4-3 클러스터링(군집) - K-mean 분석 기법 실습

In [12]:
```
iris2<-iris[,1:4]

k <- c()
km.out.withness<-c()
km.out.between<-c()
for (i in 2:25){
   set.seed(1)
    km.out<-kmeans(iris2, centers=i)
    k[i-1] <- i
    km.out.withness[i-1]<-km.out$tot.withinss
    km.out.between[i-1]<-km.out$betweenss
}
df = data.frame(k, km.out.withness, km.out.between)
df
```

k	km.out.withness	km.out.between
2	152.34795	529.0226
3	78.85144	602.5192
4	57.22847	624.1421
5	49.82228	631.5483
6	42.45606	638.9145
7	34.75675	646.6139
8	29.98894	651.3817
9	28.71578	652.6548
10	26.92642	654.4442
11	25.53673	655.8339
12	25.00831	656.3623
13	22.97014	658.4005
14	21.79146	659.5791
15	20.49144	660.8792
16	19.44148	661.9291
17	18.97767	662.3929

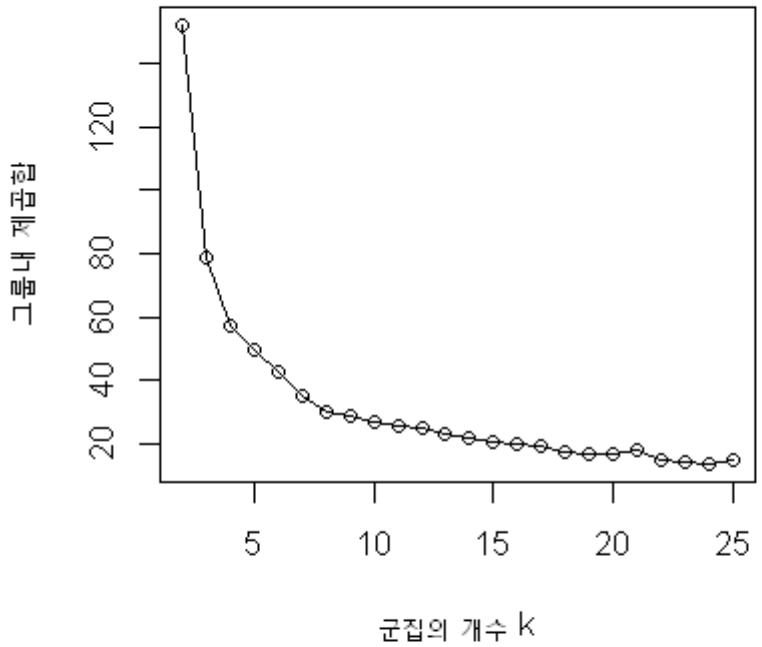
k	km.out.withness	km.out.between
18	17.16533	664.2053
19	16.54186	664.8287
20	16.86666	664.5039
21	17.60204	663.7686
22	14.95706	666.4135
23	14.06861	667.3020
24	13.30426	668.0663
25	14.49682	666.8738

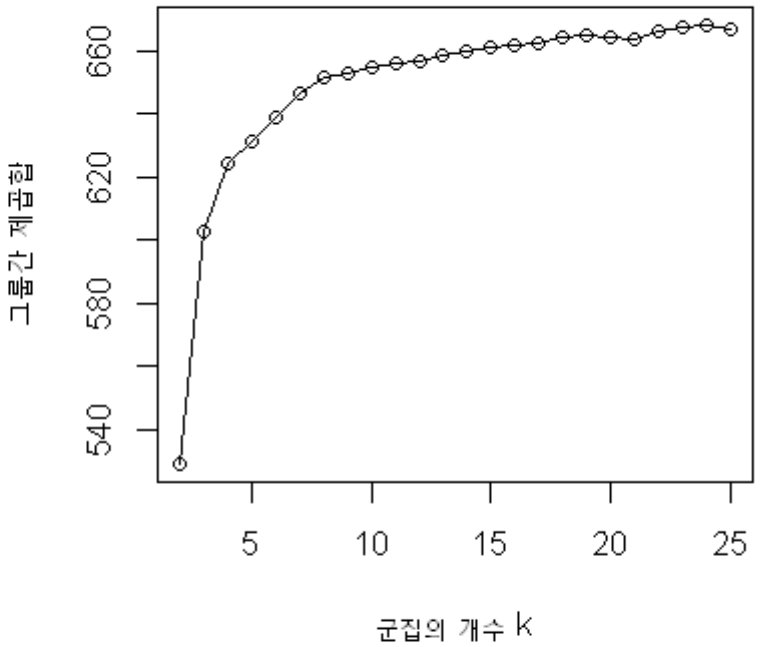
In [21]:
```
par(mfrow=c(2,2))

plot(df$k, df$km.out.withness, type='ol',
    xlab="군집의 개수 k",
    ylab="그룹내 제곱합")

plot(df$k, df$km.out.between, type='ol',
    xlab="군집의 개수 k",
    ylab="그룹간 제곱합")
```

```
Warning message in plot.xy(xy, type, ...):
"플랏 타입 'ol'은 첫번째 문자에서 잘려질 것입니다"Warning message in plot.xy(xy, type, ...):
"플랏 타입 'ol'은 첫번째 문자에서 잘려질 것입니다"
```





군집 수 K=3을 이용하여 클러스터링을 수행

In [27]:
```
km.out.k3<-kmeans(iris2, centers=3 )
km.out.k3$centers #각 군집의 중심점 출력
km.out.k3$cluster #각 관측치의 할당된 군집번호 출력
km.out.k3$size #각 군집의 데이터 관측치 개수 출력
table(km.out.k3$cluster, iris$Species) #군집결과와 원래 품종 개수 비교
```

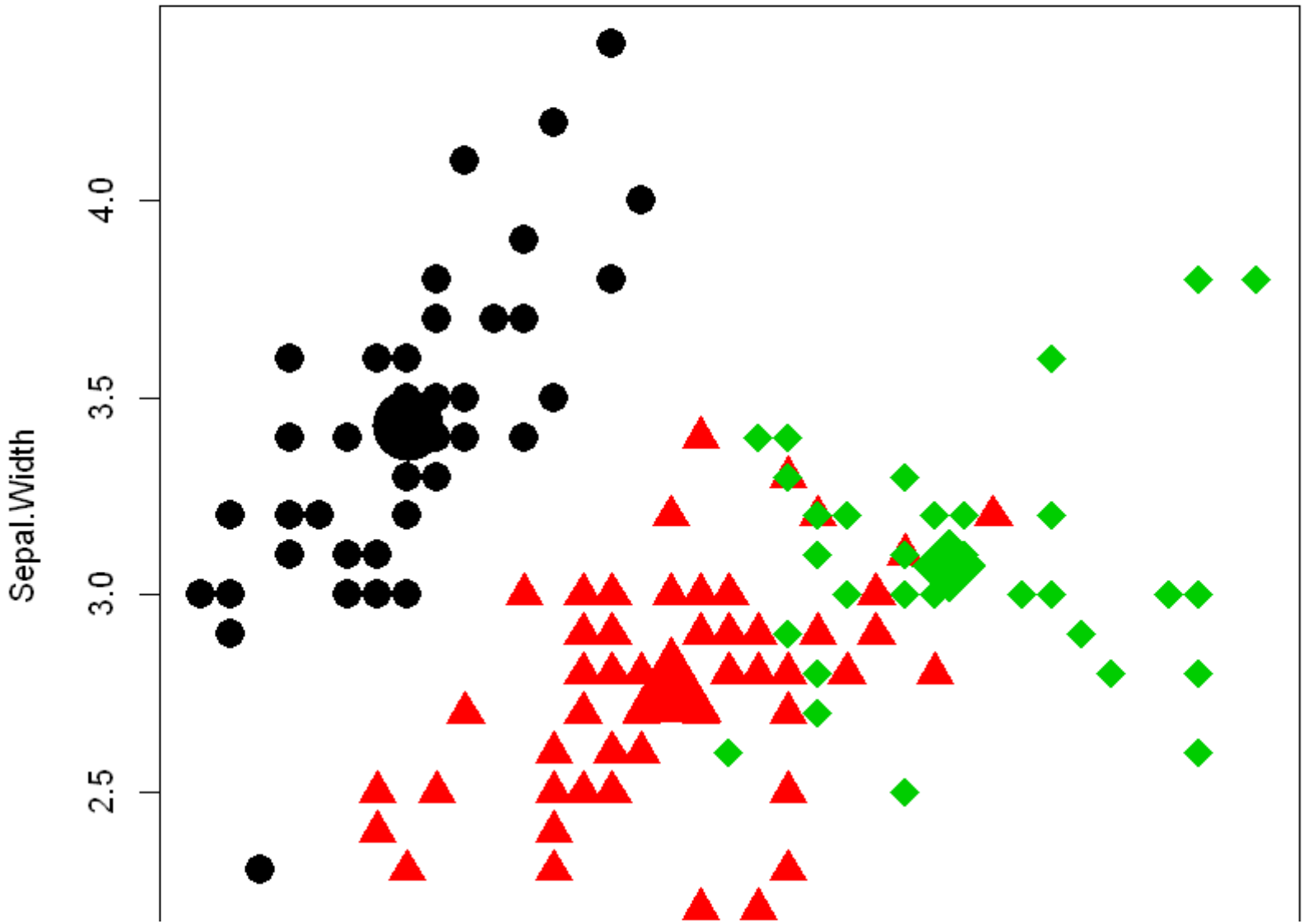
	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
1	5.006000	3.428000	1.462000	0.246000
2	5.901613	2.748387	4.393548	1.433871
3	6.850000	3.073684	5.742105	2.071053

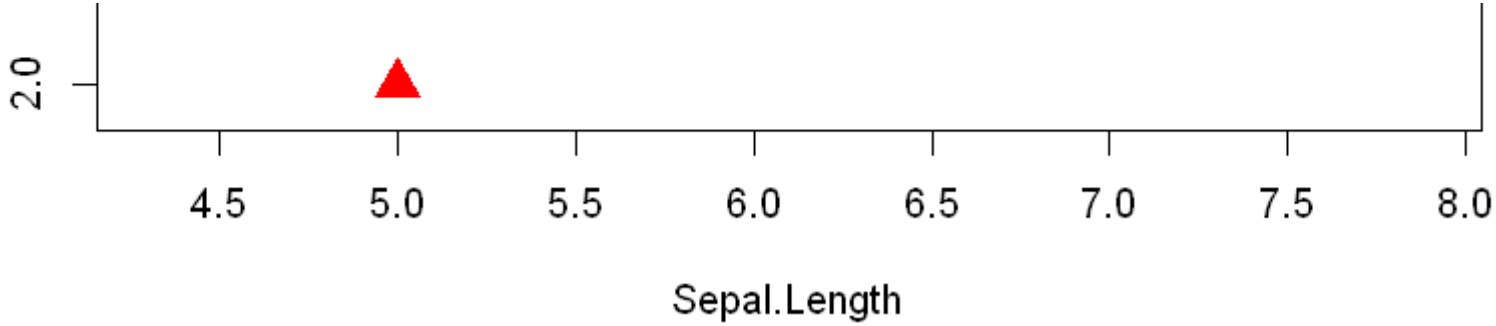
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 2 2 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 2 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 2 3 3 3 3 2 3 3 3 3 3 3 2 2 3 3 3 3 2 3 2 3 2 3 3
2 2 3 3 3 3 3 2 3 3 3 3 2 3 3 3 2 3 3 3 2 3 3 2

50 62 38

	setosa	versicolor	virginica
1	50	0	0
2	0	48	14
3	0	2	36

In [28]:
```
plot(iris2[,1:2], col=km.out.k3$cluster, pch=ifelse(km.out.k3$cluster==1, 16, ifelse(km.out.k3$cluster==2, 17,
18)), cex=2) ; points(km.out.k3$centers, col=1:3, pch=16:18, cex=5)
```





In [ ]: