TASK 1: k-Means

Corresponds (more or less) to the three expected species? ☐YES ☐NO   No as there is only 1 sample in one cluster

Number of records in each cluster: 1) 193 2) 106 3) 1

TASK 2: preprocessing

Is it better to rescale before or after detecting and filtering out the outliers?   Rescale after filtering out outliers

Corresponds (more or less) to the three expected species? ☐YES ☐ NO -> Yes

Number of records in each cluster: 1) 98 2) 95 3) 92

Coordinates of the three centroids:

	PW	PL	SW	SL
1)	0.0566	0.0652	0.0847	0.0784
2)	0.066	0.050	0.0122	0.0210
3)	0.0528	0.0608	0.0559	0.0626

TASK 3: choice of k

Which K corresponds to the best clustering? (using the Davies-Boulding index).
k=2 -> score: 0.3639

TASK 4: Hierarchical clustering

Using SingleLink, how many records are included in each of the two top clusters?

Cluster 1:   187

Cluster 2:   98

Which approaches produce a (more or less) correct clustering corresponding to the three species, if any?

SingleLink: No -> since individual points get added to the big clusters at later stages

CompleteLink: Yes, there are 3 distinct clusters which are merged in the last 2 steps

AverageLink: Yes, there are 3 distinct clusters which are merged in the last 2 steps

TASK 5: DB-Scan

How many clusters does DB-SCAN find with eps=1, min_samples=5?   1

Can you give a value for epsilon leading to two clusters (plus noise)?   0.02

K-DISTANCES:

Which K did you use?   5

According to the k-distances plot, what value(s) of epsilon would you consider as a parameter to DB-Scan and why?

From the k-distance plot, the "ellbow" occurs at around 0.02