Tuto charclust Package

## Intro

The goal of this tutorial is to show how to use the charclust package and interpret the results.  
The dataset for this example is “insertion\_master” which is provided with the package. but du clustering methodes des kmeans avec 3 classes

library(charclust)  
data(insertion\_master)  
data.illu <- insertion\_master[,c(1:6,12)]  
data.act <-insertion\_master[,7:11]  
res.kmeans<-kmeans(data.act,centers=3,nstart=5)

## Creation de l’objet

explications ….

obj <- objcharac(insertion\_master, data.act, data.illu, res.kmeans$cluster)

## Evaluation du clustering

### DB

db <- db\_index(obj)  
print(db)

## DB Index   
## 0.6033811

bof on va regarder un autre indicateur ### Silhouette expliquer en quoi consiste l’indice

silhouette <- sil(obj)  
print(silhouette$silglob)

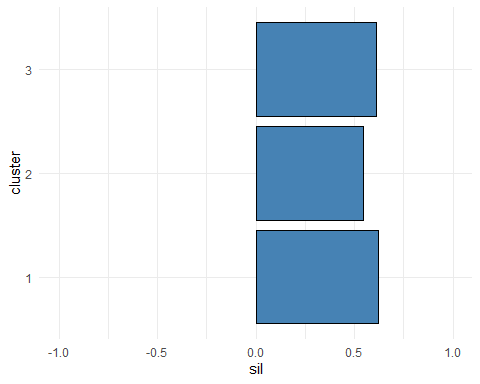
## [1] 0.5930815

il est pas mal regardons par clusters

print(silhouette$silclus)

## 1 2 3   
## 0.6226540 0.5475537 0.6090367

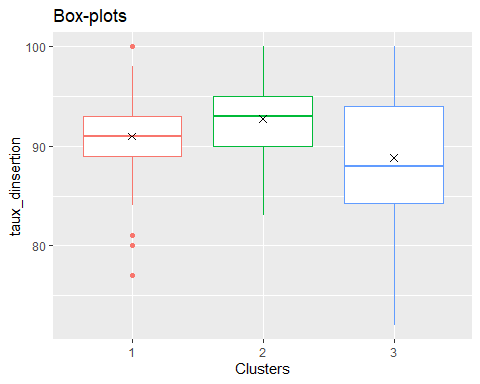
silhouette$plot

 classe 2 legerement moins bien classé mais globalement idem

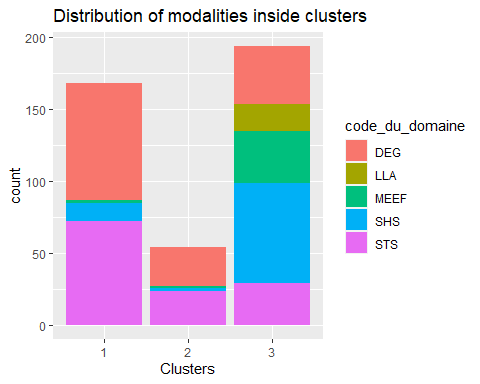
## Caractérisation univariée

expliquer vite fait on va d’abord visualiser

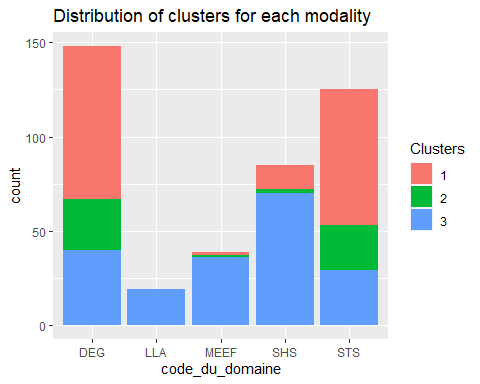
graph\_uni\_act <- charac\_graph(obj, type = "act", profile = "l")  
graph\_uni\_act$taux\_dinsertion

 hierarchie du taux d’insertion groupe 2 >>> groupe 3 >>> groupe 1

graph\_uni\_illus <- charac\_graph(obj, type = "illus", profile = "l")  
graph\_uni\_illus$code\_du\_domaine



graph\_uni\_illus2 <- charac\_graph(obj, type = "illus", profile = "c")  
graph\_uni\_illus2$code\_du\_domaine



univariate = charac\_uni(obj)  
print(univariate)

## $taux\_dinsertion  
## G 1 G 2 G 3 % epl.   
## 90.988095 92.759259 88.824742 7.517836   
##   
## $emplois\_cadre\_ou\_professions\_intermediaires  
## G 1 G 2 G 3 % epl.   
## 88.82738 93.68519 81.07216 20.18582   
##   
## $emplois\_stables  
## G 1 G 2 G 3 % epl.   
## 79.77381 83.55556 66.88660 26.96383   
##   
## $emplois\_a\_temps\_plein  
## G 1 G 2 G 3 % epl.   
## 97.36310 97.92593 88.36598 31.14120   
##   
## $salaire\_brut\_annuel\_estime  
## G 1 G 2 G 3 % epl.   
## 30773.21429 35322.22222 26122.16495 84.37342   
##   
## $diplome  
## khi2 p\_value cramer   
## 3.609294e+01 1.453840e-08 2.945537e-01   
##   
## $etablissement  
## khi2 p\_value cramer   
## 2.516395e+02 3.312681e-09 5.499557e-01   
##   
## $academie  
## khi2 p\_value cramer   
## 1.269567e+02 8.238328e-08 3.906306e-01   
##   
## $code\_du\_domaine  
## khi2 p\_value cramer   
## 1.488700e+02 3.373612e-28 4.230015e-01   
##   
## $domaine  
## khi2 p\_value cramer   
## 1.488700e+02 3.373612e-28 4.230015e-01   
##   
## $discipline  
## khi2 p\_value cramer   
## 2.022341e+02 8.239040e-26 4.930211e-01   
##   
## $femmes  
## G 1 G 2 G 3 % epl.   
## 48.06548 47.12963 70.73711 34.06873

print(univariate$taux\_dinsertion)

## G 1 G 2 G 3 % epl.   
## 90.988095 92.759259 88.824742 7.517836

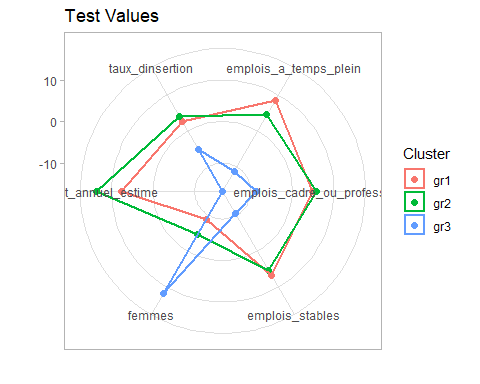
print(univariate$code\_du\_domaine)

## khi2 p\_value cramer   
## 1.488700e+02 3.373612e-28 4.230015e-01

vtest = testval(obj)  
#print result for numeric variables  
tquanti <- vtest$num[[1]]  
print(tquanti)

## taux\_dinsertion emplois\_cadre\_ou\_professions\_intermediaires  
## G 1 vt 2.539999 4.768985  
## G 2 vt 3.902054 5.878595  
## G 3 vt -5.127112 -8.651083  
## emplois\_stables emplois\_a\_temps\_plein salaire\_brut\_annuel\_estime  
## G 1 vt 6.866863 8.449083 7.572277  
## G 2 vt 5.430371 4.506094 13.793643  
## G 3 vt -10.412509 -11.346030 -16.740760  
## femmes  
## G 1 vt -8.939882  
## G 2 vt -4.570777  
## G 3 vt 11.872340

vtest$graph

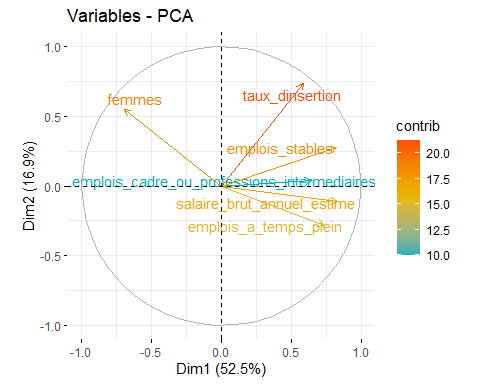


#print result for qualitative variables  
tquali <- vtest$num[[2]]  
#print(tquali)  
#print result for one qualitative variable  
print(tquali$domaine)

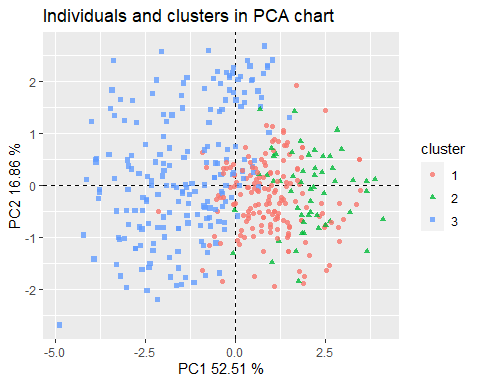
## Droit, economie et gestion Lettres, langues, arts Masters enseignement  
## 1 3.996125 -0.9749608 -1.839795  
## 2 4.560795 -0.9749608 -1.691125  
## 3 -4.730061 1.1156768 2.063950  
## Sciences humaines et sociales Sciences, technologies et sante  
## 1 -3.250187 3.730284  
## 2 -4.283108 4.192541  
## 3 4.006800 -4.397345

## Caractérisation multivariée

pca <- charac\_graph(obj, type = "pca")  
pca$var



pca$ind



### ADL

lda = clustlda(obj)  
#print(lda$coef\_LDA)  
#the confusion matrix  
print(lda$confusion\_matrix)

##   
## ytest 1 2 3  
## 1 41 0 0  
## 2 3 9 0  
## 3 2 0 49

#the model evaluation  
print(lda$eval)

## var FValue pvalue  
## 1 taux\_dinsertion 2.362846 0.095873  
## 2 emplois\_cadre\_ou\_professions\_intermediaires 2.761440 0.064781  
## 3 emplois\_stables 3.050615 0.048777  
## 4 emplois\_a\_temps\_plein 10.728466 0.000031  
## 5 salaire\_brut\_annuel\_estime 489.106243 0.000000