

R 설치

← → ↻ <https://cran.r-project.org/bin/windows/base/>

R-3.5.3 for Windows (32/64 bit)

[Download R 3.5.3 for Windows](#) (79 megabytes, 32/64 bit)

[Installation and other instructions](#)

[New features in this version](#)

If you want to double-check that the package you have downloaded matches the package distributed by CRAN, you can compare the [md5sum](#) of the .exe to the [fingerprint](#) on the master server. You will need a version of md5sum for windows: both [graphical](#) and [command line versions](#) 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 [R FAQ](#) 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](https://cran.r-project.org/bin/windows/base/release.htm).

R 프로그래밍 특성 (무료 – 오픈 소스)

- In-Memory Computing

빠른 처리 속도(H/W 메모리 크기에 영향 받음)

- Object-oriented programming

데이터, 함수, 차트 등 모든 것이 object로 관리

- Package

최신의 알고리즘 및 방법론이 Package로 제공됨

- Visualization

분석에 통찰을 부여할 수 있는 그래픽에 대한 지원

R Studio 설치

https://www.rstudio.com/products/rstudio/download/



Products Resources Pricing About Us Blogs

Choose Your Version of RStudio

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.](#)



	RStudio Desktop Open Source License	RStudio Desktop Commercial License	RStudio Server Open Source License	RStudio Server Pro Commercial License	RStudio Server Pro + RStudio Connect Commercial License
	FREE	\$995 per year	FREE	\$9,995 per year	\$29,995 per year
	DOWNLOAD Learn More	BUY Learn More	DOWNLOAD Learn More	DOWNLOAD Learn More	TALK Learn More
Integrated Tools for R	●	●	●	●	●
Priority Support		●		●	●
Access via Web Browser			●	●	●
Enterprise Security				●	●
Project Sharing				●	●
Manage Multiple R Sessions & Versions				●	●
Admin Dashboard				●	●
Load Balancing				●	●

R Studio 화면

1 #R의 corr 함수를 통해 상관계수를 파악하고 그래프를 통해 독립변수 간의 상관관계 분석을 수행한다. EuStockMarkets의 DAX와 EuStockMarkets SMI 상관분석
2 cor(EuStockMarkets[, 'DAX'], EuStockMarkets[, 'SMI'])
3 cor(EuStockMarkets)
4
5 # 상관계수 행렬 plot
6 install.packages("corrplot")
7 library(corrplot)
8 # 상관계수 행렬 계산
9 CorrEuStockMarkets <- cor(EuStockMarkets)
10 # 상관계수 행렬, plot 함수, method로는 ellipse 사용
11 corrplot(CorrEuStockMarkets, method="ellipse")
12 #ellipse 이외에도 circle, square, number, shade, color, pie 등이 있음
13 corrplot(CorrEuStockMarkets, method="square")
14 corrplot(CorrEuStockMarkets, method="shade")
15 corrplot(CorrEuStockMarkets, method="number")
16 corrplot(CorrEuStockMarkets, method="pie")
17 corrplot(CorrEuStockMarkets, method="circle")
18 (top Level) :

코딩 영역

diff <- Before -After
diff
source('C:/RStudio/데이터분석과R프로그래밍/paired_Ttest.R', encoding = 'UTF-8', asIs=TRUE)
Before와 after의 차이확이 확인. 가설을 설정해 보자 귀무가설, 대립가설
diff <- Before -After
diff
source('C:/RStudio/데이터분석과R프로그래밍/paired_Ttest.R', encoding = 'UTF-8', asIs=TRUE)
대응표본t test를 해보자
t.test(Before, After, alternative=c("two.sided", "less", "greater"),
conf.level=0.95)
data <- read.csv("diet.csv", header =T)
attach(data)
names(data)
summary(data)
대응표본t test를 해보자
t.test(Before, After, alternative=c("two.sided"), paired=TRUE,
conf.level=0.95)
diet.csv 파일에 있는 내용을 boxplot로 그린다. 다이어트전후 차이. 최저값 최저값 중앙값, 3사분위 표시
boxplot(diff)
diff

작업내역/환경

R version 3.5.2 (2018-12-20) -- "Eggshell Igloo"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-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 on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from D:/Rdata/.RData]
> |

콘솔(실행결과)

R Lagged Differences

diff (base)

Lagged Differences

Description

Returns suitably lagged and iterated differences.

Usage

diff(x, ...)
Default S3 method:
diff(x, lag = 1, differences = 1, ...)
S3 method for class 'POSIXt':
diff(x, lag = 1, differences = 1, ...)
S3 method for class 'Date':
diff(x, lag = 1, differences = 1, ...)

Arguments

x a numeric vector or matrix containing the values to be differenced.
lag an integer indicating which lag to use.
differences an integer indicating the order of the difference.

탐색기/그래프/도움말/패키지

Package 사용

The screenshot shows the RStudio interface. The top-left pane displays the R script with the following code:

```
1 # Cran site에서 제공하는 패키지 활용하기
2 dim(available.packages())
3
```

The bottom-left pane shows the console output:

```
R version 3.5.2 (2018-12-20) -- "Eggshell Igloo"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-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 on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from D:/Rdata/.RData]

> # Cran site에서 제공하는 패키지 활용하기
> dim(available.packages())
[1] 13876 17
>
```

The bottom-right pane shows the R Documentation for the `dim` function:

Dimensions of an Object

Description
Retrieve or set the dimension of an object.

Usage
`dim(x)`
`dim(x) <- value`

Arguments
`x` an R object, for example a matrix, array or data frame.
`value` For the default method, either `NULL` or a numeric vector, which is coerced to integer (by truncation).

Details
The functions `dim` and `dim<-` are [internal generic primitive](#) functions.
`dim` has a method for `data.frames`, which returns the lengths of the `row.names` attribute of `x` and of `x` (as the numbers of rows and columns respectively).

Value
For an array (and hence in particular, for a matrix) `dim` retrieves the `dim` attribute of the object. It is `NULL` or a vector of mode `integer`.
The replacement method changes the `"dim"` attribute (provided the new value is compatible) and removes any `"dimnames"` and `"names"` attributes.

Cran site에서 제공하는 패키지 활용하기
dim(available.packages()) 결과값 13876개 Package

available.packages()

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains the R script:

```
1 # Cran site에서 제공하는 패키지 활용하기
2 dim(available.packages())
3 available.packages()
4
```
- Console:** Displays the output of the `available.packages()` function as a matrix with 4 columns. The first column lists package names, and the other three columns contain `NA` values.

Package Name	Column 2	Column 3	Column 4
ABCanalysis	NA	NA	NA
abcdeFBA	NA	NA	NA
ABCoptim	NA	NA	NA
ABCp2	NA	NA	NA
abcrf	NA	NA	NA
abctools	NA	NA	NA
abd	NA	NA	NA
abe	NA	NA	NA
abf2	NA	NA	NA
ABHgenotypeR	NA	NA	NA
abind	NA	NA	NA
abjutils	NA	NA	NA
abn	NA	NA	NA
abnormality	NA	NA	NA
abodOutlier	NA	NA	NA
ABPS	NA	NA	NA
AbsFilterGSEA	NA	NA	NA
AbSim	NA	NA	NA
abstractr	NA	NA	NA
abundant	NA	NA	NA
Ac3net	NA	NA	NA
ACA	NA	NA	NA
acc	NA	NA	NA
accelerometry	NA	NA	NA
accelmissing	NA	NA	NA
AcceptanceSampling	NA	NA	NA
ACCLMA	NA	NA	NA
accrual	NA	NA	NA
accrued	NA	NA	NA
accSDA	NA	NA	NA
ACD	NA	NA	NA
- Environment/History/Connections:** Shows a 'New Connection' dialog box.
- Files/Plots/Packages/Help/Viewer:** The 'Help' tab is active, displaying the documentation for `available.packages()`. The documentation includes:
 - Description:** `available.packages` returns a matrix of details corresponding to packages currently available at one or more repositories. The current list of packages is downloaded over the internet (or copied from a local mirror).
 - Usage:**

```
available.packages(contriburl = contrib.url(repos, type), method,
  fields = NULL, type = getOption("pkgType"),
  filters = NULL, repos = getOption("repos"),
  ignore_repo_cache = FALSE, max_repo_cache_age,
  ...)
```
 - Arguments:**
 - `contriburl`: URL(s) of the 'contrib' sections of the repositories. Specify this argument only if your repository mirror is incomplete, e.g., because you burned only the 'contrib' section on a CD.
 - `method`: download method, see [download.file](#).
 - `type`: character string, indicate which type of packages: see [install.packages](#). If `type = "both"` this will use the source repository.
 - `fields`: a character vector giving the fields to extract from the 'PACKAGES' file(s) in addition to the default ones, or `NULL` (default). Unavailable fields result in `NA` values.
 - `filters`: a character vector or list or `NULL` (default). See 'Details'.

Package 사용예

(1) 패키지(Package)

패키지 개념 : 데이터+기능(함수)+알고리즘 (압축),
R 활용 및 응용은 패키지 활용 능력

(2) 패키지 설치

형식) `install.packages("패키지명"), install.packages("ggplot2")`

(3) 패키지 설치 확인

`installed.packages()` # 현재 시스템에 설치된 전체 패키지

(4) 패키지 사용 : R은 in Memory -> 메모리 탑재해야 사용가능

형식) `library(패키지명)` , `library(ggplot2)`, `require(ggplot2)`

(5) 현재 로드 된 패키지 확인. `search()`

(6) 패키지 제거 : `remove.packages("패키지명")`

`remove.packages("ggplot2")`

Package 실행

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for managing packages. It includes comments in Korean and functions like `dim(available.packages())`, `install.packages()`, `library()`, `require()`, `search()`, and `remove.packages()`.
- Console:** Shows the output of the executed code. It lists installed and available packages (e.g., MASS, Matrix, memoise) and their versions. It also shows an error message: "Error in remove.packages : there is no package called 'ggplot2'".
- Environment/History/Connections:** The Environment pane is active, showing the search path for R objects. It lists the current search path, including `.".GlobalEnv"`, `"package:ggplot2"`, `"tools:rstudio"`, `"package:stats"`, `"package:graphics"`, `"package:grDevices"`, `"package:utils"`, `"package:datasets"`, `"package:methods"`, and `"Autoloads"`.
- Documentation:** The right pane shows the documentation for the `search` function, including its description, usage, value, and references.

```
# Cron site에서 제공하는 패키지 활용하기
dim(available.packages())
available.packages()
# 패키지(Package) 패키지 개념 : 데이터+기능(함수)+알고리즘(알속), R 활용 및 응용은
# 패키지 활용 능력
# 패키지 설치 형식) install.packages("패키지명").
install.packages("ggplot2")
# 패키지 설치 확인
installed.packages()
# 현재 시스템에 설치된 전체 패키지, 패키지 사용 : R은 in Memory -> 메모리 탑재해야
# 사용가능, 형식) library(패키지명) ,
library(ggplot2)
require(ggplot2)
#현재 로드된 패키지 확인.
search()
#패키지 제거 : remove.packages("패키지명")
remove.packages("ggplot2")
```

Console Output:

```
MASS NA NA NA "yes" "3.5.1"
Matrix NA NA NA "yes" "3.5.1"
memoise NA NA NA "no" "3.5.2"
mgcv NA NA NA "yes" "3.5.1"
mime NA NA NA "yes" "3.5.2"
munsell NA NA NA "no" "3.5.2"
openssl NA NA NA "yes" "3.5.2"
pillar NA NA NA "no" "3.5.2"
pkgbuild NA NA NA "no" "3.5.2"
pkgconfig NA NA NA "no" "3.5.2"
pkgload NA NA NA "yes" "3.5.2"
plyr NA NA NA "no" "3.5.2"
prettyunits NA NA NA "yes" "3.5.2"
processx NA NA NA "yes" "3.5.2"
[ reached getOption("max.print") -- omitted 160 rows ]
> #패키지 제거 : remove.packages("패키지명")
> remove.packages("ggplot2")
Removing package from 'C:/Users/vivah/OneDrive/문서/R/win-library/3.5'
(as 'lib' is unspecified)
Error in remove.packages : there is no package called 'ggplot2'
> # 현재 시스템에 설치된 전체 패키지, 패키지 사용 : R은 in Memory -> 메모리 탑재해야 사용가능,
# 형식) library(패키지명) ,
> library(ggplot2)
> require(ggplot2)
> #현재 로드된 패키지 확인.
> search()
[1] ".GlobalEnv" "package:ggplot2" "tools:rstudio" "package:stats"
[5] "package:graphics" "package:grDevices" "package:utils" "package:datasets"
[9] "package:methods" "Autoloads" "package:base"
```

Documentation: Give Search Path for R Objects

Description
Gives a list of attached packages (see [library](#)), and **R** objects, usually **data.frames**.

Usage
`search()`
`searchpaths()`

Value
A character vector, starting with `".GlobalEnv"`, and ending with `"package:base"` which is R's **base** package required always.
`searchpaths` gives a similar character vector, with the entries for packages being the path to the package used to load the code.

References
Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) *The New S Language*. Wadsworth & Brooks/Cole. ([search](#)).
Chambers, J. M. (1998) *Programming with Data. A Guide to the S Language*. Springer. ([searchPaths](#)).

See Also
[.packages](#) to list just the packages on search path.
[loadedNamespaces](#) to list loaded namespaces.

Package : 영화 매트릭스



<https://www.youtube.com/watch?v=2svLxh03bYw>

변수(Variable)와 데이터 유형(Data type)

1) 변수(Variable)

- 값을 저장하는 메모리 이름
- R은 모두 변수가 객체(데이터+함수+차트 등)

2) 변수명 작성 규칙

- 첫번째 단어 : 영문자 시작
- 두번째 단어 : 숫자, 언더바(_), 점(name.id) 사용 가능
- 대소문자 구분
- 의미를 파악할 수 있는 이름으로 지정

3) 데이터 유형(Data Type)

유형(Type)	값(value)	예
숫자형(Numeric)	정수, 실수	1, 2, 3, 22, 23
문자형(Character)	문자, 문자열	"대한민국"
논리형(Logical)	참, 거짓	TRUE 또는 T FALSE 또는 F
기타	결측치, 정의되지 않음	NA, Null

R에서 제공하는 주요 자료구조(객체 타입)

1. Vector : 동일 데이터 타입을 갖는 1차원 배열
2. Matrix : 동일 데이터 타입을 갖는 2차원 배열
3. Array : 동일 데이터 타입을 갖는 3차원 배열
4. Data Frame : 열 단위로 서로 다른 데이터 타입을 갖는 배열
5. List : 서로 다른 데이터 구조(Vector, Data Frame, Array, List)

중첩

1. Vector

1) Vector 자료구조

- ✓ R의 기본 데이터 구조

- ✓ 1차원 배열 형태

 - 접근 : [index] : 1부터 시작

- ✓ 동일한 타입의 데이터만 저장 가능

- ✓ `X <-c(1, 10, 13)` 숫자형 벡터

- ✓ `Y <-c("대한민국", "자동차")` 문자형 벡터

- ✓ `Z <-c(TRUE, FALSE)` 논리형 벡터

- ✓여기서 C는 concentration (연결하다)



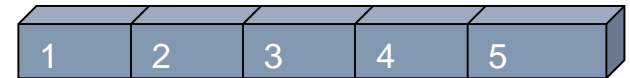
Vector

2. Matrix

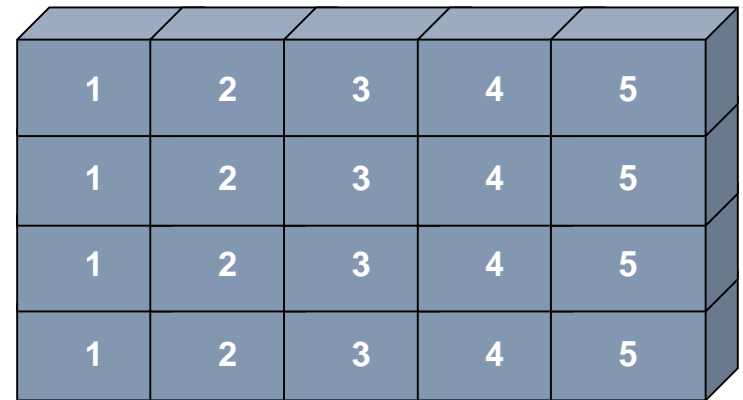
2) Matrix 자료구조

- ✓ 동일 데이터 타입을 갖는 2차원 배열
- ✓ 행렬(matrix) 객체

```
> #행렬  
> mx=matrix(c(1,2,3,4,5,6,7,8), ncol=2)  
> mx  
      [,1] [,2]  
[1,]    1    5  
[2,]    2    6  
[3,]    3    7  
[4,]    4    8
```



Vector



Matrix

3. Array

3) Array : 동일 데이터 타입을 갖는 다차원 배열

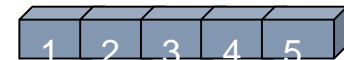
✓ 3차원 배열 객체 생성

✓ R에서 활용도 낮음

`d <- c(1:12)` # 12개 벡터 객체 생성

`arr <- array(d, c(3,2,2))` # 3행2열 구조 2개

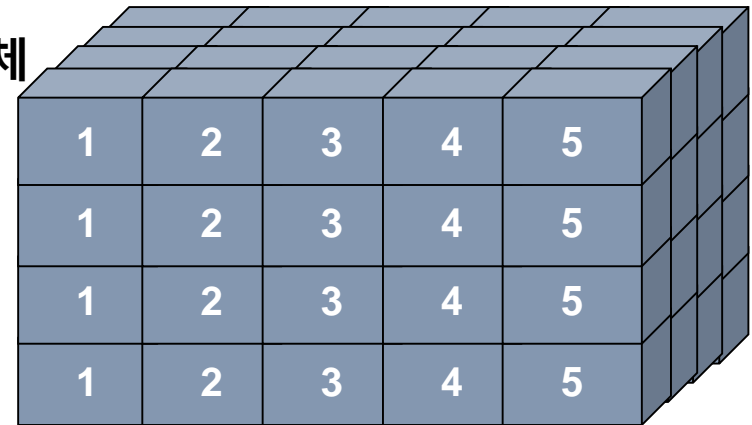
`arr` #1~6(1면), 7~12(2면) -> 3차원 배열 객체



Vector



Matrix

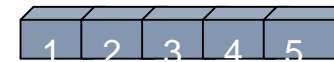


Array

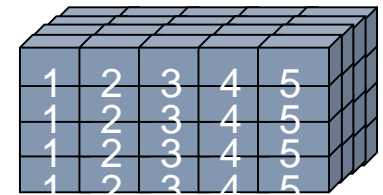
4. Data Frame

5) Data Frame

- ✓ 리스트 보다 활용범위 넓다.
- ✓ DBMS의 테이블 구조와 유사
(서로 다른 데이터 타입을 갖는 컬럼)
- ✓ 가장 많이 사용하는 객체 타입
- ✓ list와 Vector 혼합형
 - 컬럼 구성 : list, list 구성 : vector
- ✓ data frame 생성방법
 - Vector, Matrix, txt/excel/csv 파일



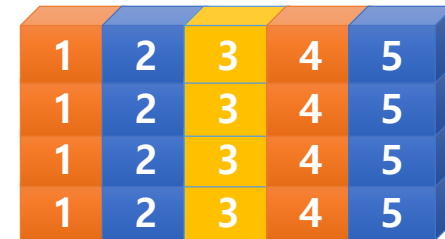
Vector



Array



Matrix



Data Frame

5. Data Frame

- Data Frame 특징

1. 형식) data.frame(컬럼1, 컬럼2.. 컬럼n)
2. 컬럼 단위로 서로 다른 자료형 가능
3. 모든 컬럼은 크기가 동일해야 함

```
> #데이터 프레임
> income = c(10,100,200,300)
> car=c("Samsung","KIA","Hyundai","BMW")
> marriage=c(FALSE, FALSE, TRUE, FALSE)
> mydat=data.frame(income,car,marriage)
> mydat
```

	income	car	marriage
1	10	Samsung	FALSE
2	100	KIA	FALSE
3	200	Hyundai	TRUE
4	300	BMW	FALSE

5. List

4) List : 서로 다른 데이터 구조

- ✓ Vector, Data Frame, Array, List의 중첩 구조
- ✓ c(구조체), python(딕셔너리)
- ✓ 함수 내에서 여러 값을 하나의 키로 묶어서 반환할 경우 유용함

```
member <- list(name="홍길동",  
               age = 25,address="한양",  
               gender="남자", htype="아파트")
```

Key	Value
name	홍길동
age	25
address	한양시
gender	남자
htype	아파트