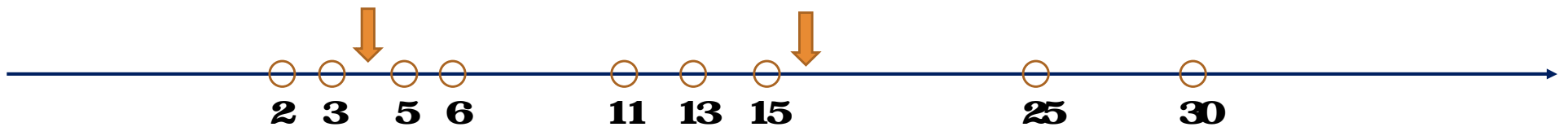
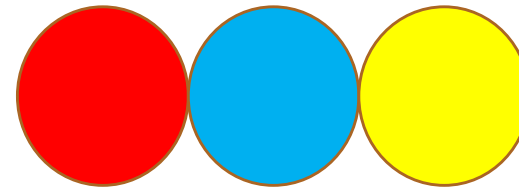


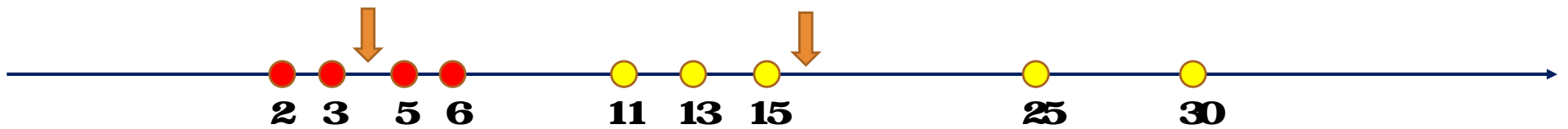
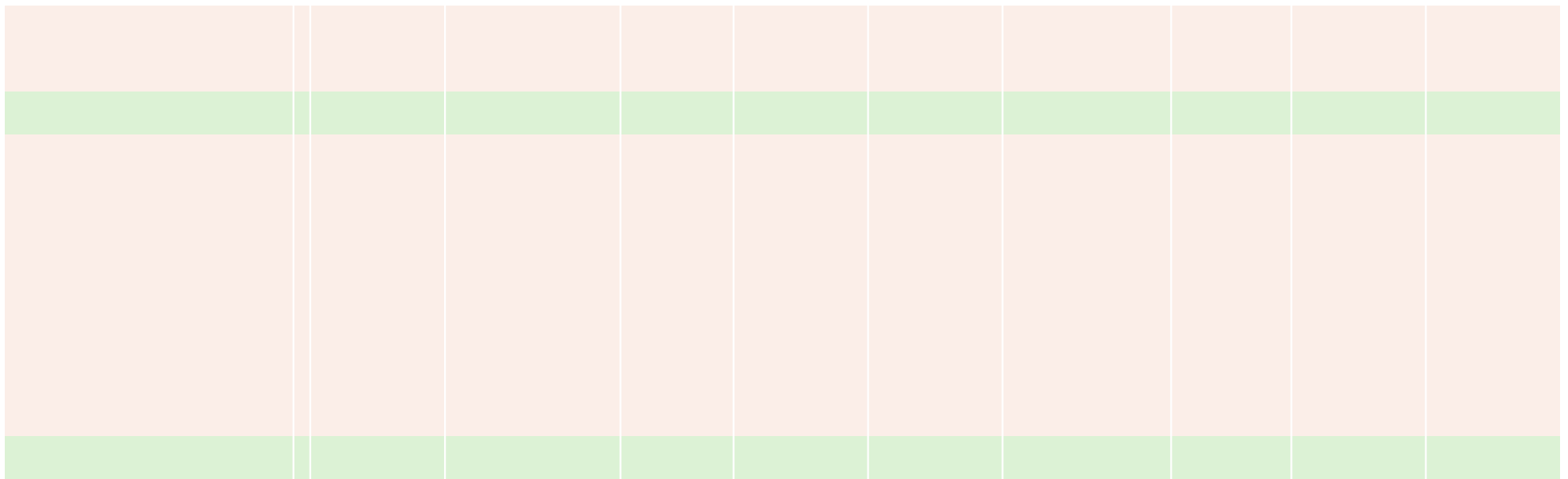
Then, find the distance between each point and each mean

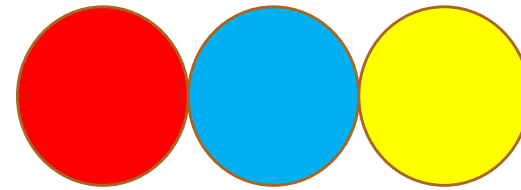
Data	2	3	5	6	11	13	15	25	30
K= 2									
1st iteration		Mean1 =	4			Mean2 =	16		
Distance to Mean1	2	1	1	2	7	9	11	21	26
Distance to Mean2	14	13	11	10	5	3	1	9	14
Cluster Number									
		C1 total =				C2 total =			
		C1 count =				C2 count =			
		Mean1 =				Mean2 =			



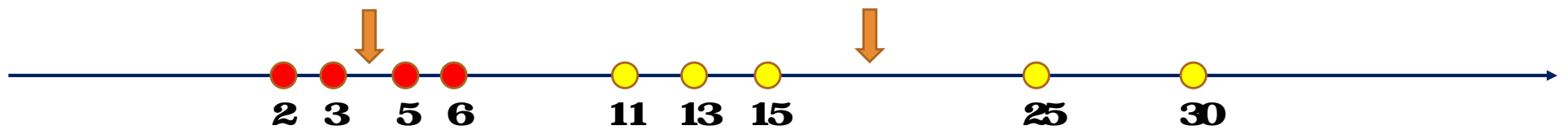
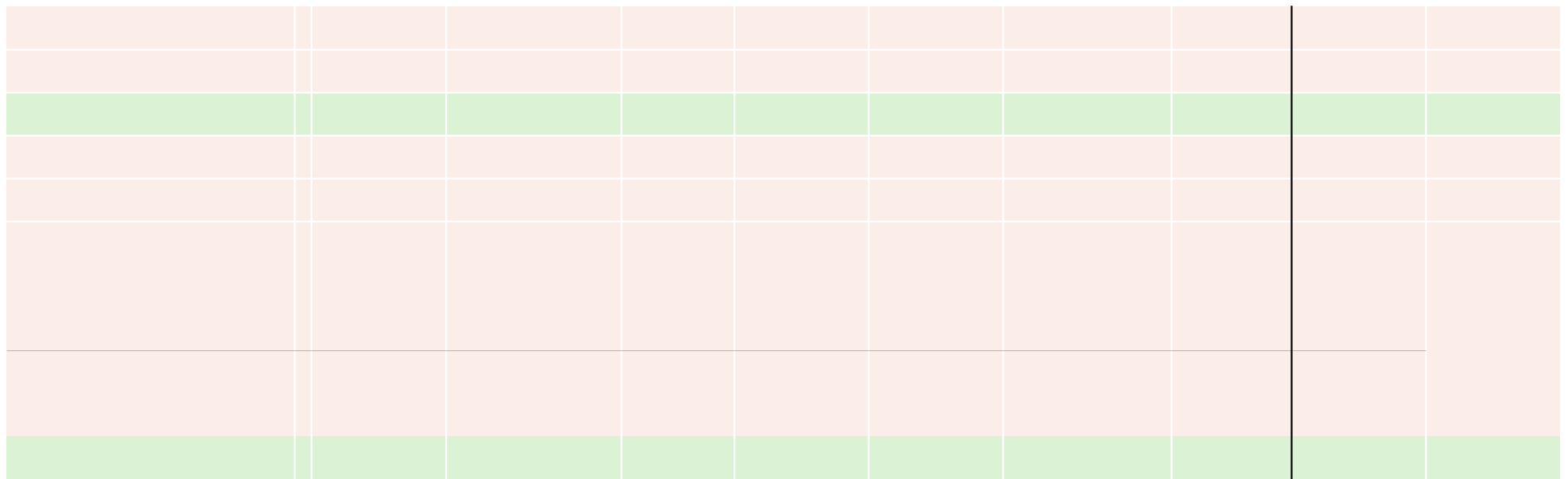


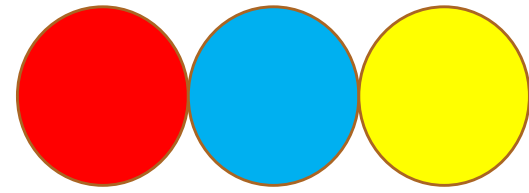
Use the shortest distance to assign a point to a cluster

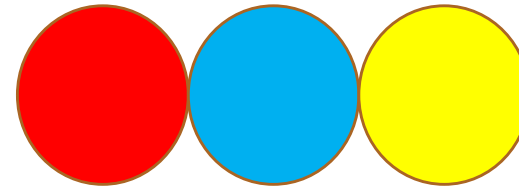




Update the mean of each cluster

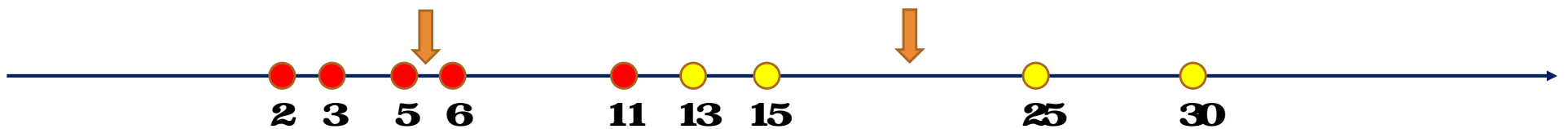


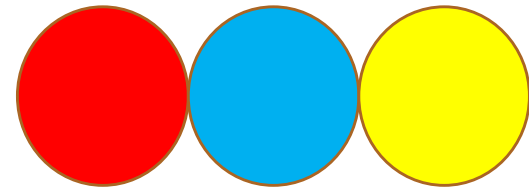


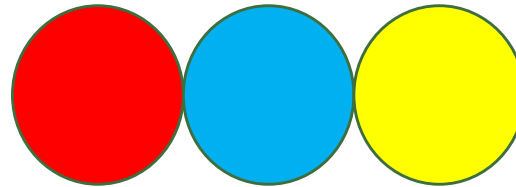


The means stay the same at the 3^d iteration

Data	2	3	5	6	10	13	15	25	30
K=2									
3^d iteration		Mean1= 54				Mean2= 20.75			
Distance to Mean1	2	1	1	2	7	9	11	21	26
Distance to Mean2	18.75	17.75	15.75	14.75	9.75	7.75	5.75	4.25	9.25
Cluster Number	C1	C1	C1	C1	C1	C2	C2	C2	C2
		C1 total= 27				C2 total= 83			
		C1 count= 5				C2 count= 4			
		Mean1= 54				Mean2= 20.75			



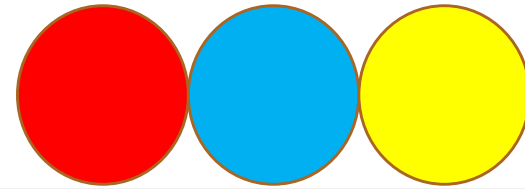




Advanced Questions

Is there any method to determine the optimal value of k ?

Other than absolute value (Manhattan distance), may we use other distance formula?

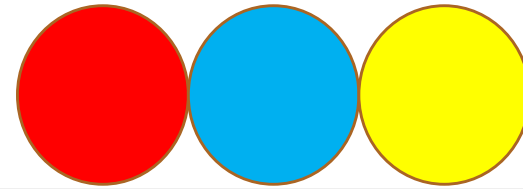


Applications

Image compression

Document analysis

Market segmentation



Further Topics

KMeans Clustering

Hierarchical Clustering

Hard/Soft Clustering

Clustering with Outliers