

Feuille de Travaux Dirigés n° 5

Analyse en composantes principales

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

Partie I

1. Récupérer les données dans R en exécutant les instructions suivantes. Penser à remplacer "C :\\..." par le répertoire dans lequel vous avez enregistré le fichier que vous souhaitez ouvrir.

```
> Chemin <- "C:\\..."
> hotels <- read.csv(paste(Chemin, "MASTER2TD5_EX1.CSV",
+   sep = ""), row.names = 1)
```

2. Quelles sont les différentes variables reproduites dans le tableau au verso ? Quelle est leur nature ? Qui sont les individus sur qui on va faire porter l'analyse en composantes principales ? Obtenir les statistiques descriptives, les covariances et les corrélations entre les variables quantitatives du jeu de données. Créer ensuite le graphique en étoile des hôtels.

	PAYS	ETOILE	CONFORT	CHAMBRE
1	Grèce : 8	Min. :0.000	Min. :2.00	Min. : 50.0
2	Maroc :12	1st Qu. :2.000	1st Qu. :4.00	1st Qu. :148.0
3	Portugal : 5	Median :3.000	Median :5.00	Median :250.0
4	Tunisie :10	Mean :2.974	Mean :5.18	Mean :261.2
5	Turquie : 4	3rd Qu. :4.000	3rd Qu. :6.00	3rd Qu. :317.0
6		Max. :5.000	Max. :9.00	Max. :800.0

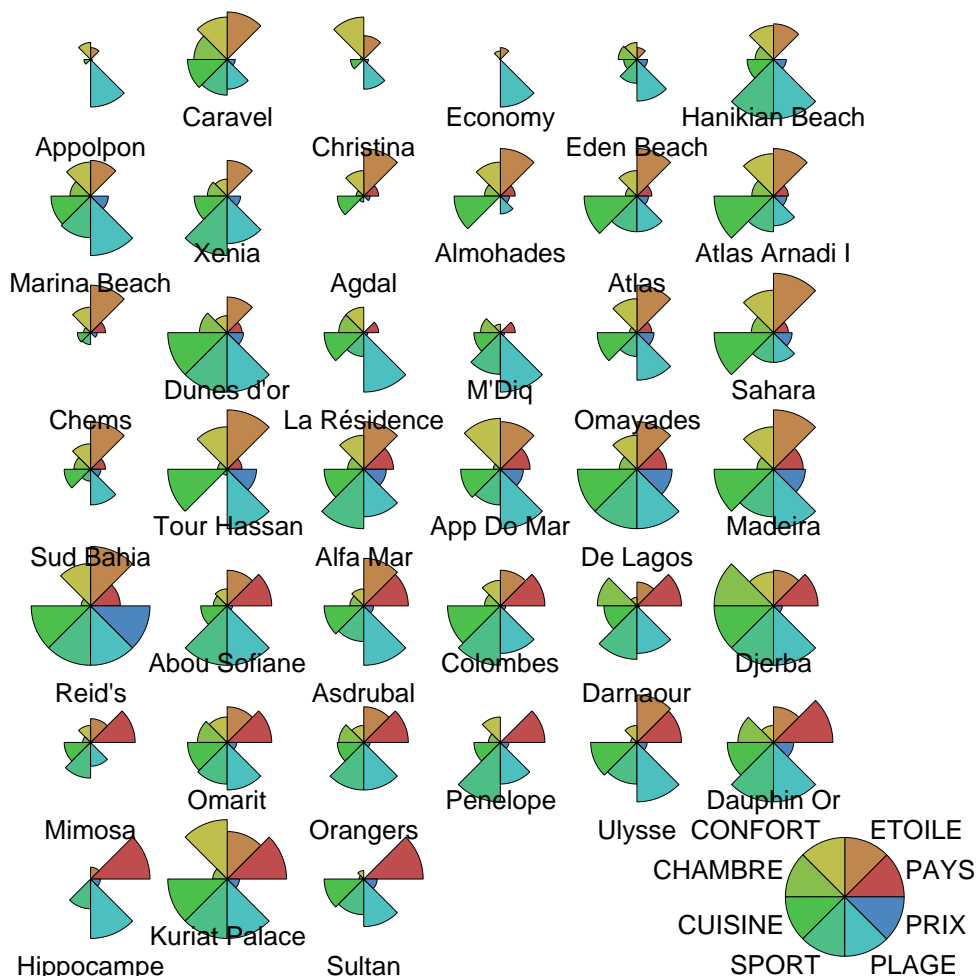
	CUISINE	SPORT	PLAGE	PRIX
1	Min. : 1.000	Min. : 0.000	Min. : 0.00	Min. : 369.0
2	1st Qu. : 5.000	1st Qu. : 4.000	1st Qu. : 6.50	1st Qu. : 447.0
3	Median : 7.000	Median : 6.000	Median : 8.00	Median : 495.0
4	Mean : 6.667	Mean : 6.231	Mean : 7.77	Mean : 529.9
5	3rd Qu. : 9.000	3rd Qu. :10.000	3rd Qu. :10.00	3rd Qu. : 574.0
6	Max. :10.000	Max. :10.000	Max. :10.00	Max. :1101.0

```
> palette(rainbow(12, s = 0.6, v = 0.75))
> stars(hotels, key.loc = c(14.5, 2), draw.segments = T,
+   main = "Diagramme en étoile des hôtels")
> palette("default")
```

	ETOILE	CONFORT	CHAMBRE	CUISINE	SPORT	PLAGE	PRIX
ETOILE	2.24	1.48	18.03	2.36	0.43	-0.51	111.63
CONFORT	1.48	2.47	17.01	2.32	0.19	-0.22	102.57
CHAMBRE	18.03	17.01	22449.75	167.10	246.90	74.76	-721.16
CUISINE	2.36	2.32	167.10	7.02	4.18	1.84	207.25
SPORT	0.43	0.19	246.90	4.18	11.87	4.98	147.87
PLAGE	-0.51	-0.22	74.76	1.84	4.98	7.39	126.37
PRIX	111.63	102.57	-721.16	207.25	147.87	126.37	19006.99

	ETOILE	CONFORT	CHAMBRE	CUISINE	SPORT	PLAGE	PRIX
ETOILE	1.00	0.63	0.08	0.60	0.08	-0.12	0.54
CONFORT	0.63	1.00	0.07	0.56	0.04	-0.05	0.47
CHAMBRE	0.08	0.07	1.00	0.42	0.48	0.18	-0.03
CUISINE	0.60	0.56	0.42	1.00	0.46	0.26	0.57
SPORT	0.08	0.04	0.48	0.46	1.00	0.53	0.31
PLAGE	-0.12	-0.05	0.18	0.26	0.53	1.00	0.34
PRIX	0.54	0.47	-0.03	0.57	0.31	0.34	1.00

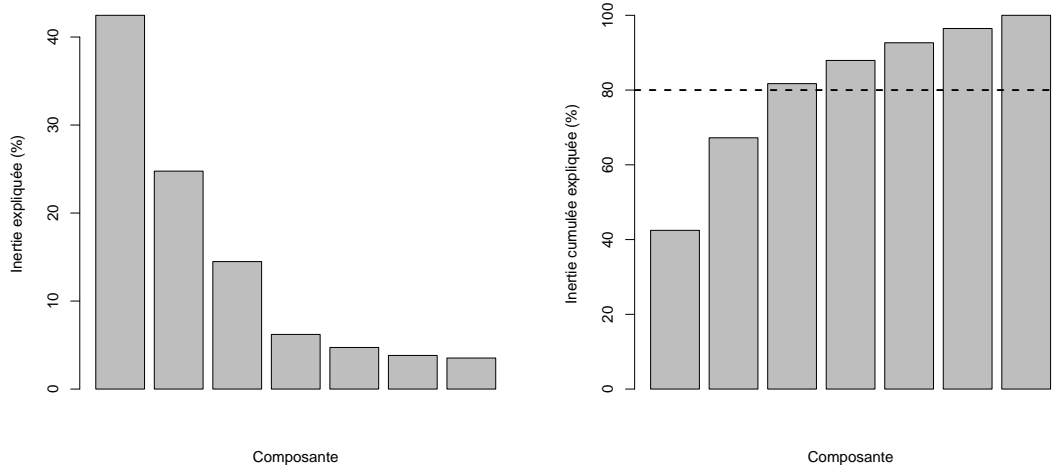
Diagramme en étoile des hôtels



3. Faire l'analyse en composantes principales du tableau de données, puis construire les diagrammes des valeurs propres suivants. Par combien d'axes l'information est-elle résumée de manière satisfaisante ?

```
> library(FactoMineR)
> res.pca <- PCA(hotels, quali.sup = 1, scale.unit = TRUE,
+   ncp = 8, graph = FALSE)

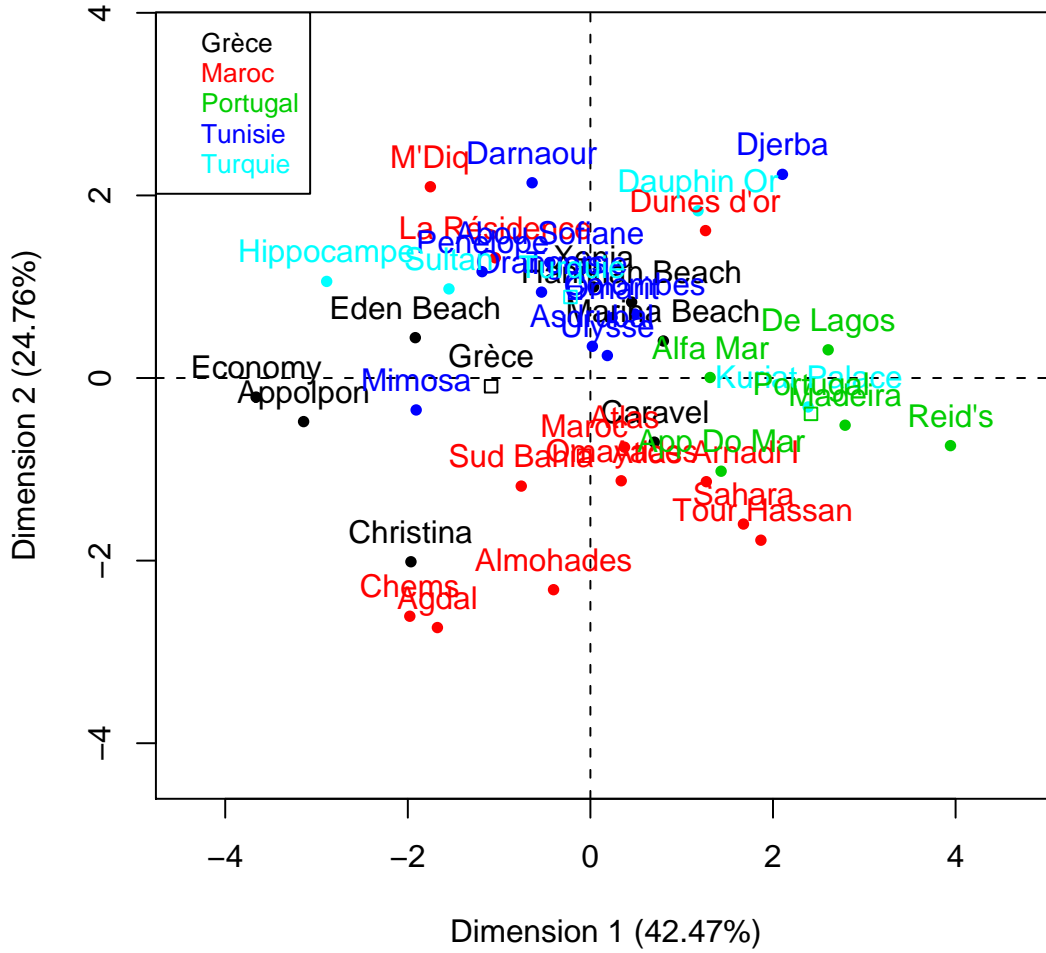
> barplot(res.pca$eig$per, ylab = "Inertie expliquée (%)",
+   xlab = "Composante")
> barplot(res.pca$eig$cum, ylab = "Inertie cumulée expliquée (%)",
+   xlab = "Composante")
> abline(h = 80, lty = 2, lwd = 2)
```



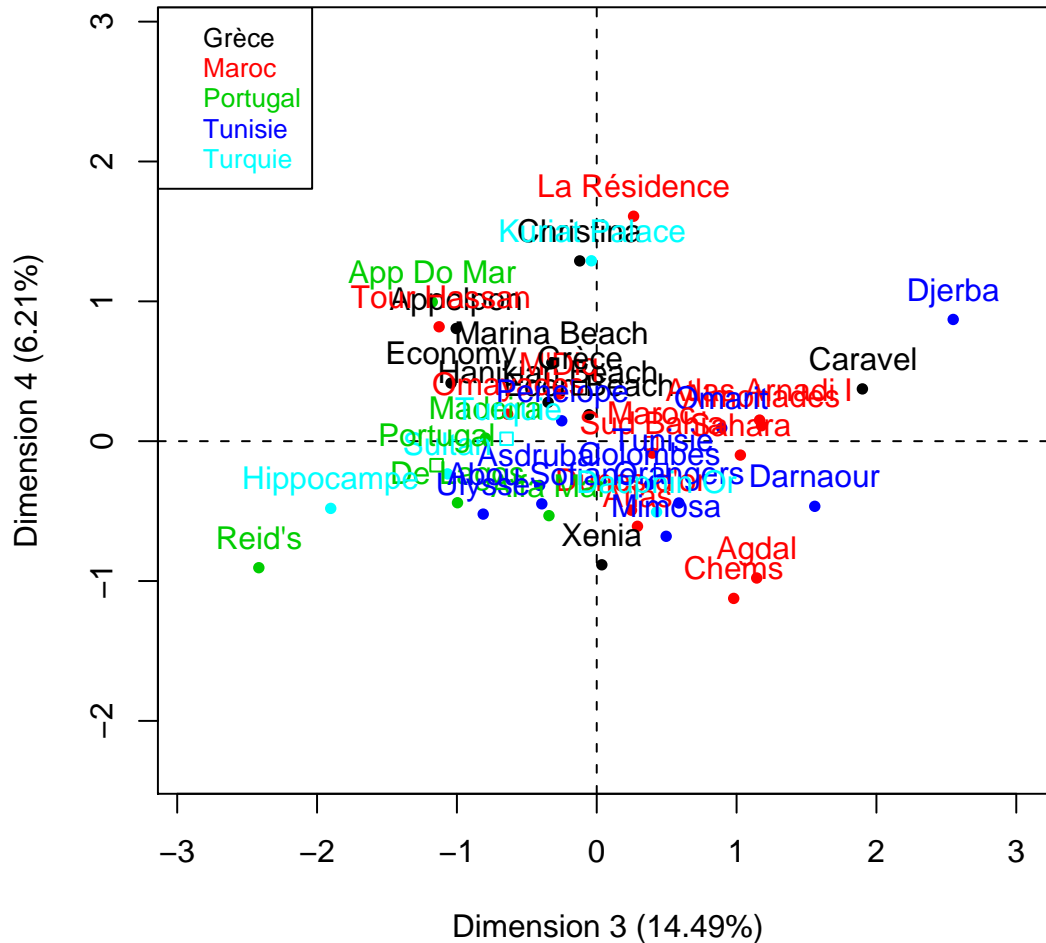
3. Représenter les individus dans le premier et le second plan factoriel en étiquetant les données de telle sorte que l'on puisse identifier à quel hôtel est associé chaque point tout en indiquant les pays où sont implantés les hôtels.

```
> plot(res.pca, choix = "ind", habillage = 1)
> plot(res.pca, choix = "ind", habillage = 1, axes = c(3,
+   4))
```

Individuals factor map (PCA)



Individuals factor map (PCA)

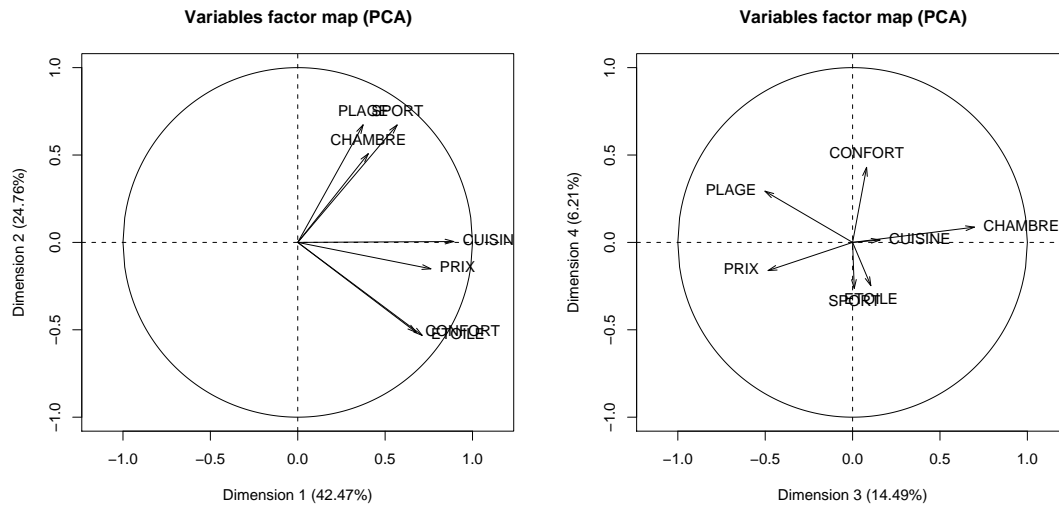


Commenter la qualité de la représentation obtenue sur les quatre premiers axes factoriels en analysant le contenu des tableaux suivants :

- `res.pcavarcoord`, `res.pcavarcor`, `res.pcavarcos2`,
`res.pcavarcontrib`.
- `res.pcaindcoord`, `res.pcaindcos2`, `res.pcaindcontrib`.

4. Construire les cercles des corrélations des variables avec le premier et le second axe factoriel puis avec le troisième et le quatrième axe factoriel. On obtiendra des graphiques similaires à ceux reproduits ci-dessous.

```
> plot(res.pca, choix = "var")
> plot(res.pca, choix = "var", axes = c(3, 4))
```



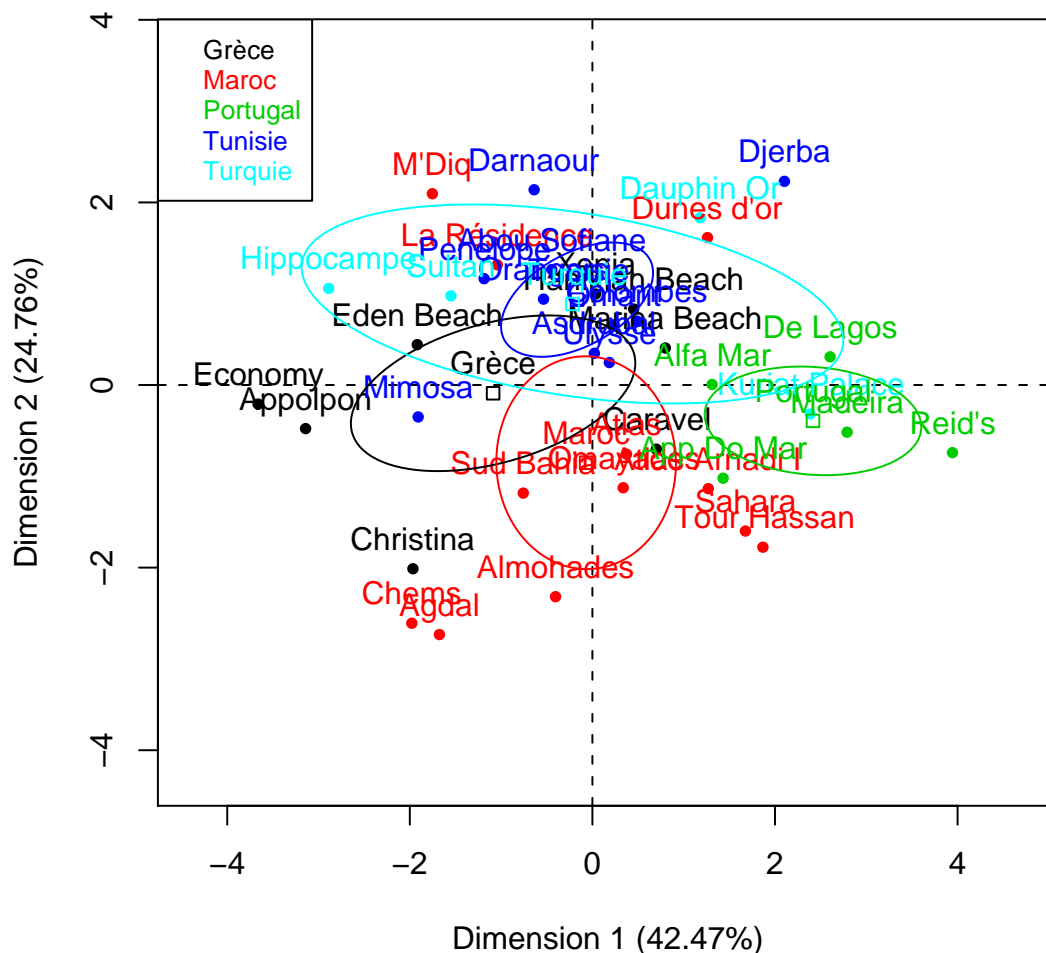
5. La fonction `dimdesc` permet d'obtenir une description automatique des axes de l'ACP. Commenter ses résultats lorsqu'elle est appliquée à `res.pca`.

6. La fonction `coord.ellipse` permet d'obtenir des régions de confiance pour les modalités d'une variable qualitative. Commenter ses résultats lorsqu'elle est appliquée à `res.pca`.

```
> elldata = cbind.data.frame(hotels[, 1], res.pca$ind$coord)
> coordell = coord.ellipse(elldata, bary = TRUE)
> plot.PCA(res.pca, habillage = 1, ellipse = coordell,
+         new.plot = F)
```

```
> plot.PCA(res.pca, habillage = 1, ellipse = coordell,
+         new.plot = F)
```

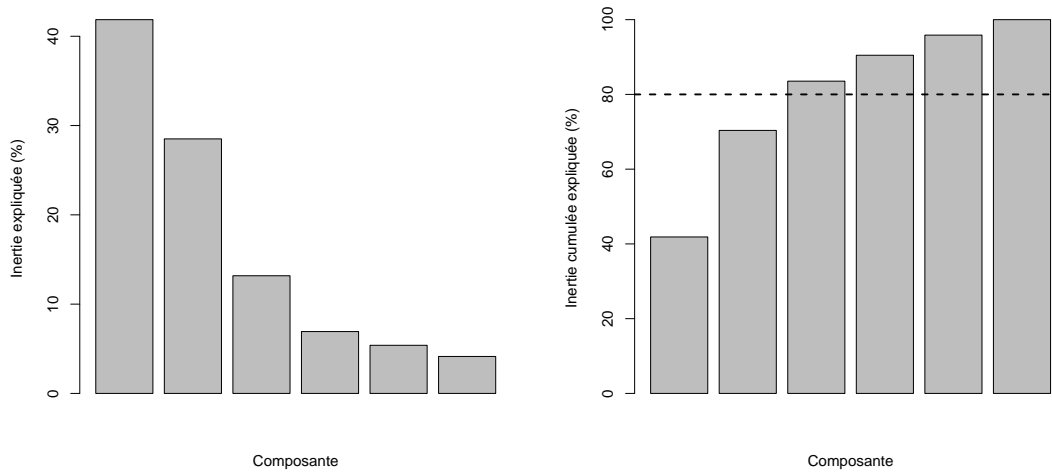
Individuals factor map (PCA)



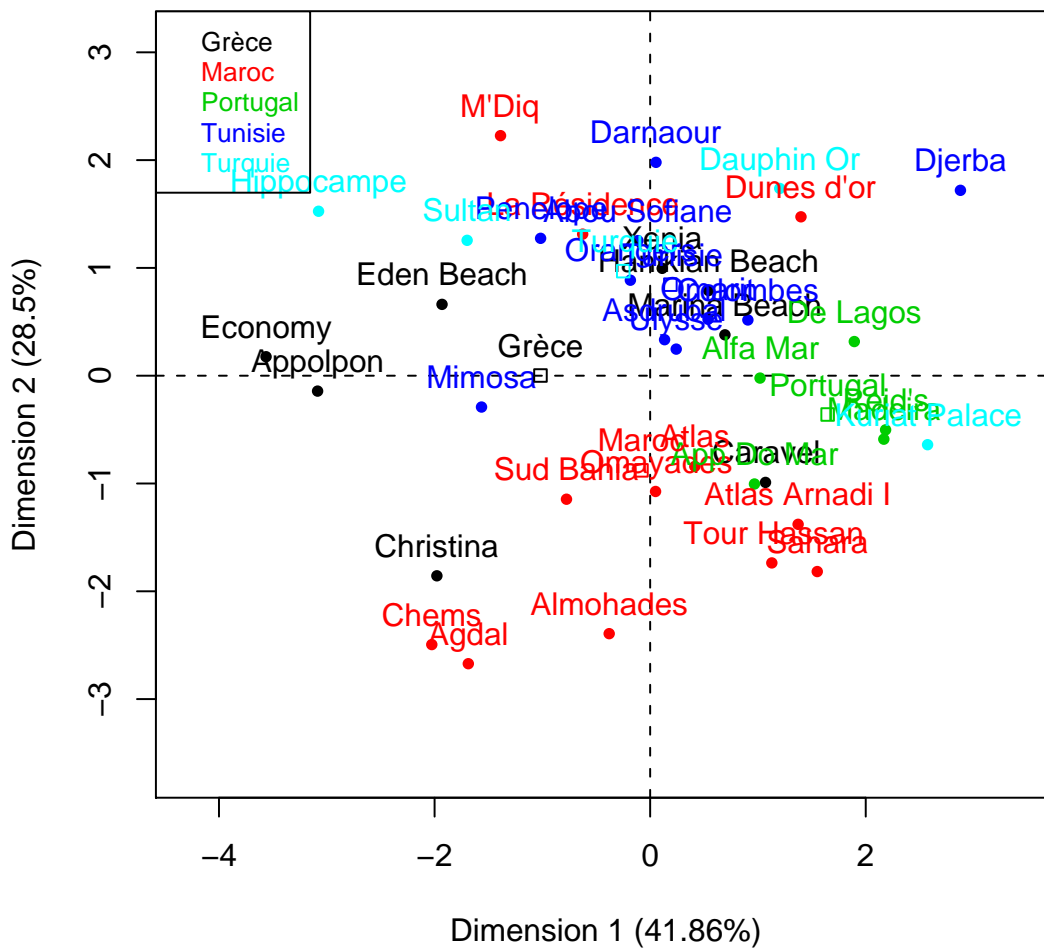
Partie II

Comparer les résultats que vous venez d'obtenir avec ceux reproduits ci-dessous. On commencera par chercher la différence existant entre l'analyse qui vient d'être faite et celle qui a été réalisée ci-dessous. On s'intéressera en particulier au rôle joué par la variable *Prix*. En quoi cette seconde manière d'analyser les données est-elle plus intéressante? Enfin commenter les résultats obtenus.

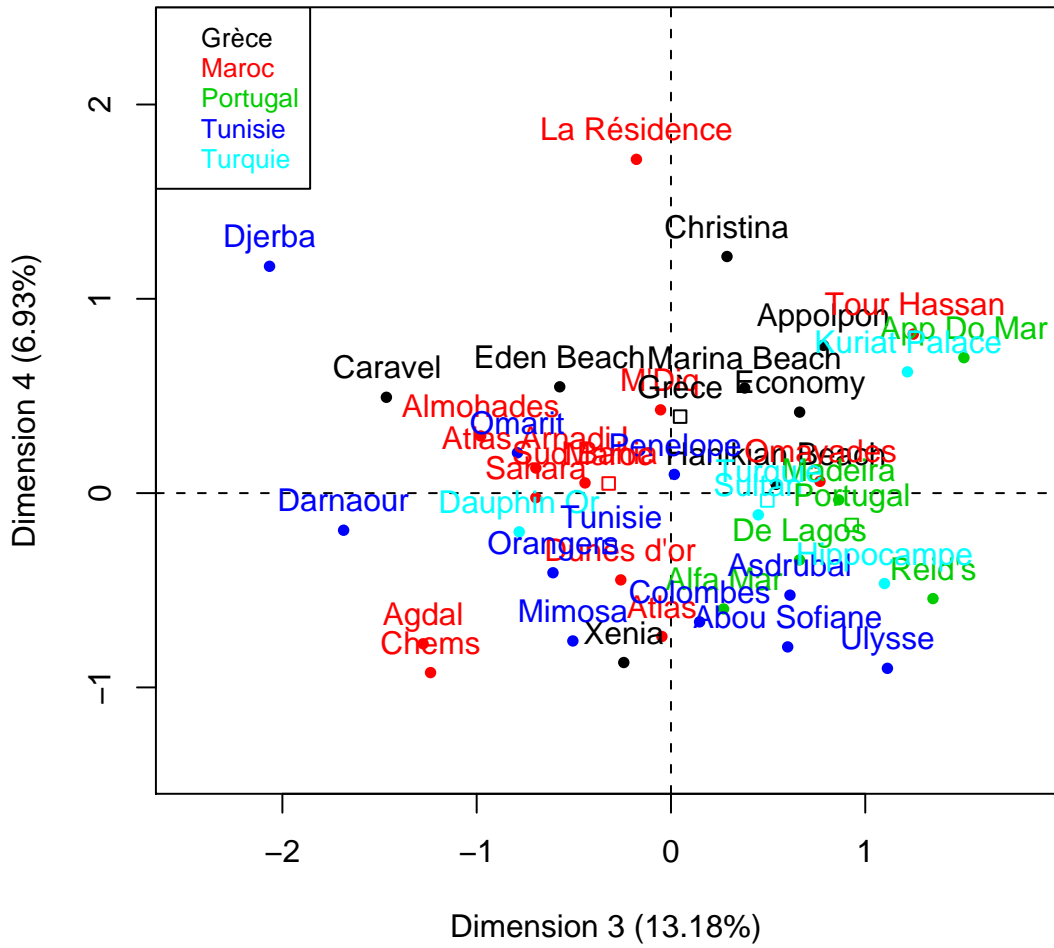
```
> library(FactoMineR)
> res.pca2 <- PCA(hotels, quali.sup = 1, quanti.sup = 8,
+   scale.unit = TRUE, ncp = 8, graph = FALSE)
```



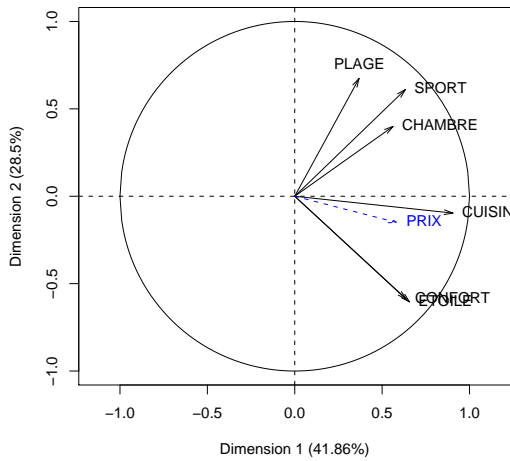
Individuals factor map (PCA)



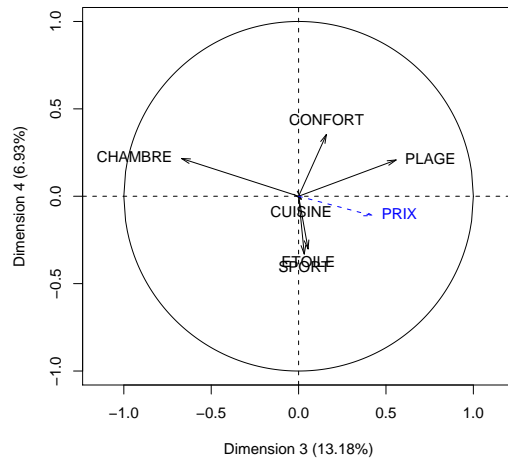
Individuals factor map (PCA)



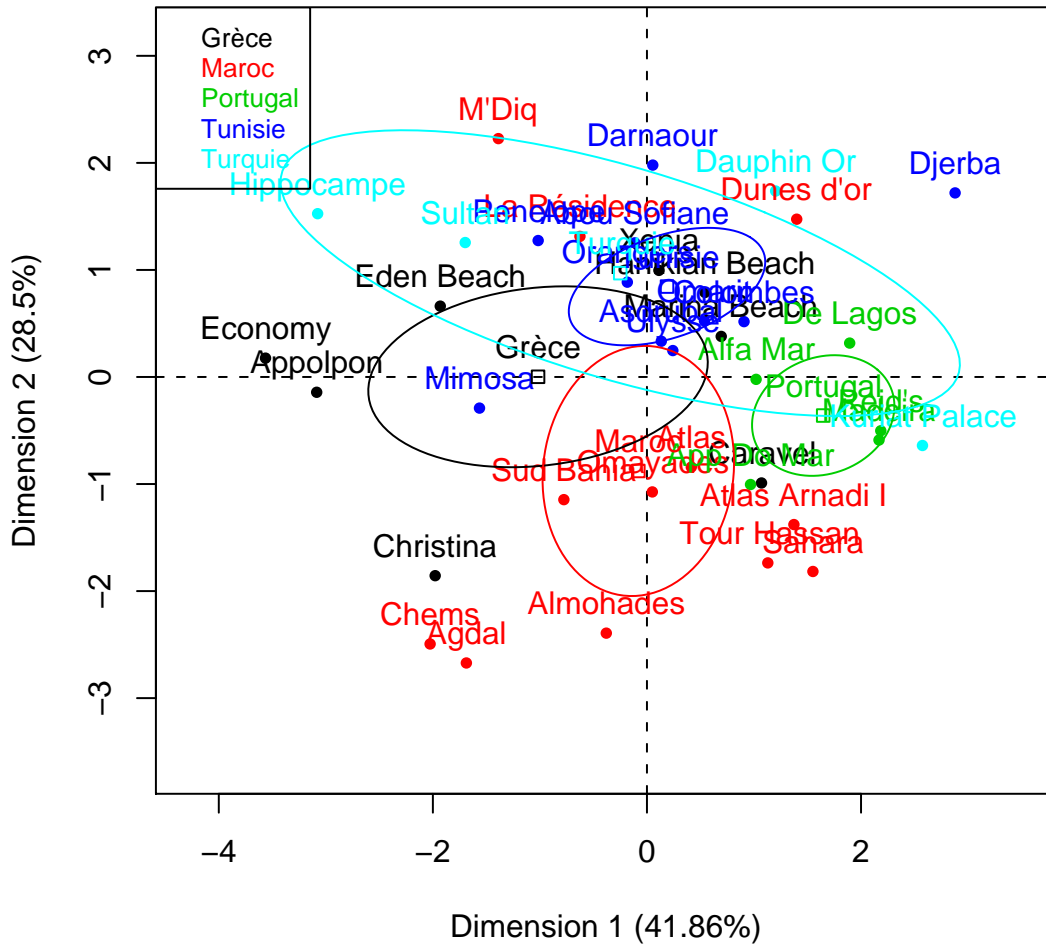
Variables factor map (PCA)



Variables factor map (PCA)



Individuals factor map (PCA)



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

.....