

Homework 1

Emer Gunne

June 5, 2024

Clustering on Iris dataset

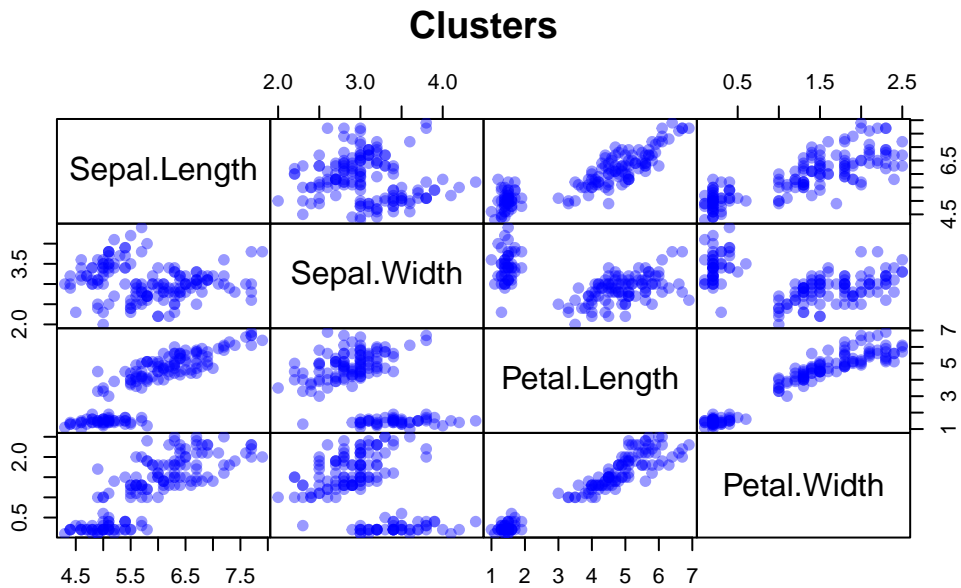
The Iris dataset in R programming contains 150 observations of 4 variables:

Sepal Length & Sepal Width , Petal Length & Petal Width

The 5th Column contains the **Species** label for each observation of which there are three: **Setosa** , **Versicolor** and **Virginica**

View pairs of clusters:

I would like to view a matrix of the 4 variables showing how we can see clusters based on the dimensions of the variables which we can then relate to species



Descriptive statistics

I will now investigate the variation in the dataset using summary statistics

All Species

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Min. :4.300	Min. :2.000	Min. :1.000	Min. :0.100
1st Qu.:5.100	1st Qu.:2.800	1st Qu.:1.600	1st Qu.:0.300
Median :5.800	Median :3.000	Median :4.350	Median :1.300
Mean :5.843	Mean :3.057	Mean :3.758	Mean :1.199
3rd Qu.:6.400	3rd Qu.:3.300	3rd Qu.:5.100	3rd Qu.:1.800
Max. :7.900	Max. :4.400	Max. :6.900	Max. :2.500

Species

setosa :50
versicolor:50
virginica :50

Setosa

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Min. :4.300	Min. :2.300	Min. :1.000	Min. :0.100
1st Qu.:4.800	1st Qu.:3.200	1st Qu.:1.400	1st Qu.:0.200
Median :5.000	Median :3.400	Median :1.500	Median :0.200
Mean :5.006	Mean :3.428	Mean :1.462	Mean :0.246
3rd Qu.:5.200	3rd Qu.:3.675	3rd Qu.:1.575	3rd Qu.:0.300
Max. :5.800	Max. :4.400	Max. :1.900	Max. :0.600

Species

setosa :50
versicolor: 0
virginica : 0

Versicolor

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
Min. :4.900	Min. :2.000	Min. :3.00	Min. :1.000	setosa : 0
1st Qu.:5.600	1st Qu.:2.525	1st Qu.:4.00	1st Qu.:1.200	versicolor:50
Median :5.900	Median :2.800	Median :4.35	Median :1.300	virginica : 0
Mean :5.936	Mean :2.770	Mean :4.26	Mean :1.326	
3rd Qu.:6.300	3rd Qu.:3.000	3rd Qu.:4.60	3rd Qu.:1.500	
Max. :7.000	Max. :3.400	Max. :5.10	Max. :1.800	

Virginica

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Min. :4.900	Min. :2.200	Min. :4.500	Min. :1.400
1st Qu.:6.225	1st Qu.:2.800	1st Qu.:5.100	1st Qu.:1.800
Median :6.500	Median :3.000	Median :5.550	Median :2.000
Mean :6.588	Mean :2.974	Mean :5.552	Mean :2.026
3rd Qu.:6.900	3rd Qu.:3.175	3rd Qu.:5.875	3rd Qu.:2.300
Max. :7.900	Max. :3.800	Max. :6.900	Max. :2.500

Species

setosa : 0

versicolor: 0

virginica :50

In summary

Species Setosa has the shortest sepals and petals, widest sepals and narrowest petals.

Species Versicolor has the narrowest sepals.

Species Virginica has the longest sepals and petals.

Petal length and petal width set the species Setosa apart as can be seen in the cluster matrix.