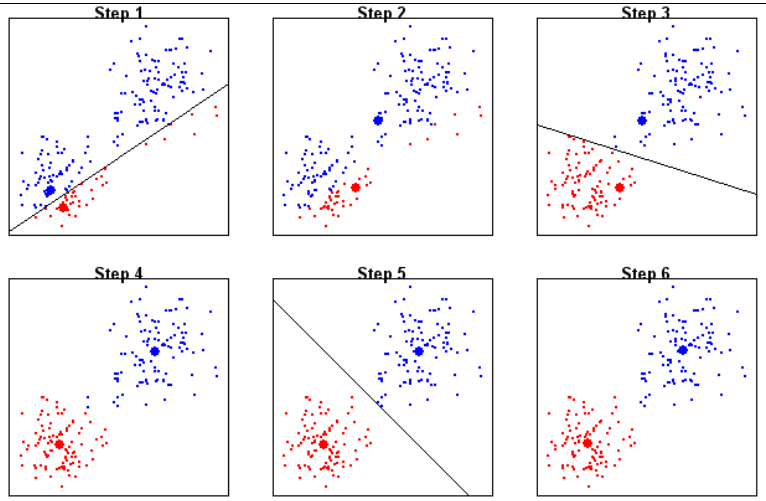
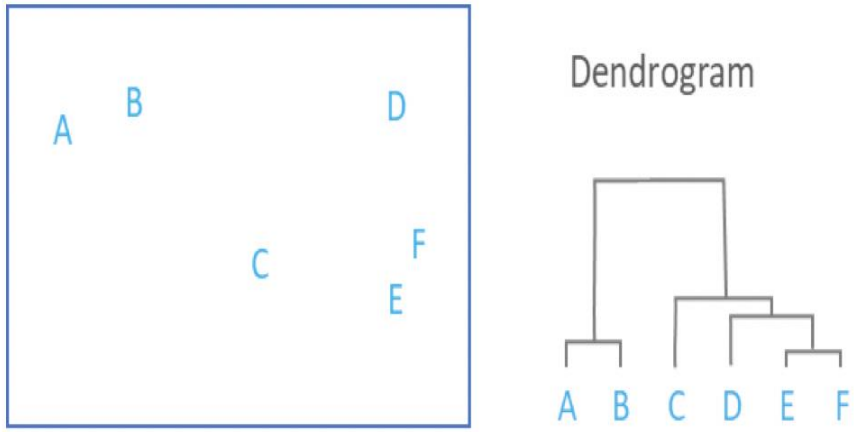
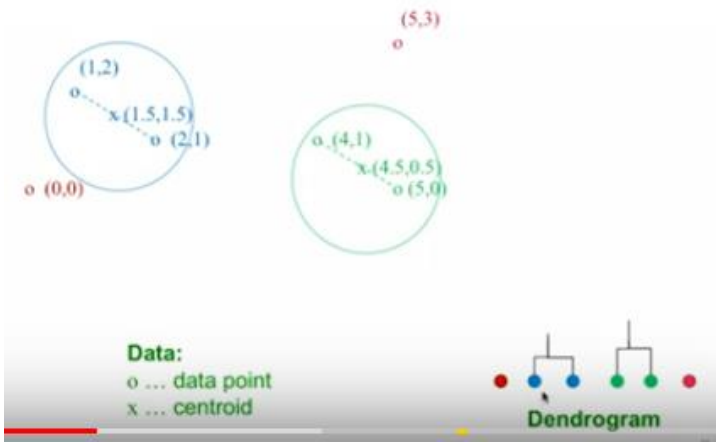
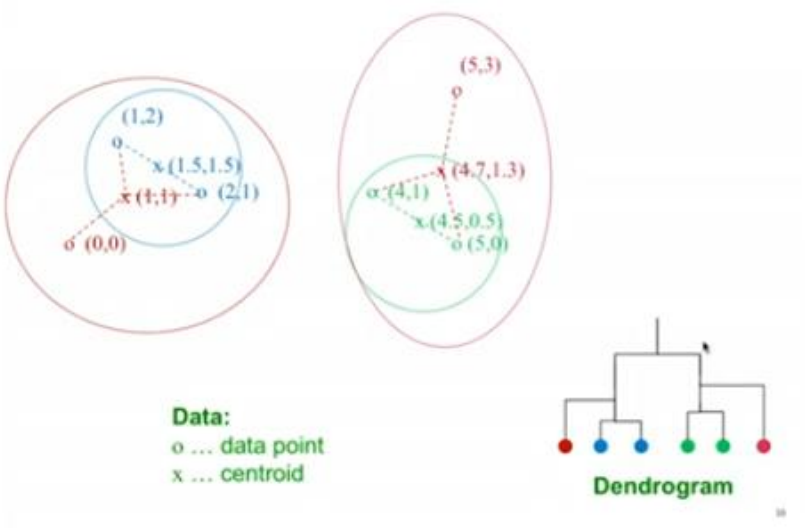
K-Mean Clustering and Hierarchical Clustering

	K-mean Clustering	**Hierarchical Clustering**
Key operation	Aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean	Repeatedly combine two nearest clusters **Centroid**: the average of all data points in the cluster **Clustroid**: an existing data point that is closest to all other points in the cluster
Operation chart		
Comparison	• time complexity • produces a single partitioning • need K – number of clusters to be specified • starts with a random choice of cluster centers, therefore it may yield different clustering results on different runs of the algorithm	• can be slow • give different partitioning depending on the level of resolution • doesn't need number of clusters to be specified • reproductable, can get the same clustering result on different runs • cannot handle big data





With python code

https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1

https://www.geeksforgeeks.org/k-means-clustering-introduction/

https://www.datascience.com/blog/k-means-clustering

http://stanford.edu/~cpiech/cs221/handouts/kmeans.html