lab 12

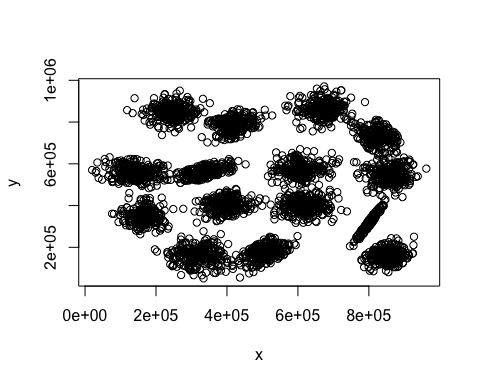
Aishat Olatunji

2022-11-14

data<-read.csv( "/Users/aishatolatunji/Downloads/PotentialClients.csv"  
)  
head(data)

## x y  
## 1 664159 550946  
## 2 665845 557965  
## 3 597173 575538  
## 4 618600 551446  
## 5 635690 608046  
## 6 588100 557588

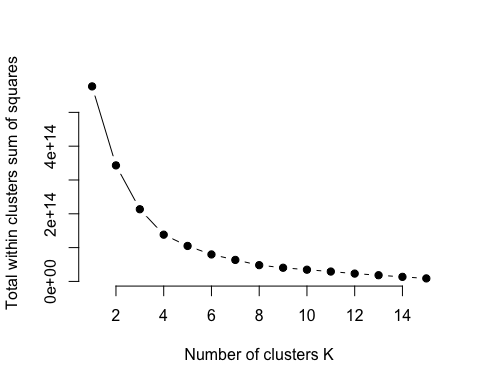
plot(data)



#lets compute and plot wss (weighted sum statistic) for k=2 to k=15  
k.max = 15  
  
wss = sapply(1:k.max, function(k) {kmeans(data, k, nstart = 50, iter.max = 15)$tot.withins})  
wss

## [1] 5.768070e+14 3.431836e+14 2.135087e+14 1.382507e+14 1.049354e+14  
## [6] 7.976902e+13 6.357671e+13 4.814692e+13 4.042723e+13 3.465658e+13  
## [11] 2.918277e+13 2.314662e+13 1.834987e+13 1.348673e+13 8.917616e+12

#lets plot the WSS  
plot(1:k.max, wss, type="b", pch=19, frame = FALSE , xlab = "Number of clusters K", ylab = "Total within clusters sum of squares")



# lets say the best number of clusters = 5  
clusters.km <- kmeans(data,5)  
  
#lets plot the color-coded clusters.km  
plot(data, col=clusters.km$cluster)

