

ggplot

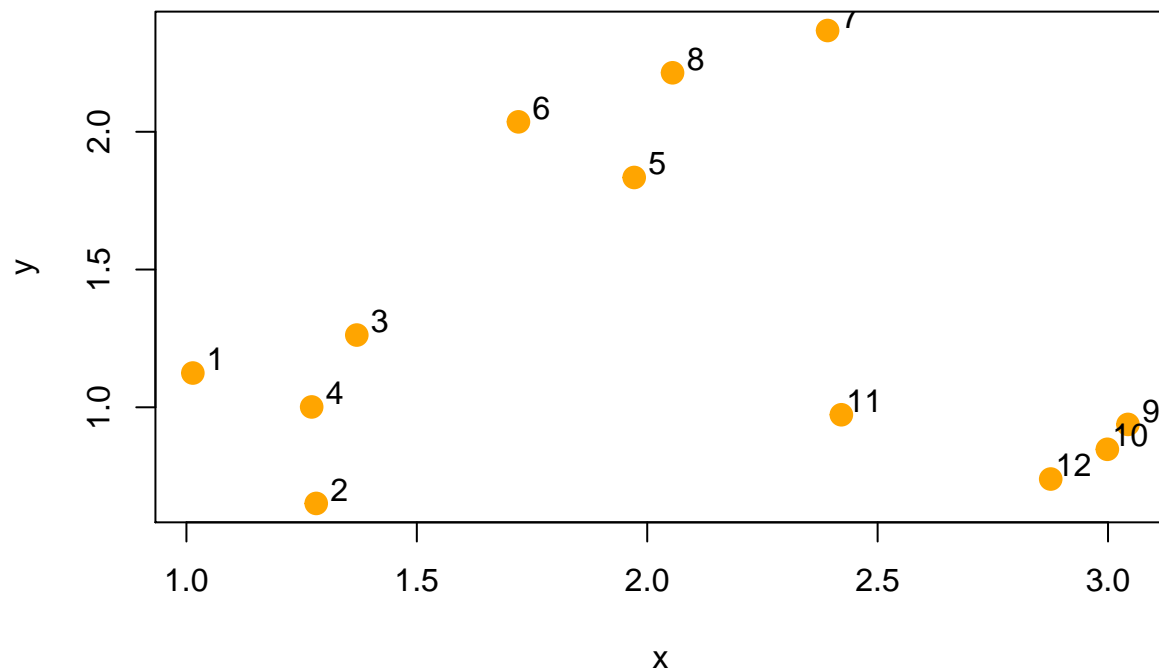
Anandu R

6/15/2020

Heirarchical 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))
```



Step 1

Find the pairwise distance between each points using the `dist()` function

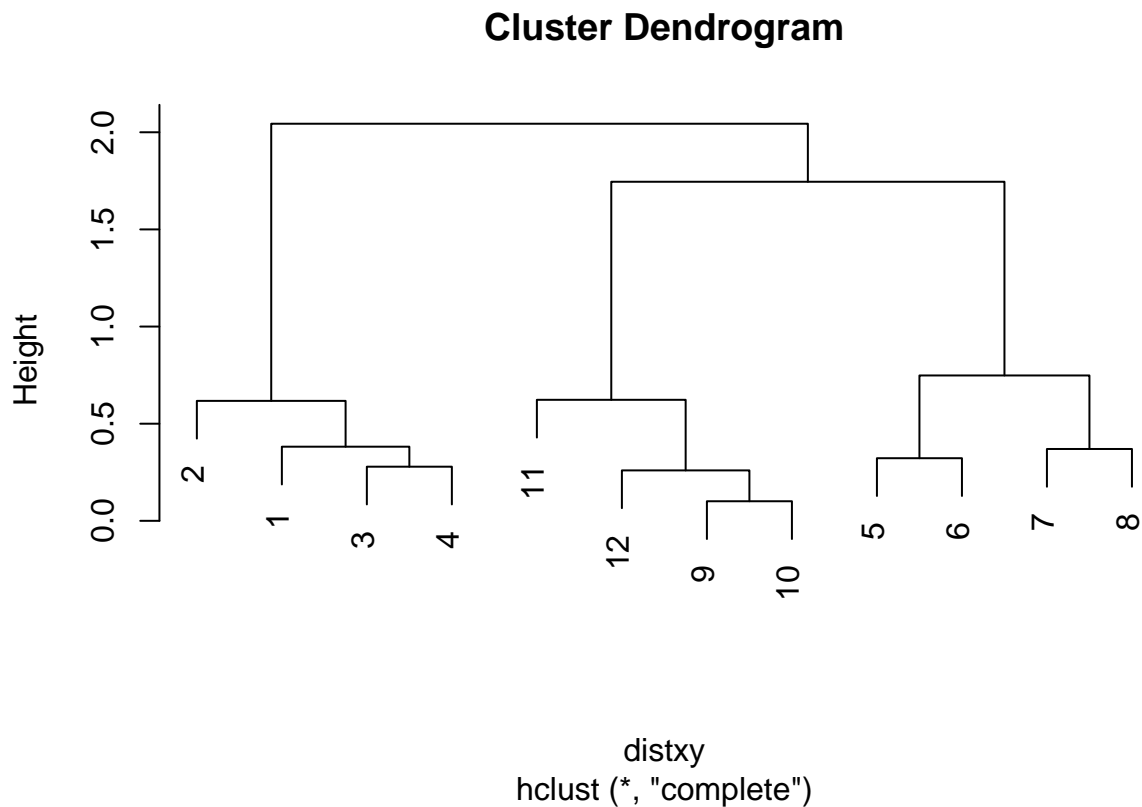
```
data = data.frame(x,y)
distxy = dist(data)
distxy
```

```
##           1           2           3           4           5           6           7           8           9
## 2  0.5437960
## 3  0.3815662 0.6174629
## 4  0.2857667 0.3503437 0.2787112
## 5  1.1921330 1.3695806 0.8303631 1.0878976
## 6  1.1534655 1.4528453 0.8496384 1.1278820 0.3223282
## 7  1.8555210 2.0439466 1.5051844 1.7663981 0.6787189 0.7483020
## 8  1.5071417 1.7439180 1.1729397 1.4437660 0.3890738 0.3789129 0.3697452
## 9  2.0378454 1.7847065 1.7046487 1.7725298 1.3968812 1.7191597 1.5712478 1.6141887
## 10 2.0038226 1.7280288 1.6807817 1.7335403 1.4239434 1.7452463 1.6367795 1.6607389 0.1007910
## 11 1.4154969 1.1841454 1.0906637 1.1498184 0.9715091 1.2733600 1.3950015 1.2942430 0.6229117 0.59069
## 12 1.9010201 1.5963685 1.5939898 1.6249871 1.4194203 1.7363256 1.6983574 1.6874935 0.2595640 0.16352
```

Step 2

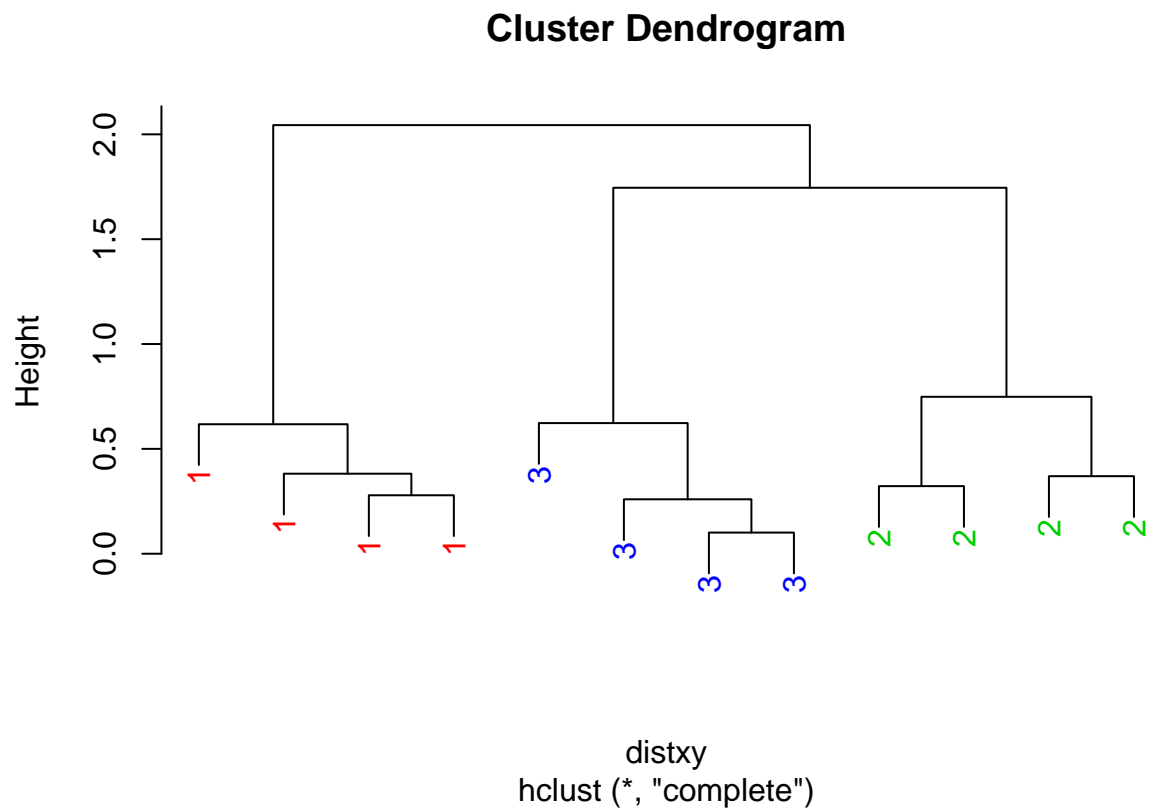
Heirarchical Clustering using hclust()

```
hClustering = hclust(distxy)
plot(hClustering)
```



Using a custom function to plot the clusters

```
myplclust(hClustering, lab = rep(1:3, each = 4), lab.col = rep(2:4, each = 4))
```



Using the heatmap function to derive correlation between points in data

Inorder to be able to organise the data with some logical way

```
dataAsMatrix = as.matrix(data)[sample(1:12),]  
heatmap(dataAsMatrix)
```

