

ANAD

2^{ème} année Cycle Supérieur (2CS)

TP 2

Réalisation d'une Analyse Factorielle des Correspondances Multiples

Auteurs :

AMMALI Mouna (SIT1)

NOUI Moussa Akram (SIT1)

Table des matières

1. Introduction et objectifs :	4
2. Présentation questionnaire	4
3. Aperçu du dataset :	7
4. Réalisation de l'AFCM sous R :	7
4.1. Résumé du jeu de données	8
4.1.1. Variables actives	8
4.1.2. Variables signalétiques :	9
4.2. Tableau complet disjonctif	11
4.3. Lancement de l'AFCM :	12
4.4. Étude du tableau de valeurs propres :	12
4.4.1. <i>Explication</i> :	13
4.5. Étude du tableau des contributions	14
4.5.1. Signification des axes :	15
4.6. Étude des associations entre modalités :	16
4.7. Visualisations	17
4.8. Étude des individus :	18
5. Réalisation d'une AFC :	19
5.1. Interprétation des résultats de l'AFC :	19
5.1.1. Étude des valeurs propres :	19
5.1.2. Interprétation des axes :	20
6. Conclusion :	22

Table des figures

Figure 1 : Aperçu du jeu de données	7
Figure 2 : Observation des variables signalétique	9
Figure 3 : Visualisation des pourcentages de variances expliquées par chaque dimension de l'ACM.....	13
Figure 4 : Graphe des variables	17
Figure 5 : Graphe des modalités	17
Figure 6 : Graphe des individus selon le \cos^2	17
Figure 7 : Biplot des individus-variables	17
Figure 8: Tableau de l'AFC.....	19
Figure 9: Décomposition de l'inertie de l'AFC	19
Figure 10: Représentation graphique des contributions à la construction du 1 axe.....	21
Figure 11 Représentation graphique des contributions à la construction du 2 -ème axe.....	21
Figure 12: Graphe de l'AC	21

Tables des abréviations

AFCM : Analyse factorielle des Correspondances Multiples

AFC : Analyse factorielle des Correspondances

ACP : Analyse des Composantes principales

1. Introduction et objectifs :

L'analyse factorielle des correspondances multiples, est une méthode de description statistique multidimensionnelle d'un tableau de données qualitatives. Elle permet des représentations graphiques du contenu du tableau de données : représentation des similitudes entre les individus et entre les modalités.

Cette méthode est particulièrement adaptée aux questionnaires et aux enquêtes d'opinion.

À travers ce travail, nous effectuerons une AFCM sur les données d'une enquête d'opinions menées aux États-Unis d'Amérique afin de mesurer la confiance dans les informations en ligne et l'accès aux données. L'objectif étant de déterminer les associations pertinentes entre les questions et les réponses et essayer de comprendre l'état des lieux.

2. Présentation questionnaire

Le questionnaire de l'enquête d'opinion se présente comme suit : il contient 6 questions relatives à l'enquête, et 4 questions sur la population. Chaque question a un nombre de modalités variant entre 2 et 11.

Question	Réponses possibles	Abréviation
En général, comment décririez-vous vos points de vue sur la plupart des questions politiques ? êtes-vous :	Très conservateur	view
	Conservateur	
	Modéré	
	Liberal	
	Très libéral	
En général, quelle confiance avez-vous dans la presse lorsqu'il s'agit de rapporter les nouvelles de manière complète, précise et équitable ?	Beaucoup	Trust
	Raisonnement	
	Pas beaucoup	
	Pas du tout	
Auquel de ces médias faites-vous confiance lorsqu'il s'agit de rapporter l'actualité de manière complète, précise et équitable ?	The New York Times The	outlets
	Wall Street Journal	
	USA TODAY	

	The Washington Post	
	Fox News	
	Breitbart	
	CNN	
	BuzzFeed News	
	Huffington Post	
	Time	
	U.S. News & World Report	
Payez-vous actuellement pour accéder à une source d'actualités en ligne (y compris les versions en ligne de magazines imprimés, de Journaux et d'autres publications) ?	Yes	Payforaccess
	no	
Comment votre impression d'un article de presse en ligne changerait-elle si vous pouviez facilement accéder aux données derrière les affirmations de l'article ?	Diminue la confiance	Onlineaccess
	Augmente la confiance	
	Pas de changement	
Approuvez-vous ou désapprouvez-vous la façon dont Donald Trump gère son travail de président ?	Tout à fait d'accord	DonaldT
	Plutôt d'accord	
	Plutôt en désaccord	
	Désapprouve fortement	
Quel est votre âge ?	18-29	age
	30-44	
	45-59	
	60+	
Quel est votre sexe ?	Femme	gender
	Homme	
Combien d'argent total combiné tous les membres de votre foyer ont-ils gagné l'année dernière ?	\$0 to \$9,999	Money
	\$10,000 to \$24,999	
	\$25,000 to \$49,999	

	\$50,000 to \$74,999	
	\$75,000 to \$99,999	
	\$100,000 to \$124,999	
	\$125,000 to \$149,999	
	\$150,000 to \$174,999	
	\$175,000 to \$199,999	
	\$200,000 and up	
	Préfère ne pas répondre	
Région	Nouvelle-Angleterre	region
	Moyen-Atlantique	
	Est Nord Centre	
	Ouest Nord Centre	
	Sud Atlantique	
	Est Centre-Sud	
	Ouest Centre-Sud	
	Montagne	
	Pacifique	

3. Aperçu du dataset :

Le jeu de donnée initial contient plus de 1000 réponses, après prétraitement, on se limitera à 825 réponses, se présentant comme suit, on s'intéressera au 6 premières variables, les variables « age », « gender », « money », « region » étant considérées comme des variables signalétiques.

Names	view	trust	outlets	payforaccess	onlineaccess	DonaldT	age	gender	money	region
1	moderate	fair amount	The new york times	no	no change	strongly disapprove	30-44	male	\$10.000 to \$24.999	East South Central
2	moderate	fair amount	The new york times	no	Increase trust	somewhat disapprove	18-29	female	\$0 to \$9.999	Middle Atlantic
3	liberal	fair amount	The new york times	yes	Increase trust	strongly disapprove	30-44	male	125.000 to \$149.999	East North Central
4	liberal	fair amount	The washington post	no	Increase trust	strongly disapprove	18-29	male	125.000 to \$149.999	Mountain

Figure 1 : Aperçu du jeu de données

[Jeu de donnée complet](#)

4. Réalisation de l'AFCM sous R :

N.B. : LE SCRIPT COMPLET DE LA RÉALISATION EST ANNEXÉ

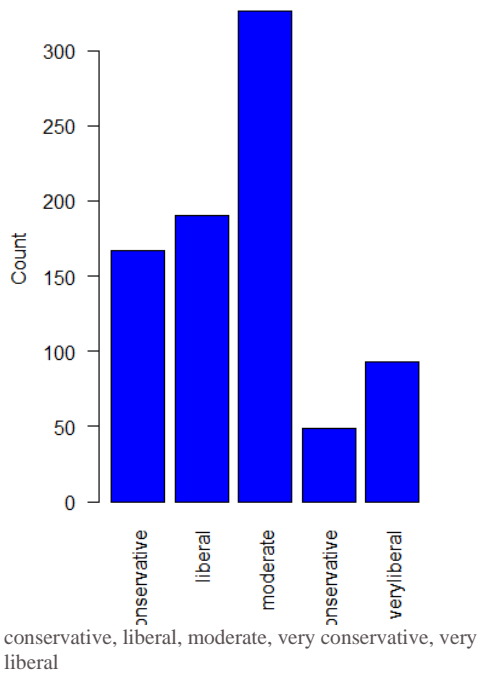
- On commence par importer le dataset sous R :

```
> data <- read.table(file.choose(),header=T)
> data
```

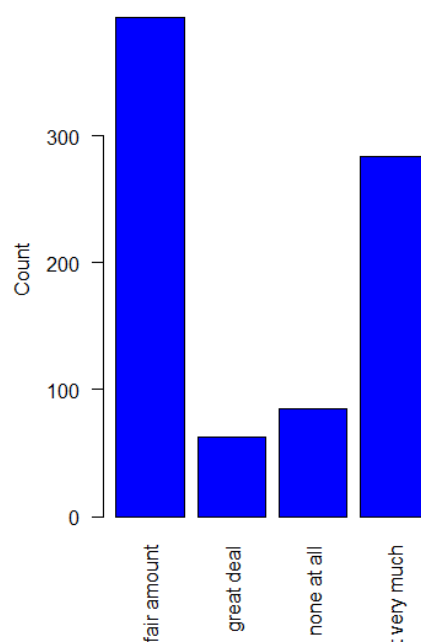
	view	trust	outlets	payforaccess	onlineaccess
1	moderate	fairamount	TheNewYorkTimes	No	nochange
2	moderate	fairamount	CNN	No	Increasetrust
3	liberal	fairamount	ThewashingtonPost	YES	Increasetrust
4	liberal	greatdeal	ThewashingtonPost	No	Increasetrust
5	veryconservative	noneatall	FoxNews	No	Increasetrust
6	veryliberal	fairamount	ThewashingtonPost	No	Increasetrust
7	conservative	fairamount	FoxNews	No	nochange
...					

4.1. Résumé du jeu de données

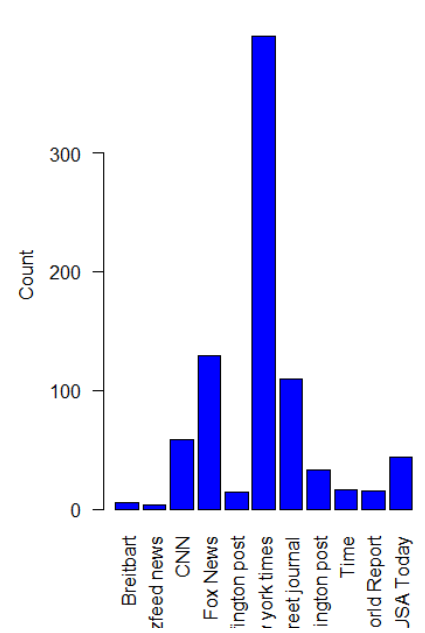
4.1.1. Variables actives



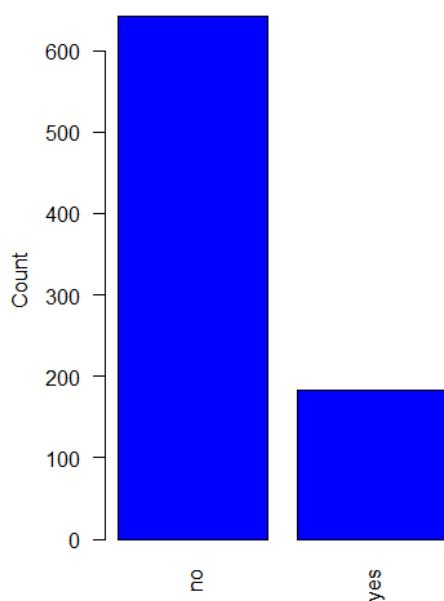
View



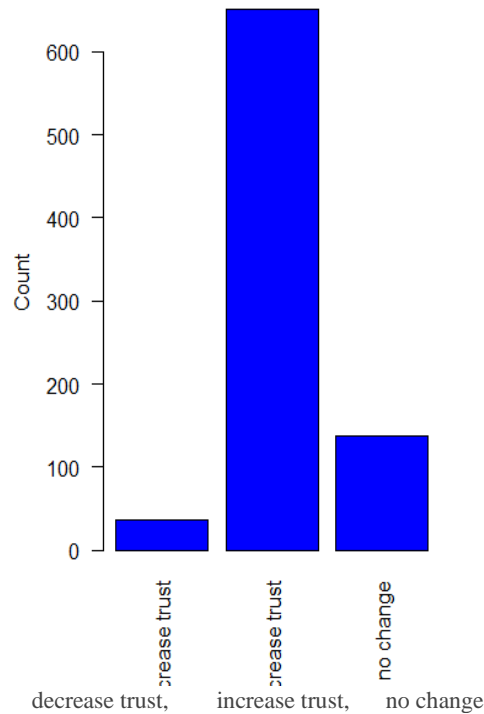
Trust



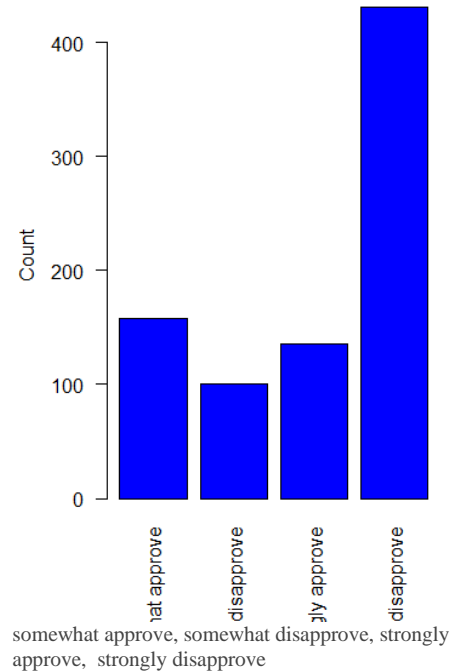
Outlets



Pay for access



Online access



DonaldT

4.1.2. Variables signalétiques :

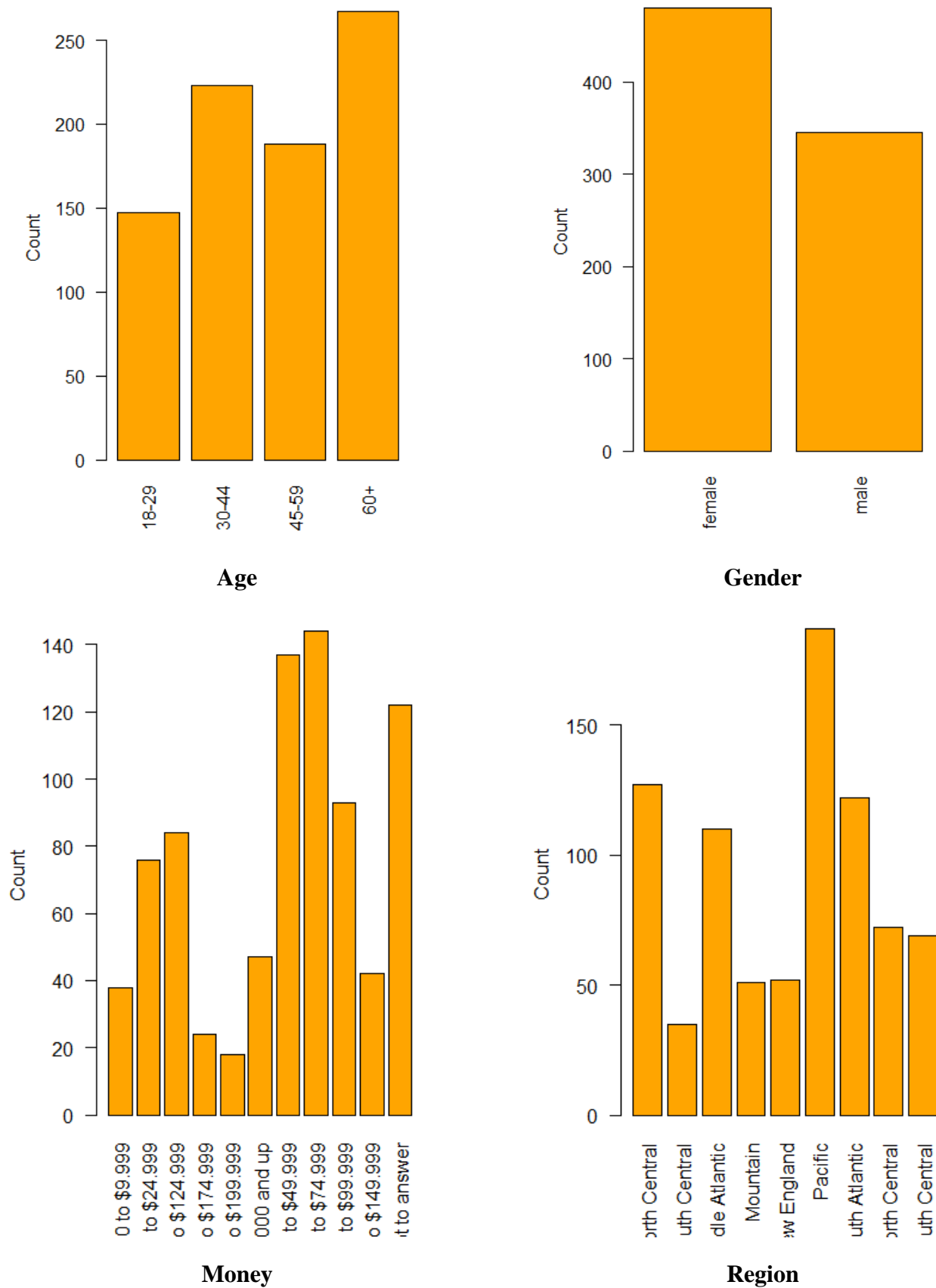


Figure 2 : Observation des variables signalétique

Tableau des fréquences et poids des modalités

	Fréquence	Poids	Poids (%)
Conservative	167	0,2024	20,2424
liberal	190	0,2303	23,0303
moderate	326	0,3952	39,5152
very conservative	49	0,0594	5,9394
Very liberal	93	0,1127	11,2727
fair amount	393	0,4764	47,6364
great deal	63	0,0764	7,6364
None at all	85	0,1030	10,3030
not very much	284	0,3442	34,4242
Breitbart	5	0,0061	0,6061
Buzzfeed news	3	0,0036	0,3636
CNN	58	0,0703	7,0303
Fox News	129	0,1564	15,6364
Huffington post	14	0,0170	1,6970
The New York times	398	0,4824	48,2424
The wall street journal	110	0,1333	13,3333
The Washington post	14	0,0170	1,6970
Time	16	0,0194	1,9394
U.S. News & World Report	15	0,0182	1,8182
USA Today	44	0,0533	5,3333
No	642	0,7782	77,8182
Yes	183	0,2218	22,1818

Decrease trust	36	0,0436	4,3636
Increase trust	651	0,7891	78,9091
No change	138	0,1673	16,7273
somewhat approve	158	0,1915	19,1515
somewhat disapprove	100	0,1212	12,1212
strongly approve	136	0,1648	16,4848
strongly disapprove	431	0,5224	52,2424

4.2. Tableau complet disjonctif

On transforme le tableau de données en tableau disjonctif à l'aide de la fonction `acm.disjonctif` de la librairie `ade4`. Voici un aperçu du résultat obtenu :

```
#Creation du tableau disjonctif complet
> disj <- acm.disjonctif(data)
> disj[1:3,1:20]
view.conservative view.liberal view.moderate view.very conservative
1                0                0                1                0
2                0                0                1                0
3                0                1                0                0
view.veryliberal trust.fair amount trust.great deal trust.none at all
1                0                1                0                0
2                0                1                0                0
3                0                1                0                0
trust.not very much outlets.Breitbart outlets.Buzzfeed news outlets.CNN
1                0                0                0                0
2                0                0                0                0
3                0                0                0                0
outlets.Fox News outlets.Huffington post outlets.The new york times
1                0                0                1
2                0                0                1
3                0                0                1
outlets.The wall street journal outlets.The Washington post outlets.Time
1                0                0                0
2                0                0                0
3                0                0                0
outlets.U.S.News & World Report outlets.USA Today
1                0                0
```

4.3. Lancement de l'AFCM :

On réalise l'ACM à l'aide de la fonction *MCA* du package *FactoMineR*. On considère les variables signalétiques (*Age*, *gender*, *region*, *money*) comme variables qualitatives supplémentaires.

```
afcm <- MCA(data, quali.sup=(7:10))
```

4.4. Étude du tableau de valeurs propres :

Les valeurs propres ainsi que les inerties obtenues sont décrites dans le tableau ci-dessous :

	Valeurs propres	Inertie	Inerties cumulées
dim 1	0.45	11.62	11.62
dim 2	0.25	6.65	18.27
dim 3	0.22	5.62	23.88
dim 4	0.21	5.35	29.23
dim 5	0.19	5.08	34.31
dim 6	0.19	4.99	39.30
dim 7	0.18	4.65	43.94
dim 8	0.17	4.50	48.44
dim 9	0.17	4.48	52.92
dim 10	0.17	4.39	57.30
dim 11	0.17	4.35	61.65
dim 12	0.16	4.26	65.91
dim 13	0.16	4.17	70.08
dim 14	0.15	3.98	74.07
dim 15	0.15	3.81	77.88

dim 16	0.14	3.69	81.57
dim 17	0.13	3.41	84.98
dim 18	0.13	3.36	88.35
dim 19	0.12	3.12	91.47
dim 20	0.11	2.91	94.38
dim 21	0.10	2.51	96.89
dim 22	0.07	1.81	98.70
dim 23	0.05	1.30	100.00

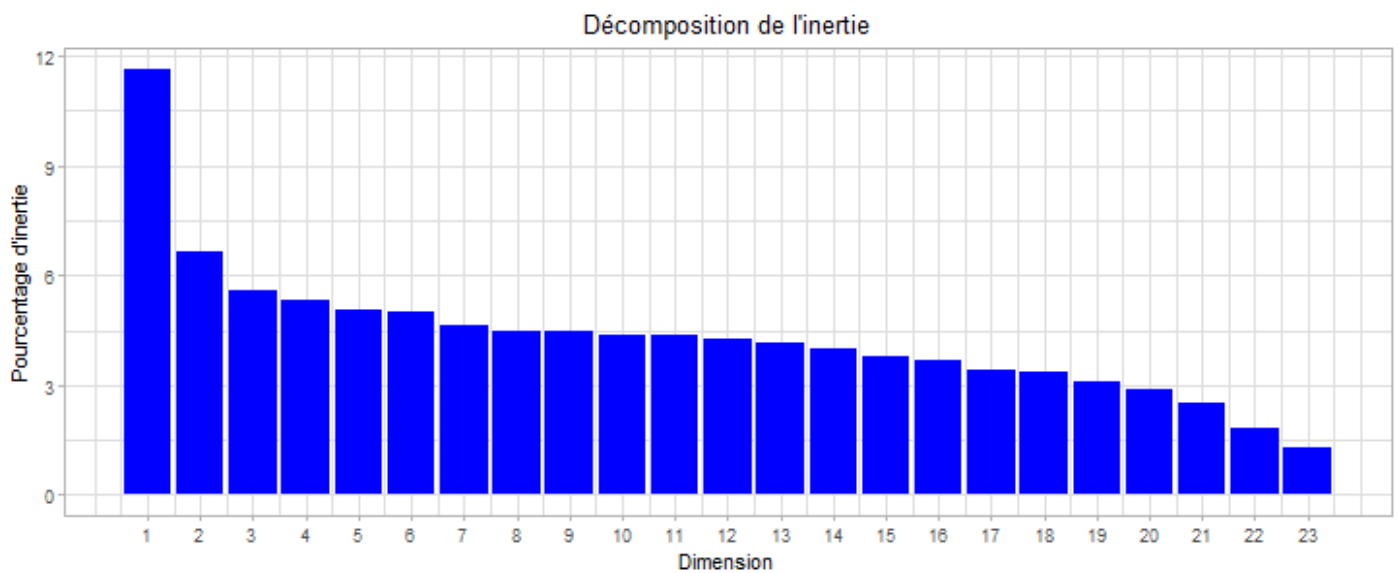


Figure 3 : Visualisation des pourcentages de variances expliquées par chaque dimension de l'ACM

4.4.1. Explication :

On remarque que les valeurs propres sont toutes inférieures à 1 ($\lambda_i < 1; i = 1, \dots, 23$) et que les taux d'inertie sont très faibles. En effet le premier plan factoriel n'apporte que 18.27%. Ceci est dû au nombre élevé de modalités de l'ACM. Ces taux d'inertie ne renseignent pas sur le taux réel d'inertie expliquée par un axe factoriel.

On se limitera par la suite à l'étude du premier plan factoriel seulement.

4.5. Étude du tableau des contributions

Tableau des contributions des variables

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
Conservative	9.00	0.00	0.47	6.38	8.05
liberal	5.58	0.45	0.20	3.59	1.31
moderate	0.04	10.37	0.17	0.01	3.26
very conservative	5.42	9.10	0.22	9.38	1.62
Very liberal	3.90	7.84	0.21	1.94	1.28
fair amount	4.75	1.51	0.68	0.52	6.55
great deal	2.68	6.45	2.34	4.07	3.42
None at all	7.91	2.96	10.61	0.14	0.71
not very much	3.23	0.48	12.04	0.09	6.77
Breitbart	0.25	0.09	2.72	0.70	0.58
Buzzfeed news	0.03	0.18	13.83	0.36	3.95
CNN	0.00	8.20	2.56	3.90	2.51
Fox News	11.97	4.37	2.62	0.00	0.89
Huffington post	0.01	0.32	5.90	0.98	8.81
The New York times	10.02	1.83	0.75	0.24	0.54
The wall street journal	2.78	0.77	7.72	5.00	1.16
The Washington post	0.01	0.51	1.57	1.28	1.13
Time	0.00	3.06	0.00	0.90	7.10
U.S. News & World Report	0.21	0.01	3.80	8.59	1.50
USA Today	0.35	5.04	0.13	2.15	4.69
No	0.24	0.59	0.09	2.32	1.31

Yes	0.83	2.09	0.30	8.15	4.61
Decrease trust	0.12	2.93	0.25	10.31	7.66
Increase trust	0.49	0.96	5.61	0.53	0.47
No change	1.78	1.57	23.87	0.00	8.45
somewhat approve	3.82	9.27	0.08	19.05	0.25
somewhat disapprove	0.02	6.82	0.93	5.56	9.68
strongly approve	13.57	10.62	0.34	3.66	0.04
strongly disapprove	11.00	1.62	0.00	0.19	1.69

4.5.1. Signification des axes :

Intéressons-nous en premier lieu à l'axe 1 : On observe que les modalités qui ont le plus contribué à sa construction sont : “*Strongly approve*”, “*Strongly disapprove*”, “*Fox News*”, “*None at all*”, “*very conservative*”. Ce sont les variables dont la contribution est très proche ou supérieure à leur poids.

Leurs signes de saturation sont comme suit :

+	-
<i>Strongly approve, Fox news, None att all, very conservative</i>	<i>Strongly disapprove</i>

L'axe 1 est donc un *axe d'opposition*, il oppose il individus qui sont en total désaccord avec la façon dont Donald Trump gère son travail de président à ceux qui sont fortement d'accord avec ce dernier.

Les variables qui ont contribué et qui sont situés du côté positif de l'axe sont très proches entre elles. Ceci signifie que les individus qui sont fortement d'accord avec la façon dont Donald Trump gère son travail sont des gens très conservateurs qui n'ont pas confiance en la presse lorsqu'il s'agit de rapporter les nouvelles de manière complète, et ont tendance à faire confiance au média Fox News

En analysant le deuxième Axe on observe que les variables qui ont le plus contribué à sa construction sont : “*Moderate*”, “*Very Conservative*”, “*Very Liberal*”, “*Somewhat Approve*”, “*Somewhat Disapprove*”, “*CNN*” Ce sont les variables dont la contribution est supérieure ou proche à leur poids.

Leurs signes de saturation sont comme suit :

+	-
Moderate, Somewhat Approve, Somewhat Disapprove, CNN	Strongly disapprove, strongly approve,

On constate que l'axe 2 donc **oppose** les individus qui ont des orientations politiques extrêmes ou penchantes, à ceux qui s'identifient plutôt modérés. Ces derniers font plutôt confiance au media CNN qu'aux autres

4.6. Étude des associations entre modalités :

En se référant à la figure n°5 (paragraphe 5) on compare les proximités entre les modalités des variables. On remarque que :

- Les modalités "*Fox news*", "*none at all*" sont superposées, ça indique que ces modalités sont associées aux mêmes individus. Ce qui veut dire que les individus qui font confiance à Fox News (Les conservateurs / droite ou d'extrême droite) ne font pas du tout confiance aux médias. On remarque aussi une proximité de ces deux modalités à "*Strongly Approve*" (Très en accord avec Donald Trump) qui est aussi une caractéristique des conservateurs républicains.
- De même pour les modalités "*liberal*", "*strongly disapprove*" et "*The New York Times*". Ce qui signifie que les individus qui sont d'orientation libérale, sont pour la majorité en total opposition à la manière dont Donald Trump gère son poste de président, et ces individus font confiance aux médias "The New York Times".
- On remarque une proximité aussi entre les modalités "*very liberal*" et "*great deal*". Ce qui signifie que les individus d'orientation très libérale ont une confiance raisonnable dans la presse lorsqu'il s'agit de rapporter les nouvelles de manière complète, précise et équitable.
- Les modalités "*CNN*" et "*Time*" sont très proches, notamment parce que ce sont des médias jugés plutôt centriste ou non-biaisés, contrairement à d'autres médias qui ont souvent un penchant politique

4.7. Visualisations

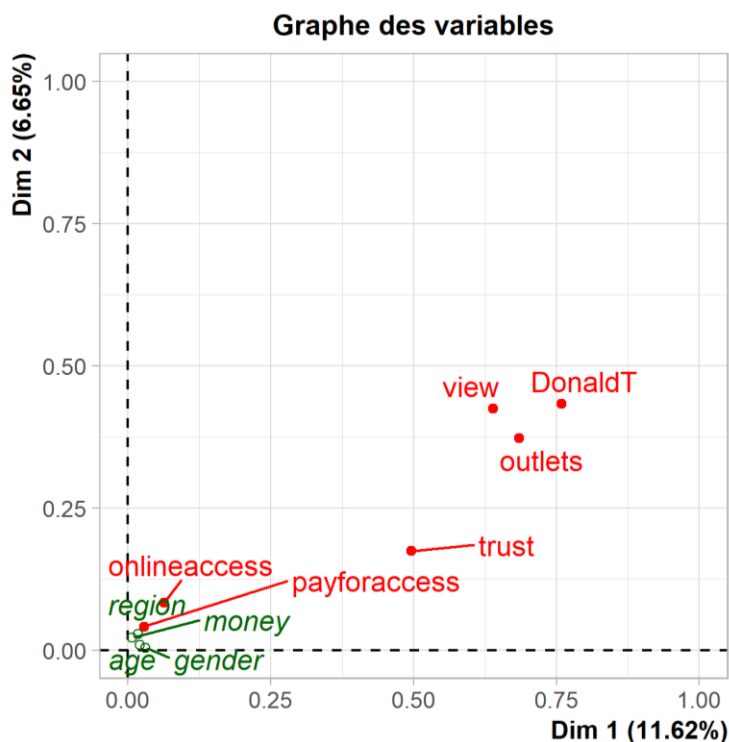


Figure 4 : Graphe des variables

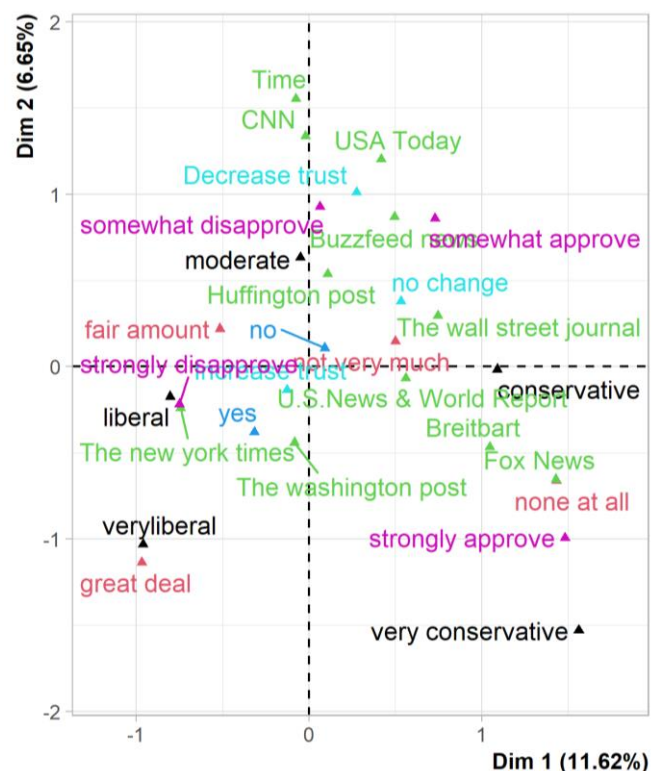


Figure 5 : Graphe des modalités

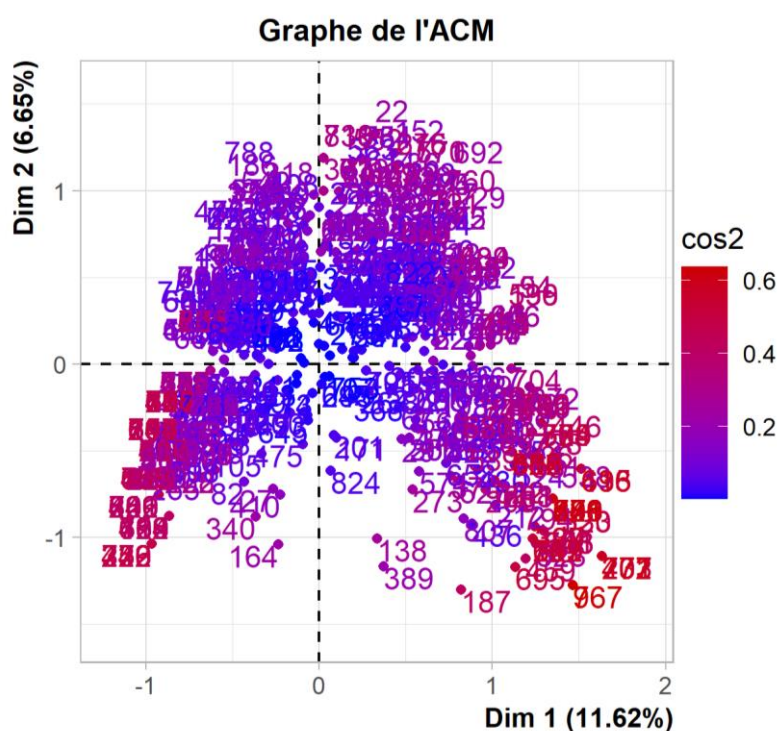
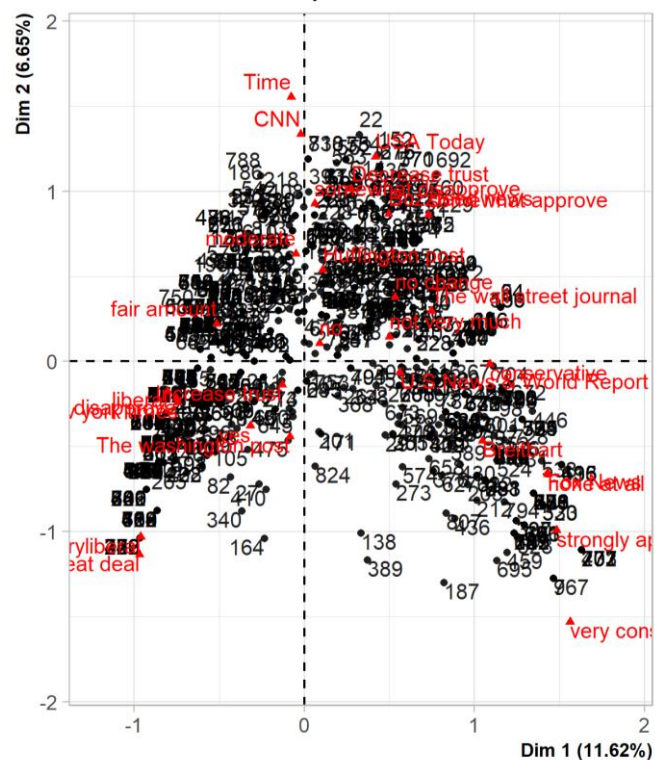
Figure 6 : Graphe des individus selon le \cos^2 

Figure 7 : Biplot des individus-variables

- La *figure 4* nous indique les corrélations des variables à chaque axe. En effet on observe que le rapport de corrélation des variables « *outlet* », « *donaldT* » et « *view* » est significativement différent de zéro. Ces variables sont donc, les variables les mieux représentées par le premier plan factoriel.

4.8. Étude des individus :

Intéressons-nous maintenant à certains individus remarquables :

- La proximité de deux individus sur le graphe des individus indique que ces derniers ont eu des réponses similaires. À l'exemple des individus 459 et 397 :

Names	view	trust	outlets	payforaccess	onlineaccess	DonaldT	age	gender	money	region
459	very conservative	none at all	The wall street journal	yes	Increase trust	strongly approve	60+	male	\$100.000 to \$124.999	South Atlantic

Names	view	trust	outlets	payforaccess	onlineaccess	DonaldT	age	gender	money	region
397	conservative	none at all	Fox News	yes	Increase trust	strongly approve	60+	male	Prefer not to answer	East North Central

- L'individu 788 se positionnant du côté positif de l'axe 2 quant à lui a des penchants politiques plutôt modérés

Names	view	trust	outlets	payforaccess	onlineaccess	DonaldT	age	gender	money	region
788	moderate	fair amount	Time	no	Decrease trust	strongly disapprove	60+	female	\$0 to \$9.999	West North Central

- À l'opposé de l'individu 749 qui a des penchants extrêmes libéral et une confiance raisonnable en les médias

Names	view	trust	outlets	payforaccess	onlineaccess	DonaldT	age	gender	money	region
749	veryliberal	great deal	The new york times	yes	Increase trust	strongly disapprove	45-59	male	\$50.000 to \$74.999	Pacific

- Ce dernier s'oppose à l'individu 794 qui se situe du côté positif de l'axe 1, il se caractérise par des penchants politiques extrêmement conservateurs, et une très faible confiance en les médias

Names	view	trust	outlets	payforaccess	onlineaccess	DonaldT	age	gender	money	region
794	very conservative	none at all	Fox News	yes	Increase trust	somewhat approve	45-59	male	Prefer not to answer	Pacific

5. Réalisation d'une AFC :

Dans le paragraphe qui suit, on réalisera une AFC en croisant les deux variables « *DonaldT* », « *View* ».

On commence par croiser les deux variables et obtenir le tableau de contingence

```
> tableau<-table(data$view,data$DonaldT)
> tableau
```

Names	somewhat.approve	somewhat.disapprove	strongly.approve	Strongly.disapprove
conservative	66	20	66	15
liberal	5	21	4	160
moderate	78	54	30	164
very conservative	6	3	35	7
very liberal	3	2	1	87

Figure 8: Tableau de l'AFC

5.1. Interprétation des résultats de l'AFC :

5.1.1. Étude des valeurs propres :

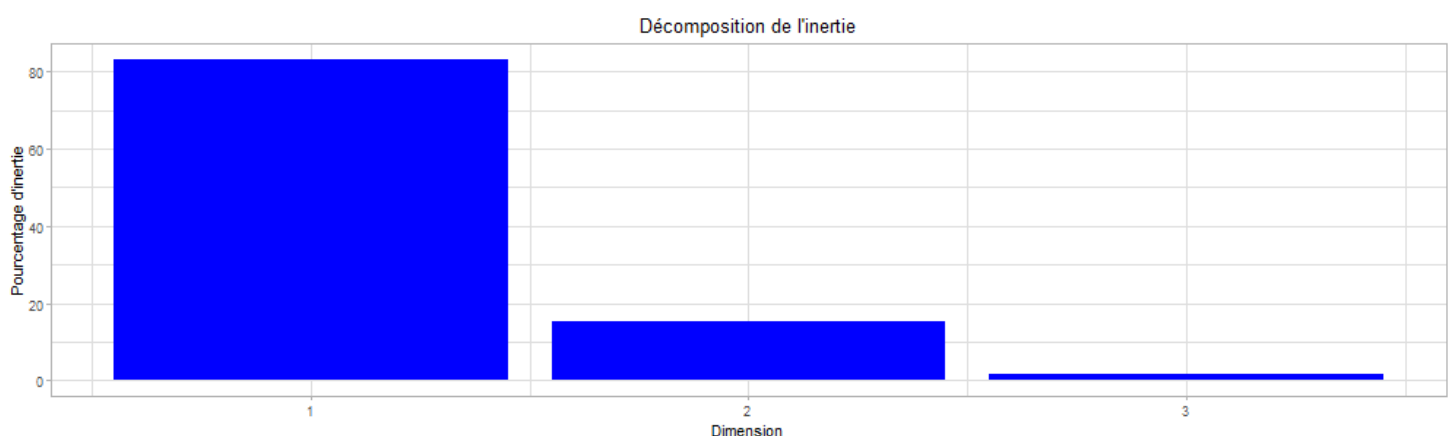


Figure 9: Décomposition de l'inertie de l'AFC

On remarque tout d'abord les taux d'inertie très élevées comparées aux résultats obtenus par l'ACP. En effet l'axe 1 à lui seul a un taux d'inertie de 82.98 %.

5.1.2. Interprétation des axes :

On dira qu'un profil ligne ou un profil colonne a contribué à la construction d'un axe si sa contribution absolue est supérieure à son poids.

Avec quelle conclusion et quelle décision ressortez-vous après l'analyse.

	Dim 1	Dim 2	Dim 3		Dim 1	Dim 2	Dim 3
somewhat.approve	14.65	42.84	23.40	conservative	41.99	2.68	17.21
somewhat.disapprove	0.09	14.62	73.20	liberal	23.59	6.42	8.59
strongly.approve	46.05	36.83	0.68	moderate	0.30	27.88	10.60
Strongly.disapprove	39.21	5.71	2.72	very conservative	18.78	53.91	11.39
				very liberal	15.33	9.11	52.20

- On obtient le poids des colonnes et des lignes comme suit :

```
>afc$call$marge.col
> afc$call$marge.row
```

conservative	liberal	moderate	very conservative	very liberal
0.20242424	0.23030303	0.39515152	0.05939394	0.11272727

somewhat approve	somewhat disapprove	strongly approve	strongly disapprove
0.1915152	0.1212121	0.1648485	0.5224242

<i>PL</i>		<i>PC</i>	
+	-	+	-
Conservative, very conservative	Very liberal	Strongly approve	

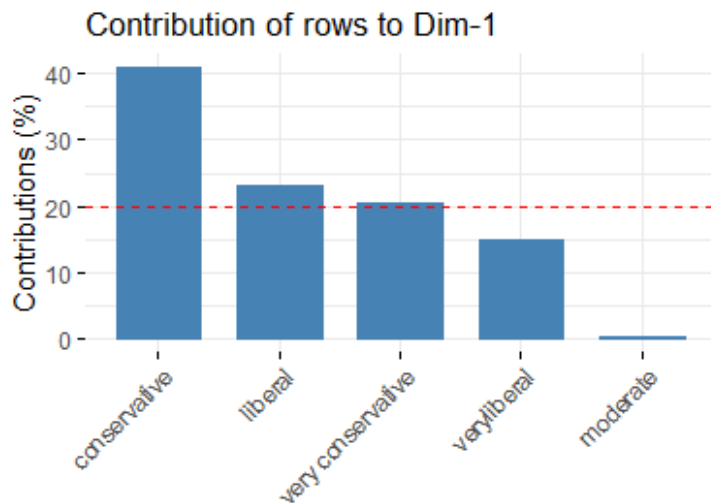


Figure 10: Représentation graphique des contributions à la construction du 1^{er} axe

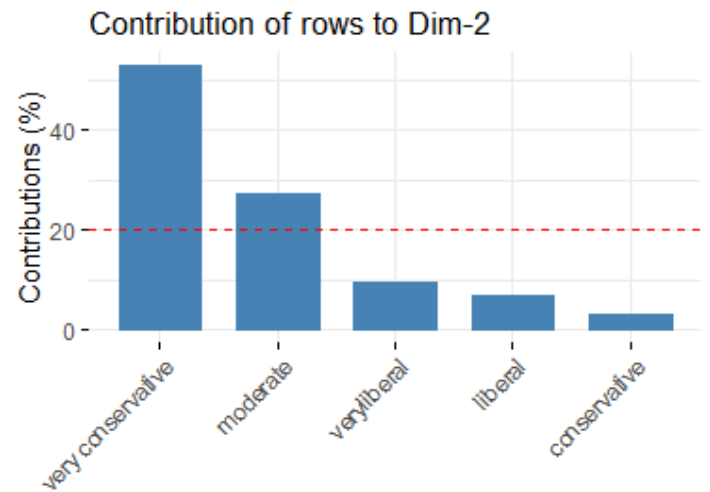


Figure 11 Représentation graphique des contributions à la construction du 2^{ème} axe

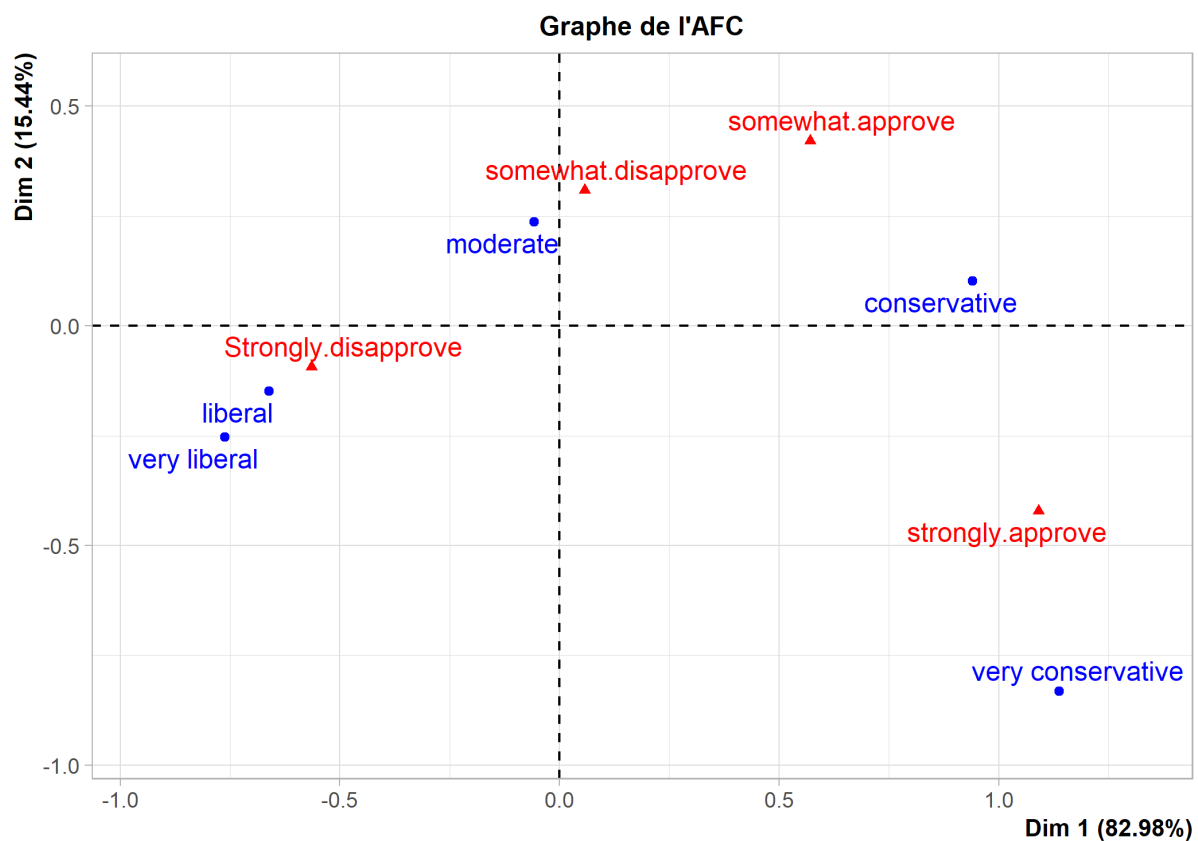


Figure 12: Graphe de l'AC

L'axe 1 est donc axe de mesure, il mesure le taux d'accord ou de désaccord d'un individu avec la manière dont le président Donald Trump répond à ses fonctions de présidents. On conclut que les individus d'orientation conservatrice et très conservatrice sont fortement d'accord avec la manière dont Donald Trump gère sa présidence, contrairement aux profils d'orientation libéral qui sont ses opposants et sont en total désaccord avec se dernier.

Notons que ces résultats sont similaires aux résultats de l'AFCM.

6. Conclusion :

À travers cette étude on a pu réaliser l'intérêt de l'ACM dans le traitement d'un nombre important de réponses dans le cadre d'enquêtes d'opinion ou de réponses à des enquêtes. Néanmoins cette méthode reste limitée, étant sensible aux modalités à faible fréquence. Les résultats numériques obtenues à travers l'ACM sont très faibles comparés à l'ACP, ou bien l'AFC, ceci est dû au nombre élevé de modalités traitées par l'ACP.

Néanmoins on a pu tirer des conclusions parlantes dans le cadre de notre étude, reflétant la réalité des choses.

En ce qui concerne l'AFC et l'AFCM on a conclu qu'on obtient des résultats similaires dans leur signification.

Tout ceci confirme l'importante utilité de l'AFCM et son large champ d'utilisation.

Annexe

```

#bibliotheques utilisées
library(FactoMineR)
library(factoextra)
library(ade4)
#chargement des données
data <- read.table(file.choose(),header=T, sep=",")
data
#Calcul des fréquences
i=0
while(i < ncol(data)){
  i=i+1
  data[,i]=as.factor(data[,i])
}
summary(data$view)
#histogramme des donnees
barplot(table(data$DonaldT), ylab = "Count", col="blue", las = 2)
#Creation du tableau disjonctif complet
disj <- acm.disjonctif(data)
#Lancement de L'AFCM
afcm <- MCA(data, quali.sup=(7:10))
#Valeurs propres
get_eigenvalue(afcm)
#Pourcentage des variances expliquées
fviz_eig(afcm, addlabels = TRUE, col="blue")
get_mca_ind(afcm)
#Représentation du biplot variable-individus
fviz_mca_biplot (afcm, repel = TRUE, ggtheme = theme_minimal())
#Représentation des variables
fviz_mca_var (afcm, choice = "mca.cor", repel = TRUE)
#Categorie des variables selon la qualite de representation cos^2
fviz_mca_var(afcm, col.var = "cos2", gradient.cols = c("#00AFBB",
"#E7B800", "#FC4E07"))
#MCA map factor
fviz_ellipses(afcm, c("view", "DonaldT"), geom = "point")
#Creation du tableau de contingence
tableau <- table(data$view, data$DonaldT)
addmargins(tableau)
#Realisation de L'AFC
afc <- CA(tableau, graph=TRUE)

```