

Analyse exploratoire des données

Loïc

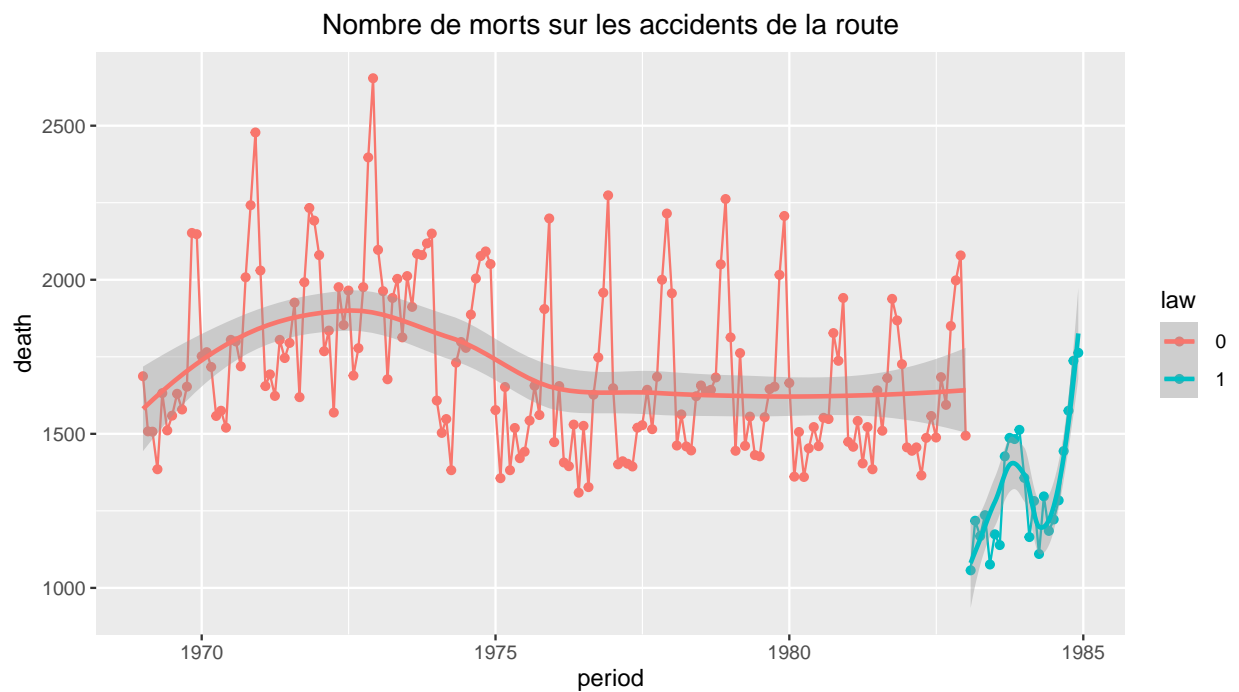
I. Contexte

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 1983.

La loi sur le port obligatoire de la ceinture de sécurité, **law**, a été introduite en Février 1983.

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

```
##      death      law
## Min.   :1057    0:169
## 1st Qu.:1462    1: 23
## Median :1631
## Mean   :1670
## 3rd Qu.:1851
## Max.   :2654
```



II. Modèle additif global

On ne prend pas en compte la variable `law`

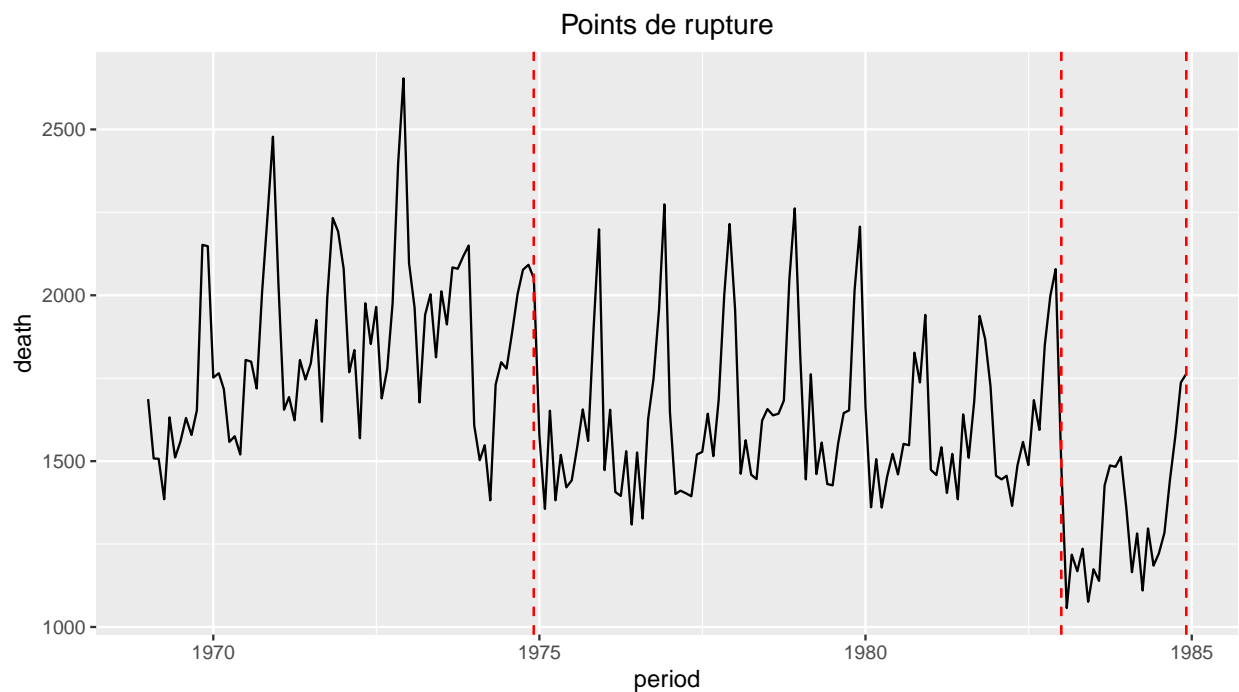
```
ts_ukdeath <- ts(data = ukdeath$death, start = c(1969, 1), frequency = 12)
```

1. Analyse des segments

En découpant la série en segments “homogènes” en moyenne et variance, on peut détecter 2 points de ruptures sans considérer celui de la limite, dont un à la promulgation de la loi.

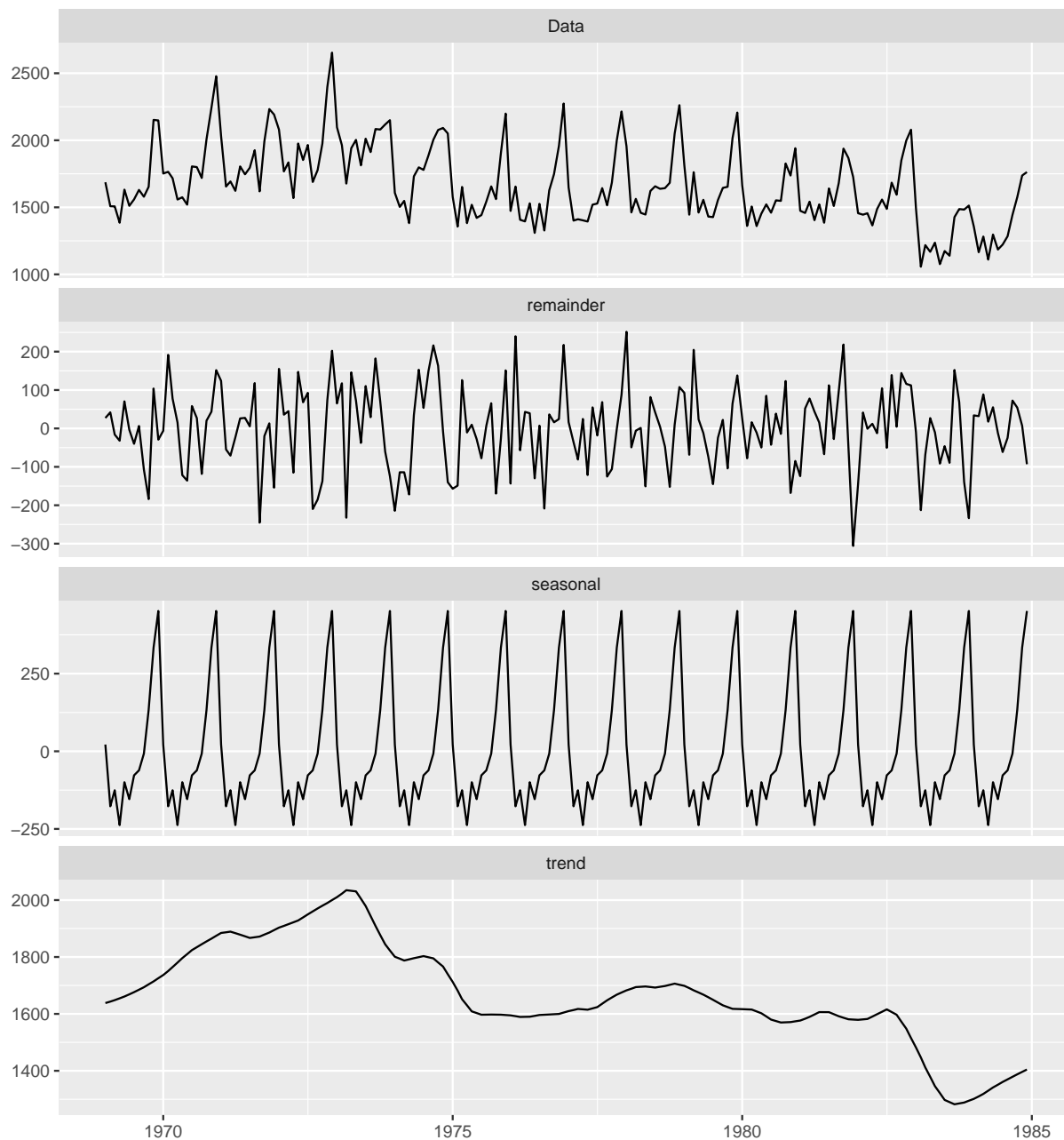
Ainsi on a trois segments:

- **1970 - 1975:** on sait pas, ça monte puis ça descend (mdr)
- **1975 - 1983:** Période stable
- **Après 1983:** Depuis la loi



2. Analyse de chaque décomposition

```
stl_Add <- stl(ts_ukdeath, "periodic")
```



En visualisant la tendance globale (*trend*), on s'aperçoit qu'on obtient bien les 3 mêmes segments qu'avec le découpage par point de rupture ci-dessus.

On remarque aussi qu'il existe une saisonnalité (*seasonal*) avec une forte hausse des accidents pendant les périodes de fêtes de fin d'année (et du nouvel An aussi du coup).

Le bruit (*remainder*) n'influe que peu sur les données (d'où la grande échelle à gauche).

3. Analyse du bruit

Procedons à l'analyse du bruit avec les fonctions d'autocorrélation et les tests de normalités.

```
noise <- stl_Add$time.series[,3]
```

a - Fonctions d'autocorrélation

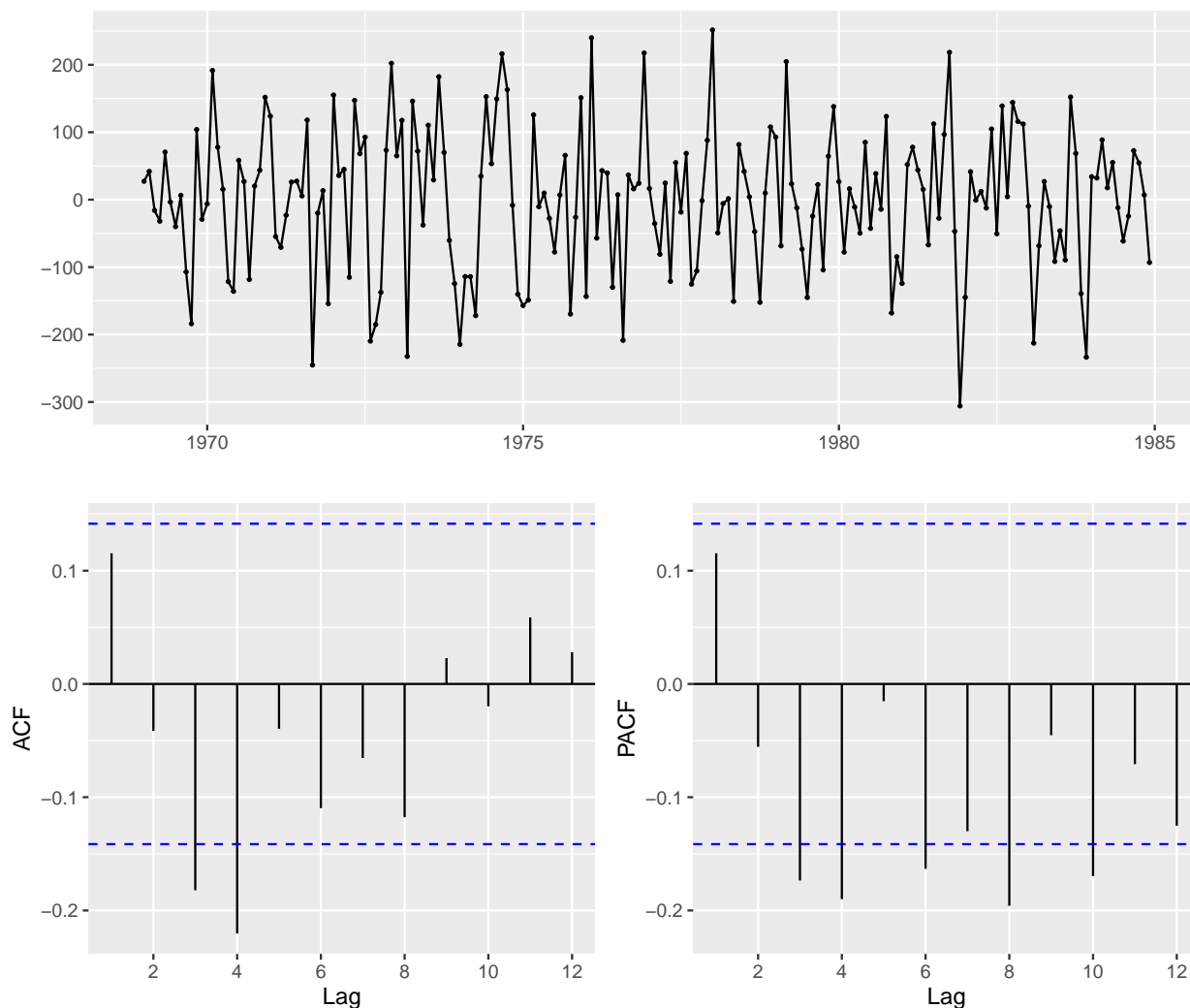
Visuellement, on ne détecte pas de motif sur les bruits. Ils seraient donc distribués aléatoirement.

L'**ACF** montre qu'il y a globalement peu d'autocorrélation. A des intervalles de 3 et 4 mois, on a des autocorrélations négatives plus fortes.

Il s'agit de la forte différence de valeurs entre les débuts d'année en *février-mars-avril* et les derniers mois en *novembre-décembre-janvier*.

le **PACF** montre que les autocorrélations partielles sont globalement négatives, compte tenu des valeurs des intervalles intermédiaires.

Il s'agit de la forte augmentation des valeurs à la fin par rapport à l'ensemble des données de l'année.



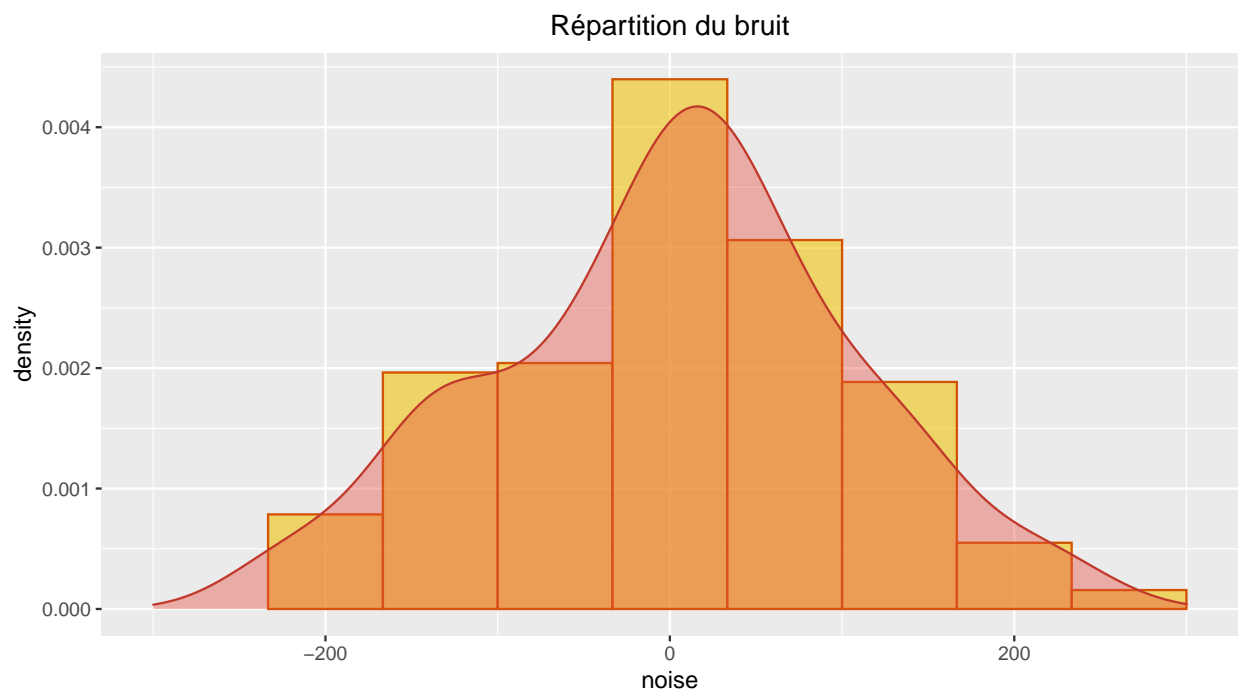
b - Normalité du bruit

Malgré la répartition du bruit presque normale, comme représentée ci-dessous, la p-value du test de Shapiro-Wilk est significative, $0.4 < 0.5$.

On rejette donc l'hypothèse de la normalité du bruit.

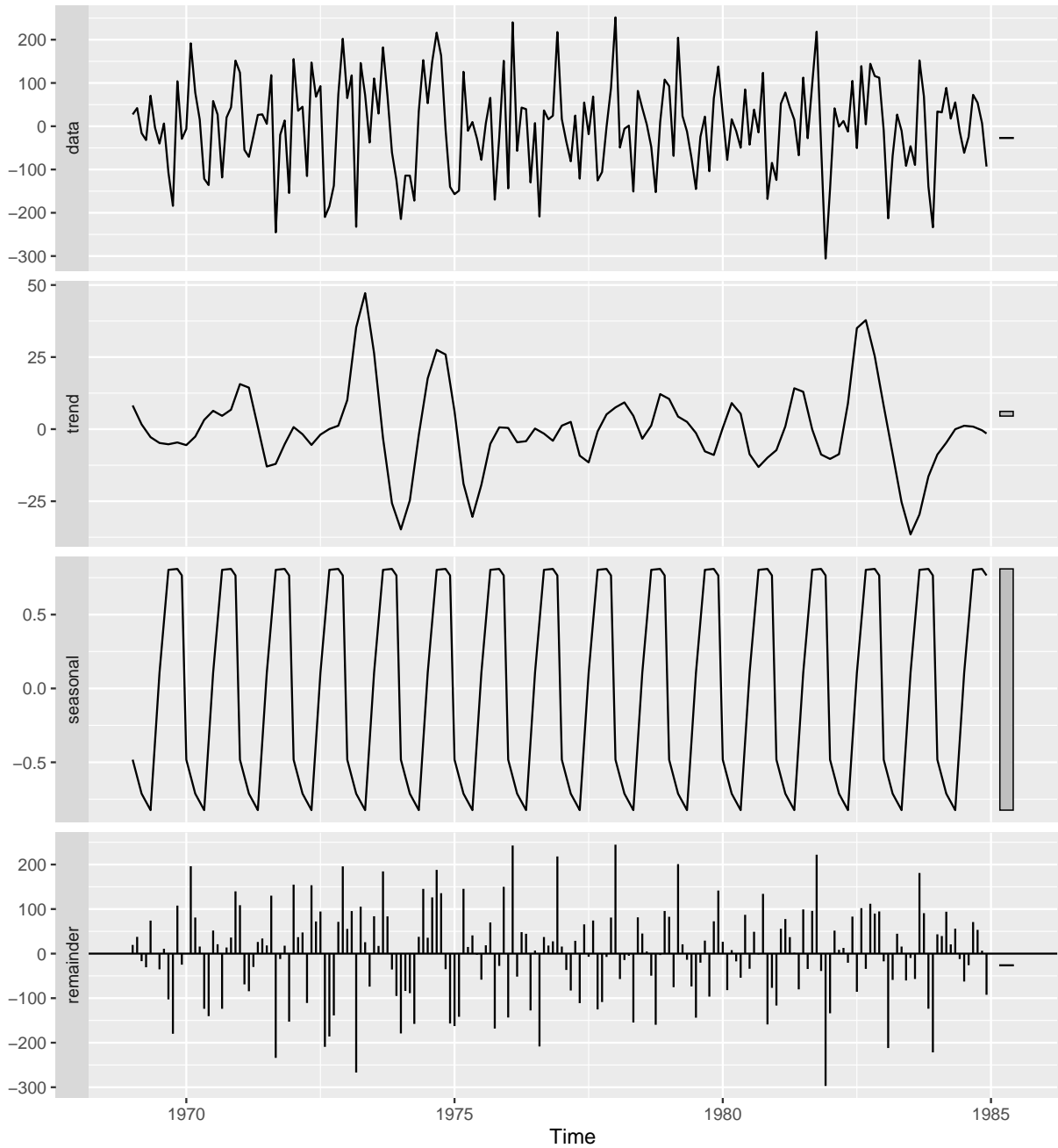
```
shapiro.test(noise)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  noise  
## W = 0.99224, p-value = 0.4
```



III. Modèle additif avec bruit

```
ts_bruit <- ts(data = noise, start = c(1969, 1), frequency = 12)
```



On remarque que la saisonnalité du bruit, même si elle existe, n'a quasiment pas d'effet sur celui-ci. Il possède en revanche une tendance globale. Toutefois, les données du bruit sont essentiellement expliquées par le bruit du bruit (*remainder*).