



문공 A0015

# R 프로그래밍

김 태 완

kimtwan21@dongduk.ac.kr

# 데이터 탐색하기

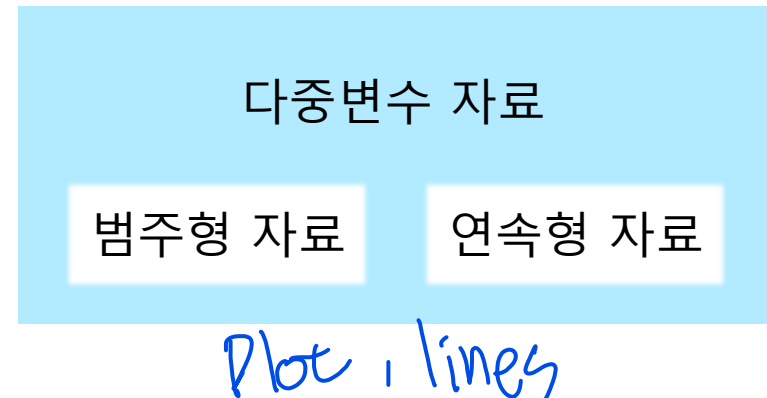
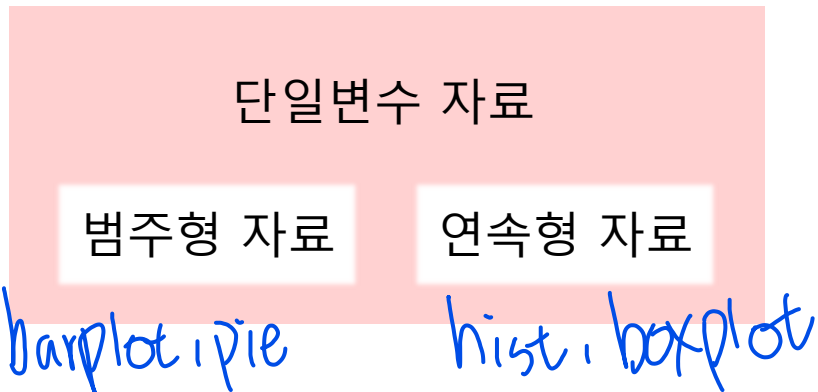
- 자료의 종류
  - 자료의 특성에 따른 종류

범주형 자료(categorical data)	연속형 자료(numerical data)
질적 자료(qualitative data)	양적 자료(quantitative data)
범주 또는 그룹으로 구분할 수 있는 값으로 구성된 자료	크기가 있는 숫자들로 구성된 자료
대소비교나 산술연산이 적용되지 않음	대소비교가 가능하고 산술연산이 가능함
ex) 성별, 혈액형, 찬성 여부 등	ex) 몸무게, 키, 일평균 온도 등

# 데이터 탐색하기

- 자료의 종류
  - 변수의 개수에 따른 종류
    - 변수(variable) : 연구, 조사, 관찰하고 싶은 대상의 특성

단일변수 자료(univariate data)	다중변수 자료(multivariate data)
일변량 자료	다변량 자료
하나의 변수로만 구성된 자료	두 개 이상의 변수로 구성된 자료
벡터에 저장하여 분석	매트릭스, 데이터프레임에 저장하여 분석



자료의 종류

단일변수 자료의 탐색

다중변수 자료의 탐색

## 데이터 탐색하기

- 단일변수 자료의 탐색
  - 단일변수 범주형 자료의 탐색 : 10명의 학생이 선호하는 계절

WINTER	SUMMER	SPRING	SUMMER	SUMMER
FALL	FALL	SUMMER	SPRING	SPRING

‘선호 계절’ -> 단일변수 자료

계절은 크기를 측정할 수 없음 -> 범주형 자료

## 데이터 탐색하기

- 단일변수 자료의 탐색
  - 단일변수 범주형 자료의 탐색 : 10명의 학생이 선호하는 계절
    - **도수분포표**의 작성 : table( ) 함수 이용

```
favorite ← c('WINTER', 'SUMMER', 'SPRING', 'SUMMER', 'SUMMER', 'FALL',  
'FALL', 'SUMMER', 'SPRING', 'SPRING')
```

favorite

```
table(favorite)
```

```
table(favorite)/length(favorite)
```

```
> table(favorite)
```

favorite

FALL	SPRING	SUMMER	WINTER
2	3	4	1

# 데이터 탐색하기

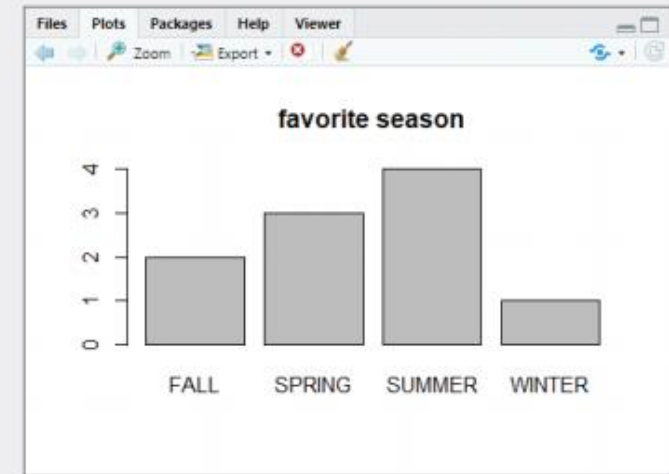
- 단일변수 자료의 탐색
  - 단일변수 범주형 자료의 탐색 : 10명의 학생이 선호하는 계절
    - **막대그래프**의 작성 : `barplot( )` 함수 이용

```
ds ← table(favorite)
ds
```

```
barplot(ds, main='favorite season')
```

*main : 막대그래프 상단의 타이틀을 지정하는 매개변수*

```
> barplot(ds, main='favorite season')
```



## 데이터 탐색하기

- 단일변수 자료의 탐색
  - 단일변수 범주형 자료의 탐색 : 10명의 학생이 선호하는 계절
    - **도수분포표 데이터의 순서 정렬**

```
ds ← table(favorite)
ds
```

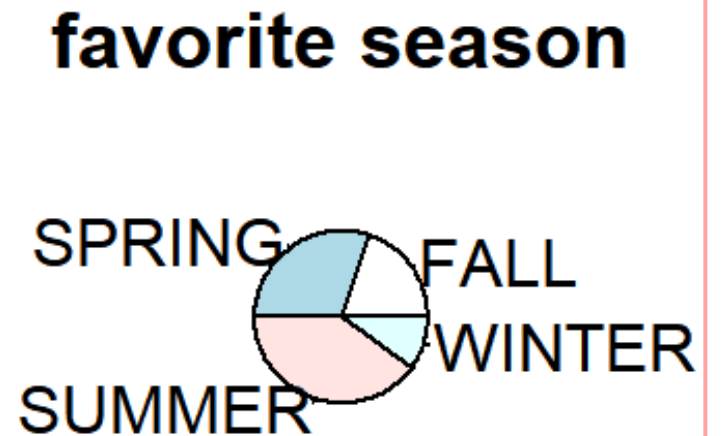
```
ds.new ← ds[c(2, 3, 1, 4)]
ds.new
```

## 데이터 탐색하기

- 단일변수 자료의 탐색
  - 단일변수 범주형 자료의 탐색 : 10명의 학생이 선호하는 계절
    - **원그래프**의 작성 : pie( ) 함수 이용

```
ds ← table(favorite)
ds
```

```
pie(ds, main='favorite season')
```





## 데이터 탐색하기

- 단일변수 자료의 탐색
  - 단일변수 범주형 자료의 탐색
    - 숫자로 표현된 범주형 자료 : 여성 1 / 남성 2 → 여성 < 남성 의미가 아님
  - 15명의 학생이 선호하는 색

2, 3, 1, 1, 2, 2, 1, 3, 2, 1, 3, 2, 1, 2  
(1 = 초록, 2 = 빨강, 3 = 파랑)

'선호 하는 색' → 단일변수 자료

색을 숫자로 표현하였지만 크기를 비교할 수 없음 → 범주형 자료

## 데이터 탐색하기

- 단일변수 자료의 탐색
  - 단일변수 범주형 자료의 탐색
    - 숫자로 표현된 범주형 자료 : 15명의 학생이 선호하는 색

```
favorite.color ← c(2, 3, 2, 1, 1, 2, 2, 1, 3, 2, 1, 3, 2, 1, 2)  
ds ← table(favorite.color)
```

```
barplot(ds, main='favorite color')  
colors ← c('green', 'red', 'blue')  
names(ds) ← colors
```

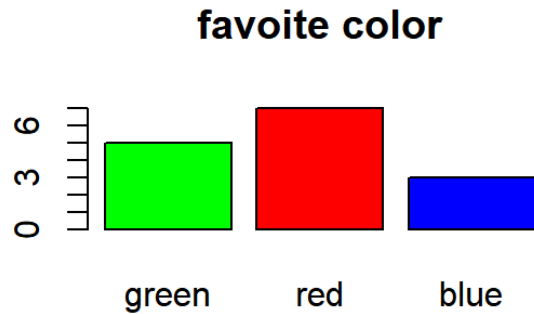
```
barplot(ds, main='favorite color', col=colors)  
pie(ds, main='favorite color', col=colors)
```

# 데이터 탐색하기

- 단일변수 자료의 탐색
  - 단일변수 범주형 자료의 탐색
    - 숫자로 표현된 범주형 자료 : 15명의 학생이 선호하는 색

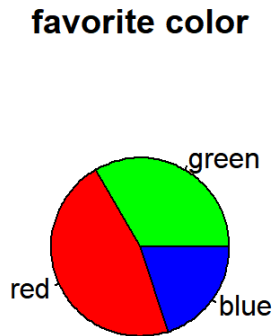
> `barplot(ds, main = 'favoite color', col = colors)`

# 색을 지정하여 막대그래프 작성



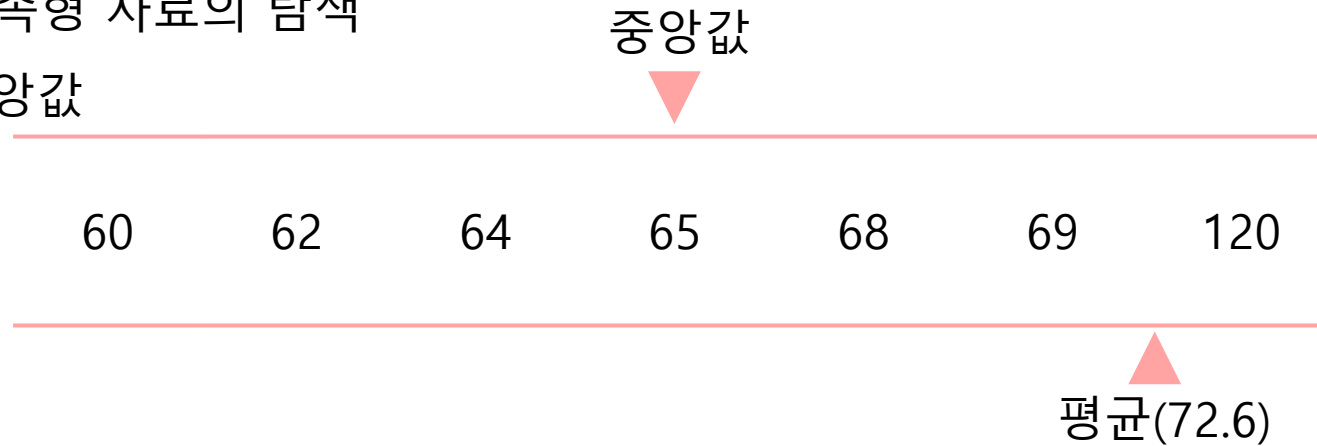
> `pie(ds, main = 'favorite color', col = colors)`

# 색을 지정하여 원그래프 작성



# 데이터 탐색하기

- 단일변수 자료의 탐색
  - 단일변수 연속형 자료의 탐색
    - 평균과 중앙값



- 평균(mean) : 자료의 값들을 모두 합산한 수 값들의 개수로 나눈 값
  - > 특이값에 영향 받음(특이값이 있으면 평균이 치우침)
- 중앙값(median) : 자료의 값들을 일렬로 줄 세웠을 때 가장 중앙에 위치하는 값
  - > 특이값에 영향 받지 않음
- 절사평균(trimmed mean) : 자료의 관측값들 중에서 작은 값들의 하위 n%와 큰 값들의 상위 n%를 제외하고 중간에 있는 나머지 값들로 평균을 계산하는 방식 (특이값에 영향을 받는 평균의 특징을 완화)

## 데이터 탐색하기

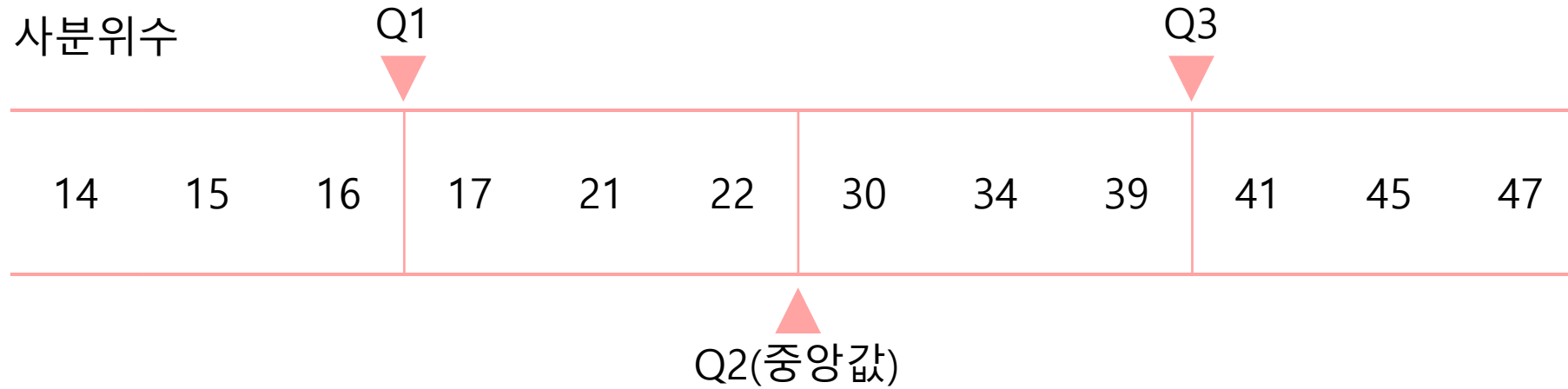
- 단일변수 자료의 탐색
  - 단일변수 연속형 자료의 탐색
    - mean( ), median( ) 함수 이용하여 평균 구하기 + trim 매개변수 이용하여 절사평균 구하기

```
weight ← c(60, 62, 64, 65, 68, 69, 120)  
weight.heavy ← c(weight, 120)
```

```
mean(weight)  
mean(weight.heavy)  
median(weight)  
median(weight.heavy)  
mean(weight, trim=0.2)  
mean(weight.heavy, trim=0.2)
```

# 데이터 탐색하기

- 단일변수 자료의 탐색
  - 단일변수 연속형 자료의 탐색
    - 사분위수



- 사분위수(quartile) : 주어진 자료에 있는 값들을 크기순으로 나열했을 때 이것을 4등분하는 지점에 있는 값들
  - > 4등분하면 생기는 3개의 등분점을 순서대로 1사분위수(Q1), 2사분위수(Q2), 3사분위수(Q3) 라고 함
  - > 2사분위수(Q2)는 중앙값과 동일

## 데이터 탐색하기

- 단일변수 자료의 탐색
  - 단일변수 연속형 자료의 탐색
    - 사분위수 : `quantile( )` 함수 이용, `summary( )` 함수 이용

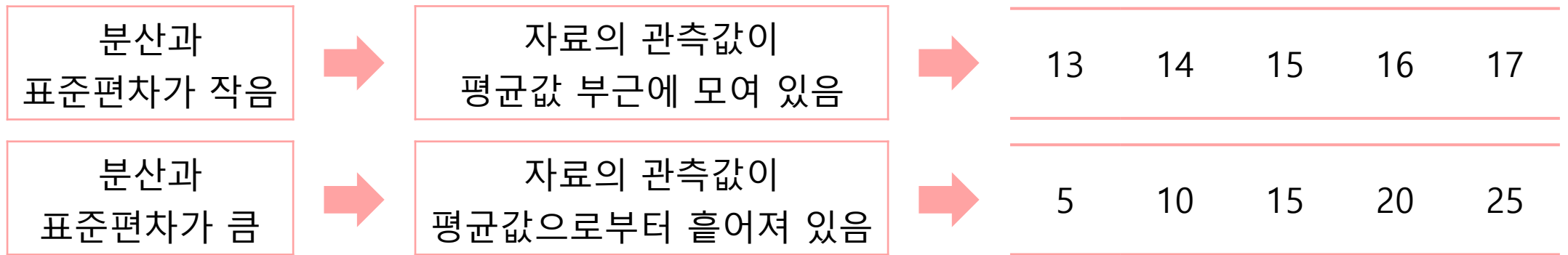
```
mydata ← c(60, 62, 64, 65, 68, 69, 120)
quantile(mydata)
```

```
quantile(mydata, (0:10)/10)  0~10의 정수를 10으로 나누라는 것을 의미
summary(mydata)             -> 0.1 ~ 1.0(10% ~ 100%)
```

# `summary( )` 함수 : 사분위수(1st Qu., Median., 3rd Qu.)  
최댓값(Max.), 최솟값(Min.), 평균(Mean)을 함께 출력함

# 데이터 탐색하기

- 단일변수 자료의 탐색
  - 단일변수 연속형 자료의 탐색
    - 산포 :



- 산포(distribution) : 주어진 자료에 있는 값들이 퍼져 있는 정도
- 분산(variance) : 주어진 자료의 각각의 값들이 평균으로 떨어져 있는 정도를 계산하여 합산한 후 값들의 개수로 나누어 계산
- 표준편차(standard deviation) : 분산의 제곱근



# 데이터 탐색하기

- 단일변수 자료의 탐색
  - 단일변수 연속형 자료의 탐색
    - 산포 : var( )함수, sd( )함수, range( )함수, diff( )함수 이용

```
mydata ← c(60, 62, 64, 65, 68, 69, 120)
```

```
var(mydata)
```

```
sd(mydata)
```

```
range(mydata)
```

```
diff(range(mydata))
```

# 데이터 탐색하기

- 단일변수 자료의 탐색
  - 단일변수 연속형 자료의 탐색
    - 히스토그램

연속형 자료



히스토그램

자료의 구간을 나누고 구간에 속  
하는 값들의 개수를 셈

막대들이 붙어 있음

hist( ) 함수 이용

범주형 자료



막대그래프

값의 종류별로 개수를 셈

막대 사이에 간격이 있음

barplot( ) 함수 이용

자료의 종류

단일변수 자료의 탐색

다중변수 자료의 탐색

## 데이터 탐색하기

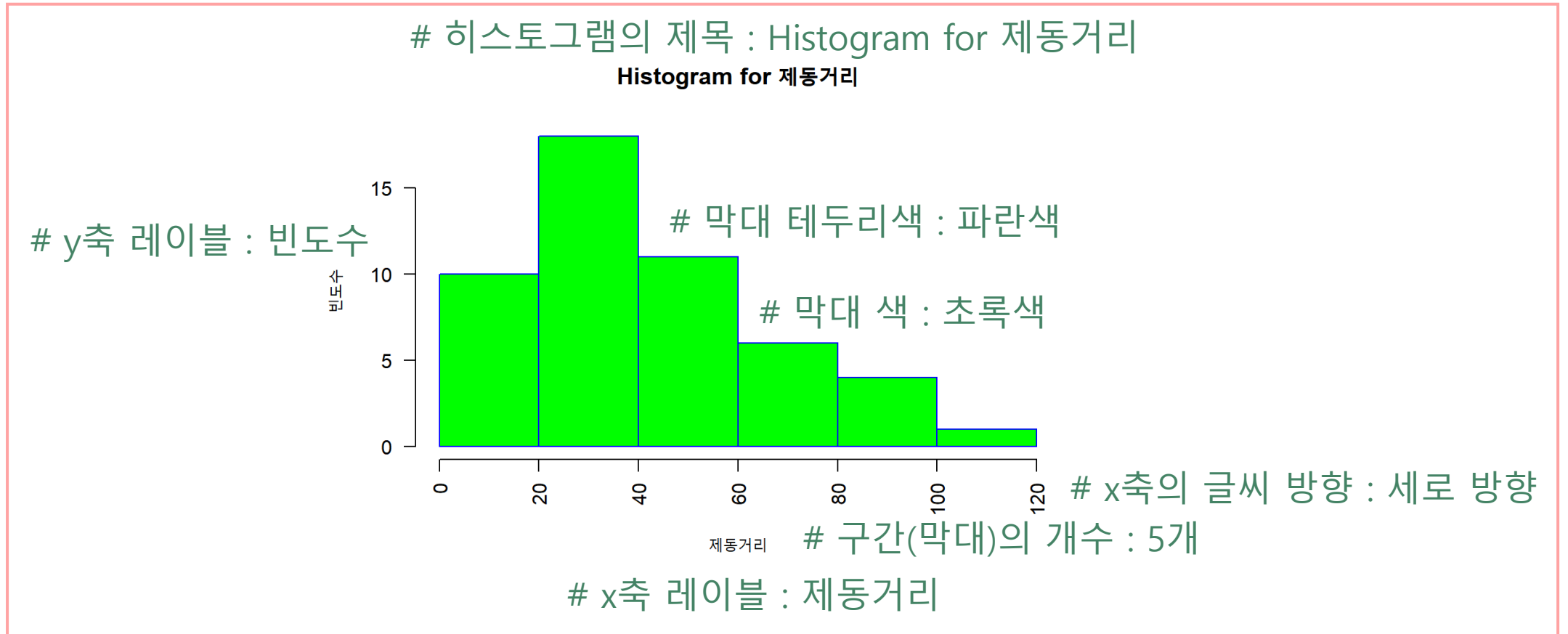
- 단일변수 자료의 탐색
  - 단일변수 연속형 자료의 탐색
    - **히스토그램** : hist( ) 함수 이용

```
dist ← cars[,2]
hist(dist,
      main="Histogram for 제동거리",
      xlab ="제동거리",
      ylab="빈도수",
      border="blue",
      col="green",
      las=2,
      breaks=5)

# 자동차 제동거리
# 자료(data)
# 제목
# x축 레이블
# y축 레이블
# 막대 테두리색
# 막대 색
# x축 글씨 방향(0~3)
# 막대 개수 조절
```

# 데이터 탐색하기

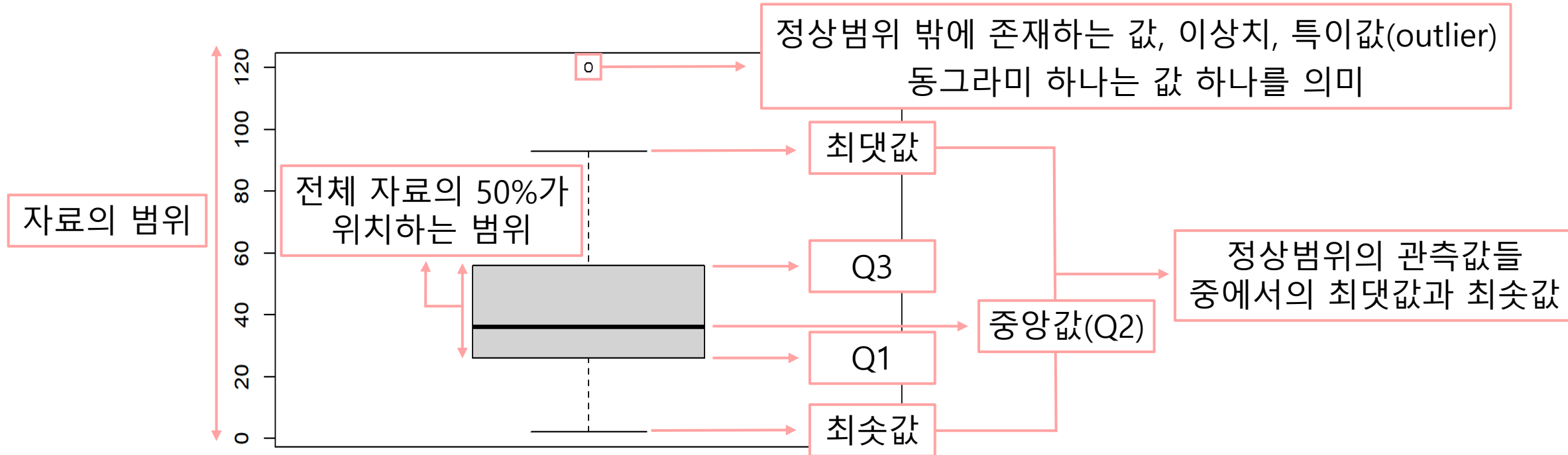
- 단일변수 자료의 탐색
  - 단일변수 연속형 자료의 탐색
    - 히스토그램 : hist( ) 함수 이용



# 데이터 탐색하기

- 단일변수 자료의 탐색
  - 단일변수 연속형 자료의 탐색
    - 상자그림

상자그림(box plot) : 사분위수를 시각화하여 그래프 형태로 나타낸 것  
(상자그림의 '상자'는 안쪽 작은 상자를 의미)



자료의 종류

단일변수 자료의 탐색

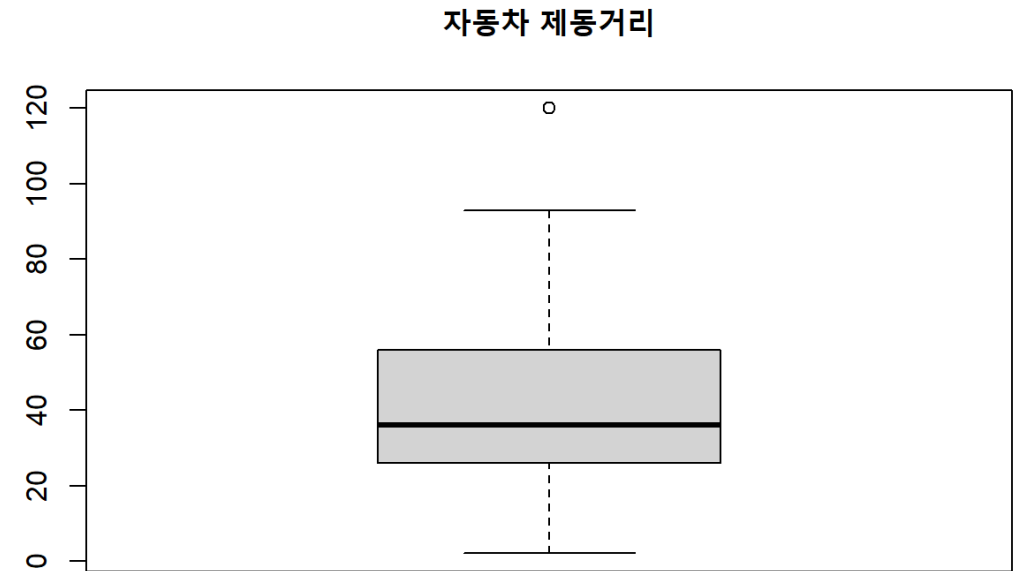
다중변수 자료의 탐색

# 데이터 탐색하기

- 단일변수 자료의 탐색
  - 단일변수 연속형 자료의 탐색
    - 상자그림 : `boxplot( )` 함수 이용

```
dist ← cars[,2]
```

```
boxplot(dist, main="자동차 제동거리")  
boxplot.stats(dist)
```



# 데이터 탐색하기

- 단일변수 자료의 탐색
  - 단일변수 연속형 자료의 탐색
    - 그룹이 있는 자료의 상자그림

```
> levels(iris[,5])
```

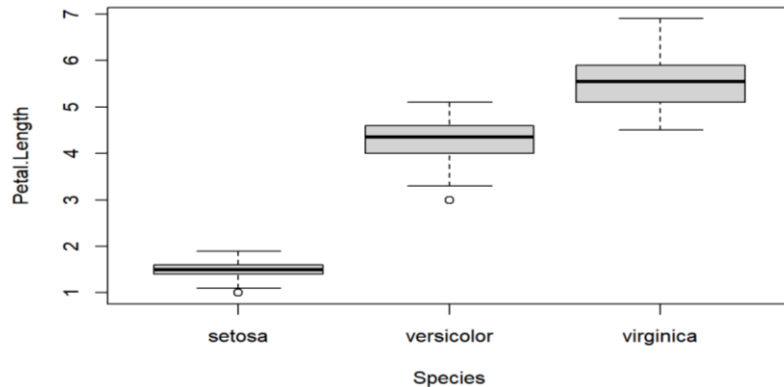
```
[1] "setosa" "versicolor" "virginica"
```

사용할 데이터셋 : iris

```
> boxplot(Petal.Length~Species, data = iris, main = '품종별 꽃잎의 길이')
```

Petal.Length(품종의 길이)를 Species(품종)별로 나누어 상자그림을 그림

품종별 꽃잎의 길이



# iris의 다섯 번째 열(품종)의 종류

# 3개의 종류로 구성

-> iris 데이터셋은 3그룹의 품종으로 구성됨

상자그림의 제목


# 상자그림을 그릴 대상 자료와 그룹의 정보  
자료가 동일한 데이터셋에 있지 않을 때

-> boxpot(iris\$Petal.Length~iris\$Species)  
와 같은 문법 사용  
(각각 자료와 데이터셋을 지정)

# 데이터 탐색하기

- 다중변수 자료의 탐색
  - 다중변수 자료는 2차원 형태를 나타내며, 이는 매트릭스나 데이터 프레임에 저장하여 분석
  - 산점도 (scatter plot)
    - 2개의 변수로 구성된 자료의 분포를 알아보는 그래프

변수



	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

관측값

자료의 종류

단일변수 자료의 탐색

다중변수 자료의 탐색



## 데이터 탐색하기

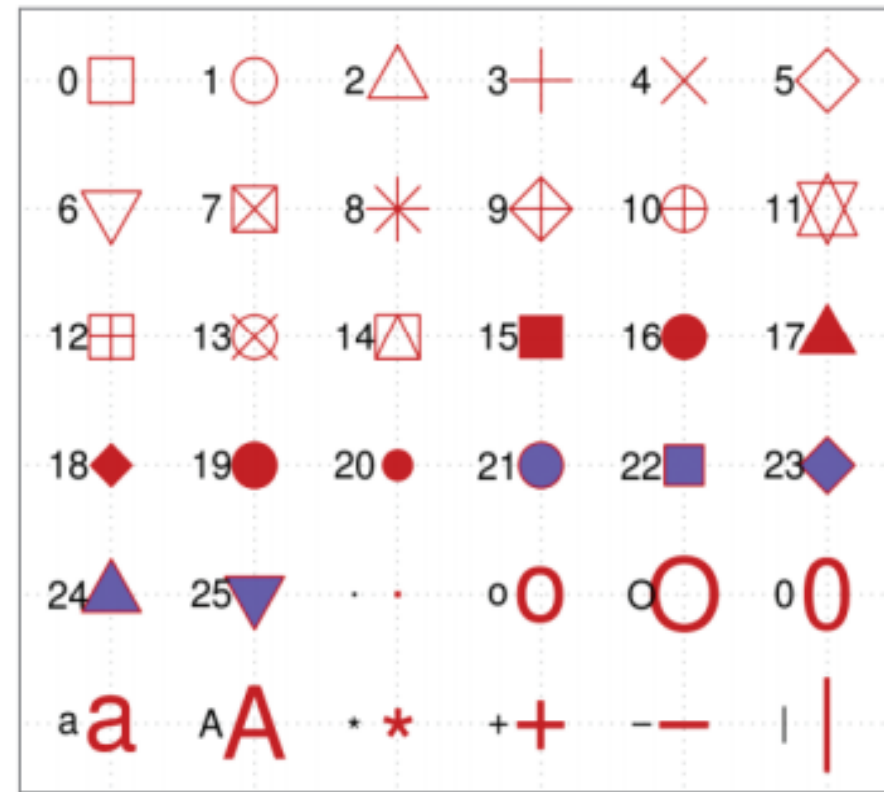
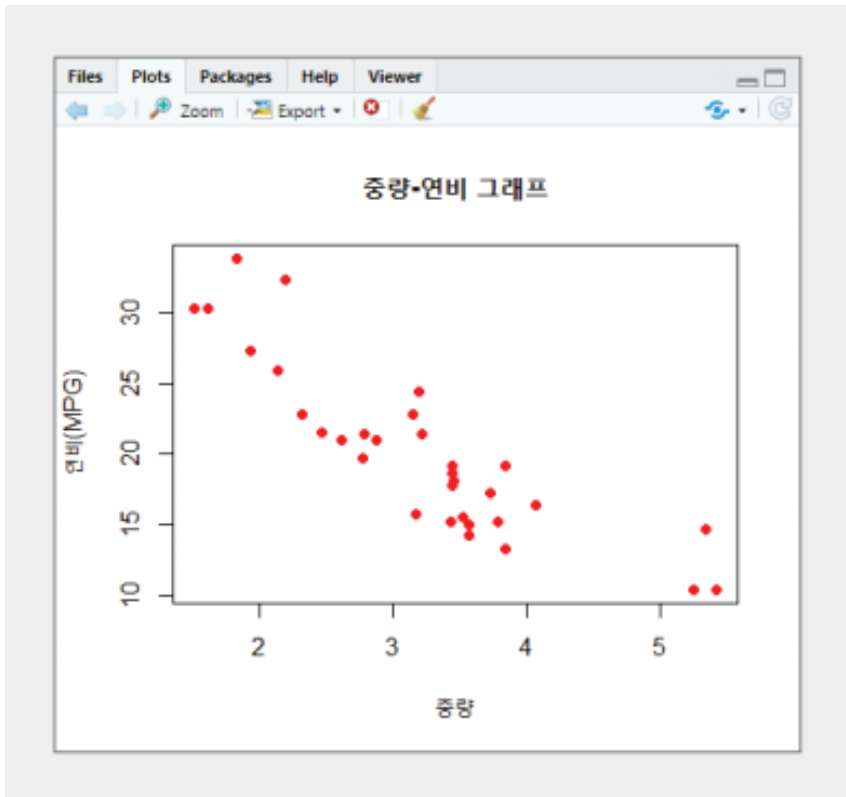
- 다중변수 자료의 탐색
  - 산점도
    - mtcars 데이터셋에서 자동차의 중량(wt)과 연비(mpg) 사이의 관계

```
wt ← mtcars$wt          # 중량 자료
mpg ← mtcars$mpg         # 연비 자료

plot(wt, mpg,            # 2개 변수(x축, y축)
      main="중량-연비 그래프", # 제목
      xlab="중량",          # x축 레이블
      ylab="연비(MPG)",     # y축 레이블
      col="red",           # point의 color
      pch=19)              # point의 종류
```

# 데이터 탐색하기

- 다중변수 자료의 탐색
  - 산점도
    - 두 변수 사이의 산점도 : `plot()` 함수의 `pch` 매개변수



## 데이터 탐색하기

- 다중변수 자료의 탐색
  - 산점도
    - 두 변수 사이의 산점도 : `plot( )` 함수의 매개변수 지정

```
plot(mtcars$wt, mtcars$mpg,  
...)
```

||

```
plot(mtcars[, c('wt', 'mpg')],  
...)
```

||

```
plot(mpg~wt, data=mtcars,  
...)
```

# 데이터프레임 열 이름으로 값 추출하기

# 데이터프레임 인덱스로 값 추출하기

# '~' 이용해 값 추출하기

# 매개변수를 지정할 때 값의 순서에 주의

## 데이터 탐색하기

- 다중변수 자료의 탐색
  - 여러 변수들 간의 산점도
    - `pairs()` 함수 이용

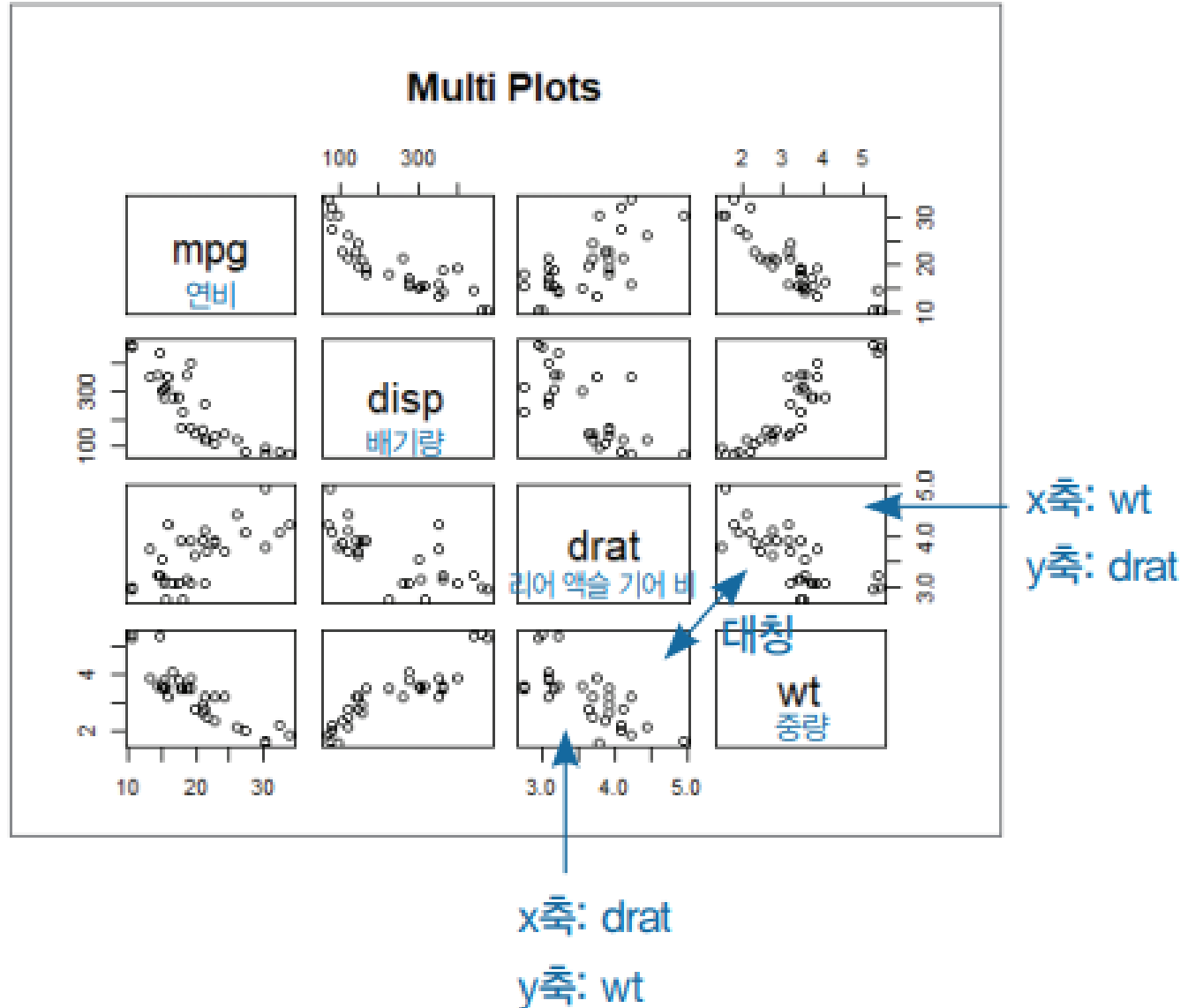
```
vars ← c("mpg", "disp", "drat", "wt")
target ← mtcars[,vars]
head(target)
pairs(target, main="Multi Plots")
```

```
> vars <- c("mpg","disp","drat","wt") # C
> target <- mtcars[,vars]
> head(target)
```

	mpg	disp	drat	wt
Mazda RX4	21.0	160	3.90	2.620
Mazda RX4 Wag	21.0	160	3.90	2.875
Datsun 710	22.8	108	3.85	2.320
Hornet 4 Drive	21.4	258	3.08	3.215
Hornet Sportabout	18.7	360	3.15	3.440
Valiant	18.1	225	2.76	3.460

## 데이터 탐색하기

- 다중변수 자료의 탐색
  - 여러 변수들 간의 산점도
    - `pairs( )` 함수 이용



## 데이터 탐색하기

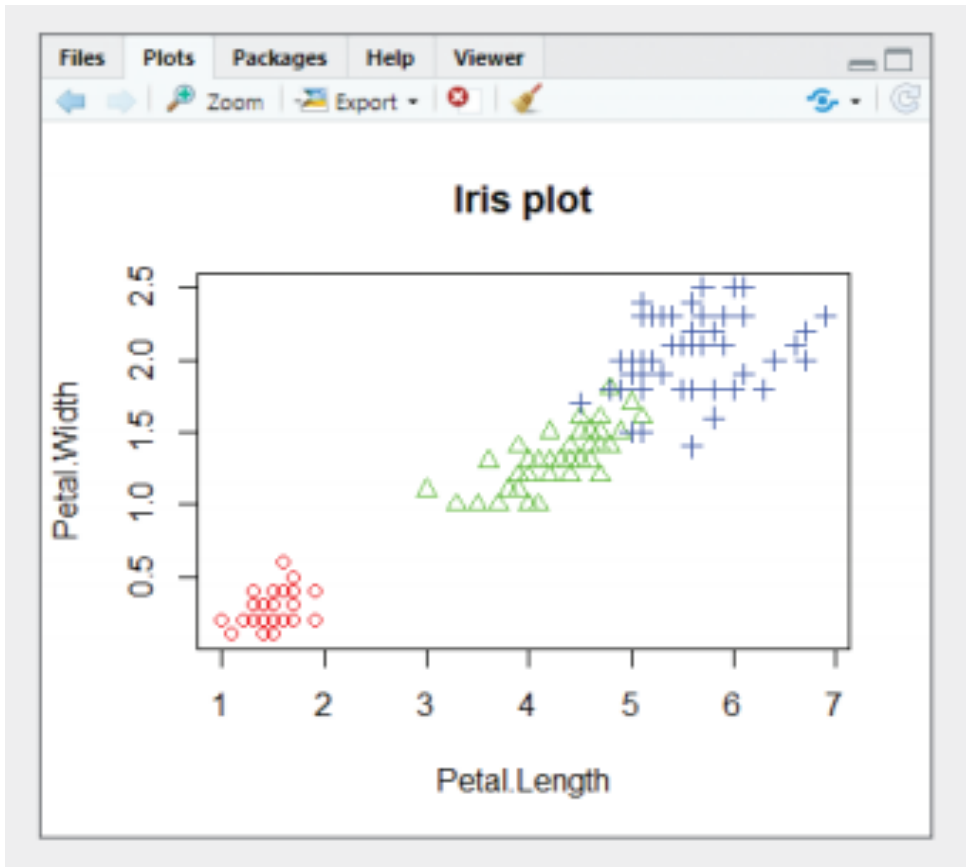
- 다중변수 자료의 탐색
  - 그룹 정보가 있는 두 변수의 산점도
    - 그룹 정보를 알고 있다면 산점도를 작성 시 각 그룹별 관측값들을 다른 색깔과 점의 모양으로 표시

```
iris.2 ← iris[,3:4] # 데이터 준비
point ← as.numeric(iris$Species) # 점의 모양

color ← c("red", "green", "blue") # 점의 컬러
plot(iris.2,
     main="Iris plot",
     pch=c(point),
     col=color[point])
```

# 데이터 탐색하기

- 다중변수 자료의 탐색
  - 그룹 정보가 있는 두 변수의 산점도
    - 그룹 정보를 알고 있다면 산점도를 작성 시 각 그룹별 관측값들을 다른 색깔과 점의 모양으로 표시



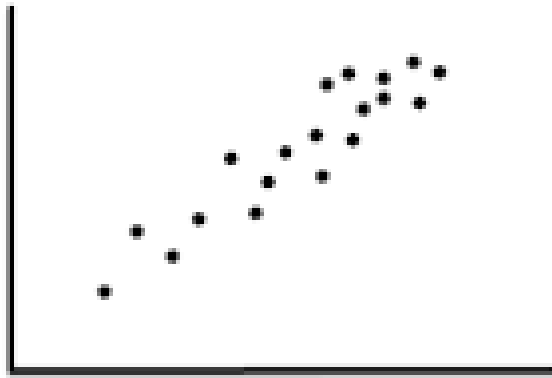
- Petal.Length의 길이가 길수록 Petal.Width도 커짐
- setosa 품종은 다른 두 품종에 비해 꽃잎의 길이와 폭이 확연히 작음
- virginica 품종은 다른 두 품종에 비해 꽃잎의 길이와 폭이 제일 큼

# 데이터 탐색하기

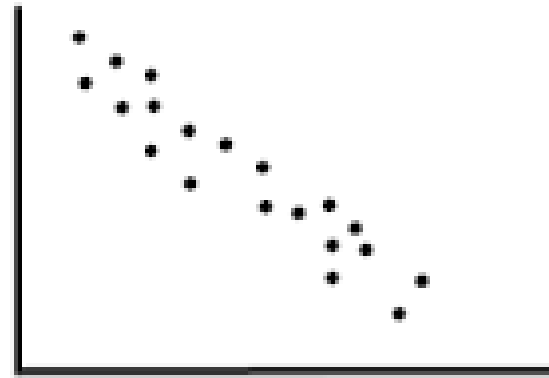
- 다중변수 자료의 탐색

- 상관분석

- 추세의 모양이 선(線, line) 모양이어서 종량과 연비는 '선형적 관계'에 있다고 표현
    - 선형적 관계라고 해도 강한 선형적 관계가 있고 약한 선형적 관계도 있음
    - 상관분석(correlation analysis) : 얼마나 선형성을 보이는지 수치상으로 나타낼 수 있는 방법



양(+)의 상관관계



음(-)의 상관관계



# 데이터 탐색하기

- 다중변수 자료의 탐색

- 상관계수

- 피어슨 상관계수(Pearson's correlation coefficient)

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \rightarrow \text{식 시험 X}$$

- 1 ≤ r ≤ 1
    - r > 0 : 양의 상관관계(x가 증가하면 y도 증가)
    - r < 0 : 음의 상관관계(x가 증가하면 y는 감소)
    - r이 1이나 -1에 가까울수록 x, y의 상관성이 높다(관측값들의 분포가 직선에 가까워짐)

# 데이터 탐색하기

- 다중변수 자료의 탐색

- 상관분석

- R을 이용한 상관계수의 계산 : 음주정도와 혈중알코농도의 상관성

```
> beers ← c(5, 2, 9, 8, 3, 7, 3, 5, 3, 5) # 맥주를 마신 잔 수 beers에 입력
> bal ← c(0.1, 0.03, 0.19, 0.12, 0.04, 0.0095, 0.07, 0.06, 0.02, 0.05) # 혈중알코농도 bal에 입력
> tbl ← data.frame(beers, bal) # 데이터프레임 생성
> tbl
  beers  bal
1     5 0.100
2     2 0.030
3     9 0.190
...
```

# 데이터 탐색하기

- 다중변수 자료의 탐색

- 상관분석

- R을 이용한 상관계수의 계산 : 음주정도와 혈중 알코농도의 상관성

beers	5	2	9	8	3	7	3	5	3	5
bal	0.10	0.03	0.19	0.12	0.04	0.095	0.07	0.06	0.02	0.05

```
beers = c(5,2,9,8,3,7,3,5,3,5)
```

```
bal ← c(0.1,0.03,0.19,0.12,0.04,0.0095,0.07,0.06,0.02,0.05)
```

```
tbl ← data.frame(beers,bal)
```

```
plot(bal~beers, data=tbl)
```

```
res ← lm(bal~beers,data=tbl)
```

```
abline(res)
```

```
cor(beers,bal)
```

```
# 산점도
```

```
# 회귀식 도출
```

```
# 회귀선 그리기
```

```
# 상관계수 계산
```

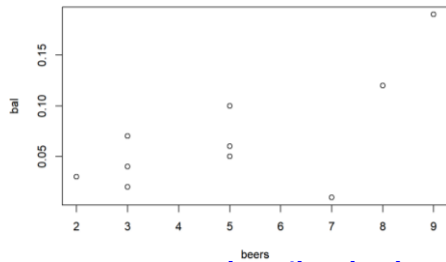
# 데이터 탐색하기

- 다중변수 자료의 탐색

- 상관분석

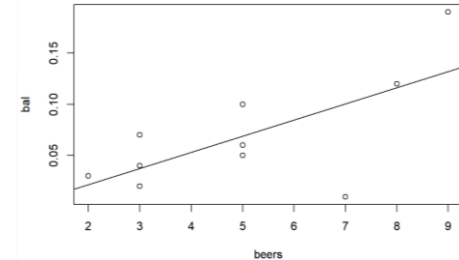
- R을 이용한 상관계수의 계산 : `lm( )` 함수와 `abline` 함수, `cor( )` 함수 이용하기

```
> plot(bal~beers, data = tbl)
```



```
> res <- lm(bal~beers, data = tbl)
```

```
> abline(res)
```



```
> cor(beers, bal)
```

```
[1] 0.6797025
```

# x축 : bal, y축 : beers 산점도 생성

두 변수의 선형 관계를 가장 잘 나타낼 수 있는 선의 식  
( $y = ax + b$  형태의 1차식)

# `lm( )` 함수 : 회귀식을 자동으로 찾음

# `abline( )` 함수 : 회귀식을 이용하여 산점도 위에  
회귀선을 그림

# `cor( )` 함수 : 상관계수를 계산함

# 반올림하여 0.68로 상관성이 높음

## 데이터 탐색하기

- 다중변수 자료의 탐색
  - 상관분석
    - R을 이용한 상관계수의 계산 : 음주정도와 혈중 알코농도의 상관성

```
cor(iris[,1:4])          # 4개 변수 간 상관성 분석
```

```
> cor(iris[,1:4])          # 4개 변수 간 상관성 분석
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	1.0000000	-0.1175698	0.8717538	0.8179411
Sepal.Width	-0.1175698	1.0000000	-0.4284401	-0.3661259
Petal.Length	0.8717538	-0.4284401	1.0000000	0.9628654
Petal.Width	0.8179411	-0.3661259	0.9628654	1.0000000

# 데이터 탐색하기

- 다중변수 자료의 탐색
  - 선그래프
  - 시계열 자료의 분석 : 학급의 월별 지각생 통계

month	1	2	3	4	5	6	7	8	9	10	11	12
late	5	8	7	9	4	6	12	13	8	6	6	4

시계열 자료(times series data) : 시간의 변화에 따라 수집한 자료  
-> 선그래프를 통해 증감 추이를 확인할 수 있음

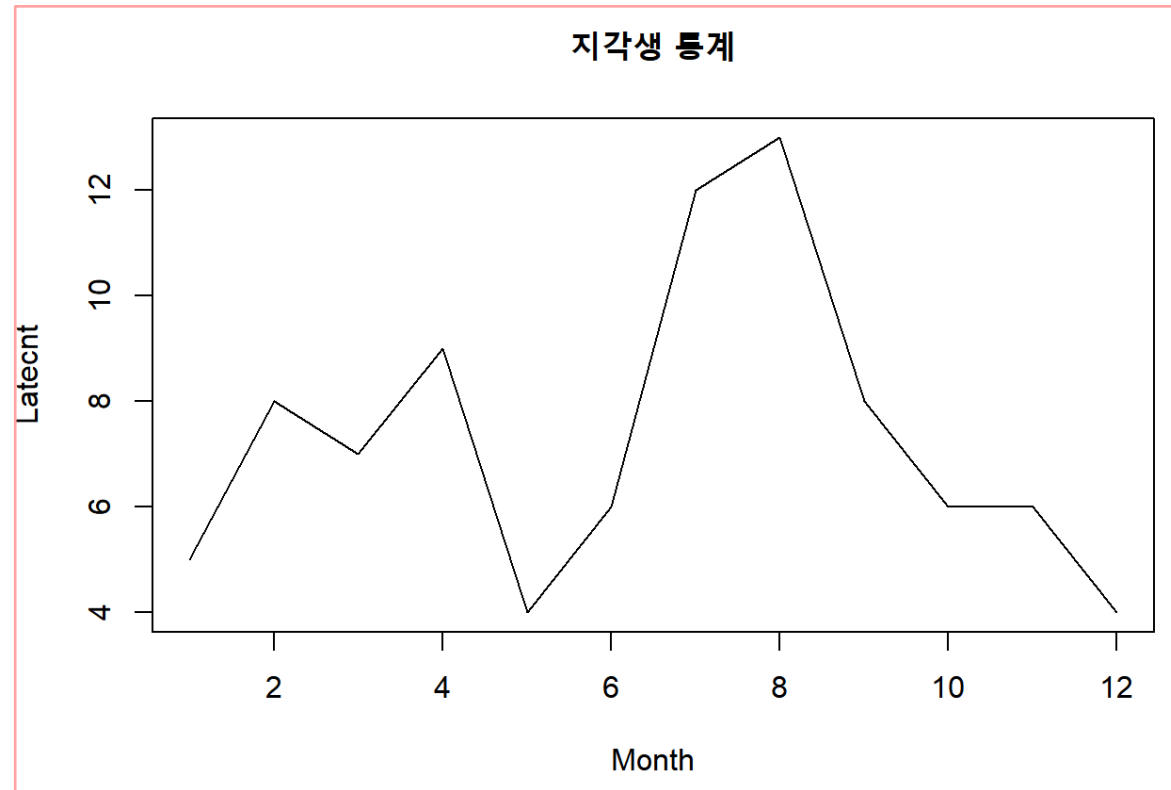
## 데이터 탐색하기

- 다중변수 자료의 탐색
  - 선그래프

```
month = 1:12 # 자료 입력
late = c(5,8,7,9,4,6,12,13,8,6,6,4)
plot(month,                                     # x data
      late,                                    # y data
      main="지각생 통계",                     # 제목
      type="l",                               # 그래프의 종류 선택(알파벳)
      lty=1,                                  # 선의 종류(line type) 선택
      lwd=1,                                  # 선의 굵기 선택
      xlab="Month",                           # x축 레이블
      ylab="Late cnt"                         # y축 레이블
)
```

# 데이터 탐색하기

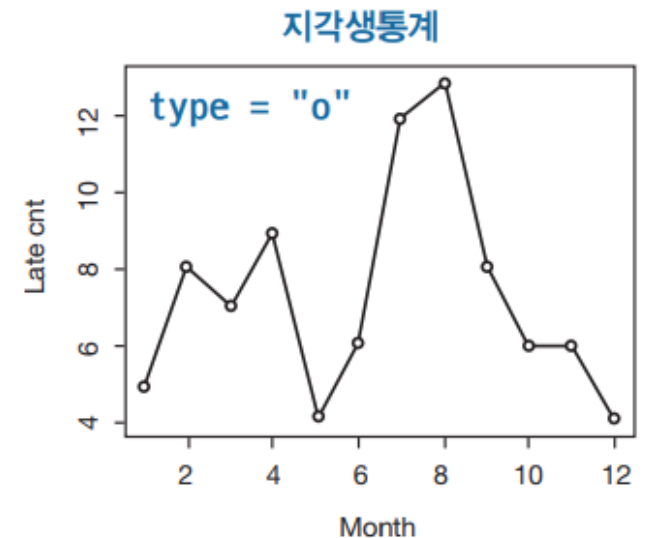
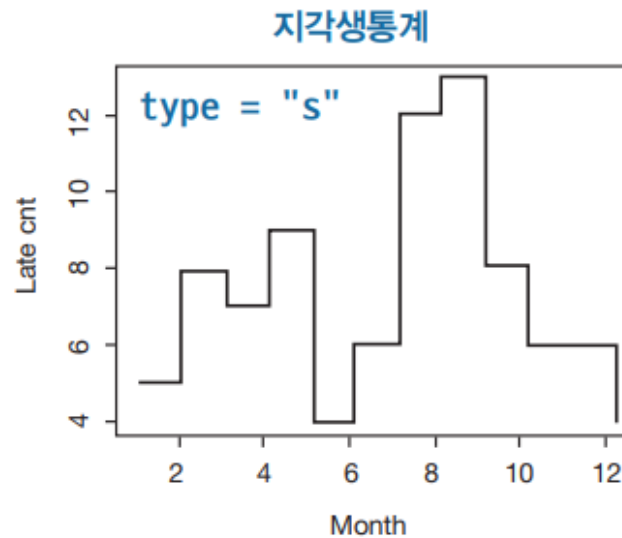
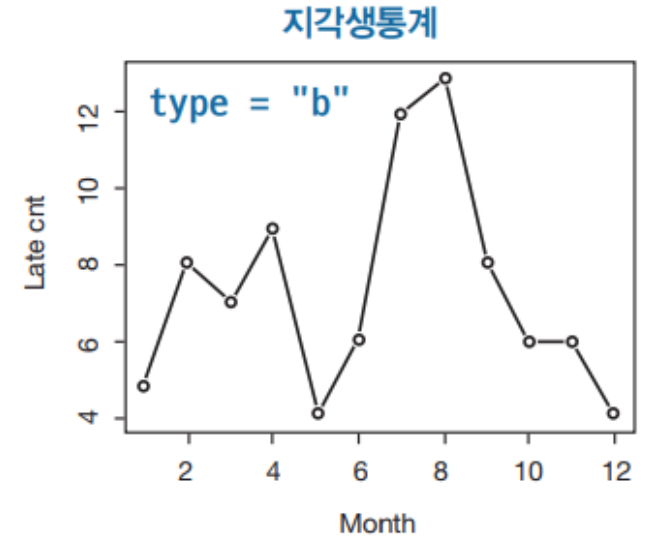
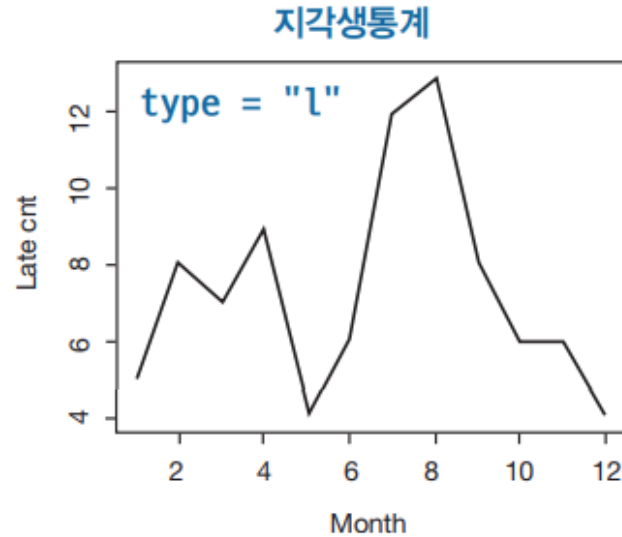
- 다중변수 자료의 탐색
  - 선그래프
    - 선그래프의 작성 : 학급의 월별 지각생 통계





# 데이터 탐색하기

- 다중변수 자료의 탐색
  - 선그래프
    - plot( )함수의 type 매개변수



# 데이터 탐색하기

- 다중변수 자료의 탐색
  - 선그래프
    - plot( )함수의 lty 매개변수



# 데이터 탐색하기

- 다중변수 자료의 탐색

- 복수의 선그래프의 작성

- 어느 학급의 월별 지각생 통계

month	1	2	3	4	5	6	7	8	9	10	11	12
late1	5	8	7	9	4	6	12	13	8	6	6	4
late2	4	6	5	8	7	8	10	11	6	5	7	3

```
month = 1:12
```

```
late1 = c(5,8,7,9,4,6,12,13,8,6,6,4)
```

```
late2 = c(4,6,5,8,7,8,10,11,6,5,7,3)
```

```
plot(month,
```

```
late1,
```

```
main="Late Students",
```

```
type= "b",
```

```
lty=1,
```

```
col="red",
```

```
xlab="Month ",
```

```
ylab="Late cnt",
```

```
ylim=c(1, 15)
```

```
)
```

```
# x data
```

```
# y data
```

```
# 그래프의 종류 선택(알파벳)
```

```
# 선의 종류(line type) 선택
```

```
# 선의 색 선택
```

```
# x축 레이블
```

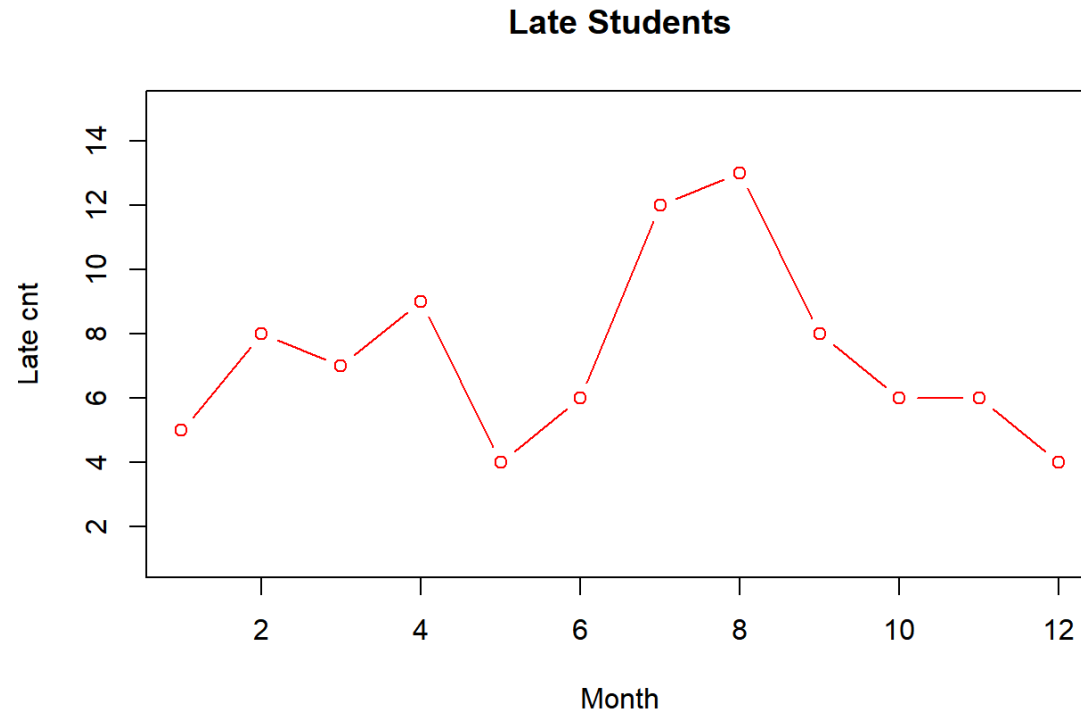
```
# y축 레이블
```

```
# y축 값의 (하한, 상한)
```

# 데이터 탐색하기

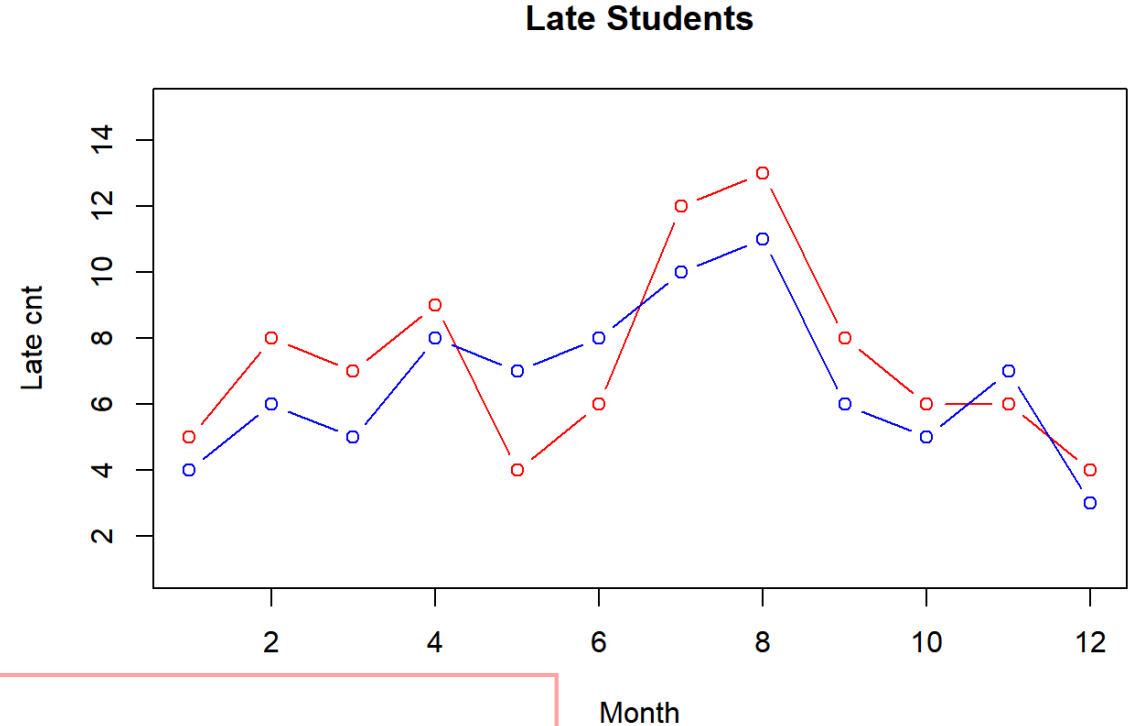
- 다중변수 자료의 탐색
  - 복수의 선그래프의 작성
    - 어느 학급의 월별 지각생 통계

# y축 : 1~15



# 데이터 탐색하기

- 다중변수 자료의 탐색
  - 복수의 선그래프의 작성: lines( ) 함수 이용
  - lines( ) 함수
    - : plot( ) 함수로 작성한 그래프 위에 선을 겹쳐서 그리는 역할



```
lines(month,          # x data
      late2,          # y data
      type = "b",     # 선의 종류(line type) 선택
      col = "blue")   # 선의 색 선택
```

## 예시

- 예시 1 : 성별 분석

```
1 gender ← c('F','F','F','M','M','F','F','F','M','M')  
# coding here #
```

- 실행 결과

```
gender ← c('F','F','F','M','M','F','F','F','M','M')
```

```
table(gender)
```

```
gender
```

```
  F  M
```

```
 6  4
```

# gender 도수분포표 작성

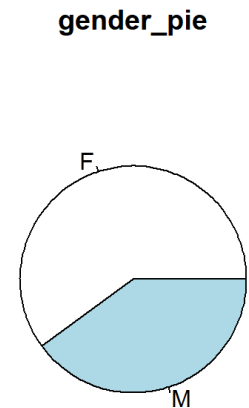
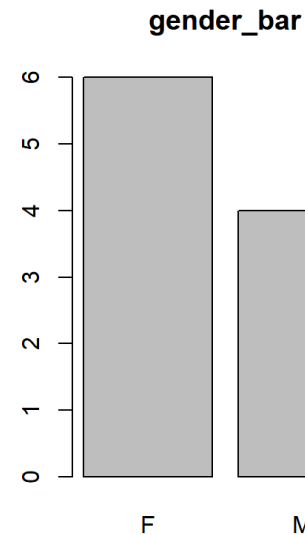
# 예시

- 예시 2 : 성별 분석

```
1 gender <- c('F','F','F','M','M','F','F','F','M','M')
2 par(mfrow=c(1,2))
3 gt <- table(gender)
# coding here #
```

- 실행 결과

```
> gender <- c('F','F','F','M','M','F','F','F','M','M')
> par(mfrow=c(1,2))
# 1행 2열로 화면 분할
> gt <- table(gender)
> 
# 막대그래프 출력 / 제목 : 'gender_bar'
> 
# 원그래프 출력 / 제목 : 'gender_pie'
```



## 예시

- 예시 3 : 과목별 성적 분석

```
1 score ← c(90, 85, 73, 80, 85, 65, 78, 50, 68, 96)
# coding here #
```

- 실행 결과

```
> score ← c(90, 85, 73, 80, 85, 65, 78, 50, 68, 96)
>  # score의 평균 구하기
[1] 77
>  # score의 중앙값 구하기
[1] 79
>  # score의 절사평균 구하기(절사범위 : 10%)
[1] 78
>  # score의 표준편차 구하기
[1] 13.49074
```



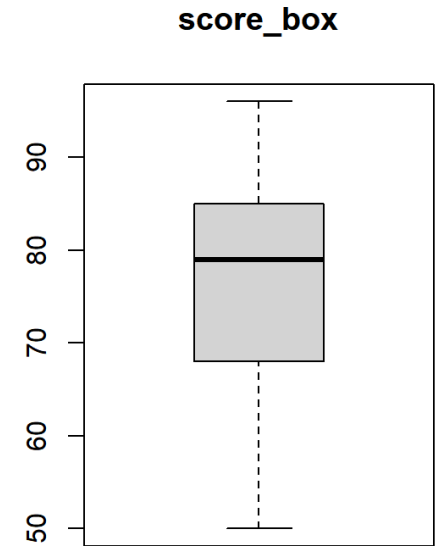
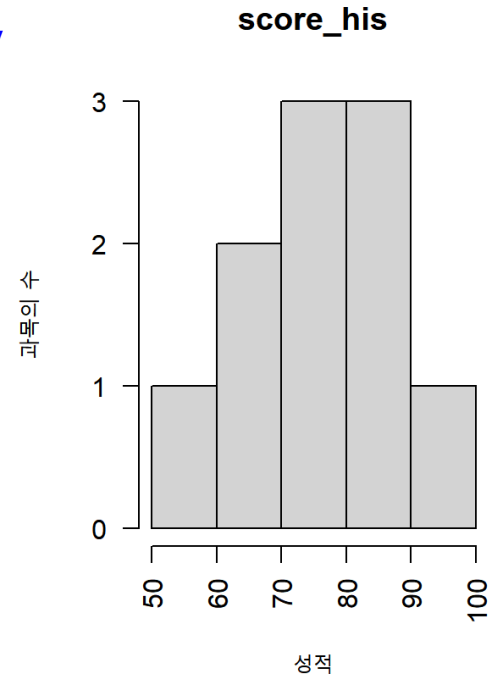
# 예시

- 예시 4 : 과목별 성적 분석

```
1 score ← c(90, 85, 73, 80, 85, 65, 78, 50, 68, 96)
# coding here #
```

- 실행 결과

```
> score ← c(90, 85, 73, 80, 85, 65, 78, 50, 68,
> 
+ 
+ 
+ 
+ 
+ las = 2,
+ 
# 히스토그램 출력
> 
# 상자그림 출력
```



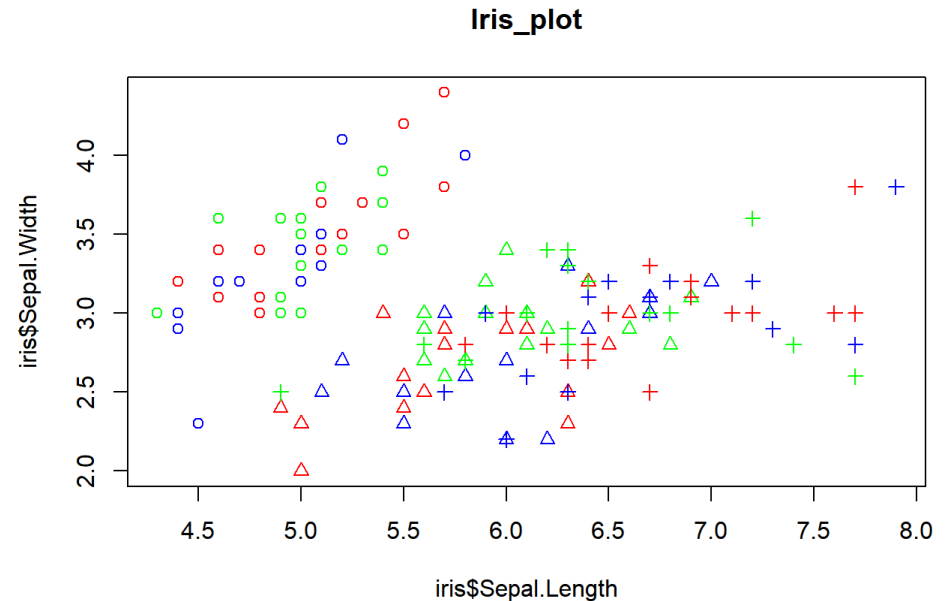
## 예시

- 예시 5 : iris 데이터셋에서 Species 정보에 따른 Sepal.Length와 Sepal.Width의 분포 분석

```
point ← as.numeric(iris$Species)
color ← c('red', 'green', 'blue')
# cording here #
```

- 실행 결과

```
> point ← as.numeric(iris$Species)
> color ← c('red', 'green', 'blue')
> 
+ 
+ 
+ 
+ 
+ 
# 산점도 출력
```



# 예시

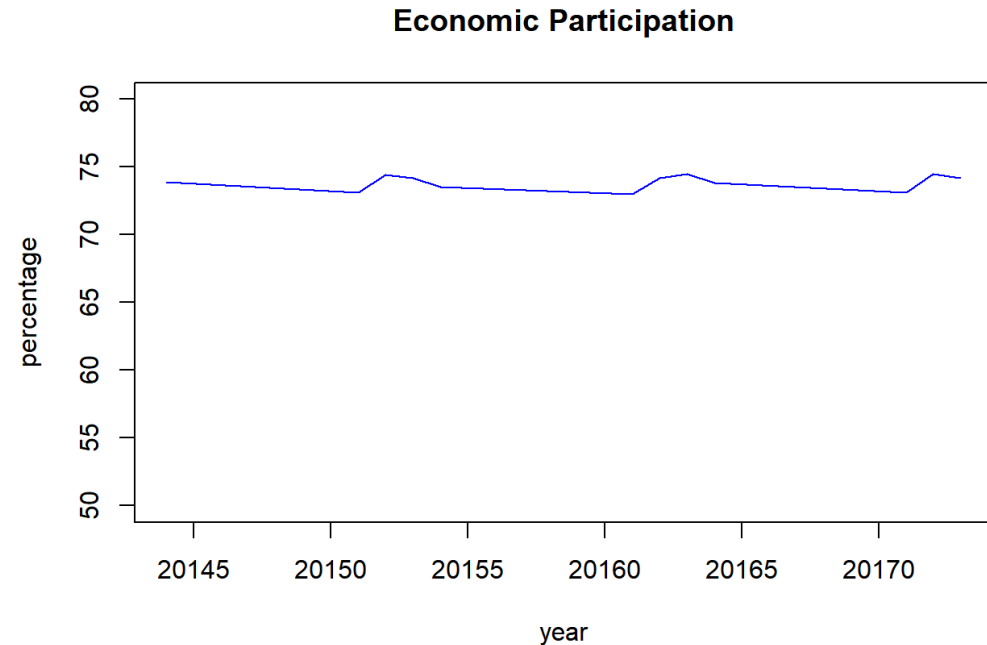
- 예시 6 : 2014년 4분기부터 2017년 3분기까지 남녀 경제활동참가율 분석

```
year ← c(20144, 20151, 20152, 20153, 20154, 20161, 20162, 20163, 20164, 20171, 20172, 20173)
men ← c(73.9, 73.1, 74.4, 74.2, 73.5, 73, 74.2, 74.5, 73.8, 73.1, 74.5, 74.2)
women ← c(51.4, 50.5, 52.4, 52.4, 51.9, 50.9, 52.6, 52.7, 52.2, 51.5, 53.2, 53.1)
# cording here #
```

- 실행 결과

(생략)

```
> 
+   men,
+   
+   type = ' | ',
+   lty = 1,
+   
+   
+   
+   
```



# 예시

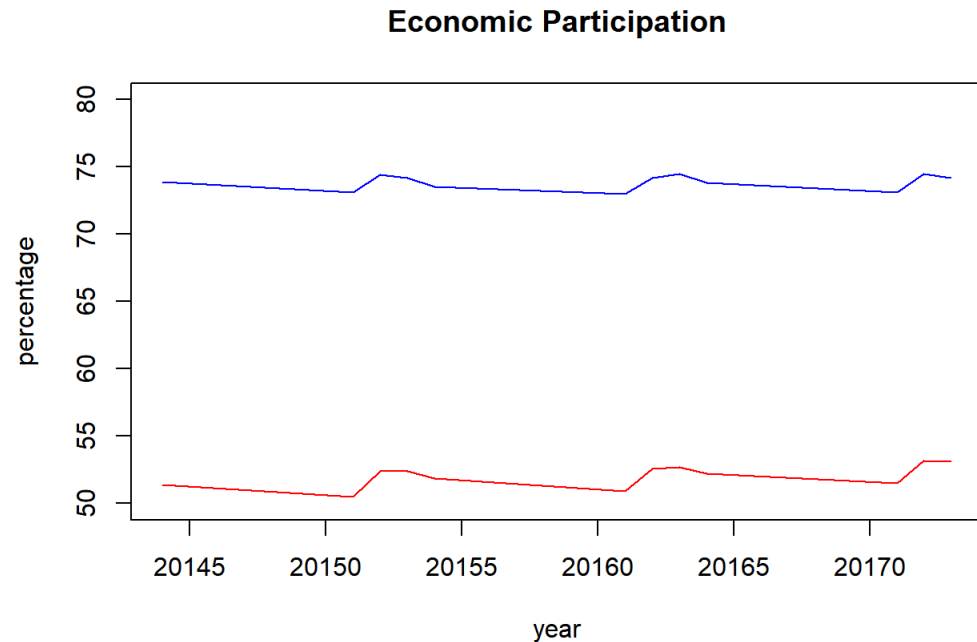
- 예시 7 : 2014년 4분기부터 2017년 3분기까지 남녀 경제활동참가율 분석

```
year ← c(20144, 20151, 20152, 20153, 20154, 20161, 20162, 20163, 20164, 20171, 20172, 20173)
men ← c(73.9, 73.1, 74.4, 74.2, 73.5, 73, 74.2, 74.5, 73.8, 73.1, 74.5, 74.2)
women ← c(51.4, 50.5, 52.4, 52.4, 51.9, 50.9, 52.6, 52.7, 52.2, 51.5, 53.2, 53.1)
# cording here #
```

- 실행 결과

(생략)

```
> 
+   women,
+   type = ' | ',
+   
```



감사합니다

[kimtwan21@dongduk.ac.kr](mailto:kimtwan21@dongduk.ac.kr)

김 태 완