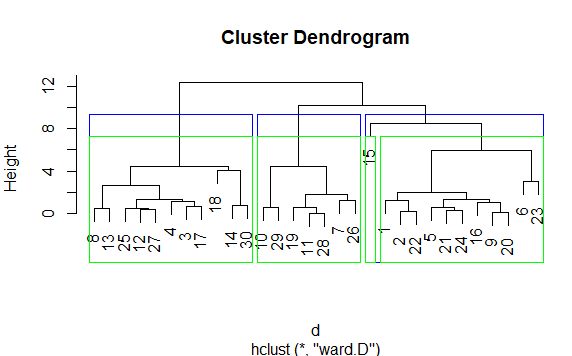
Interpretation of results

Explications script

> projdata <- read.csv2('Tourismdev2015.csv')

> View(projdata)  
> str(projdata)  
'data.frame': 30 obs. of 8 variables:  
 $ ï..Country.Name : Factor w/ 30 levels "Algeria","Argentina",..: 1 2 3 4 5 6 7 8 9 10 ...  
 $ Arrivals : int 1710000 5736000 882000 163000 17971000 119800 56886000 9139000 2622000 84452000 ...  
 $ Adjustednetnationalincome: num -3.069 2.194 4.217 -0.557 -1.578 ...  
 $ Lifeexpectancy : num 75.9 76.4 68.8 59.9 82.1 ...  
 $ GDPpercapita : num 418898 137145 21262 340269 55673 ...  
 $ Democracy : Factor w/ 2 levels "No ","Yes": 2 2 2 2 2 1 2 1 2 2 ...  
 $ highrisksnaturaldisasters: Factor w/ 2 levels "No","Yes": 2 1 1 2 1 2 2 1 1 1 ...  
 $ Conflicts : Factor w/ 2 levels "No","Yes": 1 1 1 2 1 2 2 2 1 1 ...  
  
> grpdata <- projdata[c(2:5)]  
> grpdata.scaled <- scale(grpdata)  
> d <- dist(grpdata.scaled, method = "euclidean")  
> hcward <- hclust(d, method="ward.D")  
> plot(hcward)  
> rect.hclust(tree = hcward, k = 3, border = "blue")  
>   
> rect.hclust(tree = hcward, k = 4, border = "green")

  
>   
>  
> projdata$groups3 <- cutree(hcward,k=3)  
> table(projdata$groups3)  
  
 1 2 3   
12 11 7   
> prop.table(table(projdata$groups3))  
  
 1 2 3   
0.4000000 0.3666667 0.2333333   
> anova\_results <- aov(projdata$Arrivals ~ projdata$groups3)  
> summary(anova\_results)  
 Df Sum Sq Mean Sq F value Pr(>F)   
projdata$groups3 1 6.773e+15 6.773e+15 24.79 **2.94e-05** \*\*\* pvalue de la repartition globale, \*\*\* bonne qualité < 5% on accepte H1, bonne qualité du modèle   
Residuals 28 7.649e+15 2.732e+14   
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
> summary(aov(projdata$Adjustednetnationalincome ~ projdata$groups3))  
 Df Sum Sq Mean Sq F value Pr(>F)   
projdata$groups3 1 188.0 188.0 8.318 0.00747 \*\* **< 5% H0 pas de relation**Residuals 28 632.7 22.6

---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Pvalue = 0,00747 < 0,05, on rejette H0, on accepte H1, c’est-à-dire que les clusters sont significativement différents. Ainsi La moyenne des pib dans les 3 groupes n’est pas la même, donc il y un niveau de richesse différents entre les 3 groupes.

(le niveau de richesses des pays est différent)

A l’intérieur de chaque groupe pour voir si premier groupes groupes des pays les plus riches

> summary(aov(projdata$Lifeexpectancy ~ projdata$groups3))  
 Df Sum Sq Mean Sq F value Pr(>F)  
projdata$groups3 1 5.6 5.6 0.071 0.791 **H1 relation**   
Residuals 28 2195.1 78.4

Pvalue > 5% On rejette H0 on accepte H1

Ainsi la moyenne de l’espérance de vie entre les 3 groupes n’est pas significativement différente

> summary(aov(projdata$GDPpercapita ~ projdata$groups3))  
 Df Sum Sq Mean Sq F value Pr(>F)  
projdata$groups3 1 6.522e+14 6.522e+14 0.827 0.371 **H1 relation**Residuals 28 2.207e+16 7.882e+14

Pvalue > 5% On rejette H0 on accepte H1

Ainsi la moyenne des PIB/tête entre les 3 groupes n’est pas significativement différente, or quand on regarde les valeurs moyennes pour chaque cluster elles sont signifativements différentes ??

> chisq.test(table(projdata$Democracy, projdata$groups3))  
  
 Pearson's Chi-squared test  
  
data: table(projdata$Democracy, projdata$groups3)  
X-squared = 0.35308, df = 2, p-value = 0.8382 **H1 there is a relation**  
>   
> chisq.test(table(projdata$highrisksnaturaldisasters, projdata$groups3))  
  
 Pearson's Chi-squared test  
  
data: table(projdata$highrisksnaturaldisasters, projdata$groups3)  
X-squared = 1.1662, df = 2, p-value = 0.5582 **H1 There is a relation**  
  
> chisq.test(table(projdata$Conflicts, projdata$groups3))  
  
 Pearson's Chi-squared test  
  
data: table(projdata$Conflicts, projdata$groups3)  
X-squared = 0.3824, df = 2, p-value = 0.826 **H1 There is a relation**  
  
  
> apply(grpdata, 2, mean)  
 Arrivals Adjustednetnationalincome Lifeexpectancy   
 1.610504e+07 2.608693e+00 7.329850e+01   
 GDPpercapita   
 8.142074e+06   
>   
> tapply(projdata$Arrivals, projdata$groups3, mean)  
 1 2 3   
 6665075 4813209 50032143

cluster 1 : à en moyenne 6.665075 M d’arrivée de touristes en 2015

cluster 2 : à en moyenne 4.813 209 M d’arrivée de touristes en 2015

cluster 3 : 50.032143 M d’arrivée de touriste en 2015 (Chine, France, All, Thailande, US, Angleterre)

=> Le cluster 3 regroupe les pays qui ont le plus grand nombre d’arrivées de touristes

> tapply(projdata$Adjustednetnationalincome, projdata$groups3, mean)  
 1 2 3   
-1.322879 5.874330 4.216815

cluster 1 : à en moyenne un national income de -1.322879 => comment peut être négatif ?

cluster 2 : à en moyenne un national income de 5.874330

cluster 3 : à en moyenne un national income de 4.216815

⇒ Cluster 2 : regroupement de pays avec les national income les plus élevés

=> Cluster 1 : regroupement de pays avec les national income les plus faibles

>   
> tapply(projdata$Lifeexpectancy, projdata$groups3, mean)  
 1 2 3   
77.27267 65.54491 78.66985   
cluster 1 : à en moyenne une espérance de vie de 77.27267

cluster 2 : à en moyenne une espérance de vie de 65.54491 DIFFERENCE ENTRE 1 ET 3?

cluster 3 : à en moyenne une espérance de vie de 78.66985

=> Le cluster 2 représente le groupe de pays qui a l’espérance de vie la plus faible et le cluster 3 celui qui a la plus grande, elle est comparable avec le cluster 1

> tapply(projdata$GDPpercapita, projdata$groups3, mean)  
 1 2 3   
12509775.75 8508315.08 79062.29

cluster 1: a en moyenne un GDP per capita de 1 250 977.75

cluster 2: a en moyenne un GDP per capita de 8 508 315.08

cluster 3: a en moyenne un GDP per capita de 79 062.29

=> Le cluster 3 est le groupe où le PIB par tête est le plus bas.

=> Le cluster 2 est le groupe où le PIB par tête est le plus haut.

>   
> projdata$groups3 <- factor(projdata$groups3)  
> TukeyHSD(aov(projdata$Arrivals ~ projdata$groups3))  
 Tukey multiple comparisons of means  
 95% family-wise confidence level  
  
Fit: aov(formula = projdata$Arrivals ~ projdata$groups3)  
  
$`projdata$groups3`  
 diff lwr upr p adj  
2-1 -1851866 -14278856 0.9277122 cluster 2 et 1, pvalue, h0 accepté, arrivées entre cluster 1 et cluster 2 ne sont pas significativement différents, ils sont comparables, pas indépendants

3-1 43367068 29208277 57525859 0.0000001 pvalue, H1 accepté, différence significative en terme d’arrivées

3-2 45218934 30824987 59612881 0.0000001 3 est vraiment différent   
>   
> TukeyHSD(aov(projdata$Adjustednetnationalincome ~ projdata$groups3))  
 Tukey multiple comparisons of means  
 95% family-wise confidence level  
  
Fit: aov(formula = projdata$Adjustednetnationalincome ~ projdata$groups3)  
  
$`projdata$groups3`  
 diff lwr upr p adj  
2-1 7.197209 2.7443131 11.65010 0.0012232 Le 1 serait le chef de file, le 1 est vraiment différent des 2 autres   
3-1 5.539694 0.4662518 10.61314 0.0302514  
3-2 -1.657515 -6.8152192 3.50019 0.7081500 ce la sont comparable   
  
>   
> TukeyHSD(aov(projdata$Lifeexpectancy ~ projdata$groups3))  
 Tukey multiple comparisons of means  
 95% family-wise confidence level  
  
Fit: aov(formula = projdata$Lifeexpectancy ~ projdata$groups3)  
  
$`projdata$groups3`  
 diff lwr upr p adj  
2-1 -11.727760 -18.476033 -4.979486 0.0005537  
3-1 1.397185 -6.291515 9.085884 0.8945726  
3-2 13.124945 5.308547 20.941342 0.0008130

3 et 1 significayivement semblables, 2 chef de file

>   
> TukeyHSD(aov(projdata$GDPpercapita ~ projdata$groups3))  
 Tukey multiple comparisons of means  
 95% family-wise confidence level  
  
Fit: aov(formula = projdata$GDPpercapita ~ projdata$groups3)  
  
$`projdata$groups3`  
 diff lwr upr p adj  
2-1 -4001461 -33569344 25566422 0.9399622  
3-1 -12430713 -46119117 21257691 0.6357527  
3-2 -8429253 -42677170 25818665 0.8157874

3 groupes sont pareils

> > prop.table(table(projdata$Democracy, projdata$groups3), 2)  
   
 1 2 3  
 No 0.2500000 0.1818182 0.1428571  
 Yes 0.7500000 0.8181818 0.8571429

Democracy No 0 Yes 1

The 3 clusters are democratic countries

Pas intéressant de l’analyser puisque H1 accepté, pas significativement différent entre clusters

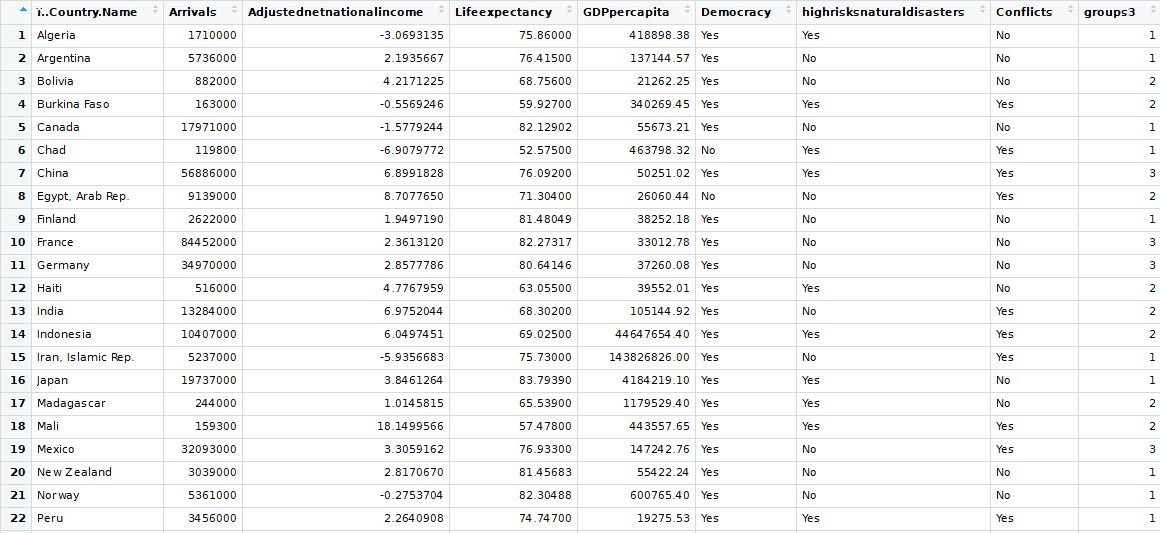
>   
> prop.table(table(projdata$highrisksnaturaldisasters, projdata$groups3), 2)  
   
 1 2 3  
 No 0.5833333 0.3636364 0.4285714  
 Yes 0.4166667 0.6363636 0.5714286  
cluster 1 : 58,33 % où il n’y a pas de grands risques de désastre naturel, 41,67 % où il y a un gros risque de désastre naturel.

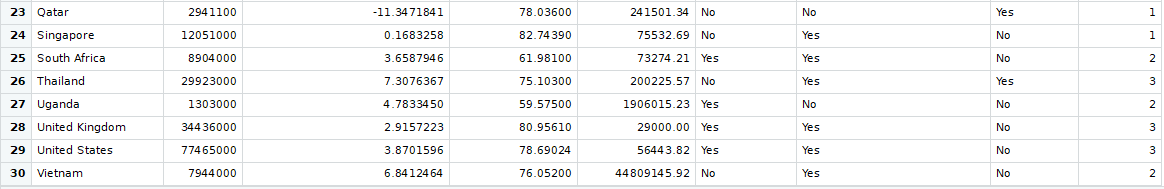
cluster 2 : 36,36 % oû il n’y a pas de grands risques de désastre naturel, 63,63 % où il y a un gros risque de désastre naturel.

cluster 3 : 42,86 % où il n’y a pas de grands risques de désastre naturel, 57,14 % où il y a un gros risque de désastre naturel.

Tous les clusters regroupent des pays avec des risques de catastrophes naturelles relativement importantes, on peut noter que le groupe 1 est moins risqué que le groupe 3, lui même moins risqué que le groupe 3

> prop.table(table(projdata$Conflicts, projdata$groups3), 2)  
   
 1 2 3  
 No 0.6666667 0.5454545 0.5714286  
 Yes 0.3333333 0.4545455 0.4285714  
>





Clusters description:

Test en enlevant des variables,

Démarchage :

30 pays, variables quantitatives (National income, GPR per capita, Life expectancy, Arrivals) et qualitatives (democracy, risque naturel, conflicts)

On prends les pays on les séparent en différents groupes, on choisit la séparation en bleu, c’est à dire de ne prendre que 3 groupes

Ensuite on réalise le test anova pour le global (pour les variables quantitatives), on obtient la p-value de la répartition globale qui est de 3 étoiles, \*\*\* donc de bonne qualité, < 5% on accepte H1, bonne qualité du modèle

Puis on réalise le test anova pour chaque variables quantitatives,

* National income : On voit que la moyenne des PIB dans les 3 groupes n’est pas la même donc les niveaux de richesses sont différents.
* Espérance de vie : On voit que la p-value est inférieur à 5% donc on accepte H1. Ainsi la moyenne de l’espérance de vie entre les 3 groupes n’est pas significativement différente.
* PIB par habitant : On voit que la p-value est supérieur à 5% donc on accepte H1. Ainsi la moyenne des PIB/tête entre les 3 groupes n’est pas significativement différente.

Et le test de chi2 pour les variables qualitatives :

* Democracy : On voit que le test n’est pas significatif car la p-value n’est pas supérieur à 5%. La variable démocratie n’influe donc pas significativement sur les arrivées touristiques de nos pays.
* Conflicts : On voit que le test n’est pas significatif car la p-value n’est pas supérieur à 5%. La variable conflits n’influe donc pas significativement sur les arrivées touristiques de nos pays.
* High natural risks : On voit que le test n’est pas significatif car la p-value n’est pas supérieur à 5%. La variable risques naturels n’influe donc pas significativement sur les arrivées touristiques de nos pays.

Ensuite on regarde pour chaque cluster les moyennes des :

* arrivées :

=> Le cluster 3 regroupe les pays qui ont le plus grand nombre d’arrivées de touristes, la plus grande espérance de vie, et le PIB par tête le plus haut.

* national income :

=>Le Cluster 2 représente le regroupement de pays avec les national income les plus élevés, qui a l’espérance de vie la plus faible ainsi que les pays où le PIB par tête est le plus haut

=> Le cluster 1 représente le regroupement de pays avec les national income les plus faibles

* espérance de vie :

=> Le cluster 2 représente le groupe de pays qui a l’espérance de vie la plus faible et le cluster 3 celui qui a la plus grande, elle est comparable avec le cluster 1

* PIB/tête :

=> Le cluster 3 est le groupe où le PIB par tête est le plus bas.

=> Le cluster 2 est le groupe où le PIB par tête est le plus haut.

TukeyHSD correspond aux intervalles de confiances :

pour arrivées :

cluster 2-1 : h0 accepté, arrivées entre cluster 1 et cluster 2 ne sont pas significativement différents, ils sont comparables, pas indépendants

cluster 3-1 :pvalue, H1 accepté, différence significative en terme d’arrivées

cluster 3-2 : significativements différents

Donc c’est le cluster 3 qui vraiment différents en terme d’arrivées, donc qui acceuille le plus de monde

pour national income :

Le 1 a vraiment un national income différent des 2 autres

pour espérance de vie :

Le 2 est significativement différent des 2 autres

Le 3 et le 1 sont comparables entre eux

pour le PIB/tête :

Les 3 groupes sont semblables ??

prop.table : permet de connaître pour les caractéristiques de chaque clusters pour les variables quantitatives

pour démocratie :

Les pays dans les 3 clusters sont des pays démocratiques

Il n’y a pas de différences significatives

pour les risques de catastrophes naturelles : on remarque que tous les pays des 3 clusters sont des pays avec un taux de potentielles catastrophes naturelles relativement important, bien que le cluster 2 soit plus sujet aux catastrophes naturelles et le cluster 1 le moins

pour les conflits :

Les pays des différents clusters sont des pays avec relativement peu de risque de conflits. On notera cependant que le cluster 2 est le cluster ou il y a le plus de risque de conflits

Est-ce-que les variables de sécurité et/ou économique influences l’afflux touristiques dans le monde ?

En conclusion les variables de sécurité et économique ne sont pas une barrière pour le tourisme sur les pays étudiés en 2015

Montre que par exemple une entreprise qui voudrait s’implanter dans des pays touristiques n’as pas besoin de prendre en compte les variables sécurité et économiques pour s’installer