

1강

# R 데이터처리(1)

정보통계학과 김성수교수

## 목 차

1. R 및 R Studio
2. 데이터파일 읽기
3. 데이터베이스 자료 읽기
4. 통계패키지 자료 읽기
5. Rdata 저장 및 가져오기



# 1

# R 및 R Studio



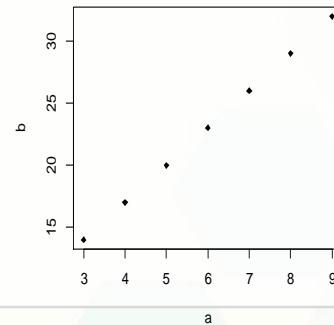
# R의 소개

- R은 자료처리, 통계분석, 통계그래프 등에 뛰어난 기능을 가지고 있는 **무료 통계시스템**.
- R은 **대화형 프로그램 언어**(interpreted programming language)
- R은 **객체지향**(object-oriented) 시스템
  - 데이터, 변수, 행렬 등은 모두 객체(object)
  - 객체는 연산자 “<-”, 또는 “=”에 의해 생성됨.

예)

```
> x = 2:10  
> y = 3*x + 5  
> x  
[1] 2 3 4 5 6 7 8 9 10  
> y  
[1] 11 14 17 20 23 26 29 32 35
```

```
> a <- 3:9  
> b <- 3*a + 5  
> plot(a,b, pch=18)
```



# R의 태동

- **S의 탄생** : Becker and Chambers (AT&T Bell Lab) 가 1980년대에 새로 개발한 통계프로그램 언어를 S 라 명함 – **S-PLUS** 시스템으로 발전.
- **R의 탄생** : Ross Ihaka and Robert Gentleman(Univ. of Auckland, New Zealand)가 교육 목적으로 S 의 축소버전 (reduced version) “R & R” 을 만듬
- **R의 발표** : 1995년 Martin Maechler 0| Ross Ihaka and Robert Gentleman 를 설득하여 Linux system 과 같이 Open Source Software 규약인 GPL(General Public Licence) 규약 하에 R의 source code를 발표
- **R Core Team 의 결성** : 1997년 8월 R 시스템의 발전을 위한 국제적인 R core team이 결성됨. 이후 확장 발전하여 현재(2015년 7월) 21명의 멤버로 구성됨. 2000년 2월 29일 R version 1.0.0 발표됨. 2015년 7월 현재 R version 3.2.1.

참고 : [www.r-project.org](http://www.r-project.org)

Peter Dalgaard (2005), Introductory Statistics with R, Springer.

# R 다운받기 : [www.r-project.org](http://www.r-project.org)



[Home]

[Download](#)

[CRAN](#)

[R Project](#)

[About R](#)

[Contributors](#)

[What's New?](#)

[Mailing Lists](#)

[Bug Tracking](#)

[Conferences](#)

[Search](#)

[R Foundation](#)

[Foundation](#)

[Board](#)

[Members](#)

[Donors](#)

[Donate](#)

## 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 Mac OS. 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

- [The R Journal Volume 7/1](#) is available.
- [R version 3.2.1 \(World-Famous Astronaut\)](#) has been released on 2015-06-18.
- [R version 3.1.3 \(Smooth Sidewalk\)](#) has been released on 2015-03-09.
- [useR! 2015](#), will take place at the University of Aalborg, Denmark, June 30 - July 3, 2015.
- [useR! 2014](#), took place at the University of California, Los Angeles, USA June 30 - July 3, 2014.

# R 다운받기 : CRAN Mirrors

Italy

<http://cran.mirror.garr.it/mirrors/CRAN/>  
<http://cran.stat.unipd.it/>  
<http://dssm.unipa.it/CRAN/>

Japan

<http://cran.ism.ac.jp/>  
<http://cran.md.tsukuba.ac.jp/>  
<http://ftp.yz.yamagata-u.ac.jp/pub/cran>

Korea

<http://cran.nexr.com/>  
<http://healthstat.snu.ac.kr/CRAN/>  
<http://cran.biodisk.org/>

Lebanon

<http://rmirror.lau.edu.lb/>

Mexico

<http://cran.itam.mx/>  
<http://www.est.colpos.mx/R-mirror/>

Netherlands

<http://cran.xl-mirror.nl/>  
<http://cran-mirror.cs.uu.nl/>

New Zealand

<http://cran.stat.auckland.ac.nz/>

Garr Mirror, Milano

University of Padua

Università degli Studi di Palermo

Institute of Statistical Mathematics, Tokyo

University of Tsukuba

Yamagata University

NexR Corporation, Seoul

Graduate School of Public Health, Seoul National University, Seoul

The Genome Institute of UNIST (Ulsan National Institute of Science and Technology)

Lebanese American University, Byblos

Instituto Tecnologico Autonomo de Mexico

Colegio de Postgraduados, Texcoco

XL-Data, Amsterdam

Utrecht University

University of Auckland

# R 다운받기 : Download

## 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.

### 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 packages (managed by Uwe Ligges). There is also information on [third party software](#) available for CRAN Windows services and corresponding environment and make variables.

[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.

### R-3.2.1 for Windows (32/64 bit)

[Download R 3.2.1 for Windows](#) (62 megabytes, 32/64 bit)

[Installation and other instructions](#)

[New features in this version](#)

# R 매뉴얼



[Home]

Download

CRAN

R Project

About R  
Contributors  
What's New?  
Mailing Lists  
Bug Tracking  
Conferences  
Search

R Foundation  
Foundation  
Board  
Members  
Donors  
Donate

Documentation  
Manuals  
FAQs  
The R Journal  
Books  
Certification  
Other

Links

Bioconductor  
Related Projects

## The F

### Getting

R is a free software variety of UNIX mirror.

If you have questions, please re-

### News

- [R 3.2.0 \(Final\)](#) scheduled
- [R version](#)
- [The R Journal](#)
- [R version](#)
- [useR! 2013](#), took place at the University of Debrecen, Debrecen, Hungary June 30 - July 3, 2013.
- [useR! 2014](#), took place at the University of California, Los Angeles, USA June 30 - July 3, 2014.

edited by the R Development Core Team.

The following manuals for R were created on Debian Linux and may differ from the manuals for Mac or Windows on platform-specific pages, but most parts will be identical for all platforms. The correct version of the manuals for each platform are part of the respective R installations. The manuals change with R, hence we provide versions for the most recent released R version (R-release), a very current version for the patched release version (R-patched) and finally a version for the forthcoming R version that is still in development (R-devel).

Here they can be downloaded as PDF files, EPUB files, or directly browsed as HTML:

### Manual

An Introduction to R is based on the former "Notes on R", gives an introduction to the language and how to use R for doing statistical analysis and graphics.

### R-release

[HTML](#) | [PDF](#) |  
[EPUB](#)

### R-patched

[HTML](#) | [PDF](#) |  
[EPUB](#)

### R-devel

[HTML](#) | [PDF](#) |  
[EPUB](#)

# 작업영역 지정

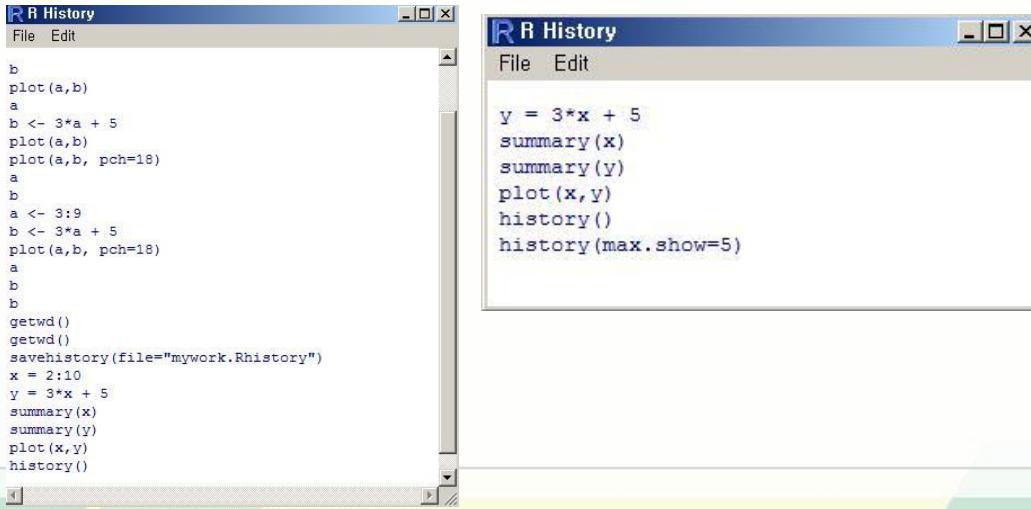
- **작업 영역(Working directory)** : R에서 데이터를 가져오고 저장하는 디폴트 폴더를 지정해두면 편리하게 작업할 수 있음. 이를 작업 영역(Working directory) 이라 함.

```
> getwd() # shows the working directory  
[1] "C:/Users/user/Documents"  
> setwd("c:/Rfolder/data") # change the working directory  
> getwd()  
[1] "c:/Rfolder/data"  
> setwd(choose.dir()) # select the working directory interactively  
> getwd()  
[1] "C:/Rfolder/data"
```

# R Command 저장

- R은 대화형으로 작업을 수행함.
- 따라서 분석 작업에 이용된 R 명령(Command)을 파일로 자동 저장해두면 다음에 따로 입력하지 않고 사용할 수 있음.

```
> savehistory(file="mywork.Rhistory")
> history() # left
> history(max.show=5) # right
```

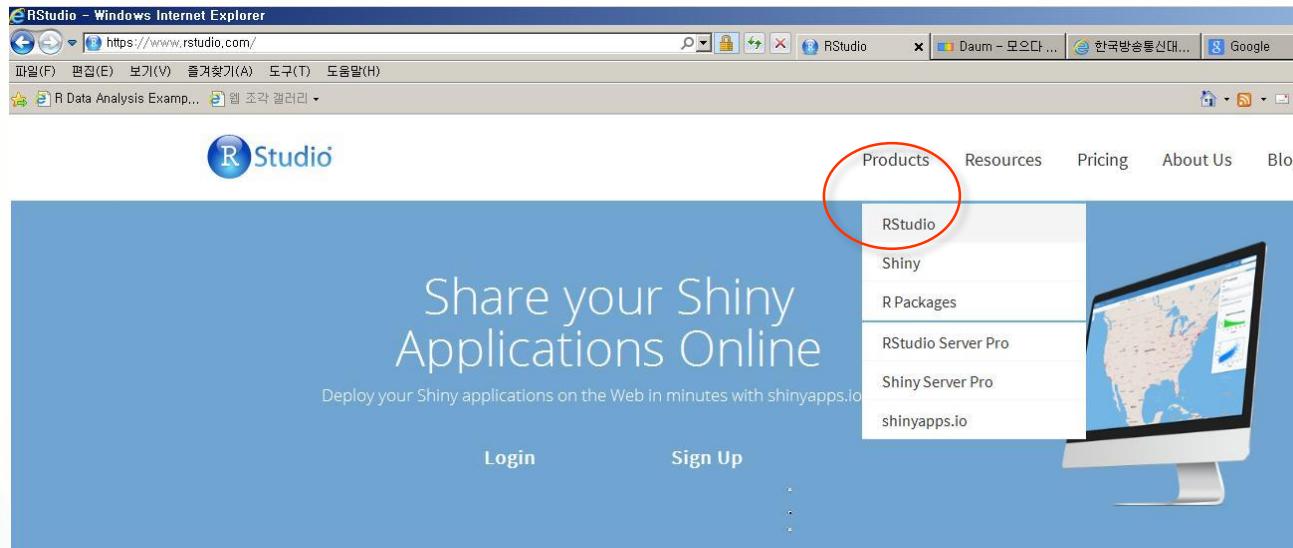


The image shows two side-by-side windows titled "R History". The window on the left displays a full history of commands entered, including variable assignments, function calls, and file operations. The window on the right shows a subset of these commands, specifically the last five entries, as indicated by the "# right" comment in the red box above. Both windows have standard Windows-style menus at the top.

```
b
plot(a,b)
a
b <- 3*a + 5
plot(a,b)
plot(a,b, pch=18)
a
b
a <- 3:9
b <- 3*a + 5
plot(a,b, pch=18)
a
b
b
getwd()
getwd()
savehistory(file="mywork.Rhistory")
x = 2:10
y = 3*x + 5
summary(x)
summary(y)
plot(x,y)
history()
history(max.show=5)
```

# R Studio 의 소개

- R Studio : 사용자가 친숙하게 R을 쉽게 사용할 수 있도록 개발된 R 통합환경 시스템
- 다운로드 : [www.rstudio.com](http://www.rstudio.com)



참고 : “<http://dss.princeton.edu/training/RStudio101.pdf>”  
“<http://www.rstudio.com>”

# R Studio 화면

RStudio  
File Edit Code View Plots Session Build Debug Tools Help  
Project: (None)

Console ~/ Go to file/function

```
t"
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for online help,
or 'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/.RData]

> x = c(1:10)
> y = 1.5*x + rnorm(10)
> plot(x,y)
> x+y
[1] 5.852384 5.493186 7.380824 11.459714
[5] 12.169048 13.478761 17.706424 20.379888
[9] 23.121829 25.681762
>
```

Environment History Import Dataset Clear List Global Environment values

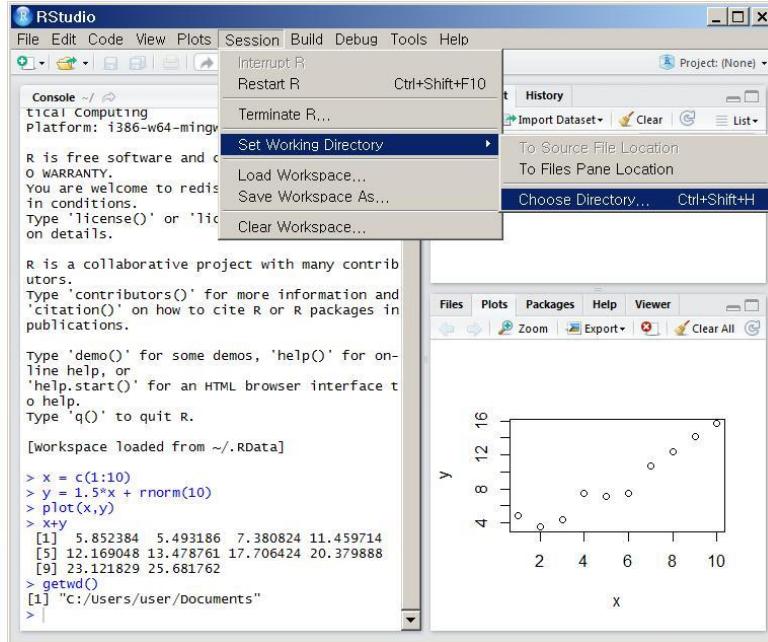
| x  | y     |
|----|-------|
| 1  | 4.85  |
| 2  | 3.49  |
| 3  | 4.38  |
| 4  | 5.12  |
| 5  | 6.05  |
| 6  | 7.02  |
| 7  | 8.05  |
| 8  | 9.12  |
| 9  | 10.25 |
| 10 | 11.45 |

Files Plots Packages Help Viewer Zoom Export Clear All

A scatter plot showing the relationship between variables x and y. The x-axis ranges from 2 to 10, and the y-axis ranges from 4 to 16. The data points show a clear positive linear trend, starting at approximately (2, 3.49) and ending at approximately (10, 11.45). The points are open circles.

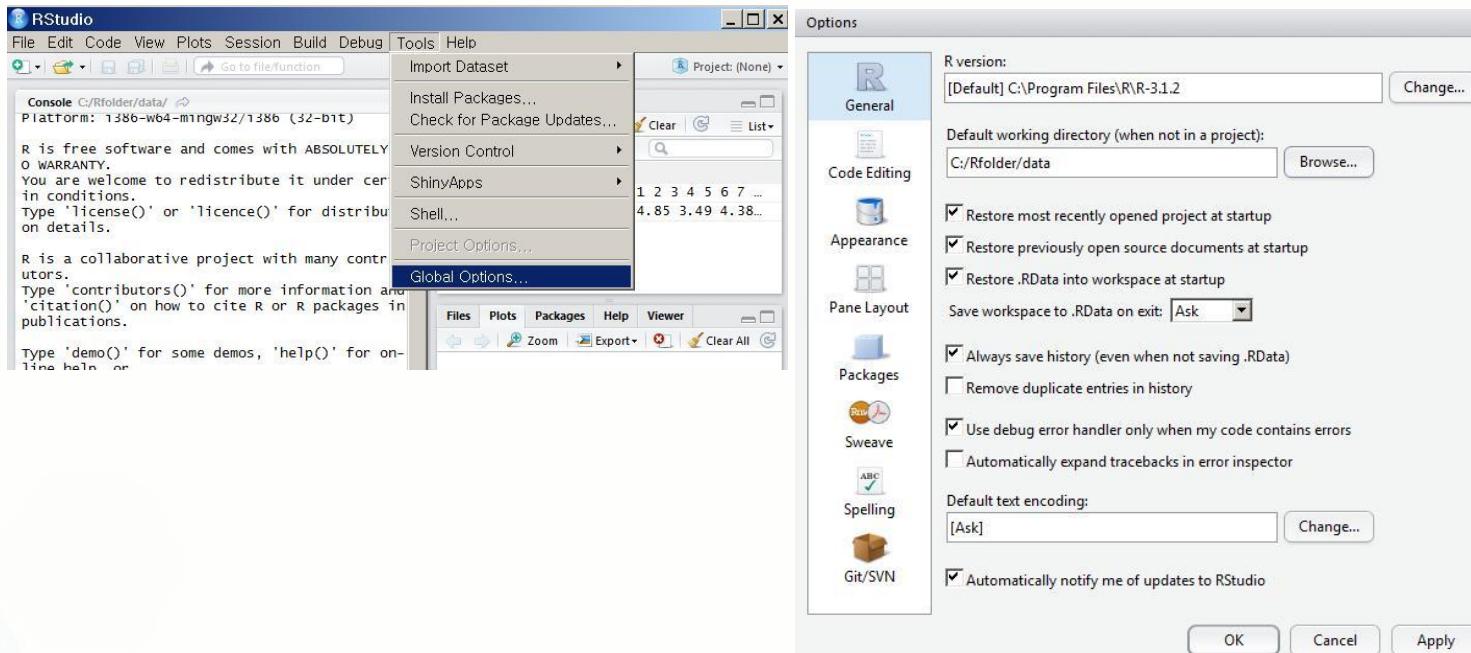
# R Studio : 작업영역 바꾸기

- 작업영역 보기 : `getwd()`
- 작업영역 바꾸기 : `setwd` 함수 이용 또는 "Session-Set Working Directory - Choose Directory"를 이용



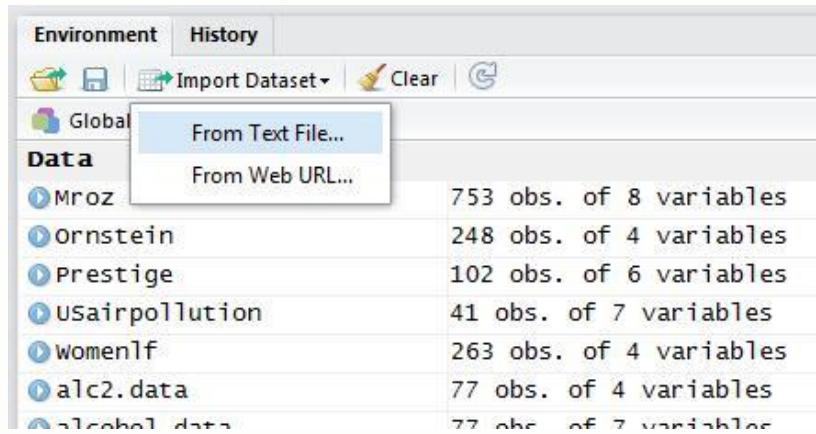
# R Studio : 작업영역 고정하기

- 작업영역 바꾸기 : "Tools—Global Options"를 클릭한 후 Options 창에서 지정



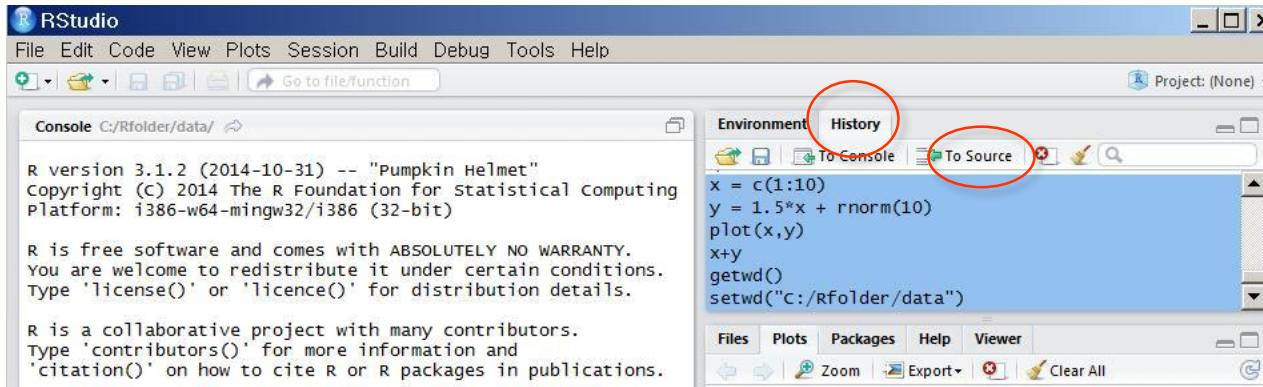
# R Studio : Import Dataset

- Environment 메뉴에서 Import Dataset :  
텍스트 파일이나 웹 URL에서 데이터를 가져오는 기능을 함.



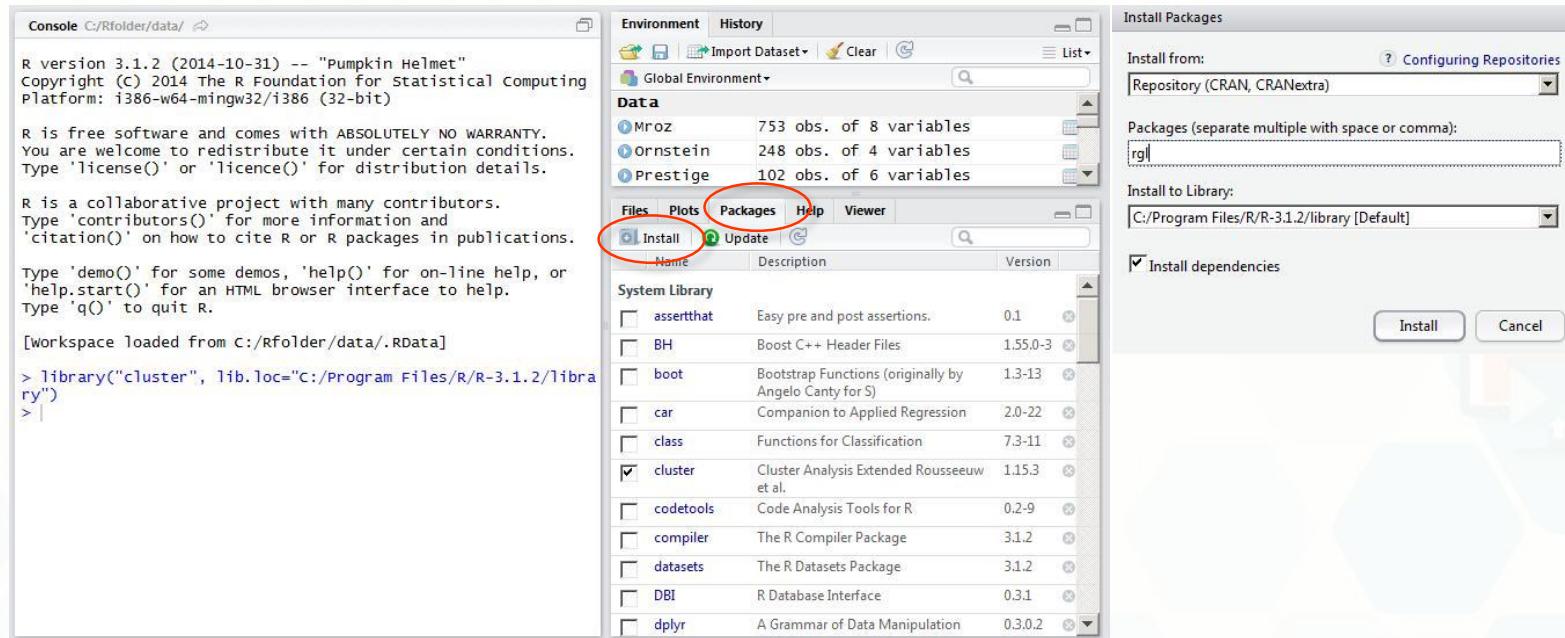
# R Studio : History

- Environment 메뉴에서 History 탭 :  
사용한 명령어를 실행하거나 저장하는 기능 (To Source)



# R Studio : Packages

- Packages 탭은 R Studio에 포함된 패키지를 보여줌. 패키지 목록을 체크하면 됨.
- 새로운 Package를 설치하기 위해서는 Install 탭을 누른 뒤 이름을 입력하여 설치.



2

## 데이터파일 읽기



# 설문지 자료

## 보험성향을 조사하기 위한 설문지

1. 성별구분은 ? (m : 남자, f : 여자)

2. 직업은 ?

(1) 근로자 (2) 사무직 (3) 전문가

3. 종교는 ?

(1) 기독교 (2) 불교 (3) 없음

4. 교육수준은 ?

(1) 무학 (2) 국졸 (3) 중졸 (4) 고졸 (5) 대졸 이상

5. 보험가입금액은 ? () 천만원

6. 월 평균 수입은 ? () 만원

| 일련번호<br>(id) | 성별<br>(sex) | 직업<br>(job) | 종교<br>(religion) | 교육수준<br>(edu) | 보험금액<br>(단위:천만원)<br>(amount) | 월수입<br>(단위:만원)<br>(salary) |
|--------------|-------------|-------------|------------------|---------------|------------------------------|----------------------------|
| 1            | m           | 1           | 1                | 3             | 7.0                          | 110                        |
| 2            | m           | 2           | 1                | 4             | 12.0                         | 135                        |
| 3            | f           | 2           | 3                | 5             | 8.5                          | 127                        |
| 4            | f           | 3           | 3                | 5             | 5.0                          | 150                        |
| 5            | m           | 1           | 3                | 3             | 4.5                          | 113                        |
| 6            | m           | 2           | 1                | 2             | 3.5                          | 95                         |
| 7            | m           | 3           | 2                | 4             | 4.0                          | 102                        |
| 8            | f           | 3           | 2                | 4             | 4.0                          | 122                        |
| 9            | f           | 2           | 3                | 4             | 4.5                          | 110                        |
| 10           | m           | 1           | 3                | 5             | 17.0                         | 200                        |
| 11           | f           | 1           | 1                | 3             | 22.0                         | (결측값)                      |
| 12           | m           | 2           | 1                | 2             | 5.5                          | 105                        |
| 13           | m           | 3           | 2                | 1             | 4.5                          | 130                        |
| 14           | m           | 3           | 2                | 5             | 7.0                          | 150                        |
| 15           | m           | (결측값)       | 3                | 4             | 6.0                          | 110                        |
| 16           | f           | 1           | 3                | (결측값)         | 7.0                          | 88                         |
| 17           | m           | 1           | 1                | 4             | 6.0                          | 138                        |
| 18           | f           | 2           | 1                | 5             | 5.0                          | 110                        |
| 19           | m           | 2           | 3                | 3             | 7.0                          | 85                         |
| 20           | m           | 3           | 3                | 4             | 9.5                          | 110                        |
| 21           | m           | 3           | 1                | 4             | 10.0                         | 95                         |
| 22           | m           | 3           | 2                | 3             | 12.0                         | 88                         |

| insurance - 메모장 |       |       |       |        |  |  |
|-----------------|-------|-------|-------|--------|--|--|
| 파일(F)           | 편집(E) | 서식(O) | 보기(V) | 도움말(H) |  |  |
| 1 m 1           | 1     | 3     | 7.0   | 110    |  |  |
| 2 m 2           | 1     | 4     | 12.0  | 135    |  |  |
| 3 f 2           | 3     | 5     | 8.5   | 127    |  |  |
| 4 f 3           | 3     | 5     | 5.0   | 150    |  |  |
| 5 m 1           | 3     | 3     | 4.5   | 113    |  |  |
| 6 m 2           | 1     | 2     | 3.5   | 95     |  |  |
| 7 m 3           | 2     | 4     | 4.0   | 102    |  |  |
| 8 f 3           | 2     | 4     | 4.0   | 122    |  |  |
| 9 f 2           | 3     | 4     | 4.5   | 110    |  |  |
| 10 m 1          | 3     | 5     | 17.0  | 200    |  |  |
| 11 f 1          | 1     | 3     | 22.0  | NA     |  |  |
| 12 m 2          | 1     | 2     | 5.5   | 105    |  |  |
| 13 m 3          | 2     | 1     | 4.5   | 130    |  |  |
| 14 m 3          | 2     | 5     | 7.0   | 150    |  |  |
| 15 m NA         | 3     | 4     | 6.0   | 110    |  |  |
| 16 f 1          | 3     | NA    | 7.0   | 88     |  |  |
| 17 m 1          | 1     | 4     | 6.0   | 138    |  |  |
| 18 f 2          | 1     | 5     | 5.0   | 110    |  |  |
| 19 m 2          | 3     | 3     | 7.0   | 85     |  |  |
| 20 m 3          | 3     | 4     | 9.5   | 110    |  |  |
| 21 m 3          | 1     | 4     | 10.0  | 95     |  |  |
| 22 m 3          | 2     | 3     | 12.0  | 88     |  |  |

# 텍스트파일 읽기/저장

```
> insurance.data = read.table("c:/Rfolder/data/insurance.txt", header=T)
> insurance.data[c(10:11),]
  id sex job religion edu amount salary
10 10  m  1        3   5    17.0  200
11 11  f  1        1   3    22.0   NA
> insurance2.data = read.table("c:/Rfolder/data/insurance2.txt", header=T, na.string="-9")
> insurance2.data[c(15:16),]
  id sex job religion edu amount salary
15 15  m NA       3   4     6.0   110
16 16  f  1       3  NA     7.0    88
> csv.data = read.csv("c:/Rfolder/data/csv.txt", header=T)  # 구분자가 콤마
> tab.data = read.table("c:/Rfolder/data/tab.txt", header=T, sep="\t") #구분자가 Tab
> write.table(tab.data, file="c:/Rfolder/data/test.txt") # 저장
```

# 고정형식 텍스트 파일 읽기

```
> fwf.data = read.fwf(file="c:/Rfolder/data/insurance3.txt",
  widths=c(2,2,3,3,3,6,6),
  col.names=c("id","sex","job","religion","edu","amount","salary"))
```

```
> fwf.data[fwf.data$job == -9, "job"] = NA
> fwf.data[fwf.data$edu == -9, "edu"] = NA
> fwf.data[fwf.data$salary == -9, "salary"] = NA
> head(fwf.data, n=3)
```

```
id sex job religion edu amount salary
1 1 m   1     1   3    7.0  110
2 2 m   2     1   4   12.0  135
3 3 f   2     3   5    8.5  127
```

```
> fwf2.data = read.fwf(file="c:/Rfolder/data/insurance3.txt",
  widths=c(2,-2,-3,3,3,6,6),
  col.names=c("id","religion","edu","amount","salary"))
```

```
> head(fwf2.data,n=3)
```

```
id religion edu amount salary
1 1           1   3    7.0  110
2 2           1   4   12.0  135
3 3           3   5    8.5  127
```

| 1  | m | 1  | 1 | 3  | 7.0  | 110 |
|----|---|----|---|----|------|-----|
| 2  | m | 2  | 1 | 4  | 12.0 | 135 |
| 3  | f | 2  | 3 | 5  | 8.5  | 127 |
| 4  | f | 3  | 3 | 5  | 5.0  | 150 |
| 5  | m | 1  | 3 | 3  | 4.5  | 113 |
| 6  | m | 2  | 1 | 2  | 3.5  | 95  |
| 7  | m | 3  | 2 | 4  | 4.0  | 102 |
| 8  | f | 3  | 2 | 4  | 4.0  | 122 |
| 9  | f | 2  | 3 | 4  | 4.5  | 110 |
| 10 | m | 1  | 3 | 5  | 17.0 | 200 |
| 11 | f | 1  | 1 | 3  | 22.0 | -9  |
| 12 | m | 2  | 1 | 2  | 5.5  | 105 |
| 13 | m | 3  | 2 | 1  | 4.5  | 130 |
| 14 | m | 3  | 2 | 5  | 7.0  | 150 |
| 15 | m | -9 | 3 | 4  | 6.0  | 110 |
| 16 | f | 1  | 3 | -9 | 7.0  | 88  |
| 17 | m | 1  | 1 | 4  | 6.0  | 138 |
| 18 | f | 2  | 1 | 5  | 5.0  | 110 |
| 19 | m | 2  | 3 | 3  | 7.0  | 85  |
| 20 | m | 3  | 3 | 4  | 9.5  | 110 |
| 21 | m | 3  | 1 | 4  | 10.0 | 95  |
| 22 | m | 3  | 2 | 3  | 12.0 | 88  |

# 엑셀파일 읽기

```
> install.packages("xlsx")
> library(xlsx)
> alcohol.data = read.xlsx("c:/Rfolder/data/alcohol.xlsx", 1)
> head(alcohol.data, n=2)

  TYPE MEOH ACET  BU1 MEPR ACAL LNPR01
1 typeA   3   15  0.2    9    9   5.86
2 typeA  23   13  0.8    9    2   6.67

> alc2.data = read.xlsx("c:/Rfolder/data/alcohol.xlsx", 1, colIndex=c(1,2,6:7))
> head(alc2.data, n=2)

  TYPE MEOH ACAL LNPR01
1 typeA   3    9   5.86
2 typeA  23    2   6.67
```

|    | A     | B     | C     | D    | E    | F    | G      |
|----|-------|-------|-------|------|------|------|--------|
| 1  | TYPE  | MEOH  | ACET  | BU1  | MEPR | ACAL | LNPRO1 |
| 2  | typeA | 3.0   | 15.0  | 0.2  | 9.0  | 9.0  | 5.86   |
| 3  | typeA | 23.0  | 13.0  | 0.8  | 9.0  | 2.0  | 6.67   |
| 4  | typeA | 65.0  | 96.0  | 0.4  | 9.0  | 4.0  | 5.31   |
| 5  | typeA | 279.0 | 66.0  | 0.9  | 36.0 | 4.8  | 5.45   |
| 6  | typeA | 292.0 | 210.0 | 1.1  | 34.0 | 8.0  | 4.08   |
| 7  | typeA | 371.0 | 414.0 | 1.2  | 39.0 | 9.0  | 6.22   |
| 8  | typeA | 393.0 | 287.0 | 1.8  | 41.0 | 9.7  | 6.47   |
| 9  | typeA | 394.0 | 217.0 | 1.0  | 31.0 | 8.6  | 6.31   |
| 10 | typeA | 418.0 | 62.0  | 0.8  | 24.0 | 7.0  | 5.83   |
| 11 | typeA | 426.0 | 204.0 | 1.3  | 37.0 | 8.6  | 6.07   |
| 12 | typeA | 465.0 | 267.0 | 1.5  | 39.0 | 17.0 | 8.01   |
| 13 | typeA | 469.0 | 226.0 | 1.6  | 35.0 | 8.0  | 6.21   |
| 14 | typeA | 475.0 | 172.0 | 1.9  | 33.0 | 14.0 | 6.30   |
| 15 | typeA | 498.0 | 343.0 | 2.3  | 42.0 | 21.0 | 6.59   |
| 16 | typeA | 507.0 | 397.0 | 3.3  | 30.0 | 21.0 | 5.98   |
| 17 | typeA | 523.0 | 367.0 | 2.6  | 45.0 | 25.0 | 6.67   |
| 18 | typeA | 583.0 | 226.0 | 2.3  | 46.0 | 11.0 | 6.49   |
| 19 | typeA | 632.5 | 342.3 | 2.2  | 38.2 | 13.3 | 7.64   |
| 20 | typeB | 595.0 | 410.0 | 16.0 | 24.0 | 11.4 | 5.09   |
| 21 | typeB | 630.0 | 243.0 | 13.0 | 13.0 | 13.0 | 5.87   |
| 22 | typeB | 707.0 | 131.0 | 15.0 | 28.0 | 9.0  | 4.89   |

※ 자료출처 : Tanagra 사이트 <http://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html>

# 3

# 데이터베이스 자료 읽기



# 데이터베이스 자료 읽기 : DBMS 연결

- Access 파일을 ODBC(Open Database Connectivity)와 연결하여 R에서 가져오는 방법

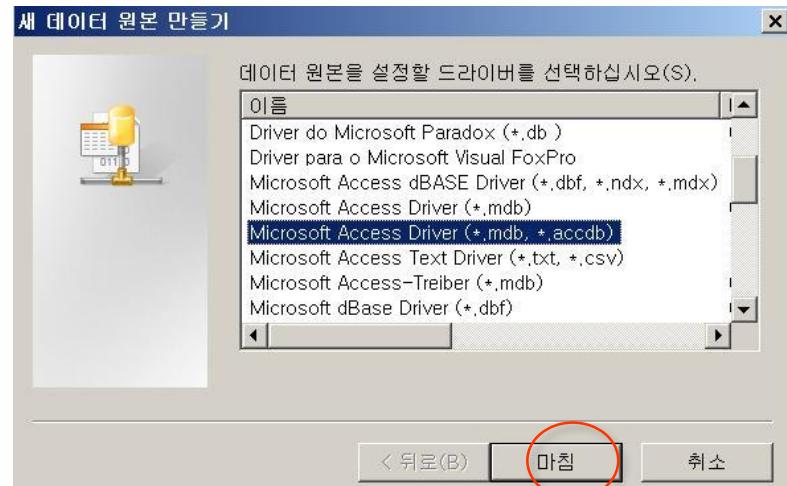
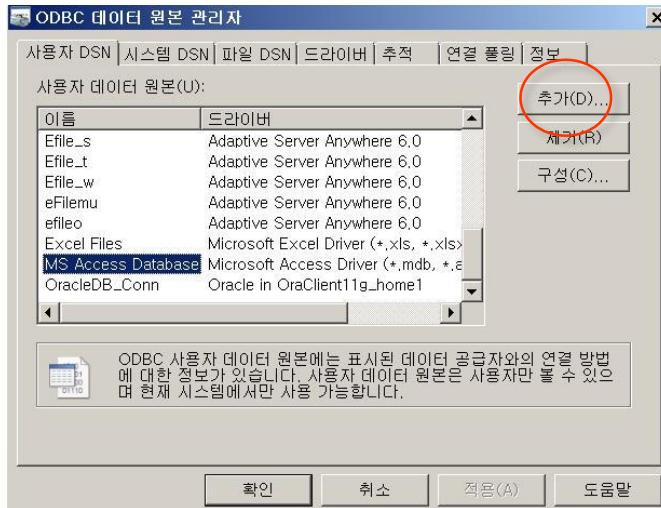
| ID1 | name | korean | english |
|-----|------|--------|---------|
| 1   | 정연도  | 84     | 75      |
| 2   | 한숙희  | 64     | 45      |
| 3   | 이명옥  | 69     | 57      |
| 4   | 박정용  | 84     | 57      |
| 5   | 안승제  | 76     | 37      |
| 6   | 박선미  | 92     | 67      |
| 7   | 이민영  | 94     | 85      |
| 8   | 이원종  | 62     | 85      |
| 9   | 김수정  | 35     | 38      |
| 10  | 김한준  | 75     | 36      |
| 11  | 심경택  | 42     | 29      |
| 12  | 최선욱  | 34     | 31      |
| 13  | 배희열  | 84     | 54      |
| 14  | 장윤정  | 63     | 40      |
| 15  | 이진홍  | 42     | 55      |

| ID1 | name | math | social | physics |
|-----|------|------|--------|---------|
| 1   | 정연도  | 91   | 84     | 94      |
| 2   | 한숙희  | 55   | 70     | 60      |
| 3   | 이명옥  | 61   | 67     | 76      |
| 4   | 박정용  | 70   | 55     | 70      |
| 5   | 안승제  | 65   | 70     | 55      |
| 6   | 박선미  | 80   | 85     | 80      |
| 7   | 이민영  | 75   | 86     | 85      |
| 8   | 이원종  | 75   | 86     | 85      |
| 9   | 김수정  | 80   | 85     | 75      |
| 10  | 김한준  | 44   | 55     | 65      |
| 11  | 심경택  | 64   | 70     | 80      |
| 12  | 최선욱  | 35   | 66     | 70      |
| 13  | 배희열  | 68   | 85     | 85      |
| 14  | 장윤정  | 60   | 75     | 65      |
| 15  | 이진홍  | 70   | 75     | 70      |

<Access 파일의 midterm 및 endterm 테이블>

# 1) DSN(Data Source Name) 설정

- ① [시작]–[제어판]–[관리도구]를 선택한 후, ‘데이터 원본(ODBC)’을 더블클릭
- ② ‘ODBC 데이터 원본 관리자’ 대화상자에서 ‘MS Access...’를 선택하고, ‘추가’버튼



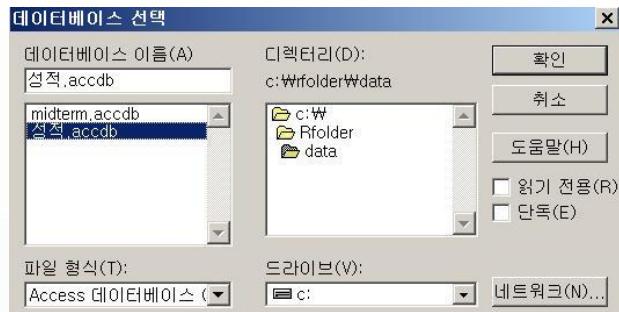
- ③ ‘새 데이터 원본 만들기’ 대화상자에서 ‘Microsoft Access Driver(\*.mdb,...)’를 선택하고 ‘마침’ 버튼

# 1) DSN(Data Source Name) 설정 : 계속

## ④ ‘ODBC Microsoft Access 설정’ 대화상자에서 ‘선택’버튼



⑤ Access 파일을 선택하고 확인. ‘데이터 원본 이름’에 DSN으로 사용할 이름(예를 들어, score) 및 설명을 입력하고 ‘확인’버튼을 누르면 설정이 끝남.



## 2) 데이터베이스 자료 가져오기

- RODBC 연결하기

```
> install.packages("RODBC")
> library(RODBC)
> channel <- odbcConnect("score")
```

1) sqlFetch : ODBC 데이터베이스의 테이블을 data frame 으로 가져오는 기능

```
> sqlFetch(channel, "midterm")
```

| ID1 | name | korean | english |
|-----|------|--------|---------|
| 1   | 1    | 정연도    | 84      |
| ... | ...  | ...    | ...     |
| 15  | 15   | 이진홍    | 42      |

|    |    |     |    |    |
|----|----|-----|----|----|
| 15 | 15 | 이진홍 | 42 | 55 |
|----|----|-----|----|----|

```
> sqlFetch(channel, "endterm")
```

| ID1 | name | math | social | physics |
|-----|------|------|--------|---------|
| 1   | 1    | 정연도  | 91     | 84      |
| ... | ...  | ...  | ...    | ...     |
| 14  | 14   | 장윤정  | 60     | 75      |

|    |    |     |    |    |    |
|----|----|-----|----|----|----|
| 14 | 14 | 장윤정 | 60 | 75 | 65 |
|----|----|-----|----|----|----|

## 2) 데이터베이스 자료 가져오기

2) sqlQuery : ODBC 데이터베이스에 SQL 문을 수행

```
> sqlQuery(channel, "select * from endterm where ID1 > 10")
```

ID1 name math social physics

1 11 심경택 64 70 80

2 12 최선욱 35 66 70

3 13 배희열 68 85 85

4 14 장윤정 60 75 65

5 15 이진홍 70 75 70

```
> sqlQuery(channel, "select * from endterm where math >= 80 ")
```

ID1 name math social physics

1 1 정연도 91 84 94

2 6 박선미 80 85 80

3 9 김수정 80 85 75

## 2) 데이터베이스 자료 가져오기

※ 두 테이블을 합친 뒤, 기술통계량을 구하는 예

```
> midterm <- sqlQuery(channel, "select * from midterm")
> endterm <- sqlQuery(channel, "select * from endterm")
> score <- cbind(midterm, endterm[,c(3:5)])
> summary(score[3:6])
```

|         | korean | english | math  | social |
|---------|--------|---------|-------|--------|
| Min.    | :34.00 | :29.00  | :35.0 | :55.00 |
| 1st Qu. | :52.00 | :37.50  | :60.5 | :68.50 |
| Median  | :69.00 | :54.00  | :68.0 | :75.00 |
| Mean    | :66.67 | :52.73  | :66.2 | :74.27 |
| 3rd Qu. | :84.00 | :62.00  | :75.0 | :85.00 |
| Max.    | :94.00 | :85.00  | :91.0 | :86.00 |

# 4

# 통계패키지 자료 읽기

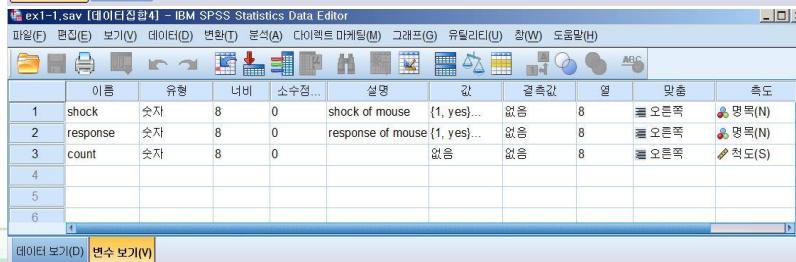


# 통계패키지 데이터 읽기

## ▪ SPSS 데이터 읽기 예

쥐의 자극에 대한 반응결과

|      | 반응 유 | 반응 무 | 계   |
|------|------|------|-----|
| 자극 유 | 25   | 19   | 44  |
| 자극 무 | 31   | 141  | 172 |
| 계    | 56   | 160  | 216 |



```
> install.packages("foreign")
> library(foreign)
> ex1 = read.spss("c:/Rfolder/data/ex1-1.sav",
+                  to.data.frame=T, use.value.label=T)
```

> **ex1**

shock response count

|   |     |     |     |
|---|-----|-----|-----|
| 1 | yes | yes | 25  |
| 2 | yes | no  | 19  |
| 3 | no  | yes | 31  |
| 4 | no  | no  | 141 |

```
> mouse.data = ex1[rep(1:nrow(ex1), ex1$count),]
```

> head(mouse.data)

shock response count

|     |     |     |    |
|-----|-----|-----|----|
| 1   | yes | yes | 25 |
| 1.1 | yes | yes | 25 |
| 1.2 | yes | yes | 25 |
| 1.3 | yes | yes | 25 |
| 1.4 | yes | yes | 25 |
| 1.5 | yes | yes | 25 |

# 통계패키지 데이터 읽기

## ■ SPSS 데이터 읽기 예

쥐의 자극에 대한 반응결과

|      | 반응 유 | 반응 무 | 계   |
|------|------|------|-----|
| 자극 유 | 25   | 19   | 44  |
| 자극 무 | 31   | 141  | 172 |
| 계    | 56   | 160  | 216 |

The screenshot shows the SPSS Data Editor window with the file 'ex1-1.sav' open. The data view displays a 2x2 contingency table with four cells containing values 25, 19, 31, and 141. The columns are labeled 'shock' and 'response'. The rows are labeled '1' and '2'. The toolbar at the top includes various data analysis and visualization tools.

The screenshot shows the SPSS Data Editor window with the file 'ex1-1.sav' open. The variable view displays three variables: 'shock', 'response', and 'count'. The 'shock' variable is a categorical variable with values 1 and 2. The 'response' variable is also categorical with values 1 and 2. The 'count' variable is a numerical variable. The toolbar at the top includes various data analysis and visualization tools.

> **attach(mouse.data)**

> **mouse.table = table(shock, response)**

> **mouse.table**

response

shock yes no

yes 25 19

no 31 141

> **summary(mouse.table)**

Number of cases in table: 216

Number of factors: 2

Test for independence of all factors:

Chisq = 27.458, df = 1, p-value = 1.605e-07

# 통계패키지 데이터 읽기

- SPSS 읽기 : `read.spss()`
  - Stata 읽기 : `read.dta()`
  - SAS XPORT file : `read.xport()`
  - Systat 데이터 : `read.systat()`
- ※ 참고 : "foreign" 패키지  
([http://cran.r-project.org/web/packages/foreign/foreign.pdf](http://cran.r-project.org/web/packages/foreign/))

# 5

# RData 저장 및 가져오기



# RData 저장 및 가져오기

- R 시스템파일로 저장/가져오기 함수 : save() , load()

```
> save(ex1, file="c:/Rfolder/data/ex1.RData")
> rm(ex1)
> load("c:/Rfolder/data/ex1.RData")
> ex1
```

```
shock response count
1   yes   yes    25
2   yes   no     19
3   no    yes    31
4   no    no    141
```

```
> load(file=file.choose())
```

