

연구조사방법론

2021.03.11



🥏 오늘의 강의 내용

□ 제2강 통계분석 프로그램의 종류와 선택



통계분석 프로그램(통계패키지)

- 복잡한 통계분석을 용이하게 할 수 있도록 도와주는 컴퓨터 프로그램
- □ SPSS와 SAS가 대표적인 통계 패키지임
 - SPSS (Statistical Product and Service Solution)
 - □ 사용이 편리함-마우스 클릭으로 사용
 - □ 대부분의 통계분석이 가능
 - □ 매우 특별한 통계분석은 다루지 못함
 - SAS (Statistical Analysis System)
 - □ 가장 강력한 통계패키지
 - □ 난이도가 높음-프로그램 언어를 사용
 - □ 가격이 높음

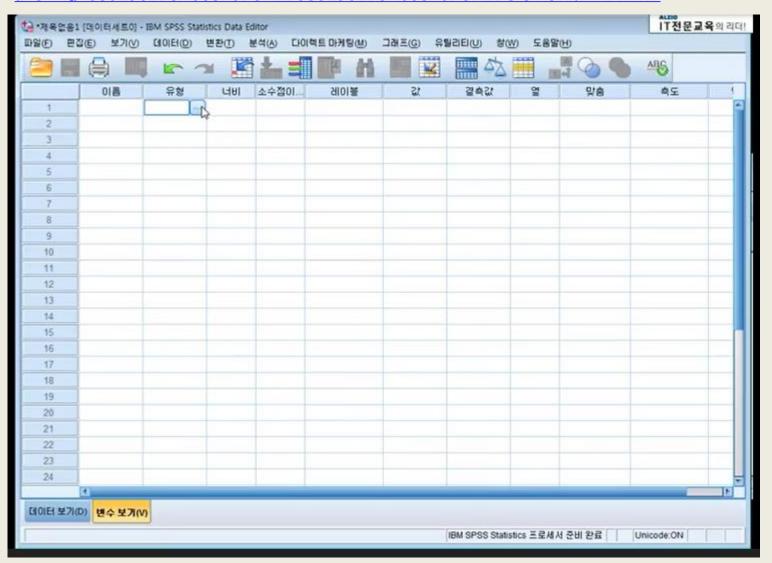


SPSS로 가능한 통계분석

모듈		주요 통계분석			
SPSS Statistics Professional Bundle					
Conjoint	Orthoplan, 직교계획 컨조인트 분석을 위한 실험카드 작성 컨조인트 분석				
Exact Tests	독립/연관된 일표본, 이표본, K-표본 검정 적합도 검정	R*C Contingency Table의 독립성 검정 Measures of Association의 검정			
Complex Samples	층화 추출 군집 추출 다단계 추출 Complex Samples Descriptives Complex Sample Tabulate	Complex Samples General Linear Models Complex Ordinals Selection Complex Samples Logistic Regression Complex Samples Cox Regression			
Neural Networks	다중 퍼셉트론 방사형 기저함수				
Direct Marketing	RFM(Recency, Frequency, Monetary) 분석 고객 및 접촉 정보의 세분화 고객 및 접촉 정보의 프로파일 분석 구매 가능성이 높은 지역(Postal Code) 식별	고객 성향 점수 분석 캠페인 비교 테스트(통제집단 패키지 테스트) 응답 가능성이 높은 고객 식별 응답 가능성 점수화			
Bootstrapping	기술통계 프로시저 - 기술통계 - 빈도분석 - 데이터탐색 - 집단별 평균분석 - 교차분석 - 일표본/독립표본 T-검정 - 상관분석의 이변량상관계수 - 상관분석의 편상관계수	모델링 프로시저 - 일원배치 분산분석 - 일변량 일반선형모형 - 일반화 선형모형 - 회귀분석 - 판별분석 - 로지스틱 회귀분석 - 이분형/다항 로지스틱 회귀분석 - 순서 회귀분석 - 선형혼합모형 - 선형혼합모형 - Cox 회귀모형			
AMOS	구조방정식 모형	53. 31120			

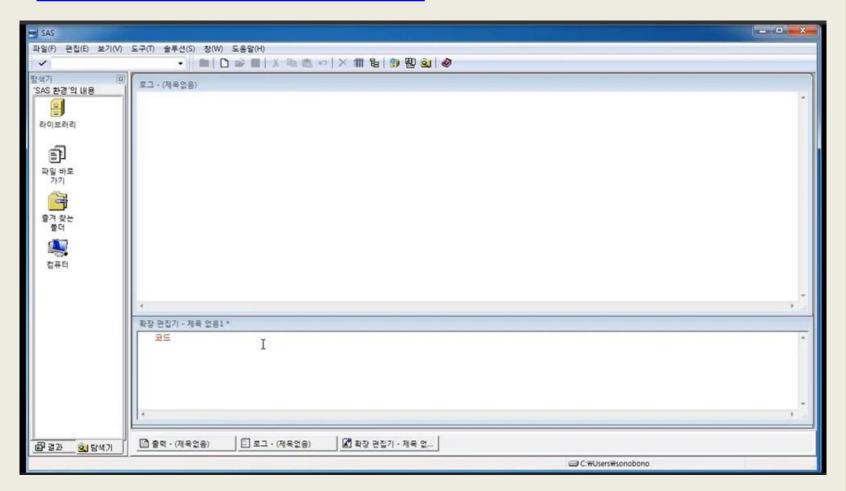


https://www.bing.com/videos/search?q=spss+%ec%82%ac%ec%9a%a9%eb%b2%95&&view=detail&mid=383E962B5A035F0D9AEB383E962B5A035F0D9AEB&&FORM=VDRVRV



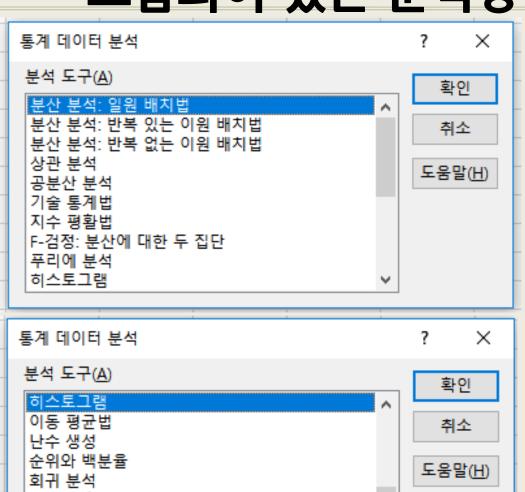


https://www.bing.com/videos/search?q=sas+%ec%82%ac%ec%9a%a9 %eb%b2%95&&view=detail&mid=24F81E762AFC0627040724F81E76 2AFC06270407&&FORM=VRDGAR





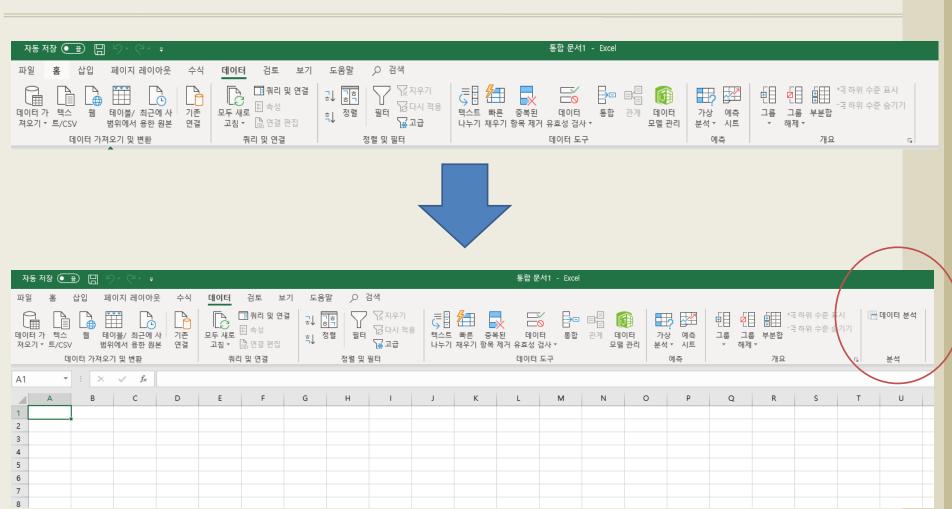
엑셀 데이터 분석에 포함되어 있는 분석방법



- □ 사용법이 익숙함
- □ 자료 정리와 간단한 통 계분석은 가능
- □ 복잡한 통계분석은 엑 셀로 분석 불가



데이터 분석도구 설치 후





Excel 분석도구 설치

→ 파일 → 아래와 같은 창에서 옵션 선택



정보

새로 만들기

열기

저장

다른 이름으로 저장

사용 기록

인쇄

공유

내보내기

게시

닫기

계정

피드백

옵션

정보



통합 문서 보호

이 통합 문서에서 다른 사용자들이 변경할 수 있는 내용을 제어합니다.



통합 문서 검사

이 파일을 게시하기 전에 다음 내용이 포함되어 있다는 점에 주의하세요. ■ 만든 이 이름, 절대 경로



통합 문서 관리 ▼ 통합 문서 관리

耐 저장되지 않은 변경 내용이 없습니다.



브라우저 보기 옵션

이 통합 문서를 웹에서 볼 때 사용자에게 표시할 항목을 선택합니다.

속성 ~

크기 아직 저장되지 않음

통합 문서1 - Excel

제목 제목 추가 태그 태그 추가 범주 범주 추가

관련 날짜

마지막으로 수정한 날짜

만든 날짜 오늘 오후 3:54

마지막으로 인쇄한 날짜

관련 사용자

만든 이



만든 이 추가

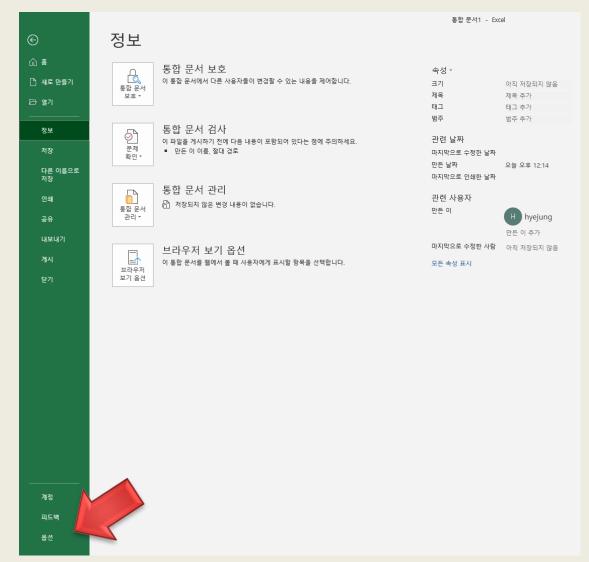
마지막으로 수정한 사람 이직 저장되지 않음

모든 속성 표시



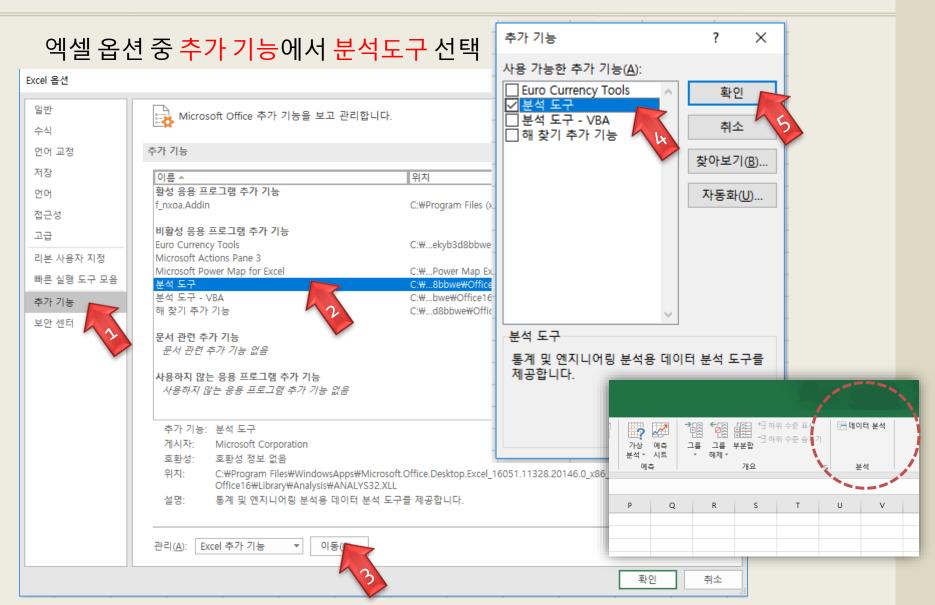
Excel 분석도구 설치 - office365

→ 파일 → 아래와 같은 창에서 옵션 선택





Excel 분석도구 설치(2)





R-language

□ 역사

- R-언어는 S-언어에서 시작됨. S-언어는 S-PLUS 소프트웨어로 개발 되어 판매되고 있음
- R-언어는 S-언어의 독립적이고 개방된 무료 언어
- 통계처리를 목적으로 개발되기는 하였으나 현재는 데이터분석, 그래픽, 머신러닝, 데이터마이닝, 자연어처리 등 다양한 분야에서 응용되고 있음

□ 특징

- □ 최대 강점은 무료이면서 대부분의 통계분석 방법이 공개되어 있음
- □ 인터프리터 언어로 실행을 위한 컴파일 과정이 필요 없고 코드를 직접 실 행
- □ 뛰어난 그래픽 처리 기능을 가지고 있음
- □ 타 프로그램(텍스트 파일, 데이터페이스, SPSS, SAS, XML 등)으로부터 데이터를 쉽게 불러들일 수 있고, 내보낼 수도 있음
- 웹페이지, 소셜미디어 사이트, 다양한 온라인서비스에 직접 접근하여 데 이터를 가져올 수 있음



https://cran.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
- · Download R for (Mac) OS X
- · Download R for Windows

R is part of many Linux distraction to the link above.

s, you should check with your Linux package management system in addition

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 (2020-02-29, Holding the Windsock) R-3.6.3.tar.gz, read what's new in the latest version.
- Sources of <u>R alpha and beta releases</u> (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are <u>available here</u>. Please read about <u>new</u> features and bug fixes before filing corresponding feature requests or bug reports.
- Source code of older versions of R is available here.
- Contributed extension <u>packages</u>

Questions About R

 If you have questions about R like how to download and install the software, or what the license terms are, please read our <u>answers to frequently asked questions</u> before you send an email.



R 프로그램 설치(2)



CRAN Mirrors What's new? Task Views Search

About R R Homepage The R Journal

Software R Sources

R Binaries

<u>Packages</u>

<u>Other</u>

Documentation Manuals

FAQs

Contributed

R for Windows

Subdirectories:

<u>base</u> Binaries for base distribution. This is what you want to <u>install R for the first time</u>.

Binaries of contributed CRAN packages (for R >= 2.13.x; managed by Uwe Lig here is

contrib also information on third party software available for CRAN Windows services

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 프로그램 설치(3)



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

R-3.6.3 for Windows (32/64 bit)

Download R 3.6.3 for Windows (83 megabytes, 32/64 bit)

<u>Installation and other instructions</u> <u>New features in this version</u>



If you want to double-check that the package you have downloaded matches the package distributed by CRAN, you can compare the <u>md5sum</u> of the .exe to the <u>fingerprint</u> on the master server. You will need a version of md5sum for windows: both <u>graphical</u> and <u>command line versions</u> 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 RFAQ for general information about R and the R Windows FAQ for Windows-specific information.

Other builds

- Patches to this release are incorporated in the r-patched snapshot build.
- A build of the development version (which will eventually become the next major release of R) is available in the r-devel snapshot build.
- Previous releases

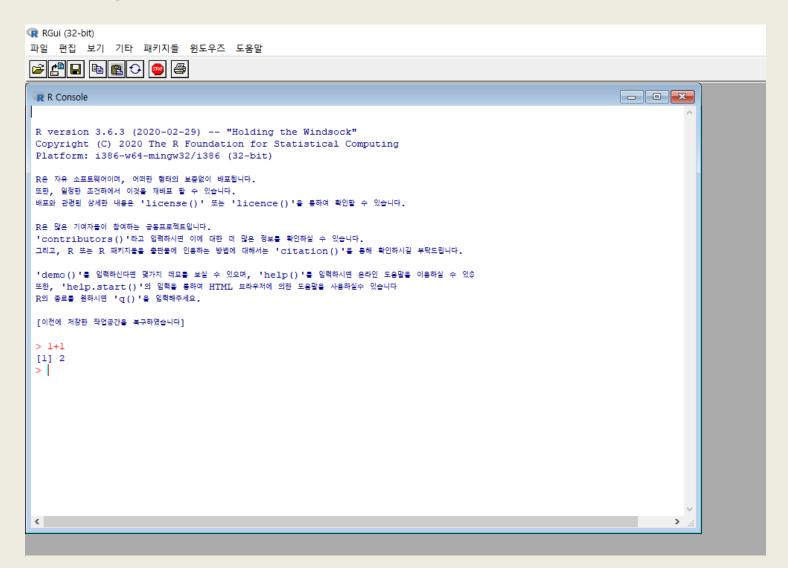
Note to webmasters: A stable link which will redirect to the current Windows binary release is CRAN MIRROR>/bin/windows/base/release.htm.

Last change: 2020-02-29



R 프로그램 설치(4)

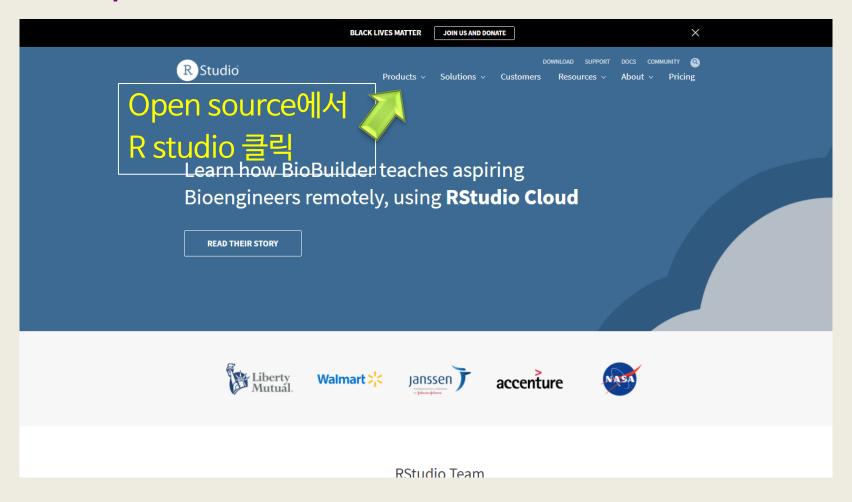
□ 설치된 모습





R Studio 설치(1) R을좀더효율적으로 사용할 수 있도록

https://www.rstudio.com/







Products • Resources • Pricing About • Blogs •

RStudio

Take control of your R code

RStudio is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management. Click here to see more RStudio features.

RStudio is available in **open source** and **commercial** editions and runs on the desktop (Windows, Mac, and Linux) or in a browser connected to RStudio Server or RStudio Server Pro (Debian/Ubuntu, Red Hat/CentOS, and SUSE Linux).

There are two versions of RStudio:





RStudio Server

Centralize access and computation



R Studio Desktop

	Open Source Edition	RStudio Desktop Pro
Overview	 Access RStudio locally Syntax highlighting, code completion, and smart indentation Execute R code directly from the source editor Quickly jump to function definitions Easily manage multiple working directories using projects Integrated R help and documentation Interactive debugger to diagnose and fix errors quickly Extensive package development tools 	All of the features of open source; plus: • A commercial license for organizations not able to use AGPL software • Access to priority support • RStudio Professional Drivers
Support	Community forums only	 Priority Email Support 8 hour response during business hours (ET)
License	AGPL v3	RStudio License Agreement
Pricing	Free	\$995/year
	DOWNLOAD RSTUDIO DESKTOP	DOWNLOAD FREE RSTUDIO DESKTOP PRO TRIAL Purchase Contact Sales



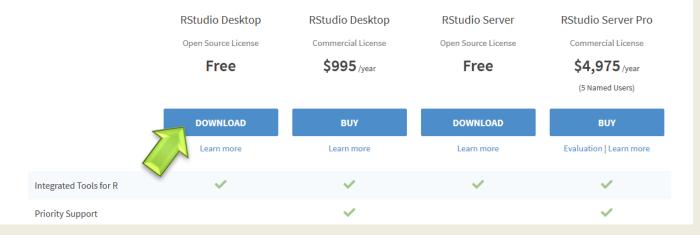
Download RStudio

Choose Your Version

RStudio is a set of integrated tools designed to help you be more productive with R. It includes a console, syntax-highlighting editor that supports direct code execution, and a variety of robust tools for plotting, viewing history, debugging and managing your workspace.

LEARN MORE ABOUT RSTUDIO FEATURES







All Installers

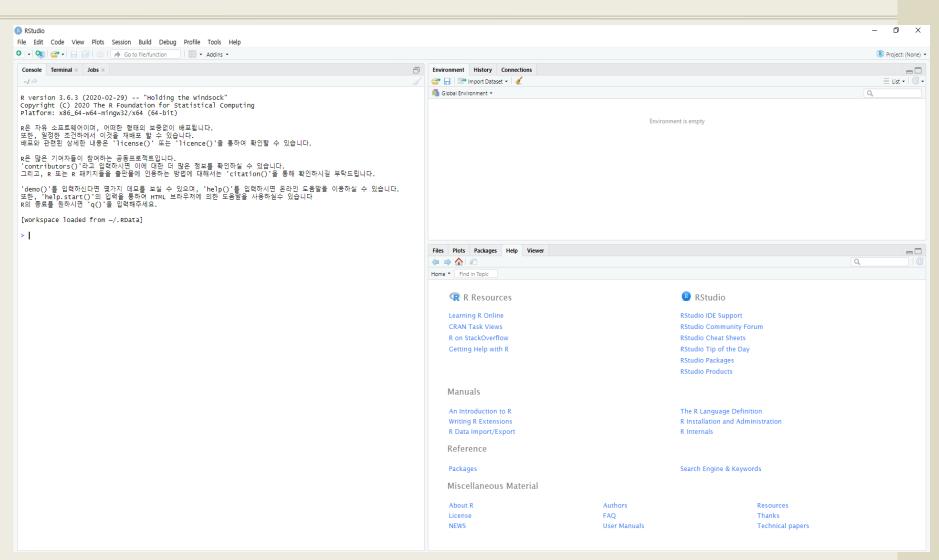
Linux users may need to import RStudio's public code-signing key prior to installation, depending on the operating system's security policy.

RStudio requires a 64-bit operating system. If you are on a 32 bit system, you can use an older version of RStudio.

os	Download	Size	SHA-256
Windows 10/8/7	₹ RStudio-1.4.1106.exe	155.97 MB	d2ff8453
macOS 10.13+	★ RStudio-1.4.1106.din	153.35 MB	c64d2cda
Ubuntu 16	♣ rstudio-1.4.1106-amd64.deb	118.45 MB	1fc82387
Ubuntu 18/Debian 10	♣ rstudio-1.4.1106-amd64.deb	121.07 MB	3b5d3835
Fedora 19/Red Hat 7	≛ rstudio-1.4.1106-x86_64.rpm	138.18 MB	a9e6ddc4
Fedora 28/Red Hat 8	≛ rstudio-1.4.1106-x86_64.rpm	138.16 MB	35e57c1c
Debian 9		121.33 MB	c7c9dd68
OpenSUSE 15	≛ rstudio-1.4.1106-x86_64.rpm	123.57 MB	3539d9c3

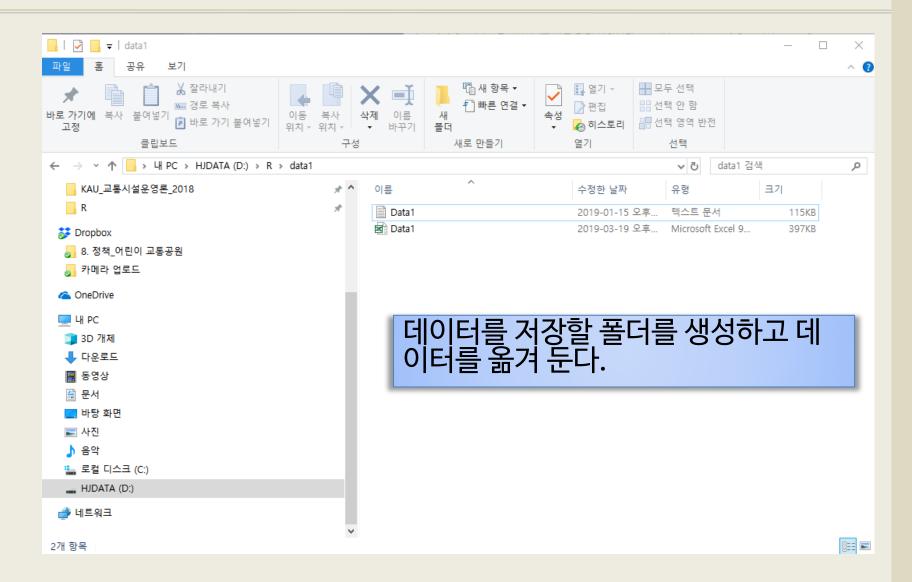


R studio 첫 화면



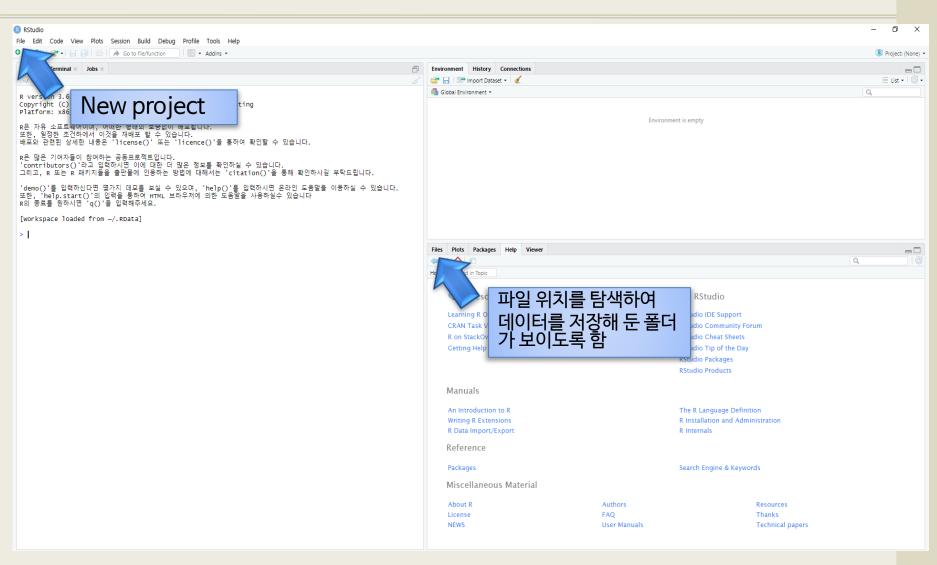


R 시작하기-데이터 준비



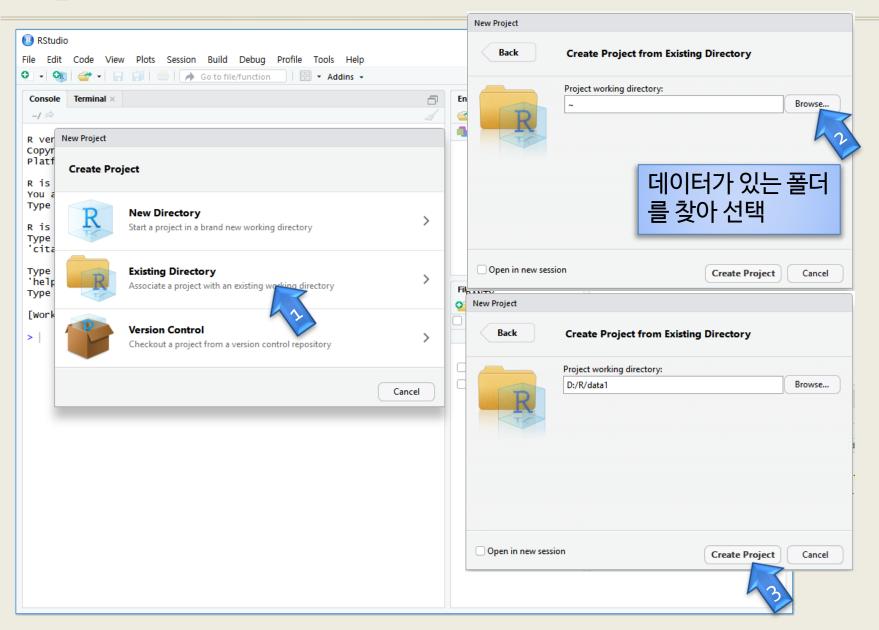


R 시작하기-프로젝트 생성



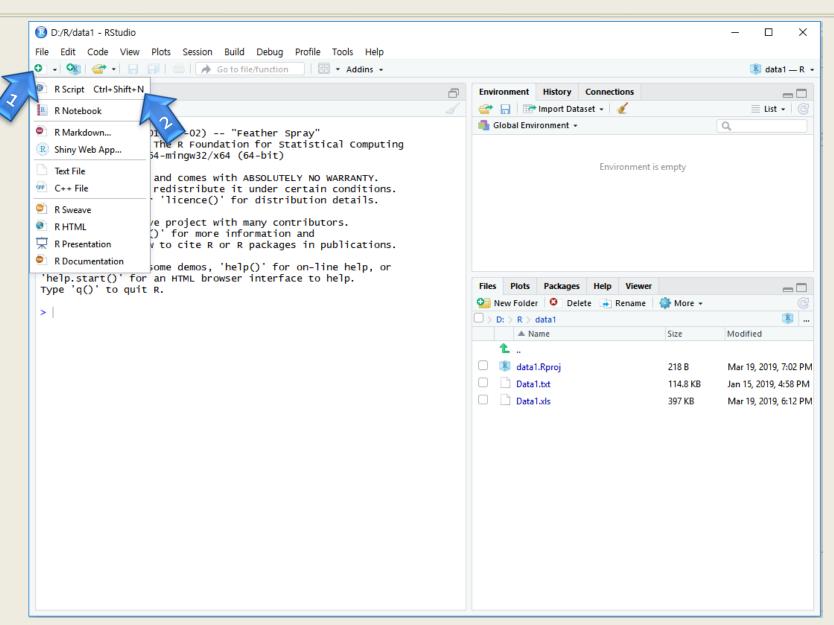


R 시작하기-프로젝트 생성 절차



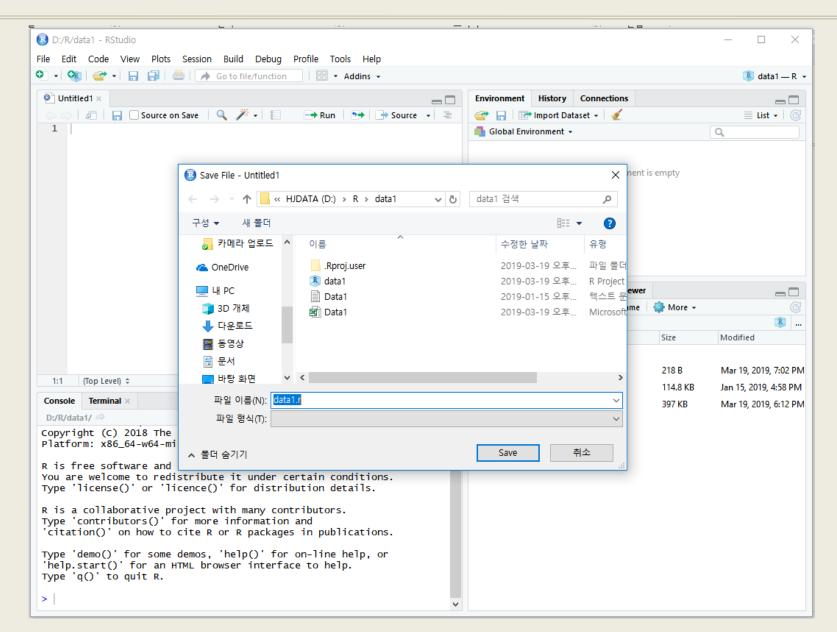


R 시작하기-R script 생성



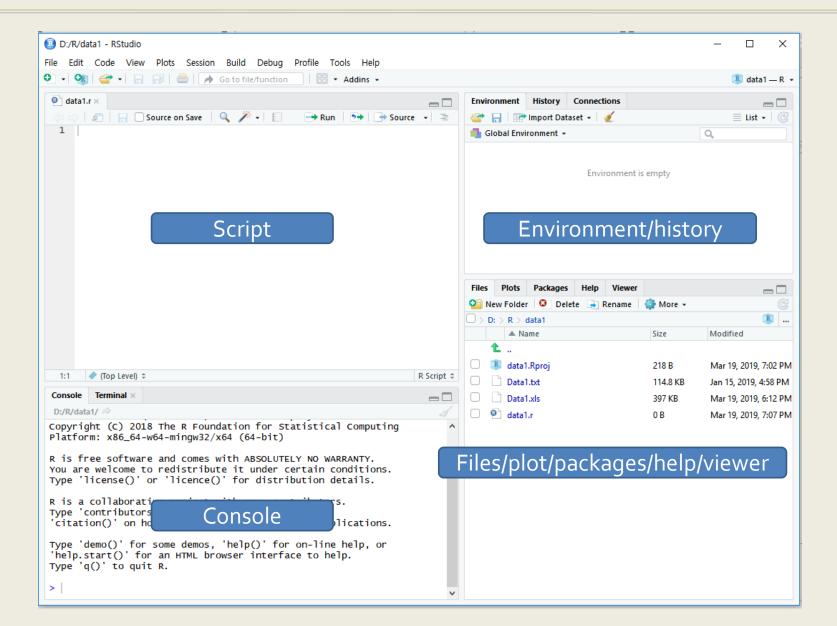


R 시작하기-R script 저장



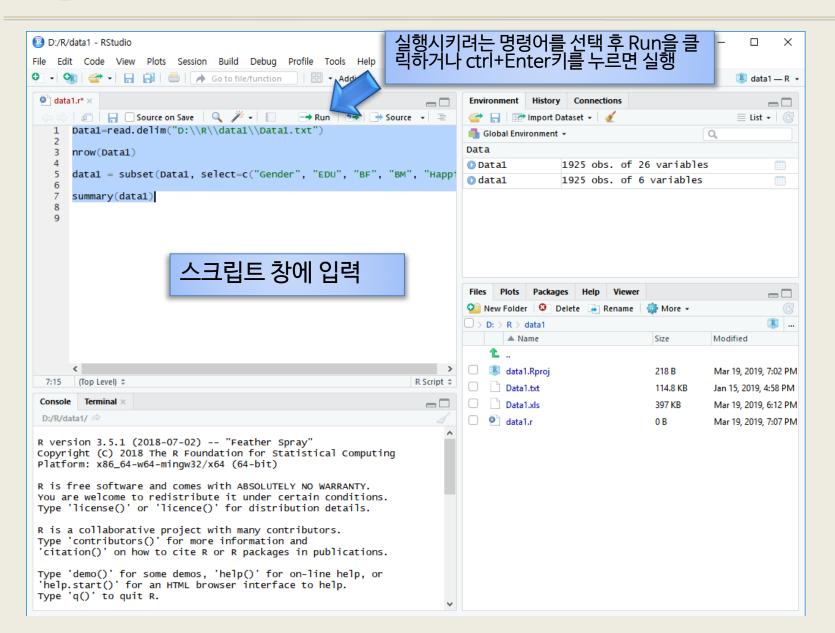


R 사용 준비 완료된 모습



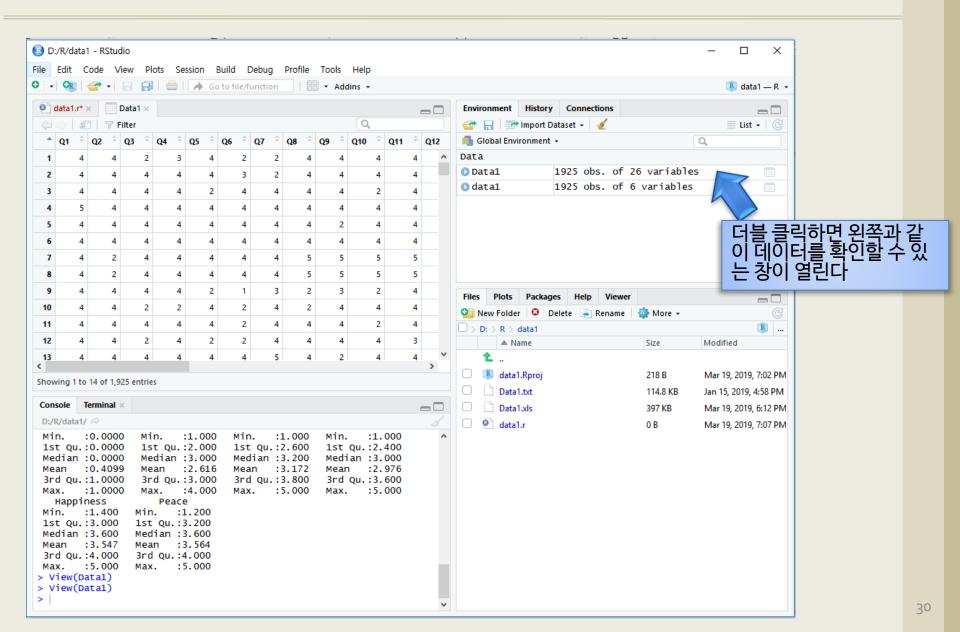


데이터 불러오기





데이터 확인





R 기초 - 객체(변수의 할당과 삭제)

http://dev.epiloum.net/1546

- □ 객체(Object): 대부분의 언어에서는 어떤 값을 보관하는 가상의 공간을 변수라고 하며, R에서는 객체(Object)라고 함
 - □ 객체는 미리 선언할 필요없이 할당하면 바로 생성됨
 - □ 문자와 순자를 사용할 수 있고 첫 글자는 반드시 문자
 - □ 영문자는 대문자와 소문자를 구분한다는 점 유의!!
- □ 객체의 할당: =, ⟨-, -⟩ 사용

```
1 x = 13
2 y <- sqrt(9)
3 4 * x -> z
```

□ 객체의 확인과 삭제: ls(), objects()로 확인. rm(객체이름)으로 삭제

```
1 ls()
2 objects()
3 rm(x)
```



R 기초 - 자료구조

- □ 7가지 자료구조를 할당할 수 있음
 - □ 스칼라: 객체에 하나의 값을 할당한 것. 실수, 문자열, 논리값(TRUE, FALSE)의 세 가지 자료형

```
1 x <- 2.7 # 실수
2 y <- "string" # 문자열
3 z <- FALSE # 논리값
```

벡터: 같은 자료형을 가진 스칼라 값들을 일렬로 나열한 자료구조

```
1  v <- c(1.2, 2.7, 3.1, 4.9, 5.4)
2  w <- c(TRUE, FALSE, FALSE)
3  x <- c("a", "b", "c", "d")
4  y <- 1.5:4.9  # 1.5 2.5 3.5 4.5
```

□ 매트릭스(행렬): 스칼라 값들을 2차원으로 나열한 구조

```
1 matrix(
2     c(1, 2, 4, 8, 16, 32),
3     nrow=2,
4     ncol=3
5 )
> matrix(
+ c(1,2,4,8,16,32),
+ nrow = 2,
+ ncol=3
+)
[,1] [,2] [,3]
[1,] 1 4 16
[2,] 2 8 32
```



R 기초 - 자료구조(계속)

- □ 배열(Array): 행렬을 3차원 이상으로 확장한 자료구조
- □ 리스트(List): 1차원 연관배열(Associative Array)에 해당

```
> y <- list(name="Mike", gender="M", company="ProgramCreek")
> y
$name
[1] "Mike"
$gender
[1] "M"
$company
[1] "ProgramCreek"
```

□ 팩터(Factor): 같은 항목 간에 그룹이 지어져 있는 자료형 태. 명목형 변수를 저장하는 자료구조

```
data(iris)
levels(iris$species)

iris$speciesf <- factor(iris$species, levels=c("versicolor"
    ,"virginica","setosa"))
levels(iris$speciesf)</pre>
```





R 기초 - 자료구조(계속)

□ 데이터프레임(Data Frame) : 모든 자료구조를 2차원 자료 구조로 모두 저장

```
> name <- c("Mike", "Lucy", "John")
> age <- c(20, 25, 30)
> student <- c(TRUE, FALSE, TRUE)
> df = data.frame(name, age, student)
> df
    name age student
1 Mike 20 TRUE
2 Lucy 25 FALSE
3 John 30 TRUE
```



R 기초 - 산술연산자

□ R에서 사용되는 산술연산자

```
1 3 + 2
2 7 - 4
3 11 * 5
4 16 / 5
5 16 %/% 5 # 정수나눗셈
6 16 %% 5 # 나머지
7 3^2 # 거듭제곱
```

■ 벡터나 행렬도 산술연산자를 사용할 수 있음

```
1 v <- c(1.2, 2.7, 3.1)
2 v <- v + 1 # 2.2 3.7 4.1
```



R 기초 - 조건문과 비교, 논리 연산자

- □ 비교연산자 : ⟨, ⟨=, ⟩, ⟩=, ==, !=
- □ 논리연산자: ||, &&
- □조건문

```
1  if(x == 3 || x >= 7 || x <= 1) {
2      x <- x+1
3  }
4  else(x != 1 && x < 3) {
5      x <- x-1
6  }</pre>
```



R 기초 - 반복문

■ while 문: while(구문) {} 의 구조, 구문이 TRUE일 동안 반복

```
1  i <- 0
2  while(i < 10) {
3     i <- i + 1
4  }</pre>
```

□ for 문 : *for(변수명 in 벡터) { }* 의 구조, 벡터의 숫자가 차례대로 들어 옴

```
1  sum <- 0
2  for(i in c(1, 4, 7)) {
3     sum <- sum + i
4  }</pre>
```

```
1    sum <- 0
2    for(i in 5:15) {
3        if(i%2 == 0)
4        {
5            next;
6        }
7        if(i%10 == 0)
8        {
9            break;
10        }
11        sum <- sum + i
12    }</pre>
```

- 〕 반복문을 중간에 중지할 때는 다른 언어와 마찬가지로 break문을 사용
- □ 반복문 내의 남은 구문을 모두 통과할 때는 다른 언어는 보통 continue문을 사용하지만, R 에서는 같은 구문이 next문 사용



R 기초 - 기초적인 수학적 계산

```
1 sin(pi) # 1.224606e-16
2 log(3) # 1.098612 (자연로그)
3 log10(100) # 2 (상용로그)
4 floor(3.14) # 3 (버림)
5 ceiling(5.87) # 6 (올림)
6 round(4.65) # 5 (반올림)
7 sqrt(9) # 3 (제곱근)
8 min(c(1,5,7)) # 1 (최소값)
9 max(c(2,6,9)) # 9 (최대값)
10 mean(c(3,6,9)) # 6 (평균)
11 sum(c(1,2,4,7)) # 14 (합계)
12 sd(c(1,2,3,4)) # 1.290994 (표준편차)
13 runif(2, 1, 10) # 난수 2개를 벡터로 생성(1 초과 10 미만)
```

```
1 # (편)미분

2 f <- expression(2*x^3 - y*x^2 + 2*y^2 + 1)

D(f, "x")

D(f, "y")

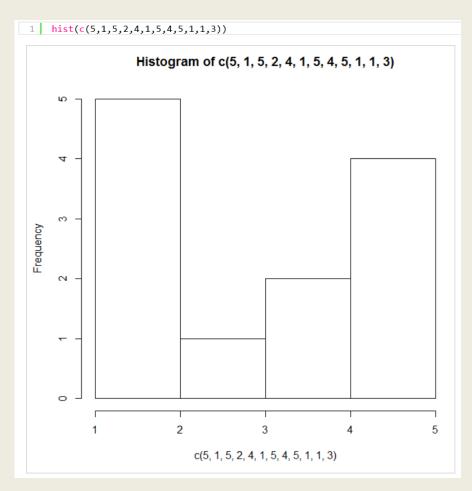
# 정적분

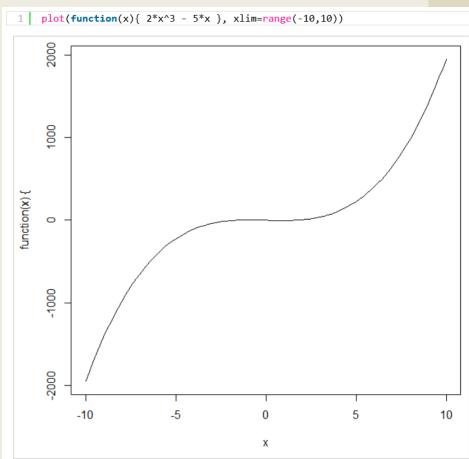
f <- function(x) 2*x^3 - 3*x^2 + 1

integrate(f, 0, 3)
```



R 기초 - 기초적인 시각화 도구



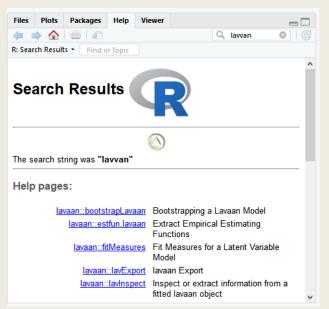




- R에서 제공되는 각종 함수와 데이터셋은 패키지(package)에 보관됨
- □ R 설치 시 기본적으로 표준패키지(standard package)가 설치 기본 함수, 통계분석 함수, 그래픽 함수, 참고용 데이터 셋이 포함되어 있음
- R 개발자들은 자신들이 개발한 코드를 패키지 형태로 공유하며 사용자들은 R에 원하는 패키지를 설치하고 메모리에 적재하여 사용하면 다양한 분석 기 능을 사용할 수 있음

install.packages("패키지 이름") #패키지 설치 library(패키지 이름) #메모리에 적재 detach(패키지 이름) #사용중지

🗅   주요 통계분석 패키지 : car, psych, mixlm, multcomp, lavvan



Help에 검색하면 관련 정보를 얻을 수 있음



R-예제 (R기초와 활용 page 8~9)

□ 10개 국가의 도시인구비율(urban.pop)과 기대수명 (life.exp)을 입력하고, 기대수명의 평균과 표준편차, 둘 간의 상관계수를 구하여라

```
ata1.r* × Difference | Ontitled1* ×
                           Data1 ×

← □ □ □ Source on Save □ □ ▼ ▼ □ □

                                                     >→ Source → =
                                             → Run
 1 urban.pop <- c(50,47,69,47,47,57,72,42,51,40)
 2 life.exp <- c(67,72,77,65,74,75,76,59,72,60)
 3 mean(life.exp)
 4 sd(life.exp)
  5 cor(life.exp, urban.pop)
Console
        Terminal ×
D:/R/data1/
> urban.pop <- c(50,47,69,47,47,57,72,42,51,40)
> life.exp <- c(67,72,77,65,74,75,76,59,72,60)
> mean(life.exp)
[1] 69.7
> sd(life.exp)
[1] 6.566751
> cor(life.exp, urban.pop)
[1] 0.7778366
```



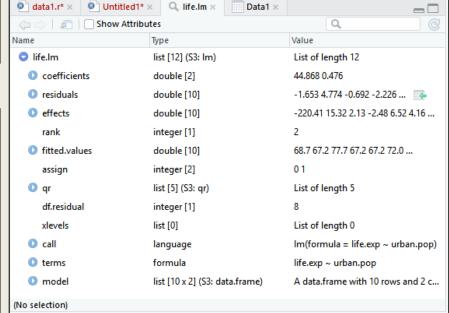
R-예제(계속)

□ 기대수명을 종속변수로, 도시인구비율을 독립변수 로 하여 회귀분석을 수행하여라

```
| Source on Save | Run | Run | I urban.pop <- c(50,47,69,47,47,57,72,42,51,40) | life.exp <- c(67,72,77,65,74,75,76,59,72,60) | mean(life.exp) | sd(life.exp) | cor(life.exp, urban.pop) | cor(life.exp, urban.pop) | life.lm <- lm(life.exp ~ urban.pop) | life.lm | life
```



```
Terminal ×
Console
D:/R/data1/ @
[1] 69.7
> sd(life.exp)
[1] 6.566751
> cor(life.exp, urban.pop)
[1] 0.7778366
> life.lm <- lm(life.exp ~ urban.pop)</pre>
> life.lm
call:
lm(formula = life.exp ~ urban.pop)
Coefficients:
(Intercept)
                urban, pop
    44.8678
                   0.4757
>
```



〈모형결과 저장형태(list 자료형)〉



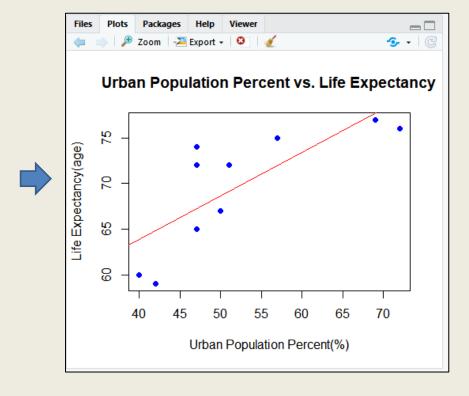
R-예제(계속)

□ 기대수명과 도시인구비율 간의 관계를 그래프로 그 려보자

```
plot(urban.pop, life.exp, col="blue", pch = 19,
main = "Urban Population Percent vs. Life Expectancy",
xlab = "Urban Population Percent(%)",
ylab = "Life Expectancy(age)")

abline(life.lm, col = "red")

17
```





- □ R과 Rstudio 설치하기
- □ Slide 41~43 page의 예제를 따라하고 R script file 과 결과물을 word로 간단히 정리해서 2강 과제 제출 에 업로드