# Le niveau d'instruction contribue t-il à augmenter notre productivité au travail ?

Projet de recherche économétrie CPES2

AUTHOR
HERPE Victor & FOUQUET Baptiste

PUBLISHED
27/05/2024

# I- Introduction

De l'avènement de la loi Ferry en 1882 (instaurant l'instruction obligatoire de 6 à 13 ans) à la massification scolaire sous la Ve République, l'âge moyen d'entrée dans le monde professionnel n'a eu de cesse de reculer à mesure que le niveau d'instruction obligatoire augmentait en France. Ainsi, le XXe siècle a vu le temps de travail des travailleurs se réduire mais leur qualification s'accroitre. Entre les deux, se pose alors la question de la productivité du travail et de ses déterminants. Dès 1992, (Mankiw et al) cherchant à explorer les déterminants de la croissance économique à long terme, avaient observé le rôle de l'accumulation de capital humain (mesurée par le taux de scolarisation) dans l'augmentation de la productivité du travail. Bien que celle ci ne se concentre pas exclusivement sur la relation entre niveau d'éducation et croissance économique, cette étude a contribué à ouvrir la porte à toute une littérature académique en la matière. On retrouve ainsi les travaux de (Barro, 2001), (Hanushek et al, 2008) ou plus récemment (Psacharopoulos et al 2018) qui, avec tout un pan de l'économie de la croissance, ont souligné le rôle central de l'Éducation au sein de nos économies.

En tant qu'étudiants engagés dans la voie de l'enseignement supérieur, nous sommes directement concernés par ces considérations. Alors qu'aujourd'hui plus de 75% des jeunes bacheliers francais décident de poursuivre leurs études, il y a seulement cinquante ans cela ne concernait pas plus de 20% des étudiants. Toutefois, il convient aussi de se demander jusqu'à où cet "investissement sur l'avenir" est-il véritablement rentable économiquement et si cette "inflation des diplômes" a t-elle un sens pour nos seules économies. En un mot, quel est le coût et le prix de l'Éducation pour nos sociétés ?

# II- Nettoyage de données

```
#On commence par charger les packages nécessaires
library(dplyr)
library(ggplot2)
library(readxl)
library(tidyr)
library(maps)
library(corrplot)
library(tidyverse)
library(stargazer)
library(plotly)
library(kableExtra)
library(scales)
library(viridis)
library(mapproj)
```

## 1- Importation du fichier du PIB par habitant

Notre première base de données est librement disponible sur le site Our World In Data (link) qui regroupe des recherches empiriques et leurs données. En l'espèce les données que nous importons ont été compilées par la Banque mondiale depuis plusieurs sources dont Feenstra et al. (2015) et Penn World Table (2021), et sont exprimés en $ international de 2017.

```
# En premier lieu, on importe un fichier csv mesurant la productivité (PIB/habitants) des pays en f
labor_productivity_vs_gdp_per_capita <- read.csv("labor-productivity-vs-gdp-per-capita.csv")
```

```
# On affiche le nom de chaque variable de cette donnée
names(labor_productivity_vs_gdp_per_capita)
```

```
[1] "Entity"
[2] "Code"
[3] "Year"
[4] "Productivity..output.per.hour.worked"
[5] "GDP.per.capita..output..multiple.price.benchmarks."
[6] "Population..historical.estimates."
[7] "Continent"
```

Notre base de donnée contient 7 variables :

* `Entity` : Pays
* `Code` : 3 lettres définissant de manière unique le pays correspondant
* `Year` : Année
* `Productivity..output.per.hour.worked` : PIB divisé par le nombre annuel d'heures de travail par travailleurs et le nombre de personnes actives
* `GDP.per.capita..output..multiple.price.benchmarks.` : PIB divisé par habitant et ajusté en fonction du coût de la vie et du niveau d'inflation par pays
* `Population..historical.estimates.` : Population estimé avec différentes bases de données
* `Continent` : Continent

Toutes ces données ne nous seront pas utiles nous pouvons donc en enlever certaines. Ainsi, nous allons uniquement garder la variable de productivité PIB/habitants (corrigée par le niveau de vie dans chaque pays et par l'inflation) ainsi que d'autres données périphériques qui vont nous être utiles par la suite.

```
# On garde toutes les variables sauf celle de la productivité par heure travaillée
GDPpercapita <- labor_productivity_vs_gdp_per_capita%>%
  select(-Productivity..output.per.hour.worked)


# On renomme les variables pour plus de clarté
GDPpercapita = rename(GDPpercapita,"GDPcapita" = GDP.per.capita..output..multiple.price.benchmarks.

# Affichons les 5 premières données
kable(head(GDPpercapita, n = 5), caption = "Observation de la donnée:")
```

Observation de la donnée:

| Country | WBcode | Year | GDPcapita | Population | Continent |
|---|---|---|---|---|---|
| Abkhazia | OWID_ABK | 2015 | NA | NA | Asia |
| Afghanistan | AFG | -10000 | NA | 14737 | |
| Afghanistan | AFG | -9000 | NA | 20405 | |
| Afghanistan | AFG | -8000 | NA | 28253 | |
| Afghanistan | AFG | -7000 | NA | 39120 | |

```
# Calcul du nombre de NA pour la variable GDPcapita
nblignes <- nrow(GDPpercapita)
nbofNAGDPcapita <- sum(is.na(GDPpercapita$GDPcapita))
pourcentageofNAGDPcapita <- round(nbofNAGDPcapita/nblignes*100)

nbofNAcontinent <- sum(is.na(GDPpercapita$Continent) | GDPpercapita$Continent == "")
pourcentageofNAcontinent <- round(nbofNAcontinent/nrow(GDPpercapita)*100)
```

Nous constatons qu'il y a 83% de NA dans la colonne `GDPcapita` ce qui est beaucoup. De même pour la colonne des continents, il y a 100% de NA. Avant de commencer l'analyse il est nécessaire de supprimer tous les NA qui sont dans notre base

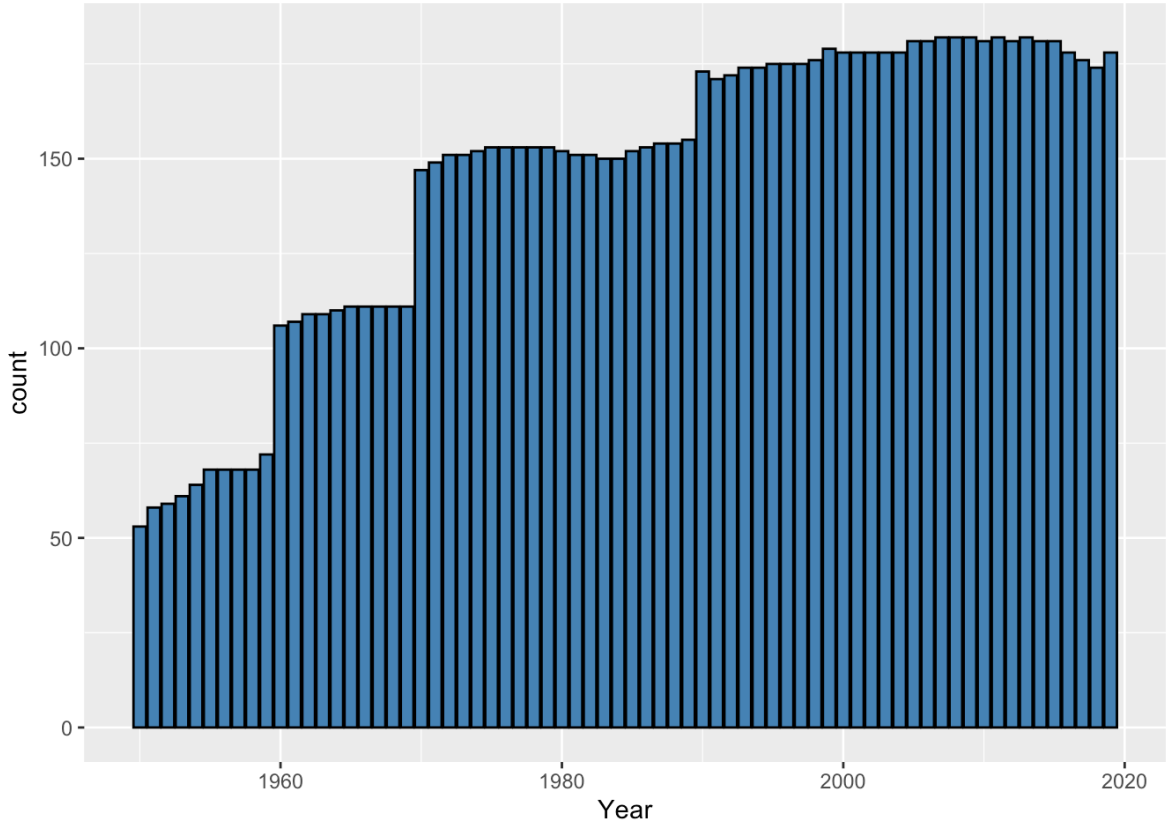
de données et plus particulièrement ceux qui sont dans la variable `GDPcapita`.

```
# Suppression des NA de la colonne `GDPcapita`
GDPpercapitaOFF <- na.omit(GDPpercapita[!is.na(GDPpercapita$GDPcapita),])

minyearGDPpercapita <- min(GDPpercapitaOFF$Year)
maxyearGDPpercapita <- max(GDPpercapitaOFF$Year)
```

Ici les données sont compris entre 1950 et 2019 ce qui est cohérent, utile et exploitable.

```
ggplot(GDPpercapitaOFF, aes(x = Year)) +
  geom_bar(color = "black", fill = "steelblue", stat = "count")
```




La représentation ci-dessus montre le nombre de pays qui sont présents dans cette base de données pour chaque année. On remarque ainsi l'introduction d'une vingtaine de pays en 1960 ainsi qu'une trentaine en 1970. Ces constats seront utiles par la suite pour expliquer des sauts de valeur dans nos graphiques.

Ensuite on cherche à associer chaque pays à son continent au regard du pourcentage de NA dans la colonne Continent. Pour ce faire on va importer une nouvelle base de données.

```
#Importation de notre base de données pour y ajouter le continent
WBcodewithcontinent <- read.csv("country-and-continent-codes-list-csv.csv", encoding = "UTF-8" )
# On supprime la variable continent de notre base de donnée initiale car elle est inexploitable
GDPpercapitaOFF <- GDPpercapitaOFF%>%select(-Continent)

# Affichons les 5 premières données de notre donnée pour les continents
kable(head(WBcodewithcontinent, n = 5), caption = "Observation de la donnée:")
```

Observation de la donnée:

| Continent_Name | Continent_Code | Country_Name | Two_Letter_Country_Code | Three_Letter_Country_Code | Country_Number |
|---|---|---|---|---|---|
| Asia | AS | Afghanistan, Islamic Republic of | AF | AFG | 4 |

| Continent_Name | Continent_Code | Country_Name | Two_Letter_Country_Code | Three_Letter_Country_Code | Country_Number |
|---|---|---|---|---|---|
| Europe | EU | Albania, Republic of | AL | ALB | 8 |
| Antarctica | AN | Antarctica (the territory South of 60 deg S) | AQ | ATA | 10 |
| Africa | AF | Algeria, People's Democratic Republic of | DZ | DZA | 12 |
| Oceania | OC | American Samoa | AS | ASM | 16 |

Seulement 2 colonnes nous intéressent (le continent et le code avec les 3 lettres du WBcode). Nous allons donc les garder.

```
# On renomme les variables qui vont nous être utiles puis on les garde
WBcodewithcontinent = rename(WBcodewithcontinent, "WBcode" = Three_Letter_Country_Code,"Continent"
WBcodewithcontinentperfect <- WBcodewithcontinent%>%select(WBcode, Continent)

# On compte le nombre de pays qui ne sont toujours pas associés à un continent
nbofNAWBcode <- sum(is.na(WBcodewithcontinentperfect$WBcode) | WBcodewithcontinentperfect$WBcode ==
```

Dès lors, il y a 4 pays qui ne sont pas associés à un continent donc nous allons les enlever :

```
# Suppression des lignes contenant des NA
WBcodewithcontinentperfect <- na.omit(WBcodewithcontinentperfect[!(is.na(WBcodewithcontinentperfect

# On vérifie que chaque WBcode n'apparait qu'une seule fois
occurrences <- table(WBcodewithcontinentperfect$WBcode)
# On trouve le nombre maximum d'occurrences
max_occurrence <- max(occurrences)
```

On constate que des pays sont asscociés à 2 continents ce qui pose problème. Ces pays sont ceux du Moyen-Orient et sont associés à la fois à l'Asie et à l'Afrique. Nous décidons donc de les classer dans les pays d'Asie.

```
# Réarrangement des pays du Moyen-Orient
WBcodewithcontinentperfect <- WBcodewithcontinentperfect[c(-17,-59,-117,-84,-235,-192),]

kable(head(WBcodewithcontinentperfect, n = 5), caption = "Observation de la donnée:")
```

Observation de la donnée:

| WBcode | Continent |
|---|---|
| AFG | Asia |
| ALB | Europe |
| ATA | Antarctica |
| DZA | Africa |
| ASM | Oceania |

Maintenant que notre base de données avec chaque continent est propre, on peut donc la regrouper avec notre base de données `GDPpercapitaOFF`

```
# On join les 2 bases de données en fonction du WBcode

GDPpercapitaperfect <- WBcodewithcontinentperfect%>%right_join(GDPpercapitaOFF, by = "WBcode")

kable(head(GDPpercapitaperfect, n = 5), caption = "Observation de la donnée:")
```

Observation de la donnée:

| WBcode | Continent | Country | Year | GDPcapita | Population |
|---|---|---|---|---|---|
| ALB | Europe | Albania | 1971 | 3159.809 | 2389820 |
| ALB | Europe | Albania | 1972 | 3214.666 | 2455181 |
| ALB | Europe | Albania | 1973 | 3267.848 | 2520442 |
| ALB | Europe | Albania | 1974 | 3330.071 | 2585457 |
| ALB | Europe | Albania | 1975 | 3385.273 | 2650128 |

**Notre base de données pour le PIB par habitant est désormais terminée**

## 2- Importation du fichier du niveau d'éducation par pays

Notre seconde base de données provient d'une mise à jour des données de Barro et Lee (2013) et sont librement disponibles sur ce site (link). Ces données recensent les niveaux moyens d'éducation par pays tous les 5 ans de 1950 à 2015.

```
# Importation des données
BL_v3_MF <- read.csv("BL_v3_MF1564.csv")

# Visualisation de chaque variable
names(BL_v3_MF)
```

```
 [1] "BLcode"      "country"     "year"        "sex"         "agefrom"
 [6] "ageto"       "lu"          "lp"          "lpc"         "ls"
[11] "lsc"         "lh"          "lhc"         "yr_sch"      "yr_sch_pri"
[16] "yr_sch_sec"  "yr_sch_ter"  "WBcode"      "region_code" "pop"
```

Il y a 20 variables dans cette base de données :
*BLcode : Code Barro-Lee du pays*
country : Pays
*year : Année*
sex : Sexe de l'individu
*agefrom : Age minimal*
ageto : Age maximal
*lu : Pourcentage de la population qui n'est jamais allé à l'école*
lp : Pourcentage de la population qui est entré à l'école primaire
*lpc : Pourcentage de la population qui a terminé l'école primaire*
ls : Pourcentage de la population qui est entré à l'école secondaire
*lsc : Pourcentage de la population qui a terminé l'école secondaire*
lh : Pourcentage de la population qui est entré dans les études supérieures
*lhc : Pourcentage de la population qui a terminé ses études supérieures*
yr_sch : Années moyennes passées à l'école
*yr_sch_pri : Années moyennes passées à l'école primaire*
yr_sch_sec : Années moyennes passées à l'école secondaire
*yr_sch_ter : Années moyennes passées dans les études supérieur*
WBcode : Code de la banque mondiale du pays
*region_code : Région du monde*
pop : Population

Nous pouvons supprimer les variables qui ne vont pas nous être utiles comme le sexe, BLcode, pop, ageto et agefrom.

```
#Suppression des variables
BL_v3_MF <- BL_v3_MF%>%select(-sex,-BLcode,-pop,-ageto,-agefrom)
```

```
# Ensuite nous renommons nos variables
BL_v3_MF = rename(BL_v3_MF,"Country" = country, "Year" = year, "Region" = region_code)

# Observation des données
kable(head(BL_v3_MF, n = 5), caption = "Observation des données:")
```

Observation des données:

| Country | Year | lu | lp | lpc | ls | lsc | lh | lhc | yr_sch | yr_sch_pri | yr_sch_sec | yr_sch_ter | WBcode | Region |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Algeria | 1950 | 81.12 | 17.10 | 3.65 | 1.48 | 0.50 | 0.30 | 0.18 | 0.834 | 0.729 | 0.095 | 0.010 | DZA | Middle East and North Africa |
| Algeria | 1955 | 81.50 | 16.54 | 3.43 | 1.66 | 0.53 | 0.26 | 0.17 | 0.823 | 0.714 | 0.100 | 0.008 | DZA | Middle East and North Africa |
| Algeria | 1960 | 82.50 | 14.29 | 3.16 | 2.88 | 1.02 | 0.33 | 0.19 | 0.896 | 0.716 | 0.169 | 0.010 | DZA | Middle East and North Africa |
| Algeria | 1965 | 80.08 | 15.00 | 4.21 | 4.46 | 1.90 | 0.45 | 0.24 | 1.151 | 0.871 | 0.267 | 0.014 | DZA | Middle East and North Africa |
| Algeria | 1970 | 72.02 | 20.25 | 6.06 | 7.25 | 3.87 | 0.38 | 0.18 | 1.690 | 1.247 | 0.432 | 0.011 | DZA | Middle East and North Africa |

**Ainsi, notre base de données quant au niveau d'éducation par pays et par années est propre**

# III- Statistiques descriptives

Maintenant que nos données sont propres, nous pouvons en prendre toute la mesure en les analysant distinctement.

**1 - Données PIB par habitant**

Tout d'abord, mesurons le nombre de pays présents dans notre base de données :

```
# Calcul du nombre de pays dans la base de données
freq_table1 <- table(GDPpercapitaperfect$Country)
nbcountryGDPpercapita <- length(freq_table1)
# Sachant qu'il y a 195 pays dans le monde
pourcentageofcountry1 <- round(nbcountryGDPpercapita/195*100)
```

Ainsi, 182 pays sont ici renseignés dans notre base de données, soit 93% des pays du monde. On peut dès lors considérer que nos données sont suffisamment représentatives pour généraliser nos conclusions.

Pour se donner un ordre d'idée des valeurs des données sur lesquelles nous travaillons, la fonction *summary* permet de donner des indicateurs de localisation et de dispersion :

```
kable(as.matrix(summary(GDPpercapitaperfect$GDPcapita)) %>% t(),
      caption = "PIB par habitant - Statistiques descriptives:")
```

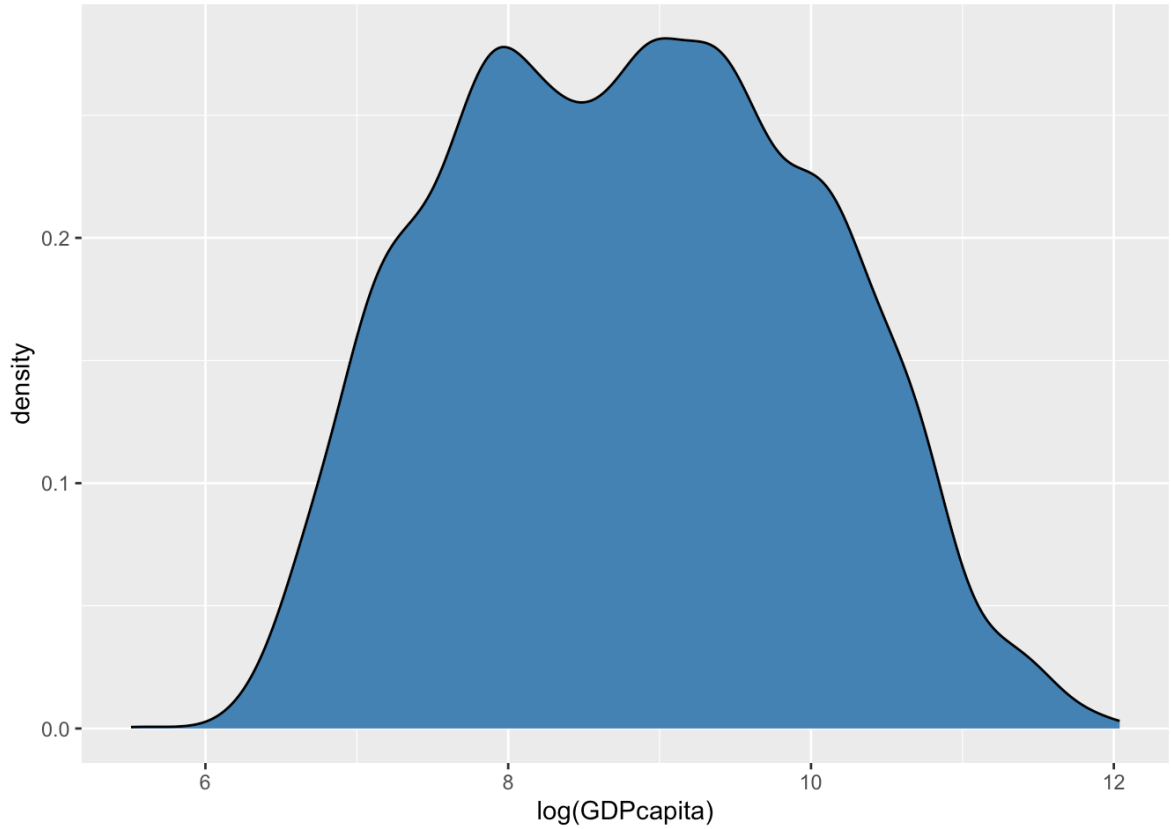
PIB par habitant - Statistiques descriptives:

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|---|---|---|---|---|---|
| 246.7417 | 2625.074 | 6724.392 | 13111.16 | 16629.46 | 169200.3 |

*On remarque ainsi l'écart significatif de PIB par habitant entre le Liberia en 1996 (246$) et le Qatar en 2012 (169 200 $). De plus, la moyenne est de 13 111$ tandis que la médiane est de 6 724$ ce qui implique que la moyenne est tirée à la hausse par des valeurs extremes. De ce fait, nous normaliserons nos données par la suite en leur appliquant la fonction logarithme*

**On peut alors faire une première visualisation de ces données**

```
# Densité
ggplot(GDPpercapitaperfect, aes(x = log(GDPcapita))) +
  geom_density(color = "black", fill = "steelblue")
```




**2 - Données niveau d'étude par pays**

De même, pour notre seconde base de données, nous utilisons ici la variable du nombre moyen d'années d'études pour la population (par pays et par années) qui est plus générale que nos autres variables de cette base de données.

Mesurons la représentativité de notre base de données :

```
# Calcul du nombre de pays reppresentés
freq_table2 <- table(BL_v3_MF$Country)
nbcountryBL <- length(freq_table2)
# Il y a 195 pays dans le monde
pourcentageofcountry2 <- round(nbcountryBL/195*100)
```

Ainsi 146 pays sont renseignés, soit 75% des pays du monde sont représentés dans cette base de données.

Indicateurs de localisation et de dispersion :

```
kable(as.matrix(summary(BL_v3_MF$yr_sch)) %>% t(),
      caption = "Niveau d'étude par pays - Statistiques descriptives:")
```

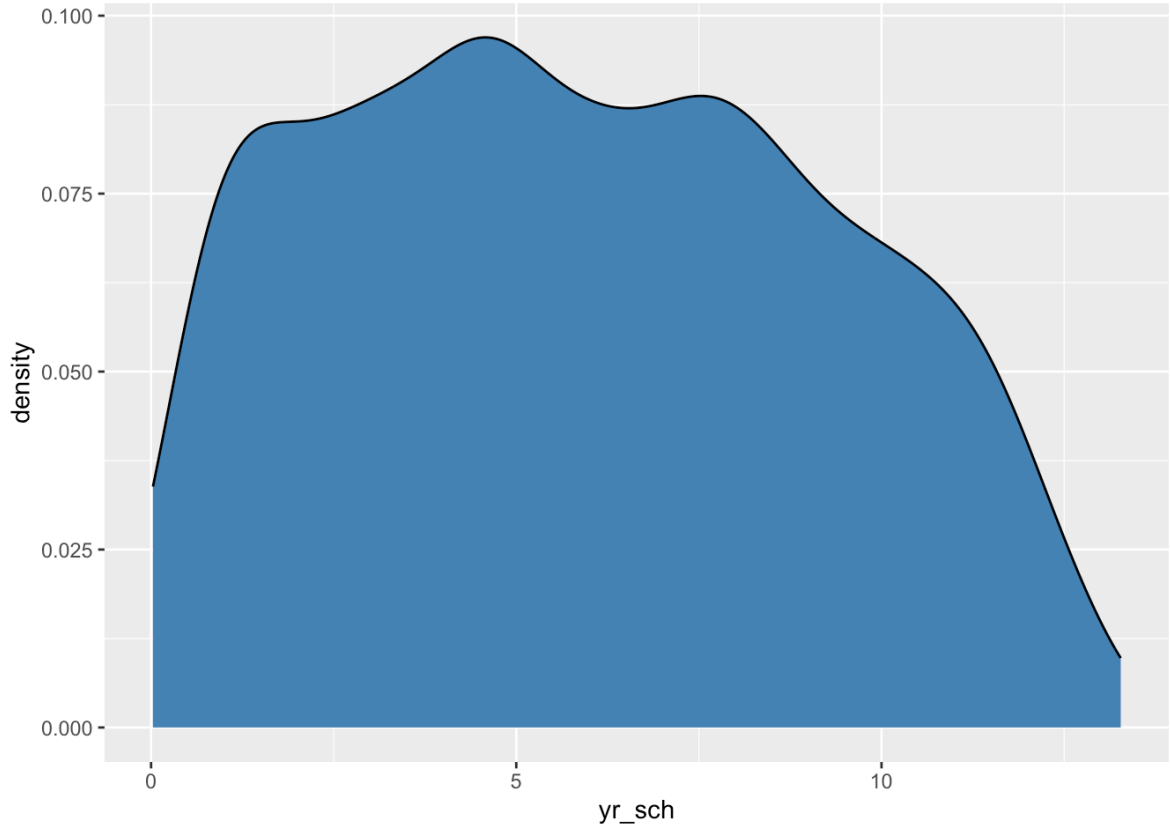
Niveau d'étude par pays - Statistiques descriptives:

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|---|---|---|---|---|---|
| 0.026 | 3.10025 | 5.787 | 5.928776 | 8.61175 | 13.275 |

*La valeur minimale correspond au Yemen en 1950 (avec moins d'une année d'éducation en moyenne) et la valeur maximale correspond aux USA en 2015 (13 ans d'études en moyenne). De plus, la moyenne (6 ans) et la médiane (5,8 années) de nos données sont très proches.*

**Data visualisation pour le niveau d'education**

```
# Densité
ggplot(BL_v3_MF, aes(x = yr_sch)) +
  geom_density(color = "black", fill = "steelblue")
```




La densité permet d'observer graphiquement la distribution des données. On remarque que le niveau d'éducation le plus atteint en moyenne par la population depuis 1950 se situe vers les 5 ans. De plus, le niveau moyen d'années d'instruction tend à décroitre plus rapidement à partir de 8 années d'études.

**Dès lors, nous pouvons joindre nos deux bases de données afin de faire des analyses jointes par la suite**

Avant de joindre par la droite sur la base de données avec le niveau d'éducation, il convient d'enlever le nom du pays de l'autre base de données (PIB) puisque nous allons regrouper nos données par le *WBcode* qui est commun à tous (même orthographe) alors que les pays peuvent être écrit différemment :

```
# Suppression de la variable Country dans GDPpercapitaperfect
GDPpercapitaperfect <- GDPpercapitaperfect%>%select(-Country)

# Join des 2 bases de données par le WBcode et Year
donnee <- GDPpercapitaperfect%>%right_join(BL_v3_MF, by = c("WBcode","Year"))

kable(head(donnee, n = 5), caption = "Observation des données:")
```

Observation des données:

| WBcode | Continent | Year | GDPcapita | Population | Country | lu | lp | lpc | ls | lsc | lh | lhc | yr_sch | yr_sch_pri | yr_sch_sec | yr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ALB | Europe | 1975 | 3385.273 | 2650128 | Albania | 26.81 | 29.25 | 22.81 | 39.89 | 14.35 | 4.05 | 2.16 | 5.765 | 3.769 | 1.873 | |

| WBcode | Continent | Year | GDPcapita | Population | Country | lu | lp | lpc | ls | lsc | lh | lhc | yr_sch | yr_sch_pri | yr_sch_sec | yr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ALB | Europe | 1980 | 3714.541 | 2941650 | Albania | 19.73 | 24.77 | 18.97 | 51.02 | 19.10 | 4.48 | 2.51 | 6.951 | 4.650 | 2.161 | |
| ALB | Europe | 1985 | 3689.646 | 3171727 | Albania | 13.64 | 20.83 | 15.98 | 60.39 | 23.49 | 5.14 | 3.07 | 7.931 | 5.370 | 2.397 | |
| ALB | Europe | 1990 | 3681.208 | 3295073 | Albania | 10.15 | 16.91 | 12.79 | 67.25 | 26.44 | 5.68 | 3.48 | 8.611 | 5.914 | 2.513 | |
| ALB | Europe | 1995 | 4391.234 | 3284370 | Albania | 13.06 | 15.27 | 11.73 | 65.88 | 26.47 | 5.79 | 3.63 | 8.537 | 5.926 | 2.422 | |

Nous obtenons donc notre base de données finale. Toutefois, comme nous venons de regrouper par la droite le fichier BL_v3_MF, il est possible qu'il y ait des NA dans les colonnes qui appartenaient à GDPpercapitaperfect et notamment la colonne GDPcapita. Vérifions :

```
# Calcul du nombre et du pourcentage de NA dans la colonne GDPcapita
nbNAGDPcapita <- sum(is.na(donnee$GDPcapita) |donnee$GDPcapita == "")
pourcentageNAGDPcapita <- round(nbNAGDPcapita/nrow(donnee)*100)
```

Nous avons donc 21% de NA dans la colonne GDPcapita de la base de données `donnee` Nous allons donc supprimer les lignes pour lesquels il y a des NA dans la colonne GDPcapita :

```
# Suppression des NA de la colonne GDPcapita
donneperfect <- na.omit(donnee[!is.na(donnee$GDPcapita),])


kable(head(donnee, n = 5), caption = "Observation des données:")
```

Observation des données:

| WBcode | Continent | Year | GDPcapita | Population | Country | lu | lp | lpc | ls | lsc | lh | lhc | yr_sch | yr_sch_pri | yr_sch_sec | yr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ALB | Europe | 1975 | 3385.273 | 2650128 | Albania | 26.81 | 29.25 | 22.81 | 39.89 | 14.35 | 4.05 | 2.16 | 5.765 | 3.769 | 1.873 | |
| ALB | Europe | 1980 | 3714.541 | 2941650 | Albania | 19.73 | 24.77 | 18.97 | 51.02 | 19.10 | 4.48 | 2.51 | 6.951 | 4.650 | 2.161 | |
| ALB | Europe | 1985 | 3689.646 | 3171727 | Albania | 13.64 | 20.83 | 15.98 | 60.39 | 23.49 | 5.14 | 3.07 | 7.931 | 5.370 | 2.397 | |
| ALB | Europe | 1990 | 3681.208 | 3295073 | Albania | 10.15 | 16.91 | 12.79 | 67.25 | 26.44 | 5.68 | 3.48 | 8.611 | 5.914 | 2.513 | |

| WBcode | Continent | Year | GDPcapita | Population | Country | lu | lp | lpc | ls | lsc | lh | lhc | yr_sch | yr_sch_pri | yr_sch_sec | yr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ALB | Europe | 1995 | 4391.234 | 3284370 | Albania | 13.06 | 15.27 | 11.73 | 65.88 | 26.47 | 5.79 | 3.63 | 8.537 | 5.926 | 2.422 | |

Ci dessus, les 5 premières lignes de notre base de données finale.

Calculons maintenant le nombre de pays présents dans notre base de données finale :

```
#Nombre de pays dans notre base de données finale
freq_table3 <- table(donneperfect$Country)
nbcountrydonnee <- length(freq_table3)
#Il y a 195 pays dans le monde
pourcentageofcountry3 <- round(nbcountrydonnee/195*100)
```

Ainsi, dans cette base de données nous avons 138 pays qui sont renseignés ce qui fait 71% des pays du monde qui sont représentés dans cette base de données. Ce nombre est suffisant pour avoir des résultats significatifs.

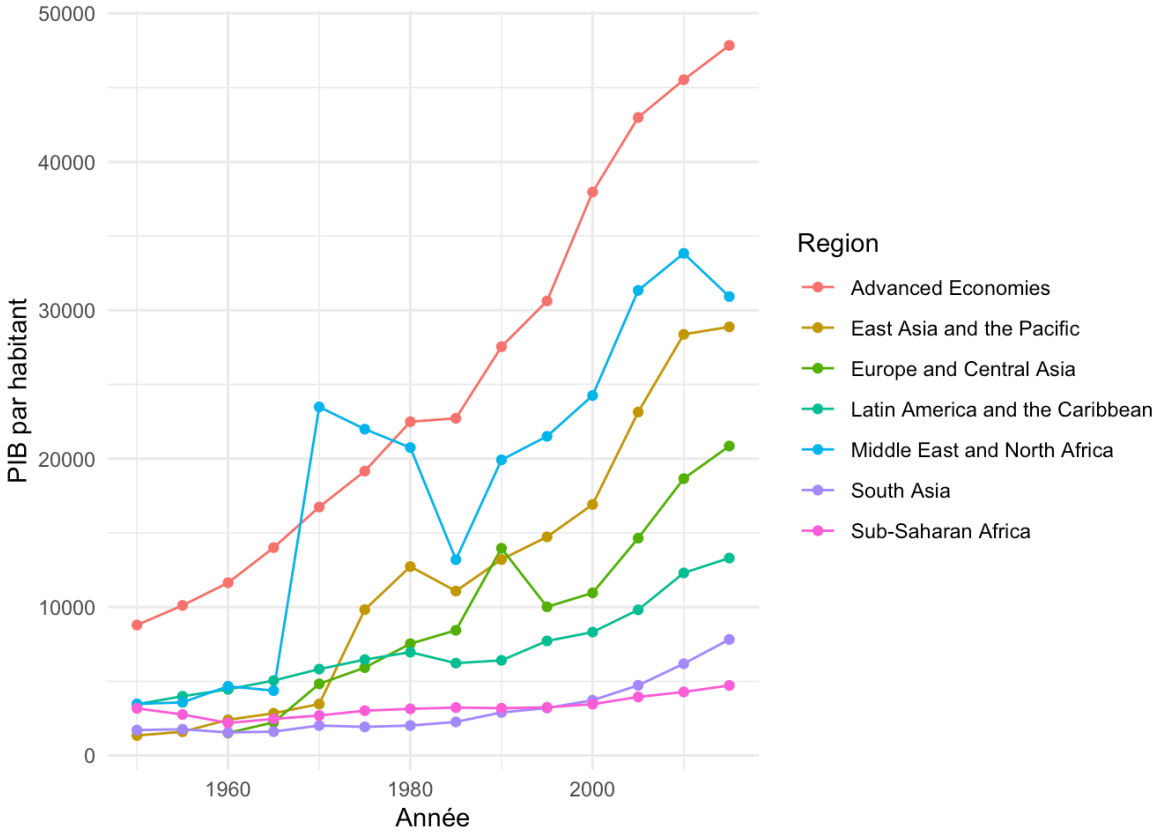
# IV- Visualisation des données

Afin d'avoir une reppresentation plus fine de la distribution de nos données finales, nous pouvons procéder à plusieurs reppréntations graphiques.

**Evolution temporelle du PIB par habitant**

```
# Calcul de la moyenne du PIB par habitant de chaque continent
moyennes_pib <- aggregate(GDPcapita ~ Year + Region, data = donneperfect, FUN = mean, na.rm = TRUE)

ggplot(moyennes_pib, aes(x = Year, y = GDPcapita, color = Region)) +
  geom_line() +
  geom_point() +
  labs(x = "Année", y = "PIB par habitant") +
  theme_minimal()
```



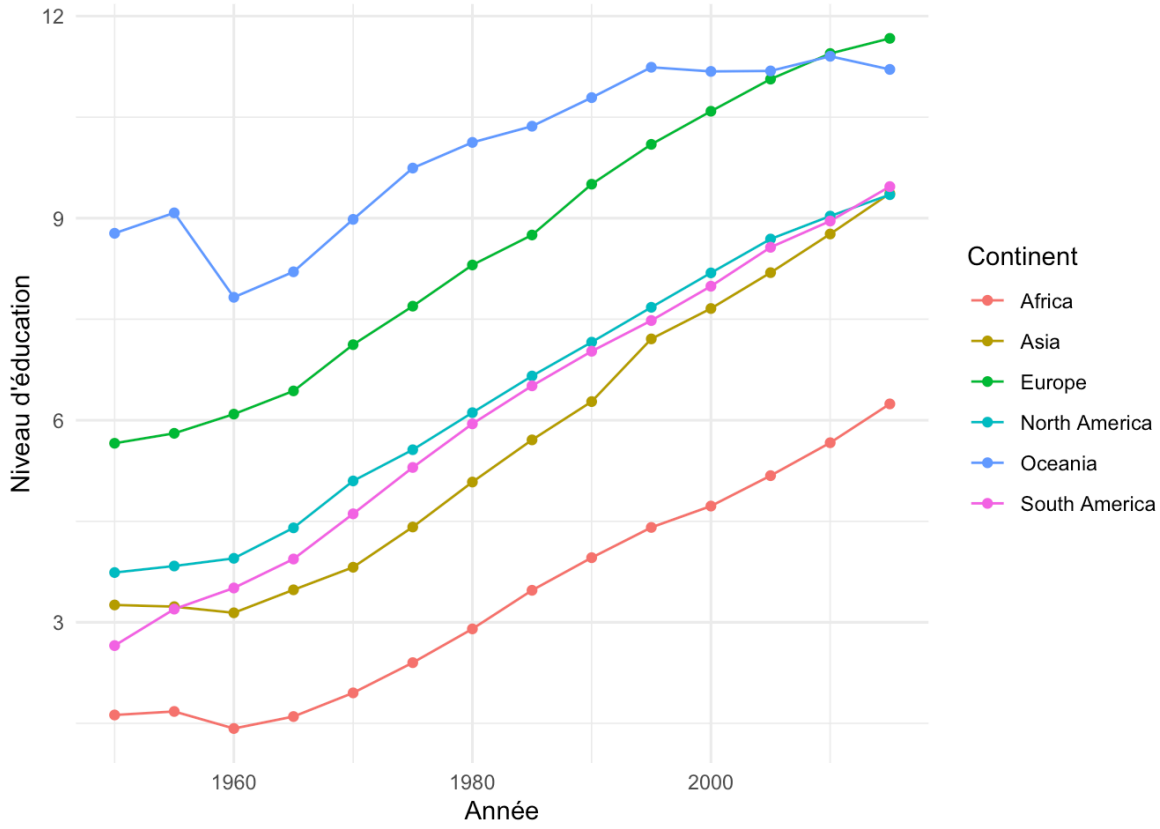

Le graphique montre l'évolution du PIB par habitant en fonction des années et des continents. On remarque un saut de la productivité des pays d'Asie entre 1969 et 1970. Celui ci correspond à l'introduction des pays du Moyen-Orient tels que le Qatar, le Koweit, le Bahrein ou l'Arabie Saoudite dans notre base de données à partir de 1970. Leur PIB par habitant est alors très élevé compte tenu de la faible population des ces pays et de leurs hauts revenus. De plus, les prix du pétrole vont s'envoler dans ces années avec les chocs pétroliers de 1973 et 1979.

Il en est de même pour l'Océanie en 1960 avec l'introduction des données pour les iles Fidji. De plus, soulignons que l'Amérique du Nord comprend ici notamment les pays des Caraibes (Haiti, Porto Rico…) ce qui explique l'allure de sa courbe malgré les données des USA.

**Evolution temporelle du niveau d'éducation**

```
# Calcul de la moyenne du niveau d'instruction par continent
moyenne_edu <- aggregate(yr_sch ~ Year + Continent, data = donneperfect, FUN = mean, na.rm = TRUE)

ggplot(moyenne_edu, aes(x = Year, y = yr_sch, color= Continent)) +
  geom_line() +
  geom_point() +
  labs(x = "Année", y = "Niveau d'éducation") +
  theme_minimal()
```



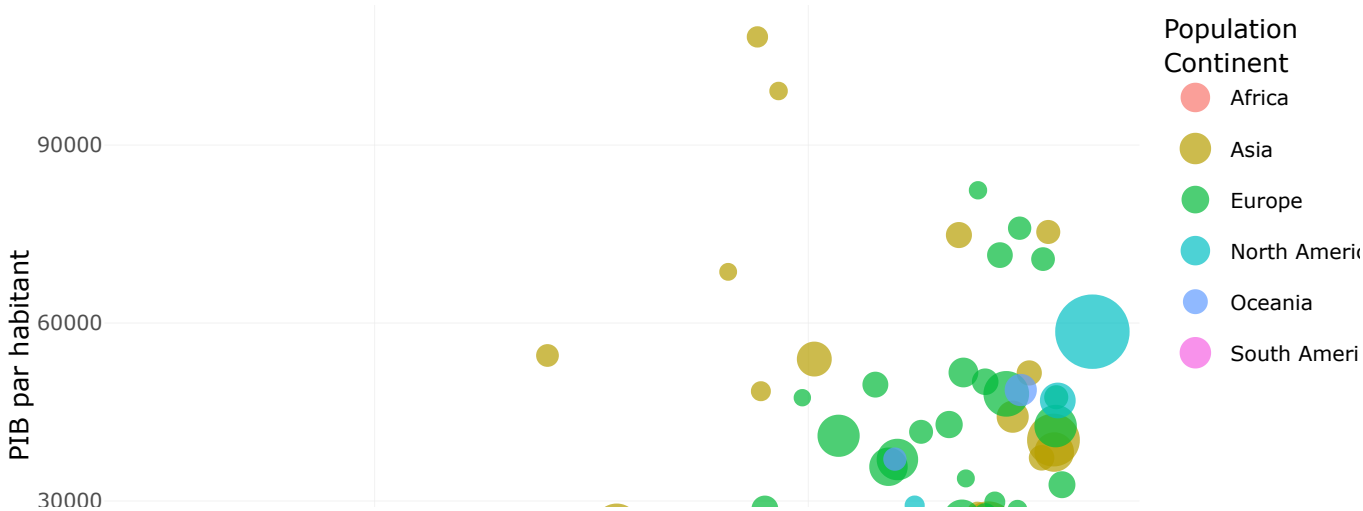

De même, pour ce graphique les baisses observées pour l'Océanie, l'Afrique et l'Asie autour de 1960 sont aussi dues à l'introduction de nouvelles valeurs telles que les Fidji en 1960 pour l'Océanie. On remarquera aussi l'avance et le caractère pionnier de l'Australie et de la Nouvelle Zélande qui dès 1950 avaient un système éducatif performant avec en moyenne 9 ans d'instruction.

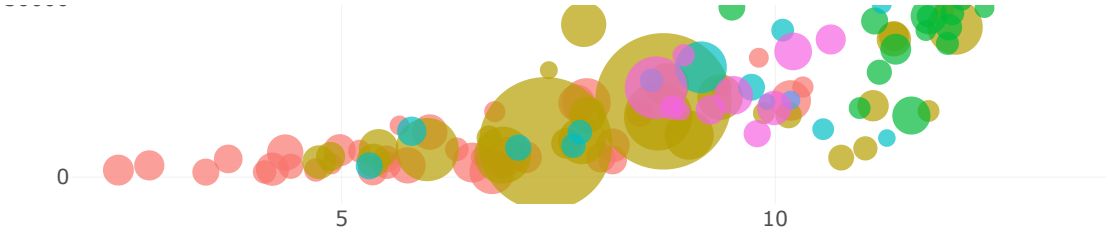
Le graphique montre la tendance mondiale à la hausse du niveau d'instruction de la population. Si l'écart d'instruction entre les régions du monde ne se résorbe pas entre 1950 et 2015, on constate cependant le rattrapage des pays du Moyen-Orient et d'Afrique du Nord sur les pays d'Asie de l'Est et d'Amérique latine.

**Graphique à bulle interactif pour l'année 2015**

```
ggbull <- ggplot(donneperfect %>% filter(Year == "2015"), aes(x = yr_sch, y = GDPcapita, size = Pop
  geom_point(alpha = 0.7) +
  scale_size_continuous(range = c(2, 20)) +
  labs(x = "Années d'éducation moyennes", y = "PIB par habitant", size = "Population", color = "Cor
  theme_minimal()

ggplotly(ggbull, tooltip = "text")
```

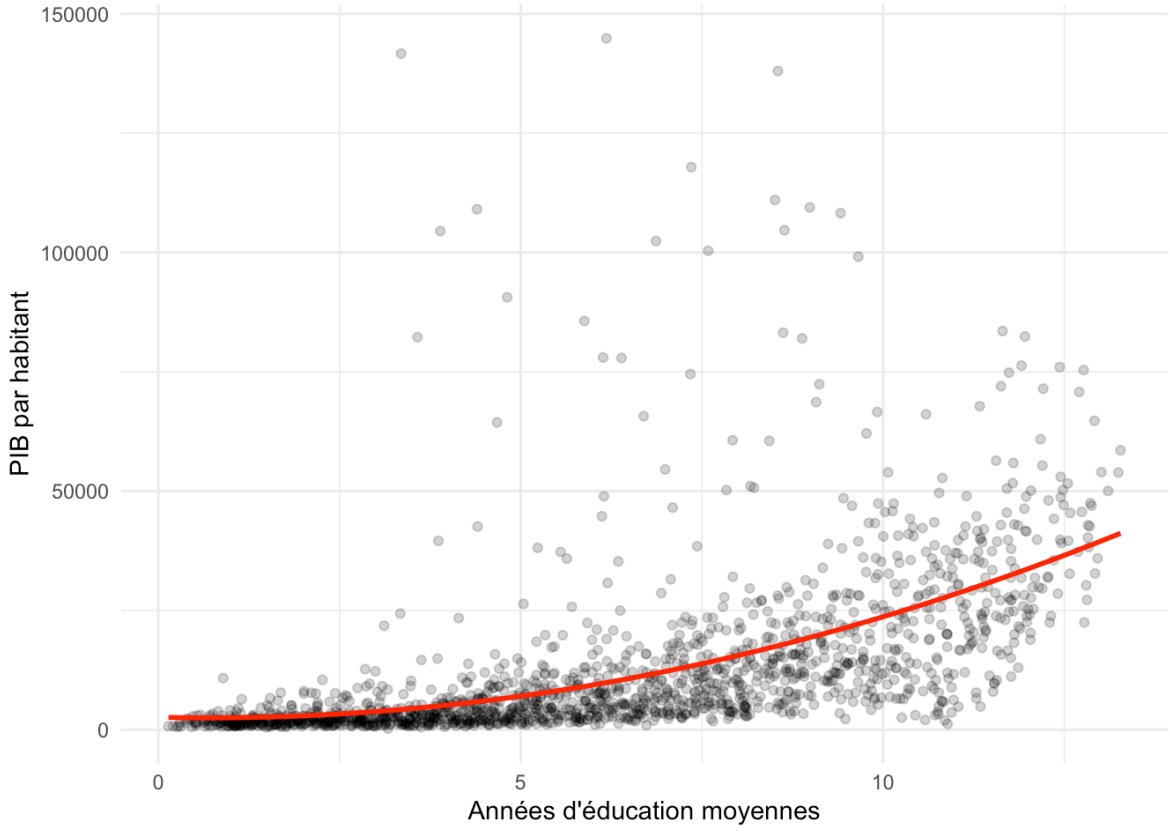





Ici, nous avons fait le choix de représenter l'évolution du PIB par habitants en fonction du nombre d'années d'éducation en 2015. On remarque une certaine homogénéité dans le distribution des données selon leur continent, notamment pour l'Europe, L'Amérique du Sud ou bien l'Afrique. Toutefois les Etats Unis, Singapour, Macao ainsi que les pays pétroliers de la péninsule arabique se détachent des pays de leur continent. Certaines valeurs comme le Qatar et la région administrative de Macao en Chine sont même aberantes au regard de leur niveau de productivité. Ce qui commence à se dessiner est aussi la croissance plus rapide des valeurs de productivité au dessus de 10 annnées d'éducation. Plus encore, au dessus de 12 années d'éducation en moyenne, pas un pays n'a de productivité en decà des pays ayant moins de 8 années d'éducation en moyenne.

**Diagramme de dispersion avec courbe polynomiale**

```
ggplot(donneperfect, aes(x = yr_sch, y = GDPcapita)) +
  geom_point(alpha = 0.2) +
  geom_smooth(method = "lm", formula = y ~ poly(x, 2), se = FALSE, color = "red") +
  labs(x = "Années d'éducation moyennes", y = "PIB par habitant") +
  theme_minimal()
```




On peut ainsi confirmer nos conjectures précédentes en repprésentant une courbe de régression polynomiale. On remarque que la courbe étant croissante et convexe, elle semble établir une correlation positive entre les deux variables.

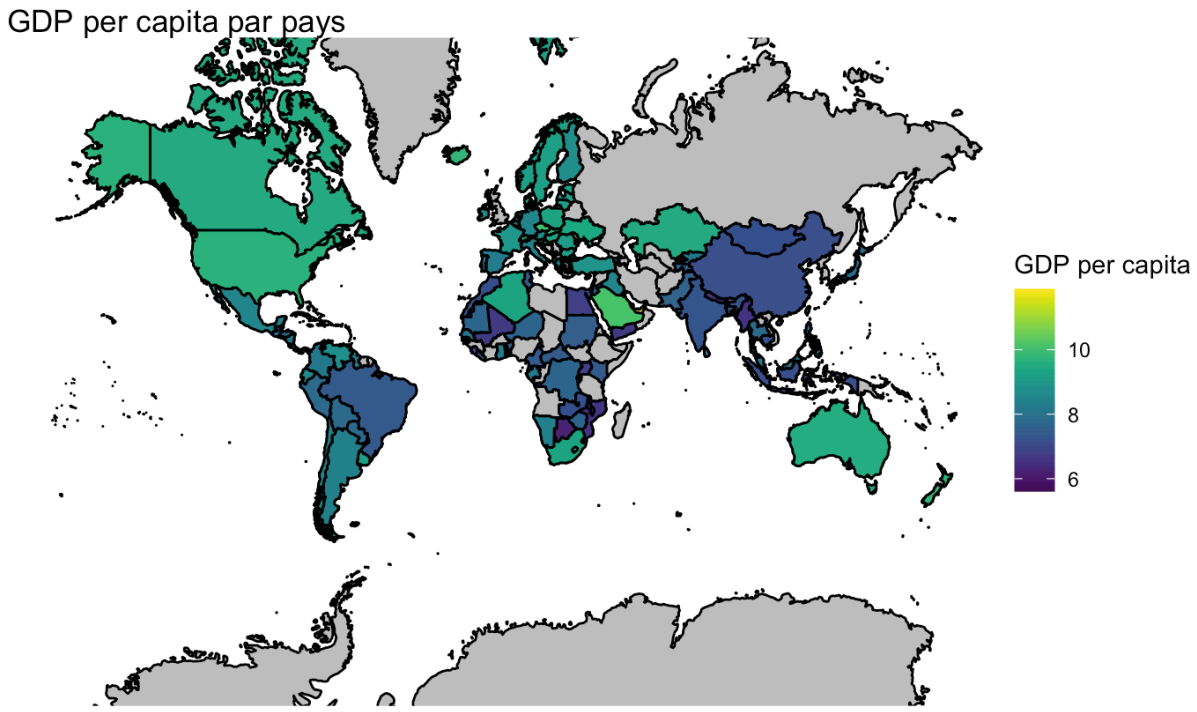
**Carte du monde avec le PIB par habitant par pays**

```
world_map <- map_data("world")


# Créer un jeu de données avec les valeurs de GDPcapita par pays et ici on prend la base de données
gdp_data <- donnee %>%
```

```
                filter(!is.na(GDPcapita))

                ggplot() +
        geom_map(data = world_map, map = world_map,
              aes(x = long, y = lat, map_id = region),
                 fill = "grey", color = "black") +
        geom_map(data = gdp_data, map = world_map,
          aes(fill = log(GDPcapita), map_id = Country),
                          color = "black") +
    expand_limits(x = world_map$long, y = world_map$lat) +
scale_fill_viridis(name = "GDP per capita",limits = c(min(log(gdp_data$GDPcapita), na.rm = TRUE
        labs(fill = "GDP per capita", title = "GDP per capita par pays") +
          coord_map("mercator", xlim = c(-180,180), ylim = c(-90,90)) +
                        theme_void()
```

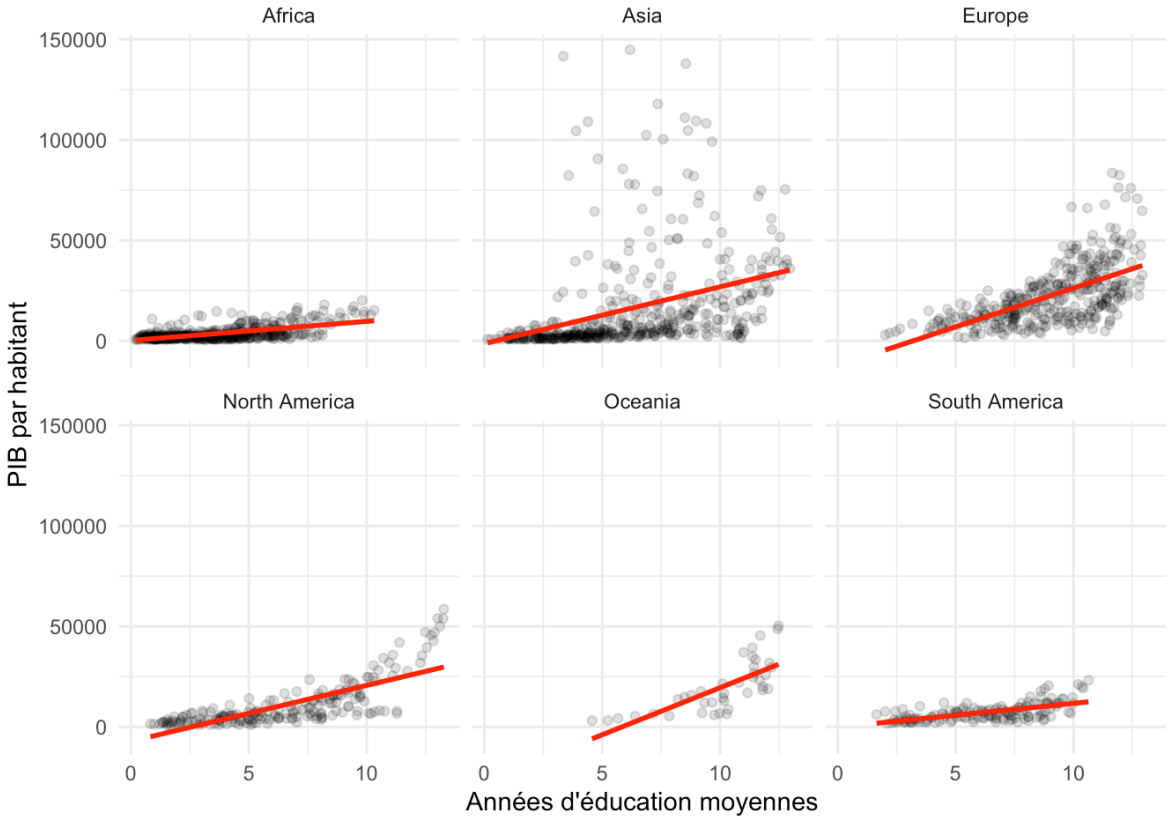



Sur la carte du monde ci dessus, les pays sont colorés en fonction de la moyenne de leur productivité (de 1950 à 2015). Plus les pays sont clairs, plus leur productivité est élevée. Ainsi il apparait clairement que les pays du Nord (ainsi que l'Australie) ont une productivité plus élevée que les pays du Sud. De même, les Emirats Arabe Unis, le Koweit et le Bahrein constituent les valeurs les plus élevées de nos données et tirent celles ci à la hausse.

Nous avons décidé de choisir une projection de la terre plane (coord_map()) car c'est la représentation la plus courante aujourd'hui et elle respecte parfaitement la superficie et les frontières de chaque territoire, permettant ainsi de comparer les différents pays.

**Graphique en fonction du continent**

```
ggplot(donneperfect, aes(x = yr_sch, y = GDPcapita)) +
  geom_point(alpha = 0.15) +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  facet_wrap(~ Continent) +
  labs(x = "Années d'éducation moyennes", y = "PIB par habitant", color = "Continent") +
  theme_minimal()
```




La mise en regard des régressions par continent souligne que l'Europe et l'Asie semblent avoir des courbes de corrélation similaires alors que les distributions de leurs données ne sont pas similaires. De plus, on remarque que ce sont les valeurs extrêmes d'Asie (pays du Moyen-Orient) qui tirent à la hausse la courbe qui n'aurait pas la même allure autrement.

# V- Analyse de régression

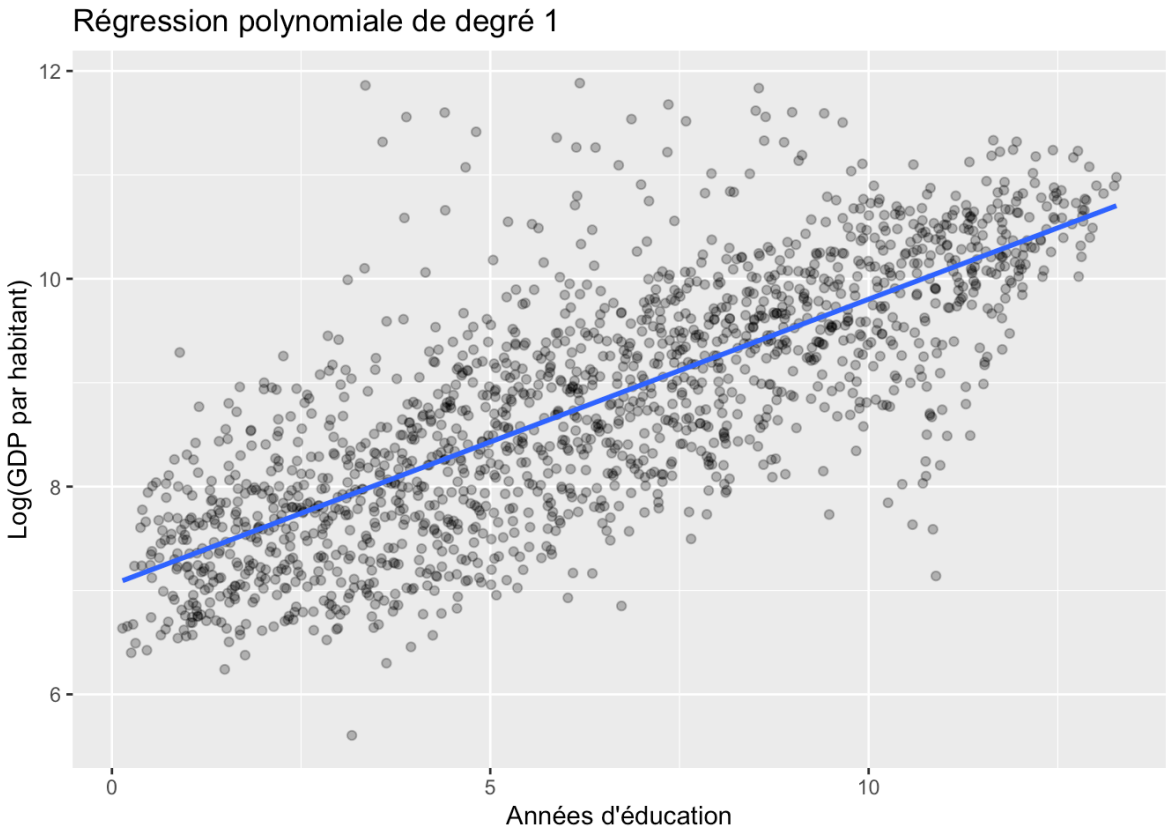
L'équation à estimer est la suivante :

$$1\{GDPpercapita_i\} = \alpha + \beta \times 1\{Year\_school_i\} + \varepsilon_i$$

De fait, on régresse le PIB/HABITANTS (la variable dépendante) sur le NIVEAU-D'EDUCATION (la variable indépendante) afin de trouver les coefficients α et β qui caractérisent cette régression.

**Tracer le graphique avec la régression**

```
ggplot(donneperfect, aes(x = yr_sch, y = log(GDPcapita))) +
  geom_point(alpha = 0.3) +
  geom_smooth(method = "lm", formula = y ~ poly(x, 1), se = FALSE) +
  labs(title = "Régression polynomiale de degré 1",
       x = "Années d'éducation", y = "Log(GDP par habitant)")
```




**Régression linéaire**

```
stargazer(lm(GDPcapita~ yr_sch, donneperfect), type = 'text')
```

```
================================================
                        Dependent variable:
                    ----------------------------
                               GDPcapita
------------------------------------------------
yr_sch                       2,823.542***
                              (106.702)

Constant                     -4,882.718***
                              (761.070)

------------------------------------------------
Observations                    1,618
R2                              0.302
Adjusted R2                     0.302
Residual Std. Error    14,289.790 (df = 1616)
F Statistic          700.233*** (df = 1; 1616)
================================================
Note:                *p<0.1; **p<0.05; ***p<0.01
```

La variable `yr_sch` a un coefficient de 2,823.542. Cela signifie que pour chaque année supplémentaire de scolarité en moyenne, le PIB par habitant augmente d'environ `2,823.542` $, toutes choses étant égales par ailleurs.

La constante, également appelée l'intercept, est -4,882.718. C'est la valeur du PIB par habitant quand le niveau d'éducation est égale à 0.

En outre, les coefficients de régression sont accompagnés de valeurs p entre parenthèses qui indiquent le niveau de significativité statistique. Dans ce cas, le coefficient de `yr_sch` est significatif au niveau de 1%, ce qui signifie que notre relation obtenue entre la durée moyenne de scolarisation et le PIB par habitant est statistiquement significative. De plus le R2 ajusté est de 0.302, ce qui signifie que le modèle explique environ 30.2% de la variation observée dans le PIB par habitant.

Autrement dit, la durée moyenne de scolarisation explique à elle seule environ 30.2% de la variation dans le PIB par habitant dans l'échantillon.

# VI- Evaluation de la causalité

Ainsi, la régression précédente tend à souligner la corrélation positive entre niveau d'éducation et la productivité du travail. Ces résultats suggèrent que le niveau d'éducation a une influence sur la productivité des travailleurs, que ce soit directement (en leur donnant les outils intellectuels nécessaires à créer, innover et produire davantage notamment dans le secteur des services) ou indirectement (en favorisant l'innovation technique et en améliorant la productivité du capital).

Cela étant, la corrélation précédente n'est pas pour autant nécessairement causale. En effet, on peut estimer que d'autres variables, corrélées à la fois au niveau d'éducation et à la productivité du travail, interviennent dans cette relation : ce sont les biais de variables ommises. Ainsi, nous pouvons conidérer que la tendance plus ou moins libérale de chaque pays (ou bien un régime plus ou moins démocratique) influence à la fois les objectifs de productivité et le niveau d'instruction de la population (liberté d'enseignement et d'apprentissage). Cependant, il est difficile de contrôler pour de telles variables dans notre régression par le seul fait que ces variables sont difficilement quantifiables. Il en est de même pour des variables telles que l'appartenance d'un pays à l'hémisphère Nord ou Sud ou bien l'inégalité initiale de dotation en ressources naturelles du pays. En effet, on peut estimer que plus un pays possède de ressources naturelles, plus son PIB par habitant va être élevé puisque dans leur chaîne de production, le coût de la matière première va être minimisé.

# VII- Robustesse

Quand bien même il semble difficile d'inclure toutes ces variables de contrôle, nous pouvons toujours inclure certaines variables pour tester la robustesse de nos résultats de référence. En effet, si les orientations politiques d'un pays peuvent influencer à la fois le niveau d'éducation moyen et sa productivité par habitant, on soulignera que l'investissement dans le capital fixe (ou formation brute de capital fixe, FBCF) influence elle même positivement à la fois la productivité du travail (en engendrant des moyens de production) et le niveau d'éducation (en investissant dans des infrastructures telles que des écoles d'une part, et en substituant le capital au travail des enfants qui permet davantage d'instruction et de spécialisation dans les domaines du tertiaires d'autre part). Il est donc important de vérifier que nos résultats précédents sont bien robustes à l'inclusion des variables qui peuvent être contrôlées (compte tenu des données disponibles), même si la prise en compte de ces variables ne saurait prouver l'absence de pertinence des résultats précédents.

```
# Mise en page de notre variable de contrôle

gross_fixed_capital_formation <- read.csv("API_NE.GDI.FTOT.CD_DS2_en_csv_v2_44965.csv", sep = ",",

# Nettoyage de notre nouvelle base de données importée
names(gross_fixed_capital_formation) <- gsub("^X\\.", "", names(gross_fixed_capital_formation))
for (i in seq_along(gross_fixed_capital_formation)) {
  gross_fixed_capital_formation[[i]] <- gsub("\"", "", gross_fixed_capital_formation[[i]])
}
names(gross_fixed_capital_formation) <- gsub("\\.", "", names(gross_fixed_capital_formation))

# Supression des variables qui ne vont pas nous être utiles pour constituer notre variable de contr
gross_fixed_capital_formation <- gross_fixed_capital_formation[, !(names(gross_fixed_capital_format

# Nous gardons uniquement les codes à 3 charactères des pays
gfcfperfect <- subset(gross_fixed_capital_formation, nchar(as.character(CountryCode)) == 3)

# Transformation du tableau en format long en regroupant les années dans une colonne "Year"
gfcfperfect2 <- pivot_longer(gfcfperfect, cols = -CountryCode, names_to = "Year", values_to = "gfcf

# Nombre de NA dans notre nouvelle variable gfcf
nblignesgfcf <- nrow(gfcfperfect2)
nbofNAgfcf <- sum(is.na(gfcfperfect2$gfcf) |gfcfperfect2$gfcf == "")
pourcentageofNAgfcf <- round(nbofNAgfcf/nblignesgfcf*100)
```

```r
# Modification nécessaire afin de join notre variable gfcf à notre base de données
gfcfperfect2 = rename(gfcfperfect2, "WBcode"=CountryCode)
donneperfect$Year[2] <- as.character(donneperfect$Year[2])

# On effectue le joint entre donneperfect et notre tableau gfcfperfect2 en fonction de l'année et c
donneperfectgfcf <- donneperfect%>%left_join(gfcfperfect2, by = c("WBcode","Year"))

# On enlève toute les lignes de la colonne gfcf où il y a des NA, rien (des espaces) ou des 0
donneperfectgfcf <- subset(donneperfectgfcf, !is.na(gfcf) & gfcf != "" & gfcf != "0")

# On réordonne pour avoir l'Europe comme variable de référence dans notre régression
donneperfectgfcf$Continent <- factor(donneperfectgfcf$Continent,
                           levels = c("Europe", "Africa","Asia", "North America",
                                      "Oceania","South America"))

# On effectue notre régression multivariée avec notre variable de contrôle cette fois-ci
# On décide de loguer le gfcf et la population puisque certaines données sont trop élevés pour que
donneperfectgfcf$gfcf <- as.numeric(donneperfectgfcf$gfcf)
model1 <- lm(GDPcapita ~ yr_sch, data = donneperfectgfcf)
model2<- lm(GDPcapita ~ yr_sch + log(gfcf), data = donneperfectgfcf)
model3<- lm(GDPcapita ~ yr_sch + log(gfcf)  + Continent, data = donneperfectgfcf)

stargazer(model1, model2, model3, type="text",
          dep.var.labels = c("Label pour modèle 1"),
          covariate.labels = c("Années de scolarité", "log(Formation brute de capital fixe)"))
```

```
================================================================================================
===============
                                                 Dependent variable:
                                    ---------------------------------------------------------
---------------
                                                 Label pour modèle 1
                                          (1)                    (2)
(3)
------------------------------------------------------------------------------------------------
---------------
Années de scolarité                     3,027.808***           2,278.976***
1,869.790***
                                          (119.859)              (143.168)
(166.360)

log(Formation brute de capital fixe)                           1,712.018***
1,535.966***
                                                                 (192.880)
(194.000)

ContinentAfrica
-7,131.966***

(1,331.933)

ContinentAsia
-5,522.668***

(1,123.091)

ContinentNorth America
-6,707.941***
```

```
(1,313.963)

ContinentOceania
-4,185.003*

(2,294.608)

ContinentSouth America
-11,318.740***

(1,471.289)

Constant                                  -7,309.641***           -40,044.970***
-28,025.550***
                                            (942.748)               (3,798.602)
(4,174.324)

-------------------------------------------------------------------------------------------------------
--------------
Observations                                 1,057                   1,057
1,057
R2                                           0.377                   0.420
0.455
Adjusted R2                                  0.376                   0.419
0.452
Residual Std. Error                  12,618.800 (df = 1055)    12,177.850 (df = 1054)
11,832.870 (df = 1049)
F Statistic                         638.143*** (df = 1; 1055) 381.989*** (df = 2; 1054) 125.218***
(df = 7; 1049)
=======================================================================================================
===============
Note:                                                                                 *p<0.1;
**p<0.05; ***p<0.01
```

Ainsi, le coefficient de base diminue de 25 % avec l'inclusion de la variable investissement dans le capital fixe et reste toujours statistiquement différent de 0 à un niveau de confiance de 99%. On peut ainsi valider nos hypothèses sur le fait que l'ommission du capital fixe biaise vers le haut notre coefficient et que ce dernier a bien une influence positive sur le PIB/habitant et le niveau d'éducation.

Par ailleurs, on peut chercher à estimer ce que représente cette diminution de 25 % lorsque l'on inclut l'investissement dans le capital fixe. Pour se rendre compte de la significativité de cette baisse, on cherche à voir quel pays serait le plus proche de la productivité moyenne par habitant francais (par exemple) diminuée de 25 %.

```r
# Calcul du pourcentage de diminution de l'effet
percentdecrease <- (2278.976/3027.808)

# Ici, on prends comme pays de référence la France qui a un gdppercapita de 40998 en 2015
neargdpcountry <- percentdecrease*40998

# Nous cherchons les données pour l'année 2015 car c'est l'année la plus récente
donnee2015 <- subset(donneperfectgfcf, Year == 2015)

# On trouve l'index de la valeur la plus proche de neargdpcountry dans GDPcapita pour l'année 2015
index_plus_proche <- which.min(abs(donnee2015$GDPcapita - neargdpcountry))

# On trouve la valeur la plus proche en 2015 puis ensuite le pays associé
valeur_plus_proche <- donnee2015$GDPcapita[index_plus_proche]
pays_plus_proche <- donnee2015$Country[index_plus_proche]
```

Ainsi on observe que cette baisse de 25% de productivité revient à passer du PIB par habitant de la France en 2015 à celui de la Slovenia en 2015. En somme, cette baisse de 25% de l'effet du niveau d'instruction sur la productivité du travail est assez

significative.

Par ailleurs, en controlant notre régression par les différents continents, on obtient l'équation suivante :

$$1\{GDPpercapita_i\} = \alpha + \beta \times 1\{Year\_school_i\} + \gamma \times 1\{log(gfcf_i)\} + \theta_i \times 1\{Continent_i\} + \varepsilon_i$$

Les coefficients pour les variables catégorielles représentant les continents dans le modèle (3) sont également significatifs, ce qui montre que ces effets sont statistiquement différents de zéro. Le coefficient 1,869.790 pour les années de scolarité dans la colonne (3) signifie qu'une année supplémentaire de scolarité est associée à une augmentation moyenne du PIB par habitant de 1,869.790 unités après contrôle pour les années de scolarité et la formation brute de capital fixe, toutes choses étant égales par ailleurs. Cet effet est un effet moyen global qui s'applique à tous les pays de l'échantillon, et signifie que chaque pays bénéficierait en moyenne d'une augmentation du PIB par habitant avec une année supplémentaire de scolarité.

Par ailleurs, les coefficients des continents montrent quant à eux des différences de niveau de PIB par habitant par rapport à l'Europe , mais ils ne changent pas l'effet positif de l'éducation sur le PIB. Par exemple, les pays d'Afrique ont un PIB par habitant en moyenne plus bas de 7,131.966 unités que ceux d'Europe.

En somme, les résultats indiquent que l'éducation a un effet positif et significatif sur le PIB par habitant pour tous les pays de l'échantillon. Les différences entre les continents reflètent des variations dans les niveaux de PIB par habitant, mais elles ne suggèrent pas que l'éducation est moins bénéfique dans certains continents. Ainsi, les coefficients de base pour `Années de scolarité` et `log(Formation brute de capital fixe)` sont robustes, car ils restent significatifs et de signe attendu à travers différents modèles.

# VIII- Hétérogénéité

Afin d'estimer si la relation calculée diffère d'un continent à un autre, la variable indépendante doit être interragie avec la variable `Continent` , ce qui est équivalent à estimer la régression séparément pour chaque continent.

```
# On contrôle et intéragit progressivement avec le Continent dans la régression
model1 <- lm(GDPcapita ~ yr_sch, data = donneperfectgfcf)
model2<- lm(GDPcapita ~ yr_sch + Continent, data = donneperfectgfcf)
model3<- lm(GDPcapita ~ yr_sch + Continent + yr_sch*Continent, data = donneperfectgfcf)


stargazer(model1, model2, model3, type="text",
          dep.var.labels = c("Label pour modèle 1", "Label pour modèle 2", "Label pour modele 3"),
          covariate.labels = c("Années de scolarité"))
```

```
================================================================================================
========
                                                Dependent variable:
                                    ------------------------------------------------------------
--------
                                                Label pour modèle 1
                                       (1)                  (2)                    (3)
------------------------------------------------------------------------------------------------
--------
Années de scolarité                3,027.808***         2,469.437***
3,869.391***
                                     (119.859)            (152.408)              (393.560)

ContinentAfrica                                         -8,463.471***
10,904.020**
                                                         (1,359.529)
(4,248.364)

ContinentAsia                                           -5,014.359***
```

```
9,970.471**
                                                                (1,153.720)
(4,366.936)

ContinentNorth America                                        −8,012.430***                453.651
                                                                (1,341.336)
(4,740.205)

ContinentOceania                                              −6,067.622***
−41,183.850**
                                                                (2,348.335)
(18,461.850)

ContinentSouth America                                        −12,330.120***               7,667.038
                                                                (1,508.172)
(6,419.880)

yr_sch:ContinentAfrica
−2,703.879***
                                                                                           (507.150)

yr_sch:ContinentAsia
−1,564.660***
                                                                                           (462.568)

yr_sch:ContinentNorth America                                                              −649.841
                                                                                           (516.593)

yr_sch:ContinentOceania                                                                   3,125.292*

(1,700.590)

yr_sch:ContinentSouth America
−2,248.058***
                                                                                           (781.886)

Constant                                −7,309.641***              2,389.073
−11,454.380***
                                          (942.748)              (1,680.609)
(3,959.660)

-------------------------------------------------------------------------------------------------
--------
Observations                                1,057                    1,057                    1,057
R2                                          0.377                    0.423                    0.446
Adjusted R2                                 0.376                    0.419                    0.440
Residual Std. Error           12,618.800 (df = 1055)    12,175.480 (df = 1050)    11,956.510 (df =
1045)
F Statistic                  638.143*** (df = 1; 1055) 128.114*** (df = 6; 1050) 76.447*** (df =
11; 1045)
=================================================================================================
========
Note:                                                                          *p<0.1; **p<0.05;
***p<0.01
```

La troisième colone (3) montre que la pente de la fonction de régression pour la catégorie de référence, l'Europe, équivaut à 3,869.391 $/habitant et est statistiquement différente de 0 à un niveau de confiance de 99%. La différence entre l'effet sur le continent européen et sur les autres continents varie de -2,703.879 (soit un effet de 1,166 pour l'Afrique) à 3,125.292 (un effet de 6,994 pour l'Océanie). Ainsi, plus le coefficient d'intrerraction est proche de 0, plus l'effet estimé est proche de celui estimé sur le continent européen.Par exemple, le fait d'être dans un pays africain fait que l'effet de l'année de scolarité sur le PIB par habitant est moins important de 2,703.879 $/habitant par rapport au même effet sur le continent européen.

Les résultats montrent que l'éducation a un effet positif et significatif sur le PIB par habitant, mais cet effet varie selon les continents. Les coefficients des interactions soulignent l'hétérogénéité des effets de l'éducation selon les continents, soulignant l'importance de prendre en compte les contextes régionaux dans l'analyse des politiques éducatives. Ainsi, on remarquera que le classement relatif de ces effets selon le continent dispose que l'effet du niveau d'éducation sur le PIB/habitant est le moins important en Afrique, puis en Amérique du Sud, puis en Asie et enfin en Amérique du Nord, en Europe puis en Océanie. Il y a donc bien comme prévu un pattern du niveau de développement du continent dans la différence d'effet du niveau d'éducation sur le PIB/habitant entre les différents continents.

Enfin, comme les coefficients d'interaction ne sont pas tous significatifs de la même manière, on peut conclure que ces différences d'effets en fonction du continent sont statistiquement significatives, à l'exception peut-être de l'Amérique du Nord et de l'Océanie : on ne peut pas rejeter l'hypothèse que le coefficient des années de scolarité est le même pour l'Europe et l'Amérique du Nord. On peut tout de même conclure qu'il existe une certaine hétérogénéité des effets selon les continents, notamment pour l'Europe, l'Asie, l'Afrique et l'Amérique du Sud.

# IX- Conclusion

Dans cette analyse, nous avons utilisé les données du PIB par habitant et du niveau d'éducation de 138 pays de tous les continents de 1950 à 2015, afin d'estimer la relation entre le niveau d'instruction et le PIB par habitant national. Cette analyse repose à la fois sur une comparaison statique et géographique, ainsi qu'une évolution dynamique qui permet de comparer les données en fonction des années. Graphiquement, on observe que plus le niveau d'éducation tend à être élevé, plus le PIB/habitant a tendance à être élevé, de même que ces données sont significativement différent d'un continent à un autre.

Les résultats montrent qu'en moyenne, toute chose égale par ailleurs, une augmentation d'une année de scolarisation augmente de 2 824 $/habitant le PIB par tête d'un pays. Cependant, ce résultat pourrait ne pas être interpété comme causal si des variables tel que l'investissement dans le capital fixe masquent des effets positifs ou négatifs à la fois sur le PIB/habitant et sur le niveau d'éducation moyen. De plus, la grande quantité des données dont nous disposons nous ont permis d'avoir des résultats significatifs et potentiellement généralisables à l'ensemble des pays du monde. En outre, les coefficients estimés semblent être relativement robustes lorsqu'on les contrôle par l'investissement en capital fixe et part les différents continents. De plus, les résultats soulignent une hétérogénéité des effets selon les différents continents considérés.

# Références


Mankiw, et al. « A Contribution to the Empirics of Economic Growth ». Quarterly Journal of Economics, vol. 107, no May, 1992, p. 407-437

Krueger, Alan B., and Mikael Lindahl. "Education for Growth: Why and for Whom?" Journal of Economic Literature, vol. 39, no. 4, 2001, pp. 1101-1136

Barro, Robert. The Contribution of Human and Social Capital to Sustained Economic Growth and Well-Being. OECD, 2001

Arnaud Chevalier, Colm Harmon, Ian Walker, Yu Zhu, Does Education Raise Productivity, or Just Reflect it?, The Economic Journal, Volume 114, Issue 499, November 2004, p. 499–517

Education and Economic Growth: It's not Just Going to School but Learning That Matters | Eric A. Hanushek, 2008, p. 62-70

Psacharopoulos, George, et Harry Anthony Patrinos. « Returns to Investment in Education: A Decennial Review of the Global Literature ». Education Economics, vol. 26, no 5, septembre 2018, p. 445-458