

World Happiness

Harkiolakis Laurent and Hedrache Hanene
31 mars 2017

Question

A travers notre analyse de deux rapports réalisés en 2015 et en 2016 sur la question du bonheur à l'échelle mondiale, nous avons voulu répondre à la problématique suivante : Quel est le facteur qui a le plus influencé le bonheur entre 2015 et 2016 dans le monde ?

```
# package
library(dplyr);

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(ggplot2);
library(gridExtra);

##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##   combine
```

Nous avons d'abord chargé les fichiers pour les afficher. Nous avons dans un premier temps choisi de représenter les pays en fonction des facteurs qui sont Economy, Family, Freedom, Government Corruption, Generosity. Ces facteurs ont à l'exception du facteur Family étaient obtenus suite à des questions binaires du type : Êtes-vous satisfait de la politique de santé dans votre pays ?

```
df <- read.csv("2015.csv", header=TRUE)
head(df);
```

	Country	Region	Happiness.Rank	Happiness.Score	Standard.Error
## 1	Switzerland	Western Europe	1	7.587	0.03411
## 2	Iceland	Western Europe	2	7.561	0.04884

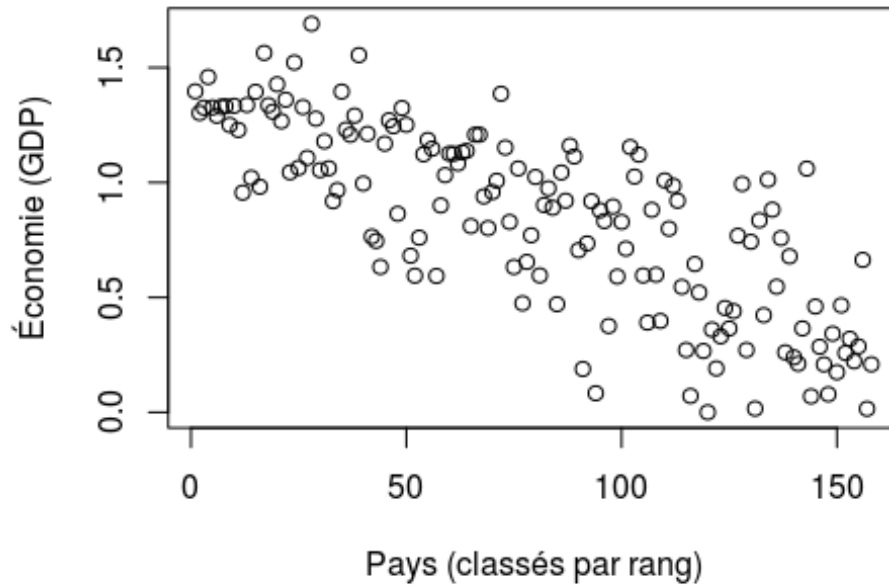
```
## 3    Denmark Western Europe      3      7.527    0.03328
## 4    Norway Western Europe      4      7.522    0.03880
## 5    Canada North America      5      7.427    0.03553
## 6    Finland Western Europe     6      7.406    0.03140
## Economy..GDP.per.Capita. Family Health..Life.Expectancy. Freedom
## 1      1.39651 1.34951      0.94143 0.66557
## 2      1.30232 1.40223      0.94784 0.62877
## 3      1.32548 1.36058      0.87464 0.64938
## 4      1.45900 1.33095      0.88521 0.66973
## 5      1.32629 1.32261      0.90563 0.63297
## 6      1.29025 1.31826      0.88911 0.64169
## Trust..Government.Corruption. Generosity Dystopia.Residual
## 1      0.41978 0.29678      2.51738
## 2      0.14145 0.43630      2.70201
## 3      0.48357 0.34139      2.49204
## 4      0.36503 0.34699      2.46531
## 5      0.32957 0.45811      2.45176
## 6      0.41372 0.23351      2.61955
```

```
df2 <- read.csv("2016.csv", header=TRUE)
head(df2);
```

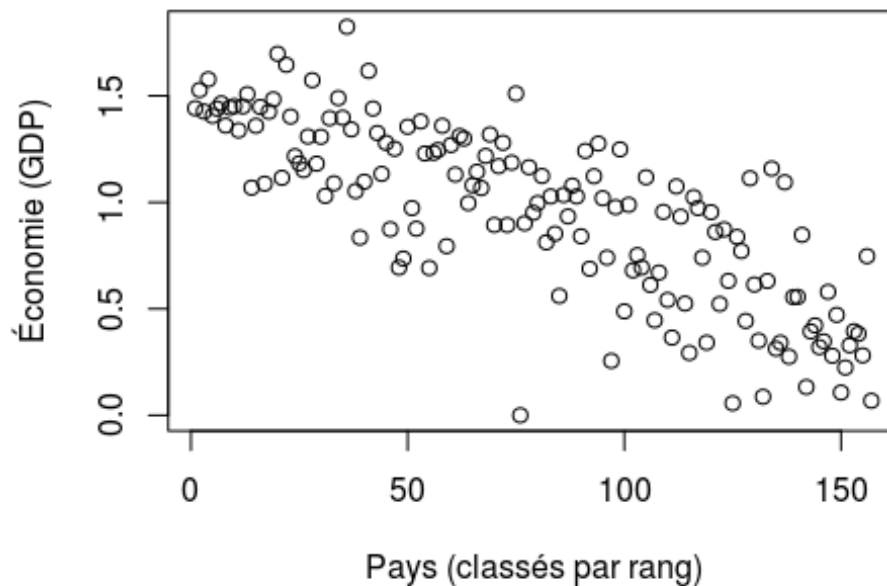
```
##      Country      Region Happiness.Rank Happiness.Score
## 1    Denmark Western Europe      1      7.526
## 2 Switzerland Western Europe      2      7.509
## 3    Iceland Western Europe      3      7.501
## 4    Norway Western Europe      4      7.498
## 5    Finland Western Europe      5      7.413
## 6    Canada North America      6      7.404
## Lower.Confidence.Interval Upper.Confidence.Interval
## 1      7.460      7.592
## 2      7.428      7.590
## 3      7.333      7.669
## 4      7.421      7.575
## 5      7.351      7.475
## 6      7.335      7.473
## Economy..GDP.per.Capita. Family Health..Life.Expectancy. Freedom
## 1      1.44178 1.16374      0.79504 0.57941
## 2      1.52733 1.14524      0.86303 0.58557
## 3      1.42666 1.18326      0.86733 0.56624
## 4      1.57744 1.12690      0.79579 0.59609
## 5      1.40598 1.13464      0.81091 0.57104
## 6      1.44015 1.09610      0.82760 0.57370
## Trust..Government.Corruption. Generosity Dystopia.Residual
## 1      0.44453 0.36171      2.73939
## 2      0.41203 0.28083      2.69463
## 3      0.14975 0.47678      2.83137
## 4      0.35776 0.37895      2.66465
## 5      0.41004 0.25492      2.82596
## 6      0.31329 0.44834      2.70485
```

Nous allons réaliser des plot pour étudier la relation entre les pays et les différents facteurs.

```
plot(df$Economy..GDP.per.Capita., ylab="Économie (GDP)", xlab="Pays  
(classés par rang)");
```



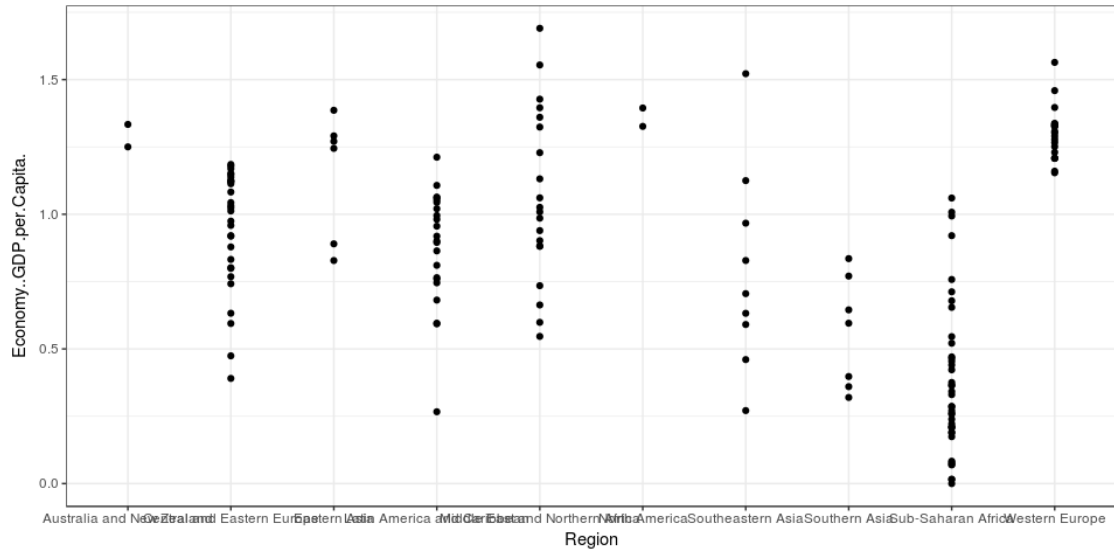
```
plot(df2$Economy..GDP.per.Capita., ylab="Économie (GDP)", xlab="Pays  
(classés par rang)");
```



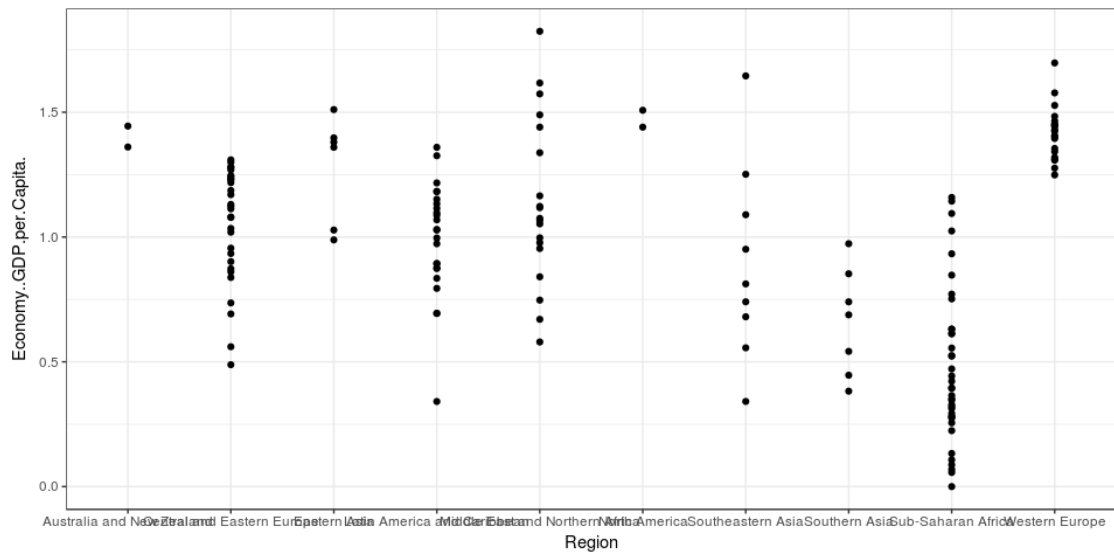
Ces graphiques représentent le PIB en fonction des pays en 2015 et en 2016. On voit que les pays les mieux classés sont ceux dont le PIB est élevé. Il y a donc une forte corrélation entre le PIB et le bonheur.

On fait l'hypothèse que les différentes régions font partie d'un intervalle restreint dans le classement, par exemple l'Europe de l'Ouest occupe le haut du classement. On peut vérifier ceci au niveau des régions à travers un ggplot.

```
library(ggplot2);
df %>%
  group_by(Region) %>%
  ggplot(aes(x=Region, y=Economy..GDP.per.Capita.)) +
    geom_point() + ylim(0,NA) + theme_bw();
```



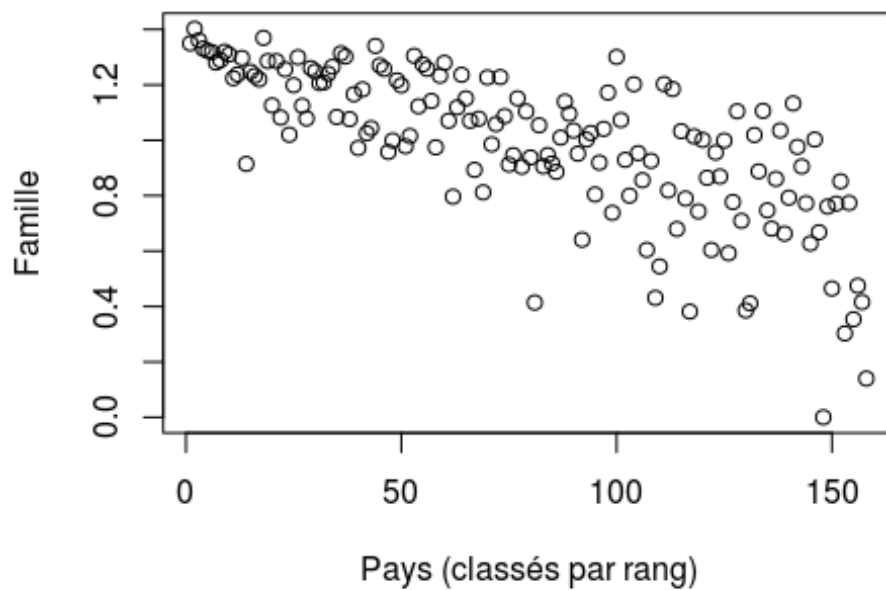
```
library(ggplot2);
df2 %>%
  group_by(Region) %>%
  ggplot(aes(x=Region, y=Economy..GDP.per.Capita.)) +
    geom_point() + ylim(0,NA) + theme_bw();
```



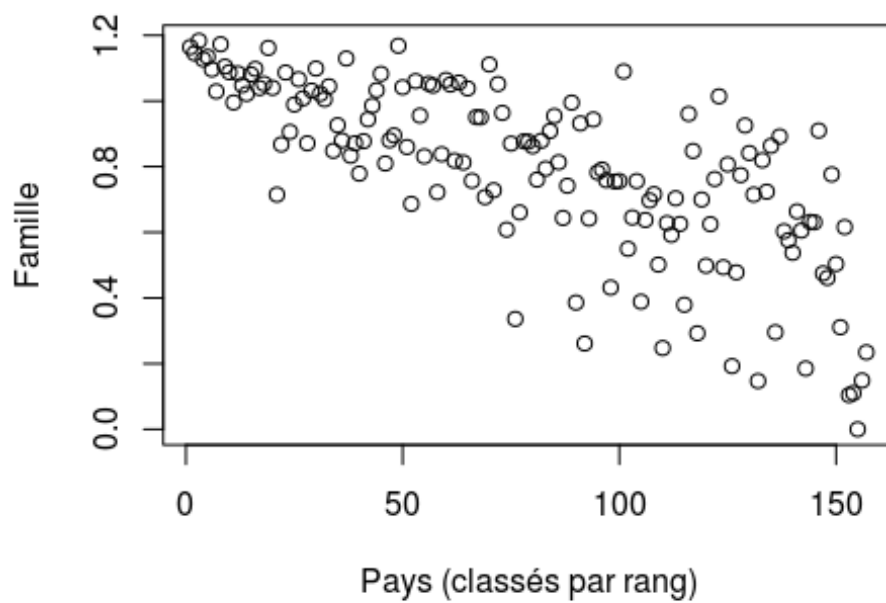
On voit ainsi que le PIB moyen en Europe de l'Ouest est assez élevé par rapport aux autres régions. Cela confirme que PIB et bonheur sont liés.

Analysons maintenant le facteur famille :

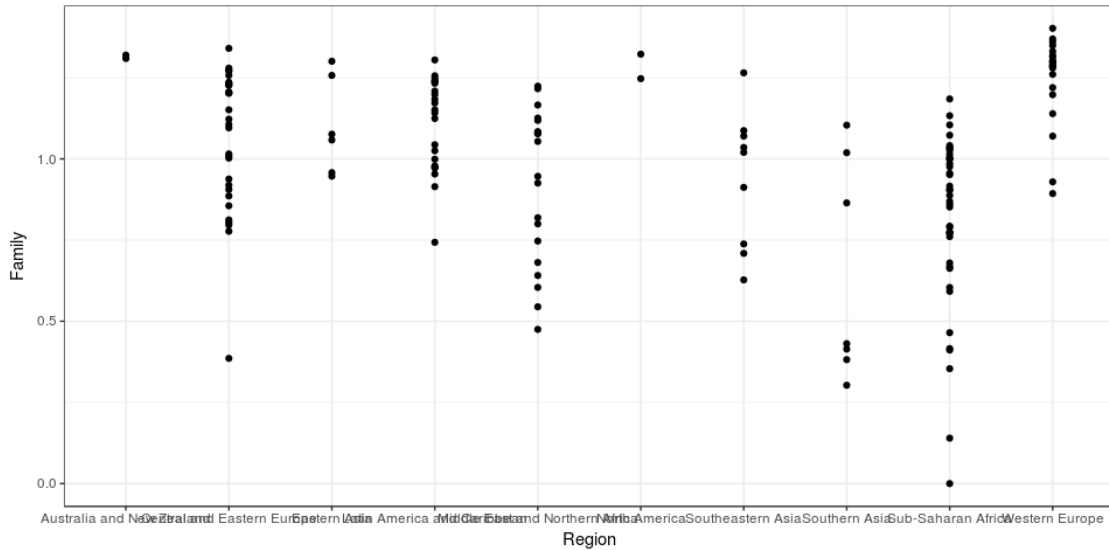
```
plot(df$Family, ylab="Famille", xlab="Pays (classés par rang)");
```



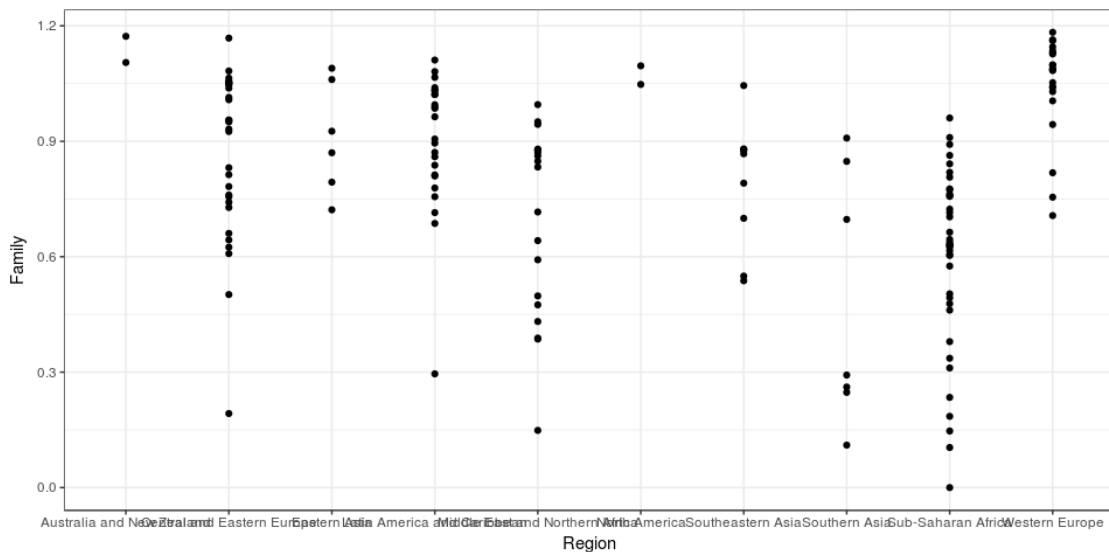
```
plot(df2$Family, ylab="Famille", xlab="Pays (classés par rang)");
```



```
library(ggplot2);
df %>%
  group_by(Region) %>%
  ggplot(aes(x=Region, y=Family)) +
    geom_point() + ylim(0,NA) + theme_bw();
```



```
library(ggplot2);
df2 %>%
  group_by(Region) %>%
  ggplot(aes(x=Region, y=Family)) +
    geom_point() + ylim(0,NA) + theme_bw();
```

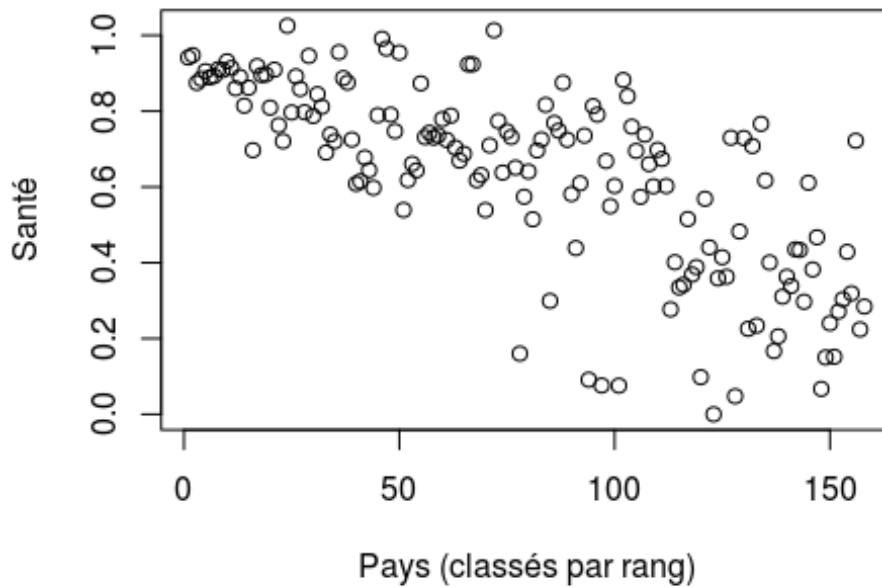


Il nous est difficile de conclure dessus étant donné que nous n'avons pas trouvé à quoi correspondait la mesure réalisée pour ce facteur. Cependant, les graphiques semblent montrer une corrélation entre ce

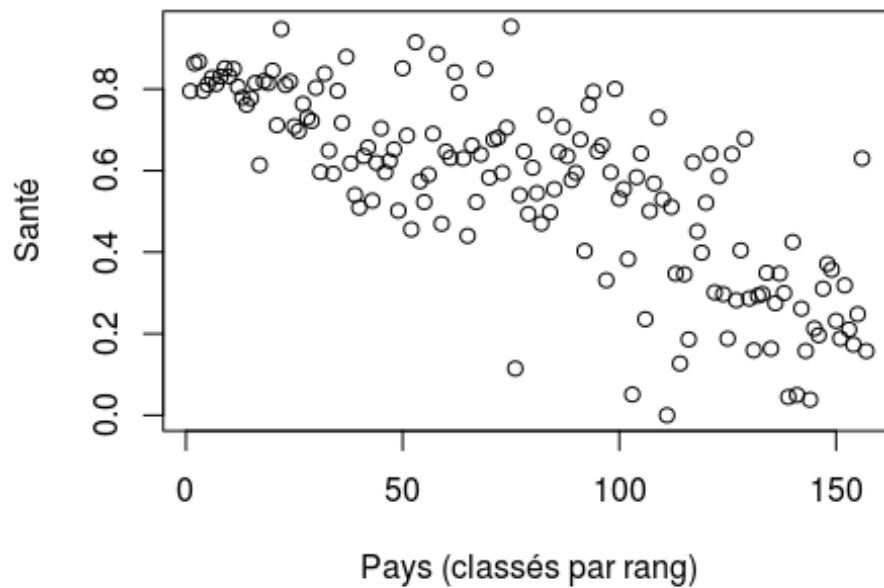
facteur et le bonheur (les pays et régions où ce facteur est le plus élevé occupent en général un rang plus important).

Nous pouvons maintenant nous intéresser au facteur Santé :

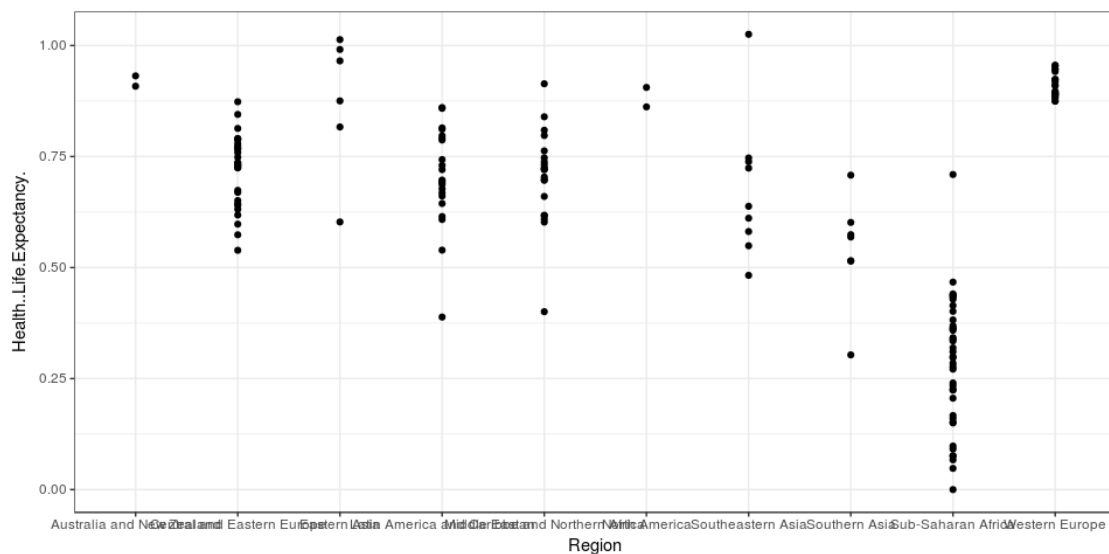
```
plot(df$Health..Life.Expectancy., ylab="Santé", xlab="Pays (classés par rang)");
```



```
plot(df2$Health..Life.Expectancy., ylab="Santé", xlab="Pays (classés par rang)");
```

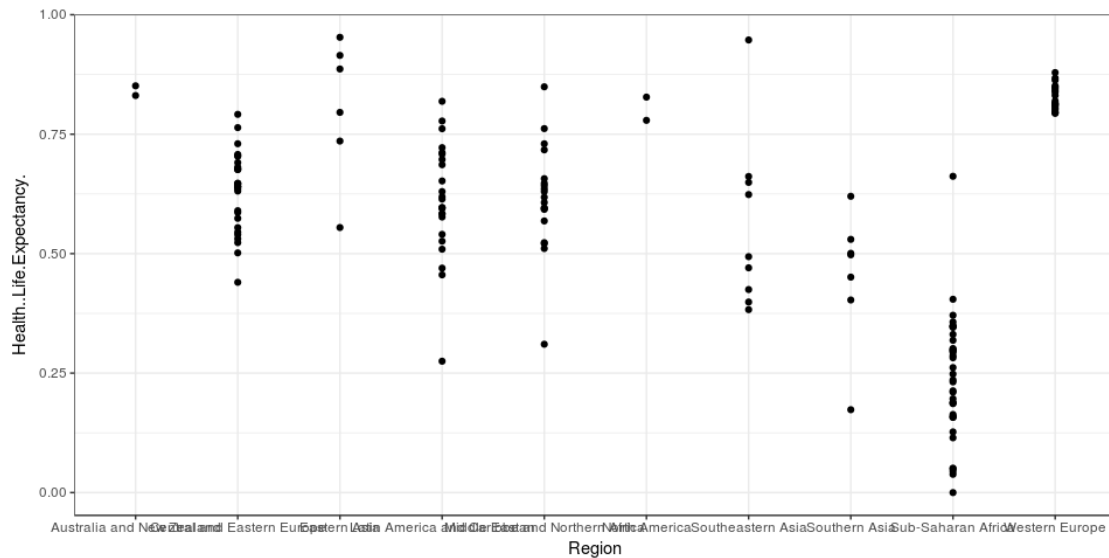



```
library(ggplot2);
df %>%
  group_by(Region) %>%
  ggplot(aes(x=Region, y=Health..Life.Expectancy.)) +
    geom_point() + ylim(0,NA) + theme_bw();
```



```
library(ggplot2);
df2 %>%
  group_by(Region) %>%
```

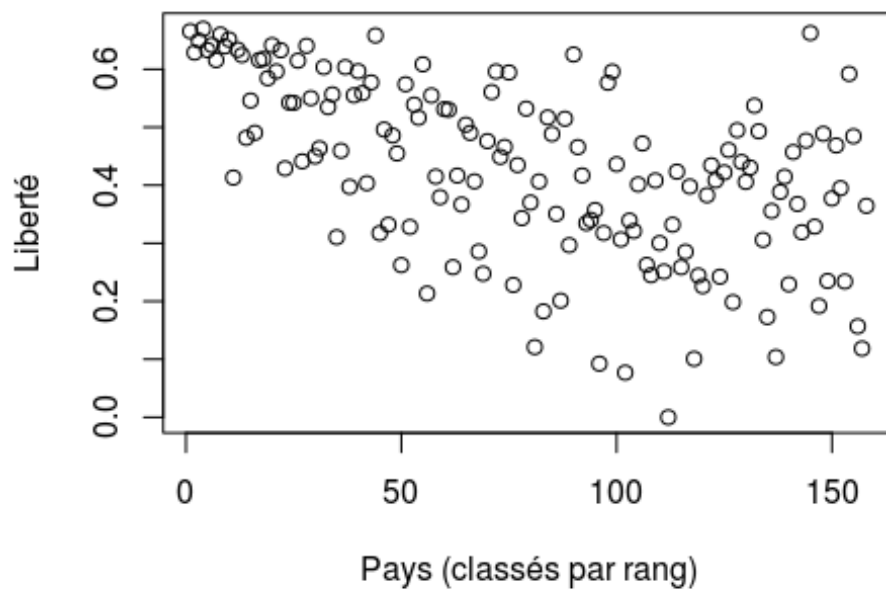
```
ggplot(aes(x=Region, y=Health..Life.Expectancy.)) +  
  geom_point() + ylim(0,NA) + theme_bw();
```



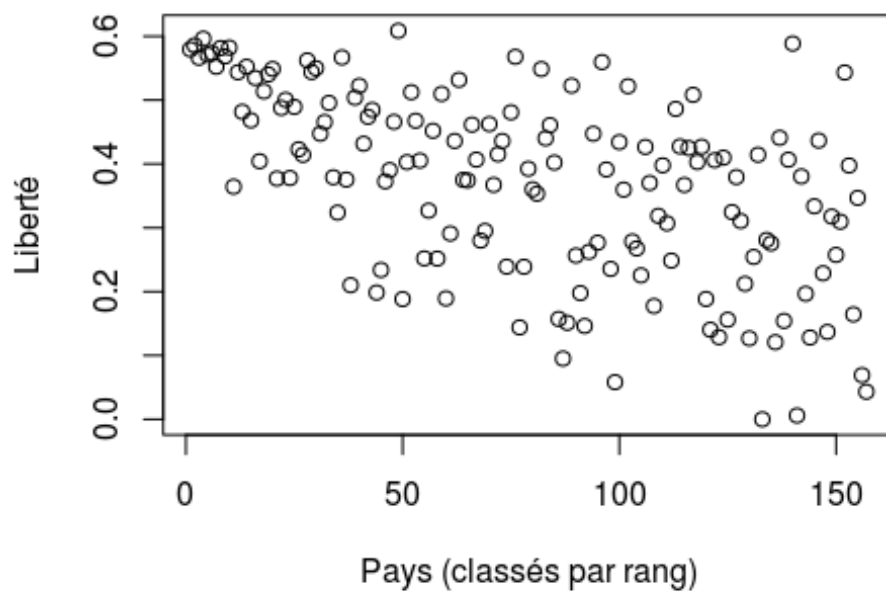
Tout comme les facteurs précédents, la santé joue un rôle dans le classement des pays. On observe à nouveau que les pays dont le rang est élevé sont satisfaits par rapport à la politique de santé dans leur pays.

Le facteur auquel nous allons maintenant nous intéresser est celui de la liberté.

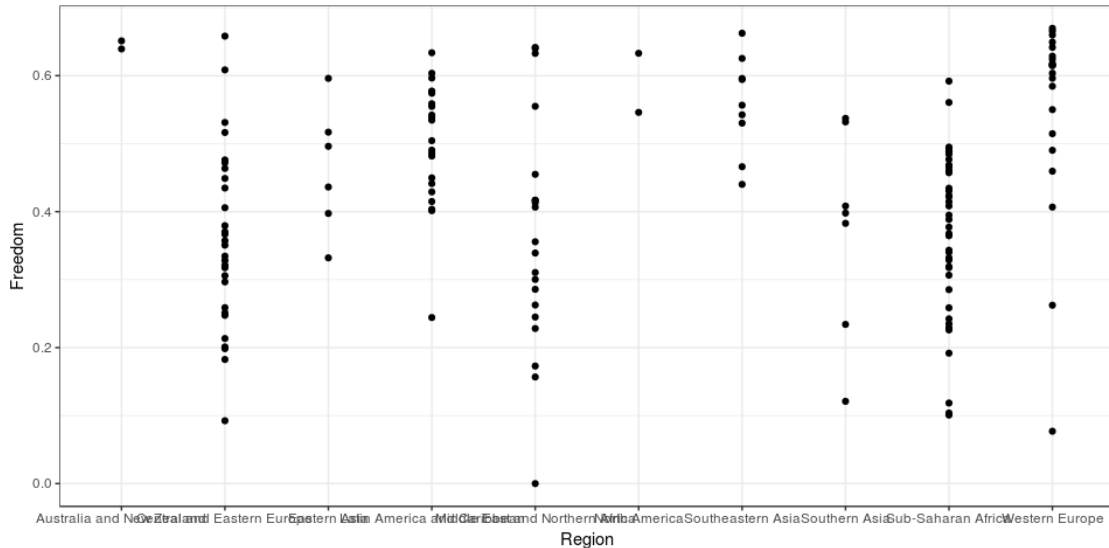
```
plot(df$Freedom, ylab="Liberté", xlab="Pays (classés par rang)");
```



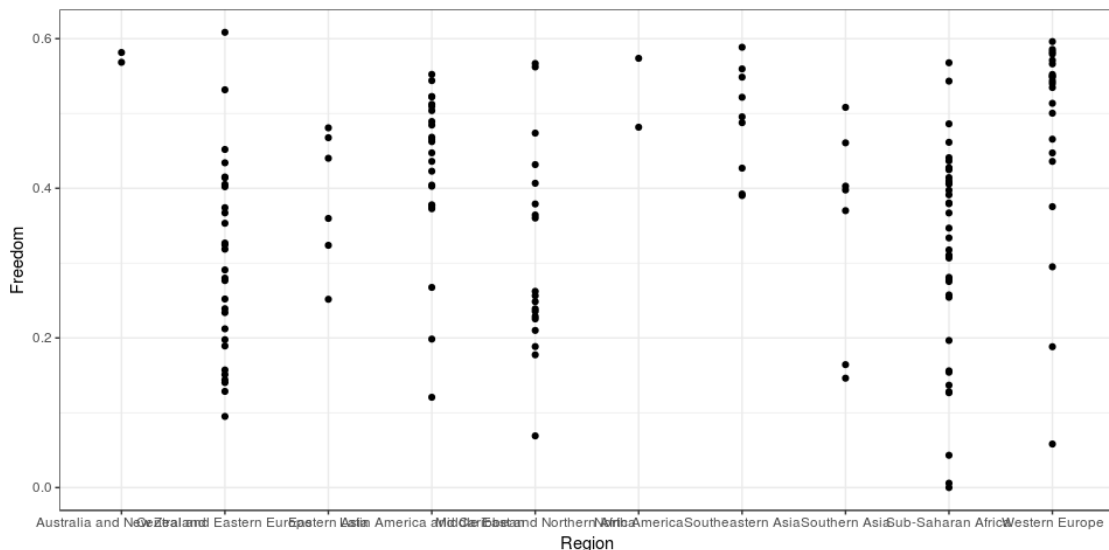
```
plot(df2$Freedom, ylab="Liberté", xlab="Pays (classés par rang)");
```



```
library(ggplot2);
df %>%
  group_by(Region) %>%
  ggplot(aes(x=Region, y=Freedom)) +
    geom_point() + ylim(0,NA) + theme_bw();
```



```
library(ggplot2);
df2 %>%
  group_by(Region) %>%
  ggplot(aes(x=Region, y=Freedom)) +
    geom_point() + ylim(0,NA) + theme_bw();
```

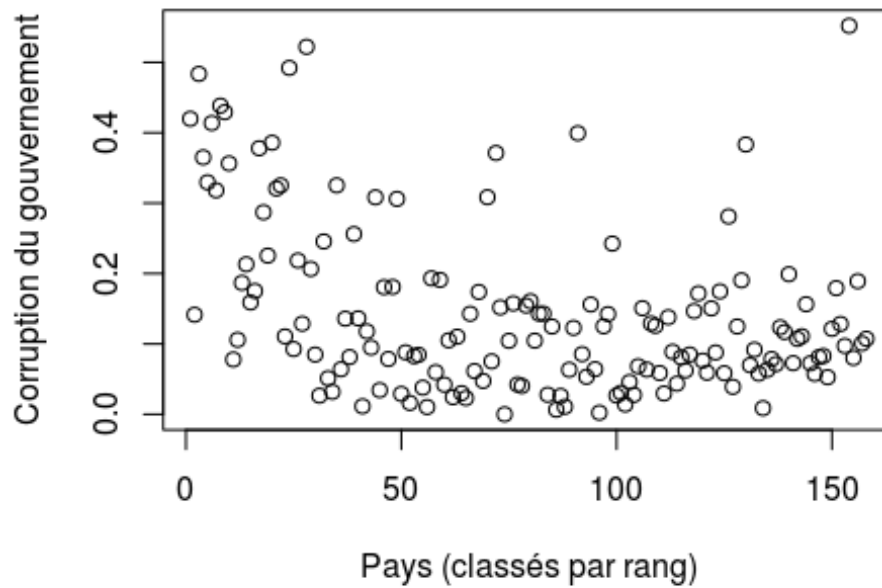


Pour ce facteur, il semble y avoir une légère influence du sentiment de liberté sur le bonheur puisque les pays en tête du classement se sentent libres. Cependant, la courbe ne semble pas suivre une régression linéaire et les valeurs sont trop dispersées pour conclure

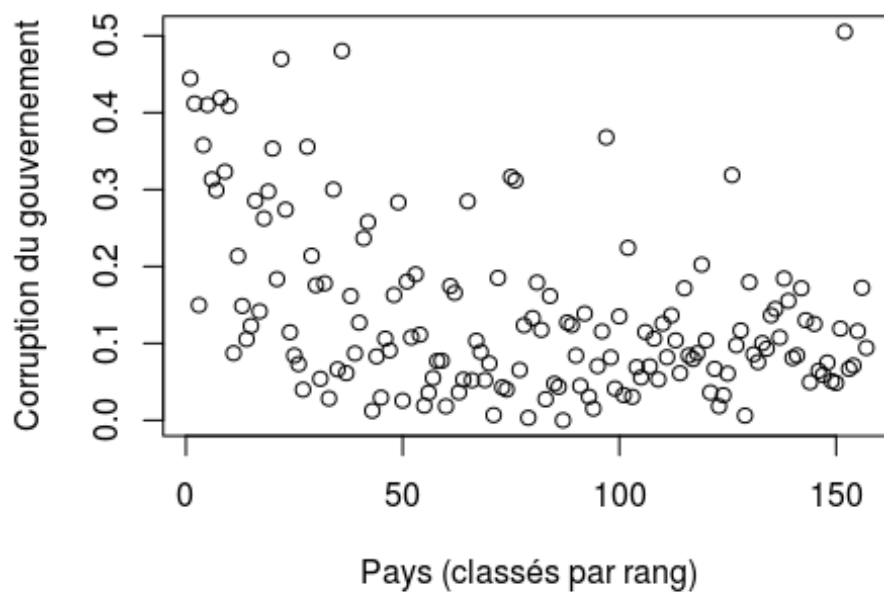
que ce facteur est vraiment important pour juger du bonheur d'un pays.

Le facteur suivant est celui de la corruption dans le gouvernement.

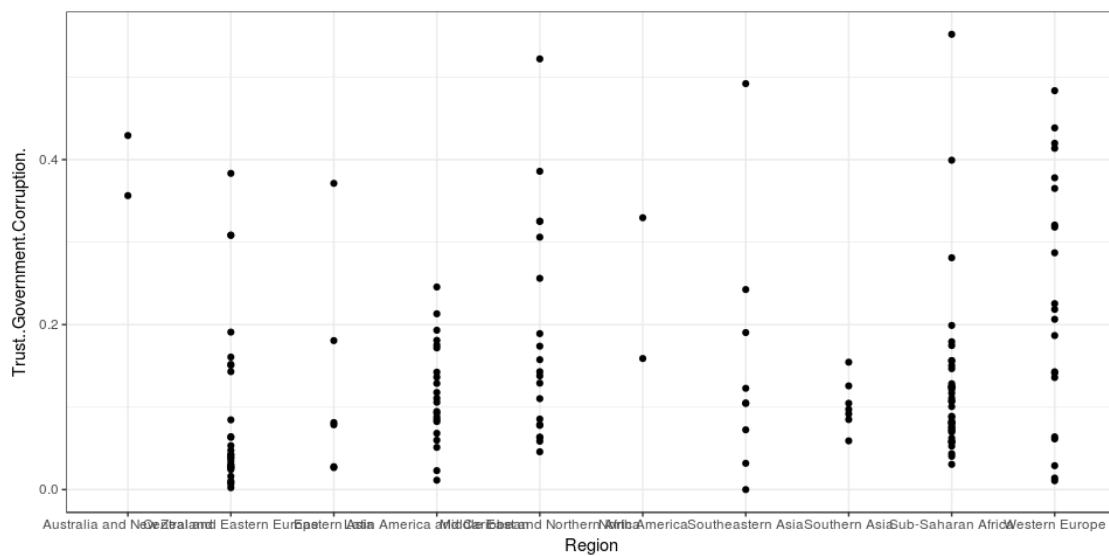
```
plot(df$Trust..Government.Corruption., ylab="Corruption du gouvernement",  
xlab="Pays (classés par rang)");
```



```
plot(df2$Trust..Government.Corruption., ylab="Corruption du gouvernement",  
xlab="Pays (classés par rang)");
```

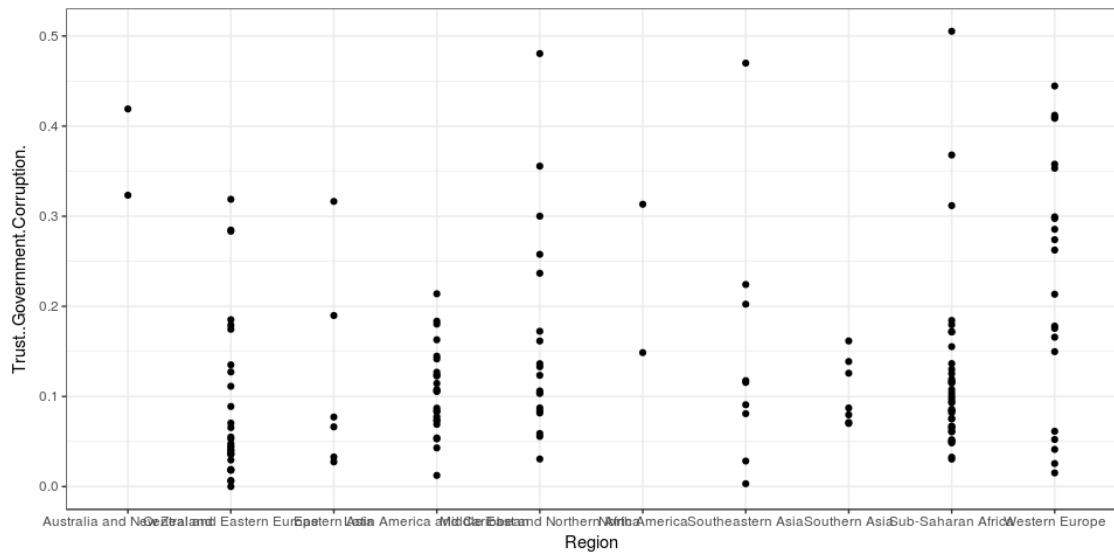


```
library(ggplot2);
df %>%
  group_by(Region) %>%
  ggplot(aes(x=Region, y=Trust..Government.Corruption.)) +
    geom_point() + ylim(0,NA) + theme_bw();
```



```
library(ggplot2);
df2 %>%
  group_by(Region) %>%
```

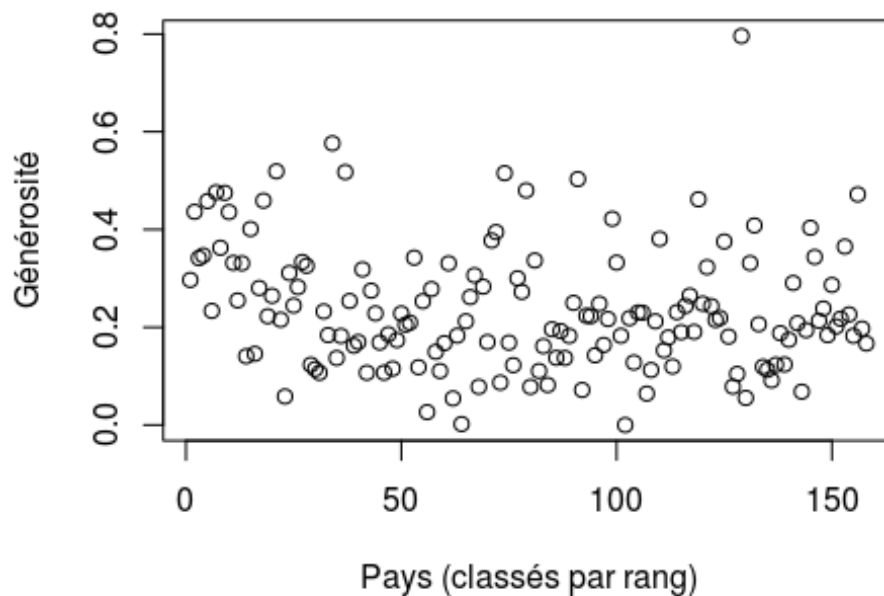
```
ggplot(aes(x=Region, y=Trust..Government.Corruption.)) +  
  geom_point() + ylim(0,NA) + theme_bw();
```



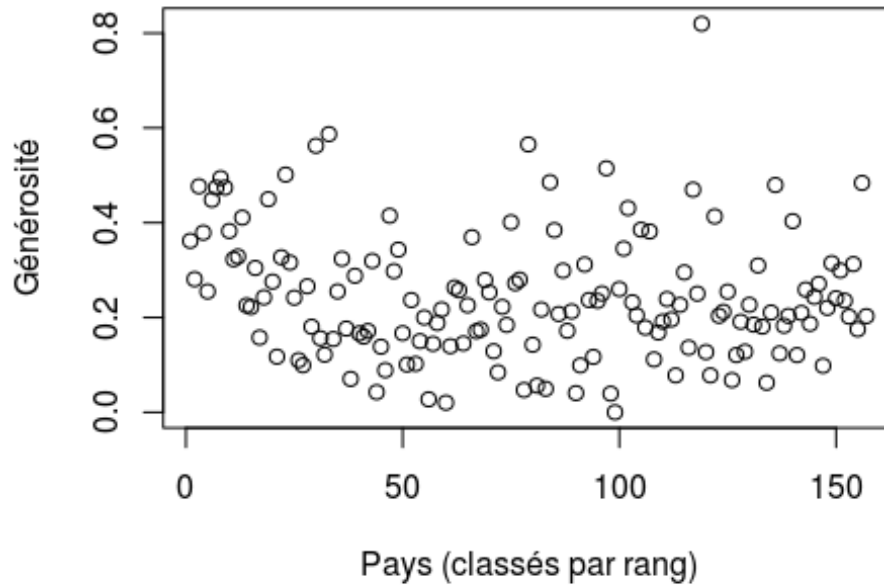
On observe des valeurs plutôt basses et assez homogènes. Il ne semble pas y avoir de corrélation entre ce facteur et le bonheur.

Enfin, analysons le facteur de la générosité (est-ce que les gens ont donné à une oeuvre de charité au cours de l'année précédente ?) :

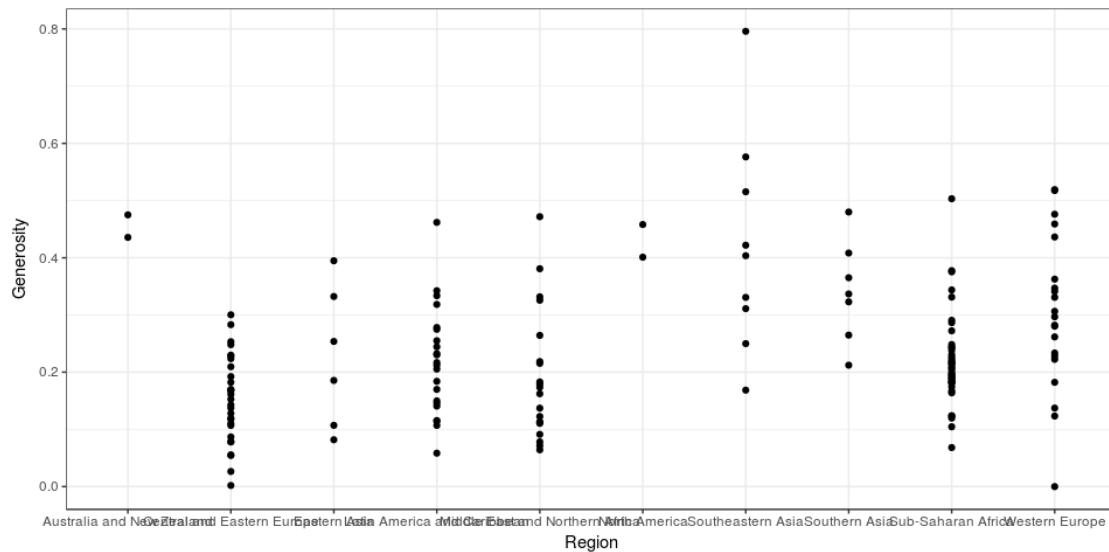
```
plot(df$Generosity, ylab="Générosité", xlab="Pays (classés par rang)");
```



```
plot(df2$Generosity, ylab="Générosité", xlab="Pays (classés par rang)");
```



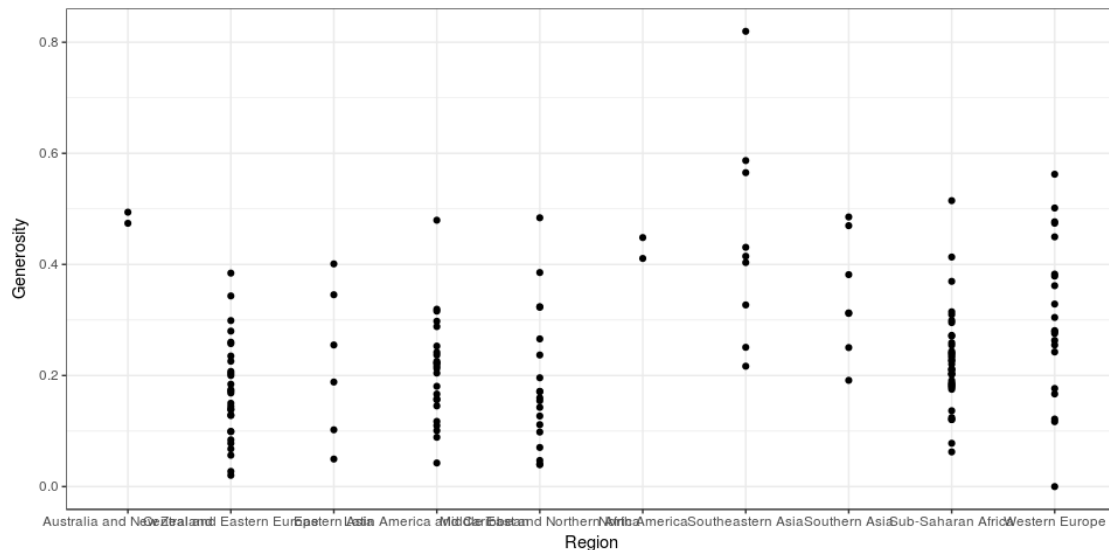
```
library(ggplot2);
df %>%
  group_by(Region) %>%
  ggplot(aes(x=Region, y=Generosity)) +
    geom_point() + ylim(0,NA) + theme_bw();
```



```
library(ggplot2);
df2 %>%
```



```
group_by(Region) %>%
ggplot(aes(x=Region, y=Generosity)) +
  geom_point() + ylim(0,NA) + theme_bw();
```



L'homogénéité des mesures ne permet pas de conclure à une quelconque corrélation entre ce facteur et le bonheur.

Nous chargeons à nouveau ce fichier pour avoir des dataframes avec des noms qui ont plus de sens pour nous.

```
df2015 <- read.csv("2015.csv", header=TRUE)
head(df2015)
```

##	Country	Region	Happiness.Rank	Happiness.Score	Standard.Error
## 1	Switzerland	Western Europe	1	7.587	0.03411
## 2	Iceland	Western Europe	2	7.561	0.04884
## 3	Denmark	Western Europe	3	7.527	0.03328
## 4	Norway	Western Europe	4	7.522	0.03880
## 5	Canada	North America	5	7.427	0.03553
## 6	Finland	Western Europe	6	7.406	0.03140
##	Economy..GDP.per.Capita.	Family Health..Life.Expectancy.	Freedom		
## 1	1.39651	1.34951	0.94143	0.66557	
## 2	1.30232	1.40223	0.94784	0.62877	
## 3	1.32548	1.36058	0.87464	0.64938	
## 4	1.45900	1.33095	0.88521	0.66973	
## 5	1.32629	1.32261	0.90563	0.63297	
## 6	1.29025	1.31826	0.88911	0.64169	
##	Trust..Government.	Corruption.	Generosity	Dystopia.Residual	
## 1	0.41978	0.29678	2.51738		
## 2	0.14145	0.43630	2.70201		
## 3	0.48357	0.34139	2.49204		
## 4	0.36503	0.34699	2.46531		
## 5	0.32957	0.45811	2.45176		
## 6	0.41372	0.23351	2.61955		

```
df2016 <- read.csv("2016.csv", header=TRUE)
head(df2016)
```

```
##      Country      Region Happiness.Rank Happiness.Score
## 1  Denmark Western Europe           1         7.526
## 2 Switzerland Western Europe           2         7.509
## 3  Iceland Western Europe           3         7.501
## 4   Norway Western Europe           4         7.498
## 5   Finland Western Europe           5         7.413
## 6    Canada North America           6         7.404
## Lower.Confidence.Interval Upper.Confidence.Interval
## 1              7.460              7.592
## 2              7.428              7.590
## 3              7.333              7.669
## 4              7.421              7.575
## 5              7.351              7.475
## 6              7.335              7.473
## Economy..GDP.per.Capita. Family Health..Life.Expectancy. Freedom
## 1          1.44178 1.16374          0.79504 0.57941
## 2          1.52733 1.14524          0.86303 0.58557
## 3          1.42666 1.18326          0.86733 0.56624
## 4          1.57744 1.12690          0.79579 0.59609
## 5          1.40598 1.13464          0.81091 0.57104
## 6          1.44015 1.09610          0.82760 0.57370
## Trust..Government.Corruption. Generosity Dystopia.Residual
## 1          0.44453 0.36171          2.73939
## 2          0.41203 0.28083          2.69463
## 3          0.14975 0.47678          2.83137
## 4          0.35776 0.37895          2.66465
## 5          0.41004 0.25492          2.82596
## 6          0.31329 0.44834          2.70485
```

Nous voulons maintenant avoir un dataframe qui regrouperait toutes nos données. Pour cela, il nous faut retirer les colonnes qui ne correspondent pas, à savoir Standard Error dans df2015 et Lower/Upper Confidence Interval dans df2016. De plus, nous avons du retirer le pays occupant le 158ème rang en 2015 car il n'y en avait que 157 en 2016. Enfin, nous avons rajouté une colonne Année afin de distinguer les données de 2015 et celles de 2016.

```
df15temp <- df2015 %>% select(-Standard.Error) %>% filter(Happiness.Rank
!= 158) %>% mutate(Annee=2015);
df16temp <- df2016 %>% select(-Lower.Confidence.Interval,-
Upper.Confidence.Interval) %>% mutate(Annee=2016);
names(df15temp);
```

```
## [1] "Country"      "Region"
## [3] "Happiness.Rank" "Happiness.Score"
## [5] "Economy..GDP.per.Capita." "Family"
## [7] "Health..Life.Expectancy." "Freedom"
```

```
## [9] "Trust..Government.Corruption." "Generosity"
## [11] "Dystopia.Residual"              "Annee"

names(df16temp);

## [1] "Country"              "Region"
## [3] "Happiness.Rank"       "Happiness.Score"
## [5] "Economy..GDP.per.Capita." "Family"
## [7] "Health..Life.Expectancy." "Freedom"
## [9] "Trust..Government.Corruption." "Generosity"
## [11] "Dystopia.Residual"              "Annee"

nrow(df15temp);

## [1] 157

nrow(df16temp);

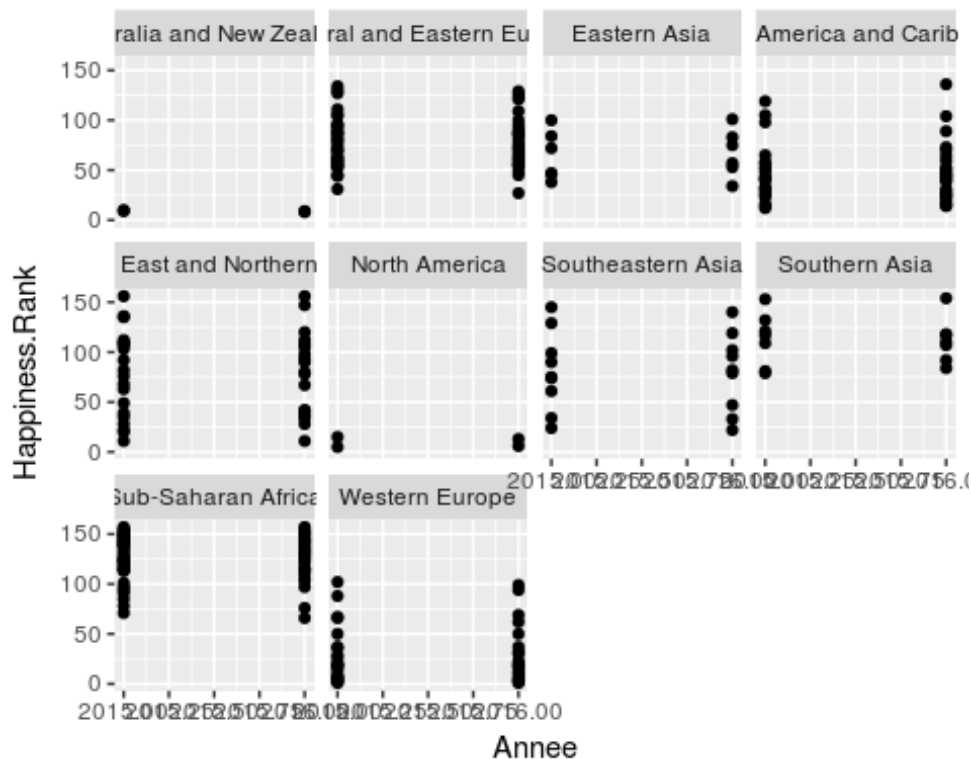
## [1] 157

d <- rbind(df15temp,df16temp);
nrow(d);

## [1] 314
```

Nous avons vu précédemment que les pays provenant de même région semblaient présenter le même profil, c'est-à-dire que les valeurs des facteurs pouvaient être restreints à un intervalle. C'est pourquoi nous avons décidé de travailler par la suite en fonction des régions.

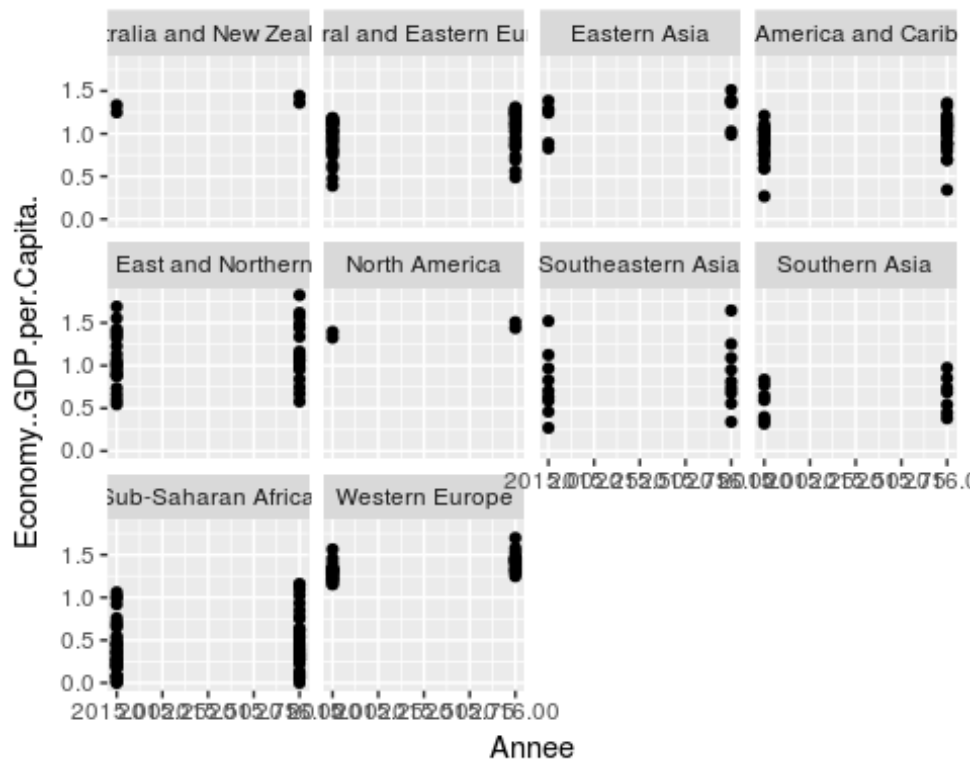
```
d %>% ggplot(aes(x=Annee,y=Happiness.Rank)) +
  geom_point() +
  facet_wrap(~Region);
```



Nous avons représenté les évolutions des rangs par région entre 2015 et 2016. Globalement, les classements sont maintenus d'une année à l'autre, mais il y a eu des changements, telles que des baisses en Central and Eastern Europe, Southeastern Asia, Western Europe, ainsi qu'une hausse en Latin America and Caribbean.

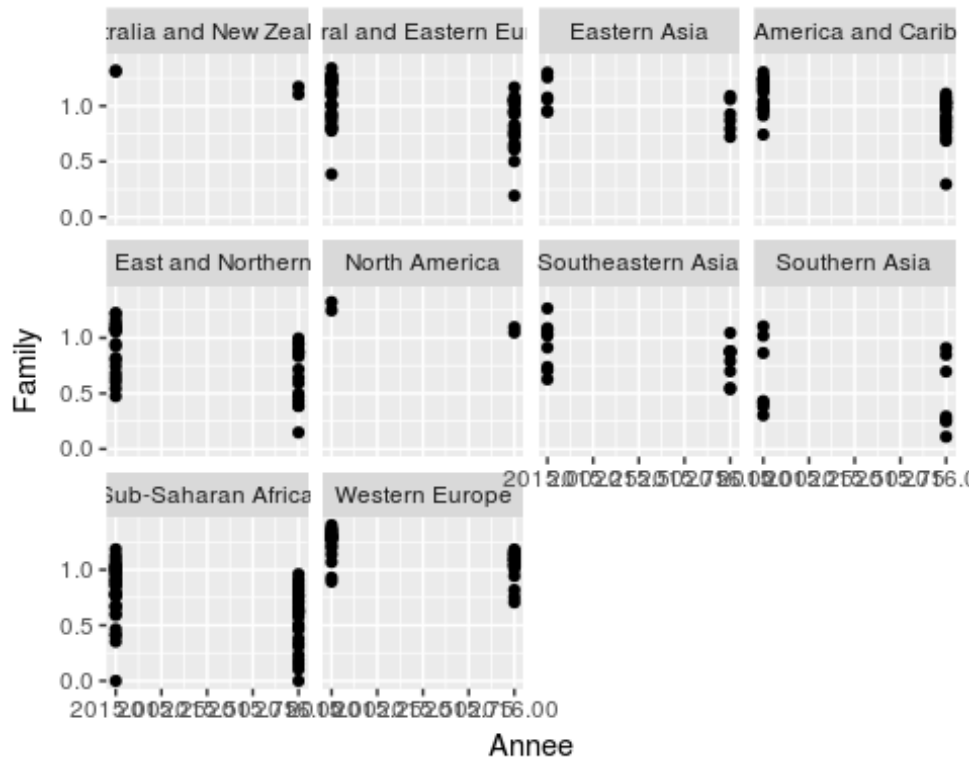
Nous allons nous focaliser sur les facteurs qui d'après nos résultats précédents ont un lien avec le bonheur.

```
d %>% ggplot(aes(x=Annee,y=Economy..GDP.per.Capita.)) +  
  geom_point() +  
  facet_wrap(~Region);
```



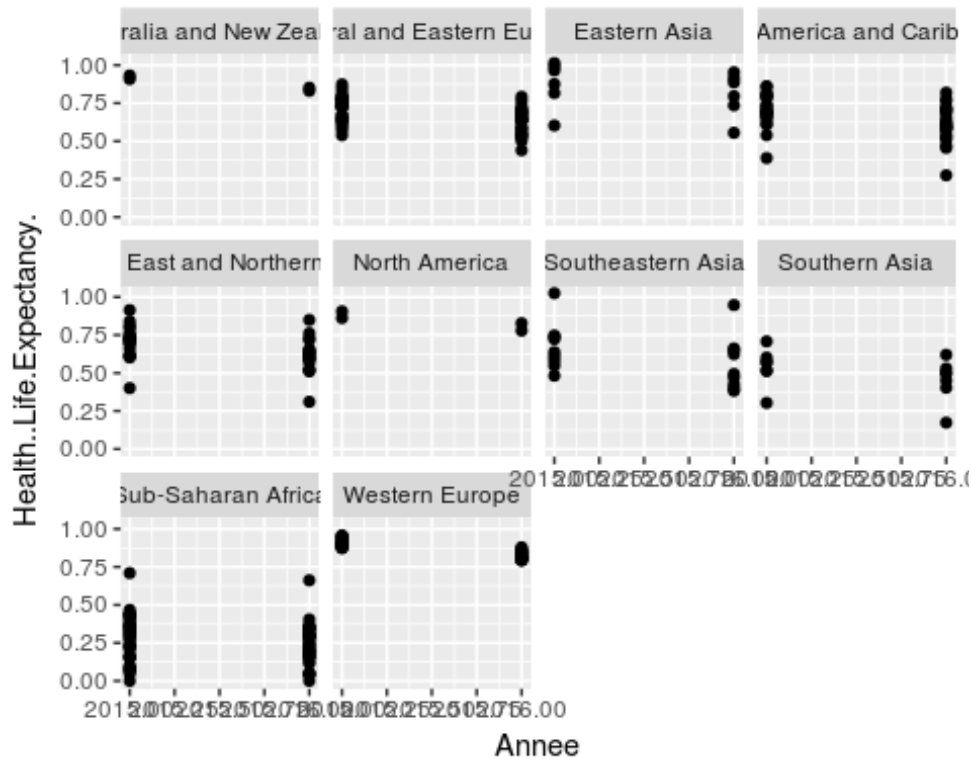
Par rapport aux régions qu'on a citées plus haut, il y a eu peu d'évolution en ce qui concerne le PIB, donc on ne le considérera pas comme le facteur principal de l'évolution du bonheur en 2016.

```
d %>% ggplot(aes(x=Annee,y=Family)) +  
  geom_point() +  
  facet_wrap(~Region);
```



Il y a eu des baisses dans les régions concernées ce qui semble confirmer notre hypothèse de départ, à savoir que le facteur famille a une influence sur le bonheur d'une année à une autre. Nous ne savons pas comment a été calculé ce facteur ni ce qu'il représente mais lorsque le score de celui-ci diminue, le classement des régions telles que Southeastern Asia diminue.

```
d %>% ggplot(aes(x=Annee,y=Health..Life.Expectancy.)) +  
  geom_point() +  
  facet_wrap(~Region);
```



Ainsi, le facteur Santé influence de la même manière le rang moyen de la région. On observe encore une fois une diminution de la valeur de ce facteur en Southeastern Asia.

Conclusion

Pour conclure, nous ne pouvons pas nous limiter à un seul et unique facteur pour déterminer celui qui influencera le plus l'évolution du bonheur d'un pays d'une année à l'autre. Cependant, nous pouvons restreindre les facteurs influençant le plus cette dernière aux facteurs Famille et Santé.