

## Exercise 16

Bilal Kudaimi

October 30, 2020

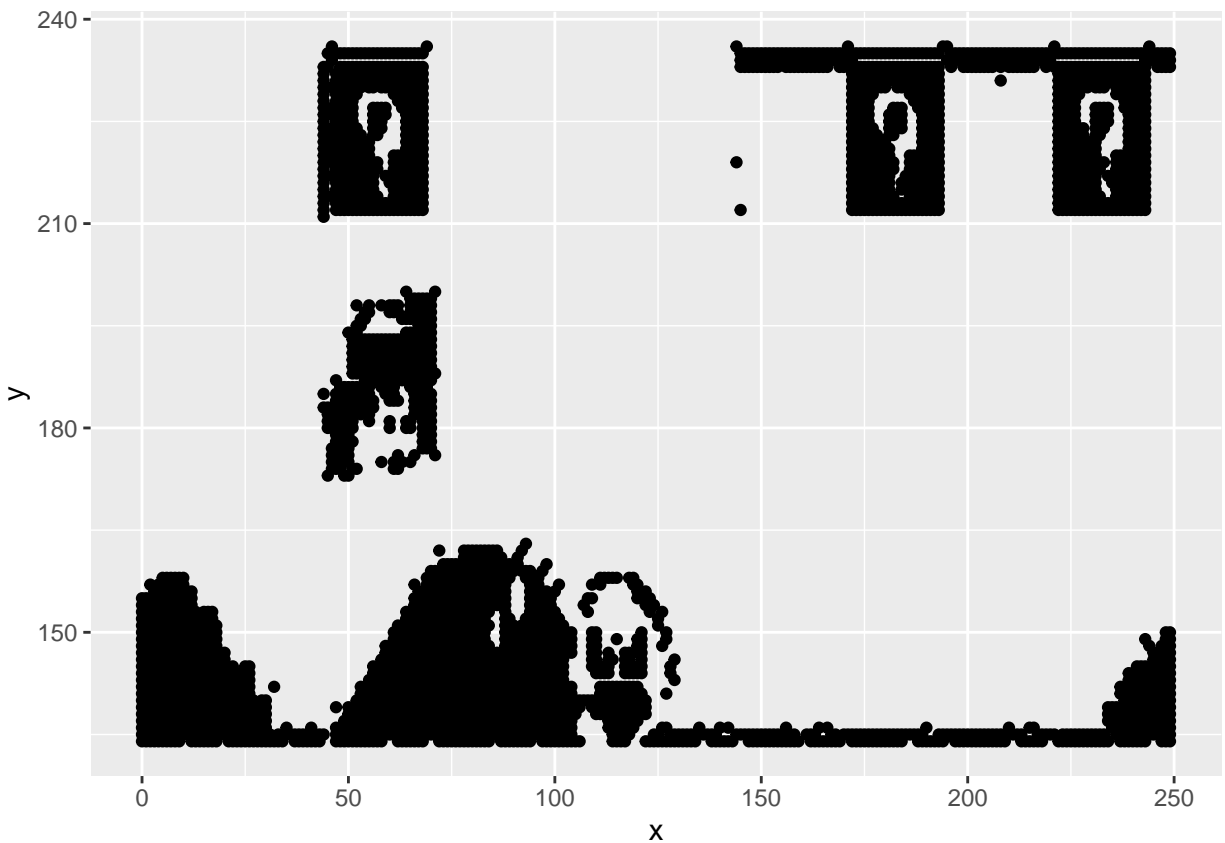
#A: Scatter plot of the data

#B: Fitting the dataset with the k-means algorithm and making a scatter plot of clusters for each K value

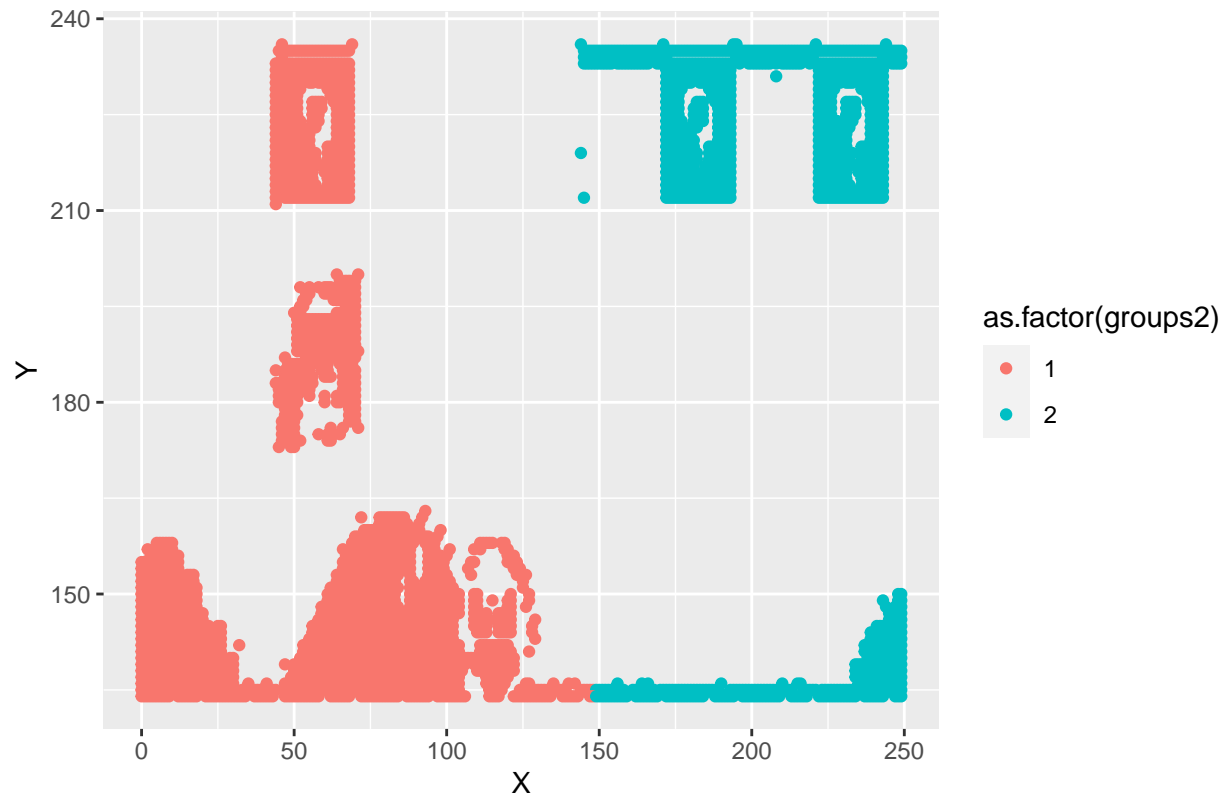
#C: Plot of average distance to cluster center versus K values

```
## Warning: package 'factoextra' was built under R version 4.0.3
```

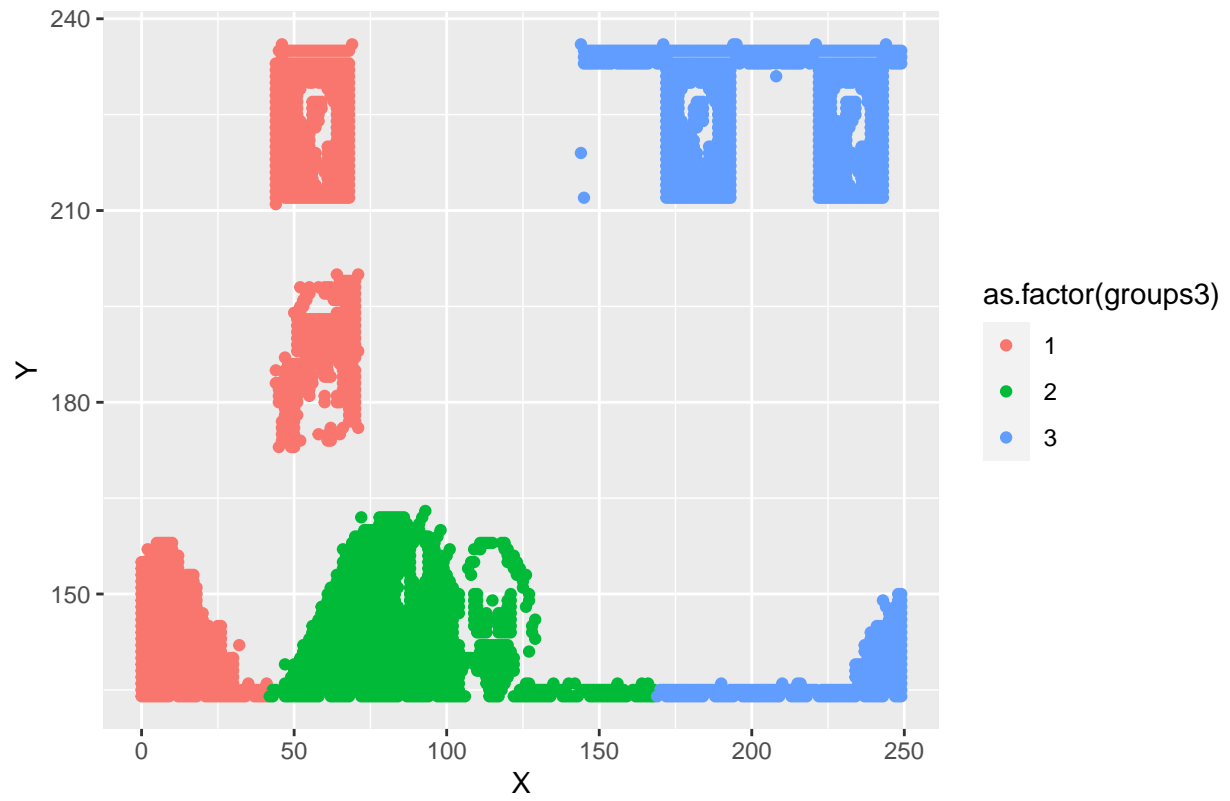
```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

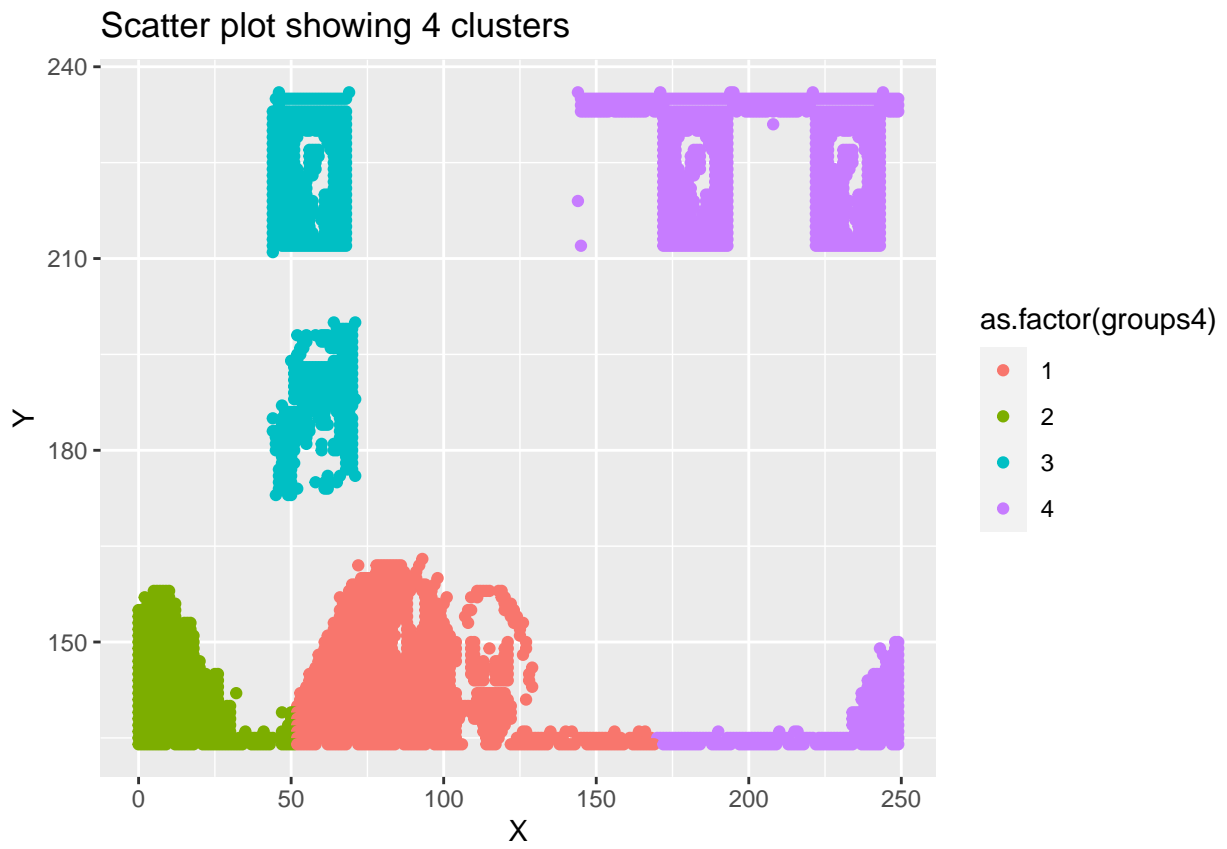


Scatter plot showing 2 clusters

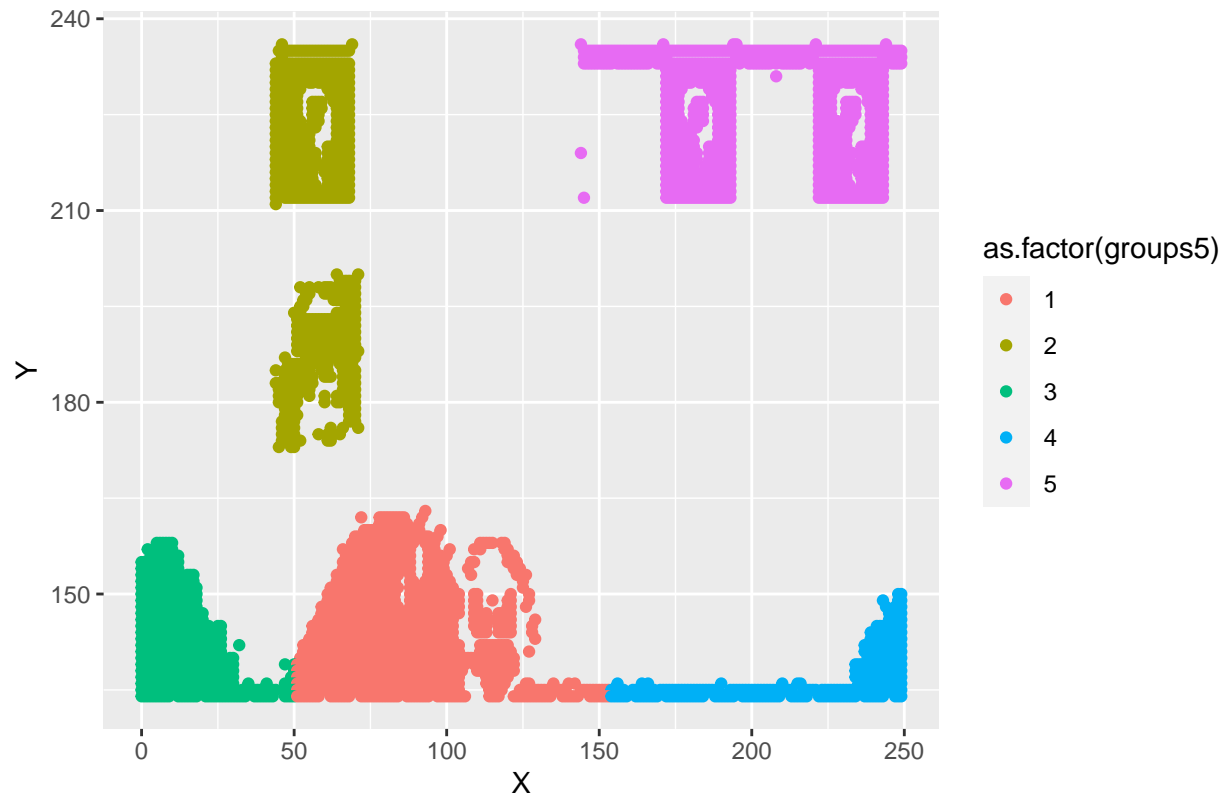


Scatter plot showing 3 clusters

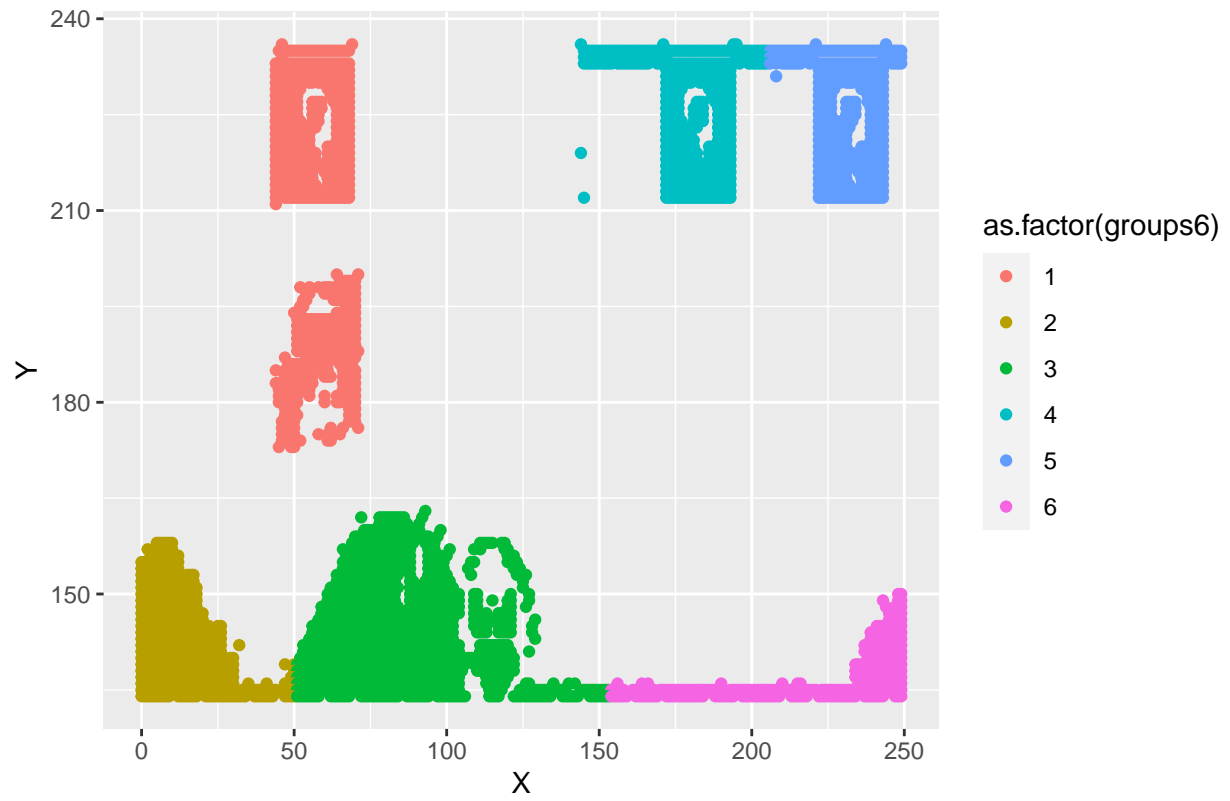




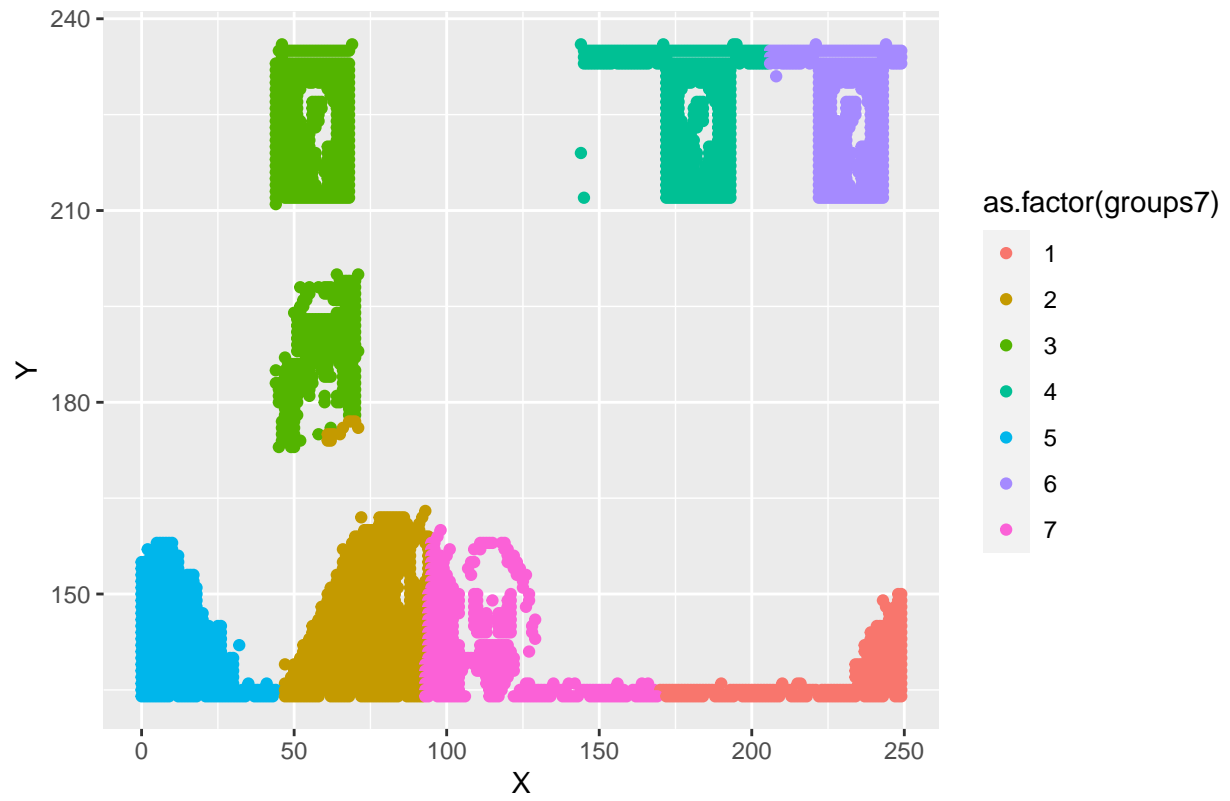
Scatter plot showing 5 clusters



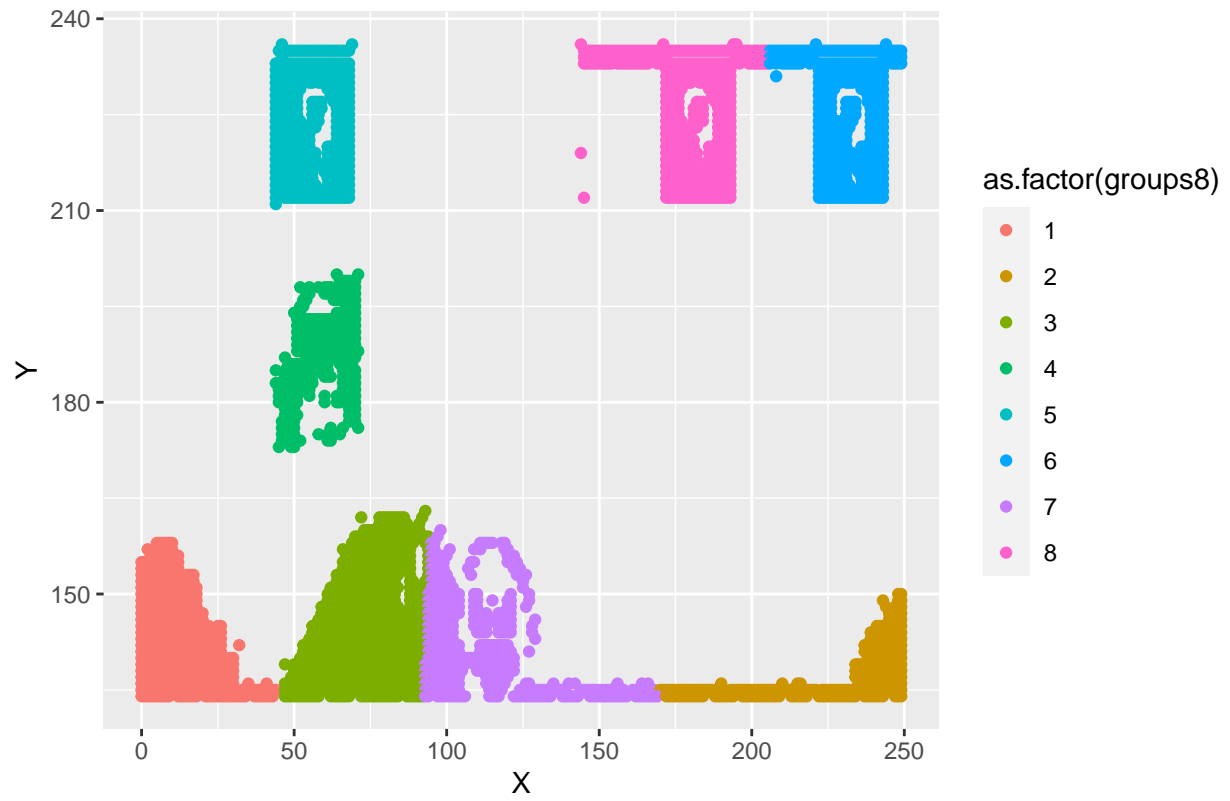
Scatter plot showing 6 clusters



Scatter plot showing 7 clusters

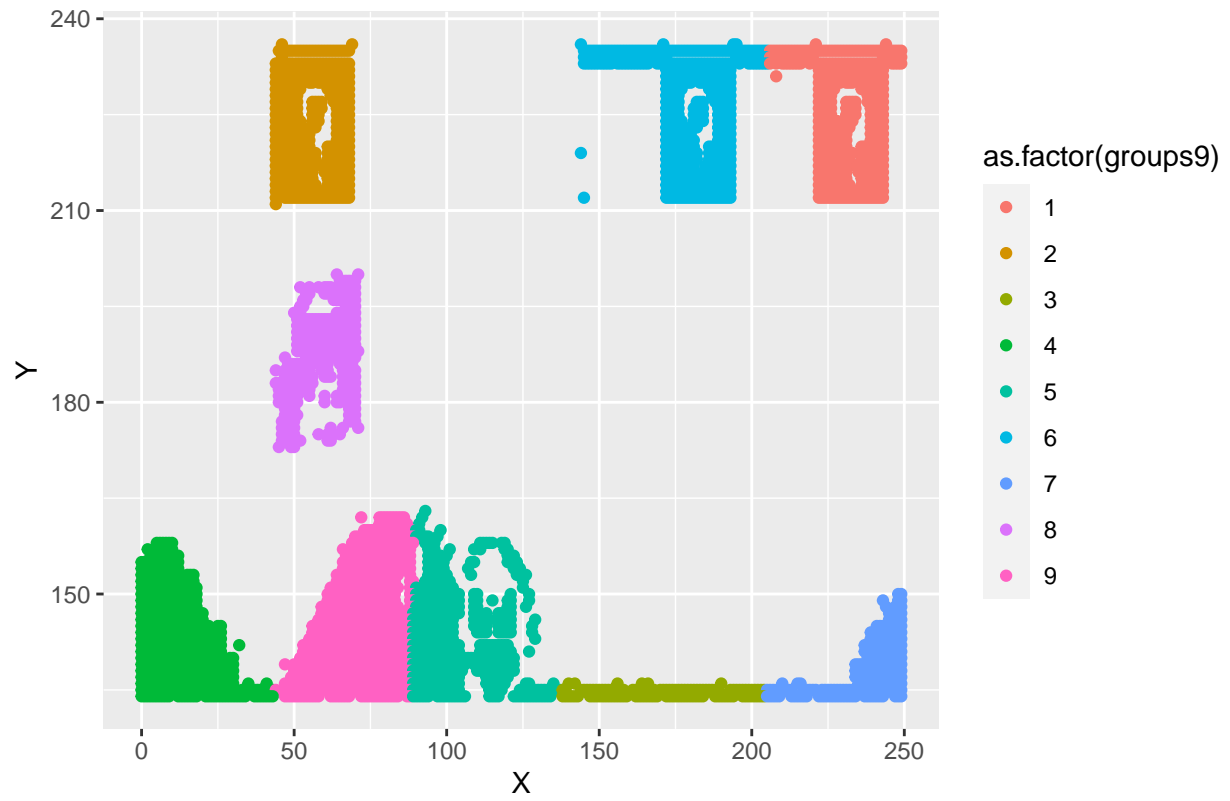


Scatter plot showing 8 clusters

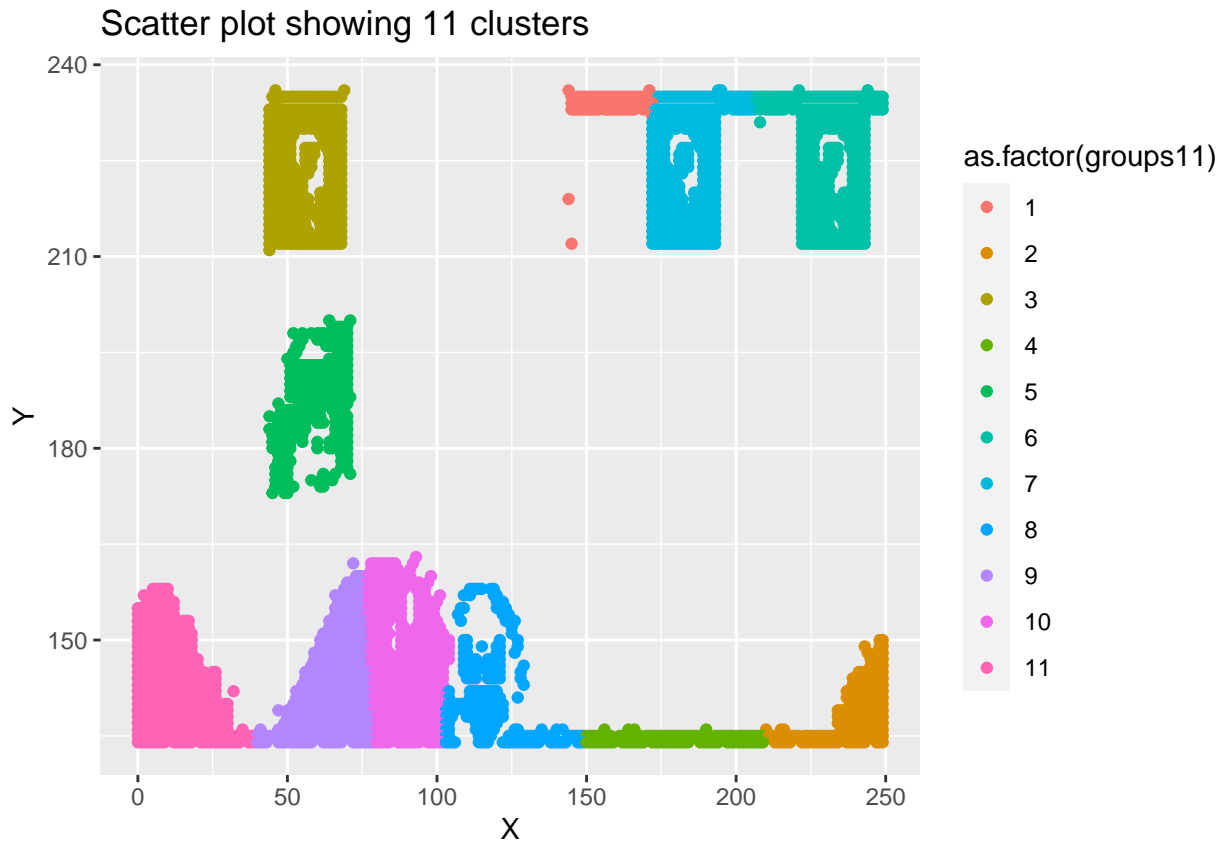




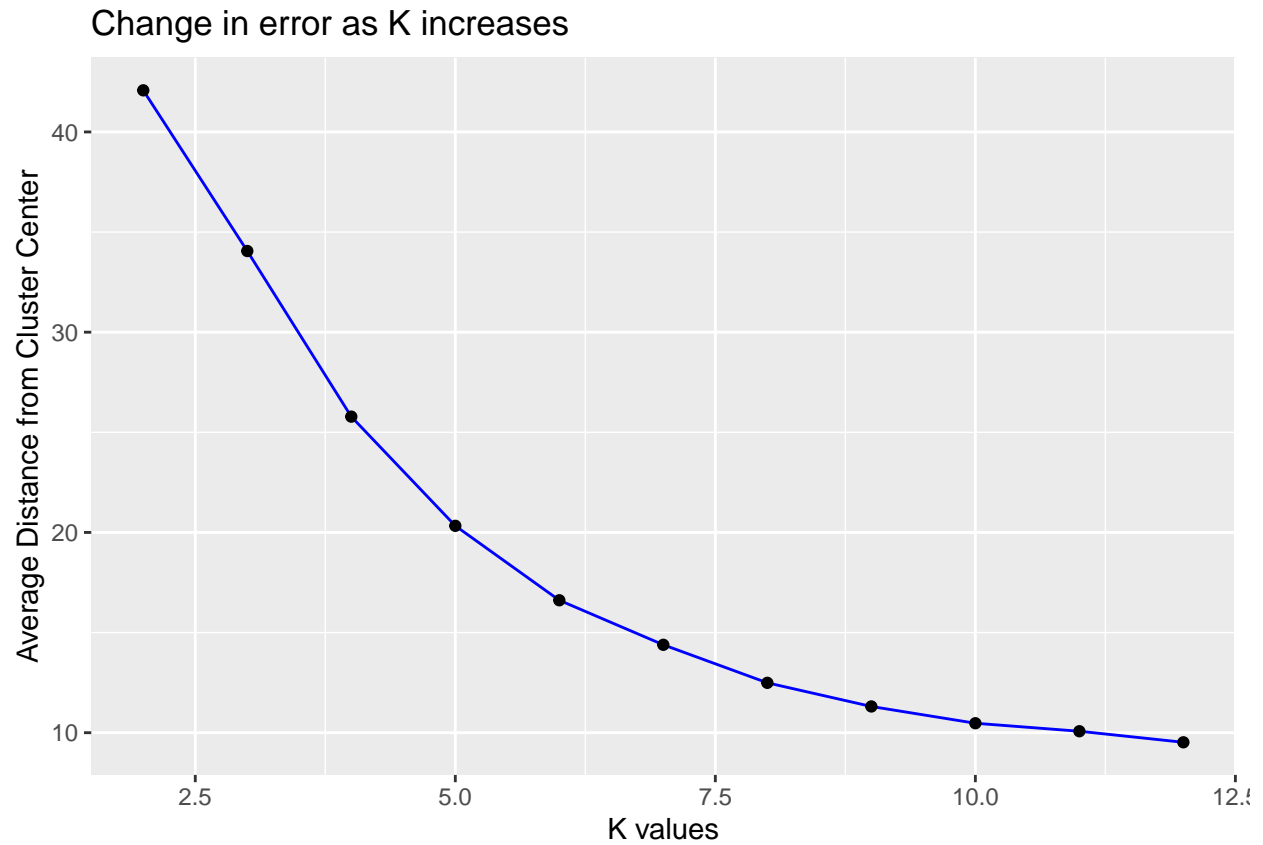
Scatter plot showing 9 clusters











We can see that the average distance from a cluster center decreases as K increases. There are many elbows to this plot, indicating there is no clear answer as to which K value is optimal. Perhaps this is because the data points form an image from the video game Super Mario Bros.