

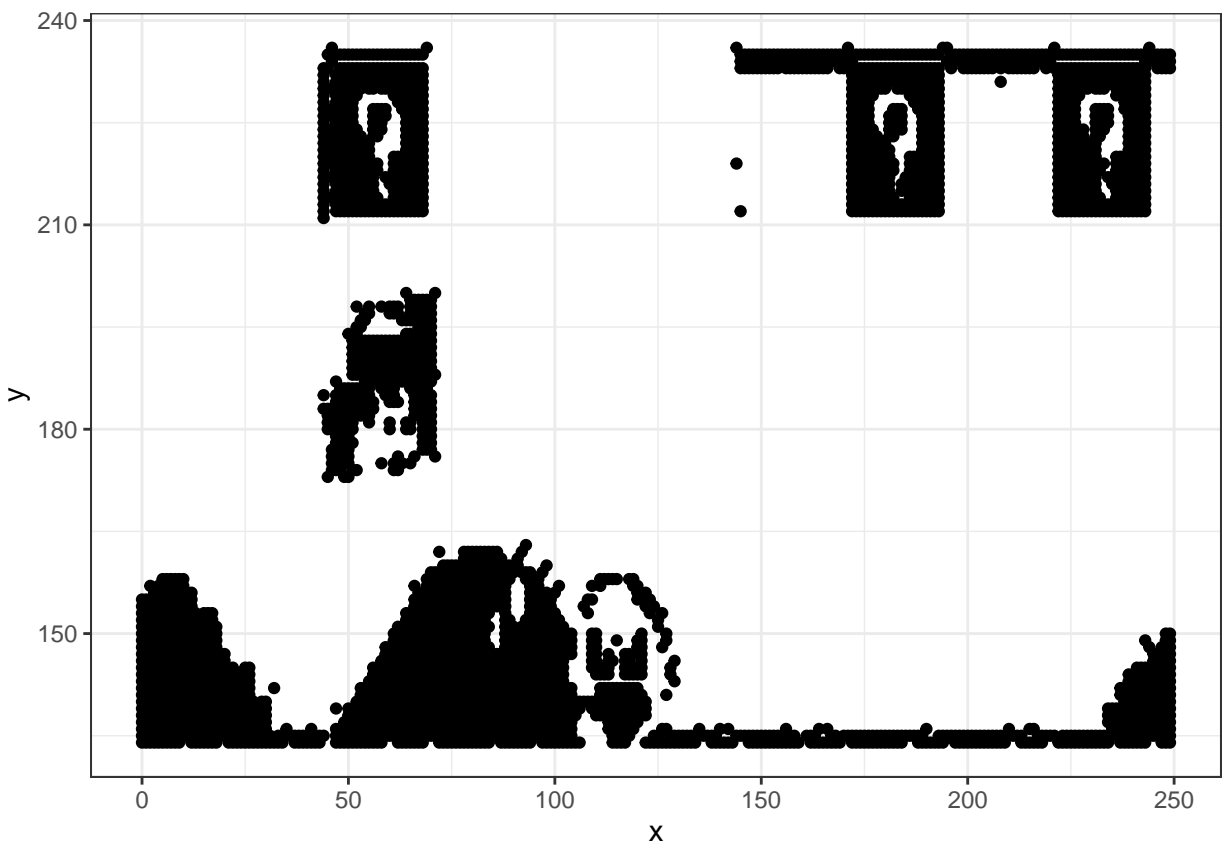
# Week11\_Clustering

Akila Selvaraj

2/27/2022

## Scatter plot of clustering data

```
library(ggplot2)
ggplot(clustering_data, aes(x = x, y = y)) + geom_point() +
  theme_bw()
```



## Scatter Plots for each value of k

```
library(purrr) # for map_dbl
# Compute clustering algorithm (e.g., k-means clustering) for different values of k. For instance, by v
# For each k, calculate the total within-cluster sum of square (wss).
tot_withinss <- map_dbl(1:20, function(k) {
```

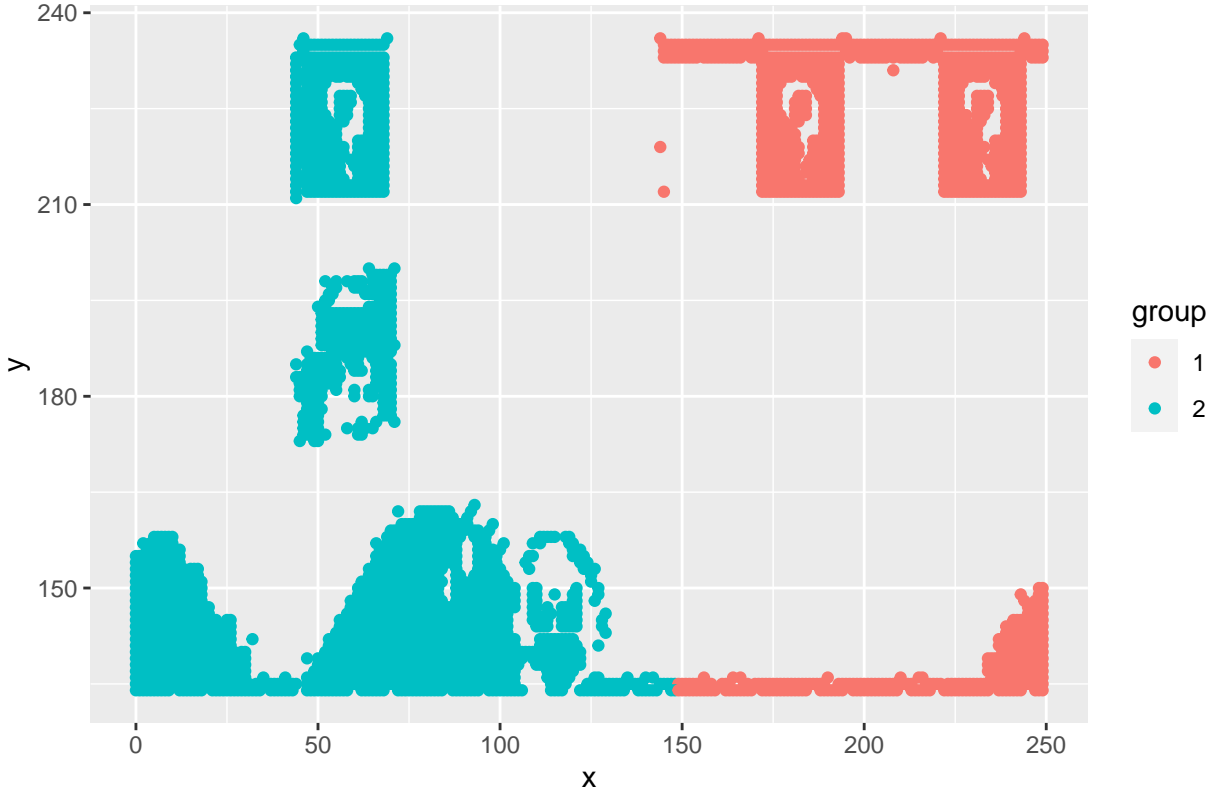
```

model <- kmeans(clustering_data, centers = k)
clustering_data$group <- as.factor(model$cluster)
Title <- paste("Scatter plot of clusters where K =",k)
cluster_plot <- clustering_data %>% ggplot(aes(x = x, y = y, n, colour = group)) + geom_point() +
print(cluster_plot)
model$tot.withinss
})

```

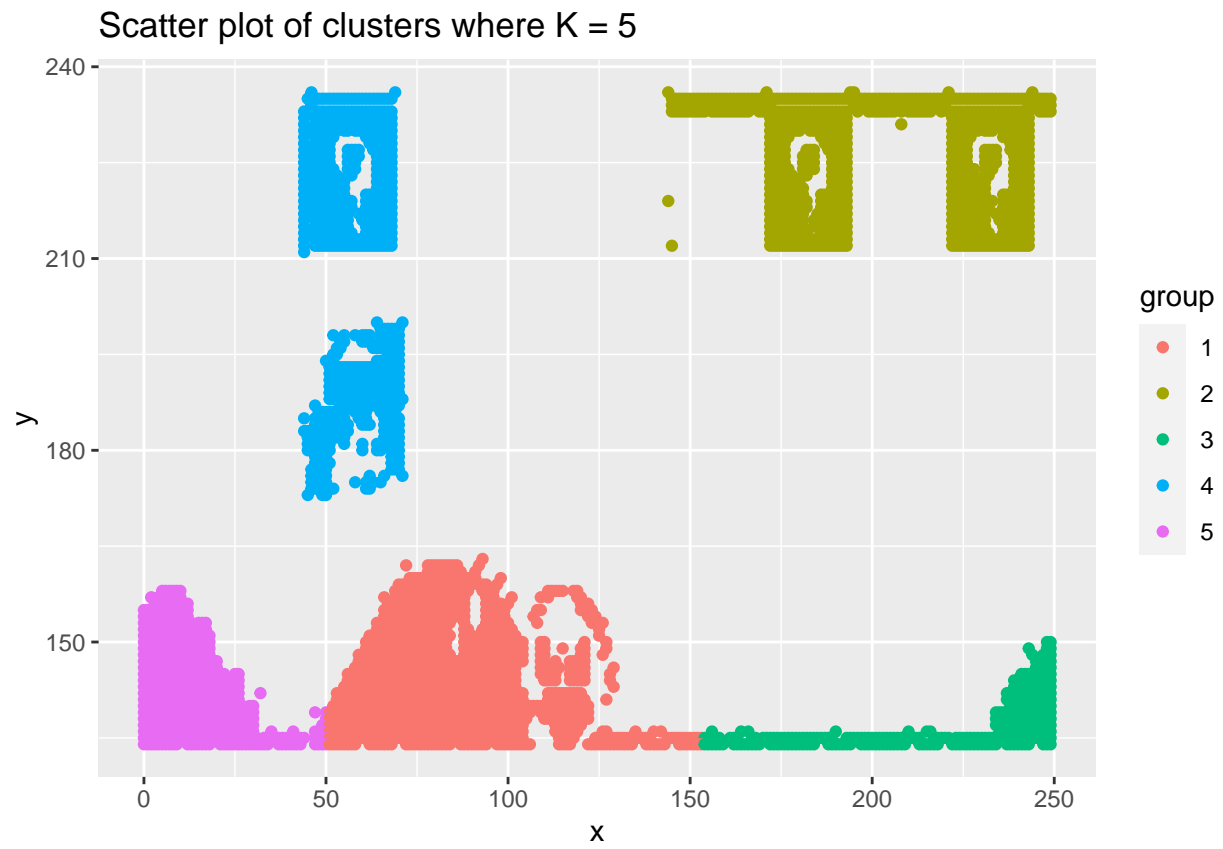


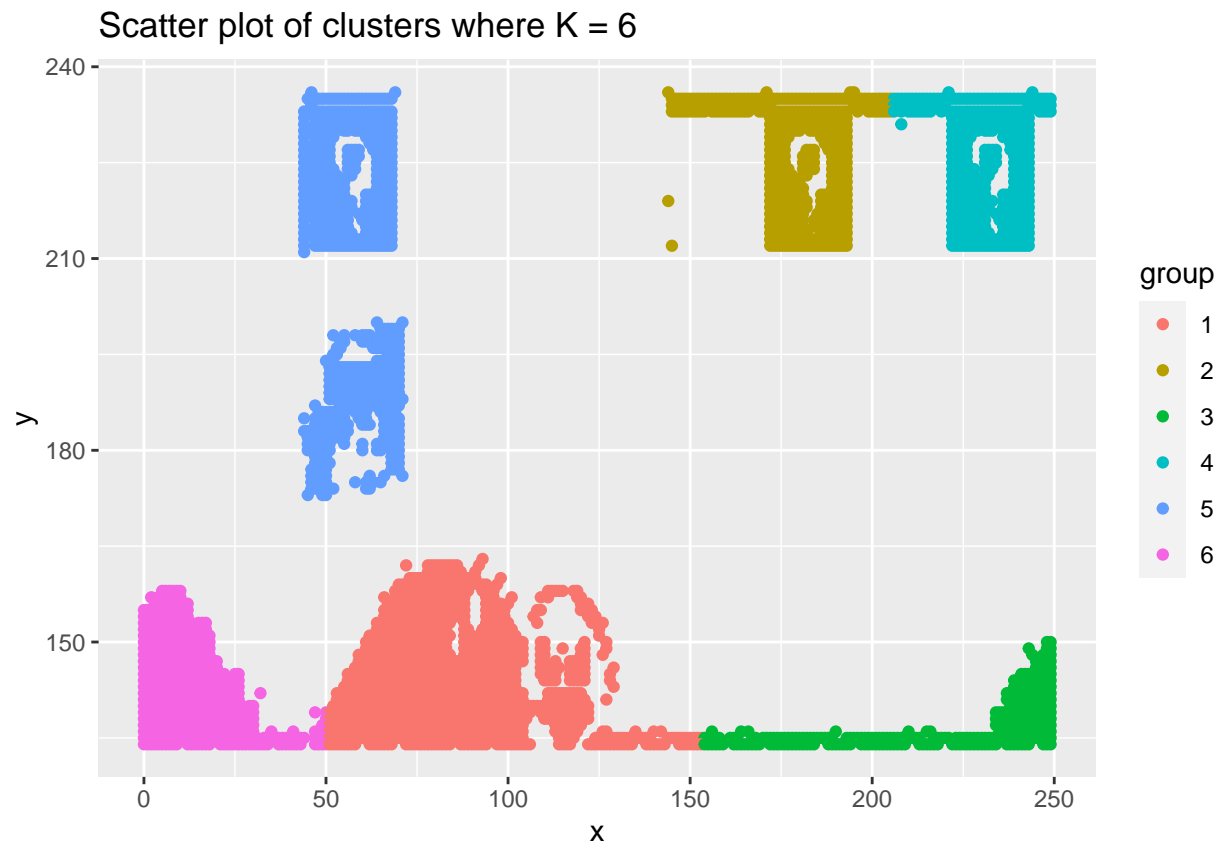
## Scatter plot of clusters where $K = 2$

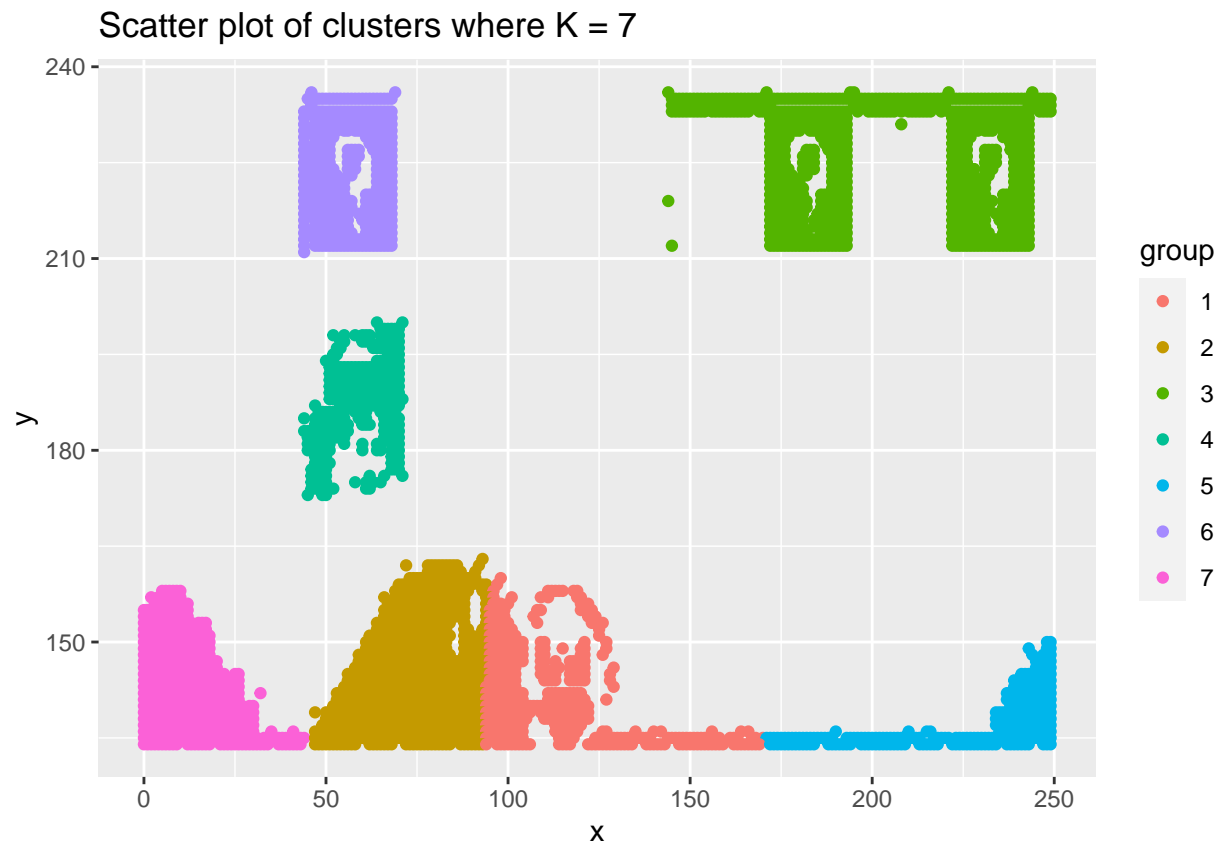




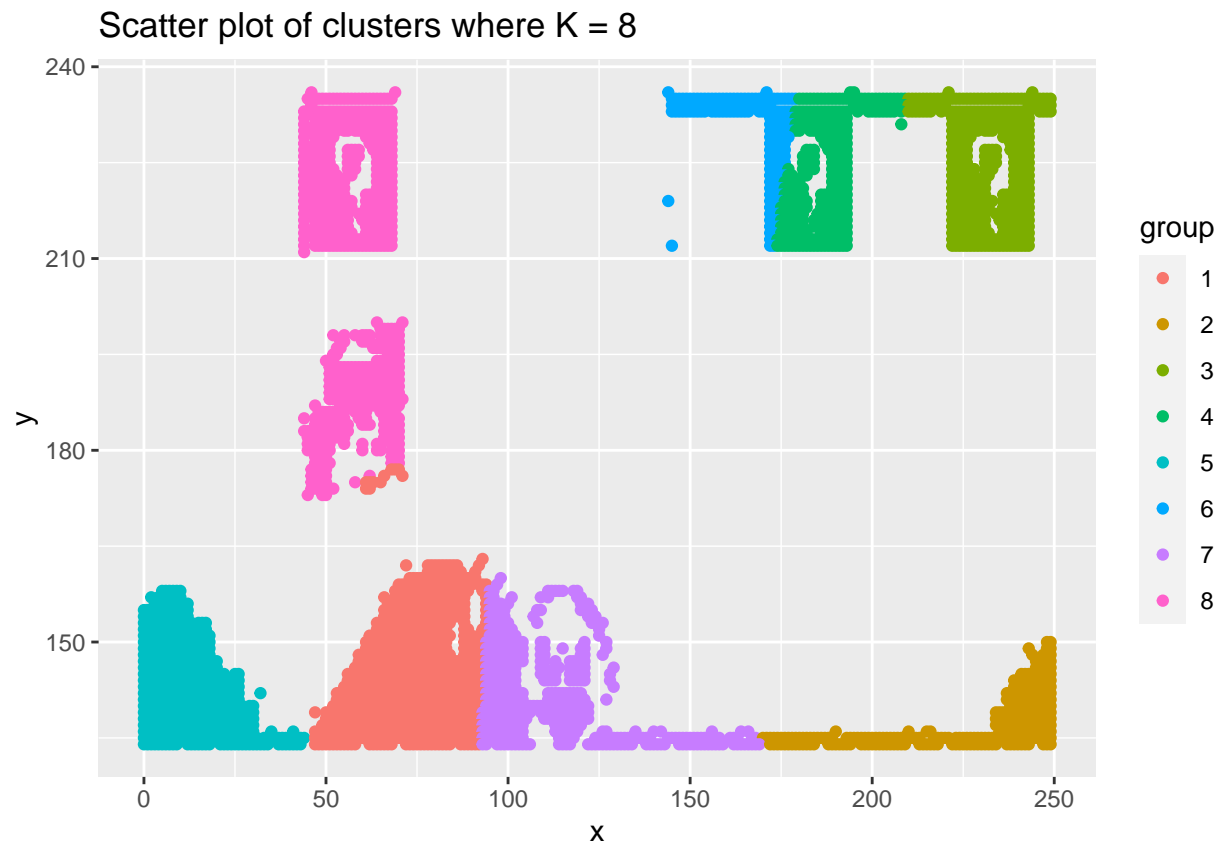


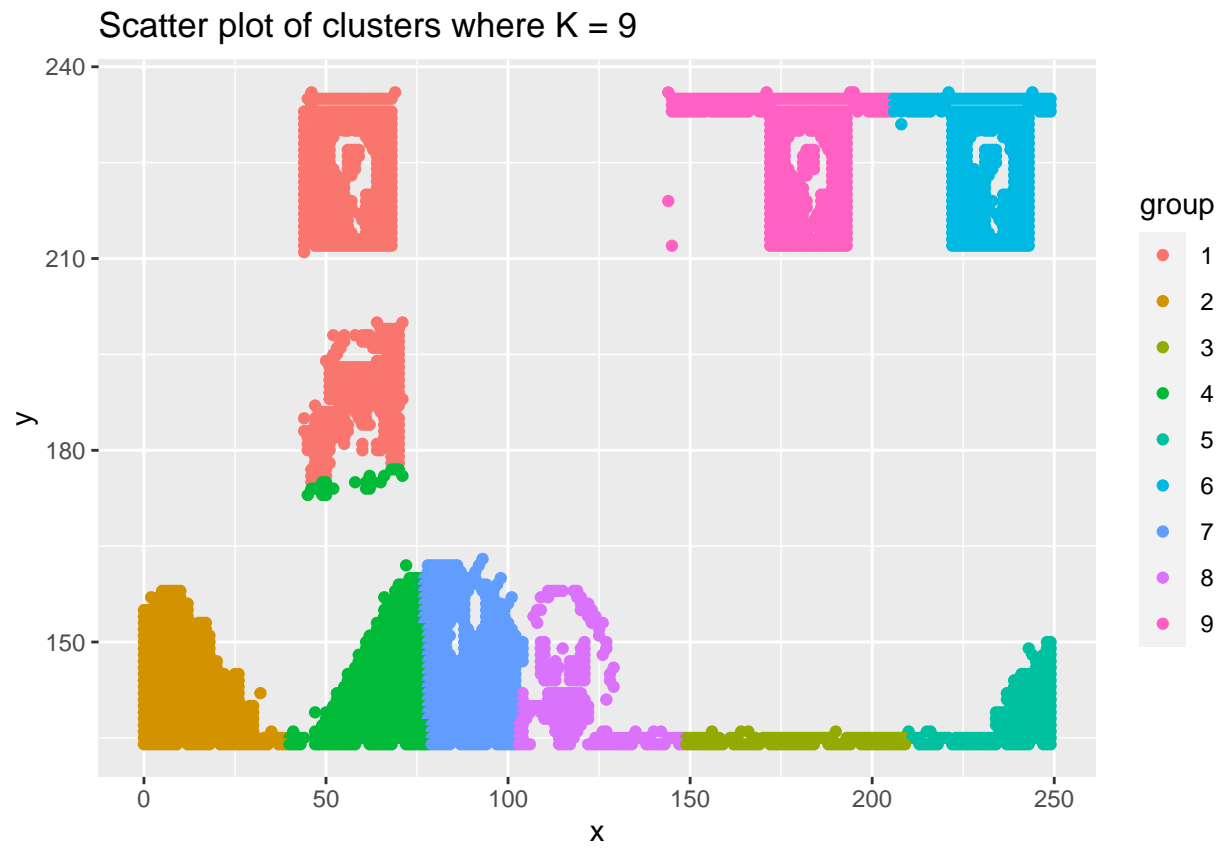


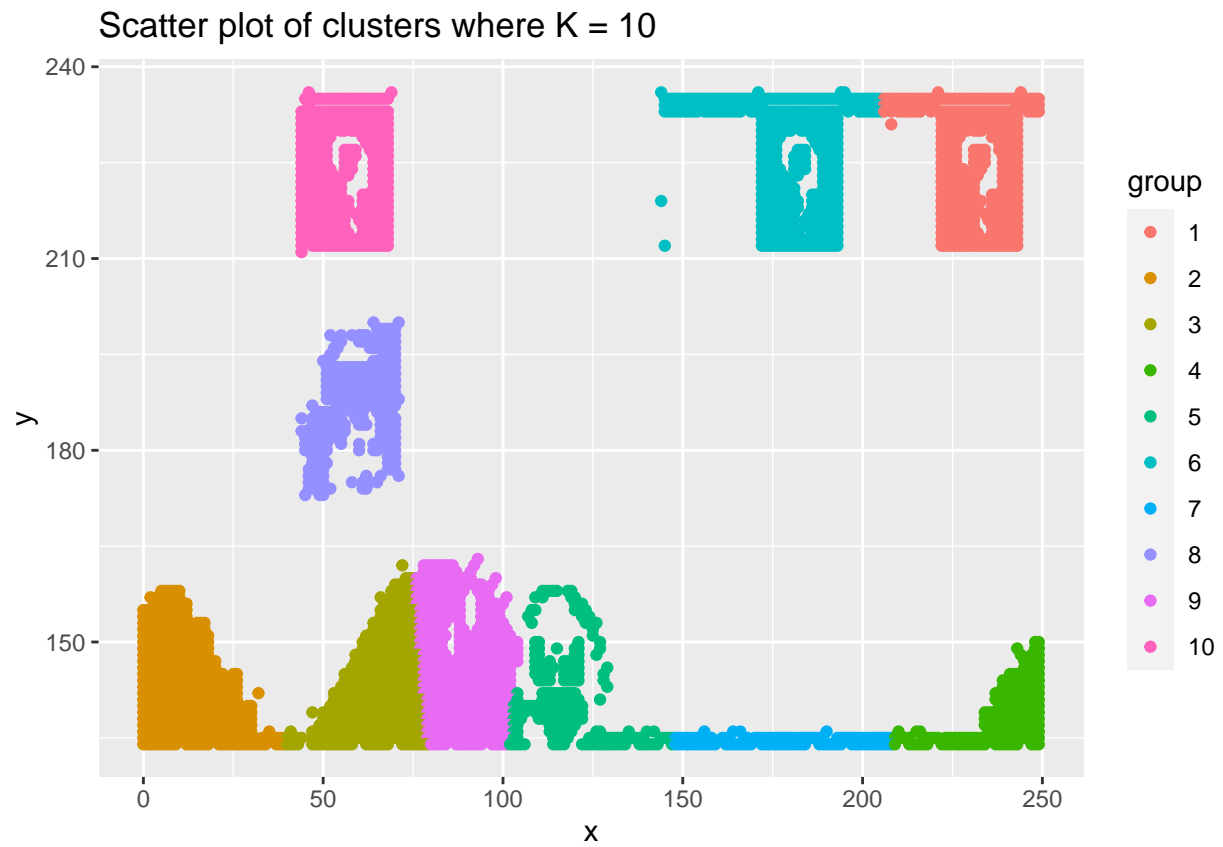


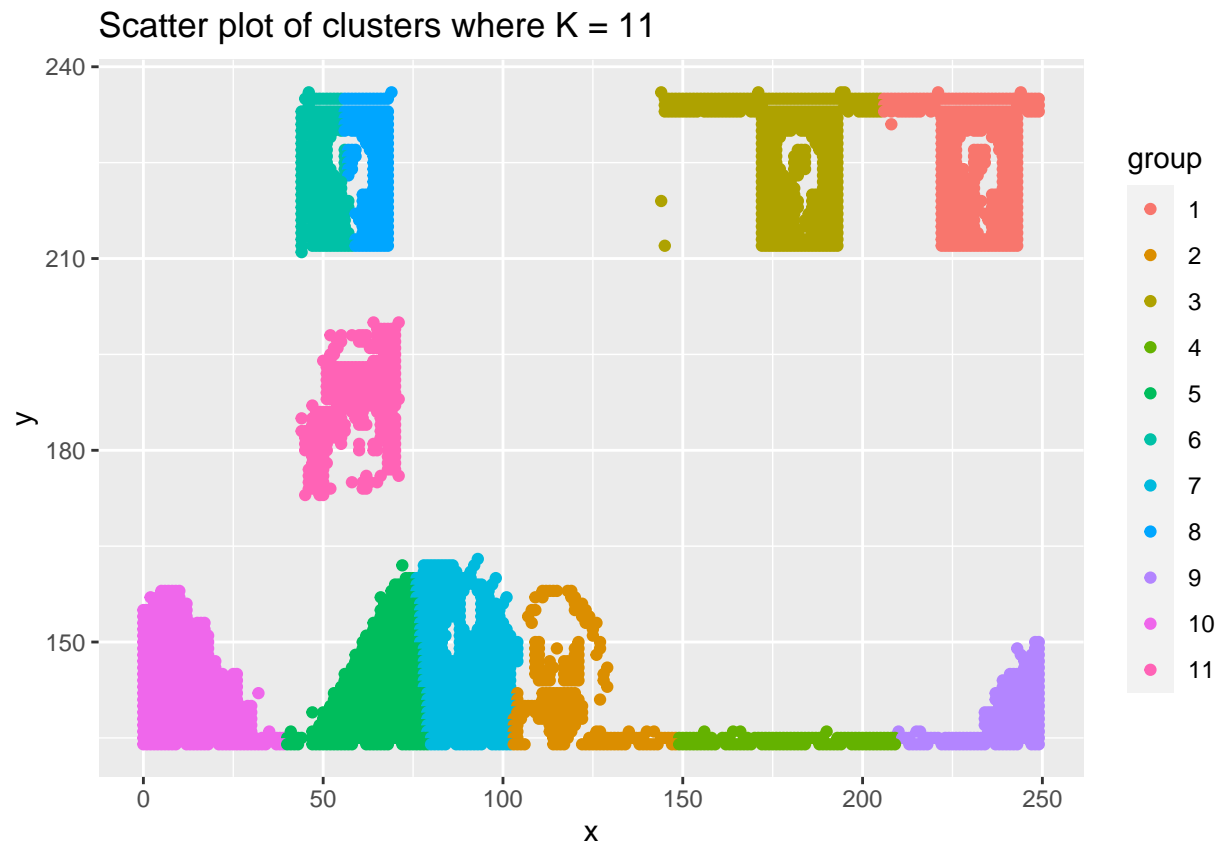


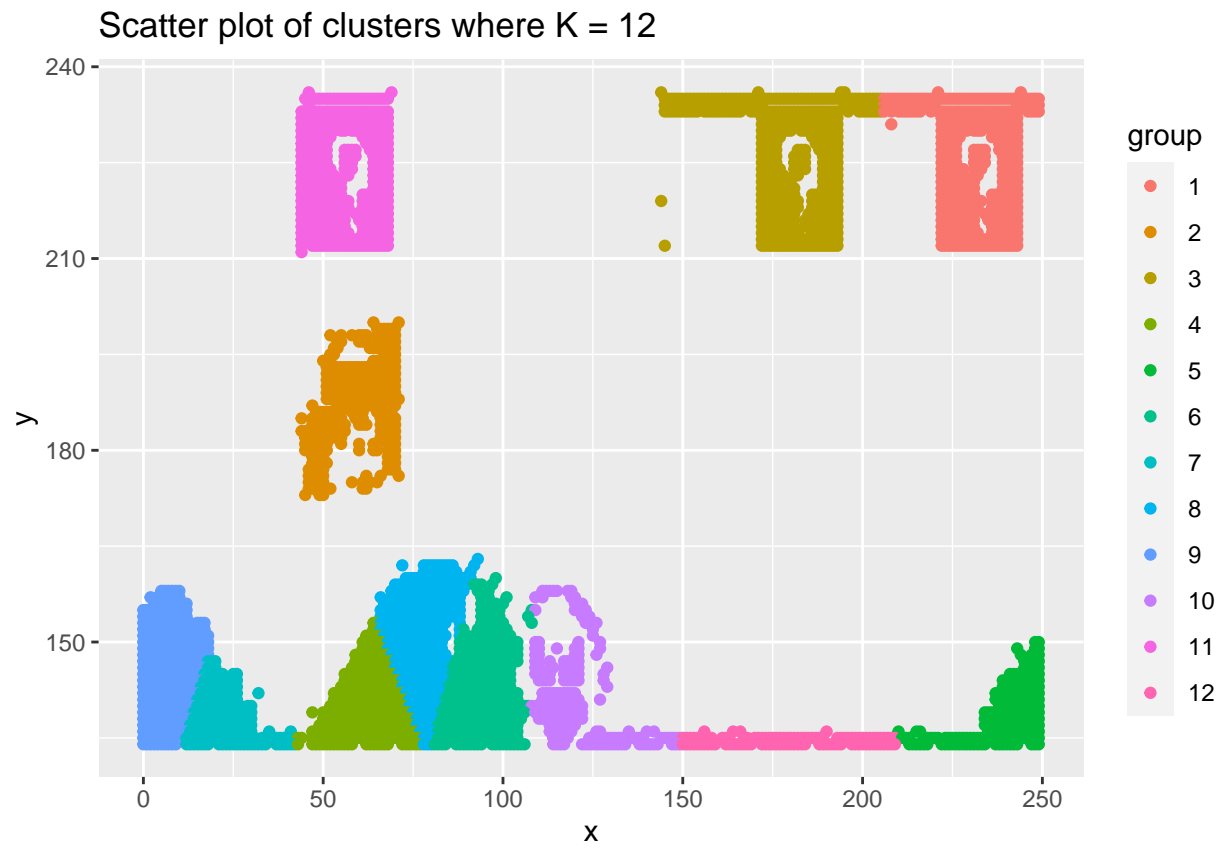


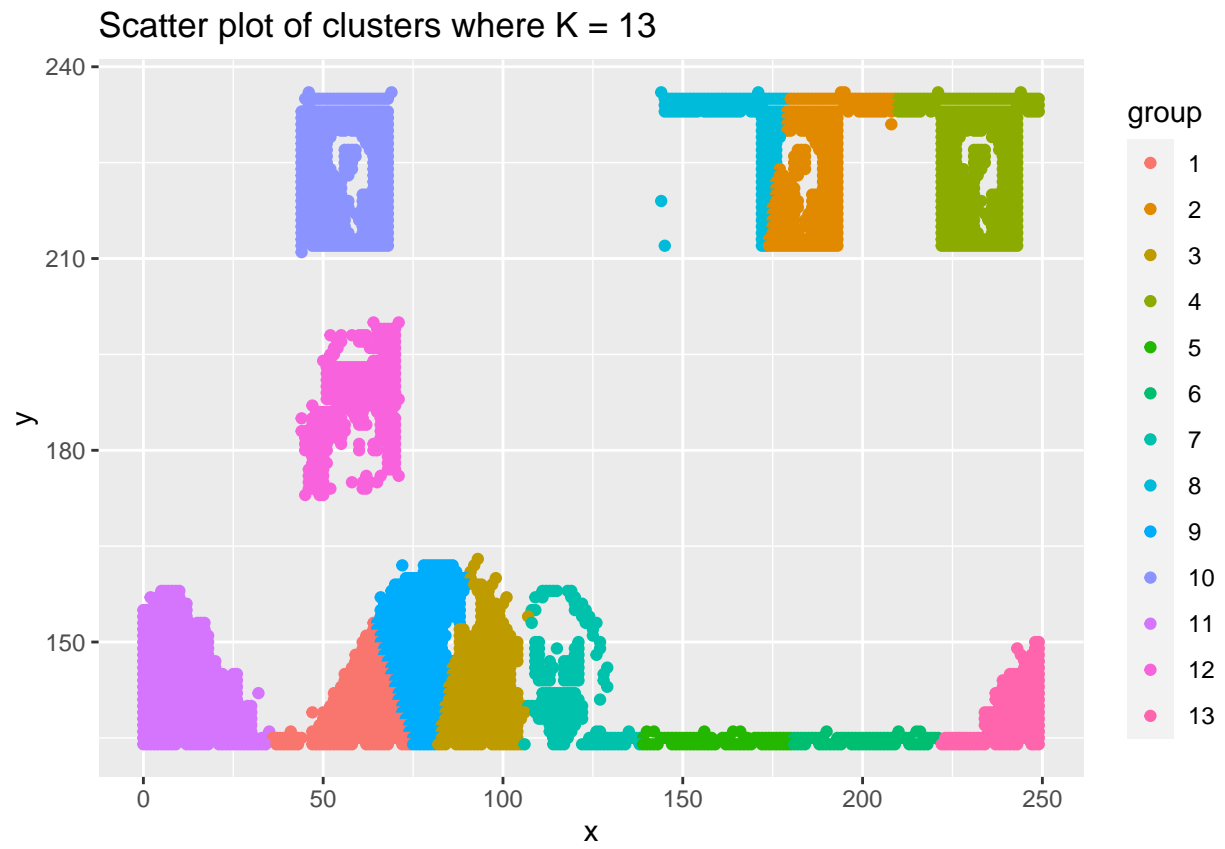


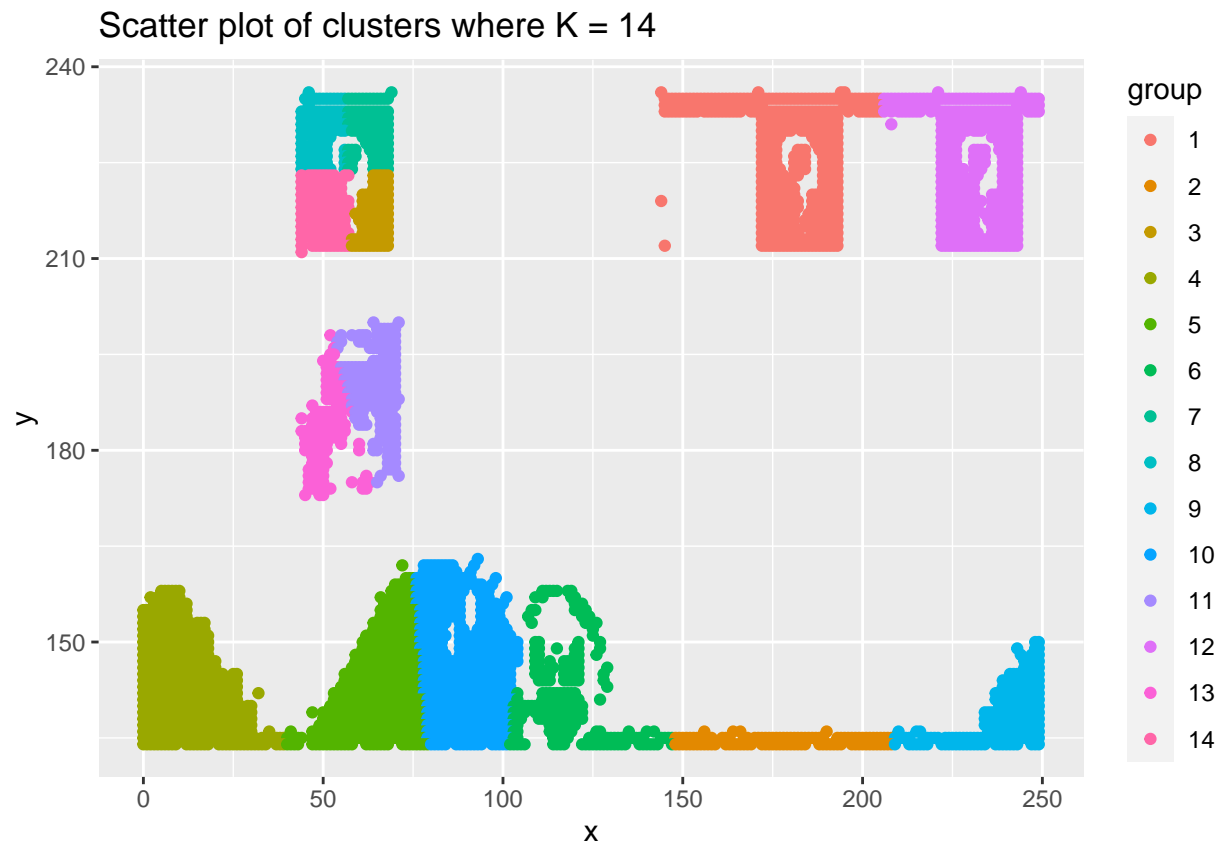


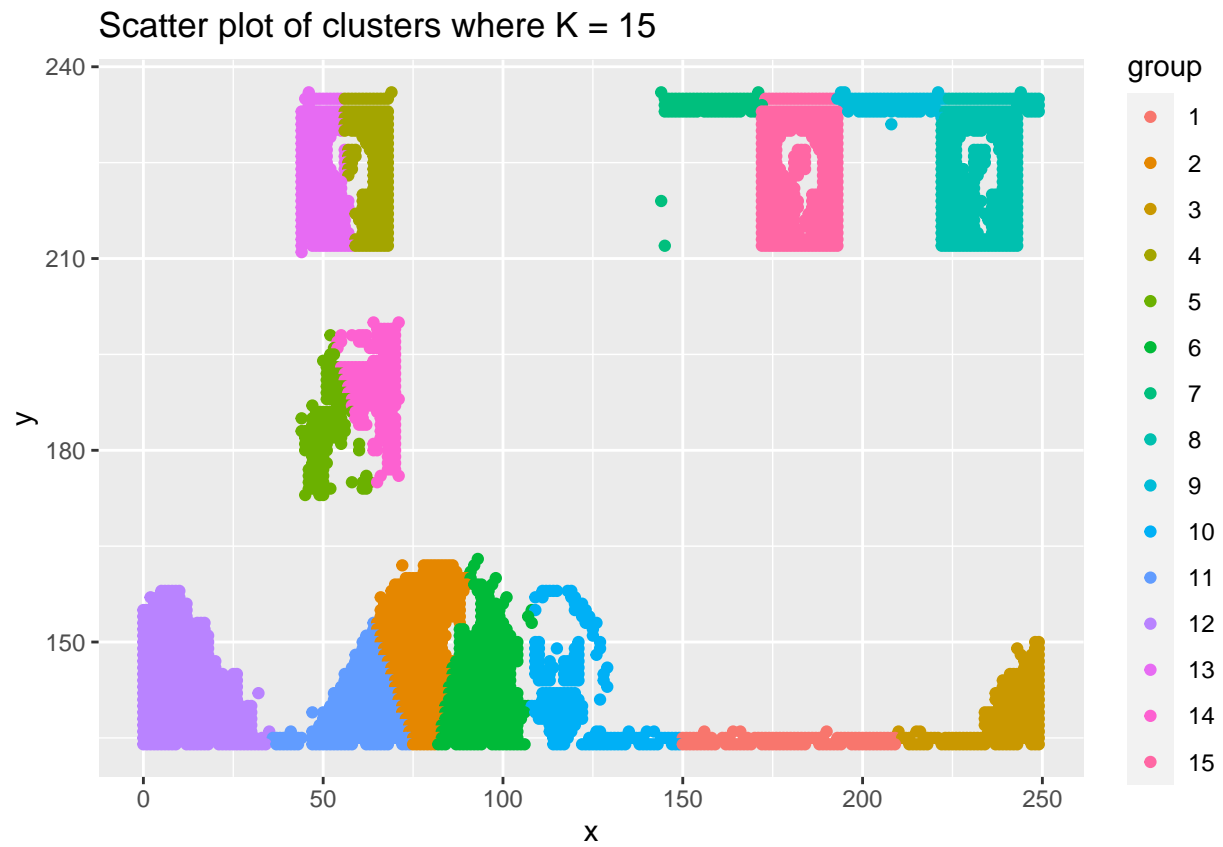




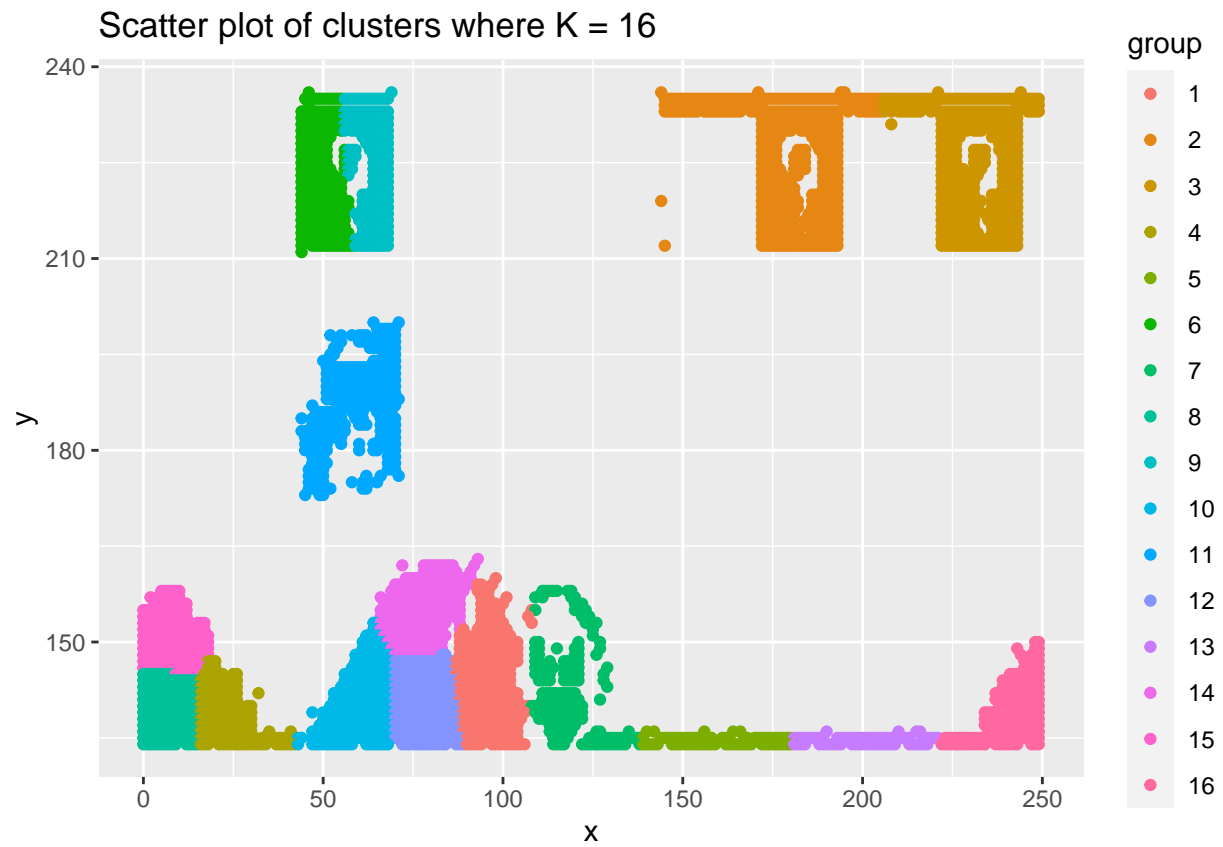


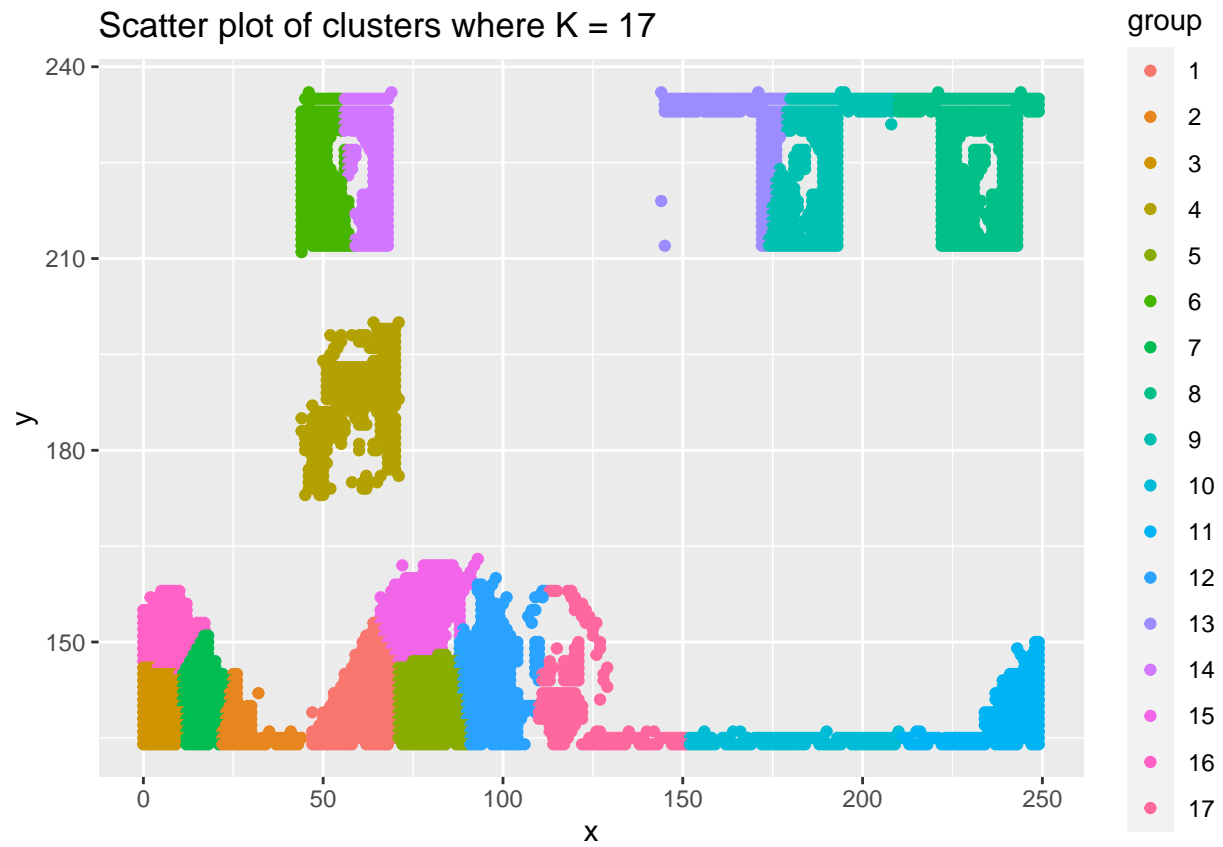


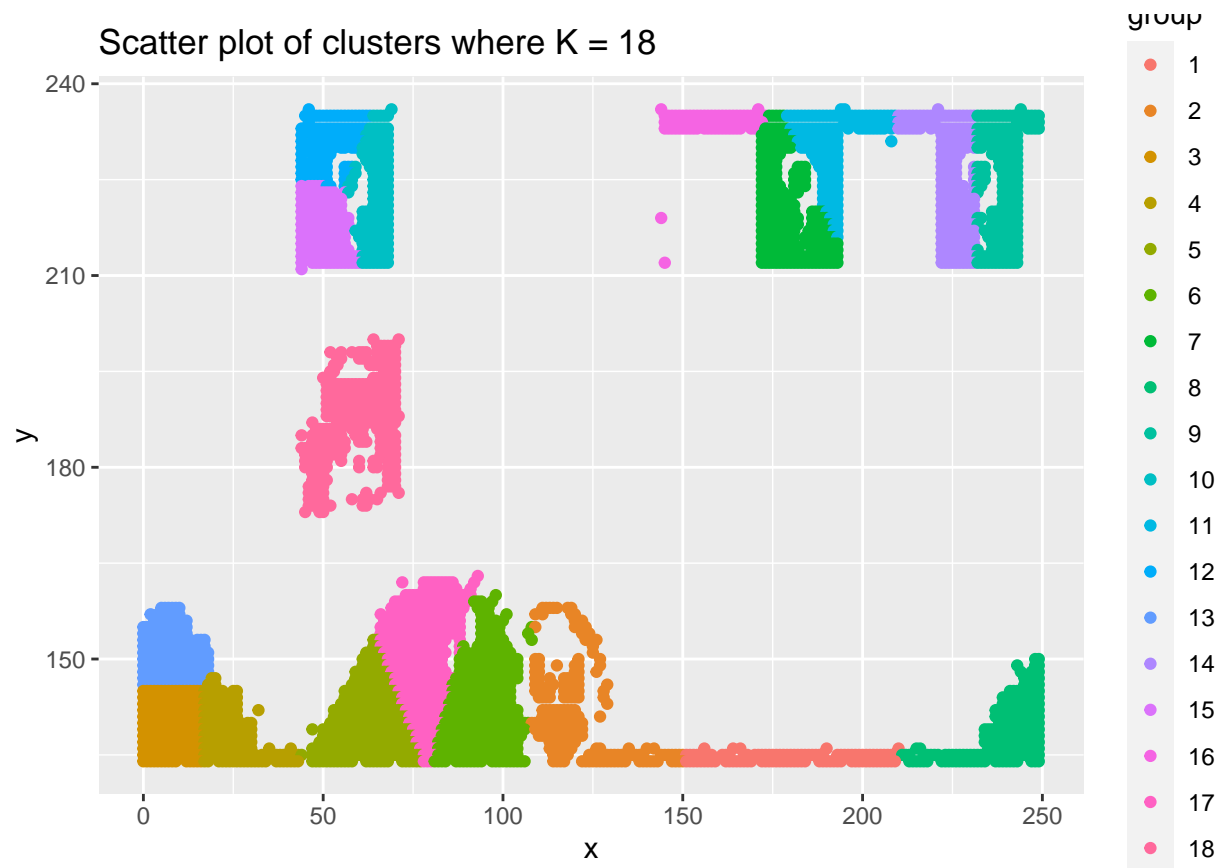


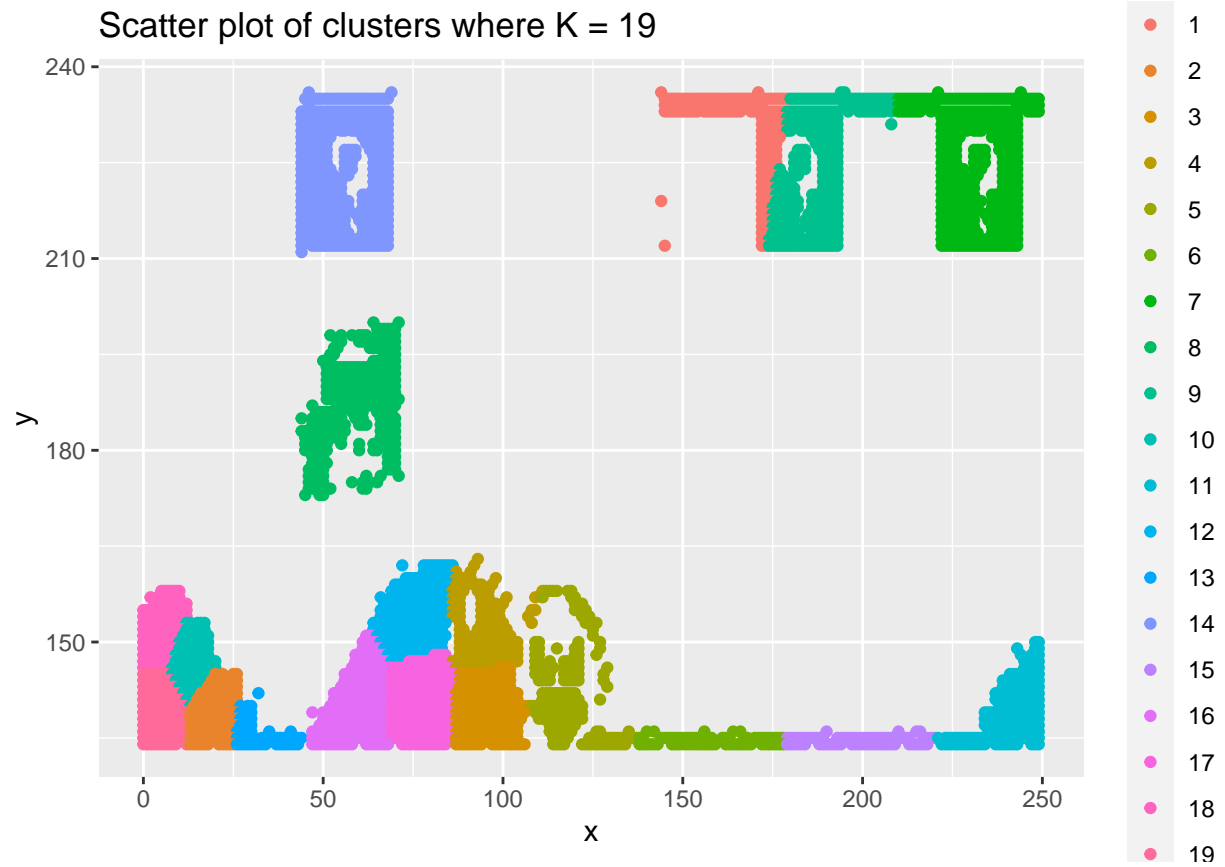


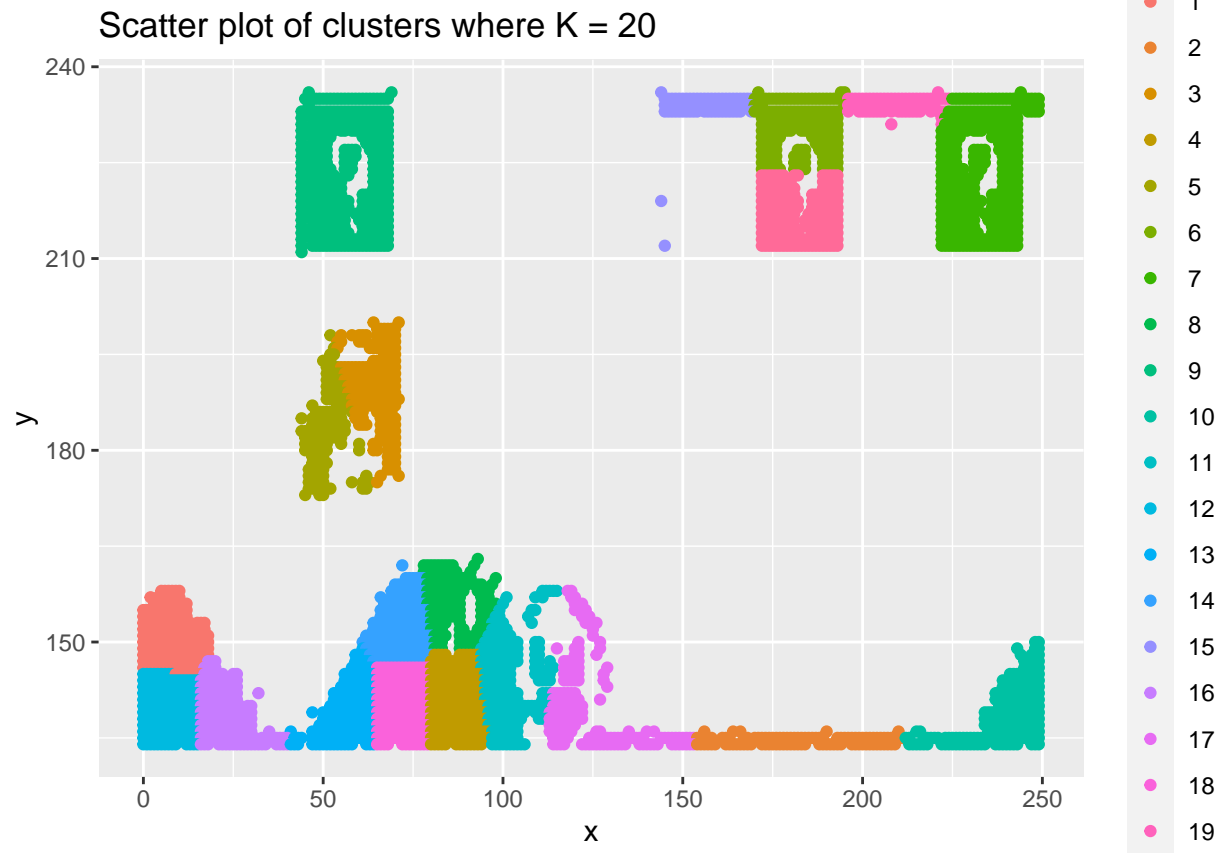




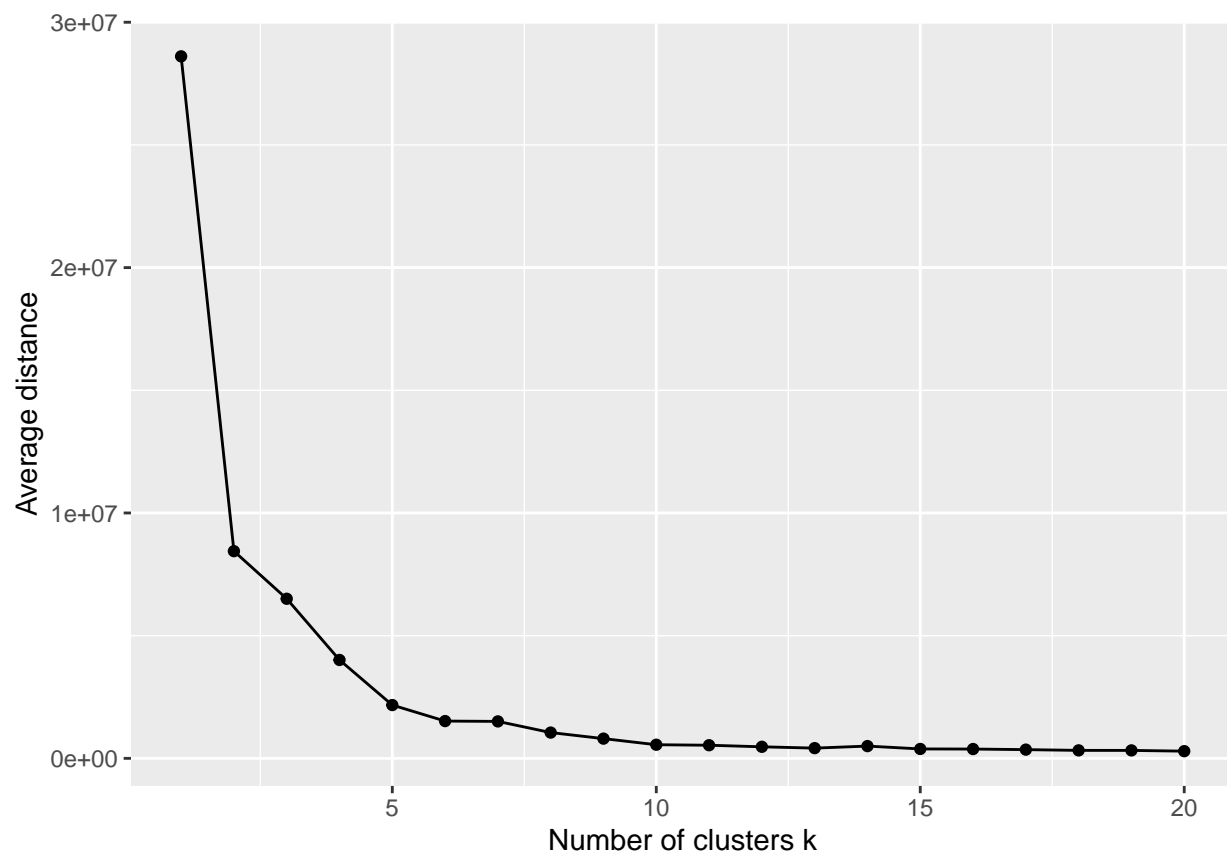








```
#Plot the curve of average distance vs number of clusters k.
data.frame(k = 1:20, tot_withinss = tot_withinss) %>%
  ggplot(aes(x = k, y = tot_withinss )) + geom_point() +
  geom_line() + labs(x = "Number of clusters k", y = "Average distance")
```



```
scale_x_continuous(breaks = 1:20)
```

```
## <ScaleContinuousPosition>
## Range:
## Limits: 0 -- 1
```

## Elbow point

The location of a bend (knee) in the plot is generally considered as an indicator of the appropriate number of clusters. Looking at the graph of number of clusters vs average, the elbow point for this dataset is 5.