Demonstration Cluster Analysis

W. Grossmann

```
#====================================
# libraries
# ====================================
library(cluster)
library(vegan) # for visualisation
```

```
## Warning: package 'vegan' was built under R version 3.4.4
```

```
## Loading required package: permute
```

```
## Warning: package 'permute' was built under R version 3.4.3
```

```
## Loading required package: lattice
```

```
## This is vegan 2.5-2
```

```
library(vegan3d) # for visualization
```

```
## Warning: package 'vegan3d' was built under R version 3.4.3
```

```
#-------------------------------------------------------------------------
#Data importance of 63 pictures for seven categories of tourist behaviour
#-------------------------------------------------------------------------

Koeffizienten3<-read.csv("coefficients3.csv", header = TRUE, sep=";")

variable.names(Koeffizienten3)
```

```
## [1] "Bild" "F1"   "F2"   "F3"   "F4"   "F5"   "F6"   "F7"
```

```
summary(Koeffizienten3)
```

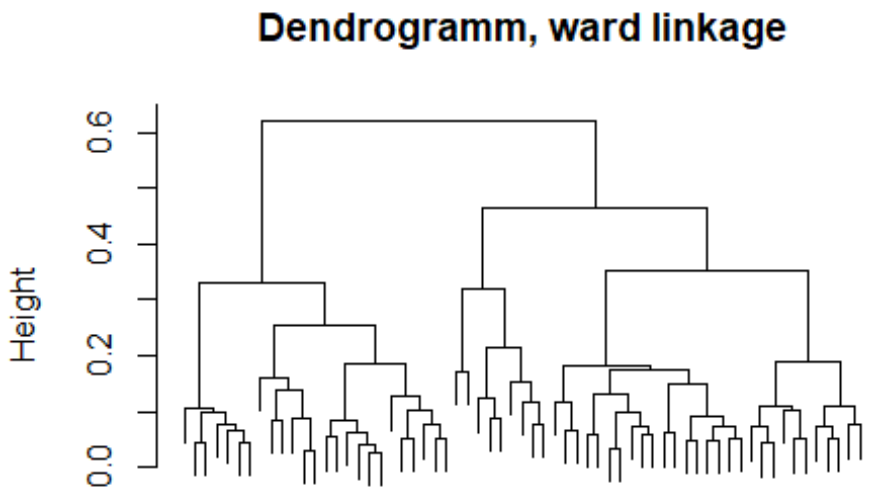
```
##       Bild          F1                  F2                  F3
##  1_1    : 1   Min.   :-0.02169   Min.   :-0.02751   Min.   :-0.04493
##  1_1_2  : 1   1st Qu.: 0.01792   1st Qu.: 0.05034   1st Qu.: 0.04238
##  1_10   : 1   Median : 0.03363   Median : 0.07464   Median : 0.06667
##  1_2    : 1   Mean   : 0.04510   Mean   : 0.07089   Mean   : 0.06410
##  1_3    : 1   3rd Qu.: 0.05802   3rd Qu.: 0.09439   3rd Qu.: 0.08767
##  1_4    : 1   Max.   : 0.16883   Max.   : 0.15896   Max.   : 0.15008
##  (Other):58
##        F4                  F5                  F6
##  Min.   :-0.02068   Min.   :-0.004305   Min.   :-0.018977
##  1st Qu.: 0.02630   1st Qu.: 0.042016   1st Qu.: 0.006603
##  Median : 0.05541   Median : 0.076484   Median : 0.028341
##  Mean   : 0.06565   Mean   : 0.075289   Mean   : 0.043371
##  3rd Qu.: 0.10534   3rd Qu.: 0.112465   3rd Qu.: 0.063542
```

```
##  Max.   : 0.17670   Max.   : 0.172912   Max.   : 0.230793
##
##        F7
##  Min.   :-0.01233
##  1st Qu.: 0.01015
##  Median : 0.01992
##  Mean   : 0.03567
##  3rd Qu.: 0.05403
##  Max.   : 0.18164
##
```

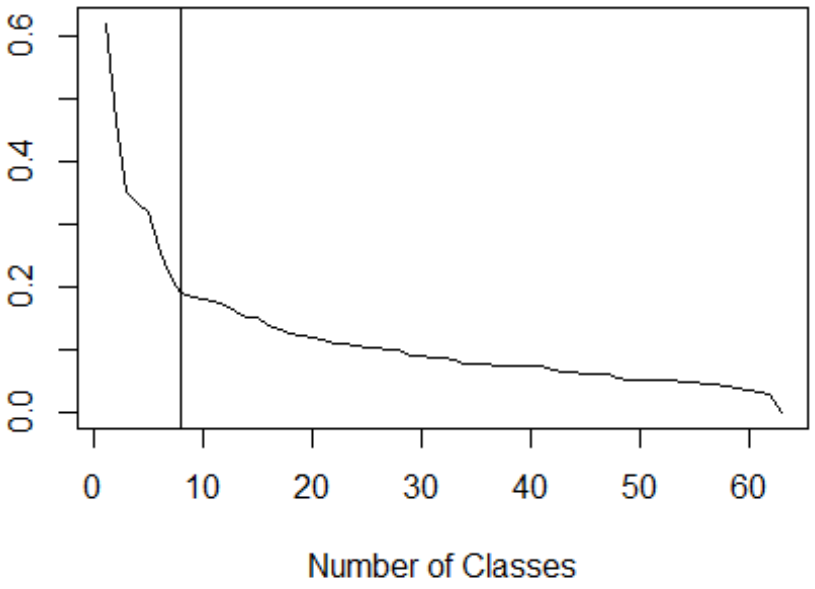
```
#-----------------------------------
#Hierarchical Clustering not standardised
#-----------------------------------

# Cluster variables and distance
clustvar3.1<-Koeffizienten3[2:64,2:8]
d3.1<-daisy(clustvar3.1)

#Hierarchical clustering

cl3.1<-hclust(d3.1, method = "ward.D2", members=NULL)
summary(cl3.1)
```

```
##             Length Class  Mode
## merge       124    -none- numeric
## height       62    -none- numeric
## order        63    -none- numeric
## labels       63    -none- character
## method        1    -none- character
## call          4    -none- call
## dist.method   0    -none- NULL
```

```
#pdf("cltree1_1.pdf",width = 7, height = 3.5)
#par(oma=c(0,2,0,2))

plot(cl3.1, xlab="", labels = FALSE, sub = "",
     main= "Dendrogramm, ward linkage")
```




```
#dev.off()
plot(1:63, c(cl3.1$height[62:1],0),type = "l",
     xlab ="Number of Classes", ylab="")
abline(v=8)
```



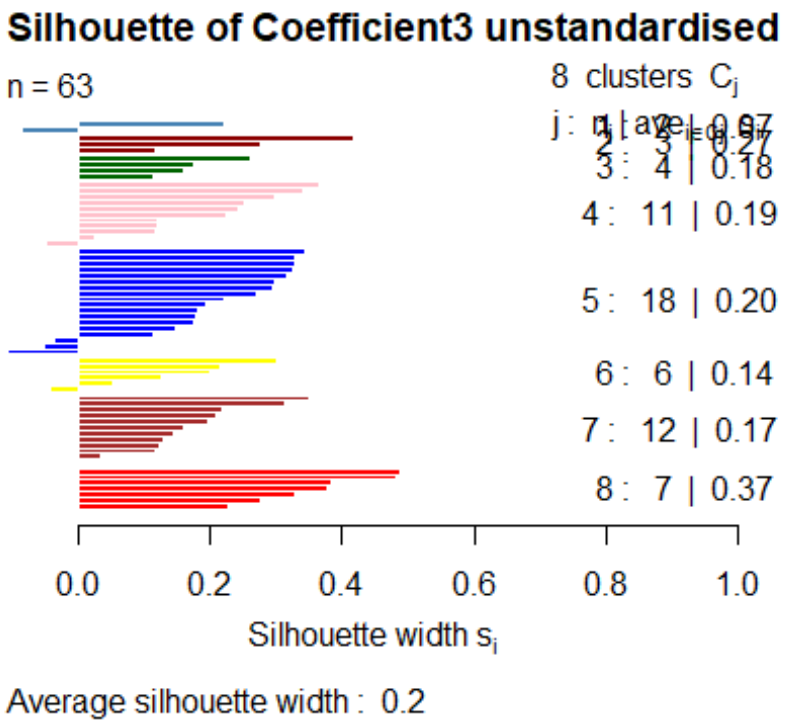
```r
cl3.1.8<-cutree(cl3.1, k=8) # eight clusters
cl3.1.8
```

```
##  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
##  1  2  3  2  4  5  3  6  4  7  8  8  7  8  7  2  5  5  7  8  8  7  4  5  5
## 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51
##  7  4  6  6  6  7  7  7  4  8  8  5  5  5  5  5  5  7  6  5  7  1  5  4  4
## 52 53 54 55 56 57 58 59 60 61 62 63 64
##  4  3  5  5  5  7  6  4  5  5  4  4  3
```

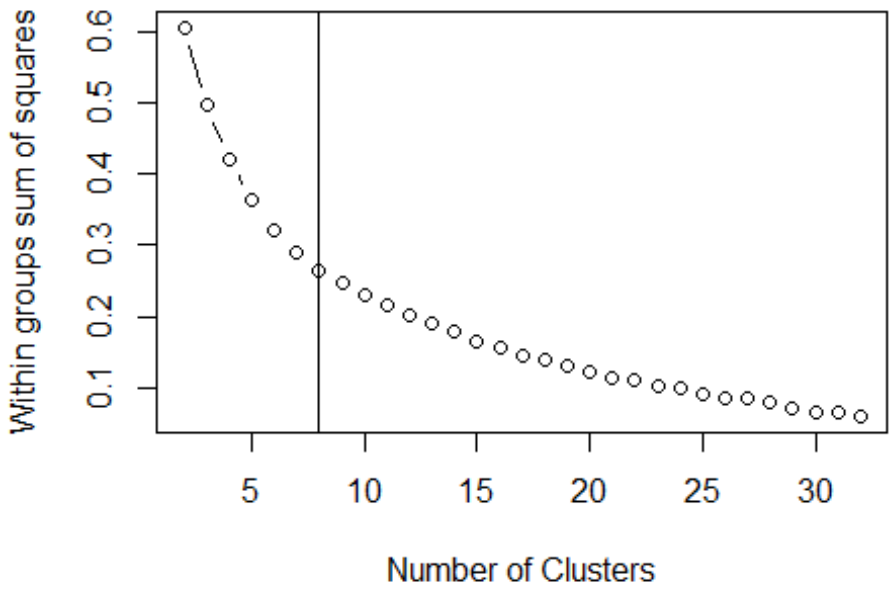
```r
table(cl3.1.8)
```

```
## cl3.1.8
##  1  2  3  4  5  6  7  8
##  2  3  4 11 18  6 12  7
```

```r
#plot(cl3.1.8) # shows distribution of indices

Co3_NoIntercept_Cl8 <- cl3.1.8
Bild3 <- as.character(Koeffizienten3$Bild[2:64])

result_ns3 <- cbind(clustvar3.1, Bild3, Co3_NoIntercept_Cl8)

#write.csv2(result_ns3, file="Co3_NoIntercept_Cl8.csv")

#-----------------------------------------
#Silhouette
#-----------------------------------------
#pdf("5_silhu.pdf",width = 7, height = 3.5)
#par(oma=c(0,2,0,2))
cols <- c("steelblue", "darkred", "darkgreen", "pink", "blue", "yellow",
"brown",
          "red")

silh3.1.8<-silhouette(cl3.1.8, daisy(clustvar3.1))
plot(silh3.1.8, col=cols,
     main = "Silhouette of Coefficient3 unstandardised")
```



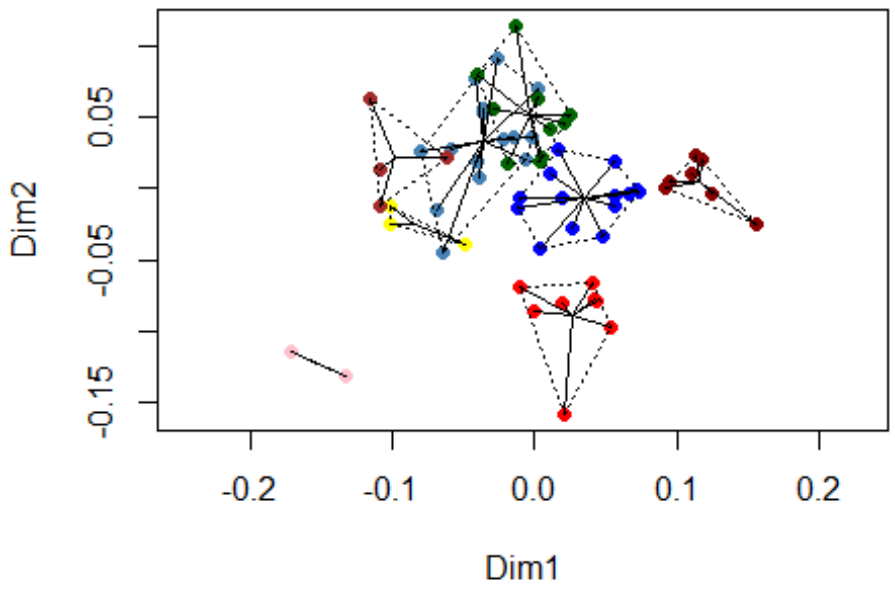

```
#dev.off()



#=====================================================
#k-Means with k = 8 unstandardised
#-----------------------------------------------

wss <- 2 :32

for(i in 2 : 32)
  wss[i] <- sum(kmeans(clustvar3.1,centers=i,nstart=25)$withinss)

plot(2:32, wss[2:32], type="b", xlab="Number of Clusters",
     ylab="Within groups sum of squares")
abline(v=8)
```



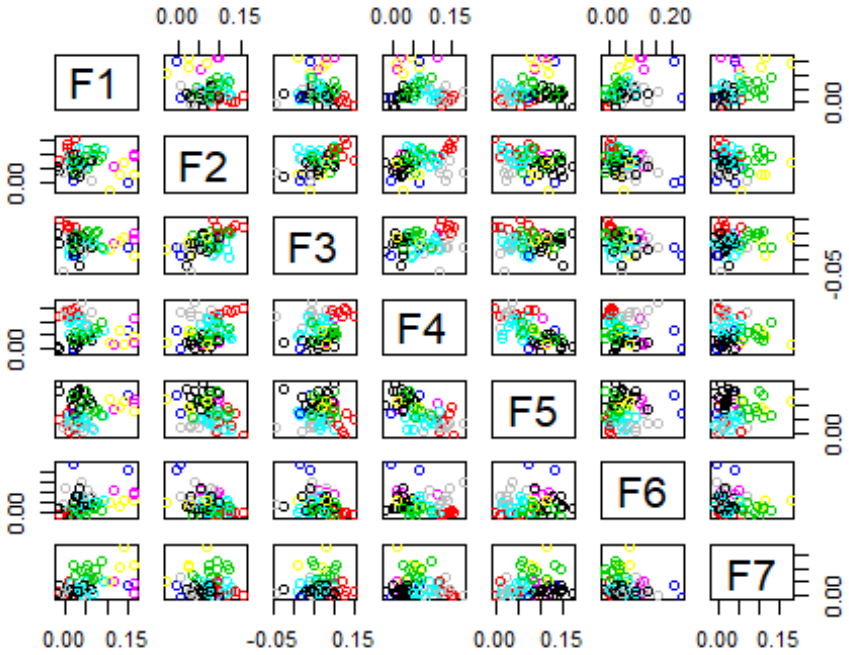

```
k_means_unstand_8<-kmeans(clustvar3.1, centers = 8)
table(k_means_unstand_8$cluster)

##
##  1  2  3  4  5  6  7  8
##  2  6 10  6  7 10 15  7

#summary for clusters
dim(k_means_unstand_8$centers)

## [1] 8 7

k_means_unstand_8$centers

##            F1           F2         F3         F4         F5            F6
## 1 0.085688421 0.0008882524 0.02863814 0.02902846 0.10097857  0.2195885346
## 2 0.025543719 0.0398770685 0.03629467 0.13764291 0.02606160  0.0852363808
## 3 0.053924901 0.0835720950 0.08075914 0.05595415 0.08986842  0.0142837813
## 4 0.024356482 0.0762057682 0.06077713 0.10344332 0.04728505  0.0091419061
## 5 0.008539344 0.1166257381 0.11843842 0.13512878 0.04131549 -0.0007737785
## 6 0.044140177 0.0957260706 0.04967511 0.05898768 0.04383836  0.0511600909
## 7 0.024964143 0.0596633863 0.05914392 0.02159938 0.12279862  0.0413321485
## 8 0.142945623 0.0477390018 0.06318994 0.03966871 0.10116764  0.0716129749
##           F7
## 1 0.01649053
## 2 0.01955312
## 3 0.09059361
## 4 0.02787721
```

```
## 5 0.01564893
## 6 0.01573883
## 7 0.01577661
## 8 0.07943539

Co3_k_means_undstand_8 <- cbind(clustvar3.1, Bild3,
k_means_unstand_8$cluster)

#write.csv2(Co3_k_means_undstand_8, file="Co3_k_means_undstand_8.csv")

#================================================

kclus <- kmeans(clustvar3.1,centers= 8, iter.max=1000, nstart=10000)

table(kclus$cluster)

##
##  1  2  3  4  5  6  7  8
## 15  7 10  2 14  3  4  8

#summary for clusters
dim(kclus$centers)

## [1] 8 7

kclus$centers

##            F1           F2         F3         F4         F5            F6
## 1 0.024964143 0.0596633863 0.05914392 0.02159938 0.12279862  0.0413321485
## 2 0.008539344 0.1166257381 0.11843842 0.13512878 0.04131549 -0.0007737785
## 3 0.053924901 0.0835720950 0.08075914 0.05595415 0.08986842  0.0142837813
## 4 0.085688421 0.0008882524 0.02863814 0.02902846 0.10097857  0.2195885346
## 5 0.035210201 0.0886022312 0.05698391 0.07435127 0.04621656  0.0267555899
## 6 0.150488363 0.0807314948 0.07889459 0.04537237 0.10287798  0.1012297428
## 7 0.137288569 0.0229946321 0.05141146 0.03539096 0.09988488  0.0494003990
## 8 0.030982520 0.0516658110 0.03517590 0.12443455 0.02892896  0.0879115465
##           F7
## 1 0.01577661
## 2 0.01564893
## 3 0.09059361
## 4 0.01649053
## 5 0.02244472
## 6 0.03310679
## 7 0.11418184
## 8 0.01596804

Co3_k_means_undstand_8_final <- cbind(clustvar3.1, Bild3, kclus$cluster)

#write.csv2(Co3_k_means_undstand_8_final, file="Co3_k_means_undstand_8_
_final.csv")

# distance matrix
```

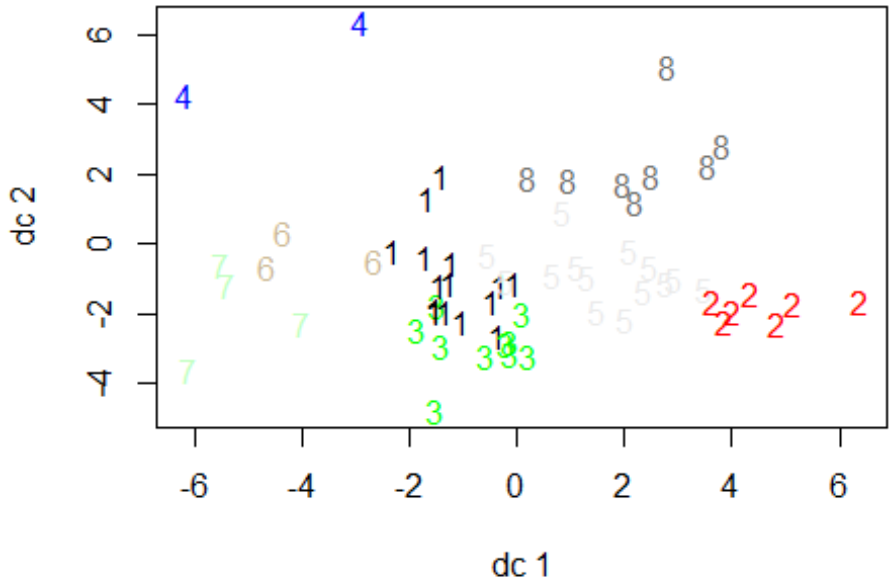
```r
stand_dist <- dist(clustvar3.1)

# Multidimensional scaling
cmd <- cmdscale(stand_dist)

# plot MDS, with colors by groups from kmeans
groups <- levels(factor(kclus$cluster))
table(kclus$cluster)

##
##  1  2  3  4  5  6  7  8
## 15  7 10  2 14  3  4  8

ordiplot(cmd, type = "n")

## species scores not available

for(i in seq_along(groups)){
  points(cmd[factor(kclus$cluster) == groups[i], ], col = cols[i], pch = 16)
}

# add spider and hull
ordispider(cmd, factor(kclus$cluster), label = FALSE)

ordihull(cmd, factor(kclus$cluster), lty = "dotted")
```



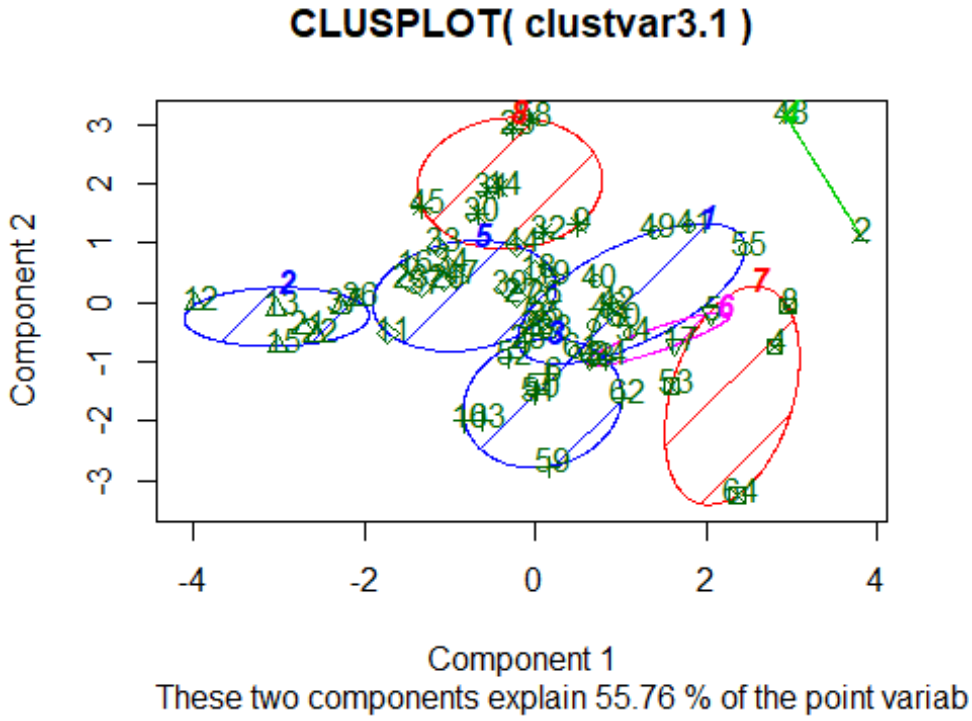

```r
#All 2dimensional Plots
plot(clustvar3.1, col=kclus$cluster)
```



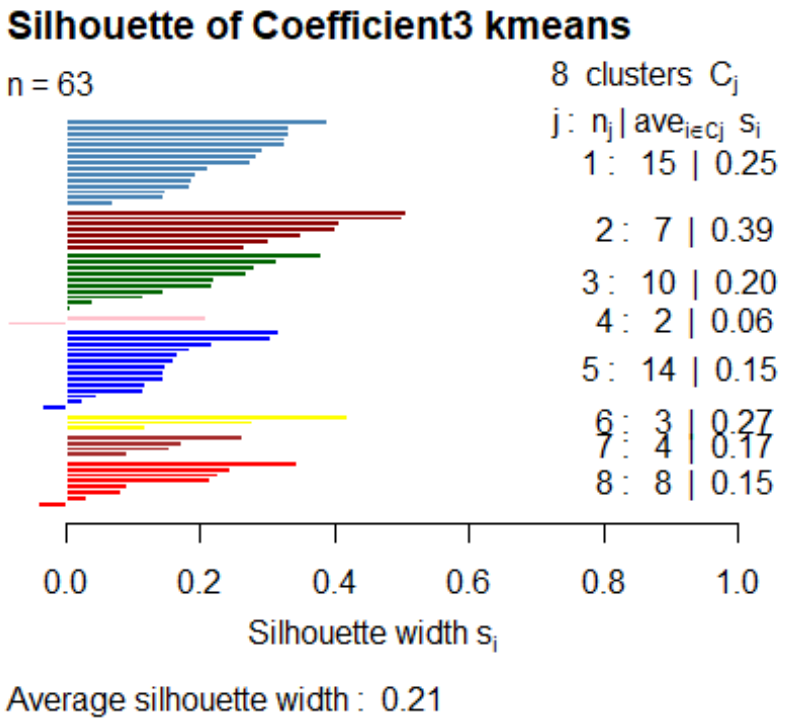

```
#Principal components Plots
library(fpc)

## Warning: package 'fpc' was built under R version 3.4.3

plotcluster(clustvar3.1, kclus$cluster)
```



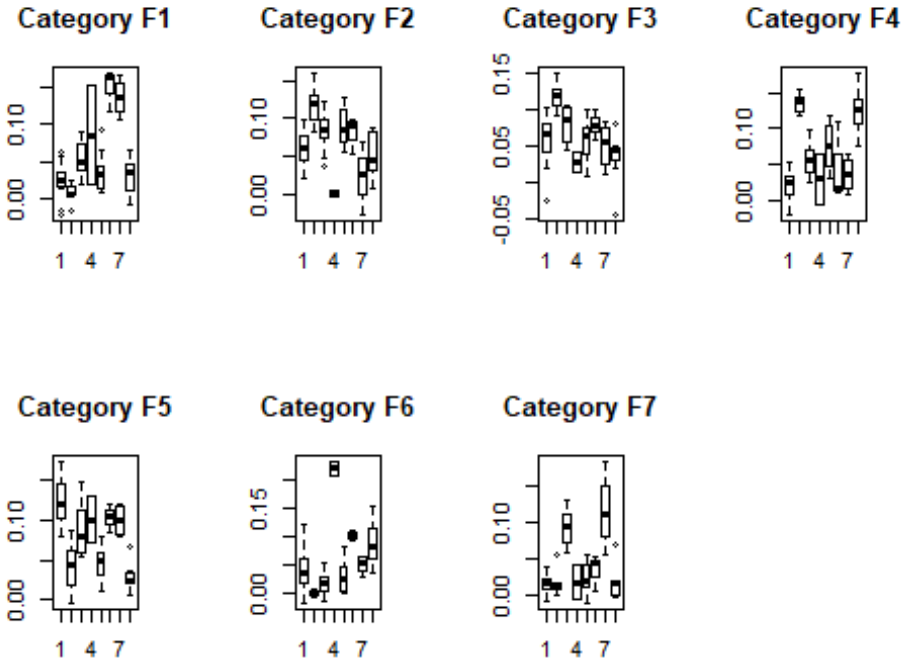

```
clusplot(clustvar3.1, kclus$cluster, color=TRUE, shade=TRUE,
         labels=2, lines=0)
```



```
#----------------------------------------
  #Silhouette
#----------------------------------------
#pdf("5_silhu.pdf",width = 7, height = 3.5)
#par(oma=c(0,2,0,2))
silh_kmeans <- silhouette(kclus$cluster, daisy(clustvar3.1))
plot(silh_kmeans, col=c("steelblue", "darkred", "darkgreen",
                        "pink", "blue", "yellow", "brown", "red"),
     main = "Silhouette of Coefficient3 kmeans")
```




```
#dev.off()

#Grouped Boxplots for Clusters

par(mfrow= c(2,4))

boxplot(clustvar3.1$F1~kclus$cluster, main = "Category F1")
boxplot(clustvar3.1$F2~kclus$cluster, main = "Category F2")
boxplot(clustvar3.1$F3~kclus$cluster, main = "Category F3")
boxplot(clustvar3.1$F4~kclus$cluster, main = "Category F4")
boxplot(clustvar3.1$F5~kclus$cluster, main = "Category F5")
boxplot(clustvar3.1$F6~kclus$cluster, main = "Category F6")
boxplot(clustvar3.1$F7~kclus$cluster, main = "Category F7")
```


Category F1
Category F2
Category F3
Category F4
Category F5
Category F6
Category F7
0.00
0.10
0.15
-0.05
0.05
1 4 7