Projet R

2024-10-31

Introduction

Ce document analyse la relation entre l'espérance de vie (`lifeExp`) et le PIB par habitant (`gdpPercap`) pour différentes années (1957, 1982 et 2007) en utilisant le dataset Gapminder.

Chargement des bibliothèques et des données

```
library(gapminder)
library(e1071)
library(ggplot2)
library(gridExtra)
library(corrplot)
```

```
## corrplot 0.95 loaded
```

```
# Chargement des données
data(gapminder)
```

Filtrage des données pour les années d'intérêt

Nous filtrons les données pour trois années spécifiques : 1957, 1982 et 2007.

```
gapminderP = subset(gapminder, year == c(1957, 1982, 2007))
gapminder1957 = subset(gapminder, year == 1957)
gapminder1982 = subset(gapminder, year == 1982)
gapminder2007 = subset(gapminder, year == 2007)
```

Analyse univariée de l'espérance de vie (`lifeExp`)

Calcul de l'IQR, de la kurtosis et de la skewness pour `lifeExp` en 1957, 1982 et 2007.

Calcul des statistiques de dispersion

```
IQR57 = IQR(gapminder1957$lifeExp)
IQR82 = IQR(gapminder1982$lifeExp)
IQR07 = IQR(gapminder2007$lifeExp)
```

Calcul de la kurtosis et de la skewness pour `lifeExp`

```
kurtosis_57 = kurtosis(gapminder1957$lifeExp)
kurtosis_82 = kurtosis(gapminder1982$lifeExp)
kurtosis_07 = kurtosis(gapminder2007$lifeExp)

skewness_57 = skewness(gapminder1957$lifeExp)
skewness_82 = skewness(gapminder1982$lifeExp)
skewness_07 = skewness(gapminder2007$lifeExp)
```

Tableau récapitulatif pour `lifeExp`

```
summary_1957 = as.numeric(summary(gapminder1957$lifeExp))
summary_1982 = as.numeric(summary(gapminder1982$lifeExp))
summary_2007 = as.numeric(summary(gapminder2007$lifeExp))

summary_57_kurt_sk_IQ = c(summary_1957, kurtosis_57, skewness_57, IQR57)
summary_82_kurt_sk_IQ = c(summary_1982, kurtosis_82, skewness_82, IQR82)
summary_07_kurt_sk_IQ = c(summary_2007, kurtosis_07, skewness_07, IQR07)

summary_table = data.frame(
  statistic = c("Min", "1st Qu", "Median", "Mean", "3rd Qu", "Max", "Kurtosis", "Skewness", "IQR"),
  "1957" = summary_57_kurt_sk_IQ,
  "1982" = summary_82_kurt_sk_IQ,
  "2007" = summary_07_kurt_sk_IQ
)

colnames(summary_table) = gsub("^X", "", colnames(summary_table))

summary_table[,c("1957","1982","2007")] <- round(summary_table[, c("1957","1982","2007")], digits = 2)
summary_table
```

```
##   statistic  1957  1982  2007
## 1       Min 30.33 38.44 39.61
## 2    1st Qu 41.25 52.94 57.16
## 3    Median 48.36 62.44 71.94
## 4      Mean 51.51 61.53 67.01
## 5    3rd Qu 63.04 70.92 76.41
## 6       Max 73.47 77.11 82.60
## 7  Kurtosis -1.31 -1.10 -0.87
## 8  Skewness  0.24 -0.34 -0.67
## 9       IQR 21.79 17.98 19.25
```

Analyse univariée du PIB par habitant (`gdpPercap`)

Calcul des statistiques pour `gdpPercap`

```r
summary_G1957 = as.numeric(summary(gapminder1957$gdpPercap))
summary_G1982 = as.numeric(summary(gapminder1982$gdpPercap))
summary_G2007 = as.numeric(summary(gapminder2007$gdpPercap))

IQR_G57 = IQR(gapminder1957$gdpPercap)
IQR_G82 = IQR(gapminder1982$gdpPercap)
IQR_G07 = IQR(gapminder2007$gdpPercap)

kurtosis_G57 = kurtosis(gapminder1957$gdpPercap)
kurtosis_G82 = kurtosis(gapminder1982$gdpPercap)
kurtosis_G07 = kurtosis(gapminder2007$gdpPercap)

skewness_G57 = skewness(gapminder1957$gdpPercap)
skewness_G82 = skewness(gapminder1982$gdpPercap)
skewness_G07 = skewness(gapminder2007$gdpPercap)

summary_G57_kurt_sk_IQ = c(summary_G1957, kurtosis_G57, skewness_G57, IQR_G57)
summary_G82_kurt_sk_IQ = c(summary_G1982, kurtosis_G82, skewness_G82, IQR_G82)
summary_G07_kurt_sk_IQ = c(summary_G2007, kurtosis_G07, skewness_G07, IQR_G07)

summary_table1 = data.frame(
  statistic = c("Min", "1st Qu", "Median", "Mean", "3rd Qu", "Max", "Kurtosis", "Skewness", "IQR"),
  "1957" = summary_G57_kurt_sk_IQ,
  "1982" = summary_G82_kurt_sk_IQ,
  "2007" = summary_G07_kurt_sk_IQ
)

colnames(summary_table1) = gsub("^X", "", colnames(summary_table1))

summary_table1[,c("1957","1982","2007")] <- round(summary_table1[, c("1957","1982","2007")], digits = 2)
summary_table1
```

```
##   statistic      1957     1982     2007
## 1       Min    336.00   424.00   277.55
## 2    1st Qu    930.54  1363.34  1624.84
## 3    Median   2173.22  4216.23  6124.37
## 4      Mean   4299.41  7518.90 11680.07
## 5    3rd Qu   4876.36 12347.95 18008.84
## 6       Max 113523.13 33693.18 49357.19
## 7  Kurtosis    102.69     0.58     0.25
## 8  Skewness      9.58     1.20     1.20
## 9       IQR   3945.82 10984.61 16383.99
```

Analyse bivariée : Relation entre `lifeExp` et `gdpPercap`

Covariance entre `lifeExp` et `gdpPercap`

```r
covariance_1957 = cov(gapminder1957$gdpPercap, gapminder1957$lifeExp)
covariance_1982 = cov(gapminder1982$gdpPercap, gapminder1982$lifeExp)
covariance_2007 = cov(gapminder2007$gdpPercap, gapminder2007$lifeExp)
```

Corrélation de Pearson entre `lifeExp` et `gdpPercap`

```
cor_1957 = cor(gapminder1957$gdpPercap, gapminder1957$lifeExp)
cor_1982 = cor(gapminder1982$gdpPercap, gapminder1982$lifeExp)
cor_2007 = cor(gapminder2007$gdpPercap, gapminder2007$lifeExp)
```

Interprétation des coefficients de corrélation :

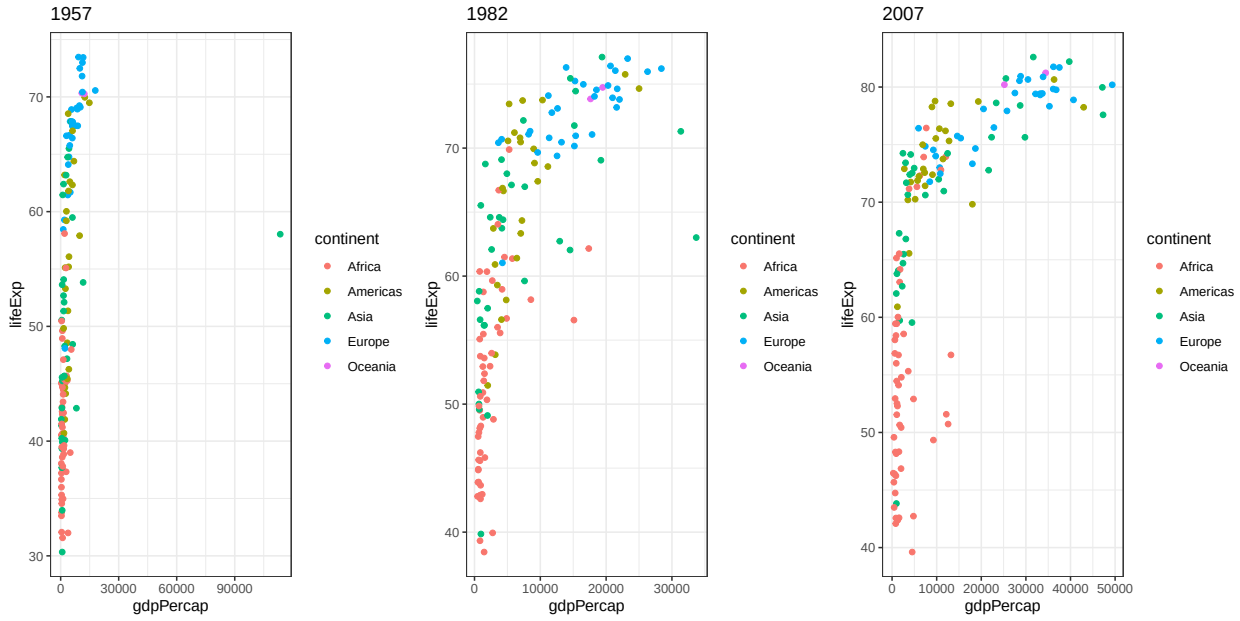
- En 1957, la corrélation de 0.7 montre une relation modérée.
- En 1982, une corrélation de 0.6 montre une relation similaire.
- En 2007, la corrélation revient à 0.7, indiquant une relation modérée persistante.

Visualisation de la relation entre `lifeExp` et `gdpPercap`

Scatterplots

```
p1 = ggplot(gapminder1957, aes(x = gdpPercap, y = lifeExp, colour = continent)) + geom_point() + theme_
p2 = ggplot(gapminder1982, aes(x = gdpPercap, y = lifeExp, colour = continent)) + geom_point() + theme_
p3 = ggplot(gapminder2007, aes(x = gdpPercap, y = lifeExp, colour = continent)) + geom_point() + theme_

grid.arrange(p1, p2, p3, ncol = 3)
```



Visualisation des matrices de corrélation

Utilisation de `corrplot` pour tracer les matrices de corrélation pour chaque année.

```r
# Matrices de corrélation
par(mfrow = c(1,3), oma = c(0, 1, 2, 1), mar = c(4, 3, 4, 3))
cor_matrix_1957 = matrix(c(1, cor_1957, cor_1957, 1), nrow = 2, ncol = 2)
cor_matrix_1982 = matrix(c(1, cor_1982, cor_1982, 1), nrow = 2, ncol = 2)
cor_matrix_2007 = matrix(c(1, cor_2007, cor_2007, 1), nrow = 2, ncol = 2)

# Noms des colonnes et lignes
colnames(cor_matrix_1957) = c("lifeExp", "gdpPercap")
rownames(cor_matrix_1957) = c("lifeExp", "gdpPercap")

colnames(cor_matrix_1982) = c("lifeExp", "gdpPercap")
rownames(cor_matrix_1982) = c("lifeExp", "gdpPercap")

colnames(cor_matrix_2007) = c("lifeExp", "gdpPercap")
rownames(cor_matrix_2007) = c("lifeExp", "gdpPercap")

# Affichage des matrices de corrélation
corrplot(cor_matrix_1957, method = "number", col = colorRampPalette(c("red", "white", "blue"))(200), ti
corrplot(cor_matrix_1982, method = "number", title = "Matrice de corrélation - 1982", tl.col = "red", c
corrplot(cor_matrix_2007, method = "number", title = "Matrice de corrélation - 2007", tl.col = "red", c
```