

## Chapter07 데이터 전처리

신봉균 20191624

2023-04-12

### 코드 7-1

```
z= c(1,2,3,NA,5,NA,8) #결측값이 포함된 벡터 z
sum(z)                  #정상계산이 안됨

## [1] NA

is.na(z)                #NA 여부 확인

## [1] FALSE FALSE FALSE TRUE FALSE TRUE FALSE

sum(is.na(z))           #NA 의 개수 확인

## [1] 2

sum(z, na.rm=TRUE)      #NA 를 제외하고 합계를 계산함

## [1] 19
```

(1)**sum(z)**는 z 안의 결측 값이 존재하기 때문에 결과가 **NA** 로 나온다. (2)**is.na(z)**는 z 에 결측값이 아닌 것은 **FALSE** 로 결측 값인 것은 **TRUE** 로 변환하여 보여준다.

(3)**sum(is.na(z))**는 **is.na(z)**에 **FALSE** 는 **0** 으로 취급하고 **TRUE** 는 **1** 로 취급하기 때문에 **is.na(z)**를 다 더한 값은 결측 값의 갯수를 의미한다. (4)**sum(z, na.rm=TRUE)**에서 **na.rm()**함수는 매개변수 안에 결측 값을 제거해주는 역할을 한다 **rm** 은 remove 를 뜻한다.

### 코드 7-2

```
z1= c(1,2,3,NA,5,NA,8) #결측값이 포함된 벡터 z1
z2= c(5,8,1,NA,3,NA,7) #결측값이 포함된 벡터 z2
z1[is.na(z1)]          #NA 를 0 으로 치환

## [1] NA NA

z1
```

```
## [1] 1 2 3 NA 5 NA 8
```

```
z3= as.vector(na.omit(z2)) #NA 를 제거하고 새로운 벡터 생성  
z3
```

```
## [1] 5 8 1 3 7
```

(1) `z1[is.na(z1)]` 는 결측 값인 NA 를 0 으로 변환해준다

(2) `z3= as.vector(na.omit(z2))` NA 를 제거한 벡터를 z3 라는 변수에 할당시킨다.

### 코드 7-3

```
x= iris  
x[1,2] = NA; x[1,3]= NA  
x[2,3]= NA; x[3,4]= NA  
head(x)  
  
## Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## 1 5.1 NA NA 0.2 setosa  
## 2 4.9 3.0 NA 0.2 setosa  
## 3 4.7 3.2 1.3 NA 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
```

(1) x 에 R 에 기본적으로 내장되어 있는 데이터 **iris** 를 저장시킨다.

(2) x 의 1 행 2 열에 있는 데이터를 NA 로 대체, x 의 1 행 3 열을 NA 로 대체, 2 행 3 열,  
3 행 4 열을 NA 로 대체

(3) x 를 6 개의 row 만 출력

### 코드 7-4

*#for 문을 이용한 방법*

*#apply 를 활용한 방법*

```
col_na <- function(y) {  
  return(sum(is.na(y)))  
}  
  
na_count = apply(x,2,FUN=col_na)  
na_count
```

```
## Sepal.Length Sepal.Width Petal.Length Petal.Width Species
##           0           1           2           1           0
```

(1) **for (i in 1:ncol(x))** : 데이터프레임 x 의 열 개수만큼 반복합니다. (2) **this.na = is.na(x[i,])** : x[i,]는 i 번째 열 모든 행을 의미. is.na() 함수를 사용하여 해당 열의 값 중 NA 값인 부분을 TRUE, 그렇지 않은 부분을 FALSE 로 구분하여 this.na 변수에 할당

(3) **cat(colnames(x)[i], ", sum(this.na), ")** : colnames(x)[i]는 x 의 i 번째 열의 이름을 의미한다. sum(this.na)는 해당 열에서 NA 값의 개수를 의미한다. **cat()** 함수를 사용하여 열 이름, tab 문자, NA 값의 개수를 출력한다.

(4) **col\_na** 는 결측 값의 갯수를 제한해주는 함수 (5) **apply(x,2,FUN=col\_na)**는 data 는 iris 데이터가 저장되어있는 x 의 각 열에 **col\_na** 함수 적용

### 코드 7-5

```
rowSums(is.na(x))           #행별 NA 개수

##      [1] 2 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##      [38] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##      [75] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##     [112] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##     [149] 0 0

sum(rowSums(is.na(x))>0)    #NA 에 포함된 행의 개수

##      [1] 3

sum(is.na(x))               #데이터셋 전체에서 NA 개수

##      [1] 4
```

(1) 행별 NA 개수를 계산한다. (2) NA 값을 가진 행들의 개수 (3) x 행열에 전체 NA 의 갯수

### 코드 7-6

```
head(x)
```

##	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
## 1	5.1	NA	NA	0.2	setosa
## 2	4.9	3.0	NA	0.2	setosa
## 3	4.7	3.2	1.3	NA	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
```

`x[!complete.cases(x),]` *#NA 가 포함된 행들 출력*

```
## Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1      5.1      NA      NA      0.2 setosa
## 2      4.9      3.0      NA      0.2 setosa
## 3      4.7      3.2      1.3      NA setosa
```

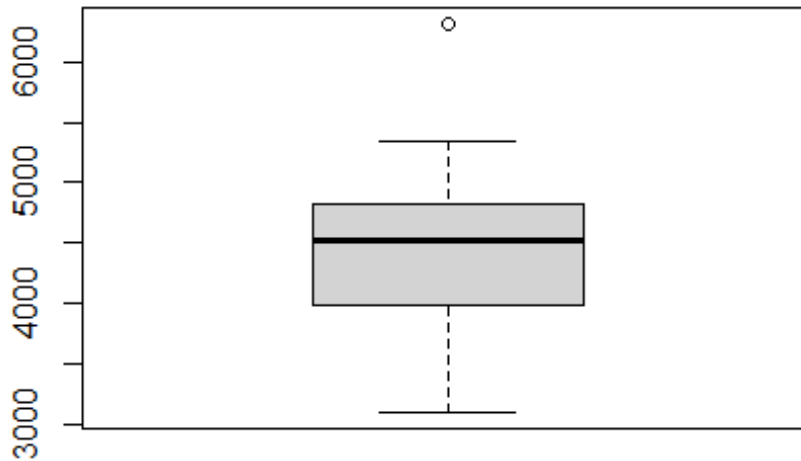
`y= x[complete.cases(x),]` *#NA 가 포함된 행들 제거*  
`head(y)`

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

- (1) x 의 6 개 행만 출력
- (2) NA 가 포함된 행들을 출력
- (3) NA 가 포함된 포함된 행을 제거한 데이터셋을 y 변수에 할당
- (4) Y 데이터셋 6 개 행 출력

## 코드 7-7

```
st= data.frame(state.x77)
boxplot(st$Income)
```



```
boxplot.stats(st$Income)$out
```

```
## [1] 6315
```

- (1) st 변수에 state.x77 데이터셋을 저장
- (2) st 데이터셋에 Income 열 데이터를 가지고 boxplot 작성
- (3) **boxplot.stats()** 함수는 리스트 형태로 여러 개의 결과값을 반환하는데 그중에서 out 은 특이값을 의미한다. 그러므로 st 데이터셋의 Income 열에 대한 데이터의 특이 값들을 추출한다.

## 코드 7-8

```
out.val = boxplot.stats(st$Income)$out #특이 값추출
st$Income[st$Income %in% out.val] = NA #특이값을 NA 로 대체
head(st)
```

```
##      Population Income Illiteracy Life.Exp Murder HS.Grad Frost  Ar
ea
## Alabama      3615   3624         2.1   69.05   15.1    41.3    20  507
08
## Alaska       365     NA         1.5   69.31   11.3    66.7   152 5664
32
```

```
## Arizona      2212  4530      1.8   70.55    7.8   58.1    15 1134
17
## Arkansas     2110  3378      1.9   70.66   10.1   39.9    65  519
45
## California   21198  5114      1.1   71.71   10.3   62.6    20 1563
61
## Colorado     2541  4884      0.7   72.06    6.8   63.9   166 1037
66
```

```
newdata= st[complete.cases(st),] #st 데이터셋에 NA 값이 포함 안되어있는 행들을 newdata에 할당
head(newdata)
```

```
##      Population Income Illiteracy Life.Exp Murder HS.Grad Frost   Area
rea
## Alabama      3615   3624        2.1   69.05   15.1   41.3    20   50
708
## Arizona      2212   4530        1.8   70.55    7.8   58.1    15  113
417
## Arkansas     2110   3378        1.9   70.66   10.1   39.9    65   51
945
## California   21198   5114        1.1   71.71   10.3   62.6    20  156
361
## Colorado     2541   4884        0.7   72.06    6.8   63.9   166  103
766
## Connecticut   3100   5348        1.1   72.48    3.1   56.0   139    4
862
```

(1)**out.val** 에 특이 값들을 저장 (2)**stIncome[stIncome %in% out.val]= NA** st 데이터셋에 저장되어 있는 특이 값들을 NA 로 대체해버린다 (3)**newdata** st 데이터셋에서 결측값에 포함된 행들을 제거한 데이터셋을 **newdata** 에 할당한다.

### 코드 7-9

```
v1= c(1,7,6,8,4,2,3)
order(v1)

## [1] 1 6 7 5 3 2 4

v1= sort(v1) #오름차순
v1

## [1] 1 2 3 4 6 7 8

v2= sort(v1, decreasing = T) #내림차순
v2
```

```
## [1] 8 7 6 4 3 2 1
```

(1) **sort(v1)** 기본은 오름차순이 기본

(2) **sort(v1, decreasing= T)** 내림차순

### 코드 7-10

```
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
```

```
order(iris$Sepal.Length)
```

```
## [1] 14 9 39 43 42 4 7 23 48 3 30 12 13 25 31 46 2
## 10
## [19] 35 38 58 107 5 8 26 27 36 41 44 50 61 94 1 18 20
## 22
## [37] 24 40 45 47 99 28 29 33 60 49 6 11 17 21 32 85 34
## 37
## [55] 54 81 82 90 91 65 67 70 89 95 122 16 19 56 80 96 97
## 100
## [73] 114 15 68 83 93 102 115 143 62 71 150 63 79 84 86 120 139
## 64
## [91] 72 74 92 128 135 69 98 127 149 57 73 88 101 104 124 134 137
## 147
## [109] 52 75 112 116 129 133 138 55 105 111 117 148 59 76 66 78 87
## 109
## [127] 125 141 145 146 77 113 144 53 121 140 142 51 103 110 126 130 108
## 131
## [145] 106 118 119 123 136 132
```

```
iris[order(iris$Sepal.Length),]
```

*#오름 차순으로 정렬*

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 14           4.3         3.0         1.1         0.1   setosa
## 9           4.4         2.9         1.4         0.2   setosa
## 39           4.4         3.0         1.3         0.2   setosa
## 43           4.4         3.2         1.3         0.2   setosa
## 42           4.5         2.3         1.3         0.3   setosa
## 4           4.6         3.1         1.5         0.2   setosa
## 7           4.6         3.4         1.4         0.3   setosa
## 23           4.6         3.6         1.0         0.2   setosa
## 48           4.6         3.2         1.4         0.2   setosa
```

## 3	4.7	3.2	1.3	0.2	setosa
## 30	4.7	3.2	1.6	0.2	setosa
## 12	4.8	3.4	1.6	0.2	setosa
## 13	4.8	3.0	1.4	0.1	setosa
## 25	4.8	3.4	1.9	0.2	setosa
## 31	4.8	3.1	1.6	0.2	setosa
## 46	4.8	3.0	1.4	0.3	setosa
## 2	4.9	3.0	1.4	0.2	setosa
## 10	4.9	3.1	1.5	0.1	setosa
## 35	4.9	3.1	1.5	0.2	setosa
## 38	4.9	3.6	1.4	0.1	setosa
## 58	4.9	2.4	3.3	1.0	versicolor
## 107	4.9	2.5	4.5	1.7	virginica
## 5	5.0	3.6	1.4	0.2	setosa
## 8	5.0	3.4	1.5	0.2	setosa
## 26	5.0	3.0	1.6	0.2	setosa
## 27	5.0	3.4	1.6	0.4	setosa
## 36	5.0	3.2	1.2	0.2	setosa
## 41	5.0	3.5	1.3	0.3	setosa
## 44	5.0	3.5	1.6	0.6	setosa
## 50	5.0	3.3	1.4	0.2	setosa
## 61	5.0	2.0	3.5	1.0	versicolor
## 94	5.0	2.3	3.3	1.0	versicolor
## 1	5.1	3.5	1.4	0.2	setosa
## 18	5.1	3.5	1.4	0.3	setosa
## 20	5.1	3.8	1.5	0.3	setosa
## 22	5.1	3.7	1.5	0.4	setosa
## 24	5.1	3.3	1.7	0.5	setosa
## 40	5.1	3.4	1.5	0.2	setosa
## 45	5.1	3.8	1.9	0.4	setosa
## 47	5.1	3.8	1.6	0.2	setosa
## 99	5.1	2.5	3.0	1.1	versicolor
## 28	5.2	3.5	1.5	0.2	setosa
## 29	5.2	3.4	1.4	0.2	setosa
## 33	5.2	4.1	1.5	0.1	setosa
## 60	5.2	2.7	3.9	1.4	versicolor
## 49	5.3	3.7	1.5	0.2	setosa
## 6	5.4	3.9	1.7	0.4	setosa
## 11	5.4	3.7	1.5	0.2	setosa
## 17	5.4	3.9	1.3	0.4	setosa
## 21	5.4	3.4	1.7	0.2	setosa
## 32	5.4	3.4	1.5	0.4	setosa
## 85	5.4	3.0	4.5	1.5	versicolor
## 34	5.5	4.2	1.4	0.2	setosa
## 37	5.5	3.5	1.3	0.2	setosa
## 54	5.5	2.3	4.0	1.3	versicolor
## 81	5.5	2.4	3.8	1.1	versicolor
## 82	5.5	2.4	3.7	1.0	versicolor
## 90	5.5	2.5	4.0	1.3	versicolor



## 91	5.5	2.6	4.4	1.2 versicolor
## 65	5.6	2.9	3.6	1.3 versicolor
## 67	5.6	3.0	4.5	1.5 versicolor
## 70	5.6	2.5	3.9	1.1 versicolor
## 89	5.6	3.0	4.1	1.3 versicolor
## 95	5.6	2.7	4.2	1.3 versicolor
## 122	5.6	2.8	4.9	2.0 virginica
## 16	5.7	4.4	1.5	0.4 setosa
## 19	5.7	3.8	1.7	0.3 setosa
## 56	5.7	2.8	4.5	1.3 versicolor
## 80	5.7	2.6	3.5	1.0 versicolor
## 96	5.7	3.0	4.2	1.2 versicolor
## 97	5.7	2.9	4.2	1.3 versicolor
## 100	5.7	2.8	4.1	1.3 versicolor
## 114	5.7	2.5	5.0	2.0 virginica
## 15	5.8	4.0	1.2	0.2 setosa
## 68	5.8	2.7	4.1	1.0 versicolor
## 83	5.8	2.7	3.9	1.2 versicolor
## 93	5.8	2.6	4.0	1.2 versicolor
## 102	5.8	2.7	5.1	1.9 virginica
## 115	5.8	2.8	5.1	2.4 virginica
## 143	5.8	2.7	5.1	1.9 virginica
## 62	5.9	3.0	4.2	1.5 versicolor
## 71	5.9	3.2	4.8	1.8 versicolor
## 150	5.9	3.0	5.1	1.8 virginica
## 63	6.0	2.2	4.0	1.0 versicolor
## 79	6.0	2.9	4.5	1.5 versicolor
## 84	6.0	2.7	5.1	1.6 versicolor
## 86	6.0	3.4	4.5	1.6 versicolor
## 120	6.0	2.2	5.0	1.5 virginica
## 139	6.0	3.0	4.8	1.8 virginica
## 64	6.1	2.9	4.7	1.4 versicolor
## 72	6.1	2.8	4.0	1.3 versicolor
## 74	6.1	2.8	4.7	1.2 versicolor
## 92	6.1	3.0	4.6	1.4 versicolor
## 128	6.1	3.0	4.9	1.8 virginica
## 135	6.1	2.6	5.6	1.4 virginica
## 69	6.2	2.2	4.5	1.5 versicolor
## 98	6.2	2.9	4.3	1.3 versicolor
## 127	6.2	2.8	4.8	1.8 virginica
## 149	6.2	3.4	5.4	2.3 virginica
## 57	6.3	3.3	4.7	1.6 versicolor
## 73	6.3	2.5	4.9	1.5 versicolor
## 88	6.3	2.3	4.4	1.3 versicolor
## 101	6.3	3.3	6.0	2.5 virginica
## 104	6.3	2.9	5.6	1.8 virginica
## 124	6.3	2.7	4.9	1.8 virginica
## 134	6.3	2.8	5.1	1.5 virginica
## 137	6.3	3.4	5.6	2.4 virginica

```
## 147      6.3      2.5      5.0      1.9 virginica
## 52       6.4      3.2      4.5      1.5 versicolor
## 75       6.4      2.9      4.3      1.3 versicolor
## 112      6.4      2.7      5.3      1.9 virginica
## 116      6.4      3.2      5.3      2.3 virginica
## 129      6.4      2.8      5.6      2.1 virginica
## 133      6.4      2.8      5.6      2.2 virginica
## 138      6.4      3.1      5.5      1.8 virginica
## 55       6.5      2.8      4.6      1.5 versicolor
## 105      6.5      3.0      5.8      2.2 virginica
## 111      6.5      3.2      5.1      2.0 virginica
## 117      6.5      3.0      5.5      1.8 virginica
## 148      6.5      3.0      5.2      2.0 virginica
## 59       6.6      2.9      4.6      1.3 versicolor
## 76       6.6      3.0      4.4      1.4 versicolor
## 66       6.7      3.1      4.4      1.4 versicolor
## 78       6.7      3.0      5.0      1.7 versicolor
## 87       6.7      3.1      4.7      1.5 versicolor
## 109      6.7      2.5      5.8      1.8 virginica
## 125      6.7      3.3      5.7      2.1 virginica
## 141      6.7      3.1      5.6      2.4 virginica
## 145      6.7      3.3      5.7      2.5 virginica
## 146      6.7      3.0      5.2      2.3 virginica
## 77       6.8      2.8      4.8      1.4 versicolor
## 113      6.8      3.0      5.5      2.1 virginica
## 144      6.8      3.2      5.9      2.3 virginica
## 53       6.9      3.1      4.9      1.5 versicolor
## 121      6.9      3.2      5.7      2.3 virginica
## 140      6.9      3.1      5.4      2.1 virginica
## 142      6.9      3.1      5.1      2.3 virginica
## 51       7.0      3.2      4.7      1.4 versicolor
## 103      7.1      3.0      5.9      2.1 virginica
## 110      7.2      3.6      6.1      2.5 virginica
## 126      7.2      3.2      6.0      1.8 virginica
## 130      7.2      3.0      5.8      1.6 virginica
## 108      7.3      2.9      6.3      1.8 virginica
## 131      7.4      2.8      6.1      1.9 virginica
## 106      7.6      3.0      6.6      2.1 virginica
## 118      7.7      3.8      6.7      2.2 virginica
## 119      7.7      2.6      6.9      2.3 virginica
## 123      7.7      2.8      6.7      2.0 virginica
## 136      7.7      3.0      6.1      2.3 virginica
## 132      7.9      3.8      6.4      2.0 virginica
```

```
iris[order(iris$Sepal.Length, decreasing = T),] #내림 차순으로 정렬
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width  Species
## 132          7.9         3.8         6.4         2.0  virginica
## 118          7.7         3.8         6.7         2.2  virginica
```

## 119	7.7	2.6	6.9	2.3	virginica
## 123	7.7	2.8	6.7	2.0	virginica
## 136	7.7	3.0	6.1	2.3	virginica
## 106	7.6	3.0	6.6	2.1	virginica
## 131	7.4	2.8	6.1	1.9	virginica
## 108	7.3	2.9	6.3	1.8	virginica
## 110	7.2	3.6	6.1	2.5	virginica
## 126	7.2	3.2	6.0	1.8	virginica
## 130	7.2	3.0	5.8	1.6	virginica
## 103	7.1	3.0	5.9	2.1	virginica
## 51	7.0	3.2	4.7	1.4	versicolor
## 53	6.9	3.1	4.9	1.5	versicolor
## 121	6.9	3.2	5.7	2.3	virginica
## 140	6.9	3.1	5.4	2.1	virginica
## 142	6.9	3.1	5.1	2.3	virginica
## 77	6.8	2.8	4.8	1.4	versicolor
## 113	6.8	3.0	5.5	2.1	virginica
## 144	6.8	3.2	5.9	2.3	virginica
## 66	6.7	3.1	4.4	1.4	versicolor
## 78	6.7	3.0	5.0	1.7	versicolor
## 87	6.7	3.1	4.7	1.5	versicolor
## 109	6.7	2.5	5.8	1.8	virginica
## 125	6.7	3.3	5.7	2.1	virginica
## 141	6.7	3.1	5.6	2.4	virginica
## 145	6.7	3.3	5.7	2.5	virginica
## 146	6.7	3.0	5.2	2.3	virginica
## 59	6.6	2.9	4.6	1.3	versicolor
## 76	6.6	3.0	4.4	1.4	versicolor
## 55	6.5	2.8	4.6	1.5	versicolor
## 105	6.5	3.0	5.8	2.2	virginica
## 111	6.5	3.2	5.1	2.0	virginica
## 117	6.5	3.0	5.5	1.8	virginica
## 148	6.5	3.0	5.2	2.0	virginica
## 52	6.4	3.2	4.5	1.5	versicolor
## 75	6.4	2.9	4.3	1.3	versicolor
## 112	6.4	2.7	5.3	1.9	virginica
## 116	6.4	3.2	5.3	2.3	virginica
## 129	6.4	2.8	5.6	2.1	virginica
## 133	6.4	2.8	5.6	2.2	virginica
## 138	6.4	3.1	5.5	1.8	virginica
## 57	6.3	3.3	4.7	1.6	versicolor
## 73	6.3	2.5	4.9	1.5	versicolor
## 88	6.3	2.3	4.4	1.3	versicolor
## 101	6.3	3.3	6.0	2.5	virginica
## 104	6.3	2.9	5.6	1.8	virginica
## 124	6.3	2.7	4.9	1.8	virginica
## 134	6.3	2.8	5.1	1.5	virginica
## 137	6.3	3.4	5.6	2.4	virginica
## 147	6.3	2.5	5.0	1.9	virginica

## 69	6.2	2.2	4.5	1.5 versicolor
## 98	6.2	2.9	4.3	1.3 versicolor
## 127	6.2	2.8	4.8	1.8 virginica
## 149	6.2	3.4	5.4	2.3 virginica
## 64	6.1	2.9	4.7	1.4 versicolor
## 72	6.1	2.8	4.0	1.3 versicolor
## 74	6.1	2.8	4.7	1.2 versicolor
## 92	6.1	3.0	4.6	1.4 versicolor
## 128	6.1	3.0	4.9	1.8 virginica
## 135	6.1	2.6	5.6	1.4 virginica
## 63	6.0	2.2	4.0	1.0 versicolor
## 79	6.0	2.9	4.5	1.5 versicolor
## 84	6.0	2.7	5.1	1.6 versicolor
## 86	6.0	3.4	4.5	1.6 versicolor
## 120	6.0	2.2	5.0	1.5 virginica
## 139	6.0	3.0	4.8	1.8 virginica
## 62	5.9	3.0	4.2	1.5 versicolor
## 71	5.9	3.2	4.8	1.8 versicolor
## 150	5.9	3.0	5.1	1.8 virginica
## 15	5.8	4.0	1.2	0.2 setosa
## 68	5.8	2.7	4.1	1.0 versicolor
## 83	5.8	2.7	3.9	1.2 versicolor
## 93	5.8	2.6	4.0	1.2 versicolor
## 102	5.8	2.7	5.1	1.9 virginica
## 115	5.8	2.8	5.1	2.4 virginica
## 143	5.8	2.7	5.1	1.9 virginica
## 16	5.7	4.4	1.5	0.4 setosa
## 19	5.7	3.8	1.7	0.3 setosa
## 56	5.7	2.8	4.5	1.3 versicolor
## 80	5.7	2.6	3.5	1.0 versicolor
## 96	5.7	3.0	4.2	1.2 versicolor
## 97	5.7	2.9	4.2	1.3 versicolor
## 100	5.7	2.8	4.1	1.3 versicolor
## 114	5.7	2.5	5.0	2.0 virginica
## 65	5.6	2.9	3.6	1.3 versicolor
## 67	5.6	3.0	4.5	1.5 versicolor
## 70	5.6	2.5	3.9	1.1 versicolor
## 89	5.6	3.0	4.1	1.3 versicolor
## 95	5.6	2.7	4.2	1.3 versicolor
## 122	5.6	2.8	4.9	2.0 virginica
## 34	5.5	4.2	1.4	0.2 setosa
## 37	5.5	3.5	1.3	0.2 setosa
## 54	5.5	2.3	4.0	1.3 versicolor
## 81	5.5	2.4	3.8	1.1 versicolor
## 82	5.5	2.4	3.7	1.0 versicolor
## 90	5.5	2.5	4.0	1.3 versicolor
## 91	5.5	2.6	4.4	1.2 versicolor
## 6	5.4	3.9	1.7	0.4 setosa
## 11	5.4	3.7	1.5	0.2 setosa

## 17	5.4	3.9	1.3	0.4	setosa
## 21	5.4	3.4	1.7	0.2	setosa
## 32	5.4	3.4	1.5	0.4	setosa
## 85	5.4	3.0	4.5	1.5	versicolor
## 49	5.3	3.7	1.5	0.2	setosa
## 28	5.2	3.5	1.5	0.2	setosa
## 29	5.2	3.4	1.4	0.2	setosa
## 33	5.2	4.1	1.5	0.1	setosa
## 60	5.2	2.7	3.9	1.4	versicolor
## 1	5.1	3.5	1.4	0.2	setosa
## 18	5.1	3.5	1.4	0.3	setosa
## 20	5.1	3.8	1.5	0.3	setosa
## 22	5.1	3.7	1.5	0.4	setosa
## 24	5.1	3.3	1.7	0.5	setosa
## 40	5.1	3.4	1.5	0.2	setosa
## 45	5.1	3.8	1.9	0.4	setosa
## 47	5.1	3.8	1.6	0.2	setosa
## 99	5.1	2.5	3.0	1.1	versicolor
## 5	5.0	3.6	1.4	0.2	setosa
## 8	5.0	3.4	1.5	0.2	setosa
## 26	5.0	3.0	1.6	0.2	setosa
## 27	5.0	3.4	1.6	0.4	setosa
## 36	5.0	3.2	1.2	0.2	setosa
## 41	5.0	3.5	1.3	0.3	setosa
## 44	5.0	3.5	1.6	0.6	setosa
## 50	5.0	3.3	1.4	0.2	setosa
## 61	5.0	2.0	3.5	1.0	versicolor
## 94	5.0	2.3	3.3	1.0	versicolor
## 2	4.9	3.0	1.4	0.2	setosa
## 10	4.9	3.1	1.5	0.1	setosa
## 35	4.9	3.1	1.5	0.2	setosa
## 38	4.9	3.6	1.4	0.1	setosa
## 58	4.9	2.4	3.3	1.0	versicolor
## 107	4.9	2.5	4.5	1.7	virginica
## 12	4.8	3.4	1.6	0.2	setosa
## 13	4.8	3.0	1.4	0.1	setosa
## 25	4.8	3.4	1.9	0.2	setosa
## 31	4.8	3.1	1.6	0.2	setosa
## 46	4.8	3.0	1.4	0.3	setosa
## 3	4.7	3.2	1.3	0.2	setosa
## 30	4.7	3.2	1.6	0.2	setosa
## 4	4.6	3.1	1.5	0.2	setosa
## 7	4.6	3.4	1.4	0.3	setosa
## 23	4.6	3.6	1.0	0.2	setosa
## 48	4.6	3.2	1.4	0.2	setosa
## 42	4.5	2.3	1.3	0.3	setosa
## 9	4.4	2.9	1.4	0.2	setosa
## 39	4.4	3.0	1.3	0.2	setosa

```
## 43      4.4      3.2      1.3      0.2      setosa
## 14      4.3      3.0      1.1      0.1      setosa
```

```
iris.new = iris[order(iris$Sepal.Length),] #정렬된 데이터를 저장
head(iris.new)
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 14      4.3      3.0      1.1      0.1      setosa
## 9      4.4      2.9      1.4      0.2      setosa
## 39     4.4      3.0      1.3      0.2      setosa
## 43     4.4      3.2      1.3      0.2      setosa
## 42     4.5      2.3      1.3      0.3      setosa
## 4      4.6      3.1      1.5      0.2      setosa
```

```
iris[order(iris$Species, -iris$Petal.Length, decreasing = T),]# 정렬 기준이 2
개
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 107      4.9      2.5      4.5      1.7      virginica
## 127      6.2      2.8      4.8      1.8      virginica
## 139      6.0      3.0      4.8      1.8      virginica
## 122      5.6      2.8      4.9      2.0      virginica
## 124      6.3      2.7      4.9      1.8      virginica
## 128      6.1      3.0      4.9      1.8      virginica
## 114      5.7      2.5      5.0      2.0      virginica
## 120      6.0      2.2      5.0      1.5      virginica
## 147      6.3      2.5      5.0      1.9      virginica
## 102      5.8      2.7      5.1      1.9      virginica
## 111      6.5      3.2      5.1      2.0      virginica
## 115      5.8      2.8      5.1      2.4      virginica
## 134      6.3      2.8      5.1      1.5      virginica
## 142      6.9      3.1      5.1      2.3      virginica
## 143      5.8      2.7      5.1      1.9      virginica
## 150      5.9      3.0      5.1      1.8      virginica
## 146      6.7      3.0      5.2      2.3      virginica
## 148      6.5      3.0      5.2      2.0      virginica
## 112      6.4      2.7      5.3      1.9      virginica
## 116      6.4      3.2      5.3      2.3      virginica
## 140      6.9      3.1      5.4      2.1      virginica
## 149      6.2      3.4      5.4      2.3      virginica
## 113      6.8      3.0      5.5      2.1      virginica
## 117      6.5      3.0      5.5      1.8      virginica
## 138      6.4      3.1      5.5      1.8      virginica
## 104      6.3      2.9      5.6      1.8      virginica
## 129      6.4      2.8      5.6      2.1      virginica
## 133      6.4      2.8      5.6      2.2      virginica
## 135      6.1      2.6      5.6      1.4      virginica
## 137      6.3      3.4      5.6      2.4      virginica
```

## 141	6.7	3.1	5.6	2.4	virginica
## 121	6.9	3.2	5.7	2.3	virginica
## 125	6.7	3.3	5.7	2.1	virginica
## 145	6.7	3.3	5.7	2.5	virginica
## 105	6.5	3.0	5.8	2.2	virginica
## 109	6.7	2.5	5.8	1.8	virginica
## 130	7.2	3.0	5.8	1.6	virginica
## 103	7.1	3.0	5.9	2.1	virginica
## 144	6.8	3.2	5.9	2.3	virginica
## 101	6.3	3.3	6.0	2.5	virginica
## 126	7.2	3.2	6.0	1.8	virginica
## 110	7.2	3.6	6.1	2.5	virginica
## 131	7.4	2.8	6.1	1.9	virginica
## 136	7.7	3.0	6.1	2.3	virginica
## 108	7.3	2.9	6.3	1.8	virginica
## 132	7.9	3.8	6.4	2.0	virginica
## 106	7.6	3.0	6.6	2.1	virginica
## 118	7.7	3.8	6.7	2.2	virginica
## 123	7.7	2.8	6.7	2.0	virginica
## 119	7.7	2.6	6.9	2.3	virginica
## 99	5.1	2.5	3.0	1.1	versicolor
## 58	4.9	2.4	3.3	1.0	versicolor
## 94	5.0	2.3	3.3	1.0	versicolor
## 61	5.0	2.0	3.5	1.0	versicolor
## 80	5.7	2.6	3.5	1.0	versicolor
## 65	5.6	2.9	3.6	1.3	versicolor
## 82	5.5	2.4	3.7	1.0	versicolor
## 81	5.5	2.4	3.8	1.1	versicolor
## 60	5.2	2.7	3.9	1.4	versicolor
## 70	5.6	2.5	3.9	1.1	versicolor
## 83	5.8	2.7	3.9	1.2	versicolor
## 54	5.5	2.3	4.0	1.3	versicolor
## 63	6.0	2.2	4.0	1.0	versicolor
## 72	6.1	2.8	4.0	1.3	versicolor
## 90	5.5	2.5	4.0	1.3	versicolor
## 93	5.8	2.6	4.0	1.2	versicolor
## 68	5.8	2.7	4.1	1.0	versicolor
## 89	5.6	3.0	4.1	1.3	versicolor
## 100	5.7	2.8	4.1	1.3	versicolor
## 62	5.9	3.0	4.2	1.5	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
## 75	6.4	2.9	4.3	1.3	versicolor
## 98	6.2	2.9	4.3	1.3	versicolor
## 66	6.7	3.1	4.4	1.4	versicolor
## 76	6.6	3.0	4.4	1.4	versicolor
## 88	6.3	2.3	4.4	1.3	versicolor
## 91	5.5	2.6	4.4	1.2	versicolor

## 52	6.4	3.2	4.5	1.5	versicolor
## 56	5.7	2.8	4.5	1.3	versicolor
## 67	5.6	3.0	4.5	1.5	versicolor
## 69	6.2	2.2	4.5	1.5	versicolor
## 79	6.0	2.9	4.5	1.5	versicolor
## 85	5.4	3.0	4.5	1.5	versicolor
## 86	6.0	3.4	4.5	1.6	versicolor
## 55	6.5	2.8	4.6	1.5	versicolor
## 59	6.6	2.9	4.6	1.3	versicolor
## 92	6.1	3.0	4.6	1.4	versicolor
## 51	7.0	3.2	4.7	1.4	versicolor
## 57	6.3	3.3	4.7	1.6	versicolor
## 64	6.1	2.9	4.7	1.4	versicolor
## 74	6.1	2.8	4.7	1.2	versicolor
## 87	6.7	3.1	4.7	1.5	versicolor
## 71	5.9	3.2	4.8	1.8	versicolor
## 77	6.8	2.8	4.8	1.4	versicolor
## 53	6.9	3.1	4.9	1.5	versicolor
## 73	6.3	2.5	4.9	1.5	versicolor
## 78	6.7	3.0	5.0	1.7	versicolor
## 84	6.0	2.7	5.1	1.6	versicolor
## 23	4.6	3.6	1.0	0.2	setosa
## 14	4.3	3.0	1.1	0.1	setosa
## 15	5.8	4.0	1.2	0.2	setosa
## 36	5.0	3.2	1.2	0.2	setosa
## 3	4.7	3.2	1.3	0.2	setosa
## 17	5.4	3.9	1.3	0.4	setosa
## 37	5.5	3.5	1.3	0.2	setosa
## 39	4.4	3.0	1.3	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
## 1	5.1	3.5	1.4	0.2	setosa
## 2	4.9	3.0	1.4	0.2	setosa
## 5	5.0	3.6	1.4	0.2	setosa
## 7	4.6	3.4	1.4	0.3	setosa
## 9	4.4	2.9	1.4	0.2	setosa
## 13	4.8	3.0	1.4	0.1	setosa
## 18	5.1	3.5	1.4	0.3	setosa
## 29	5.2	3.4	1.4	0.2	setosa
## 34	5.5	4.2	1.4	0.2	setosa
## 38	4.9	3.6	1.4	0.1	setosa
## 46	4.8	3.0	1.4	0.3	setosa
## 48	4.6	3.2	1.4	0.2	setosa
## 50	5.0	3.3	1.4	0.2	setosa
## 4	4.6	3.1	1.5	0.2	setosa
## 8	5.0	3.4	1.5	0.2	setosa
## 10	4.9	3.1	1.5	0.1	setosa
## 11	5.4	3.7	1.5	0.2	setosa



## 16	5.7	4.4	1.5	0.4	setosa
## 20	5.1	3.8	1.5	0.3	setosa
## 22	5.1	3.7	1.5	0.4	setosa
## 28	5.2	3.5	1.5	0.2	setosa
## 32	5.4	3.4	1.5	0.4	setosa
## 33	5.2	4.1	1.5	0.1	setosa
## 35	4.9	3.1	1.5	0.2	setosa
## 40	5.1	3.4	1.5	0.2	setosa
## 49	5.3	3.7	1.5	0.2	setosa
## 12	4.8	3.4	1.6	0.2	setosa
## 26	5.0	3.0	1.6	0.2	setosa
## 27	5.0	3.4	1.6	0.4	setosa
## 30	4.7	3.2	1.6	0.2	setosa
## 31	4.8	3.1	1.6	0.2	setosa
## 44	5.0	3.5	1.6	0.6	setosa
## 47	5.1	3.8	1.6	0.2	setosa
## 6	5.4	3.9	1.7	0.4	setosa
## 19	5.7	3.8	1.7	0.3	setosa
## 21	5.4	3.4	1.7	0.2	setosa
## 24	5.1	3.3	1.7	0.5	setosa
## 25	4.8	3.4	1.9	0.2	setosa
## 45	5.1	3.8	1.9	0.4	setosa

- (1) **order()** 함수를 사용하여 iris 데이터셋의 **Sepal.Length** 열을 오름차순으로 정렬
- (2) **Sepal.Length** 열을 기준으로 iris 데이터셋을 오름차순으로 정렬
- (2) **Sepal.Length** 열을 기준으로 iris 데이터셋을 내림차순으로 정렬한 데이터셋을 iris.new 변수에 할당 (4) Species 와 Petal.Length 를 기준으로 내림차순으로 데이터 정렬

## 코드 7-11

```
sp = split(iris, iris$Species) #품종별로 데이터 분리
sp
#분리 결과 확인

## $setosa
##      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
##
## $versicolor
##      Sepal.Length Sepal.Width Petal.Length Petal.Width   Species
## 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
##
## $virginica
##      Sepal.Length Sepal.Width Petal.Length Petal.Width  Species
## 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

```
summary(sp)    #분리 결과 요약
```

```
##           Length Class      Mode
## setosa      5      data.frame list
## versicolor 5      data.frame list
## virginica   5      data.frame list
```

```
sp$setosa      #setosa 품종의 데이터 확인
```

```
##      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
```

(1) iris 의 Species 별로 데이터를 분리 (2) **summary()** 함수를 사용하여 데이터 요약 (3) sp 의 setosa 품종 데이터 호출

### 코드 7-12

```
subset(iris, Species == 'setosa')
```

```
##      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
```

```
subset(iris, Sepal.Length>7.5)
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width  Species
## 106          7.6         3.0         6.6         2.1 virginica
## 118          7.7         3.8         6.7         2.2 virginica
## 119          7.7         2.6         6.9         2.3 virginica
## 123          7.7         2.8         6.7         2.0 virginica
## 132          7.9         3.8         6.4         2.0 virginica
## 136          7.7         3.0         6.1         2.3 virginica
```

```
subset(iris, Sepal.Length > 5.1 & Sepal.Width > 3.9)
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width  Species
## 15          5.8         4.0         1.2         0.2 setosa
## 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
```

```
subset(iris, Sepal.Length > 7.6, select= c(Petal.Length, Petal.Width))
```

```
##      Petal.Length Petal.Width
## 118          6.7         2.2
## 119          6.9         2.3
## 123          6.7         2.0
## 132          6.4         2.0
## 136          6.1         2.3
```

- (1) **subset()** 함수를 이용하여 iris 데이터셋에서 품종이 setosa 인 행들만 추출한다.
- (2) iris 데이터셋에서 꽃받침의 길이가 7.5 보다 큰 행들만 추출한다.
- (3) iris 데이터셋에서 꽃받침의 길이가 5.1 보다 크고 꽃받침의 폭이 3.9 보다 큰 행들만 추출한다.

### 코드 7-13

```
x= 1:100
y= sample(x, size=10, replace= FALSE) #비복원 추출
y
## [1] 92 17 80 75 94 45 3 47 100 8
```

(1)x 변수에 1 부터 100 까지 저장 (2)**sample()** 함수에서 size 는 추출할 값의 개수를 지정하는 매개변수이고, replace= FALSE 는 비복원 추출이다.

### 코드 7-14

```
idx= sample(1:nrow(iris), size= 50,
           replace=FALSE)
iris.50 = iris[idx,] #50 개의 행 추출
dim(iris.50)         #행과 열의 갯수 확인
## [1] 50 5
head(iris.50)
##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 125           6.7         3.3          5.7         2.1  virginica
## 42            4.5         2.3          1.3         0.3   setosa
## 57            6.3         3.3          4.7         1.6 versicolor
## 9             4.4         2.9          1.4         0.2   setosa
## 72            6.1         2.8          4.0         1.3 versicolor
## 55            6.5         2.8          4.6         1.5 versicolor
```

- (1) **sample()** 함수를 이용하여 iris 데이터 셋에서 비복원 추출 방식으로 50 개의 행 인덱스 추출
- (2) iris.50 에 비복원 추출한 50 개의 인덱스를 이용하여 해당하는 row 를 iris.50 에 할당
- (3) 몇 by 몇인지 확인 (4)iris.50 데이터셋의 6 개의 상단 데이터 출력



### 코드 7-15

```
sample(1:20, size=5)
## [1] 10  4  5  2 17

sample(1:20, size=5)
## [1]  5 17 19  4  7

sample(1:20, size=5)
## [1]  9 10  4 20  7

set.seed(100)
sample(1:20, size=5)
## [1] 10  6 16 14 12

set.seed(100)
sample(1:20, size=5)
## [1] 10  6 16 14 12

set.seed(100)
sample(1:20, size=5)
## [1] 10  6 16 14 12
```

- (1) 비복원 추출 함수 **sample()**을 사용하였을때는 출력할때마다 답이 달라진다.  
(2)**set.seed()**함수는 다음번에 다시 추출해도 같은 값이 나오게끔 만들어주는 함수이다.

### 코드 7-16

```
combn(1:5,3)  #1~5 에서 3 개를 뽑는 조합

##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
## [1,]    1    1    1    1    1    1    2    2    2    3
## [2,]    2    2    2    3    3    4    3    3    4    4
## [3,]    3    4    5    4    5    5    4    5    5    5

x= c('red','green','blue','black','white')
com = combn(x,2)      #x 의 원소를 2 개씩 뽑는 조합
com

##      [,1]    [,2]    [,3]    [,4]    [,5]    [,6]    [,7]    [,8]    [,9]
## [1,] "red"    "red"    "red"    "red"    "green" "green" "green" "blue"  "blue"
```

```
## [2,] "green" "blue" "black" "white" "blue" "black" "white" "black" "white"
"
##      [,10]
## [1,] "black"
## [2,] "white"

for(i in 1:ncol(com)){
  cat(com[,i], '\n')
}

## red green
## red blue
## red black
## red white
## green blue
## green black
## green white
## blue black
## blue white
## black white
```

### 코드 7-17

```
agg= aggregate(iris[, -5], by=list(iris$Species), FUN=mean)
agg
```

	Group.1	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
## 1	setosa	5.006	3.428	1.462	0.246
## 2	versicolor	5.936	2.770	4.260	1.326
## 3	virginica	6.588	2.974	5.552	2.026

(1)**aggregate(iris[, -5], by=list(iris\$Species), FUN=mean)**는 품종별로 iris 데이터셋에서 Species 열이 빠진 데이터셋을 기반으로 평균을 계산한다.

### 코드 7-18

```
agg= aggregate(iris[, -5], by=list(표준편차= iris$Species), FUN=sd)
agg
```

	표준편차	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
## 1	setosa	0.3524897	0.3790644	0.1736640	0.1053856
## 2	versicolor	0.5161711	0.3137983	0.4699110	0.1977527
## 3	virginica	0.6358796	0.3224966	0.5518947	0.2746501

(1)**aggregate(iris[, -5], by=list(표준편차= iris\$Species), FUN=sd)**는 iris 데이터셋에서 Species 열을 삭제시킨 데이터셋을 가지고 품종별로 표준편차를 구하는 코드이다.

## 7-19

```
head(mtcars)

##           mpg cyl  disp  hp drat   wt  qsec vs  am gear carb
## Mazda RX4      21.0   6  160 110 3.90 2.620 16.46  0   1    4    4
## Mazda RX4 Wag  21.0   6  160 110 3.90 2.875 17.02  0   1    4    4
## Datsun 710      22.8   4  108  93 3.85 2.320 18.61  1   1    4    1
## Hornet 4 Drive  21.4   6  258 110 3.08 3.215 19.44  1   0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0   0    3    2
## Valiant         18.1   6  225 105 2.76 3.460 20.22  1   0    3    1

agg= aggregate(mtcars, by=list(cyl=mtcars$cyl, vs= mtcars$vs), FUN=max)

agg
##    cyl vs  mpg cyl  disp  hp drat   wt  qsec vs  am gear carb
## 1   4  0 26.0   4 120.3  91 4.43 2.140 16.70  0   1    5    2
## 2   6  0 21.0   6 160.0 175 3.90 2.875 17.02  0   1    5    6
## 3   8  0 19.2   8 472.0 335 4.22 5.424 18.00  0   1    5    8
## 4   4  1 33.9   4 146.7 113 4.93 3.190 22.90  1   1    5    2
## 5   6  1 21.4   6 258.0 123 3.92 3.460 20.22  1   0    4    4
```

## 코드 7-20

```
x= data.frame(name=c('a','b','c'), math=c(90,80,40))
y= data.frame(name=c('a','b','c'), korean=c(75,60,90))
x

##    name math
## 1     a   90
## 2     b   80
## 3     c   40

y

##    name korean
## 1     a     75
## 2     b     60
## 3     c     90
```

두개의 데이터프레임 생성

## 코드 7-21

```
z= merge(x,y, by=c('name'))
z

##    name math korean
## 1     a   90     75
```

```
## 2    b    80    60
## 3    c    40    90
```

**merge(x,y, by=c('name'))** 함수는 x,y 를 병합하고 병합의 기준이 되는 열이 name 임을 의미한다.

## 코드 7-22

```
merge(x,y, all.x=T)

##    name math korean
## 1    a    90     75
## 2    b    80     60
## 3    c    40     90

merge(x,y, all.y=T)

##    name math korean
## 1    a    90     75
## 2    b    80     60
## 3    c    40     90

merge(x,y, all=T)

##    name math korean
## 1    a    90     75
## 2    b    80     60
## 3    c    40     90
```

- (1) **merge()** 함수에서 **all.x=T** 의 의미는 첫 번째 데이터셋의 행들을 모두 출력하고 이행들과 대응도는 행이 두 번째 데이터셋에 있으면 병합해서 출력, 없으면 NA 로 출력하라는 의미
- (2) **merge()** 함수에서 **all.y=T** 의 의미는 두 번째 데이터셋의 행들을 모두 출력하고 이행들과 대응도는 행이 두 번째 데이터셋에 있으면 병합해서 출력, 없으면 NA 로 출력하라는 의미
- (3) \***merge()** 함수에서 **all=T**의 의미는 두 데이터셋에서 공통열의 값들이 어느 쪽에 있더라도 모두 출력하고, 두 데이터셋에서 대응되는 행들이 없으면 NA 로 출력

## 코드 7-23

```
x= data.frame(name=c('a','b','c'), math= c(90,80,40))
y= data.frame(sname= c('a','b','c'), korean=c(75,60,90))
x

##    name math
## 1    a    90
```

```
## 2    b    80
## 3    c    40

y

##      sname korean
## 1      a      75
## 2      b      60
## 3      c      90

merge(x,y, by.x= c('name'), by.y= c('sname'))

##      name math korean
## 1      a    90      75
## 2      b    80      60
## 3      c    40      90
```

**merge()** 함수에서 매개변수 `by.x` 는 두개의 데이터셋 중에서 첫 번째 데이터셋의 병합 기준 열의 이름을 지정하고, `by.y` 는 두번째 데이터셋의 병합 기준 열의 이름을 지정하는 역할을 한다.