

R 프로그래밍 기초 Part1

한상곤(sangkon@pusan.ac.kr)

2023.06.13(화)

Contents

1 R 프로그래밍 관련 주요 Topic	1
1.1 R 프로그래밍 언어 소개	1
1.2 R의 주요 특징	1
1.3 작업 환경 구축하기	2
1.4 패키지와 Session 보기	2
1.5 R Session	2
1.6 작업 공간	3
1.7 데이터의 유형과 구조	3
1.8 Vector(벡터)(1차원)	3
1.9 List(중첩 자료구조)	4
1.10 apply 함수에 대해서	6
1.11 팩터(factor)	6
1.12 Matrix(2차원)	7
1.13 Array(다차원)	8
1.14 Data Frame(2차원 테이블 구조)	8
1.15 문자열 처리	9
1.16 데이터 불러오기 및 저장하기	9
2 데이터 분석 과정	10
3 예제를 사용한 데이터 분석 과정에 필요한 기초적인 함수	10
3.1 iris를 사용한 예제	10

세부적인 사항은 교재를 참고하세요. 해당 노트는 복습을 위해서 별도로 제작된 것입니다.

1 R 프로그래밍 관련 주요 Topic

1.1 R 프로그래밍 언어 소개

- 1993년 뉴질랜드 오클랜드 대학의 통계학과 교수인 Ross Ihaka, Robert Gentleman에 의해서 개발
- 통계 언어인 S의 계보를 잇고 있음
- R의 S+(S+PLUS)의 무료 버전으로 1993년 배포되어 현재는 GNU 프로젝트로 개발 및 개선

1.2 R의 주요 특징

- 통계분석(Statistical Analysis)과 시각화(Visualization)을 위한 공개 소프트웨어로 활용되고 있음
 - 풍부한 통계 패키지 제공
 - 다양한 시각화 도구를 제공

1.3 작업 환경 구축하기

- R 공식 사이트에서 무료로 다운로드 할 수 있음
 - 교재 p.18을 참고해서 진행
 - 패키지 설치를 위해서 Rtools 다운로드 후 설치
- RStudio 다운로드 후 설치

1.4 패키지와 Session 보기

- R 패키지는 통계학 관련 교수와 학자 그리고 관련 개발자에 의해서 꾸준히 관리
 - 대부분의 패키지는 무료로 사용할 수 있음

패키지는 `install.packages("패키지명")` 명령어를 사용해서 설치할 수 있다. RStudio의 Packages 탭을 사용해도 설치가 가능하기 때문에 두 개 중 자신에게 편한 것을 선택하자.

```
install.packages("stringr") # 패키지 설치
```

패키지를 사용하는 방법은 `library(패키지명)`, `require(패키지명)`으로 가능하다. 패키지가 없을 경우 library는 경고를 출력한다. 가능하면 library를 사용하자. 그리고 RStudio에서도 가능하다.

```
library(stringr)  
search() # 현재 사용하는 패키지 확인
```

패키지 삭제는 `remove.packages("패키지명")`으로 가능하다. RStudio에서도 가능하다.

1.5 R Session

R Session이란 사용자가 R 프로그램을 기동한 이후 R 콘솔 시작과 종료 전까지의 기간에 수행된 정보를 의미한다.

```
sessionInfo()
```

```
## R version 4.3.0 (2023-04-21 ucrt)  
## Platform: x86_64-w64-mingw32/x64 (64-bit)  
## Running under: Windows 11 x64 (build 22621)  
##  
## Matrix products: default  
##  
##  
## locale:  
## [1] LC_COLLATE=Korean_Korea.utf8  LC_CTYPE=Korean_Korea.utf8  
## [3] LC_MONETARY=Korean_Korea.utf8 LC_NUMERIC=C  
## [5] LC_TIME=Korean_Korea.utf8  
##  
## time zone: Asia/Seoul  
## tzcode source: internal  
##  
## attached base packages:  
## [1] stats      graphics   grDevices utils      datasets   methods    base  
##  
## other attached packages:  
## [1] stringr_1.5.0  
##  
## loaded via a namespace (and not attached):  
## [1] digest_0.6.31   fastmap_1.1.1   xfun_0.39       magrittr_2.0.3  
## [5] glue_1.6.2     knitr_1.43    htmltools_0.5.5 rmarkdown_2.22  
## [9] lifecycle_1.0.3 cli_3.6.1     compiler_4.3.0  rstudioapi_0.14  
## [13] tools_4.3.0    evaluate_0.21  yaml_2.3.7     rlang_1.1.1  
## [17] stringi_1.7.12
```

1.6 작업 공간

- `setwd()`, 지정된 작업공간을 변경하기 위해서 사용
- `getwd()`, 지정된 작업공간을 확인하기 위해서 사용

```
getwd()
```

```
## [1] "C:/Users/sigma/works/practice-r"
```

1.7 데이터의 유형과 구조

R에서 제공하는 5가지 주요 자료구조는 다음과 같다.

- Vector(1차원)
- List(중첩 자료구조)
- Matrix(2차원)
- Array(다차원)
- Data Frame(2차원 테이블 구조)

1.8 Vector(벡터)(1차원)

일반적인 벡터는 `c` 함수를 사용해서 생성할 수 있다.

```
c(1,2,3,4,5)
c(1:20)
seq(1, 10, 2)
rep(1:3, 3)
```

벡터는 동일한 자료형으로 생성한다.

```
v1 <- c(33, -5, 20:23, 12, -2:3)
v2 <- c("홍길동", "이순신", "유관순")
v3 <- c(T, TRUE, FALSE, T, TRUE, F, T)
v4 <- c(33, 05, 20:23, 12, "4") # 주의!
v1; v2; v3; v4
```

```
## [1] 33 -5 20 21 22 23 12 -2 -1 0 1 2 3
## [1] "홍길동" "이순신" "유관순"
## [1] TRUE TRUE FALSE TRUE TRUE FALSE TRUE
## [1] "33" "5" "20" "21" "22" "23" "12" "4"
```

벡터는 인덱스를 사용해서 벡터 참조 가능하다.

```
v1 <- c(13, -5, 20:23, 12, -2:3) # 생성
v1[1]
v1[c(2, 4)]
v1[c(3:5)]
v1[c(4, 5:8, 7)]
v1[-1]; v1[-c(2, 4)]; v1[-c(2:5)]; v1[-c(2, 5:10, 1)]
```

벡터에 사용 가능한 기본적인 함수는 아래와 같다.

```
vec <- 1:10
min(vec)
max(vec)
sum(vec)
range(vec) # 범위(m-M)
mean(vec) # 평균
```

```
median(vec) # 중앙  
sd(rnorm(10)) # 표준 편차  
table(vec) # 분할표
```

1.9 List(중첩 자료구조)

리스트의 경우 딱셔너리 형태뿐만 아니라 서로 다른 자료형 값들을 저장하고 다룰 수 있다.

- 리스트를 생성하는 방법은 매우 간단

```
list <- list("lee", "이순신", 95)
```

- 기존의 리스트를 벡터 구조로 변경하는 것도 가능

```
unlist <- unlist(list)
```

- 1개 이상의 값을 갖는 리스트 객체 생성

```
num <- list(c(1:5), c(6, 10))
```

- 리스트를 제대로 활용하는 방법은 key와 value 형식으로 리스트를 생성하는 것

```
member <- list(name = c("홍길동", "유관순"),  
                age = c(35, 25),  
                address = c("한양", "충남"),  
                gender = c("남자", "여자"),  
                htype = c("아파트", "오피스텔"))
```

```
member$name
```

```
## [1] "홍길동" "유관순"
```

```
member$name[1]
```

```
## [1] "홍길동"
```

```
member$name[2]
```

```
## [1] "유관순"
```

- 리스트의 key를 이용하여 value에 접근할 수 있음

```
member
```

```
## $name  
## [1] "홍길동" "유관순"  
##  
## $age  
## [1] 35 25  
##  
## $address  
## [1] "한양" "충남"  
##  
## $gender  
## [1] "남자" "여자"  
##  
## $htype  
## [1] "아파트" "오피스텔"
```

```
member$age[1] <- 45
```

```
member$id <- "hong"
```

```
member$pwd <- "1234"
```

```

member$age <- NULL
length(member)

## [1] 6
mode(member); class(member)

## [1] "list"
## [1] "list"


- 리스트 객체에 함수 적용하는 것도 유사


a <- list(c(1:5))
b <- list(c(6:10))
lapply(c(a, b), max)

## [[1]]
## [1] 5
##
## [[2]]
## [1] 10


- 결과값을 벡터형식으로 확인하기 위해서는 별도의 함수를 사용


sapply(c(a, b), max)

## [1] 5 10


- 다차원 리스트 객체 생성하는 방법은 내포된 리스트 형태


multi_list <- list(c1 = list(1, 2, 3),
                    c2 = list(10, 20, 30),
                    c3 = list(100, 200, 300))
multi_list$c1; multi_list$c2; multi_list$c3

## [[1]]
## [1] 1
##
## [[2]]
## [1] 2
##
## [[3]]
## [1] 3

## [[1]]
## [1] 10
##
## [[2]]
## [1] 20
##
## [[3]]
## [1] 30

## [[1]]
## [1] 100
##
## [[2]]
## [1] 200
##

```

```

## [[3]]
## [1] 300

• do.call을 사용해서 다차원 리스트를 열 단위로 바인딩할 수 있음

do.call(cbind, multi_list)

##      c1 c2 c3
## [1,]  1 10 100
## [2,]  2 20 200
## [3,]  3 30 300

class(do.call(cbind, multi_list))

## [1] "matrix" "array"

```

1.10 apply 함수에 대해서

데이터 분석의 특성상, 데이터를 다룰 때 원자별/그룹별로 함수를 적용할 경우가 많다. apply 계열의 함수는 데이터 구조를 갖는 R object 를 input 으로 받아 원소 별 혹은 그룹별로 함수를 적용시키는 것으로, input 과 output 의 형태에 따라 여러 종류로 나뉜다. - apply (input : array, output : array) - lapply (input : list or vector, output : list) - sapply (input : list or vector, output : vector or array) - vapply (input : list or vector, output : vector or array) - tapply (input : list or vector and factor, output : vector or array) - mapply (input : list or vector, output : vector or array)

1.11 팩터(factor)

팩터는 일종의 범주형 자료를 저장하는 데 사용한다. 범주형 자료란 성별, 혈액형 등 같은 것으로 팩터는 벡터를 만든 다음 factor 함수를 이용하여 변환한다.

```

bt <- c('A', 'B', 'C', 'B', 'C')
bt.new <- factor(bt)
bt

## [1] "A" "B" "C" "B" "C"

bt.new

## [1] A B C B C
## Levels: A B C

bt[5]

## [1] "C"

bt.new[5]

## [1] C
## Levels: A B C

levels(bt.new)

## [1] "A" "B" "C"
as.integer(bt.new)

## [1] 1 2 3 2 3
as.integer(bt)

## [1] NA NA NA NA NA

```

```

bt.new[6] <- 'B'
bt.new[8] <- 'A'
bt.new

## [1] A     B     C     B     C     B     <NA> A
## Levels: A B C

levels(bt.new)

## [1] "A" "B" "C"
as.integer(bt.new)

## [1] 1 2 3 2 3 2 NA 1

```

1.12 Matrix(2차원)

2차원 데이터는 테이블 형태로 표현할 수 있는데, 이때 사용되는 몇 가지 용어들이 있다. 데이터 테이블에서 가로줄 방향을 행(row) 또는 관측값(observation)이라고 부르고, 세로줄 방향을 열(column) 또는 변수(variable)이라고 부른다. 이 때, 행과 열의 상호 교차에 의해서 만들어지는 사각형 영역을 셀(cell)이라고 부른다.

모든 경우에 통용되는 단어는 아니지만, 일반적인 경우 모든 셀 값들이 동일한 자료형일 때는 매트릭스를 사용하고, 자료형이 서로 다른 값들로 구성된 셀로 데이터 형태가 구성된다면 데이터프레임을 사용한다.

- 행렬은 2차원 배열의 구조를 가지며, 벡터를 활용해서 행렬 객체를 생성할 수 있음

```
m <- matrix(c(1:5))
```

- 벡터를 사용해서 열 우선으로 행렬 객체를 생성

```
m <- matrix(c(1:10), nrow = 2)
```

- 행과 열의 수가 일치하지 않는 경우는 경우 부족한 데이터는 첫번째 데이터를 재사용

```
m <- matrix(c(1:11), nrow = 2)
```

- 행 우선으로 데이터를 생성할 경우 가능하면 byrow를 사용해서 가독성을 확보

```
m <- matrix(c(1:10), nrow = 2, byrow = T)
```

- 기준에 만들어준 벡터를 활용해서 행렬을 생성

```
x1 <- c(m, 40, 50:52)
```

```
x2 <- c(30, 5, 6:8)
```

```
mr <- rbind(x1, x2) # 행방향
```

- 열을 기준으로 행렬을 생성할 수 있음

```
mc <- cbind(x1, x2) # 열방향
```

- 행렬 생성 함수를 사용해서 행렬 객체를 생성할 수 있음

```
m3 <- matrix(10:19, 2)
```

```
m4 <- matrix(10:20, 2)
```

```
m3; m4
```

```
mode(m3); class(m3)
```

- 인덱스(첨자)를 사용하여 행렬 객체에 접근할 수 있음

```
m3[1, ]
```

```
m3[ , 5]
```

```
m3[2, 3]
```

```
m3[1, c(2:5)]
```

- 3행 3열의 행렬 객체 생성하는 방법인 행과 열의 크기를 정하는 것

```
x <- matrix(c(1:9), nrow = 3, ncol = 3)
```

- 행렬의 크기와 형태를 확인하는 방법은 length와 ncol을 사용할 수 있음

```
length(x)
ncol(x)
nrow(x)
```

행렬에서 *MARGIN*이 1인 경우 행, 2인 경우 열을 나타낸다. 다른 언어와 혼동하지 않도록 주의하자.

- apply를 사용해서 함수나 사용자 정의 함수를 적용할 수 있음

```
apply(x, 1, max)
apply(x, 1, min)
apply(x, 1, mean)
apply(x, 2, max)
apply(x, 2, min)
apply(x, 2, mean)
```

- 사용자 정의 함수도 가능

```
f <- function(x) {
  x * c(1, 2, 3)
}
result <- apply(x, 1, f)

result <- apply(x, 2, f) # 1은 행, 2는 열
```

- 행렬 객체에 칼럼명을 지정할 수 있음

```
colnames(x) <- c("one", "two", "three")
```

1.13 Array(다차원)

- 배열을 생성하는 방법은 array를 활용하는 것

```
vec <- c(1:12)
arr <- array(vec, c(3, 2, 2))
```

- 배열 객체의 자료 조회하는 방법은 행렬과 유사

```
arr[ , , 1]
arr[ , , 2]
```

1.14 Data Frame(2차원 테이블 구조)

- 데이터 프레임은 열(!) 단위로 서로 다른 자료형을 포함할 수 있음

```
no <- c(1, 2, 3)
name <- c("hong", "lee", "kim")
pay <- c(150, 250, 300)
vemp <- data.frame(No = no, Name = name, Pay = pay)
vemp
```

```
##   No Name Pay
## 1   1 hong 150
## 2   2 lee 250
```

```

## 3 3 kim 300
m <- matrix(
  c(1, "hong", 150,
    2, "lee", 250,
    3, "kim", 300), 3, by = T)
memp <- data.frame(m)
memp

```

```

## X1 X2 X3
## 1 1 hong 150
## 2 2 lee 250
## 3 3 kim 300

```

- 텍스트 파일을 활용해서 데이터프레임 객체를 생성

```
txtemp <- read.table('~/data/emp.txt', header = 1, sep = "", fileEncoding = "euc-kr")
```

- CSV 파일도 가능

```

name <- c("사번", "이름", "급여")
csvtmp <- read.csv('~/data/emp.csv', header = F, col.names = name, fileEncoding = "euc-kr")

```

- 데이터 프레임을 생성하는 간단한 예제

```

df <- data.frame(x = c(1:5), y = seq(2, 10, 2), z = c('a', 'b', 'c', 'd', 'e'))
df$x
str(df)
ncol(df)
nrow(df)
names(df)
df[c(2:3), 1]
summary(df)

```

- 데이터프레임 자료에 함수 적용하는 방법도 유사

```
apply(df[, c(1, 2)], 2, sum)
```

- 데이터프레임의 부분 객체 만들기는 방법은 subset을 활용

```

x1 <- subset(df, x >= 3)
y1 <- subset(df, y <= 8)
xyand <- subset(df, x >= 2 & y <= 6)
xyor <- subset(df, x >= 2 | y <= 6)

```

- 데이터 프레임 병합은 merge를 사용

```

height <- data.frame(id = c(1, 2), h = c(180, 175))
weight <- data.frame(id = c(1, 2), w = c(80, 75))
user <- merge(height, weight, by.x = "id", by.y = "id")
user

```

1.15 문자열 처리

- 문자열 처리는 교재를 참고

1.16 데이터 불러오기 및 저장하기

- 교재를 참고

2 데이터 분석 과정

데이터를 분석하는 과정을 간단하게 요약하였다. 해당 과정은 일반화된 형태로 개별 데이터 분석 절차와 차이가 날 수 있지만, 일반적인 상황에서 데이터 분석을 하는 전형적인 과정을 소개한다.

1. 문제 정의(problem definition) 및 계획 모든 데이터 분석은 문제를 명확하게 정의하는 것으로 시작한다. 데이터 분석의 문제 정의가 명확해야 그 문제를 해결하기 위한 데이터가 어떤 것인지 추정할 수 있고, 어떤 기법을 적용해야 할지 계획할 수 있다.
2. 데이터 수집(data acquisition) 문제 해결에 필요한 데이터를 파악하고 관련 데이터를 수집하는 것으로, 대부분의 경우 여러 경로를 통해서 데이터를 수집해야 하는 경우가 많다. 많은 경우 해당 데이터가 문서 형태로 기록된 것이 많기 때문에 데이터 수집 비용의 경우 비용 및 시간을 고려해서 진행해야 한다.
3. 데이터 정제 및 전처리 데이터 수집 후 데이터 정제 및 전처리를 진행한다. 이 때 정제 및 전처리를 진행하는 대표적인 이유는 데이터에 사용된 단위가 단일하지 않기 때문이다. 이와 같이 데이터를 분석하는데 있어서 데이터의 일관성을 유지하기 위해서 정제 및 전처리가 필요하다. 또한 이상치나 오류 등이 포함되어 있으므로 정제와 전처리를 진행해야 한다.
4. 데이터 탐색(data exploration) 가벼운 데이터 분석 단계이다. 정돈된 데이터를 이해하고, 수집된 데이터에 대한 지표 등을 파악하는 단계이다. 탐색 단계를 거쳐서 분석 과정에 사용될 기법 등을 결정하는 경우가 많다.
5. 데이터 분석(data analysis) 탐색 단계에서 파악한 정보를 바탕으로 심화된 분석을 수행한다. 전통적인 통계 기법을 포함하여 머신러닝, 딥러닝 등을 활용해서 데이터를 분석하고 해당 분석 결과에 대한 해석(interpretation) 과정을 진행한다.
6. 결과 보고 분석과 해석이 종료되면 보고서를 작성한다. 결과 보고를 위해서 기준에 탐색과 분석에서 사용된 시각화 자료를 바탕으로 결과를 해석하고 이해하기 쉽도록 보고서를 작성한다.

3 예제를 사용한 데이터 분석 과정에 필요한 기초적인 함수

3.1 iris를 사용한 예제

```
data(iris)  
iris
```

##	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
## 1	5.1	3.5	1.4	0.2	setosa
## 2	4.9	3.0	1.4	0.2	setosa
## 3	4.7	3.2	1.3	0.2	setosa
## 4	4.6	3.1	1.5	0.2	setosa
## 5	5.0	3.6	1.4	0.2	setosa
## 6	5.4	3.9	1.7	0.4	setosa
## 7	4.6	3.4	1.4	0.3	setosa
## 8	5.0	3.4	1.5	0.2	setosa
## 9	4.4	2.9	1.4	0.2	setosa
## 10	4.9	3.1	1.5	0.1	setosa
## 11	5.4	3.7	1.5	0.2	setosa
## 12	4.8	3.4	1.6	0.2	setosa
## 13	4.8	3.0	1.4	0.1	setosa
## 14	4.3	3.0	1.1	0.1	setosa
## 15	5.8	4.0	1.2	0.2	setosa
## 16	5.7	4.4	1.5	0.4	setosa
## 17	5.4	3.9	1.3	0.4	setosa
## 18	5.1	3.5	1.4	0.3	setosa
## 19	5.7	3.8	1.7	0.3	setosa
## 20	5.1	3.8	1.5	0.3	setosa

## 21	5.4	3.4	1.7	0.2	setosa
## 22	5.1	3.7	1.5	0.4	setosa
## 23	4.6	3.6	1.0	0.2	setosa
## 24	5.1	3.3	1.7	0.5	setosa
## 25	4.8	3.4	1.9	0.2	setosa
## 26	5.0	3.0	1.6	0.2	setosa
## 27	5.0	3.4	1.6	0.4	setosa
## 28	5.2	3.5	1.5	0.2	setosa
## 29	5.2	3.4	1.4	0.2	setosa
## 30	4.7	3.2	1.6	0.2	setosa
## 31	4.8	3.1	1.6	0.2	setosa
## 32	5.4	3.4	1.5	0.4	setosa
## 33	5.2	4.1	1.5	0.1	setosa
## 34	5.5	4.2	1.4	0.2	setosa
## 35	4.9	3.1	1.5	0.2	setosa
## 36	5.0	3.2	1.2	0.2	setosa
## 37	5.5	3.5	1.3	0.2	setosa
## 38	4.9	3.6	1.4	0.1	setosa
## 39	4.4	3.0	1.3	0.2	setosa
## 40	5.1	3.4	1.5	0.2	setosa
## 41	5.0	3.5	1.3	0.3	setosa
## 42	4.5	2.3	1.3	0.3	setosa
## 43	4.4	3.2	1.3	0.2	setosa
## 44	5.0	3.5	1.6	0.6	setosa
## 45	5.1	3.8	1.9	0.4	setosa
## 46	4.8	3.0	1.4	0.3	setosa
## 47	5.1	3.8	1.6	0.2	setosa
## 48	4.6	3.2	1.4	0.2	setosa
## 49	5.3	3.7	1.5	0.2	setosa
## 50	5.0	3.3	1.4	0.2	setosa
## 51	7.0	3.2	4.7	1.4	versicolor
## 52	6.4	3.2	4.5	1.5	versicolor
## 53	6.9	3.1	4.9	1.5	versicolor
## 54	5.5	2.3	4.0	1.3	versicolor
## 55	6.5	2.8	4.6	1.5	versicolor
## 56	5.7	2.8	4.5	1.3	versicolor
## 57	6.3	3.3	4.7	1.6	versicolor
## 58	4.9	2.4	3.3	1.0	versicolor
## 59	6.6	2.9	4.6	1.3	versicolor
## 60	5.2	2.7	3.9	1.4	versicolor
## 61	5.0	2.0	3.5	1.0	versicolor
## 62	5.9	3.0	4.2	1.5	versicolor
## 63	6.0	2.2	4.0	1.0	versicolor
## 64	6.1	2.9	4.7	1.4	versicolor
## 65	5.6	2.9	3.6	1.3	versicolor
## 66	6.7	3.1	4.4	1.4	versicolor
## 67	5.6	3.0	4.5	1.5	versicolor
## 68	5.8	2.7	4.1	1.0	versicolor
## 69	6.2	2.2	4.5	1.5	versicolor
## 70	5.6	2.5	3.9	1.1	versicolor
## 71	5.9	3.2	4.8	1.8	versicolor
## 72	6.1	2.8	4.0	1.3	versicolor
## 73	6.3	2.5	4.9	1.5	versicolor
## 74	6.1	2.8	4.7	1.2	versicolor

## 75	6.4	2.9	4.3	1.3 versicolor
## 76	6.6	3.0	4.4	1.4 versicolor
## 77	6.8	2.8	4.8	1.4 versicolor
## 78	6.7	3.0	5.0	1.7 versicolor
## 79	6.0	2.9	4.5	1.5 versicolor
## 80	5.7	2.6	3.5	1.0 versicolor
## 81	5.5	2.4	3.8	1.1 versicolor
## 82	5.5	2.4	3.7	1.0 versicolor
## 83	5.8	2.7	3.9	1.2 versicolor
## 84	6.0	2.7	5.1	1.6 versicolor
## 85	5.4	3.0	4.5	1.5 versicolor
## 86	6.0	3.4	4.5	1.6 versicolor
## 87	6.7	3.1	4.7	1.5 versicolor
## 88	6.3	2.3	4.4	1.3 versicolor
## 89	5.6	3.0	4.1	1.3 versicolor
## 90	5.5	2.5	4.0	1.3 versicolor
## 91	5.5	2.6	4.4	1.2 versicolor
## 92	6.1	3.0	4.6	1.4 versicolor
## 93	5.8	2.6	4.0	1.2 versicolor
## 94	5.0	2.3	3.3	1.0 versicolor
## 95	5.6	2.7	4.2	1.3 versicolor
## 96	5.7	3.0	4.2	1.2 versicolor
## 97	5.7	2.9	4.2	1.3 versicolor
## 98	6.2	2.9	4.3	1.3 versicolor
## 99	5.1	2.5	3.0	1.1 versicolor
## 100	5.7	2.8	4.1	1.3 versicolor
## 101	6.3	3.3	6.0	2.5 virginica
## 102	5.8	2.7	5.1	1.9 virginica
## 103	7.1	3.0	5.9	2.1 virginica
## 104	6.3	2.9	5.6	1.8 virginica
## 105	6.5	3.0	5.8	2.2 virginica
## 106	7.6	3.0	6.6	2.1 virginica
## 107	4.9	2.5	4.5	1.7 virginica
## 108	7.3	2.9	6.3	1.8 virginica
## 109	6.7	2.5	5.8	1.8 virginica
## 110	7.2	3.6	6.1	2.5 virginica
## 111	6.5	3.2	5.1	2.0 virginica
## 112	6.4	2.7	5.3	1.9 virginica
## 113	6.8	3.0	5.5	2.1 virginica
## 114	5.7	2.5	5.0	2.0 virginica
## 115	5.8	2.8	5.1	2.4 virginica
## 116	6.4	3.2	5.3	2.3 virginica
## 117	6.5	3.0	5.5	1.8 virginica
## 118	7.7	3.8	6.7	2.2 virginica
## 119	7.7	2.6	6.9	2.3 virginica
## 120	6.0	2.2	5.0	1.5 virginica
## 121	6.9	3.2	5.7	2.3 virginica
## 122	5.6	2.8	4.9	2.0 virginica
## 123	7.7	2.8	6.7	2.0 virginica
## 124	6.3	2.7	4.9	1.8 virginica
## 125	6.7	3.3	5.7	2.1 virginica
## 126	7.2	3.2	6.0	1.8 virginica
## 127	6.2	2.8	4.8	1.8 virginica
## 128	6.1	3.0	4.9	1.8 virginica

```

## 129      6.4      2.8      5.6      2.1  virginica
## 130      7.2      3.0      5.8      1.6  virginica
## 131      7.4      2.8      6.1      1.9  virginica
## 132      7.9      3.8      6.4      2.0  virginica
## 133      6.4      2.8      5.6      2.2  virginica
## 134      6.3      2.8      5.1      1.5  virginica
## 135      6.1      2.6      5.6      1.4  virginica
## 136      7.7      3.0      6.1      2.3  virginica
## 137      6.3      3.4      5.6      2.4  virginica
## 138      6.4      3.1      5.5      1.8  virginica
## 139      6.0      3.0      4.8      1.8  virginica
## 140      6.9      3.1      5.4      2.1  virginica
## 141      6.7      3.1      5.6      2.4  virginica
## 142      6.9      3.1      5.1      2.3  virginica
## 143      5.8      2.7      5.1      1.9  virginica
## 144      6.8      3.2      5.9      2.3  virginica
## 145      6.7      3.3      5.7      2.5  virginica
## 146      6.7      3.0      5.2      2.3  virginica
## 147      6.3      2.5      5.0      1.9  virginica
## 148      6.5      3.0      5.2      2.0  virginica
## 149      6.2      3.4      5.4      2.3  virginica
## 150      5.9      3.0      5.1      1.8  virginica

```

iris 각 열의 이름별 의미와 자료형은 아래와 같다. - Sepal.Length, 꽃받침의 길이, 숫자 - Sepal.Width, 꽃받침의 폭, 숫자 - Petal.Length, 꽃잎의 길이, 숫자 - Petal.Width, 꽃잎의 폭, 숫자 - Species, 붓꽃의 품종, 문자(Factor)

- 먼저 데이터의 전체 규모와 해당 데이터의 특징을 알아보자.

```

dim(iris)

## [1] 150   5

nrow(iris)

## [1] 150

ncol(iris)

## [1] 5

colnames(iris)

## [1] "Sepal.Length" "Sepal.Width"  "Petal.Length" "Petal.Width"  "Species"

head(iris)

##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1      5.1      3.5       1.4       0.2   setosa
## 2      4.9      3.0       1.4       0.2   setosa
## 3      4.7      3.2       1.3       0.2   setosa
## 4      4.6      3.1       1.5       0.2   setosa
## 5      5.0      3.6       1.4       0.2   setosa
## 6      5.4      3.9       1.7       0.4   setosa

tail(iris)

##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 145       6.7       3.3       5.7       2.5 virginica
## 146       6.7       3.0       5.2       2.3 virginica

```

```

## 147      6.3      2.5      5.0      1.9 virginica
## 148      6.5      3.0      5.2      2.0 virginica
## 149      6.2      3.4      5.4      2.3 virginica
## 150      5.9      3.0      5.1      1.8 virginica
str(iris)

## 'data.frame':   150 obs. of  5 variables:
## $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
## $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
## $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
## $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
## $ Species     : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...
unique(iris[,5])

## [1] setosa      versicolor virginica
## Levels: setosa versicolor virginica
table(iris[, "Species"])

```

```

##
##      setosa versicolor virginica
##      50      50      50

```

1. 데이터를 조작하는 방법은 아래와 같다.

```
iris[,c(1:2)] # 1~2열의 모든 데이터
```

```

##      Sepal.Length Sepal.Width
## 1          5.1        3.5
## 2          4.9        3.0
## 3          4.7        3.2
## 4          4.6        3.1
## 5          5.0        3.6
## 6          5.4        3.9
## 7          4.6        3.4
## 8          5.0        3.4
## 9          4.4        2.9
## 10         4.9        3.1
## 11         5.4        3.7
## 12         4.8        3.4
## 13         4.8        3.0
## 14         4.3        3.0
## 15         5.8        4.0
## 16         5.7        4.4
## 17         5.4        3.9
## 18         5.1        3.5
## 19         5.7        3.8
## 20         5.1        3.8
## 21         5.4        3.4
## 22         5.1        3.7
## 23         4.6        3.6
## 24         5.1        3.3
## 25         4.8        3.4
## 26         5.0        3.0
## 27         5.0        3.4

```

## 28	5.2	3.5
## 29	5.2	3.4
## 30	4.7	3.2
## 31	4.8	3.1
## 32	5.4	3.4
## 33	5.2	4.1
## 34	5.5	4.2
## 35	4.9	3.1
## 36	5.0	3.2
## 37	5.5	3.5
## 38	4.9	3.6
## 39	4.4	3.0
## 40	5.1	3.4
## 41	5.0	3.5
## 42	4.5	2.3
## 43	4.4	3.2
## 44	5.0	3.5
## 45	5.1	3.8
## 46	4.8	3.0
## 47	5.1	3.8
## 48	4.6	3.2
## 49	5.3	3.7
## 50	5.0	3.3
## 51	7.0	3.2
## 52	6.4	3.2
## 53	6.9	3.1
## 54	5.5	2.3
## 55	6.5	2.8
## 56	5.7	2.8
## 57	6.3	3.3
## 58	4.9	2.4
## 59	6.6	2.9
## 60	5.2	2.7
## 61	5.0	2.0
## 62	5.9	3.0
## 63	6.0	2.2
## 64	6.1	2.9
## 65	5.6	2.9
## 66	6.7	3.1
## 67	5.6	3.0
## 68	5.8	2.7
## 69	6.2	2.2
## 70	5.6	2.5
## 71	5.9	3.2
## 72	6.1	2.8
## 73	6.3	2.5
## 74	6.1	2.8
## 75	6.4	2.9
## 76	6.6	3.0
## 77	6.8	2.8
## 78	6.7	3.0
## 79	6.0	2.9
## 80	5.7	2.6
## 81	5.5	2.4

## 82	5.5	2.4
## 83	5.8	2.7
## 84	6.0	2.7
## 85	5.4	3.0
## 86	6.0	3.4
## 87	6.7	3.1
## 88	6.3	2.3
## 89	5.6	3.0
## 90	5.5	2.5
## 91	5.5	2.6
## 92	6.1	3.0
## 93	5.8	2.6
## 94	5.0	2.3
## 95	5.6	2.7
## 96	5.7	3.0
## 97	5.7	2.9
## 98	6.2	2.9
## 99	5.1	2.5
## 100	5.7	2.8
## 101	6.3	3.3
## 102	5.8	2.7
## 103	7.1	3.0
## 104	6.3	2.9
## 105	6.5	3.0
## 106	7.6	3.0
## 107	4.9	2.5
## 108	7.3	2.9
## 109	6.7	2.5
## 110	7.2	3.6
## 111	6.5	3.2
## 112	6.4	2.7
## 113	6.8	3.0
## 114	5.7	2.5
## 115	5.8	2.8
## 116	6.4	3.2
## 117	6.5	3.0
## 118	7.7	3.8
## 119	7.7	2.6
## 120	6.0	2.2
## 121	6.9	3.2
## 122	5.6	2.8
## 123	7.7	2.8
## 124	6.3	2.7
## 125	6.7	3.3
## 126	7.2	3.2
## 127	6.2	2.8
## 128	6.1	3.0
## 129	6.4	2.8
## 130	7.2	3.0
## 131	7.4	2.8
## 132	7.9	3.8
## 133	6.4	2.8
## 134	6.3	2.8
## 135	6.1	2.6

```

## 136      7.7      3.0
## 137      6.3      3.4
## 138      6.4      3.1
## 139      6.0      3.0
## 140      6.9      3.1
## 141      6.7      3.1
## 142      6.9      3.1
## 143      5.8      2.7
## 144      6.8      3.2
## 145      6.7      3.3
## 146      6.7      3.0
## 147      6.3      2.5
## 148      6.5      3.0
## 149      6.2      3.4
## 150      5.9      3.0

```

```
iris[,c(1,3,5)] # 1,3,5열의 모든 데이터
```

```

##   Sepal.Length Petal.Length Species
## 1      5.1        1.4  setosa
## 2      4.9        1.4  setosa
## 3      4.7        1.3  setosa
## 4      4.6        1.5  setosa
## 5      5.0        1.4  setosa
## 6      5.4        1.7  setosa
## 7      4.6        1.4  setosa
## 8      5.0        1.5  setosa
## 9      4.4        1.4  setosa
## 10     4.9        1.5  setosa
## 11     5.4        1.5  setosa
## 12     4.8        1.6  setosa
## 13     4.8        1.4  setosa
## 14     4.3        1.1  setosa
## 15     5.8        1.2  setosa
## 16     5.7        1.5  setosa
## 17     5.4        1.3  setosa
## 18     5.1        1.4  setosa
## 19     5.7        1.7  setosa
## 20     5.1        1.5  setosa
## 21     5.4        1.7  setosa
## 22     5.1        1.5  setosa
## 23     4.6        1.0  setosa
## 24     5.1        1.7  setosa
## 25     4.8        1.9  setosa
## 26     5.0        1.6  setosa
## 27     5.0        1.6  setosa
## 28     5.2        1.5  setosa
## 29     5.2        1.4  setosa
## 30     4.7        1.6  setosa
## 31     4.8        1.6  setosa
## 32     5.4        1.5  setosa
## 33     5.2        1.5  setosa
## 34     5.5        1.4  setosa
## 35     4.9        1.5  setosa
## 36     5.0        1.2  setosa

```

## 37	5.5	1.3	setosa
## 38	4.9	1.4	setosa
## 39	4.4	1.3	setosa
## 40	5.1	1.5	setosa
## 41	5.0	1.3	setosa
## 42	4.5	1.3	setosa
## 43	4.4	1.3	setosa
## 44	5.0	1.6	setosa
## 45	5.1	1.9	setosa
## 46	4.8	1.4	setosa
## 47	5.1	1.6	setosa
## 48	4.6	1.4	setosa
## 49	5.3	1.5	setosa
## 50	5.0	1.4	setosa
## 51	7.0	4.7	versicolor
## 52	6.4	4.5	versicolor
## 53	6.9	4.9	versicolor
## 54	5.5	4.0	versicolor
## 55	6.5	4.6	versicolor
## 56	5.7	4.5	versicolor
## 57	6.3	4.7	versicolor
## 58	4.9	3.3	versicolor
## 59	6.6	4.6	versicolor
## 60	5.2	3.9	versicolor
## 61	5.0	3.5	versicolor
## 62	5.9	4.2	versicolor
## 63	6.0	4.0	versicolor
## 64	6.1	4.7	versicolor
## 65	5.6	3.6	versicolor
## 66	6.7	4.4	versicolor
## 67	5.6	4.5	versicolor
## 68	5.8	4.1	versicolor
## 69	6.2	4.5	versicolor
## 70	5.6	3.9	versicolor
## 71	5.9	4.8	versicolor
## 72	6.1	4.0	versicolor
## 73	6.3	4.9	versicolor
## 74	6.1	4.7	versicolor
## 75	6.4	4.3	versicolor
## 76	6.6	4.4	versicolor
## 77	6.8	4.8	versicolor
## 78	6.7	5.0	versicolor
## 79	6.0	4.5	versicolor
## 80	5.7	3.5	versicolor
## 81	5.5	3.8	versicolor
## 82	5.5	3.7	versicolor
## 83	5.8	3.9	versicolor
## 84	6.0	5.1	versicolor
## 85	5.4	4.5	versicolor
## 86	6.0	4.5	versicolor
## 87	6.7	4.7	versicolor
## 88	6.3	4.4	versicolor
## 89	5.6	4.1	versicolor
## 90	5.5	4.0	versicolor

## 91	5.5	4.4 versicolor
## 92	6.1	4.6 versicolor
## 93	5.8	4.0 versicolor
## 94	5.0	3.3 versicolor
## 95	5.6	4.2 versicolor
## 96	5.7	4.2 versicolor
## 97	5.7	4.2 versicolor
## 98	6.2	4.3 versicolor
## 99	5.1	3.0 versicolor
## 100	5.7	4.1 versicolor
## 101	6.3	6.0 virginica
## 102	5.8	5.1 virginica
## 103	7.1	5.9 virginica
## 104	6.3	5.6 virginica
## 105	6.5	5.8 virginica
## 106	7.6	6.6 virginica
## 107	4.9	4.5 virginica
## 108	7.3	6.3 virginica
## 109	6.7	5.8 virginica
## 110	7.2	6.1 virginica
## 111	6.5	5.1 virginica
## 112	6.4	5.3 virginica
## 113	6.8	5.5 virginica
## 114	5.7	5.0 virginica
## 115	5.8	5.1 virginica
## 116	6.4	5.3 virginica
## 117	6.5	5.5 virginica
## 118	7.7	6.7 virginica
## 119	7.7	6.9 virginica
## 120	6.0	5.0 virginica
## 121	6.9	5.7 virginica
## 122	5.6	4.9 virginica
## 123	7.7	6.7 virginica
## 124	6.3	4.9 virginica
## 125	6.7	5.7 virginica
## 126	7.2	6.0 virginica
## 127	6.2	4.8 virginica
## 128	6.1	4.9 virginica
## 129	6.4	5.6 virginica
## 130	7.2	5.8 virginica
## 131	7.4	6.1 virginica
## 132	7.9	6.4 virginica
## 133	6.4	5.6 virginica
## 134	6.3	5.1 virginica
## 135	6.1	5.6 virginica
## 136	7.7	6.1 virginica
## 137	6.3	5.6 virginica
## 138	6.4	5.5 virginica
## 139	6.0	4.8 virginica
## 140	6.9	5.4 virginica
## 141	6.7	5.6 virginica
## 142	6.9	5.1 virginica
## 143	5.8	5.1 virginica
## 144	6.8	5.9 virginica

```

## 145      6.7      5.7  virginica
## 146      6.7      5.2  virginica
## 147      6.3      5.0  virginica
## 148      6.5      5.2  virginica
## 149      6.2      5.4  virginica
## 150      5.9      5.1  virginica

```

```
iris[,c('Sepal.Length', 'Sepal.Width')] # 1,3,5열의 모든 데이터
```

```

##   Sepal.Length Sepal.Width
## 1      5.1       3.5
## 2      4.9       3.0
## 3      4.7       3.2
## 4      4.6       3.1
## 5      5.0       3.6
## 6      5.4       3.9
## 7      4.6       3.4
## 8      5.0       3.4
## 9      4.4       2.9
## 10     4.9       3.1
## 11     5.4       3.7
## 12     4.8       3.4
## 13     4.8       3.0
## 14     4.3       3.0
## 15     5.8       4.0
## 16     5.7       4.4
## 17     5.4       3.9
## 18     5.1       3.5
## 19     5.7       3.8
## 20     5.1       3.8
## 21     5.4       3.4
## 22     5.1       3.7
## 23     4.6       3.6
## 24     5.1       3.3
## 25     4.8       3.4
## 26     5.0       3.0
## 27     5.0       3.4
## 28     5.2       3.5
## 29     5.2       3.4
## 30     4.7       3.2
## 31     4.8       3.1
## 32     5.4       3.4
## 33     5.2       4.1
## 34     5.5       4.2
## 35     4.9       3.1
## 36     5.0       3.2
## 37     5.5       3.5
## 38     4.9       3.6
## 39     4.4       3.0
## 40     5.1       3.4
## 41     5.0       3.5
## 42     4.5       2.3
## 43     4.4       3.2
## 44     5.0       3.5
## 45     5.1       3.8

```

## 46	4.8	3.0
## 47	5.1	3.8
## 48	4.6	3.2
## 49	5.3	3.7
## 50	5.0	3.3
## 51	7.0	3.2
## 52	6.4	3.2
## 53	6.9	3.1
## 54	5.5	2.3
## 55	6.5	2.8
## 56	5.7	2.8
## 57	6.3	3.3
## 58	4.9	2.4
## 59	6.6	2.9
## 60	5.2	2.7
## 61	5.0	2.0
## 62	5.9	3.0
## 63	6.0	2.2
## 64	6.1	2.9
## 65	5.6	2.9
## 66	6.7	3.1
## 67	5.6	3.0
## 68	5.8	2.7
## 69	6.2	2.2
## 70	5.6	2.5
## 71	5.9	3.2
## 72	6.1	2.8
## 73	6.3	2.5
## 74	6.1	2.8
## 75	6.4	2.9
## 76	6.6	3.0
## 77	6.8	2.8
## 78	6.7	3.0
## 79	6.0	2.9
## 80	5.7	2.6
## 81	5.5	2.4
## 82	5.5	2.4
## 83	5.8	2.7
## 84	6.0	2.7
## 85	5.4	3.0
## 86	6.0	3.4
## 87	6.7	3.1
## 88	6.3	2.3
## 89	5.6	3.0
## 90	5.5	2.5
## 91	5.5	2.6
## 92	6.1	3.0
## 93	5.8	2.6
## 94	5.0	2.3
## 95	5.6	2.7
## 96	5.7	3.0
## 97	5.7	2.9
## 98	6.2	2.9
## 99	5.1	2.5

```

## 100      5.7      2.8
## 101      6.3      3.3
## 102      5.8      2.7
## 103      7.1      3.0
## 104      6.3      2.9
## 105      6.5      3.0
## 106      7.6      3.0
## 107      4.9      2.5
## 108      7.3      2.9
## 109      6.7      2.5
## 110      7.2      3.6
## 111      6.5      3.2
## 112      6.4      2.7
## 113      6.8      3.0
## 114      5.7      2.5
## 115      5.8      2.8
## 116      6.4      3.2
## 117      6.5      3.0
## 118      7.7      3.8
## 119      7.7      2.6
## 120      6.0      2.2
## 121      6.9      3.2
## 122      5.6      2.8
## 123      7.7      2.8
## 124      6.3      2.7
## 125      6.7      3.3
## 126      7.2      3.2
## 127      6.2      2.8
## 128      6.1      3.0
## 129      6.4      2.8
## 130      7.2      3.0
## 131      7.4      2.8
## 132      7.9      3.8
## 133      6.4      2.8
## 134      6.3      2.8
## 135      6.1      2.6
## 136      7.7      3.0
## 137      6.3      3.4
## 138      6.4      3.1
## 139      6.0      3.0
## 140      6.9      3.1
## 141      6.7      3.1
## 142      6.9      3.1
## 143      5.8      2.7
## 144      6.8      3.2
## 145      6.7      3.3
## 146      6.7      3.0
## 147      6.3      2.5
## 148      6.5      3.0
## 149      6.2      3.4
## 150      5.9      3.0

```

iris[1:5,] # 1~5행의 모든 데이터

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
```

```

## 1      5.1      3.5      1.4      0.2  setosa
## 2      4.9      3.0      1.4      0.2  setosa
## 3      4.7      3.2      1.3      0.2  setosa
## 4      4.6      3.1      1.5      0.2  setosa
## 5      5.0      3.6      1.4      0.2  setosa

iris[1:5,c(1,3,5)] # 1~5행의 데이터 중 1,3열의 데이터

## Sepal.Length Petal.Length Species
## 1      5.1      1.4  setosa
## 2      4.9      1.4  setosa
## 3      4.7      1.3  setosa
## 4      4.6      1.5  setosa
## 5      5.0      1.4  setosa

colSums(iris[,-5])

## Sepal.Length Sepal.Width Petal.Length Petal.Width
##     876.5      458.6      563.7      179.9

colMeans(iris[,-5])

## Sepal.Length Sepal.Width Petal.Length Petal.Width
##     5.843333    3.057333    3.758000    1.199333

rowSums(iris[,-5])

## [1] 10.2 9.5 9.4 9.4 10.2 11.4 9.7 10.1 8.9 9.6 10.8 10.0 9.3 8.5 11.2
## [16] 12.0 11.0 10.3 11.5 10.7 10.7 10.7 9.4 10.6 10.3 9.8 10.4 10.4 10.2 9.7
## [31] 9.7 10.7 10.9 11.3 9.7 9.6 10.5 10.0 8.9 10.2 10.1 8.4 9.1 10.7 11.2
## [46] 9.5 10.7 9.4 10.7 9.9 16.3 15.6 16.4 13.1 15.4 14.3 15.9 11.6 15.4 13.2
## [61] 11.5 14.6 13.2 15.1 13.4 15.6 14.6 13.6 14.4 13.1 15.7 14.2 15.2 14.8 14.9
## [76] 15.4 15.8 16.4 14.9 12.8 12.8 12.6 13.6 15.4 14.4 15.5 16.0 14.3 14.0 13.3
## [91] 13.7 15.1 13.6 11.6 13.8 14.1 14.1 14.7 11.7 13.9 18.1 15.5 18.1 16.6 17.5
## [106] 19.3 13.6 18.3 16.8 19.4 16.8 16.3 17.4 15.2 16.1 17.2 16.8 20.4 19.5 14.7
## [121] 18.1 15.3 19.2 15.7 17.8 18.2 15.6 15.8 16.9 17.6 18.2 20.1 17.0 15.7 15.7
## [136] 19.1 17.7 16.8 15.6 17.5 17.8 17.4 15.5 18.2 18.2 17.2 15.7 16.7 17.3 15.8

rowMeans(iris[,-5])

## [1] 2.550 2.375 2.350 2.350 2.550 2.850 2.425 2.525 2.225 2.400 2.700 2.500
## [13] 2.325 2.125 2.800 3.000 2.750 2.575 2.875 2.675 2.675 2.675 2.350 2.650
## [25] 2.575 2.450 2.600 2.600 2.550 2.425 2.425 2.675 2.725 2.825 2.425 2.400
## [37] 2.625 2.500 2.225 2.550 2.525 2.100 2.275 2.675 2.800 2.375 2.675 2.350
## [49] 2.675 2.475 4.075 3.900 4.100 3.275 3.850 3.575 3.975 2.900 3.850 3.300
## [61] 2.875 3.650 3.300 3.775 3.350 3.900 3.650 3.400 3.600 3.275 3.925 3.550
## [73] 3.800 3.700 3.725 3.850 3.950 4.100 3.725 3.200 3.200 3.150 3.400 3.850
## [85] 3.600 3.875 4.000 3.575 3.500 3.325 3.425 3.775 3.400 2.900 3.450 3.525
## [97] 3.525 3.675 2.925 3.475 4.525 3.875 4.525 4.150 4.375 4.825 3.400 4.575
## [109] 4.200 4.850 4.200 4.075 4.350 3.800 4.025 4.300 4.200 5.100 4.875 3.675
## [121] 4.525 3.825 4.800 3.925 4.450 4.550 3.900 3.950 4.225 4.400 4.550 5.025
## [133] 4.250 3.925 3.925 4.775 4.425 4.200 3.900 4.375 4.450 4.350 3.875 4.550
## [145] 4.550 4.300 3.925 4.175 4.325 3.950

```

1. 조건에 맞는 데이터를 추출하는 방법

`subset`은 행렬이 아닌 데이터프레임에서 주로 활용된다.

- `is.matrix()` # 데이터셋인지 확인하는 함수

- `is.data.frame()` # 데이터프레임인지 확인하는 함수

```

is.data.frame(iris)

## [1] TRUE

IR.1 <- subset(iris, Species=="setosa")
IR.1

##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
## 7          4.6         3.4          1.4         0.3  setosa
## 8          5.0         3.4          1.5         0.2  setosa
## 9          4.4         2.9          1.4         0.2  setosa
## 10         4.9         3.1          1.5         0.1  setosa
## 11         5.4         3.7          1.5         0.2  setosa
## 12         4.8         3.4          1.6         0.2  setosa
## 13         4.8         3.0          1.4         0.1  setosa
## 14         4.3         3.0          1.1         0.1  setosa
## 15         5.8         4.0          1.2         0.2  setosa
## 16         5.7         4.4          1.5         0.4  setosa
## 17         5.4         3.9          1.3         0.4  setosa
## 18         5.1         3.5          1.4         0.3  setosa
## 19         5.7         3.8          1.7         0.3  setosa
## 20         5.1         3.8          1.5         0.3  setosa
## 21         5.4         3.4          1.7         0.2  setosa
## 22         5.1         3.7          1.5         0.4  setosa
## 23         4.6         3.6          1.0         0.2  setosa
## 24         5.1         3.3          1.7         0.5  setosa
## 25         4.8         3.4          1.9         0.2  setosa
## 26         5.0         3.0          1.6         0.2  setosa
## 27         5.0         3.4          1.6         0.4  setosa
## 28         5.2         3.5          1.5         0.2  setosa
## 29         5.2         3.4          1.4         0.2  setosa
## 30         4.7         3.2          1.6         0.2  setosa
## 31         4.8         3.1          1.6         0.2  setosa
## 32         5.4         3.4          1.5         0.4  setosa
## 33         5.2         4.1          1.5         0.1  setosa
## 34         5.5         4.2          1.4         0.2  setosa
## 35         4.9         3.1          1.5         0.2  setosa
## 36         5.0         3.2          1.2         0.2  setosa
## 37         5.5         3.5          1.3         0.2  setosa
## 38         4.9         3.6          1.4         0.1  setosa
## 39         4.4         3.0          1.3         0.2  setosa
## 40         5.1         3.4          1.5         0.2  setosa
## 41         5.0         3.5          1.3         0.3  setosa
## 42         4.5         2.3          1.3         0.3  setosa
## 43         4.4         3.2          1.3         0.2  setosa
## 44         5.0         3.5          1.6         0.6  setosa
## 45         5.1         3.8          1.9         0.4  setosa

```

## 46	4.8	3.0	1.4	0.3	setosa
## 47	5.1	3.8	1.6	0.2	setosa
## 48	4.6	3.2	1.4	0.2	setosa
## 49	5.3	3.7	1.5	0.2	setosa
## 50	5.0	3.3	1.4	0.2	setosa
## 51	7.0	3.2	4.7	1.4	versicolor
## 52	6.4	3.2	4.5	1.5	versicolor
## 53	6.9	3.1	4.9	1.5	versicolor
## 54	5.5	2.3	4.0	1.3	versicolor
## 55	6.5	2.8	4.6	1.5	versicolor
## 56	5.7	2.8	4.5	1.3	versicolor
## 57	6.3	3.3	4.7	1.6	versicolor
## 58	4.9	2.4	3.3	1.0	versicolor
## 59	6.6	2.9	4.6	1.3	versicolor
## 60	5.2	2.7	3.9	1.4	versicolor
## 61	5.0	2.0	3.5	1.0	versicolor
## 62	5.9	3.0	4.2	1.5	versicolor
## 63	6.0	2.2	4.0	1.0	versicolor
## 64	6.1	2.9	4.7	1.4	versicolor
## 65	5.6	2.9	3.6	1.3	versicolor
## 66	6.7	3.1	4.4	1.4	versicolor
## 67	5.6	3.0	4.5	1.5	versicolor
## 68	5.8	2.7	4.1	1.0	versicolor
## 69	6.2	2.2	4.5	1.5	versicolor
## 70	5.6	2.5	3.9	1.1	versicolor
## 71	5.9	3.2	4.8	1.8	versicolor
## 72	6.1	2.8	4.0	1.3	versicolor
## 73	6.3	2.5	4.9	1.5	versicolor
## 74	6.1	2.8	4.7	1.2	versicolor
## 75	6.4	2.9	4.3	1.3	versicolor
## 76	6.6	3.0	4.4	1.4	versicolor
## 77	6.8	2.8	4.8	1.4	versicolor
## 78	6.7	3.0	5.0	1.7	versicolor
## 79	6.0	2.9	4.5	1.5	versicolor
## 80	5.7	2.6	3.5	1.0	versicolor
## 81	5.5	2.4	3.8	1.1	versicolor
## 82	5.5	2.4	3.7	1.0	versicolor
## 83	5.8	2.7	3.9	1.2	versicolor
## 84	6.0	2.7	5.1	1.6	versicolor
## 85	5.4	3.0	4.5	1.5	versicolor
## 86	6.0	3.4	4.5	1.6	versicolor
## 87	6.7	3.1	4.7	1.5	versicolor
## 88	6.3	2.3	4.4	1.3	versicolor
## 89	5.6	3.0	4.1	1.3	versicolor
## 90	5.5	2.5	4.0	1.3	versicolor
## 91	5.5	2.6	4.4	1.2	versicolor
## 92	6.1	3.0	4.6	1.4	versicolor
## 93	5.8	2.6	4.0	1.2	versicolor
## 94	5.0	2.3	3.3	1.0	versicolor
## 95	5.6	2.7	4.2	1.3	versicolor
## 96	5.7	3.0	4.2	1.2	versicolor
## 97	5.7	2.9	4.2	1.3	versicolor
## 98	6.2	2.9	4.3	1.3	versicolor
## 99	5.1	2.5	3.0	1.1	versicolor

```

## 100      5.7      2.8      4.1      1.3 versicolor
## 101      6.3      3.3      6.0      2.5 virginica
## 102      5.8      2.7      5.1      1.9 virginica
## 103      7.1      3.0      5.9      2.1 virginica
## 104      6.3      2.9      5.6      1.8 virginica
## 105      6.5      3.0      5.8      2.2 virginica
## 106      7.6      3.0      6.6      2.1 virginica
## 107      4.9      2.5      4.5      1.7 virginica
## 108      7.3      2.9      6.3      1.8 virginica
## 109      6.7      2.5      5.8      1.8 virginica
## 110      7.2      3.6      6.1      2.5 virginica
## 111      6.5      3.2      5.1      2.0 virginica
## 112      6.4      2.7      5.3      1.9 virginica
## 113      6.8      3.0      5.5      2.1 virginica
## 114      5.7      2.5      5.0      2.0 virginica
## 115      5.8      2.8      5.1      2.4 virginica
## 116      6.4      3.2      5.3      2.3 virginica
## 117      6.5      3.0      5.5      1.8 virginica
## 118      7.7      3.8      6.7      2.2 virginica
## 119      7.7      2.6      6.9      2.3 virginica
## 120      6.0      2.2      5.0      1.5 virginica
## 121      6.9      3.2      5.7      2.3 virginica
## 122      5.6      2.8      4.9      2.0 virginica
## 123      7.7      2.8      6.7      2.0 virginica
## 124      6.3      2.7      4.9      1.8 virginica
## 125      6.7      3.3      5.7      2.1 virginica
## 126      7.2      3.2      6.0      1.8 virginica
## 127      6.2      2.8      4.8      1.8 virginica
## 128      6.1      3.0      4.9      1.8 virginica
## 129      6.4      2.8      5.6      2.1 virginica
## 130      7.2      3.0      5.8      1.6 virginica
## 131      7.4      2.8      6.1      1.9 virginica
## 132      7.9      3.8      6.4      2.0 virginica
## 133      6.4      2.8      5.6      2.2 virginica
## 134      6.3      2.8      5.1      1.5 virginica
## 135      6.1      2.6      5.6      1.4 virginica
## 136      7.7      3.0      6.1      2.3 virginica
## 137      6.3      3.4      5.6      2.4 virginica
## 138      6.4      3.1      5.5      1.8 virginica
## 139      6.0      3.0      4.8      1.8 virginica
## 140      6.9      3.1      5.4      2.1 virginica
## 141      6.7      3.1      5.6      2.4 virginica
## 142      6.9      3.1      5.1      2.3 virginica
## 143      5.8      2.7      5.1      1.9 virginica
## 144      6.8      3.2      5.9      2.3 virginica
## 145      6.7      3.3      5.7      2.5 virginica
## 146      6.7      3.0      5.2      2.3 virginica
## 147      6.3      2.5      5.0      1.9 virginica
## 148      6.5      3.0      5.2      2.0 virginica
## 149      6.2      3.4      5.4      2.3 virginica
## 150      5.9      3.0      5.1      1.8 virginica

```

```
IR.2 <- subset(iris, Sepal.Length > 5.0 & Sepal.Width > 4.0)
```

```
IR.2
```

```

##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 16          5.7       4.4        1.5       0.4   setosa
## 33          5.2       4.1        1.5       0.1   setosa
## 34          5.5       4.2        1.4       0.2   setosa
IR.2[,c(2,4)]
```

```

##      Sepal.Width Petal.Width
## 16          4.4       0.4
## 33          4.1       0.1
## 34          4.2       0.2
```

```

iris.m <- as.matrix(iris[,1:4])
iris.m
```

```

##      Sepal.Length Sepal.Width Petal.Length Petal.Width
## [1,]          5.1       3.5        1.4       0.2
## [2,]          4.9       3.0        1.4       0.2
## [3,]          4.7       3.2        1.3       0.2
## [4,]          4.6       3.1        1.5       0.2
## [5,]          5.0       3.6        1.4       0.2
## [6,]          5.4       3.9        1.7       0.4
## [7,]          4.6       3.4        1.4       0.3
## [8,]          5.0       3.4        1.5       0.2
## [9,]          4.4       2.9        1.4       0.2
## [10,]         4.9       3.1        1.5       0.1
## [11,]         5.4       3.7        1.5       0.2
## [12,]         4.8       3.4        1.6       0.2
## [13,]         4.8       3.0        1.4       0.1
## [14,]         4.3       3.0        1.1       0.1
## [15,]         5.8       4.0        1.2       0.2
## [16,]         5.7       4.4        1.5       0.4
## [17,]         5.4       3.9        1.3       0.4
## [18,]         5.1       3.5        1.4       0.3
## [19,]         5.7       3.8        1.7       0.3
## [20,]         5.1       3.8        1.5       0.3
## [21,]         5.4       3.4        1.7       0.2
## [22,]         5.1       3.7        1.5       0.4
## [23,]         4.6       3.6        1.0       0.2
## [24,]         5.1       3.3        1.7       0.5
## [25,]         4.8       3.4        1.9       0.2
## [26,]         5.0       3.0        1.6       0.2
## [27,]         5.0       3.4        1.6       0.4
## [28,]         5.2       3.5        1.5       0.2
## [29,]         5.2       3.4        1.4       0.2
## [30,]         4.7       3.2        1.6       0.2
## [31,]         4.8       3.1        1.6       0.2
## [32,]         5.4       3.4        1.5       0.4
## [33,]         5.2       4.1        1.5       0.1
## [34,]         5.5       4.2        1.4       0.2
## [35,]         4.9       3.1        1.5       0.2
## [36,]         5.0       3.2        1.2       0.2
## [37,]         5.5       3.5        1.3       0.2
## [38,]         4.9       3.6        1.4       0.1
## [39,]         4.4       3.0        1.3       0.2
## [40,]         5.1       3.4        1.5       0.2
```

## [41,]	5.0	3.5	1.3	0.3
## [42,]	4.5	2.3	1.3	0.3
## [43,]	4.4	3.2	1.3	0.2
## [44,]	5.0	3.5	1.6	0.6
## [45,]	5.1	3.8	1.9	0.4
## [46,]	4.8	3.0	1.4	0.3
## [47,]	5.1	3.8	1.6	0.2
## [48,]	4.6	3.2	1.4	0.2
## [49,]	5.3	3.7	1.5	0.2
## [50,]	5.0	3.3	1.4	0.2
## [51,]	7.0	3.2	4.7	1.4
## [52,]	6.4	3.2	4.5	1.5
## [53,]	6.9	3.1	4.9	1.5
## [54,]	5.5	2.3	4.0	1.3
## [55,]	6.5	2.8	4.6	1.5
## [56,]	5.7	2.8	4.5	1.3
## [57,]	6.3	3.3	4.7	1.6
## [58,]	4.9	2.4	3.3	1.0
## [59,]	6.6	2.9	4.6	1.3
## [60,]	5.2	2.7	3.9	1.4
## [61,]	5.0	2.0	3.5	1.0
## [62,]	5.9	3.0	4.2	1.5
## [63,]	6.0	2.2	4.0	1.0
## [64,]	6.1	2.9	4.7	1.4
## [65,]	5.6	2.9	3.6	1.3
## [66,]	6.7	3.1	4.4	1.4
## [67,]	5.6	3.0	4.5	1.5
## [68,]	5.8	2.7	4.1	1.0
## [69,]	6.2	2.2	4.5	1.5
## [70,]	5.6	2.5	3.9	1.1
## [71,]	5.9	3.2	4.8	1.8
## [72,]	6.1	2.8	4.0	1.3
## [73,]	6.3	2.5	4.9	1.5
## [74,]	6.1	2.8	4.7	1.2
## [75,]	6.4	2.9	4.3	1.3
## [76,]	6.6	3.0	4.4	1.4
## [77,]	6.8	2.8	4.8	1.4
## [78,]	6.7	3.0	5.0	1.7
## [79,]	6.0	2.9	4.5	1.5
## [80,]	5.7	2.6	3.5	1.0
## [81,]	5.5	2.4	3.8	1.1
## [82,]	5.5	2.4	3.7	1.0
## [83,]	5.8	2.7	3.9	1.2
## [84,]	6.0	2.7	5.1	1.6
## [85,]	5.4	3.0	4.5	1.5
## [86,]	6.0	3.4	4.5	1.6
## [87,]	6.7	3.1	4.7	1.5
## [88,]	6.3	2.3	4.4	1.3
## [89,]	5.6	3.0	4.1	1.3
## [90,]	5.5	2.5	4.0	1.3
## [91,]	5.5	2.6	4.4	1.2
## [92,]	6.1	3.0	4.6	1.4
## [93,]	5.8	2.6	4.0	1.2
## [94,]	5.0	2.3	3.3	1.0

## [95,]	5.6	2.7	4.2	1.3
## [96,]	5.7	3.0	4.2	1.2
## [97,]	5.7	2.9	4.2	1.3
## [98,]	6.2	2.9	4.3	1.3
## [99,]	5.1	2.5	3.0	1.1
## [100,]	5.7	2.8	4.1	1.3
## [101,]	6.3	3.3	6.0	2.5
## [102,]	5.8	2.7	5.1	1.9
## [103,]	7.1	3.0	5.9	2.1
## [104,]	6.3	2.9	5.6	1.8
## [105,]	6.5	3.0	5.8	2.2
## [106,]	7.6	3.0	6.6	2.1
## [107,]	4.9	2.5	4.5	1.7
## [108,]	7.3	2.9	6.3	1.8
## [109,]	6.7	2.5	5.8	1.8
## [110,]	7.2	3.6	6.1	2.5
## [111,]	6.5	3.2	5.1	2.0
## [112,]	6.4	2.7	5.3	1.9
## [113,]	6.8	3.0	5.5	2.1
## [114,]	5.7	2.5	5.0	2.0
## [115,]	5.8	2.8	5.1	2.4
## [116,]	6.4	3.2	5.3	2.3
## [117,]	6.5	3.0	5.5	1.8
## [118,]	7.7	3.8	6.7	2.2
## [119,]	7.7	2.6	6.9	2.3
## [120,]	6.0	2.2	5.0	1.5
## [121,]	6.9	3.2	5.7	2.3
## [122,]	5.6	2.8	4.9	2.0
## [123,]	7.7	2.8	6.7	2.0
## [124,]	6.3	2.7	4.9	1.8
## [125,]	6.7	3.3	5.7	2.1
## [126,]	7.2	3.2	6.0	1.8
## [127,]	6.2	2.8	4.8	1.8
## [128,]	6.1	3.0	4.9	1.8
## [129,]	6.4	2.8	5.6	2.1
## [130,]	7.2	3.0	5.8	1.6
## [131,]	7.4	2.8	6.1	1.9
## [132,]	7.9	3.8	6.4	2.0
## [133,]	6.4	2.8	5.6	2.2
## [134,]	6.3	2.8	5.1	1.5
## [135,]	6.1	2.6	5.6	1.4
## [136,]	7.7	3.0	6.1	2.3
## [137,]	6.3	3.4	5.6	2.4
## [138,]	6.4	3.1	5.5	1.8
## [139,]	6.0	3.0	4.8	1.8
## [140,]	6.9	3.1	5.4	2.1
## [141,]	6.7	3.1	5.6	2.4
## [142,]	6.9	3.1	5.1	2.3
## [143,]	5.8	2.7	5.1	1.9
## [144,]	6.8	3.2	5.9	2.3
## [145,]	6.7	3.3	5.7	2.5
## [146,]	6.7	3.0	5.2	2.3
## [147,]	6.3	2.5	5.0	1.9
## [148,]	6.5	3.0	5.2	2.0

```
## [149,]      6.2      3.4      5.4      2.3
## [150,]      5.9      3.0      5.1      1.8
iris$Sepal.Length

##  [1] 5.1 4.9 4.7 4.6 5.0 5.4 4.6 5.0 4.4 4.9 5.4 4.8 4.8 4.3 5.8 5.7 5.4 5.1
##  [19] 5.7 5.1 5.4 5.1 4.6 5.1 4.8 5.0 5.0 5.2 5.2 4.7 4.8 5.4 5.2 5.5 4.9 5.0
##  [37] 5.5 4.9 4.4 5.1 5.0 4.5 4.4 5.0 5.1 4.8 5.1 4.6 5.3 5.0 7.0 6.4 6.9 5.5
##  [55] 6.5 5.7 6.3 4.9 6.6 5.2 5.0 5.9 6.0 6.1 5.6 6.7 5.6 5.8 6.2 5.6 5.9 6.1
##  [73] 6.3 6.1 6.4 6.6 6.8 6.7 6.0 5.7 5.5 5.5 5.8 6.0 5.4 6.0 6.7 6.3 5.6 5.5
##  [91] 5.5 6.1 5.8 5.0 5.6 5.7 5.7 6.2 5.1 5.7 6.3 5.8 7.1 6.3 6.5 7.6 4.9 7.3
## [109] 6.7 7.2 6.5 6.4 6.8 5.7 5.8 6.4 6.5 7.7 7.7 6.0 6.9 5.6 7.7 6.3 6.7 7.2
## [127] 6.2 6.1 6.4 7.2 7.4 7.9 6.4 6.3 6.1 7.7 6.3 6.4 6.0 6.9 6.7 6.9 5.8 6.8
## [145] 6.7 6.7 6.3 6.5 6.2 5.9
```