

Analyse de l'impact covid

Charlotte Gaudin

Table des matières

Préparation des données	1
Visualisation des données	5

Préparation des données

Les données de l'évolution du nombre de cas de Covid_19 au cours du temps sont disponibles sur [Github](#). Nous les récupérons sous forme d'un fichier en format CSV dont chaque ligne correspond à un Pays/Province. L'URL est : "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv"

Téléchargement :

```
library(readr)
data = read_csv("https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_cov

##
## -- Column specification -----
## cols(
##   .default = col_double(),
##   'Province/State' = col_character(),
##   'Country/Region' = col_character()
## )
## i Use 'spec()' for the full column specifications.

View(data)
```

Sélection des données :

Je garde uniquement la Belgique (Belgium), la Chine - toutes les provinces sauf Hong-Kong (China), Hong Kong (China, Hong-Kong), la France métropolitaine

(France), l'Allemagne (Germany), l'Iran (Iran), l'Italie (Italy), le Japon (Japan), la Corée du Sud (Korea, South), la Hollande sans les colonies (Netherlands), le Portugal (Portugal), l'Espagne (Spain), le Royaume-Unis sans les colonies (United Kingdom), les États-Unis (US).

```
data2 = subset(data,
data$`Country/Region`=='Belgium'|data$`Country/Region`=='China'|
data$`Country/Region`=='France'|data$`Country/Region`=='Germany'|
data$`Country/Region`=='Iran'|data$`Country/Region`=='Italy'|
data$`Country/Region`=='Japan'|data$`Country/Region`=='Korea, South'|
data$`Country/Region`=='Netherlands'|data$`Country/Region`=='Portugal'|
data$`Country/Region`=='Spain'|data$`Country/Region`=='United Kingdom'|
data$`Country/Region`=='US')
View(data2)
less = c(35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 52, 53, 54,
        55, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69)
data3 = data2[-less,]
View(data3)
data4 = data3[,-3]
data4 = data4[,-3]
View(data4)
```

Modification du tableau :

Les dates sont rassemblées dans une seule colonne. Une colonne Pays_Date a été ajoutée. Les colonnes Long, Lat et Province/State ont été éliminées.

```
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

library(reshape2)
table_pays = melt(data4, id.vars = 'Country/Region',
                  measure.vars = colnames(select(data4,
                                                  ends_with("/20"))))
table_pays = table_pays[!is.na(table_pays$value),]
table_pays$Pays_Date = paste(table_pays$'Country/Region',
                             table_pays$'variable',sep="_")
```

```

table_pays2 = melt(data4, id.vars = 'Country/Region',
                   measure.vars = colnames(select(data4,
                                                  ends_with("/21"))))

table_pays2$Pays_Date = paste(table_pays2$'Country/Region',
                              table_pays2$'variable', sep="_")
table_pays2 = table_pays2[!is.na(table_pays2$value),]
table_pays3 = bind_rows(table_pays, table_pays2)
View(table_pays3)

```

Fusion des lignes par mêmes pays et mêmes dates. Par exemple, toutes les provinces de Chine ont été rassemblées ensemble pour les mêmes dates.

```

nb_cas_pays = aggregate(value ~ Pays_Date, data= table_pays3, sum)
library(tidyr)

```

```

##
## Attaching package: 'tidyr'

## The following object is masked from 'package:reshape2':
##
## smiths

```

```

nb_cas_pays = separate(nb_cas_pays, "Pays_Date", c("Pays", "Date"),
                      sep = "_")
nb_cas_pays$Pays_Date = paste(nb_cas_pays$'Pays',
                              nb_cas_pays$'Date', sep="_")
View(nb_cas_pays)

```

Regardons ce que nous avons obtenu :

```
head(nb_cas_pays)
```

```

##   Pays   Date  value   Pays_Date
## 1 Belgium 1/1/21 648289 Belgium_1/1/21
## 2 Belgium 1/10/21 664263 Belgium_1/10/21
## 3 Belgium 1/11/21 665223 Belgium_1/11/21
## 4 Belgium 1/12/21 667322 Belgium_1/12/21
## 5 Belgium 1/13/21 670249 Belgium_1/13/21
## 6 Belgium 1/14/21 672886 Belgium_1/14/21

```

```
tail(nb_cas_pays)
```

```

##   Pays   Date  value   Pays_Date
## 5299  US 9/4/20 6202700 US_9/4/20

```

```
## 5300 US 9/5/20 6245682 US_9/5/20
## 5301 US 9/6/20 6276981 US_9/6/20
## 5302 US 9/7/20 6300449 US_9/7/20
## 5303 US 9/8/20 6327884 US_9/8/20
## 5304 US 9/9/20 6361638 US_9/9/20
```

Les deux colonnes qui nous intéressent sont value et Pays. Vérifions leurs classes :

```
class(nb_cas_pays$'value')
```

```
## [1] "numeric"
```

```
class(nb_cas_pays$'Pays')
```

```
## [1] "character"
```

La variable value est numeric et Pays est character.

Y a-t-il des points manquants dans nos données?

```
na_records = apply(nb_cas_pays, 1, function (x) any(is.na(x)))
nb_cas_pays[na_records,]
```

```
## [1] Pays      Date      value     Pays_Date
## <0 rows> (or 0-length row.names)
```

Conversion des numéros de semaine :

```
library("lubridate")
```

```
##
## Attaching package: 'lubridate'

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

```
library("magrittr")
```

```
##
## Attaching package: 'magrittr'

## The following object is masked from 'package:tidyr':
##
##   extract
```

```
library(parsedate)
```

```
##  
## Attaching package: 'parsedate'  
  
## The following object is masked from 'package:readr':  
##  
## parse_date
```

Modification du format de la date dans la colonne 'convert_date'

```
nb_cas_pays$convert_date = mdy(nb_cas_pays$Date)
```

Trie des points sont dans l'ordre chronologique :

```
nb_cas_pays = nb_cas_pays[order(nb_cas_pays$convert_date),]
```

Je divise le jeu de données en petits tableaux en fonction des différents pays.

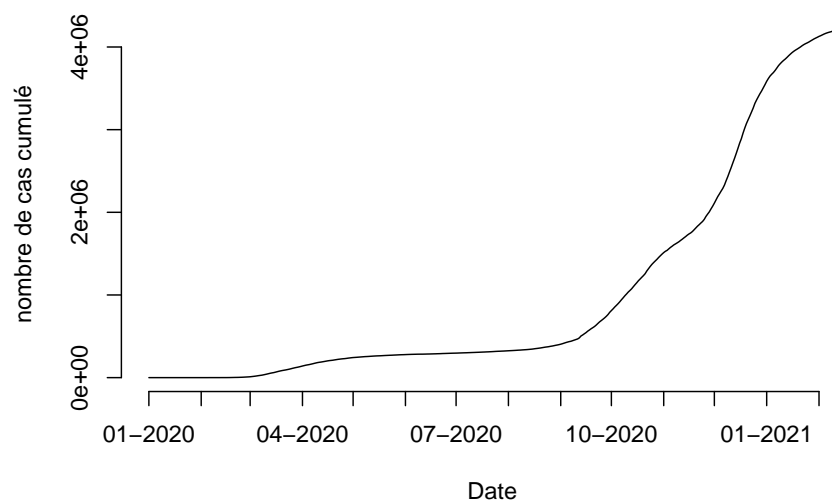
```
library(dplyr)  
table_China = subset(nb_cas_pays, nb_cas_pays$Pays == "China")  
table_Belgium = subset(nb_cas_pays, nb_cas_pays$Pays == "Belgium")  
table_Hong_Kong = subset(data4, data4$`Province/State` == "Hong_Kong")  
table_France = subset(nb_cas_pays, nb_cas_pays$Pays == "France")  
table_Germany = subset(nb_cas_pays, nb_cas_pays$Pays == "Germany")  
table_Iran = subset(nb_cas_pays, nb_cas_pays$Pays == "Iran")  
table_Italy = subset(nb_cas_pays, nb_cas_pays$Pays == "Italy")  
table_Japan = subset(nb_cas_pays, nb_cas_pays$Pays == "Japan")  
table_Korea = subset(nb_cas_pays, nb_cas_pays$Pays == "Korea, South")  
table_Netherlands = subset(nb_cas_pays, nb_cas_pays$Pays == "Netherlands")  
table_Portugal = subset(nb_cas_pays, nb_cas_pays$Pays == "Portugal")  
table_Spain = subset(nb_cas_pays, nb_cas_pays$Pays == "Spain")  
table_UK = subset(nb_cas_pays, nb_cas_pays$Pays == "United Kingdom")  
table_US = subset(nb_cas_pays, nb_cas_pays$Pays == "US")
```

Visualisation des données

Visualisation d'une partie des données :

```
plot_UK = plot(x=table_UK$convert_date, y=table_UK$value,  
type="l", axes=F, xlab="Date", ylab="nombre de cas cumulé",  
main="Nombre de cas cumulé en fonction de la date pour les UK")  
axis.Date(side=1, at=seq(min(table_UK$convert_date),  
max(table_UK$convert_date), by="months"), format="%m-%Y")  
axis(side=2, at=seq(0e+00, 4e+06, by=1e+06))
```

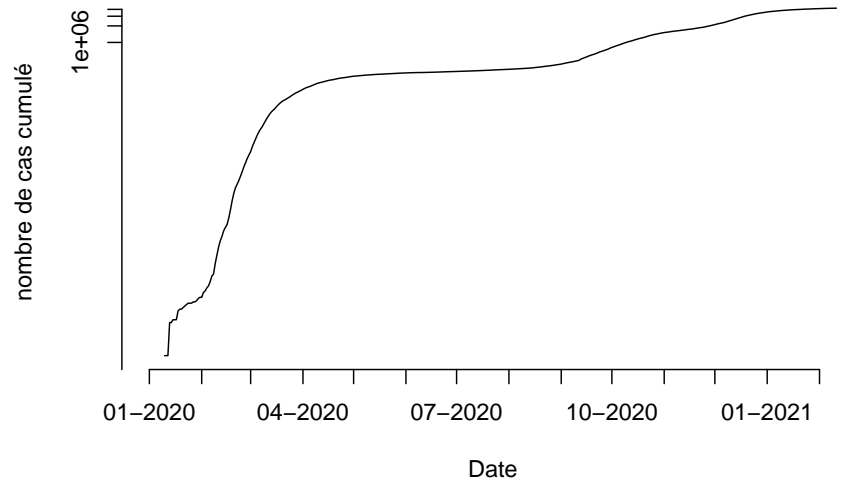
Nombre de cas cumulé en fonction de la date pour les UK



Plot UK avec une échelle linéaire :

```
plot_UK = plot(x=table_UK$convert_date, y=table_UK$value,  
type="l", axes=F, xlab="Date", ylab="nombre de cas cumulé",  
main="Nombre de cas cumulé en fonction de la date pour les UK", log = "y")  
axis.Date(side=1, at=seq(min(table_UK$convert_date),  
max(table_UK$convert_date), by="months"), format="%m-%Y")  
axis(side=2, at=seq(0e+00, 4e+06, by=1e+06))
```

Nombre de cas cumulé en fonction de la date pour les UK



Plot UK échelle logarithmique :

Visualisation de toutes les données :

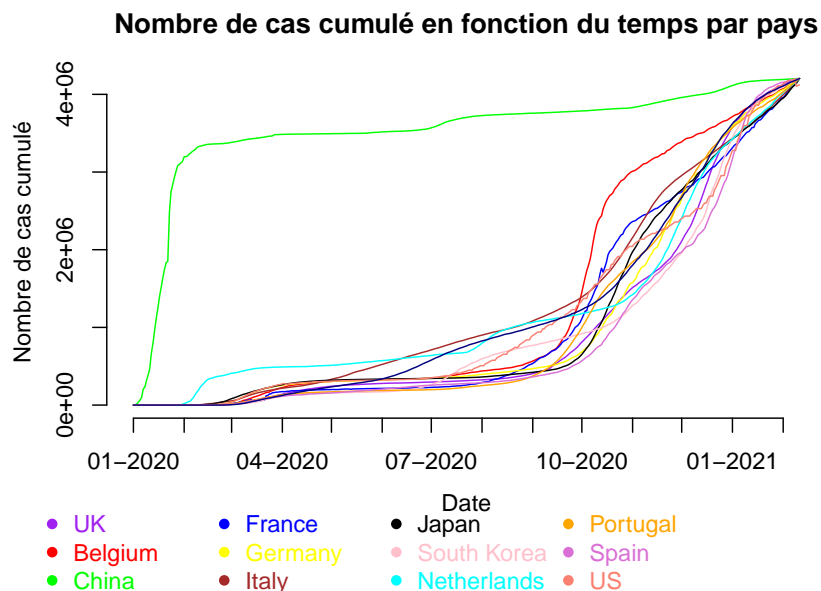
```
plot.new()
par(mar=c(7,4,3,3))
plot_UK = plot(x=table_UK$convert_date, y=table_UK$value,
              type="l", axes=F, xlab="", ylab="", col="purple")
axis(side=2, at=seq(0e+00, 4e+06, by=1e+06))
mtext("Nombre de cas cumulé", side=2, line=2.5)
par(new=T)
plot_Belgium = plot(x=table_Belgium$convert_date,
                   y=table_Belgium$value, type="l", axes=F, xlab="", ylab="", col="red")
par(new=T)
plot_China = plot(x=table_China$convert_date,
                  y=table_China$value, type="l", axes=F, xlab="", ylab="", col="green")
par(new=T)
plot_France = plot(x=table_France$convert_date,
                   y=table_France$value, type="l", axes=F, xlab="", ylab="", col="blue")
par(new=T)
plot_Germany = plot(x=table_Germany$convert_date,
                    y=table_Germany$value, type="l", axes=F, xlab="", ylab="", col="yellow")
par(new=T)
plot_Iran = plot(x=table_Iran$convert_date,
                 y=table_Iran$value, type="l", axes=F, xlab="", ylab="", col="brown")
par(new=T)
plot_Italy = plot(x=table_Italy$convert_date,
                  y=table_Italy$value, type="l", axes=F, xlab="", ylab="", col="black")
```

```

par(new=T)
plot_Japan = plot(x=table_Japan$convert_date,
y=table_Japan$value, type="l", axes=F, xlab="", ylab="", col="pink")
par(new=T)
plot_Korea = plot(x=table_Korea$convert_date,
y=table_Korea$value, type="l", axes=F, xlab="", ylab="", col="cyan")
par(new=T)
plot_Netherlands = plot(x=table_Netherlands$convert_date,
y=table_Netherlands$value, type="l", axes=F, xlab="", ylab="", col="orange")
par(new=T)
plot_Portugal = plot(x=table_Portugal$convert_date,
y=table_Portugal$value, type="l", axes=F, xlab="", ylab="", col="orchid")
par(new=T)
plot_Spain = plot(x=table_Spain$convert_date,
y=table_Spain$value, type="l", axes=F, xlab="", ylab="", col="salmon")
par(new=T)
plot_US = plot(x=table_US$convert_date,
y=table_US$value, type="l", axes=F, xlab="", ylab="", col="navy")
axis.Date(side=1, at=seq(min(table_UK$convert_date),
max(table_UK$convert_date), by="months"), format="%m-%Y")
mtext("Date", side=1, line=2.5)
legend(x="top",
legend = c("UK", "Belgium", "China", "France", "Germany", "Italy",
"Japan", "South Korea", "Netherlands", "Portugal", "Spain", "US"),
text.col = c("purple", "red", "green", "blue", "yellow", "brown",
"black", "pink", "cyan", "orange", "orchid", "salmon", "navy"),
pch=c(16), col = c("purple", "red", "green", "blue", "yellow",
"brown", "black", "pink", "cyan", "orange", "orchid", "salmon",
"navy"), ncol = 5, inset = 1.22, xpd = 1, box.lty = 0, bg = 'transparent')
title(main = "Nombre de cas cumulé en fonction du temps par pays")

```


Représentation graphique des courbes des différents pays avec échelle linéaire :



Représentation graphique des courbes des différents pays avec échelle

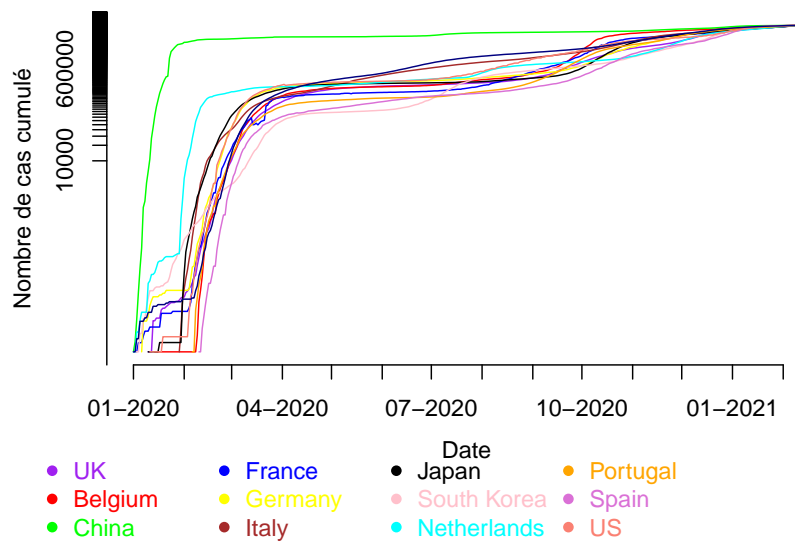
```
plot.new()
par(mar=c(7,4,3,3))
plot_UK = plot(x=table_UK$convert_date, y=table_UK$value, type="l",
              axes=F, xlab="", ylab="", col="purple", log = "y")
axis(side=2, log = "y", at=seq(0e+00, 1e+07, by=1e+04))
mtext("Nombre de cas cumulé", side=2, line=2.5)
par(new=T)
plot_Belgium = plot(x=table_Belgium$convert_date,
                   y=table_Belgium$value, type="l", axes=F, xlab="", ylab="", col="red", log = "y")
par(new=T)
plot_China = plot(x=table_China$convert_date,
                  y=table_China$value, type="l", axes=F, xlab="", ylab="", col="green", log = "y")
par(new=T)
plot_France = plot(x=table_France$convert_date,
                   y=table_France$value, type="l", axes=F, xlab="", ylab="", col="blue", log = "y")
par(new=T)
plot_Germany = plot(x=table_Germany$convert_date,
                    y=table_Germany$value, type="l", axes=F, xlab="", ylab="", col="yellow", log = "y")
par(new=T)
plot_Iran = plot(x=table_Iran$convert_date,
                 y=table_Iran$value, type="l", axes=F, xlab="", ylab="", col="brown", log = "y")
par(new=T)
```

```

plot_Italy = plot(x=table_Italy$convert_date,
y=table_Italy$value, type="l", axes=F, xlab="", ylab="", col="black", log = "y")
par(new=T)
plot_Japan = plot(x=table_Japan$convert_date,
y=table_Japan$value, type="l", axes=F, xlab="", ylab="", col="pink", log = "y")
par(new=T)
plot_Korea = plot(x=table_Korea$convert_date,
y=table_Korea$value, type="l", axes=F, xlab="", ylab="", col="cyan", log = "y")
par(new=T)
plot_Netherlands = plot(x=table_Netherlands$convert_date,
y=table_Netherlands$value, type="l", axes=F, xlab="", ylab="",
col="orange", log = "y")
par(new=T)
plot_Portugal = plot(x=table_Portugal$convert_date, y=table_Portugal$value,
type="l", axes=F, xlab="", ylab="", col="orchid", log = "y")
par(new=T)
plot_Spain = plot(x=table_Spain$convert_date, y=table_Spain$value, type="l",
axes=F, xlab="", ylab="", col="salmon", log = "y")
par(new=T)
plot_US = plot(x=table_US$convert_date, y=table_US$value, type="l", axes=F,
xlab="", ylab="", col="navy", log = "y")
axis.Date(side=1, at=seq(min(table_UK$convert_date), max(table_UK$convert_date),
by="months"), format="%m-%Y")
mtext("Date", side=1, line=2.5)
legend(x="top",
legend = c("UK", "Belgium", "China", "France", "Germany", "Italy", "Japan",
"South Korea", "Netherlands", "Portugal", "Spain", "US"),
text.col = c("purple", "red", "green", "blue", "yellow", "brown", "black",
"pink", "cyan", "orange", "orchid", "salmon", "navy"), pch=c(16),
col = c("purple", "red", "green", "blue", "yellow", "brown", "black", "pink",
"cyan", "orange", "orchid", "salmon", "navy"), ncol = 5, inset = 1.22, xpd = 1,
box.lty = 0, bg = 'transparent')
title(main = "Nombre de cas cumulé en fonction du temps par pays")

```

Nombre de cas cumulé en fonction du temps par pays



logarithmique :