

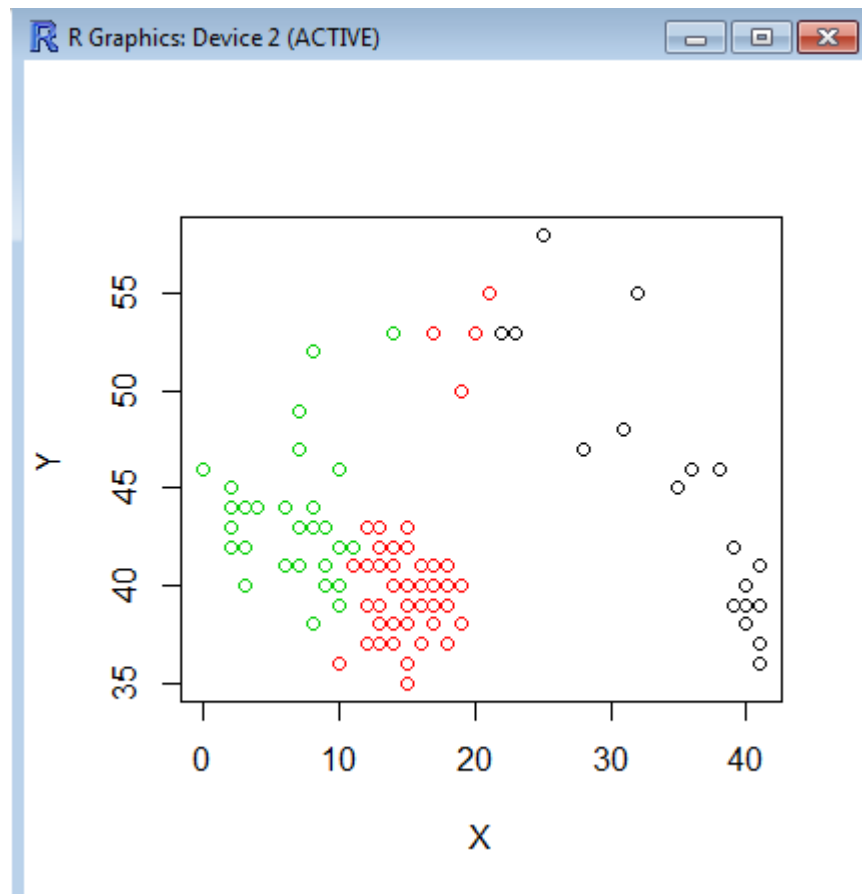
## REPORT LAB-3

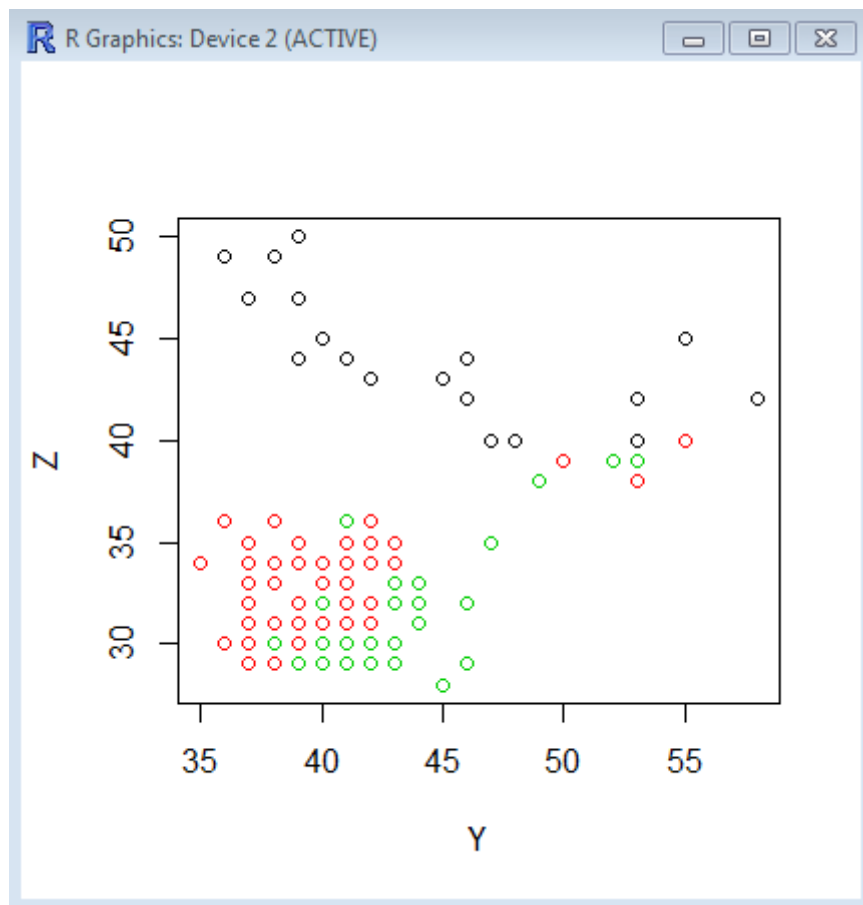
### QUESTION 1

#### Dataset:

The chosen data set is **Steps** data. I have performed clustering analysis on this data using the K-Means, K-Medians, Hierarchical, Expectation Maximization clustering.

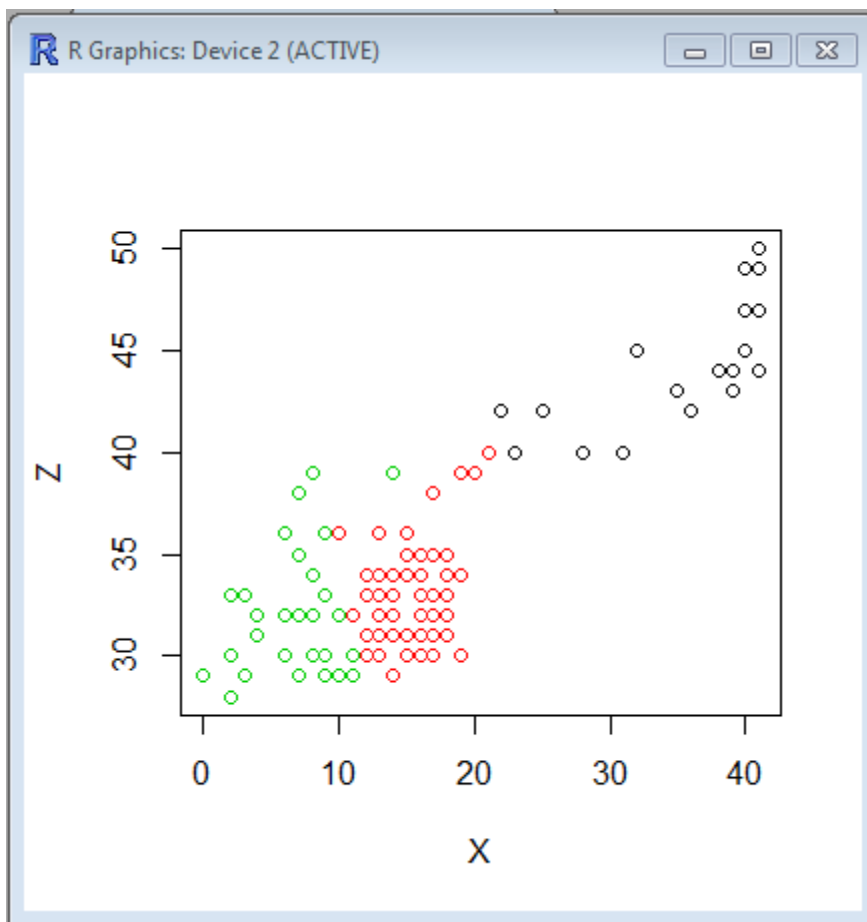
#### K Means:





Code:

```
Datain <- read.csv(file.choose(), sep = ",", header=FALSE)
YY <- Datain[,dim(Datain)[2]]
XX <- Datain[,1:(dim(Datain)[2]-1)]
pdata <- princomp(X, cor=T)
pc.data <- pdata$scores
pc.data1 <- -1*pc.data[,1]
pc.data2 <- -1*pc.data[,2]
X <- cbind(pc.data1, pc.data2)
cl <- kmeans(X,10)
cl$cluster
plot(pc.data1, pc.data2,col=cl$cluster)
points(cl$centers, pch=16)
```



## 2. K- Medians

K-Medians is calculated based on 1-norm distance. So for every pair of values we need to calculate the distance.

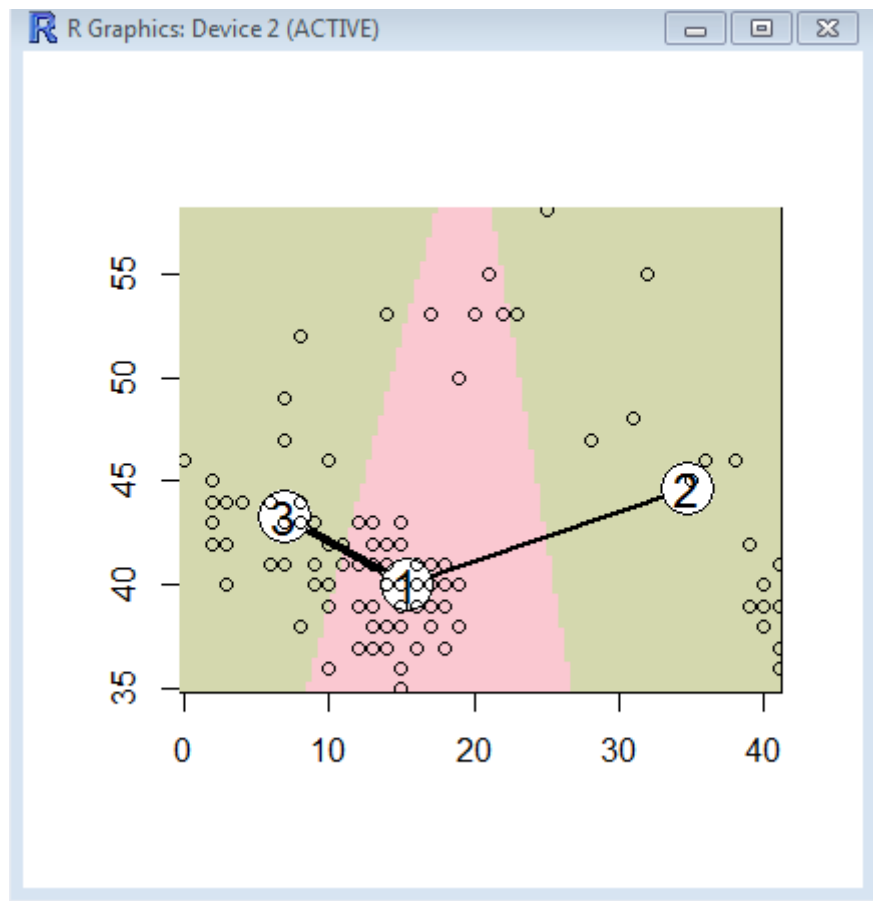
Algorithm:

```
Data1=kcca (myData, k=3)
```

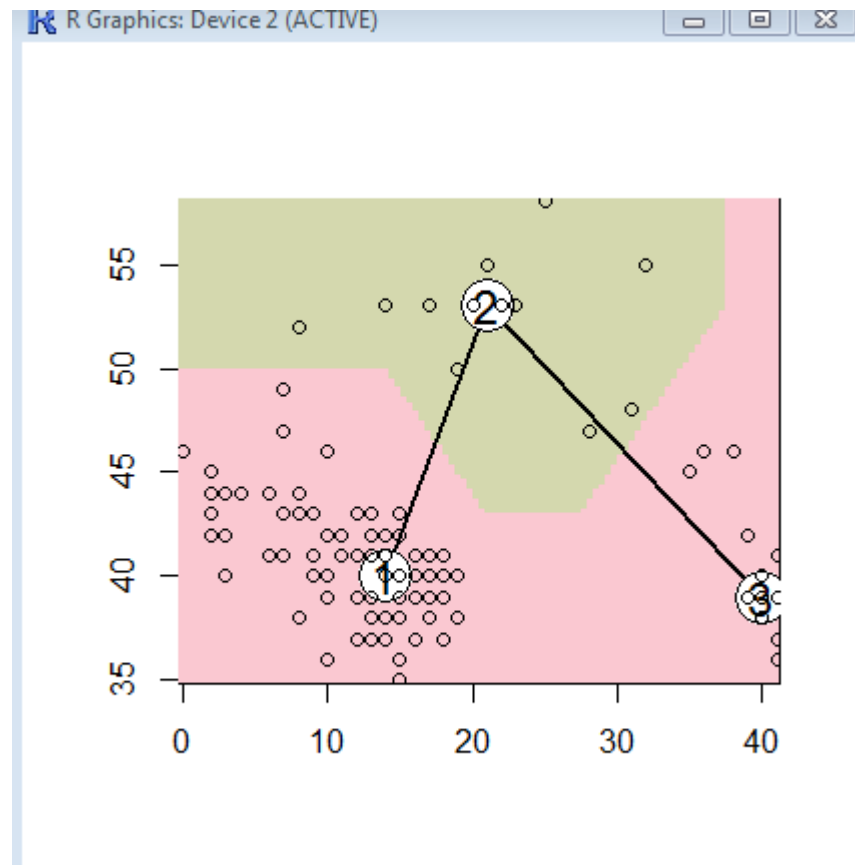
```
image (data1)
```

```
points (myData)
```

```
barplot (data1)
```



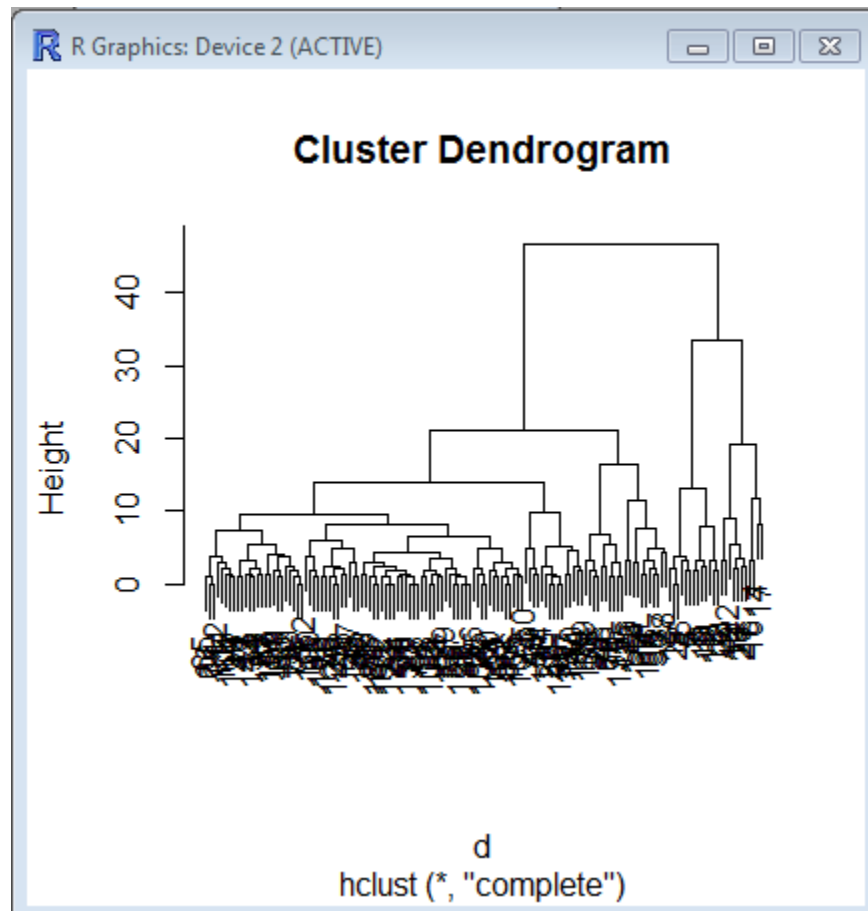
```
data2=kcca(myData,k=3,family=kccaFamily("kmedians"),control=list  
(initcent="kmeanspp"))  
  
image(data2)  
  
points(myData)
```



## Hierarchical Clustering:

### Algorithm:

```
distancevector<-dist(as.matrix(myData))  
hirachialc<-hclust(distancevector)  
plot(hirachialc)
```



#### 4. Expectation Maximization:

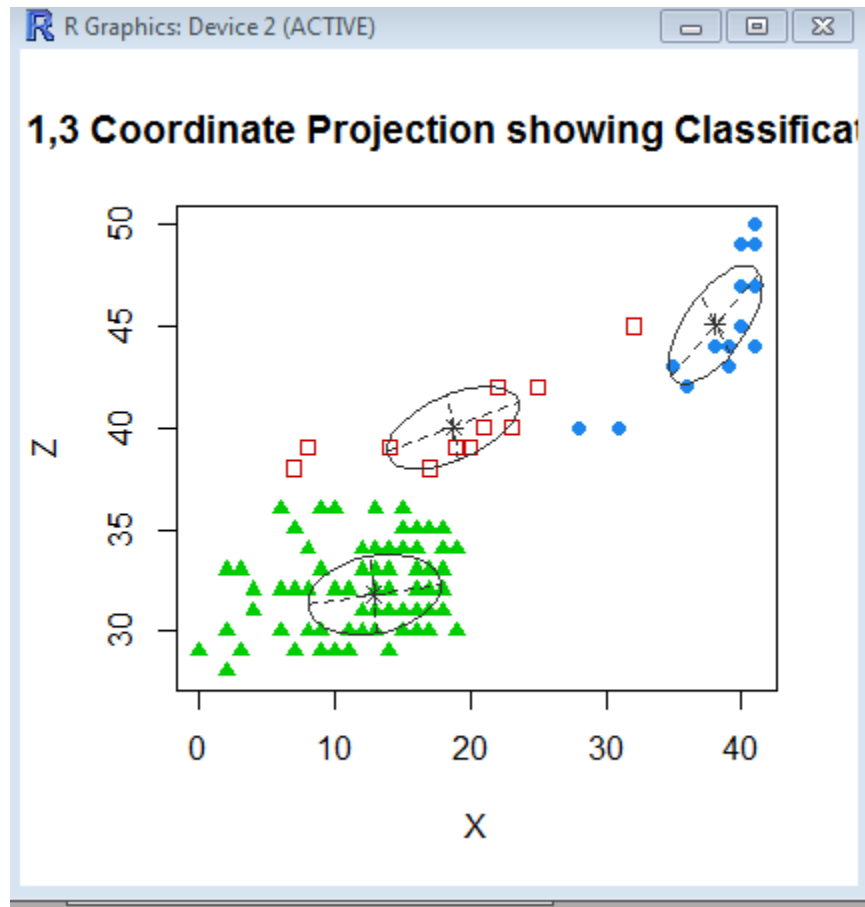
```
expectclust<-Mclust(myData[,1:3],G=3)
```

```
plot(expectclust,what=c("classification"),dims=c(1,2,3))
```

```
plot(expectclust,what=c("classification"),dims=c(1,2))
```

```
plot(expectclust,what=c("classification"),dims=c(2,3))
```

```
plot(expectclust,what=c("classification"),dims=c(1,3))
```



## Question 2:

Collected steps data from android application and stored as CSV files in the Android Internal Storage.

