

04-02

Visualisation des séries temporelles

Hiver 2023

Spécialisation technique en intelligence artificielle
Algorithmes d'apprentissage non supervisé — 420-A58-SF — M. Swawola, M.Sc.

**NOUS ÉCLAIRONS.
VOUS BRILLEZ.**

FORMATION CONTINUE
ET SERVICES AUX ENTREPRISES



Sommaire

1. Objets “ts”
2. Graphiques
3. Décomposition d’une séries temporelle
4. Graphiques de saison
5. Autocorrélations
6. Exemple - Série Google
7. Références



Objets “ts”

Objets “ts”

- Permet de fabriquer **des séries temporelles**

```
mydata <- c(1,2,3,2,1) # Les données
```

```
mydata <- as.ts(mydata) # Objet “série temporelle”
```

```
mydata <- ts(mydata) # Équivalent
```

Datation

■ Dater une série temporelle

```
# Dater une série temporelle annuelle qui débute en 2020
```

```
mydata <- ts(mydata, start=2020)
```

```
# Dater une série temporelle trimestrielle débutant au 3ème trimestre 2020
```

```
mydata <- ts(mydata, start=c(2020,3), frequency=4)
```

```
# Dater une série temporelle mensuelle débutant en août 2020
```

```
mydata <- ts(mydata, start=c(2020,8), frequency=12)
```

Question

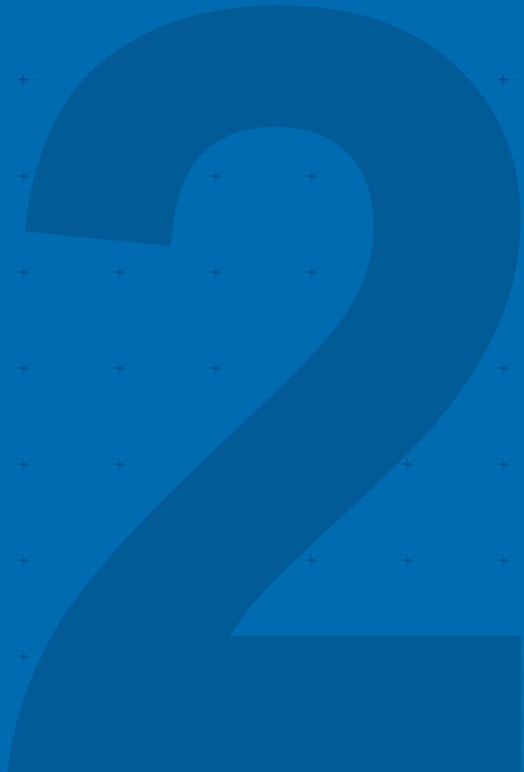
- Soit la ligne de code suivante:

```
x <- ts(c(299,7,92,4,58), start=2015)
```

S'agit-il d'une série annuelle, trimestrielle ou mensuelle ?

Objets et fonctions ts

Type de données	Fréquence	Exemple de “start”
Annuel	1	1998
Trimestriel	4	c(1998,3)
Journalier	7 ou 365.25	1 ou c(1998,280)
Hebdomadaire	52.18	c(1998,22)
Horaire	24 ou 168 ou 8.766	1
Demi-horaire	48 ou 336 ou 17.532	1

A large, dark blue number 2 is positioned on the left side of the image. The background is a solid blue color with a grid of small white plus signs (+) spaced evenly across it.

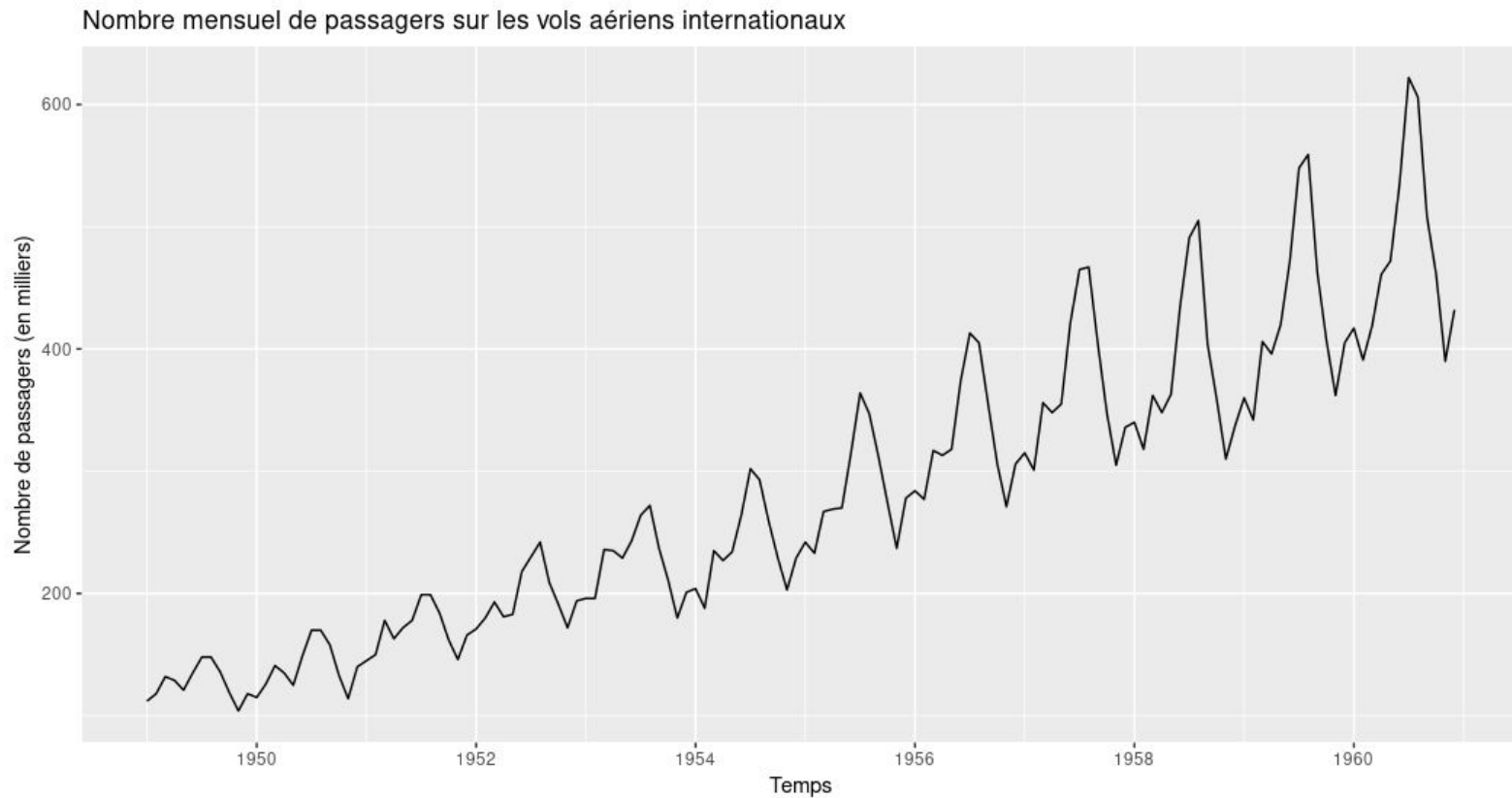
Graphiques

Exemple de graphique

■ Afficher une série temporelle avec **ggplot2**

```
library(ggplot2)
library(fpp2)
airline <- read.table("../..data/airline49.dat")
airline <- ts(airline, start=c(1949,1), frequency=12)
autoplot(airline) +
  ggtitle("Nombre mensuel de passagers sur les vols aériens internationaux") +
  xlab("Temps") +
  ylab("Nombre de passagers (en milliers)")
```

Exemple de graphique



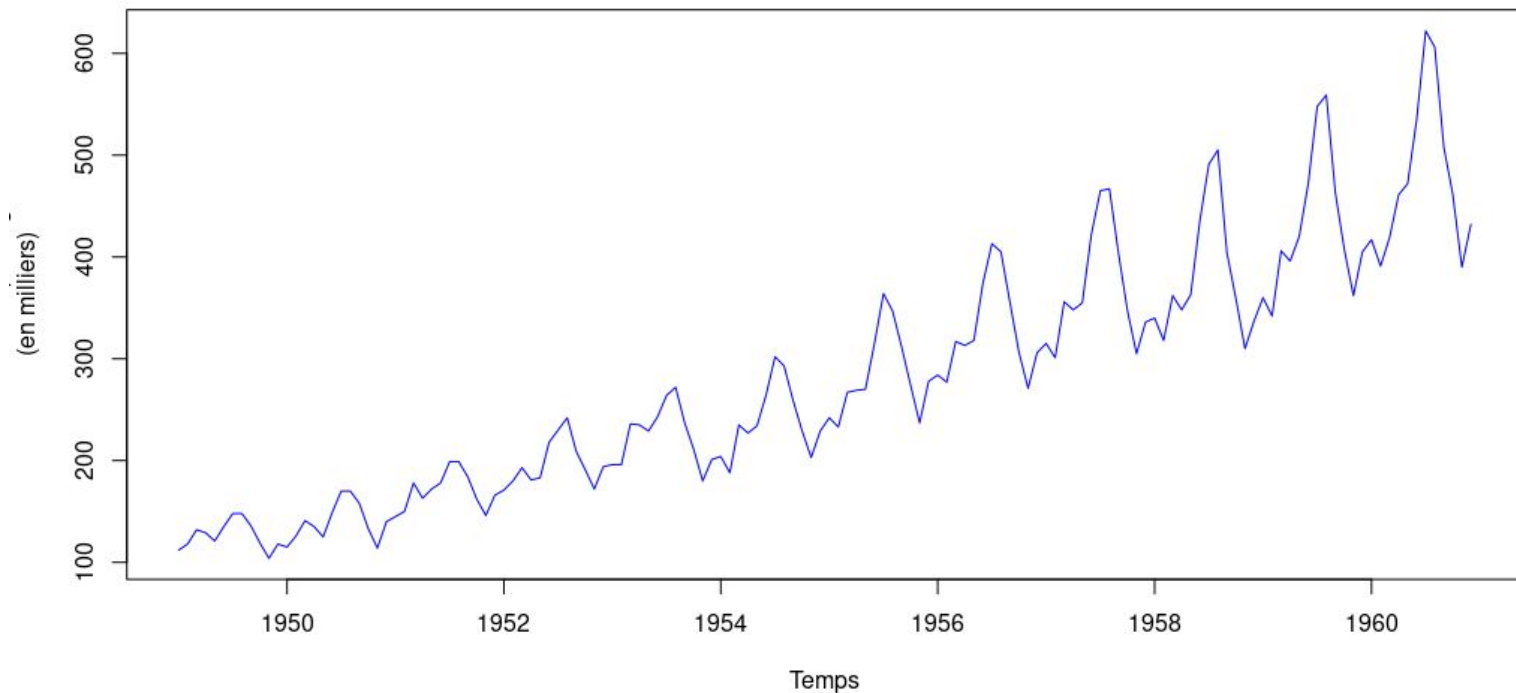
Exemple de graphique

- Autre manière de faire sans ggplot2

```
plot.ts(airline, main="Nombre mensuel de passagers sur les vols aériens internationaux",  
xlab="Temps", ylab="Nombre de passagers (en milliers)", col="blue")
```

Exemple de graphique

**Nombre mensuel de passagers sur les vols
aériens internationaux**





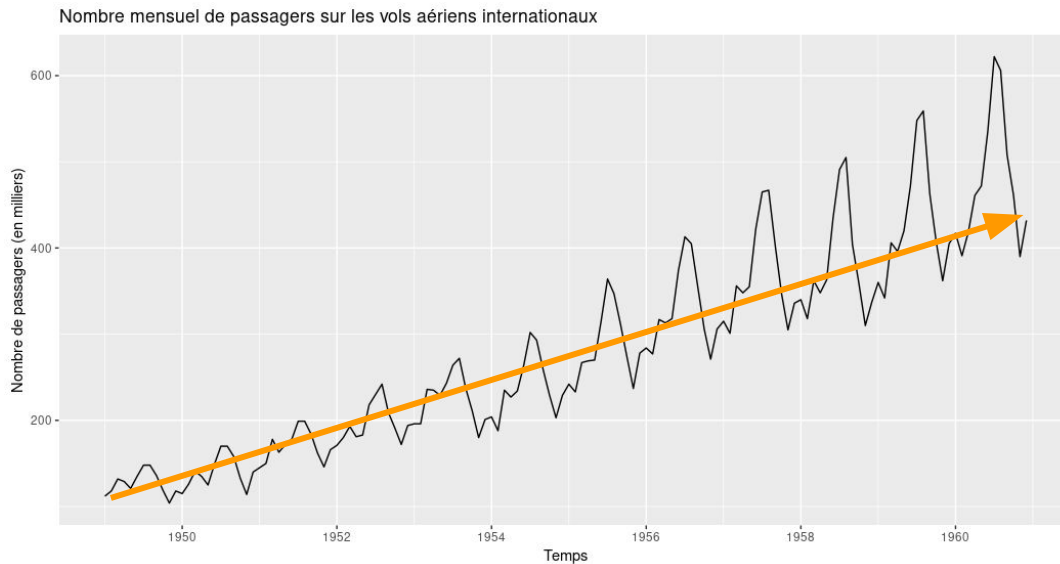
Décomposition d'une série temporelle

Décomposition d'une série temporelle

- La **décomposition** classique d'une série temporelle se décline en quatre composantes
 - La **tendance** (trend)
 - La **saisonnalité** (season)
 - Le **cycle**
 - La **partie résiduelle**
- La décomposition peut être **additive** ou **multiplicative**

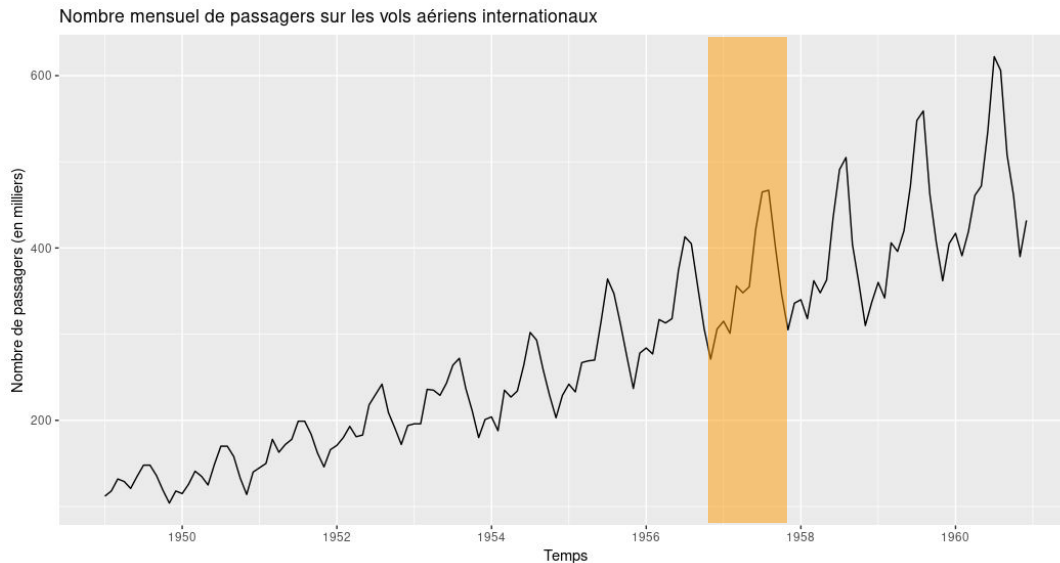
La tendance

- La **tendance** est le comportement à long terme de la série



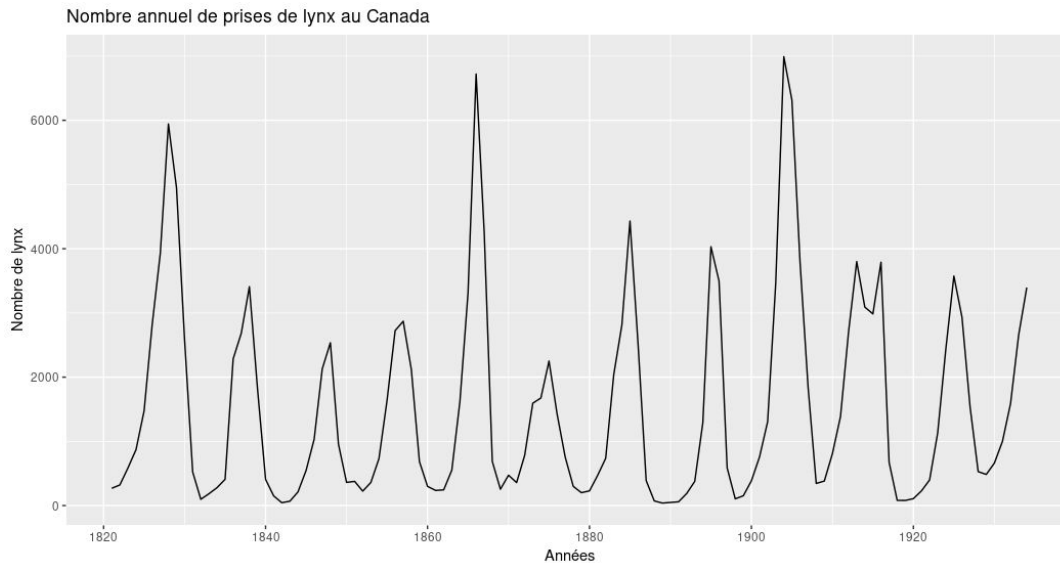
La saisonnalité

- La **saisonnalité** ou partie saisonnière apparaît lorsqu'un facteur saisonnier survient par exemple chaque année de manière similaire



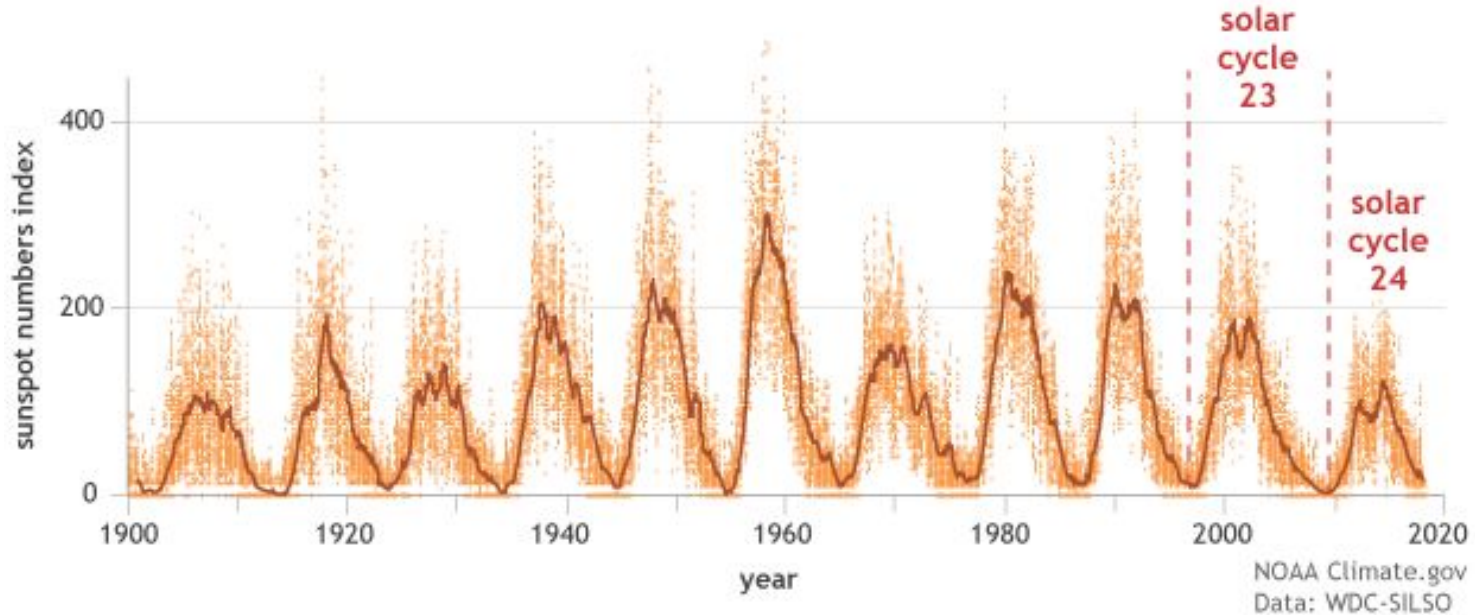
Le cycle

- Le **cycle** apparaît lorsque la série montre des périodes de croissance et décroissance qui reviennent de façon non régulière.



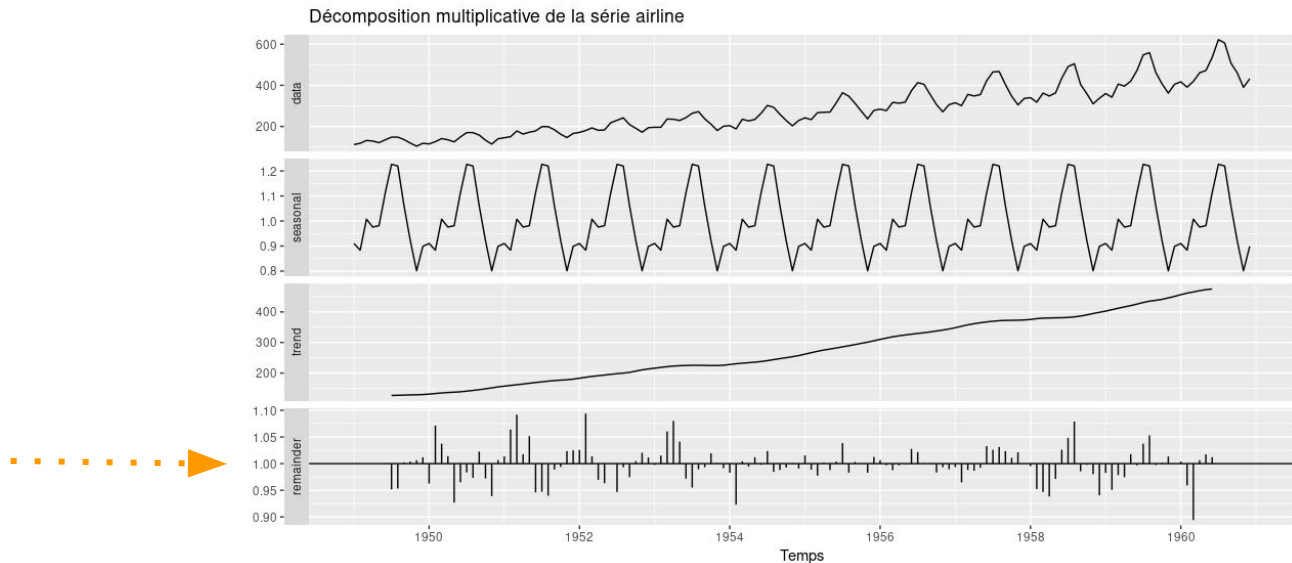
Ne pas confondre saisonnalité et cycle !

Solar activity over the past 11 sunspot cycles



La partie résiduelle

- La **partie résiduelle**, ou **résidu**, est ce qu'il reste après avoir ôté de la série initiale les précédentes composantes (tendance, saisonnalité et cycle)



Décomposition sous R

- Pour décomposer une série, utiliser la fonction `decompose` avec comme argument `type` les valeurs `additive` ou `multiplicative`

```
# Décomposition de la série "airline"
```

```
airline %>% decompose(type="multiplicative") %>%
```

```
  autoplot() + xlab("Temps") +
```

```
  ggtitle("Décomposition multiplicative de la série airline")
```

4

Graphiques de saison

Graphiques de saison

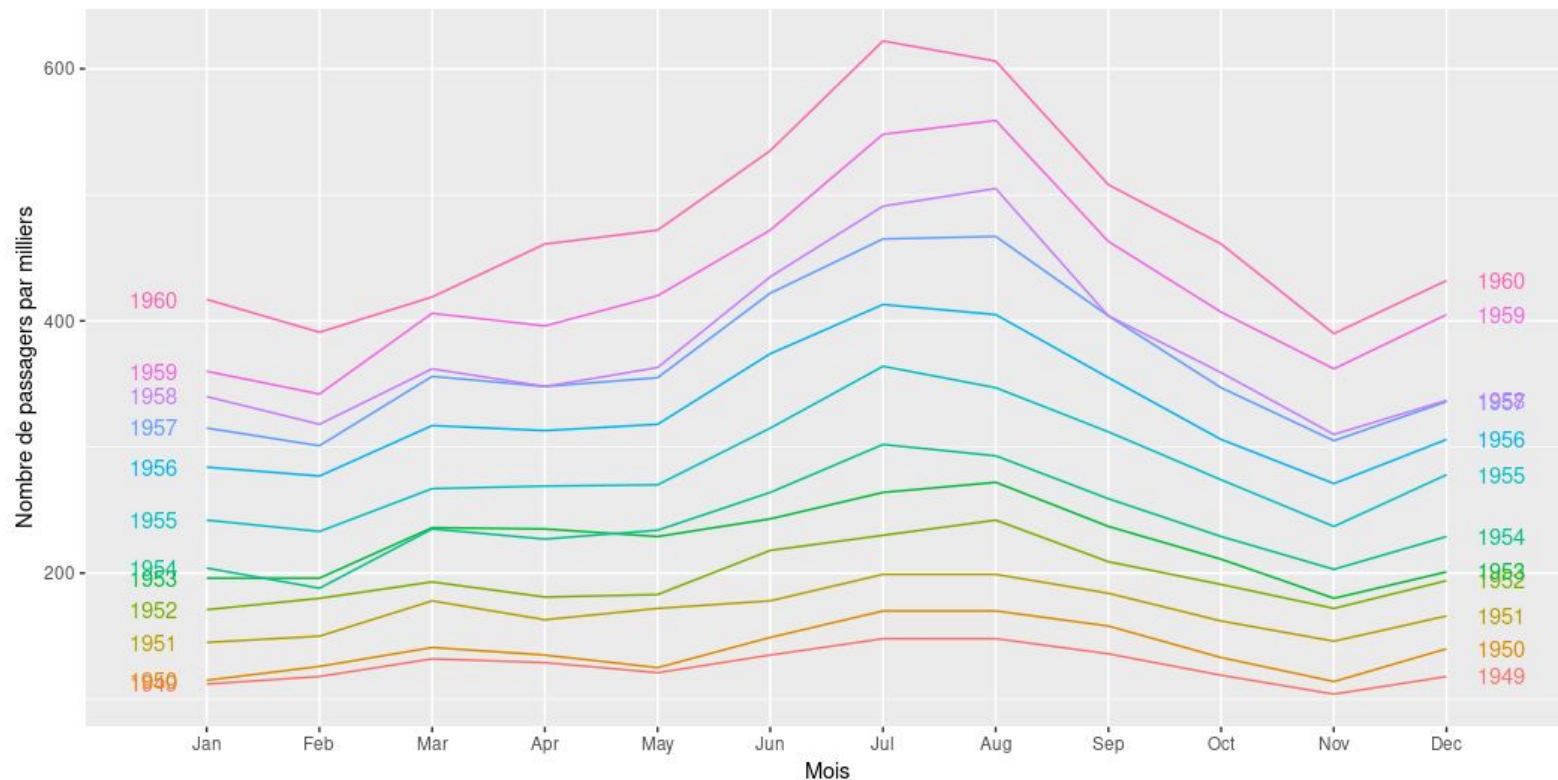
- La fonction `ggseasonplot` permet de montrer la série temporelle découpée sur chaque saison pour mieux en appréhender les caractéristiques

```
# Graphe de saison
```

```
ggseasonplot(airline, year.labels=TRUE, year.labels.left=TRUE) +  
  ylab("Nombre de passagers par milliers") +  
  xlab("Mois") +  
  ggtitle("Graphe de saison : série airline")
```

Graphiques de saison

Graphe de saison : série airline



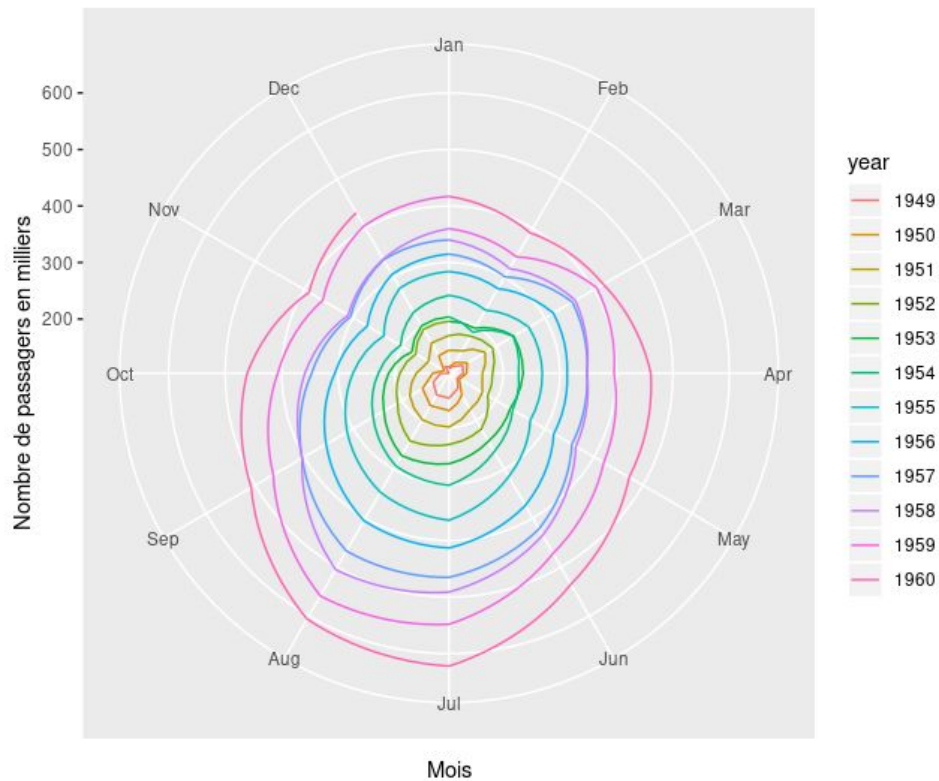
Graphiques de saison

- Il est possible de changer l'axe horizontal pour un axe de "rotation"

```
# Graphe polaire de saison  
ggseasonplot(airline, polar=TRUE) +  
  ylab("Nombre de passagers par milliers") +  
  xlab("Mois") +  
  ggtitle("Graphe polaire de saison : série airline")
```


Graphiques de saison

Graphe polaire de saison: série airline



Graphiques de saison

- On peut aussi tracer la série sur chaque saison séparément en regroupant toutes les données d'une même saison sur le même sous-graphique

```
# Sous-série saisonnière
```

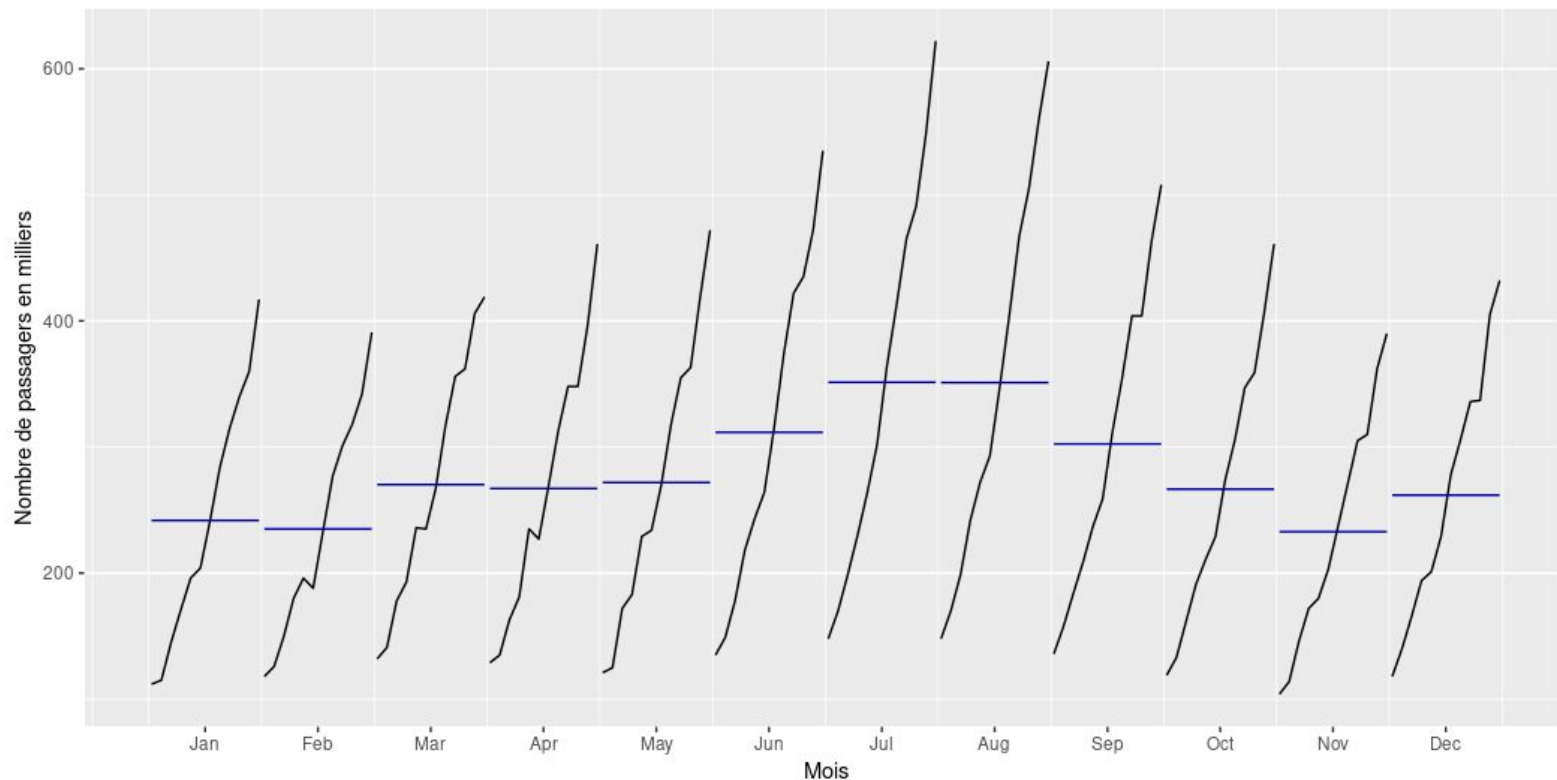
```
ggsubseriesplot(airline) +
```

```
  ylab("Nombre de passagers en milliers") + xlab("Mois")+
```

```
  ggtitle("sous séries saisonnières: airline")
```

Graphiques de saison

sous séries saisonnières: airline



5

Autocorrélations

Autocorrélations

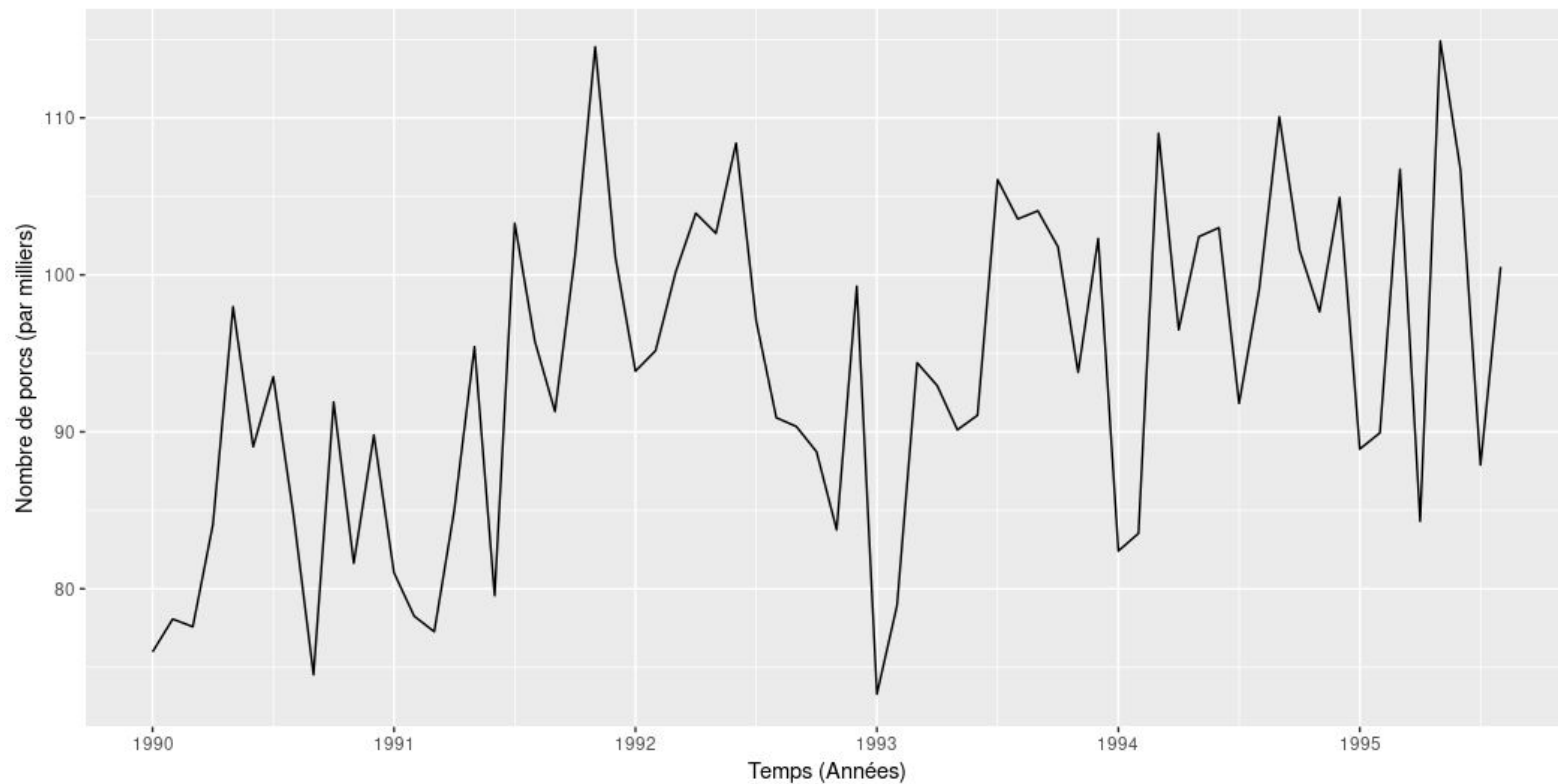
- L'**autocorrélation** mesure le degré de dépendance linéaire qu'il y a entre deux variables du processus $(X_t)_{t \in 1, \dots, T}$; elle est définie par :

$$\rho(k) = \frac{\sum_{t=k+1}^T (X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=1}^T (X_t - \bar{X})^2}$$

- En R, la fonction pour obtenir l'autocorrélation est **ggAcf** ou plus simplement **acf**

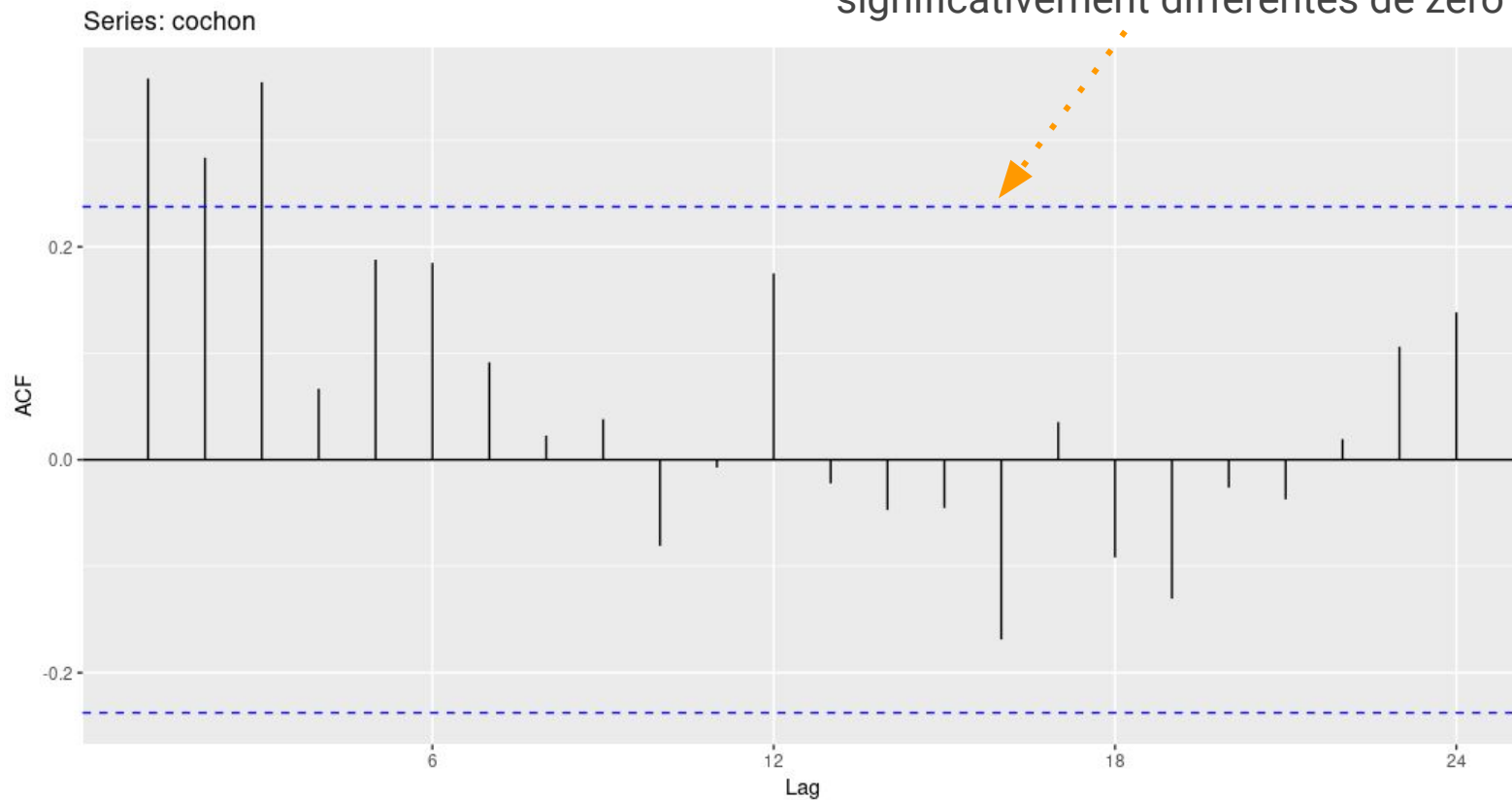
Autocorrélations

Série mensuelle du nombre de porcs abattus dans une province



Corrélogramme

autocorrelations (statistiquement)
significativement différentes de zéro

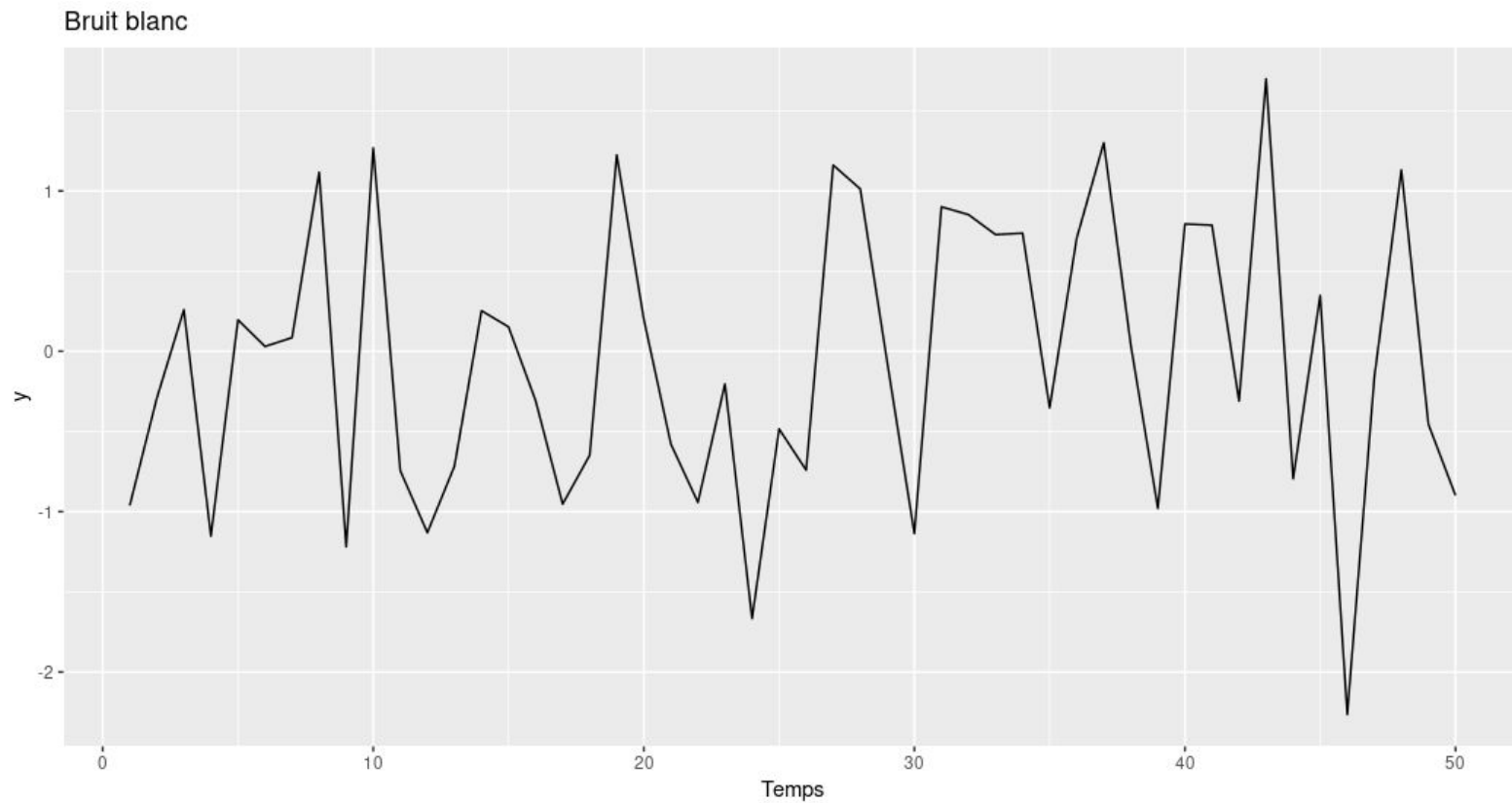


Exemple: bruit blanc

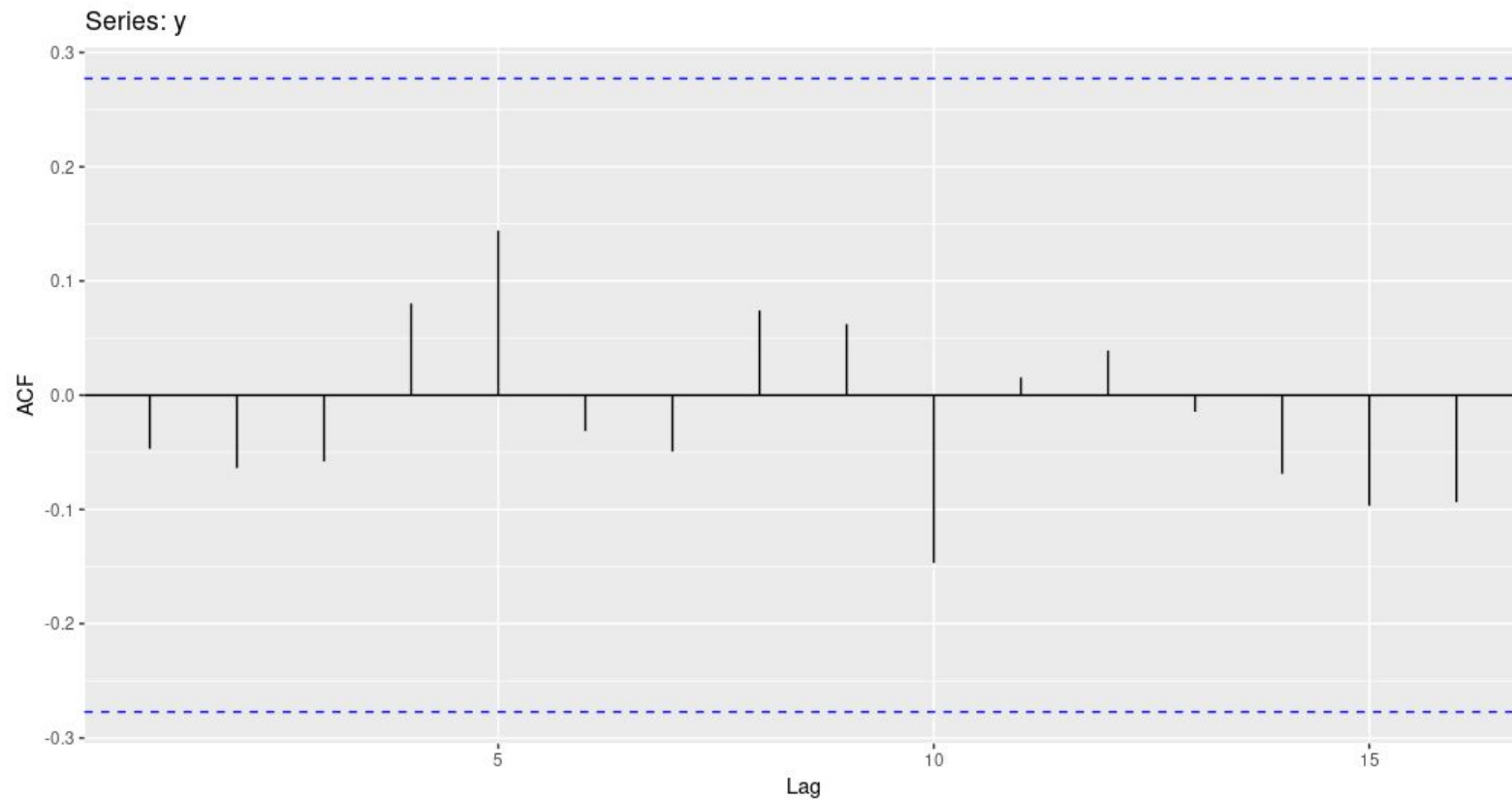
- Un **bruit blanc** est une suite de variables **i.i.d.**.

```
# Bruit blanc généré  
set.seed(3)  
y <- ts(rnorm(50))  
autoplot(y)  
  + ggtitle("Bruit blanc")  
  + xlab("Temps")
```


Exemple: bruit blanc



Exemple: bruit blanc



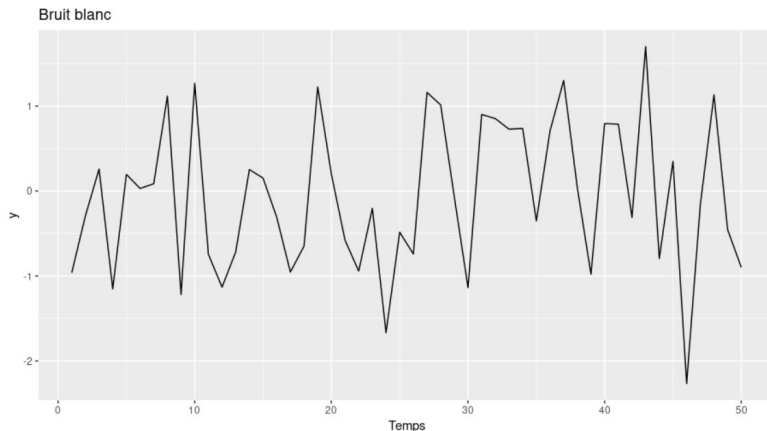
Test de Ljung et Box

- Le **test de Ljung et Box** est un test bâti à partir de la valeur des **h** premières autocorrélations
- Si la **p-value** est très petite, cela signifie que l'on rejette le fait que le processus soit un bruit blanc

#Test de Ljung et Box

```
Box.test(y, lag=24, fitdf=0, type="Lj")
```

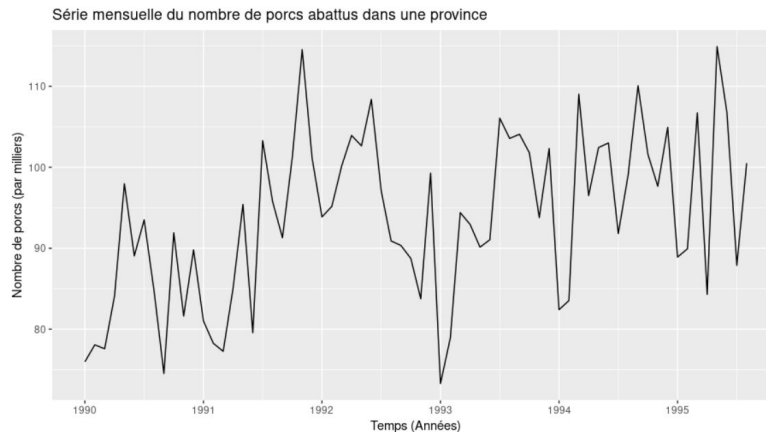
- Donc, on ne rejette pas le fait que la série **y** soit un bruit blanc



Box-Ljung test

```
data: y  
X-squared = 13.353, df = 24, p-value = 0.9599
```

Autre exemple



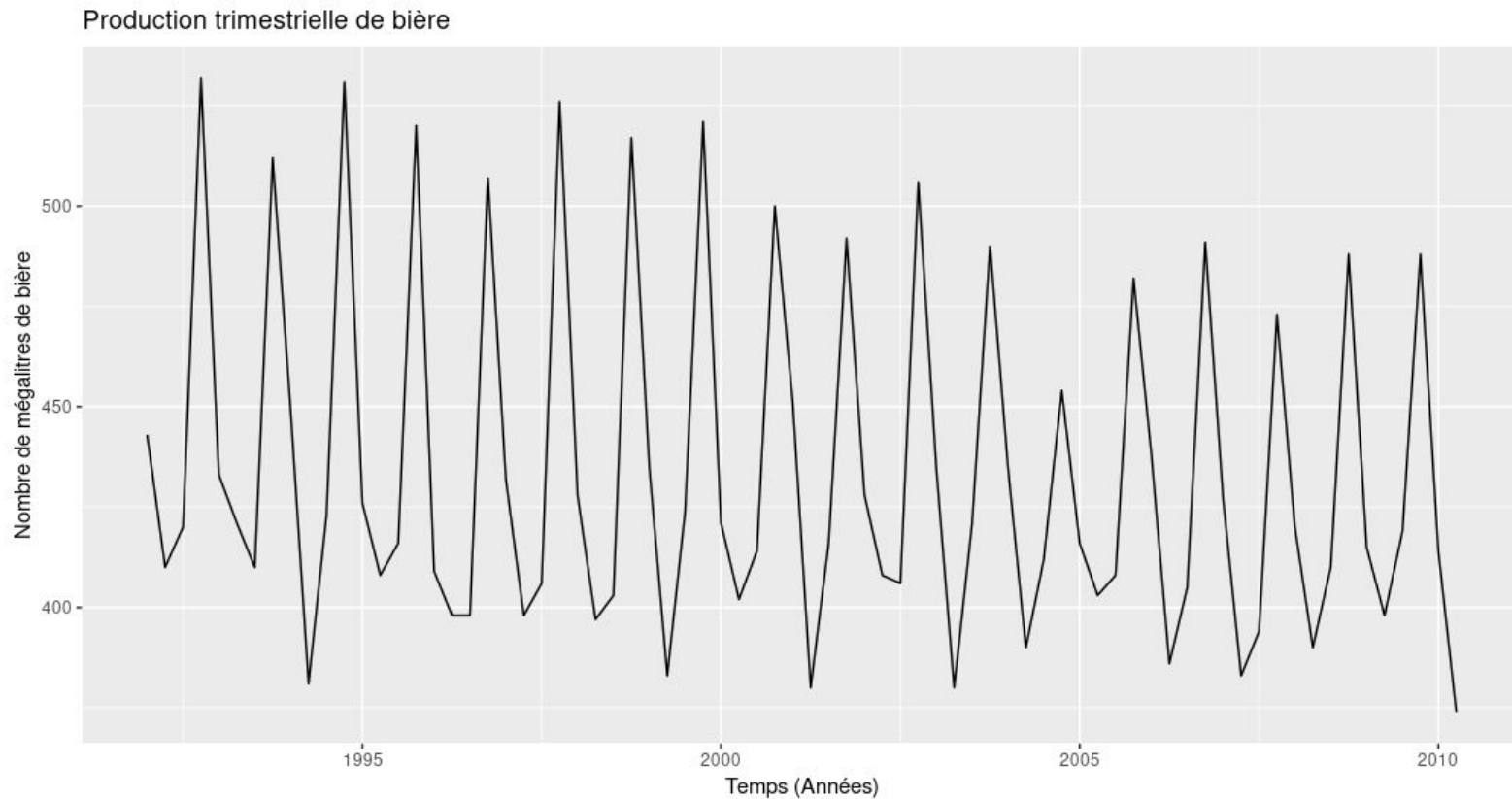
Box-Ljung test

```
data: cochon
```

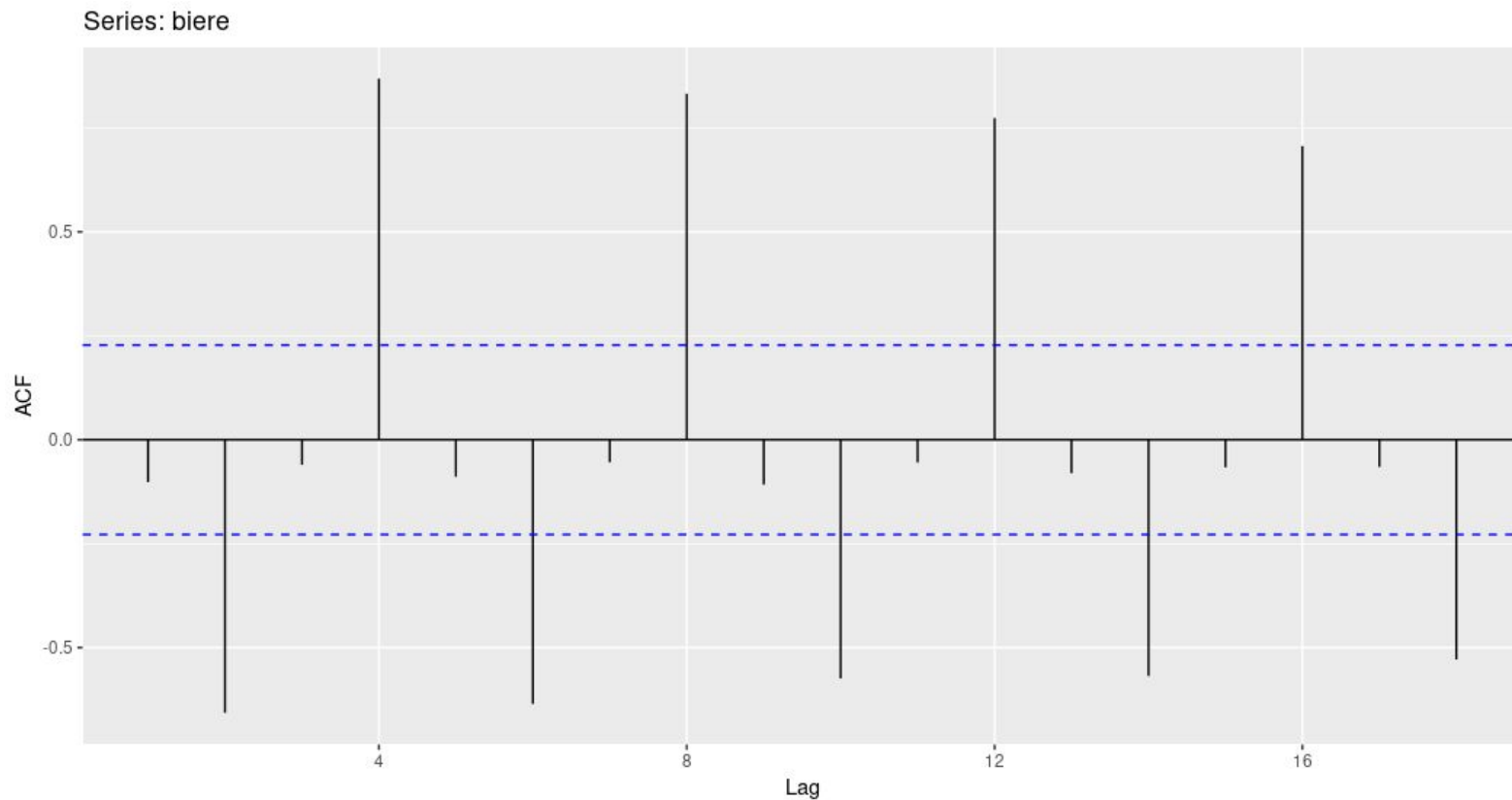
```
X-squared = 42.815, df = 24, p-value = 0.01044
```

On rejette le fait que la série soit un bruit blanc

Tendance et saisonnalité sur les ACF



Tendance et saisonnalité sur les ACF

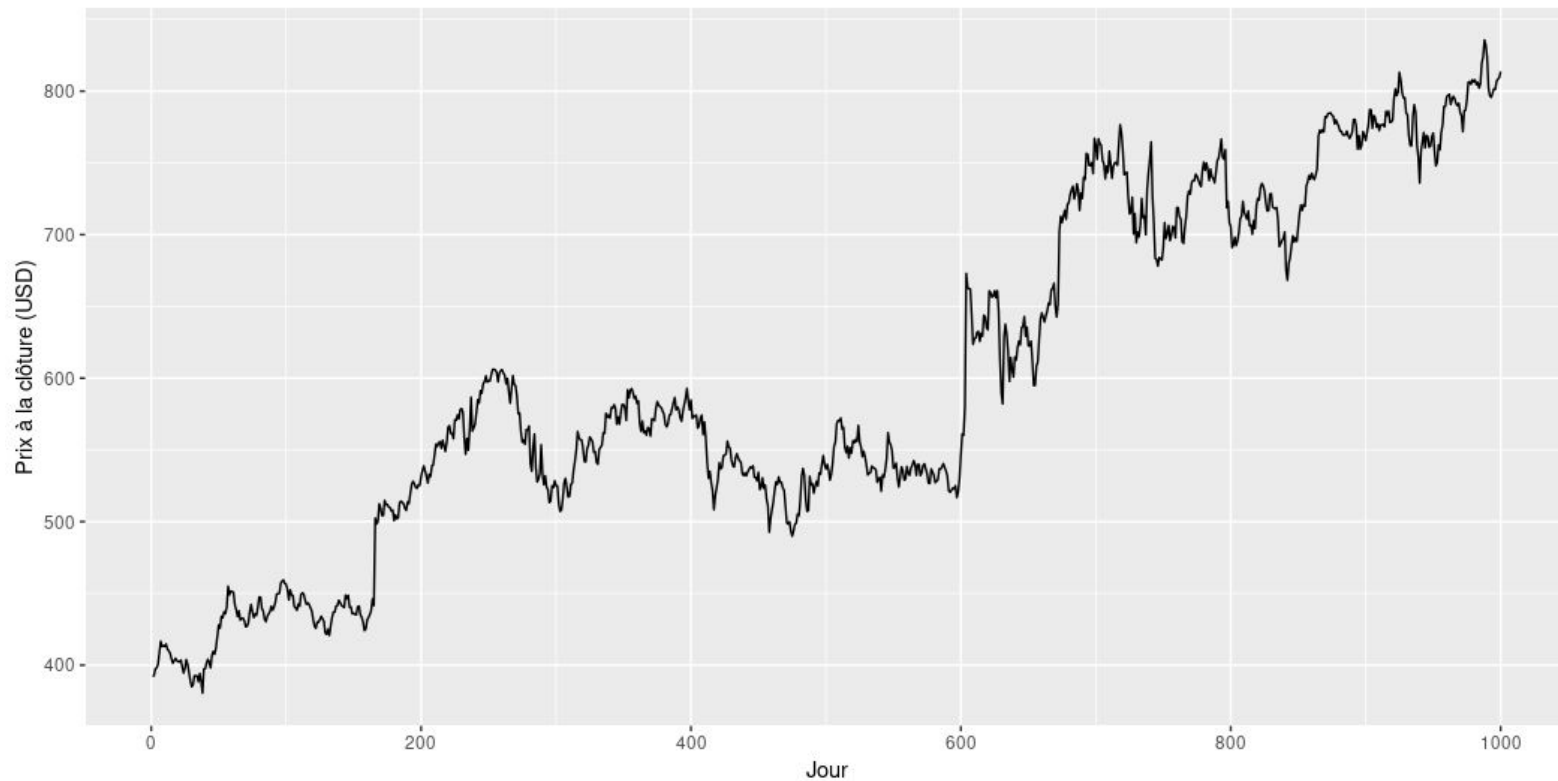


6

Exemple Série Google

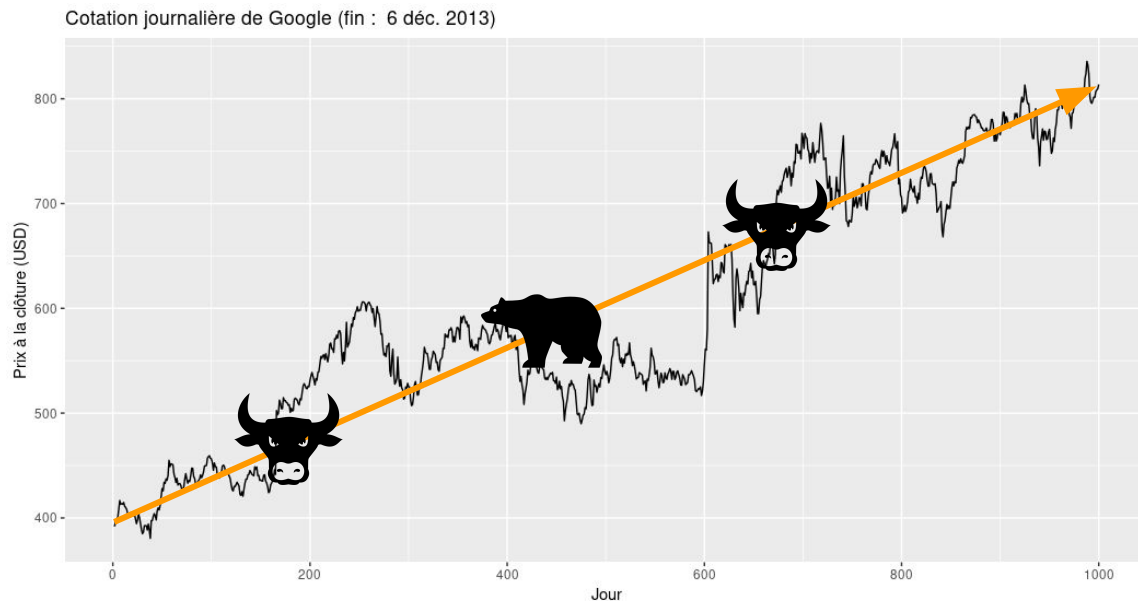
Série Google

Cotation journalière de Google (fin : 6 déc. 2013)

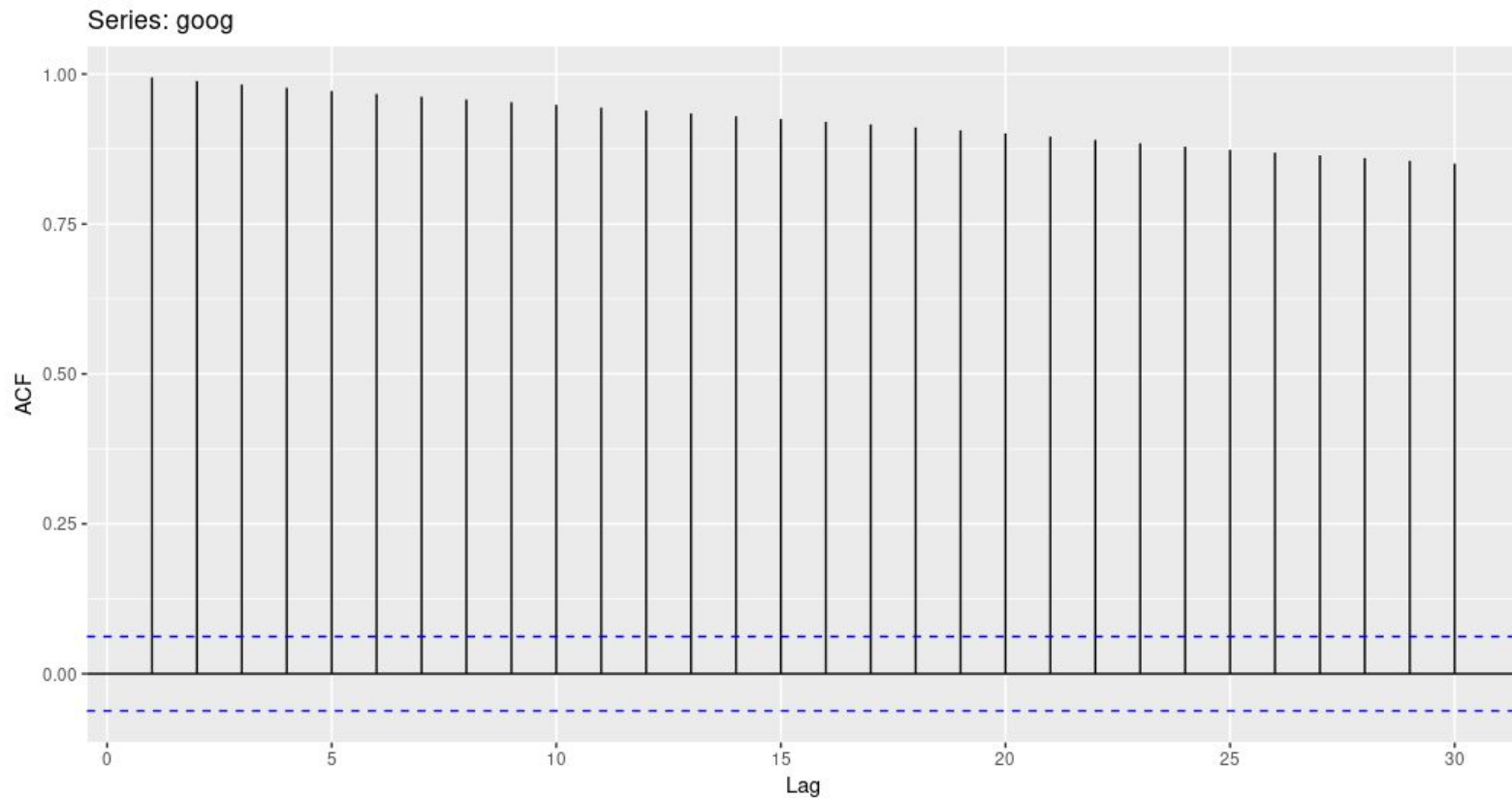


Question

- La série **Google** présente t-elle une tendance et/ou une saisonnalité et/ou des cycles ?



Corrélogramme de l'action Google



Série différenciée

- Pour “retirer” la tendance, il faut **différencier** la série

```
# Série différenciée
```

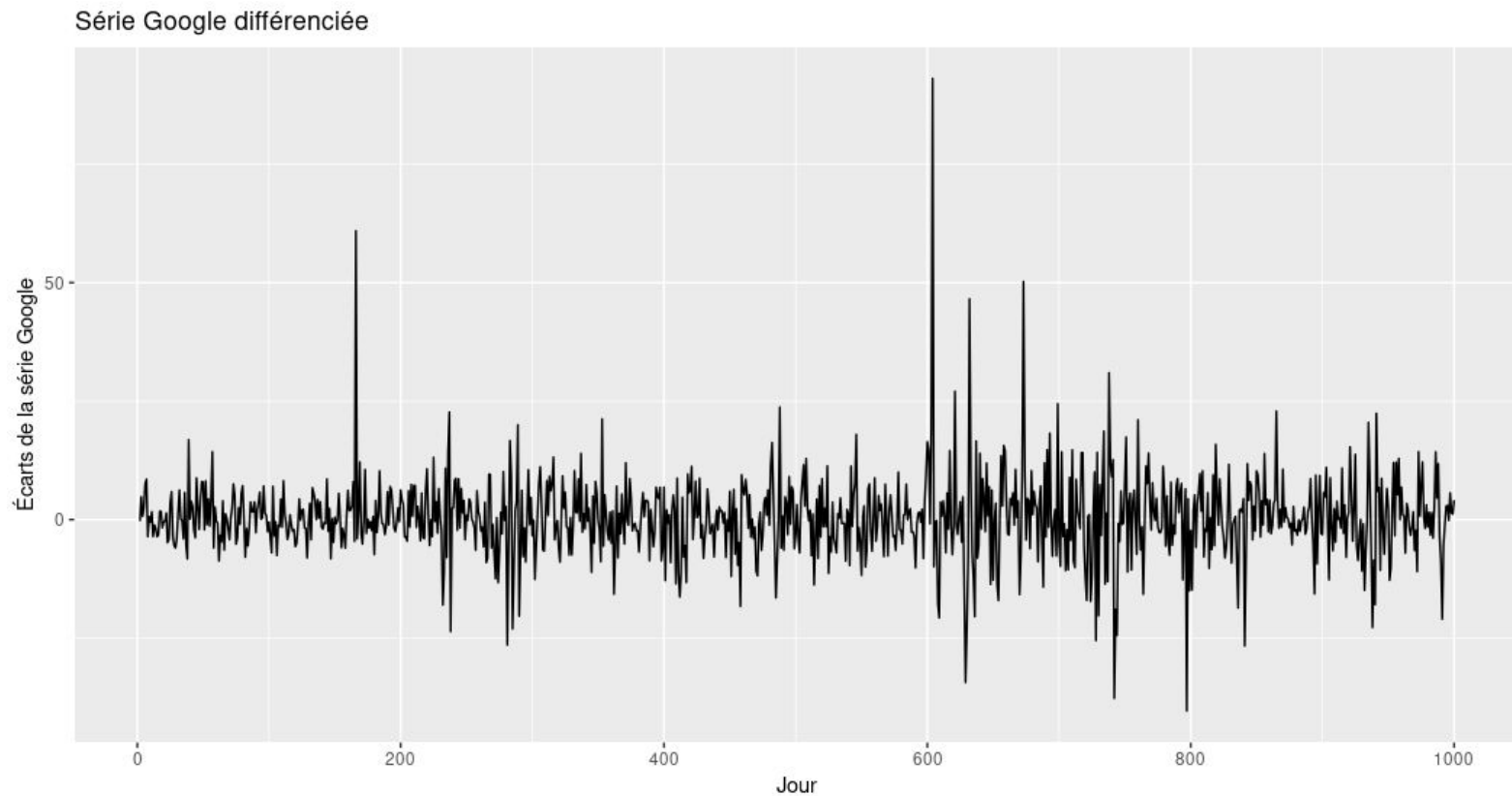
```
googdiff <- diff(goog, differences=1)
```

```
autoplot(googdiff) +
```

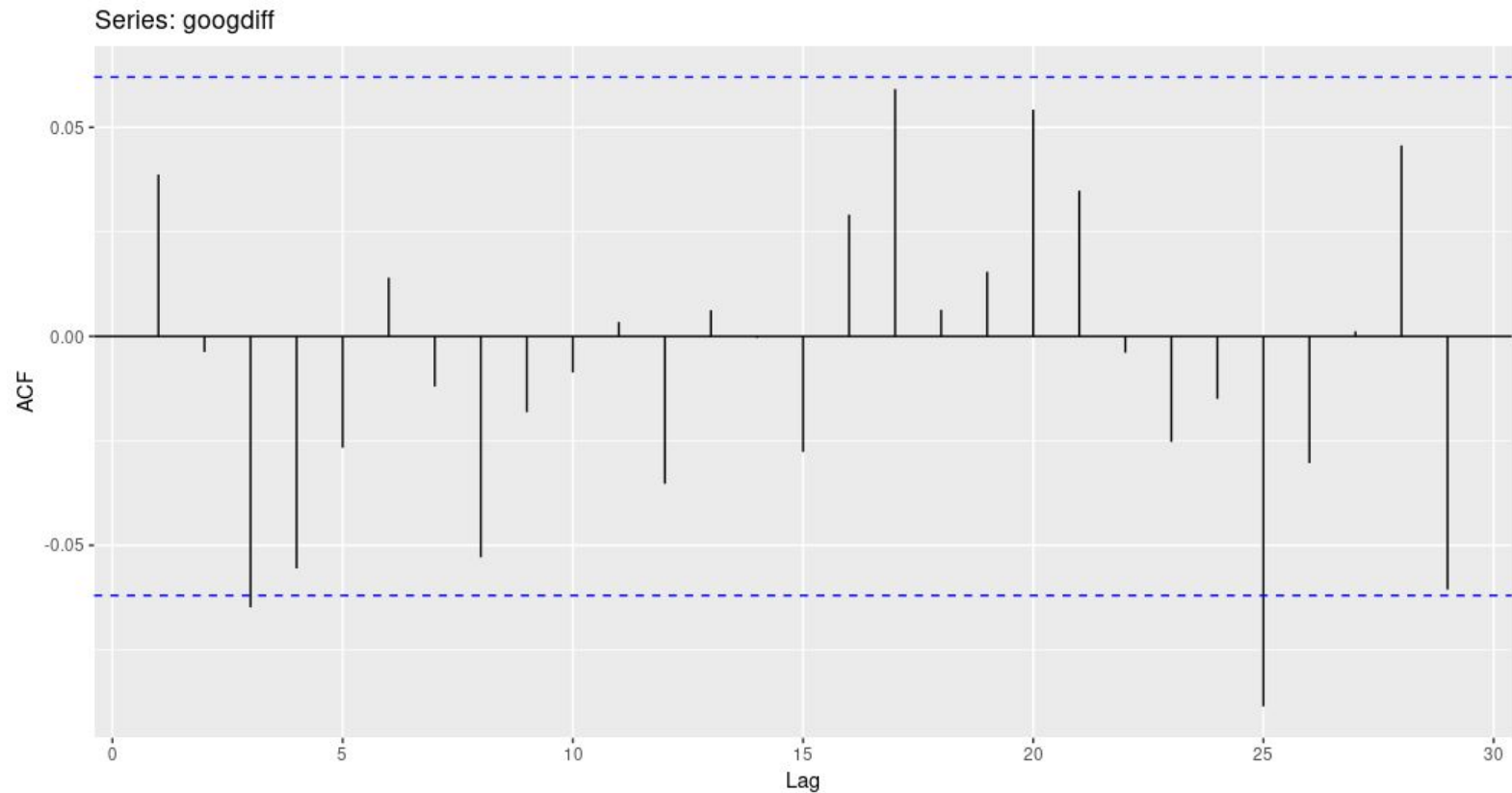
```
  ggtitle("Série Google différenciée") +
```

```
  xlab("Jour") + ylab("Écarts de la série Google")
```

Série différenciée



Corrélogramme de la série Google différenciée



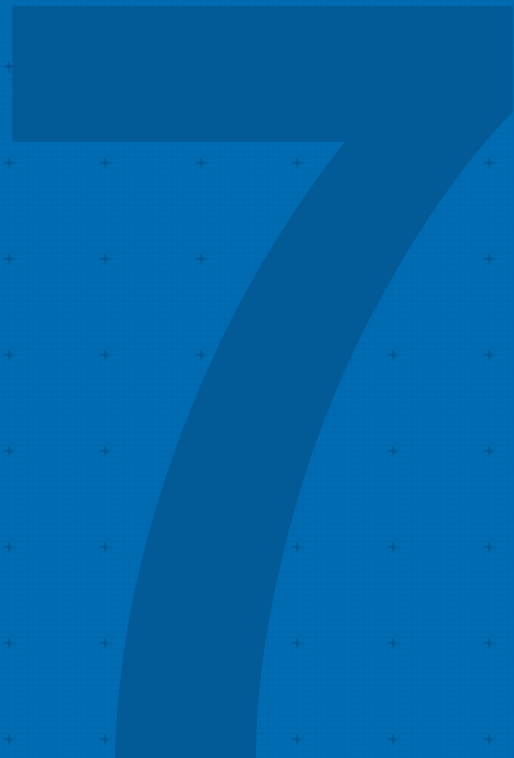
En résumé

- Initialement, la série présente une tendance linéaire croissante
- Cela est corroboré par l'examen du corrélogramme où les autocorrélations successives décroissent lentement
- Après avoir différencié la série : $Y_t = X_t - X_{t-1}$, la série différenciée Y_t ne présente plus de tendance, et le corrélogramme correspondant est quasiment celui d'un bruit blanc !



Pull de <https://github.com/mswawola-cegep/420-a58-sf.git>

04-02-TP



Références

Références

[1] Cours “R et la prévision de séries temporelles” de Michel Carbon - Université Laval