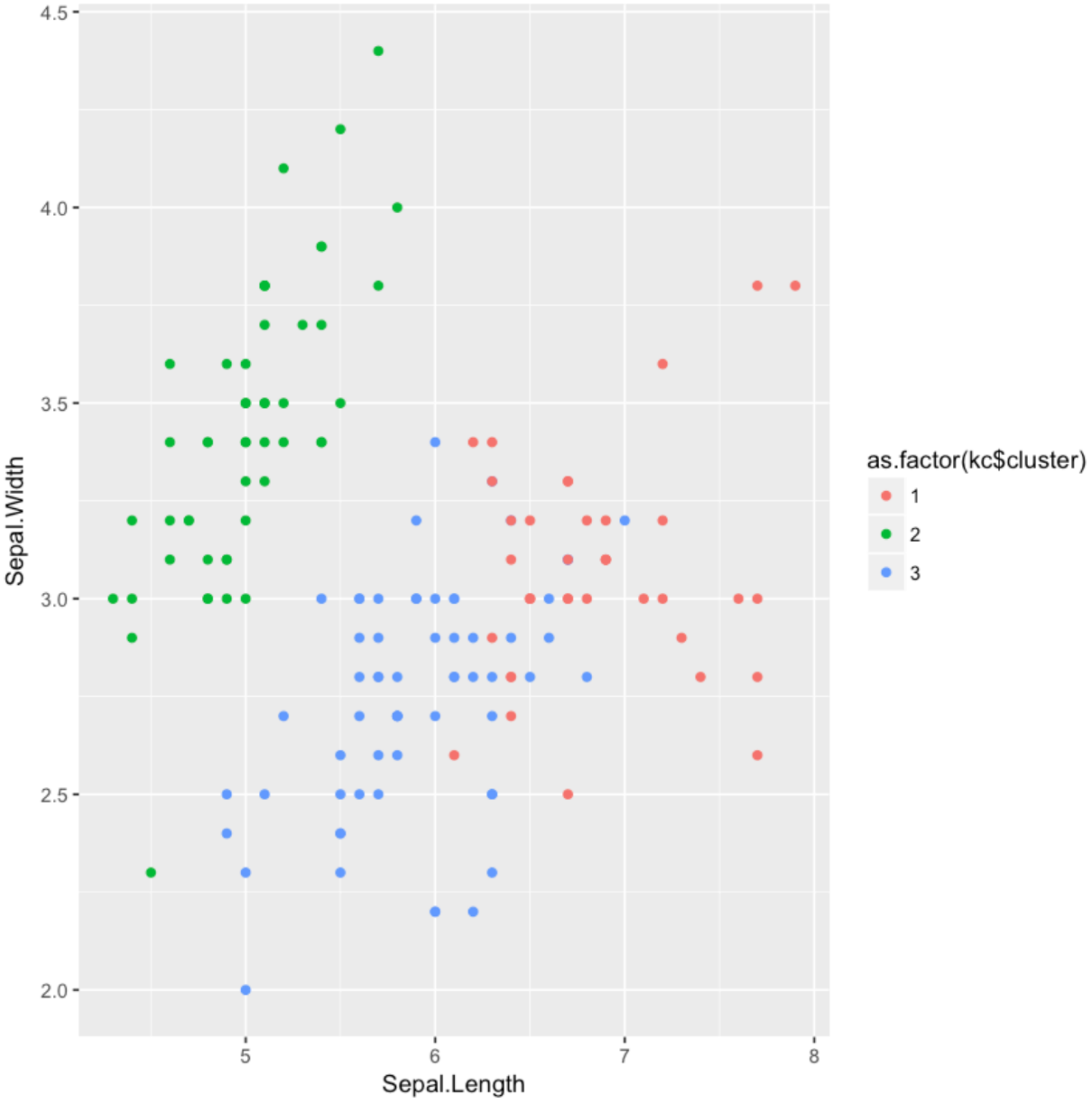
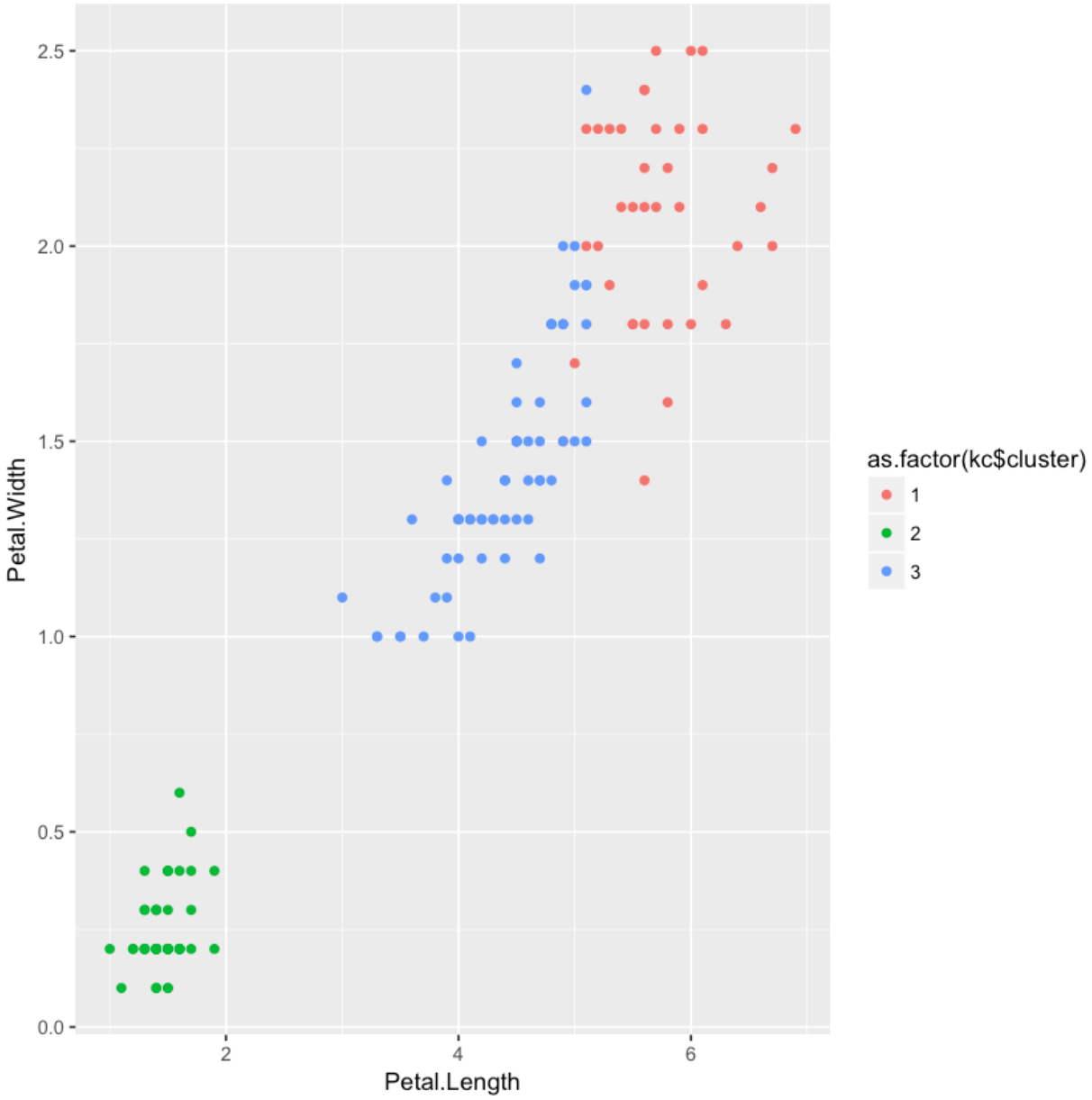
In [1]:
```
## Question 2
library(ggplot2)

data(iris)

head(iris)

x = iris[,-5]
y = iris$Species

kc <- kmeans(x,3)
kc

table(y,kc$cluster)

## Sepal Width vs Sepal Length
ggplot(x,aes(x = Sepal.Length, y= Sepal.Width,color = as.factor(kc$clu
ster))) + geom_point()

## Petal Width vs Petal Length
ggplot(x,aes(x = Petal.Length, y= Petal.Width,color = as.factor(kc$clu
ster))) + geom_point()
```

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa

```
K-means clustering with 3 clusters of sizes 38, 50, 62

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

Clustering vector:
  [1] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
 [38] 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
 [75] 3 3 3 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 1 1 1 1 3 1 1 1 1
[112] 1 1 3 3 1 1 1 1 3 1 3 1 3 1 1 3 3 1 1 1 1 1 3 1 1 1 1 3 1 1 1 3 1 1 1 3 1
[149] 1 3

Within cluster sum of squares by cluster:
[1] 23.87947 15.15100 39.82097
 (between_SS / total_SS =  88.4 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot
.withinss"
[6] "betweenss"    "size"         "iter"         "ifault"


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


4.5
4.0
3.5
3.0
2.5
2.0
Sepal.Width
5
6
7
8
Sepal.Length
as.factor(kc$cluster)
1
2
3


Petal.Width
Petal.Length
2.5
2.0
1.5
1.0
0.5
0.0
2
4
6
as.factor(kc$cluster)
1
2
3