CONTENIDO

1. Estacionariedad y estacionalidad

```
library(tseries)

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

library(forecast)
library(zoo)

##
## Attaching package: 'zoo
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

1.1. Estacionariedad

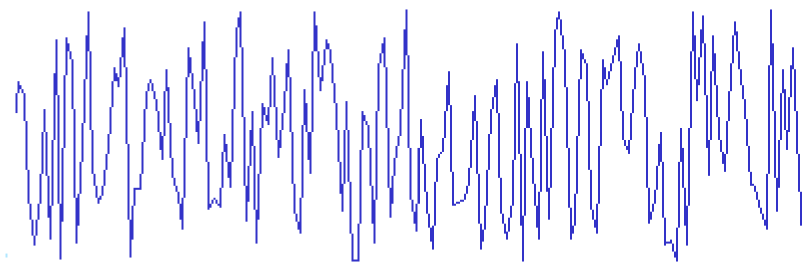
Dada una serie de tiempo X, se define su **función de autocorrelación de retraso** k como:

$$\varphi_k(t) = corr(X_t, X_{t-k})$$

1.1.1. Ruido blanco

El *ruido blanco* es una serie de tiempo $\{\xi_t\}_{t\in T}$ que cumple las siguientes condiciones:

1. Media constante: $E[\xi_t] = cte$ para toda t.
2. Supongamos que que cada hora, durante 3 días (72 horas) lanzo una moneada. Si cae cara, gano 1 peso, si cae cruz, pierdo 2 .
   Esto me va a generar una lista (una secuencia) de 72 datos. Anoto esos 72 datos: $\{x_1, x_2, \ldots, x_{72}\}$
   En la hora t, en promedio ganó: $E[X] = 1 \cdot \frac{1}{2} + (-2) \cdot \frac{1}{2} = -0.5$ Esto significa que $E[X_t] = -0.5$ para cada t (es decir, en cada lanzamiento espero perder 50 centavos).
   Recordemos que la esperanza de una variable aleatoria se define como:
   $$E[X = x] = \sum xP(X = x)$$
3. Varianza constante: $Var(\xi_t) = cte$ para toda t.
4. **Autocorrelación de un retraso nula:** $corr(\xi_t, \xi_{t-1}) = 0$ (Esto significa que mi futuro no depende de mi pasado) para toda t. Esta condición nos dice que la serie es estacionaria (si esto pasa, NO se puede hacer pronósticos)



Típicamente, la gráfica de un ruido blanco se ve así:

Nota: En Teoría de Funciones Generalizadas, el Ruido Blanco es la *derivada distribucional del proceso de Wiener* (Si f es una función con soporte compacto, su derivada distribucional se define como el único funcional continuo f' dado por:

$$f'(\phi) = -\int_R \phi' f$$

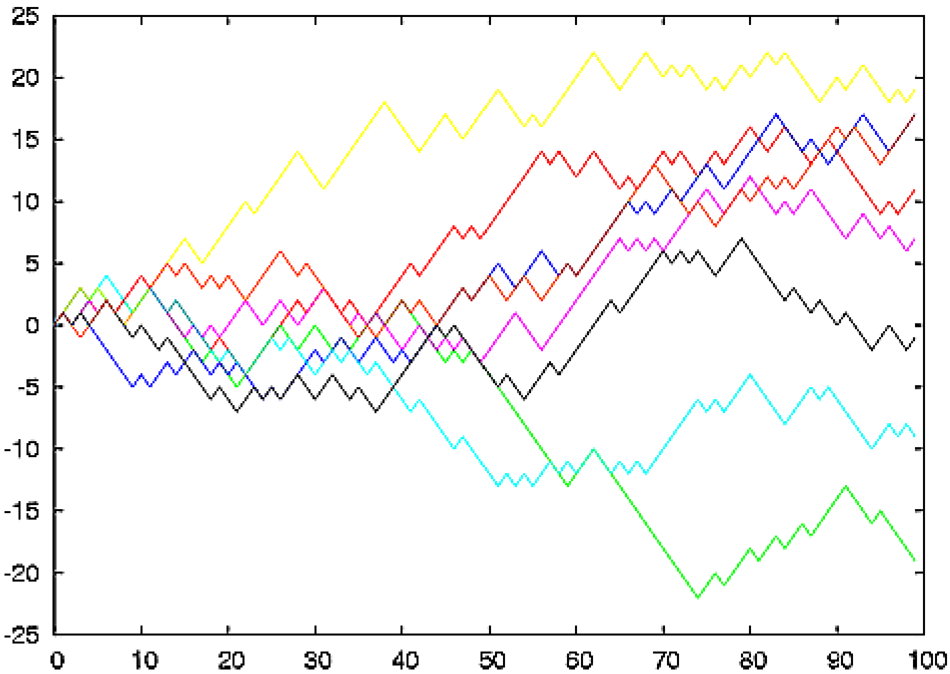
para toda ϕ infinitamente derivable con soporte compacto).

1.1.2. Caminata Aleatoria

$$Y_i = \begin{cases} 1 & p &= 0.5 \\ -1 & p &= 0.5 \end{cases}$$

$$Me = Y_1 + Y_2 + \cdots + Y_n$$

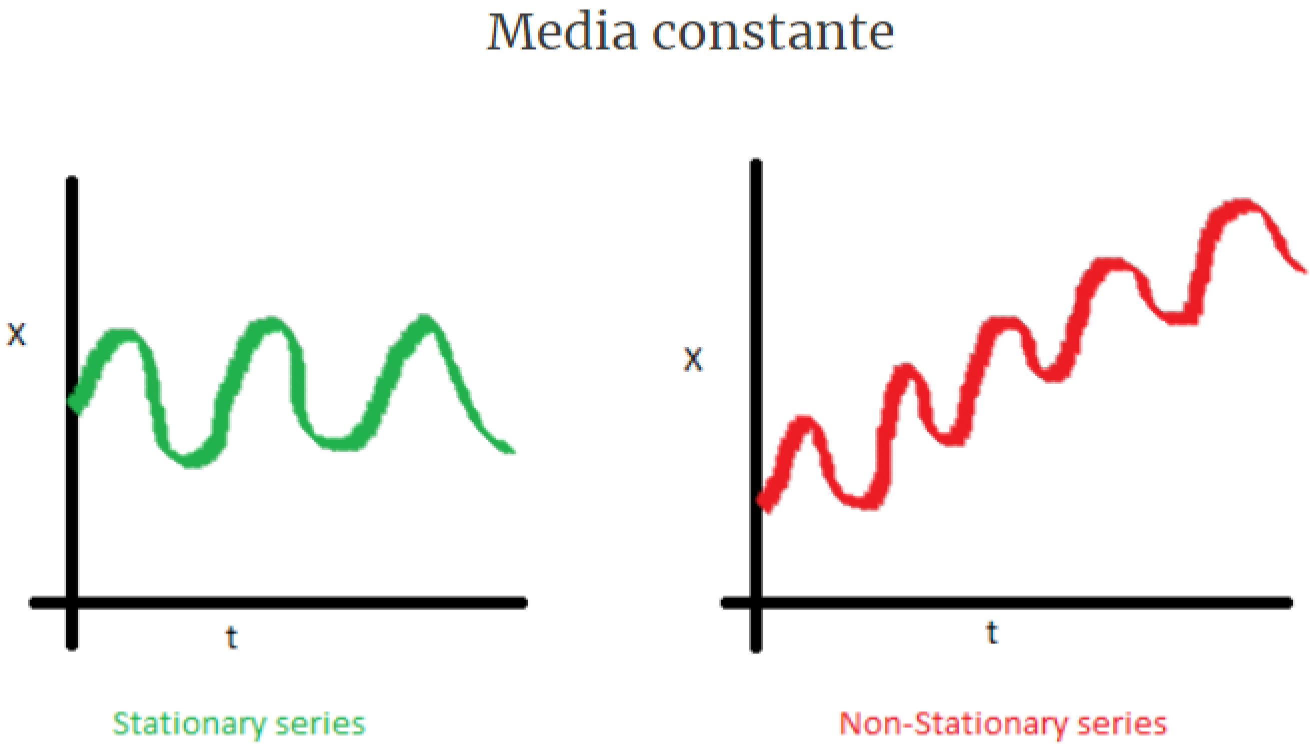
Una **caminata aleatoria** es una serie de tiempo donde las diferencias entre períodos son ruido blanco. Intuitivamente, esto nos dice que el valor X_t únicamente depende del valor X_{t-1} junto con un ruido blanco ξ_t, lo que significa que no somos capaces de predecir el valor a futuro. Esto implica que la mejor predicción para el siguiente paso es el valor actual, pero para períodos de tiempos largos, no habrán buenas precisiones.

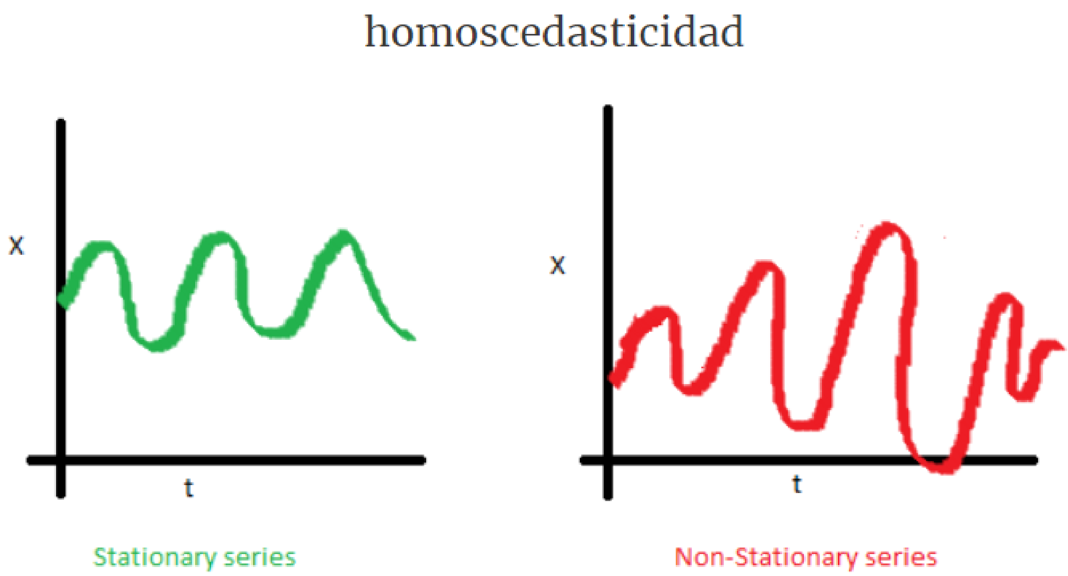


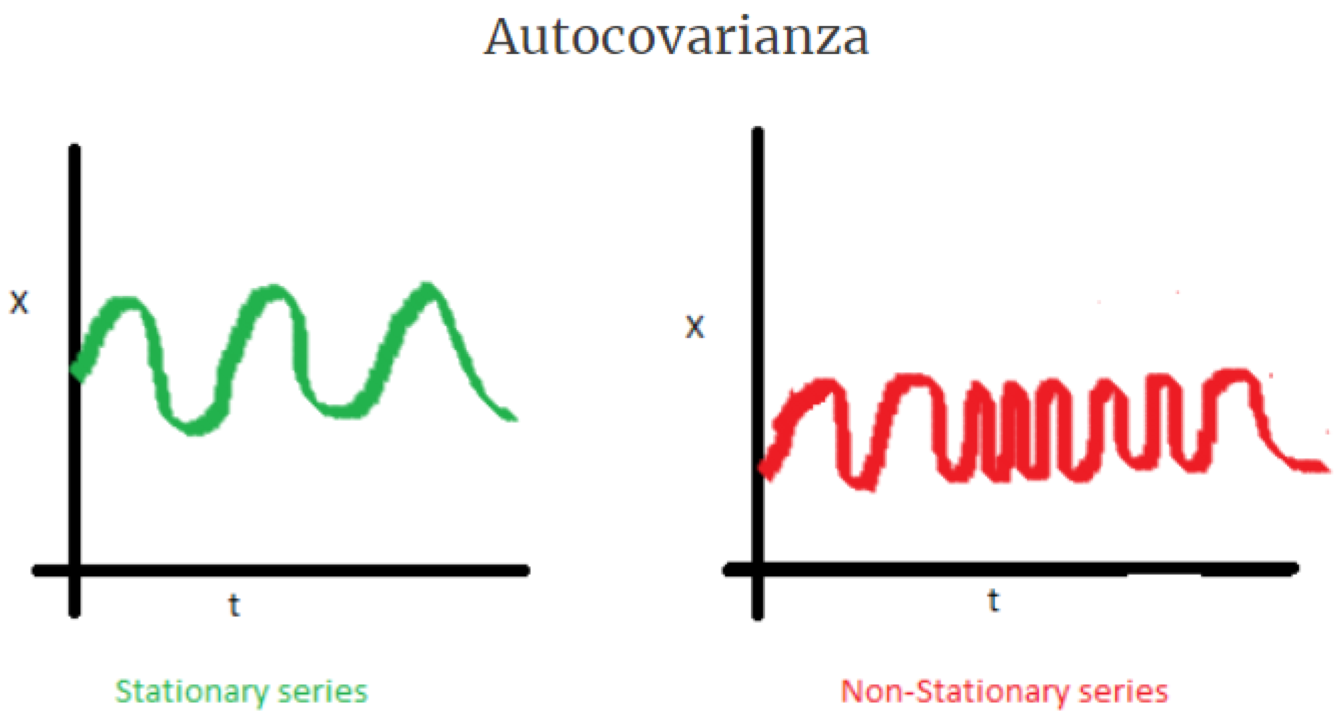
Esto es: $X_{t+1} = X_t + \xi_t$

1.1.3. Series estacionarias

Una **serie estacionaria** si es *estable* a lo largo del tiempo. Esto equivale a que la media, la varianza y la autocorrelación "de cualquier tamaño de paso" se mantienen constantes, lo que significa que los posibles valores en cada tiempo siempre tienen el mismo valor promedio, el mismo proemdio de desviación y la misma relación con el tiempo inmediato anterior, por lo cual no se presentan tendencias.


Media constante
X
t
Stationary series
X
t
Non-Stationary series


homoscedasticidad
x
t
Stationary series
x
t
Non-Stationary series



Formalmente:

1. $E[X_t] = cte.$
2. $Var[X_t] = cte.$
3. Para todas t y s, se tiene $Cov(X_{t+k}, X_t) = Cov(X_{s+k}, X_s)$ toda k

1.1.4. Prueba de Dickey-Fuller (DF)

Es una prueba estadística para verificar si un conjunto de datos proviene de una serie estacionaria o no. Se trata de un contraste de hipótesis.

$$\begin{cases} H_0 & : \textit{La serie no es estacionaria} \\ H_1 & : \textit{La serie es estacionaria} \end{cases}$$

Algo para recordar: Prueba de hipótesis

Se dispone de una muestra tomada de una población. Queremos extender las características de esa muestra a toda nuestra población (es el problema central de la Inferencia Estadística). Uno de los métodos usados son las pruebas o contraste de hipótesis. Estas están formadas por dos hipótesis:

$$\begin{cases} H_0 & : \textit{hipótesis nula} \\ H_1 & : \textit{hipótesis alternativa} \end{cases}$$

y queremos decidir si aceptamos o rechazamos la hipótesis nula en favor de la alternativa.
Supongamos que somos los jueces en un juicio con un sospechoso de un crimen. Nuestra hipótesis nula es que el sospechoso es inocente (toda persona es inocente hasta que se demuestre lo contrario). El trabajo del fiscal consiste en dar pruebas suficientes para que rechacemos la inocencia del sospechosos y, por lo tanto, aceptemos la hipótesis alternativa.
En este caso:

- **Juicio:** Es el contraste de hipótesis.
- **Evidencia:** Son los datos con los que contamos.

- ¿ Cómo tomamos la decisión? típicamente, se establece un criterio que debe cumplir un cierto estadístico para poder decir que tenemos evidencia suficiente para rechazar la hipótesis nula. A este estadístico se le llama *estadístico de contraste.*

Las pruebas DF toma como estadístico de contraste a φ_1(la autocorrelación de un retraso) y tiene la forma:

$$\begin{cases} H_0 : & \varphi_1 & = 1 \\ H_1 : & \varphi_1 & < 1 \end{cases}$$

Si $\varphi_0 < c$ donde c es cierto valor que **nosotros definimos de manera arbitraria**, entonces rechazamos H_0 y por lo tanto la serie es estacionaria.
El p-valor es la probabilidad de que la hipótesis nula es verdadera dadas las observaciones que obtuvimos en el experimento.
O bien, con p-valores: si el p-valor es pequeño (menor que un nivel de significación α), significa que la probabilidad de que H_0 sea cierta dados los datos es pequeña así que es poco probable que la serie NO sea estacionaria, y por lo tanto aceptamos que es estacionaria.

$$\textit{SI} \quad \textit{p-valor} < \alpha \;\; \textit{Decisión: Rechazar } H_0$$

FORMULA DE AUTOCORRELACIÓN (φ_k)

La siguiente ecuación es la fórmula pata calcular k el coeficiente de autocorrelación φ_k entre las observaciones Y_t y Y_{t-k} que se encuentra a k períodos de distancia:

$$\varphi_k = \frac{\sum_{t=k+1}^{n} \left(Y_t - \bar{Y}\right)\left(Y_{t-k} - \bar{Y}\right)}{\sum_{t=1}^{n} \left(Y_t - \bar{Y}\right)^2}$$

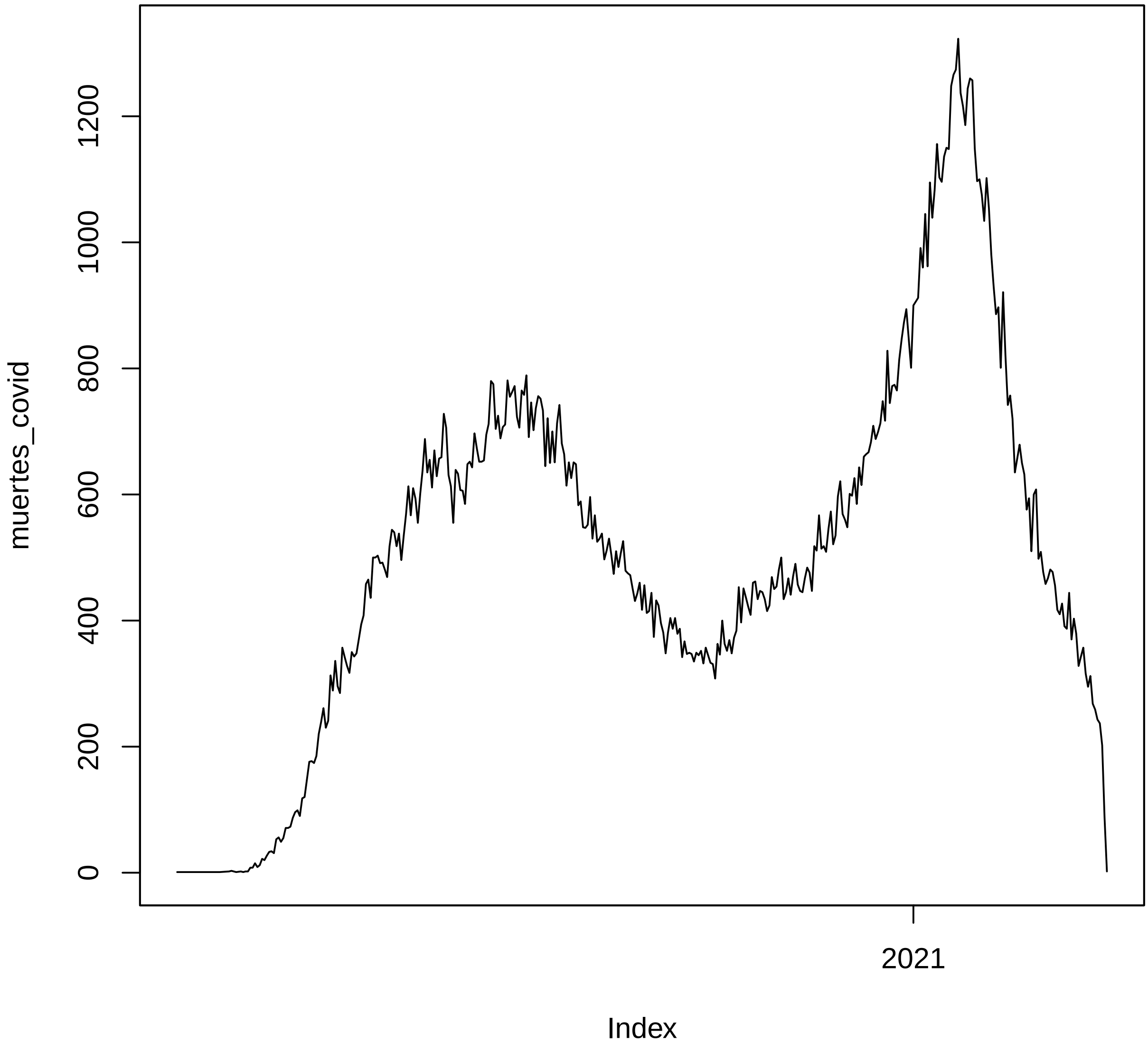
donde:

$$\begin{aligned} \varphi_k &: \textit{Coeficiente de autocorrelación para un retraso de } k \textit{ períodos.} \\ \bar{Y} &: \textit{Media de los valores de la serie.} \\ Y_t &: \textit{Observación en el período } t \\ Y_{t-k} &: \textit{Observación } k \textit{ períodos anteriores o durante un período t-k} \end{aligned}$$

1.1.5. R Prueba de Dickey-Fuller (DF)

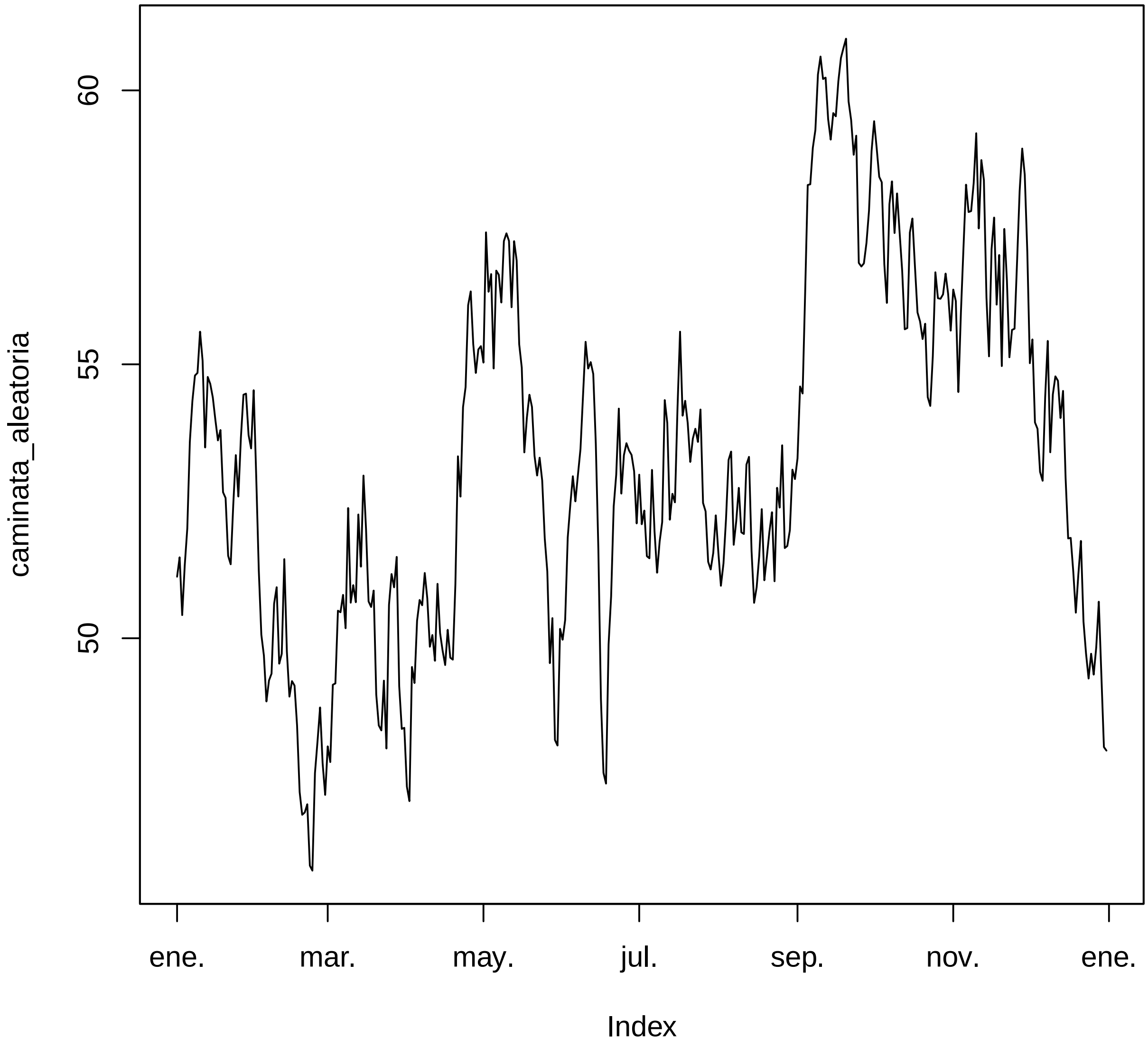
```
setwd("C:\\Users\\81799\\Downloads\\Pronosticos_y_series_de_tiempo\\data")
wn_rw <- read.csv("wn_rw.csv") #Tabla con ruido blanco y caminata aleaotoria
tabla_df <- read.csv("Tabla_rp.csv") #Tabla de delitos.
serie_covid <- read.csv("muertes_covid.csv") #Tabla de muertes covid
tabla_df <- tabla_df[-c(1:24),] #Quitar los primeros 2 años
tiempo_wn_rw <- seq(as.Date("2021/1/1"),as.Date("2021/12/31"),"days")
tiempo_tabla_df <- seq(as.Date("2012/1/1"),as.Date("2019/12/1"),"months")#Hago una secuencia de
head(tiempo_tabla_df,3)

## [1] "2012-01-01" "2012-02-01" "2012-03-01"

#serie de tiempo de ruido blanco
ruido_blanco <- zoo(x=wn_rw$ruido,order.by = tiempo_wn_rw)
#Serie de tiempo de caminata aleatoria
caminata_aleatoria <- zoo(x=wn_rw$caminata,order.by = tiempo_wn_rw)
#Serie de tiempo del total de delitos
total_delitos <- zoo(x=tabla_df$totales,order.by = tiempo_tabla_df,frequency=12)
#Serie de tiempo de muertes covid
muertes_covid <- zoo(serie_covid$Total,as.Date(serie_covid$Fecha))
```

```r
plot(muertes_covid)
```



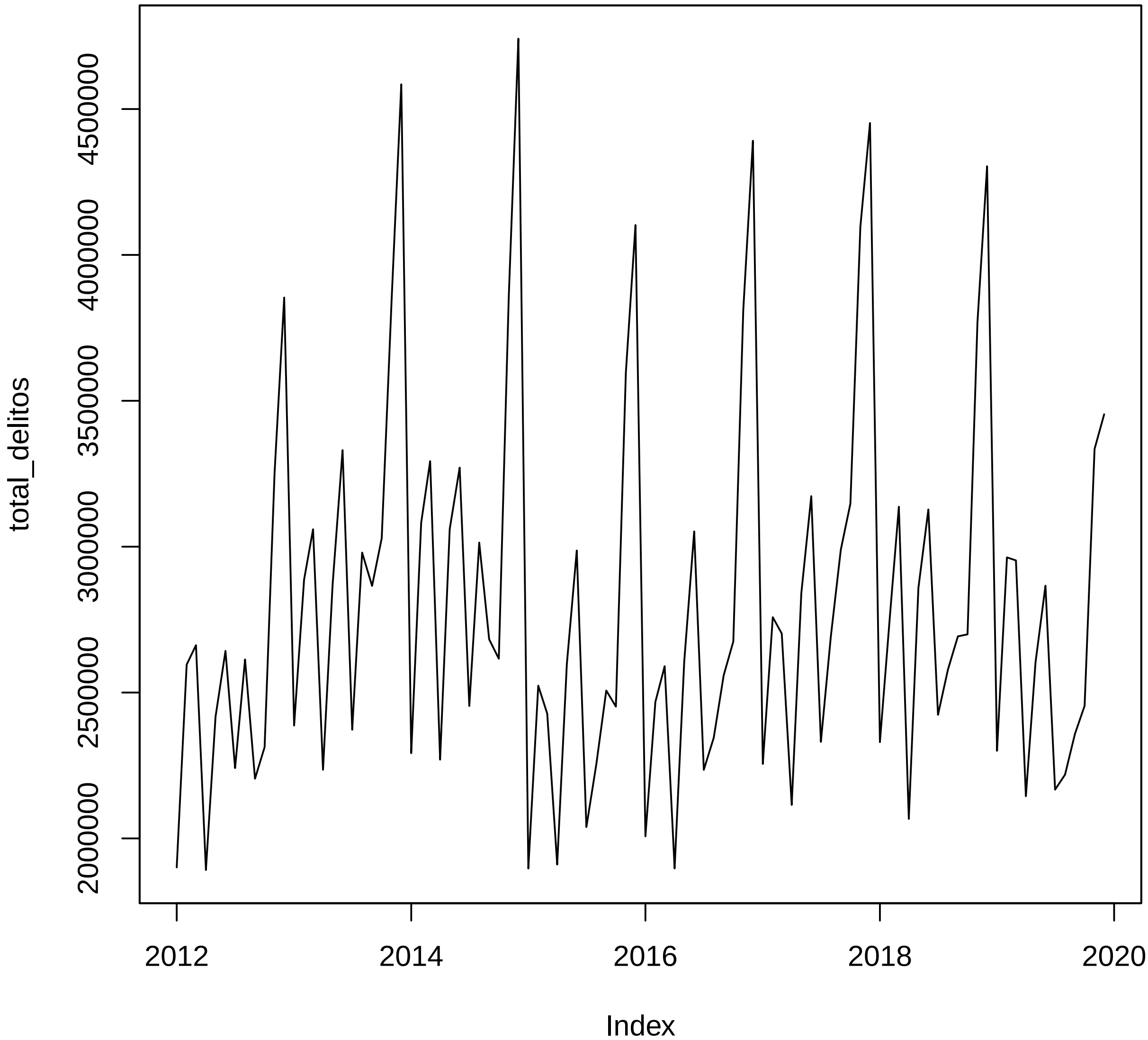

```r
adf.test(muertes_covid) #La serie NO es estacionaria.

##
##  Augmented Dickey-Fuller Test
##
## data:  muertes_covid
## Dickey-Fuller = -0.38009, Lag order = 7, p-value = 0.987
## alternative hypothesis: stationary
```

```r
plot(caminata_aleatoria)
```




```r
adf.test(caminata_aleatoria) #La caminta aletoria NO es estacionaria

##
##  Augmented Dickey-Fuller Test
##
## data:  caminata_aleatoria
## Dickey-Fuller = -1.9664, Lag order = 7, p-value = 0.5911
## alternative hypothesis: stationary

#Ya que el p-valor es un valor alto
```

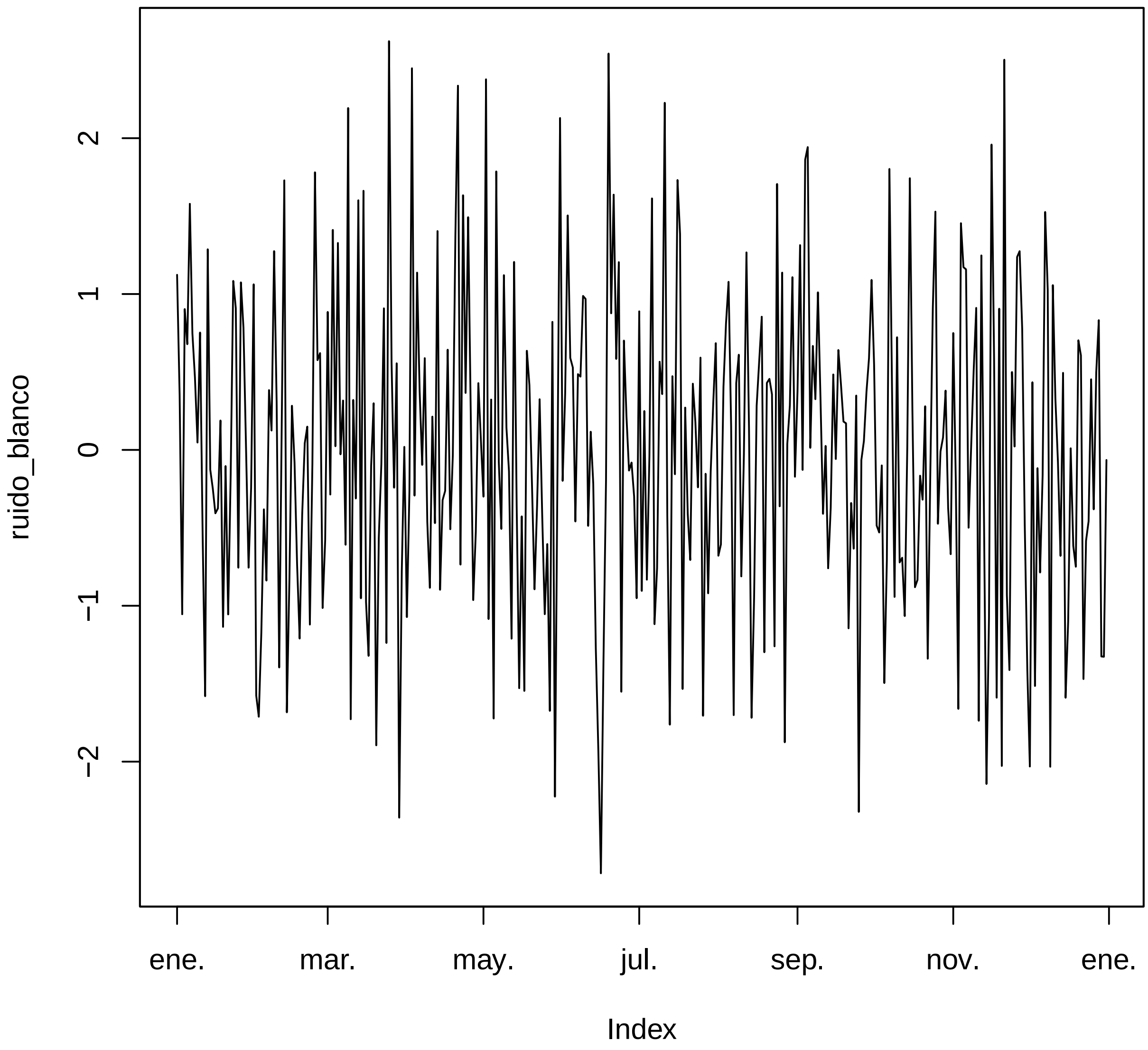
```r
plot(total_delitos)
```




```
adf.test(total_delitos)  #El total de delitos es estacionario

## Warning in adf.test(total_delitos): p-value smaller than printed p-value

##
##  Augmented Dickey-Fuller Test
##
## data:  total_delitos
## Dickey-Fuller = -4.694, Lag order = 4, p-value = 0.01
## alternative hypothesis: stationary

#En forma de como lo haría python sería
adf.test(total_delitos, k = ceiling(12*(length(total_delitos)/100)^(1/4))) #No es estacionaria

##
##  Augmented Dickey-Fuller Test
##
## data:  total_delitos
## Dickey-Fuller = -1.582, Lag order = 12, p-value = 0.7492
## alternative hypothesis: stationary
```

```
#k = ceiling(12*(length(total_delitos)/100)^(1/4)) son los retrasos de la serie, con esto lo hac
```

```
plot(ruido_blanco)
```




```
adf.test(ruido_blanco) #El ruido blanco es estacionario

## Warning in adf.test(ruido_blanco): p-value smaller than printed p-value

##
##  Augmented Dickey-Fuller Test
##
## data:  ruido_blanco
## Dickey-Fuller = -7.7765, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```

1.2. Estacionalidad

La **estacionalidad** sugiere que ciertas tendencias aparecerán de forma cíclica. Por ejemplo, las temperaturas suben y bajan según las horas del día y los meses del año. Estos son dos ejemplos distintos de patrones estacionales que observamos en nuestra vida, el cambio durante el día y el cambio durante un año completo.
Hay varias formas de comprobar si existe estacionalidad. Un enfoque es descomponer la secuencia de la serie temporal en tres efectos.

1. **Tendencia.**Representa el patrón consistente en todos los datos.
2. **Estacional.** Expresa todos los efectos cíclicos debido a la estacionalidad.
3. **Residual.** Son el error de predicción, o la diferencia entre los datos reales y el modelo que ajustamos.

El tipo más simple de descomposición es la descomposición clásica. Con ella esperamos una relación lineal entre las tres partes y la serie de tiempo observada. Hay dos enfoques principales en la descomposición clásica.

1. **Aditivo:** La realidad es el resultado de sumar la tendencia con la estacionalidad y con los residuos.

$$X_t = Ten_t + Cic_t + \epsilon_t$$

2. **Multiplicativo:** La realidad es el resultado de multiplicar la tendencia con la estacionalidad y con los residuos.

$$X_t = Ten_t \cdot Cic_t \cdot \epsilon_t$$

1.3. Autocorrelación

La autocorrelación representa la correlación entre una secuencia y consigo misma. Más precisamente, mide el nivel de semejanza entre una secuencia de hace varios períodos y datos actuales.
La secuencia de hace varios períodos atrás se llama ”retraso” o ”lag” en inglés, porque es una versión retrasada de la original. Por ejemplo, si calculamos la autocorrelación para una serie de tiempo con frecuencia diaria, estamos determinando cuántos de los valores de ayer se parecen a los valores de hoy. Si la frecuencia es, en cambio, anual, entonces la autocorrelación medirá las similitudes de año en año.

Función de auto-correlación.

La función de autocorrelación proporciona la autocorrelación para cualquier retraso que consideremos. La alternativa sería encontrar manualmente la correlación entre nuestros datos originales y múltiples retrasos de si misma.
En R usaremos la función acf

Auto-correlación parcial.

La autocorrelación mide la similitud entre una serie temporal y una versiones anteriores de sí misma. Sin embargo, los coeficientes también capturan efectos de los momentos anteriores de manera indirecta. Por indirecta nos referimos a todos los demás canales a través de los cuales los datos del pasado afectan a los datos actuales. Si deseamos determinar solo la relación directa entre la serie de tiempo y su versión retrasada, necesitamos calcular la autocorrelación parcial.