

# Causes of death in the world

N'Djo-Soro Coulibaly, Ciré Keïta, Higo Aragao, Thi Tran

2024-03-05

## Contexte du problème

Il est question d'analyser des données concernant les causes de mortalité dans le monde. Pour mener à bien cette analyse, nous avons défini des questions sur lesquelles travailler: 1) Représentation sur carte graphique des décès en fonction des Causes et par année 2) Les 5 principales causes de mortalité dans des sous-régions sélectionnées 3) L'évolution des décès par rapport à la démographie des régions de 1990 à 2019 4) Nuage de point présentant une région à une date donnée, le taux de mortalité comparée au PIB, mais aussi le ratio de décès par cause

D'autres questions pourront être ajoutées au fur et à mesure pour plus de précisions et d'interprétations

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2     3.4.4      v tibble    3.2.1
## v lubridate   1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggplot2)
```

```
library(shiny)
```

```
## Warning: le package 'shiny' a été compilé avec la version R 4.3.3
```

```
library(bslib)
```

```
##
```

```
## Attachement du package : 'bslib'
```

```
##
```

```
## L'objet suivant est masqué depuis 'package:utils':
```

```
##
```

```
##      page
```

```
library(maps)
```

```
## Warning: le package 'maps' a été compilé avec la version R 4.3.3
```

```
##
```

```
## Attachement du package : 'maps'
```

```
##
```

```
## L'objet suivant est masqué depuis 'package:purrr':
```

```
##
```

```
## map
```

```
library(patchwork)
```

```
## Warning: le package 'patchwork' a été compilé avec la version R 4.3.3
```

```
theme_set(theme_minimal())
```

## Chargement des datasets

### Causes de mortalité

Les données proviennent de Our World In Data Source: <https://www.kaggle.com/datasets/ivanchvez/causes-of-death-our-world-in-data>

```
CausesData <- read_csv("20222703-causes-f-death-clean-output-v2.0.csv") %>%  
  rename(iso_code=Code) %>%  
  filter(Year %in% 1990:2019) %>%  
  filter(!is.na(iso_code));
```

```
## Rows: 201762 Columns: 6
```

```
## -- Column specification -----
```

```
## Delimiter: ","
```

```
## chr (4): Causes name, Causes Full Description, Entity, Code
```

```
## dbl (2): Death Numbers, Year
```

```
##
```

```
## i Use `spec()` to retrieve the full column specification for this data.
```

```
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

### Données démographiques et économiques des pays

```
CountriesData <- read_csv("owid-co2-data.csv") %>%  
  select(year, iso_code, population, gdp) %>%  
  filter(year %in% 1990:2019) %>%  
  filter(!is.na(iso_code));
```

```
## Rows: 50598 Columns: 79
```

```
## -- Column specification -----
```

```
## Delimiter: ","
```

```
## chr (2): country, iso_code
```

```
## dbl (77): year, population, gdp, cement_co2, cement_co2_per_capita, co2, co2...
```

```
##
```

```
## i Use `spec()` to retrieve the full column specification for this data.
```

```
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

### Données géographique des pays

source: <https://github.com/luke/ISO-3166-Countries-with-Regional-Codes/blob/master/all/all.csv>. Ces données sont le résultat de la fusion de données provenant de deux sources, l'article ISO 3166-1 de Wikipedia pour les codes alpha et numériques des pays, et le site des statistiques de l'ONU pour les codes régionaux et sous-régionaux des pays. Outre les pays, elles incluent les territoires dépendants.

```
ContinentsData <- read_csv("continents2.csv") %>%  
  select('alpha-3', region, 'sub-region') %>%  
  rename(iso_code='alpha-3') %>%  
  filter(!is.na(region));
```

```
## Rows: 249 Columns: 11
## -- Column specification -----
## Delimiter: ","
## chr (7): name, alpha-2, alpha-3, iso_3166-2, region, sub-region, intermediat...
## dbl (4): country-code, region-code, sub-region-code, intermediate-region-code
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
subRegions <- ContinentsData %>% distinct(`sub-region`);
```

## fusion des datasets

```
#data <- merge(CausesData, CountriesData, by="iso_code");
data <- left_join(CausesData, CountriesData, by=c("iso_code"="iso_code", "Year"="year")) %>%
  left_join(ContinentsData, by=c("iso_code"="iso_code")) %>%
  arrange(Year);
```

## Représentation sur la carte géographique des Décès en million par cause et par année

```
world_data <- ggplot2::map_data('world')
world_data <- fortify(world_data) %>% select(long, lat, group, region)
head(world_data)
```

```
##      long      lat group region
## 1 -69.89912 12.45200     1  Aruba
## 2 -69.89571 12.42300     1  Aruba
## 3 -69.94219 12.43853     1  Aruba
## 4 -70.00415 12.50049     1  Aruba
## 5 -70.06612 12.54697     1  Aruba
## 6 -70.05088 12.59707     1  Aruba
```

## modification des noms de pays

Ce code consiste à modifier le nom de certains pays dans notre jeu de données qui sont différents de ceux fournis par ggplot2::map\_data

```
data <- data %>%
  mutate(Entity = recode(str_trim(Entity), "United States" = "USA",
                        "Congo"="Republic of Congo",
                        "Democratic Republic of Congo"="Democratic Republic of the Congo",
                        "United Kingdom"="UK"))
```

## Représentation sur carte graphique avec Shiny

Shiny est un package R qui facilite la création d'applications web interactives (apps) directement à partir de R.

```
causesName <- distinct(data, `Causes name`) %>% arrange(`Causes name`)
year <- distinct(data, Year) %>% arrange(Year)

# Define UI for app that draws a histogram ----
```

```

ui <- page_sidebar(
  title = "Représentation sur carte graphique des nombres de Décès en million dans le monde par causes",
  sidebar = sidebar(
    helpText("Variables"),
    selectInput(
      "cause",
      label="Causes de Décès en million",
      choices = causesName,
      selected = "Acute hepatitis"
    ),
    sliderInput(
      "year",
      label="Années",
      min = 1990,
      max = 2019,
      value = year,
    )
  ),
  card(plotOutput("map"))
)

server <- function(input, output) {
  output$map <- renderPlot({
    plain <- theme(
      axis.text = element_blank(),
      axis.line = element_blank(),
      axis.ticks = element_blank(),
      panel.border = element_blank(),
      panel.grid = element_blank(),
      axis.title = element_blank(),
      panel.background = element_rect(fill = "white"),
      plot.title = element_text(hjust = 0.5)
    )

    selectedData <- data %>%
      filter(`Causes name`==input$cause) %>%
      filter(Year==input$year[1]) %>%
      left_join(world_data, by=c("Entity"="region"))

    worldDeath <- ggplot(data = selectedData, mapping = aes(x = long, y = lat, group = group)) +
      coord_fixed(1.3) +
      geom_polygon(aes(fill = `Death Numbers`)) +
      scale_fill_distiller(palette="Spectral", direction = -1) +
      plain

    worldDeath
  })
}

shinyApp(ui = ui, server = server)

```

## Les 5 principales causes de mortalité dans des régions sélectionnées

Ces sous-régions de chaque continent avec 5 causes de mortalité en million

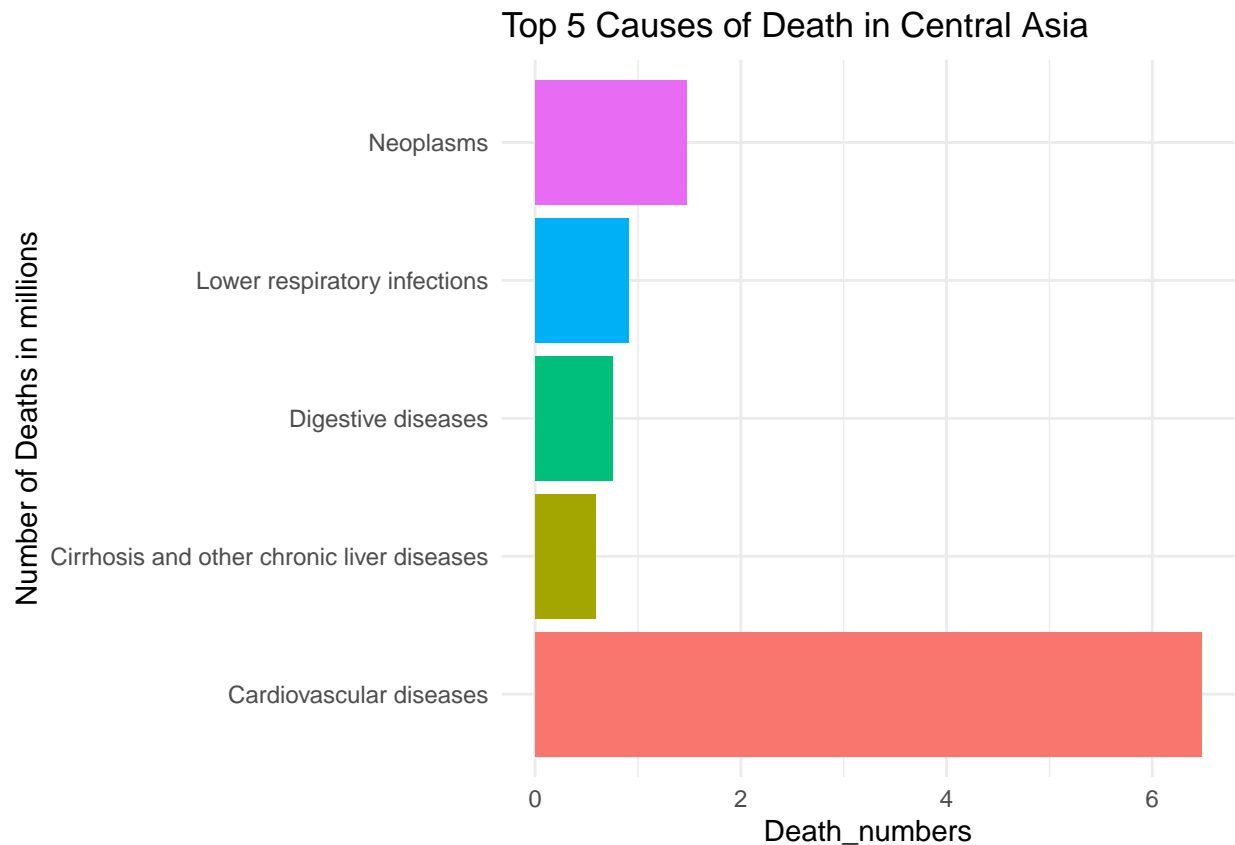
Asie (5 sub): Asie central, Asie de l'est, Asie du sud-est, Asie du sud et Asie de l'ouest

```
RegionCentralAsia <- data %>%
  filter(!is.na(`Death Numbers`)) %>%
  filter(`sub-region` == c("Central Asia")) %>%
  group_by(`sub-region`, `Causes name`) %>%
  summarise(Death_numbers = sum(as.numeric(`Death Numbers`)/1000000, na.rm = TRUE)) %>%
  arrange(`sub-region`, desc(Death_numbers)) %>%
  slice_max(order_by = Death_numbers, n = 5)
```

```
## `summarise()` has grouped output by 'sub-region'. You can override using the
## `.groups` argument.
```

```
ggplot(RegionCentralAsia, aes(y=`Causes name`, x=Death_numbers, fill=`Causes name`)) +
  geom_bar(stat="identity") +
  theme_minimal() +
  labs(title="Top 5 Causes of Death in Central Asia",
       y="Number of Deaths in millions") +
  guides(fill=FALSE)
```

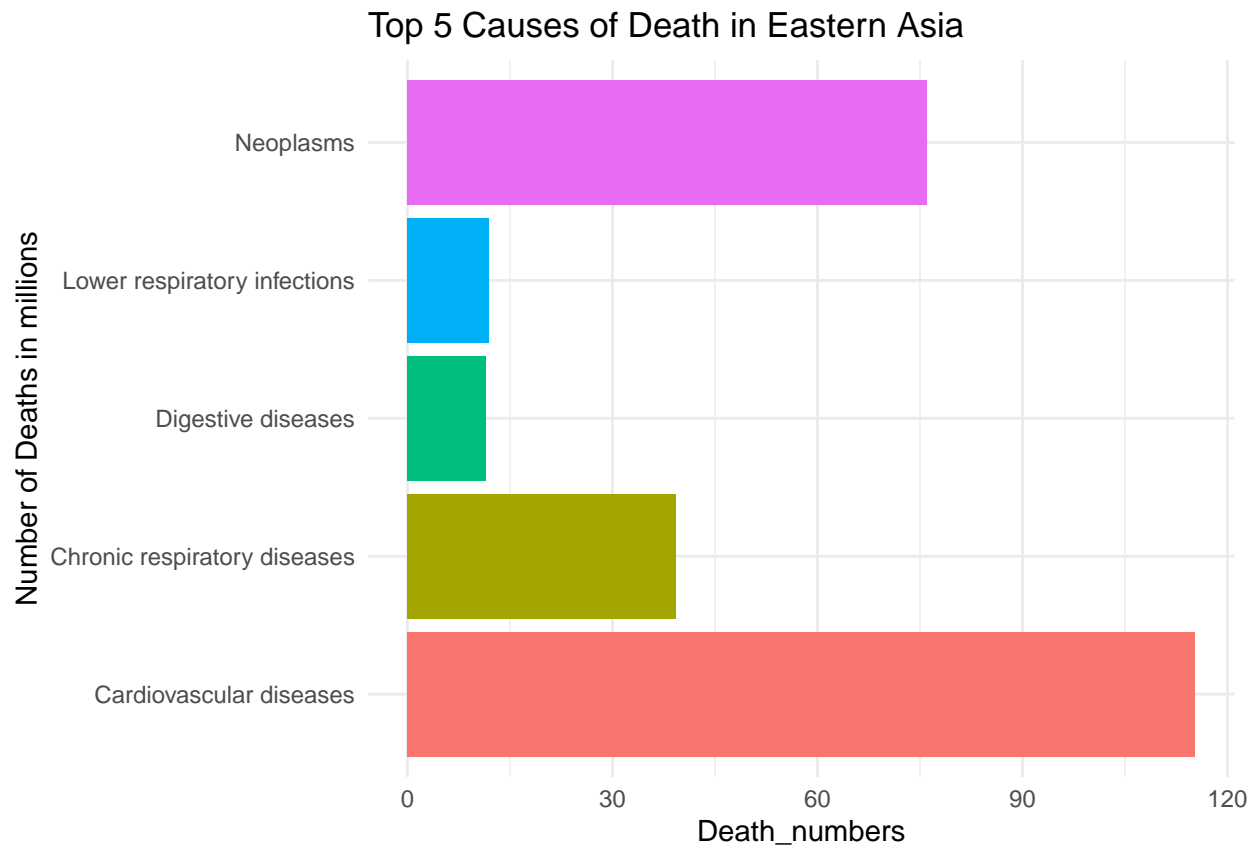
```
## Warning: The `scale` argument of `guides()` cannot be `FALSE`. Use "none" instead as
## of ggplot2 3.3.4.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



```
RegionEasternAsia <- data %>%
  filter(!is.na(`Death Numbers`)) %>%
  filter(`sub-region` == c("Eastern Asia")) %>%
  group_by(`sub-region`, `Causes name`) %>%
  summarise(Death_numbers = sum(as.numeric(`Death Numbers`)/1000000, na.rm = TRUE)) %>%
  arrange(`sub-region`, desc(Death_numbers)) %>%
  slice_max(order_by = Death_numbers, n = 5)
```

## `summarise()` has grouped output by 'sub-region'. You can override using the  
## `.groups` argument.

```
ggplot(RegionEasternAsia, aes(y=`Causes name`, x=Death_numbers, fill=`Causes name`)) +
  geom_bar(stat="identity") +
  theme_minimal() +
  labs(title="Top 5 Causes of Death in Eastern Asia",
       y="Number of Deaths in millions") +
  guides(fill=FALSE)
```



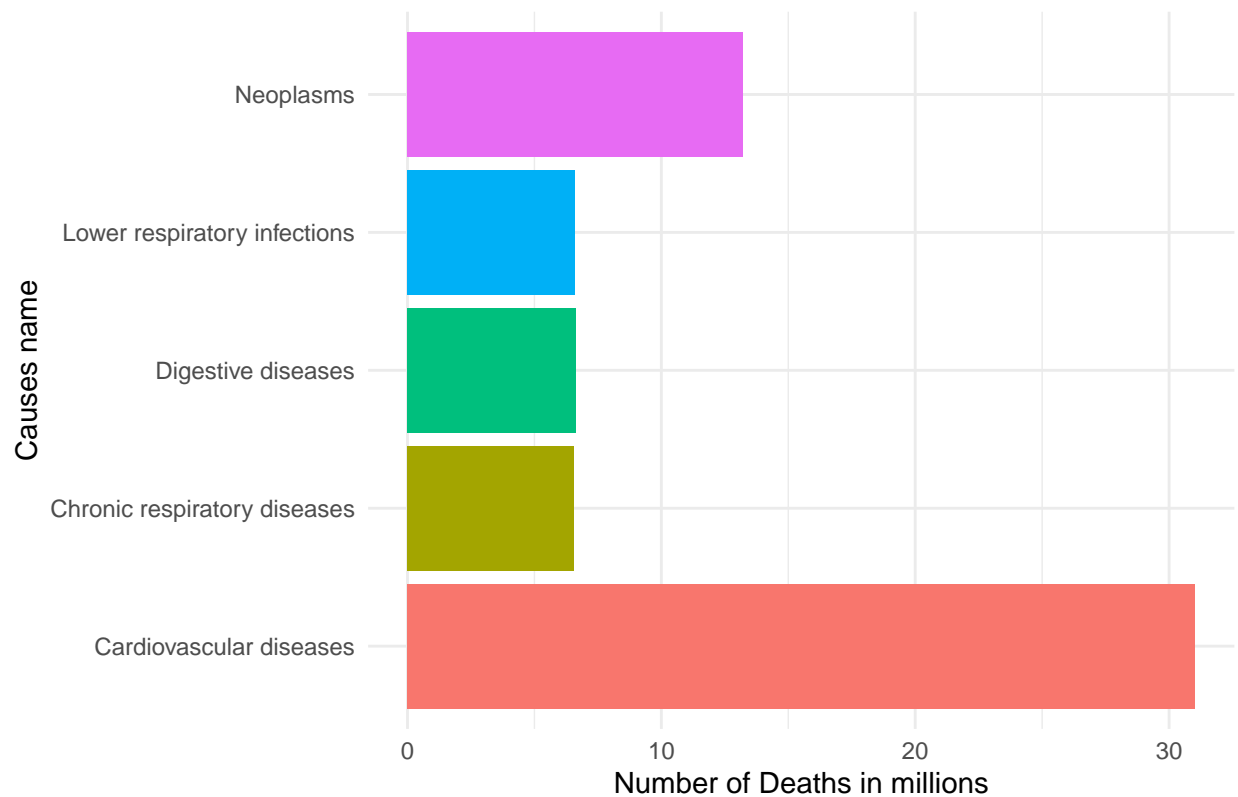
```
RegionSoutheasternAsia <- data %>%
  filter(!is.na(`Death Numbers`)) %>%
  filter(`sub-region` == c("South-eastern Asia")) %>%
  group_by(`sub-region`, `Causes name`) %>%
  summarise(Death_numbers = sum(as.numeric(`Death Numbers`)/1000000, na.rm = TRUE)) %>%
  arrange(`sub-region`, desc(Death_numbers)) %>%
  slice_max(order_by = Death_numbers, n = 5)
```

```
## `summarise()` has grouped output by 'sub-region'. You can override using the
## `.groups` argument.
```

```
ggplot(RegionSoutheasternAsia, aes(x=`Causes name`, y=Death_numbers, fill=`Causes name`)) +
  geom_bar(stat="identity") +

  theme_minimal() +
  labs(title="Top 5 Causes of Death in South-eastern Asia",
       y="Number of Deaths in millions") +
  coord_flip()+
  guides(fill=FALSE)
```

### Top 5 Causes of Death in South-eastern Asia

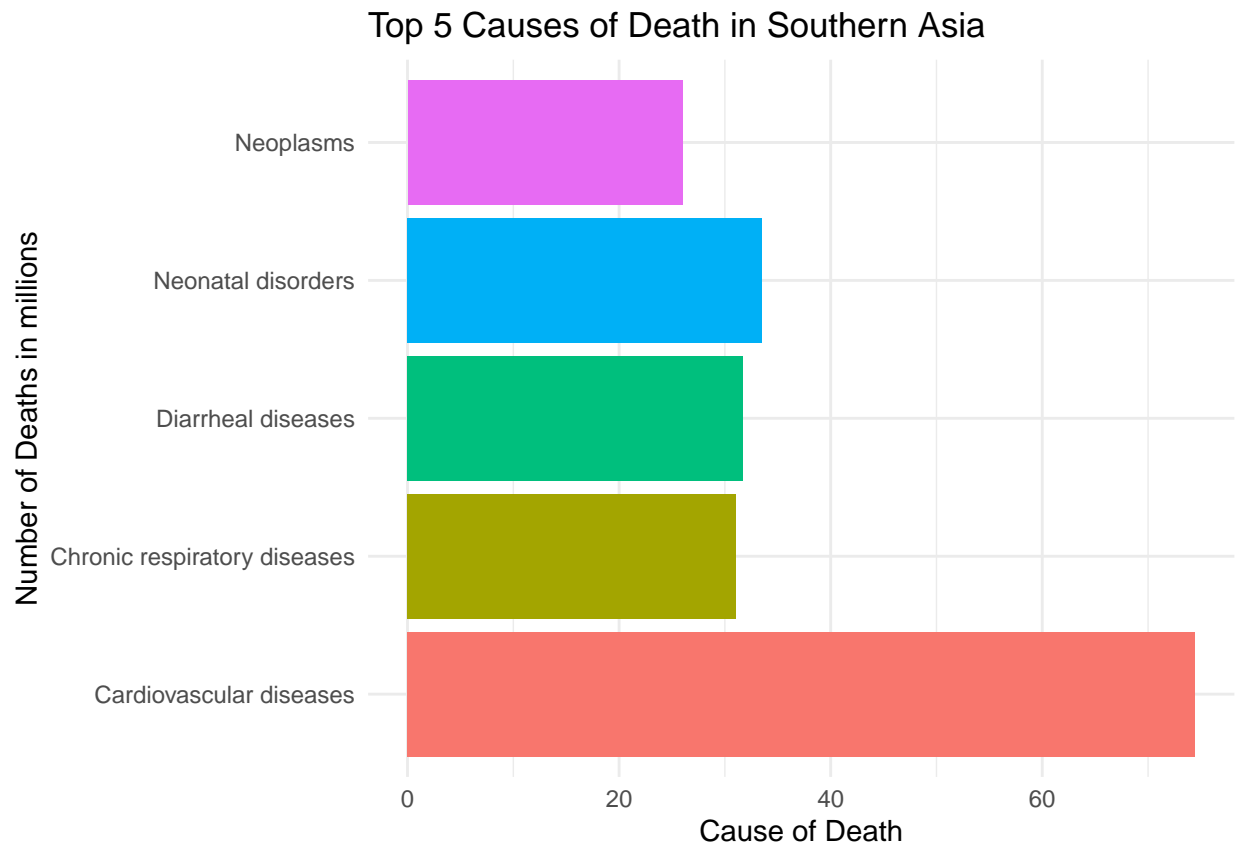


```
RegionSouthernAsia <- data %>%
  filter(!is.na(`Death Numbers`)) %>%
  filter(`sub-region` == c("Southern Asia")) %>%
  group_by(`sub-region`, `Causes name`) %>%
  summarise(Death_numbers = sum(as.numeric(`Death Numbers`)/1000000, na.rm = TRUE)) %>%
  arrange(`sub-region`, desc(Death_numbers)) %>%
  slice_max(order_by = Death_numbers, n = 5)
```

## `summarise()` has grouped output by 'sub-region'. You can override using the  
## `.groups` argument.

```
ggplot(RegionSouthernAsia, aes(y=`Causes name`, x=Death_numbers, fill=`Causes name`)) +
  geom_bar(stat="identity") +
  theme_minimal() +
  labs(title="Top 5 Causes of Death in Southern Asia",
       x="Cause of Death",
       y="Number of Deaths in millions") +
  guides(fill=FALSE)
```



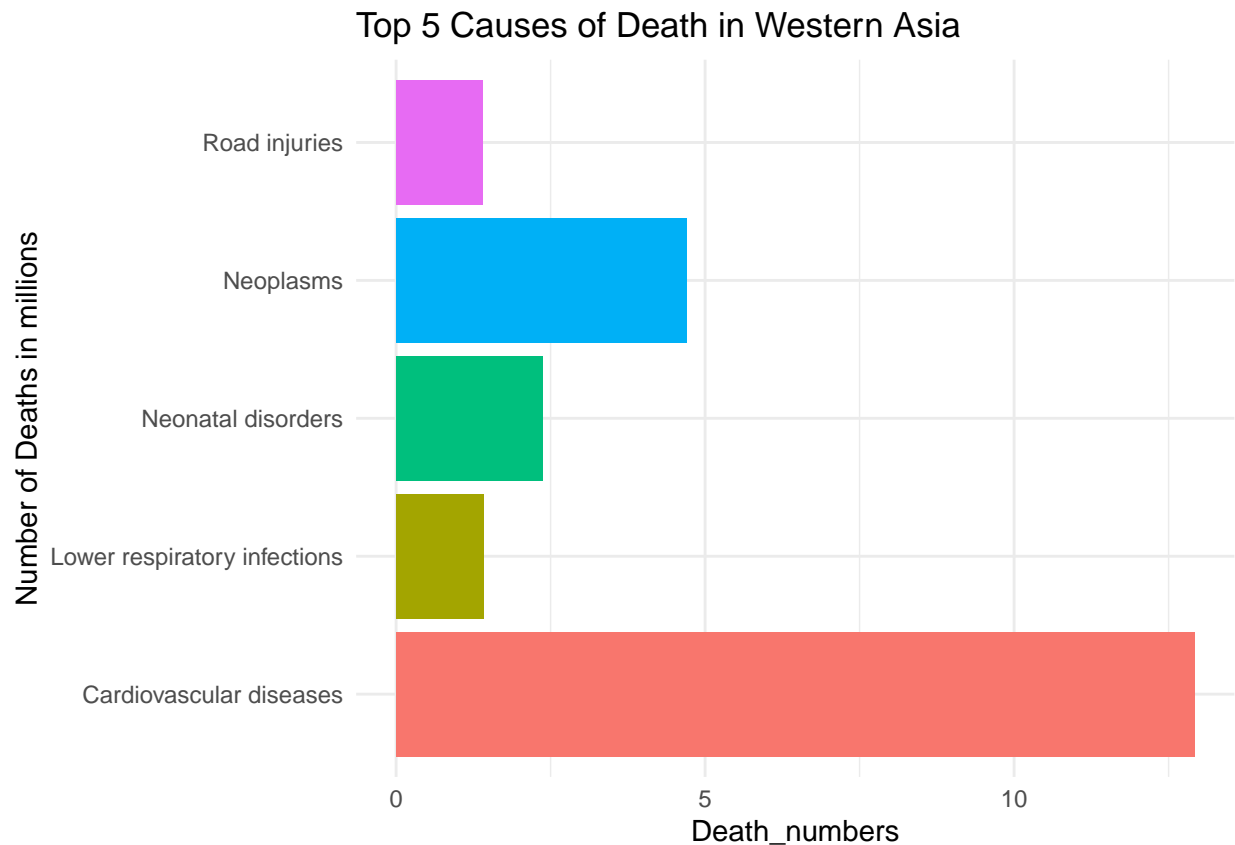


```
RegionWesternAsia <- data %>%
  filter(!is.na(`Death Numbers`)) %>%
  filter(`sub-region` == c("Western Asia")) %>%
  group_by(`sub-region`, `Causes name`) %>%
  summarise(Death_numbers = sum(as.numeric(`Death Numbers`)/1000000, na.rm = TRUE)) %>%
  arrange(`sub-region`, desc(Death_numbers)) %>%
  slice_max(order_by = Death_numbers, n = 5)
```

## `summarise()` has grouped output by 'sub-region'. You can override using the  
## `.groups` argument.

```
ggplot(RegionWesternAsia, aes(y=`Causes name`, x=Death_numbers, fill=`Causes name`)) +
  geom_bar(stat="identity") +

  theme_minimal() +
  labs(title="Top 5 Causes of Death in Western Asia",
       y="Number of Deaths in millions") +
  guides(fill=FALSE)
```

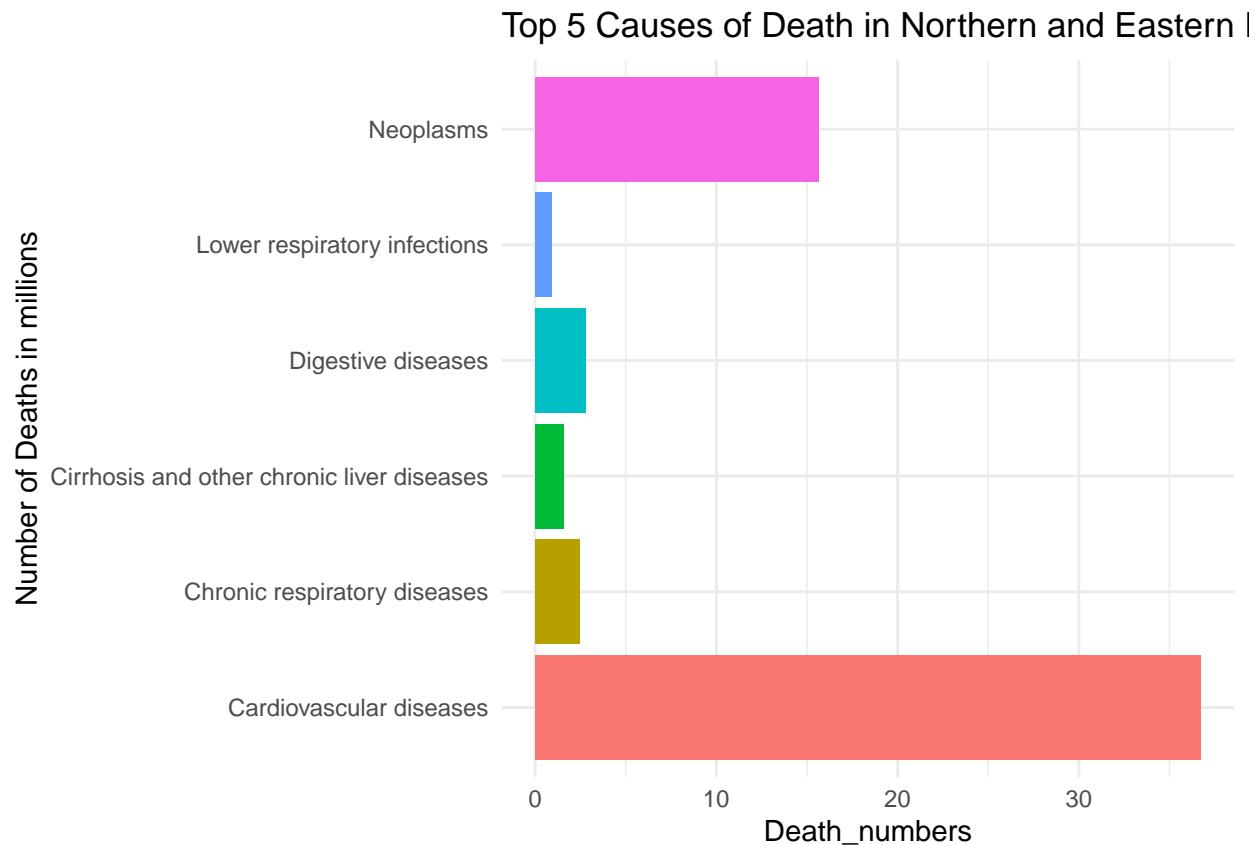


### Europe (2 sub) : Europe du nord et Europe du sud

```
RegionNorthernEurope <- data %>%
  filter(!is.na(`Death Numbers`)) %>%
  filter(`sub-region` == c("Northern Europe", "Eastern Europe")) %>%
  group_by(`sub-region`, `Causes name`) %>%
  summarise(Death_numbers = sum(as.numeric(`Death Numbers`)/1000000, na.rm = TRUE)) %>%
  arrange(`sub-region`, desc(Death_numbers)) %>%
  slice_max(order_by = Death_numbers, n = 5)
```

## `summarise()` has grouped output by 'sub-region'. You can override using the  
## `.groups` argument.

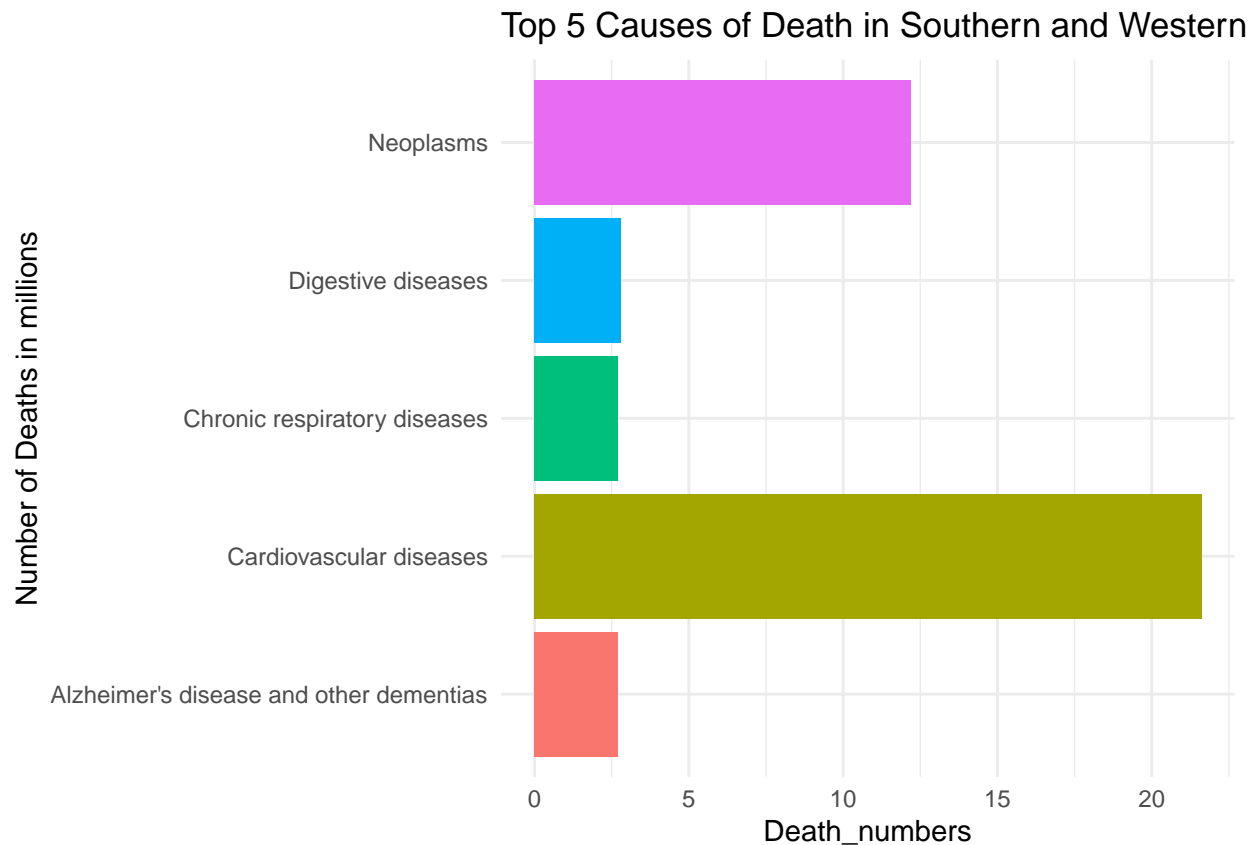
```
ggplot(RegionNorthernEurope, aes(y=`Causes name`, x=Death_numbers, fill=`Causes name`)) +
  geom_bar(stat="identity") +
  theme_minimal() +
  labs(title="Top 5 Causes of Death in Northern and Eastern Europe",
       y="Number of Deaths in millions") +
  guides(fill=FALSE)
```



```
RegionSouthernEurope <- data %>%
  filter(!is.na(`Death Numbers`)) %>%
  filter(`sub-region` == c("Western Europe", "Southern Europe")) %>%
  group_by(`sub-region`, `Causes name`) %>%
  summarise(Death_numbers = sum(as.numeric(`Death Numbers`)/1000000, na.rm = TRUE)) %>%
  arrange(`sub-region`, desc(Death_numbers)) %>%
  slice_max(order_by = Death_numbers, n = 5)
```

## `summarise()` has grouped output by 'sub-region'. You can override using the  
## `.groups` argument.

```
ggplot(RegionSouthernEurope, aes(y=`Causes name`, x=Death_numbers, fill=`Causes name`)) +
  geom_bar(stat="identity") +
  theme_minimal() +
  labs(title="Top 5 Causes of Death in Southern and Western Europe",
       y="Number of Deaths in millions") +
  guides(fill=FALSE)
```



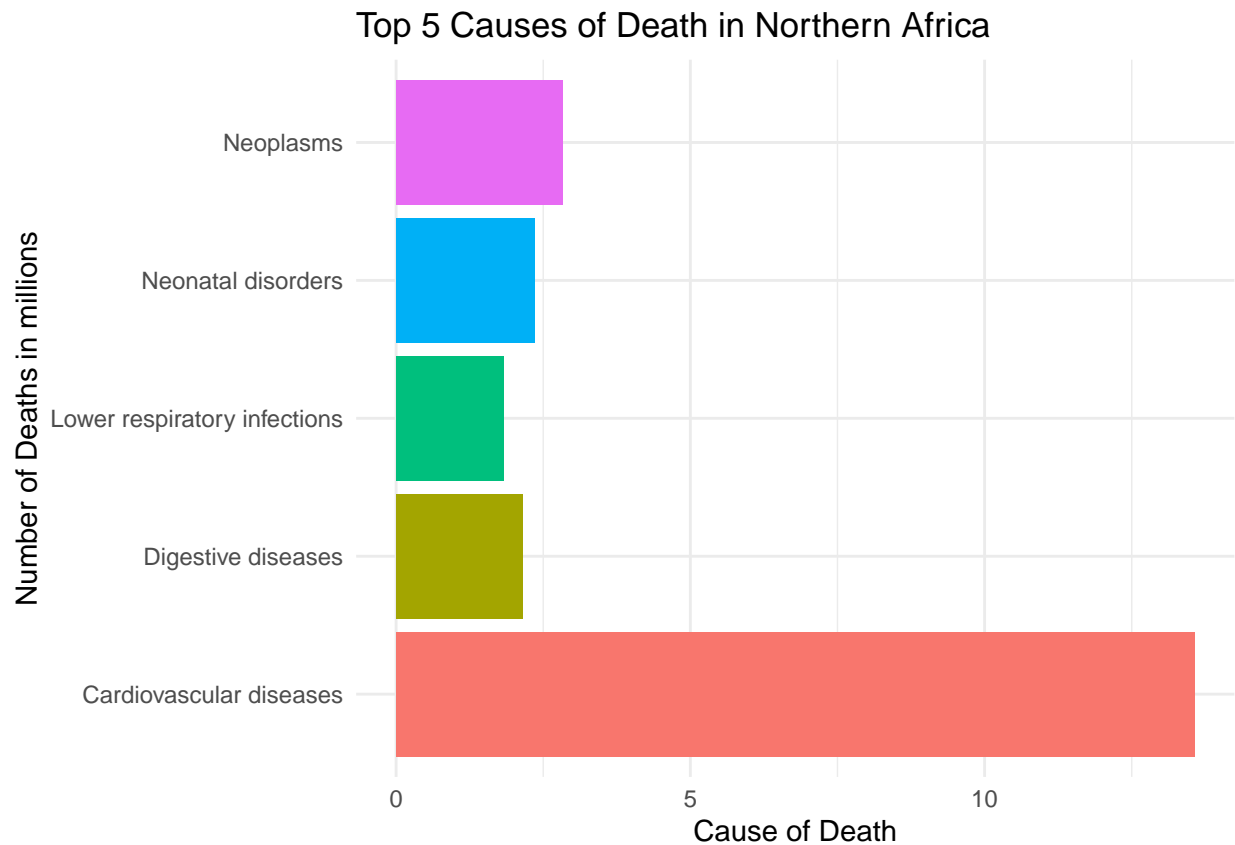
Afrique (2-sub): Afrique du nord et Afrique sub-saharienne

```
RegionNorthernAfrica <- data %>%
  filter(!is.na(`Death Numbers`)) %>%
  filter(`sub-region` == c("Northern Africa")) %>%
  group_by(`sub-region`, `Causes name`) %>%
  summarise(Death_numbers = sum(as.numeric(`Death Numbers`)/1000000, na.rm = TRUE)) %>%
  arrange(`sub-region`, desc(Death_numbers)) %>%
  slice_max(order_by = Death_numbers, n = 5)
```

## `summarise()` has grouped output by 'sub-region'. You can override using the  
## `.groups` argument.

```
ggplot(RegionNorthernAfrica, aes(y=`Causes name`, x=Death_numbers, fill=`Causes name`)) +
  geom_bar(stat="identity") +

  theme_minimal() +
  labs(title="Top 5 Causes of Death in Northern Africa",
       x="Cause of Death",
       y="Number of Deaths in millions") +
  guides(fill=FALSE)
```

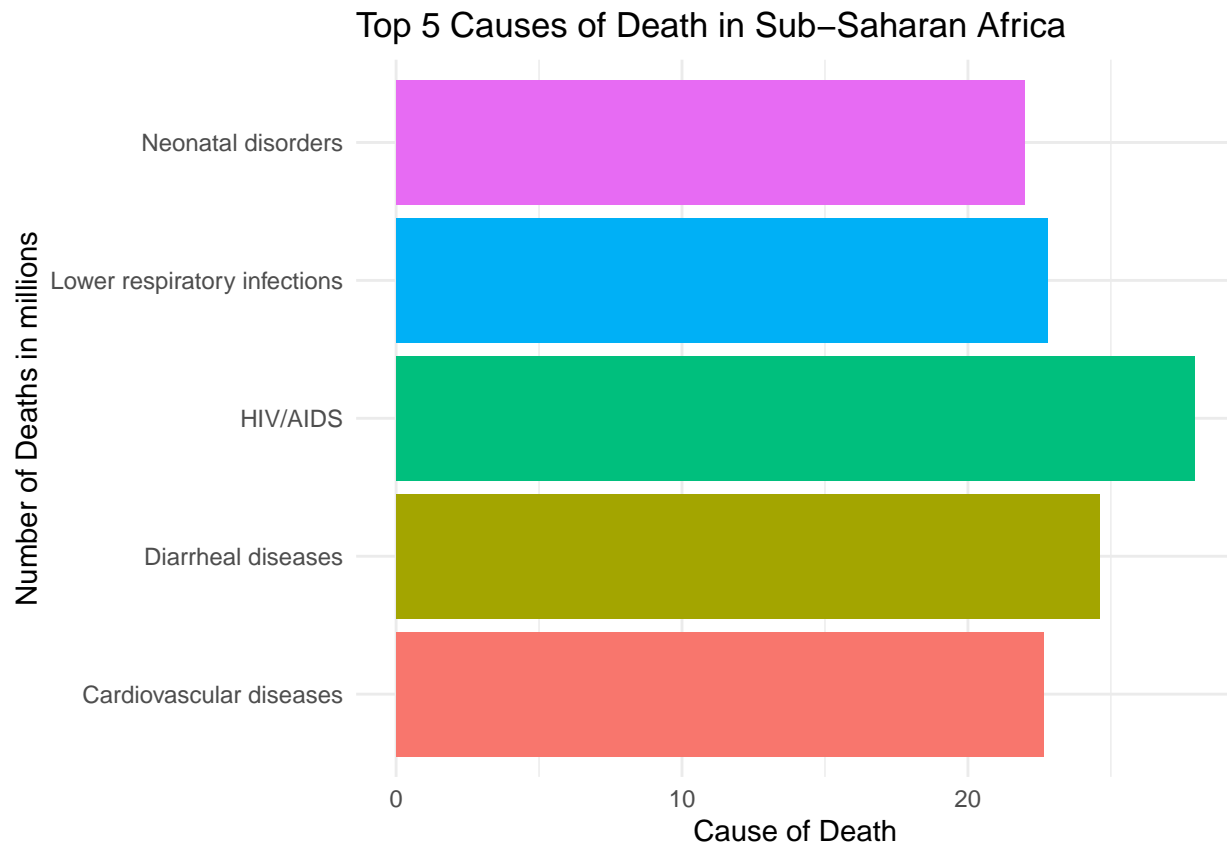


```
RegionSubSaharan <- data %>%
  filter(!is.na(`Death Numbers`)) %>%
  filter(`sub-region` == c("Sub-Saharan Africa")) %>%
  group_by(`sub-region`, `Causes name`) %>%
  summarise(Death_numbers = sum(as.numeric(`Death Numbers`)/1000000, na.rm = TRUE)) %>%
  arrange(`sub-region`, desc(Death_numbers)) %>%
  slice_max(order_by = Death_numbers, n = 5)
```

## `summarise()` has grouped output by 'sub-region'. You can override using the  
## `.groups` argument.

```
ggplot(RegionSubSaharan, aes(y=`Causes name`, x=Death_numbers, fill=`Causes name`)) +
  geom_bar(stat="identity") +

  theme_minimal() +
  labs(title="Top 5 Causes of Death in Sub-Saharan Africa",
       x="Cause of Death",
       y="Number of Deaths in millions") +
  guides(fill=FALSE)
```



Amerique (2-sub): Amerique du nord et America Latine et les caraïbes

```
sousRegionAmericas<- data %>%
  filter(!is.na(`Death Numbers`)) %>%
  filter(region=="Americas") %>%
  group_by(`sub-region`, `Causes name`) %>%
  summarise(Death_numbers=sum(as.numeric(`Death Numbers`))) %>%
  arrange(`sub-region`, desc(Death_numbers)) %>%
  group_by(`sub-region`) %>%
  slice_max(order_by = Death_numbers, n=5)
```

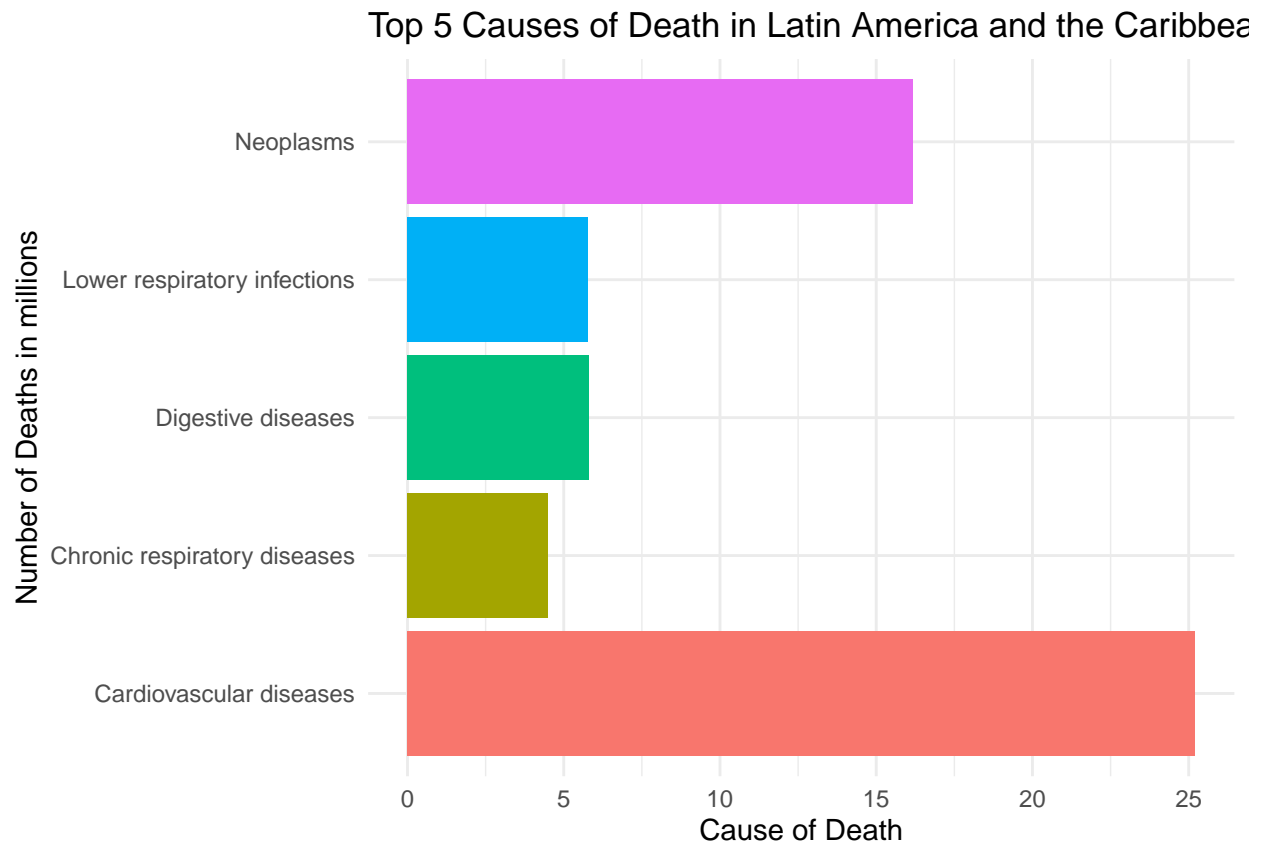
## `summarise()` has grouped output by 'sub-region'. You can override using the  
## `.groups` argument.

```
RegionLatinAndCari <- data %>%
  filter(!is.na(`Death Numbers`)) %>%
  filter(`sub-region` == c("Latin America and the Caribbean")) %>%
  group_by(`sub-region`, `Causes name`) %>%
  summarise(Death_numbers = sum(as.numeric(`Death Numbers`))/1000000, na.rm = TRUE)) %>%
  arrange(`sub-region`, desc(Death_numbers)) %>%
  slice_max(order_by = Death_numbers, n = 5)
```

## `summarise()` has grouped output by 'sub-region'. You can override using the  
## `.groups` argument.

```
ggplot(RegionLatinAndCari , aes(y=`Causes name`, x=Death_numbers, fill=`Causes name`)) +
  geom_bar(stat="identity") +
```

```
theme_minimal() +
labs(title="Top 5 Causes of Death in Latin America and the Caribbean",
x="Cause of Death",
y="Number of Deaths in millions") +
guides(fill=FALSE)
```

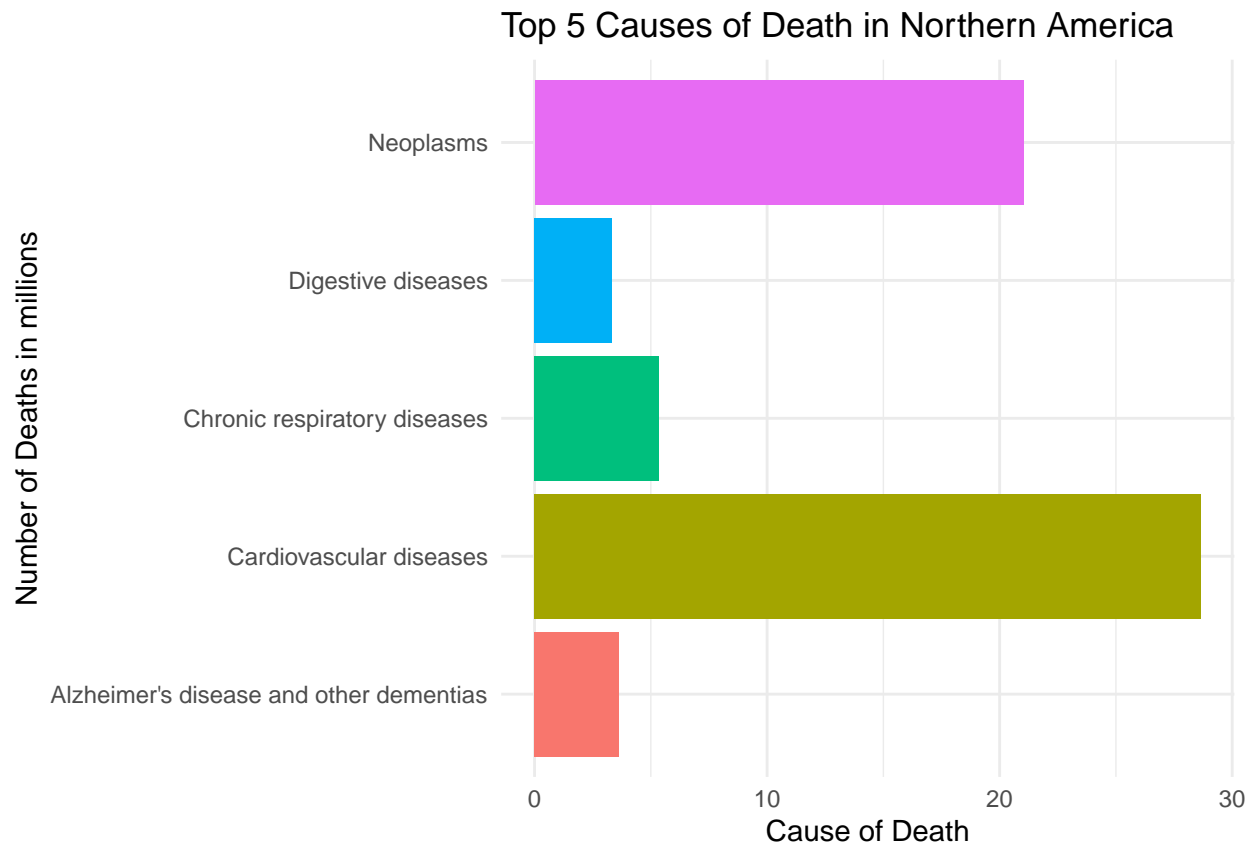


```
RegionNorthernAmerica <- data %>%
  filter(!is.na(`Death Numbers`)) %>%
  filter(`sub-region` == c("Northern America")) %>%
  group_by(`sub-region`, `Causes name`) %>%
  summarise(Death_numbers = sum(as.numeric(`Death Numbers`)/1000000, na.rm = TRUE)) %>%
  arrange(`sub-region`, desc(Death_numbers)) %>%
  slice_max(order_by = Death_numbers, n = 5)
```

## `summarise()` has grouped output by 'sub-region'. You can override using the  
## `.groups` argument.

```
ggplot(RegionNorthernAmerica, aes(y=`Causes name`, x=Death_numbers, fill=`Causes name`)) +
  geom_bar(stat="identity") +

  theme_minimal() +
  labs(title="Top 5 Causes of Death in Northern America",
x="Cause of Death",
y="Number of Deaths in millions") +
  guides(fill=FALSE)
```



# Courbes d'évolution du nombre de Décès en million au cours des années de certaines régions

## L'Afrique sub-saharienne et l'Afrique du nord

En Afrique subsaharienne, on observe un pic de décès en 1994. En effet, nous avons assisté à un évènement, le génocide au Rwanda qui d'après l'ONU a fait environ 800000 morts.

De 1996 à 2003, on observe également une forte augmentation du nombre de décès. Cela peut être dû à l'accès difficile aux soins, la mauvaise qualité de vie, conflits, également les maladies infectieuses et néonatales. L'épidémie de VIH/SIDA a également eu un impact significatif en Afrique subsaharienne

Depuis 2003 on observe une baisse de mortalité, cela est dû à une amélioration de la qualité de vie avec des programmes de prévention, de traitement et de sensibilisation et e développement socio-économique

```
sumAfricaPopulation <- data %>%
  filter(region=="Africa") %>%
  distinct(Entity,Year,population, `sub-region`) %>%
  group_by(Year, `sub-region`) %>%
  summarise(population=sum(population, na.rm = TRUE) / 100000000);
```

## `summarise()` has grouped output by 'Year'. You can override using the  
## `.groups` argument.

```
sumAfricaDeaths <- data %>%
  filter(region=="Africa") %>%
```



```

group_by(Year, `sub-region`) %>%
summarise(death_numbers=sum(`Death Numbers`, na.rm = TRUE) / 1000000);

## `summarise()` has grouped output by 'Year'. You can override using the
## `.groups` argument.

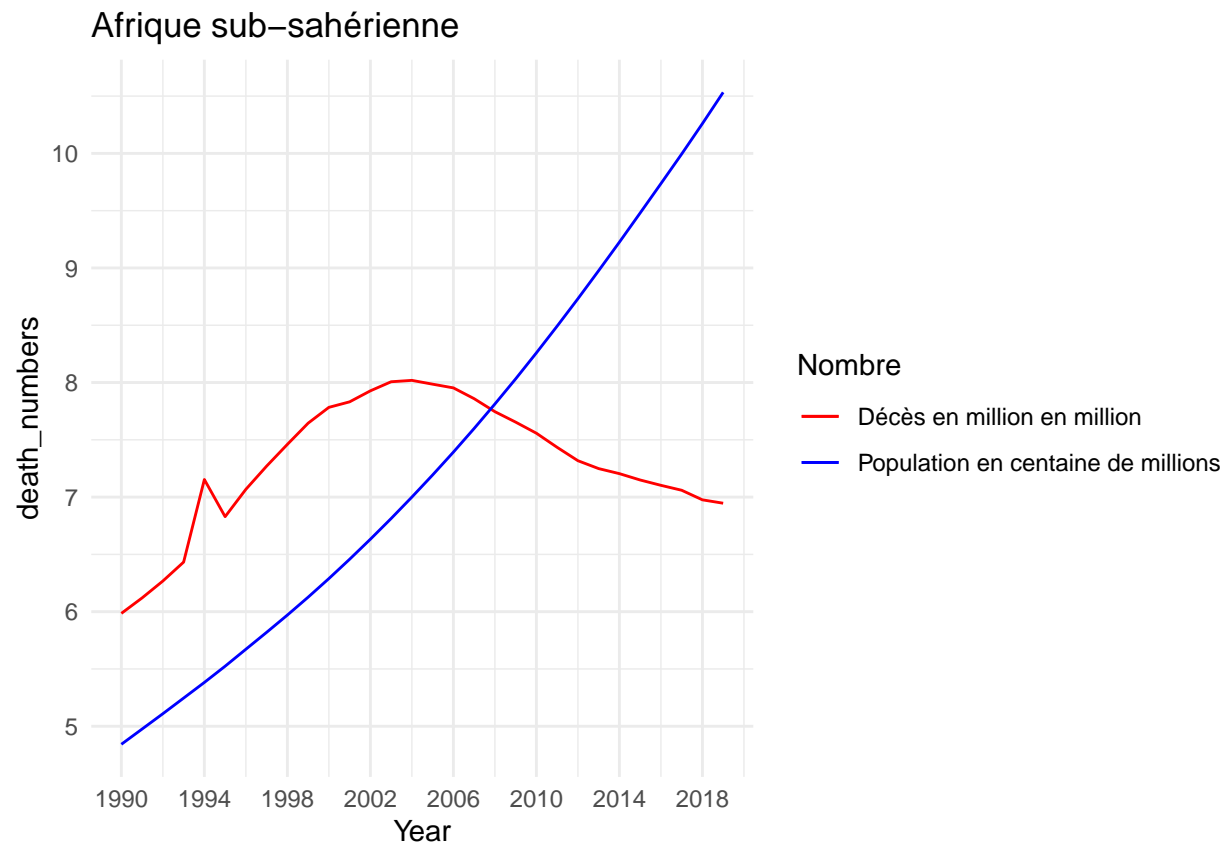
africaData <- left_join(sumAfricaPopulation, sumAfricaDeaths, by=c("Year"="Year", "sub-region"="sub-reg

plotSubSaharanAfrica <- africaData %>%
  filter(`sub-region`=="Sub-Saharan Africa") %>%
  ggplot(aes(x=Year)) +
  geom_line(aes(y = death_numbers, color="Décès en million en million")) +
  geom_line(aes(y = population, color="Population en centaine de millions"))+
  scale_color_manual(values = c("Décès en million en million" = "red", "Population en centaine de milli
  scale_x_continuous(breaks = seq(1990, 2019, by = 4))+
  labs(color = "Nombre", title = "Afrique sub-sahérienne");

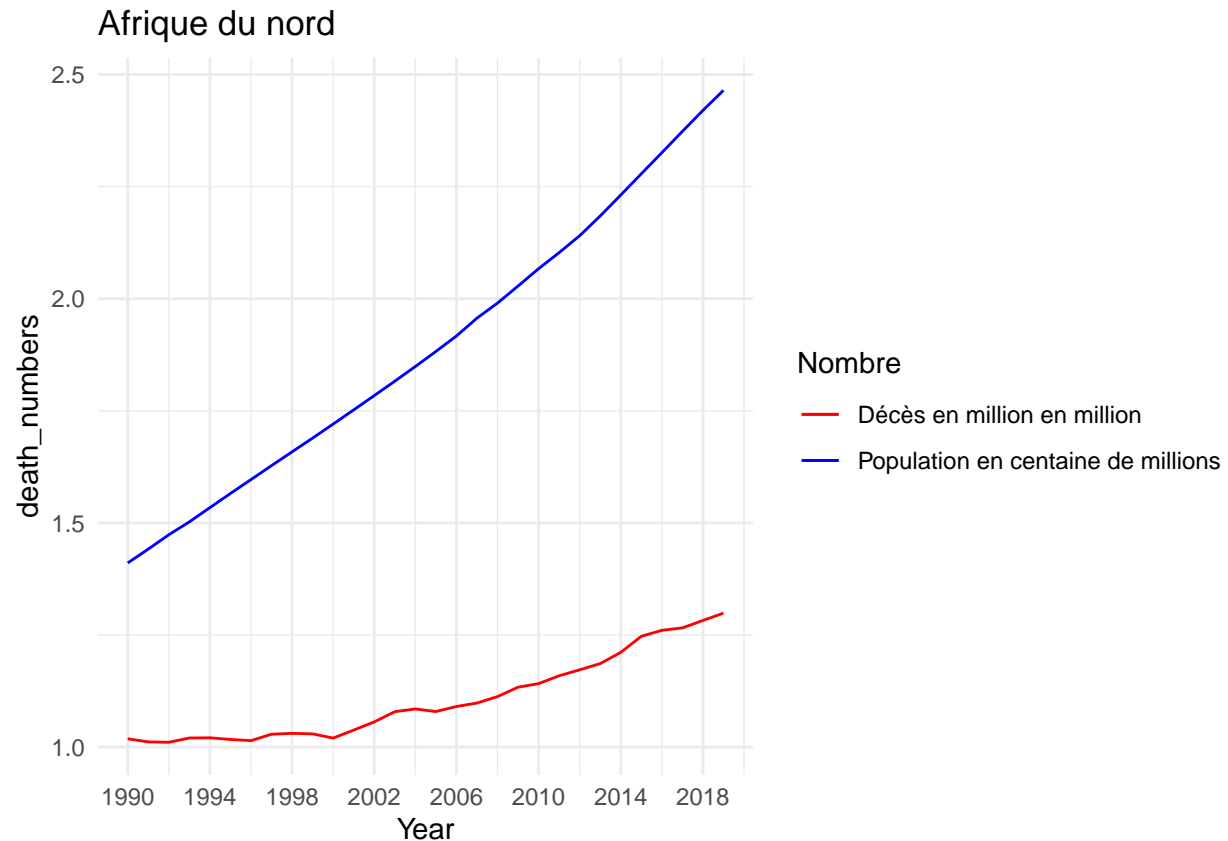
plotNorthernAfrica <- africaData %>%
  filter(`sub-region`=="Northern Africa") %>%
  ggplot(aes(x=Year)) +
  geom_line(aes(y = death_numbers, color="Décès en million en million")) +
  geom_line(aes(y = population, color="Population en centaine de millions"))+
  scale_color_manual(values = c("Décès en million en million" = "red", "Population en centaine de milli
  scale_x_continuous(breaks = seq(1990, 2019, by = 4))+
  labs(color = "Nombre",title = "Afrique du nord")

plotSubSaharanAfrica

```



plotNorthernAfrica



## Amérique du nord et Amérique Latine (+ caraïbes)

```
sumAmericaPopulation <- data %>%
  filter(region=="Americas") %>%
  distinct(Entity,Year,population, `sub-region`) %>%
  group_by(Year, `sub-region`) %>%
  summarise(population=sum(population, na.rm = TRUE) / 100000000);
```

## `summarise()` has grouped output by 'Year'. You can override using the  
## `.groups` argument.

```
sumAmericaDeaths <- data %>%
  filter(region=="Americas") %>%
  group_by(Year, `sub-region`) %>%
  summarise(death_numbers=sum(`Death Numbers`, na.rm = TRUE) / 1000000);
```

## `summarise()` has grouped output by 'Year'. You can override using the  
## `.groups` argument.

```
americaData <- left_join(sumAmericaPopulation, sumAmericaDeaths, by=c("Year"="Year", "sub-region"="sub-region"))
```

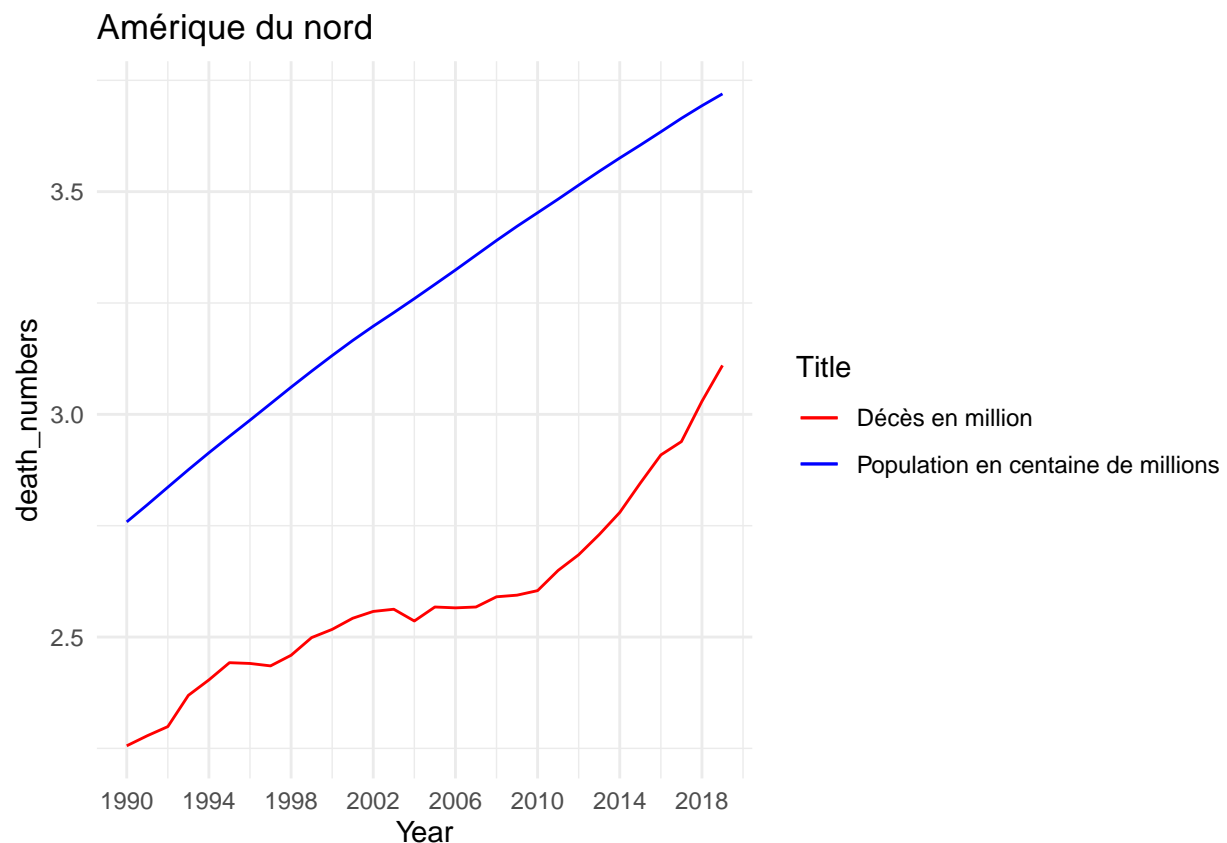
```
plotNorthernAmerica <- americaData %>%
  filter(`sub-region`=="Northern America") %>%
  ggplot(aes(x = Year)) +
  geom_line(aes(y = death_numbers, color = "Décès en million"), size = 0.5) +
  geom_line(aes(y = population, color = "Population en centaine de millions"), size = 0.5) +
  scale_color_manual(values = c("Décès en million" = "red", "Population en centaine de millions" = "blue"))
```

```
labs(color = "Title", title = "Amérique du nord")+
scale_x_continuous(breaks = seq(1990, 2019, by = 4))
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

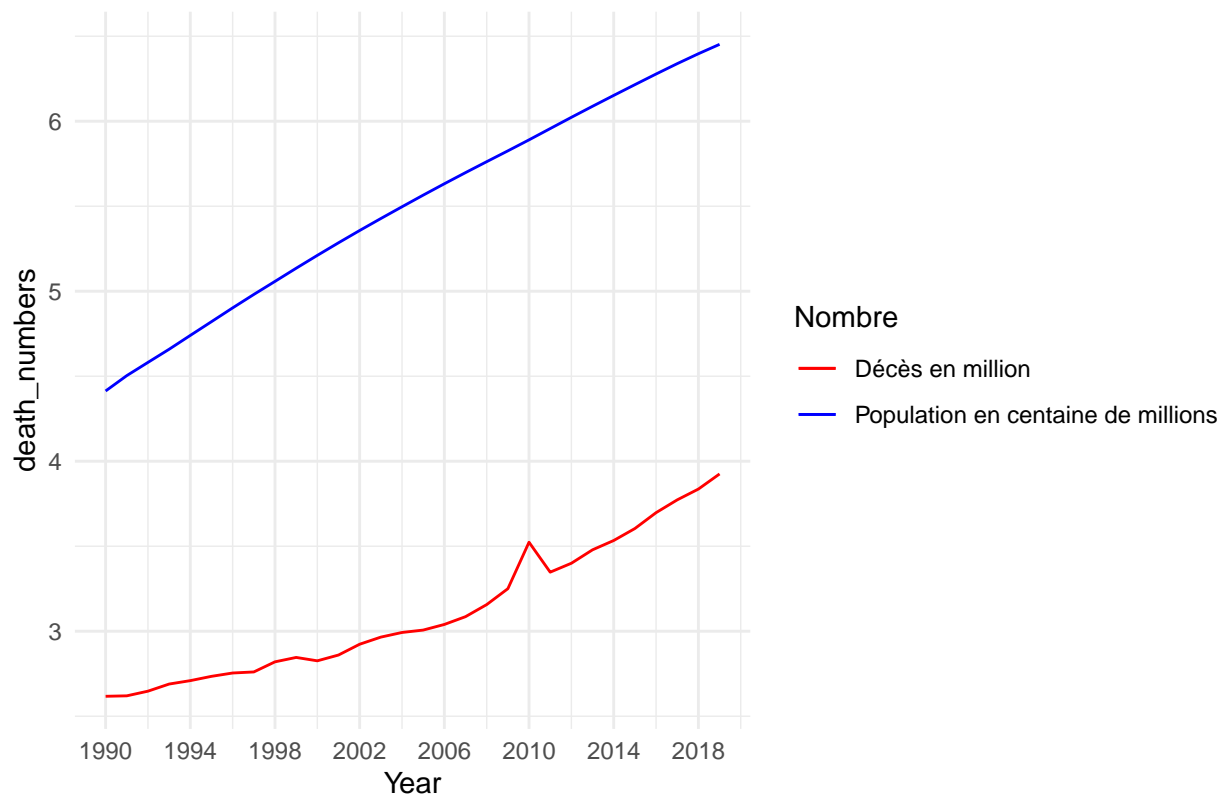
```
plotLatinAndCaribbeanAmerica <- americaData %>%
  filter(`sub-region`=="Latin America and the Caribbean") %>%
  ggplot(aes(x = Year)) +
  geom_line(aes(y = death_numbers, color = "Décès en million"), size = 0.5) +
  geom_line(aes(y = population, color = "Population en centaine de millions"), size = 0.5) +
  scale_color_manual(values = c("Décès en million" = "red", "Population en centaine de millions" = "blue")) +
  labs(color = "Nombre", title = "Amérique Latine et les caraïbes")+
  scale_x_continuous(breaks = seq(1990, 2019, by = 4))
```

```
plotNorthernAmerica
```



```
plotLatinAndCaribbeanAmerica
```

## Amérique Latine et les caraïbes



## Europe du sud et de l'ouest et Europe de l'est

```
sumEuropePopulation <- data %>%
  filter(region=="Europe") %>%
  distinct(Entity,Year,population, `sub-region`) %>%
  group_by(Year, `sub-region`) %>%
  summarise(population=sum(population, na.rm = TRUE) / 10000000);
```

## `summarise()` has grouped output by 'Year'. You can override using the  
## `.groups` argument.

```
sumEuropeDeaths <- data %>%
  filter(region=="Europe") %>%
  group_by(Year, `sub-region`) %>%
  summarise(death_numbers=sum(`Death Numbers`, na.rm = TRUE) / 1000000);
```

## `summarise()` has grouped output by 'Year'. You can override using the  
## `.groups` argument.

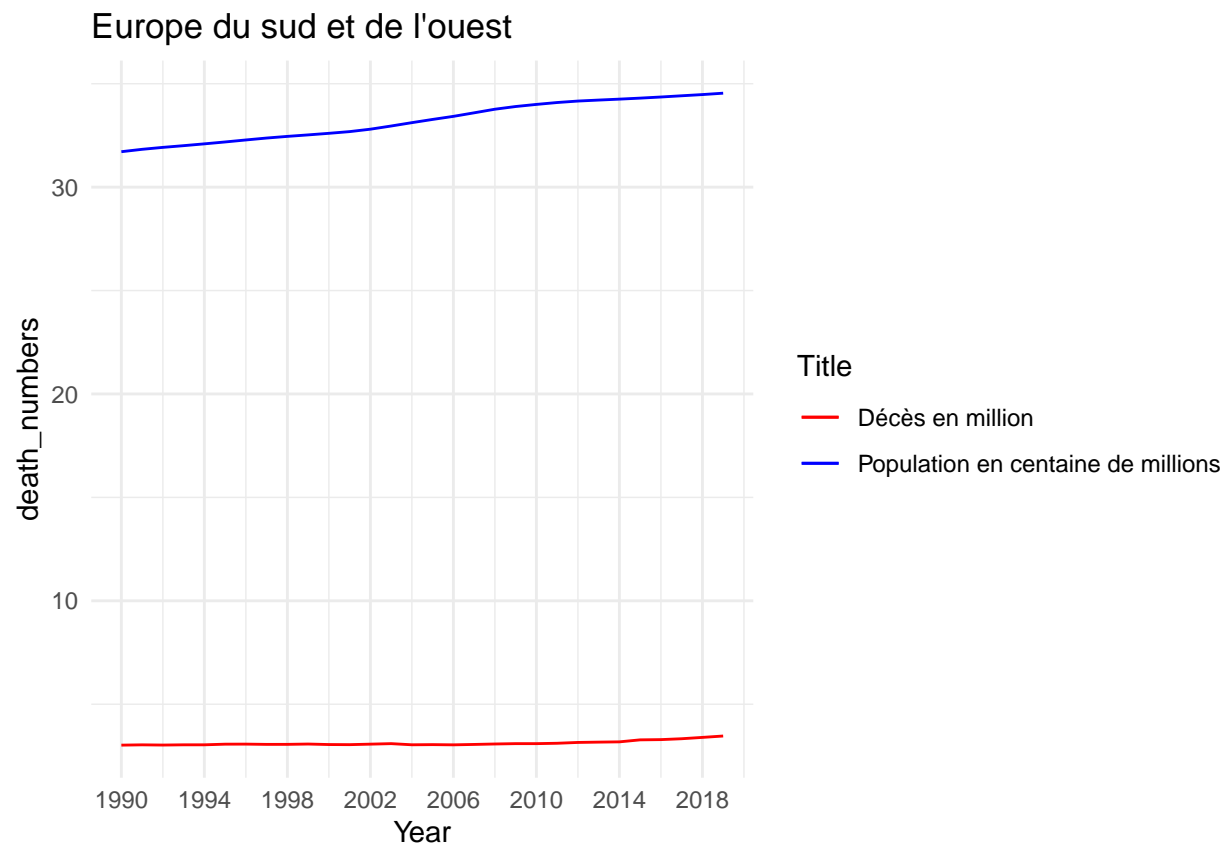
```
europeData <- left_join(sumEuropePopulation, sumEuropeDeaths, by=c("Year"="Year", "sub-region"="sub-regi
```

```
plotSouthernAndWesternEurope <- europeData %>%
  filter(`sub-region`==c("Southern Europe", "Western Europe")) %>%
  group_by(Year) %>%
  summarise(death_numbers=sum(death_numbers), population=sum(population)) %>%
  ggplot(aes(x = Year)) +
  geom_line(aes(y = death_numbers, color = "Décès en million"), size = 0.5) +
```

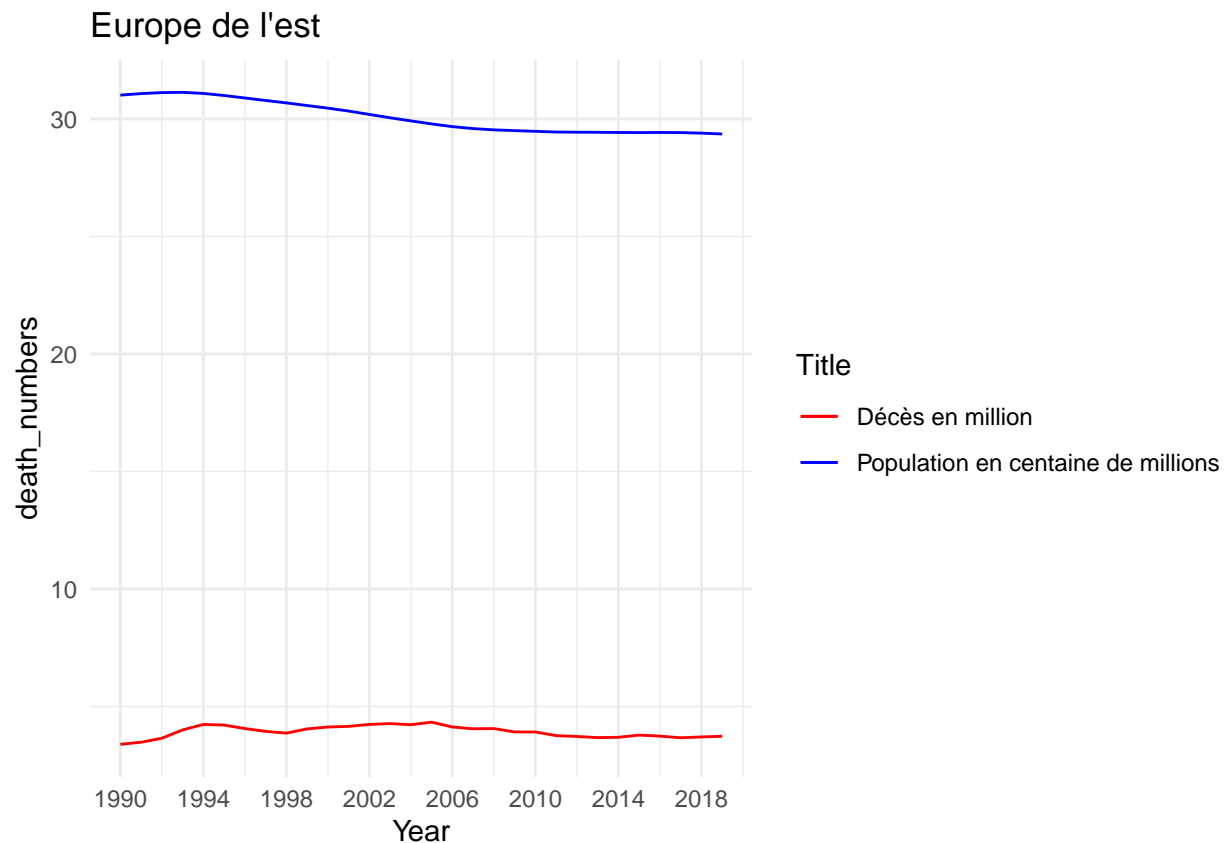
```
geom_line(aes(y = population, color = "Population en centaine de millions"), size = 0.5) +
scale_color_manual(values = c("Décès en million" = "red", "Population en centaine de millions" = "blue")) +
labs(color = "Title", title = "Europe du sud et de l'ouest") +
scale_x_continuous(breaks = seq(1990, 2019, by = 4))
```

```
plotEasternEurope <- europeData %>%
  filter(`sub-region`=="Eastern Europe") %>%
  ggplot(aes(x = Year)) +
  geom_line(aes(y = death_numbers, color = "Décès en million"), size = 0.5) +
  geom_line(aes(y = population, color = "Population en centaine de millions"), size = 0.5) +
  scale_color_manual(values = c("Décès en million" = "red", "Population en centaine de millions" = "blue")) +
  labs(color = "Title", title = "Europe de l'est") +
  scale_x_continuous(breaks = seq(1990, 2019, by = 4))
```

```
plotSouthernAndWesternEurope
```



```
plotEasternEurope
```



### Asie de l'ouest, Asie du sud-est et Asie de l'est

```
sumAsiaPopulation <- data %>%
  filter(region=="Asia") %>%
  distinct(Entity,Year,population, `sub-region`) %>%
  group_by(Year, `sub-region`) %>%
  summarise(population=sum(population, na.rm = TRUE) / 100000000)
```

## `summarise()` has grouped output by 'Year'. You can override using the  
## `.groups` argument.

```
sumAsiaDeaths <- data %>%
  filter(region=="Asia") %>%
  group_by(Year, `sub-region`) %>%
  summarise(death_numbers=sum(`Death Numbers`, na.rm = TRUE) / 1000000)
```

## `summarise()` has grouped output by 'Year'. You can override using the  
## `.groups` argument.

```
asiaData <- left_join(sumAsiaPopulation, sumAsiaDeaths, by=c("Year"="Year", "sub-region"="sub-region"))
```

```
plotWesternAsia <- asiaData %>%
  filter(`sub-region`=="Western Asia") %>%
  ggplot(aes(x = Year)) +
  geom_line(aes(y = death_numbers, color = "Décès en million"), size = 0.5) +
  geom_line(aes(y = population, color = "Population en centaine de millions"), size = 0.5) +
```

```

scale_color_manual(values = c("Décès en million" = "red", "Population en centaine de millions" = "blue"),
labs(color = "Title", title = "Asie de l'ouest")+
scale_x_continuous(breaks = seq(1990, 2019, by = 4))

plotCentralAsia <- asiaData %>%
  filter(`sub-region`=="Central Asia") %>%
  ggplot(aes(x = Year)) +
  geom_line(aes(y = death_numbers, color = "Décès en million"), size = 0.5) +
  geom_line(aes(y = population, color = "Population en centaine de millions"), size = 0.5) +
  scale_color_manual(values = c("Décès en million" = "red", "Population en centaine de millions" = "blue"),
labs(color = "Title", title = "Asie central")+
  scale_x_continuous(breaks = seq(1990, 2019, by = 4))

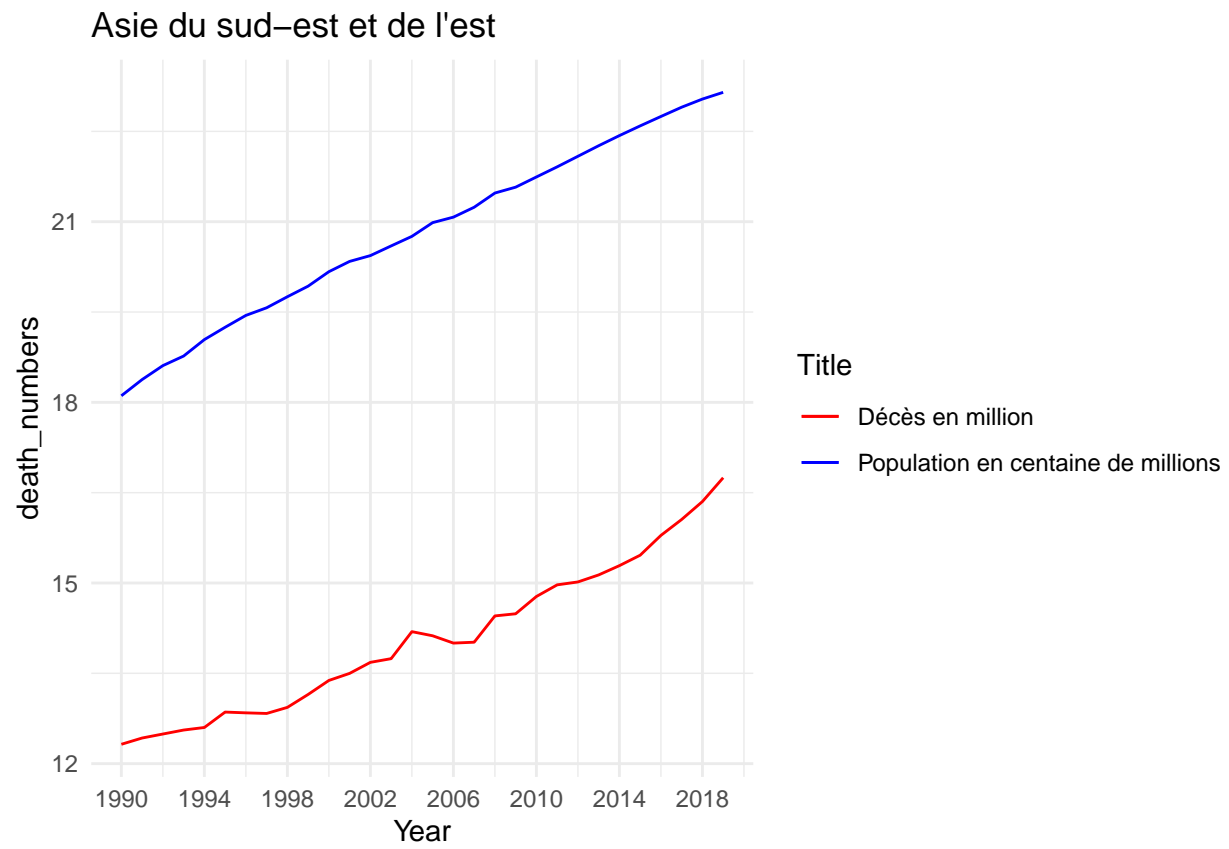
plotSouthEasternAndEsternAsia <- asiaData %>%
  filter(`sub-region`==c("South-eastern Asia", "Eastern Asia")) %>%
  group_by(Year) %>%
  summarise(death_numbers=sum(death_numbers), population=sum(population)) %>%
  ggplot(aes(x = Year)) +
  geom_line(aes(y = death_numbers, color = "Décès en million"), size = 0.5) +
  geom_line(aes(y = population, color = "Population en centaine de millions"), size = 0.5) +
  scale_color_manual(values = c("Décès en million" = "red", "Population en centaine de millions" = "blue"),
labs(color = "Title", title = "Asie du sud-est et de l'est") +
  scale_x_continuous(breaks = seq(1990, 2019, by = 4))

## Warning: There were 30 warnings in `filter()`.
## The first warning was:
## i In argument: ``sub-region` == c("South-eastern Asia", "Eastern Asia")`.
## i In group 1: `Year = 1990`.
## Caused by warning in `` `sub-region` == c("South-eastern Asia", "Eastern Asia") ``:
## ! la taille d'un objet plus long n'est pas multiple de la taille d'un objet plus court
## i Run `dplyr::last_dplyr_warnings()` to see the 29 remaining warnings.

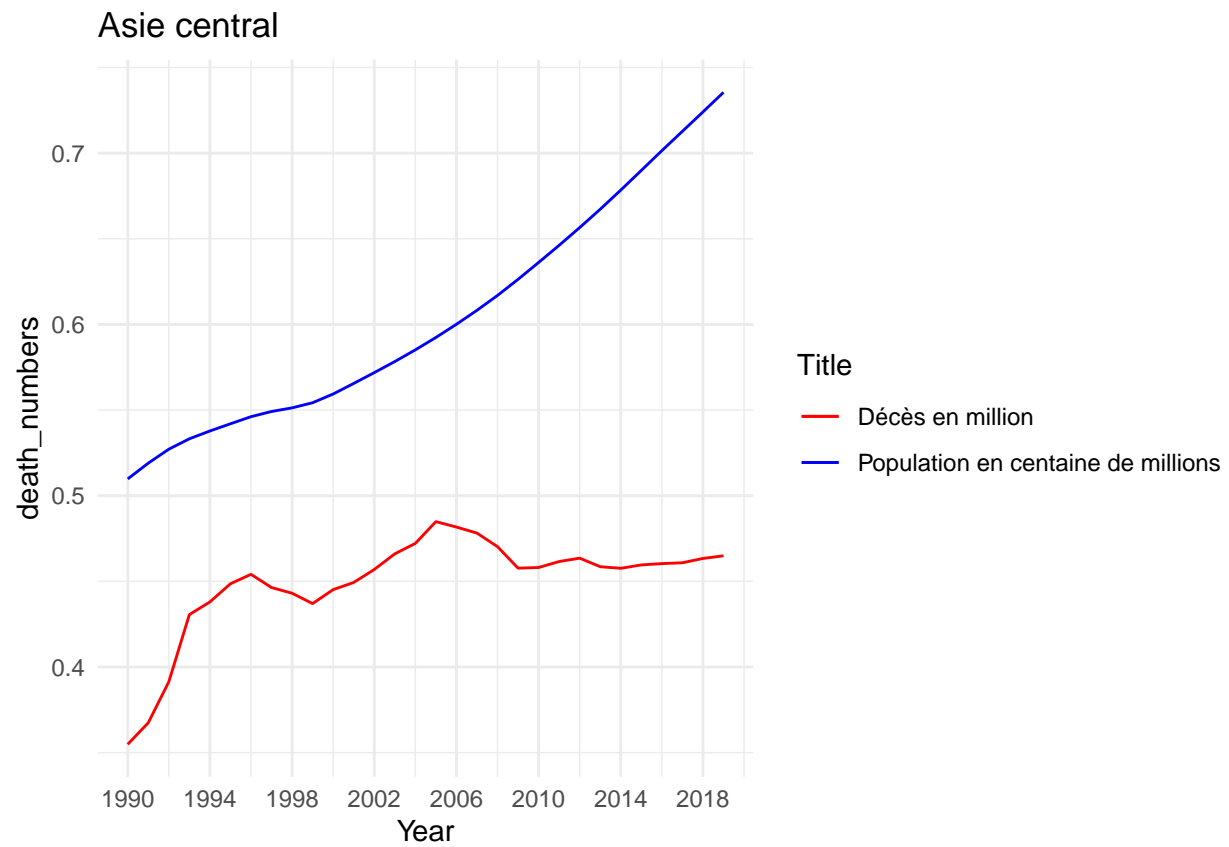
plotSouthEasternAndEsternAsia

```

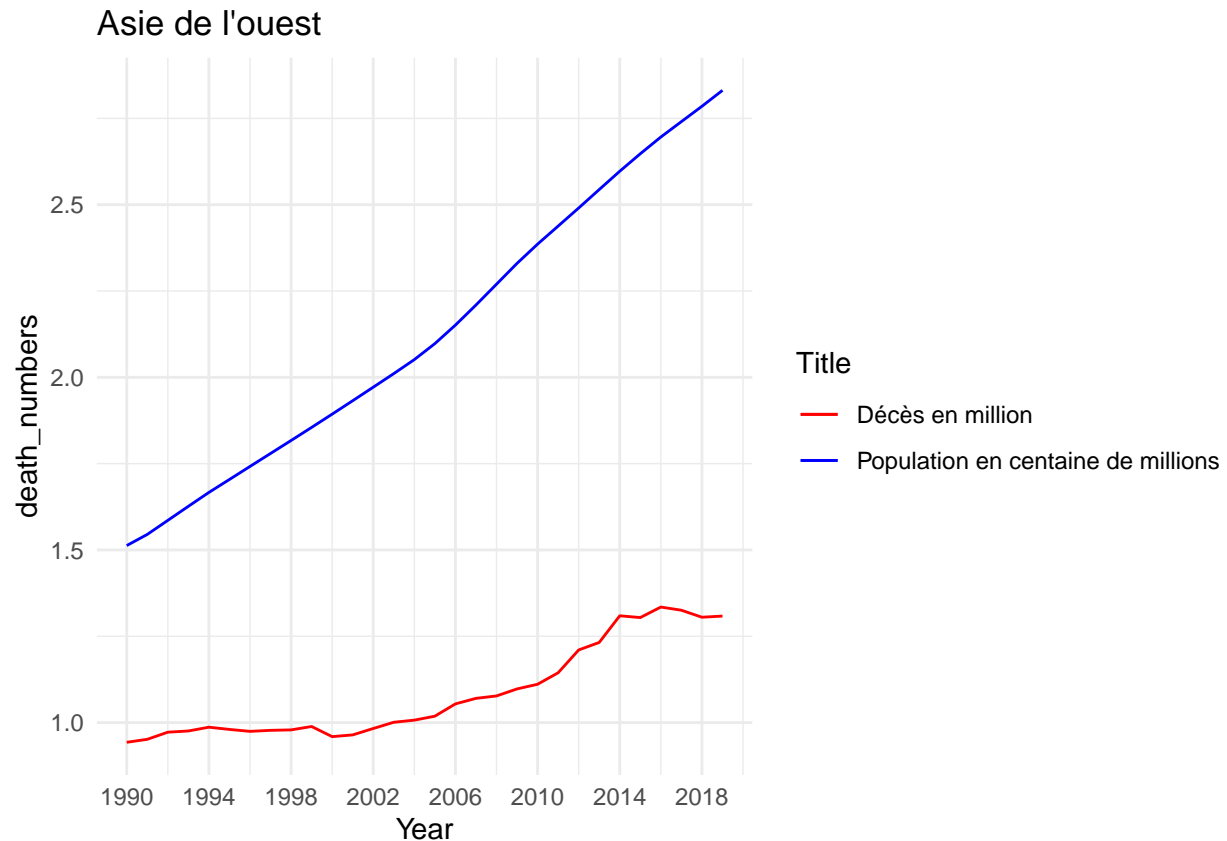




plotCentralAsia



plotWesternAsia



## Nuages de point

Ces nuages de point représentent en abscisse le PIB d'une sous-région, en ordonnée le taux de mortalité à une date donnée. Les formes des points représentent les causes de mortalité et la couleur le ratio de décès par cause.

### Sub-Sahara Africa en 2016

A la sortie de l'épidémie d'Ebola qui a sévi en Afrique de l'Ouest, principalement en Guinée, au Liberia et en #####Sierra Leone, a été l'une des plus graves de l'histoire de la maladie. Elle a également touché d'autres pays de #####la région, bien que dans une moindre mesure.

```
subSahara2015 <- data %>%
  filter(`sub-region` == "Sub-Saharan Africa") %>%
  group_by(Year)%>%
  filter(!is.na(gdp)) %>%
  filter(Year == 2015) %>%
  filter(!is.na(`Death Numbers`))%>%
  group_by(`Causes name`) %>%
  summarize(`Death Numbers`=sum(as.numeric(`Death Numbers`))) %>%
  arrange(desc(`Death Numbers`))%>%
  top_n(5);
```

## Selecting by Death Numbers

```

subSahara2 <- data %>%
  filter(`sub-region` == "Sub-Saharan Africa") %>%
  group_by(Year)%>%
  filter(!is.na(gdp)) %>%
  filter(!is.na(`Death Numbers`))%>%
  filter(Year == 2015) %>%
  distinct(Entity, population, gdp) %>%
  summarise(PIB = sum(gdp), population = sum(population))

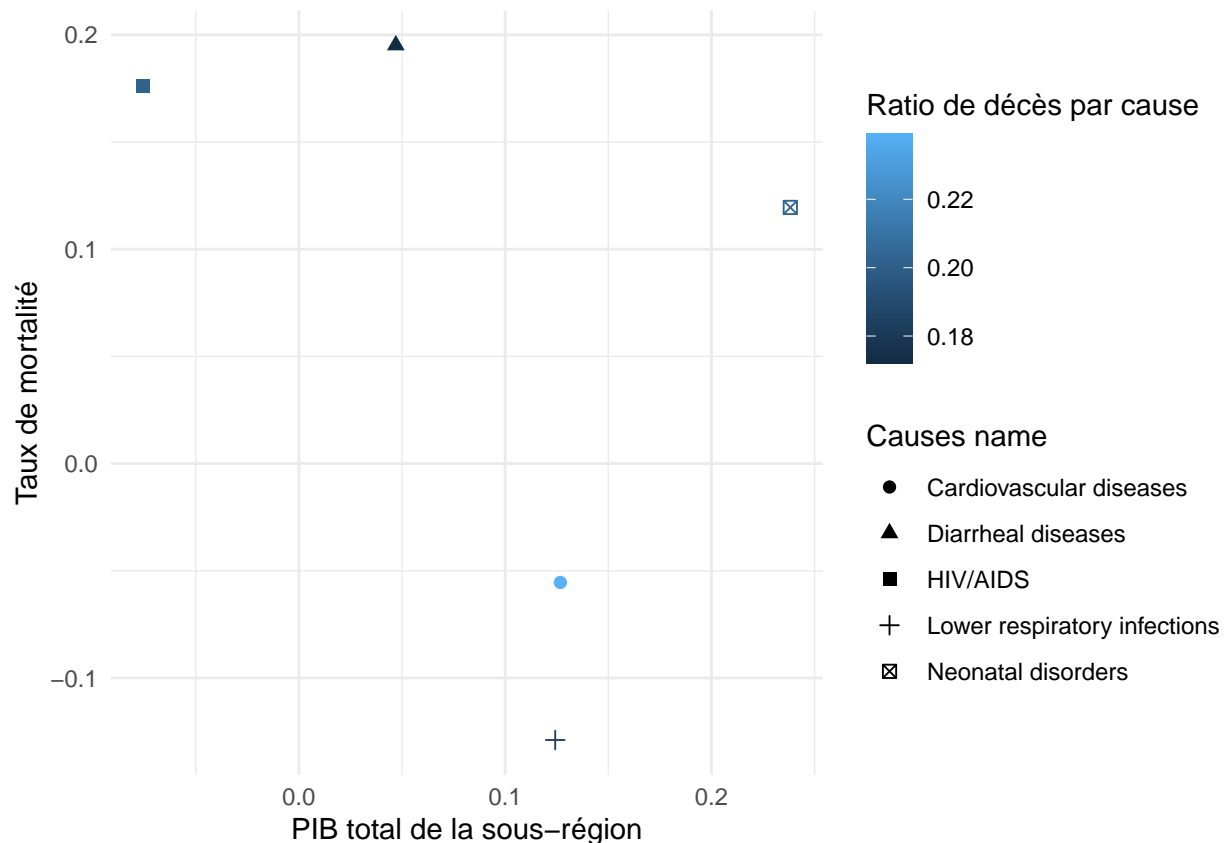
PIB_total <- sum(subSahara2$PIB) / 1000000000000000
total_population <- sum(subSahara2$population)
total_deces <- sum(subSahara2015$`Death Numbers`)

taux_mortalite = total_deces / total_population

ratio_deces_par_cause <- subSahara2015$`Death Numbers` / total_deces

ggplot(subSahara2015, aes(x = PIB_total, y = taux_mortalite, shape = `Causes name`, color = ratio_deces,
  geom_point(position = position_jitter(width = 0.3, height = 0.3), size = 2) +
  labs(x = "PIB total de la sous-région", y = "Taux de mortalité", color = "Ratio de décès par cause")

```



## Western Europe en 2010

2009 - 2010 une période qui pourrait être considérée comme particulièrement significative en Europe occidentale est celle de la pandémie de grippe H1N1

```
westernEurope2010 <- data %>%
  filter(`sub-region` == "Western Europe") %>%
  group_by(Year)%>%
  filter(!is.na(gdp)) %>%
  filter(Year == 2010) %>%
  filter(!is.na(`Death Numbers`))%>%
  group_by(`Causes name`) %>%
  summarize(`Death Numbers`=sum(as.numeric(`Death Numbers`))) %>%
  arrange(desc(`Death Numbers`))%>%
  top_n(5);
```

## Selecting by Death Numbers

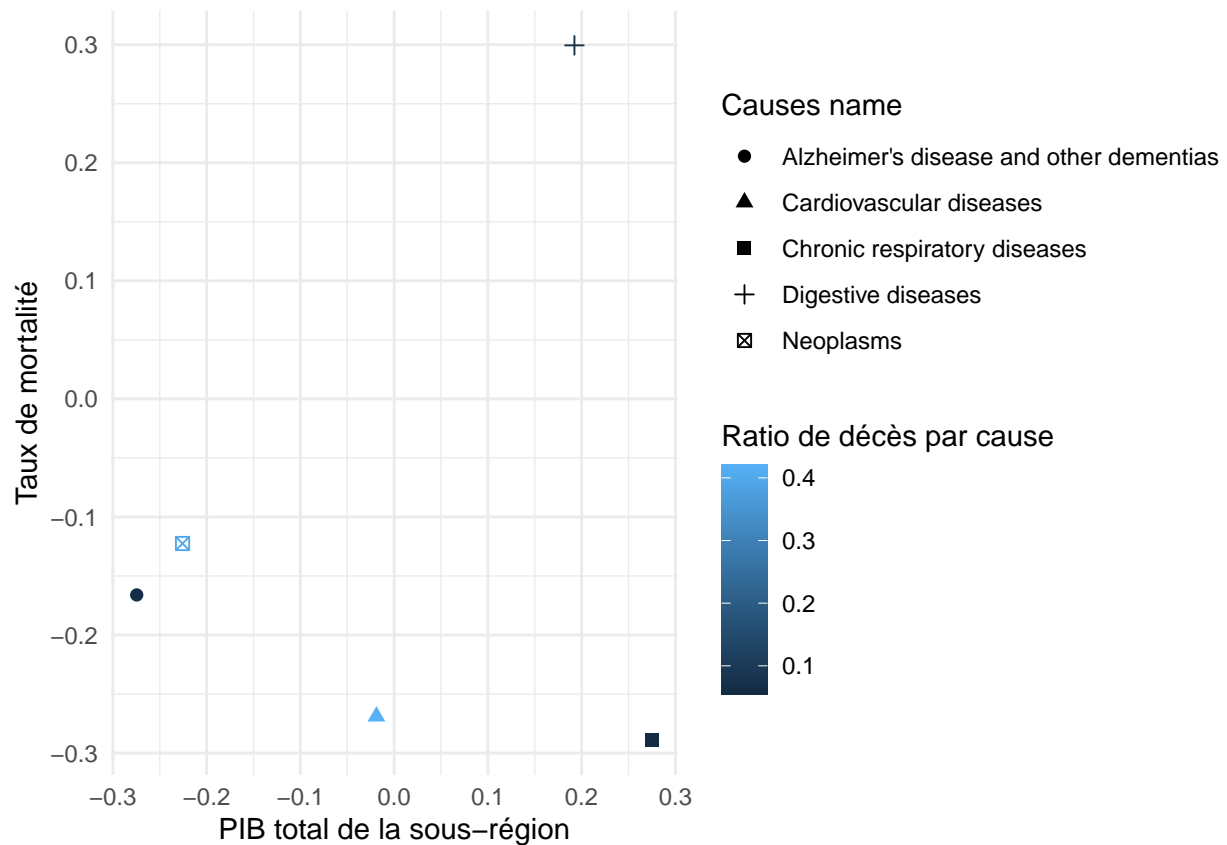
```
westernEurope2 <- data %>%
  filter(`sub-region` == "Western Europe") %>%
  group_by(Year)%>%
  filter(!is.na(gdp)) %>%
  filter(!is.na(`Death Numbers`))%>%
  filter(Year == 2010) %>%
  distinct(Entity, population, gdp) %>%
  summarise(PIB = sum(gdp), population = sum(population))
```

```
PIB_total <- sum(westernEurope2$PIB) / 1000000000000000
total_population <- sum(westernEurope2$population)
total_deces <- sum(westernEurope2010$`Death Numbers`)
```

```
taux_mortalite = total_deces / total_population
```

```
ratio_deces_par_cause <- westernEurope2010$`Death Numbers` / total_deces
```

```
ggplot(westernEurope2010, aes(x = PIB_total, y = taux_mortalite, shape = `Causes name`, color = ratio_d
  geom_point(position = position_jitter(width = 0.3, height = 0.3), size = 2) +
  labs(x = "PIB total de la sous-région", y = "Taux de mortalité", color = "Ratio de décès par cause")
```



## Eastern Asia en 2003

Une période qui pourrait être considérée comme particulièrement critique pour l'Asie de l'Est est celle

```
easternAsia2003 <- data %>%
  filter(`sub-region` == "Eastern Asia") %>%
  group_by(Year)%>%
  filter(!is.na(gdp)) %>%
  filter(Year == 2003) %>%
  filter(!is.na(`Death Numbers`))%>%
  group_by(`Causes name`) %>%
  summarize(`Death Numbers`=sum(as.numeric(`Death Numbers`))) %>%
  arrange(desc(`Death Numbers`))%>%
  top_n(5);
```

de l'épidémie de syndrome respiratoire aigu sévère (SRAS) en 2002-2003. Le SRAS, causé par un coronavirus, a ##### émergé en Chine continentale en novembre 2002 avant de se propager rapidement à d'autres pays de la région et ##### au-delà.

## Selecting by Death Numbers

```
easternAsia2 <- data %>%
  filter(`sub-region` == "Eastern Asia") %>%
  group_by(Year)%>%
  filter(!is.na(gdp)) %>%
```

```

filter(!is.na(`Death Numbers`))>%
filter(Year == 2003) %>%
  distinct(Entity, population, gdp) %>%
  summarise(PIB = sum(gdp), population = sum(population))

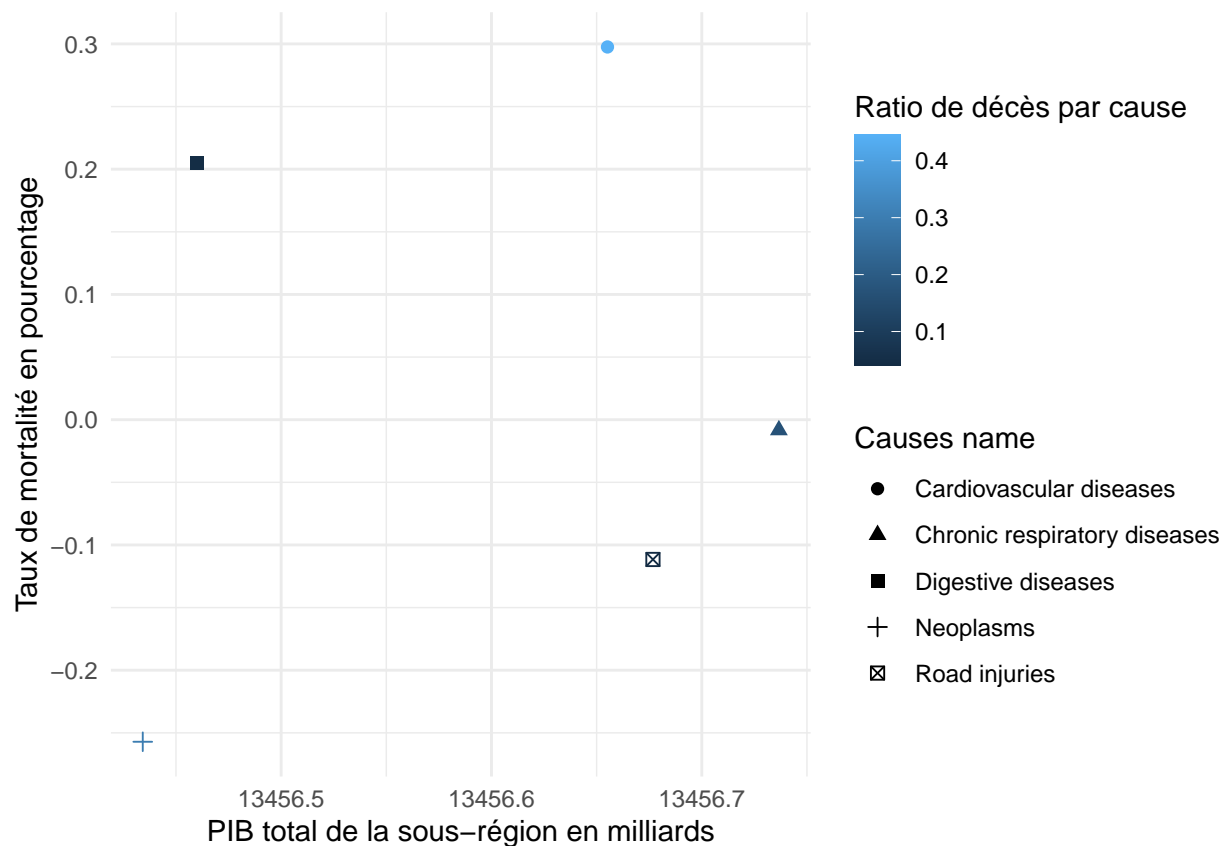
PIB_total <- sum(easternAsia2$PIB) / 1000000000
total_population <- sum(easternAsia2$population)
total_deces <- sum(easternAsia2003$`Death Numbers`)

taux_mortalite = (total_deces / total_population)

ratio_deces_par_cause <- (easternAsia2003$`Death Numbers` / total_deces)

ggplot(easternAsia2003, aes(x = PIB_total, y = taux_mortalite, shape = `Causes name`, color = ratio_deces)) +
  geom_point(position = position_jitter(width = 0.3, height = 0.3), size = 2) +
  labs(x = "PIB total de la sous-région en milliards", y = "Taux de mortalité en pourcentage", color = "Ratio de décès par cause")

```



## Bilan

Au terme de cette analyse, nous constatons diverses causes de Décès en million à travers le monde. Cela peut être dû au climat, aux habitudes qu'elles soient comportementales ou alimentaires. La cause qui revient le plus souvent est celle des maladies cardiovasculaires, cela s'explique par plusieurs raisons en fonction des sous régions entre-autres: Le vieillissement de la population, Urbanisation et changements de mode de vie (une alimentation riche en graisses et en sucres dans les zones urbaines), accès inégal aux soins de santé. Nous

observons également le maintien de l'écart entre la démographie en centaine de millions et les décès en million dans une région, c'est à dire plus la population augmente, plus le nombre de Décès en million augmente.