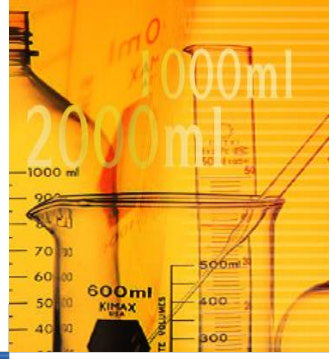
R 프로그래밍



Chapter 4

일변량 자료 탐색

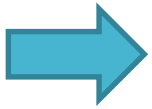
Sejong Oh

Bio Information technology Lab.

개요

- 통계 기법은 자료를 정리하고 분석할 수 있는 강력한 수단
- 데이터 분석 에서도 많은 부분에서 통계적 기법을 필요로 한다

  - 여론조사 결과 분석
  - 제조업 불량율 분석
  - 학습 효과 분석
  - ...



통계학을 알아야 한다.

자료의 종류

- 질적 자료(qualitative data) 또는 범주형 자료(categorical data) : 원칙적으로 숫자로 표시될 수 없는 자료
  - 예) 교육수준 : 초졸, 중졸, 고졸, 대졸 / 성별 : M, F

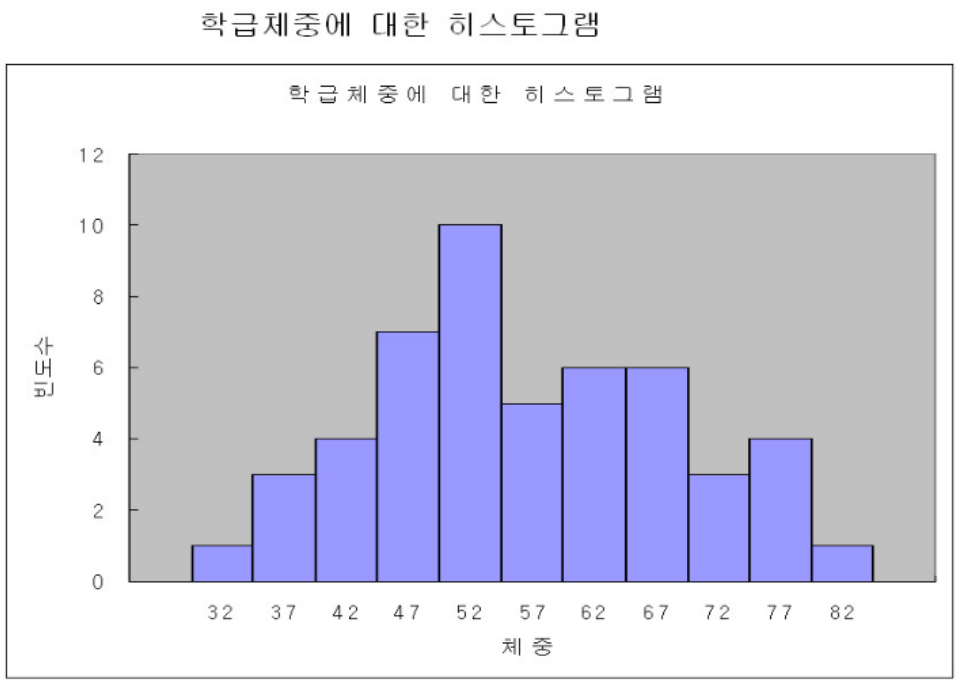
```
> iris$Species
  [1] setosa     setosa     setosa     setosa     setosa     setosa     setosa     setosa     setosa
 [10] setosa     setosa     setosa     setosa     setosa     setosa     setosa     setosa     setosa
 [19] setosa     setosa     setosa     setosa     setosa     setosa     setosa     setosa     setosa
 [28] setosa     setosa     setosa     setosa     setosa     setosa     setosa     setosa     setosa
 [37] setosa     setosa     setosa     setosa     setosa     setosa     setosa     setosa     setosa
 [46] setosa     setosa     setosa     setosa     setosa     versicolor versicolor versicolor versicolor
 [55] versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor
 [64] versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor
 [73] versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor
 [82] versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor
 [91] versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor
[100] versicolor virginica  virginica  virginica  virginica  virginica  virginica  virginica  virginica
[109] virginica  virginica  virginica  virginica  virginica  virginica  virginica  virginica  virginica
[118] virginica  virginica  virginica  virginica  virginica  virginica  virginica  virginica  virginica
[127] virginica  virginica  virginica  virginica  virginica  virginica  virginica  virginica  virginica
[136] virginica  virginica  virginica  virginica  virginica  virginica  virginica  virginica  virginica
[145] virginica  virginica  virginica  virginica  virginica  virginica
```

- 양적 자료(quantitative data) : 자료자체가 숫자로 표현됨.
  - 이산자료(discrete data) : 정수값을 취할 수 있는 자료(각 세대의 자녀수)
  - 연속자료(continuous data) : 실수 값을 취할 수 있는 자료(키, 몸무게, 온도)

```
> 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 5.7 5.1 5.4 5.1 4.6 5.1 4.8 5.0
 [27] 5.0 5.2 5.2 4.7 4.8 5.4 5.2 5.5 4.9 5.0 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
 [53] 6.9 5.5 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 6.3 6.1 6.4 6.6 6.8 6.7
 [79] 6.0 5.7 5.5 5.5 5.8 6.0 5.4 6.0 6.7 6.3 5.6 5.5 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
[105] 6.5 7.6 4.9 7.3 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 6.2 6.1 6.4 7.2
[131] 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 6.7 6.7 6.3 6.5 6.2 5.9
```

자료의 종류

- 질적 자료의 분석방법
  - 도수분포표(frequency distribution table)
  - 상대적 도수분포표(relative frequency distribution table) : 관찰수가 다른 집단들을 비교하는데 더욱 유용
  - 막대그래프(bar graph)
  - 원형그래프(pie graph)



자료의 종류

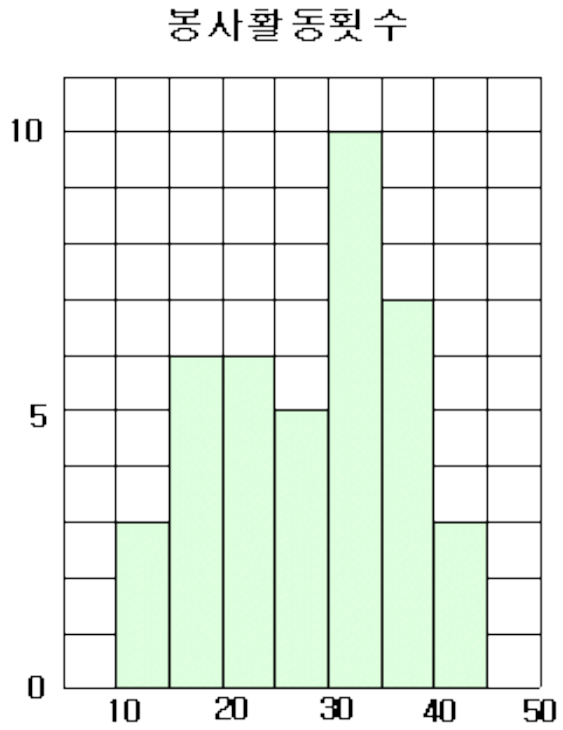
- 질적자료의 분석방법
  - 줄기.잎 그림(stem and leaf display)
  - 도수분포표
    - 이산자료 : 누적빈도와 상대적 누적 빈도도 계산
    - 연속자료 : ..
  - 히스토그램(histogram)
  - 도수다각형(frequency polygons) : 히스토그램에서 각 구간의 중점을 이용하여 작성
  - 상대도수다각형(relative frequency polygons) : 두 개 이상의 자료집단 분포를 같은 그림 위에 놓고 비교할 수 있어서 편리
  - 누적 상대도수(cumulative relative frequency)

자료의 종류

봉사활동횟수

봉사활동횟수(번)	학생 수(명)
10이상 ~15미만	3
15~20	6
20~25	6
25~30	5
30~35	10
35~40	7
45~50	3
합계	40

도수분포표

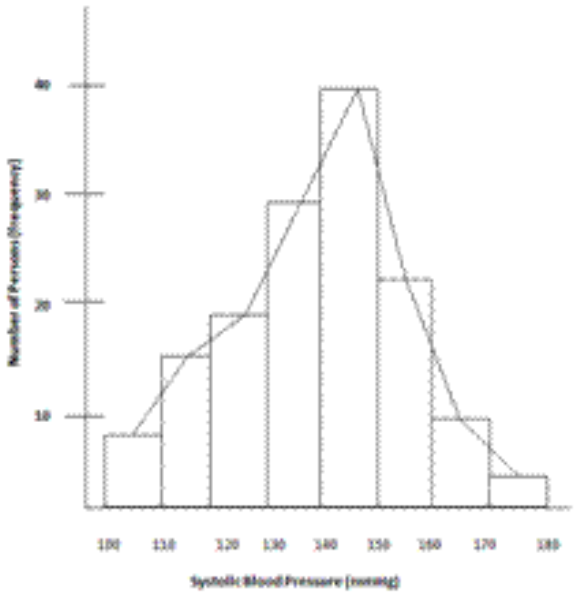


히스토그램

자료의 종류

```
4|899
4|6
4|4455
4|333
4|01
3|99
3|677777
3|55
3|223
3|111
2|88999
2|666667
2|444455
2|22333
2|000000
1|88888888888899999999999
1|666666777777
1|444444444444555555555555
1|22222222222222222222333333333
1|00000000000000011111111111111111111111111111
```

줄기잎그림



도수 다각형

자료의 종류

- 일변량 자료(univariate data)
  - 분석대상이 되는 변수의 개수가 1개
  - ex) 단국대 학생들의 몸무게 분포를 분석해보자

- 다변량 자료 (multivariate data)
  - 분석대상이 되는 변수의 개수가 2개 이상인 경우
  - 변수가 2개인 경우를 특별히 이변량 자료(bivariate data) 라고함
  - ex) 출생 지역과 몸무게가 상관관계가 있는지 분석해보자

모집단과 표본

- 모집단 (population)
  - 관심을 가지는 조사대상 전체

- 표본(sample)
  - 모집단에서 실제 조사가 이루어지는 집단, 표본은 모집단의 부분집합

> 단국대 학생 중 100명을 선별하여 외국어 실력을 조사해보자

  - 모집단 : 단국대 학생 전체
  - 표본 : 선발된 100명

- 모수(parameter)
  - 모집단의 특성을 나타내는 척도로 보통 평균과 표준편차 등이 많이 사용됨

자료요약 : 평균

모든 국민들의 소득자료를 가지고 있다. 이 자료를 요약해서 설명할수 있는 값은 무엇이 있을까?

- 평균(mean)
  - 균형점, 무게중심

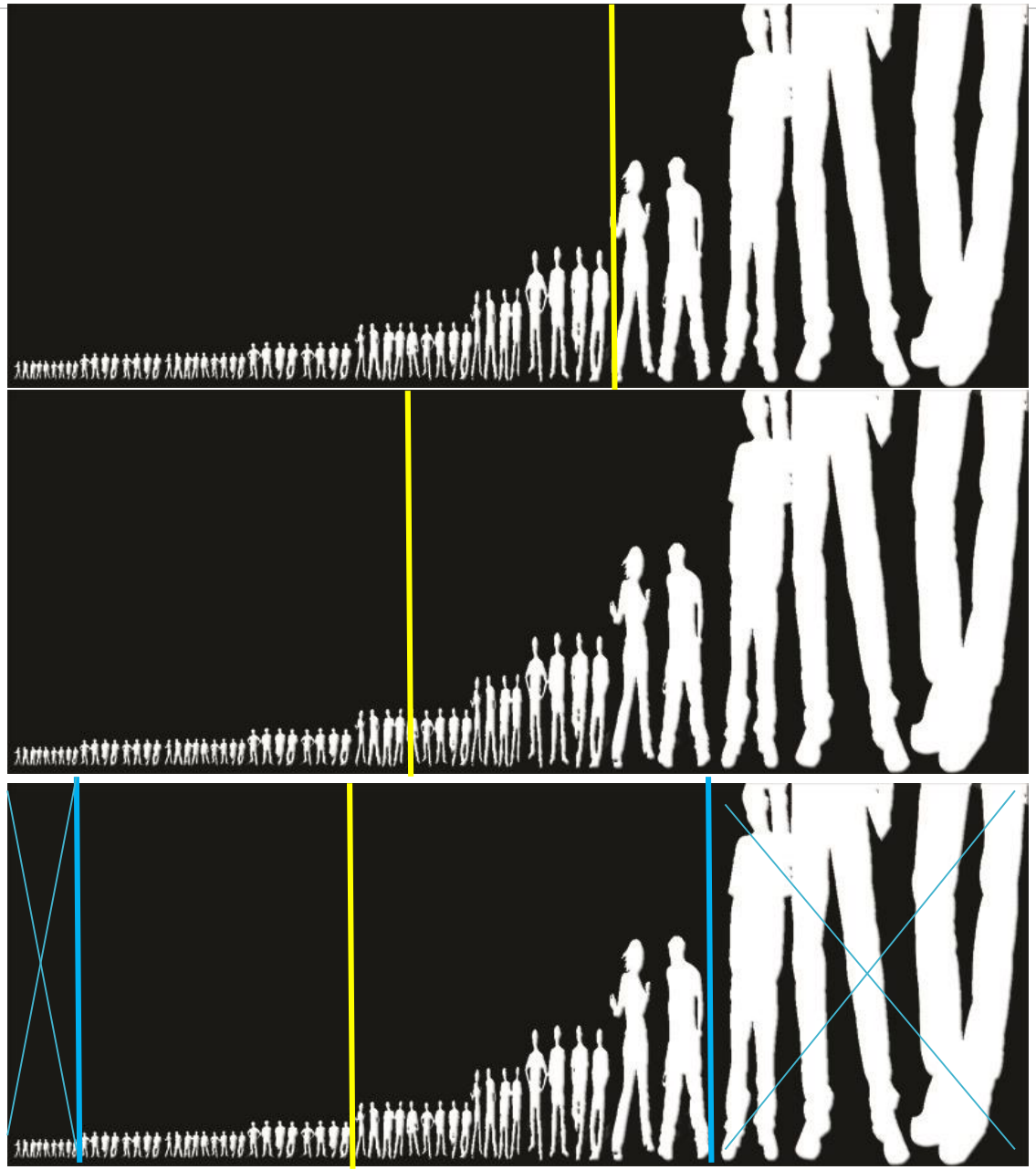
$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

- 중앙값(median)
  - 어떤 주어진 값들을 정렬했을 때 가장 중앙에 위치하는 값을 의미

- 절사평균(trimmed mean)
  - 표본중에서 작은값 n% 와 큰값 n%를 제외하고 나머지 (100-2n)% 의 자료만 사용하여 구한 평균

EBS – 평균의 함정(https://youtu.be/Pp_Pd6GZLOE)

자료요약 : 평균



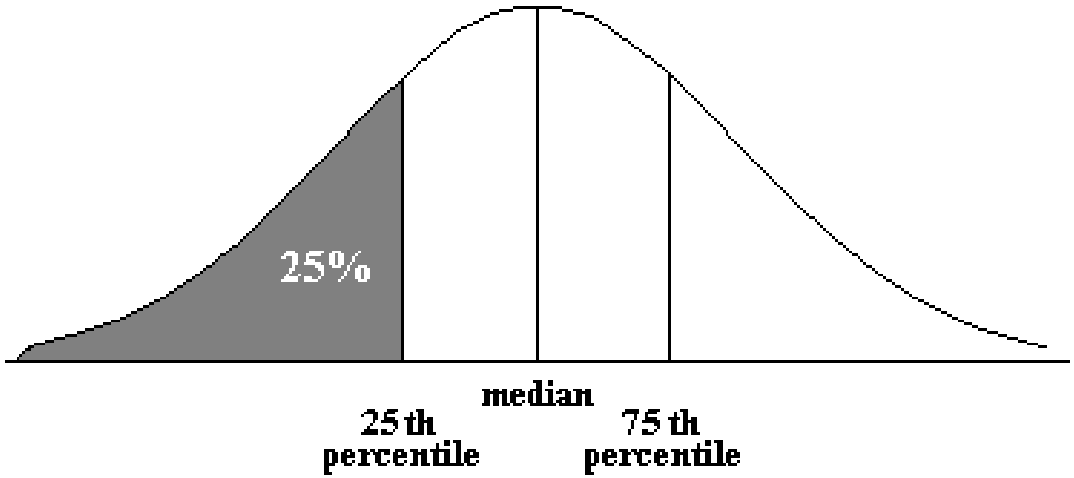
평균

중앙값

절사평균

자료요약 : 4 분위 수

- 4분위수 (quartile)
  - **측정값을 4등분하는 백분위수**
  - 제 1 사분위수 (Q1) : 제 25 백분위수
  - 제 2 사분위수 (Q2) : 제 50 백분위수, 중앙값
  - 제 3 사분위수 (Q3) : 제 75 백분위수



R 실습 : mean(), median(), quantile(), summary()

```
mydata = c(50,60,100,75,200)
mydata.big = c(mydata, 50000)
mean(mydata)                    # 평균
mean(mydata.big)
median(mydata)                  # 중앙값
median(mydata.big)
mean(mydata, trim=0.2)          # 절사평균
mean(mydata.big, trim=0.2)
quantile(mydata)                # 사분위수
quantile(mydata, (0:10)/10)
summary(mydata)
fivenum(mydata)                 # quantile()과 비슷
```

R 실습 : mean(), median(), quantile(), summary()

- quantile()

```
> quantile(mydata)
  0%  25%  50%  75% 100%
  50   60   75  100  200
```

최소 중앙 최대

```
> quantile(mydata, (0:10)/10)
  0%  10%  20%  30%  40%  50%  60%  70%  80%  90% 100%
  50   54   58   63   69   75   85   95  120  160  200
```

- summary()

```
> summary(mydata)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
     50      60      75      97     100     200
```

R 실습 : table()

- 범주형 자료(categorical data)의 분석

```
ans=c("Y","Y","N","Y","Y")
table(ans)                          # 도수분포표 출력
```

```
> ans=c("Y","Y","N","Y","Y")
> table(ans)
ans
N Y
1 4
```

```
> table(iris$Species)

    setosa versicolor  virginica
        50         50         50
```

R 실습 : 줄기-잎 그림

- stem()

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

```
> stem(iris$Sepal.Length)

  The decimal point is 1 digit(s) to the left of the |

  42 | 0
  44 | 0000
  46 | 000000
  48 | 00000000000
  50 | 00000000000000000000
  52 | 00000
  54 | 0000000000000
  56 | 000000000000000
  58 | 0000000000
  60 | 000000000000
  62 | 0000000000000
  64 | 000000000000
  66 | 0000000000
  68 | 0000000
  70 | 00
  72 | 0000
  74 | 0
  76 | 00000
  78 | 0

> 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 5.7 5.1 5.4 5.1 4.6
 [24] 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 5.5 4.9 4.4 5.1 5.0 4.5 4.4 5.0 5.1 4.8
 [47] 5.1 4.6 5.3 5.0 7.0 6.4 6.9 5.5 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
 [70] 5.6 5.9 6.1 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 5.5 6.1
 [93] 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 6.7 7.2 6.5 6.4 6.8 5.7 5.8
[116] 6.4 6.5 7.7 7.7 6.0 6.9 5.6 7.7 6.3 6.7 7.2 6.2 6.1 6.4 7.2 7.4 7.9 6.4 6.3 6.1 7.7 6.3 6.4
[139] 6.0 6.9 6.7 6.9 5.8 6.8 6.7 6.7 6.3 6.5 6.2 5.9
```

자료요약 : 산포(distribution)

- 산포
  - 데이터가 퍼져 있는 정도, 흩어져 있는 정도
  - 분산과 표준편차를 가지고 표현

- 분산 (variance)

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

- 표준편차(standard deviation)

$$S = \sqrt{(\text{분산})}$$

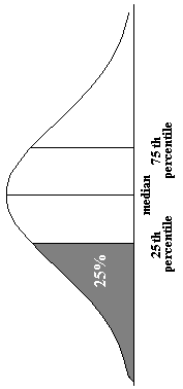
R 실습 : diff(), var(), sd()

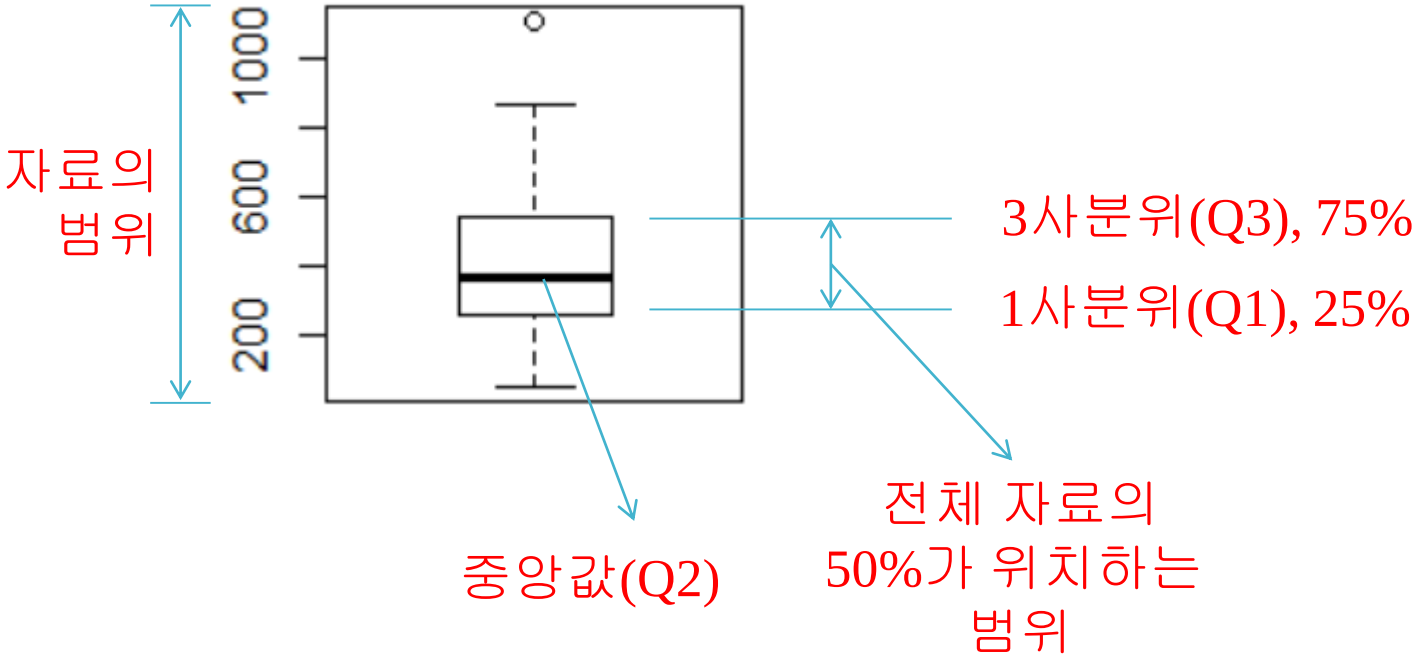
```
diff(range(mydata))        # 최대값-최소값
var(mydata)                # 분산
sd(mydata)                 # 표준편차
```

```
> diff(range(mydata))
[1] 150
> var(mydata)
[1] 3670
> sd(mydata)
[1] 60.58052
```

자료요약 : box plot

- Box plot 또는 Box whisker plot



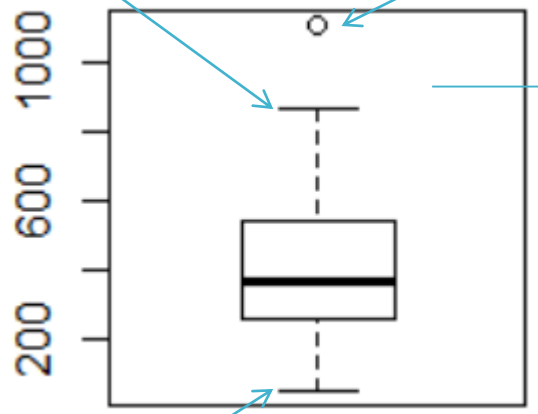


☞ box 의 넓이는 아무 의미가 없음

자료요약 : box plot

이상치를 제외한 값중
최대값

정상범위 밖에 존재하는
데이터 표시. 이상치.특이값(outlier)



Q3 + 1.5 * IQR

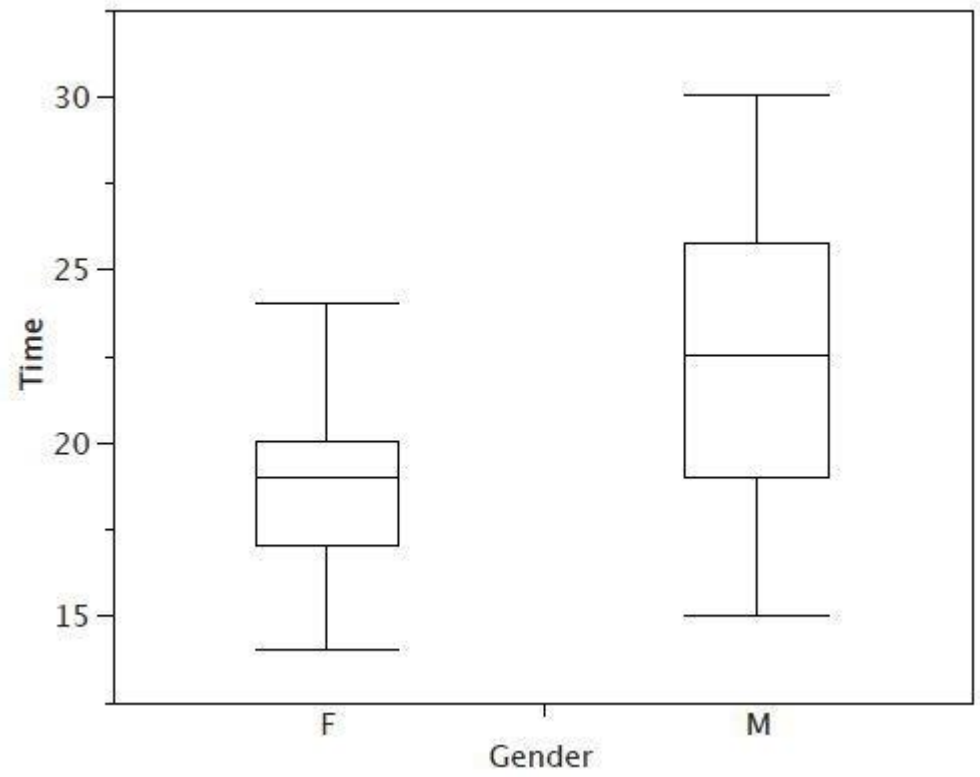
정상적인 데이터가
분포할 것으로 기대되는 범위
(표시되지 않음)

Q1 - 1.5 * IQR

이상치를 제외한 값중
최소값

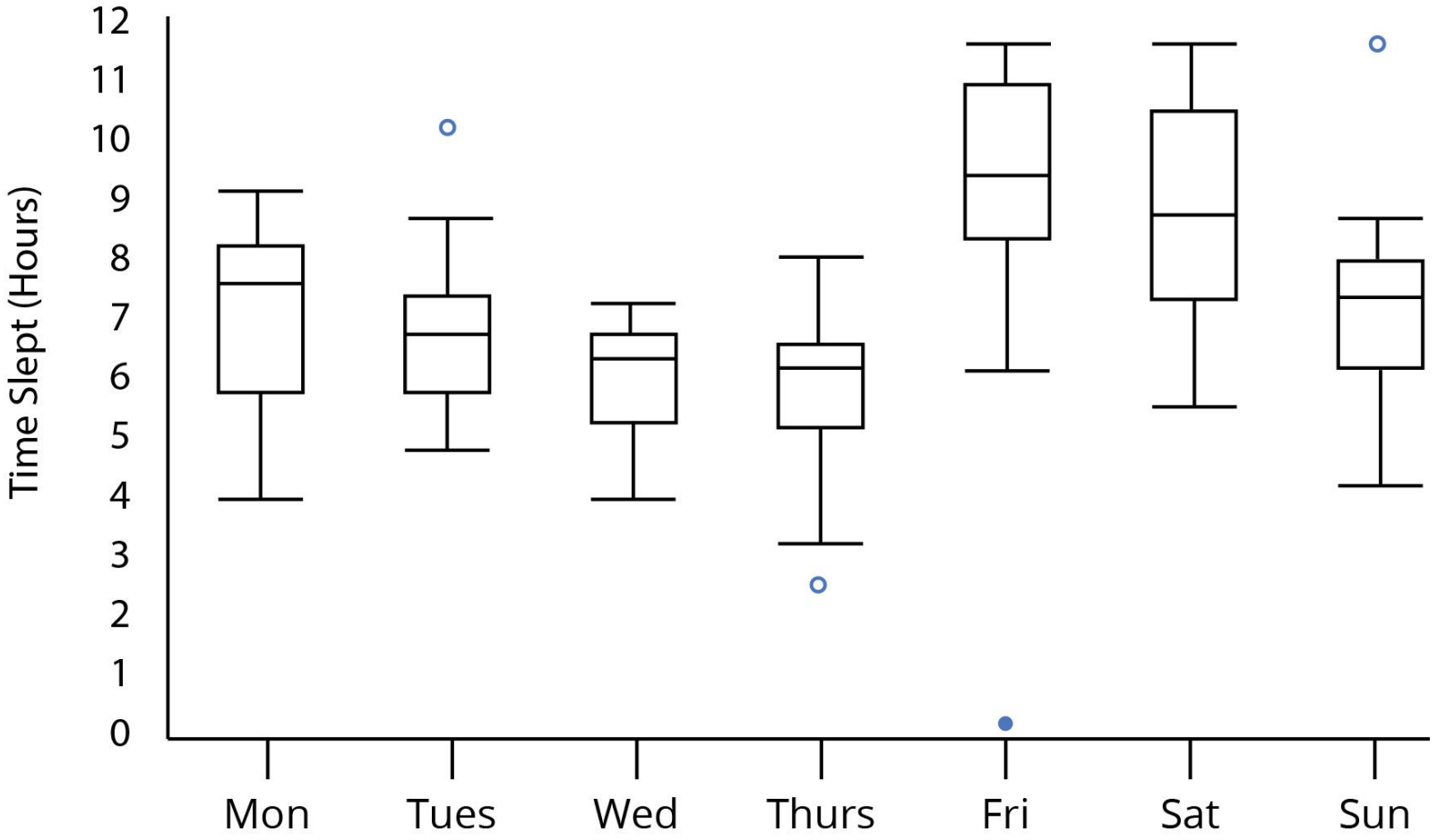
(IQR = Q3-Q1)

자료요약 : box plot



시간제 일자리 근무시간(남녀)

자료요약 : box plot



20명의 학생에 대한 수면시간

R 실습 : barplot()

- 막대그래프 그리기

```
head(mtcars)                          # 자동차 모델별 제원
carb <- mtcars[,"carb"]               # 기화기 수
table(carb)                           # 도수분포표
barplot(table(carb),
        main="Barplot of Carburetors",
        xlab="#of carburetors",
        ylab="frequency")
```

- table() 함수 : 주어진 자료로 부터 도수 분포표를 그려준다.

```
> table(mtcars$carb)

 1  2  3  4  6  8
 7 10  3 10  1  1
```

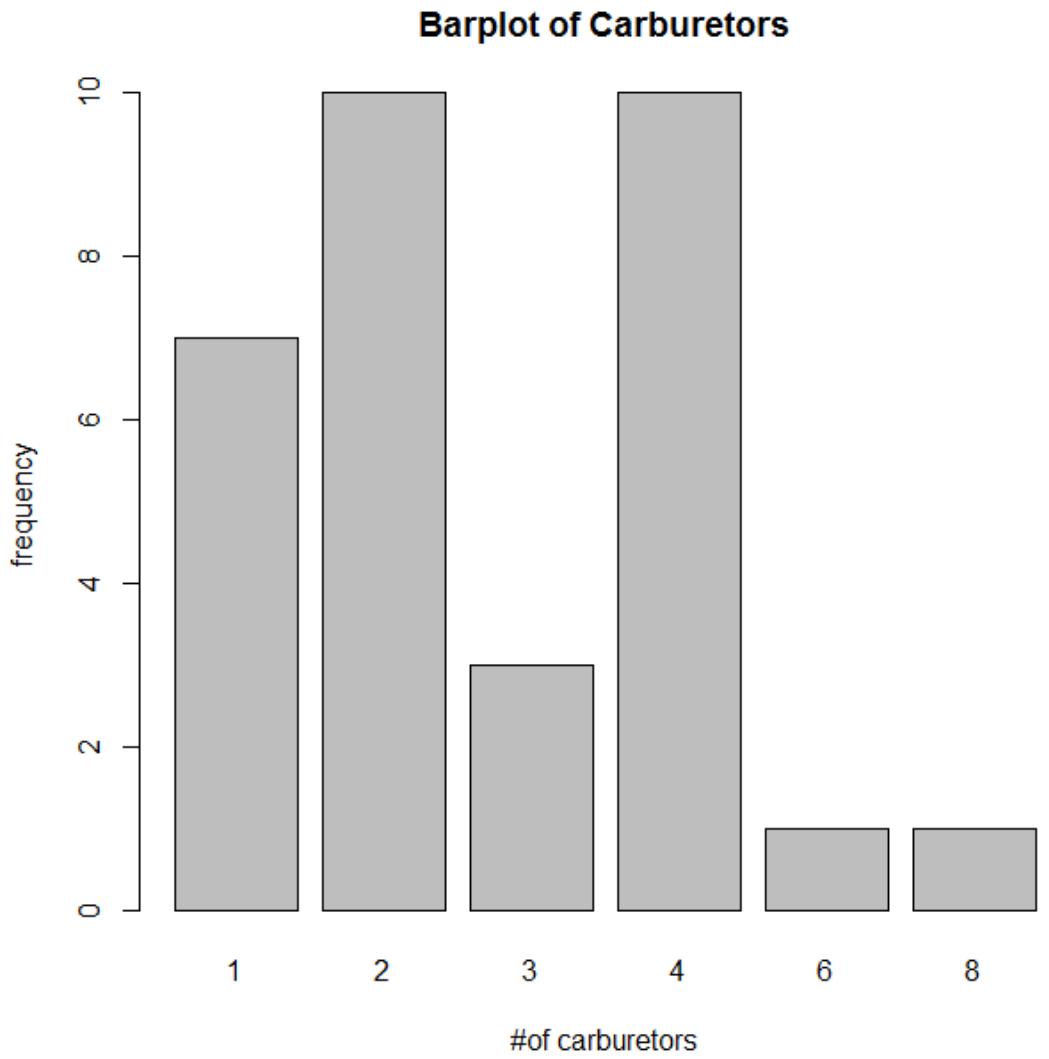
R 실습 : barplot()

- 옵션↓

인수	설명
angle, density, col	막대를 칠하는 선분의 각도, 선분의 수, 선분의 색 지정
legend	오른쪽 상단에 범례추가
names	각 막대의 라벨을 정하는 문자열 벡터를 지정
width	각 막대의 상대적인 폭을 벡터로 지정
space	각 막대 사이의 간격을 지정
beside	TRUE를 지정하면 각각의 값마다 막대를 그림
horiz	TRUE를 지정하면 막대를 옆으로 눕혀서 그림

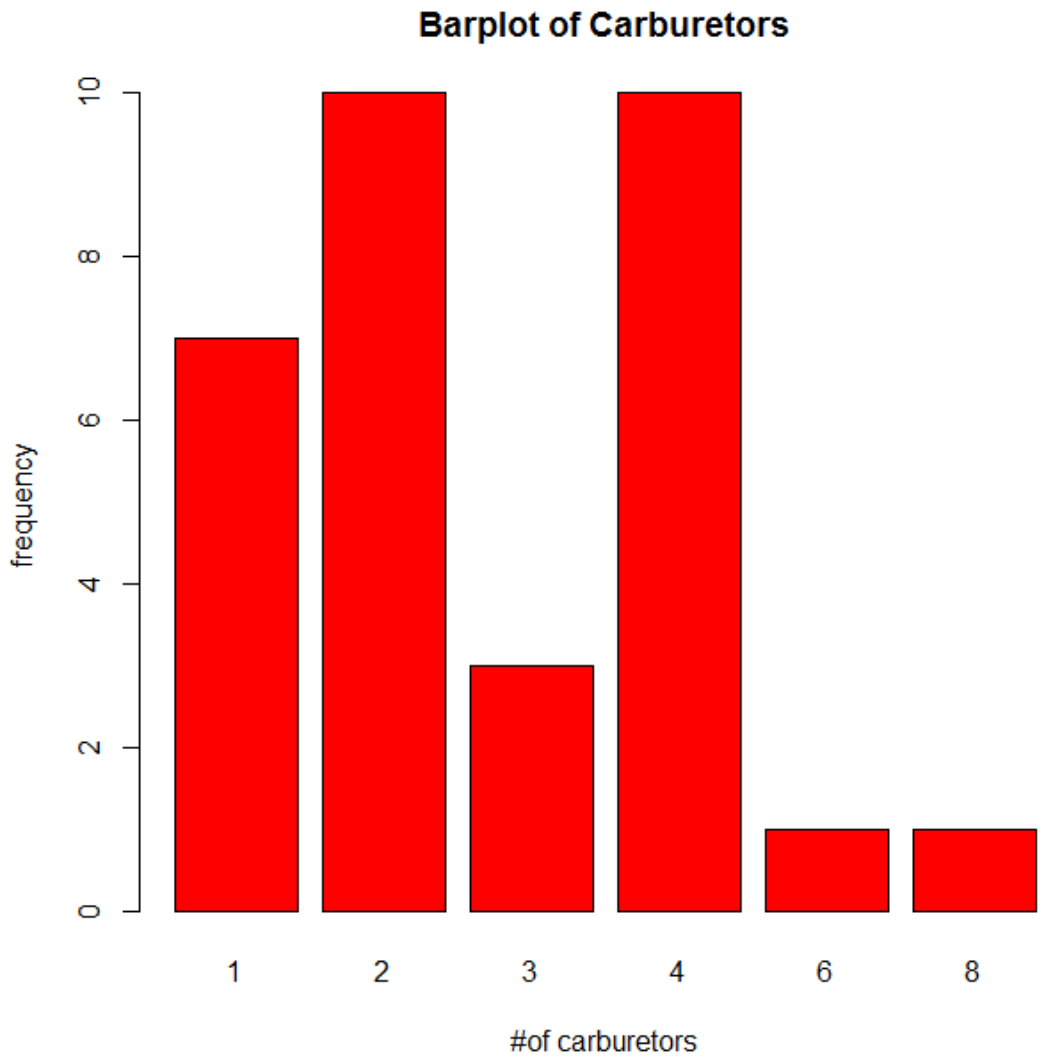
R 실습 : barplot()

```
> barplot(table(carb),
+         main="Barplot of Carburetors",
+         xlab="#of carburetors",
+         ylab="frequency")
```



R 실습 : barplot()

```
> barplot(table(carb),
+         main="Barplot of Carburetors",
+         xlab="#of carburetors",
+         ylab="frequency",
+         col="red")
```

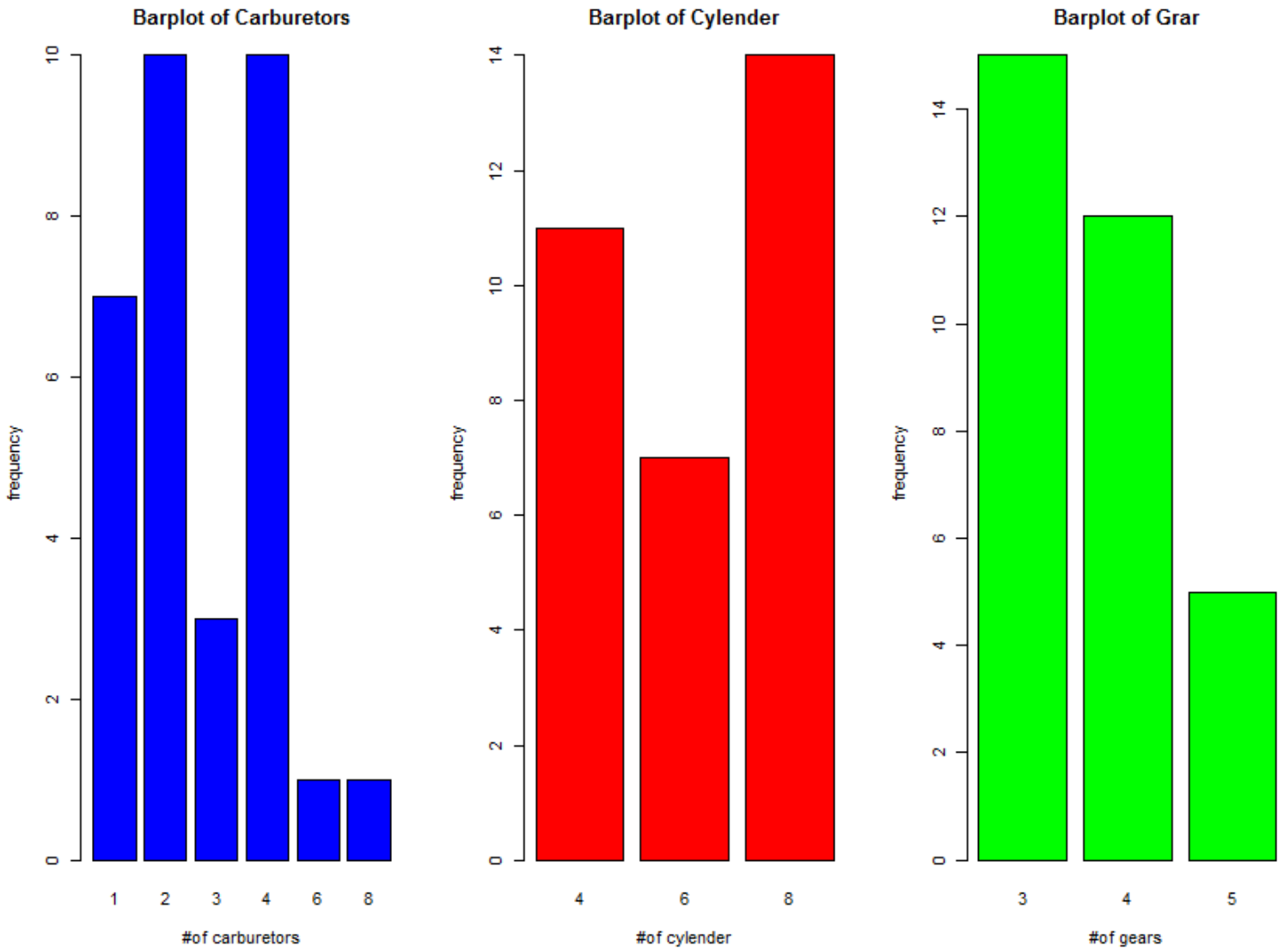


R 실습 : barplot()

- 한 화면에 그래프 여러 개 그리기

```
par(mfrow=c(1,3))              # 1x3 윈도우 생성
barplot(table(mtcars$carb),
        main="Barplot of Carburetors",
        xlab="#of carburetors",
        ylab="frequency",
        col="blue")
barplot(table(mtcars$cyl),
        main="Barplot of Cylender",
        xlab="#of cylender",
        ylab="frequency",
        col="red")
barplot(table(mtcars$gear),
        main="Barplot of Grar",
        xlab="#of gears",
        ylab="frequency",
        col="green")
```

R 실습 : barplot()


Barplot of Carburetors
frequency
0
2
4
6
8
10
1
2
3
4
6
8
#of carburetors
Barplot of Cylender
frequency
0
2
4
6
8
10
12
14
4
6
8
#of cylender
Barplot of Grar
frequency
0
2
4
6
8
10
12
14
3
4
5
#of gears

R 실습 : barplot()

- Barplot에 대한 보다 상세한 옵션을 보려면

```
? barplot
```

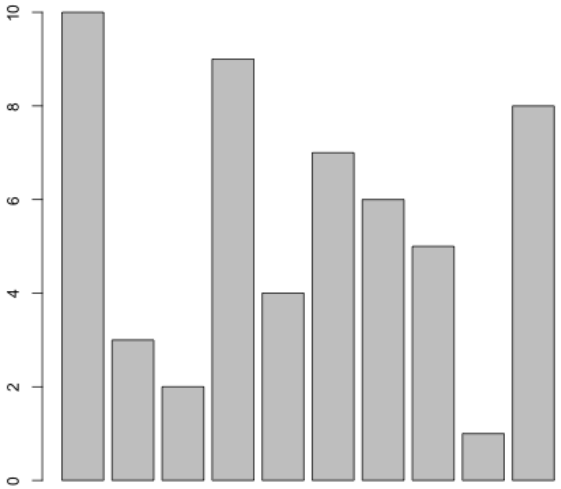
또는 Rstudio 의 help 탭에서 barplot 검색

- 다양한 막대 그래프 예제
  http://www.theanalysisfactor.com/r-11-bar-charts/

- R 에서 지원하는 color 이름
  http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf

R 실습 : hist()

- 막대 그래프는 도수 분포표를 만들 수 있는 정수형, 문자형 자료의 경우에 사용하고, 실수형 자료에 대해서는 히스토그램을 사용한다



barplot

histogram

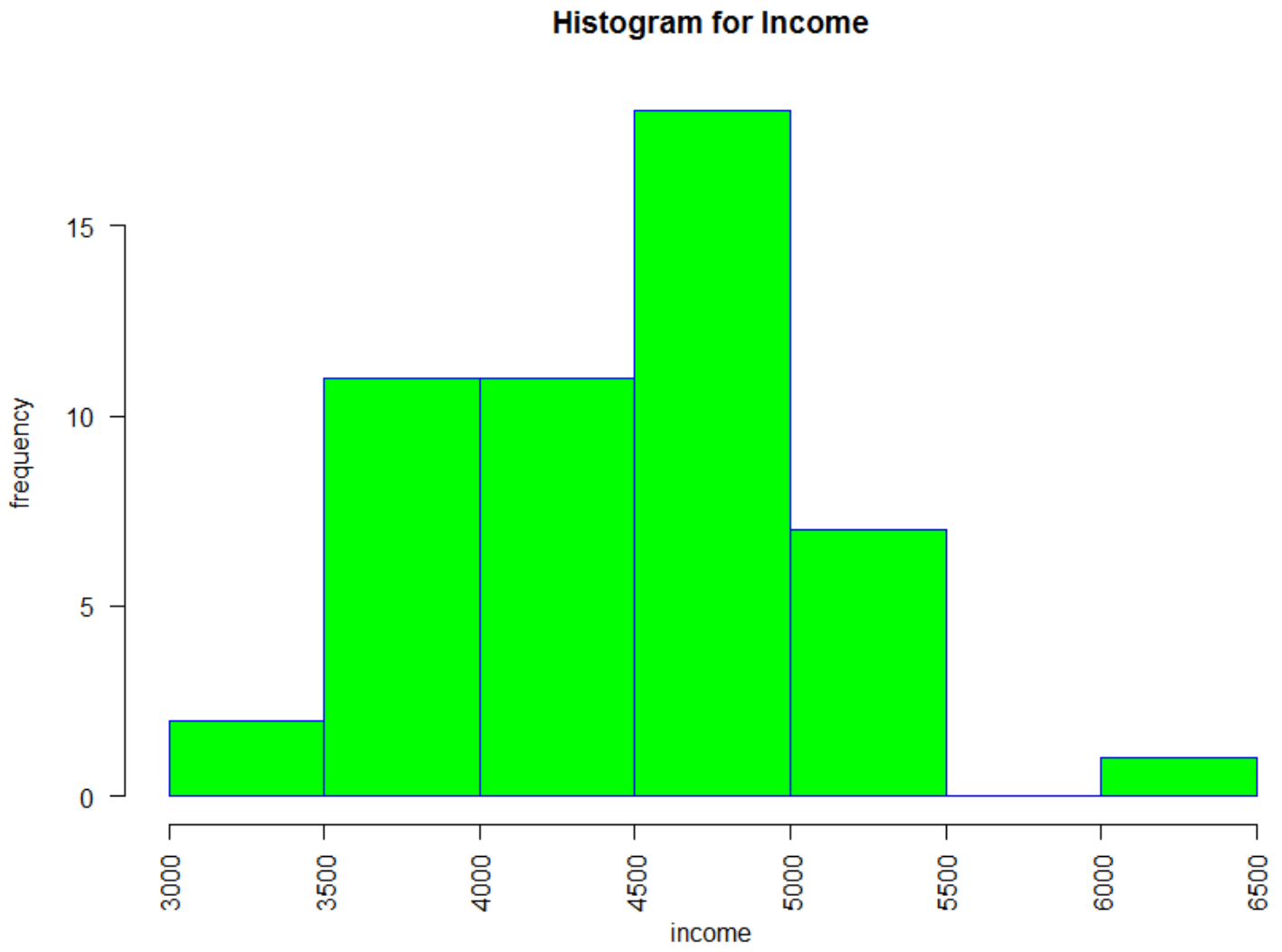
막대그래프는 막대간 간격이 있고 막대의 면적은 의미가 없다
히스토그램은 막대가 붙어 있고 막대의 면적이 의미가 있다

R 실습 : hist()

- 히스토그램 그리기

```
st.income <- state.x77[,"Income"]
hist(st.income,                                  # data
     main="Histogram for Income",                # 제목
     xlab ="income",          # x축 레이블
     ylab="frequency",        # y축 레이블
     border="blue",           # 막대 테두리색
     col="green",             # 막대 색
     las=2,                   # x축 글씨방향(0~3)
     breaks=5)                # x축 막대 개수 조절
```

breaks=n 일때 막대 개수는 $\log_2(n)+1$ 로 계산
N이 커질수록 막대의 개수가 늘어난다

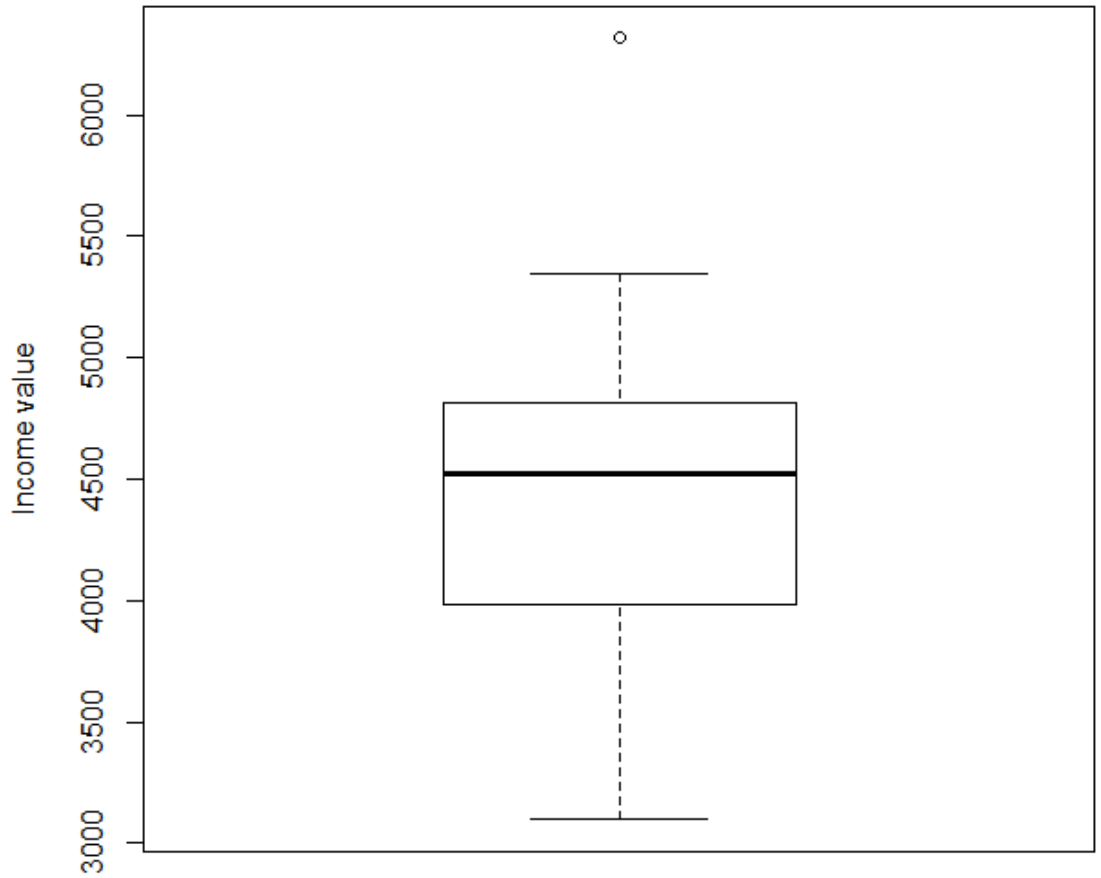

Histogram for Income
frequency
0
5
10
15
3000
3500
4000
4500
5000
5500
6000
6500
income

R 실습 : boxplot()

- state.x77 데이터셋에서 주별 소득에 대해 boxplot 을 그려보자

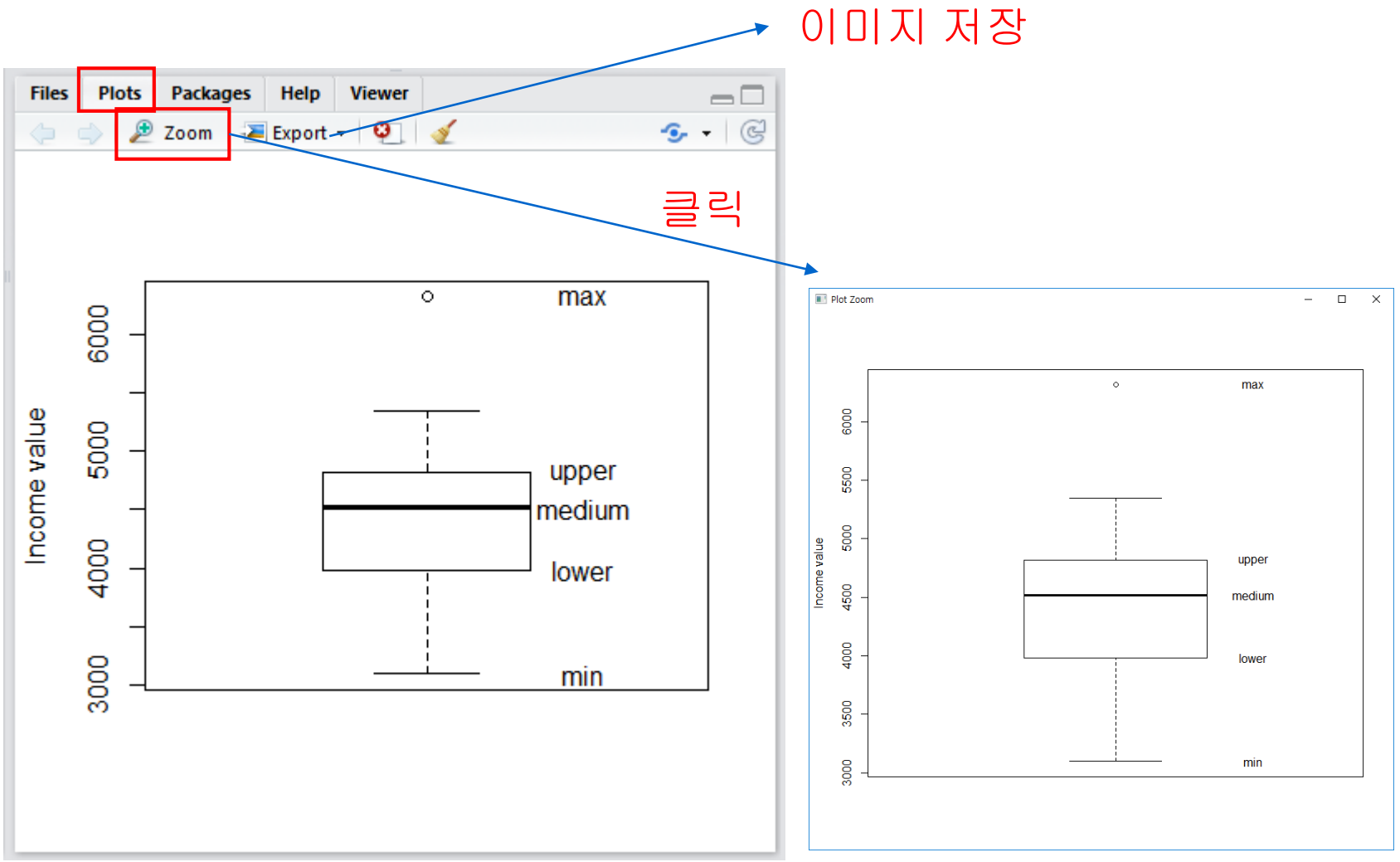
```
head(state.x77)
st.income <- state.x77[,"Income"]
boxplot(st.income, ylab="Income value")
```

R 실습 : boxplot()



R 실습 : boxplot()

- Rstudio 에서 그래프 보기



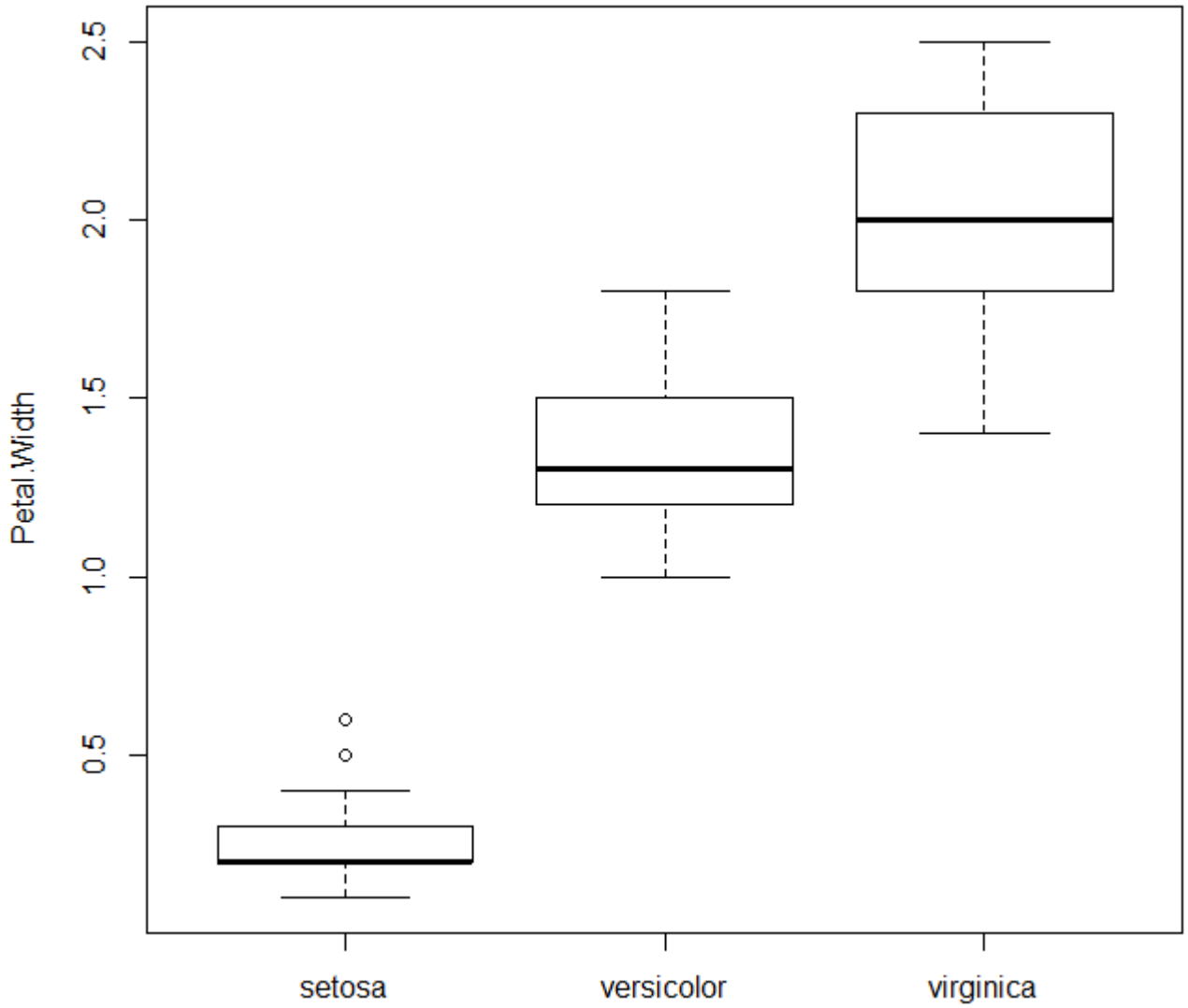
R 실습 : boxplot()

- iris dataset에서 품종(Species)에 따른 Petal.Width 자료에 대한 boxplot 을 그려 보시오

```
boxplot(Petal.Width~Species,data=iris,
          ylab="Petal.Width")
```

```
> 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
```

R 실습 : boxplot()



[연습1]

- 홍길동군의 과목별 성적은 다음과 같다.

KOR	ENG	MATH	HIST	SOC	MUSIC	BIO	EARTH	PHY	ART
90	85	73	80	85	65	78	50	68	96

1. 이 데이터를 score 벡터에 저장하시오. (과목명은 데이터 이름으로 저장하시오)
2. score 벡터의 내용을 보이시오
3. 전체 성적의 평균은 얼마인가
4. 전체 성적의 중앙값은 얼마인가
5. 전체 성적의 표준편차를 보이시오
6. 가장 성적이 높은 과목의 이름을 보이시오
7. 성적에 대한 boxplot 을 그리시오. 이상치에 해당하는 과목이 있으면 제시하시오
8. 성적에 대한 histogram 을 그리되 다음조건을 만족하도록 하시오
   (그래프 title : Hong's score, 막대색: 보라색)

[연습2]

- mtcars 데이터셋을 이용하여 다음 문제를 해결하시오

1. 중량(wt) 의 평균값, 중앙값, 절사평균값(절사범위:15%), 표준편차 를 구하시오
2. 중량(wt)에 대해 summary() 함수의 적용 결과를 보이시오
3. 실린더수(cyl) 에 대해 도수분포표를 구하시오
4. 앞에서 구한 도수분포표를 막대그래프로 그려 보시오
5. 중량(wt)의 히스토그램, 실린더(cyl), 기어(gear) 에 대한 막대 그래프를 한 화면에 보이게 작성하시오
6. 중량(wt)에 대해 boxplot 을 그려 보시오. Boxplot 으로 부터 관찰할 수 있는 정보를 적으시오
7. 배기량(disp)에 대해 boxplot 을 그려 보시오. Boxplot 으로 부터 관찰할 수 있는 정보를 적으시오
8. 인터넷에서 수업시간에 배우지 않은 막대 그래프의 예를 찾아보고 mtcars 에 적용하여 작성하여 보시오

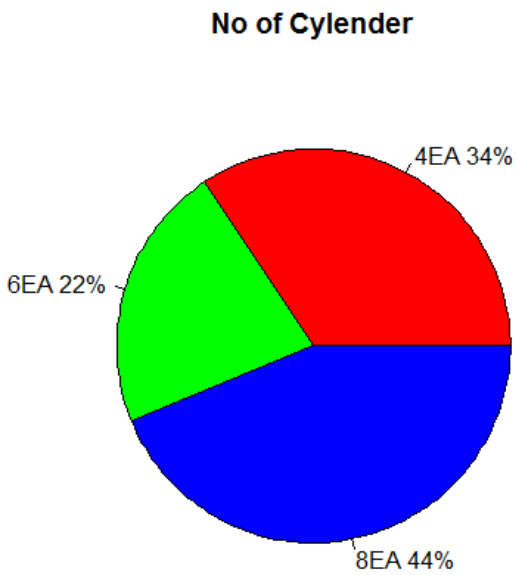
[연습3]

- mtcars 데이터셋을 이용하여 다음 문제를 해결하시오

1. 실린더수(cyl) 에 대해 다음과 같이 빈도 테이블 구하여 freq 에 저장하시오

```
 4  6  8
11  7 14
```

2. http://www.statmethods.net/graphs/pie.html 의 파이그래프 예제를 참고하여 freq 에 대해 다음과 같이 파이 그래프를 그리시오



[tip]

- paste() 함수 : 여러 문자열을 연결하여 하나로 만들 때 사용

```
paste("Good", "Morning", "Tom", sep=" ")
paste("Good", "Morning", "Tom", sep="/")
paste(1:10, "is good", sep=" ")
```

sep : 연결하는 단어 사이사이에 넣을 값을 지정

```
> paste("Good", "Morning", "Tom", sep=" ")
[1] "Good Morning Tom"
> paste("Good", "Morning", "Tom", sep="/")
[1] "Good/Morning/Tom"
> paste(1:10, "is good", sep=" ")
 [1] "1 is good"  "2 is good"  "3 is good"  "4 is good"  "5 is good"
 [6] "6 is good"  "7 is good"  "8 is good"  "9 is good"  "10 is good"
```

시계열(time series) 데이터 시각화

- 시계열 데이터란
  - 일정한 간격 또는 불규칙한 시간 간격으로 측정 또는 생성된 한 변수의 값들.
  - 예를 들면 자기기압계로 측정한 기압 값은 연속 시계열이고 반면에 연도별 인구수는 이산 시계열이다.
  - 시계열자료를 이용하여 그래프 작성시 가로축은 시간 그리고 세로축은 변수값을 표시한다.

  - R에서 시계열 데이터는 matrix 나 data frame 처럼 보이지만 time series 라는 특별한 객체에 저장해야 분석이 가능하다

시계열(time series) 데이터 시각화

- 시계열 자료 객체의 작성

년도/분기별 GNP 데이터

년＼분기	1	2	3	4
1970	13,217.2	16,099.8	16,182.4	23,546.7
1971	14,653.9	17,536.2	17,885.1	24,662.2
1972	15,459.3	17,967.4	18,545.5	26,104.5
1973	17,147.2	20,357.7	21,360.1	28,607.7
1974	19,201.2	22,177.9	22,331.1	30,044.9

데이터 저장시
년도, 분기값은 제외한다

실제 저장된 데이터 (gdp.csv)

	A	B	C	D
1	13217.2	16099.8	16182.4	23546.7
2	14653.9	17536.2	17885.1	24662.2
3	15459.3	17967.4	18545.5	26104.5
4	17147.2	20357.7	21360.1	28607.7
5	19201.2	22177.9	22331.1	30044.9
6	19891.7	23351.8	23494.1	32593.7
7	21750.9	26097.2	26072.3	35912.6
8	23089.1	27839.3	29384	40498
9	26320.4	31899.4	31897.8	41922.4
10	29683.1	34249.3	33641.6	43422.3
11	29733.3	33820.1	34438.5	40906
12	30770.4	34976.1	36385.3	45326.4
13	32878.3	37443.7	39041.7	48895.9
14	36362.7	41767.9	43968.7	53212.6
15	40492.9	45610.9	47273.7	56138.7
16	43096.4	48662	50393.1	60256.6
17	47079.2	53761.4	56626	66434.9
18	53267.4	60487.5	62676.9	72332
19	60889.8	65847.8	68620	79877.7
20	64435.4	70639.5	73072.4	85651.2

데이터 출처
http://blog.naver.com/PostView.nhn?blogId=dev000&logNo=11004603

시계열(time series) 데이터 시각화

- 시계열 자료 객체의 작성

```
ds <- read.csv("gdp.csv", header=F)
m.gdp <- as.matrix(ds)
v.gdp <- as.vector(t(m.gdp))
GDP <- ts(data=v.gdp,
          start=c(1970,1), # 첫 데이터는 1970년 1분기
          frequency=4)     # 1년당 4개의 데이터 존재
GDP
```

t() : 배열의 행과 열을 바꾸는 함수
ts() : matrix로 부터 시계열 데이터를 생성하는 함수

시계열(time series) 데이터 시각화

- 완성된 GDP 시계열 자료

```
> GDP
        Qtr1     Qtr2     Qtr3     Qtr4
1970 13217.2  16099.8  16182.4  23546.7
1971 14653.9  17536.2  17885.1  24662.2
1972 15459.3  17967.4  18545.5  26104.5
1973 17147.2  20357.7  21360.1  28607.7
1974 19201.2  22177.9  22331.1  30044.9
1975 19891.7  23351.8  23494.1  32593.7
1976 21750.9  26097.2  26072.3  35912.6
1977 23089.1  27839.3  29384.0  40498.0
1978 26320.4  31899.4  31897.8  41922.4
1979 29683.1  34249.3  33641.6  43422.3
1980 29733.3  33820.1  34438.5  40906.0
1981 30770.4  34976.1  36385.3  45326.4
1982 32878.3  37443.7  39041.7  48895.9
1983 36362.7  41767.9  43968.7  53212.6
1984 40492.9  45610.9  47273.7  56138.7
1985 43096.4  48662.0  50393.1  60256.6
1986 47079.2  53761.4  56626.0  66434.9
1987 53267.4  60487.5  62676.9  72332.0
1988 60889.8  65847.8  68620.0  79877.7
1989 64435.4  70639.5  73072.4  85651.2
1990 70952.6  78051.7  80159.5  91532.5
```

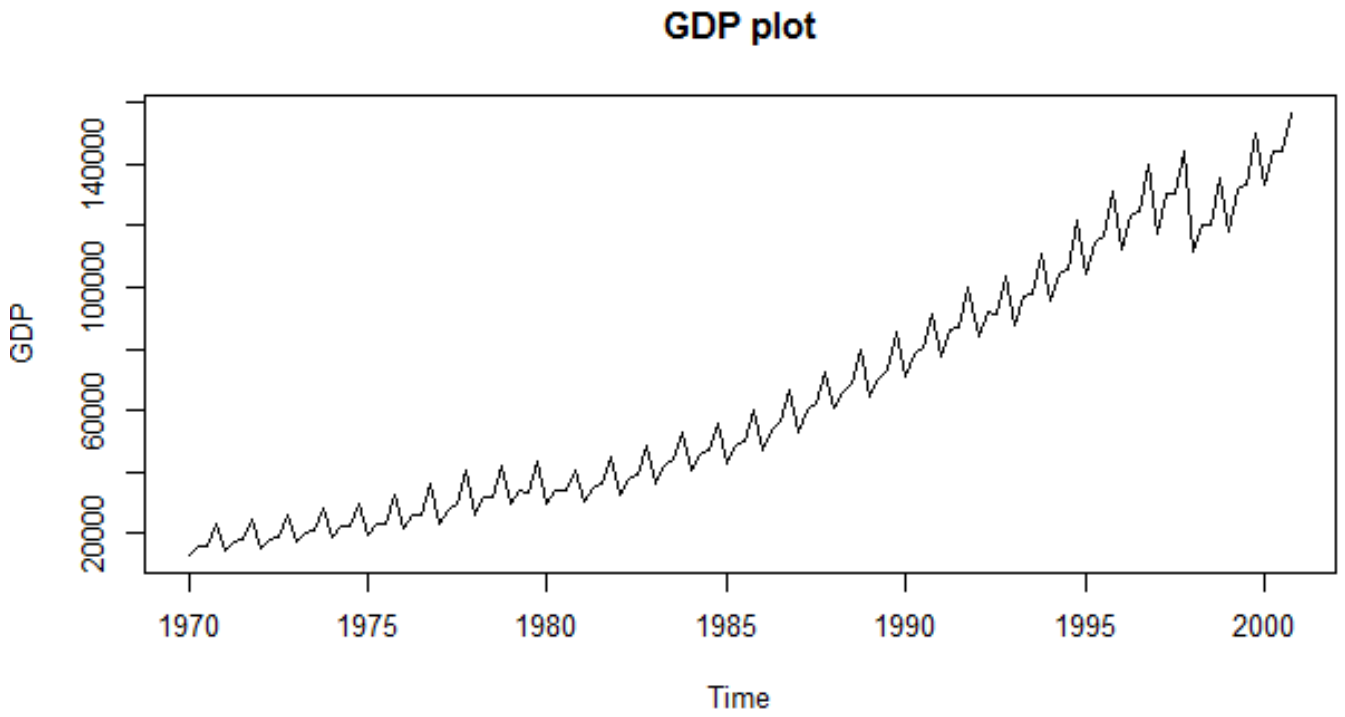
```
> class(GDP)
[1] "ts"
```

GDP 가 시계열 자료임을 보여줌

시계열(time series) 데이터 시각화

- 시계열 그래프 그리기

```
plot(GDP, main="GDP plot")
```



[연습 4]

- **1. gdp.csv** 에서 4분기 데이터만 추출하여 시계열 그래프를 그리시오

- 2. R 에서 제공하는 **AirPassengers** 데이터셋에 대해 시계열 그래프를 그리시오 (이 데이터셋 자체가 시계열 객체이므로 별도의 변환작업이 필요 없음)