

I. Le responsable marketing d'une banque a besoin de lancer une action marketing pour la vente ciblée d'un nouveau type de carte bancaire nommé GOLD CARD. Pour cela, le responsable marketing a construit, avec son équipe, un fichier nommé *Cartes_GOLD.txt* composé d'un échantillon de 34 clients. Les variables du fichier *Cartes_GOLD.txt* sont : *Montants_compte* : montants au compte bancaire des clients en dinars, *Salaires* : salaires des clients en dinars, *Nbre_mois* : nombre de mois à partir de la date d'ouverture des comptes, *Consom_hebdo* : consommation hebdomadaire des clients en dinars, *Gold* : achat **positif** ou **négatif**.

- Analyser le comportement des clients.
- Etudier la dépendance entre les caractéristiques des clients.

II. Le responsable marketing souhaite avoir un code d'une méthode lui permettant un ciblage optimal des clients dans le cadre d'une action marketing de la carte bancaire GOLD CARD.

- Proposer le code d'une méthode en quantifiant ses performances dans le ciblage des clients avec les données du fichier *Cartes_GOLD.txt*.

III. Afin d'étudier la satisfaction des clients pour la carte Gold, le responsable marketing a effectué une enquête sur un échantillon de clients qui ont déjà la carte. Le questionnaire comprend les questions suivantes :

- Etes-vous globalement satisfait de la carte Gold ? (Oui / Non)
- Comptez-vous renouveler votre carte Gold l'année prochaine ? (Oui / Non / Ne sait pas)
- Consultez-vous vos comptes bancaires via internet ? (Oui / Non)
- Le rapport qualité prix de l'abonnement annuel d'une carte vous semble-t-il correct ? (Oui / Non)
- Avez-vous profité des avantages de la carte Gold tel que la gratuité de l'assurance voyage ? (Oui / Non)

Les réponses au questionnaire pour un échantillon de 28 clients sont stockées dans le fichier *Satisfaction_cls_gold.txt*.

- Analyser les réponses des clients et étudier particulièrement leur satisfaction.

Réponses :

- I. Dans cette partie on s'intéresse à l'étude des profils des clients en effectuant une analyse en composante principale (ACP) et déterminer une structure de corrélation entre eux.

Importation des données :

```
> setwd("C:/Users/Mahdi/Documents/2A MIndS/AD/Atelier_AD_2MINDS")
> don1<-read.table(file="Cartes_GOLD.txt",header=TRUE)
> don1
```

	Montants_compte	Salaires	Nbre_mois	Consom_hebdo	Gold
1	3064	1201	10	361	negatif
2	3000	1053	11	338	negatif
3	3155	1133	19	393	negatif
4	3085	970	4	467	negatif
5	3245	1258	36	294	positif
6	3267	1386	35	225	positif
7	3080	966	13	417	negatif
8	2974	1189	12	488	negatif
9	3038	1103	14	677	negatif
10	3318	1310	29	427	negatif
11	3317	1362	25	326	positif
12	3182	1171	28	326	negatif
13	2998	1102	9	349	negatif
14	3221	1424	21	382	positif
15	3019	1230	16	275	negatif
16	3022	1285	9	303	negatif
17	3094	1329	11	339	negatif
18	3009	1210	15	536	negatif
19	3227	1331	21	414	negatif
20	3308	1366	24	282	positif
21	3212	1289	17	302	negatif
22	3361	1444	25	253	positif
23	3061	1175	12	261	negatif
24	3478	1317	42	259	positif
25	3126	1248	11	315	negatif
26	3458	1508	43	286	positif
27	3252	1361	26	346	negatif
28	3052	1186	14	443	negatif
29	3270	1399	24	306	positif
30	3198	1259	20	367	positif
31	2904	1164	6	311	negatif
32	3247	1277	19	375	positif
33	3083	1195	5	441	negatif

ACP :

```
> library(FactoMineR)
> res1.acp=PCA(don1,quali.sup=5)
> summary(don1)
```

Montants_compte	Salaires	Nbre_mois	Consom_hebdo	Gold
Min. :2904	Min. : 966	Min. : 4.00	Min. :225.0	negatif:23
1st Qu.:3045	1st Qu.:1178	1st Qu.:11.25	1st Qu.:302.2	positif:11
Median :3140	Median :1253	Median :16.50	Median :342.5	
Mean :3158	Mean :1247	Mean :18.82	Mean :360.4	
3rd Qu.:3251	3rd Qu.:1330	3rd Qu.:24.75	3rd Qu.:408.8	
Max. :3478	Max. :1508	Max. :43.00	Max. :677.0	

C'est une ACP en tenant compte que la variable d'indice 5 (ici Gold) est qualitative (quali.sup=5).

Maintenant, on doit choisir les axes comportant le maximum d'information ou d'inertie à l'aide du code suivant :

```
> summary(res1.acp)
```

Call:
PCA(X = don1, quali.sup = 5)

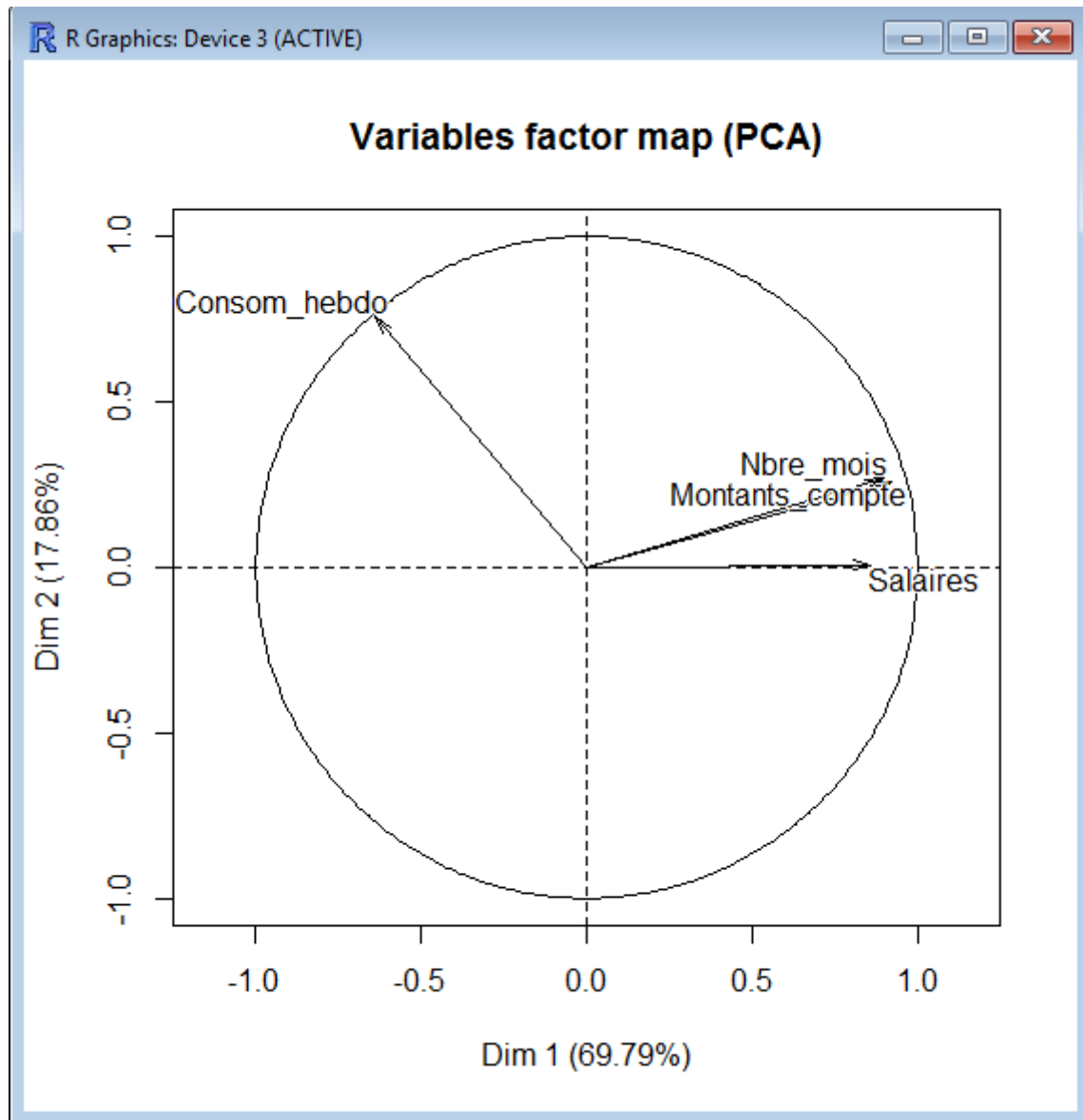
Eigenvalues

	Dim.1	Dim.2	Dim.3	Dim.4
Variance	2.791	0.714	0.366	0.128
% of var.	69.787	17.862	9.148	3.203
Cumulative % of var.	69.787	87.648	96.797	100.000

Ce tableau montre que les deux premières composantes ont 87.64% d'informations dont le premier axe fournit 69.78% de l'information totale.

Donc on se base sur ces 2 axes pour notre analyse.

Cercle de corrélation :



-La figure ci-dessus montre que les variables sont bien corrélées avec les deux axes parce qu'elles sont proches de cercle de corrélation.

-La variable « salaires » est parfaitement corrélée avec la première composante et on peut dire aussi que les variables « Montants_compte » et « Nbre_Mois » dépendent de la variable « salaires »

- La variable « consommation_hebdo » est liée avec la deuxième composante, ce qui explique l'indépendance de la consommation hebdomadaire d'un client de son salaire par exemple.

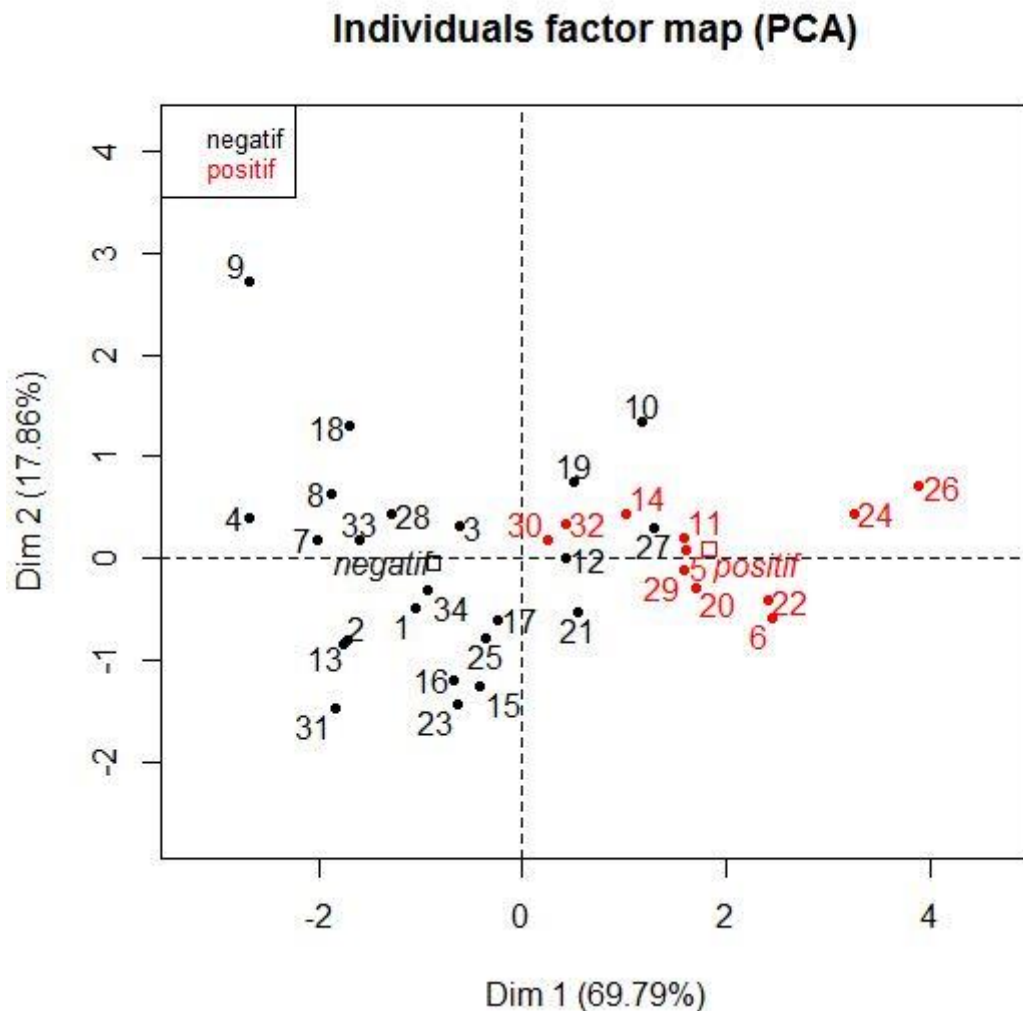
On peut conclure que toutes les variables sont bien corrélées avec les deux composantes et essentiellement le premier axe ainsi elles sont bien représentées, ce qui nous aide à analyser les données d'une manière efficace.

-L'axe 1 modélise les clients ayant un salaire de moins élevé au plus élevé.

Traçage du nuage des points :

On écrit ce code : `plot(res1.acp,choix="ind",habillage=5)`

On obtient la figure suivante où les clients sont distingués selon la variable qualitative Gold qui illustre leurs réponses (positive ou négative)



On remarque, dès le début, que le nuage est réparti en deux parties : Les clients qui achètent la carte Gold prennent toutes les valeurs positive de l'axe 1 tandis que les autres prennent les valeurs négatives avec un chevauchement dû à la récupération incomplète des données (on garde seulement 88% de l'inertie totale).

Conclusion : Pour faire une action Marketing, il faut cibler les clients avec un salaire important et qui sont anciens dans la banque qui est caractérisé par le nombre de mois . On ne peut pas juger les clients en se basant sur leurs consommation hebdomadaire car, comme on a déjà expliqué, cette consommation est indépendante des autres variables.

II . Dans cette partie, on va essayer de trouver un moyen de ciblage des clients. Ce pur cela, on pense d'une méthode de classification adéquate pour grouper les individus selon un critère concret. D'après ce qu'on a vu dans la séance de TP3, on constate que la méthode AHA (arbre hiérarchique ascendante) est performante pour donner une bonne Classification.

On analysera cette méthode par le calcul du taux de bonne classification.

Centrage et réduction des données :

Avant faire cette étape, il faut tout d'abord créer un nouveau tableau des données sans lire la variable qualitative « Gold ».Sinon, on trouve des erreurs de compilation.

```
> don1_quant=don1
> don1_quant$Gold=NULL
> don1_qual=don1$Gold
> scale(don1_quant,center=TRUE,scale=TRUE)
```

	Montants_compte	Salaires	Nbre_mois	Consom_hebdo
[1,]	-0.66496296	-0.36583827	-0.88089384	0.006113935
[2,]	-1.11827105	-1.53466252	-0.78105921	-0.245522774
[3,]	-0.02041553	-0.90286563	0.01761788	0.356217184
[4,]	-0.51622125	-2.19015180	-1.47990166	1.165830946
[5,]	0.61704897	0.08431701	1.71480668	-0.726914741
[6,]	0.77287362	1.09519204	1.61497205	-1.481824870
[7,]	-0.55163594	-2.22174164	-0.58138994	0.618794620
[8,]	-1.30242746	-0.46060781	-0.68122457	1.395586203
[9,]	-0.84911937	-1.13978947	-0.48155530	3.463383514
[10,]	1.13410350	0.49498499	1.01596423	0.728201885
[11,]	1.12702056	0.90565297	0.61662569	-0.376811493
[12,]	0.17082382	-0.60276211	0.91612960	-0.376811493
[13,]	-1.13243693	-1.14768693	-0.98072848	-0.125174783
[14,]	0.44705843	1.39529556	0.21728715	0.235869192
[15,]	-0.98369521	-0.13681190	-0.28188603	-0.934788545
[16,]	-0.96244640	0.29754846	-0.98072848	-0.628448202
[17,]	-0.45247480	0.64503675	-0.78105921	-0.234582048
[18,]	-1.05452460	-0.29476112	-0.38172067	1.920741075
[19,]	0.48955607	0.66083167	0.21728715	0.585972441
[20,]	1.06327411	0.93724281	0.51679106	-0.858203459
[21,]	0.38331198	0.32913831	-0.18205139	-0.639388929
[22,]	1.43866987	1.55324478	0.61662569	-1.175484528
[23,]	-0.68621178	-0.57117226	-0.68122457	-1.087958716
[24,]	2.26737372	0.55026722	2.31381450	-1.109840169
[25,]	-0.22582076	0.00534240	-0.78105921	-0.497159484
[26,]	2.12571494	2.05868230	2.41364913	-0.814440553
[27,]	0.66662954	0.89775551	0.71646033	-0.157996962
[28,]	-0.74995823	-0.48430019	-0.48155530	0.903253510
[29,]	0.79412244	1.19785903	0.51679106	-0.595626023

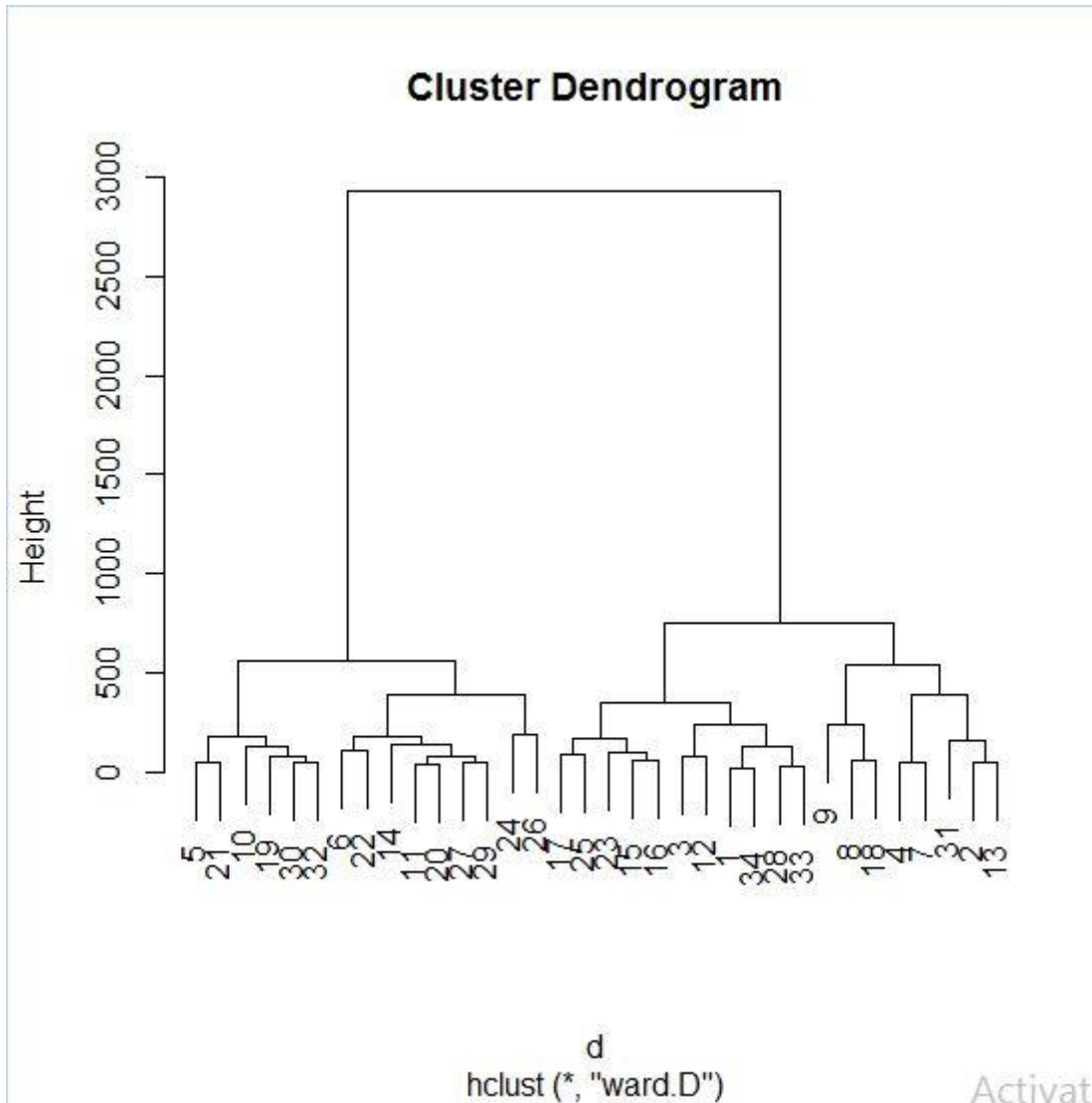
Partitionnement des groupes :

```
> d=dist(don1_quant)
> hc=hclust(d,method="ward.D")
> plot(hc)
```

La commande « dist » permet de calculer la distance entre un individu et les autres de telle façon on obtient un tableau symétrique.

Puis on utilise la commande « hclust » qui assure le fusionnement des individus puis les classes en considérant les distances calculées auparavant.

AHA :



-La méthode ward présente une répartition égale des individus sans avoir une perte pertinente d'inertie interclasse et l'effet de chaînage.

-On voit également deux clusters remarquables qui pourraient de partitionner les clients cibles.

Tableau de contingence :

Le tableau de contingence nous renseigne sur la qualité de classification par une variable qualitative (ici on interprète selon la classification par « positive » et « négative »)

```
> classe<-cutree(hc,k=2)
> table(don1_qual,classe)
      classe
don1_qual 1  2
negatif  19  4
positif   0 11
```

La méthode AHA fournit un regroupement par deux clusters traduisant la réalité d'une façon pertinente.

Pour illustrer cette hypothèse, on calcule le taux de bonne classification :

- Taux de bonne classification total : $(19+11)/(19+15)= 88.23\%$
- Taux de bonne classification positif : $11/11=100\%$
- Taux de bonne classification négatif : $19/23= 82.6 \%$

Conclusion : Grâce à la méthode ward, on peut cibler les gents qui ont tendance d'acheter la carte Gold avec beaucoup de précision.

Ainsi, l'algorithme suivi est :

- 1) Importation des données
- 2) Centrage et réduction
- 3) Application de l'AHA
- 4) Identifier les personnes ciblées.

Pour accomplir la dernière étape de cet algorithme, on tape le code suivant :

```
> clients_c=which(classe==2)
> clients_c
[1] 5 6 10 11 14 19 20 21 22 24 26 27 29 30 32
```

Les valeurs affichées sont les indices des clients qu'on doit cibler dans le tableau des données

On peut rencontrer 4 clients mal choisis d'après la classification mais ça n'affecte que rarement sur l'action de Marketing.

III- Cette partie est consacrée à l'étude de la satisfaction des clients pour la carte Gold. Donc, On procède l'AFM (Analyse Factorielle Multiple) vu qu'on des variables qualitatives multiples (ici il s'agit des questions à choix multiples).

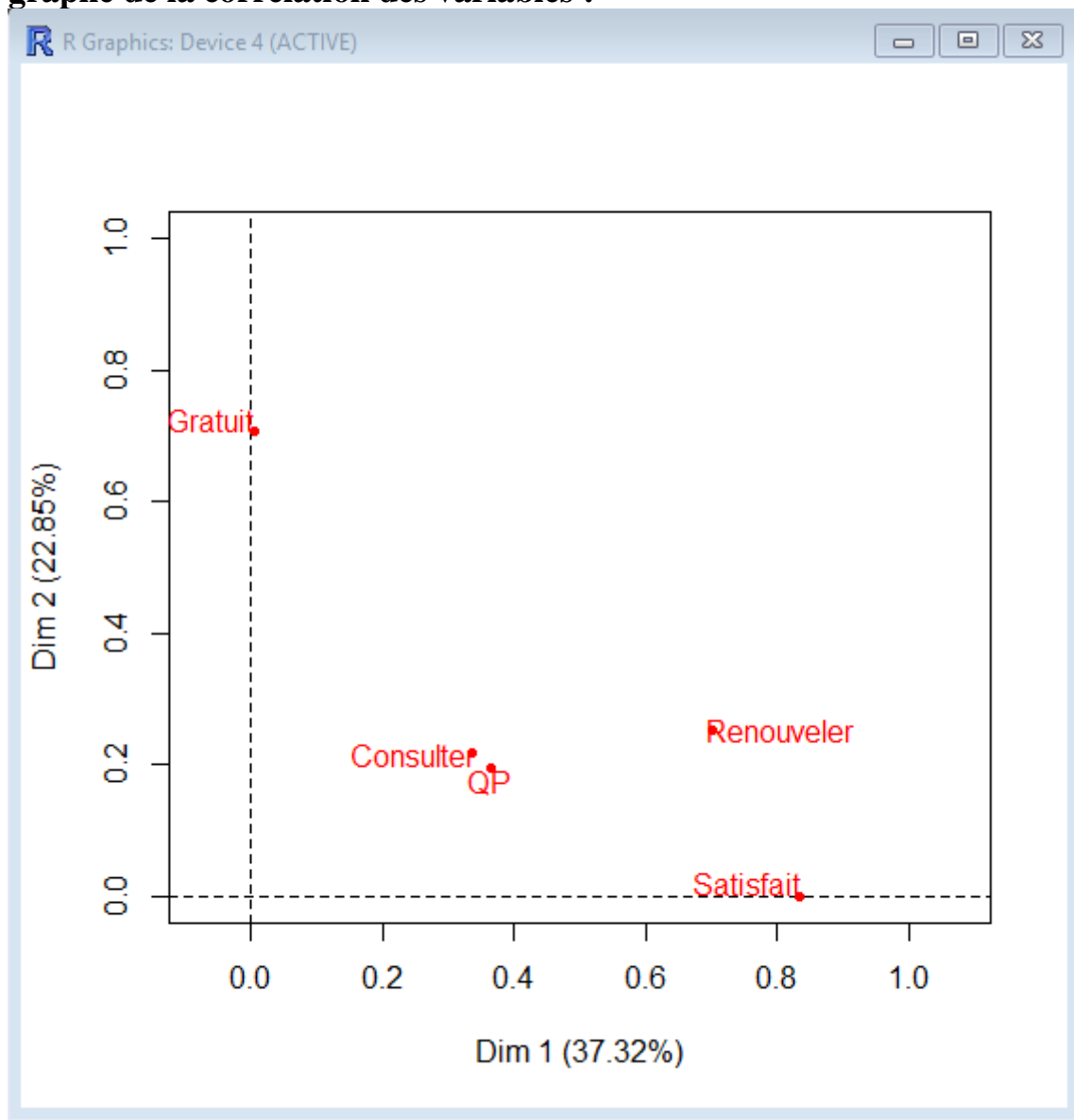
Saisi des données :

```
> don2=read.table(file="Satisfaction_cls_gold.txt",header=TRUE)
> don2
```

	Satisfait	Renouveler	Consulter	QP	Gratuit
1	Oui	Oui	Oui	Oui	Oui
2	Oui	Oui	Oui	Oui	Non
3	Oui	Oui	Non	Oui	Non
4	Oui	NSP	Oui	Oui	Non
5	Oui	NSP	Oui	Oui	Oui
6	Oui	NSP	Non	Oui	Oui
7	Oui	NSP	Oui	Oui	Non
8	Oui	NSP	Oui	Oui	Non
9	Oui	Oui	Non	Oui	Non
10	Oui	Oui	Oui	Non	Non
11	Oui	Oui	Oui	Oui	Non
12	Oui	Oui	Non	Non	Oui
13	Oui	Oui	Non	Oui	Non
14	Oui	Oui	Oui	Non	Non
15	Oui	NSP	Oui	Non	Non
16	Non	Non	Non	Oui	Oui
17	Non	Oui	Oui	Non	Non
18	Non	Non	Oui	Oui	Non
19	Non	Non	Non	Non	Non

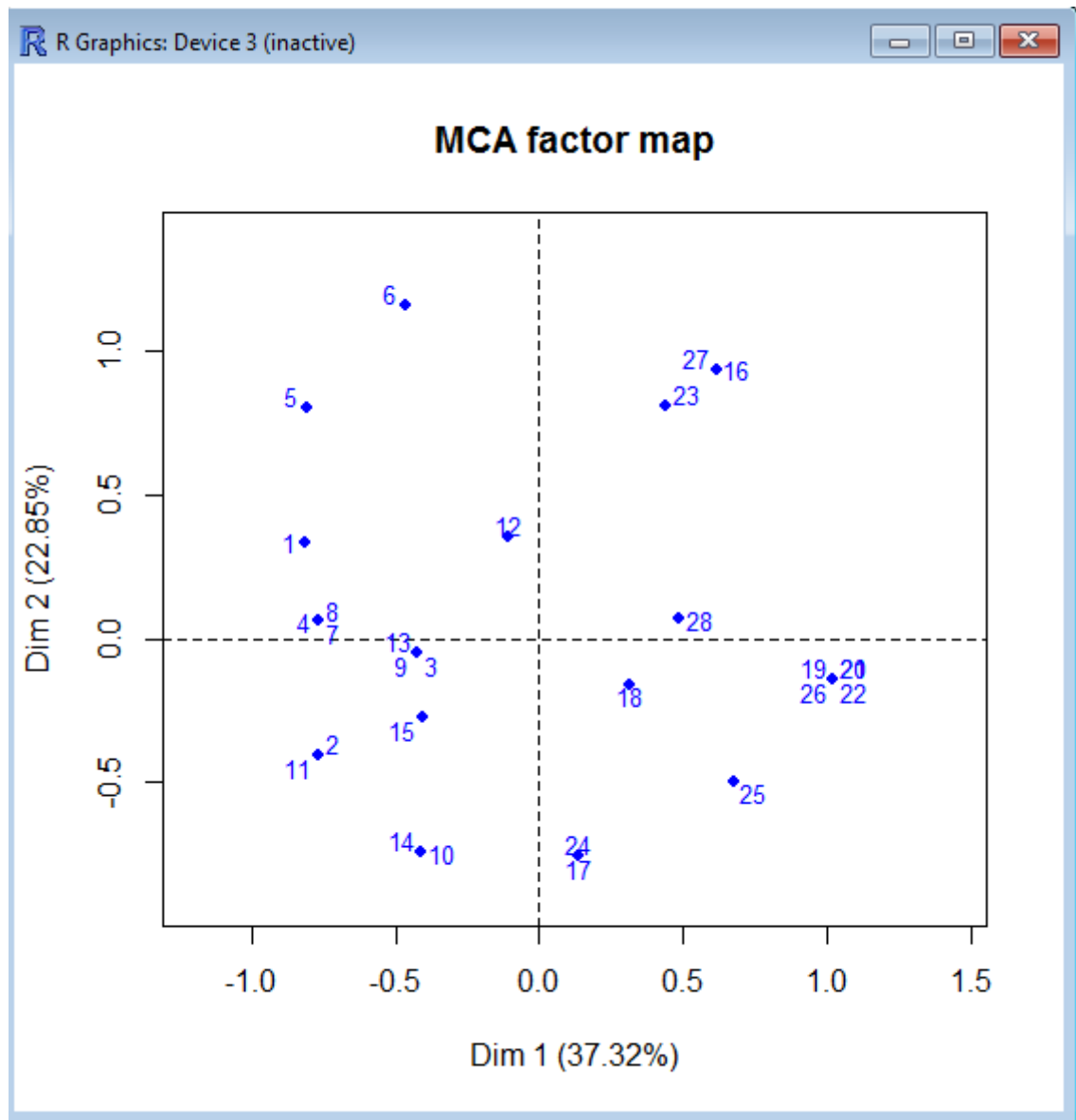
Saisi des données :

L'instruction : `resafm=MCA(don2,2,graph=T)` donne trois graphes significatives
graphe de la corrélation des variables :



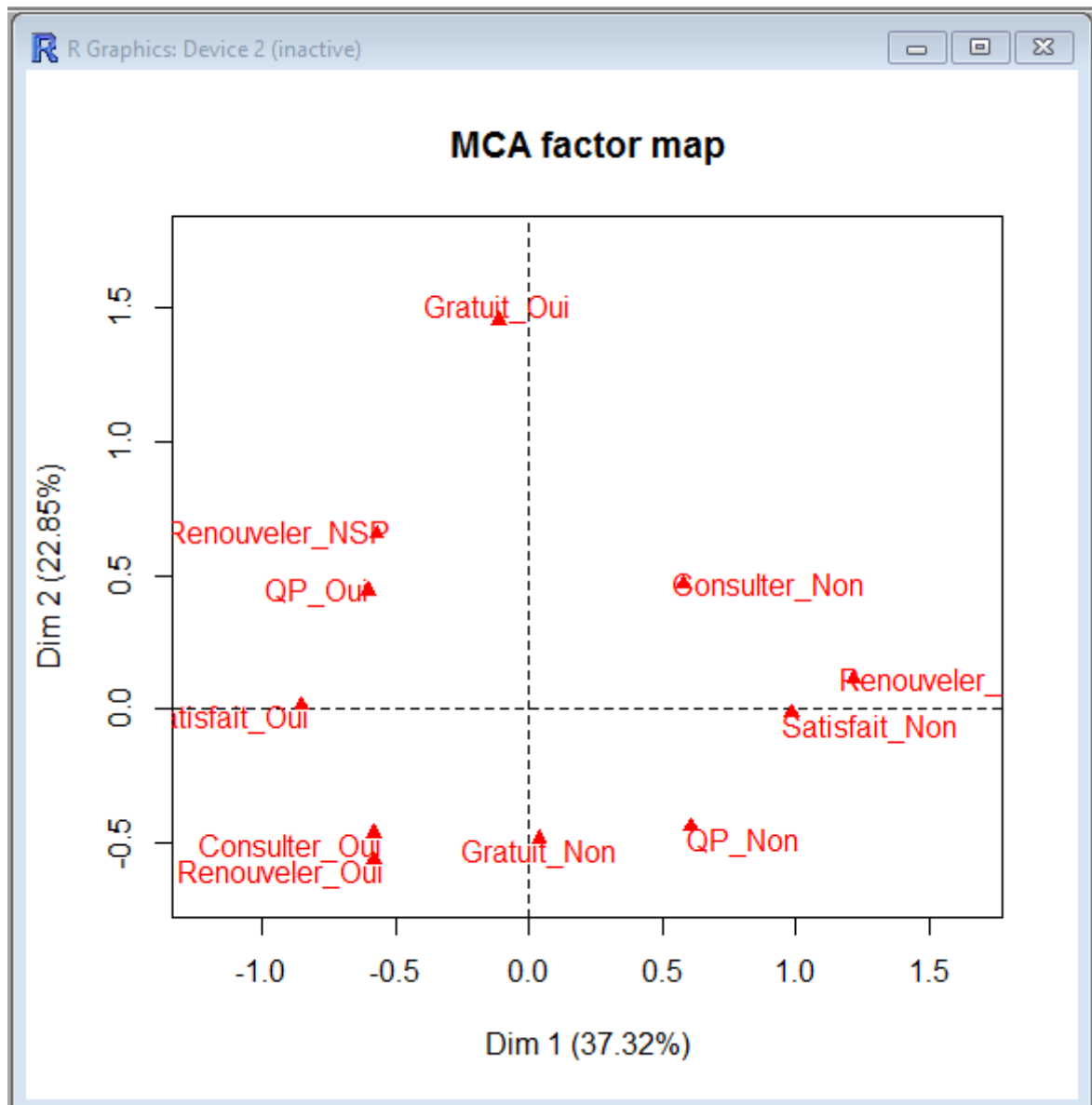
- La variable « Satisfait » est bien corrélée avec le premier facteur de même pour « gratuit »
- Les variables « consulter » et « QP » présentent un coefficient de corrélation inférieur à 40%. Donc, on ne peut pas focaliser notre analyse sur ces deux modalités.

Nuage des points des individus dans le plan factoriel :



On observe une dispersion pratiquement égale des individus sur le plan factoriel sans trouver des données aberrantes.

Nuage des points des variables fournit par la commande MCA :



- Le demi-plan gauche concerne les modalités mettant en valeur la satisfaction des clients alors que le demi-plan droit regroupe les modalités décrivant la non satisfaction des clients.
- Les points de la variable « satisfait » sont superposés au premier axe.
- les points « satisfait_oui » « QP_oui » « Consulter_oui » sont proches.

On peut conclure que

- La satisfaction des clients par le service bancaire est due à la satisfaction de la qualité/prix de l'abonnement annuel de la carte et la consultation parfaite des comptes en ligne.
- Les clients satisfaits ont une certaine ambiguïté de renouveler l'abonnement. Donc, il faut forger la confiance et la fidélité des clients.
- Les assurances voyages gratuites n'intéressent pas les clients. On peut expliquer ça en revenant de l'analyse de la première partie de l'atelier lorsqu'on dit que ces clients ont des salaires élevés. Donc le voyage, pour eux, ne représente pas un grand souci.

Afin de vérifier nos analyses, on visualisera la bonne projection des variables à l'aide de calcul de contribution et le \cos^2 des modalités.

```
> resafm$var$contrib
      Dim 1      Dim 2
Satisfait_Non 19.95492008 0.009406291
Satisfait_Oui 17.29426407 0.008152119
Renouveler_Non 21.25081317 0.284294616
Renouveler_NSP 4.14870679 8.948459769
Renouveler_Oui 5.91829609 9.201294418
Consulter_Non 7.48307143 7.906492380
Consulter_Oui 7.48307143 7.906492380
QP_Non 8.14015048 7.113959010
QP_Oui 8.14015048 7.113959010
Gratuit_Non 0.04663899 12.876872502
Gratuit_Oui 0.13991698 38.630617507

> resafm$var$cos2
      Dim 1      Dim 2
Satisfait_Non 0.83416737 0.0002407671
Satisfait_Oui 0.83416737 0.0002407671
Renouveler_Non 0.70132031 0.0057449340
Renouveler_NSP 0.13007002 0.1717862115
Renouveler_Oui 0.21829422 0.2078117089
Consulter_Non 0.33515548 0.2168332248
Consulter_Oui 0.33515548 0.2168332248
QP_Non 0.36458506 0.1950982306
QP_Oui 0.36458506 0.1950982306
Gratuit_Non 0.00417778 0.7062888716
Gratuit_Oui 0.00417778 0.7062888716
```

- Les données montrent qu'il y a dominance dans la projection des modalités de « satisfait » et la modalité « non » de renouveler
- Les autres modalités prennent presque toutes des valeurs de contribution et des angles de valeurs proches. Ce qui permettra d'analyser leurs interdépendances.

Donc, notre analyse est bien précisée.