



ANAD

Rapport du TP N°2

Application de l'AFCM et interprétation des résultats

Réalisé par :

- BOUAZIZ Sofiane
- OUSSAIDENE Smail

Groupe :

SID 01

Table des matières

1	Introduction.....	4
2	Importation des données	4
3	L'étude réalisée.....	4
4	Explication des données.....	5
5	Lecture des données	6
6	Prétraitement des données.....	6
5.1	Suppression des colonnes non nécessaires	6
5.2	Suppression des données manquantes	6
6	Étude statistiques des données	7
7	AFCM	14
7.1	Création du tableau disjonctif complet :.....	14
7.2	Application de l'AFCM :.....	15
7.3	Etude du tableau des valeurs propres :	15
7.4	Qualité de représentation globale	18
7.5	Tableau des projections des modalités	18
7.6	Biplot des individus-modalités	19
7.7	Signification des axes :	19
7.8	Tableau des rapports de corrélation des variables	22
7.9	Tableau des contributions absolues des variables	23
7.10	Différentes visualisation	24
7.11	Interprétation des résultats.....	30
7.12	Les associations possibles :	32
7.13	Qualité de représentation des modalités	32
7.14	Les questions les mieux représentées par l'AFCM	36
8	AFCM ajustée	36
8.1	Etude du tableau des valeurs propres ajustées :.....	37
8.2	Qualité de représentation globale	38
8.3	Tableau des projections des modalités	38
8.4	Biplot des individus-modalités	39
8.5	Signification des axes :	39
8.6	Interprétation des résultats.....	41
9	AFC.....	41
9.1	Création du tableau de contingence	41

9.2	Application de l'AFC sur le tableau de contingence	42
9.3	Etude du tableau des valeurs propres.....	42
9.4	Tableaux de projection	42
9.5	Signification du 1 ^{er} axe :.....	43
9.6	Visualisation des résultats.....	45
9.7	Interprétation des résultats	45
9.8	Associations possibles	45
10	Conclusion.....	46
11	Annexe : Script complet.....	47

1 Introduction

Les questionnaires sont un moyen très populaire aujourd'hui pour recueillir de l'information, ainsi analyser leurs données devient une tâche pertinente. C'est là où l'Analyse des Correspondances Multiples intervient, qui peut être vue comme une extension de l'Analyse des Correspondances. L'AFCM permet de résumer un tableau de données contenant plus de deux variables catégorielles et d'aboutir à des cartes de représentation sur lesquelles on peut visuellement observer les proximités entre les catégories des variables qualitatives et les observations.

L'objectif de cette méthode d'analyse factorielle est d'une part identifier un groupe de personnes ayant un profil similaire dans leurs réponses aux questions et d'autre part trouver les associations entre les catégories des variables.

Dans le cadre de ce TP, nous allons appliquer l'AFCM sur des données réelles et enfin en tirer des conclusions.

2 Importation des données

L'ensemble des données utilisées dans ce TP proviennent du site web de Kaggle¹ partagé par Michael Lomuscio. Les données représentent les réponses d'une enquête sur les habitudes et les croyances des jeunes adultes concernant les masques faciaux sous forme csv² disponible sur ce [lien](#).

3 L'étude réalisée

L'ensemble des données utilisé dans ce TP a été créé dans le cadre du module AP Statistics³ où tous les étudiants de ce module issus d'écoles aux États-Unis, ont répondu à un questionnaire dont le but est d'en savoir plus sur les habitudes et les croyances des jeunes adultes concernant le port des masques.

Les questions introduites dans ce questionnaire sont présentées ci-dessous :

- Êtes-vous un étudiant externe ou en internat ?
- Quel est votre âge ?
- Quel est votre genre ?
- Vivez-vous avec une personne âgée de plus de 65 ans ?
- Avez-vous eu des contacts avec une personne âgée de plus de 65 ans au cours du mois dernier ?
- Quel est le nombre de fois que vous avez mangé au restaurant au cours des deux dernières semaines ?
- Pensez-vous que les masques sont efficaces pour prévenir la propagation du coronavirus ?

¹ Kaggle est une plateforme web organisant des compétitions en science des données, fondée en 2010 par Anthony Goldbloom. Elle permet aux utilisateurs de trouver et de publier des ensembles de données, d'explorer et de construire des modèles dans un environnement de science des données basé sur le web.

² Comma-separated values, connu sous le sigle CSV, est un format texte ouvert représentant des données tabulaires sous forme de valeurs séparées par des virgules.

³ Advanced Placement Statistics (également connu sous le nom de AP Statistics ou AP Stat) est un cours de statistiques de niveau collégial offert aux États-Unis dans le cadre du programme Advanced Placement du College Board. Ce cours équivaut à un cours d'introduction aux statistiques universitaires d'un semestre, non basé sur le calcul, et est normalement proposé aux étudiants de deuxième année, juniors et seniors du secondaire.

- Quelle est votre principale raison de porter un masque ?
- Portez-vous un masque dans les lieux publics lorsque cela n'est pas nécessaire ?

4 Explication des données

Ci-dessous l'explication de chaque variable de notre ensemble de données :

- **Timestamp** : l'heure et la date auxquelles l'individu a répondu au questionnaire sous format MM/DD/YYYY hh:mm:ss
- **Boarding** : si l'individu était un étudiant externe ou un étudiant en internat. L'ensemble des valeurs prises par cet attribut est :
 - Day (étudiant externe)
 - Boarding (étudiant en internat)
- **Age** : l'âge de l'individu sous format entier
- **Gender** : Le genre déclaré par l'individu. L'ensemble des valeurs prises par cet attribut est :
 - Male (Homme)
 - Female (Femme)
 - Other (Autre)
- **ResidentialElder** : La réponse à la question « Vivez-vous avec une personne âgée de plus de 65 ans ? ». L'ensemble des valeurs prises par cet attribut est :
 - Yes (Oui)
 - No (Non)
- **InteractedElder** : La réponse à la question « Avez-vous eu des contacts avec une personne âgée de plus de 65 ans au cours du mois dernier ? ». L'ensemble des valeurs prises par cet attribut est :
 - Yes
 - No
- **Restaurant** : Le nombre de fois où le répondant a déclaré avoir mangé dans un restaurant au cours de la semaine dernière sous format entier
- **PreventSpread** : La réponse à la question « Pensez-vous que les masques sont efficaces pour prévenir la propagation du coronavirus ? ». L'ensemble des valeurs prises par cet attribut est :
 - Yes
 - No
- **Raison** : La réponse à la question « Quelle est votre principale raison de porter un masque ? » L'ensemble des valeurs prises par cet attribut est :
 - To protect yourself AND others (Pour se protéger et protéger les autres)
 - To protect other people (Pour protéger d'autres personnes)
 - Because you are required to (Parceque vous êtes obligé)
 - To protect yourself (Pour vous protéger)
 - To protect others but also because I'm required to (Pour protéger les autres mais aussi parce que je suis obligé de le faire)
- **Public** : La réponse à la question « Portez-vous un masque dans les lieux publics lorsque cela n'est pas nécessaire ? » L'ensemble des valeurs prises par cet attribut est :
 - Yes
 - No

5 Lecture des données

Nous avons utilisé la fonction « *read.csv* » pour récupérer les données représentant les réponses du questionnaire décrit auparavant du fichier csv *MaskBeliefs.csv*.

```
# na.strings=c("") : remplacer les chaines de caracteres vides avec NA
data <- read.csv("MaskBeliefs.csv", header = TRUE, na.strings=c(""))
```

Voici ci-dessous, une prise d'écran représentant une partie des données récupérées :

```
> head(data)
```

	Timestamp	Boarding	Age	Gender	ResidentialElder	InteractedElder	Restaurant	PreventsSpread	Reason	Public
1	9/25/2020 15:04:43	Day	16	Female	No	Yes	1	Yes	To protect yourself AND others	Yes
2	9/25/2020 15:04:46	Boarding	17	Male	No	No	2	Yes	To protect yourself AND others	Yes
3	9/25/2020 15:04:58	Boarding	17	Male	No	Yes	0	Yes	To protect yourself AND others	Yes
4	9/25/2020 15:05:12	Day	17	Female	No	Yes	2	Yes	To protect yourself AND others	Yes
5	9/25/2020 15:05:12	Day	17	Female	No	Yes	2	Yes	To protect yourself AND others	Yes
6	9/25/2020 15:05:15	Boarding	18	Male	No	No	0	Yes	To protect other people	No

Nous avons tiré quelques remarques :

1. Des colonnes non nécessaires pour notre étude (des variables quantitatives)
2. Quelques lignes de notre ensemble de données contiennent des données manquantes

6 Prétraitement des données

Pour aboutir à une analyse de bonne qualité, nous devons d'abord prétraiter les données. De ce fait nous avons procédé comme suit :

5.1 Suppression des colonnes non nécessaires

L'AFCM exige que l'observation des individus se repose sur des variables qualitatives. De ce fait, nous avons supprimé les colonnes suivantes :

- Timestamp
- Age
- Restaurant

Pour cela nous avons utilisé ces instructions :

```
# Suppression des colones qui ne representent pas des variables
qualitatifs
data$Timestamp<- NULL
data$Age<- NULL
data$Restaurant <- NULL
```

Pour vérifier le résultat, nous allons afficher la liste des noms des colonnes retenues :

```
> colnames(data)
```

[1]	"Boarding"	"Gender"	"ResidentialElder"	"InteractedElder"	"PreventSpread"
[6]	"Reason"	"Public"			

5.2 Suppression des données manquantes

Nous remarquons que les valeurs de certaines variables ne sont pas disponibles. Pour vérifier cela, nous avons utilisé l'instruction suivante :

```
# Affichage du nombre de valeurs manquantes
sum na <- sum(is.na(data))
```

Nous obtenons le résultat suivant :

```
> print(paste("Le nombre des NAs dans notre dataset est :", sum_na))
[1] "Le nombre des NAs dans notre dataset est : 1"
```

Nous avons utilisé l'instruction suivante pour supprimer l'individu ayant une valeur manquante pour l'un des variables qualitatives.

```
# Suppression des NAs
data = na.omit(data)
```

Pour vérifier le résultat, nous allons réafficher le nombre de valeurs manquantes :

```
> print(paste("Le nombre des NAs dans notre dataset est :", sum_na_apres))
[1] "Le nombre des NAs dans notre dataset est : 0"
```

6 Étude statistiques des données

Dans ce qui suit, nous allons étudier différentes statistiques sur notre ensemble de données pour mieux se familiariser.

Nous allons représenter pour chaque variable qualitative le nombre de personnes ayant choisie chacune de ses modalités et leurs fréquences d'apparition ainsi qu'un histogramme résumant cela.

Nous allons présenter aussi pour certaines variables qualitatives un diagramme de pie chart (diagramme circulaire) pour mieux illustrer la distribution de leurs modalités.

6.1 Variable 'Boarding'

Pour afficher le tableau d'occurrence et de fréquences de chaque modalité, nous avons utilisé la library « tidyverse » comme suit :

```
# Afficher le tableau des statistiques de la variable Boarding
Boardingstat <- fct_count(data$Boarding, prop = TRUE, sort = TRUE) %>%
  rename("Reason" = "f", "freq" = "p", "nb" = "n") %>%
  mutate(freq = round(freq, 2))
```

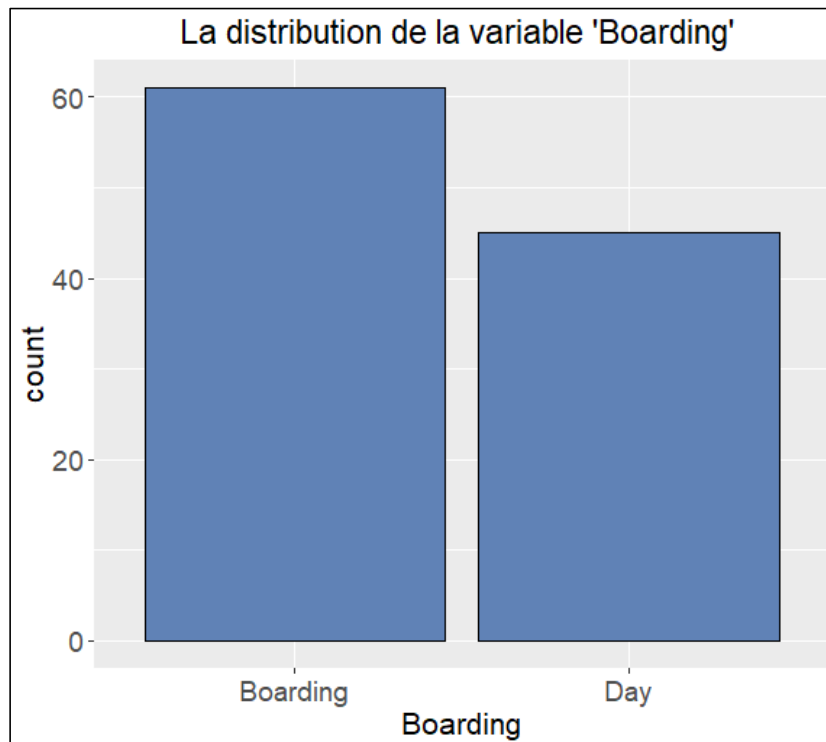
Le résultat est présenté ci-dessous :

	Boarding	nb	freq
	<fct>	<int>	<dbl>
1	Boarding	61	0.58
2	Day	45	0.42

Pour afficher l'histogramme, nous avons utilisé la fonction « ggplot » comme suit :

```
# Plotter la distribution de la variable Boarding
library(tidyverse)
ggplot(data.frame(data), aes(x = Gender)) +
  geom_bar(colour = "black", fill = "#6082B6") +
  labs(title = "La distribution de la variable 'Gender' ") +
  theme(plot.title = element_text(hjust = 0.5, size=17), axis.text.x=
element_text(angle = 45, size=14),
        axis.text.y= element_text(size=14),
        axis.title=element_text(size=16))
```

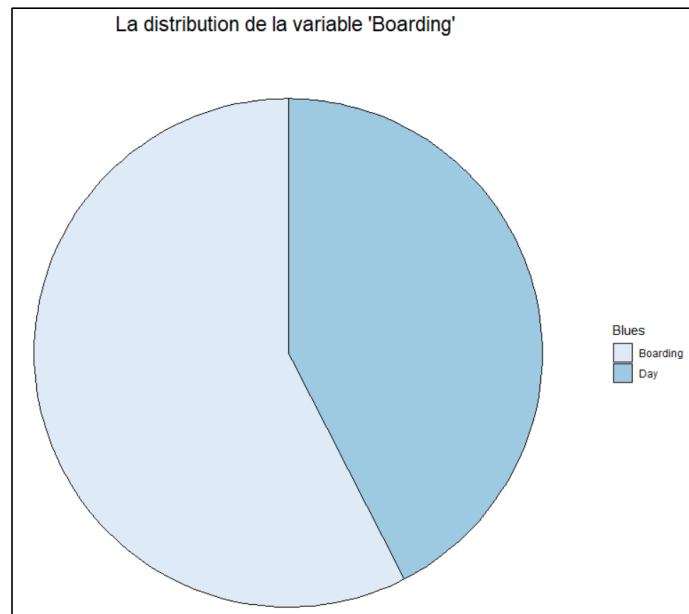
Voici ci-dessous l'histogramme :



Pour afficher le diagramme de pie chart, nous avons utilisé la fonction « ggplot » comme suit :

```
# Plotter le diagramme de pie chart
ggplot(Boardingstat , aes(x="", y= nb, fill= Boarding)) +
  scale_fill_brewer("Blues") +
  geom_bar(stat="identity", width=1, , color = "black") +
  labs(title = "La distribution de la variable 'Boarding' ") +
  coord_polar("y", start=0) +
  theme_void() +
  theme(plot.title = element_text(hjust = 0.5, size=17))
```

Voici ci-dessous le diagramme de pie chart :

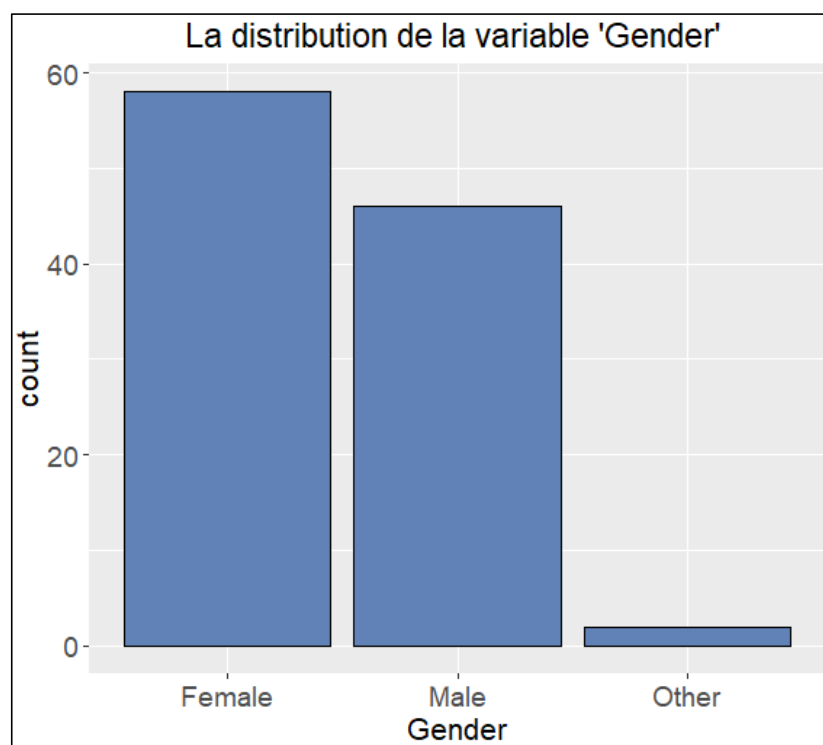


6.2 Variable 'Gender'

Voici ci-dessous le tableau d'occurrence et de fréquences de chaque modalité :

	Gender	nb	freq
	<i><fct></i>	<i><int></i>	<i><dbl></i>
1	Female	58	0.55
2	Male	46	0.43
3	other	2	0.02

Voici ci-dessous l'histogramme :

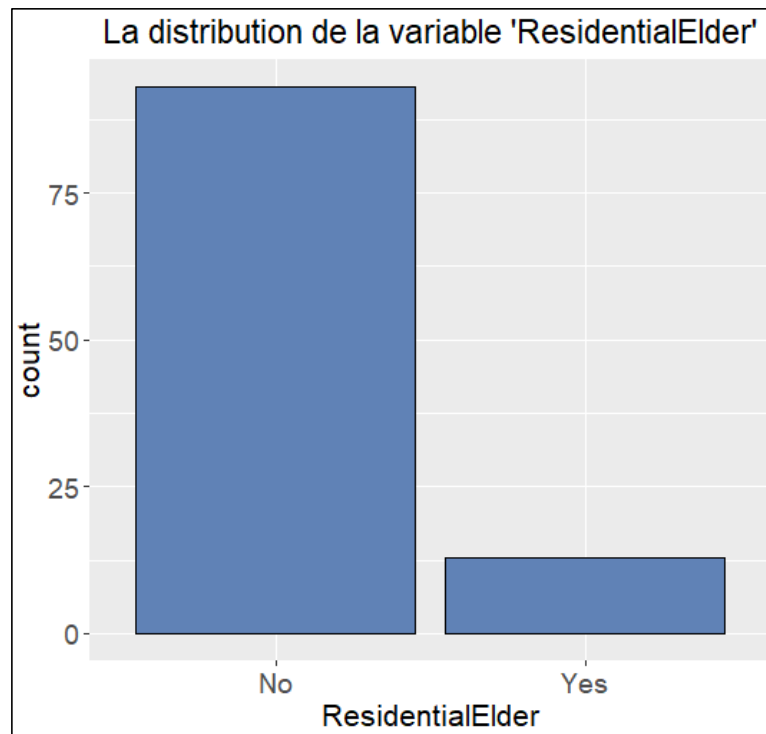


6.3 Variable 'ResidentialElder'

Voici ci-dessous le tableau d'occurrence et de fréquences de chaque modalité :

	ResidentialElder	nb	freq
	<fct>	<int>	<dbl>
1	No	93	0.88
2	Yes	13	0.12

Voici ci-dessous l'histogramme :

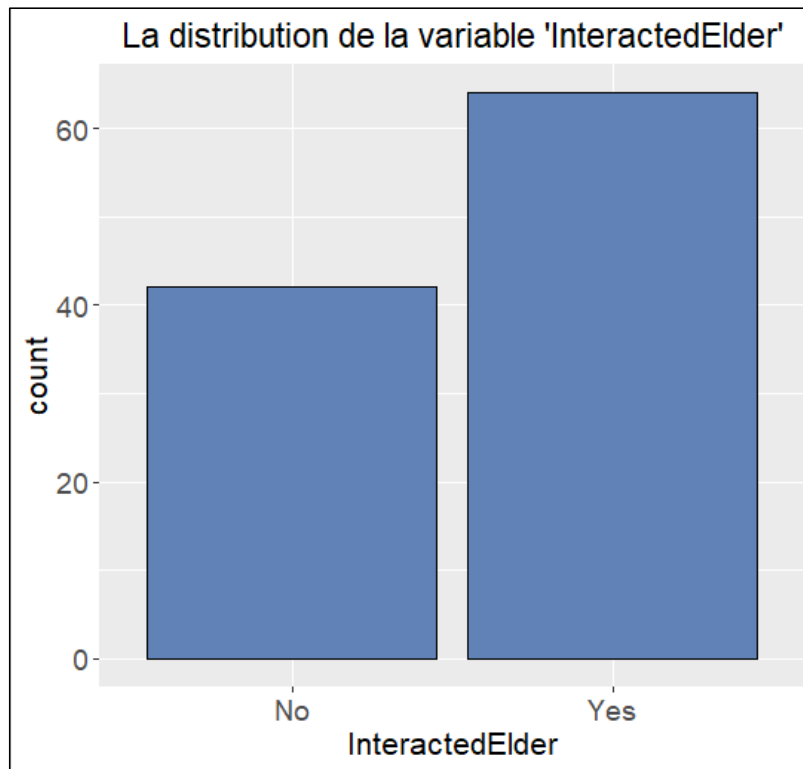


6.4 Variable 'InteractedElder'

Voici ci-dessous le tableau d'occurrence et de fréquences de chaque modalité :

	InteractedElder	nb	freq
	<fct>	<int>	<dbl>
1	Yes	64	0.6
2	No	42	0.4

Voici ci-dessous l'histogramme :

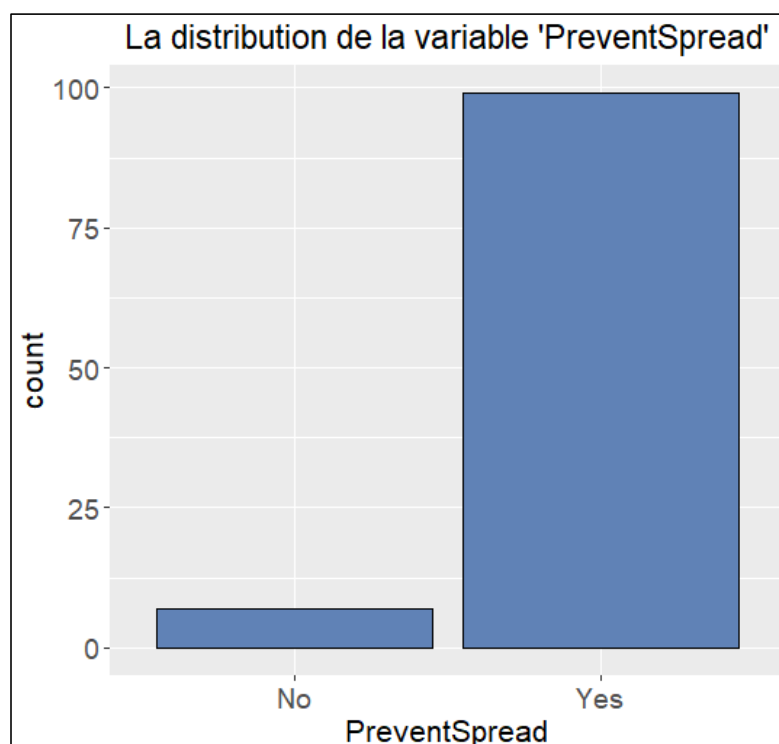


6.5 Variable 'PreventSpread'

Voici ci-dessous le tableau d'occurrence et de fréquences de chaque modalité :

	PreventSpread	nb	freq
	<i><fct></i>	<i><int></i>	<i><dbl></i>
1	Yes	99	0.93
2	No	7	0.07

Voici ci-dessous l'histogramme :

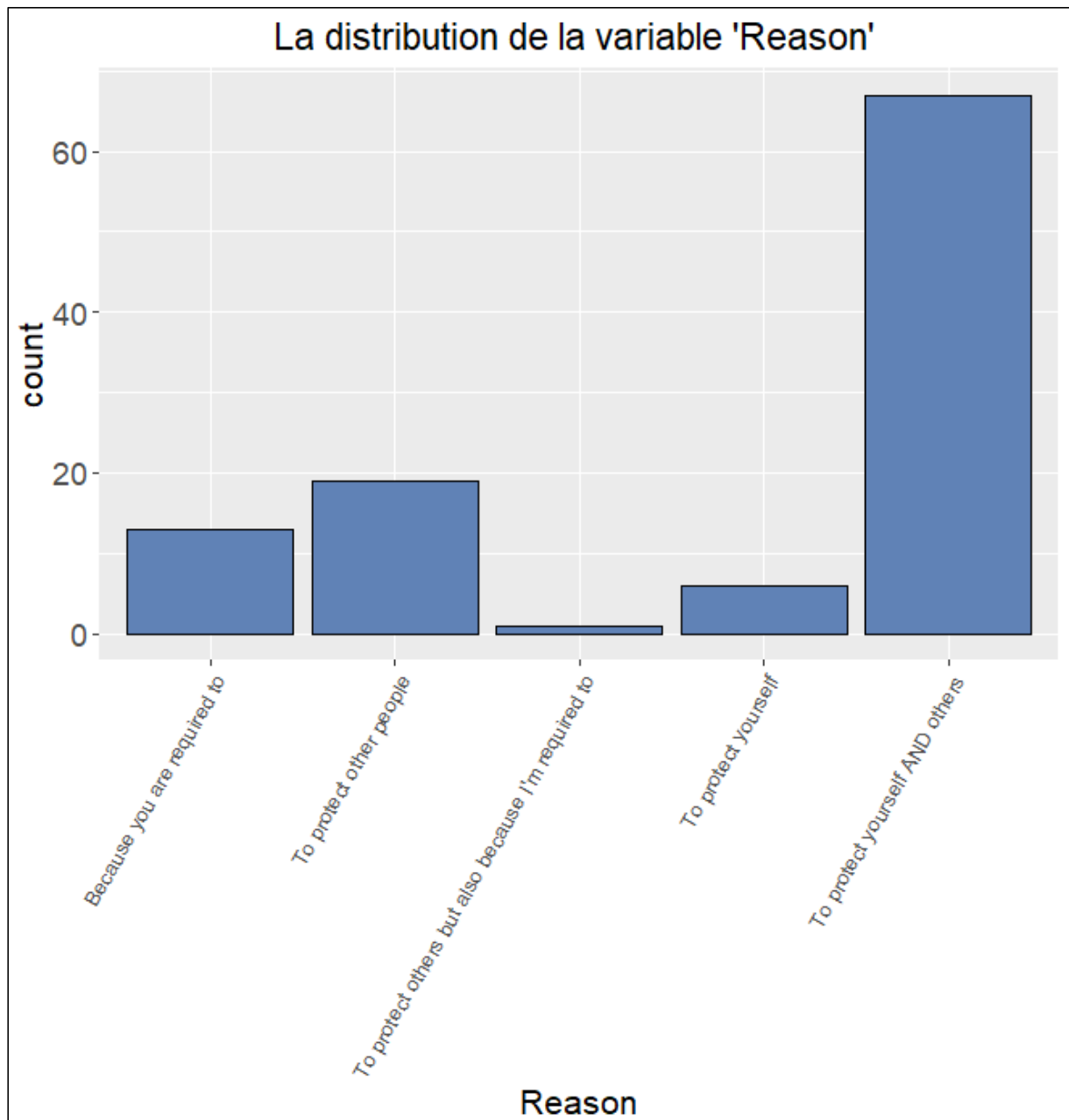


6.6 Variable 'Reason'

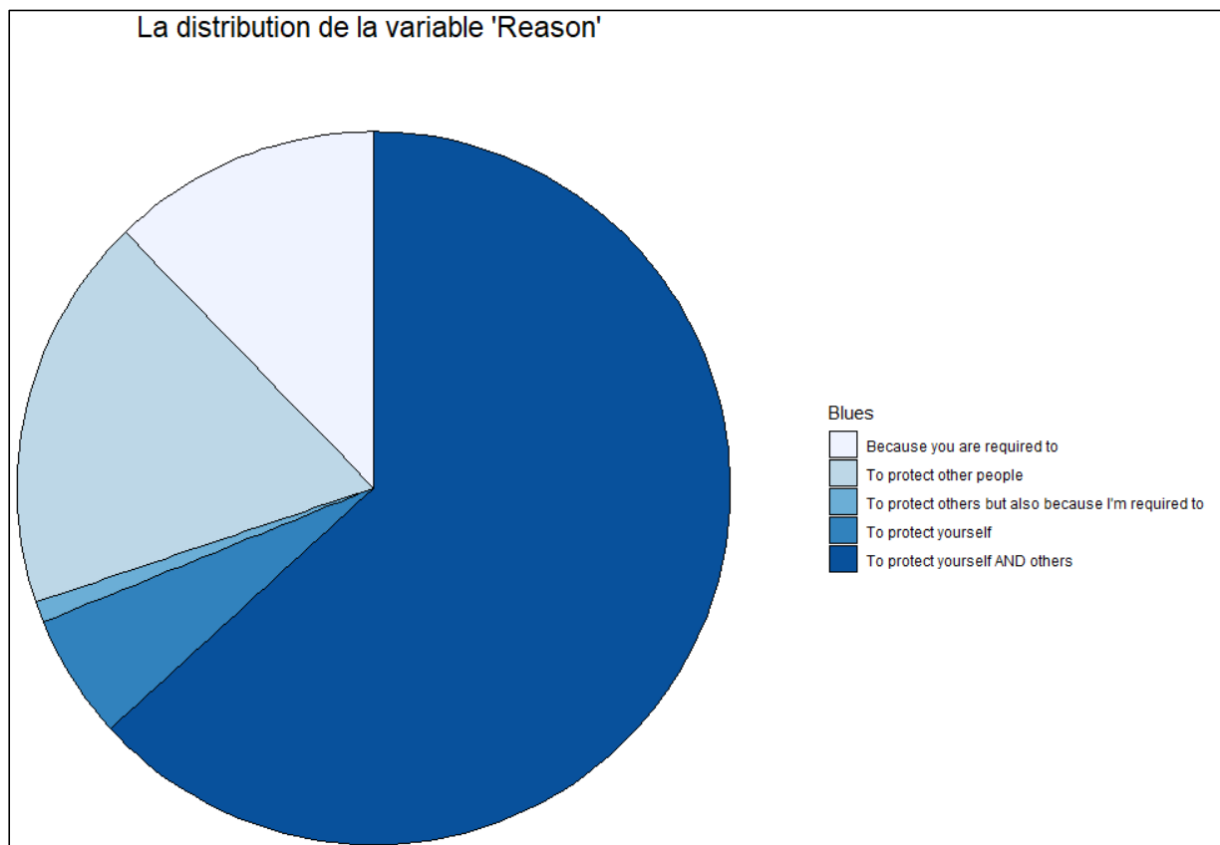
Voici ci-dessous le tableau d'occurrence et de fréquences de chaque modalité :

Reason	nb	freq
<fct>	<int>	<dbl>
1 To protect yourself AND others	67	0.63
2 To protect other people	19	0.18
3 Because you are required to	13	0.12
4 To protect yourself	6	0.06
5 To protect others but also because I'm required to	1	0.01

Voici ci-dessous l'histogramme :



Voici ci-dessous le diagramme de pie chart :

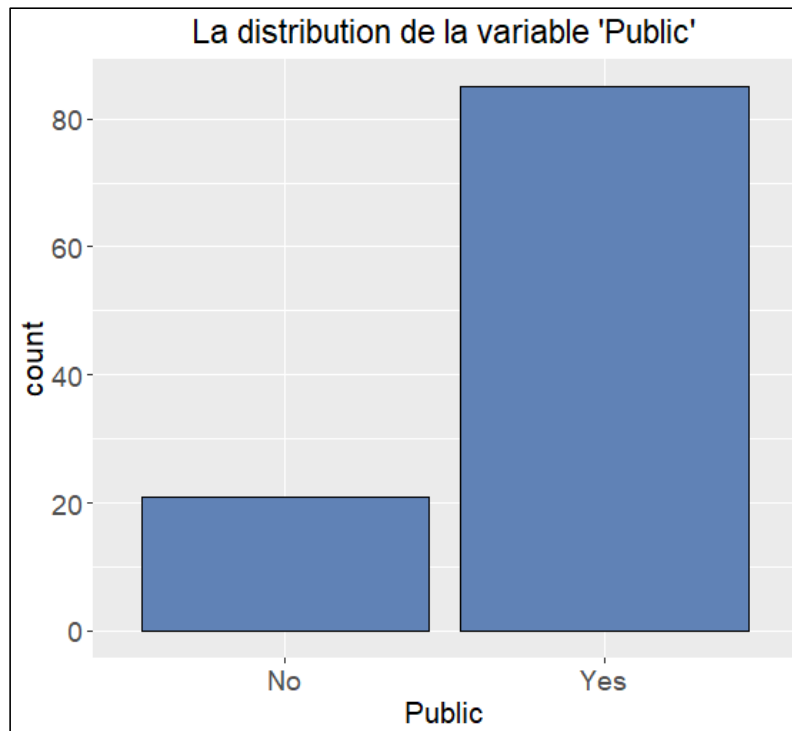


6.7 Variable 'Public'

Voici ci-dessous le tableau d'occurrence et de fréquences de chaque modalité :

	Public	nb	freq
	<fct>	<int>	<dbl>
1	Yes	85	0.8
2	No	21	0.2

Voici ci-dessous l'histogramme :



7 AFCM

7.1 Création du tableau disjonctif complet :

Nous avons utilisé l'instruction « *tab.disjonctif* » pour créer le tableau disjonctif complet à partir de notre tableau de données de cette façon :

```
# Appliquer l'AFCM sur le tableau de donnée (data)
Z <- tab.disjonctif(data)
```

Voici ci-dessous le résultat obtenu :

```
> head(Z)
Boarding Day Female Male Other No Yes No Yes No Yes Because you are required to To protect other people
1 0 1 1 0 0 1 0 0 1 0 1 0 0
2 1 0 0 1 0 1 0 1 0 0 1 0 0
3 1 0 0 1 0 1 0 0 1 0 1 0 0
4 0 1 1 0 0 1 0 0 1 0 1 0 0
5 0 1 1 0 0 1 0 0 1 0 1 0 0
6 1 0 0 1 0 1 0 1 0 0 1 0 1
To protect others but also because I'm required to To protect yourself To protect yourself AND others No Yes
1 0 0 0 1 0 1
2 0 0 0 1 0 1
3 0 0 0 1 0 1
4 0 0 0 1 0 1
5 0 0 0 1 0 1
6 0 0 0 0 1 0
```

7.2 Application de l'AFCM :

Nous avons utilisé la library «*FactoMineR*» pour appliquer l'AFCM sur notre tableau de données de la façon suivante :

```
# Appliquer l'AFCM sur le tableau de donnée (data)
afcm <- MCA(data)
```

7.3 Etude du tableau des valeurs propres :

7.3.1 Etude des valeurs propres initiales

Nous avons utilisé l'instruction suivante pour extraire le tableau des valeurs propres issus de l'AFCM appliquée à notre tableau de données

```
# Afficher le tableau des valeurs propres issue de notre AFCM
afcm$eig
```

Nous obtenons le résultat suivant :

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0.25830540	16.437616	16.43762
dim 2	0.22740393	14.471159	30.90878
dim 3	0.17086102	10.872974	41.78175
dim 4	0.16038490	10.206312	51.98806
dim 5	0.15031362	9.565412	61.55347
dim 6	0.14232389	9.056975	70.61045
dim 7	0.13646569	8.684180	79.29463
dim 8	0.10276408	6.539532	85.83416
dim 9	0.08673749	5.519658	91.35382
dim 10	0.07711875	4.907557	96.26138
dim 11	0.05874979	3.738623	100.00000

Remarque

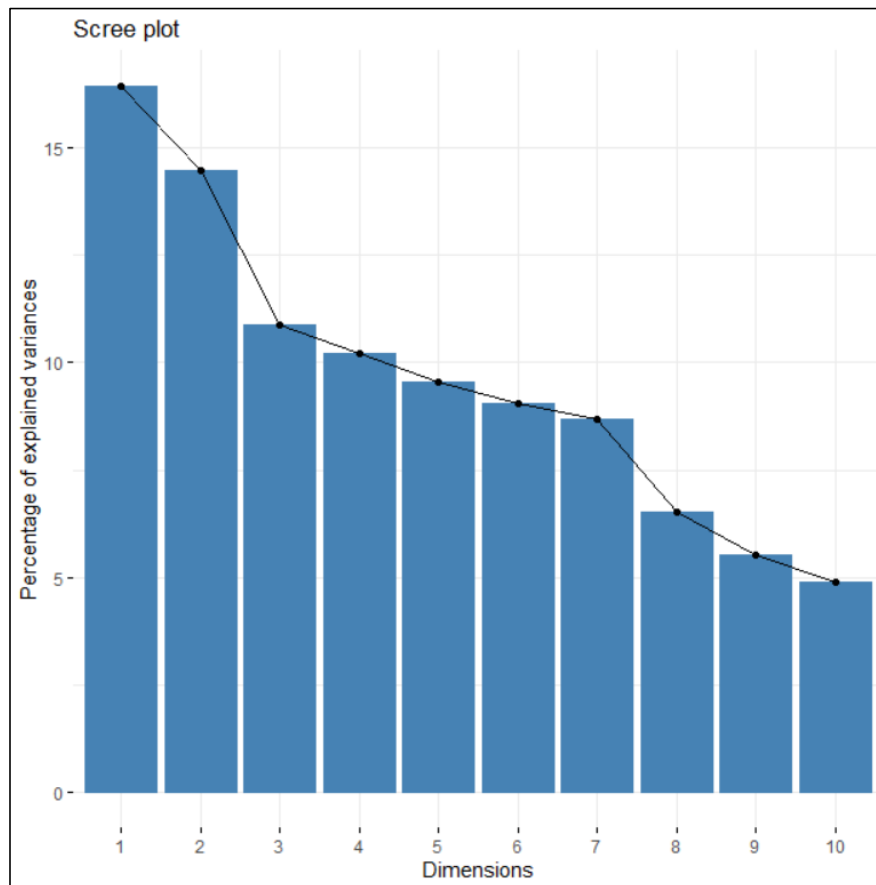
En réalité, on a 18 valeurs propres (car J (le nombre totale de modalités) = 18). Mais, en appliquant l'AFCM sur R , on remarque que seulement 11 valeurs propres sont affichées. On peut justifier cela, par le fait que les autres valeurs propres restantes sont toutes nulles et donc leur pourcentage d'inertie est égal à 0 et de ce fait elles ne sont pas représentées dans le tableau des valeurs propres.

De plus, le nombre de valeurs propres nulles est égal aux nombre de variables et ceci revient au fait que pour une variable donnée X^j à m_j modalité, on peut déduire une modalité en utilisant les $m_j - 1$ modalités restantes. Ce qui conduit à $J - P = 18 - 7 = 11$ valeurs propres non nulles.

Nous avons utilisé la library «*factoextra*» pour représenter l'éboullis des valeurs propres :

```
library("factoextra")
#Représentation de l'eboullis des valeurs propres
fviz_eig(afcm)
```

Nous avons obtenu le résultat suivant :



On remarque que l'écoulement des valeurs propres descend en pente douce, on ne pourra pas donc se baser sur ces valeurs propres pour choisir le nombre d'axes à retenir. Ainsi, il faut utiliser un indicateur corrigé.

7.3.2 Etude des valeurs propres avec correction de Benzeckri

On a vu dans la section précédente qu'il n'est pas possible en AFCM d'interpréter l'inertie totale de façon aussi intéressante qu'en ACP ou AFC. Par contre, Benzeckri a proposé une inertie corrigée pour pouvoir mieux apprécier la qualité de représentation d'un axe.

L'algorithme proposé par Benzeckri est décrit comme suit :

1. Prendre les valeurs propres supérieures à la moyenne des valeurs propres $1/p$
 - a. Calculer la nouvelle valeur propre en utilisant la formule suivante :

$$\lambda'_h = \left[\left(\frac{p}{p-1} \right) \times \left(\lambda_h - \frac{1}{p} \right) \right]^2$$

- b. Calculer l'inertie de cette valeur propre
2. Utiliser la méthode du coude pour déterminer le nombre d'axes à retenir

Nous avons utilisé les instructions suivantes pour extraire le tableau des valeurs propres corrigées

```
# Tableau de valeurs propres corrigé
p = ncol(data) # Le nombre de questions

# Correction des valeurs propres
eigs_corr = subset(afcm$eig, afcm$eig[, "eigenvalue"] > 1/p)
eigs_corr[, "eigenvalue"] = ((p/(p-1)) * (eigs_corr[, "eigenvalue"] -
(1/p))) ** 2

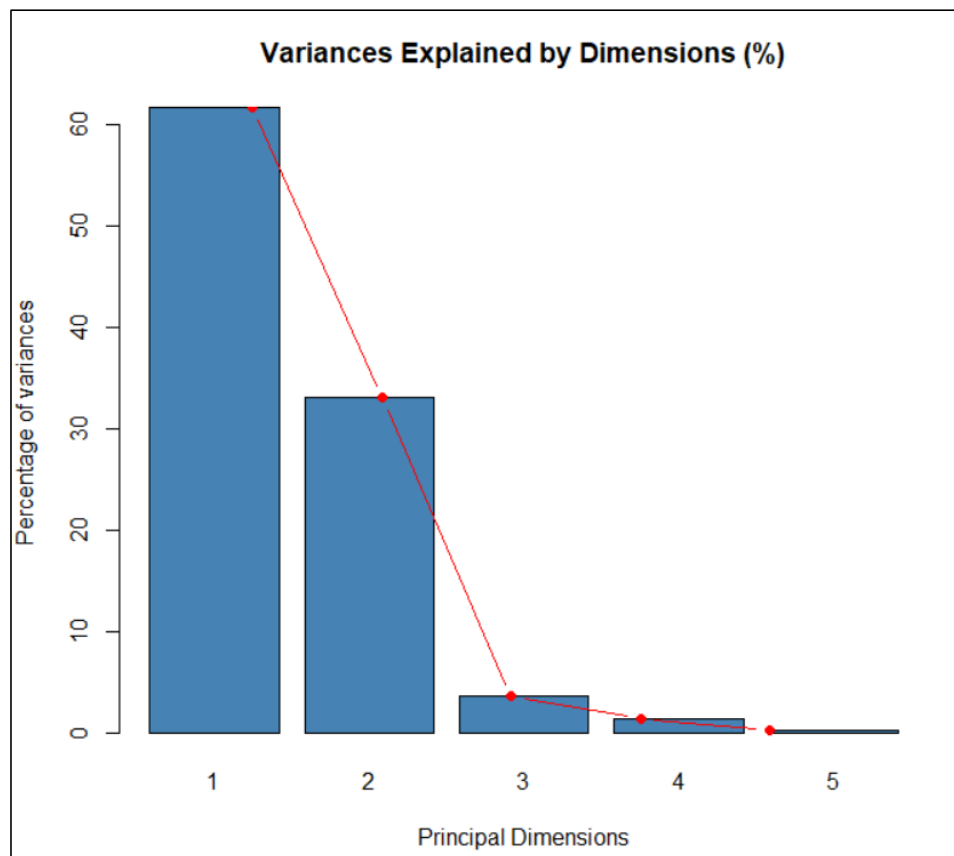
# Calcul de l'inertie des nouvelle valeur propore
eigs_corr[, "percentage of variance"] = (eigs_corr[, "eigenvalue"] /
sum(eigs_corr[, "eigenvalue"])) * 100

# Calcul de l'inertie cummulée
eigs_corr[1, "cumulative percentage of variance"] = eigs_corr[1,
"percentage of variance"]
for(i in 2:nrow(eigs_corr)) {
  eigs_corr[i, "cumulative percentage of variance"] = eigs_corr[i-1,
"cumulative percentage of variance"] + eigs_corr[i, "percentage of
variance"]
}
```

Nous obtenons le résultat suivant :

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	1.814130e-02	61.6380377	61.63804
dim 2	9.729440e-03	33.0573690	94.69541
dim 3	1.067407e-03	3.6266905	98.32210
dim 4	4.181635e-04	1.4207791	99.74288
dim 5	7.567658e-05	0.2571236	100.00000

L'éboulis des valeurs propres corrigées est représenté ci-dessus :



D'après l'éboulis des valeurs propres corrigées, le coude se trouve entre la 2ème et la 3ème valeur. Donc on retient les 2 premiers axes. (On a aussi $I_{12}(\%) = 94.69541 > 80 \%$).

7.4 Qualité de représentation globale

$$I_{12} = \frac{\lambda'1 + \lambda'2}{\sum \lambda' i}$$

$I_{12} = 94.69541 \%$, donc 94.69541 % de l'information est contenue dans le plan engendré par le premier et deuxième axe factoriel.

7.5 Tableau des projections des modalités

Nous avons utilisé l'instruction suivante pour extraire **le tableau des projections** de chaque modalité de nos variables qualitatives issues de l'AFCM appliquée à notre tableau de données.

```
# Afficher le tableau des projections des modalités issue de notre AFCM
afcm$var$coord
```

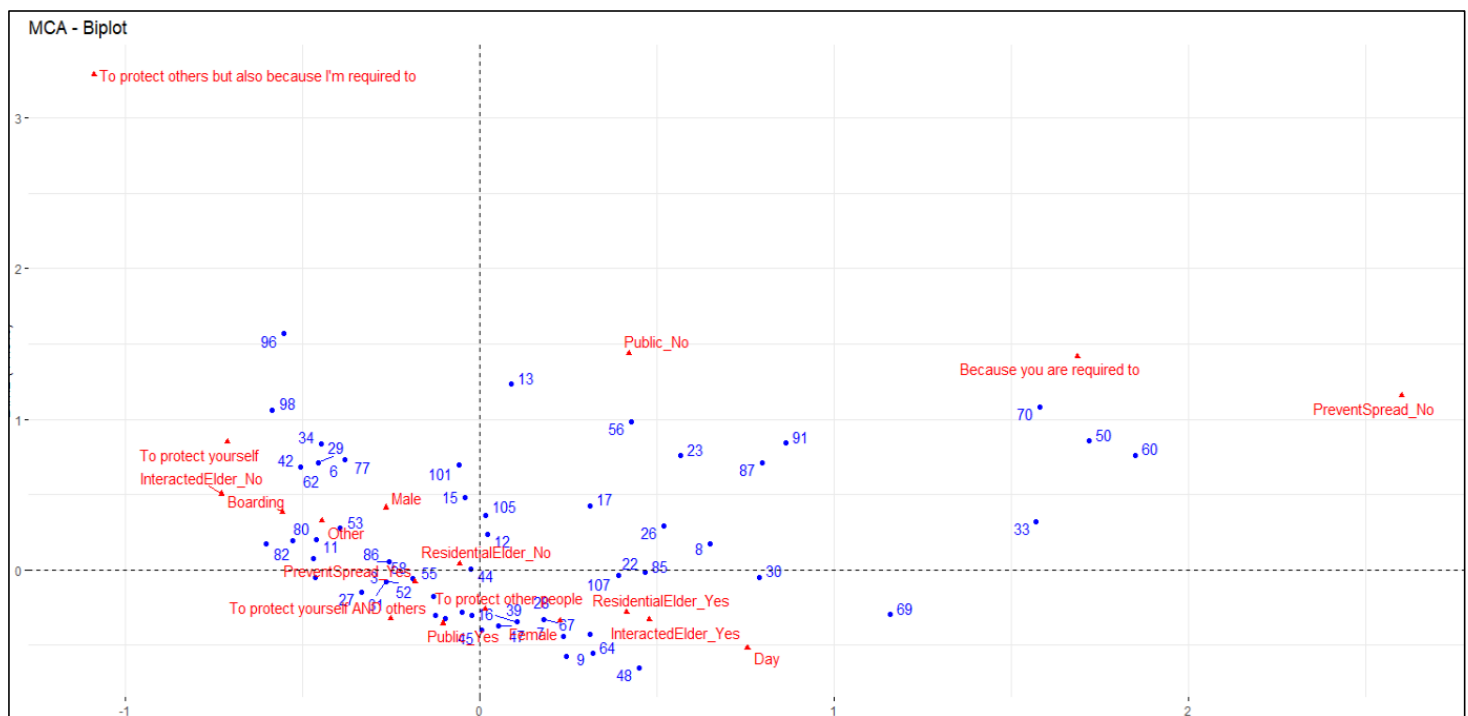
Nous obtenons le résultat suivant :

	Dim 1	Dim 2
Boarding	-0.55707025	0.38326565
Day	0.75513967	-0.51953788
Female	0.22646757	-0.33894570
Male	-0.26611811	0.41323773
Other	-0.44684311	0.32495757
ResidentialElder_No	-0.05779891	0.03961951
ResidentialElder_Yes	0.41348449	-0.28343187

InteractedElder_No	-0.72915449	0.50406906
InteractedElder_Yes	0.47850764	-0.33079532
PreventSpread_No	2.60273765	1.15647202
PreventSpread_Yes	-0.18403195	-0.08177075
Because you are required to	1.68750018	1.41925116
To protect other people	0.01512280	-0.26356598
To protect others but also because I'm required to	-1.08907455	3.28718462
To protect yourself	-0.71241592	0.85037987
To protect yourself AND others	-0.25166068	-0.32585038
Public_No	0.42146573	1.43920858
Public_Yes	-0.10412683	-0.35556918

7.6 Biplot des individus-modalités

En appliquant l'AFCM, nous sommes parvenus à représenter nos données multi-dimensionnelles dans un plan 2D facile à visualiser et à comprendre.



7.7 Signification des axes :

Afin d'extraire le **tableau des contributions** de chaque modalités de nos variables qualitatives issues de l'AFCM appliquée à notre tableau de données, nous utiliserons l'instruction suivante

```
# Afficher le tableau des contributions issue de notre AFCM
afcm$var$contrib
```

Nous obtenons le résultat suivant :

	Dim 1	Dim 2
Boarding	9.876711734	5.31040637
Day	13.388431462	7.19855086
Female	1.552039094	3.94899313
Male	1.699688360	4.65538806

Other	0.208354429	0.12516458
ResidentialElder_No	0.162100668	0.08651656
ResidentialElder_Yes	1.159643238	0.61892618
InteractedElder_No	11.650673033	6.32451897
InteractedElder_Yes	7.645754178	4.15046557
PreventSpread_No	24.741238603	5.54838291
PreventSpread_Yes	1.749380507	0.39230990
Because you are required to	19.314951392	15.51886634
To protect other people	0.002267151	0.78222275
To protect others but also because I'm required to	0.618839214	6.40392619
To protect yourself	1.588843383	2.57143361
To protect yourself AND others	2.213951690	4.21608523
Public_No	1.946284985	25.77893054
Public_Yes	0.480846879	6.36891225

Une modalité m_j d'une variable qualitative X^j contribue à la construction d'un axe α , si sa contribution absolue par rapport à cet axe est supérieure à son poids. Mais comme les contributions absolues sous R sont données en pourcentage, donc si $\text{cab}^\alpha(l) > 100 * f_l$, alors cette modalité contribue à la construction de l'axe α .

Pour afficher le poids de chaque modalité de nos variables qualitatives, nous avons utilisé l'instruction suivante

```
# Afficher les poids des modalités
afcm$call$marge.col
```

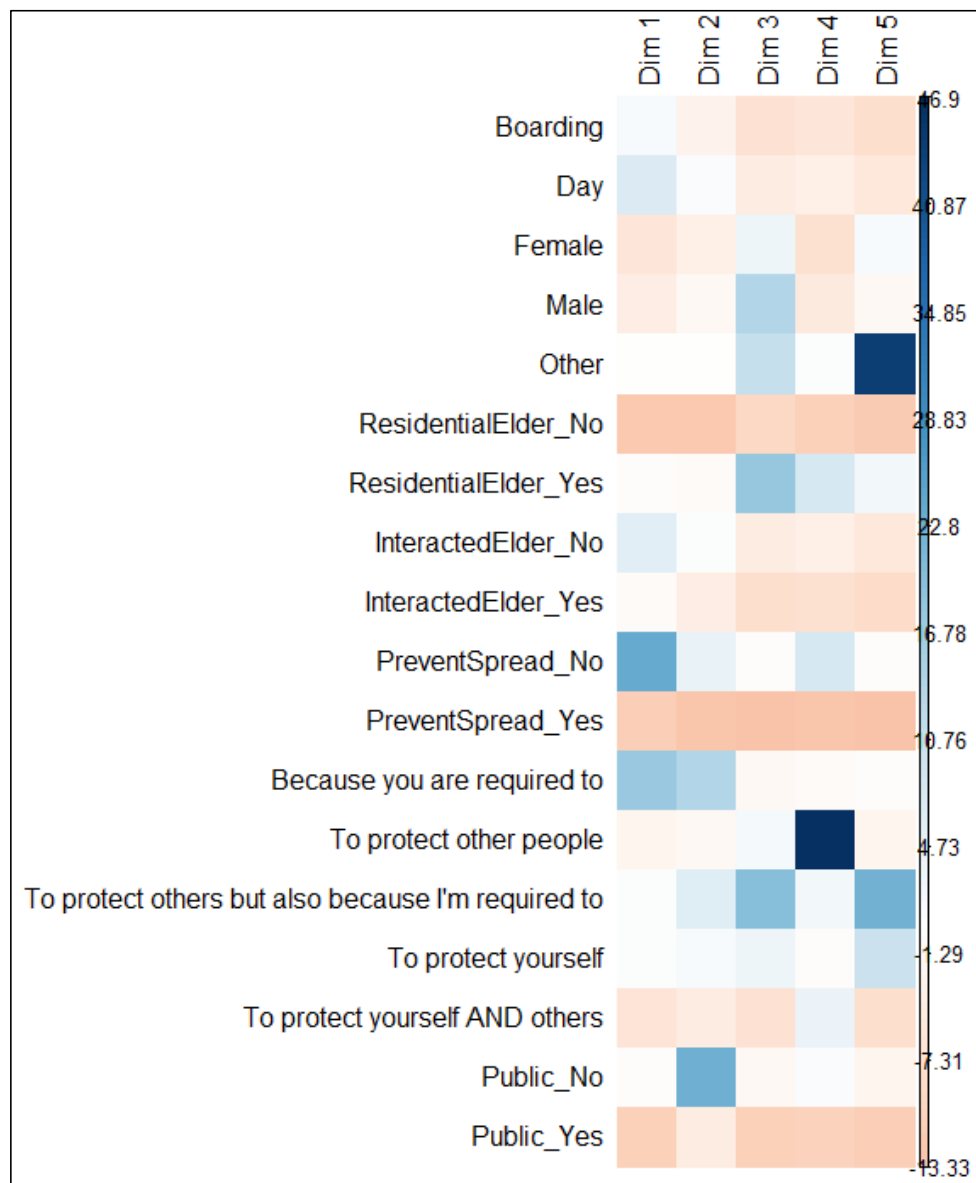
Nous obtenons le résultat suivant :

Modalité	Poids
Boarding	0.082210243
Day	0.060646900
Female	0.078167116
Male	0.061994609
Other	0.002695418
ResidentialElder_No	0.125336927
ResidentialElder_Yes	0.017520216
InteractedElder_No	0.056603774
InteractedElder_Yes	0.086253369
PreventSpread_No	0.009433962
PreventSpread_Yes	0.133423181
Because you are required to	0.017520216
To protect other people	0.025606469
To protect others but also because I'm required to	0.001347709
To protect yourself	0.008086253
To protect yourself AND others	0.090296496
Public_No	0.028301887
Public_Yes	0.114555256

Pour représenter graphiquement la différence entre la contribution absolue des modalités et leurs poids afin de visualiser les modalités qui contribuent à la construction des axes, nous avons utilisé la library « corrplot » de cette façon :

```
# Illustration de la différence entre les contribution absolues et les
# poids des modalités
library("corrplot")
corrplot(afcm$var$contrib - 100*afcm$call$marge.col, is.corr=FALSE,
method = 'color', tl.col="black", col = COL2('RdBu', 200) )
```

Voici ci-dessous le résultat :



7.7.1 Signification du 1^{er} axe

+	-
Day, PreventSpread_No, Because you are required to	Boarding, InteractedElder_No, To protect others but also because I'm required to, To protect yourself

L'axe 1 est un axe à effet opposition. Il oppose les étudiants externes qui ne pensent pas que les masques sont efficaces pour prévenir la propagation du coronavirus et qu'ils les portent, car ils

le sont obligés, aux étudiants en internat et qui n'ont pas eu des contacts avec une personne âgée de plus de 65 ans au cours du mois dernier et ils portent le masque afin de protéger les autres, mais aussi parce qu'ils sont obligés de le faire ou pour se protéger soi-même

7.7.2 Signification du 2^{ème} axe :

+	-
InteractedElder_No, PreventSpread_No, Because you are required to, To protect others but also because I'm required to, To protect yourself, Public_No	Day

L'axe 2 est un axe à effet opposition. Il oppose les étudiants en internat, aux étudiants qui n'ont pas eu des contacts avec une personne âgée de plus de 65 ans au cours du mois dernier, et qui ne pensent pas que les masques sont efficaces pour prévenir la propagation du coronavirus, et qu'ils les portent, car ils le sont obligés de le faire, pour protéger les autres, mais aussi parce qu'ils sont obligés de le faire ou pour se protéger, et qui ne portent pas un masque dans les lieux publics lorsque cela n'est pas nécessaire.

7.8 Tableau des rapports de corrélation des variables

Nous avons utilisé l'instruction suivante pour extraire le tableau des rapports de corrélation entre les variables qualitatives et les composantes principales issu de l'AFCM appliquée à notre tableau de données.

```
# Afficher le tableau des projections des modalités issue de notre AFCM
afcm$var$eta2
```

Nous obtenons le résultat suivant :

Variables	Dim 1	Dim 2
Boarding	0.42066585	0.19912102
Gender	0.06256305	0.13895931
ResidentialElder	0.02389895	0.01122943
InteractedElder	0.34890599	0.16674369
PreventSpread	0.47898690	0.09456558
Reason	0.42923117	0.46947028
Public	0.04388589	0.51173821

Lorsque le rapport de corrélation entre une variable qualitative et une composante principale est proche de 1, les individus correspondant à une même modalité sont très regroupés, et les modalités sont nettement séparées les unes des autres.

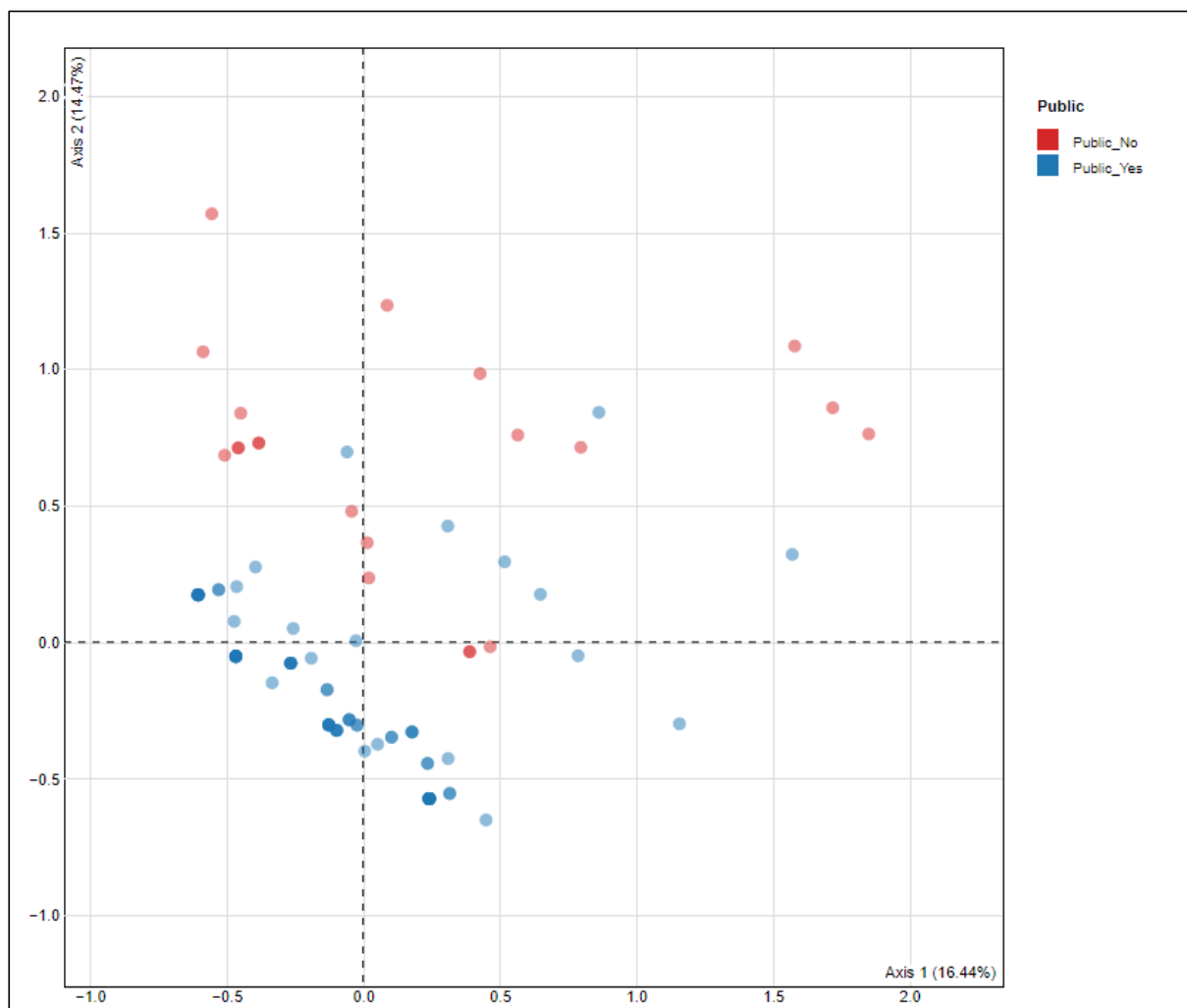
D'après le tableau présenté précédemment, on remarque que le rapport de corrélation de la variable public avec la deuxième composante principale est assez proche de 1, on peut ainsi dire que les individus correspondant à une même modalité de cette variable qualitative sont très regroupés, et ses modalités sont nettement séparées les unes des autres.

Cependant, on remarque aussi que le rapport de corrélation de la variable public avec la première composante principale est très loin de 1, on peut ainsi dire que les moyennes des groupes définis par les différentes modalités sont proches les unes des autres et les individus d'un même groupe sont dispersés.

Afin de visualiser l'exactitude de notre interprétation, nous avons utilisé la library « explor » comme suit :

```
# Visualiser les individus coloré par rapport la variable public
explor (afcm)
```

Le résultat est donné ci-dessous :



7.9 Tableau des contributions absolues des variables

Le tableau ci-dessous représente l'apport des variables qualitatives dans la construction de l'axe 1 et 2.

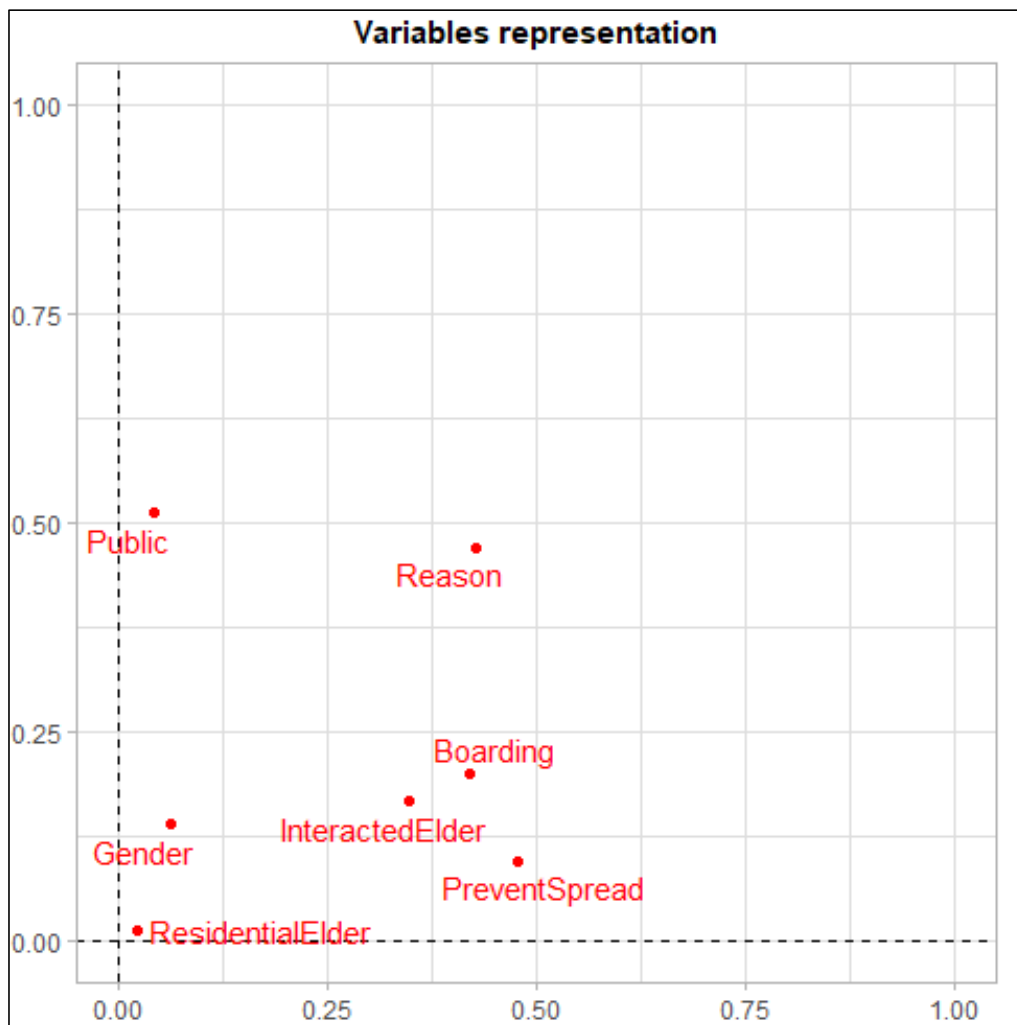
	Dim 1	Dim 2
Boarding	23.26514	12.50896
Gender	3.460082	8.729546
ResidentialElder	1.321744	0.7054427
InteractedElder	19.29643	10.47498
PreventSpread	26.49062	5.940693
Reason	23.73885	29.49253
Public	2.427132	32.14784

La contribution absolue d'une variable qualitative est égale à la somme des contributions absolues de ses modalités.

7.10 Différentes visualisation

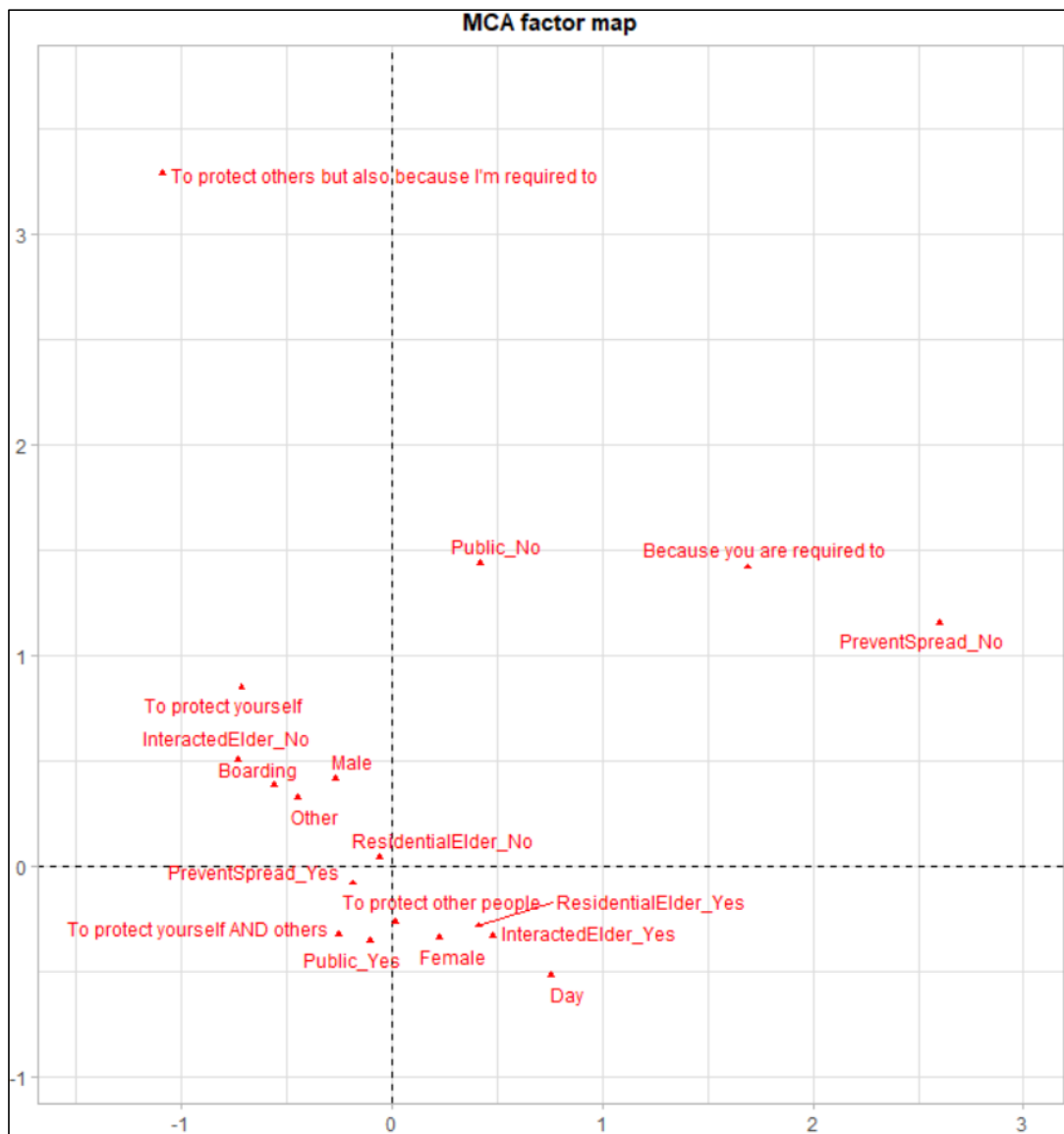
7.10.1 Représentations des variables

Ci-dessous la représentation des variables sur le 1^{er} plan factoriel



7.10.2 Représentation des modalités

Ci-dessous la représentation des modalités sur le 1^{er} plan factoriel

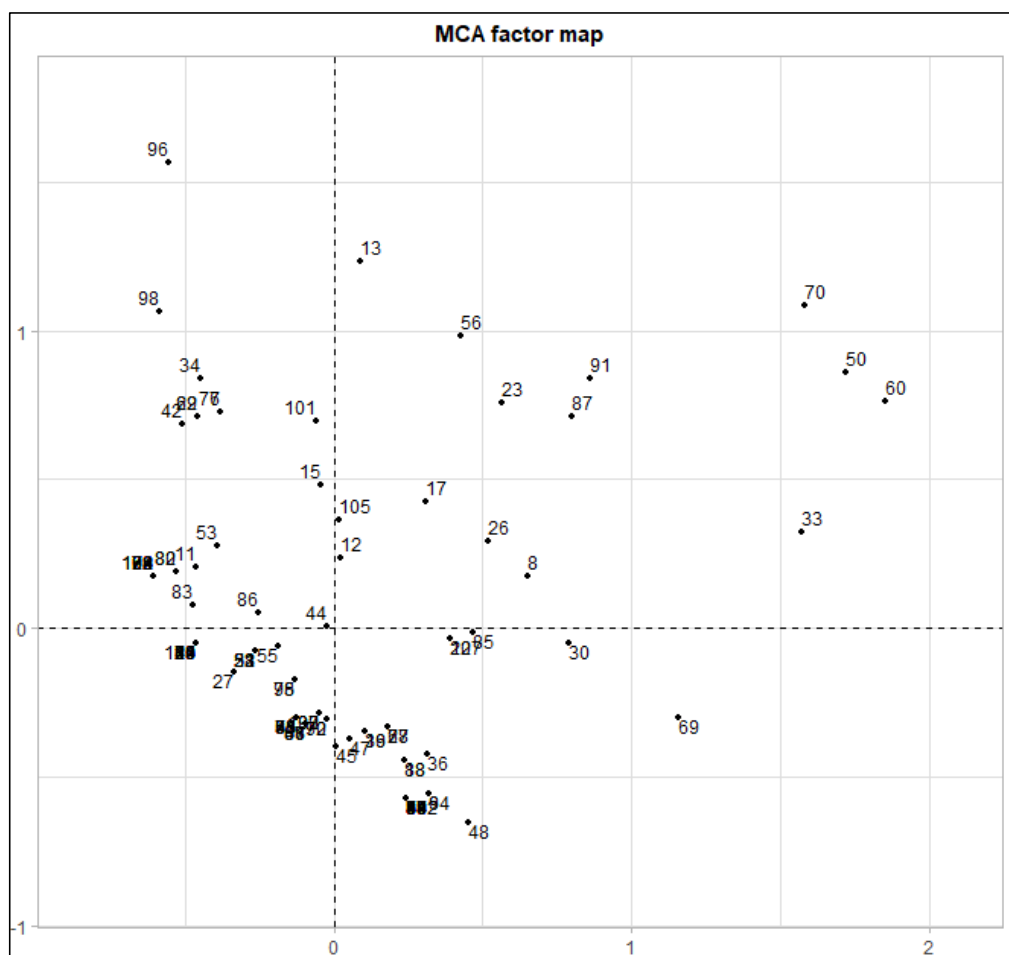


Remarque

Nous remarquons que la modalité « *To protect others but also because I'm required to* » est très loin de l'origine, en effet cette modalité a un poids très faible (0.001347709), mais elle contribue à la construction du premier et deuxième axe. En effet, cela revient au fait qu'un seul individu a choisi cette réponse est donc cet individu est considéré comme étant une donnée aberrante. Ce cas sera traité dans la section « 8. AFCM ajustée ».

7.10.3 Représentation des individus

Ci-dessous la représentation des individus sur le 1^{er} plan factoriel

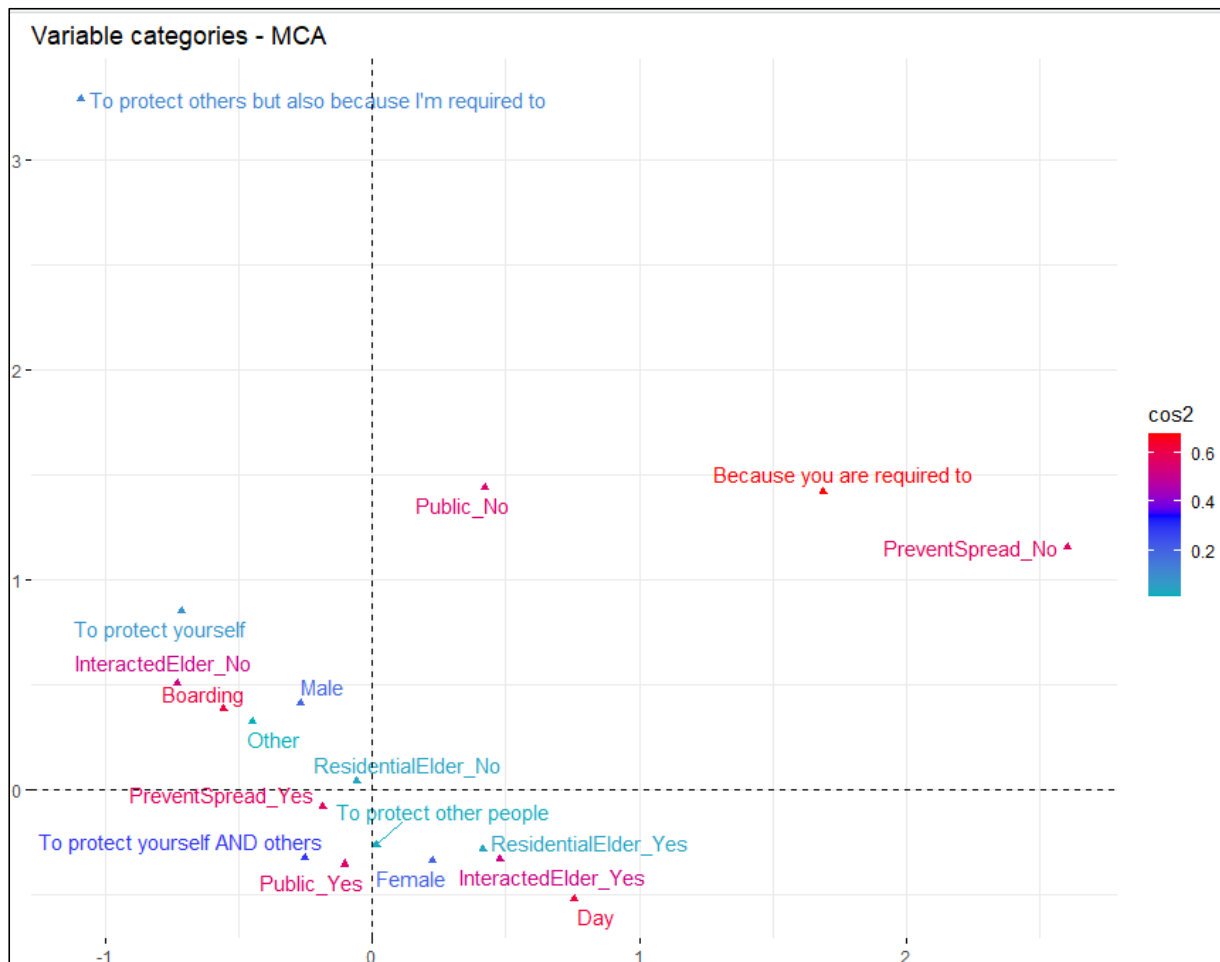


7.10.4 Illustration de la qualité de représentation des modalités

Pour visualiser la qualité de représentation des modalités sur le plan factoriel engendré par l'axe 1 et 2, nous avons utilisé la library « factoextra ».

```
# Illustration de la qualité de représentation des modalités
fviz_mca_var(afc, col.var = "cos2",
             gradient.cols = c("#00AFBB", "blue", "red"),
             repel = TRUE, # Avoid text overlapping
             ggtheme = theme_minimal())
```

Le résultat obtenu est donné ci-dessous :



7.10.5 Illustration de la contribution absolue des modalités

Pour visualiser la contribution absolue des modalités sur le plan factoriel engendré par l'axe 1 et 2, nous avons utilisé la library « factoextra».

```
# Illustration de la contribution absolue des modalités
fviz_mca_var(afc, col.var = "contrib",
             gradient.cols = c("#00AFBB", "blue", "red"),
             repel = TRUE, # Avoid text overlapping
             ggtheme = theme_minimal())
```

Le résultat obtenu est donné ci-dessous :

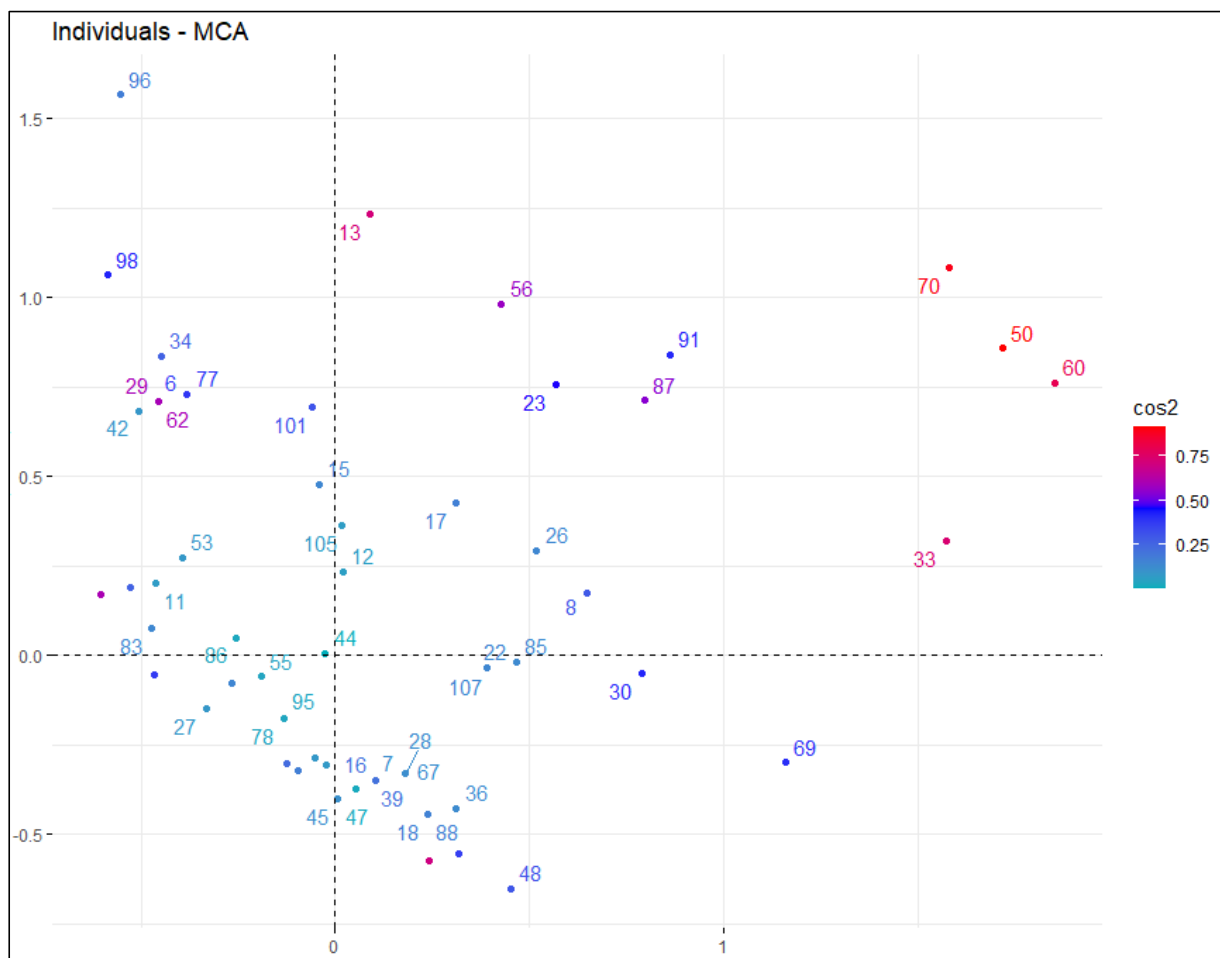


7.10.6 Illustration de la qualité de représentation des individus

Pour visualiser la qualité de représentation des individus sur le plan factoriel engendré par l'axe 1 et 2, nous avons utilisé la library « factoextra».

```
# Illustration de la contribution absolue des individus
fviz_mca_ind(afc, col.ind = "cos2",
             gradient.cols = c("#00AFBB", "blue", "red"),
             repel = TRUE, # Avoid text overlapping (slow if manypoints)
             ggtheme = theme_minimal())
```

Le résultat obtenu est donné ci-dessous :

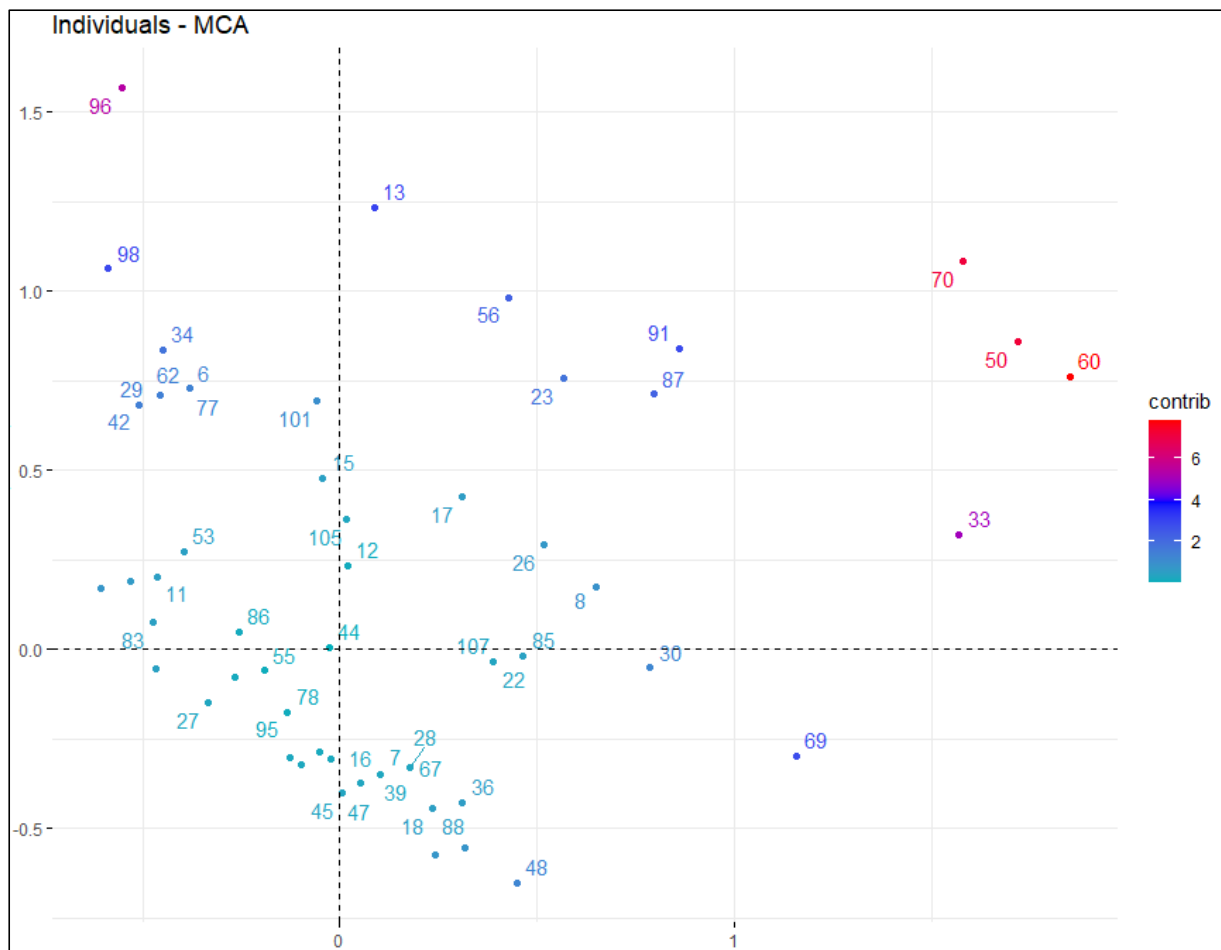


7.10.7 Illustration de la contribution absolue des individus

Pour visualiser la contribution absolue des individus sur le plan factoriel engendré par l'axe 1 et 2, nous avons utilisé la library « factoextra ».

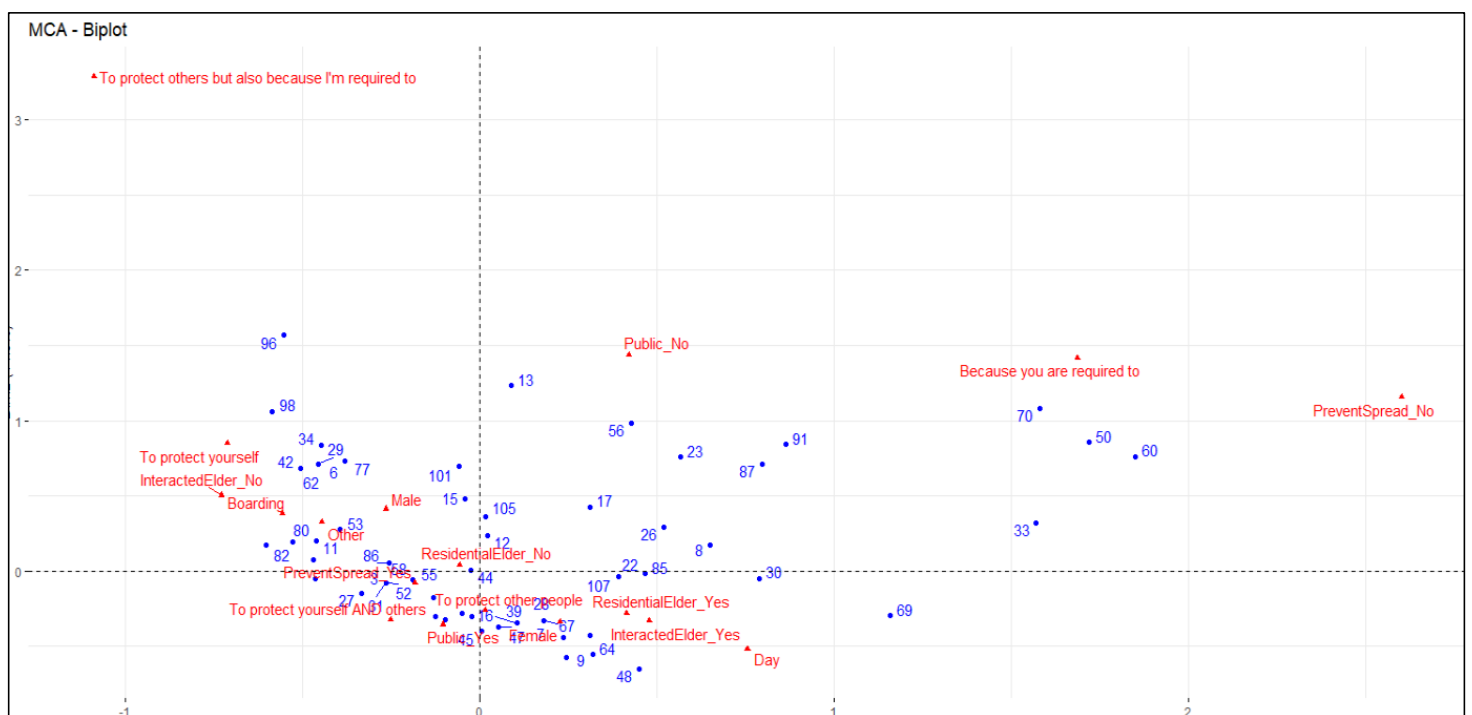
```
# Illustration de la contribution absolue des individus
fviz_mca_ind(afcm, col.ind = "contrib",
              gradient.cols = c("#00AFBB", "blue", "red"),
              repel = TRUE, # Avoid text overlapping (slow if manypoints)
              ggtheme = theme_minimal())
```

Le résultat obtenu est donné ci-dessous :



7.11 Interprétation des résultats

On révoque dans cette partie, le biplot des individus-modalités montré ci-dessus :



L'interprétation des résultats d'une AFCM est souvent assez délicate, en raison de la faible décroissance des valeurs propres, et du grand nombre de modalités, ce qui rend les graphiques assez peu lisibles.

L'interprétation des proximités entre les modalités devra tenir compte des remarques suivantes:

- Si deux modalités d'une même variable sont proches, cela signifie que les individus qui possèdent l'une des modalités et ceux qui possèdent l'autre sont globalement similaires du point de vue des autres variables
- Si deux modalités de deux variables différentes sont proches, cela peut signifier que ce sont globalement les mêmes individus qui possèdent l'une et l'autre

L'interprétation des proximités entre les individus devra aussi tenir compte de la remarque suivante :

- La proximité de deux individus indique qu'ils ont à peu près choisi les mêmes modalités.

En se basant sur le biplot des individus-modalités illustré auparavant, on peut tirer les résultats suivant :

➤ Par rapport aux modalités :

- Les étudiants qui vivent avec une personne âgée de plus de 65 ans ont eu des contacts avec une personne âgée de plus de 65 ans au cours du mois dernier
- Les étudiants qui portent le masque en public, il le porte afin de protéger les autres
- Les étudiants qui portent le masque en public, il le porte pour se protéger et protéger les autres
- Les étudiants en internat n'ont pas eu des contacts avec une personne âgée de plus de 65 ans au cours du mois dernier
- Les étudiants qui n'ont pas eu des contacts avec une personne âgée de plus de 65 ans au cours du mois dernier, portent le masque pour protéger soit même
- Les étudiants en internat ne sont généralement pas de genre féminin
- Les étudiants de genre féminin, portent le masque pour protéger autres soi
- Les étudiants en externe sont généralement plus de genre féminin
- Les étudiants en externe ont généralement eu des contacts avec une personne âgée de plus de 65 ans au cours du mois dernier
- Les étudiants qui généralement portent le masque car ils sont obligés de le faire ne pensent pas que les masques sont efficaces pour prévenir la propagation du coronavirus et ils ne portent pas un masque dans les lieux publics lorsque cela n'est pas nécessaire
- Les étudiants qui portent le masque afin de protéger les autres, et ceux qui le portent pour se protéger et protéger les autres ont globalement répondu de manière similaires aux autres questions

➤ Par rapport aux individus :

- Les individus 52, 31 et 3 ont exactement la même projection dans le plan factoriel engendré par l'axe 1 et 2, on peut ainsi déduire qu'ils ont répondu exactement de la même manière au questionnaire
- Les individus 62, 42, 29 et 6 ont à peu près répondu de manière similaire au questionnaire
- Les individus 85 et 22 ont à peu près répondu de manière similaire au questionnaire

7.12 Les associations possibles :

Comme l'étude réalisée par ce questionnaire a pour but d'en savoir plus sur les habitudes et les croyances des jeunes adultes concernant le port des masques, on peut ainsi généraliser les résultats d'interprétation comme suit :

- Les jeunes adultes qui vivent avec une personne âgée de plus de 65 ans ont eu des contacts avec une personne âgée de plus de 65 ans au cours du mois dernier
- Les jeunes adultes qui portent le masque en public, il le porte afin de protéger les autres
- Les jeunes adultes qui portent le masque en public, il le porte pour se protéger et protéger les autres
- Les jeunes adultes en internat n'ont pas eu des contacts avec une personne âgée de plus de 65 ans au cours du mois dernier
- Les jeunes adultes qui n'ont pas eu des contacts avec une personne âgée de plus de 65 ans au cours du mois dernier, portent le masque pour protéger soit même
- Les jeunes adultes en internat ne sont généralement pas de genre féminin
- Les jeunes adultes de genre féminin, portent le masque pour protéger autres soi
- Les jeunes adultes en externe sont généralement plus de genre féminin
- Les jeunes adultes en externe ont généralement eu des contacts avec une personne âgée de plus de 65 ans au cours du mois dernier
- Les jeunes adultes qui généralement portent le masque car ils sont obligés de le faire ne pensent pas que les masques sont efficaces pour prévenir la propagation du coronavirus et ils ne portent pas un masque dans les lieux publics lorsque cela n'est pas nécessaire
- Les jeunes adultes qui portent le masque afin de protéger les autres, et ceux qui le portent pour se protéger et protéger ont des opinions similaires pour les autres questions

Remarque

Malheureusement, il est difficile de tirer des conclusions générales sur les jeunes porteurs de masques, car les données ne concernent que les étudiants en statistiques AP. Il faut donc prendre les résultats de cette enquête avec un grain de sel, car il est probable qu'ils ne représentent pas les convictions d'une large population de jeunes adultes.

7.13 Qualité de représentation des modalités

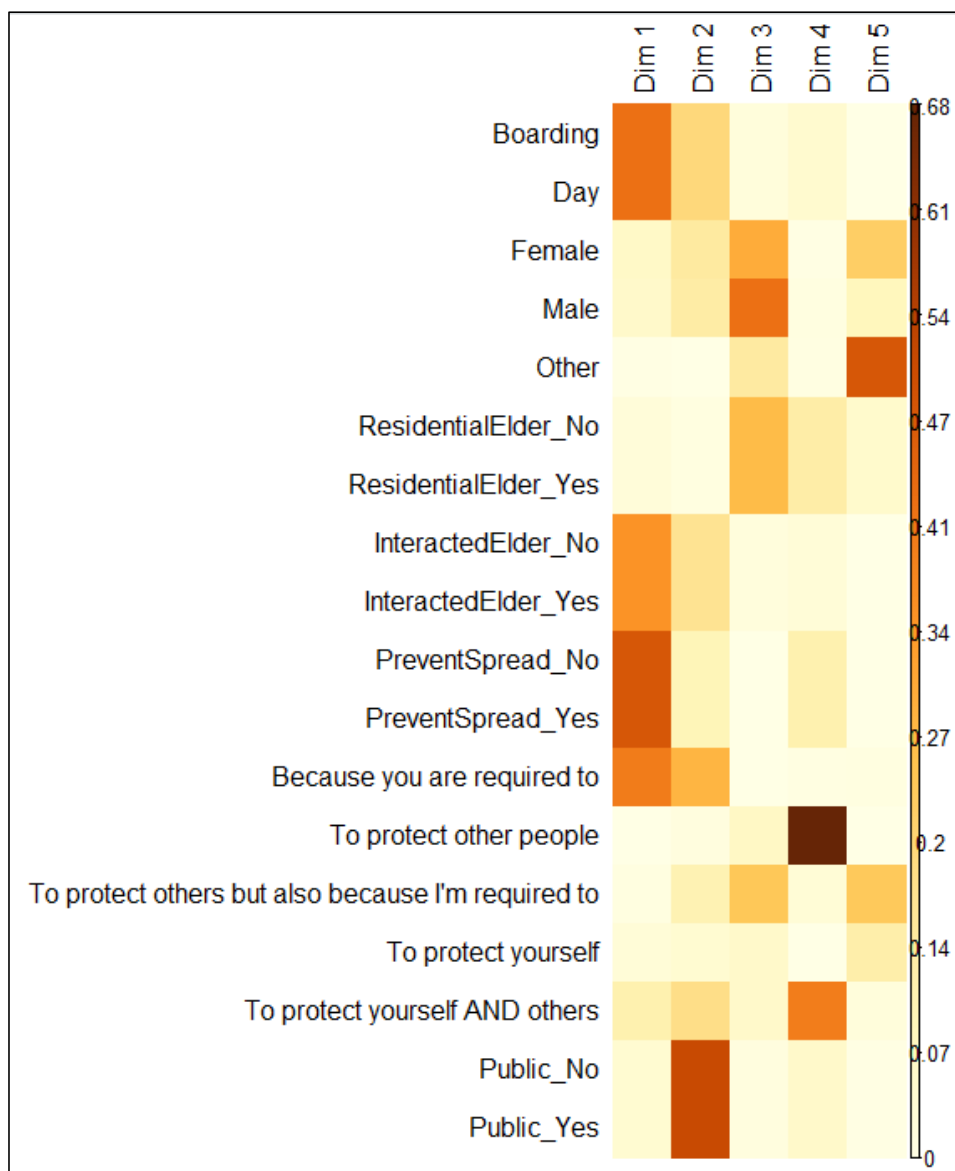
Nous avons utilisé l'instruction suivante pour extraire le **tableau des contributions relatives** de chaque modalité de nos variables qualitatives issus de l'AFCM appliquée à notre tableau de données.

```
# Afficher le tableau des contributions relatives des modalités issue de  
notre AFCM  
afcm$var$cos2
```

Nous obtenons le résultat suivant :

Modalités	Dim 1	Dim 2
Boarding	4.206658e-01	0.19912102
Day	4.206658e-01	0.19912102
Female	6.197247e-02	0.13881839
Male	5.429445e-02	0.13092016
Other	3.839784e-03	0.00203072
ResidentialElder_No	2.389895e-02	0.01122943
ResidentialElder_Yes	2.389895e-02	0.01122943
InteractedElder_No	3.489060e-01	0.16674369
InteractedElder_Yes	3.489060e-01	0.16674369
PreventSpread_No	4.789869e-01	0.09456558
PreventSpread_Yes	4.789869e-01	0.09456558
Because you are required to	3.980596e-01	0.28156516
To protect other people	4.994576e-05	0.01517096
To protect others but also because I'm required to	1.129603e-02	0.10291031
To protect yourself	3.045219e-02	0.04338876
To protect yourself AND others	1.088030e-01	0.18240916
Public_No	4.388589e-02	0.51173821
Public_Yes	4.388589e-02	0.51173821

Nous avons également représenté graphiquement les contributions relatives comme suit :



7.13.1 Qualité de représentation sur le 1^{er} axe :

Modalités	Qualité de représentation
Boarding	Mal représenté
Day	Mal représenté
Female	Très mal représenté
Male	Très mal représenté
Other	Très mal représenté
ResidentialElder_No	Très mal représenté
ResidentialElder_Yes	Très mal représenté
InteractedElder_No	Mal représenté
InteractedElder_Yes	Mal représenté
PreventSpread_No	Mal représenté
PreventSpread_Yes	Mal représenté
Because you are required to	Mal représenté
To protect other people	Très mal représenté
To protect others but also because I'm required to	Très mal représenté

To protect yourself	Très mal représenté
To protect yourself AND others	Mal représenté
Public_No	Très mal représenté
Public_Yes	Très mal représenté

7.13.2 Qualité de représentation sur le 2^{ème} axe :

Modalité	Qualité de représentation
Boarding	Très mal représentée
Day	Très mal représentée
Female	Très mal représentée
Male	Très mal représentée
Other	Très mal représentée
ResidentialElder_No	Très mal représentée
ResidentialElder_Yes	Très mal représentée
InteractedElder_No	Très mal représentée
InteractedElder_Yes	Très mal représentée
PreventSpread_No	Très mal représentée
PreventSpread_Yes	Très mal représentée
Because you are required to	Mal représentée
To protect other people	Très mal représentée
To protect others but also because I'm required to	Très mal représentée
To protect yourself	Très mal représentée
To protect yourself AND others	Très mal représentée
Public_No	Moyennement assez bien représentée
Public_Yes	Moyennement assez bien représentée

7.13.3 Sur le plan engendré par l'axe 1 et 2

Modalité	cos12	Qualité de représentation
Boarding	0.619786873	Bien représenté
Day	0.619786873	Bien représenté
Female	0.200790865	Très mal représentée
Male	0.185214606	Très mal représentée
Other	0.005870504	Très mal représentée
ResidentialElder_No	0.035128383	Très mal représentée
ResidentialElder_Yes	0.035128383	Très mal représentée
InteractedElder_No	0.515649682	Moyennement assez bien représentée
InteractedElder_Yes	0.515649682	Moyennement assez bien représentée
PreventSpread_No	0.573552481	Moyennement assez bien représentée
PreventSpread_Yes	0.573552481	Moyennement assez bien représentée
Because you are required to	0.679624726	Bien représentée
To protect other people	0.015220906	Très mal représentée

To protect others but also because I'm required to	0.114206344	Très mal représentée
To protect yourself	0.073840942	Très mal représentée
To protect yourself AND others	0.291212177	Mal représentée
Public_No	0.555624100	Moyennement assez bien représentée
Public_Yes	0.555624100	Moyennement assez bien représentée

7.14 Les questions les mieux représentées par l'AFCM

Une question (variable qualitative) est aussi bien représentée que ses modalités.

En se basant sur le tableau de la qualité de représentation des modalités sur le plan engendré par l'axe 1 et 2 (section « 17.3 Sur le plan engendré par l'axe 1 et 2 »), on peut déduire que les questions les mieux représentées sont :

- Êtes-vous un étudiant externe ou en internat ? (variable Boarding)
- Pensez-vous que les masques sont efficaces pour prévenir la propagation du coronavirus ? (variable PreventSpread)
- Portez-vous un masque dans les lieux publics lorsque cela n'est pas nécessaire ? (variable Public)
- Avez-vous eu des contacts avec une personne âgée de plus de 65 ans au cours du mois dernier ? (variable InteractedElder)

8 AFCM ajustée

D'après l'étude déjà faite sur notre tableau de données en utilisant l'AFCM, nous avons remarqué que la modalité « *To protect others but also because I'm required to* » est très loin de l'origine dans le graphe de représentation des modalités, en effet cette modalité a un poids très faible (0.001347709), mais elle contribue à la construction du premier et deuxième axe. En effet, cela revient au fait qu'un seul individu a choisi cette réponse est donc cet individu est considéré comme étant une donnée aberrante. On conclut, que la modalité « *To protect others but also because I'm required to* » est une modalité rare.

Nous allons dans ce qui suit, supprimer l'individu qui a choisi cette modalité (donnée aberrante) et effectuer l'AFCM une autre fois sur ce nouveau tableau de données.

Nous avons utilisé l'instruction suivante pour afficher la donnée aberrante :

```
# Afficher qui a choisit la modalité « To protect others but also
because I'm required to »
data[data$Reason == "To protect others but also because I'm required
to",]
```

On obtient le résultat suivant :

```
> data[data$Reason == "To protect others but also because I'm required to",]
  Boarding Gender ResidentialElder InteractedElder PreventSpread Reason Public
96 Boarding Female                No                No      Yes To protect others but also because I'm required to      No
```

Nous avons utilisé l'instruction suivante pour supprimer la donnée aberrante

```
# Supprimer la donnée aberrante  
data <- subset(data, data$Reason != "To protect others but also because  
I'm required to")
```

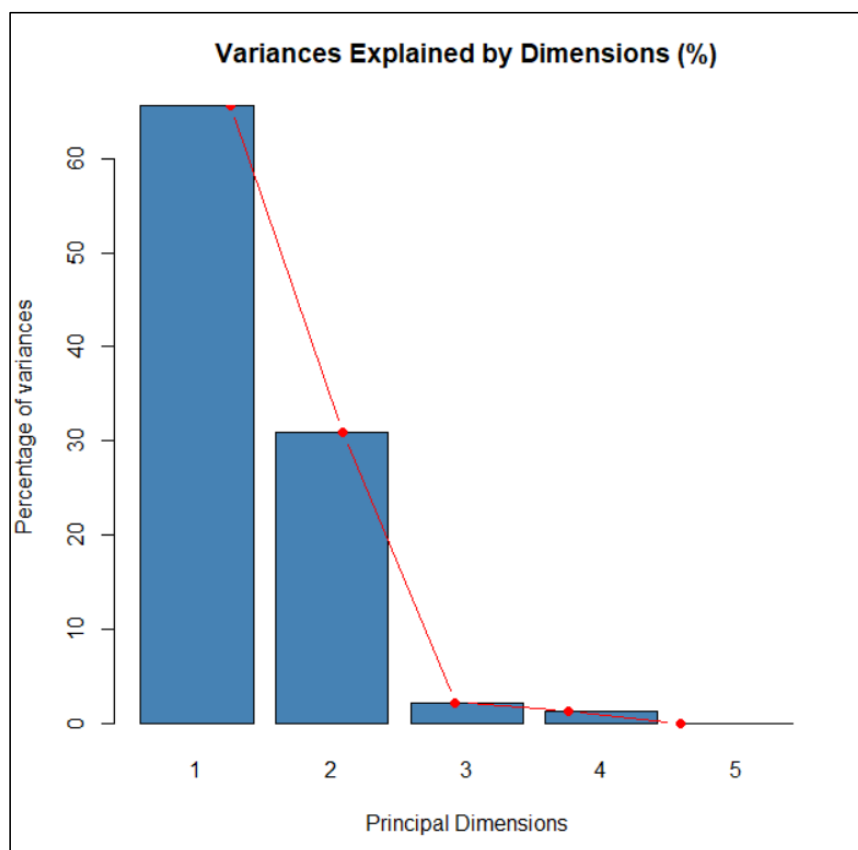
8.1 Etude du tableau des valeurs propres ajustées :

De la même manière, nous allons utiliser l'inertie corrigée de Benzeckri pour le choix des axes à retenir.

Voici ci-dessous le tableau de valeurs propres corrigées :

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	1.833929e-02	65.576507967	65.57651
dim 2	8.632789e-03	30.868600274	96.44511
dim 3	6.230906e-04	2.228009492	98.67312
dim 4	3.687546e-04	1.318570320	99.99169
dim 5	2.324540e-06	0.008311948	100.00000

L'éboulis des valeurs propres corrigés est représenté ci-dessus :



D'après l'éboulis des valeurs propres corrigées, le coude se trouve entre la 2ème et la 3ème valeur. Donc on retient les 2 premiers axes. (On a aussi $I_{12}(\%) = 96.44511 > 80\%$).

8.2 Qualité de représentation globale

$$I_{12} = \frac{\lambda'1 + \lambda'2}{\sum \lambda'i}$$

$I_{12} = 96.44511\%$, donc 96.44511% de l'information est contenue dans le plan engendré par le premier et le deuxième axe.

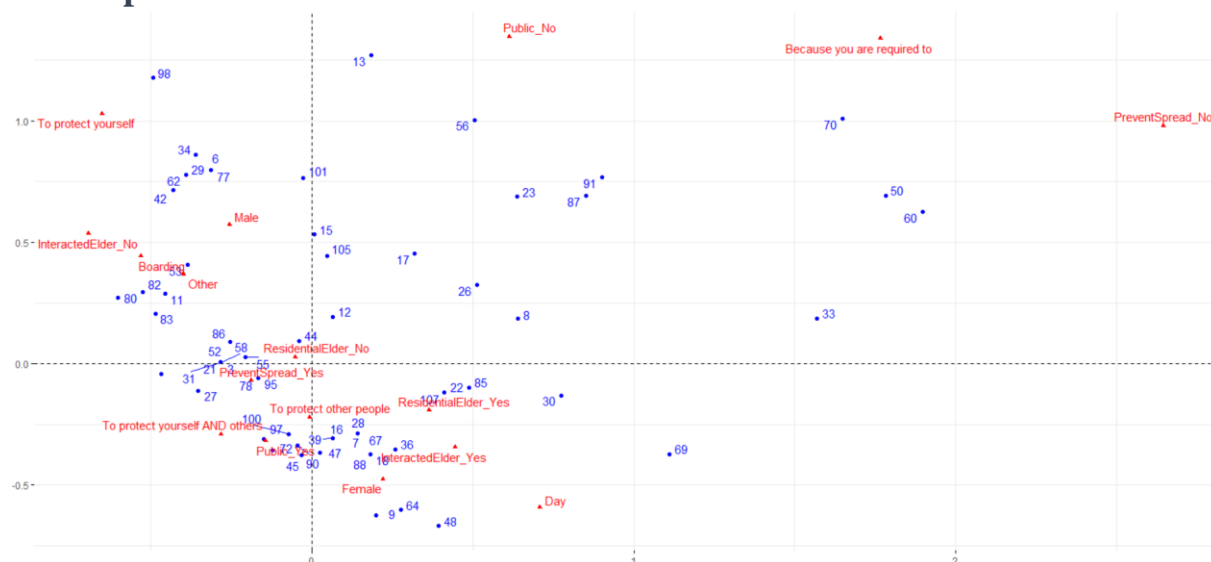
On remarque une fine amélioration de cette représentation globale par rapport à la qualité de représentation globale avec la donnée aberrante.

8.3 Tableau des projections des modalités

	Dim 1	Dim 2
Boarding	-0.531139877	0.44254592
Day	0.708186502	-0.59006123
Female	0.220736328	-0.47405925
Male	-0.256150413	0.57137712
Other	-0.399525857	0.36901485
ResidentialElder_No	-0.051377078	0.02719213
ResidentialElder_Yes	0.363591626	-0.19243662
InteractedElder_No	-0.695080370	0.53603057
InteractedElder_Yes	0.445285862	-0.34339458
PreventSpread_No	2.645507652	0.97797838
PreventSpread_Yes	-0.188964832	-0.06985560
Because you are required to	1.766426469	1.33752175
To protect other people	-0.006669977	-0.22039102
To protect yourself	-0.652446809	1.02938236
To protect yourself AND others	-0.282419906	-0.28920369
Public_No	0.613325232	1.34698598
Public_Yes	-0.144311819	-0.31693788

On remarque une fine différence entre ces projections des modalités par rapport à la qualité les projections des modalités avec la donnée aberrante.

8.4 Biplot des individus-modalités



On remarque que ce biplot des individus-modalités est presque identique au biplot des individus-modalités en considérant la donnée aberrante.

8.5 Signification des axes :

Voici ci-dessous le **tableau des contributions** de chaque modalités de nos variables qualitatives issus de l'AFCM appliquée au nouveau tableau de données

	Dim 1	Dim 2
Boarding	8.893919e+00	7.18550080
Day	1.185856e+01	9.58066774
Female	1.459307e+00	7.83301971
Male	1.585886e+00	9.18317028
Other	1.677428e-01	0.16653549
ResidentialElder_No	1.276000e-01	0.04159713
ResidentialElder_Yes	9.030155e-01	0.29437972
InteractedElder_No	1.040826e+01	7.20364129
InteractedElder_Yes	6.667790e+00	4.61483270
PreventSpread_No	2.574188e+01	4.09397947
PreventSpread_Yes	1.838705e+00	0.29242711
Because you are required to	2.131369e+01	14.22113194
To protect other people	4.441466e-04	0.56432658
To protect yourself	1.342040e+00	3.88770948
To protect yourself AND others	2.807951e+00	3.42666448
Public_No	3.953082e+00	22.18938445
Public_Yes	9.301368e-01	5.22103164

On remarque que les contributions de chaque modalité sont presque identiques aux contributions de modalités en considérant la donnée aberrante.

Voici ci-dessous le tableau qui résume le poids de chaque modalité de nos variables qualitatives :

Modalité	Poids
Boarding	0.081632653
Day	0.061224490
Female	0.077551020
Male	0.062585034
Other	0.002721088
ResidentialElder_No	0.125170068
ResidentialElder_Yes	0.017687075
InteractedElder_No	0.055782313
InteractedElder_Yes	0.087074830
PreventSpread_No	0.009523810
PreventSpread_Yes	0.133333333
Because you are required to	0.017687075
To protect other people	0.025850340
To protect yourself	0.008163265
To protect yourself AND others	0.091156463
Public_No	0.027210884
Public_Yes	0.115646259

8.5.1 Signification du 1^{er} axe

+	-
Day, PreventSpread_No, Because you are required to, Public_No	Boarding, InteractedElder_No, To protect yourself

L'axe 1 est un axe à effet opposition. Il oppose les étudiants externes qui ne pensent pas que les masques sont efficaces pour prévenir la propagation du coronavirus et qu'ils les portent, car ils le sont obligés et qui ne portent pas un masque dans les lieux publics lorsque cela n'est pas nécessaire, aux étudiants en internat et qui n'ont pas eu des contacts avec une personne âgée de plus de 65 ans au cours du mois dernier et ils portent le masque pour se protéger soi-même.

8.5.2 Signification du 2^{ème} axe :

+	-
Male , InteractedElder_No, PreventSpread_No, Because you are required to, To protect others but also because I'm required to, To protect yourself, Public_No	Day, Female

L'axe 2 est un axe à effet opposition. Il oppose les étudiantes de genre féminin en internat, aux étudiants de genre masculin qui n'ont pas eu des contacts avec une personne âgée de plus de 65 ans au cours du mois dernier, et qui ne pensent pas que les masques sont efficaces pour prévenir la propagation du coronavirus, et qu'ils portent les car ils le sont obligés de le faire, pour protéger les autres mais aussi parce qu'ils sont obligés de le faire ou pour se protéger, et qui ne portent pas un masque dans les lieux publics lorsque cela n'est pas nécessaire.

Remarque

Nous remarquons que contrairement à l'étude faite en prenant en compte la donnée aberrante :

- La modalité Public_No contribue positivement à la construction du 1^{er} axe factoriel
- La modalité Male contribue positivement à la construction du 2^{ème} axe factoriel
- La modalité Female contribue négativement à la construction du 2^{ème} axe factoriel

Mais on remarque que leurs contributions absolues est très proche de leur poids.

8.6 Interprétation des résultats

On révoque dans cette partie, le biplot des individus-modalités montré ci-dessus :



Nous remarquons, qu'on utilisant ce biplot d'individus-modalités on aboutira aux mêmes interprétations et associations obtenues dans l'étude qui prend en compte la donnée aberrante.

De ce fait dans l'AFC, nous allons considérer le tableau de données avec la donnée aberrante vu que l'interprétation et les associations dégagées restent inchangées.

9 AFC

9.1 Création du tableau de contingence

Pour notre étude, les questions les plus pertinentes sont :

- Portez-vous un masque dans les lieux publics lorsque cela n'est pas nécessaire ? (variable Public)
- Quelle est votre principale raison de porter un masque ? (variable Reason)

Pour vérifier que les variables choisies (les questions choisies) sont corrélées eux, on calcule la distance de KHI2 entre ces deux variables.

On pose alors,

H0 : " Les 2 variables ne sont pas corrélées entre eux" VS **H1 :** " Les 2 variables sont corrélées entre eux"

Nous avons utilisé l'instruction suivante pour calculer la distance de KHI2 entre les deux variables Public et Reason :

```
# Calcul de la distance de khi 2 entre Public et Reason
khisq <- chisq.test(N)
```

On obtient le résultat suivant :

```
data: N
X-squared = 17.933, df = 4, p-value = 0.001272
```

Comme $d^2(\text{Public}, \text{Reason}) = 17.933 \gg p - \text{value} = 0.001272$, On rejette alors H0 et donc les deux variables choisies pour l'AFC sont fortement corrélées entre eux

Nous avons aussi utilisé l'instruction suivante pour créer le tableau de contingence croisant les deux questions pertinentes choisies auparavant :

```
# Tableau de contingence entre la variable Public et Reason
N = t(Z[, 12:16]) %*% Z[, 17:18]
```

On obtient le résultat suivant :

	No	Yes
Because you are required to	7	6
To protect other people	4	15
To protect others but also because I'm required to	1	0
To protect yourself	2	4
To protect yourself AND others	7	60

9.2 Application de l'AFC sur le tableau de contingence

Nous avons utilisé la library «*FactoMineR*» pour appliquer l'AFC sur le tableau de contingence

```
# Application de l'AFC sur le tableau de contingence)
afc <- CA(N)
```

9.3 Etude du tableau des valeurs propres

L'AFC appliquée sur le tableau de contingence n'a produit qu'une seule valeur propre. En effet, en AFC Le nombre de valeurs propres est égale au $\min(n - 1; p - 1)$.

```
> afc$eig
      eigenvalue percentage of variance cumulative percentage of variance
dim 1  0.1691825                100                100
```

De ce fait, on va retenir qu'un seul axe pour représenter nos deux variables (questions) avec une qualité de représentation globale de 100% .

9.4 Tableaux de projection

En appliquant l'AFC sur notre tableau de contingence on obtient le tableau de projection suivant :

9.4.1 Tableau des projections des profils lignes

Nous avons utilisé l'instruction suivante pour extraire **le tableau des projections** des profils lignes issus de l'AFC appliquée à tableau de contingence.

```
# Tableau des projections de profils lignes  
afc$row$coord
```

Nous obtenons le résultat suivant :

Profil ligne	Dim 1
Because you are required to	-0.85390662
To protect other people	-0.03114349
To protect others but also because I'm required to	-2.01186954
To protect yourself	-0.33925643
To protect yourself AND others	0.23492419

9.4.2 Tableau des projections des profils colonnes

Nous avons utilisé l'instruction suivante pour extraire **le tableau des projections** des profils colonnes issus de l'AFC appliquée à tableau de contingence.

```
# Tableau des projections de profils colonnes  
afc$col$coord
```

Nous obtenons le résultat suivant :

Profil ligne	Dim 1
No	-0.8275180
Yes	0.2044456

9.5 Signification du 1^{er} axe :

Afin d'extraire **le tableau des contributions des profils lignes** issus de l'AFC appliquée à tableau de contingence.

```
# Tableau des contributions de profils lignes  
afc$row$contrib
```

Nous obtenons le résultat suivant :

Profil ligne	Dim 1
Because you are required to	52.8570518
To protect other people	0.1027607
To protect others but also because I'm required to	22.5703584
To protect yourself	3.8507583
To protect yourself AND others	20.6190709

Afin d'extraire **le tableau des contributions des profils colonnes** issus de l'AFC appliquée à tableau de contingence.

```
# Tableau des contributions de profils colonnes  
afc$col$contrib
```

Nous obtenons le résultat suivant :

Profil ligne	Dim 1
No	80.18868
Yes	19.81132

On dira qu'un profil ligne a contribué à la construction d'un axe α si $\text{cab}^\alpha(i) > 100 * f_i$.
On dira qu'un profil colonne a contribué à la construction d'un axe α si $\text{cab}^\alpha(j) > 100 * f_j$

Pour afficher le poids des profils lignes, nous avons utilisé l'instruction suivante :

```
# Poids des profils lignes
afc$call$marge.row
```

Nous obtenons le résultat suivant :

Profil ligne	Poids
Because you are required to	0.122641509
To protect other people	0.179245283
To protect others but also because I'm required to	0.009433962
To protect yourself	0.056603774
To protect yourself AND others	0.632075472

Pour afficher le poids des profils colonnes, nous avons utilisé l'instruction suivante :

```
# Poids des profils colonnes
afc$call$marge.col
```

Nous obtenons le résultat suivant :

Profil ligne	Dim 1
No	0.1981132
Yes	0.8018868

PL's		PC's	
+	-	+	-
	Because you are required to, To protect others but also because I'm required to		No

L'axe 1 est un axe à effet taille, il qualifie les étudiants qui ne portent pas de masque dans les lieux publics lorsque cela n'est pas nécessaire, représentés par les étudiants qu'ils portent les masques, car ils le sont obligés de le faire ou bien pour protéger les autres, mais aussi parce qu'ils sont obligés de le faire.

9.6 Visualisation des résultats

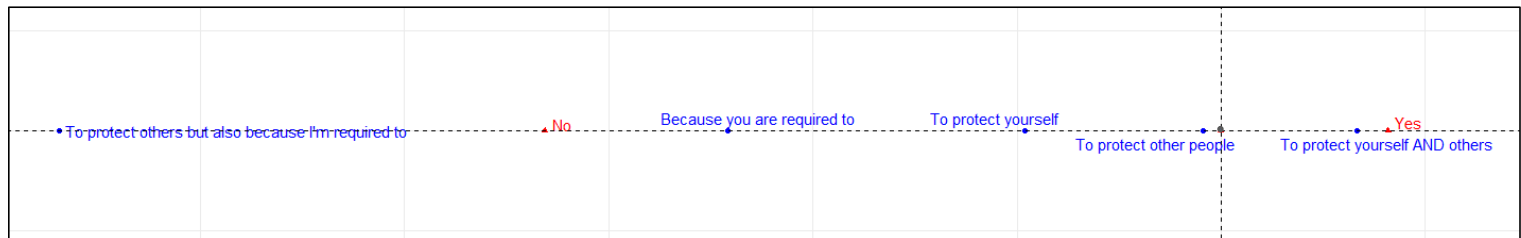
Vu qu'on a retenu un seul axe pour étude, les profils lignes et colonnes seront représentés sur un plan d'une dimension (un axe), mais comme R ne permet pas de visualiser CA factor map sur une dimension (un axe), nous avons dû ajouter un profil colonne représentant la somme des profils colonnes pour chaque ligne, afin d'avoir une deuxième dimension.

La représentation sur ce nouveau plan est identique à celle sur une dimension (un axe) car le profil colonne ajouté est linéairement dépendant donc il va générer une valeur propre nulle. Pour cela, nous avons procédé comme suit :

```
# Représentation graphique de l'AFC
Z2 = Z[, 17:18]
Z2 <- data.frame(Z2)
# Comme la somme est linéairement dépendante elle ne va pas influencer les
résultats
Z2$` ` = Z2$No + Z2$Yes
N = t(Z[, 12:16]) %*% as.matrix(Z2)

# Application de l'AFC sur le nouveau tableau de contingence
afc <- CA(N)
```

Le résultat est donné ci-dessous :



9.7 Interprétation des résultats

En se basant sur la représentation illustrée auparavant, on peut tirer les résultats suivants :

- Les étudiants qui ne portent pas de masque dans les lieux publics lorsque cela n'est pas nécessaire généralement les portent car ils le sont obligés de le faire ou bien pour protéger les autres mais, aussi parce qu'ils sont obligés de le faire
- Les étudiants qui portent les masques dans les lieux publics lorsque cela n'est pas nécessaire généralement ne les portent pas car ils le sont obligés de le faire ou bien pour protéger les autres, mais aussi parce qu'ils sont obligés de le faire, mais ils les portent pour protéger les autres ou soi-même, les autres seulement ou uniquement soi-même

9.8 Associations possibles

- Les jeunes adultes qui ne portent pas de masque dans les lieux publics lorsque cela n'est pas nécessaire généralement les portent, car ils le sont obligés de le faire ou bien pour protéger les autres, mais aussi parce qu'ils sont obligés de le faire
- Les jeunes adultes qui portent les masques dans les lieux publics lorsque cela n'est pas nécessaire généralement ne les portent pas, car ils le sont obligés de le faire ou bien pour

protéger les autres, mais aussi parce qu'ils sont obligés de le faire, mais ils les portent pour protéger les autres ou soi-même, les autres seulement ou uniquement soi-même

10 Conclusion

Afin d'en savoir plus sur les habitudes et les croyances des jeunes adultes concernant le port des masques et répondre à la problématique lancée par le questionnaire utilisé. Nous avons effectué une AFCM et une AFC sur notre ensemble de données, ensuite nous avons analysé et interprété les différents résultats issus de ses analyses afin d'aboutir à des conclusions et des associations qui vont nous aider à prendre des décisions judicieuses et convenables.

Parmi les associations les plus pertinentes qu'on a dégagé depuis cette étude sont les suivantes :

- Les jeunes adultes qui ne portent pas de masque dans les lieux publics lorsque cela n'est pas nécessaire généralement les portent, car ils le sont obligés de le faire ou bien pour protéger les autres, mais aussi parce qu'ils sont obligés de le faire.
- Les jeunes adultes qui portent les masques dans les lieux publics lorsque cela n'est pas nécessaire généralement ne les portent pas, car ils le sont obligés de le faire ou bien pour protéger les autres, mais aussi parce qu'ils sont obligés de le faire, mais ils les portent pour protéger les autres ou soi-même, les autres seulement ou uniquement soi-même.

11 Annexe : Script complet

```
##          Library's utilisees          ##

library(tidyverse)
library("FactoMineR")
library("factoextra")
library("corrplot")
library(explor)

##          Lecture des donnees          ##

# na.strings=c("") : remplacer les chaines de caracteres vides avec NA
data <- read.csv("MaskBeliefs.csv", header = TRUE, na.strings=c(""))

# Suppression des colonnes qui ne representent pas des variables
qualitatifs
data$Timestamp<- NULL
data$Age<- NULL
data$Restaurant <- NULL

# Affichage des noms de colonnes retenues
colnames(data)

# Affichage du nombre de valeurs manquantes
sum_na <- sum(is.na(data))
print(paste("Le nombre des NAs dans notre dataset est :", sum_na))

# Suppression des NAs
data = na.omit(data)

# Reaffichage du nombre de valeurs manquantes pour verification
sum_na_apres <- sum(is.na(data))
print(paste("Le nombre des NAs dans notre dataset est :", sum_na_apres))

##          Etude statistique          ##

# Afficher le tableau des statistiques de la variable Boarding
Boardingstat <- fct_count(data$Boarding , prop = TRUE, sort = TRUE) %>%
  rename("Boarding" = "f", "freq" = "p", "nb" = "n") %>%
  mutate(freq = round(freq, 2))
Boardingstat

# Plotter la distribution de la variable Boarding
ggplot(data.frame(data), aes(x = Boarding)) +
  geom_bar(colour = "black", fill = "#6082B6")+
  labs(title = "La distribution de la variable 'Boarding' ") +
  theme(plot.title = element_text(hjust = 0.5, size=17), axis.text.x=
element_text(size=14),
        axis.text.y= element_text(size=14),
axis.title=element_text(size=16))

# Plotter le diagramme de pie chart de la variable Boarding
ggplot(Boardingstat , aes(x="", y= nb, fill= Boarding)) +
  scale_fill_brewer("Blues") +
  geom_bar(stat="identity", width=1, , color = "black") +
  labs(title = "La distribution de la variable 'Boarding' ") +
  coord_polar("y", start=0) +
```

```

theme_void() +
theme(plot.title = element_text(hjust = 0.5, size=17))

# Afficher le tableau des statistiques de la variable Gender
fct_count(data$Gender , prop = TRUE, sort = TRUE) %>%
  rename("Gender" = "f", "freq" = "p", "nb" = "n") %>%
  mutate(freq = round(freq, 2))

# Plotter la distribution de la variable Gender
ggplot(data.frame(data), aes(x = Gender)) +
  geom_bar(colour = "black", fill = "#6082B6")+
  labs(title = "La distribution de la variable 'Gender' ") +
  theme(plot.title = element_text(hjust = 0.5, size=17), axis.text.x=
element_text(size=14),
        axis.text.y= element_text(size=14),
axis.title=element_text(size=16))

# Afficher le tableau des statistiques de la variable ResidentialElder
fct_count(data$ResidentialElder , prop = TRUE, sort = TRUE) %>%
  rename("ResidentialElder" = "f", "freq" = "p", "nb" = "n") %>%
  mutate(freq = round(freq, 2))

# Plotter la distribution de la variable ResidentialElder
ggplot(data.frame(data), aes(x = ResidentialElder)) +
  geom_bar(colour = "black", fill = "#6082B6")+
  labs(title = "La distribution de la variable 'ResidentialElder'") +
  theme(plot.title = element_text(hjust = 0.5, size=17), axis.text.x=
element_text(size=14),
        axis.text.y= element_text(size=14),
axis.title=element_text(size=16))

# Afficher le tableau des statistiques de la variable InteractedElder
fct_count(data$InteractedElder , prop = TRUE, sort = TRUE) %>%
  rename("InteractedElder" = "f", "freq" = "p", "nb" = "n") %>%
  mutate(freq = round(freq, 2))

# Plotter la distribution de la variable InteractedElder
ggplot(data.frame(data), aes(x = InteractedElder)) +
  geom_bar(colour = "black", fill = "#6082B6")+
  labs(title = "La distribution de la variable 'InteractedElder'") +
  theme(plot.title = element_text(hjust = 0.5, size=17), axis.text.x=
element_text(size=14),
        axis.text.y= element_text(size=14),
axis.title=element_text(size=16))

# Afficher le tableau des statistiques de la variable PreventSpread
fct_count(data$PreventSpread , prop = TRUE, sort = TRUE) %>%
  rename("PreventSpread" = "f", "freq" = "p", "nb" = "n") %>%
  mutate(freq = round(freq, 2))

```



```

# Plotter la distribution de la variable PreventSpread
ggplot(data.frame(data), aes(x = PreventSpread)) +
  geom_bar(colour = "black", fill = "#6082B6")+
  labs(title = "La distribution de la variable 'PreventSpread'") +
  theme(plot.title = element_text(hjust = 0.5, size=17), axis.text.x=
element_text(size=14),
        axis.text.y= element_text(size=14),
axis.title=element_text(size=16))

# Afficher le tableau des statistiques de la variable Reason
Reasonstat <- fct_count(data$Reason , prop = TRUE, sort = TRUE) %>%
  rename("Reason" = "f", "freq" = "p", "nb" = "n") %>%
  mutate(freq = round(freq, 2))
Reasonstat

# Plotter la distribution de la variable Reason
ggplot(data.frame(data), aes(x = Reason)) +
  geom_bar(colour = "black", fill = "#6082B6")+
  labs(title = "La distribution de la variable 'Reason' ") +
  theme(plot.title = element_text(hjust = 0.5, size=17), axis.text.x=
element_text(angle = 60, hjust = 1, size=10),
        axis.text.y= element_text(size=14),
axis.title=element_text(size=16))

# Plotter le diagramme de pie chart de la variable Reason
ggplot(Reasonstat , aes(x="", y= nb, fill= Reason)) +
  scale_fill_brewer("Blues") +
  geom_bar(stat="identity", width=1, , color = "black") +
  labs(title = "La distribution de la variable 'Reason' ") +
  coord_polar("y", start=0) +
  theme_void() +
  theme(plot.title = element_text(hjust = 0.5, size=17))

# Afficher le tableau des statistiques de la variable Public
fct_count(data$Public , prop = TRUE, sort = TRUE) %>%
  rename("Public" = "f", "freq" = "p", "nb" = "n") %>%
  mutate(freq = round(freq, 2))

# Plotter la distribution de la variable Public
ggplot(data.frame(data), aes(x = Public)) +
  geom_bar(colour = "black", fill = "#6082B6")+
  labs(title = "La distribution de la variable 'Public'") +
  theme(plot.title = element_text(hjust = 0.5, size=17), axis.text.x=
element_text(size=14),
        axis.text.y= element_text(size=14),
axis.title=element_text(size=16))

##      Le tableau disjonctif complet      ##

# Rendre le type des colonnes comme factor
data<- data %>%
  mutate(Boarding = factor(Boarding),

```

```

        Gender = factor(Gender),
        ResidentialElder = factor(ResidentialElder),
        InteractedElder = factor(InteractedElder),
        PreventSpread = factor(PreventSpread),
        Reason = factor(Reason),
        Public = factor(Public))

# Tableau disjonctif complet de data
Z <- tab.disjonctif(data)
head(Z)
colnames(Z)

##          Application de l'AFCM          ##

# Appliquer l'AFCM sur le tableau de donnee (data)
afcm <- MCA(data)

# Afficher le tableau des valeurs propres issue de notre AFCM
afcm$eig

# Representation de l'eboulis des valeurs propres
fviz_eig(afcm)

# Tableau de valeurs propres corrig  s
p = ncol(data) # Le nombre de questions

# Correction des valeurs propres
eigs_corr = subset(afcm$eig, afcm$eig[, "eigenvalue"] > 1/p)
eigs_corr[, "eigenvalue"] = ((p/(p-1)) * (eigs_corr[, "eigenvalue"] -
(1/p))) ** 2

# Calcul de l'inertie des nouvelle valeur propore
eigs_corr[, "percentage of variance"] = (eigs_corr[, "eigenvalue"] /
sum(eigs_corr[, "eigenvalue"])) * 100

# Calcul de l'inertie cummul  e
eigs_corr[1, "cumulative percentage of variance"] = eigs_corr[1,
"percentage of variance"]
for(i in 2:nrow(eigs_corr)) {
    eigs_corr[i, "cumulative percentage of variance"] = eigs_corr[i-1,
"cumulative percentage of variance"] + eigs_corr[i, "percentage of
variance"]
}

# Affichage de l'eboulie des valeurs propres corrig  es
eig.val <- eigs_corr
barplot(eig.val[, 2],
        names.arg = 1:nrow(eig.val),
        main = "Variances Explained by Dimensions (%)",
        xlab = "Principal Dimensions",
        ylab = "Percentage of variances",
        col = "steelblue")

```

```

# Add connected line segments to the plot
lines(x = 1:nrow(eig.val), eig.val[, 2],
      type = "b", pch = 19, col = "red")

# Illustration des modalités dans le plan factoriel
plot(afcm,
      invisible = c("ind", "quali.sup", "quanti.sup"),
      cex = 0.8,
      autoLab = "yes")

# Illustration du biplot individus-modalités
fviz_mca_biplot(afcm, repel = TRUE,
                ggtheme = theme_minimal())

# Afficher le tableau des projections des modalités issue de notre AFCM
afcm$var$coord

# Afficher le tableau des contributions issue de notre AFCM
afcm$var$contrib

# Afficher les poids des modalités
afcm$call$marge.col

# Afficher le tableau des contributions relatives des modalités issue de
notre AFCM
afcm$var$cos2

# Afficher le tableau des projections des modalités issue de notre AFCM
afcm$var$eta2

# Visualiser les individus colorés par rapport à la variable publique
explor(afcm)

##          Visualisations          ##

# Illustration de la qualité de représentation des modalités
fviz_mca_var(afcm, col.var = "cos2",
             gradient.cols = c("#00AFBB", "blue", "red"),
             repel = TRUE, # Avoid text overlapping
             ggtheme = theme_minimal())

# Illustration de la contribution absolue des modalités
fviz_mca_var(afcm, col.var = "contrib",
             gradient.cols = c("#00AFBB", "blue", "red"),
             repel = TRUE, # Avoid text overlapping
             ggtheme = theme_minimal())

# Illustration de la qualité de représentation des individus
fviz_mca_ind(afcm, col.ind = "cos2",
             gradient.cols = c("#00AFBB", "blue", "red"),
             repel = TRUE, # Avoid text overlapping (slow if many points)
             ggtheme = theme_minimal())

# Illustration de la contribution absolue des individus

```

```

fviz_mca_ind(afcm, col.ind = "contrib",
             gradient.cols = c("#00AFBB", "blue", "red"),
             repel = TRUE, # Avoid text overlapping (slow if many points)
             ggtheme = theme_minimal())

# Illustration de la différence entre les contribution absolues et les
# poids des modalités

corrplot(afcm$var$contrib - 100*afcm$call$marge.col, is.corr=FALSE,
method = 'color', tl.col="black", col = COL2('RdBu', 200) )

# Illustration des contribution relatives des modalités
corrplot(afcm$var$cos2, is.corr=FALSE, method = 'color', tl.col="black" )

# Tableau de contingence entre la variable Public et Reason
N = t(Z[, 12:16]) %*% Z[, 17:18]
N

# Calcul de la distance de khi 2 entre Public et Reason
khisq <- chisq.test(N)
khisq

# Application de l'AFC sur le tableau de contingence
afc <- CA(N)

# Tableau des valeurs propres
afc$eig

# Tableaux des projections

# Tableau des projections de profils lignes
afc$row$coord

# Tableau des projections de profils colonnes
afc$col$coord

# Tableaux des contributions

# Tableau des contributions de profils lignes
afc$row$contrib

# Tableau des contributions de profils colonnes
afc$col$contrib

# Poids

# Poids des profils lignes
afc$call$marge.row

# Poids des profils colonnes
afc$call$marge.col

```

```

# ***** #
# Représentation graphique de l'AFC
Z2 = Z[, 17:18]
Z2 <- data.frame(Z2)
# Comme la somme est linéairement dépendante elle ne va pas influencer les
résultats
Z2$` ` = (Z2$No + Z2$Yes)
N = t(Z[, 12:16]) %*% as.matrix(Z2)

# Application de l'AFC sur le nouveau tableau de contingence
afc <- CA(N)
afc$col$coord = -afc$col$coord
afc$row$coord = -afc$row$coord
fviz_ca_biplot(afc) + ylim(-3, 3) + xlim(-1, 0.3)
fviz_ca(afc, repel = TRUE) + ylim(-3, 3) + xlim(-1.5, 0.3)

#          AFCM ajustée          #

# Afficher qui a choisi la modalité « To protect others but also
because I'm required to »
data[data$Reason == "To protect others but also because I'm required
to",]

# Supprimer la donnée aberrante
data <- subset(data, data$Reason != "To protect others but also because
I'm required to")

# Vérifier que la donnée aberrante a été supprimée
data[data$Reason == "To protect others but also because I'm required
to",]

# Appliquer l'AFCM sur le nouveau tableau de données
afcm_ajustée <- MCA(data)

# Afficher le tableau des valeurs propres issue de notre AFCM
afcm_ajustée$eig

# Représentation de l'éclat des valeurs propres
fviz_eig(afcm_ajustée)

# Tableau de valeurs propres corrigées
p = ncol(data) # Le nombre de questions

# Correction des valeurs propres
eigs_corr = subset(afcm_ajustée$eig, afcm_ajustée$eig[, "eigenvalue"] >
1/p)
eigs_corr[, "eigenvalue"] = ((p/(p-1)) * (eigs_corr[, "eigenvalue"] -
(1/p))) ** 2

# Calcul de l'inertie des nouvelles valeurs propres
eigs_corr[, "percentage of variance"] = (eigs_corr[, "eigenvalue"] /
sum(eigs_corr[, "eigenvalue"])) * 100

```

```

# Calcul de l'inertie cumulée
eigs_corr[1, "cumulative percentage of variance"] = eigs_corr[1,
"percentage of variance"]
for(i in 2:nrow(eigs_corr)) {
  eigs_corr[i, "cumulative percentage of variance"] = eigs_corr[i-1,
"cumulative percentage of variance"] + eigs_corr[i, "percentage of
variance"]
}

eigs_corr

# Affichage de l'eboulie des valeurs propres corrigées
eig.val <- eigs_corr
barplot(eig.val[, 2],
        names.arg = 1:nrow(eig.val),
        main = "Variances Explained by Dimensions (%)",
        xlab = "Principal Dimensions",
        ylab = "Percentage of variances",
        col = "steelblue")
# Add connected line segments to the plot
lines(x = 1:nrow(eig.val), eig.val[, 2],
      type = "b", pch = 19, col = "red")

# Illustration des modalités dans le plan factoriel
plot(afcm_ajustee,
     invisible = c("ind", "quali.sup", "quanti.sup"),
     cex = 0.8,
     autoLab = "yes")

# Illustration du biplot individus-modalités
fviz_mca_biplot(afcm_ajustee, repel = TRUE,
                ggtheme = theme_minimal())

# Afficher le tableau des projections des modalités issue de notre AFCM
afcm_ajustee$var$coord

# Afficher le tableau des contributions issue de notre AFCM
afcm_ajustee$var$contrib

# Afficher les poids des modalités
afcm_ajustee$call$marge.col

# Afficher le tableau des contributions relatives des modalités issue de
notre AFCM
afcm_ajustee$var$cos2

```