

# R을 이용한 통계 기초 (1일차)

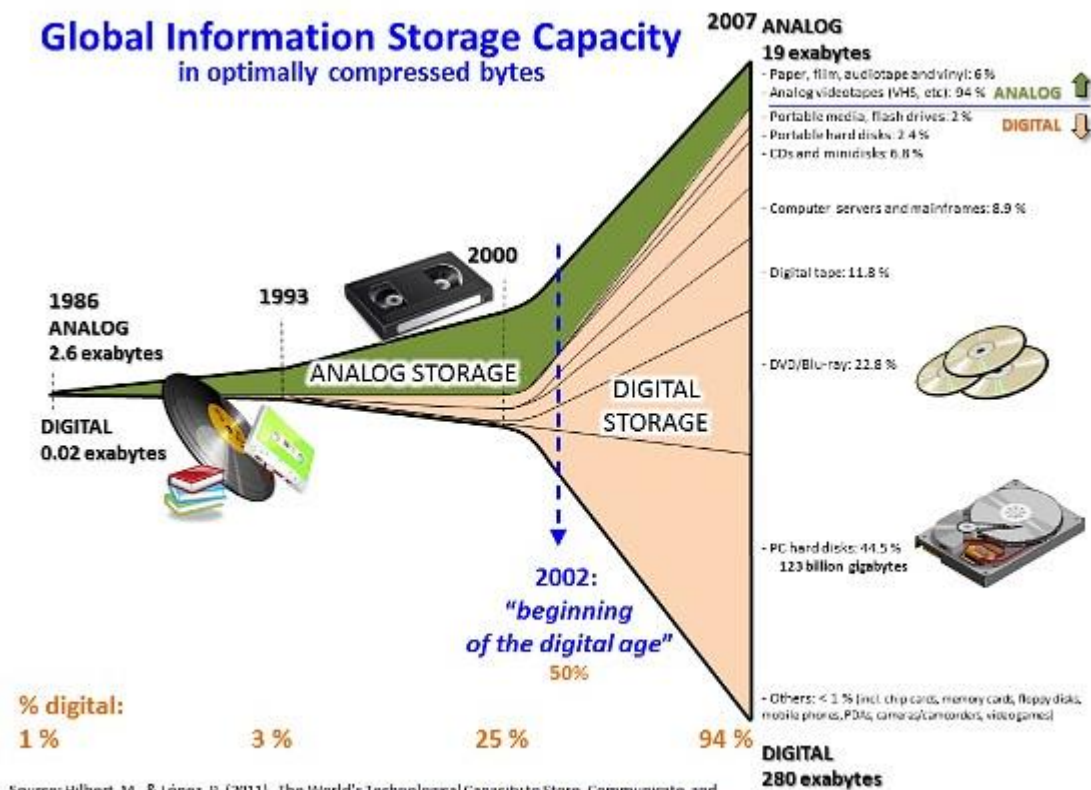
인하대학교 대학원 통계학과 국성희

# Contents

1. Big Data에 대한 소개 및 활용 사례
2. R이란 무엇인가?
3. R의 기본 연산
4. 데이터 벡터 만들기
5. 데이터 벡터에 함수 적용하기
6. 구조적인 데이터 만들기 및 다루기
7. 연습문제



# 1. Big Data에 대한 소개 및 활용 사례



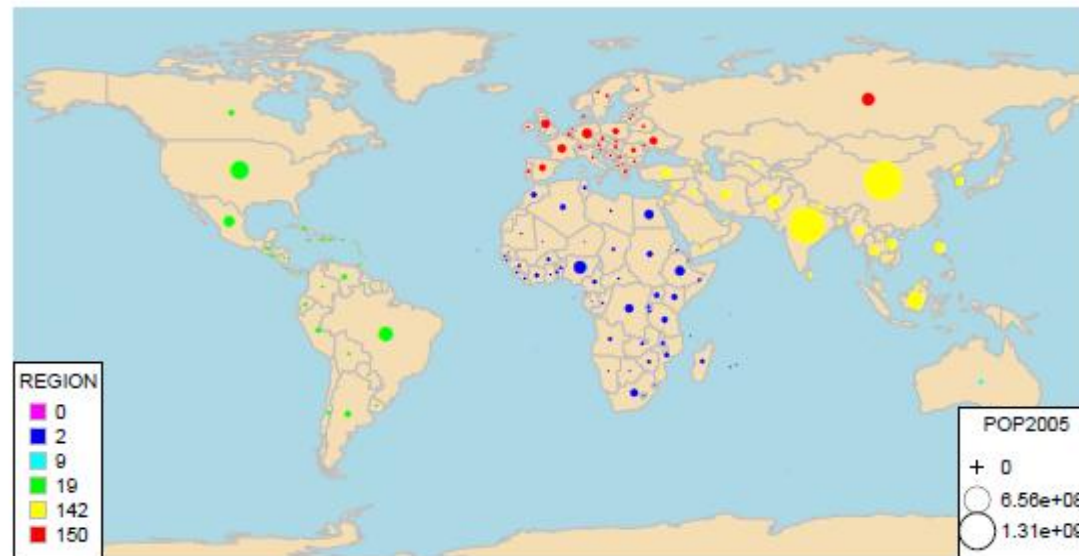
(이미지 출처 : <http://www.martinhilbert.net/WorldInfoCapacity.html> )

# 1. Big Data에 대한 소개 및 활용 사례

\* 한국 석유공사, 국가 유가 예보 서비스

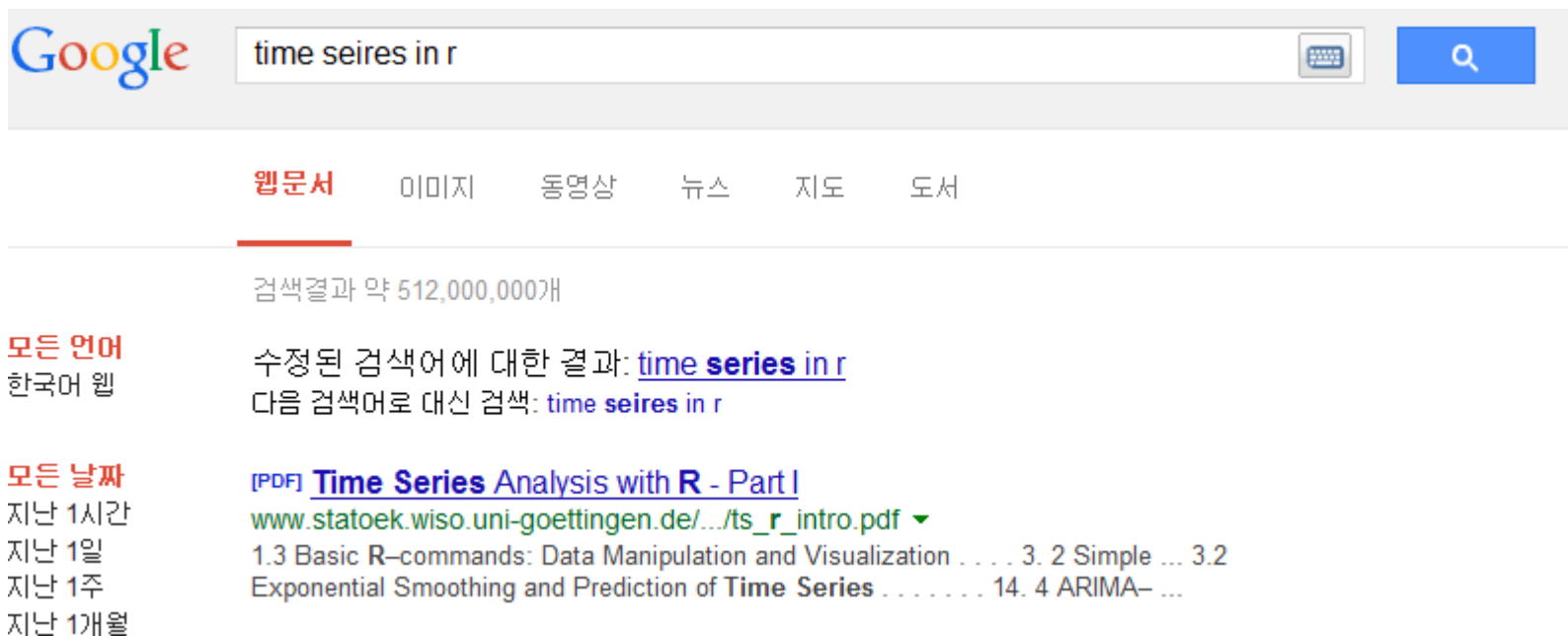


\* 공간적 특성 확인



## 2. R이란 무엇인가?

- Open Source Package .
- The language is very powerful for writing programs.
- Many statistical functions are already built in.



Google search results for "time seires in r".

Search bar: time seires in r

Filters: 웹문서 (selected), 이미지, 동영상, 뉴스, 지도, 도서

검색결과 약 512,000,000개

모든 언어  
한국어 웹

수정된 검색어에 대한 결과: [time series in r](#)  
다음 검색어로 대신 검색: [time seires in r](#)

모든 날짜  
지난 1시간  
지난 1일  
지난 1주  
지난 1개월

[PDF] [Time Series Analysis with R - Part I](#)  
[www.statোক.wiso.uni-goettingen.de/.../ts\\_r\\_intro.pdf](#) ▼  
1.3 Basic R-commands: Data Manipulation and Visualization . . . 3. 2 Simple ... 3.2  
Exponential Smoothing and Prediction of Time Series . . . . . 14. 4 ARIMA- ...

## 2. R이란 무엇인가?

Go to [www.rseek.org](http://www.rseek.org)



검색결과 약 1,500,000,000개(0.63초)

### [CRAN Task View: Time Series Analysis](#)

Base R ships with a lot of functionality useful for **time series**, in particular in the stats package. This is complemented by many packages on CRAN, which are ...  
[cran.r-project.org/web/views/TimeSeries.html](http://cran.r-project.org/web/views/TimeSeries.html)

### [Quick-R: Time Series](#)

R has extensive facilities for analyzing **time series** data. This section describes the creation of a **time series**, seasonal decomposition, modeling with exponential ...  
[www.statmethods.net/advstats/timeseries.html](http://www.statmethods.net/advstats/timeseries.html)

### [CRAN - Package timeSeries](#)

May 1, 2013 ... **timeSeries**: Rmetrics - Financial **Time Series** Objects. Environment teaching "Financial Engineering and Computational Finance" ...  
[cran.r-project.org/package=timeSeries](http://cran.r-project.org/package=timeSeries)

### [Temporal Disaggregation of Time Series](#)

파일 형식: PDF/Adobe Acrobat  
higher frequency series, where either the sum, the average, the first or the last ... Not **time series** at the desired frequency is a common problem for ...  
[journal.r-project.org/archive/2013-2/sax-steiner.pdf](http://journal.r-project.org/archive/2013-2/sax-steiner.pdf)

### [Package 'timeSeries'](#)

파일 형식: PDF/Adobe Acrobat  
May 1, 2013 ... Package **'timeSeries'**: August 29, 2013. Version 3010.97. Revision 54  
Date 2013-03-25. Title Rmetrics - Financial **Time Series** Objects.  
[cran.r-project.org/web/packages/timeSeries/timeSeries.pdf](http://cran.r-project.org/web/packages/timeSeries/timeSeries.pdf)

### [Time Series in R 1.5.0](#)

파일 형식: PDF/Adobe Acrobat  
ments in the standard **time series** package ts: much improved handling of missing values, ARIMA support, structural equation modeling, and exponential smoothing.  
[www.r-project.org/doc/Rnews/Rnews\\_2002-2.pdf](http://www.r-project.org/doc/Rnews/Rnews_2002-2.pdf)

[Introductions](#)[Task Views](#)[Support Lists](#)[Functions](#)[Books](#)[Blogs](#)[Related Tools](#)

검색결과 약 495,000,000개(0.29초)

### [Quick-R: Time Series](#)

R has extensive facilities for analyzing **time series** data. This section describes the creation of a **time series**, seasonal decomposition, modeling with exponential ...  
[www.statmethods.net/advstats/timeseries.html](http://www.statmethods.net/advstats/timeseries.html)

### [Stationary Time Series](#)

파일 형식: PDF/Adobe Acrobat  
**Stationary Time Series**. Alberto Lopez Moreno. TSTutorial version 1.2.1. A **Time Series** is stationary if has the following conditions: 1. Constant  $\mu$  (mean) for all t.  
[cran.r-project.org/web/packages/TSTutorial/.../Stationary.pdf](http://cran.r-project.org/web/packages/TSTutorial/.../Stationary.pdf)

### [Analysis of multivariate time-series using the MARSS package](#)

파일 형식: PDF/Adobe Acrobat  
Mar 19, 2014 ... versions of the workshops and the **time-series** analysis course on EEH's website ... and Steven Katz on analysis of marine plankton **time series**.  
[cran.r-project.org/web/packages/MARSS/vignettes/UserGuide.pdf](http://cran.r-project.org/web/packages/MARSS/vignettes/UserGuide.pdf)

### [Periodic Autoregressive Time Series Models in R: The partsm ...](#)

파일 형식: PDF/Adobe Acrobat  
Keywords: **Time series**, PAR models, periodic integration, R. 1. Introduction. It is well-known that some of the macroeconomic **time series** display stochastic ...  
[cran.r-project.org/web/packages/partsm/.../partsm.pdf](http://cran.r-project.org/web/packages/partsm/.../partsm.pdf)

### [Dealing With Stochastic Volatility in Time Series Using the R ...](#)

파일 형식: PDF/Adobe Acrobat  
tivity, financial **time series**. 1. Introduction. When analyzing (financial) returns, focus is often laid on estimating and predicting potentially time varying volatilities.  
[cran.r-project.org/web/packages/stochvol/vignettes/article.pdf](http://cran.r-project.org/web/packages/stochvol/vignettes/article.pdf)

### [Distributed lag linear and non-linear models for time series data](#)

파일 형식: PDF/Adobe Acrobat  
and DLNMs and the original software implementation for **time series** data are illustrated ... basis for the analysis of this data set or other **time series** data sources.  
[cran.r-project.org/web/packages/dlnm/vignettes/dlnmTS.pdf](http://cran.r-project.org/web/packages/dlnm/vignettes/dlnmTS.pdf)

### [The glarma Package for Observation Driven Time Series ...](#)

파일 형식: PDF/Adobe Acrobat  
erage observation driven models for **time series** of counts with explanatory variables and describe the estimation of these models using the glarma R- package.  
[cran.r-project.org/web/packages/glarma/.../glarma.pdf](http://cran.r-project.org/web/packages/glarma/.../glarma.pdf)

## 2. R이란 무엇인가?

Where to get R? -> Go to [www.r-project.org](http://www.r-project.org)



About R  
[What is R?](#)  
[Contributors](#)  
[Screenshots](#)  
[What's new?](#)

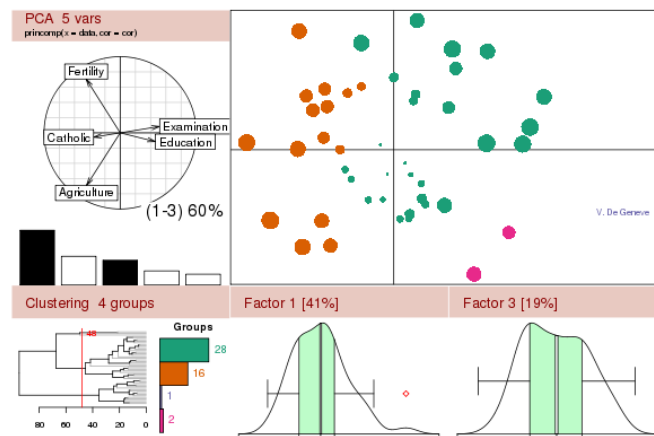
Download, Packages  
[CRAN](#)

R Project  
[Foundation](#)  
[Members & Donors](#)  
[Mailing Lists](#)  
[Bug Tracking](#)  
[Developer Page](#)  
[Conferences](#)  
[Search](#)

Documentation  
[Manuals](#)  
[FAQs](#)  
[The R Journal](#)  
[Wiki](#)  
[Books](#)  
[Certification](#)  
[Other](#)

Misc  
[Bioconductor](#)  
[Related Projects](#)  
[User Groups](#)  
[Links](#)

### The R Project for Statistical Computing



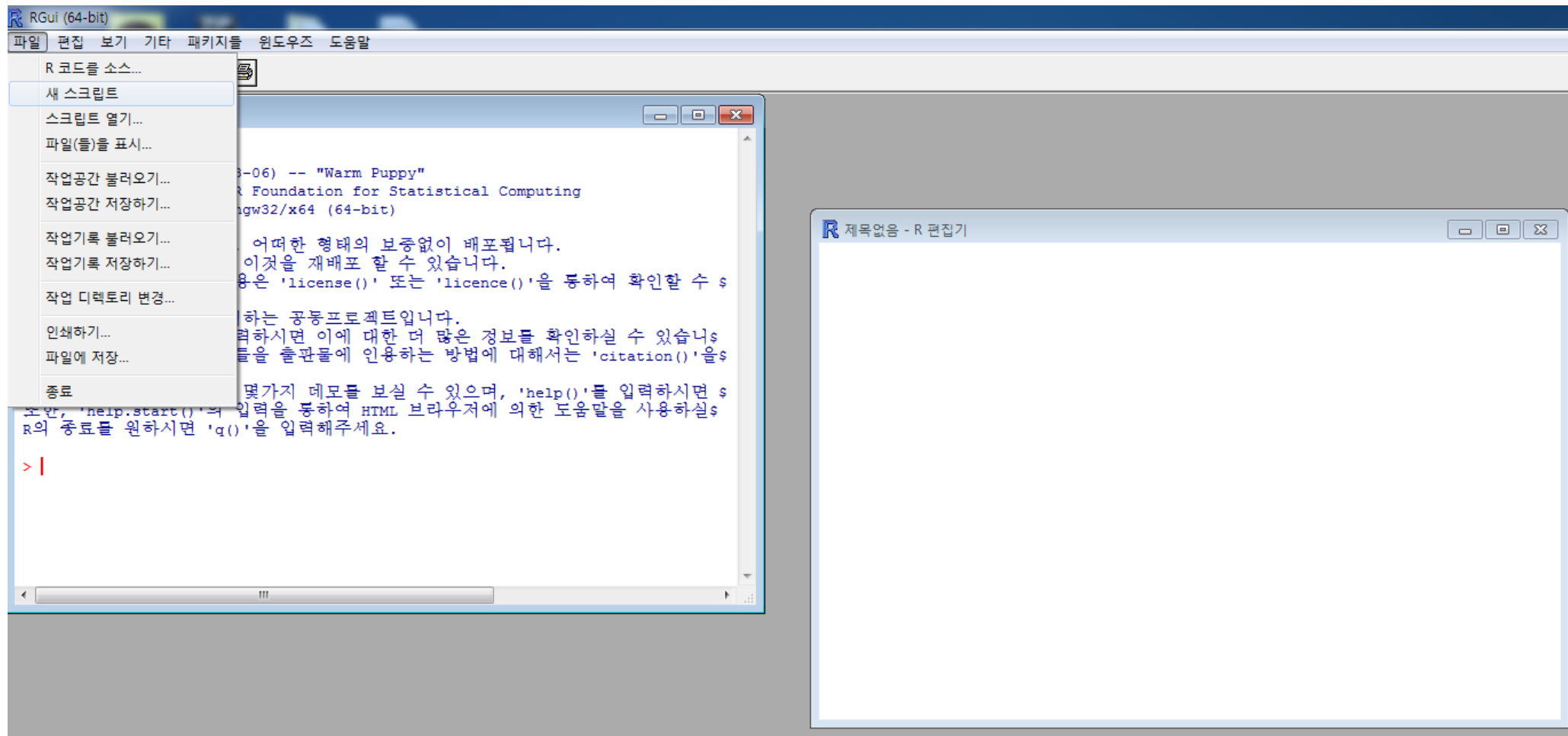
#### Getting Started:

- 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 3.1.0 (Spring Dance) has been released on 2014-04-10.
- R version 3.0.3 (Warm Puppy) has been released on 2014-03-06.
- [The R Journal Vol.5/2](#) is available.
- [useR! 2013](#), took place at the University of Castilla-La Mancha, Albacete, Spain, July 10-12 2013.
- R version 2.15.3 (Security Blanket) has been released on 2013-03-01.

## 2. R이란 무엇인가?





# 3. R의 기본 연산

- 3.1 단순계산
  - 덧셈, 뺄셈, 곱셈, 나눗셈, 제곱 등 간단한 수학 계산을 사용할 수 있다. (+, -, \*, /, ^)
- 3.2 함수를 이용한 계산

R 내장 수학 함수	내용
sqrt(x)	X의 제곱근
sin(x)	sin x
cos(x)	cos x
tan(x)	tan x
abs(x)	X의 절대값(absolute value)
log(x) log(x, base=a)	밑이 e인 log 밑이 a인 log
factorial(n)	N!

## 4. 데이터 벡터 만들기

- 4.1 c()를 이용하여 데이터 벡터 만들기

```
k_score=c(96,80,76,96,88,75,78,89,92,70)
k_score_2=c(67,83,96,90,85,75,82,89,92,75)
score=c(k_score,k_score_2)
```

(1) 숫자형 데이터

```
y=c(1,5,7)
```

(2) 문자형 데이터

```
family=c("kim","lee","chulsu","suhee")
names(family)=c("father","mother","son","daughter")
```

```
x=c("kim",2,3)
```

## 4. 데이터 벡터 만들기

- 4.1 c()를 이용하여 데이터 벡터 만들기

(3) 논리형 데이터

```
c(T,F,F,F,T,T,F)
```

```
x=-3:3
```

```
w=w<2
```

```
sum(w)
```

## 4. 데이터 벡터 만들기

- 4.2 데이터의 형태 변환

변환	변환 함수	변환 규칙
numeric	as.numeric	FALSE -> 0 THRE -> 1 "1","2",... -> 1,2,... "A" -> NA
logical	as.logical	0 -> FALSE 그 외 다른 수 -> TRUE "FALSE", "F" -> FALSE "TRUE", "T" -> TRUE
character	as.character	1,2,... -> "1", "2",... FALSE -> "FALSE" TRUE -> "TRUE"
factor	as.factor	범주형 factor형식으로

## 4. 데이터 벡터 만들기

- 4.2 데이터의 형태 변환

```
fact=c(1,2,3)
f.log=as.logical(fact)
```

```
character=as.character(factor)
numeric=as.numeric(character)
```

- 3.5 데이터의 벡터에 함수 적용하기

```
#합 ; sum(k_score)
```

```
#평균 ; mean(k_score)
```

```
#최댓값 ; max(k_score)
```

```
#최솟값 ; min(k_score)
```

```
#범위 ; range(k_score)
```

```
#분산 ; var(k_score)
```

```
#표준편차 ; sd(k_score)
```

```
#중앙값 ; median(k_score)
```

```
#벡터길이; length(k_score)
```

## 5. 데이터의 벡터에 함수 적용하기

- 5.1 데이터의 벡터에 함수 적용하기

apply() 함수는 행렬의 행과 열에 대해 원하는 함수를 적용하는 함수이다.

apply(data,dim,function)

#데이터 생성

x=cbind(x1=3,x2=c(4:1,2:8))

dimnames(x)[[1]]=letters[1:11]

#행별 합과, 열별 합

apply(x,1,sum)

apply(x,2,sum)

## 5. 데이터의 벡터에 함수 적용하기

- 5.2 필요한 데이터만 선택하기

MASS 패키지에 포함되어있는 cabbages 데이터 셋에서 필요한 데이터만 선택한다.

```
library(MASS)
```

```
#Date=d16인 경우의 표본
```

```
cabbages1=cabbages[(cabbages$Date=="d16"),]
```

```
#Vitamin C 함량이 50 이상인 경우의 표본
```

```
cabbages2=cabbages[(cabbages$VitC>=50),]
```

```
#Date=d16인 경우와 Vitamin C함량이 50인 경우의 표본
```

```
cabbages3=cabbages[(cabbages$Date=="d16")&(cabbages$VitC>=60),]
```

## 5. 데이터의 벡터에 함수 적용하기

- 5.3 순서 함수

데이터에 대해 특정 변수를 기준으로 순서대로 정렬하거나 배열한다.

```
id=c(1,2,3,4,5)
a=c(4,9,7,2,8)
da=cbind(id,a)
```

```
#sorting or order
a_s=sort(a)
da_s=da[order(a),]
```

```
#rank
x=c(11,24,24,30,30,30)
rank(x) #0.5씩 더해져서 순위값을 준다.
order(x)
```



## 6. 구조적인 데이터 만들기 및 다루기

- 6.1 구조적인 데이터 만들기

구조(structure) 또는 패턴(pattern)이 있는 데이터를 생성한다.

#1 부터 10까지 일련 번호를 생성하고자 하는 경우

1:10

seq(1,10)

#1 부터 10까지 역으로 번호를 생성하고자 할 경우

rev(1:10)

10:1

#1부터 10까지 2씩 증가한 번호를 생성하고자 할 경우

seq(1,10,by=2)

seq(from=1,to=10,by=2)

## 6. 구조적인 데이터 만들기 및 다루기

- 6.1 구조적인 데이터 만들기

# 1부터 3까지 3번 반복해서 생성할 경우

```
rep(1:3,time=3)
```

```
rep(1:3,3)
```

#1부터 3까지 각각 3번씩 반복해서 생성할 경우

```
rep(1:3,each=3)
```

#벡터나 행렬값의 초기화

```
a=rep(0,12)
```

## 6. 구조적인 데이터 만들기 및 다루기

- 6.1 구조적인 데이터 만들기

```
# 1부터 3까지 3번 반복해서 생성할 경우  
rep(1:3,time=3)  
rep(1:3,3)
```

```
#1부터 3까지 각각 3번씩 반복해서 생성할 경우  
rep(1:3,each=3)
```

```
#벡터나 행렬값의 초기화  
a=rep(0,12)
```

## 6. 구조적인 데이터 만들기 및 다루기

- 6.3 데이터 벡터 다루기

A 회사의 2007년 월별매출액이 다음과 같다.

100, 120, 130, 124, 150, 167, 170, 163, 160, 155, 145, 157

#원하는 위치의 벡터의 값 선택하기

`x[1]`

`x[1:3]`

`x[c(6,7)]`

`xm1=x[-1]`

`xm2=x[-c(1,12)]`

`xx=x[x!=150]`

#각 벡터에 일련번호를 준다.

`names(x)=seq(1,12)`

# 연습문제

- 연습문제

다음의 데이터를 사용하시오

2 3 5 7 9 10

- (1) 데이터 벡터  $x$ 를 만드시오.
- (2) 각 데이터의 제곱으로 구성된 벡터  $x_2$ 를 만드시오.
- (3) 각 데이터의 제곱의 합을 구하시오.
- (4) 각 데이터에서 2를 뺀 값들을 구하시오.
- (5) 최대값과 최소값을 구하시오.
- (6) 5보다 큰 값들만으로 구성된 데이터 벡터  $x_{up}$ 을 만드시오.
- (7) 벡터  $x$ 의 길이를 구하시오.
- (8) 3 5 6 2 1 9를 두 번째 열로 만드시오.

Q&A

