

# Mortality rates

Safar Farah, Benkhaled Mohammed Iyad, Ben Touati Oumaima, Belharet Ferhat

2025-05-20

## Table de matières :

- Introduction
- Pourquoi ce jeu de données
- Problématique
- Méthodologie
- Agrégation des données
- Analyse comparative

## Introduction

Notre travail présente une étude sur les taux de mortalité dans le monde après l'année 2021. Les données utilisées sont données par l'Organisation Mondiale de la Santé. Elles contiennent les données de mortalité détaillées entre l'année 2021 et fin 2023

SOURCE : <https://www.who.int/data/data-collection-tools/who-mortality-database>

Ce jeu de données nous permet d'analyser les tendances et les causes de décès, ce qui nous donne une meilleure compréhension de la santé publique.

## Pourquoi ce jeu de données

Nous avons choisi ces données car elles viennent de l'OMS, une source sérieuse et fiable. Ce sont des données complètes et récentes, ce qui est très pertinent pour nous aider à analyser les tendances actuelles. Le jeu de données contient des informations détaillées sur les décès par pays, année, cause, sexe et âge. De plus, les causes de décès sont classées selon la classification ICD-10, ce qui garantit des données standardisées.

## Problématique

Y a-t-il des différences importantes dans les taux de mortalité selon les causes des décès entre les hommes et les femmes ?

## Méthodologie

### Nettoyage des données

D'abord, nous allons charger les données et gérer les valeurs manquantes ou incorrectes, surtout pour les colonnes de décès et de sexe. Nous nous assurerons que les sexes (homme/femme) sont bien identifiés, et que toutes les données sont standardisées.

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

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

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

```
# On charge et organise les données
data <- read.csv("~/Desktop/Morticd10_part6")

data$SexeFM <- case_when(
  data$Sex == 1 ~ "Male",
  data$Sex == 2 ~ "Female",
  TRUE ~ NA
)

data$SexeFM <- factor(data$SexeFM, levels = c("Male", "Female"))

# Nb de morts
data <- data %>%
  mutate(across(starts_with("Deaths"), ~ as.numeric(as.character(.))))

data$TotalDeaths <- rowSums(data %>% select(starts_with("Deaths")), na.rm = TRUE)
```

## Agrégation des données

```
# Filtré les données
dataFiltered <- data %>%
  filter(!is.na(SexeFM), TotalDeaths > 0)

plotData <- dataFiltered %>%
  group_by(Cause, SexeFM) %>%
  summarise(TotalMorts = sum(TotalDeaths, na.rm = TRUE)) %>%
  ungroup()
```

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

```
top15 <- plotData %>%
  group_by(Cause) %>%
  summarise(Total = sum(TotalMorts)) %>%
  arrange(desc(Total)) %>%
```

```

head(15) %>%
  pull(Cause)

plotTop15 <- plotData %>%
  filter(Cause %in% top15)

plotTop15$Cause <- factor(plotTop15$Cause,
  levels = top15)

# Maladies et codes
codeMaladieMap <- c(
  "U071" = "Covid-19",
  "I219" = "Ischaemic heart disease",
  "C349" = "Trachea, bronchus, lung cancers",
  "R99" = "Ill-defined and unknown cause of mortality",
  "I251" = "Atherosclerotic heart disease of native coronary artery",
  "J189" = "Pneumonia",
  "F03" = "Dementia",
  "I509" = "Heart failure",
  "J449" = "Chronic obstructive pulmonary disease",
  "I259" = "Chronic ischemic heart disease",
  "G309" = "Alzheimer's disease",
  "I64" = "Family history of stroke",
  "R54" = "Age-related physical debility",
  "C509" = "Malignant neoplasm of breast",
  "C61" = "Malignant neoplasm of prostate"
)

cle <- paste(
  paste(top15, codeMaladieMap[top15], sep = ": ", collapse = "\n"),
  sep = ""
)

```

## Analyse comparative

```

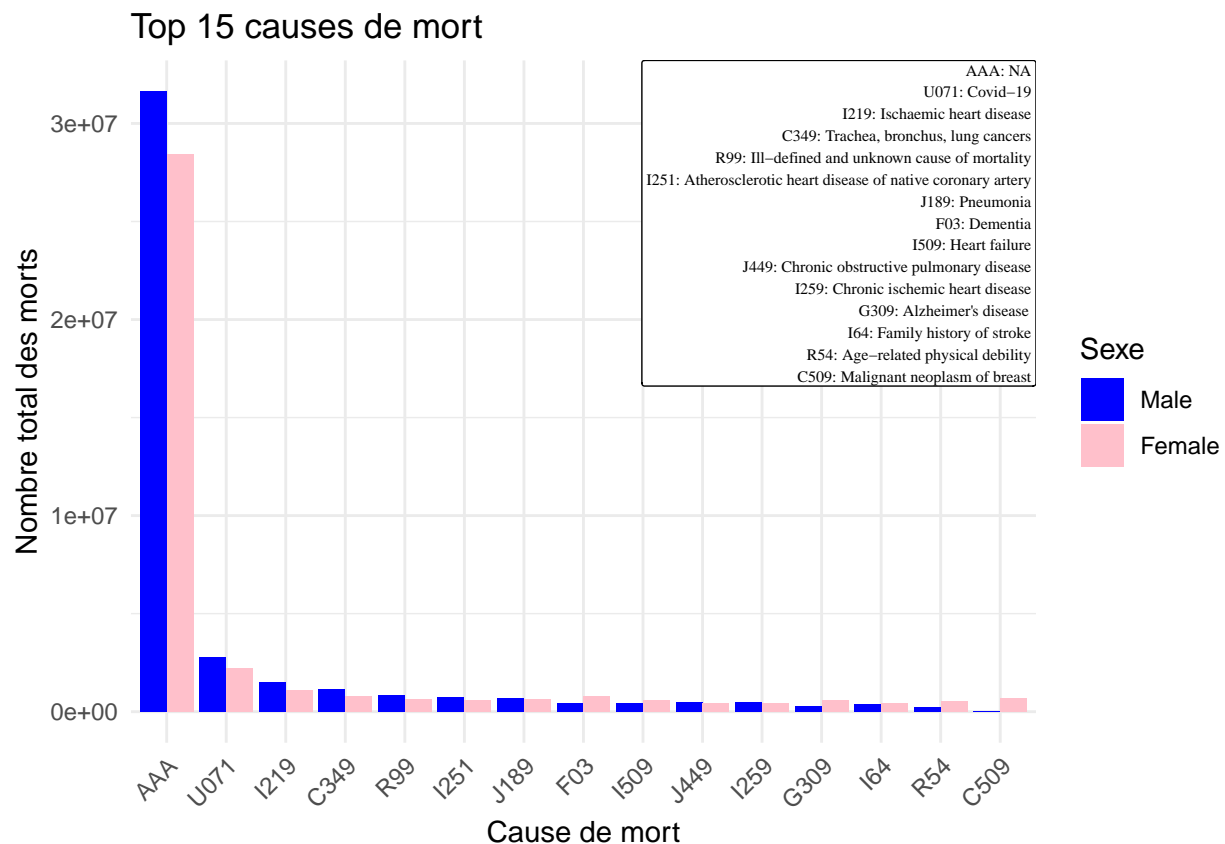
# Dessiner le graph
ggplot(plotTop15, aes(x = Cause, y = TotalMorts, fill = SexeFM)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "Top 15 causes de mort",
    x = "Cause de mort",
    y = "Nombre total des morts",
    fill = "Sexe"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_fill_manual(values = c("Male" = "blue", "Female" = "pink")) +
  annotate(
    geom = "label",
    x = Inf,
    y = Inf,
    label = cle,

```

```

hjust = 1,
vjust = 1,
color = "black",
size = 2,
label.size = 0.25,
fill = "white",
family = "serif"
)

```



Nous remarquons ici que la première chose dans le graphique est AAA qui a une valeur largement supérieure au reste des valeurs. Après avoir consulté la documentation de l'OMS, il s'est avéré que c'est le nombre total de morts, ce qui est intéressant, mais cela pollue notre graphique. Donc, nous l'avons supprimé et nous modifions notre code en changeant `head(15)` par `slice(2:16)`:

```

top15 <- plotData %>%
  group_by(Cause) %>%
  summarise(Total = sum(TotalMorts)) %>%
  arrange(desc(Total)) %>%
  slice(2:16) %>%
  pull(Cause)

plotTop15 <- plotData %>%
  filter(Cause %in% top15)

plotTop15$Cause <- factor(plotTop15$Cause,
  levels = top15)

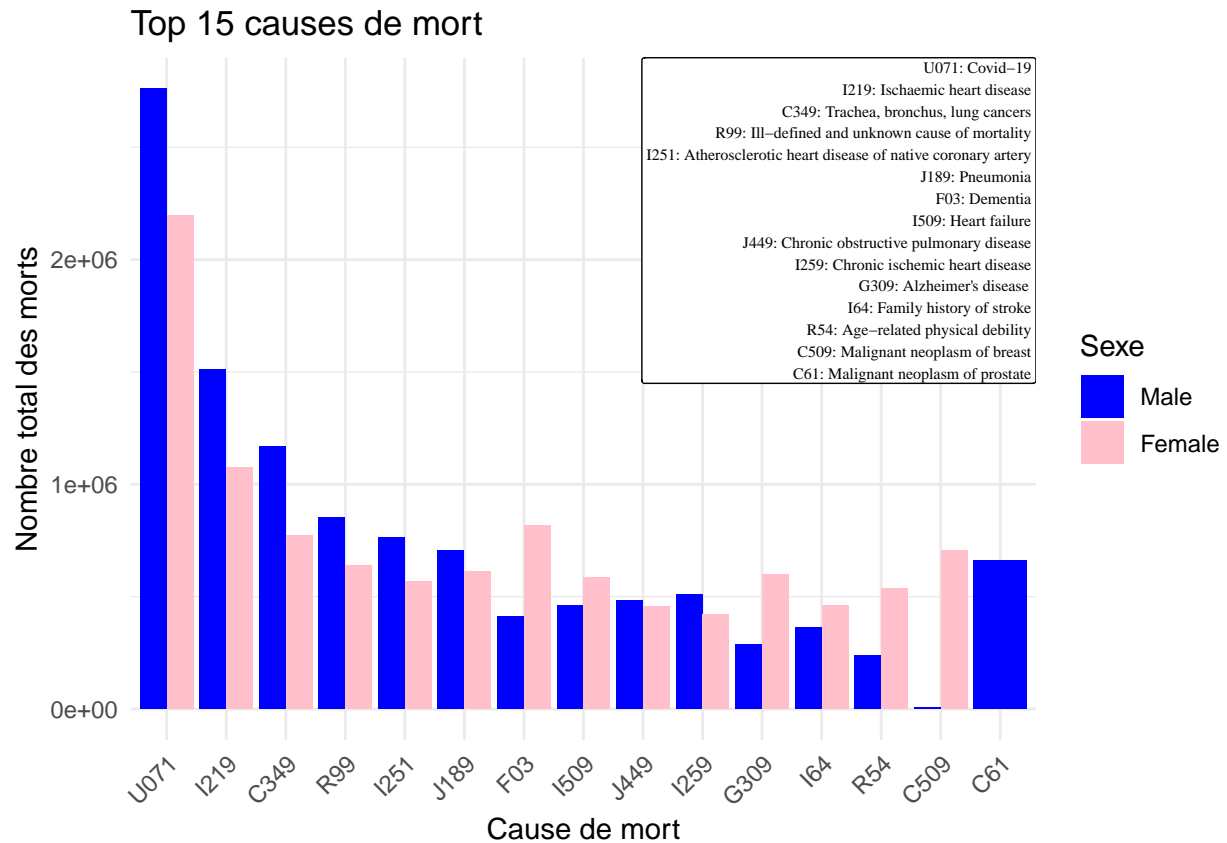
```

```

cle <- paste(
  paste(top15, codeMaladieMap[top15], sep = ": ", collapse = "\n"),
  sep = ""
)

# Dessiner le graph
ggplot(plotTop15, aes(x = Cause, y = TotalMorts, fill = SexeFM)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "Top 15 causes de mort",
    x = "Cause de mort",
    y = "Nombre total des morts",
    fill = "Sexe"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_fill_manual(values = c("Male" = "blue", "Female" = "pink")) +
  annotate(
    geom = "label",
    x = Inf,
    y = Inf,
    label = cle,
    hjust = 1,
    vjust = 1,
    color = "black",
    size = 2,
    label.size = 0.25,
    fill = "white",
    family = "serif"
  )

```



Le nouveau graphique illustre bien ce que nous voulons explorer.

Il montre des différences de mortalité entre les hommes et les femmes pour les 15 causes de mort les plus fréquentes. Les hommes ont généralement plus de décès dans la quasi totalité des catégories, notamment pour les maladies cardiovasculaires, les cancers et les maladies respiratoires. Cela correspond aux tendances de santé publique connues liées au tabagisme et aux expositions professionnelles.

Par ailleurs, les femmes présentent une mortalité plus élevée pour des maladies comme l'alzheimer et la démence. Cela s'explique en partie par une espérance de vie généralement plus longue chez les femmes, augmentant leur probabilité d'atteindre les âges où ces maladies sont plus fréquentes.

On remarque aussi que les deux cancers, le cancer du sein et le cancer du prostate sont présents dans le graphique, il occupe la 14ème et la 15ème cause respectivement, chacune des deux causes est dominée par un seul genre.