

Clustering

Julia Lee

5/19/2022

```
clust <- read_csv("clust.csv")

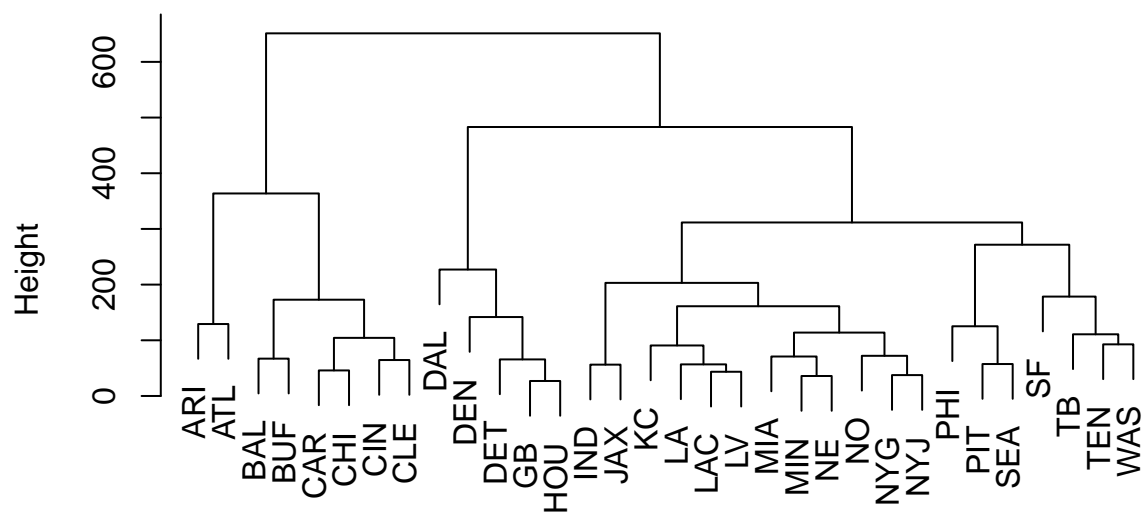
## Rows: 32 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr  (1): Team
## dbl (12): MPasstoRush, MPasses, MTD, MScore, Rushes, Passes, Wins, favorby, ...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Hclust

```
dist.map <- dist(clust)

## Warning in dist(clust): NAs introduced by coercion
hclustavg <- hclust(dist.map,method= "complete")
labels(hclustavg) <- clust$Team
p1<-plot(hclustavg)
```

Cluster Dendrogram



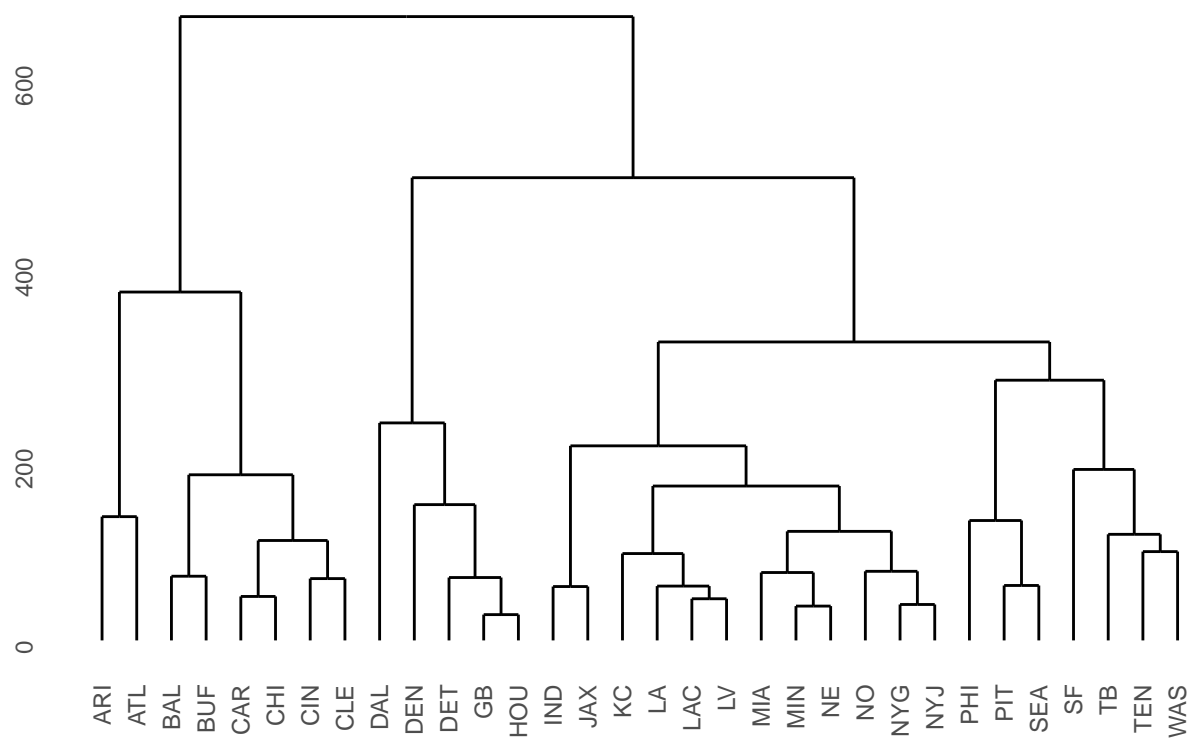
dist.map
hclust (*, "complete")

p1

NULL

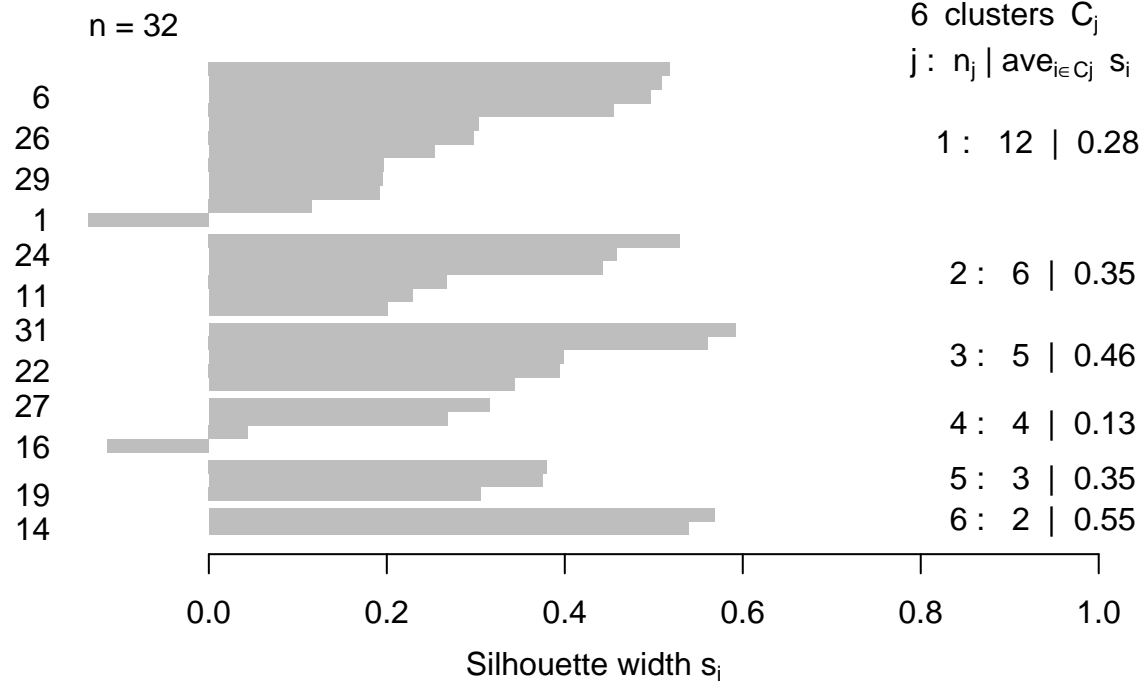
ggdendrogram(hclustavg, rotate = FALSE, size = 2) + labs(title = "complete Linkage")

complete Linkage



```
p<-cutree(hclustavg,k=6)
s<-silhouette(p,dist.map)
plot(s)
```

Silhouette plot of (x = p, dist = dist.map)



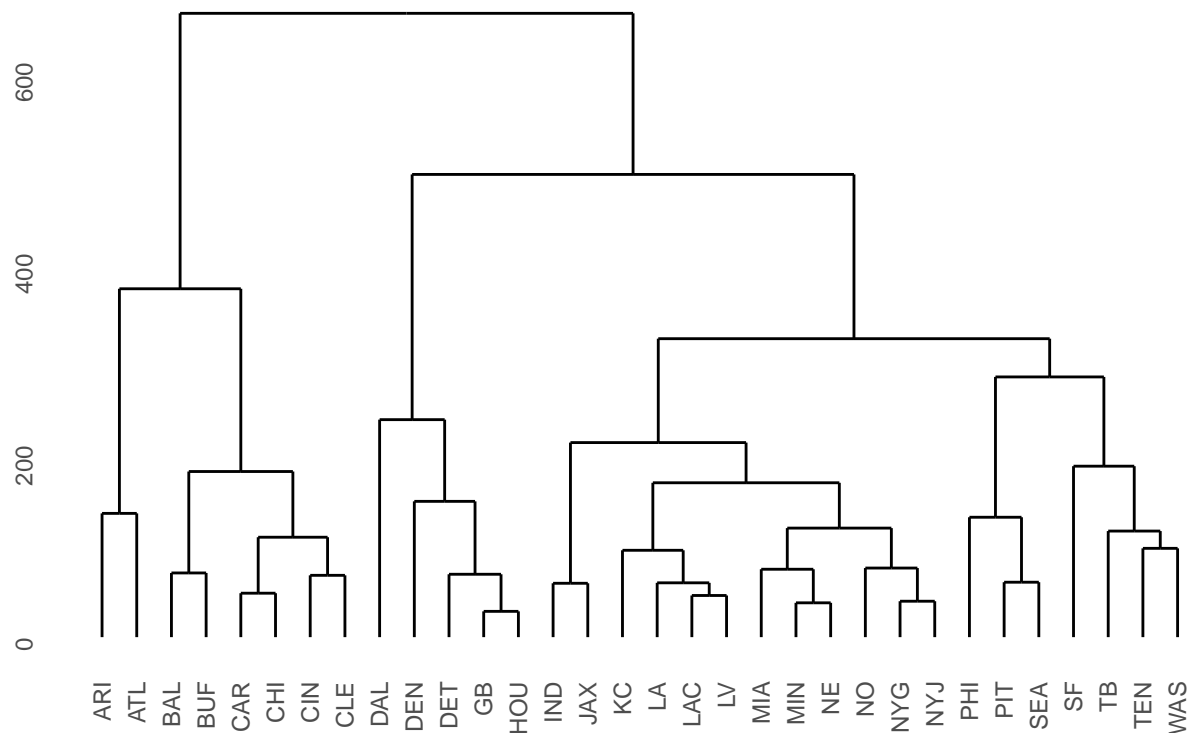
Average silhouette width : 0.33

so 2 is the

best number of clusters for complete Linkage using the silhouette diagram

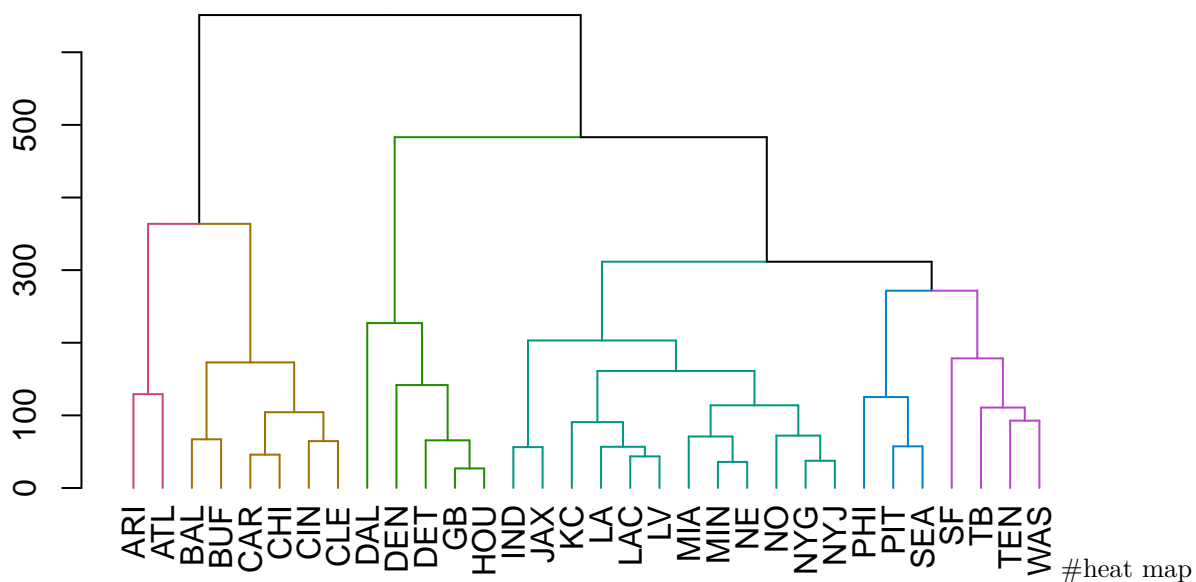
```
ggdendrogram(hclustavg, rotate = FALSE, size = 2) + labs(title = "complete Linkage")
```

complete Linkage



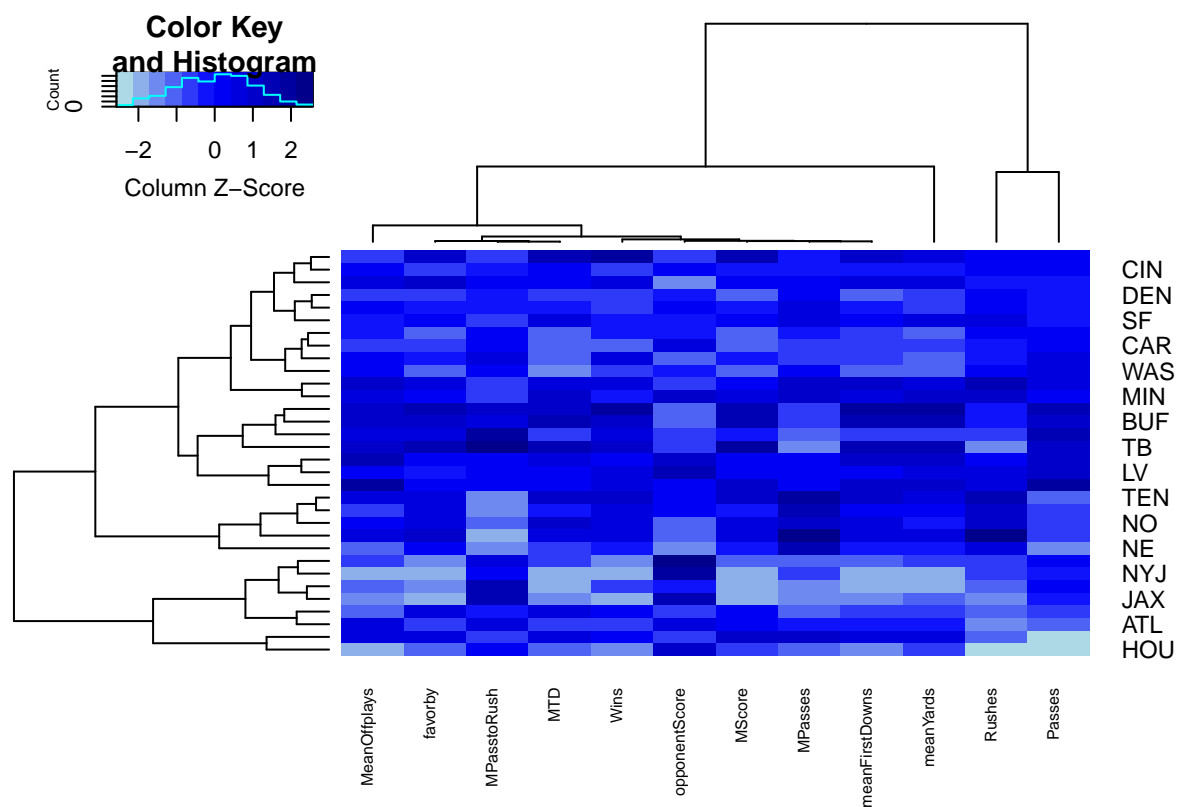
```
dend<-as.dendrogram(hclustavg)%>%
  color_branches(k=6)
```

```
plot(dend)
```

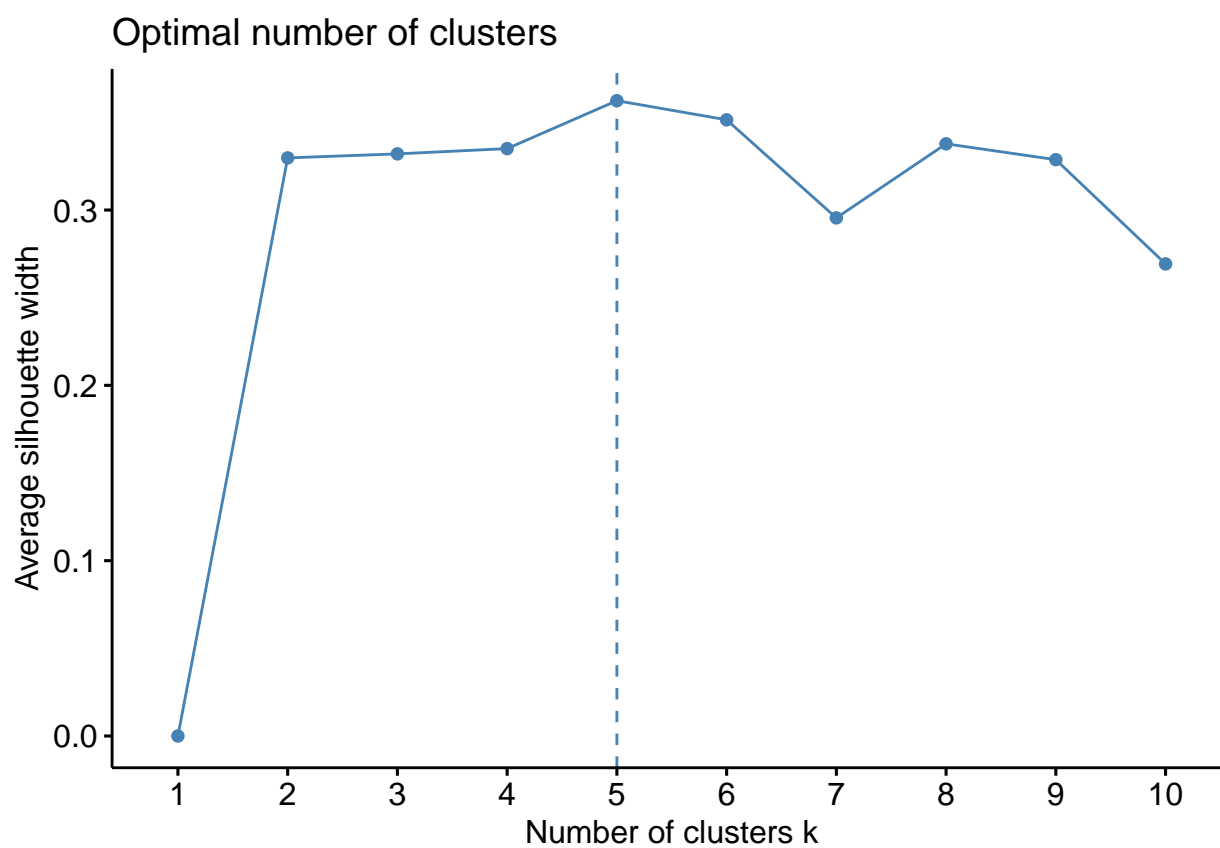


```
mycol <- colorRampPalette(c("lightblue","blue","darkblue"))(12)
C<-clust[,2:13]
c<-data.matrix(C)
rownames(c) <- clust$Team
```

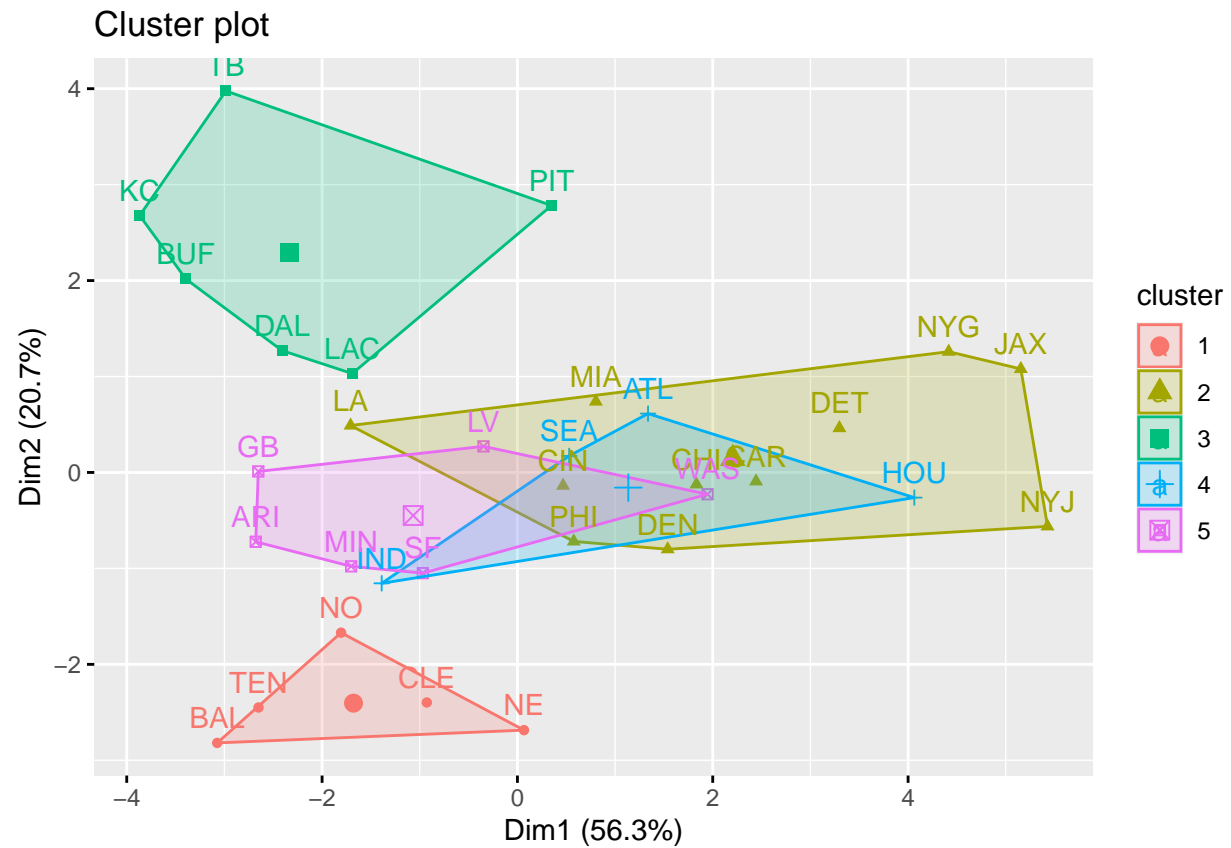
```
heatmap.2(c, col=mycol, trace="none",scale="column",cexCol =0.65 ,margins=c(7,5))
```



```
fviz_nbclust(c, kmeans, method='silhouette')
```



```
k<-kmeans(c, centers = 5)
fviz_cluster(k, data = c)
```



```
SOM<-som(c,grid=somgrid(2,3,topo="hexagonal"),rlen=1000)

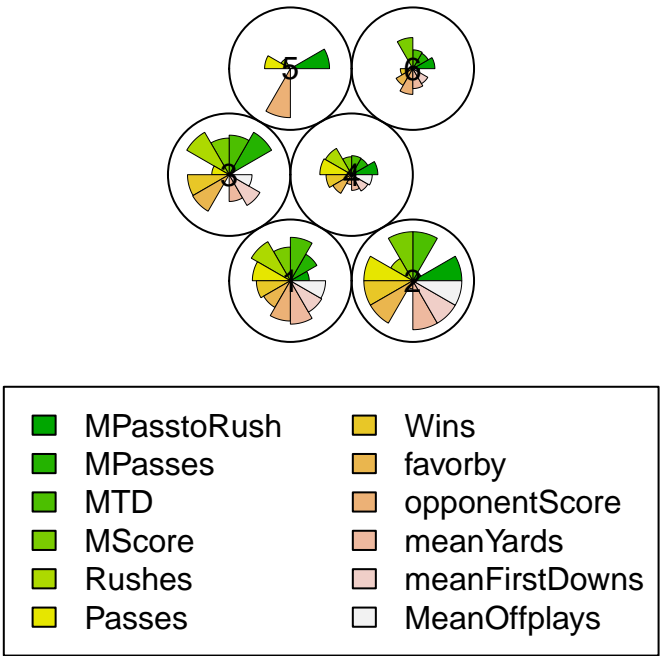
fitted<-SOM$unit.classif

wts<-SOM$codes

plot(SOM,type="codes")

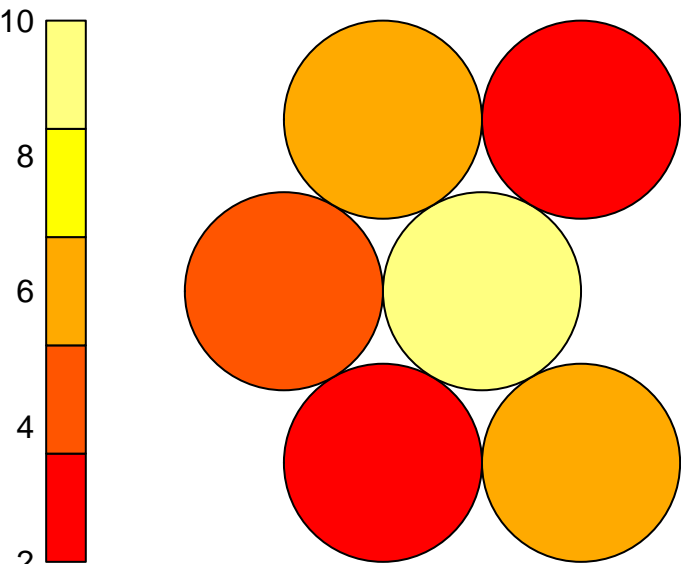
coords<-SOM$grid$pts
text(coords,labels=seq(1,6))
```

Codes plot



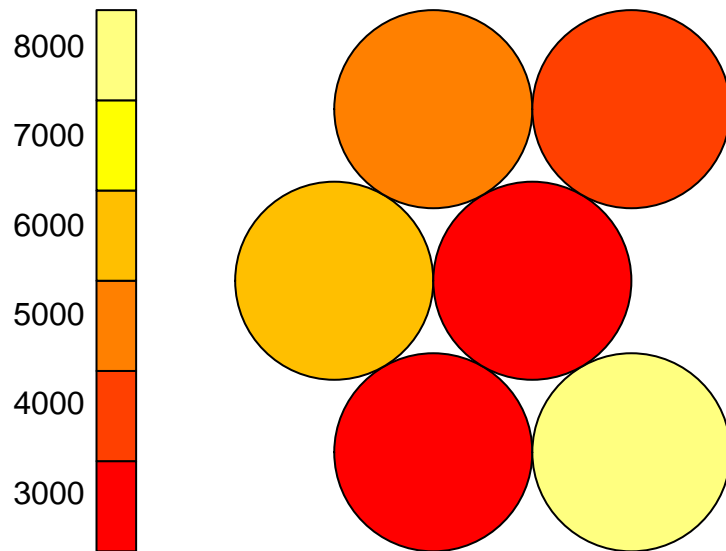
```
plot(SOM,type="counts")
```

Counts plot



```
plot(SOM,type="quality")
```

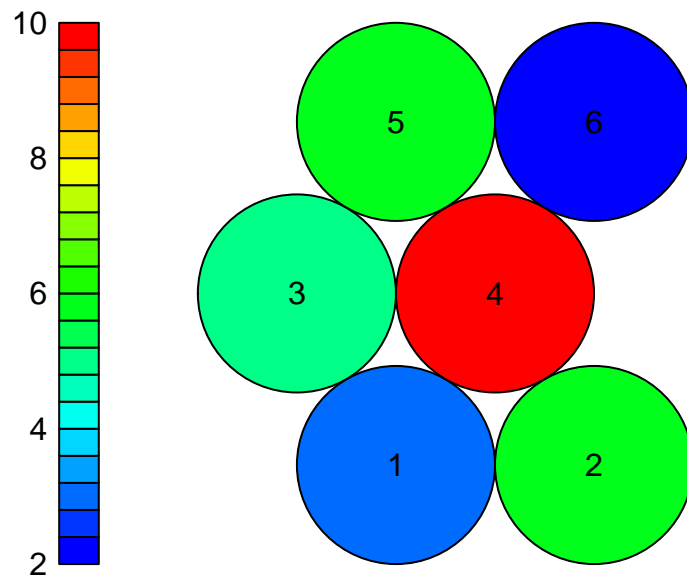

Quality plot



```
coolBlue.hotRed<-colorRampPalette(c("blue","cyan","green","yellow","red"))

plot(SOM,type="counts",palette.name=coolBlue.hotRed,ncolors=20)
coords<-SOM$grid$pts
text(coords,labels=seq(1,6))
```

Counts plot



```
par(bg = rgb(0,0.5,0.5), font = 1, pch = 16, pch = 16, cex = 0.5)
plot(SOM,type="mapping",labels = rownames(c), cex.lab = 0.65,
```

```

main = "NFL Team mapping", cex.main = 5.4)
text(coords, labels=seq(1,6), col = "blue")

```

