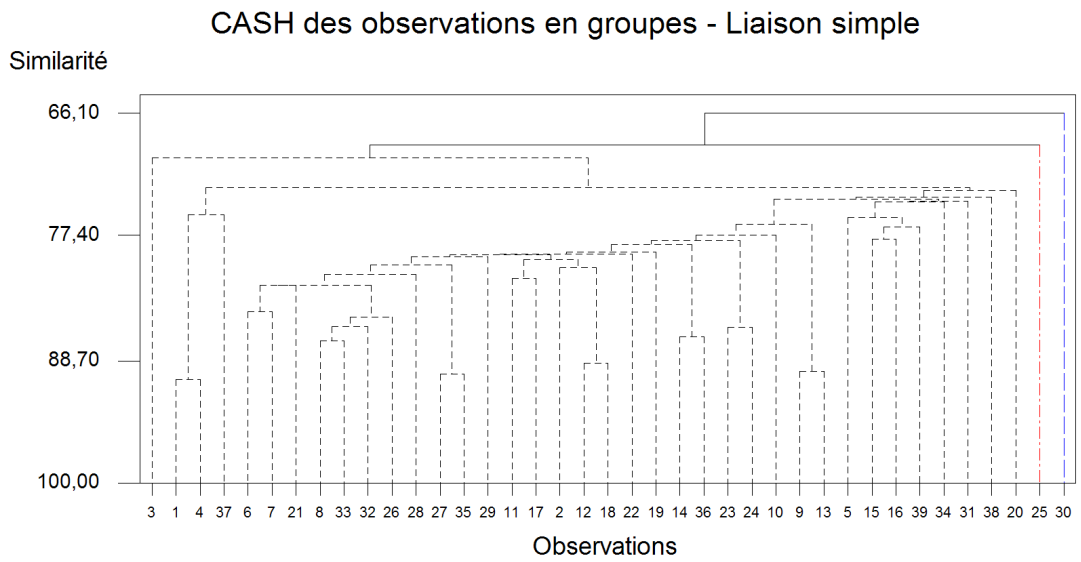
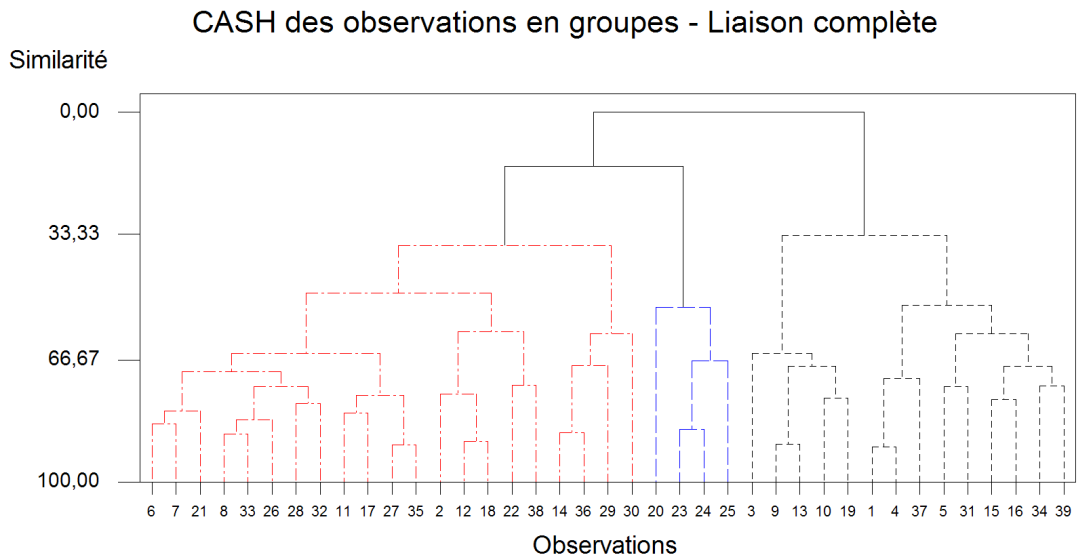
Feuille de Travaux Dirigés n° 6
Classifications

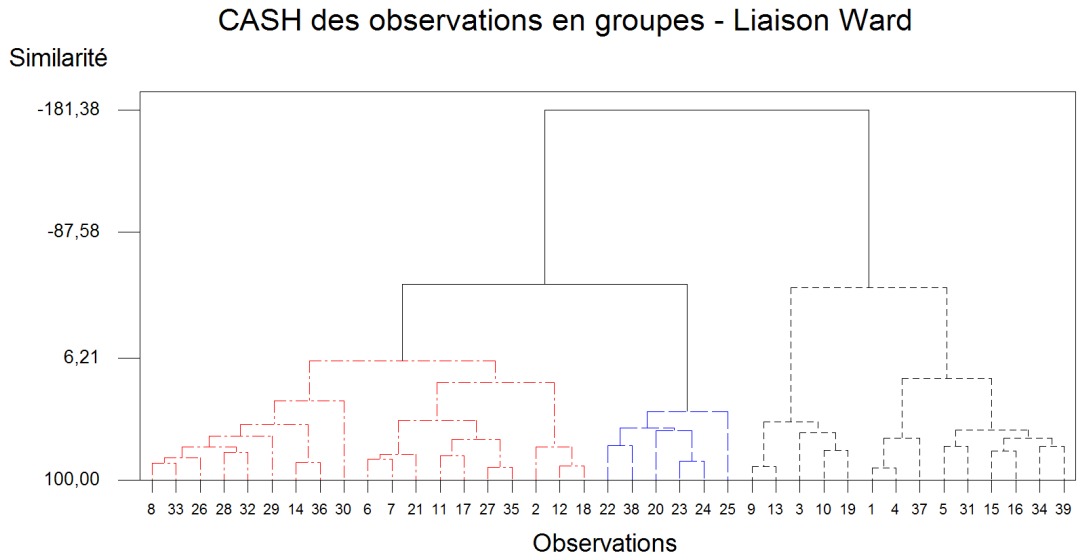
Exercice VI.1. Étude des caractéristiques d'un ensemble d'hôtels

Partie I : Classification hiérarchique ascendante

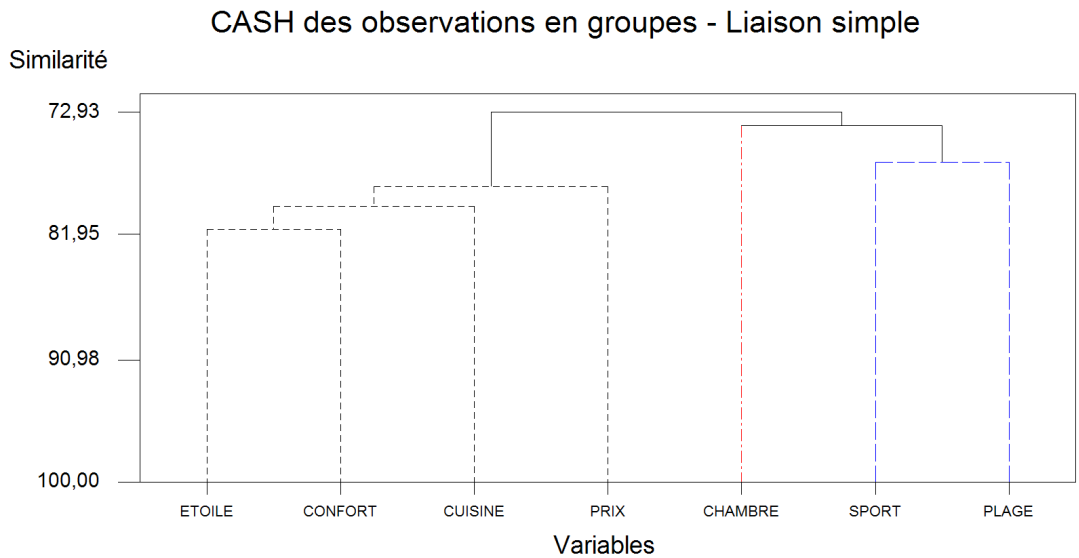
1. Quelles sont les différentes variables reproduites dans le tableau au verso ? Quelle est leur nature ? Qui sont les individus et les variables sur qui on va faire porter la classification hiérarchique ascendante ?
2. Faire la classification hiérarchique ascendante des observations en utilisant la mesure de distance euclidienne et les liaisons simple, complète et de Ward.

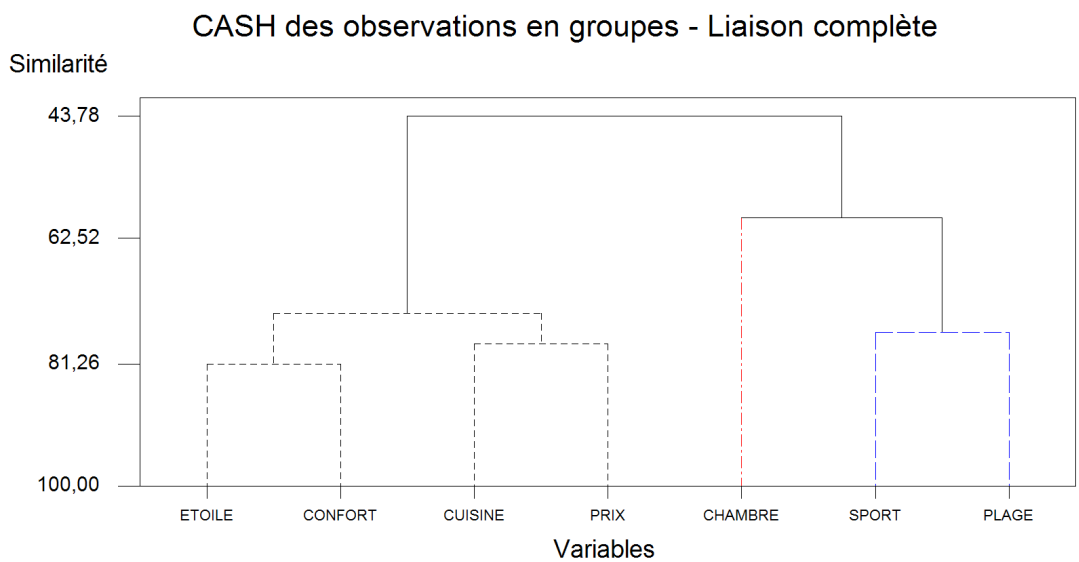


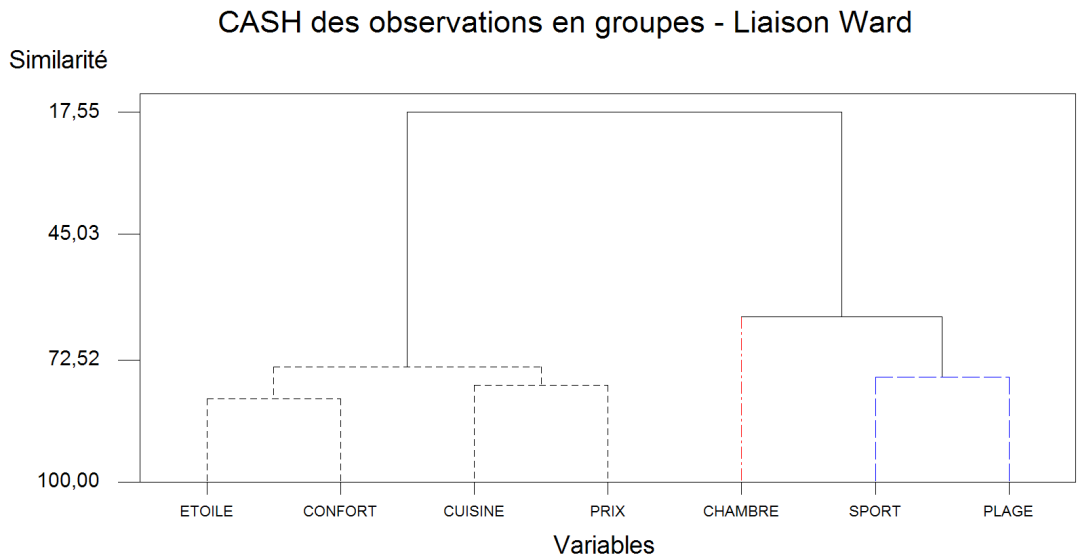




3. Faire la classification hiérarchique ascendante des variables en utilisant la mesure de distance euclidienne et les liaisons simple, complète et de Ward.


CASH des observations en groupes - Liaison simple
Similarité
72,93
81,95
90,98
100,00
ETOILE
CONFORT
CUISINE
PRIX
CHAMBRE
SPORT
PLAGE
Variables


CASH des observations en groupes - Liaison complète
Similarité
43,78
62,52
81,26
100,00
ETOILE
CONFORT
CUISINE
PRIX
CHAMBRE
SPORT
PLAGE
Variables


CASH des observations en groupes - Liaison Ward
Similarité
17,55
45,03
72,52
100,00
ETOILE
CONFORT
CUISINE
PRIX
CHAMBRE
SPORT
PLAGE
Variables

NUMERO	NOM	PAYS	ETOILE	CONFORT	CHAMBRE	CUISINE	SPORT	PLAGE	PRIX
1	Appolpon	Grèce	1	4	56	2	0	8	390
2	Caravel	Grèce	4	7	471	7	6	5	468
3	Christina	Grèce	2	7	93	3	0	5	427
4	Economy	Grèce	1	3	56	1	0	8	369
5	Eden Beach	Grèce	1	4	286	3	4	7	499
6	Hanikian Beach	Grèce	3	6	282	5	10	10	526
7	Marina Beach	Grèce	3	6	310	7	7	10	587
8	Xenia	Grèce	3	4	300	6	10	8	534
9	Agdal	Maroc	4	5	146	5	1	0	447
10	Almohades	Maroc	4	6	250	8	0	3	482
11	Atlas	Maroc	4	5	196	9	6	6	511
12	Atlas Arnadi I	Maroc	4	7	324	10	6	5	532
13	Chems	Maroc	4	5	138	3	2	0	450
14	Dunes d'or	Maroc	3	4	400	10	10	10	569
15	La Résidence	Maroc	0	5	366	7	4	10	419
16	M'Diq	Maroc	0	3	300	5	7	10	421
17	Omayades	Maroc	4	6	144	7	4	8	579
18	Sahara	Maroc	5	7	330	10	5	5	598
19	Sud Bahia	Maroc	4	5	260	5	2	6	495
20	Tour Hassan	Maroc	5	7	170	10	1	10	730
21	Alfa Mar	Portugal	4	6	254	7	10	8	646
22	App,Do Mar	Portugal	4	8	140	7	6	10	652
23	De Lagos	Portugal	4	6	273	10	10	10	802
24	Madeira	Portugal	5	7	260	10	8	10	761
25	Reid's	Portugal	5	7	169	10	10	10	1101
26	Abou Sofiane	Tunisie	3	4	225	5	10	10	434
27	Asdrubal	Tunisie	4	4	225	7	6	10	489
28	Colombes	Tunisie	3	5	250	9	10	8	436

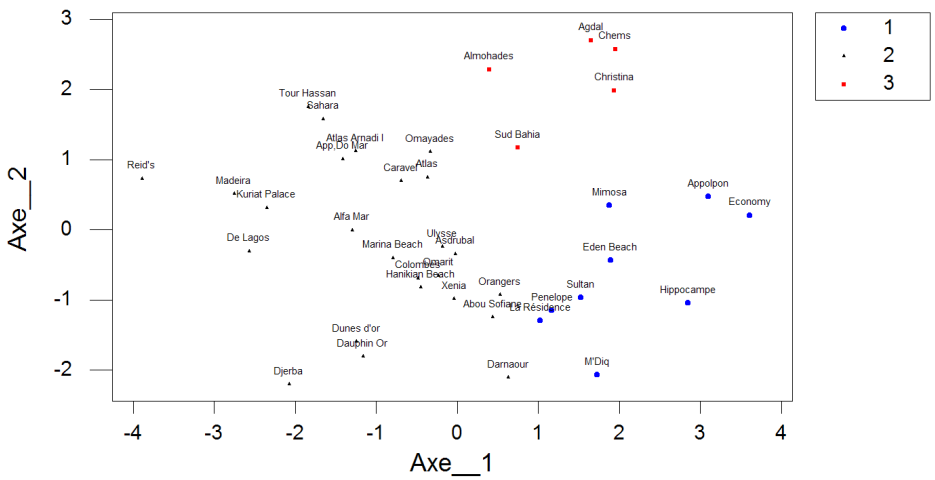
NUMERO	NOM	PAYS	ETOILE	CONFORT	CHAMBRE	CUISINE	SPORT	PLAGE	PRIX
29	Darnaour	Tunisie	2	3	550	6	9	8	399
30	Djerba	Tunisie	3	6	800	10	10	10	477
31	Mimosa	Tunisie	2	4	150	5	6	4	375
32	Omarit	Tunisie	3	5	425	7	7	8	486
33	Orangers	Tunisie	3	4	366	5	8	8	447
34	Penelope	Tunisie	0	5	200	5	10	7	473
35	Ulysse	Tunisie	4	4	130	8	7	10	495
36	Dauphin Or	Turquie	3	4	500	8	10	10	617
37	Hippocampe	Turquie	1	2	50	1	5	10	489
38	Kuriat Palace	Turquie	4	9	232	10	10	10	520
39	Sultan	Turquie	0	3	110	7	6	8	534

. . . . . .

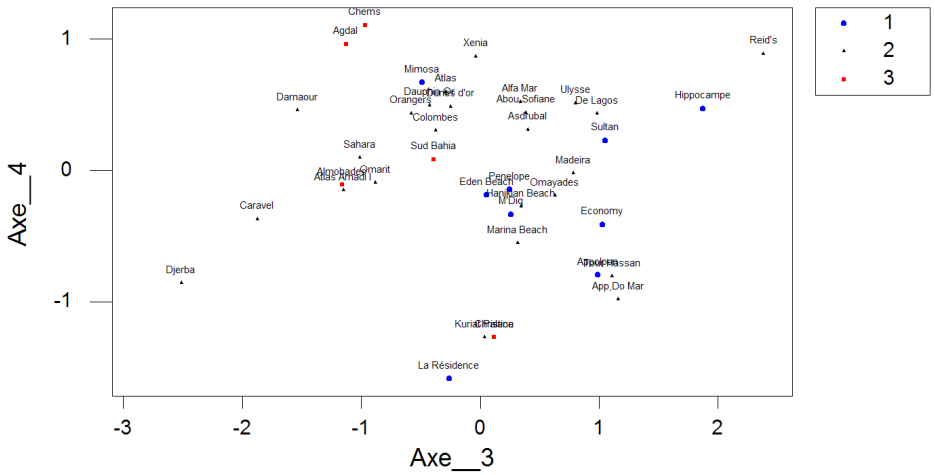
Partie II : K-moyennes

1. Obtenir la classification des hôtels en trois groupes à l'aide de la méthode des $K-$moyennes qui portera sur toutes les variables du tableau. Représenter graphiquement les trois groupes sur les premier et second plans factoriels qui ont été déterminés au TD 7. Qu'observe-t-on ? Comment se répartissent les groupes ?

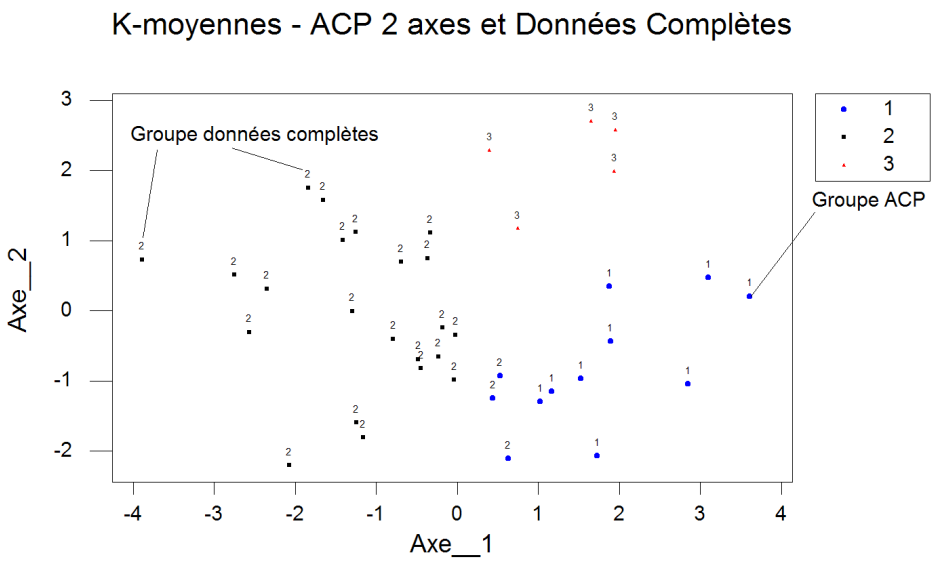
Représentation des groupes obtenus par K-moyennes sur le premier plan factoriel



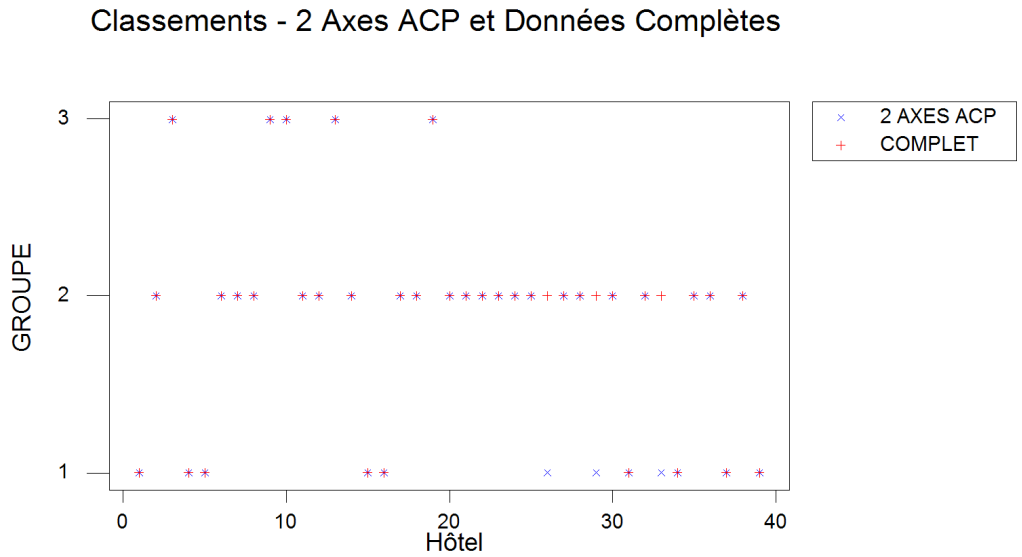
Représentation des groupes obtenus par K-moyennes sur le second plan factoriel



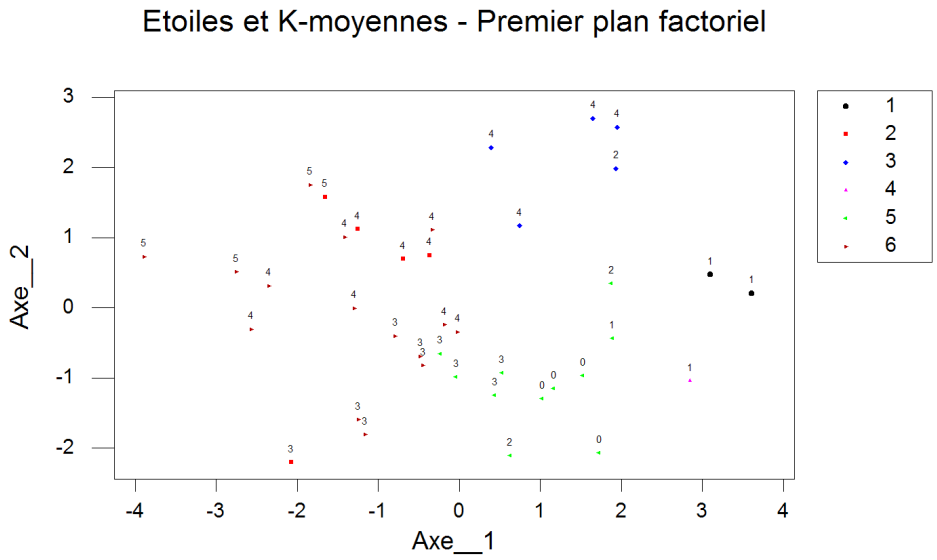
2. Obtenir la classification des hôtels en trois groupes à l'aide de la méthode des $K-$moyennes qui portera sur les coordonnées des hôtels dans le premier plan factoriel. Représenter graphiquement ces trois nouveaux groupes sur le premier plan factoriel. Qu'observe-t-on ? Comment se répartissent les groupes ?

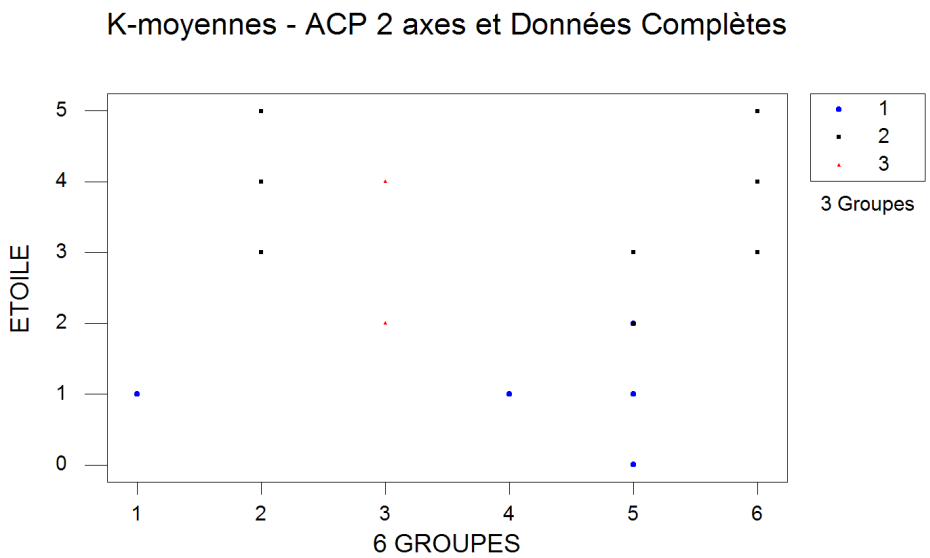


3. Quelles sont les différences de classement entre les deux classements de la question 1. et de la question 2. ? Le premier plan factoriel traduit-il fidèlement l'ensemble des données ? On pourra se réferrer au diagramme des valeurs propres obtenu au TD 7.



4. On décide de vérifier si l'attribution des étoiles est conforme aux critères de constitution des groupes par la méthode des $K-$moyennes. Puisqu'il existe 6 catégories d'étoiles, de 0 à 5, classer les hôtels en 6 groupes à l'aide de la méthode des $K-$moyennes portant cette fois-ci sur toutes les variables à l'exclusion de la variable prix.




K-moyennes - ACP 2 axes et Données Complètes
ETOILE
0
1
2
3
4
5
1
2
3
4
5
6
6 GROUPES
1
2
3
3 Groupes