

userknowledgeKMeans.R

Yug

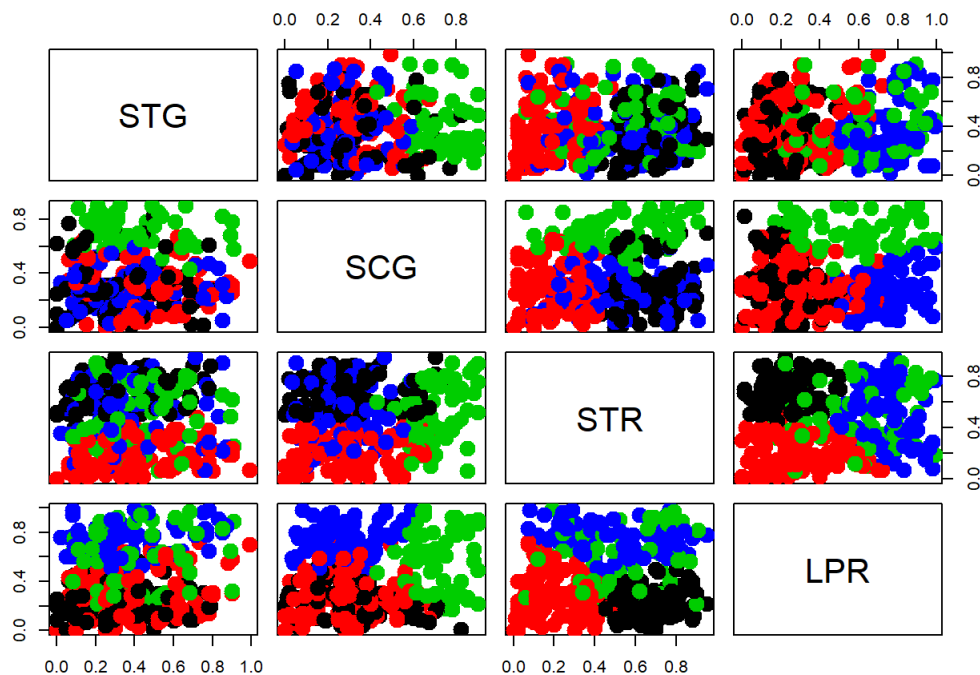
2019-04-21

```
userdatakm = read.csv("C:/Users/Yug/Desktop/Bigdata Project/CompleteUserDataSet.csv")
userdatakm$X <- NULL
userdatakm$X.1 <- NULL
userdatakm$X.2 <- NULL

user.features<-userdatakm[, 1:4]
#View(user.features)
results<-kmeans(user.features, 4)
results
```

```
## K-means clustering with 4 clusters of sizes 108, 135, 66, 94
##
## Cluster means:
##      STG      SCG      STR      LPR
## 1 0.3028333 0.3042315 0.6719444 0.2413889
## 2 0.3828963 0.2877704 0.2207037 0.2946741
## 3 0.3841970 0.7042424 0.5174242 0.5753030
## 4 0.3464043 0.2687021 0.5097872 0.7447872
##
## Clustering vector:
##  [1] 2 2 2 4 4 1 1 2 4 1 1 1 4 1 1 1 1 2 2 2 4 2 2 2 4 1 1 1 4 1 1 4 4 2 2
## [36] 2 4 2 2 2 4 1 1 4 1 1 1 1 4 2 2 3 4 3 3 3 4 1 1 3 3 1 3 3 3 2 2 2 2 2
## [71] 2 2 4 1 1 1 4 1 1 1 4 2 2 2 4 2 2 2 4 1 1 4 4 1 1 4 4 2 2 2 2 2 2 2 4
## [106] 1 1 1 4 1 1 1 4 2 2 3 3 3 3 3 3 3 1 3 3 3 1 3 3 3 2 2 2 4 2 2 2 4 1 1
## [141] 4 4 1 1 1 4 2 2 2 2 2 2 2 4 1 1 1 4 1 1 4 4 2 2 2 2 2 4 4 2 1 1 4 4 1
## [176] 1 4 4 2 2 3 3 3 2 3 3 1 3 3 3 1 3 3 3 2 2 2 2 2 2 4 1 1 4 4 1 1 1 4
## [211] 2 2 2 4 2 2 2 4 1 1 4 4 1 1 4 4 2 2 2 2 2 4 4 1 1 4 1 1 1 4 4 2 3 3
## [246] 3 2 2 3 3 1 3 3 3 1 3 3 3 1 4 4 4 1 1 1 4 4 2 2 2 2 4 1 2 4 1 1 1 1 4
## [281] 4 1 1 1 1 4 4 2 2 2 4 4 1 1 4 4 2 2 2 3 2 3 3 3 3 1 1 1 3 1 3 1 1
## [316] 3 3 3 2 4 4 2 2 2 4 4 4 1 1 1 4 4 1 1 4 4 2 2 4 2 2 2 2 4 4 1 1 1 4 4
## [351] 1 2 3 3 3 4 2 2 2 2 2 4 2 2 2 2 2 4 1 1 2 4 3 2 2 2 2 4 4 2 2 1 2 2 2
## [386] 4 2 2 2 2 4 2 3 1 2 2 3 3 3 3 1 3 3
##
## Within cluster sum of squares by cluster:
## [1] 10.71785 14.80796 10.50895 11.68633
## (between_SS / total_SS =  45.7 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"
## [5] "tot.withinss" "betweenss"    "size"         "iter"
## [9] "ifault"
```

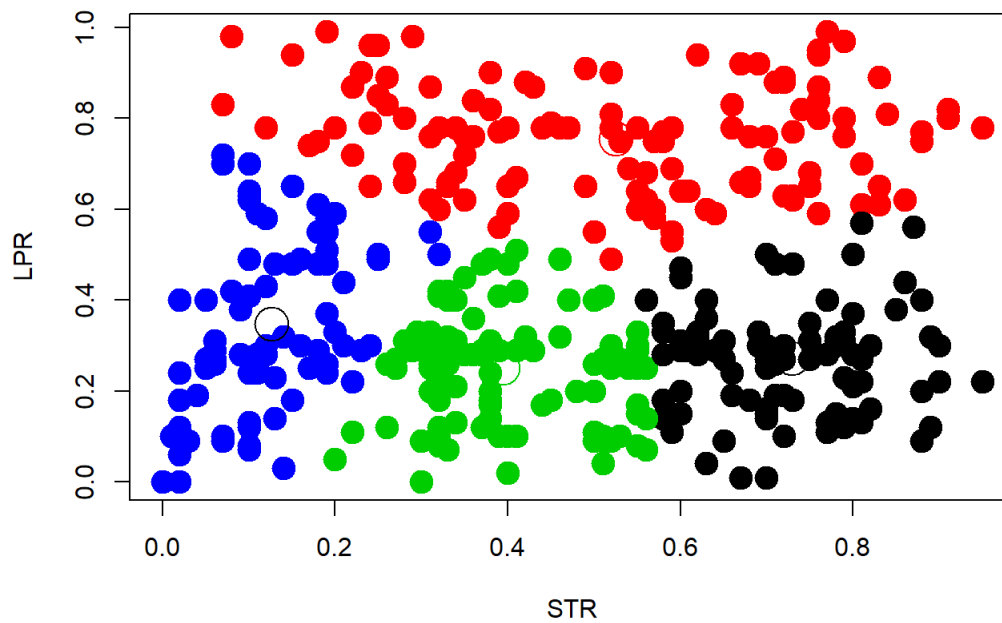
```
plot(user.features,col=results$cluster, pch = 20, cex = 3)
points(results$centers, col=1:3, pch=21, cex=3)
```



```
user.features1<-userdatakm[, 3:4]
#View(user.features)
results<-kmeans(user.features1, 4)
results
```

```
## K-means clustering with 4 clusters of sizes 92, 125, 105, 81
##
## Cluster means:
##      STR      LPR
## 1 0.7292391 0.2719565
## 2 0.5251200 0.7540000
## 3 0.3948095 0.2509619
## 4 0.1265432 0.3482716
##
## Clustering vector:
##  [1] 4 4 4 4 2 3 3 3 2 3 3 3 2 1 1 1 1 4 4 4 2 3 3 3 2 1 3 1 2 1 1 2 2 4 3
## [36] 4 2 3 3 3 2 1 3 2 1 1 1 1 2 4 4 4 2 3 3 3 2 1 1 2 2 1 1 1 2 4 4 4 4 3
## [71] 3 3 2 3 1 3 2 1 1 1 2 3 3 4 4 3 3 3 2 3 1 2 2 1 1 2 2 3 3 4 4 3 3 3 2
## [106] 3 1 1 2 1 1 1 2 4 3 4 2 2 3 3 2 2 3 3 2 2 1 1 2 2 4 3 4 2 3 3 3 2 1 3
## [141] 2 2 1 1 1 2 4 4 4 3 3 3 2 3 3 1 2 1 1 2 2 4 3 4 4 3 3 2 3 1 1 2 2 1
## [176] 1 2 2 4 4 4 2 3 3 2 2 1 1 2 2 1 1 1 2 3 3 4 4 3 3 4 2 3 1 2 2 1 1 1 2
## [211] 4 4 4 2 3 3 4 2 3 1 2 2 1 1 2 2 4 4 4 3 3 2 2 3 3 2 1 1 1 2 2 4 4 2
## [246] 2 3 3 2 2 1 1 2 2 1 1 2 2 3 2 2 2 1 1 1 2 2 4 3 4 4 2 3 3 2 3 1 1 1 2
## [281] 2 1 1 1 1 2 2 4 4 4 2 2 1 1 2 2 4 4 4 2 3 3 3 2 2 2 3 3 3 2 1 1 1 1
## [316] 1 2 2 4 2 2 3 3 3 2 2 3 1 3 2 2 1 1 2 2 3 4 2 3 3 3 3 2 2 3 3 1 2 2
## [351] 1 3 3 3 3 2 4 4 4 4 4 2 3 3 4 4 4 2 1 1 4 2 2 4 4 4 4 2 2 3 3 1 4 4 4
## [386] 2 3 4 4 4 2 3 1 1 4 4 4 3 1 2 1 1 2
##
## Within cluster sum of squares by cluster:
## [1] 2.151095 7.396523 2.384800 3.218990
## (between_SS / total_SS =  70.4 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"
## [5] "tot.withinss" "betweenss"    "size"         "iter"
## [9] "ifault"
```

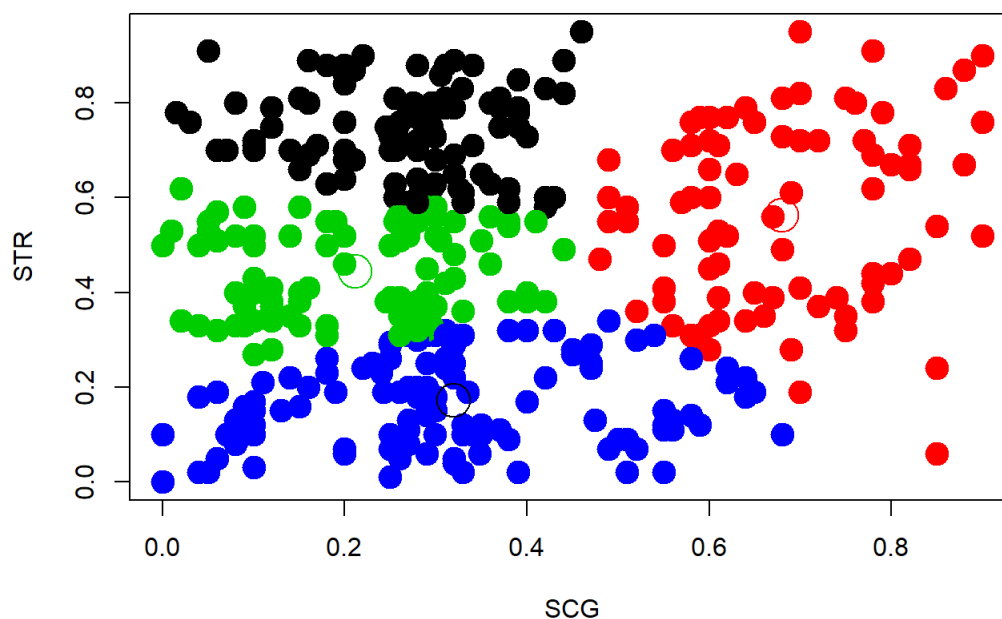
```
plot(user.features1,col=results$cluster, pch = 20, cex = 3)
points(results$centers, col=1:3, pch=21, cex=3)
```



```
user.features2<-userdatakm[, 2:3]
results<-kmeans(user.features2, 4)
results
```

```
## K-means clustering with 4 clusters of sizes 95, 86, 100, 122
##
## Cluster means:
##      SCG      STR
## 1 0.2636105 0.7414737
## 2 0.6794186 0.5636047
## 3 0.2107300 0.4454000
## 4 0.3188361 0.1720082
##
## Clustering vector:
##  [1] 4 4 4 4 4 3 3 3 3 3 3 3 1 1 1 1 4 4 4 4 3 4 4 3 1 3 1 1 1 1 1 4 4
## [36] 4 4 4 4 3 3 2 2 1 1 1 1 1 1 4 4 2 4 2 2 2 2 2 2 2 2 2 2 2 4 4 4 4 3
## [71] 3 3 3 3 1 3 3 1 1 1 1 4 4 4 4 4 4 3 4 3 3 1 3 3 1 1 1 1 4 4 4 4 4 3 4
## [106] 3 1 1 3 1 1 1 1 4 2 4 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 4 4 4 3 3 3 3 3
## [141] 3 1 1 1 1 1 4 4 4 4 3 3 3 3 3 3 1 3 1 1 1 1 4 4 4 4 3 3 4 4 1 1 1 1
## [176] 1 1 1 4 4 4 4 2 2 2 2 2 2 2 2 2 2 2 2 2 4 3 4 4 3 3 3 3 3 1 3 1 1 1
## [211] 4 4 4 4 3 3 3 4 3 1 3 1 1 1 1 1 4 4 4 4 4 3 3 3 3 3 3 1 1 1 1 4 2 4
## [246] 4 2 2 2 2 2 2 2 2 2 2 2 2 3 3 1 1 1 1 1 1 1 4 4 4 4 4 3 3 3 3 3 1 3
## [281] 3 1 1 1 1 1 1 4 4 4 3 2 1 1 3 2 4 4 4 4 2 4 2 2 2 2 2 2 2 2 2 2 2 2
## [316] 2 2 2 4 4 4 3 3 3 3 3 3 3 1 3 1 3 1 1 1 1 4 4 4 4 3 3 3 3 3 3 1 3 1
## [351] 1 4 2 2 2 2 4 4 4 4 4 4 3 3 4 4 4 4 1 1 4 2 2 4 4 4 4 4 3 3 3 4 4 4
## [386] 4 3 4 4 4 4 4 1 1 4 4 4 2 2 2 2 2 2
##
## Within cluster sum of squares by cluster:
## [1] 1.795902 4.249453 2.134206 4.284816
## (between_SS / total_SS = 71.1 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"
## [5] "tot.withinss" "betweenss"    "size"         "iter"
## [9] "ifault"
```

```
plot(user.features2,col=results$cluster, pch = 20, cex = 3)
points(results$centers, col=1:3, pch=21, cex=3)
```



```
user.features3<-userdatakm[, 2:3]
results<-kmeans(user.features3, 4)
results
```

```
## K-means clustering with 4 clusters of sizes 141, 57, 138, 67
##
## Cluster means:
##      SCG      STR
## 1 0.2477801 0.6753901
## 2 0.6921053 0.6749123
## 3 0.2197971 0.2367029
## 4 0.5779851 0.2697015
##
## Clustering vector:
##  [1] 3 3 3 3 3 3 3 3 1 1 1 1 1 1 1 1 1 1 3 3 3 3 3 3 3 3 1 1 1 1 1 1 3 4
## [36] 3 4 4 3 3 3 2 2 1 1 1 1 1 1 4 4 4 4 4 4 4 2 2 2 2 2 2 2 3 3 3 3
## [71] 3 3 3 1 1 1 1 1 1 1 1 3 3 3 3 3 3 3 1 1 1 1 1 1 1 1 3 3 3 3 3 3
## [106] 1 1 1 1 1 1 1 1 4 4 4 4 4 2 4 4 4 2 2 2 2 2 2 2 3 3 3 3 3 3 1 1
## [141] 1 1 1 1 1 1 3 3 3 3 3 3 3 1 1 1 1 1 1 1 3 4 3 3 4 1 3 4 1 1 1 1
## [176] 1 1 1 4 4 4 4 4 4 4 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 1 1 1 1 1 1
## [211] 3 3 3 3 1 3 3 3 1 1 1 1 1 1 1 4 3 3 4 4 3 4 4 1 1 1 1 1 1 4 4 4
## [246] 4 4 4 2 4 2 2 2 2 2 2 1 1 1 1 1 1 1 3 3 3 3 3 3 3 3 1 1 1 1
## [281] 1 1 1 1 1 1 3 3 3 4 4 1 1 1 2 4 4 4 4 4 4 4 4 2 2 2 2 2 2 2 2
## [316] 2 2 2 3 3 3 3 3 3 3 1 1 1 1 1 1 1 1 3 3 3 3 3 3 3 3 1 1 1 1
## [351] 1 4 4 4 2 4 3 3 3 3 3 3 3 3 3 3 4 1 1 4 2 2 3 3 3 3 3 3 1 3 3 3
## [386] 3 3 4 3 3 4 4 1 1 4 4 4 4 2 2 2 2
##
## Within cluster sum of squares by cluster:
## [1] 3.777756 1.678572 3.495177 1.754447
## (between_SS / total_SS =  75.2 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"      "withinss"
## [5] "tot.withinss" "betweenss"    "size"      "iter"
## [9] "ifault"
```

```
plot(user.features3,col=results$cluster, pch = 20, cex = 3)
points(results$centers, col=1:3, pch=21, cex=3)
```

