

rapport

Vivian Monachon

2024-04-02

Contributeurs : - OUKKAL Yacine - FARINA Alessandro - BOUQUETY Andy - MONACHON Vivian

Problématique : Est-ce que certains indicateurs de développement favorisent la mortalité par maladies infectieuses ?

III) Methodology III.A) Clean up procedures

a) Lecture des fichiers et installation des librairies nécessaires

```
library(tidyverse)
library(dplyr)
library(ggplot2)
library(scales)

annualdeath <- read.csv("../Dataset/annual-number-of-deaths-by-cause.csv", header = TRUE)
governmentEducationExpenditure <-
  read.csv("../Dataset/total-government-expenditure-on-education-gdp.csv", header = TRUE)
enrolment_ratio_tertiary_education <-
  read.csv("../Dataset/enrolment-ratio-tertiary-education.csv", header = TRUE)
waterAccess <- read.csv("../Dataset/water-withdrawals-per-capita.csv")
poptotal <- read.csv("../Dataset/number-of-deaths-per-year.csv")
airPollution <- read.csv("../Dataset/air-pollution.csv", header = TRUE)
cleanWater <- read.csv("../Dataset/share-without-improved-water.csv", header = TRUE)
medics <- read.csv("../Dataset/SYB66_154_202310_Health_Personnel.csv", header = TRUE)
```

b) Enlever les régions en trouvant les entités sans données dans la colonne Code

```
annualdeath <- annualdeath %>% filter(Code != "");
governmentEducationExpenditure <- governmentEducationExpenditure %>% filter(Code != "");
waterAccess <- waterAccess %>% filter(Code != "");

exclude <- c("Entity", "Year")
for (col in names(airPollution)) {
  if (!(col %in% exclude)) {
    airPollution[[col]] <- round(airPollution[[col]])
  }
}
```

c) Changer le nom des colonnes

```

colnames(annualdeath) <- c("Country","Code","Year","Meningitis","Alzheimers disease",
    ,"Parkinsons disease","Nutritional deficiencies","Malaria",
    ,"Drowning","Interpersonal violence", "Maternal disorders",
    ,"HIV AIDS" , "Drug use disorders", "Tuberculosis", "Cardiovascular diseases",
    ,"Lower respiratory infections", "Neonatal disorders","Alcohol use disorders",
    ,"Self harm", "Exposure to forces of nature", "Diarrheal diseases",
    ,"Environmental heat and cold exposure", "Neoplasms", "Conflict and terrorism",
    ,"Diabetes mellitus", "Chronic kidney disease", "Poisonings", "Protein energy",
    ,"Road injuries", "Chronic respiratory diseases", "Cirrhosis and other chronic",
    ,"Digestive diseases","Fire heat and hot substances","Acute hepatitis", "Measles")
)
colnames(governmentEducationExpenditure) <- c("Country","Code","Year", "GPD spent in Education (%)")

colnames(enrolment_ratio_tertiary_education) <- c("Country","Year","Gender","Age", "Units", "Enrolment")

colnames(waterAccess) <- c("Country","Code", "Year", "Water withdrawal per capita ()")

colnames(poptotal) <- c("Country","Code", "Year","Deaths", "Death_mean")

colnames(airPollution) <- c("Nitrogen oxide (NOx)", "Sulphur dioxide (SO2)", "Carbon monoxide (CO)",
    ,"Organic carbon (OC)", "NMVOCs", "Black carbon (BC)", "Ammonia (NH3)",
    ,"s1", "s2", "s3", "s3", "s4", "s5", "s6", "Country", "Year")

airPollution <- airPollution[, c("Country", "Year", names(airPollution)[-which(names(airPollution) %in%

```

d) Supprimer les colonnes inutiles et selectionner celles qui seront pertinentes pour nos analyses

```

annualdeath <- select(annualdeath,-Code, -`Interpersonal violence`, -`Fire heat and hot substances`,
    -`Road injuries`, -`Conflict and terrorism`, -`Exposure to forces of nature`,
    -`Self harm`, -`Alzheimers disease`, -`Parkinsons disease`, -`Drowning`,
    -`Maternal disorders`, -`Drug use disorders`, -`Alcohol use disorders`,
    -`Environmental heat and cold exposure`, -`Poisonings`, -`Neonatal disorders`,
    -`Nutritional deficiencies`, -`Meningitis`)

governmentEducationExpenditure <- select(governmentEducationExpenditure, -Code)

enrolment_ratio_tertiary_education <- select(enrolment_ratio_tertiary_education, -Units, -Age,)

waterAccess <- select(waterAccess, -Code)

poptotal <- select(poptotal,Country, Year, Deaths)

airPollution <- select(airPollution, -s1, -s2, -s3, -s4, -s5, -s6, -s3.1)

airPollution <- subset(airPollution, !(Country %in% c("Asia", "Europe", "Africa",
    "World", "High-income countries", "Upper-middle-income countries",
    "North America", "Lower-middle-income countries",
    "Oceania", "Low-income countries", "Gibraltar", "Cayman Islands")))

Polluants <- rowSums(airPollution[, -(1:2)])
airPollution$Polluants <- Polluants
airPollution <- airPollution[, c("Country", "Year", "Polluants")]

```

- e) Sélection d'une seule catégorie de maladie : maladies infectieuses Nous aurions eu trop de graphes et données à analyser si nous avions sélectionné tous les différents types de maladies Unifier certaines maladies sous la même catégorie

```
annualdeath$`Infectious diseases` <- annualdeath$`Malaria` + annualdeath$`Tuberculosis` +
  annualdeath$`HIV AIDS` + annualdeath$`Measles` + annualdeath$`Diarrheal diseases`

annualdeath <- select(annualdeath, Country, Year, `Infectious diseases`)
```

- f) Filter seulement les données de 1990 à 2010

```
annualdeath <- annualdeath %>% filter(Year<2011);

governmentEducationExpenditure <- governmentEducationExpenditure %>% filter(Year<2011) %>%filter(Year>1989)

enrolment_ratio_tertiary_education <- enrolment_ratio_tertiary_education %>% filter(Year<2011) %>%
  filter(Year>1989) %>%filter(Gender == "All genders") ;

waterAccess <- waterAccess %>% filter(Year<2011) %>%filter(Year>1989)

poptotal <- poptotal %>% filter(Year<2011) %>%filter(Year>1989)

airPollution <- airPollution %>% filter(Year<2011) %>%filter(Year>1989) %>% filter(Polluants > 10000)

cleanWater <- cleanWater %>% filter(Year < 2011) %>% filter(Code != "") %>%
  filter(!Entity %in% c("Gibraltar", "Sint Maarten (Dutch part)", "Cayman Islands"))

cleanWater <- select(cleanWater, -`Code`)
```

```
colnames(cleanWater) <- c("Country", "Year", "Proportion of population unable to access clean water (%)")
cleanWater$`Proportion of population unable to access clean water (%)` <-
  round((cleanWater$`Proportion of population unable to access clean water (%)`), 2)
```

- g) Enlever les pays ayant moins de 20% de données et la colonne "World" Cependant dans la table waterAccess nous avons des données qui représentent le retrait d'eau par tête seulement tous les 5 ans, nous avons donc décidé que nous garderons seulement les pays ayant la totalité des données sur les 20 années de nos analyses

```
governmentEducationExpenditure <- governmentEducationExpenditure %>% group_by(Country) %>%
  filter(n() > 5) %>% filter(Country != "World")

annualdeath <- annualdeath %>% group_by(Country) %>% filter(n() > 5) %>% filter(Country != "World")

enrolment_ratio_tertiary_education <- enrolment_ratio_tertiary_education %>%
  group_by(Country) %>% filter(n() > 5) %>% filter(!grepl("^MDG_", Country))

waterAccess <- waterAccess %>% group_by(Country) %>% filter(n() == 5)
```

- h) Merge la table poptotal et annualdeath pour obtenir la part des morts des maladies infectieuses par pays sur le nombre total de morts par pays

```
annualdeath <- merge(annualdeath, poptotal[, c("Country", "Year", "Deaths")], by = c("Country", "Year"))
annualdeath$`Infectious diseases prop (%)` <- annualdeath$`Infectious diseases` / annualdeath$Deaths *
```

III.B) Scientific workflow

Nous avons initialement l'intention de segmenter nos données par régions et de les présenter sous forme de graphiques à points. Cependant, face à plusieurs échecs lors de l'utilisation de graphiques linéaires, nous avons choisi de nous concentrer sur l'analyse des cinq pays présentant les valeurs les plus extrêmes pour chaque variable dans nos tableaux. Nous avons comparé ces pays avec leurs taux respectifs de mortalité due aux maladies infectieuses, ce qui a généré six ensembles de données distincts et, par conséquent, six hypothèses potentielles sur les origines des maladies infectieuses.

Afin d'optimiser la clarté et la compréhension de notre analyse, nous avons calculé les moyennes pour chaque variable et chaque pays sélectionné. Cette démarche a simplifié la présentation des résultats, nous permettant ainsi de formuler des hypothèses de manière plus aisée.

Nous testerons 6 hypothèses: -Il existe une corrélation négative entre les dépenses en éducation en pourcentage du PIB et la proportion de décès dus aux maladies infectieuses. C'est-à-dire que les pays qui investissent davantage dans l'éducation ont tendance à avoir un pourcentage plus faible de décès dus à des maladies infectieuses.

-Une plus forte proportion d'inscription universitaire est associée à un pourcentage plus faible de décès dus à des maladies infectieuses. Cela pourrait indiquer que les pays ayant un taux d'inscription élevé dans les établissements éducatifs bénéficient également de meilleurs résultats en termes de santé publique, potentiellement grâce à une meilleure éducation sur les questions de santé et l'accès à des informations sur la prévention et le traitement des maladies infectieuses

-Les pays avec un retrait d'eau par habitant plus élevé connaissent une proportion plus faible de décès dus aux maladies infectieuses. Cela peut suggérer que des ressources en eau plus abondantes et plus accessibles contribuent à une meilleure hygiène et à des conditions sanitaires qui réduisent la transmission des maladies infectieuses.

-Il existe une corrélation positive entre la proportion de la population n'ayant pas accès à l'eau courante et la proportion de décès dus aux maladies infectieuses. Cela signifie que les pays où une plus grande partie de la population manque d'accès à l'eau courante ont tendance à avoir des taux plus élevés de mortalité liée aux maladies infectieuses

-Il existe une relation entre la quantité de gaz toxiques polluants émis dans un pays et la proportion de décès dus à des maladies infectieuses. Les pays qui émettent une grande quantité de polluants atmosphériques pourraient avoir une incidence plus élevée de certaines maladies infectieuses, en raison de l'affaiblissement de la santé respiratoire de la population et de la création d'environnements propices à la propagation de maladies infectieuses.

-Les pays avec un plus grand nombre de personnel médical par rapport à leur population ont une proportion plus faible de décès dus aux maladies infectieuses, indiquant que l'accès à un personnel médical suffisant peut être un facteur important dans la lutte contre les maladies infectieuses.

- a) Ce programme nous permet de trouver les pays avec les pourcentages les plus et moins élevés en moyenne sur 20 ans dans la table governmentEducationExpenditure

```
governmentEducationExpenditure_mean <- governmentEducationExpenditure %>%
  group_by(Country) %>%
  summarise(Mean_GPD_spent = mean(`GPD spent in Education (%)`))

governmentEducationExpenditure_valmax <- governmentEducationExpenditure_mean %>% top_n(5, wt = `Mean_GPD_spent`)

governmentEducationExpenditure_valmin <- governmentEducationExpenditure_mean %>% slice_min(`Mean_GPD_spent`)
```

```
governmentEducationExpenditure_minmax <- rbind(governmentEducationExpenditure_valmin, governmentEducationExpenditure_valmax)
governmentEducationExpenditure_minmax$Country <- factor(governmentEducationExpenditure_minmax$Country, levels = governmentEducationExpenditure_valmin$Country)
```

b) Une fois ces pays trouvés on les selectionne, au sein de la table annualdeath, les données de ces pays

```
pays_h1_death <- select(annualdeath, Country, Year, `Infectious diseases prop (%)`) %>%
  filter(Country %in% c("Cuba", "Denmark", "Djibouti", "Lesotho", "Norway", "Monaco",
    "Sudan", "Dominican Republic", "Central African Republic", "Cambodia"))

pays_h1_death_mean <- pays_h1_death %>%
  group_by(Country) %>%
  summarise(Mean_death_ID = mean(`Infectious diseases prop (%)`))

death_valmax <- pays_h1_death_mean %>% top_n(5, wt = `Mean_death_ID`)

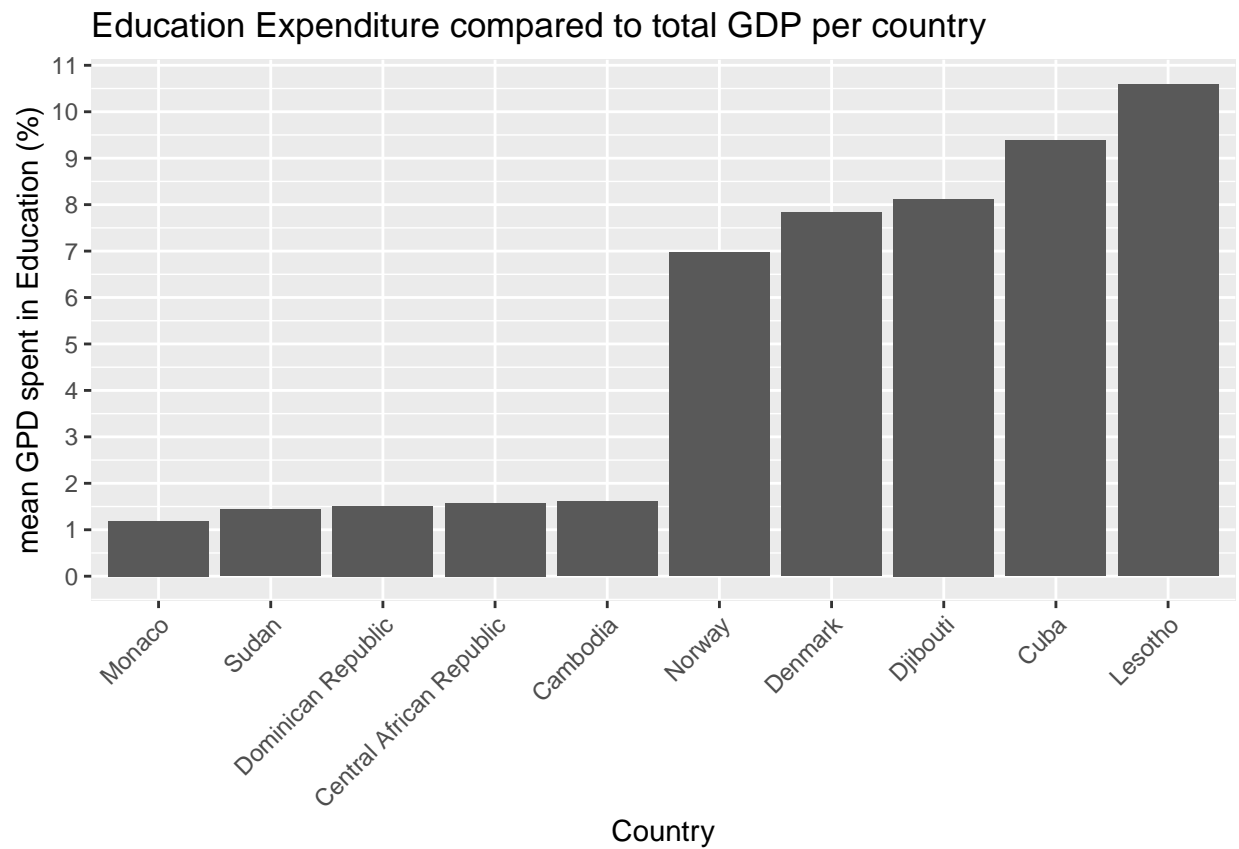
death_valmin <- pays_h1_death_mean %>% slice_min(`Mean_death_ID`, n = 5)

death_minmax <- rbind(death_valmin, death_valmax)

death_minmax$Country <- factor(death_minmax$Country, levels = death_minmax$Country[order(death_minmax$`Mean_death_ID`)])
```

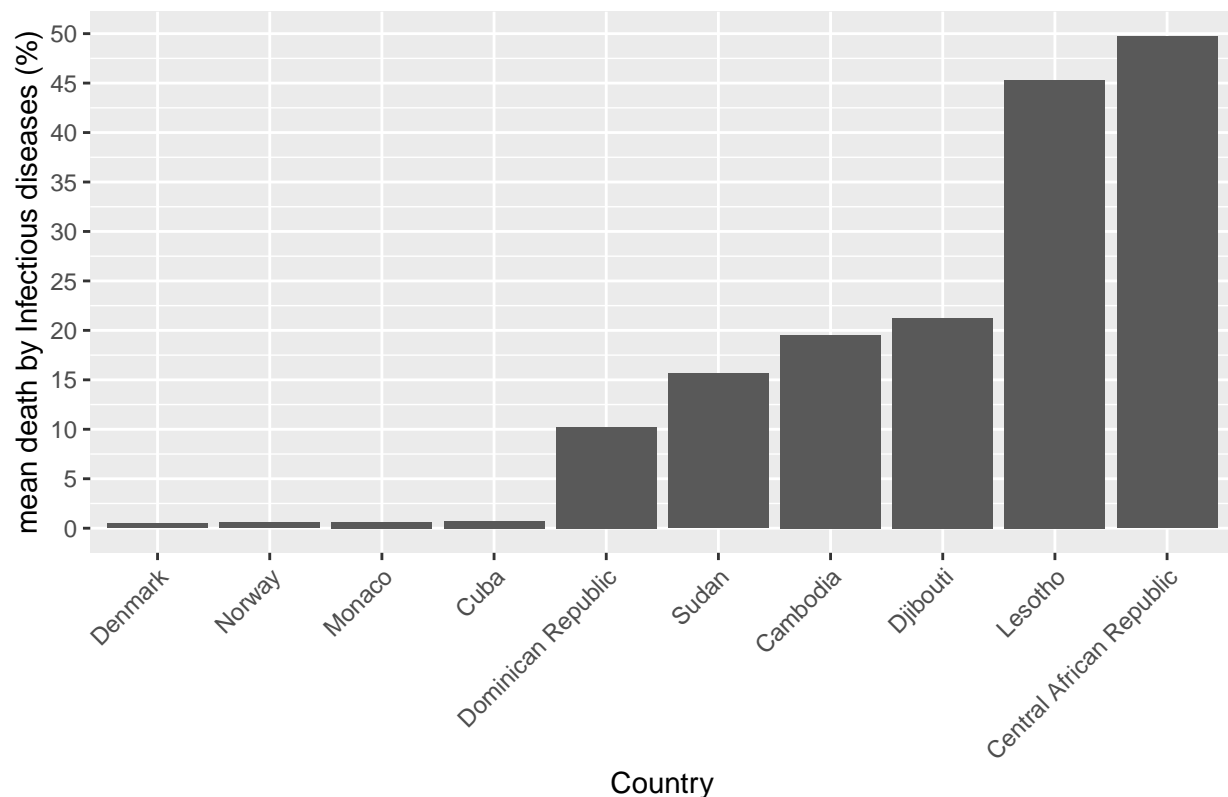
nous pouvons analyser les données de ces pays à l'aide de graphes à barres.

```
ggplot(governmentEducationExpenditure_minmax, aes(x= Country, y= `Mean_GPD_spent`)) +
  geom_bar(stat = "identity") +
  labs(title="Education Expenditure compared to total GDP per country", x="Country", y="mean GPD spent") +
  scale_y_continuous(limits = c(0, NA), breaks = pretty_breaks(n = 10), labels = comma_format()) +
  scale_fill_discrete(name = "Country") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```
ggplot(death_minmax, aes(x= Country, y= `Mean_death_ID`)) +
  geom_bar(stat = "identity") +
  labs(title="Infectious diseases per total number of death(10 countries education exp", x="Country", y=
  scale_fill_discrete(name = "Country") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Infectious diseases per total number of death(10 countries education exp



Après visualisation des données, nous ne pouvons conclure de la corrélation entre dépenses éducationnelles et la santé publique de nombreux pays divergent de cette hypothèse

Maintenant nous nous pencherons sur le jeu de données `enrolment_ratio_tertiary_education`, nous procéderons à la même méthodologie que faite précédemment. trouver les pays avec les pourcentages les plus et moins élevés en moyenne sur 20 ans

```
enrolmentratio_mean <- enrolment_ratio_tertiary_education %>%
  group_by(Country) %>%
  summarise(Enrolment_ratio = mean(`Enrolment ratio (%)`))

enrolmentratio_mean <- enrolmentratio_mean %>% filter(Enrolment_ratio > 0.1)

enrolmentratio_valmax <- enrolmentratio_mean %>% top_n(5, wt = Enrolment_ratio)

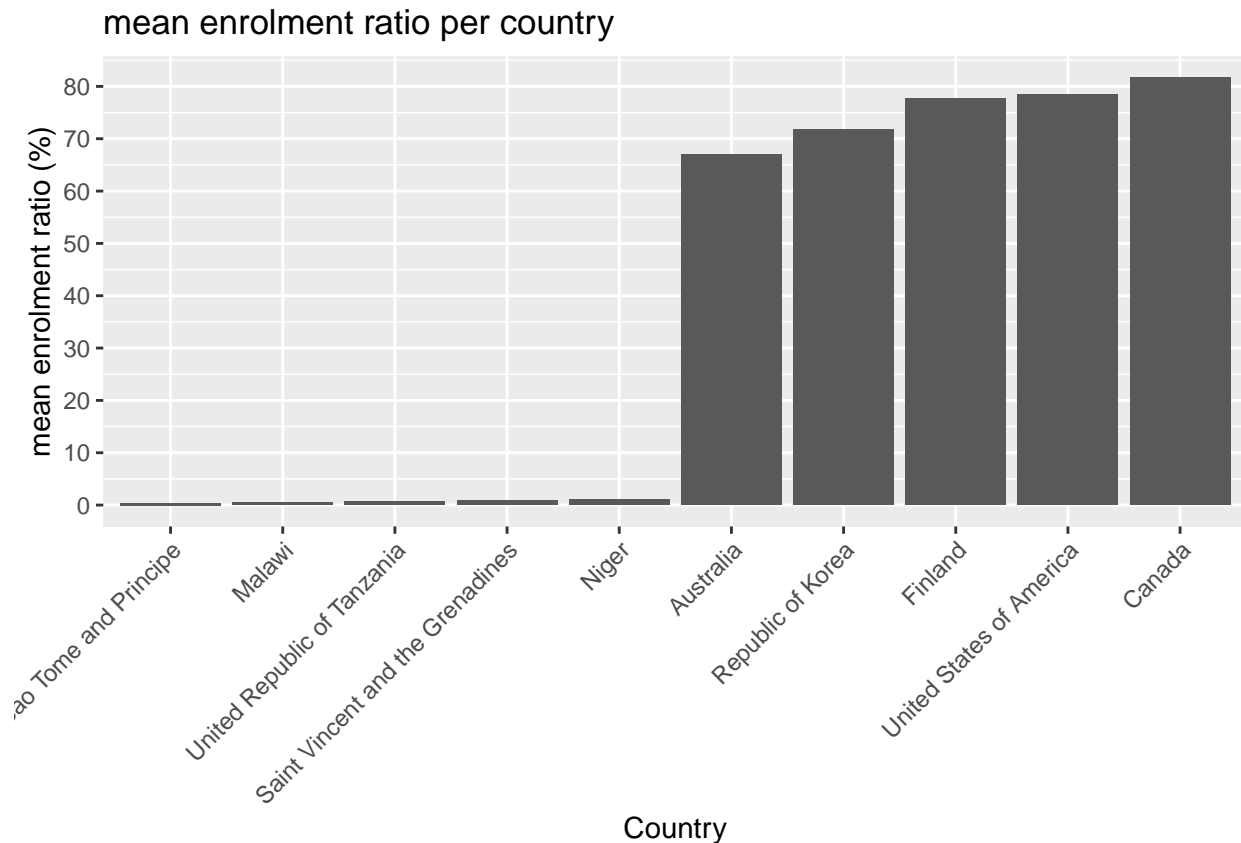
enrolmentratio_valmin <- enrolmentratio_mean %>% slice_min(Enrolment_ratio, n = 5)

enrolmentratio_minmax <- rbind(enrolmentratio_valmin, enrolmentratio_valmax)

enrolmentratio_minmax$Country <-
  factor(enrolmentratio_minmax$Country, levels =
    enrolmentratio_minmax$Country[order(enrolmentratio_minmax$Enrolment_ratio)])

ggplot(enrolmentratio_minmax, aes(x= Country, y= `Enrolment_ratio`)) +
  geom_bar(stat = "identity") +
  labs(title="mean enrolment ratio per country", x="Country", y="mean enrolment ratio (%)") +
```

```
scale_y_continuous(limits = c(0, NA), breaks = pretty_breaks(n = 10), labels = comma_format()) +
scale_fill_discrete(name = "Country") +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



On selectionne au sein de la table principale les données de ces pays

```
pays_h2_death <- select(annualdeath, Country, Year, `Infectious diseases prop (%)`) %>%
  filter(Country %in% c("Sao Tome and Principe", "Malawi", "Tanzania",
    "Saint Vincent and the Grenadines", "Niger", "Australia", "Canada",
    "Finland", "North Korea", "United States"))

pays_h2_death_mean <- pays_h2_death %>%
  group_by(Country) %>%
  summarise(Mean_death_ID = mean(`Infectious diseases prop (%)`))

death_valmax2 <- pays_h2_death_mean %>% top_n(5, wt = `Mean_death_ID`)

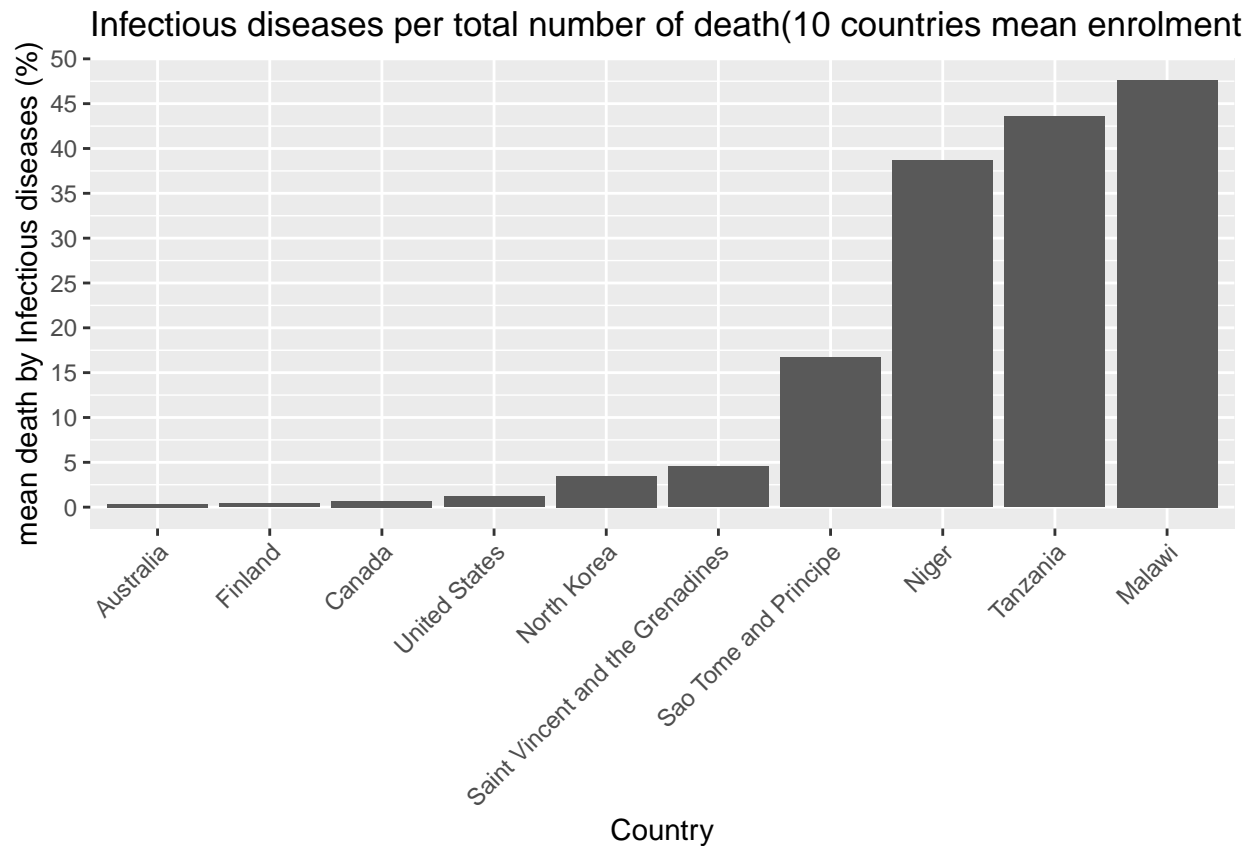
death_valmin2 <- pays_h2_death_mean %>% slice_min(`Mean_death_ID`, n = 5)

death_minmax2 <- rbind(death_valmin2, death_valmax2)

death_minmax2$Country <-
  factor(death_minmax2$Country, levels = death_minmax2$Country[order(death_minmax2$`Mean_death_ID`)])
```



```
ggplot(death_minmax2, aes(x= Country, y= `Mean_death_ID`)) +
  geom_bar(stat = "identity") +
  labs(title="Infectious diseases per total number of death(10 countries mean enrolment", x="Country", y="Mean_death_ID") +
  scale_fill_discrete(name = "Country") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Les graphiques suggèrent une tendance selon laquelle une éducation plus accessible et plus répandue pourrait coïncider avec de meilleurs résultats de santé en ce qui concerne les maladies infectieuses

Maintenant, nous combiennons les données des table waterAccess, représentant le nombre de m3 débité par habitant, tout en sélectionnant les pays les plus extremes

```
waterAccess_mean <- waterAccess %>%
  group_by(Country) %>%
  summarise(water_access = mean(`Water withdrawal per capita ()`))

waterAccess_mean <- waterAccess_mean %>% filter(water_access > 0.1)

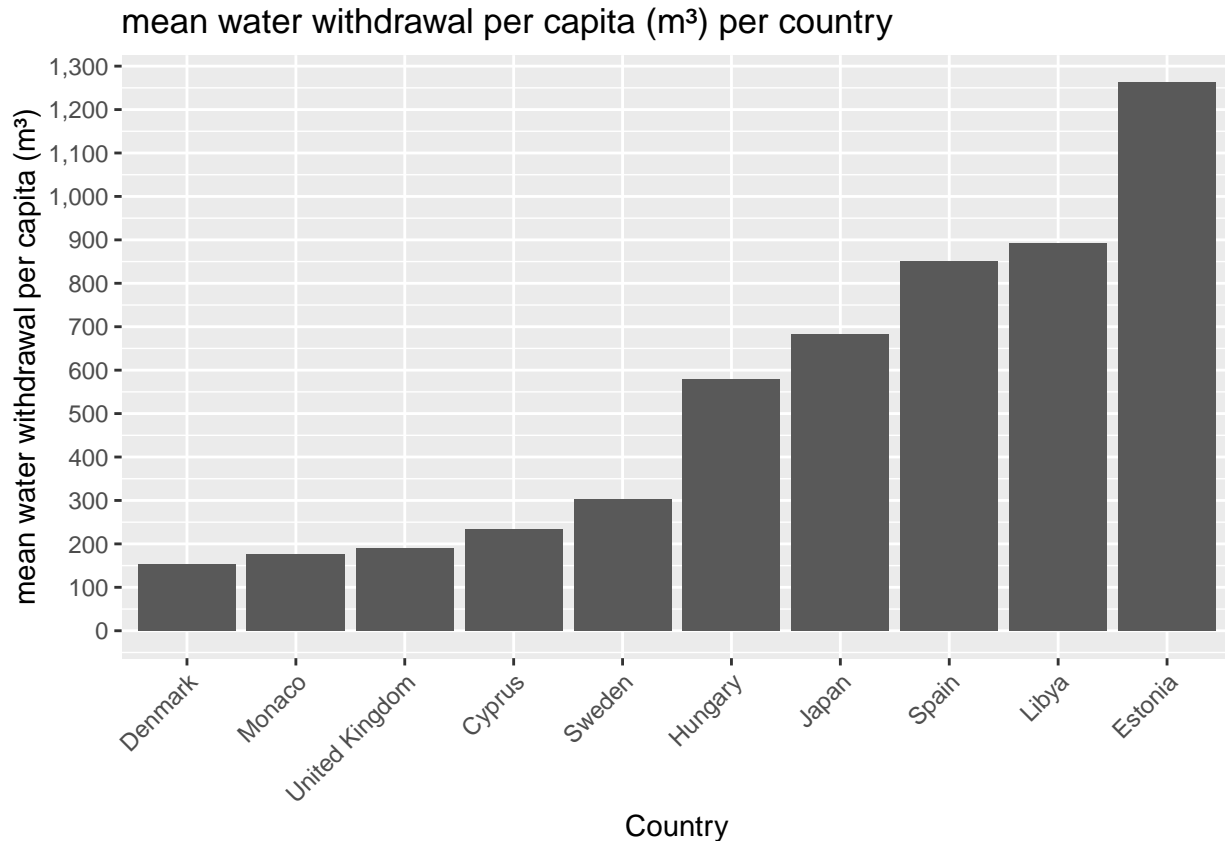
waterAccess_valmax <- waterAccess_mean %>% top_n(5, wt = water_access)

waterAccess_valmin <- waterAccess_mean %>% slice_min(water_access, n = 5)

waterAccess_minmax <- rbind(waterAccess_valmin, waterAccess_valmax)

waterAccess_minmax$Country <-
  factor(waterAccess_minmax$Country, levels = waterAccess_minmax$Country[order(waterAccess_minmax$water_access)])
```

```
ggplot(waterAccess_minmax, aes(x= Country, y= `water_access`)) +
  geom_bar(stat = "identity") + labs(title="mean water withdrawal per capita (m³) per country ", x="Country") +
  scale_fill_discrete(name = "Country") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```
pays_h3_death <- select(annualdeath, Country, Year, `Infectious diseases prop (%)`) %>%
  filter(Country %in% c("Denmark", "Monaco", "United Kingdom", "Cyprus", "Sweden",
    "Estonia", "Hungary", "Japan", "Libya", "Spain"))

pays_h3_death_mean <- pays_h3_death %>%
  group_by(Country) %>%
  summarise(Mean_death_ID = mean(`Infectious diseases prop (%)`))

death_valmax3 <- pays_h3_death_mean %>% top_n(5, wt = `Mean_death_ID`)

death_valmin3 <- pays_h3_death_mean %>% slice_min(`Mean_death_ID`, n = 5)

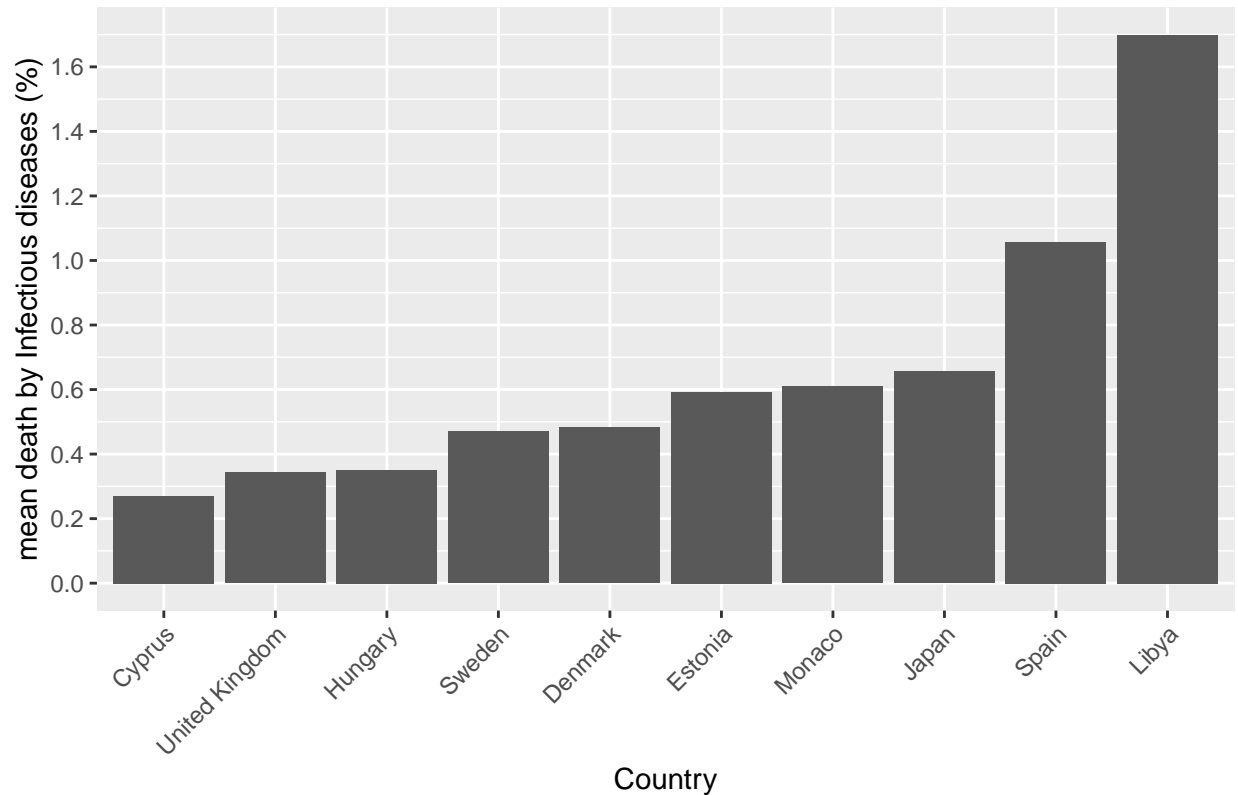
death_minmax3 <- rbind(death_valmin3, death_valmax3)

death_minmax3$Country <-
  factor(death_minmax3$Country, levels = death_minmax3$Country[order(death_minmax3$`Mean_death_ID`)])

ggplot(death_minmax3, aes(x= Country, y= `Mean_death_ID`)) +
  geom_bar(stat = "identity") +
  labs(title="Infectious diseases per total number of death(10 countries mean water withdrawal", x="Country")
```

```
scale_fill_discrete(name = "Country") +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Infectious diseases per total number of death(10 countries mean water with



Il ne semble pas y avoir une corrélation directe et uniforme entre le retrait d'eau par habitant et le pourcentage de décès dus aux maladies infectieuses sur la base des pays présentés. Par exemple, tandis que la Libye présente à la fois le retrait d'eau le plus élevé et le pourcentage le plus élevé de décès dus aux maladies infectieuses, l'Espagne a également un retrait d'eau élevé mais un pourcentage beaucoup plus faible de décès dus aux maladies infectieuses

Analyse : Clean Water access Trouver les 10 pays dont la population a le plus accès à l'eau courante et les 10 pays dont la population a le moins accès à l'eau courante

```
cleanWater_mean <- cleanWater %>%
  group_by(Country) %>%
  summarise(`Proportion of population unable to access clean water (%)` =
    mean(`Proportion of population unable to access clean water (%)`))

cleanWater_max <- cleanWater_mean %>%
  top_n(wt = `Proportion of population unable to access clean water (%)`, n = 10)

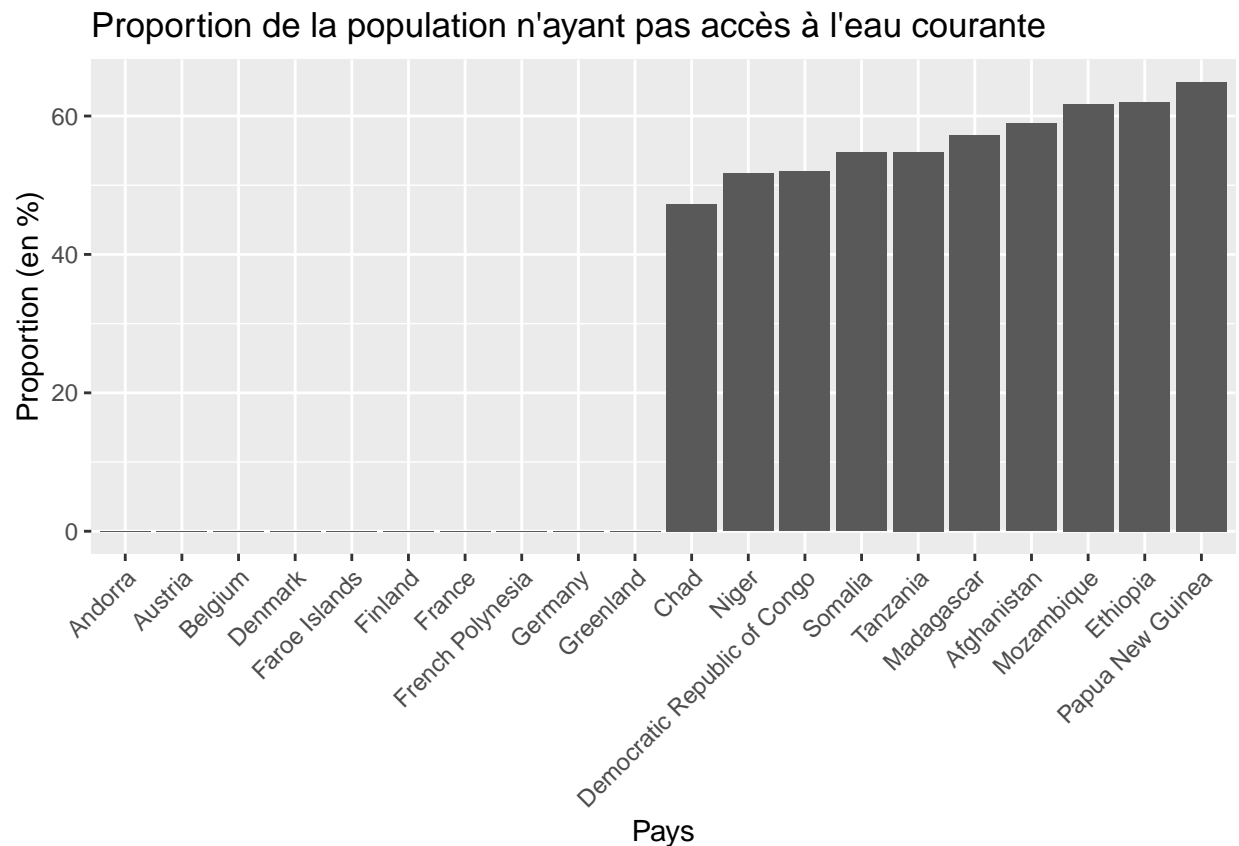
cleanWater_min <- cleanWater_mean %>%
  slice_min(`Proportion of population unable to access clean water (%)`, n = 10) %>% slice_head(n = 10)

cleanWater_minmax <- rbind(cleanWater_min, cleanWater_max)

cleanWater_minmax$Country <-
```

```
factor(cleanWater_minmax$Country, levels =
      cleanWater_minmax$Country[order(cleanWater_minmax$`Proportion of population unable to access
```

```
ggplot(cleanWater_minmax, aes(x= Country, y=`Proportion of population unable to access clean water (%)`
  geom_bar(stat = "identity")+
  labs(title="Proportion de la population n'ayant pas accès à l'eau courante", x="Pays", y="Proportion
  scale_fill_discrete(name = "Country") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Fusion avec la table des morts afin de filtrer les données

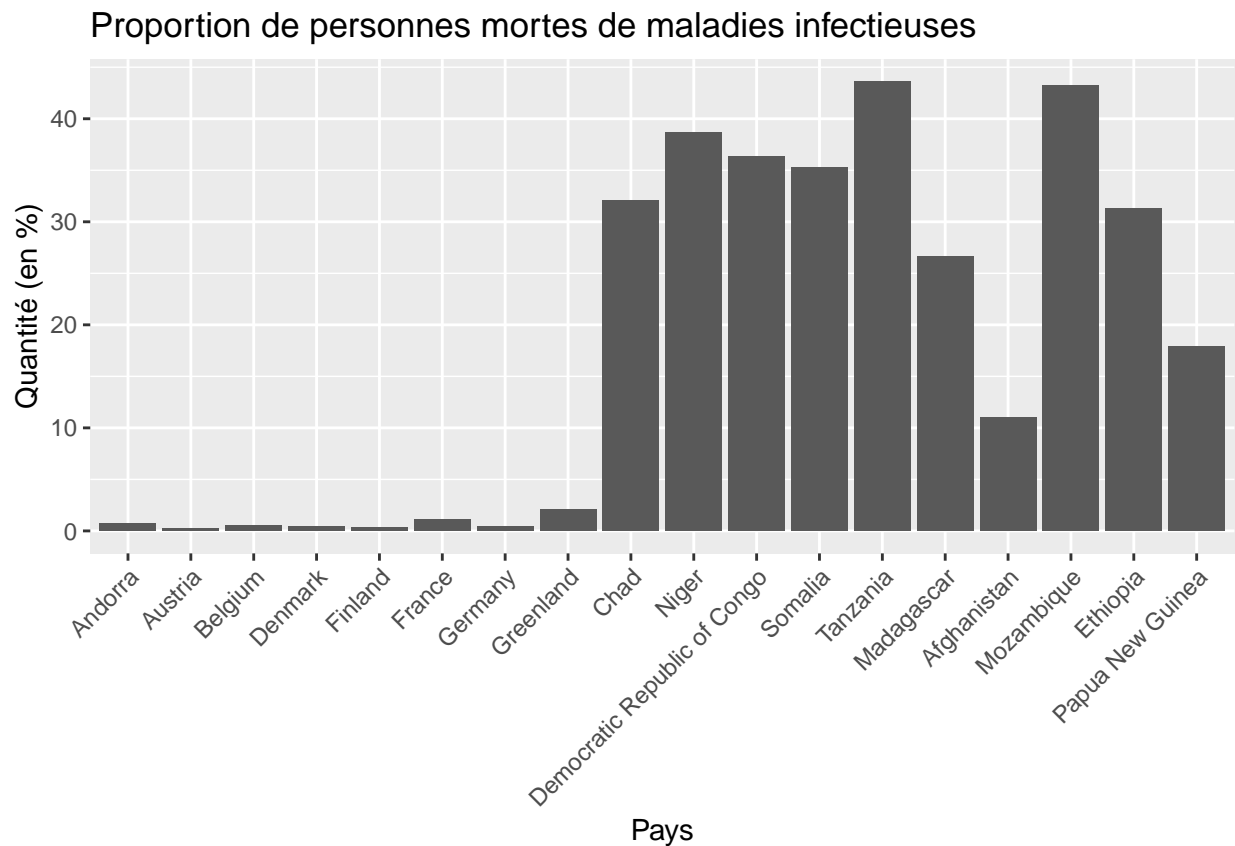
```
ProportionCleanWater <-
  merge(cleanWater_minmax, annualdeath[c("Country", "Infectious diseases prop (%)")], by = "Country")

ProportionCleanWater <-
  ProportionCleanWater %>% group_by(Country) %>%
  summarize(`Infectious diseases %` = mean(`Infectious diseases prop (%)`), `Proportion of population u
    mean(`Proportion of population unable to access clean water (%)`))
```

Affichage du nombre de morts des pays concernés

```
ggplot(ProportionCleanWater, aes(x= Country, y=`Infectious diseases %`)) +
  geom_bar(stat = "identity") +
  labs(title="Proportion de personnes mortes de maladies infectieuses", x="Pays", y="Quantité (en %)") +
  scale_y_continuous(labels = comma_format()) +
```

```
scale_fill_discrete(name = "Country") +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Il apparaît que les pays avec un accès limité à l'eau courante ont tendance à présenter des taux plus élevés de mortalité due aux maladies infectieuses. Ceci est conforme à l'idée que l'accès à l'eau propre est un facteur important dans la prévention et le contrôle des maladies infectieuses.

Analyse : Air Pollution Trouver les 10 pays qui polluent le plus et les 10 pays qui polluent le moins

```
Polluants_mean <- airPollution %>%
  group_by(Country) %>%
  summarise(Polluants = sum(`Polluants`))

Polluants_max <- Polluants_mean %>% top_n(10, wt = `Polluants`)

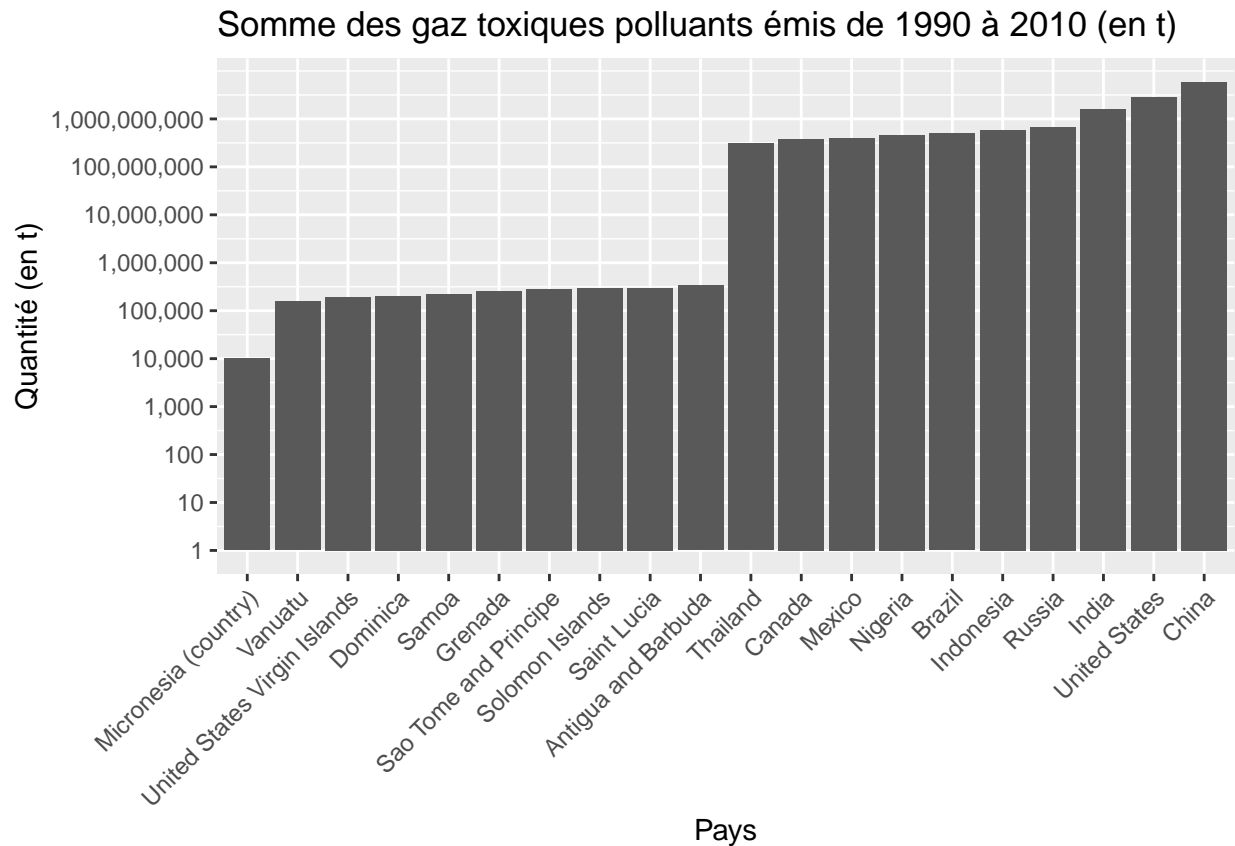
Polluants_min <- Polluants_mean %>% slice_min(`Polluants`, n = 10)

Polluants_minmax <- rbind(Polluants_min, Polluants_max)

Polluants_minmax$Country <-
  factor(Polluants_minmax$Country, levels = Polluants_minmax$Country[order(Polluants_minmax$Polluants)])

ggplot(Polluants_minmax, aes(x= Country, y= Polluants)) +
  geom_bar(stat = "identity") +
  labs(title="Somme des gaz toxiques polluants émis de 1990 à 2010 (en t)", x="Pays", y="Quantité (en t)") +
  scale_y_log10(breaks = 10^(0:9), labels = comma_format()) +
```

```
scale_fill_discrete(name = "Country") +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



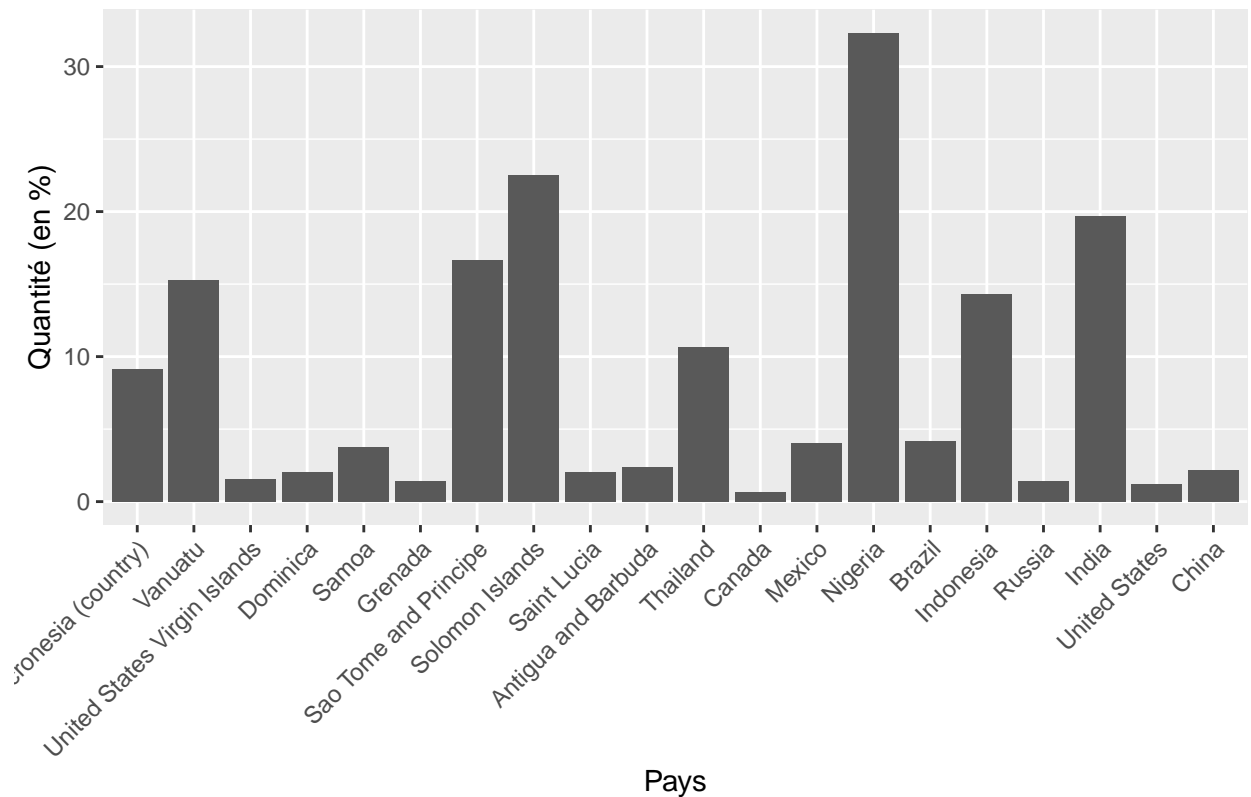
Fusion avec la table des morts afin de filtrer les données

```
ProportionPolluants <-
  merge(Polluants_minmax, annualdeath[c("Country", "Infectious diseases prop (%)")], by = "Country")
ProportionPolluants <-
  ProportionPolluants %>%
  group_by(Country) %>%
  summarize(`Infectious diseases %` = mean(`Infectious diseases prop (%)`), `Polluants` = mean(`Polluants`))
```

Affichage du nombre de morts des pays concernés

```
ggplot(ProportionPolluants, aes(x= Country, y=`Infectious diseases %`)) +
  geom_bar(stat = "identity") +
  labs(title="Proportion de personnes mortes de maladies infectieuses", x="Pays", y="Quantité (en %)") +
  scale_y_continuous(labels = comma_format()) +
  scale_fill_discrete(name = "Country") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Proportion de personnes mortes de maladies infectieuses



Il n'est pas évident qu'il existe une corrélation directe et cohérente entre les émissions totales de gaz toxiques et la proportion de décès dus à des maladies infectieuses pour les pays représentés. Par exemple, bien que la Chine et les États-Unis aient des émissions très élevées, la proportion de décès dus aux maladies infectieuses n'est pas la plus haute par rapport à d'autres pays avec des émissions plus faibles

Analyse : Medics Presence Formatage des données

```
medics <- medics %>% filter(X < 2011) %>%
  filter(row_number() %% 2 == 1)

medics <- select(medics, -`T10`, -`X.1`, -`X.3`, -`X.4`)

colnames(medics) <- c("Country", "Year", "Quantity")

medics$Quantity <- as.integer(gsub(",", "", medics$Quantity))

medics <- medics %>% group_by(Country) %>% summarise(`Quantity` = sum(`Quantity`))
```

Trouver les 10 pays avec la plus forte présence médicale et les 10 pays avec le moins de présence médicale

```
medics <- medics %>% group_by(Country) %>%
  summarise(`Quantity` = sum(`Quantity`))

medics_max <- medics %>%
  top_n(wt = Quantity, n = 20) %>% slice_head(n = 20)

medics_min <- medics %>%
```

```

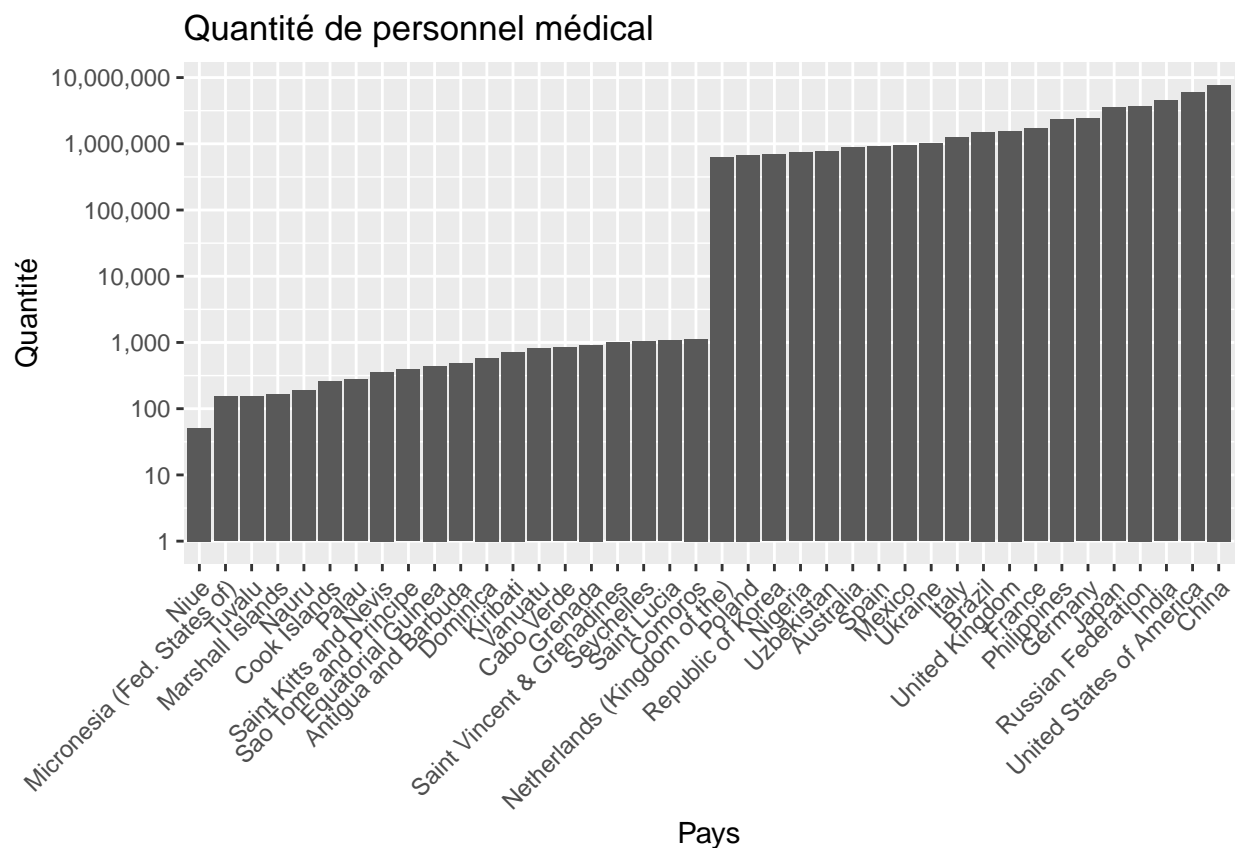
slice_min(Quantity, n = 20) %>% slice_head(n = 20)

medics_minmax <- rbind(medics_min, medics_max)

medics_minmax$Country <-
  factor(medics_minmax$Country, levels = medics_minmax$Country[order(medics_minmax$Quantity)])

ggplot(medics_minmax, aes(x= Country, y=`Quantity`)) +
  geom_bar(stat = "identity") +
  labs(title="Quantité de personnel médical", x="Pays", y="Quantité") +
  scale_fill_discrete(name = "Country") +
  scale_y_log10(breaks = 10^(0:9), labels = comma_format()) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```



Fusion avec la table des morts afin de filtrer les données

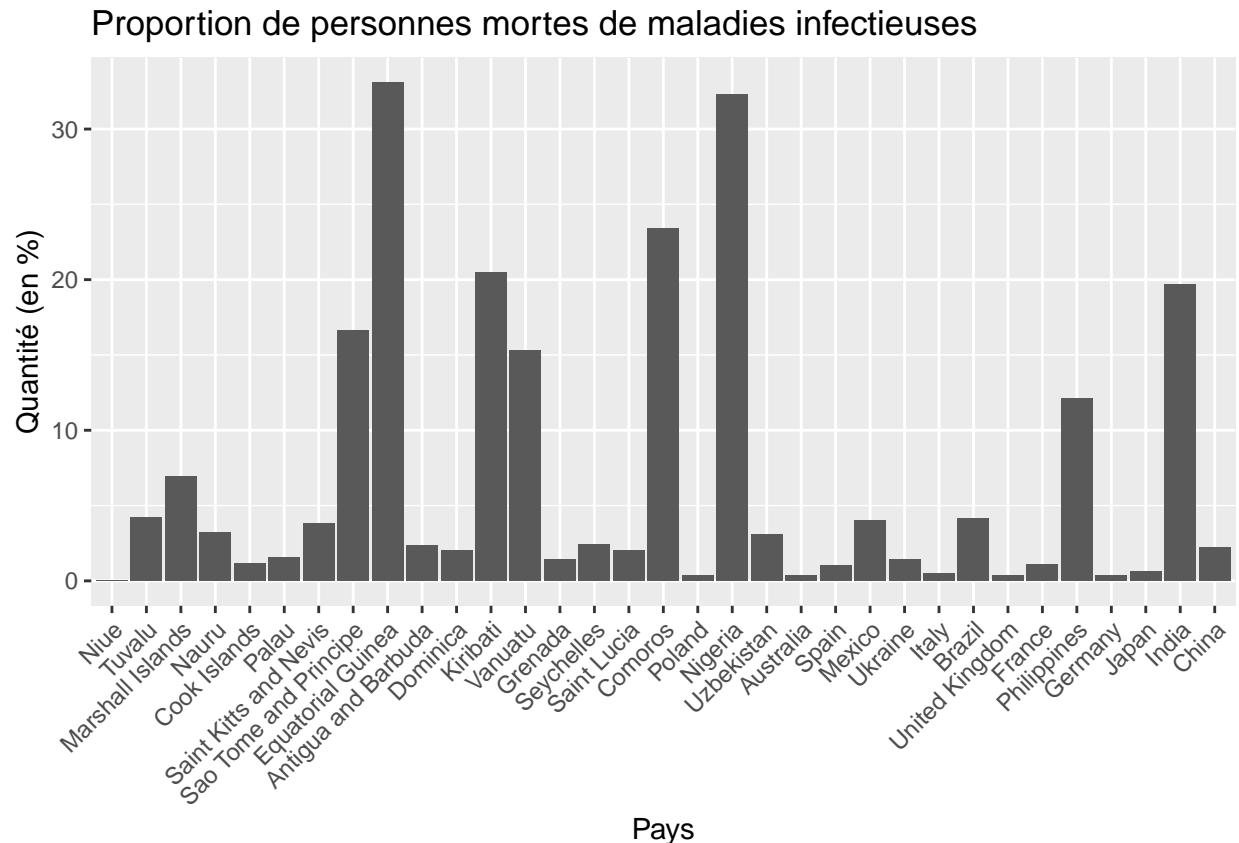
```

ProportionMedics <-
  merge(medics_minmax, annualdeath[c("Country", "Infectious diseases prop (%)")], by = "Country")
ProportionMedics <-
  ProportionMedics %>%
  group_by(Country) %>%
  summarize(`Infectious diseases %` = mean(`Infectious diseases prop (%)`), `Quantity` = mean(`Quantity`))

```

Affichage du nombre de morts des pays concernés


```
ggplot(ProportionMedics, aes(x= Country, y=`Infectious diseases %`)) +
  geom_bar(stat = "identity") +
  labs(title="Proportion de personnes mortes de maladies infectieuses", x="Pays", y="Quantité (en %)") +
  scale_y_continuous(labels = comma_format()) +
  scale_fill_discrete(name = "Country") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



- résultat : Les pays avec un plus grand nombre de personnel médical par rapport à leur population ont une proportion plus faible de décès dus aux maladies infectieuses, indiquant que l'accès à un personnel médical suffisant peut être un facteur important dans la lutte contre les maladies infectieuses.

Conclusion Générale : L'analyse des données disponibles suggère que les indicateurs de développement peuvent avoir un impact complexe et multidimensionnel sur les taux de mortalité dus aux maladies infectieuses. Les dépenses en éducation et les taux d'inscription à l'enseignement supérieur peuvent améliorer la sensibilisation aux questions de santé et favoriser des comportements qui réduisent la transmission des maladies infectieuses. Cependant, les données ne montrent pas une corrélation systématique directe entre les dépenses en éducation et les taux de mortalité infectieuse, indiquant que la qualité de l'éducation et son contenu en matière de santé publique pourraient être des facteurs déterminants.

De même, l'accès à l'eau et la qualité de l'eau semblent être des facteurs significatifs, comme en témoigne l'association entre l'accès limité à l'eau propre et des taux de mortalité plus élevés dus aux maladies infectieuses. Cela souligne l'importance de l'infrastructure sanitaire et des ressources en eau dans la prévention des maladies.

En ce qui concerne la pollution atmosphérique, bien que nous puissions intuitivement supposer que la pollution contribue à une mauvaise santé et potentiellement à une incidence accrue de maladies infectieuses,

les données ne fournissent pas une corrélation claire pour étayer cette hypothèse. Cela pourrait être dû à des facteurs atténuants, tels que l'efficacité des services de santé publics et les réponses aux crises de santé publique.

Quant à la présence de personnel médical, il semble que les pays avec un plus grand nombre de personnel médical soient associés à une proportion plus faible de décès dus aux maladies infectieuses. Cela met en évidence le rôle crucial du personnel médical dans la prévention, le diagnostic et le traitement des maladies infectieuses.