

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`),]

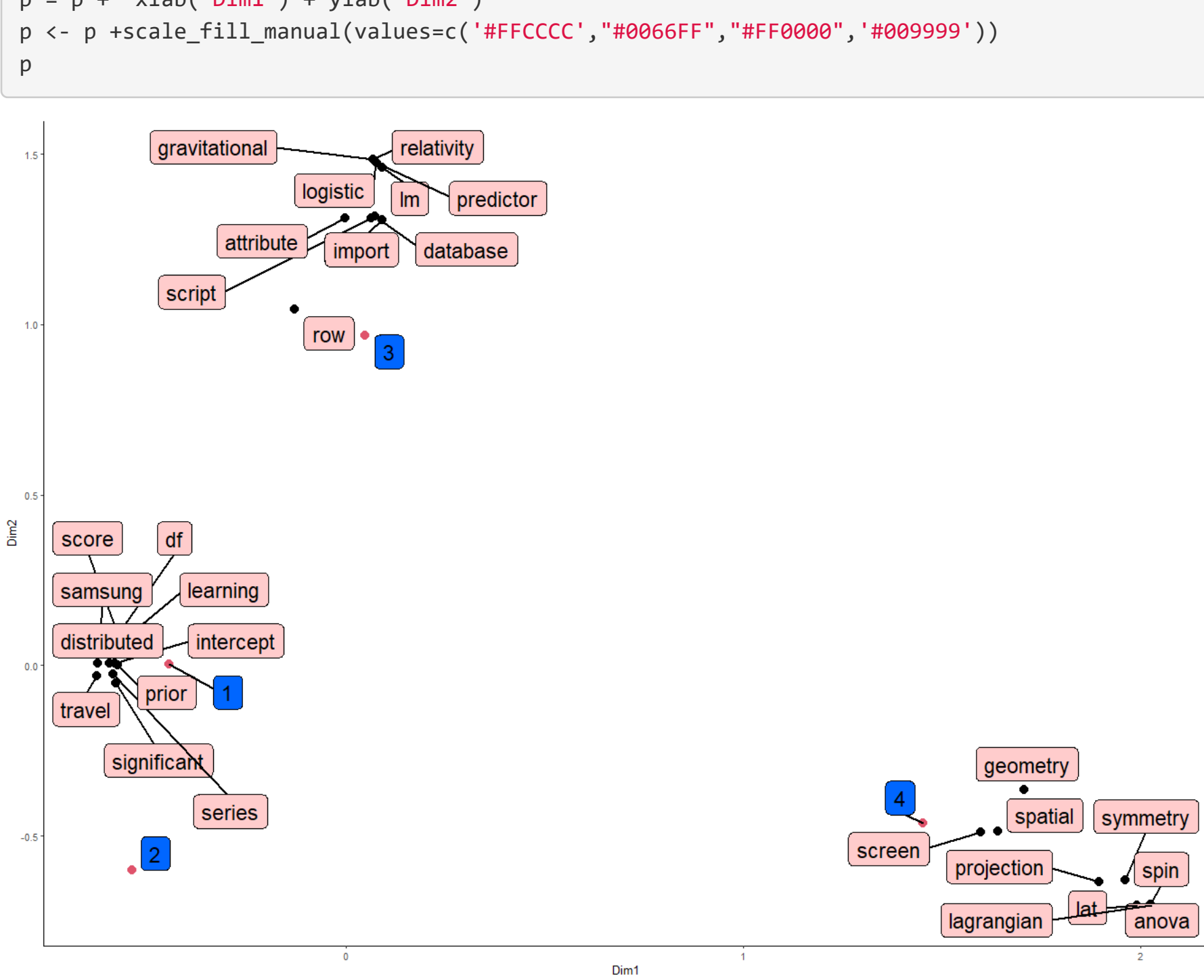
n_top_terms = 10
# mat coord terms
mat_coord_col = res.ca$col$coord
vect_index_top = c(order(-mat_terms_contrib$`Dim 1`)[1:n_top_terms],
                   order(-mat_terms_contrib$`Dim 2`)[1:n_top_terms],
                   order(-mat_terms_contrib$`Dim 3`)[1:n_top_terms])
unique_vect_index_top = unique(vect_index_top)
mat_coord_col_top = mat_coord_col[unique_vect_index_top, 1:2]
#plot(mat_coord_col_top)
# mat coord topics
mat_coord_row = res.ca$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)

set.seed(42)
p <- ggplot(all_coord, aes(`Dim 1`, `Dim 2`)) +
  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'))
p
```



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`),]

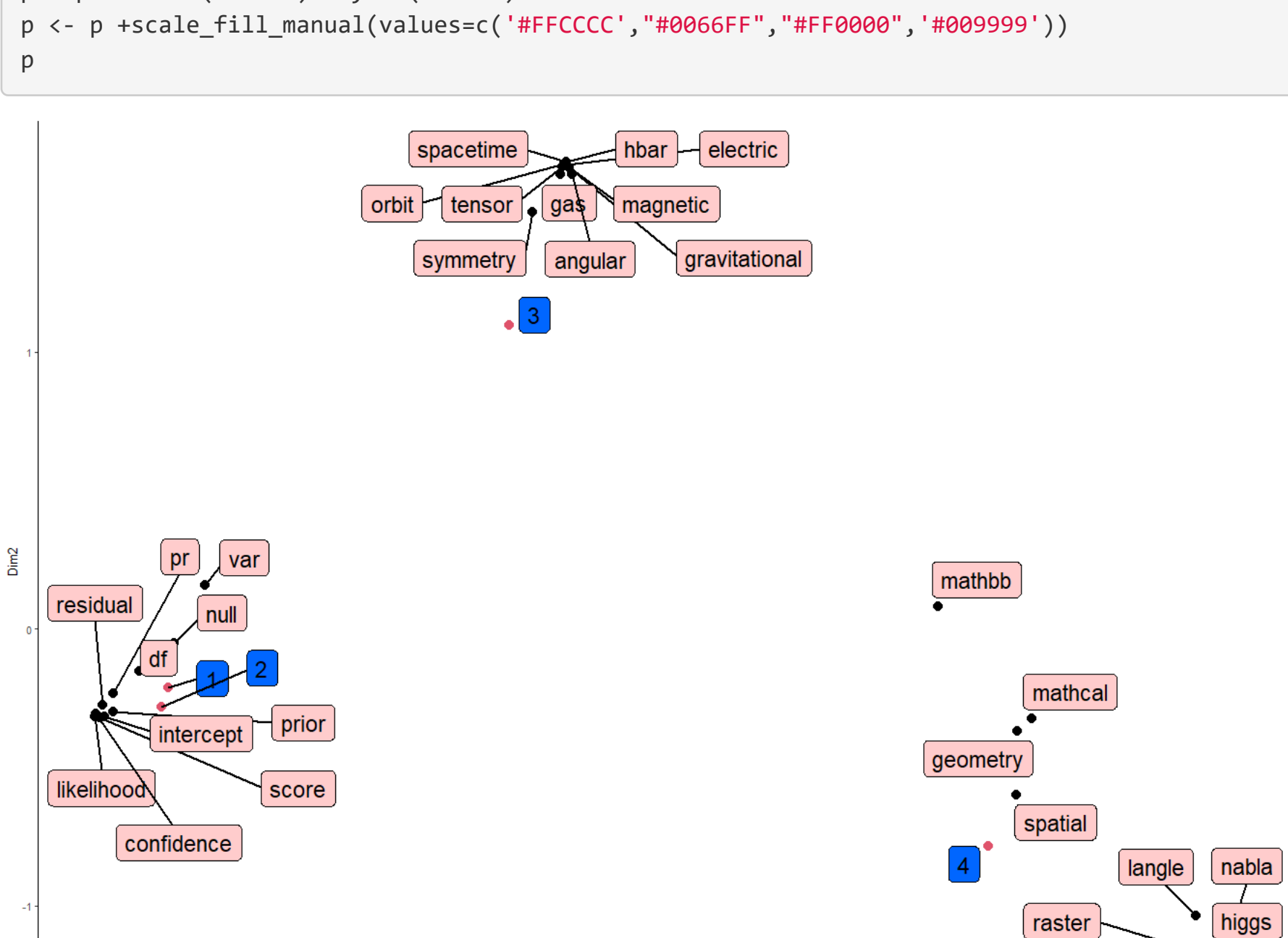
n_top_terms = 10
# mat coord terms
mat_coord_col = res.ca$col$coord
vect_index_top = c(order(-mat_terms_contrib$`Dim 1`)[1:n_top_terms],
                   order(-mat_terms_contrib$`Dim 2`)[1:n_top_terms],
                   order(-mat_terms_contrib$`Dim 3`)[1:n_top_terms])
unique_vect_index_top = unique(vect_index_top)
mat_coord_col_top = mat_coord_col[unique_vect_index_top, 1:2]
#plot(mat_coord_col_top)
# mat coord topics
mat_coord_row = res.ca$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)

set.seed(42)
p <- ggplot(all_coord, aes(`Dim 1`, `Dim 2`)) +
  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'))
p
```



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`),]

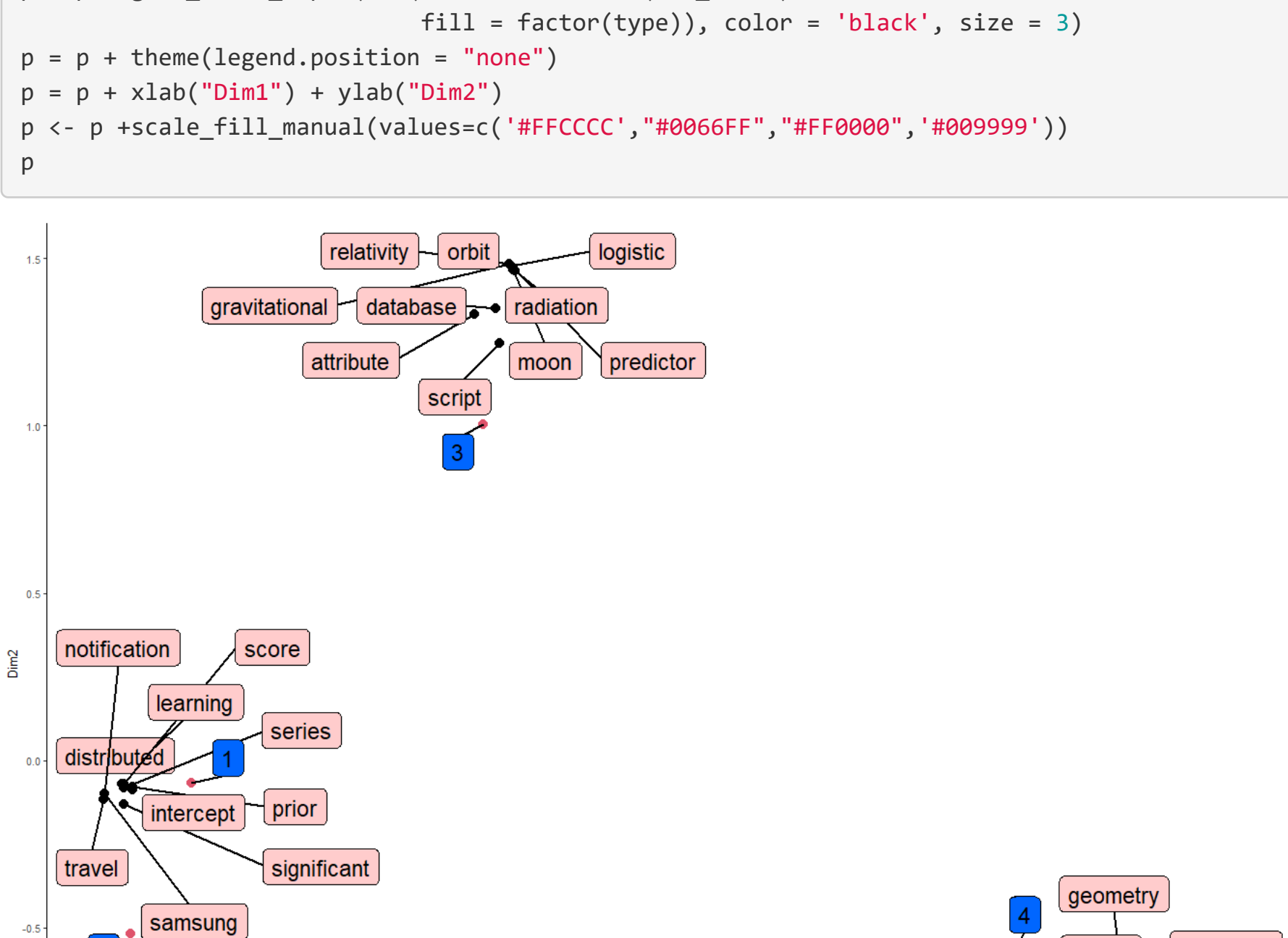
n_top_terms = 10
# mat coord terms
mat_coord_col = res.ca$col$coord
vect_index_top = c(order(-mat_terms_contrib$`Dim 1`)[1:n_top_terms],
                   order(-mat_terms_contrib$`Dim 2`)[1:n_top_terms],
                   order(-mat_terms_contrib$`Dim 3`)[1:n_top_terms])
unique_vect_index_top = unique(vect_index_top)
mat_coord_col_top = mat_coord_col[unique_vect_index_top, 1:2]
#plot(mat_coord_col_top)
# mat coord topics
mat_coord_row = res.ca$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)

set.seed(42)
p <- ggplot(all_coord, aes(`Dim 1`, `Dim 2`)) +
  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'))
p
```



4. apply afc to TTM_answers_tfidf

```
#data
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)
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)

set.seed(42)
p <- ggplot(all_coord, aes(`Dim 1`, `Dim 2`)) +
  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'))
p
```

