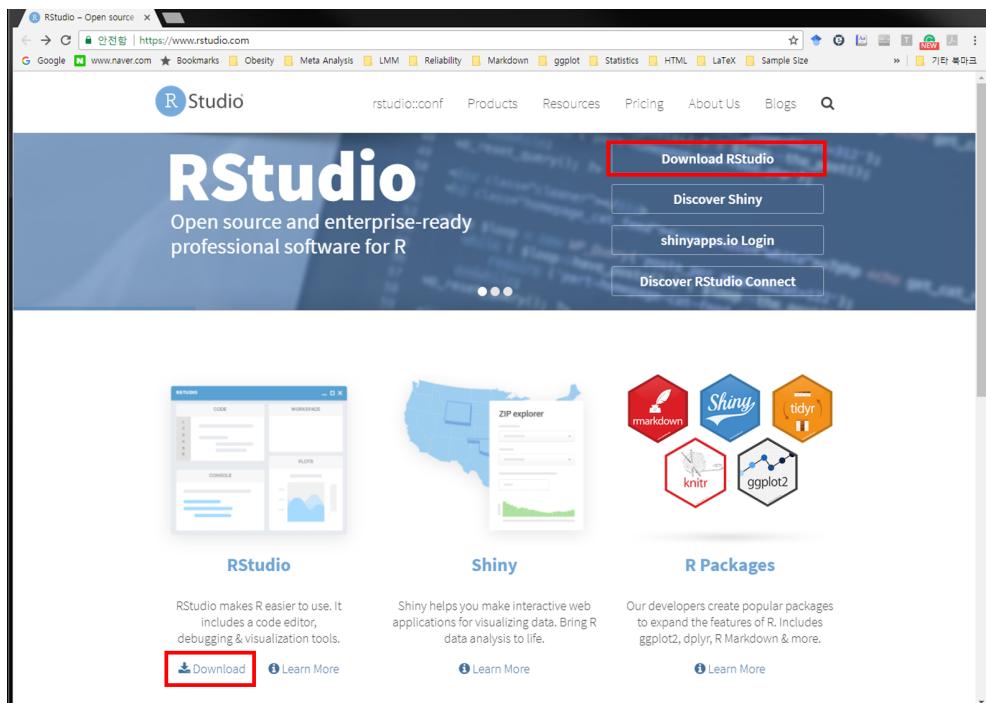


Figure 1.16: RStudio 메인 페이지 화면

2. Desktop 또는 Server 버전 중 택일

- 서버용 설치를 위해서는 Server 클릭 → 소규모 자료 분석용으로는 불필요
- 여기서는 “Desktop” 버전 선택 후 다음 링크로 이동

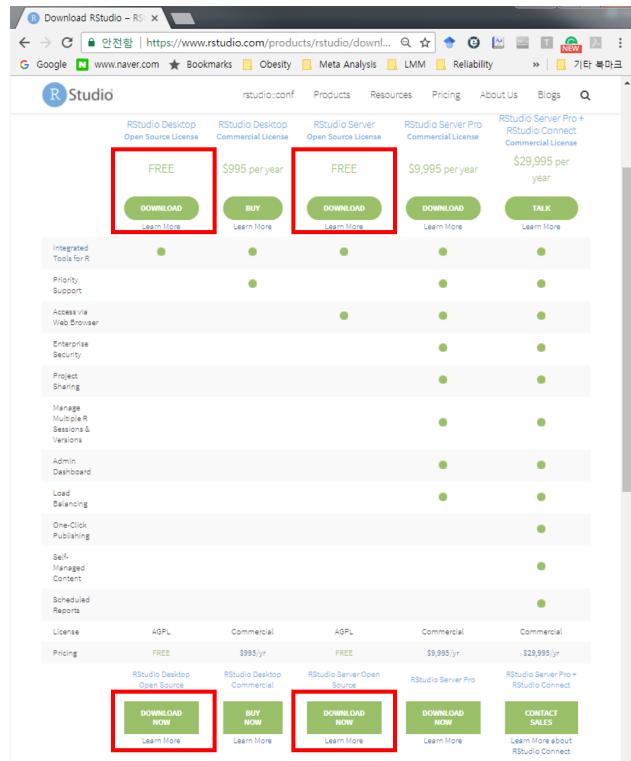


Figure 1.17: RStudio 다운로드 페이지

3. 운영체제에 맞는 Rstudio installer 다운로드(여기서는 Windows 버전 다운로드)

The screenshot shows the RStudio 1.1.383 Release Notes page. It includes a "Have Questions?" section and a "TALK TO THE SALES TEAM" button.

Installers for Supported Platforms

| Installer | Size | Date | MD5 |
|---|---------|------------|------------------------------------|
| RStudio 1.1.383 - Windows Vista/7/8/10 | 83.8 MB | 2017-10-09 | 45f759b053dcdbaa60be641b52e13c01 |
| RStudio 1.1.383 - Mac OS X 10.9+ (64-bit) | 14.3 MB | 2017-10-09 | 8cf1211930d00801700fc4859779536e03 |
| RStudio 1.1.383 - Ubuntu 12.04-18 (Debian 8 (32-bit)) | 88.2 MB | 2017-10-09 | 958b0ce14612a5e0a29ca4ba35941a |
| RStudio 1.1.383 - Ubuntu 12.04-18 (Debian 8 (64-bit)) | 97.4 MB | 2017-10-09 | 3ee6de231b7203a7e7ebf1091049913581 |
| RStudio 1.1.383 - Ubuntu 16.04+ (Debian 9+ (64-bit)) | 65 MB | 2017-10-09 | 1ccce7cb1773c3464ea6c91fc2e2c8 |
| RStudio 1.1.383 - Fedora 19+ (Red Hat 7+ / openSUSE 13.1+ (32-bit)) | 88.1 MB | 2017-10-09 | 36b4f00b6ec5c6a39194287b468cbe44 |
| RStudio 1.1.383 - Fedora 19+ (Red Hat 7+ / openSUSE 13.1+ (64-bit)) | 90.6 MB | 2017-10-09 | ae400e2504e9c59b2343c24fe3cd61d |

Zip/Tarballs

| Zip/tar archives | Size | Date | MD5 |
|---|----------|------------|----------------------------------|
| RStudio 1.1.383 - Windows Vista/7/8/10 | 121.0 MB | 2017-10-09 | ab1c9ac40cbx591c141ede7ef33000be |
| RStudio 1.1.383 - Ubuntu 12.04-18 (Debian 8 (32-bit)) | 90 MB | 2017-10-09 | b4c776c224514057656a3016ad9 |
| RStudio 1.1.383 - Ubuntu 12.04-18 (Debian 8 (64-bit)) | 99.3 MB | 2017-10-09 | 3701655b9b37110ad19430a58a944 |
| RStudio 1.1.383 - Fedora 19+ (Red Hat 7+ / openSUSE 13.1+ (32-bit)) | 88.8 MB | 2017-10-09 | 4e01a3242bc3e672061c74399e902b |
| RStudio 1.1.383 - Fedora 19+ (Red Hat 7+ / openSUSE 13.1+ (64-bit)) | 91.4 MB | 2017-10-09 | a1c6621661d7fc7b4a6011d174b4d |

Source Code

A tarball containing source code for RStudio v1.1.383 can be downloaded from here.

Figure 1.18: RStudio 운영체제 선택

4. RStudio installer 다운로드 시 파일이 저장된 폴더에서 보통 RStudio-xx.xx.xxx.exe

형식의 파일명 확인

- 더블 클릭 후 실행
- 다음> 몇 번 누르면 설치 종료

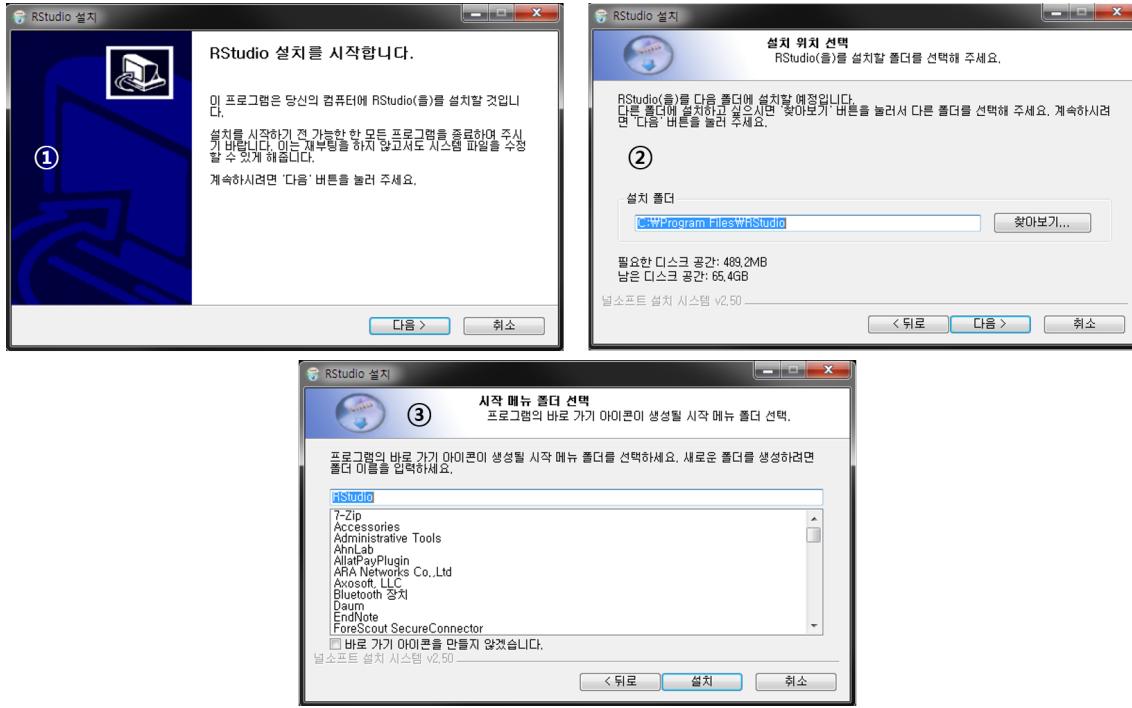


Figure 1.19: RStudio 설치화면

5. 아래와 같은 실행화면이 나타나면 RStudio 설치 성공

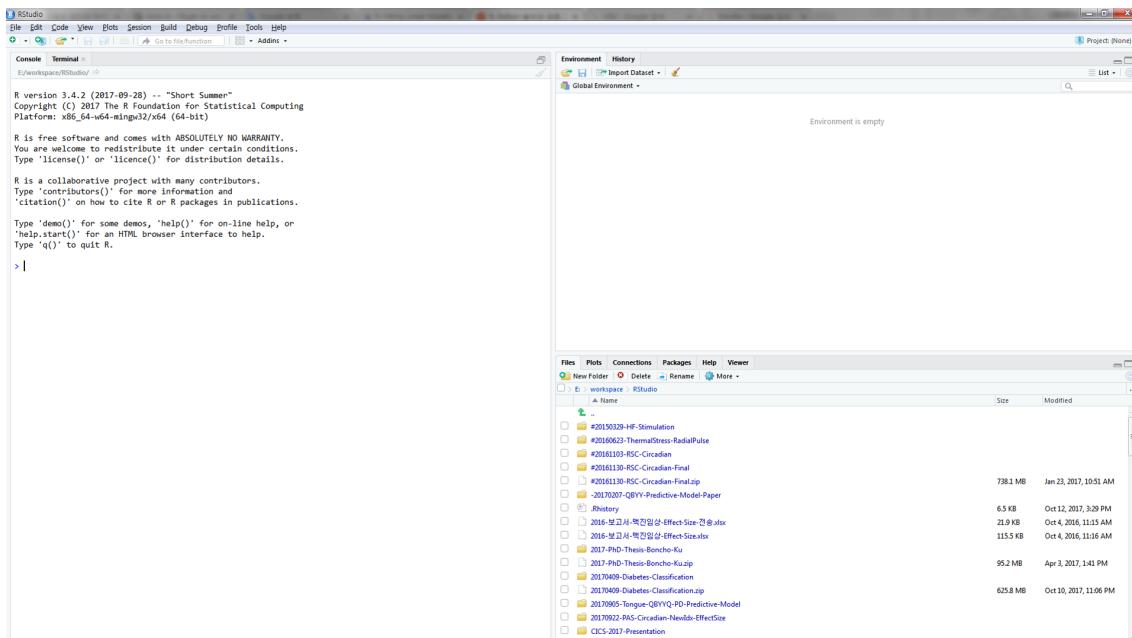


Figure 1.20: RStudio 초기 실행화면

1.5 RStudio의 구성

1.5.1 RStudio IDE 화면 구성

RStudio는 아래 그림 1.21과 같이 크게 4개 창으로 구성

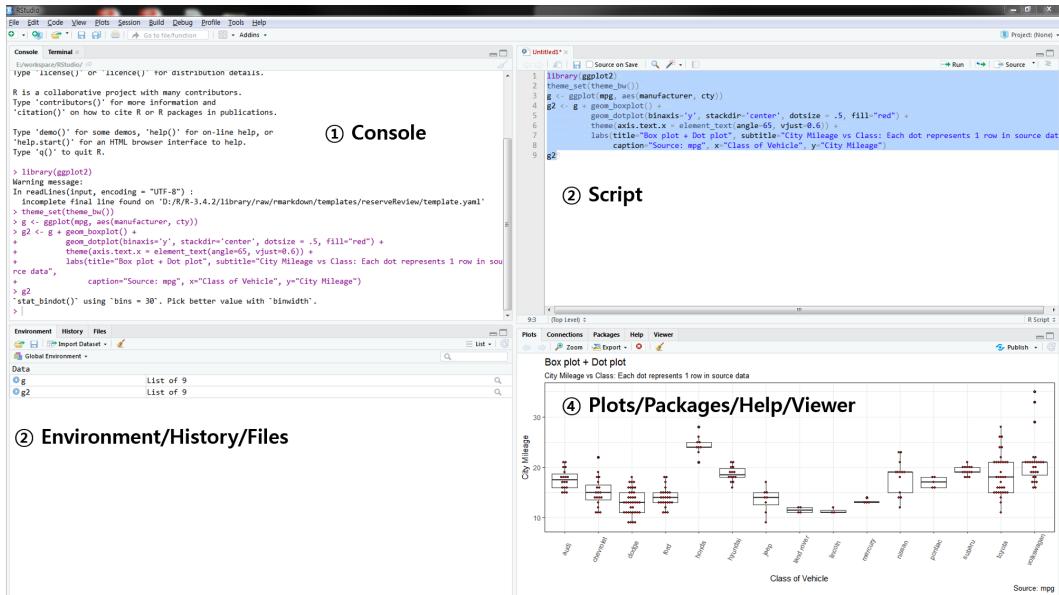


Figure 1.21: RStudio 화면구성: 우하단 그림은 <http://r-statistics.co/Top50-Ggplot2-Visualizations-MasterList-R-Code.html>에서 발췌

1. 콘솔(console)

- R 명령어 실행 공간(RGui, 정확하게는 Rterm.exe가 실행되고 있는 창)
- R 스크립트 또는 콘솔 창에서 작성한 명령어(프로그램) 실행 결과 출력
- 경고, 에러/로그 등의 메세지 확인

2. 스크립트(script) R 명령어 입력 공간으로 일괄처리(batch processing) 가능

- 일괄 처리를 위한 RStudio 제공 단축키

- **Ctrl + Enter**: 선택한 블럭 내 명령어 실행
- **Alt + Enter**: 선택 없이 커서가 위치한 라인의 명령어 실행

- 새 R 스크립트 파일 열기

- R 상단 메뉴: **File** → **New File** → **R Script**
- 단축키: **Ctrl + Shift + N**

- R 스크립트 이외 R Markdown, R Notebook, Shiny Web Application 등 새 문서의 목적에 따라 다양한 종류의 문서 생성 가능

3. Environment/History

- 1) Environment: 현재 R 작업환경에 저장되어 있는 객체의 특성을 요약 제시
- 좌측 화살표 버튼 클릭: 해당 객체의 상세 정보 확인(그림 1.22)
 - 우측 사각형 버튼 클릭: 객체가 데이터프레임인 경우 스프레드 시트 형태로 데이터셋 확인(그림 1.23)
 - 디스크 아이콘: 작업공간 저장 여부
 - 빗자루 아이콘: 작업공간 객체 일괄 삭제
 - Global Environment 아이콘: 현재 작업공간에서 사용중인 패키지 및 선택 패키지 내장 함수 목록 확인

Figure 1.22: RStudio Environment 창: 객체 상세정보

| | manufacturer | model | displ | year | cyl | trans | drv |
|----|--------------|------------|-------|------|-----|------------|-----|
| 1 | audi | a4 | 1.8 | 1999 | 4 | auto(l5) | f |
| 2 | audi | a4 | 1.8 | 1999 | 4 | manual(m5) | f |
| 3 | audi | a4 | 2.0 | 2008 | 4 | manual(m6) | f |
| 4 | audi | a4 | 2.0 | 2008 | 4 | auto(av) | f |
| 5 | audi | a4 | 2.8 | 1999 | 6 | auto(l5) | f |
| 6 | audi | a4 | 2.8 | 1999 | 6 | manual(m5) | f |
| 7 | audi | a4 | 3.1 | 2008 | 6 | auto(av) | f |
| 8 | audi | a4 quattro | 1.8 | 1999 | 4 | manual(m5) | 4 |
| 9 | audi | a4 quattro | 1.8 | 1999 | 4 | auto(l5) | 4 |
| 10 | audi | a4 quattro | 2.0 | 2008 | 4 | manual(m6) | 4 |
| 11 | audi | a4 quattro | 2.0 | 2008 | 4 | auto(s6) | 4 |
| 12 | audi | a4 quattro | 2.8 | 1999 | 6 | auto(l5) | 4 |
| 13 | audi | a4 quattro | 2.8 | 1999 | 6 | manual(m5) | 4 |
| 14 | audi | a4 quattro | 3.1 | 2008 | 6 | auto(s6) | 4 |
| 15 | audi | a4 quattro | 3.1 | 2008 | 6 | manual(m6) | 4 |

Figure 1.23: RStudio Environment 창: 스프레드 시트

- History: R 콘솔에서 실행된 명령어(스크립트)들의 이력 확인

```

Environment History Connections Git
ReadExcel <- function(filename) {
  require(XLConnect)
  require(plyr)
  WB <- loadWorkbook(filename)
  SheetName <- getSheets(WB)
  DF1 <- lapply(SheetName, function(name) readWorksheet(WB, sheet=name))
  names(DF1) <- SheetName
  return(DF1)
}
ReadExcel <- function(filename) {
  require(XLConnect)
  require(plyr)
  WB <- loadWorkbook(filename)
  SheetName <- getSheets(WB)
  DF1 <- lapply(SheetName, function(name) readWorksheet(WB, sheet=name))
  names(DF1) <- SheetName
  return(DF1)
}

```

Figure 1.24: RStudio History 창

4. File/Plots/Packages/Help/Viewer

1) File: Windows 탐색기와 유사

- 파일 및 폴더 생성, 삭제 수정, 그리고 작업경로 설정

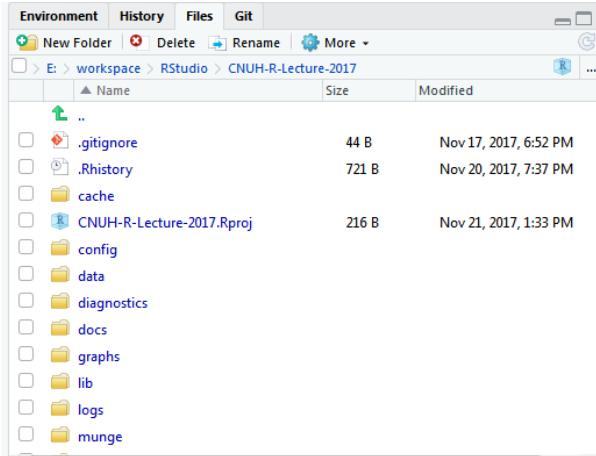


Figure 1.25: RStudio File 창

2) Plots: 콘솔 또는 스크립트 창으로부터 생성한 그래프 출력

- 작업 중 생성한 그래프가 이력에 따라 저장: 이전, 최근
- 기능으로 그래프 확대 가능
- 를 통해 선택 그래프를 이미지 파일(.png, .jpeg 등), PDF 파일 및 PDF 출력 파일로 저장 가능
- 엑스박스는 현재 화면에 출력된 그래프 삭제, 빗자루 아이콘은 생성한 모든 그래프 삭제

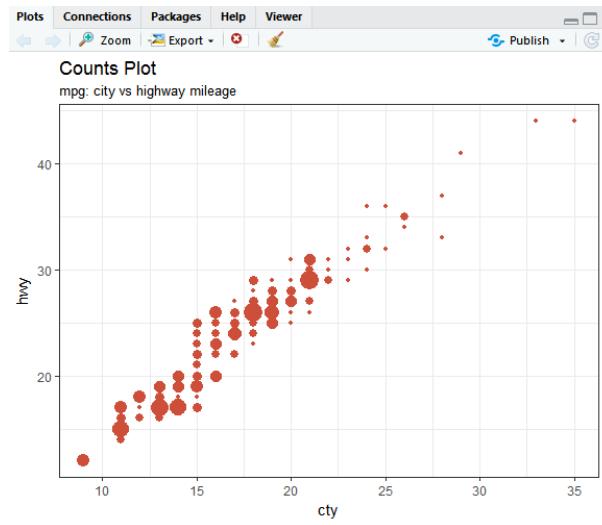


Figure 1.26: RStudio Plots 창 화면

- 3) Packages: 현재 컴퓨터에 설치된 R 패키지 목록 출력: 신규 설치 및 업데이트 가능

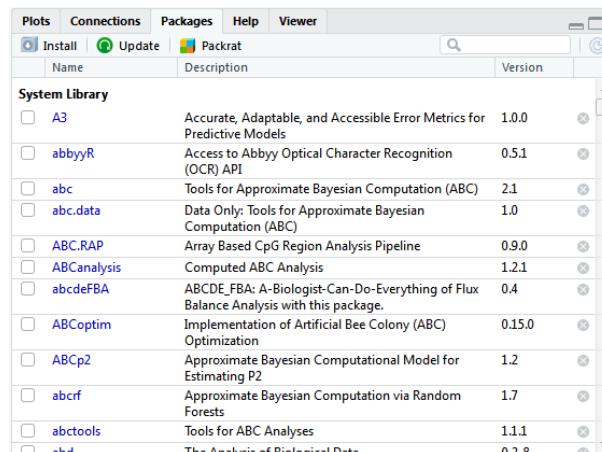


Figure 1.27: RStudio Packages 창 화면

- 4) Help: `help(topic)` 입력 시 도움말 창이 출력되는 공간

```
> help(lm)
```

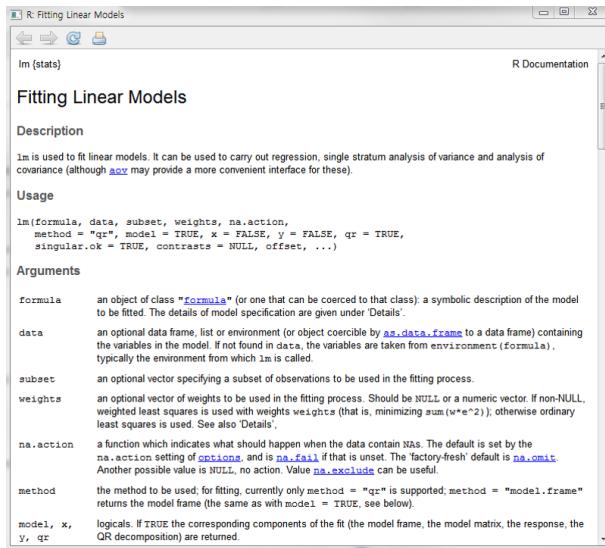


Figure 1.28: `help(lm)` 실행 후 RStudio Help 창 화면

5. RStudio의 창 layout은 [Tools] > [Global Options] > [Pane Layout]에서 변경 가능

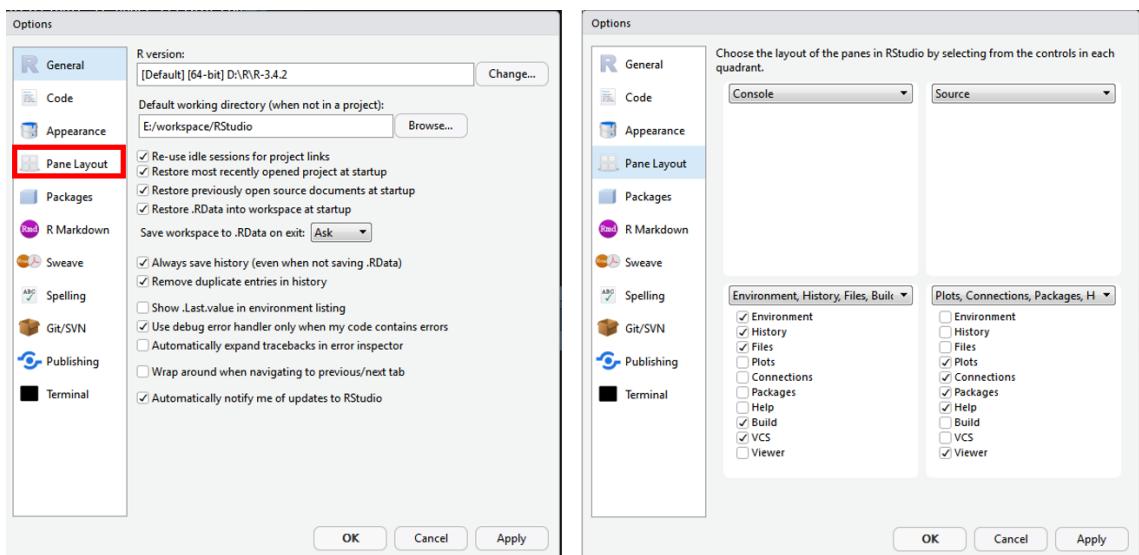


Figure 1.29: RStudio Option 선택화면 및 Pane 레이아웃 선택 화면

1.6 RStudio에서 배치 파일 생성 및 실행

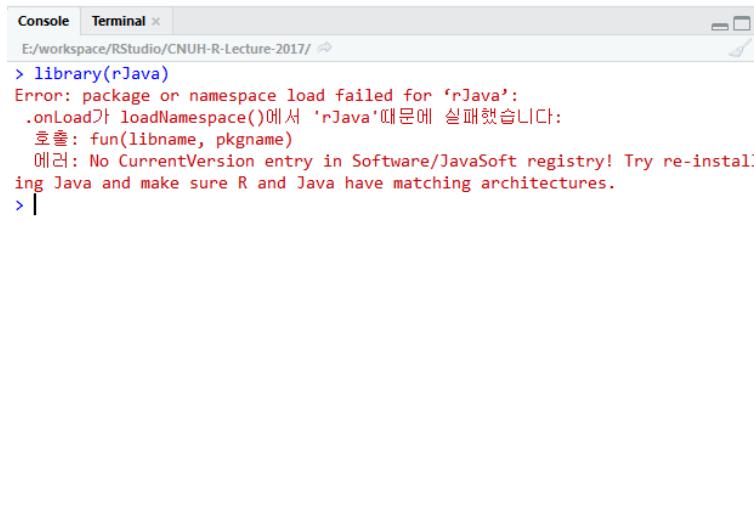
1.7 R 패키지 설치하기

1. RStudio

1.8 rJava 설치하기

1. 자바 명령을 호출하는 패키지를 정상적으로 사용하기 위한 패키지

- Java 계열 머신러닝 어플리케이션 활용 또는 최근까지 Excel 파일을 불러오기 위해 설치 필수(Excel에 관해서는 Hadley Wickham의 `readxl` 개발 이전에만...)
- 다음은 Java가 설치되지 않은 환경에서 `rJava`를 불렀을 때 발생한 오류 예시임



The screenshot shows the RStudio interface with the 'Console' tab selected. The console window displays the following text:

```
Console Terminal ×
E:/workspace/RStudio/CNUH-R-Lecture-2017/
> library(rJava)
Error: package or namespace load failed for 'rJava':
.onLoad가 loadNamespace()에서 'rJava'때문에 실패했습니다:
호출: fun(libname, pkgname)
에러: No CurrentVersion entry in Software/JavaSoft registry! Try re-installing Java and make sure R and Java have matching architectures.
> |
```

Figure 1.30: `rJava` 에러 메세지

2. `rJava` 설치 전 자바 설치 및 Windows 시스템 환경 설정 필요

3. 자바 개발 키트(Java Development Kit, JDK)는 <http://www.oracle.com/technetwork/java/javase/downloads/index.html>에서 다운로드 가능

- [Java Platform (JDK)] ▷ [Download] 클릭

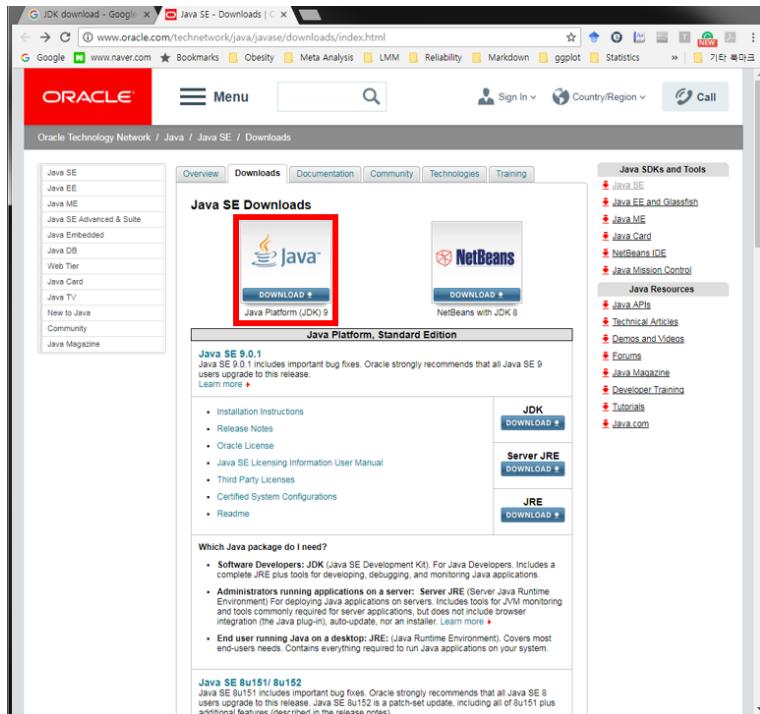


Figure 1.31: JDK 다운로드 페이지

4. JDK 다운로드 전 현재 주로 작업하고 있는 R 버전 정보(32 bit vs. 64 bit) 확인

- R 세션 시작 시 또는 RStudio에서 [Tools] > [Global Options]에서 R 버전 확인 가능
- 라이센스 계약에 동의 후 확인한 R 버전에 대응하는 링크 클릭 후 다운로드 진행

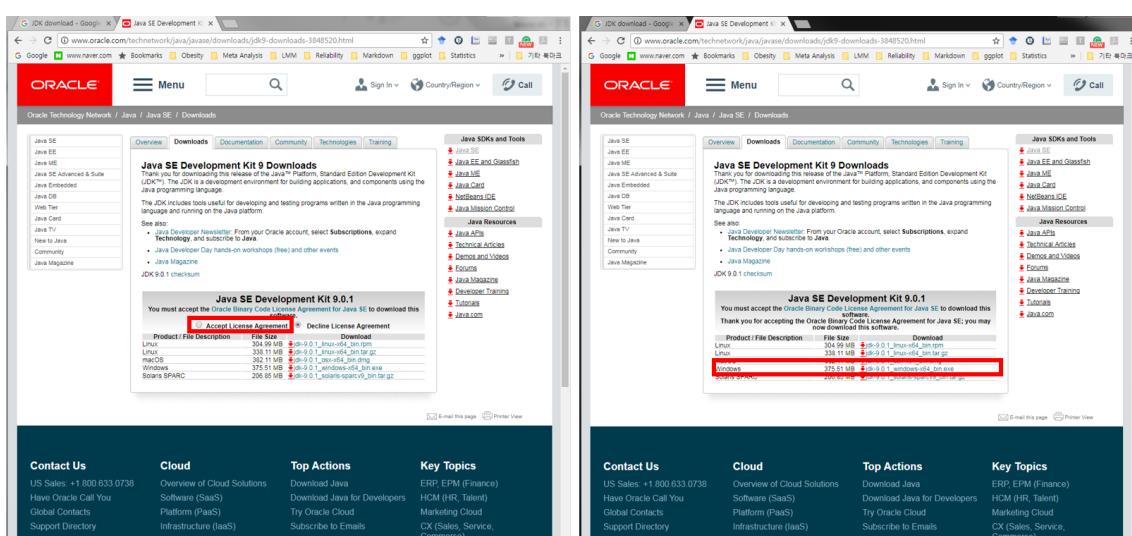


Figure 1.32: JDK 라이센스 및 다운로드 링크 페이지

5. 다운로드 완료 후 설치 확인

- 정상적으로 설치한 경우라면 C:\Program Files\Java\jdk-9.0.1 디렉토리에서 설치 확인 가능
- Java runtime environment (JRE)는 C:\Program Files\Java\jre-9.0.1에서 설치 확인 가능

6. 새 사용자 변수 생성

- Windows 시작** 클릭 → [컴퓨터] 우 클릭 → [속성] 클릭 후 다음 그림 1.33에서 [고급 시스템 설정] 클릭



Figure 1.33: Windows 시스템 설정 화면

- 시스템 속성 창에서 [환경변수(N)...] 클릭 후 환경변수 창에서 사용자 변수 [새로 만들기(N)...] 클릭
- 변수 이름에 JAVA_HOME, 변수 값에 JDK 설치 디렉토리 이름(예: C:\Program Files\Java\jdk-9.0.1) 입력

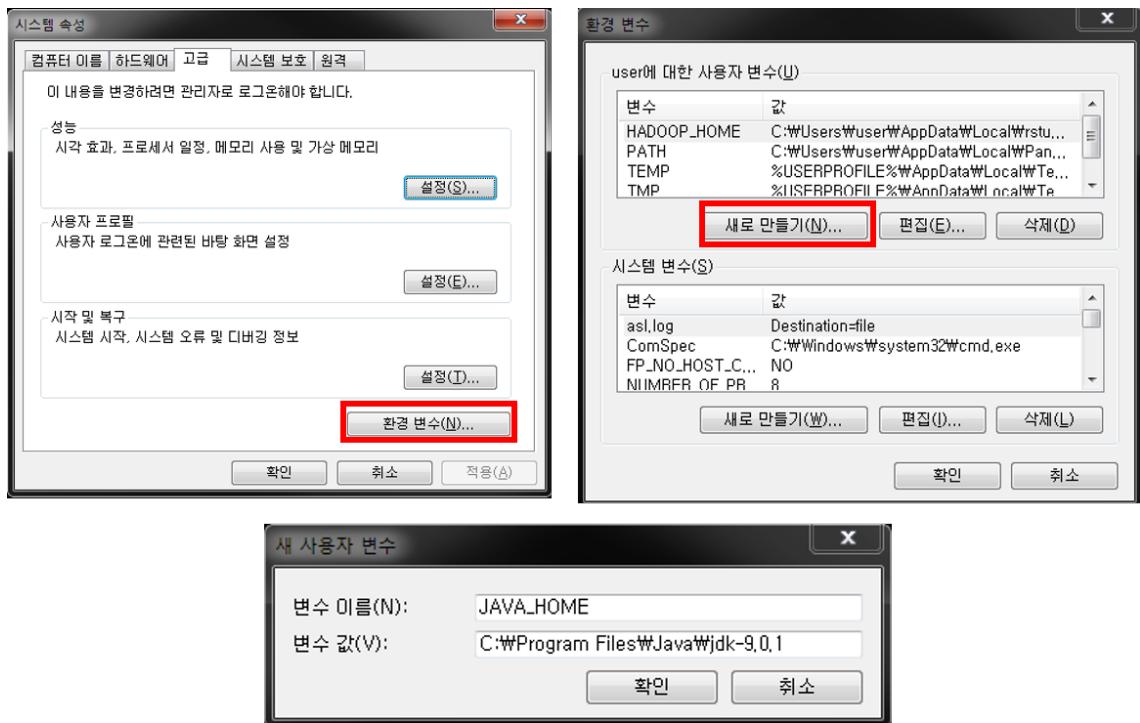


Figure 1.34: Windows 시스템 설정 화면

7. 시스템 변수 PATH 갱신

- JRE가 설치된 디렉토리(아래 나열)를 기존 PATH에 추가(디렉토리 구분자: ;)
 - C:\Program Files\Java\jre-9.0.1\bin
 - C:\Program Files\Java\jre-9.0.1\bin\server

8. 설정 완료 후 R 또는 RStudio 재실행 후 rJava 설치

9. 별 다른 에러 메세지 출력이 없으면 성공적으로 설치

1.9 RStudio 프로젝트 생성 및 ProjectTemplate 패키지

연동

1.10 R + RStudio를 잘 활용 하려면?

1. 주기적인 update
2. RStudio 제

Chapter 2

R의 기본 사용

- R은 객체지향언어(object-oriented language)임
- 객체: 숫자, 데이터셋, 단어, 테이블, 분석결과 등 모든 것을 칭함
 - “객체지향”의 의미는 R의 모든 명령어는 객체를 대상으로 이루어진다는 것을 의
- R에서 사용자가 데이터 입력을 위해 생성 또는 읽어온 객체(object)는 종종 변수(variable)라는 말과 혼용됨.
- 본 문서에서는 최상위 데이터 저장장소를 객체라고 명명하며 데이터프레임과 같이 여러 종류의 데이터타입으로 이루어진 객체의 1차원 속성을 변수라고 칭함.
- R을 다룰 수 있으려면 기본적인 R의 데이터 할당 방식과 데이터 타입, 그리고 연산 방법에 대한 이해가 있어야 함.
- 특히 데이터 타입에 따라 사용하는 함수 또는 분석 방법이 조금씩 다른 것이 R의 특징임
- 따라서 R의 데이터 타입에 좀 더 익숙해져야 R을 보다 쉽게 다룰 수 있음

Tips: 들어가기 앞서 알아두면 유용한 단축키 및 명령어

- R console에 저장된 모든 객체들을 일괄 삭제: `rm(list = ls())`
- R console 화면 청소: `ctrl + L`

2.1 R 객체 입력 방법 및 변수 설정 규칙

- R의 데이터 타입은 1)스칼라, 2)벡터, 3)요인, 4)행렬, 5)배열, 6)데이터프레임, 7)리스트로 나눌 수 있음.
 - 자세히 들어가기 전 R 데이터 타입의 구조 다이어그램을 먼저 살펴봄

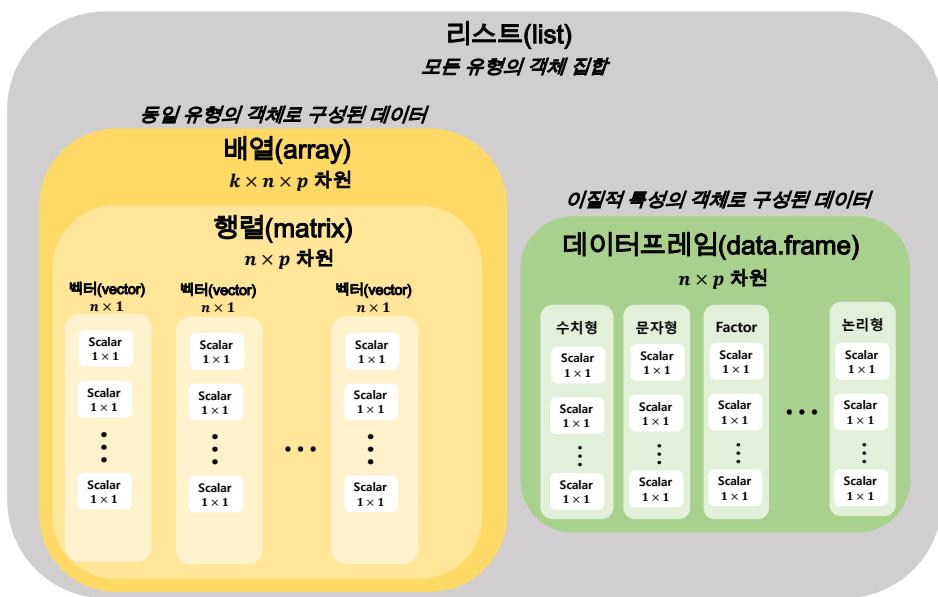


Figure 2.1: R 데이터 타입 구조 다이어그램

2.1.1 객체 입력 및 할당

- R의 가장 기본적인

앞서 열거한 7 가지 데이터 타입을 설명하기 전,

- ### 1. 객체를 할당하는 두 가지 방법

$$1) =$$

2) <- 또는 ->

3) 두 할당지시자의 차이점:

2.2 스칼라

2.3 벡터

2.4 리스트

2.5 행렬

2.6 배열

2.7 데이터 프레임

2.8 데이터 연산

Chapter 3

데이터 조작

3.1 실습 데이터 설명

3.2 파일 입출력

3.2.1 외부 파일 읽어오기/내보내기

`read.table()`, `write.table()` `readxl::read_excel()`

RStudio 메뉴 명령을 통한 파일 입출력

3.3 벡터화 연산(apply 계열 함)

3.4 dplyr + tidyr을 이용한 data wrangling

Chapter 4

ggplot2를 이용한 자료 시각화

4.1 ggplot() vs. plot()

4.2 Grammar of Graphics

Chapter 5

재현가능한 데이터 연동 문서 작성

References

- 고석범. *R과 Knitr를 활용한 데이터 연동형 문서 만들기*(에이콘 데이터 과학 13). 에이콘출판, 2014. ISBN 9788960775510. URL <https://books.google.co.kr/books?id=7UXFoAEACAAJ>.
- 서민구. *R을 이용한 데이터 처리 및 분석 실무*. 길벗, 2014. ISBN 9788966188260. URL <https://books.google.co.kr/books?id=0GjAoQEACAAJ>.