

R et la prévision des séries temporelles

Chapitre 1 - Graphiques de séries temporelles

Michel CARBON

Université Laval de Québec

13 mai 2019



Les bons prévisionnistes ne sont pas plus intelligents que les autres, par contre, leur ignorance est simplement mieux organisée.

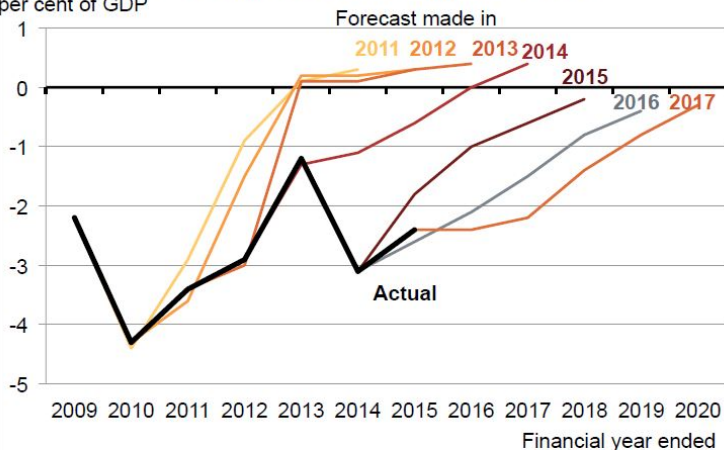


La prévision est difficile...

Commonwealth plans to drift back to surplus
show the triumph of experience over hope

GRATTAN
Institute

Actual and forecast Commonwealth underlying cash balance
per cent of GDP



Encore difficile ...



Que pouvons-nous prévoir ?...



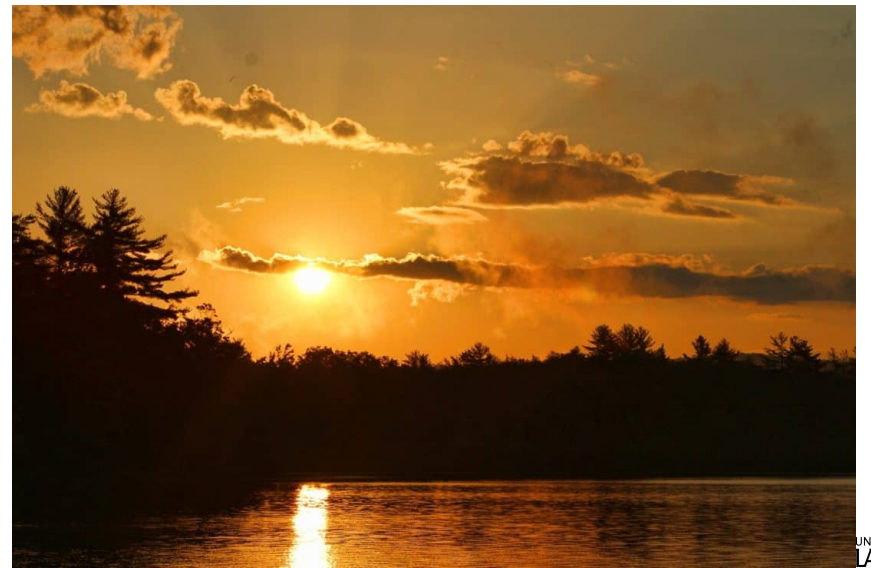
Que pouvons-nous prévoir ?...



Que pouvons-nous prévoir ?...



Que pouvons-nous prévoir ?...



Ce qui est plus facile à prévoir...

- ❶ La demande d'électricité pour les trois jours qui suivent.
- ❷ L'heure du lever de soleil ce jour l'an prochain.
- ❸ La date de la nouvelle apparition de la comète de Halley.
- ❹ Le prix de l'action de Google demain.
- ❺ La température maximale demain à Québec.
- ❻ Le taux de change *US/CAD*.
- ❼ La vente totale d'un certain médicament le mois prochain.



Qu'est-ce qui peut affecter les prévisions?...

Il est plus aisé de prévoir si :

- ❶ On a une assez bonne compréhension des facteurs qui régissent le phénomène.
- ❷ Il y a beaucoup de données disponibles.
- ❸ Les prévisions n'affectent pas ce que l'on cherche à prévoir.
- ❹ Il y a relativement peu de variations aléatoires inexplicables.
- ❺ Le futur est similaire au passé.

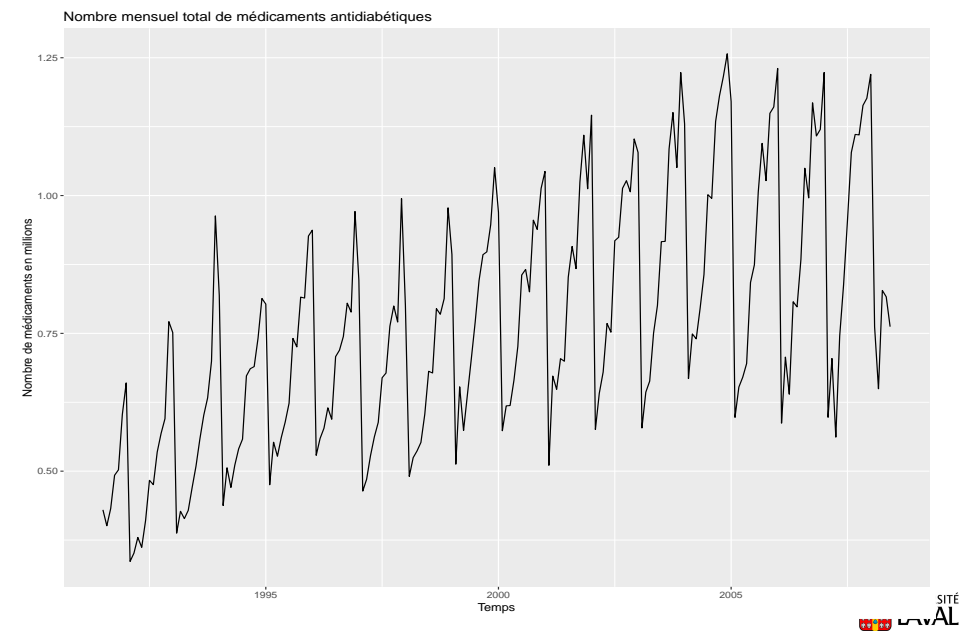


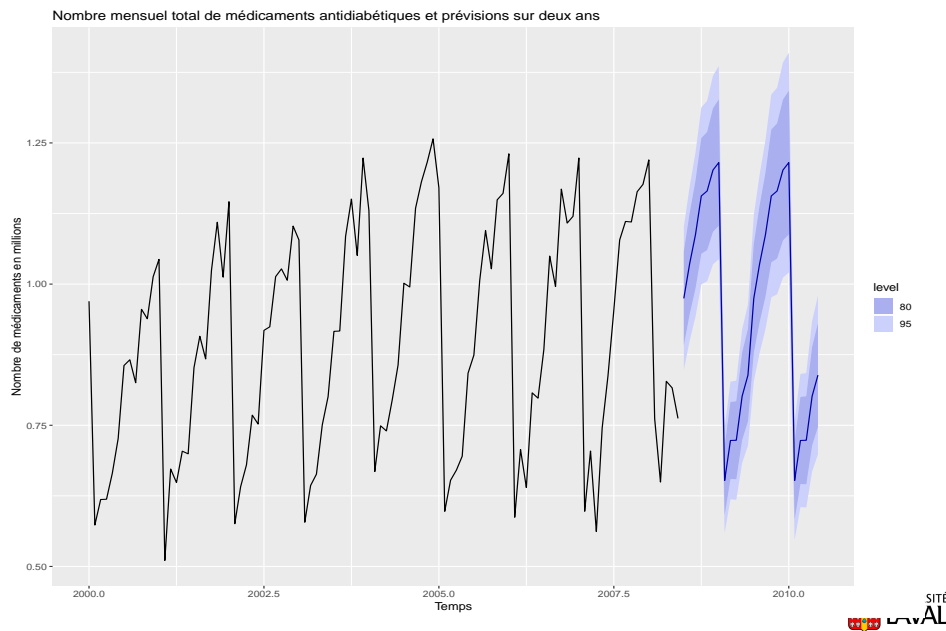
Préliminaire :

Une **série temporelle** est une suite de données recueillies durant un certain laps de temps à intervalles réguliers. On notera : $(X_t)_{t \in \mathbb{Z}}$.

Prévoir cette série, après l'avoir observée un certain temps, c'est estimer comment la suite des observations se poursuivra dans le futur.

La **première étape** dans l'étude d'une série temporelle est la visualisation du graphe des données successives.





Quelques questions...

- Quelles méthodes de prévision sont disponibles pour prendre en compte la tendance, la saison, le cycle ?
- Comment mesurer la pertinence de nos prévisions ?
- Comment choisir un bon modèle de prévision ?

Plan du chapitre 1 :

On détaillera les points suivants :

- Objets "ts"
- Graphiques
- Décomposition d'une série temporelle
- Graphiques de saison
- Autres graphiques

Objets "ts"

Fabriquer des séries temporelles :

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

```
mydata<-as.ts(mydata) # on en fait un objet "série temporelle"
```

Datation

```
# Dater une série temporelle annuelle qui débute en 2010 :
mydata<-ts(mydata,start=2010)

# Dater une série temporelle trimestrielle débutant en 2010-3 :
mydata<-ts(mydata,start=c(2010,3),frequency=4)

# Dater une série temporelle mensuelle débutant en 2010-9 :
mydata<-ts(mydata,start=c(2010,9),frequency=12)
```



Objets et fonctions ts

Type de données	Fréquence	Exemple de "start"
Annuel	1	1997
Trimestriel	4	c(1997,3)
Journalier	7 ou 365.25	1 ou c(1997,227)
Hebdomadaire	52.18	c(1997,22)
Horaire	24 ou 168 ou 8,766	1
Demi-horaire	48 ou 336 ou 17.532	1



Un exemple de code :

```
x<-ts(c(123,45,49,107,153),start=2014)
```

```
Time Series:
Start = 2014
End = 2018
Frequency = 1
[1] 123 45 49 107 153
```



Un exemple

```
library(ggplot2)

airline<- read.table("airline49.dat")

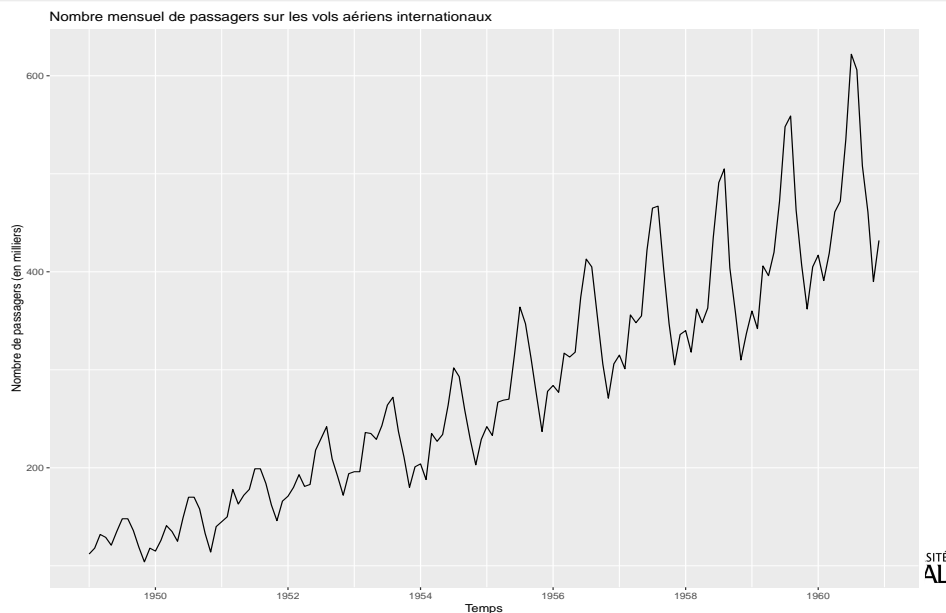
airline<-ts(airline,start=c(1949,1),frequency=12)

# Graphique :

autoplot(airline)+
ggtitle("Nombre mensuel de passagers sur les vols aériens
internationaux")+
xlab("Temps")+
ylab("Nombre de passagers (en milliers)")
```



Un exemple - suite



Michel CARBON R et la prévision des séries temporelles Chapitre 1 - Graph

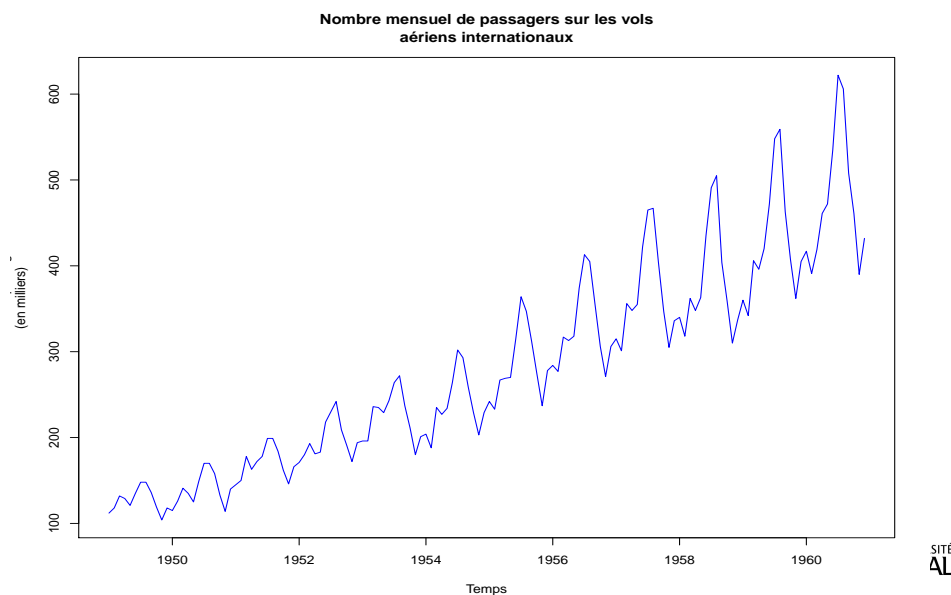
Un exemple - suite

Autre manière plus simple sans l'utilisation de "ggplot2" :

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

Michel CARBON R et la prévision des séries temporelles Chapitre 1 - Graph

Un exemple - suite



Michel CARBON R et la prévision des séries temporelles Chapitre 1 - Graph

Décomposition d'une série temporelle

La décomposition classique d'une série temporelle se décline en quatre composantes :

la tendance, la saisonnalité, le cycle et la partie résiduelle.

Michel CARBON R et la prévision des séries temporelles Chapitre 1 - Graph

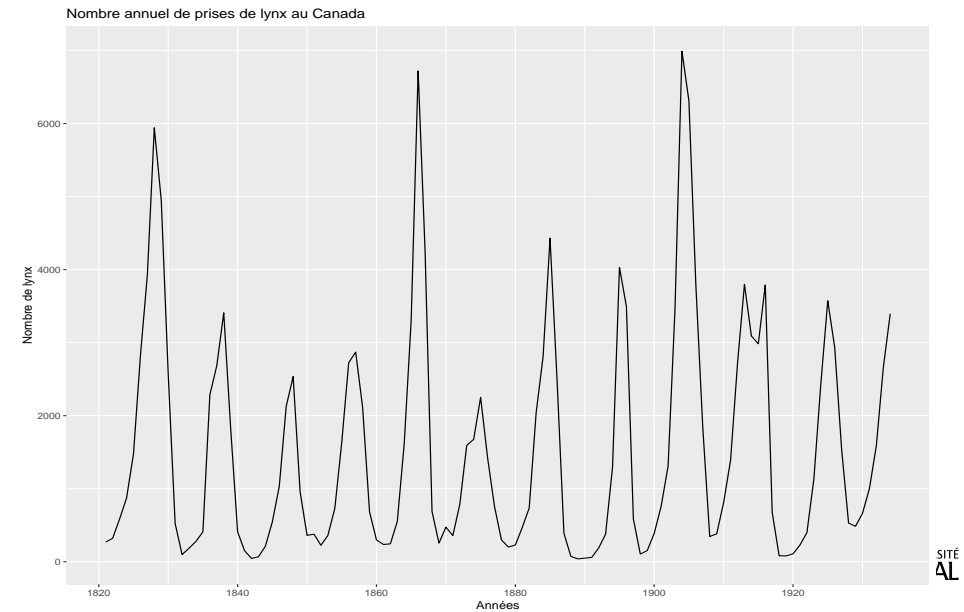
Décomposition d'une série temporelle - suite

- La **tendance** est le comportement à long terme de la série (voir la série "airline")
- La **saisonnalité** ou **partie saisonnière** apparaît lorsqu'un facteur saisonnier survient par exemple chaque année de manière similaire. La saisonnalité est connue et fixée (voir la série "airline").
- Le **cycle** apparaît lorsque la série a des périodes de croissance et décroissance qui reviennent de façon non régulière.
- La **partie résiduelle**, ou **résidu**, est ce qu'il reste après avoir ôté de la série initiale les précédentes composantes.

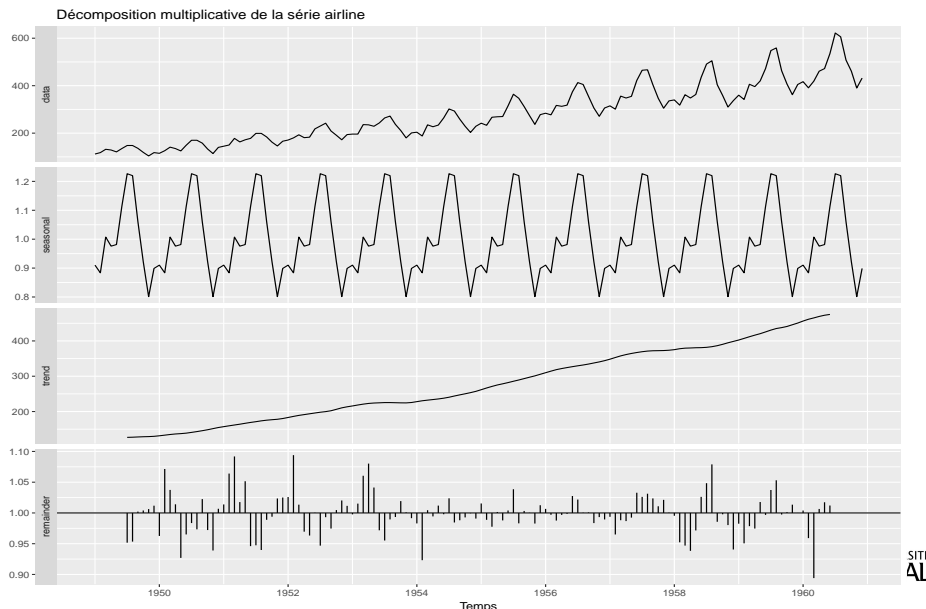
La décomposition peut être **additive** ou **multiplicative**.



Exemple : saison ou cycle ?



Décomposition d'une série temporelle - suite

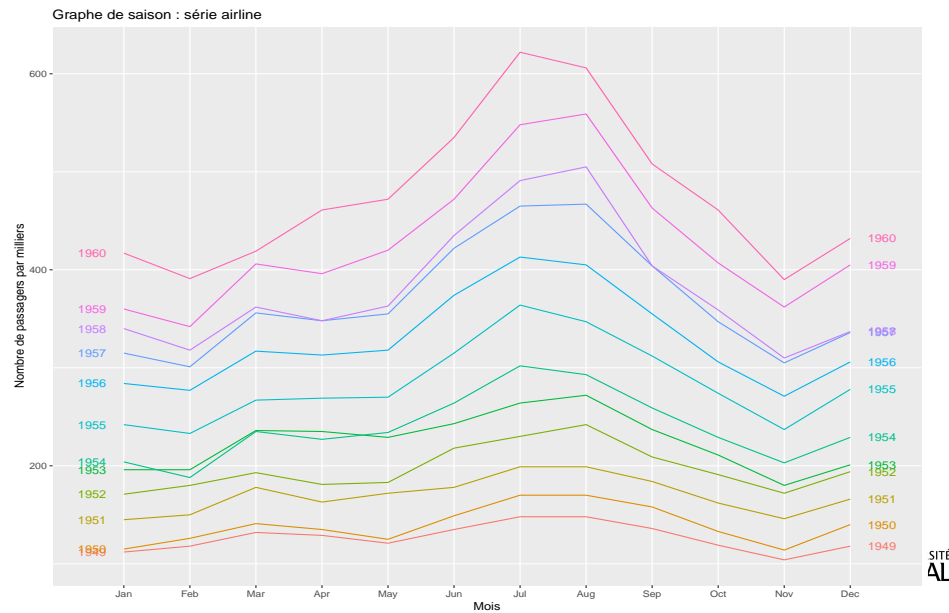


Graphique de saison

Le graphe montre la série temporelle découpée sur chaque saison pour mieux en appréhender les caractéristiques.

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

Graphique de saison - suite



Michel CARBON R et la prévision des séries temporelles Chapitre 1 - Graph

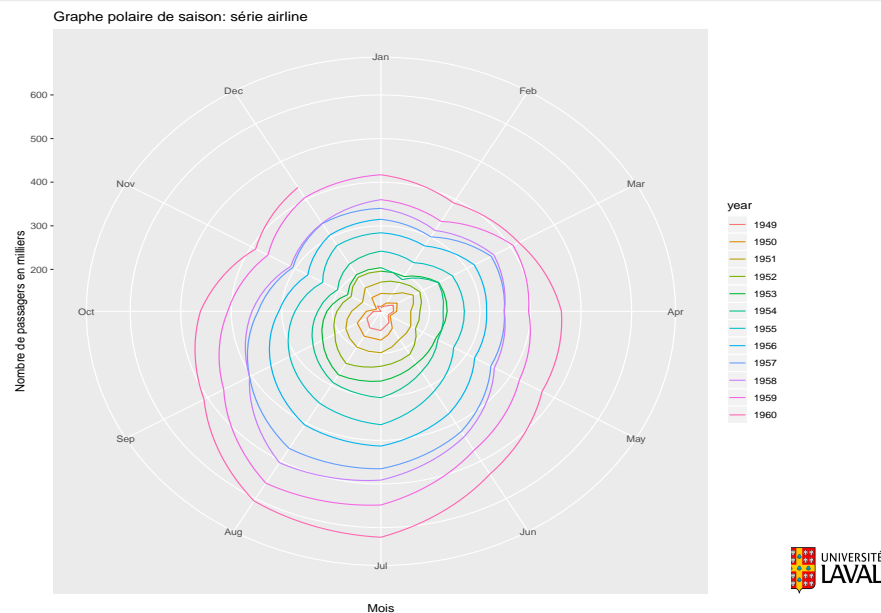
Graphique de saison - suite

On peut, au lieu de faire un axe horizontal, faire un axe de rotation :

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

Michel CARBON R et la prévision des séries temporelles Chapitre 1 - Graph

Graphique de saison - suite



Michel CARBON R et la prévision des séries temporelles Chapitre 1 - Graph

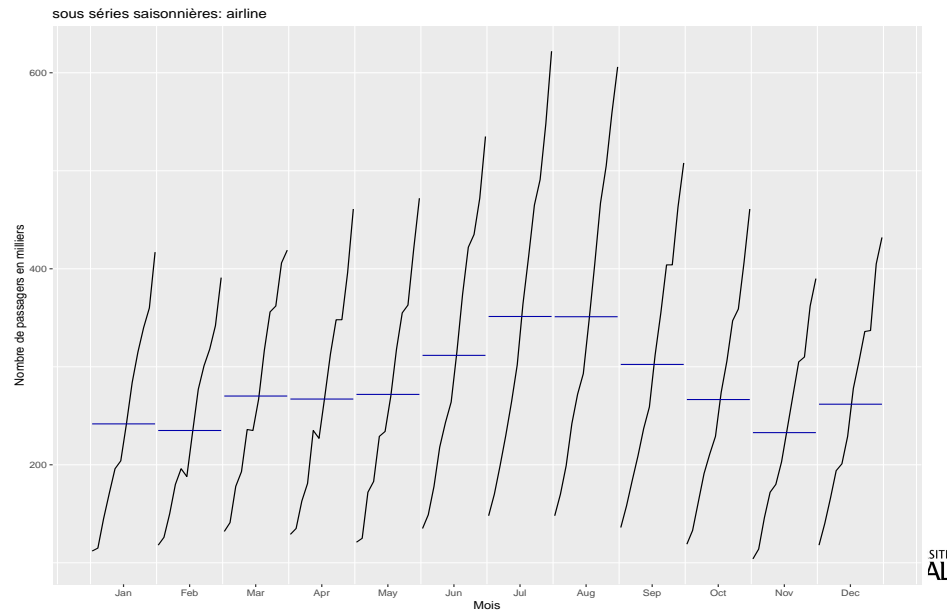
Graphique de saison - suite

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

```
ggsubseriesplot(airline) + ylab("Nombre de passagers en milliers") + ggtitle("sous séries saisonnières : airline")
```

Michel CARBON R et la prévision des séries temporelles Chapitre 1 - Graph

Graphique de saison - suite



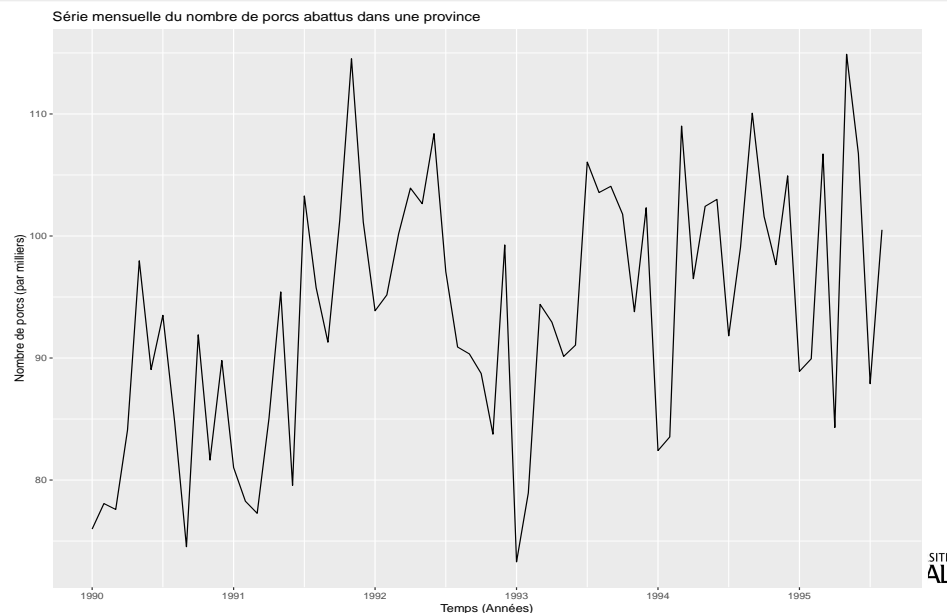
Autocorrélations

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 :

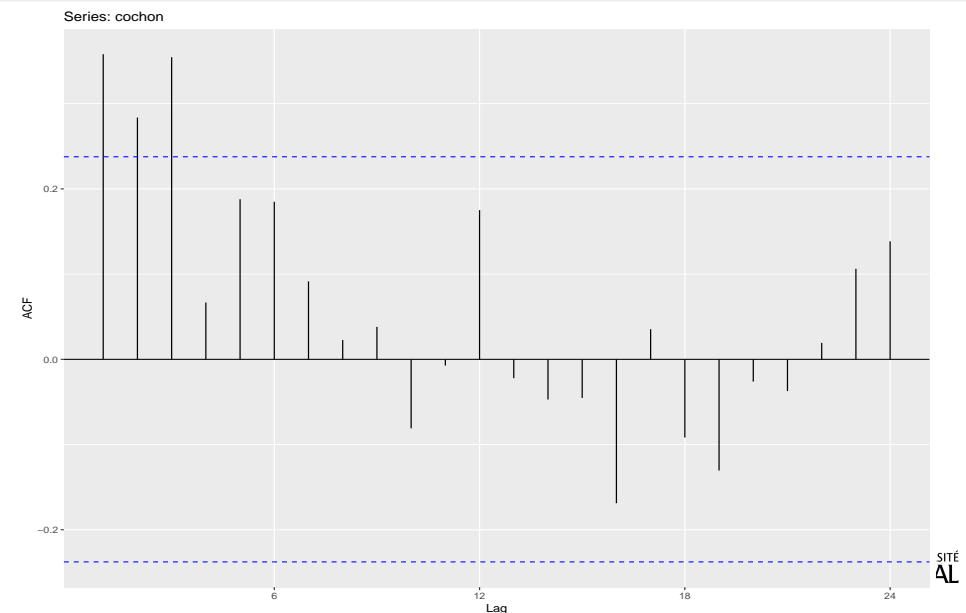
$$\rho(k) = \frac{\sum_{t=k+1}^T (X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=1}^T (X_t - \bar{X})^2}$$

Le code "R" est : **ggAcf(X)** ou plus simplement : **acf(X)**.

Exemple



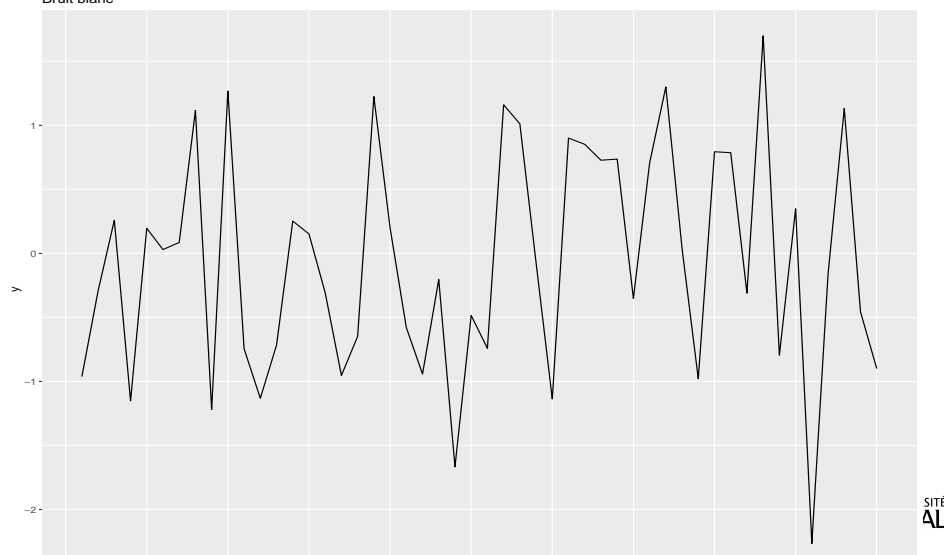
Exemple de corrélogramme



Bruit blanc

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

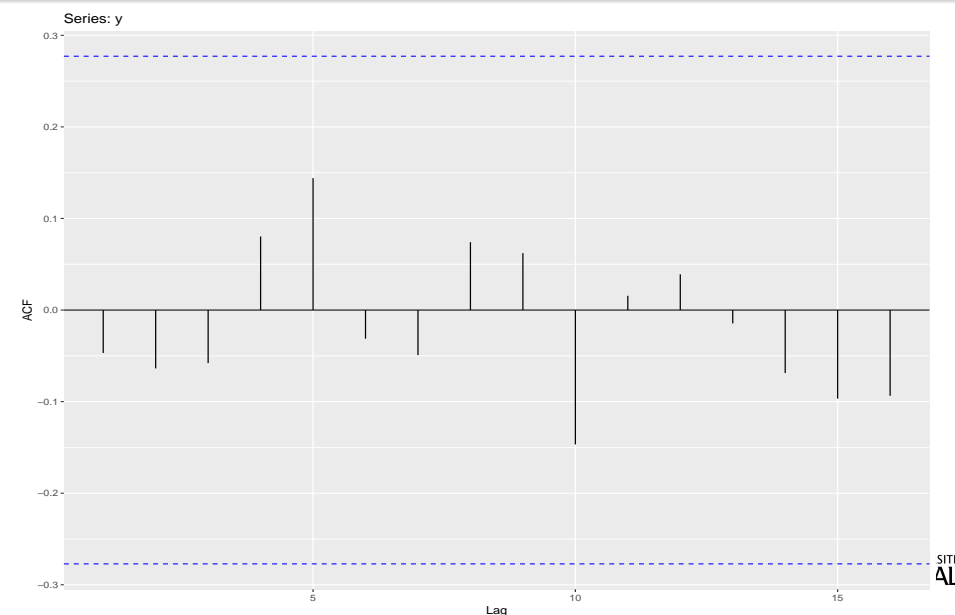
Bruit blanc



Michel CARBON

R et la prévision des séries temporelles Chapitre 1 - Graphiques

Bruit blanc et corrélogramme



Michel CARBON

R et la prévision des séries temporelles Chapitre 1 - Graphiques

Test de Ljung et Box

Le **test de Ljung et Box** est un test bâti à partir de la valeur des premières h autocorrélations.

Si la p -value est très petite, cela signifie que l'on rejette le fait que le processus est un bruit blanc.

Box-Ljung test

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

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

Autre exemple

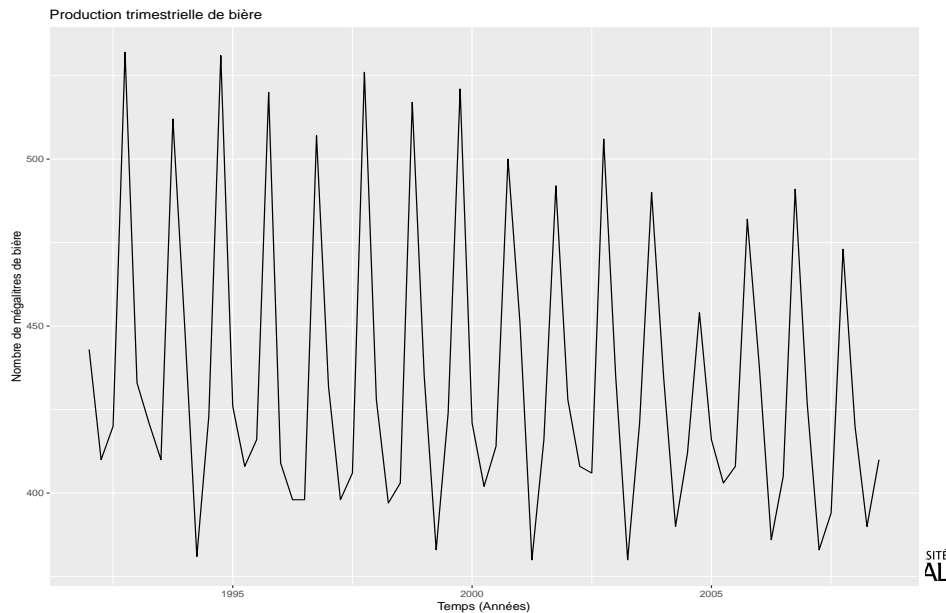
Box-Ljung test

data : cochon, X-squared = 42.815, df = 24,

p-value = 0.01044

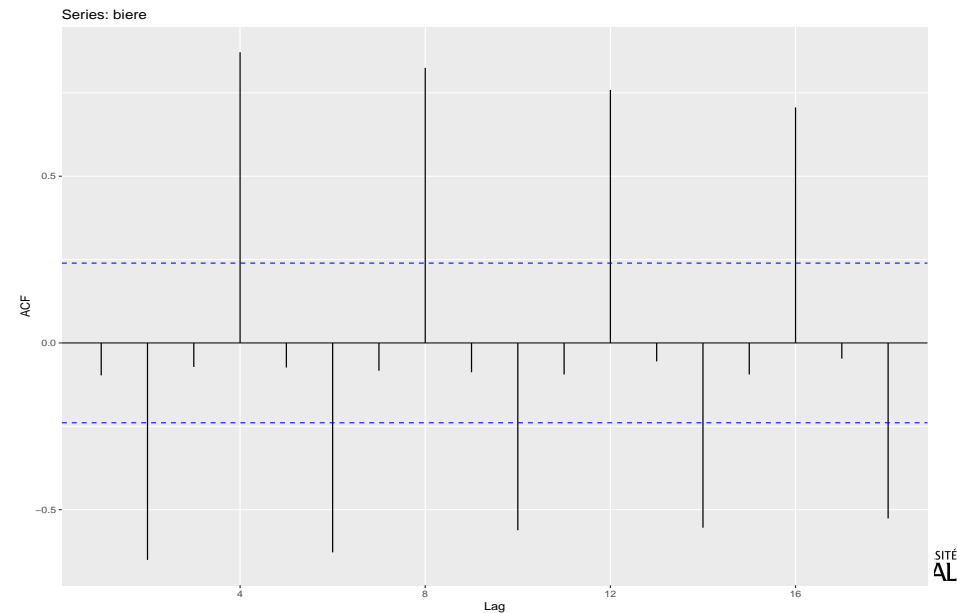
On rejette le fait que la série "cochon" soit un bruit blanc.

Tendance et saisonnalité sur les ACF - Exemple



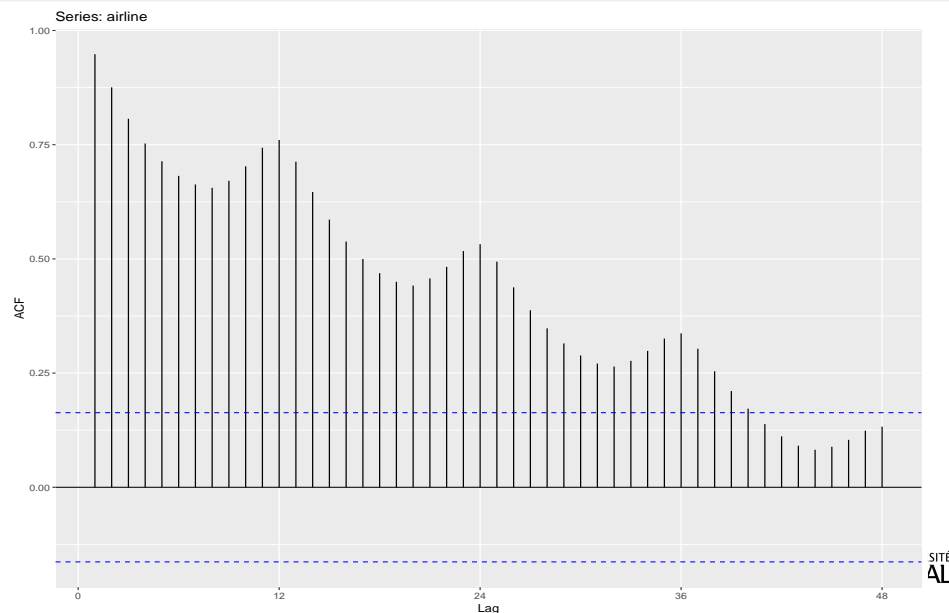
Michel CARBON R et la prévision des séries temporelles Chapitre 1 - Graphiques

Corrélogramme de l'exemple - suite



Michel CARBON R et la prévision des séries temporelles Chapitre 1 - Graphiques

Corrélogramme de la série airline



Michel CARBON R et la prévision des séries temporelles Chapitre 1 - Graphiques

Corrélogramme de la série airline - suite

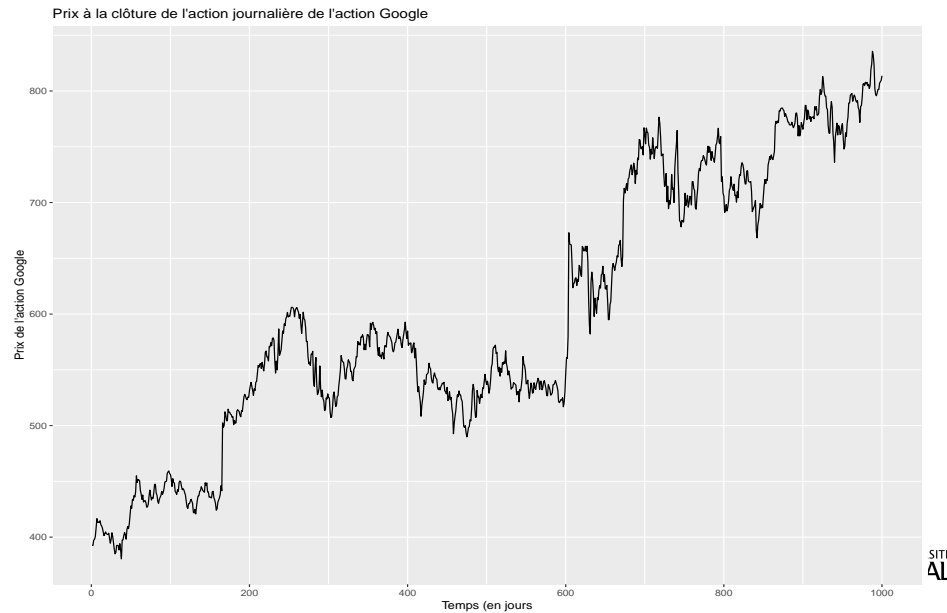
La série **airline** montre une tendance croissante et une saisonnalité marquée.

On retrouve ces particularités sur le corrélogramme :

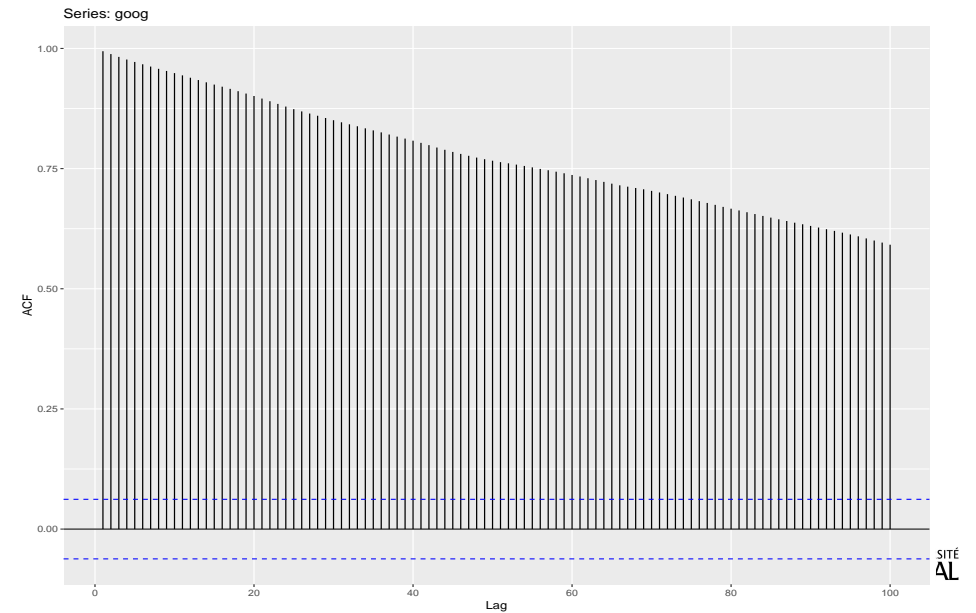
- Décroissance lente des autocorrélations successives.
- Pics sur les multiples de la saison (ici 12, 24, ...).

Michel CARBON R et la prévision des séries temporelles Chapitre 1 - Graphiques

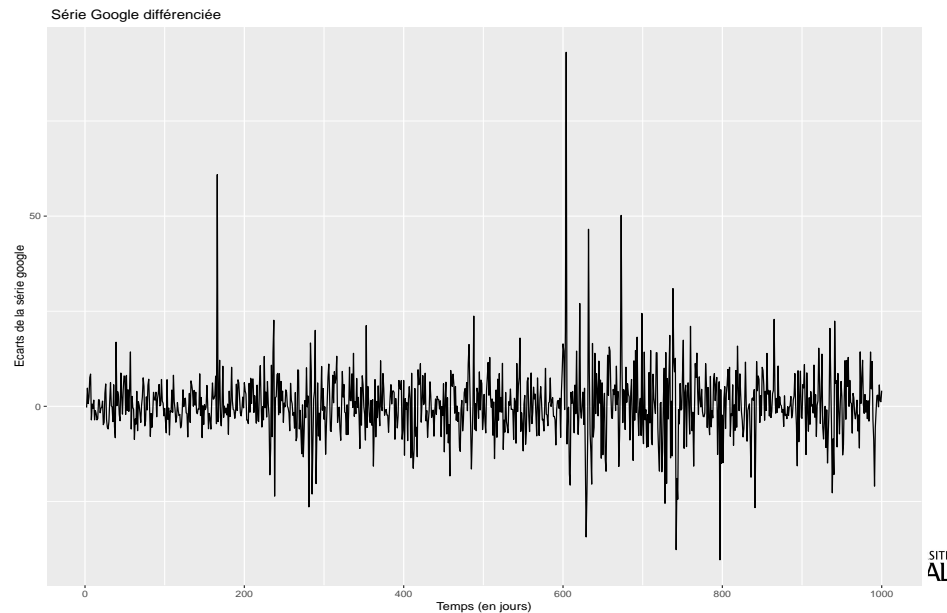
Exemple : Action Google



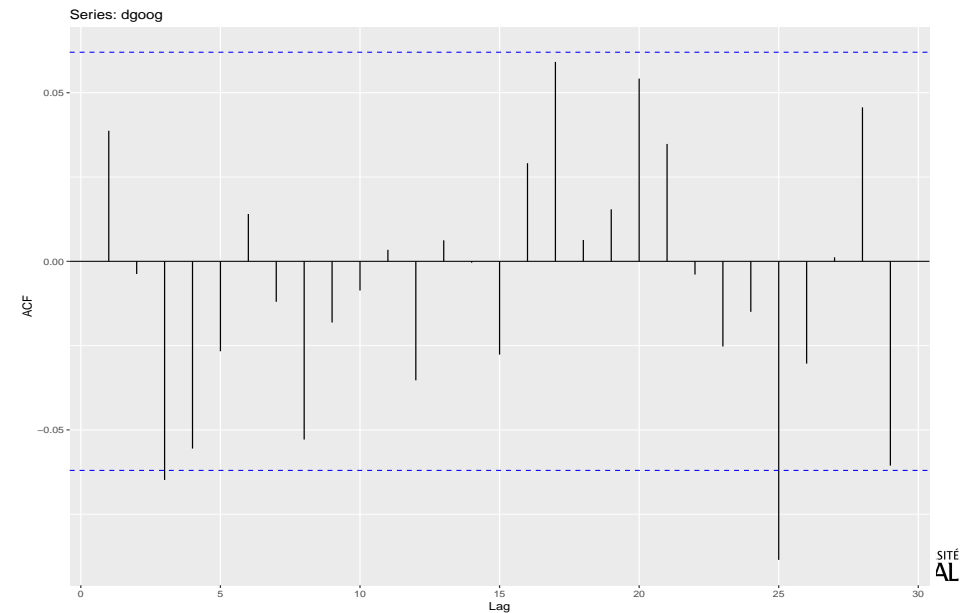
Exemple : Action Google - Corrélogramme



Exemple : Action Google - Série différenciée



Exemple : Action Google - Série différenciée



Résumé sur l'action Google

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 !



Exercice

À vous !

- 1 Importez le fichier "tute1.csv". Il s'agit de ventes trimestrielles d'une petite compagnie (variable "Ventes") entre 1981 et 1995. Il y a aussi une variable Budget Publicitaire ("BudgPubl") et le Produit Intérieur Brut ("PIB").
- 2 Convertissez-les en séries temporelles, après avoir ôté la première colonne inutile ici.
- 3 Tracez les trois séries temporelles.
- 4 Décomposez la série en tendance, saison et résidus. Tracez également l'autocorrélogramme.

