```
Apply AFC to Thematic Therm Matrix (PPD Project)
                                   Lilia HARIRECHE - Antoine RODRIGUEZ
                                               MLDS-FA
 library(FactoMineR)
 library(ggplot2)
 library(ggrepel)
# 1. apply afc to TTM questions occ
 setwd("E:/M2 MLDS FA/PPD/mixkmeans/AFC")
 TT_occ_q = read.csv("TT_occ_q.csv")
 #head(TT_occ_q)
 #AFC
 res.ca = CA(TT_occ_q, ncp = 4, graph = FALSE)
 #contribution of any therm
 mat_terms_contrib = as.data.frame(res.ca$col$contrib)
 mat_terms_contrib_order = mat_terms_contrib[order(-mat_terms_contrib$`Dim 1`,
                                                     -mat_terms_contrib$`Dim 2`,
                                                      -mat_terms_contrib$`Dim 3`),]
 n_top_terms = 10
 # mat coord terms
 mat_coord_col = res.ca$col$coord
 vect_index_top = c(order(-mat_terms_contrib$`Dim 1`)[1:n_top_terms],
                     order(-mat_terms_contrib$`Dim 2`)[1:n_top_terms],
                     order(-mat_terms_contrib$`Dim 3`)[1:n_top_terms])
 unique_vect_index_top = unique(vect_index_top)
 mat_coord_col_top = mat_coord_col[unique_vect_index_top, 1:2]
 #plot(mat_coord_col_top)
 # mat coord topics
 mat_coord_row = res.ca$row$coord
 mat_coord_row_ = mat_coord_row[,1:2]
 #points(mat_coord_row_, col= 'red')
 mat_coord_col_top = as.data.frame(mat_coord_col_top)
 mat_coord_row_ = as.data.frame(mat_coord_row_)
 mat_coord_col_top$type = "terms"
 mat_coord_row_$type = "topics"
 all_coord = rbind(mat_coord_col_top,mat_coord_row_ )
 all_coord$type = as.factor(all_coord$type)
 levels(all coord$type)<-c(1,2)</pre>
 set.seed(42)
 p <- ggplot(all coord, aes(`Dim 1`, `Dim 2`)) +</pre>
   geom_point(color =factor(all_coord$type)) +
   theme_classic(base_size = 5)
       + geom_label_repel(aes(label = rownames(all_coord),
                                fill = factor(type)), color = 'black', size = 3)
      p + theme(legend.position = "none")
     p + xlab("Dim1") + ylab("Dim2")
   <- p +scale_fill_manual(values=c('#FFCCCC',"#0066FF","#FF0000",'#009999'))</pre>
 р
                                relativity
            gravitational
                        logistic
                                     predictor
                  attribute
                           import
                                  database
             script
                         row
     score
               learning
     samsung
               intercept
     distributed
            prior
    travel
         significant
                                                                               geometry
                series
                                                                                  spatial
                                                                                         symmetry
                                                                     screen
                                                                             projection
                                                                                            spin
                                                                            lagrangian
                                                                                           anova
# 2. apply afc to TTM_answers_occ
 #data
 TT_occ_a = read.csv("TT_occ_a.csv")
 #head(TT_occ_a)
 #AFC
 res.ca = CA(TT_occ_a, ncp = 4, graph = FALSE)
 #contribution of any therm
 mat_terms_contrib = as.data.frame(res.ca$col$contrib)
 mat_terms_contrib_order = mat_terms_contrib[order(-mat_terms_contrib$`Dim 1`,
                                                      -mat_terms_contrib$`Dim 2`,
                                                      -mat_terms_contrib$`Dim 3`),]
 n_top_terms = 10
 # mat coord terms
 mat_coord_col = res.ca$col$coord
 vect_index_top = c(order(-mat_terms_contrib$`Dim 1`)[1:n_top_terms],
                     order(-mat_terms_contrib$`Dim 2`)[1:n_top_terms],
                     order(-mat_terms_contrib$`Dim 3`)[1:n_top_terms])
 unique_vect_index_top = unique(vect_index_top)
 mat_coord_col_top = mat_coord_col[unique_vect_index_top, 1:2]
 #plot(mat_coord_col_top)
 # mat coord topics
 mat_coord_row = res.ca$row$coord
 mat_coord_row_ = mat_coord_row[,1:2]
 #points(mat_coord_row_, col= 'red')
 mat_coord_col_top = as.data.frame(mat_coord_col_top)
 mat_coord_row_
                   = as.data.frame(mat_coord_row_)
 mat_coord_col_top$type = "terms"
 mat_coord_row_$type = "topics"
 all_coord = rbind(mat_coord_col_top,mat_coord_row_ )
 all_coord$type = as.factor(all_coord$type)
 levels(all_coord$type)<-c(1,2)</pre>
 set.seed(42)
 p <- ggplot(all_coord, aes(`Dim 1`, `Dim 2`)) +</pre>
   geom_point(color =factor(all_coord$type)) +
   theme_classic(base_size = 5)
 p = p + geom_label_repel(aes(label = rownames(all_coord),
                               fill = factor(type)), color = 'black', size = 3)
 p=p + theme(legend.position = "none")
 p = p + xlab("Dim1") + ylab("Dim2")
 p <- p +scale_fill_manual(values=c('#FFCCCC',"#0066FF","#FF0000",'#009999'))</pre>
 р
                               spacetime
                            orbit tensor
                                               magnetic
                                symmetry
                                         angular
                                                   gravitational
                                                                       mathbb
    residual
                                                                              mathcal
            intercept
                                                                      geometry
    likelihood
                    score
                                                                             spatial
         confidence
                                                                                     langle
                                                                                            nabla
                                                                                            higgs
                                                                              raster
                                                                                         relativity
# 3. apply afc to TTM_questions_tfidf
 #data
 TT_tfidf_q = read.csv("TT_tfidf_q.csv")
 #head(TT_tfidf_q)
 #AFC
 res.ca = CA(TT_tfidf_q, ncp = 4, graph = FALSE)
 #contribution of any therm
 mat_terms_contrib = as.data.frame(res.ca$col$contrib)
 mat_terms_contrib_order = mat_terms_contrib[order(-mat_terms_contrib$`Dim 1`,
                                                      -mat_terms_contrib$`Dim 2`,
                                                      -mat_terms_contrib$`Dim 3`),]
 n_top_terms = 10
 # mat coord terms
 mat_coord_col = res.ca$col$coord
 vect_index_top = c(order(-mat_terms_contrib$`Dim 1`)[1:n_top_terms],
                     order(-mat_terms_contrib$`Dim 2`)[1:n_top_terms],
                     order(-mat terms contrib$`Dim 3`)[1:n top terms])
 unique vect index top = unique(vect index top)
 mat_coord_col_top = mat_coord_col[unique_vect_index_top, 1:2]
 #plot(mat_coord_col_top)
 # mat coord topics
 mat_coord_row = res.ca$row$coord
 mat_coord_row_ = mat_coord_row[,1:2]
 #points(mat_coord_row_, col= 'red')
 mat_coord_col_top = as.data.frame(mat_coord_col_top)
 mat_coord_row_
                    = as.data.frame(mat_coord_row_)
 mat_coord_col_top$type = "terms"
 mat coord row $type = "topics"
 all_coord = rbind(mat_coord_col_top,mat_coord_row_ )
 all coord$type = as.factor(all coord$type)
 levels(all_coord$type)<-c(1,2)</pre>
 set.seed(42)
 p <- ggplot(all_coord, aes(`Dim 1`, `Dim 2`)) +</pre>
   geom point(color =factor(all coord$type)) +
   theme_classic(base_size = 5)
 p = p + geom_label_repel(aes(label = rownames(all_coord),
                               fill = factor(type)), color = 'black', size = 3)
     p + theme(legend.position = "none")
 p = p + xlab("Dim1") + ylab("Dim2")
 p <- p +scale_fill_manual(values=c('#FFCCCC',"#0066FF","#FF0000",'#009999'))</pre>
                         relativity -
                                             logistic
                           database radiation
                gravitational
                       attribute
                                              predictor
                                       moon
                                script
     notification
                    series
    distributed
                    prior
           intercept
                    significant
     travel
                                                                                geometry
           samsung
                                                                                        symmetry
                                                                                spatial
                                                                     screen
                                                                                      higgs
                                                                            projection
                                                                                             spin
                                                 Dim1
# 4. apply afc to TTM_answers_tfidf
 TT_tfidf_a = read.csv("TT_tfidf_a.csv")
 #head(TT tfidf a)
 #AFC
 res.ca = CA(TT_tfidf_a, ncp = 4, graph = FALSE)
 #contribution of any therm
 mat terms contrib = as.data.frame(res.ca$col$contrib)
 mat_terms_contrib_order = mat_terms_contrib[order(-mat_terms_contrib$`Dim 1`,
                                                  -mat_terms_contrib$`Dim 2`,
                                                      -mat_terms_contrib$`Dim 3`),]
 n_top_terms = 10
 # mat coord terms
 mat coord col = res.ca$col$coord
 vect_index_top = c(order(-mat_terms_contrib$`Dim 1`)[1:n_top_terms],
                     order(-mat_terms_contrib$`Dim 2`)[1:n_top_terms],
                     order(-mat_terms_contrib$`Dim 3`)[1:n_top_terms])
 unique_vect_index_top = unique(vect_index_top)
 mat_coord_col_top = mat_coord_col[unique_vect_index_top, 1:2]
 #plot(mat_coord_col_top)
 # mat coord topics
 mat_coord_row = res.ca$row$coord
 mat_coord_row_ = mat_coord_row[,1:2]
 #points(mat_coord_row_, col= 'red')
 mat_coord_col_top = as.data.frame(mat_coord_col_top)
                   = as.data.frame(mat_coord_row_)
 mat_coord_row_
 mat_coord_col_top$type = "terms"
 mat_coord_row_$type = "topics"
 all_coord = rbind(mat_coord_col_top,mat_coord_row_ )
 all_coord$type = as.factor(all_coord$type)
 levels(all_coord$type)<-c(1,2)</pre>
 set.seed(42)
 p <- ggplot(all_coord, aes(`Dim 1`, `Dim 2`)) +</pre>
   geom_point(color =factor(all_coord$type)) +
   theme_classic(base_size = 5)
 p = p + geom_label_repel(aes(label = rownames(all_coord),
                                fill = factor(type)), color = 'black', size = 3)
 p = p + theme(legend.position = "none")
 p = p + xlab("Dim1") + ylab("Dim2")
 p <- p + scale_fill_manual(values=c('#FFCCCC',"#0066FF","#FF0000",'#009999'))</pre>
 р
                           spacetime - hbar
                                            electric - gravitational
                                 tensor | gas
                                                magnetic
                                 symmetry
                                           angular
```

null

score

outlier

confidence

geometry

market

esri

relativity

setting

spatial

raster

openlayers

relativistic

validation

residual

deviation

likelihood