

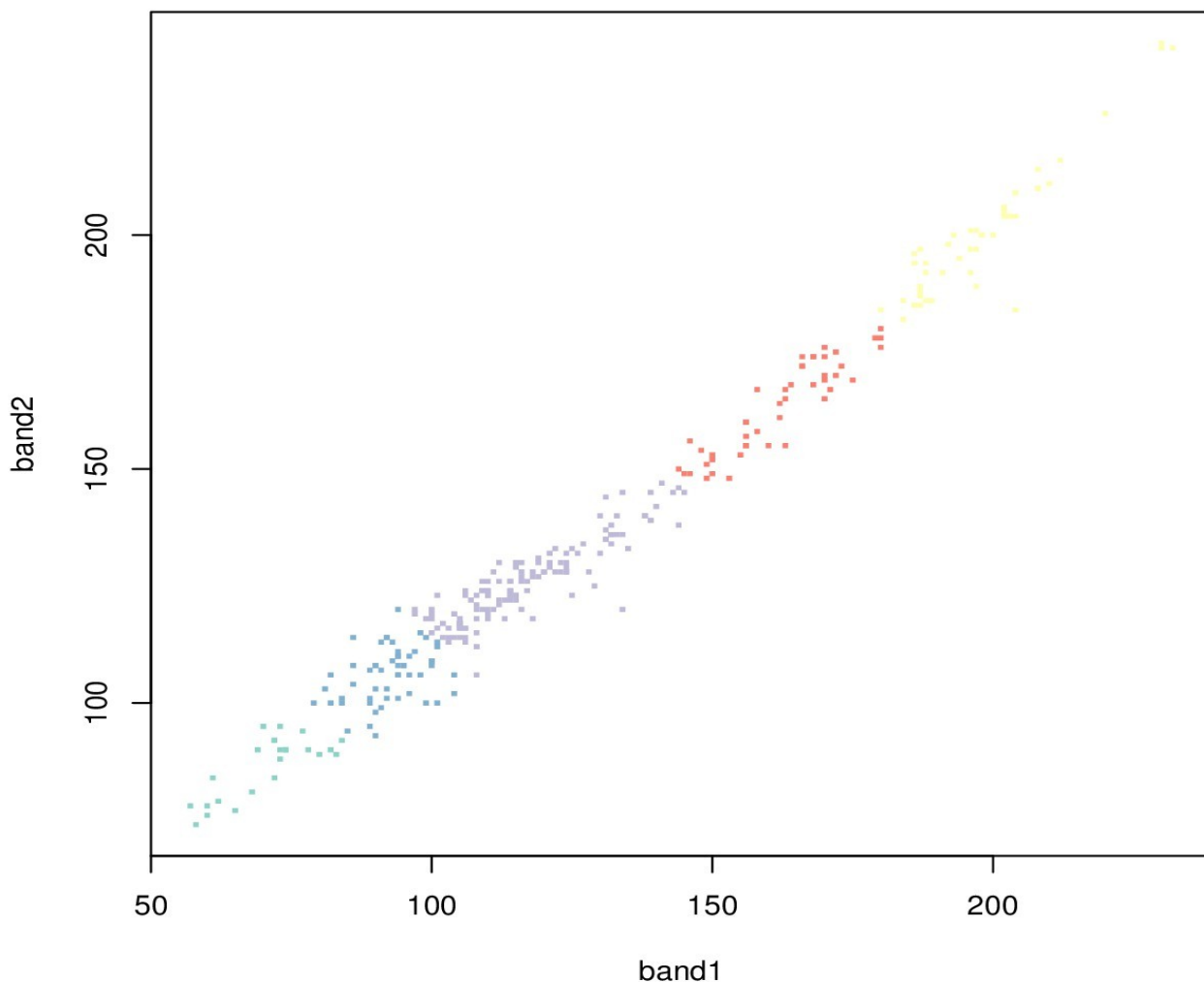
Q4 b)

The clusters obtained through, bisecting Kmeans and through the use of regular Kmeans with provided centers vary, both in terms of cluster assignment as well as cluster SSEs.

In the case of bisecting Kmeans, clusters with large values of SSE are bisected into smaller clusters, The output of the bisecting Kmeans algorithm produces three large clusters with large values of SSE and the remaining are small clusters with low SSE. The largest SSE is 20363.244.

We can also observe that there are 3 clusters which are much larger in size and the other 2 are very small, this is a result of bisection of cluster at the previous iteration, producing 2 small clusters from the largest available. This is because, Since SSE is a quadratic operation, On dividing the cluster size by 2 on an average, The SSE decrease by a factor of 4.

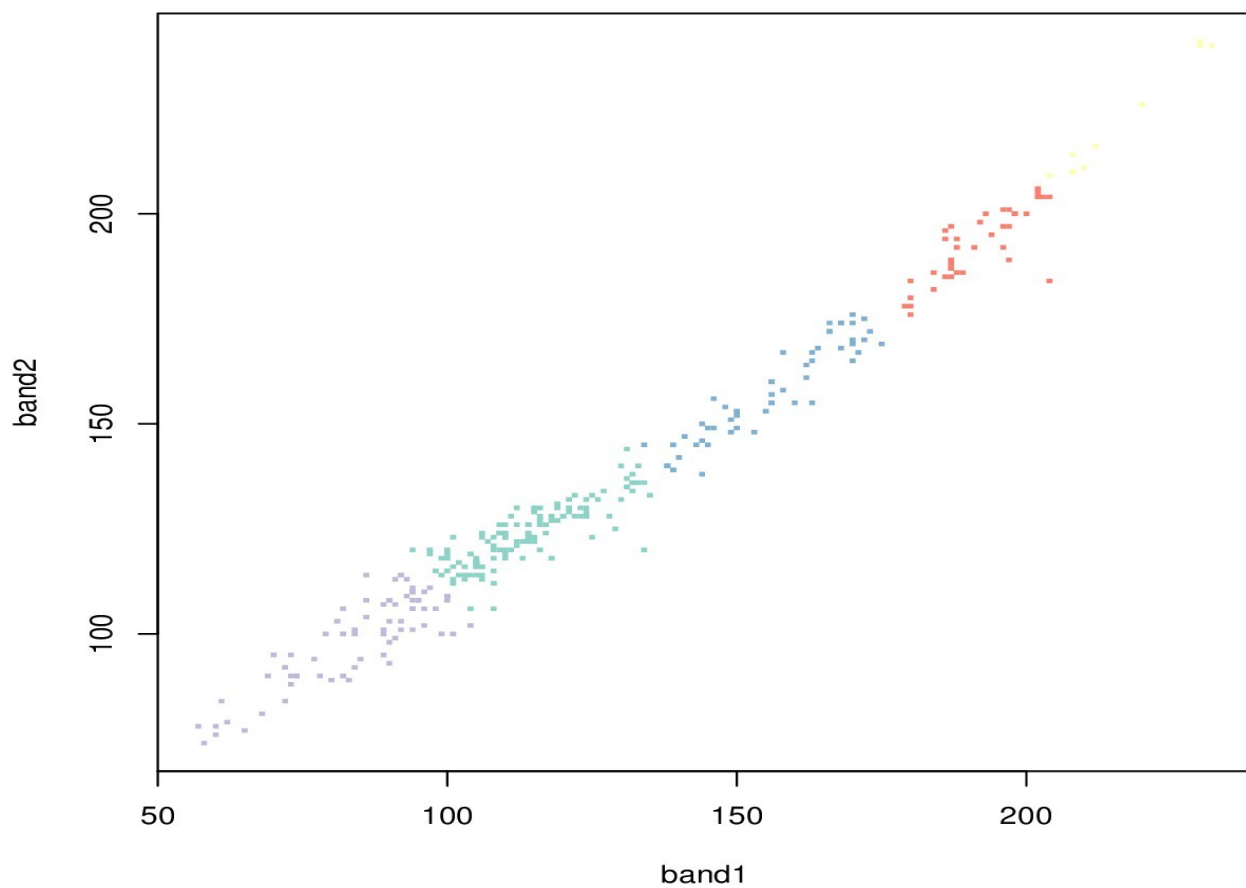
The Assignment of points of the bisecting Kmeans algorithm can be seen below:



Thus we can see, that there are 3 fairly large clusters and the other two are very small in size. This is reflected by the SSE values as well which are as follows:

	SSE
C1	3629.870
C2	9113.739
C3	20363.244
C4	16067.778
C5	2448.917

Where as in the case of Kmeans with given initial centers, The clusters assignments are a bit more even in terms of size of clusters and can be seen below



Here it can be seen that, the clusters are more even in size than in the case of bisecting Kmeans. The SSEs of the clusters is given below:

	SSE
C1	19626.720
C2	2753.700
C3	17846.222
C4	4860.462
C5	14214.926

However, when we see the values of SSE, there again a sharp divide in the quality of clusters can be seen, with 3 clusters having very high values of SSE, and 2 having fairly low values. This sharp divide in the values SSE is due to the presence of noise points in some clusters which have drastically increased the value of SSE.

Obviously, the numbering of the clusters vary in both cases, as in the first case, as the clusters are divided, the cluster number is incremented, whereas in the case of Kmeans, initial cluster centres have been provided to us.

The shape of the cluster in each case is elongated, which is because the given data is distributed in that manner. Thus the cluster shape is due to the distribution of points in the given space and not due to the clustering algorithm as both of them generally produce globular clusters.