

## Rapport du TP AFCM

### *Objectif*

**Déterminer des associations pertinentes entre les questions et les réponses qui aideront à comprendre l'état des lieux (dataset :2SLTPGO).**

**Réalisé par :**

- IDRES Aghiles (SIT4)

**Encadré par :** Mme HAMDAD et Mme IMLOUL

**Année universitaire : 2019-2020**

## Table des matières

|   |    |
|---|----|
| 1. Introduction :   | 3  |
| 2. Importation des données :                                    | 3  |
| a) Préparation du fichier « 2SLTPGO.csv » :                     | 3  |
| b) Importation du fichier dans le logiciel R :                  | 3  |
| 3. Etude des statistiques des données :                         | 4  |
| a) Fréquences :   | 4  |
| b) Représentation avec diagramme circulaire :                   | 5  |
| c) Histogramme :  | 5  |
| 4. Visualisation de la fréquence des catégories des variables : | 6  |
| 5. Transformation en tableau disjonctif :                       | 7  |
| 6. Etablissement de l'AFCM :                                    | 8  |
| 7. Etude du tableau de valeurs propres :                        | 9  |
| 8. Représentation du biplot individus-variables :               | 10 |
| 9. Etude du tableau de contributions :                          | 11 |
| a) Contribution des individus :                                 | 11 |
| b) Contribution des variables :                                 | 12 |
| 10. Création des visualisations :                               | 12 |
| a) Contribution des individus sur l'axe 1 :                     | 12 |
| b) Contribution des individus sur l'axe 2 :                     | 13 |
| c) Contribution des variables sur l'axe 1 :                     | 14 |
| d) Contribution des variables sur l'axe 2 :                     | 14 |
| 11. Ressort des associations entre modalités :                  | 15 |
| 12. Interprétation des résultats :                              | 16 |
| 13. Les questions les mieux représentées par l'AFCM :           | 16 |
| 14. Etablissement d'un tableau de contingence :                 | 17 |
| 15. Etablissement de l'AFC :                                    | 17 |
| a) Visualisation:   | 17 |
| b) Interprétation des résultats :                               | 18 |
| 16. Conclusion :  | 18 |

## 1. Introduction :

L'AFCM est une méthode d'analyse de données qui consiste à appliquer l'AFC à l'étude de tableau qui sont résultats d'observations d'individus sur plus de deux variables qualitatives mis sous forme d'un tableau disjonctif complet, qu'on décrira plus tard. Elle est bien adaptée à l'étude des questionnaires. Un questionnaire est la donnée de p variables qualitatives ayant chacune un nombre de modalités et le but est d'étudier la relation entre ces variables, entre les modalités et éventuellement entre les individus. Dans ce TP nous allons étudier un cas réel qui est l'évaluation de l'enseignement à l'ESI.

## 2. Importation des données :

### a) Préparation du fichier « 2SLTPGO.csv » :

J'ai télécharger le fichier « Evaluation 2CS SIL 2018-2019 S1.xls » envoyé par mail, puis j'ai sélectionné la feuille « TPGO », ensuite j'ai remplacé les signe « --, -, +, ++ » par « M (mauvais), AB (assez bien), B (bien), TB (très bien) » respectivement, j'ai supprimé le tableau d'observations et enfin j'ai enregistré le fichier au format csv nommé « 2SLTPGO ».

### b) Importation du fichier dans le logiciel R :

```
> setwd("C:/Users/HP 15/Documents/anad")
> data = read.csv(file="2SLTPGO.csv", sep=",")
> data
```

|    | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 |
|----|----|----|----|----|----|----|----|----|----|-----|
| 1  | B  | B  | B  | TB | TB | TB | B  | B  | AB | B   |
| 2  | B  | B  | AB | B  | TB | TB | B  | TB | B  | M   |
| 3  | B  | TB | TB | TB | TB | TB | TB | TB | B  | TB  |
| 4  | AB | B  | B  | B  | B  | B  | TB | AB | AB | B   |
| 5  | AB | TB | B  | B  | B  | B  | B  | TB | AB | AB  |
| 6  | AB | B  | AB | AB | B  | TB | TB | B  | B  | B   |
| 7  | B  | B  | B  | B  | B  | B  | B  | B  | B  | B   |
| 8  | B  | B  | B  | AB | AB | TB | TB | B  | B  | B   |
| 9  | B  | TB | AB | B  | B  | B  | TB | TB | B  | B   |
| 10 | B  | B  | B  | B  | TB | TB | B  | B  | B  | B   |
| 11 | B  | B  | B  | AB | B  | B  | AB | B  | B  | AB  |
| 12 | B  | TB | B  | B  | TB | TB | B  | TB | AB | AB  |
| 13 | TB | TB | B  | B  | B  | B  | B  | B  | B  | B   |
| 14 | B  | B  | AB | B  | B  | AB | AB | B  | AB | B   |
| 15 | B  | B  | B  | AB | AB | B  | B  | TB | B  | B   |
| 16 | B  | B  | B  | AB | AB | B  | B  | B  | B  | B   |

Sachant que :

Q1 : Les objectifs et le programme de l'UE étaient clairs ?

Q2 : Le programme réalisé correspond au programme annoncé ?

Q3 : Je maîtrisais les antécédents et prérequis nécessaires ?

Q4 : La matière est bien exposée et capte mon attention ?

Q5 : Je suis bien encadré dans la résolution des exercices ?

Q6 : Les TD/TP/Projet étaient utiles pour une bonne assimilation du cours ?

Q7 : Les supports pédagogiques (transparents, polycopié, séries, ...) étaient disponibles et utiles pour l'apprentissage ?

Q8 : Le volume horaire prévu dans l'UE a été respecté ?

Q9 : Le contenu des contrôles et évaluations correspondait aux aspects traités en Cours/TD/TP ?

Q10 : La quantité de travail demandée est adaptée à l'UE ?

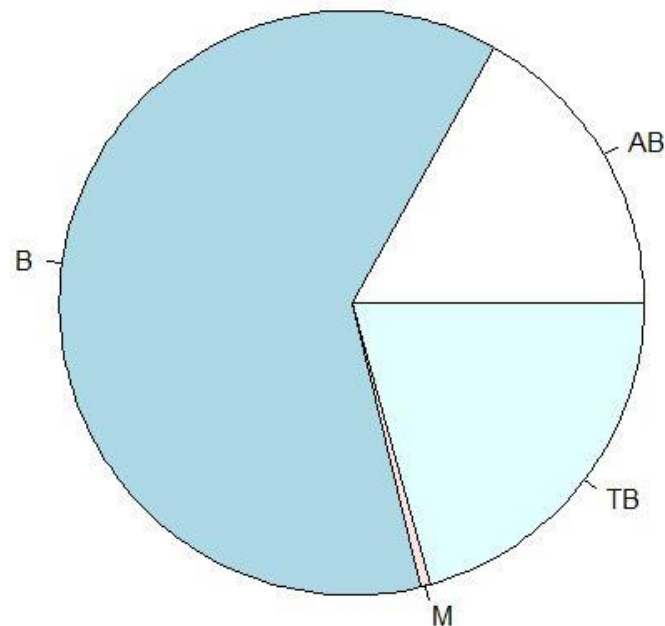
### **3. Etude des statistiques des données :**

#### **a) Fréquences :**

```
> frequence=as.matrix(data)
> prop.table(table(frequence))
frequence
      AB      B      M      TB
0.16875 0.61875 0.00625 0.20625
```

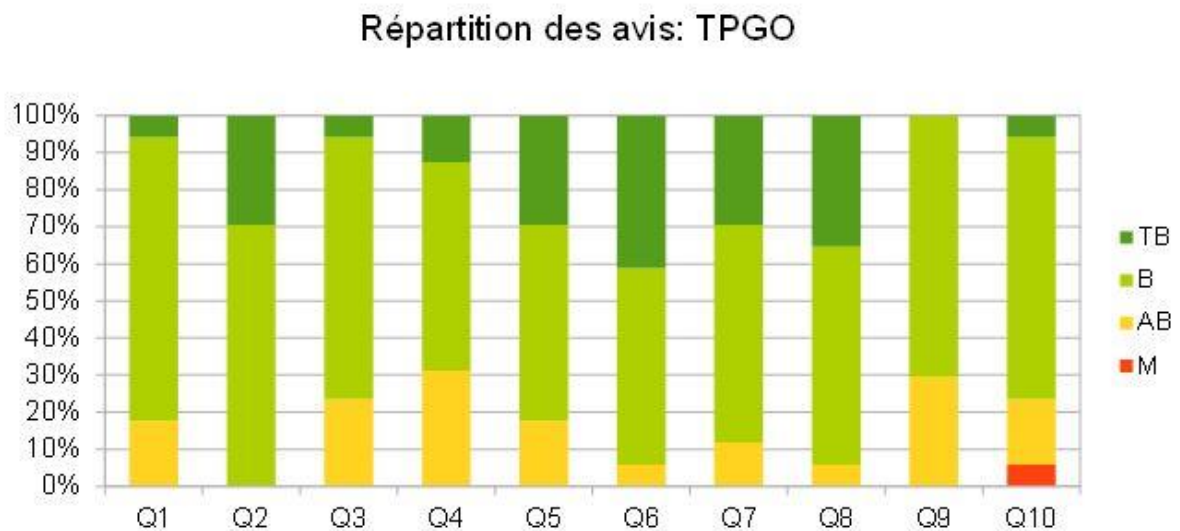
- 0.625% des réponses sont M (mauvais)
- 16.875% des réponses sont AB (assez bien)
- 61.875% des réponses sont B (bien)
- 20.625% des réponses sont TB (très bien)

**b) Représentation avec diagramme circulaire :**



On voit que la plupart des réponses sont B (bien), et la réponse M (mauvais) n'est presque jamais utilisée.

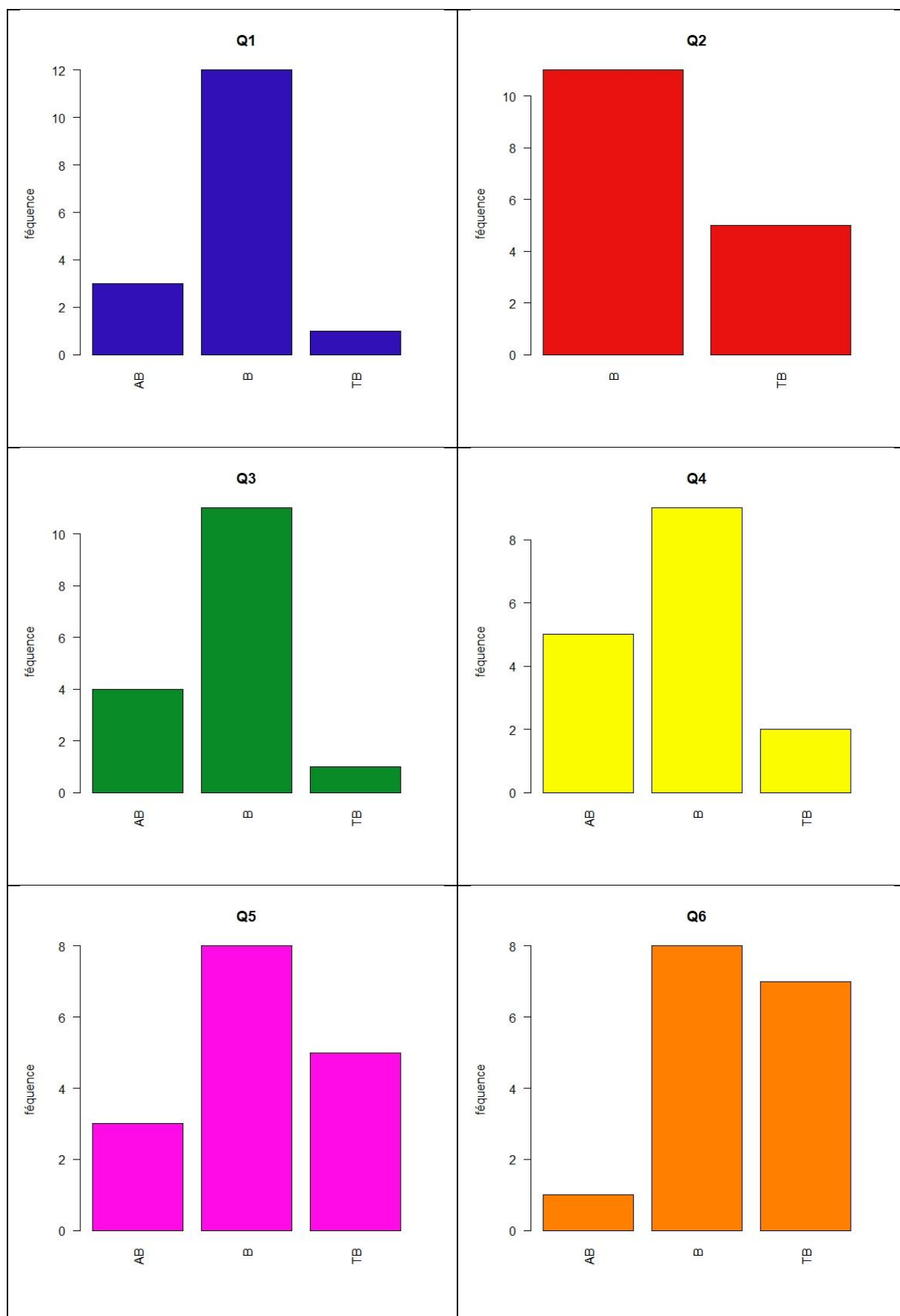
**c) Histogramme :**

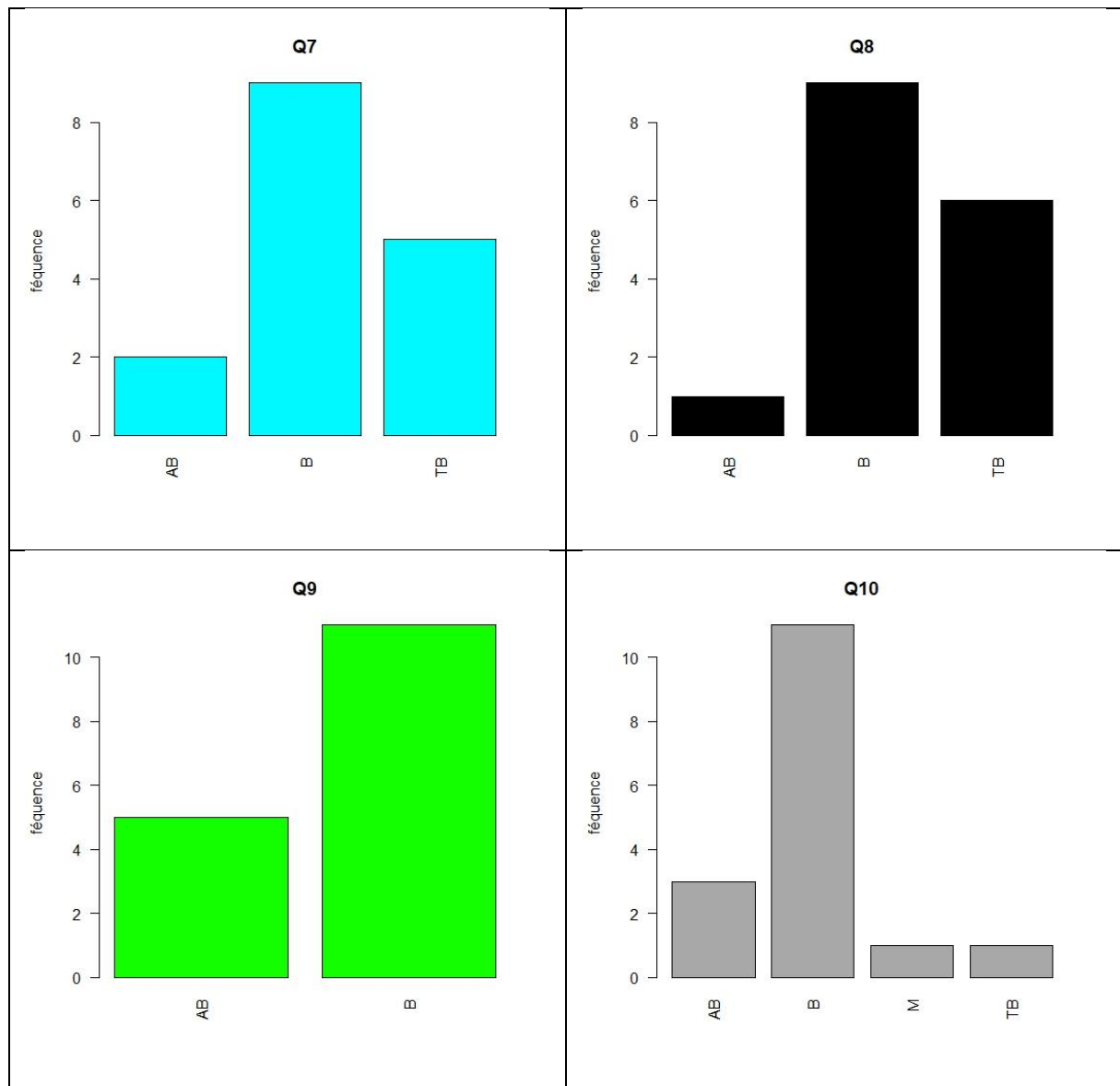


On remarque que:

- Toutes les questions ont la réponse B (bien) majoritaire.
- La question 2 n'a aucune réponse M ou AB.
- Seule la question 10 a une réponse M.

#### 4. Visualisation de la fréquence des catégories des variables :





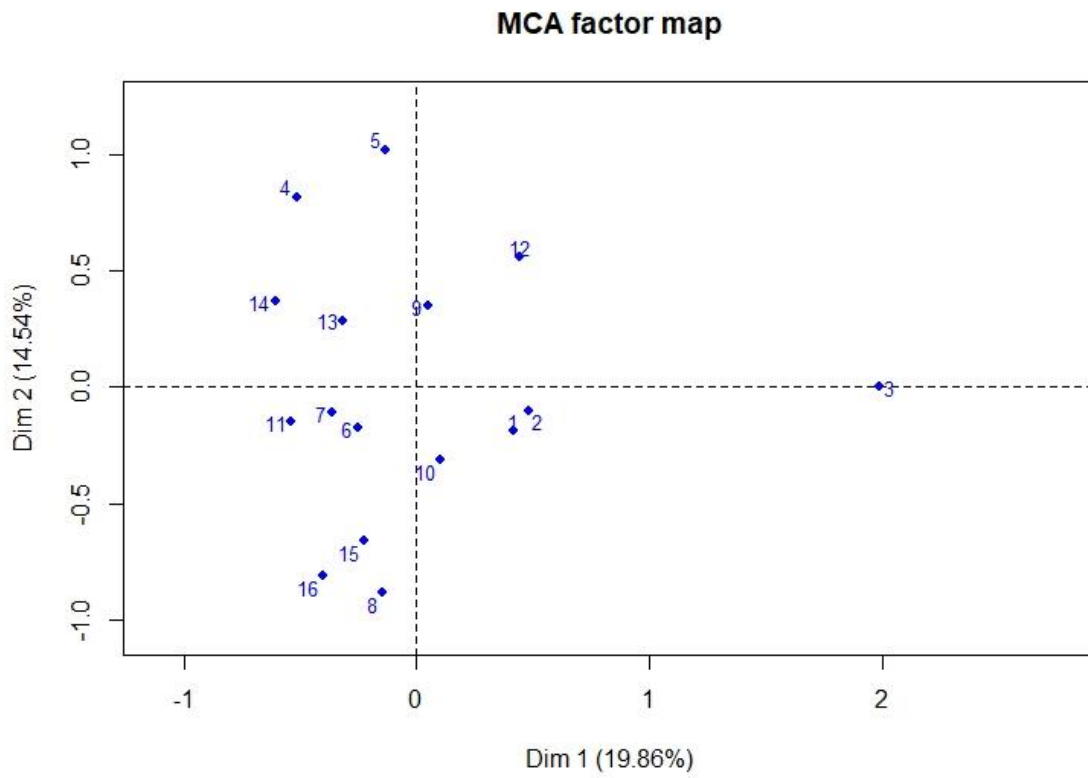
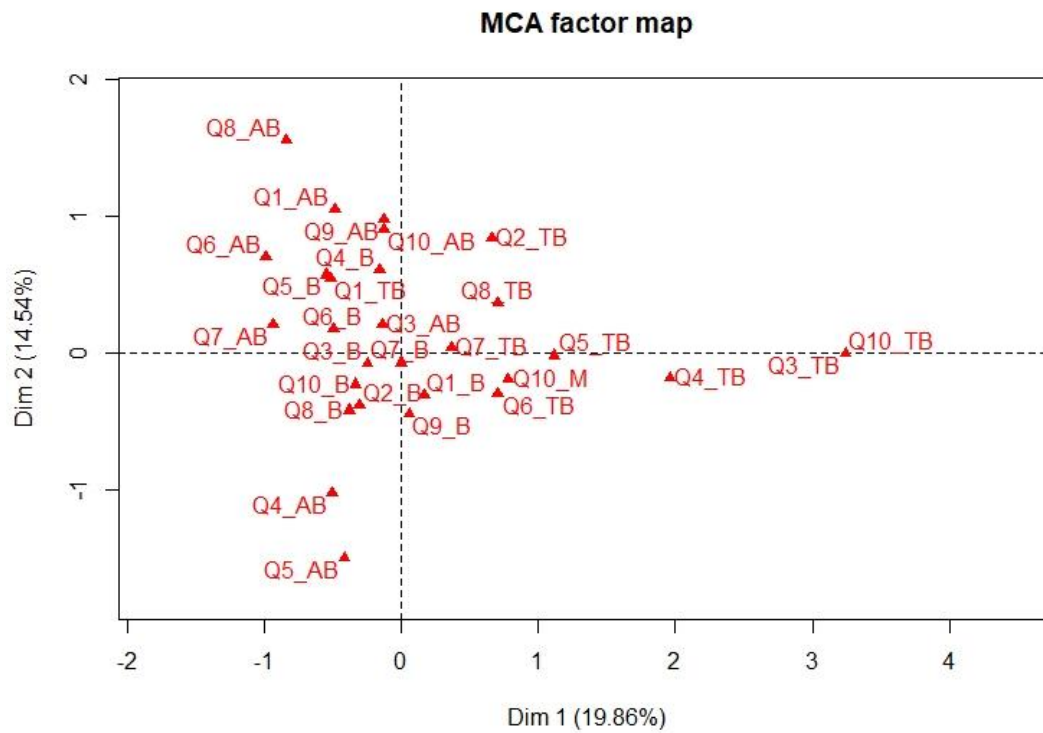
## 5. Transformation en tableau disjonctif :

```
> library(FactoMineR)
> tab_disj <- tab.disjonctif(data)
> tab_disj
```

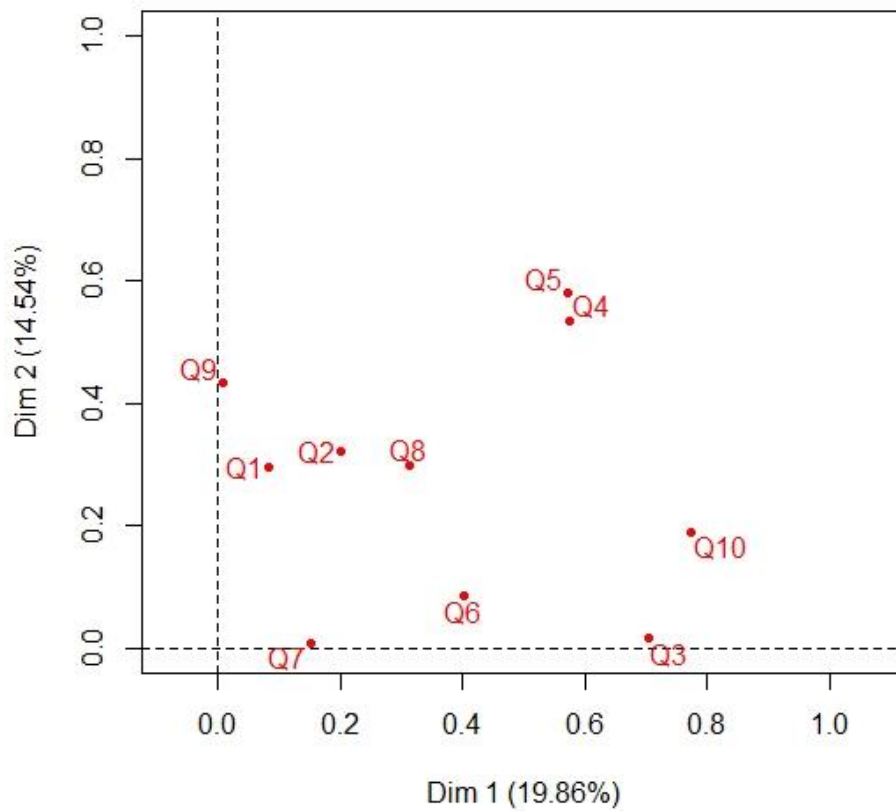
|    | AB | B | TB | B | TB | AB | B | TB | AB | B | TB | AB | B | TB | AB | B | TB | AB | B | TB | AB | B | AB | B | M | TB |
|----|----|---|----|---|----|----|---|----|----|---|----|----|---|----|----|---|----|----|---|----|----|---|----|---|---|----|
| 1  | 0  | 1 | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 0 | 1  | 0  | 0 | 1  | 0  | 0 | 1  | 0  | 0 | 1  | 0  | 1 | 0  | 1 | 0 | 0  |
| 2  | 0  | 1 | 0  | 1 | 0  | 1  | 0 | 0  | 0  | 1 | 0  | 0  | 0 | 1  | 0  | 0 | 1  | 0  | 1 | 0  | 0  | 0 | 1  | 0 | 0 | 0  |
| 3  | 0  | 1 | 0  | 0 | 1  | 0  | 0 | 1  | 0  | 0 | 1  | 0  | 0 | 1  | 0  | 0 | 1  | 0  | 0 | 1  | 0  | 0 | 0  | 0 | 1 | 0  |
| 4  | 1  | 0 | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 0 | 1  | 1  | 0 | 0  | 1 | 0 | 0  |
| 5  | 1  | 0 | 0  | 0 | 1  | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 0 | 1  | 1 | 0 | 0  |
| 6  | 1  | 0 | 0  | 1 | 0  | 1  | 0 | 0  | 1  | 0 | 0  | 0  | 1 | 0  | 0  | 0 | 1  | 0  | 0 | 1  | 0  | 1 | 0  | 0 | 1 | 0  |
| 7  | 0  | 1 | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 0 | 1 | 0  |
| 8  | 0  | 1 | 0  | 1 | 0  | 0  | 1 | 0  | 1  | 0 | 0  | 1  | 0 | 0  | 0  | 0 | 1  | 0  | 0 | 1  | 0  | 1 | 0  | 0 | 1 | 0  |
| 9  | 0  | 1 | 0  | 0 | 1  | 1  | 0 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 0 | 1  | 0  | 0 | 1  | 0 | 1 | 0  |
| 10 | 0  | 1 | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 0 | 1  | 0  | 0 | 1  | 0  | 1 | 0  | 0  | 1 | 0  | 1 | 0 | 0  |
| 11 | 0  | 1 | 0  | 1 | 0  | 0  | 1 | 0  | 1  | 0 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 1  | 0 | 0  | 0  | 1 | 0  | 1 | 0 | 0  |
| 12 | 0  | 1 | 0  | 0 | 1  | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 0 | 1  | 0  | 0 | 1  | 0  | 1 | 0  | 0  | 0 | 1  | 1 | 0 | 0  |
| 13 | 0  | 0 | 1  | 0 | 1  | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 1 | 0 | 0  |
| 14 | 0  | 1 | 0  | 1 | 0  | 1  | 0 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 1  | 0 | 0  | 1  | 0 | 0  | 1  | 0 | 1  | 0 | 1 | 0  |
| 15 | 0  | 1 | 0  | 1 | 0  | 0  | 1 | 0  | 1  | 0 | 0  | 1  | 0 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 0 | 1  | 0 | 1 | 0  |
| 16 | 0  | 1 | 0  | 1 | 0  | 0  | 1 | 0  | 1  | 0 | 0  | 1  | 0 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 0  | 1 | 0  | 1 | 0 | 0  |

## 6. Etablissement de l'AFCM :

```
> afcm = MCA(data)
```



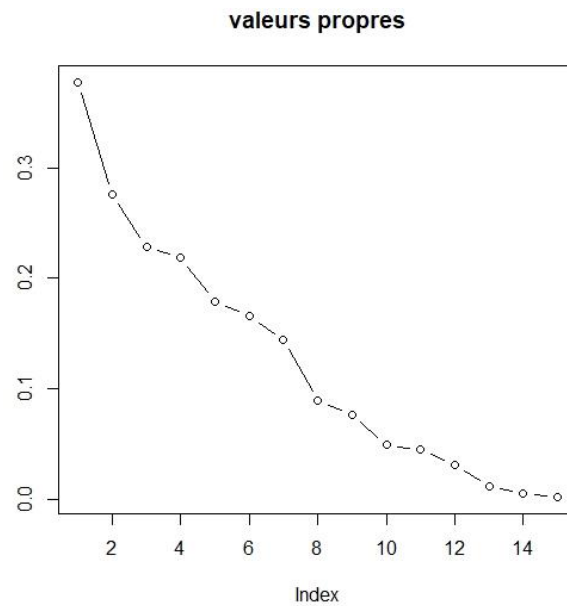




## 7. Etude du tableau de valeurs propres :

```
> afcm$eig
```

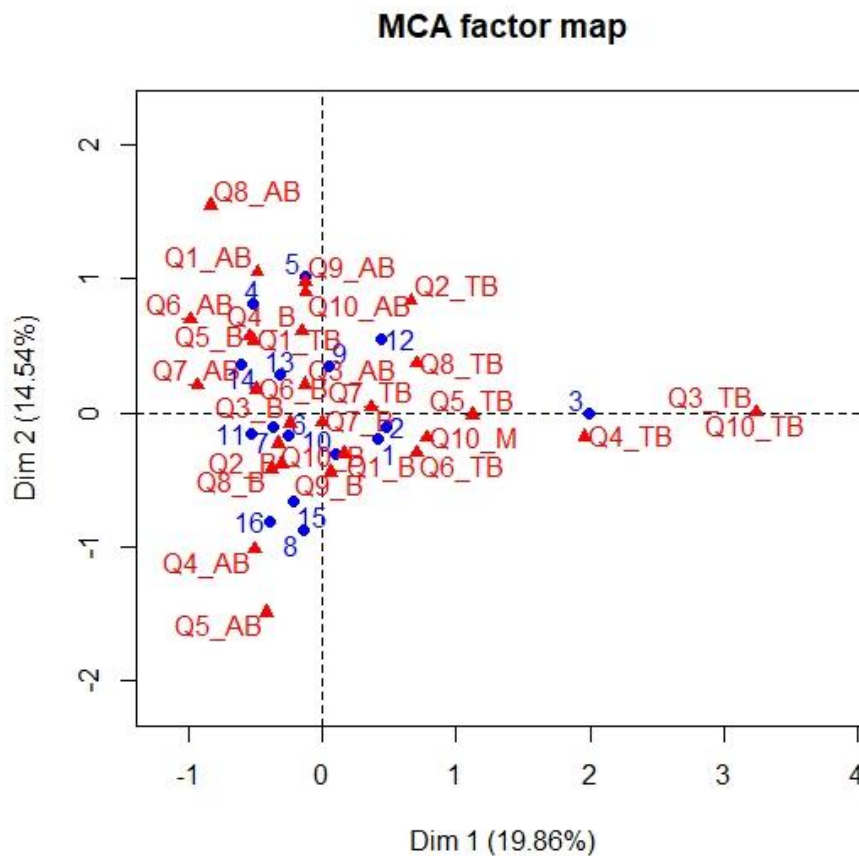
|        | eigenvalue  | percentage of variance | cumulative percentage of variance |
|--------|-------------|------------------------|-----------------------------------|
| dim 1  | 0.377294785 | 19.8576202             | 19.85762                          |
| dim 2  | 0.276177107 | 14.5356372             | 34.39326                          |
| dim 3  | 0.228999299 | 12.0525947             | 46.44585                          |
| dim 4  | 0.218818663 | 11.5167717             | 57.96262                          |
| dim 5  | 0.179249093 | 9.4341628              | 67.39679                          |
| dim 6  | 0.165563352 | 8.7138607              | 76.11065                          |
| dim 7  | 0.144329967 | 7.5963140              | 83.70696                          |
| dim 8  | 0.089091755 | 4.6890398              | 88.39600                          |
| dim 9  | 0.076732440 | 4.0385495              | 92.43455                          |
| dim 10 | 0.049177870 | 2.5883090              | 95.02286                          |
| dim 11 | 0.044642974 | 2.3496302              | 97.37249                          |
| dim 12 | 0.030802741 | 1.6211969              | 98.99369                          |
| dim 13 | 0.011640988 | 0.6126836              | 99.60637                          |
| dim 14 | 0.005230800 | 0.2753052              | 99.88168                          |
| dim 15 | 0.002248167 | 0.1183246              | 100.00000                         |



Pour le premier plan principal l'inertie cumulée est 34.39%, ce qui ne suffit pas pour établir une bonne étude. L'idéal c'est d'atteindre 80% d'inertie cumulée donc prendre les 7 premières valeurs propres.

## 8. Représentation du biplot individus-variables :

```
> plot(afcm)
```



## 9. Etude du tableau de contributions :

### a) Contribution des individus :

```
> afcm$ind$contrib
```

|    | Dim 1       | Dim 2        | Dim 3       | Dim 4       | Dim 5       |
|----|-------------|--------------|-------------|-------------|-------------|
| 1  | 2.89163082  | 8.215261e-01 | 1.20469324  | 1.17459123  | 0.13149679  |
| 2  | 3.81791573  | 2.224645e-01 | 14.30744048 | 14.94724787 | 31.32271539 |
| 3  | 65.64484812 | 2.339286e-06 | 0.40810664  | 14.12737738 | 6.43149773  |
| 4  | 4.36548493  | 1.502002e+01 | 5.45731380  | 18.35042297 | 15.75774609 |
| 5  | 0.28321781  | 2.335117e+01 | 5.90501717  | 3.23279993  | 0.18601930  |
| 6  | 1.05368629  | 6.872449e-01 | 0.35619742  | 14.23397856 | 8.57994448  |
| 7  | 2.17530752  | 2.700521e-01 | 0.36370122  | 0.67087855  | 0.36110154  |
| 8  | 0.34072158  | 1.751366e+01 | 0.88123022  | 2.89514854  | 0.61362723  |
| 9  | 0.04000971  | 2.740647e+00 | 0.01631056  | 1.18047714  | 0.90069212  |
| 10 | 0.17796314  | 2.185679e+00 | 0.51001337  | 3.87093738  | 0.55069386  |
| 11 | 4.78711185  | 5.174348e-01 | 0.41558243  | 0.03841914  | 15.44304891 |
| 12 | 3.22509568  | 6.990940e+00 | 0.06037269  | 16.52273665 | 0.20427106  |
| 13 | 1.65504939  | 1.832614e+00 | 8.88648610  | 3.70537153  | 9.02954754  |
| 14 | 6.09758208  | 3.065329e+00 | 53.71159888 | 4.07804096  | 10.09176096 |
| 15 | 0.81333184  | 9.902003e+00 | 4.38802731  | 0.82067593  | 0.01909174  |
| 16 | 2.63104352  | 1.487921e+01 | 3.12790849  | 0.15089625  | 0.37674526  |

Les individus qui ont le plus contribué sont ceux qui ont une contribution absolue  $> (1 / \text{nombre d'individus}) 1/16 = 6,25\%$

**L'axe 1 :** seul l'individu 3 contribue ( $65,64\% > 6,25\%$ ) positivement (sa projection sur l'axe 1 est positive)

**L'axe 2 :** est un axe d'opposition, il oppose les individus 4, 5 et 12 positifs (sa projection est positive) aux individus 8, 15 et 16 négatifs (leurs projections sont négatives).



## b) Contribution des variables :

```
> afcm$var$contrib
```

|        | Dim 1        | Dim 2        | Dim 3        | Dim 4        | Dim 5       |
|--------|--------------|--------------|--------------|--------------|-------------|
| Q1_AB  | 1.175761e+00 | 7.492365e+00 | 2.530282e+00 | 5.966776359  | 7.77837611  |
| Q1_B   | 5.378121e-01 | 2.572054e+00 | 1.860521e+00 | 0.715207752  | 0.55737901  |
| Q1_TB  | 4.386621e-01 | 6.635648e-01 | 3.880573e+00 | 1.693352607  | 5.03742998  |
| Q2_B   | 1.651658e+00 | 3.620185e+00 | 1.637908e+00 | 0.359740176  | 1.52035997  |
| Q2_TB  | 3.633648e+00 | 7.964407e+00 | 3.603397e+00 | 0.791428388  | 3.34479193  |
| Q3_AB  | 1.193069e-01 | 4.013453e-01 | 1.464052e+01 | 1.036863553  | 5.53228792  |
| Q3_B   | 1.101181e+00 | 1.461559e-01 | 4.752652e+00 | 1.904810866  | 0.71780460  |
| Q3_TB  | 1.739882e+01 | 8.470240e-07 | 1.782131e-01 | 6.456203141  | 3.58802248  |
| Q4_AB  | 2.118679e+00 | 1.174902e+01 | 1.107248e+00 | 1.750211367  | 0.05356746  |
| Q4_B   | 3.567630e-01 | 7.487551e+00 | 4.668388e-01 | 2.524356044  | 0.79958635  |
| Q4_TB  | 1.273428e+01 | 1.482302e-01 | 4.595085e-02 | 1.634880657  | 2.34373757  |
| Q5_AB  | 8.532013e-01 | 1.511041e+01 | 3.356643e+00 | 0.025252222  | 0.01760902  |
| Q5_B   | 3.912217e+00 | 5.882006e+00 | 4.657901e-04 | 2.654252631  | 0.75750924  |
| Q5_TB  | 1.035163e+01 | 3.220375e-03 | 1.937247e+00 | 4.769280863  | 0.99625798  |
| Q6_AB  | 1.616132e+00 | 1.109914e+00 | 2.345492e+01 | 1.863662318  | 5.63002067  |
| Q6_B   | 3.243691e+00 | 5.462396e-01 | 7.467866e+00 | 0.004168267  | 0.86595193  |
| Q6_TB  | 5.788221e+00 | 1.412069e+00 | 1.190126e+00 | 0.199775247  | 3.57829304  |
| Q7_AB  | 2.874438e+00 | 1.926207e-01 | 1.388134e+01 | 1.121500205  | 14.08726622 |
| Q7_B   | 8.805163e-05 | 9.667320e-02 | 9.952080e-01 | 14.363269381 | 0.06340946  |
| Q7_TB  | 1.122935e+00 | 1.948019e-02 | 1.036241e+00 | 19.491301552 | 4.14510906  |
| Q8_AB  | 1.157049e+00 | 5.438546e+00 | 2.383114e+00 | 8.386132489  | 8.79097676  |
| Q8_B   | 2.142049e+00 | 3.547721e+00 | 8.118948e-01 | 0.115168303  | 3.24399698  |
| Q8_TB  | 4.980227e+00 | 1.835462e+00 | 2.240434e-01 | 2.553200913  | 0.99094272  |
| Q9_AB  | 1.347800e-01 | 1.077212e+01 | 1.018289e+00 | 0.037841635  | 0.02290040  |
| Q9_B   | 6.126364e-02 | 4.896419e+00 | 4.628587e-01 | 0.017200743  | 0.01040927  |
| Q10_AB | 7.547434e-02 | 5.510624e+00 | 6.004777e-01 | 4.891829933  | 4.30782926  |
| Q10_B  | 2.009286e+00 | 1.301039e+00 | 4.914268e-02 | 1.385245770  | 0.15574550  |
| Q10_M  | 1.011919e+00 | 8.055138e-02 | 6.247810e+00 | 6.830883475  | 17.47440666 |
| Q10_TB | 1.739882e+01 | 8.470240e-07 | 1.782131e-01 | 6.456203141  | 3.58802248  |

Les variables qui ont le plus contribué sont ceux qui ont une contribution absolue  $> (1 / \text{nombre de variables}) 1/29 = 3,45\%$

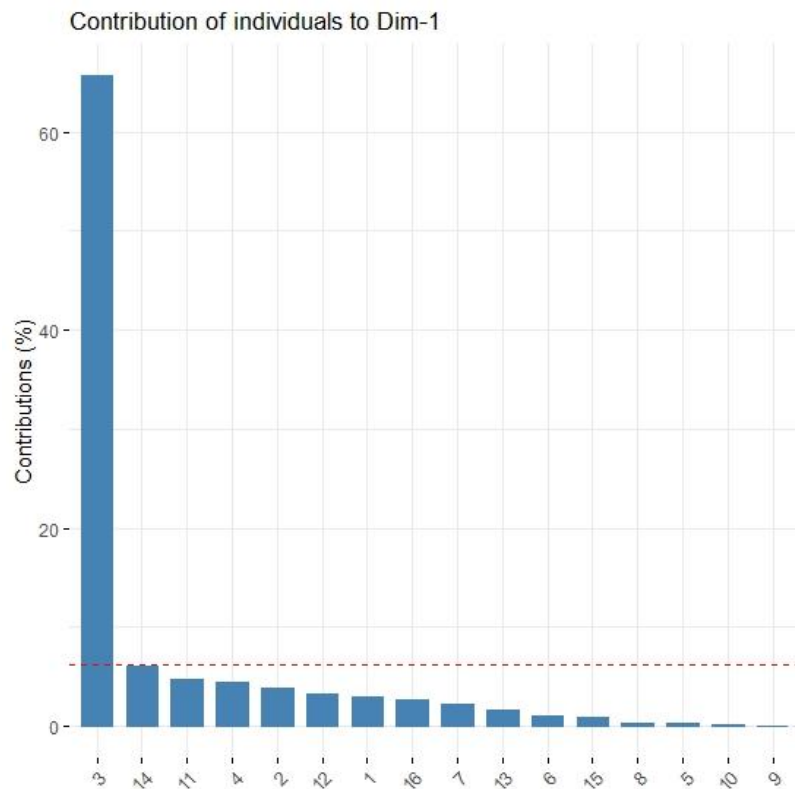
**L'axe 1 :** est un axe d'opposition, il oppose les variables Q2\_TB, Q3\_TB, Q4\_TB, Q5\_TB, Q6\_TB, Q8\_TB et Q10\_TB positives (leurs projections sont positives) à la variable Q5\_B négative (sa projection est négative).

**L'axe 2 :** est un axe d'opposition, il oppose les variables Q1\_AB, Q2\_TB, Q4\_B, Q5\_B, Q8\_AB, Q9\_AB et Q10\_AB positives (leurs projections sont positives) aux variables Q2\_B, Q4\_AB, Q5\_AB, Q8\_B et Q9\_B négatives (leurs projections sont négatives).

## 10.Création des visualisations :

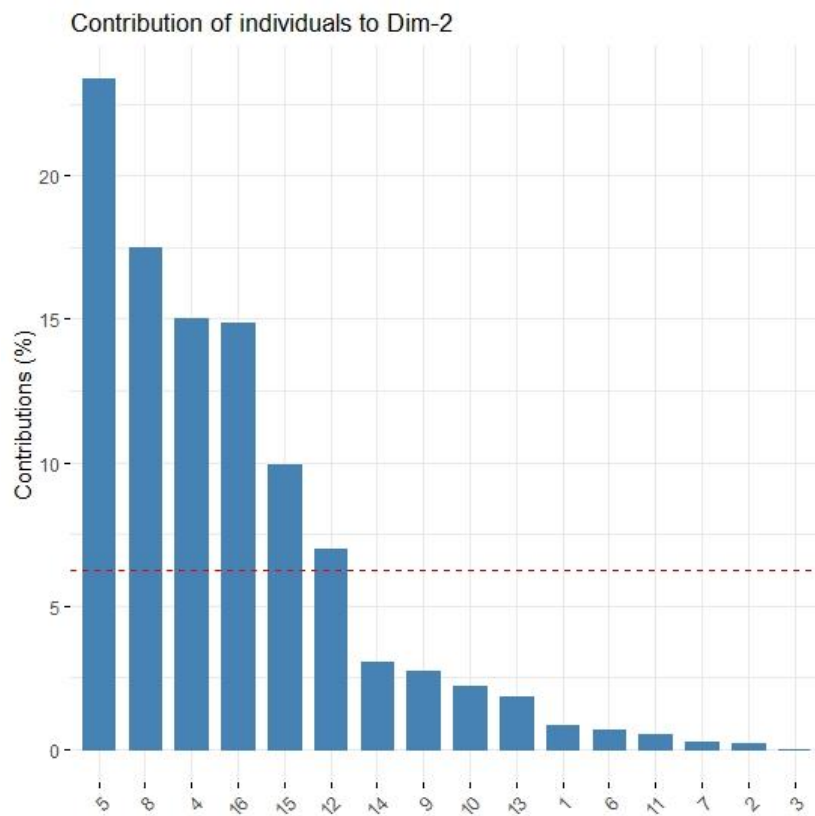
### a) Contribution des individus sur l'axe 1 :

```
> library(factoextra)
> fviz_contrib(afcm, choice = "ind", axes = 1, top = 16)
```



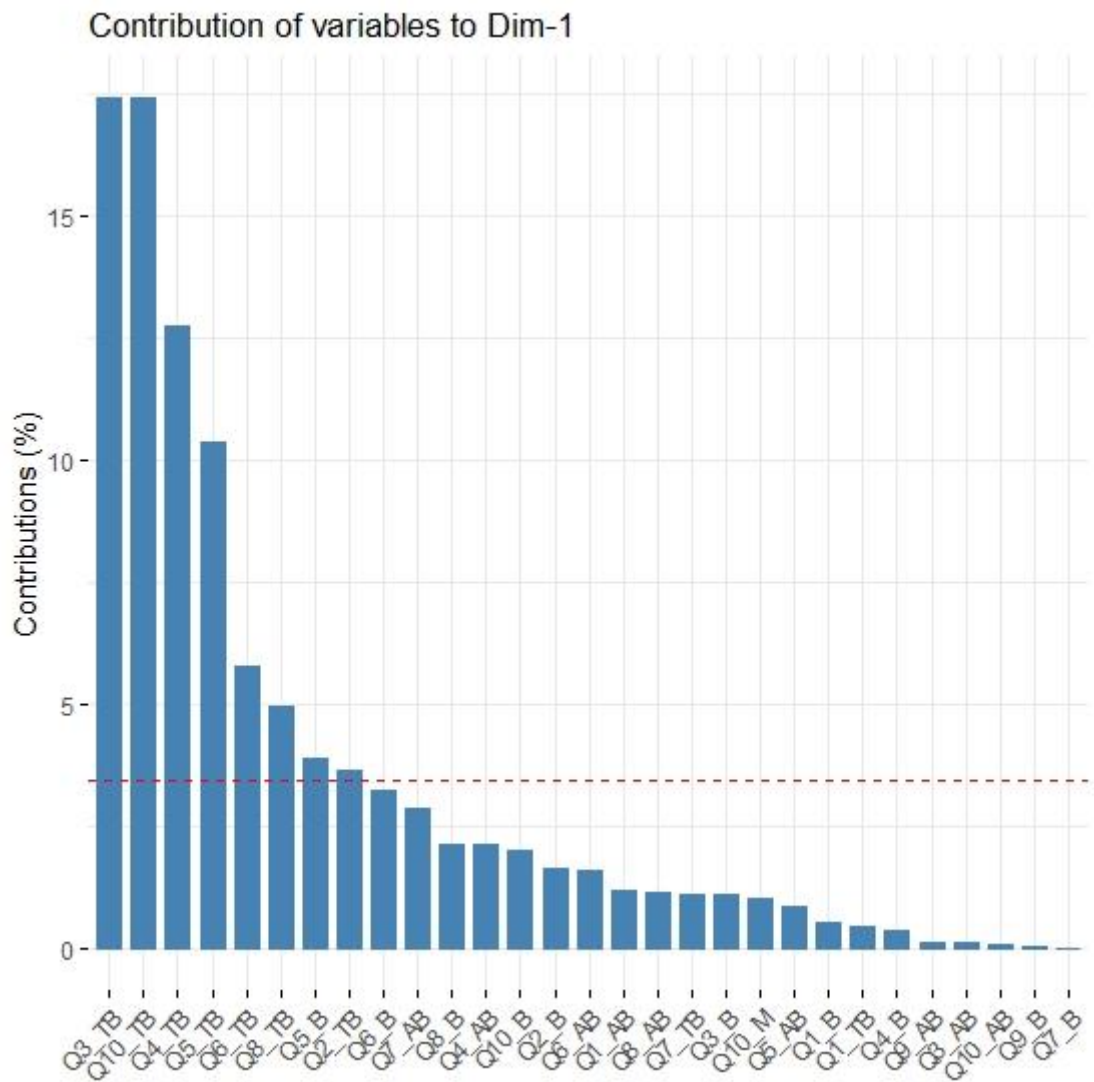
**b) Contribution des individus sur l'axe 2 :**

```
> fviz_contrib(afcm, choice = "ind", axes = 2, top = 16)
```



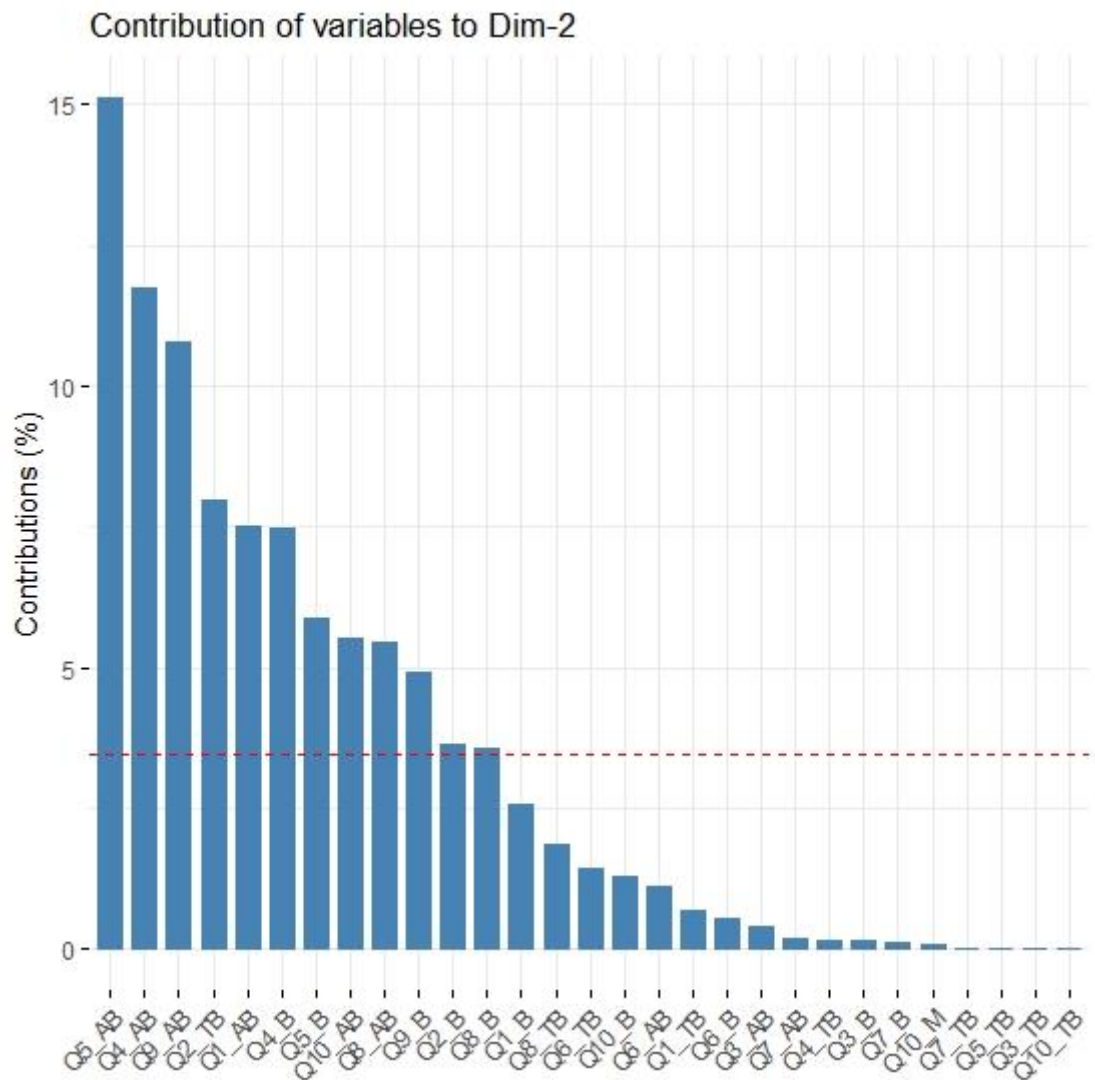
**c) Contribution des variables sur l'axe 1 :**

```
> fviz_contrib(afcm, choice = "var", axes = 1, top = 40)
```



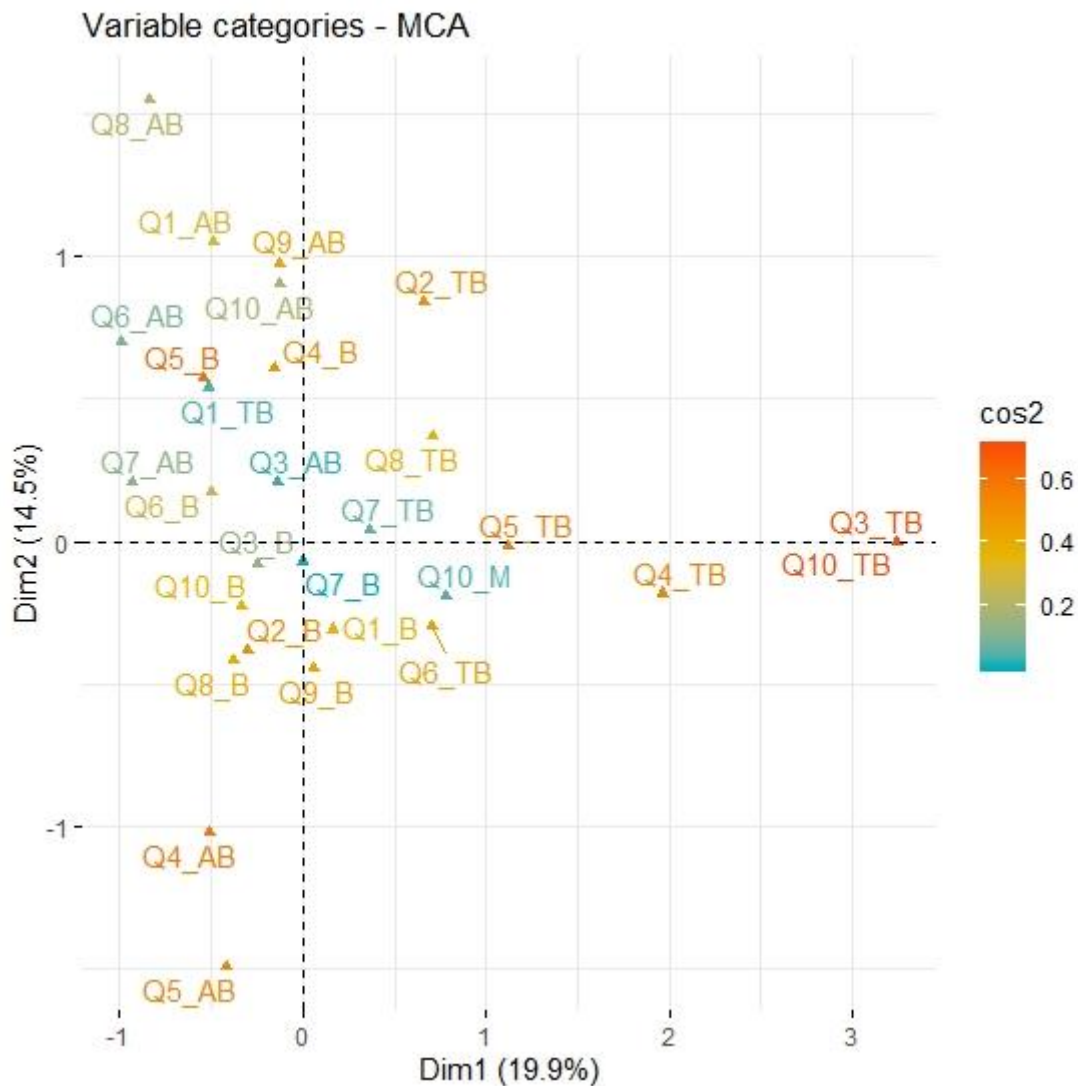
**d) Contribution des variables sur l'axe 2 :**

```
> fviz_contrib(afcm, choice = "var", axes = 2, top = 40)
```



## 11. Ressort des associations entre modalités :

```
> library(factoextra)
> fviz_mca_var(afcm, col.var = "cos2",
+               gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
+               repel = TRUE,
+               ggtheme = theme_minimal())
```



## 12. Interprétation des résultats :

- Les modalités qui ont la même couleur ou presque sont dépendantes (par exemple Q3\_TB et Q10\_TB).
- En se rapprochant de la couleur rouge la qualité de représentation (cos2) augmente.
- En se rapprochant de la couleur bleue la qualité de représentation (cos2) diminue.

## 13. Les questions les mieux représentées par l'AFCM :

Les questions les mieux représentées par l'AFCM sont Q4, Q5, Q3 et Q10. Cela ne veut pas dire que les autres sont mal représentées sauf Q7 qui est proche de l'origine.



#### 14.Etablissement d'un tableau de contingence :

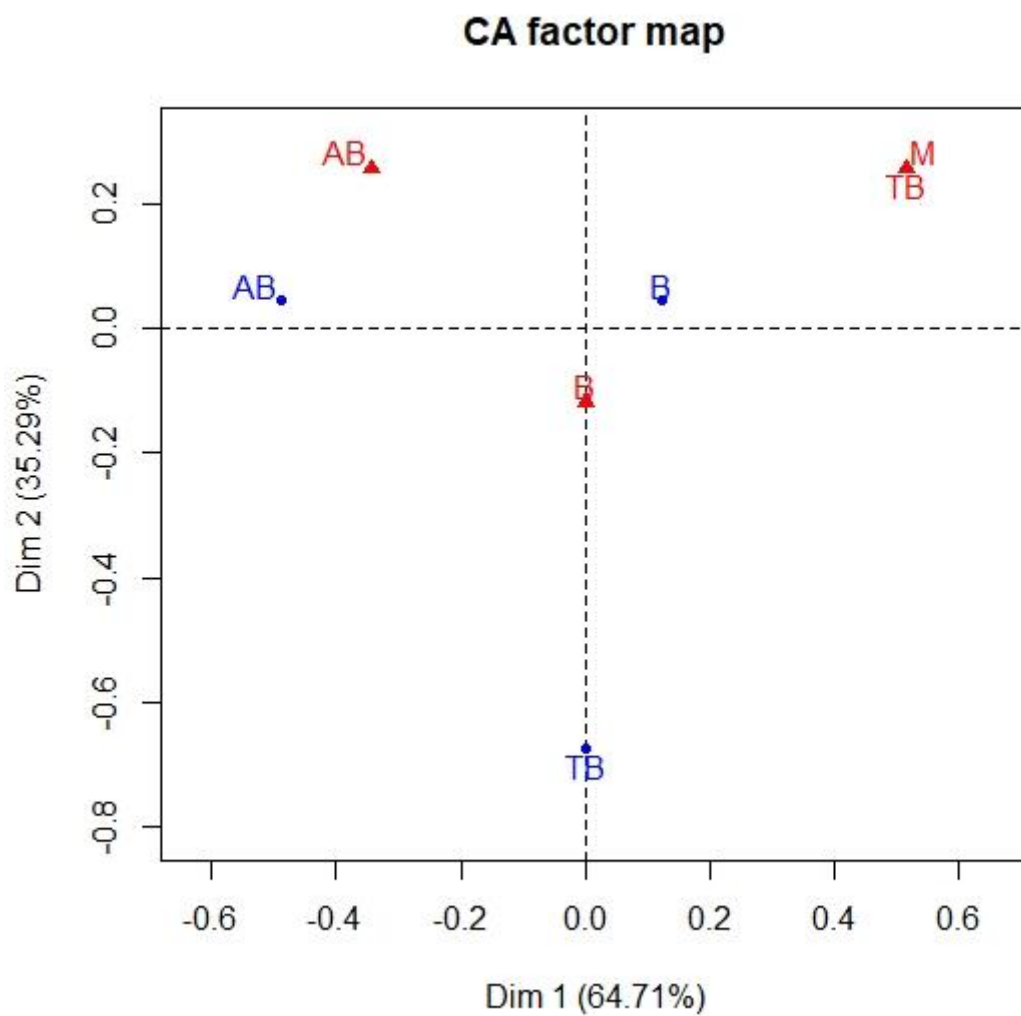
```
> X <- table(data$Q1,data$Q10)
> X
```

|    | AB | B | M | TB |
|----|----|---|---|----|
| AB | 1  | 2 | 0 | 0  |
| B  | 2  | 8 | 1 | 1  |
| TB | 0  | 1 | 0 | 0  |

#### 15.Etablissement de l'AFC :

```
> afc = CA(X)
```

##### a) Visualisation:



### **b) Interprétation des résultats :**

On remarque que :

- Q1\_AB et Q10\_AB sont proches donc les individus qui répondent par AB pour la question 1 ont plus de chance de répondre par la même réponse pour la question 10.
- Q1\_B et Q10\_B sont proches donc les individus qui répondent par B pour la question 1 ont plus de chance de répondre par la même réponse pour la question 10.
- Q1\_TB et Q10\_TB sont loin entre elles donc les individus qui répondent par TB pour la question 1 ont plus de chance de ne pas répondre par la même réponse pour la question 10.

### **16. Conclusion :**

D'après ce TP, j'ai appris à utiliser R et ses bibliothèques professionnellement, en important des données puis les traiter et extraire des visualisations qui sont ensuite interprétées en utilisant les connaissances apprises dans le module d'analyse de données et spécialement le cours d'analyse factorielle des composantes multiples.