PART-I. *R* 설치 및 기초 문법



- 1. R 설치 및 개요
- 2. 데이터 유형과 구조
- 3. 데이터 입출력
- 4. 제어문과 함수

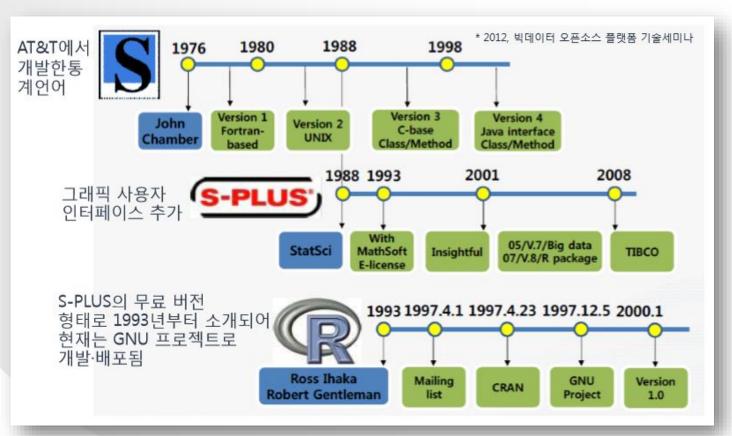


chap01_Basic 수업내용

- 1. R 설치 및 개요(Overview)
- 2. 변수와 데이터 유형
- 3. 패키지 사용
 - 패키지 개념, 기본 설치 패키지
 - 패키지 설치 확인, 사용/제거
- 4. 기본 함수 사용
- 5. 작업공간
- 6. 한글 저장 및 깨짐 현상 처리 방법 : UTF-8



- R 프로그램 탄생과 발전과정
 - ✓ 1993년 뉴질랜드 오클랜드 대학의 통계학과 교수2명(Ross Ihaka, Robert Gentleman) 개발



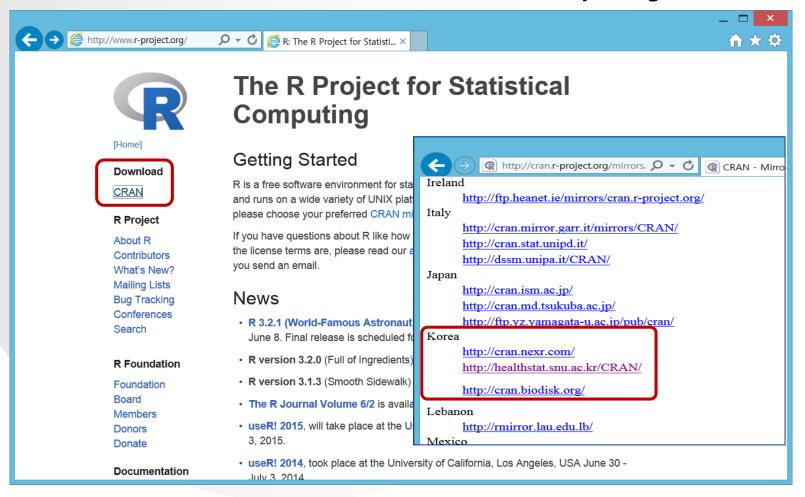


- R 프로그램 특징
 - I. In-Memory Computing
 - ✓ 빠른 처리 속도(H/W 메모리 크기에 영향 받음)
 - II. Object-oriented programming
 - ✓ 데이터, 함수, 차트 등 모든 것이 object로 관리
 - III. Package
 - ✓ 최신의 알고리즘 및 방법론이 Package로 제공됨
 - IV. Visualization
 - ✓ 분석에 통찰을 부여할 수 있는 그래픽에 대한 강력한 지원



http://www.r-project.org/

✓ CRAN Site에 패키지 0,000개 등록: dim(available.packages())



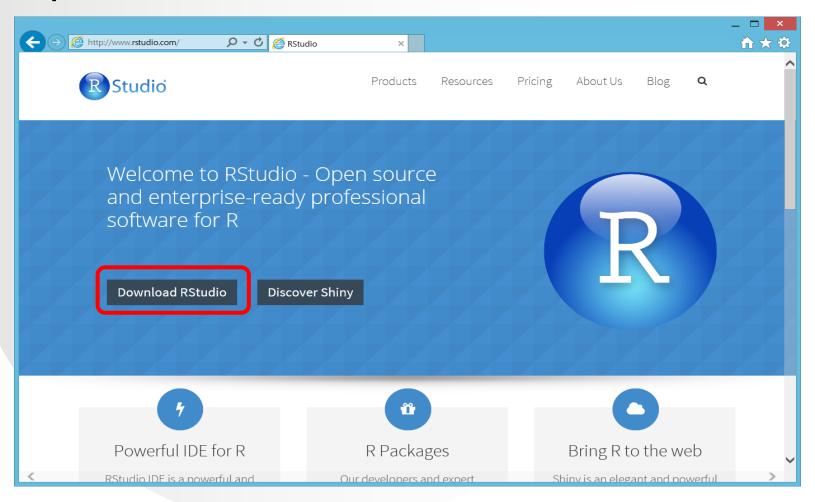


R console 화면

```
R Console
                                                                               - 0 X
R version 3.1.1 (2014-07-10) -- "Sock it to Me"
Copyright (C) 2014 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()'음 통하여 확인할 수 있습니다.
R은 많은 기여자들이 참여하는 공동프로젝트입니다.
'contributors()'라고 입력하시면 이에 대한 더 많은 정보를 확인하실 수 있습니다.
그리고, R 또는 R 패키지들음 출판물에 인용하는 방법에 대해서는 'citation()'음 통해 확인하시길 부탁드립니다.
'demo()'를 입력하신다면 몇가지 데모를 보실 수 있으며, 'help()'를 입력하시면 온라인 도움말을 이용하실 수 있습니다.
또한, 'help.start()'의 입력을 통하여 HTML 브라우저에 의한 도움말을 사용하실수 있습니다
R의 종료를 원하시면 'q()'음 입력해주세요.
> v <- c(11,12,13,14,15)
> w < -c(1,2,3,4,5)
[1] 12 14 16 18 20
[1] 11 24 39 56 75
> v/w
[1] 11.000000 6.000000 4.333333 3.500000 3.000000
> v^w
[1]
    11 144 2197 38416 759375
> 1
```

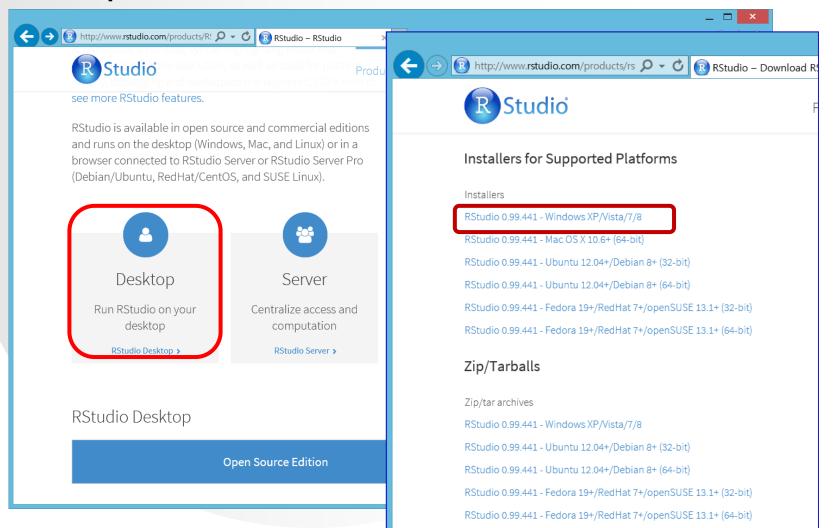


http://www.rstudio.com/



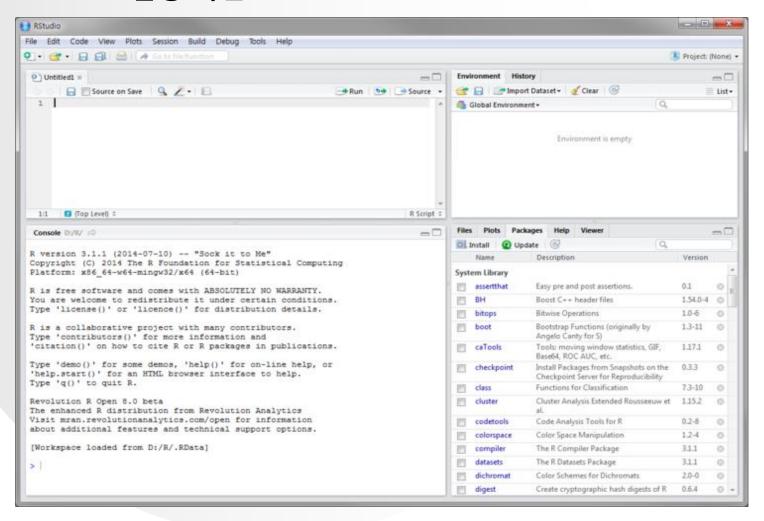


Desktop 선택



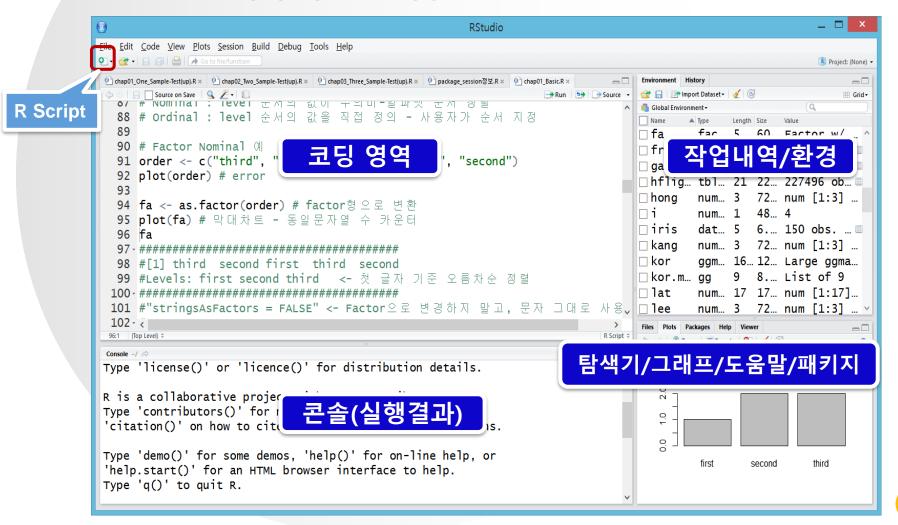


R Studio 실행 화면





● R Studio 4가지 패널 창 기능





- CRAN Site에서 제공하는 패키지 보기 dim(available.packages()) # 9650 17 -> 6,650개 available.packages() # 패키지 상세보기
- R session 정보
 - ✓ Session : R 콘솔 시작 ~ 종료 전까지 기간 정보

sessionInfo() # 세션 정보 보기

```
#R version 3.1.2 (2014-10-31)
```

#Platform: x86_64-w64-mingw32/x64 (64-bit)

#locale:

#attached base packages: 기본 패키지 - 7개

#[1] stats graphics

#[3] grDevices utils

#[5] datasets methods

#[7] base

loaded via a namespace (and not attached): 로딩된 패키지



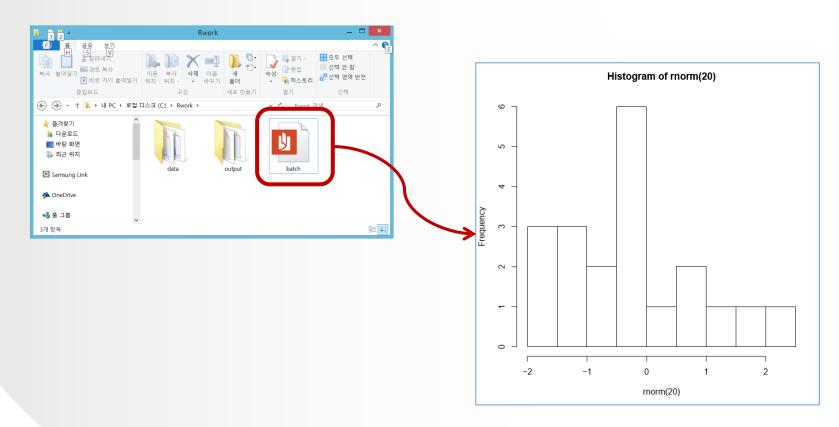
- 변수(Variable)
- 1) 변수
 - 값을 저장하는 메모리 이름
 - R은 모두 변수가 객체(데이터+함수+차트 등)
- 2) 변수명 작성 규칙
 - 영문자 시작
 - 두번째 단어 : 숫자, 언더바(_), 점(member.id) 사용 가능
 - 대소문자 구분
 - 의미를 파악할 수 있는 이름으로 지정
 - 정의된 변수는 재사용이 가능



- 명령어 실행 방법
- 1) Interaction 방식
 - ✓ 줄 단위 실행
 - ✓ 줄 선택 + Ctrl + R 또는 Ctrl + Enter
- 2) Batch 방식
 - ✓ 여러 줄 소스코드 일괄처리
 - ✓ Block 지정 + Ctrl + R
- Batch 방식 히스토그램 그리기 예 pdf("c:/Rwork/batch.pdf") # 지정된 경로의 파일에 결과 출력 hist(rnorm(20)) # 난수에 대한 히스토그램 그리기 dev.off() # 출력할 파일 닫음



■ Batch 방식 히스토그램 결과





- 데이터 유형(Data Type)
 - ✓ R은 변수 선언 시 별도의 Type을 선언하지 않음

유형(Type)	값(value)	예
숫자형(Numeric)	정수, 실수	10,25, 55.23
문자형(Character)	문자, 문자열	"홍길동"
논리형(Logical)	참, 거짓	TRUE 또는 T FALSE 또는 F
기타	결측치, 비숫자	NA, NaN



● 데이터 유형 보기

✓ mod()함수 : 데이터 유형 보기 함수

is.numeric(x)	수치형 여부	is.integer(x)	정수형 여부
is.logical(x)	논리형 여부	is.double(x)	실수형 여부
is.character(x)	문자형 여부	is.complex(x)	복소수형 여부
is.data.frame(x)	데이터프레임 여부	is.factor(x)	범주형 여부
is.na(x)	NA 여부	is.nan(x)	NaN 여부



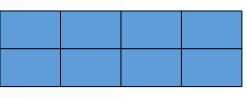
- 데이터 타입 보기 함수
 - √is.character(string) # TRUE
 - √x<- is.numeric(int)
 - √x # TRUE
 - ✓is.logical(boolean) # TRUE
 - ✓is.logical(x) # TRUE -> TRUE
 - ✓is.na(x) # NA 확인
- 현재 사용중인 변수객체 보기
 - √ls()



- 자료형과 자료구조(객체 타입) 보기
 - ✓ mode변수) : 자료형(Data Type) 보기
 - ✓ class(변수) : 자료구조(객체) 보기
- mode() vs class()
 - ✓ ex) 변수 <- 데이터
 </p>
 - ✓ mode : 데이터 성격(숫자형,문자형,논리형)
 - ✓ class : 데이터 자료구조 성격(Array, List, Table 등)



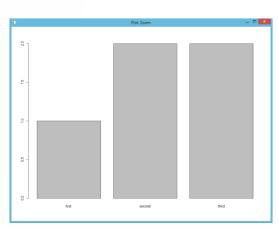
Array 자료구조



Matrix 자료구조



- Factor : 범주(값의 목록)를 갖는 자료구조
 - ✓ Nominal : level 순서의 값 무의미-알파벳 순서 정렬
 - ✓ Ordinal : level 순서의 값 의미 사용자가 직접 순서 지정
 - √# Factor Nominal 예
 - ✓order <- c("third", "second", "first", "third", "second")
 - √plot(order) # error
 - √fa <- as.factor(order) # factor형으로 변환
 - ✓plot(fa) # 막대차트 동일문자열 수 카운터



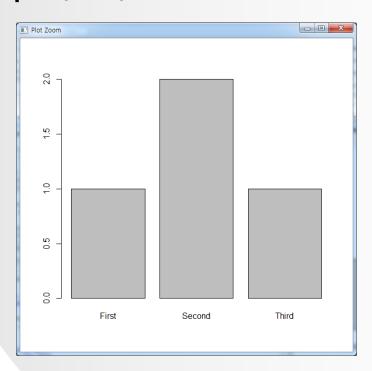


● 데이터 형 변환 함수

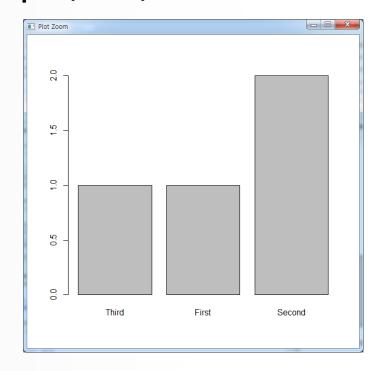
as.numeric(x)	수치형 변환	as.integer(x)	정수형 변환
as.logical(x)	논리형 변환	as.double(x)	실수형 변환
as.character(x)	문자형 변환	as.complex(x)	복소수형 변환
as.data.frame(x)	데이터프레임 변환	as.factor(x)	범주형 변환
as.list(x)	리스트형 변환	as.vector(x)	벡터형 변환



plot(level) # Nominal인 경우



plot(level2) # Ordinal인 경우





- 패키지(Package)
 - ✓ 패키지 개념: 데이터+기능(함수)+알고리즘 꾸러미(압축)
 - ✓ R 실력은 패키지 활용 능력
- 패키지 설치

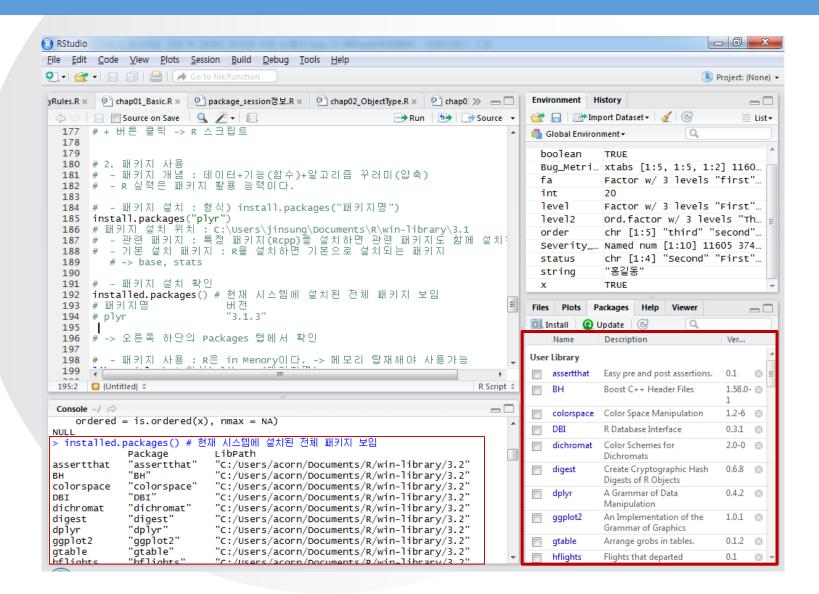
형식) install.packages("패키지명")

- ✓install.packages("plyr")
- 패키지 설치 위치
 - ✓ C:₩Users₩jinsung₩Documents₩R₩win-library₩3.1
 - ✓ 관련 패키지 : 특정 패키지 설치하면 관련 패키지도 함께 설치됨
 - ✓ 기본 설치 패키지 : R을 설치하면 기본으로 설치되는 패키지예) base, stats



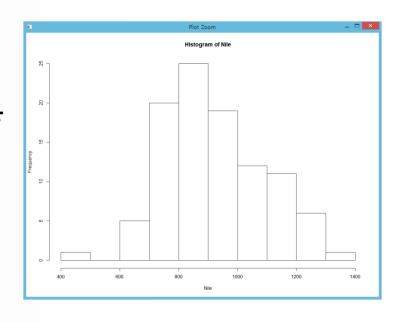
- 패키지 설치 확인
 - ✓ installed.packages() # 현재 시스템에 설치된 전체 패키지 보임
 - ✓ 패키지명 버전
 - ✓ plyr "3.1.3"
- 패키지 사용 : R은 in Menory -> 메모리 탑재해야 사용가능
 - 형식) library(패키지명)
 - √library(plyr)
 - √require(plyr)
- 현재 로드 된 패키지 확인
 - ✓ search()
- 패키지 제거 : 설치 폴더 제거 or remove.packages("패키지명")
 - ✓ remove.packages("plyr")







- 기본 함수 사용
 - ✓ 함수 도움말 : help(함수명), ?함수명
 - ✓ help(mean) # 브라우저 화면으로 제공
 - ✓ ?sum
 - ✓ google에서 검색 : 함수() in r
- 데이터 세트와 리스트 보기
 - √data()
 - ✓ hist(Nile) # 나일강 흐름에 대한 데이터 셋 히스토그램 그리기





data()

Data sets in package ; @arules; -:

Adult Adult Data Set
AdultUCI Adult Data Set
Epub Epub Data Set
Groceries Groceries Data Set
Income Income Data Set
IncomeESL Income Data Set

Data sets in package ; @datasets; -:

AirPassengers Monthly Airline Passenger Numbers 1949-1960

BJsales Sales Data with Leading Indicator

BJsales.lead (BJsales)

Sales Data with Leading Indicator

BOD Biochemical Oxygen Demand

CO2 Carbon Dioxide Uptake in Grass Plants
ChickWeight Weight versus age of chicks on different

diets

DNase Elisa assay of DNase

EuStockMarkets Daily Closing Prices of Major European Stock

Indices, 1991-1998

Formaldehyde Determination of Formaldehyde

HairEyeColor Hair and Eye Color of Statistics Students

Harman23.cor Harman Example 2.3 Harman74.cor Harman Example 7.4

Indometh Pharmacokinetics of Indomethacin
InsectSprays Effectiveness of Insect Sprays

JohnsonJohnson Quarterly Earnings per Johnson & Johnson

Share

LakeHuron Level of Lake Huron 1875-1972

Loblolly Growth of Loblolly pine trees

Nile Flow of the River Nile
Orange Growth of Orange Trees
OrchardSprays Potency of Orchard Sprays

PlantGrowth Results from an Experiment on Plant Growth Puromycin Reaction Velocity of an Enzymatic Reaction Seatbelts Road Casualties in Great Britain 1969-84

Theoph Pharmacokinetics of Theophylline
Titanic Survival of passengers on the Titanic

ToothGrowth The Effect of Vitamin C on Tooth Growth in

Guinea Pigs

UCBAdmissions Student Admissions at UC Berkeley
UKDriverDeaths Road Casualties in Great Britain 1969-84

R 제공 데이터 셋



● 함수 파라미터 보기

```
형식) args(함수명)
```

- √args(max) #
- ✓# max()함수 파라미터 : (na.rm = FALSE) <- NA 제거 안됨</p>
- √max(10,20,NA,30) # NA 출력

●함수 예제 보기

형식) example(함수명) : 해당 함수의 사용 예 제공

- ✓ example(seq)
- ✓ seq(10,20, by=2) # 10 12 14 16 18 20
- √example(mean)
- √mean(10:20) # 15



- 작업공간 보기
 - ✓ getwd() # 기본함수
- 작업공간 지정
 - ✓ setwd("C:/Rwork/Part-I") # 구분자 : / or ₩₩
 - ✓data <- read.csv("test.csv", header=T) # 지정된 경로에서 파일 로딩
 - ✓args(read.csv) # (file, header = TRUE)
 - √ data

test.csv 파일 내용



- 스크립트 파일 저장 및 깨짐 현상 처리 방법
 - ✓ 문자 인코딩 방식 지정 : UTF-8
 - ✓ 메뉴 : Tools > General Options > General Default text encoding-UTF-8
 - ✓ 메뉴 : File > Save with Encoding > UTF-8

Options	
General	R version: [Default] [64-bit] C:\Program Files\R\R-3.1.2 Change
Tode	Default working directory (when not in a project):
Appearance Pane Layout	Restore most recently opened project at startup Restore previously open source documents at startup Restore .RData into workspace at startup Save workspace to .RData on exit: Ask
Packages Sweave	✓ Always save history (even when not saving .RData) Remove duplicate entries in history ✓ Use debug error handler only when my code contains errors Automatically expand tracebacks in error inspector
Spelling Git/SVN	Default text encoding: UTF-8 Change Automatically notify me of updates to RStudio
Publishing	OK Cancel Apply