

# Projet Covid

Margaux Bailleul / Oriane Duclos / Marie Guibert

2023-05-02

```
library(dplyr)
library(forecast)
```

## Importations des données

Tout d'abord, on importe les données et on sélectionne les données concernant la France.

```
donnees_fr <- read.csv("covid_france.csv", sep=";", stringsAsFactors = T)
summary(donnees_fr)
```

```
##           date           new_cases
## 2020-01-03:    1   Min.      :    0
## 2020-01-04:    1   1st Qu.:  2968
## 2020-01-05:    1   Median : 12174
## 2020-01-06:    1   Mean   : 32315
## 2020-01-07:    1   3rd Qu.: 32913
## 2020-01-08:    1   Max.    :500563
## (Other)      :1197   NA's   :1
```

Les données ci-dessus comprennent une variable temporelle et une variable caractérisée par un enregistrement journalier des nouveaux cas de Covid-19 en France.

```
min(as.character(donnees_fr$date))
```

```
## [1] "2020-01-03"
```

```
max(as.character(donnees_fr$date))
```

```
## [1] "2023-04-19"
```

Grâce à cette étape, nous pouvons observer que notre série temporelle débute le 1er Mars 2020 et se termine le 19 Septembre 2023. Notre étude a donc une plage d'environ de 3 ans et demi.

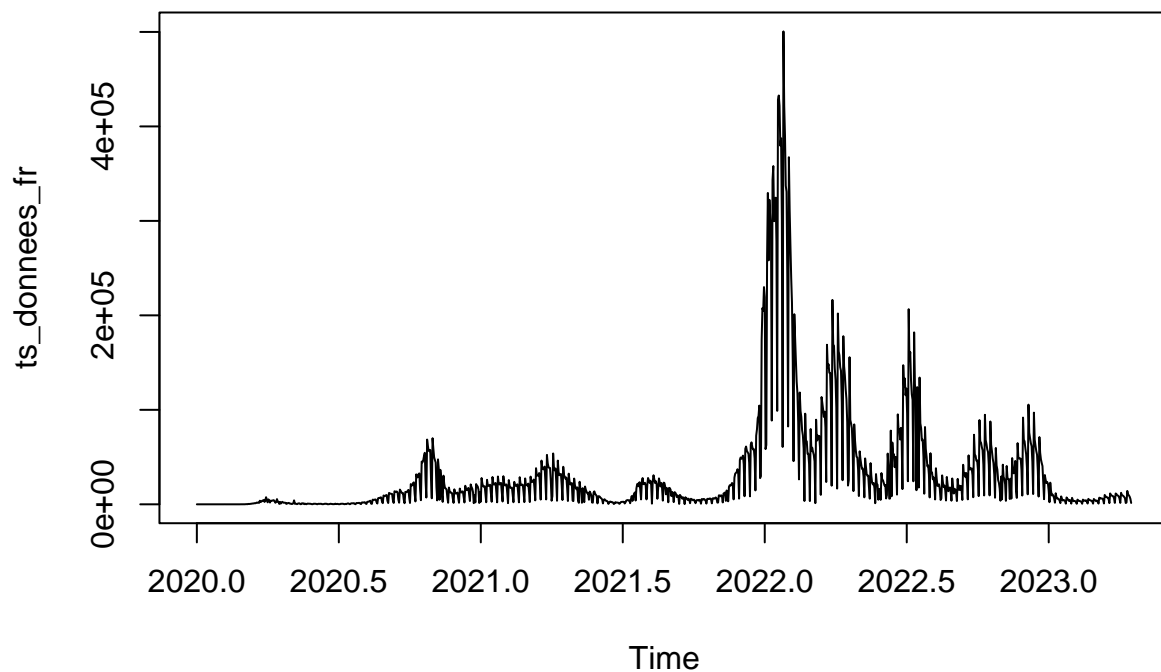
## Transformation des données en série temporelle

Premièrement, nous allons transformer nos données en séries temporelles pour pouvoir réaliser notre analyse.

```
# class(donnees_fr)
ts_donnees_fr <- ts(donnees_fr$new_cases, start = c(2020,1,3), frequency = 365)
# class(ts_donnees_fr)
```

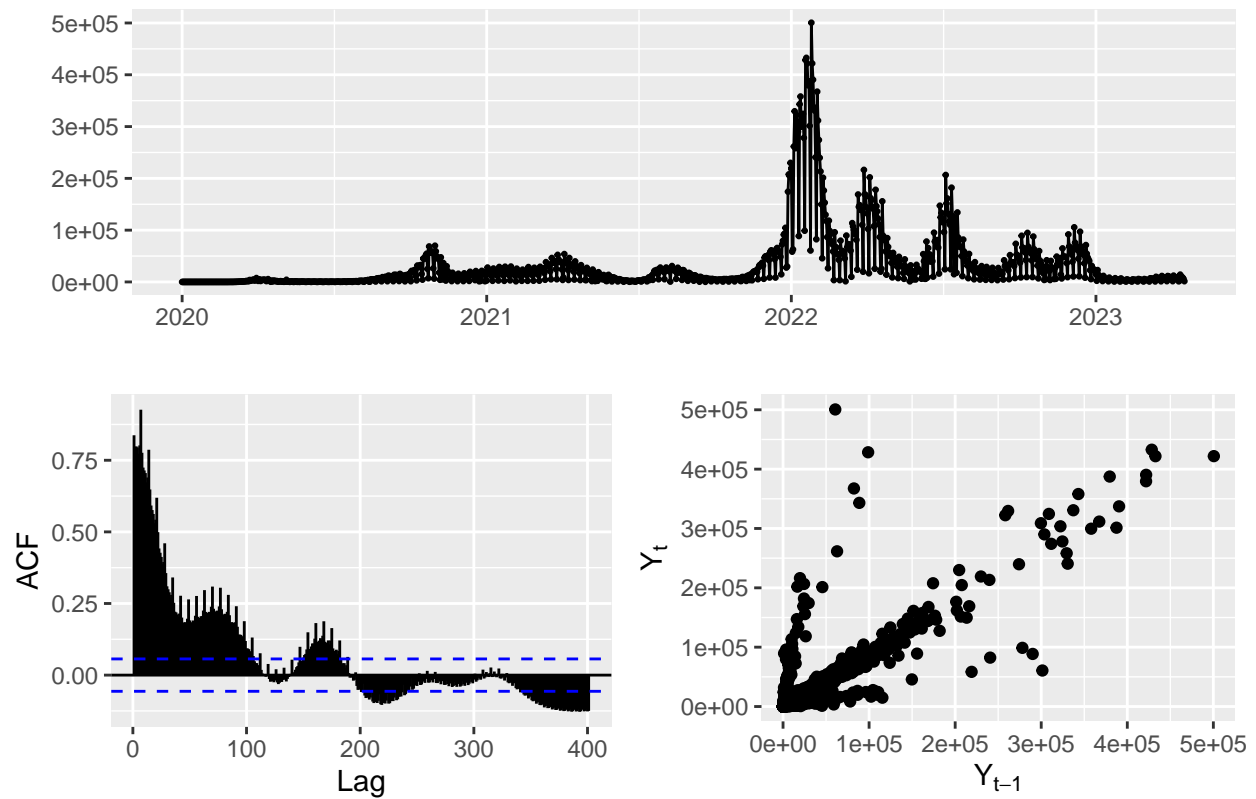
## Première partie

```
plot(ts_donnees_fr)
```



```
ts_donnees_fr |>
  ggtsdisplay(
    plot.type = "scatter")
```

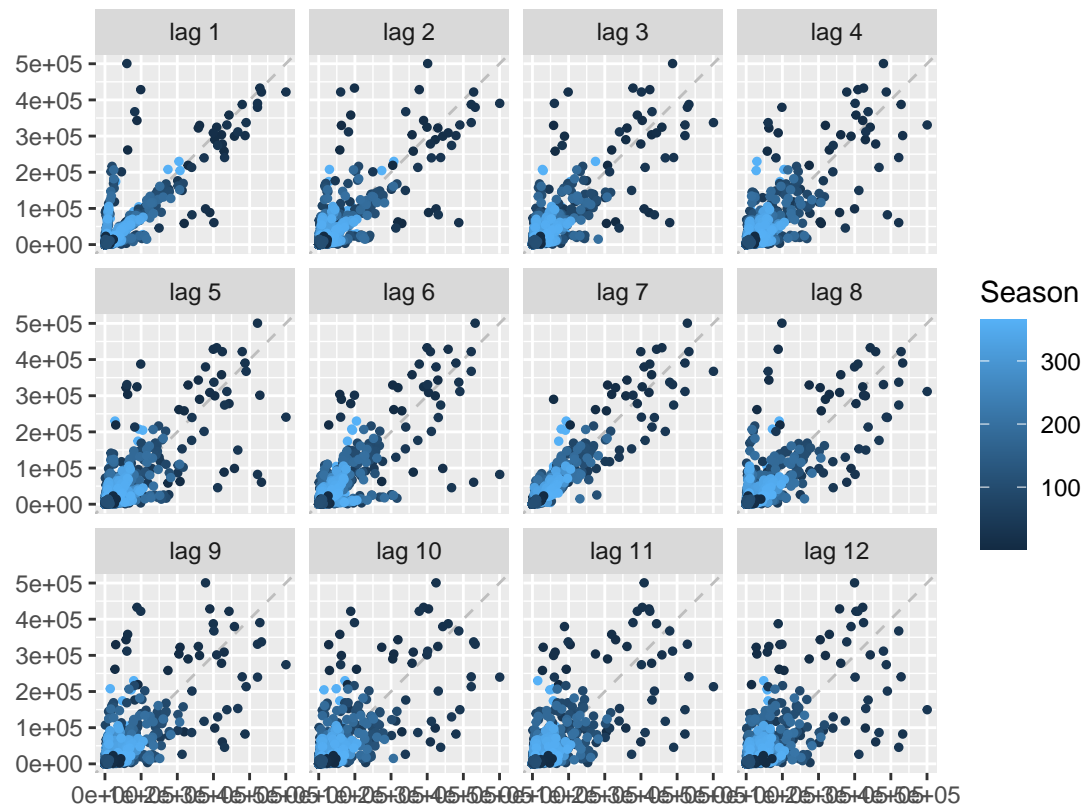
```
## Warning: Removed 1 rows containing missing values ('geom_point()').
## Removed 1 rows containing missing values ('geom_point()').
```



- ACF : pas de corrélation par mois (plus sur une période, sur plusieurs mois)
- Nuage de points à moitié aligné → corrélation

```
forecast::gglagplot(ts_donnees_fr, lags=12, do.lines = FALSE)
```

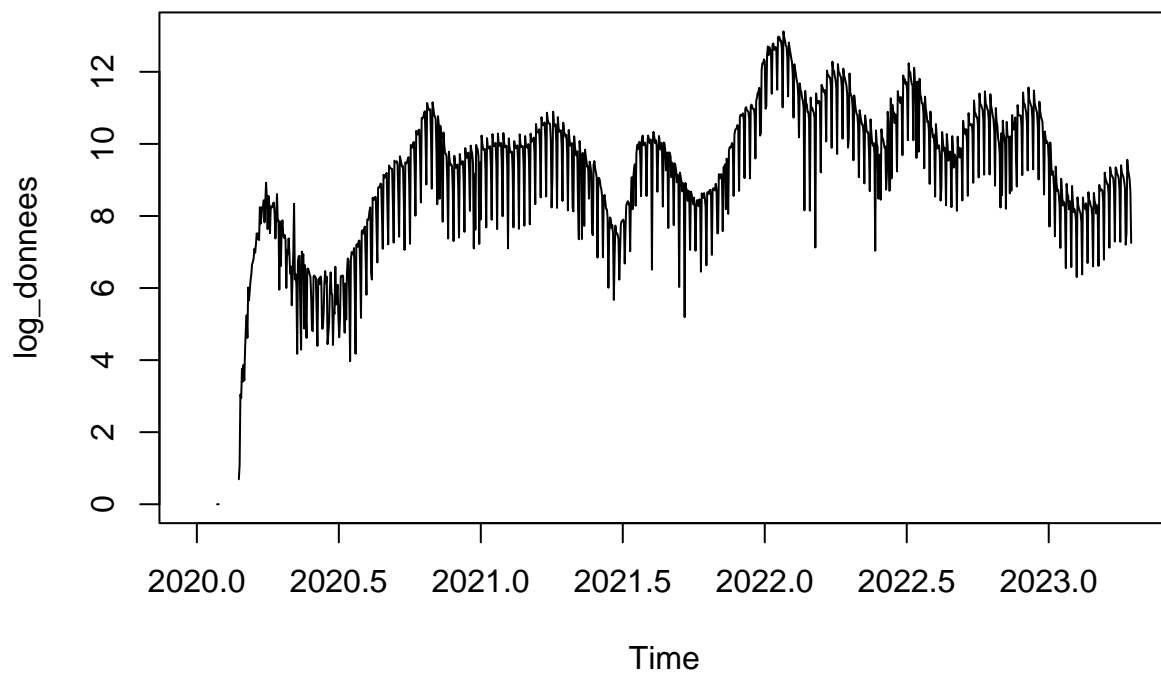
```
## Warning: Removed 12 rows containing missing values ('geom_point()').
```



On essaie de transformer notre variable :

```
log_donnees <- log(ts_donnees_fr)
```

```
plot(log_donnees)
```

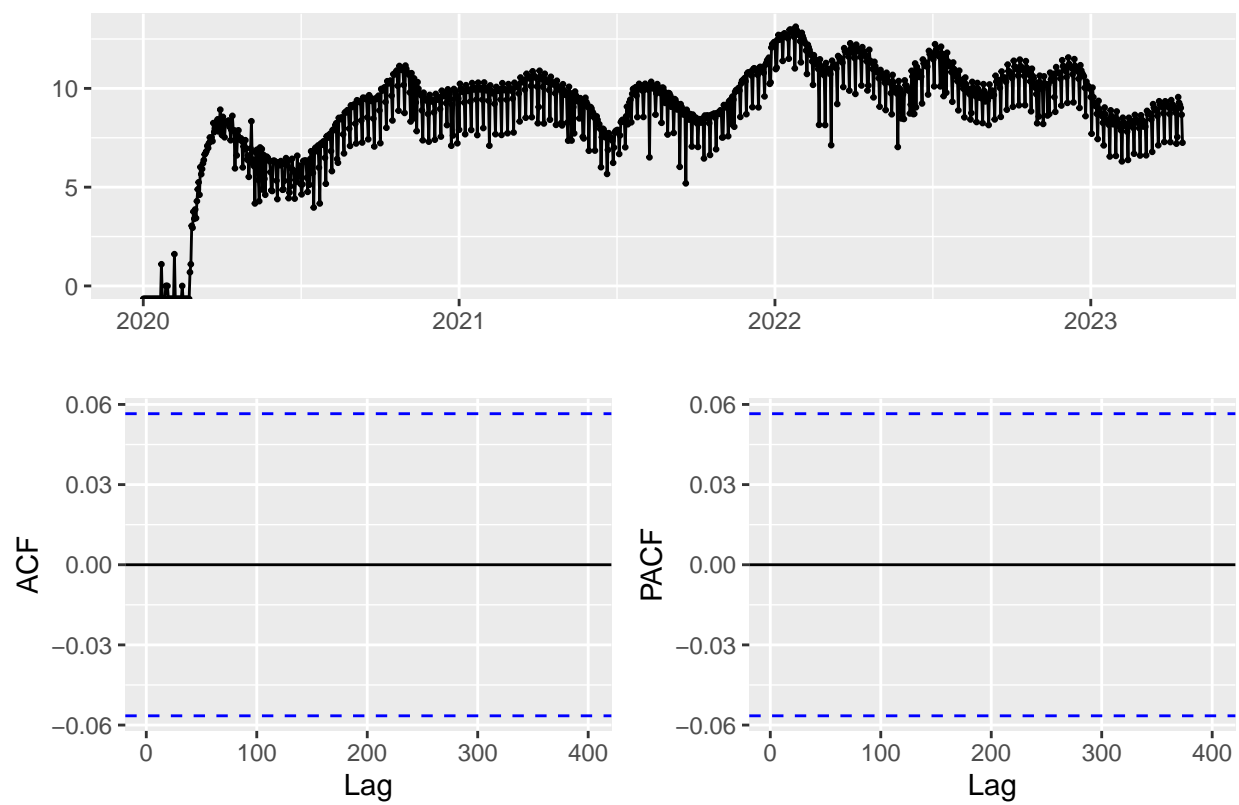


```
ggtsdisplay(log_donnees)
```

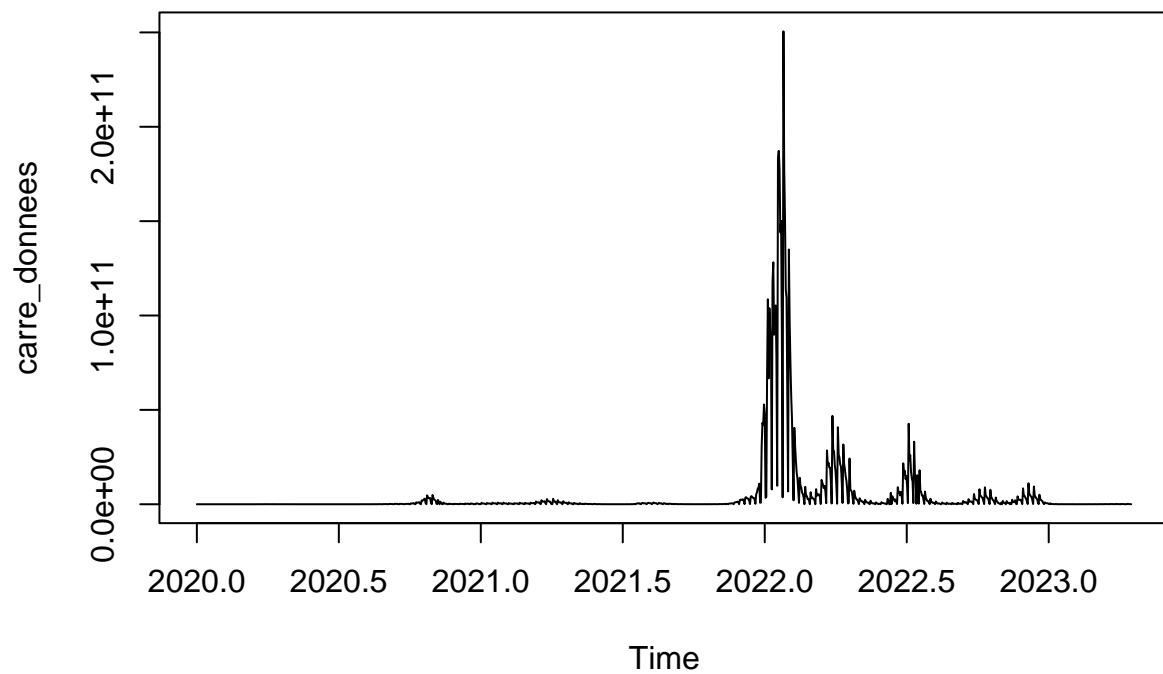
```
## Warning: Removed 1 rows containing missing values ('geom_point()').
```

```
## Warning: Removed 401 rows containing missing values ('geom_segment()').
```

```
## Removed 401 rows containing missing values ('geom_segment()').
```

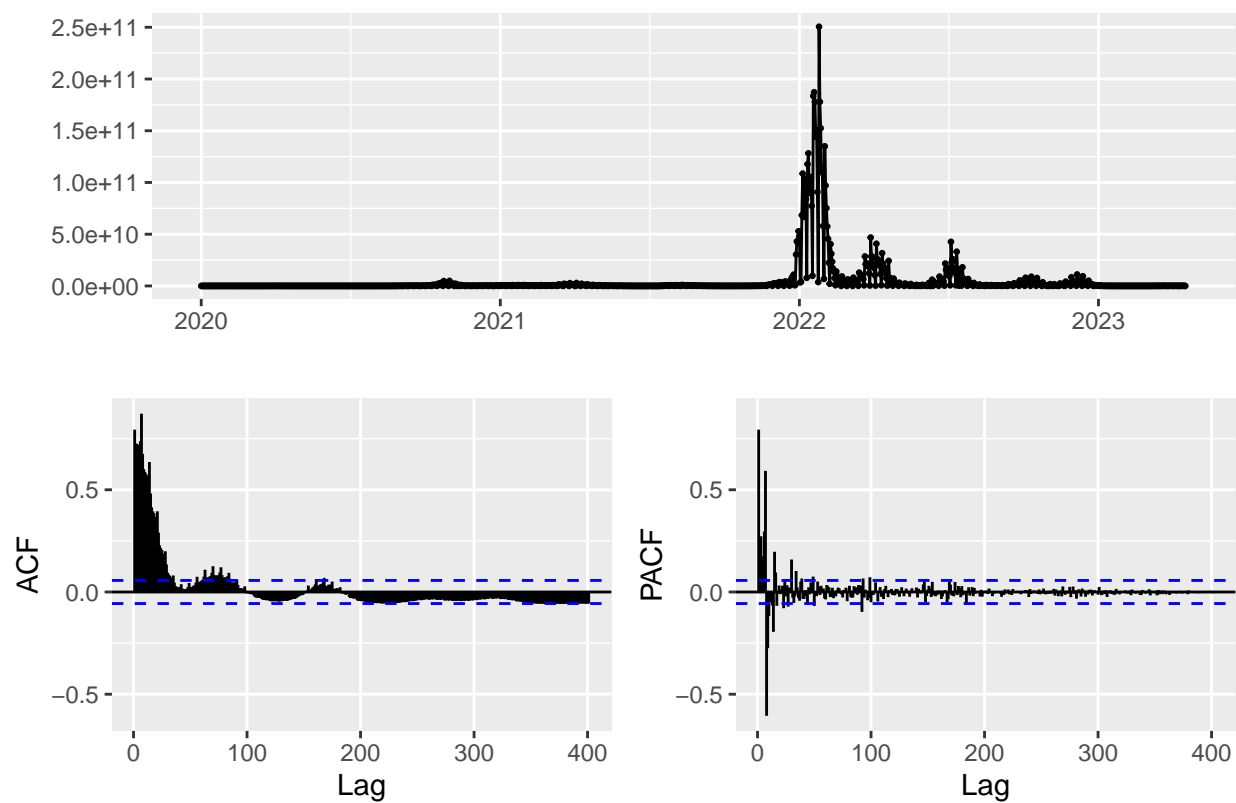


```
carre_donnees <- ts_donnees_fr^2  
plot(carre_donnees)
```



```
ggtsdisplay(carre_donnees)
```

```
## Warning: Removed 1 rows containing missing values ('geom_point()').
```



- Essayer de trouver un modèle qui lisse les données