# ggplot

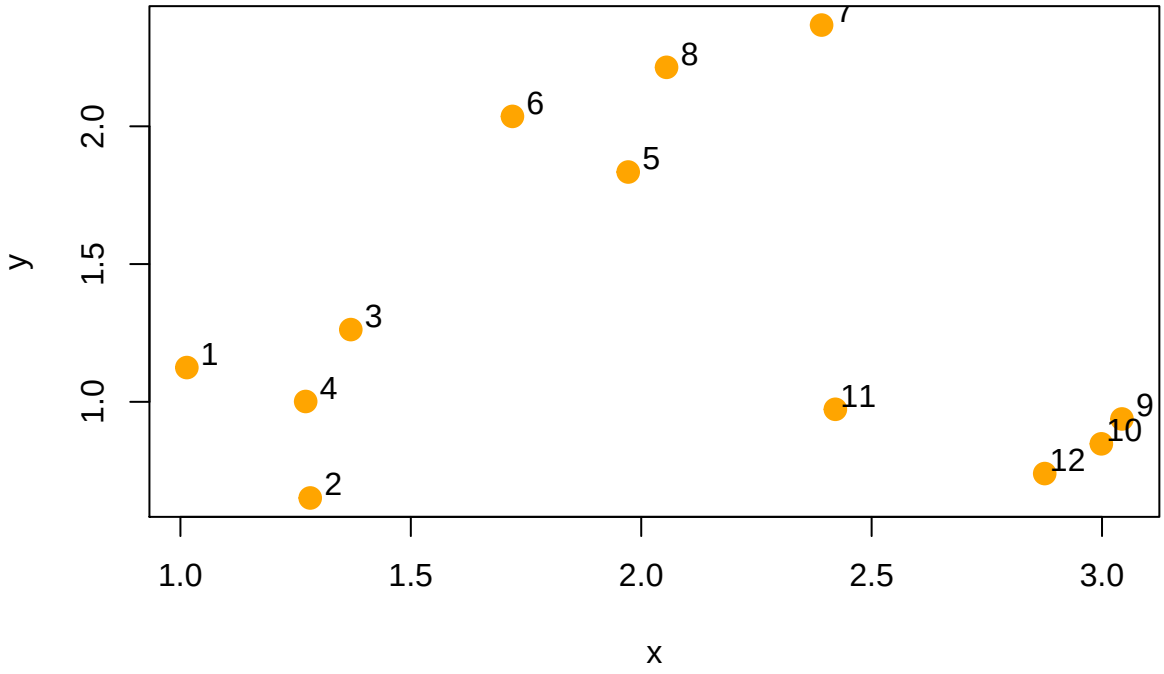
Anandu R

6/15/2020

## K-Means Clustering

Illustrating clusters

```
set.seed(96)
x = rnorm(12, mean = rep(1:3, each = 4), sd = 0.2)
y = rnorm(12, mean = rep(c(1,2,1),each = 4), sd = 0.2)
plot(x, y, col = "orange", pch = 19,cex = 1.5)
text(x+0.05, y+0.05, labels = as.character(1:12))
```



### Using the kmeans()

This function is used to find various clusters in the dataset, it returns an object with various elements the element "cluster" gives a list of integers denoting to which cluster each record in the datset belongs to

```
data = data.frame(x,y)
kmeansobj = kmeans(data, centers = 3)
names(kmeansobj)
```

```
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss" "betweenss"    "size"
```

```
kmeansobj$cluster
```

```
##  [1] 3 3 3 3 2 2 2 2 1 1 1 1
```

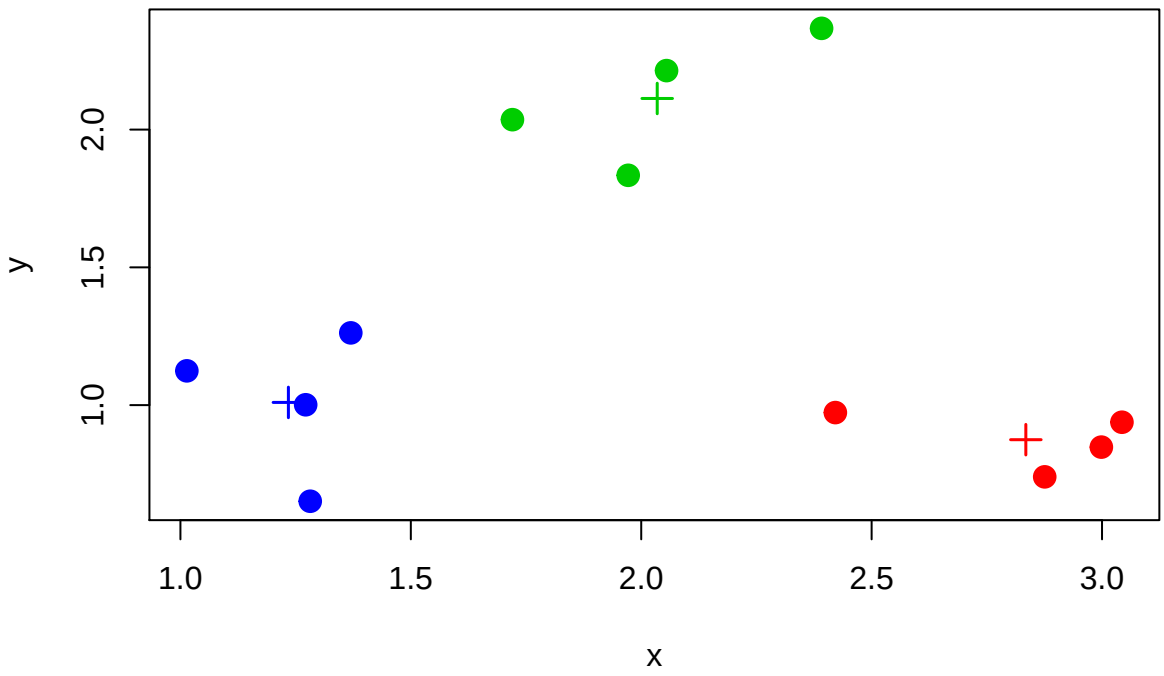
The coordinates of centroids of each cluster is stored in the "centers" element

```
kmeansobj$centers
```

```
##          x         y
## 1 2.834689 0.8744583
## 2 2.034709 2.1129590
## 3 1.234319 1.0097312
```

### Plotting the data using the identified clusters

```
plot(x, y, col = (kmeansobj$cluster+1), pch = 19, cex = 1.5)
points(kmeansobj$centers, pch = 3, col = 2:4, cex = 1.5, lwd = 1.5)
```


y
2.0
1.5
1.0
1.0
1.5
2.0
2.5
3.0
x