

# Etude dans le temps des décès dûs aux accidents de la route au Royaume-Uni

Master 1 MAS Rennes - Série Temporelle

BERNARD Baptiste, MONFRET Dylan, RAKOTOSON Loïc

Pour le 15 mai 2020

## Table des matières

<b>1</b>	<b>Contextualisation &amp; Méthodologie</b>	<b>1</b>
<b>2</b>	<b>Visualiser et confirmer la saisonnalité</b>	<b>5</b>
2.1	<i>Lag Plot, Month Plot</i> et autocorrélation de la série . . . . .	5
2.2	. . . . .	6

# 1 Contextualisation & Méthodologie

Les données sont les enregistrements mensuels du nombre de morts, **death**, sur les accidents de la route au Royaume-Uni entre Janvier 1969 et Décembre 1984. La loi sur le port obligatoire de la ceinture de sécurité, **law**, a été introduite en Février 1983. La variable **law** prend alors la valeur 0 pour les mois où la loi n'est pas en vigueur, 1 lorsqu'elle est en vigueur.

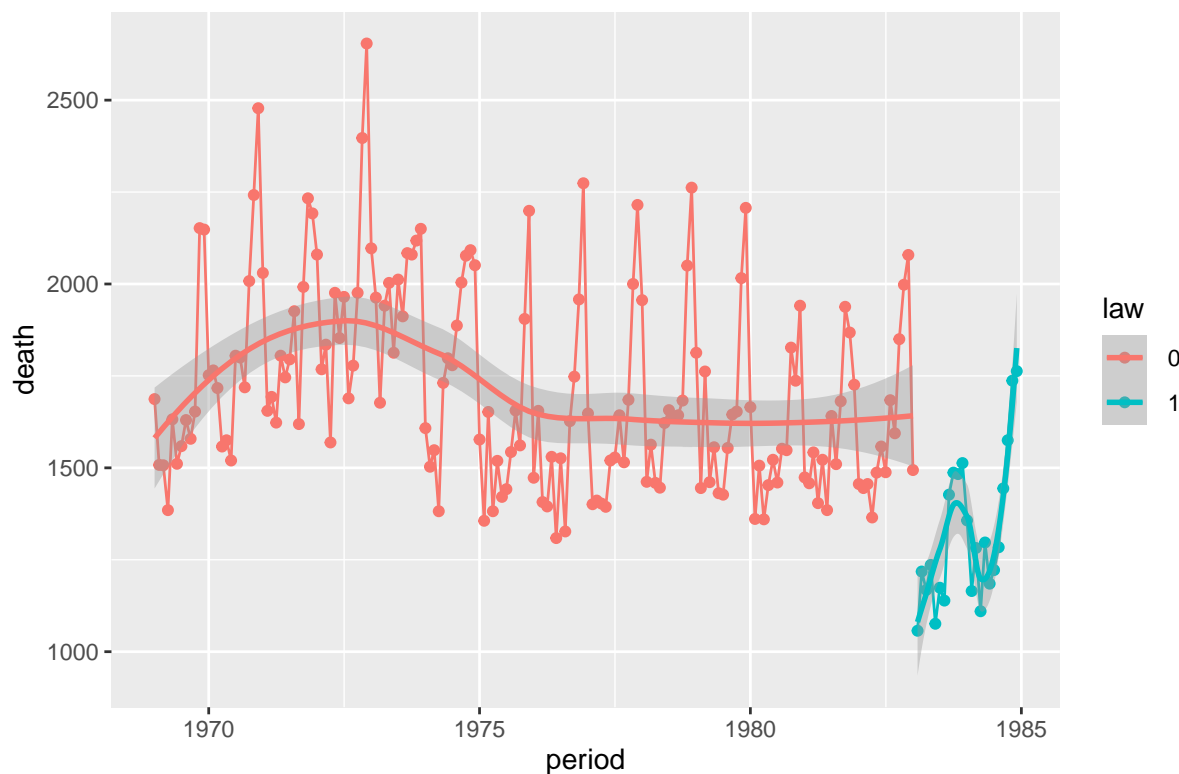
L'objectif de cette étude sera de proposer diverses modélisations de la série temporelle, qui permettraient par exemple de "simuler" le nombre de décès sur les routes britanniques après 1984.

```
period <-  
  seq(as.Date('1969-01-01'), as.Date('1984-12-31'), by = "month")  
  
ukdeath <-  
  read_delim("data.txt", delim = " ", col_types = "if") %>%  
  mutate(death_log = log(death),  
         period = period)
```

```
##      death      law      death_log      period  
## Min.    :1057    0:169   Min.     :6.963   Min.     :1969-01-01  
## 1st Qu.:1462    1: 23   1st Qu.:7.287   1st Qu.:1972-12-24  
## Median :1631                    Median :7.397   Median :1976-12-16  
## Mean   :1670                    Mean   :7.406   Mean   :1976-12-15  
## 3rd Qu.:1851                    3rd Qu.:7.523   3rd Qu.:1980-12-08  
## Max.   :2654                    Max.   :7.884   Max.   :1984-12-01
```

```
## # A tibble: 10 x 4  
##   death law death_log period  
##   <int> <fct>    <dbl> <date>  
## 1  1687 " 0"      7.43 1969-01-01  
## 2  1508 " 0"      7.32 1969-02-01  
## 3  1507 " 0"      7.32 1969-03-01  
## 4  1385 " 0"      7.23 1969-04-01  
## 5  1632 " 0"      7.40 1969-05-01  
## 6  1511 " 0"      7.32 1969-06-01  
## 7  1559 " 0"      7.35 1969-07-01  
## 8  1630 " 0"      7.40 1969-08-01  
## 9  1579 " 0"      7.36 1969-09-01  
## 10 1653 " 0"      7.41 1969-10-01
```

## Nombre de décès lors d'accidents de la route au Royaume-Uni

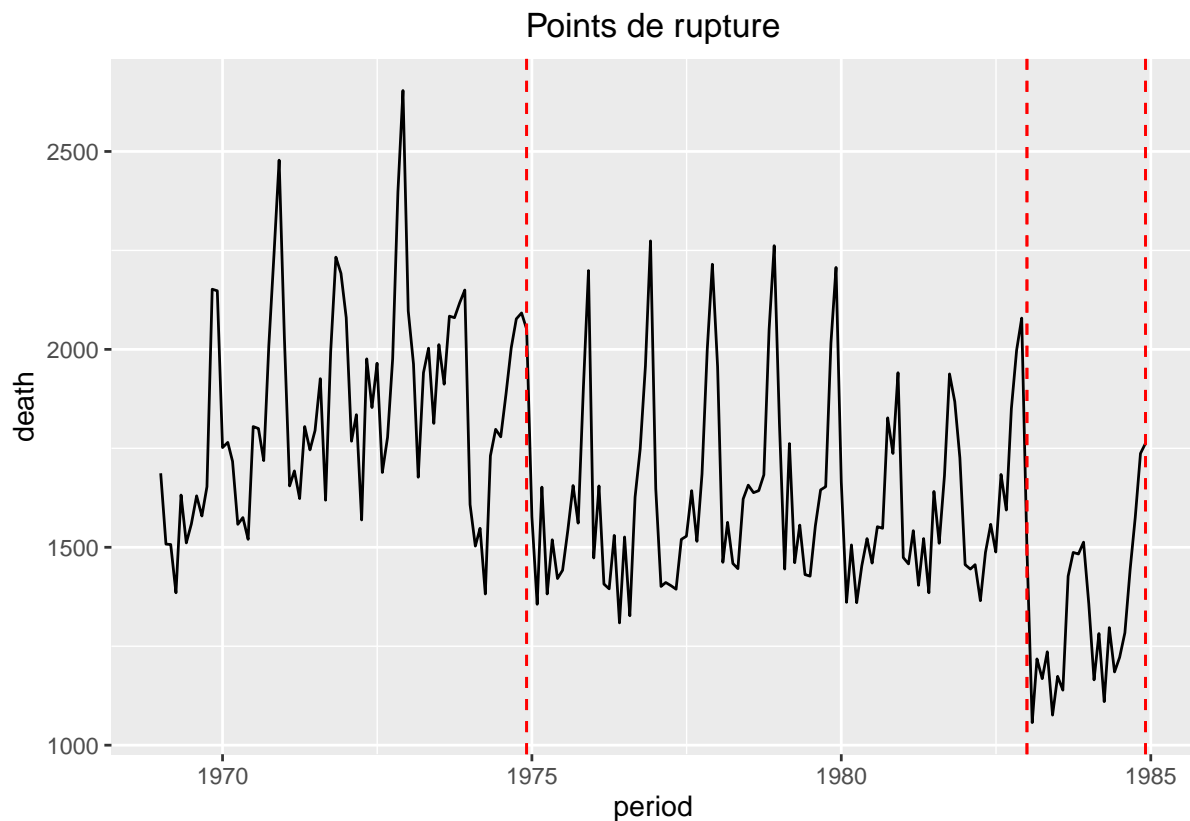


Une première observation de la série, avec mise en évidence des périodes d'application de la loi sur le port de la ceinture, permet de visualiser deux éléments constitutifs de la série :

- Le premier étant la périodicité du nombre de mort sur les routes, avec des accroissements significatifs à chaque fin d'année.
- Le second étant la tendance générale de la série, bien mis en relief par la [régression locale \(méthode non paramétrique, LOESS\)](#). Le nombre de décès chaque année à tendance à **s'accroître entre 1969 et 1973**, à **décroître entre 1973 et 1976**, à **stagner de 1976 à 1983**, avant de **chuter brutalement avec la mise en application de la loi sur le port de la ceinture en Février 1983**.

Intuitivement, nous pourrions construire nos modèles autour d'une saisonnalité des décès sur les routes par rapport fin d'année (période des fêtes), soit **un pique de décès tous les 12 mois** ; et en prenant en compte 2 à 3 phases de la série temporelle (2 phases si l'on se ramène uniquement à la période sans la loi et la période avec application de celle-ci).

C'est point de rupture sont aussi visualisable avec le package **changepoint** et sa fonction `cpt.meanvar`, et cela sans prendre en compte la variable `law`.



Pour la suite de l'étude, nous travaillerons avec le nombre de décès passé au log. Une variabilité réduite peut nous garantir un meilleur ajustement en conservant la forme de la série.

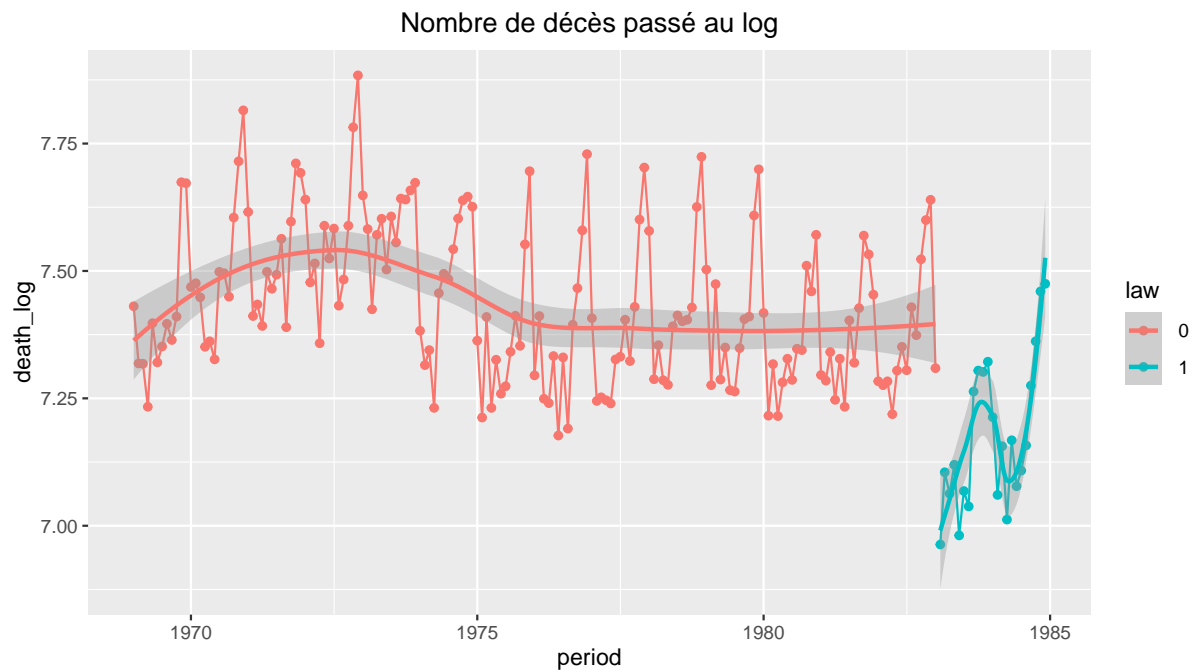
```
var(ukdeath$death)
```

```
## [1] 83874.51
```

```
var(ukdeath$death_log)
```

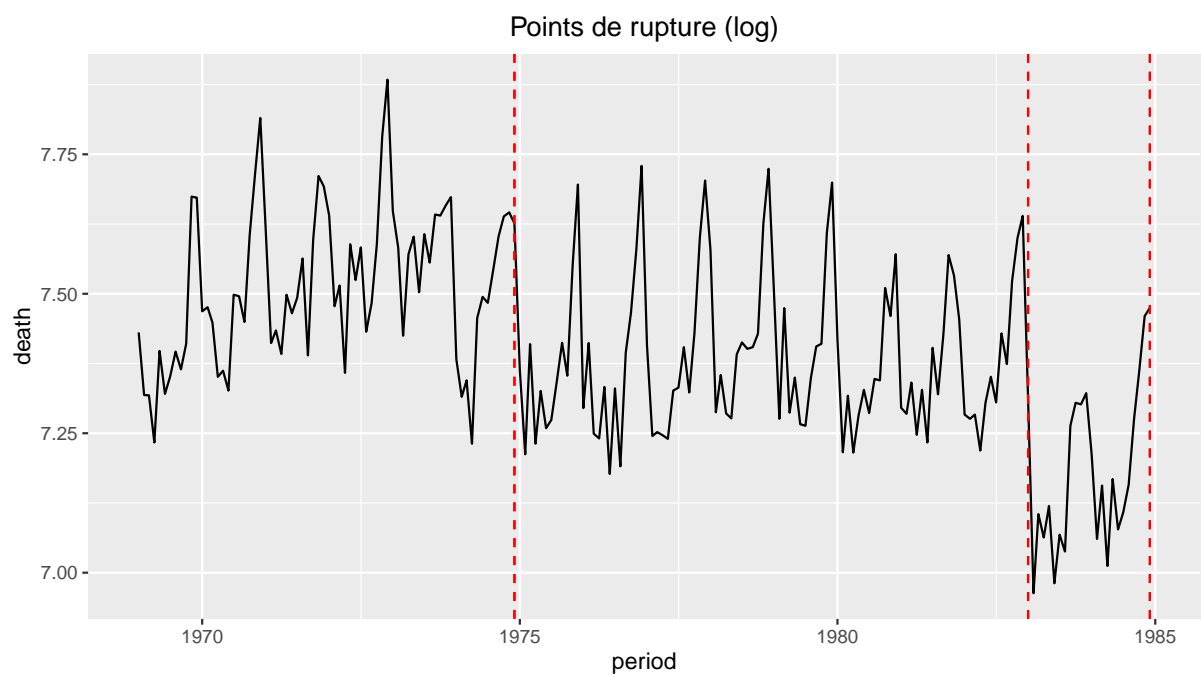
```
## [1] 0.02935256
```

```
ggplot(ukdeath) +  
  aes(x = period, y = death_log, color = law) +  
  geom_point() + geom_line() + stat_smooth(method = "loess") +  
  labs(title = "Nombre de décès passé au log") +  
  theme(plot.title = element_text(hjust = 0.5))
```



```
ts_ukdeath_log <-
  ts(
    data = ukdeath$death_log,
    start = c(1969, 1),
    frequency = 12
  )

ts_ukdeath_log %>%
  changepoint::cpt.meanvar(method = "PELT", minseglen = 11) %>%
  autoplot() +
  labs(title = "Points de rupture (log)", x = "period", y = "death") +
  theme(plot.title = element_text(hjust = 0.5))
```

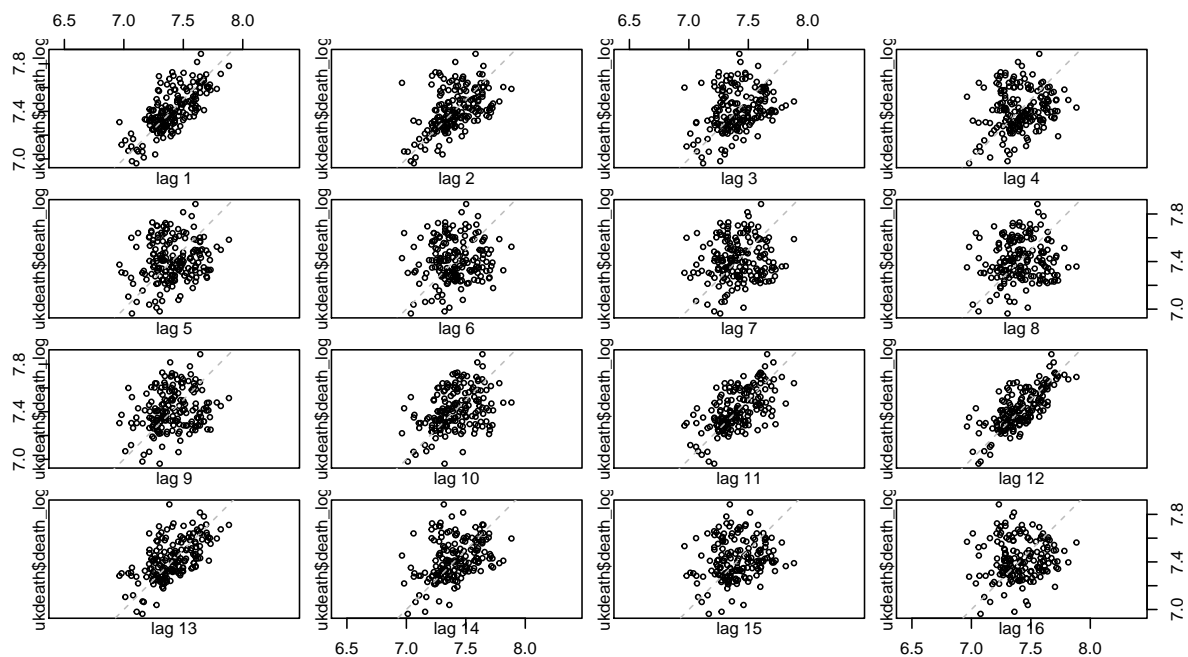


## 2 Visualiser et confirmer la saisonnalité

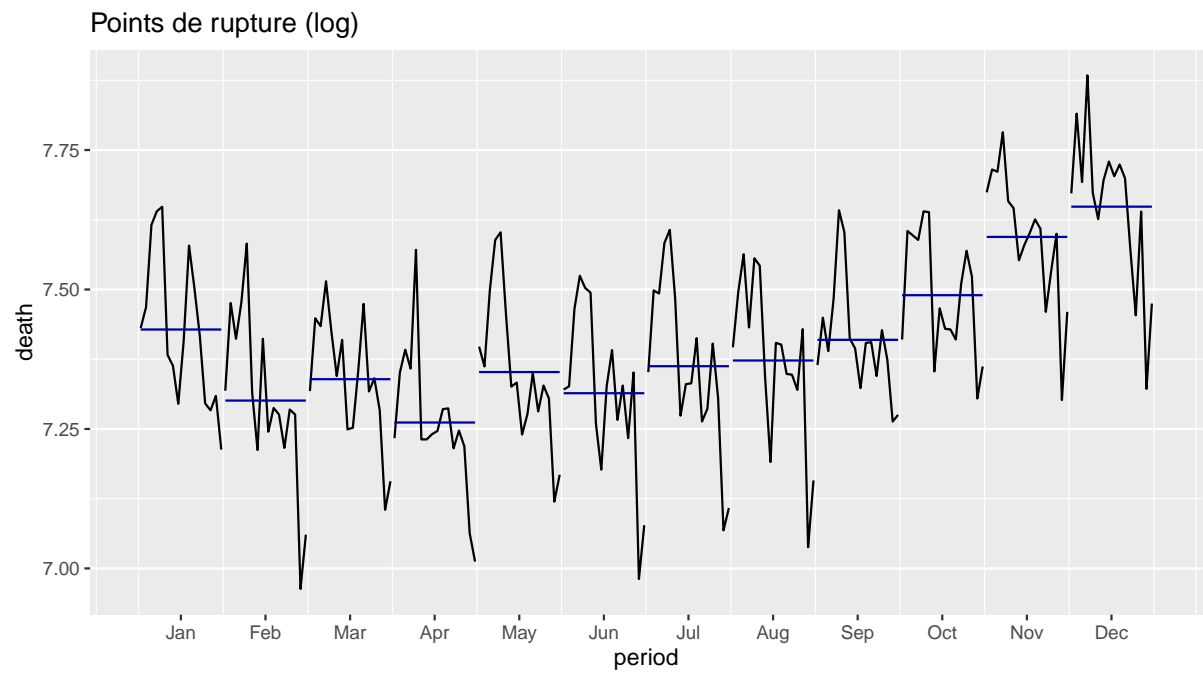
Pour construire les modèles adéquats à notre problème, nous devons avoir une idée précise des caractéristiques de la série. Il s'agit ici de confirmer si la saisonnalité des décès est bien de 12 mois, et de l'autocorrélation des données (lien ou similitude entre l'été 1963 et l'été 1973, par exemple).

### 2.1 *Lag Plot, Month Plot* et autocorrélation de la série

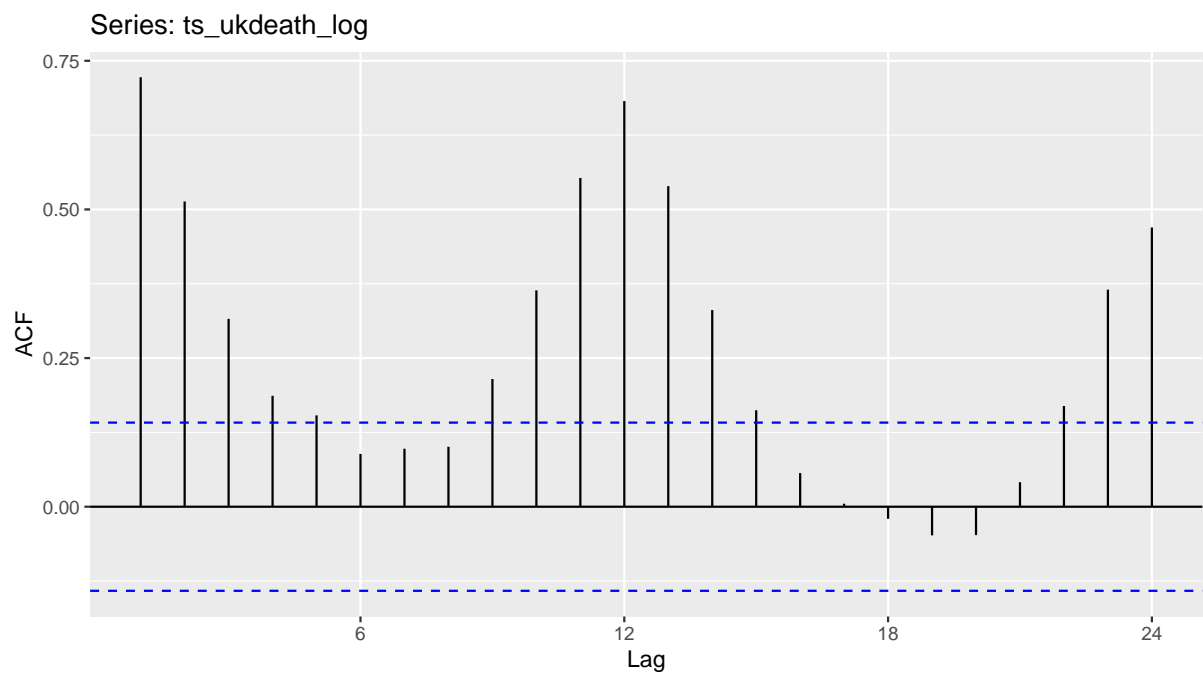
```
lag.plot(ukdeath$death_log, lags = 16)
```



```
ggmonthplot(ts_ukdeath_log) +  
  labs(title = "Points de rupture (log)", x = "period", y = "death")
```



```
ggAcf(ts_ukdeath_log)
```



## 2.2