

R 기초

R 기초

통계 계산과 그래프 작성을 위한 언어와 환경인 R은 뉴질랜드 Auckland 대학의 Ross Ihaka와 Robert Gentleman이 1995년 개발하고 R-Core 팀이 1997년 결성 2002년 R Foundation이 설립되어 전세계에 무료로 배포되고 있다.

1) GNU(GNU is Not Unix) Project

- Unix(멀티 테스킹, 범용 다중 사용자)-like Operating System
- 다양한 인터페이스, DBMS 데이터 수용
- 그래픽 기능이 강한 S 언어와 환경 하에서 개발
- S 코드 대부분은 수정 없이 R에서 실행 가능

2) S language 기반

- 설계자 John Chambers, 1998 (ACM award)
- 데이터 통합, 시각화, 분석을 위한 객체지향형 프로그래밍 언어
- 통계연구 프로젝트 착수 1976, 버전 4 출시 1998
- 대화형 기반 기술 데이터 입출력, 변환 용이
- 데이터 시각적 표현 중심으로 개발
- Java의 Eclipse 제공으로 프로그래머 개발 환경 개선
- S 기반 통계시스템 상용 프로그램

장점

- 사용 환경 ; small and effective

- 대화식 프로그램 수행
- 대용량 데이터 관리 및 처리, 행렬 연산, 그래픽 환경
- Freeware and Open source
 - GPL(General Public License) 개념으로 CRAN (Comprehensive R Archive Network)에서 배포
 - 전세계 연구자의 개발 알고리즘, 함수 활용 가능
- 다양한 확장성 및 범용성
 - UNIX, Mac, Windows 운영체제 사용이 가능함
 - Java, C, Fortran 프로그래밍 언어에 인터페이스 제공
 - DBMS 데이터 접근 용이
- Bio-informatics, Microarray 분석 표준시스템 - 빅데이터 platform
- Embedded R in Excel

R 사용준비

- Be smart: copy-left 소프트웨어 사용자, Bulky hater
- easy to use: 적은 시간 투자로 R 사용자 가능
- Be economical: free and get world-wide help
- Be challenge; 데이터, 프로그램 개별 관리, 행렬 개념의 데이터 관리,
- 데이터 분석: 함수 중심 알고리즘

R 설치

<http://www.r-project.org> 접속하여 설치 (2016.03.01 홈 페이지)



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 MacOS. 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

- [R version 3.2.4 \(Very Secure Dishes\) prerelease versions](#) will appear starting Monday 2016-02-29. Final release is scheduled for Thursday 2016-03-10.
- [R version 3.3.0 \(Supposedly Educational\) prerelease versions](#) will appear starting Monday 2016-03-14. Final release is scheduled for Thursday 2016-04-14.

Comprehensive R Archive Network for the R programming language (CRAN)은 R 배포판, 기여된 확장 프로그램들 (즉, 패키지들), R과 관련된 문서들, 그리고 바이너리들로 구성되어 있는 동일한 자료들을 전달하는 사이트들의 모음을 의미한다

Korea

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

Download and Install R

Precompiled binary distributions of want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

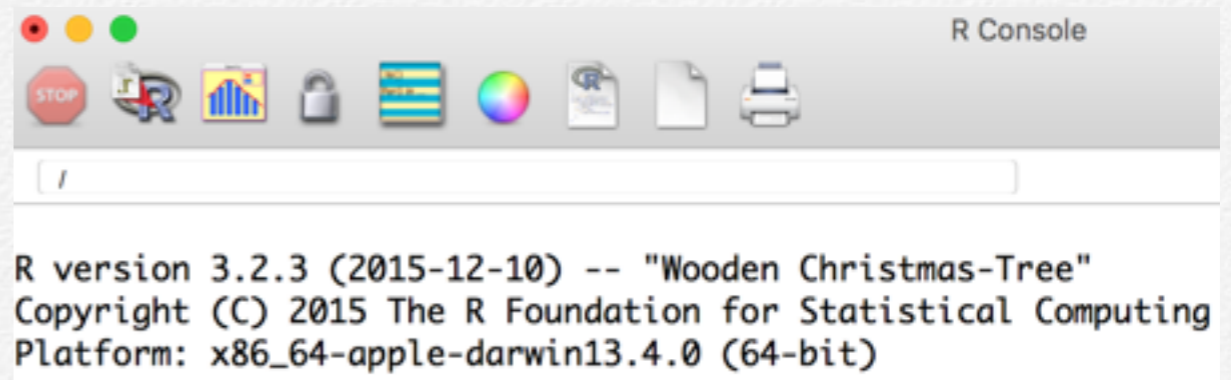
Subdirectories:

[base](#) Binaries for base distribution (managed by Duncan Murdoch). This is what you want to [install R for the first time](#).

R 초기화면

R은 3개의 윈도우로 구성 되어 있다.

- (1) Console 화면 : Unix 콘솔과 유사, 프로그램 직접 실행 혹은 Script 실행 된 로그와 텍스트 결과가 출력됨, 실행 초기에는 콘솔 화면만 나타난다.
- (2) Script 화면 : 프로그램 작업 공간 - 실행 hot-key “CTRL+R”
- (3) Graph 화면 : 프로그램이 실행 후 그래프가 출력 공간 (필요 시 팝업)



```
R version 3.2.3 (2015-12-10) -- "Wooden Christmas-Tree"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin13.4.0 (64-bit)
```

R 자체로 도움말 기능(“help(hist)”)이 있지만 오프소스로 사용법에 대한 도움 말은 전세계 홍수처럼 넘쳐나는데 그 중 백미 사이트는

- <http://www.statmethods.net/>
- <http://www.burns-stat.com/documents/tutorials/impatient-r/>

R 스튜디오 설치

R 사용을 용이하게 하는 IDE(통합 개발 환경) 도구인 R-Studio를 설치하는 것은 강추한다. R-Studio는 프로그램(script) 직접실행, 구문강조, 괄호 자동입력 지원, 명령어 완성, 다양한 단축키, 데이터 보기 및 가져오기, 그래픽 조작, 프로젝트 관리, 버전 관리 등의 다양한 기능을 제공한다.

<https://www.rstudio.com>



주요 CRAN 사이트

사이트	내용
Bayesian	Bayesian Inference
ChemPhys	Chemometrics and Computational Physics
ClinicalTrials	Clinical Trial Design, Monitoring, and Analysis
Cluster	Cluster Analysis & Finite Mixture Models
DifferentialEquations	Differential Equations
Distributions	Probability Distributions
Econometrics	Econometrics
Environmetrics	Analysis of Ecological and Environmental Data
ExperimentalDesign	Design of Experiments (DoE) & Analysis of Experimental Data
Finance	Empirical Finance
Genetics	Statistical Genetics
Graphics	Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization
HighPerformanceComputing	High-Performance and Parallel Computing with R
MachineLearning	Machine Learning & Statistical Learning
MedicalImaging	Medical Image Analysis
MetaAnalysis	Meta-Analysis
Multivariate	Multivariate Statistics
NaturalLanguageProcessing	Natural Language Processing
NumericalMathematics	Numerical Mathematics
OfficialStatistics	Official Statistics & Survey Methodology
Optimization	Optimization and Mathematical Programming
Pharmacokinetics	Analysis of Pharmacokinetic Data
Phylogenetics	Phylogenetics, Especially Comparative Methods
Psychometrics	Psychometric Models and Methods
ReproducibleResearch	Reproducible Research
Robust	Robust Statistical Methods
SocialSciences	Statistics for the Social Sciences
Spatial	Analysis of Spatial Data
SpatioTemporal	Handling and Analyzing Spatio-Temporal Data
Survival	Survival Analysis
TimeSeries	Time Series Analysis
WebTechnologies	Web Technologies and Services
gR	gRaphical Models in R