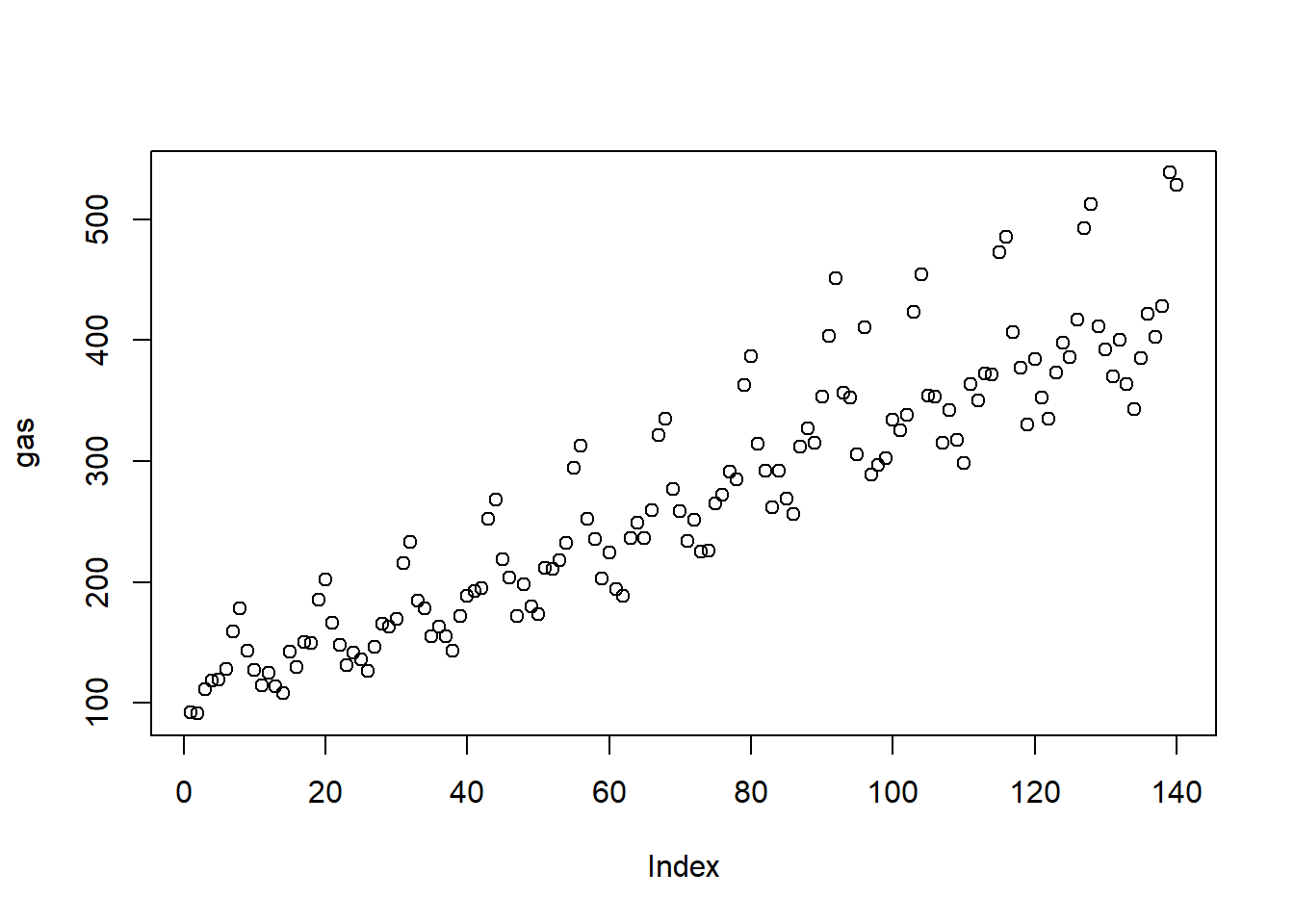
# Análisis básico de series temporales con R

## **Definición de una serie temporal**

En esta primera parte vamos a trabajar con el fichero gas6677.dat que contiene datos mensuales del consumo de gasolina en España entre enero de 1966 y agosto de 1977. Una vez situado el fichero en el directorio de trabajo lo leemos y creamos un vector llamado gas con el comando

gas = scan('http://verso.mat.uam.es/~joser.berrendero/datos/gas6677.dat')

plot(gas)



Vemos que el gráfico resultante no es el más apropiado para describir una serie temporal. Si queremos que **R** trate a un objeto como serie temporal, tenemos que determinar apropiadamente sus características con el comando ts. Para definir la serie correctamente escribimos:

gas.ts = ts(gas, start = c(1966,1), frequency = 12)

El argumento frequency se utiliza para indicar la periodicidad de la serie (en este caso mensual), mientras que el argumento start indica la fecha de la primera observación (enero de 1966).

print(gas.ts)

## Jan Feb Mar Apr May Jun Jul Aug

## 1966 92.718 91.380 111.643 118.888 119.432 127.796 158.943 178.013

## 1967 113.661 108.224 142.256 129.835 150.735 149.554 185.792 201.758

## 1968 135.951 126.615 146.647 165.822 163.365 169.294 215.538 233.427

## 1969 154.844 143.552 171.573 188.322 192.756 195.296 252.288 268.379

## 1970 179.759 173.821 211.387 210.551 218.371 232.057 294.173 312.700

## 1971 193.916 188.375 236.187 249.037 235.957 258.980 321.085 334.562

## 1972 225.010 225.742 265.159 271.986 290.953 285.108 362.687 386.347

## 1973 268.578 256.063 312.041 326.741 315.157 353.016 403.662 451.098

## 1974 289.186 296.881 302.589 334.091 325.790 337.782 423.297 454.172

## 1975 317.760 298.188 363.429 350.203 372.149 371.877 472.458 485.517

## 1976 352.200 334.938 372.891 397.388 385.657 416.961 492.480 512.209

## 1977 363.367 342.979 384.936 421.718 402.877 427.615 538.254 528.007

## Sep Oct Nov Dec

## 1966 143.385 127.179 114.403 124.900

## 1967 166.565 148.048 131.581 141.315

## 1968 184.402 178.432 155.179 163.355

## 1969 218.810 203.545 172.148 198.381

## 1970 251.891 235.560 202.876 224.383

## 1971 276.932 258.269 233.532 251.755

## 1972 314.205 292.124 261.740 291.810

## 1973 356.811 352.566 305.580 410.614

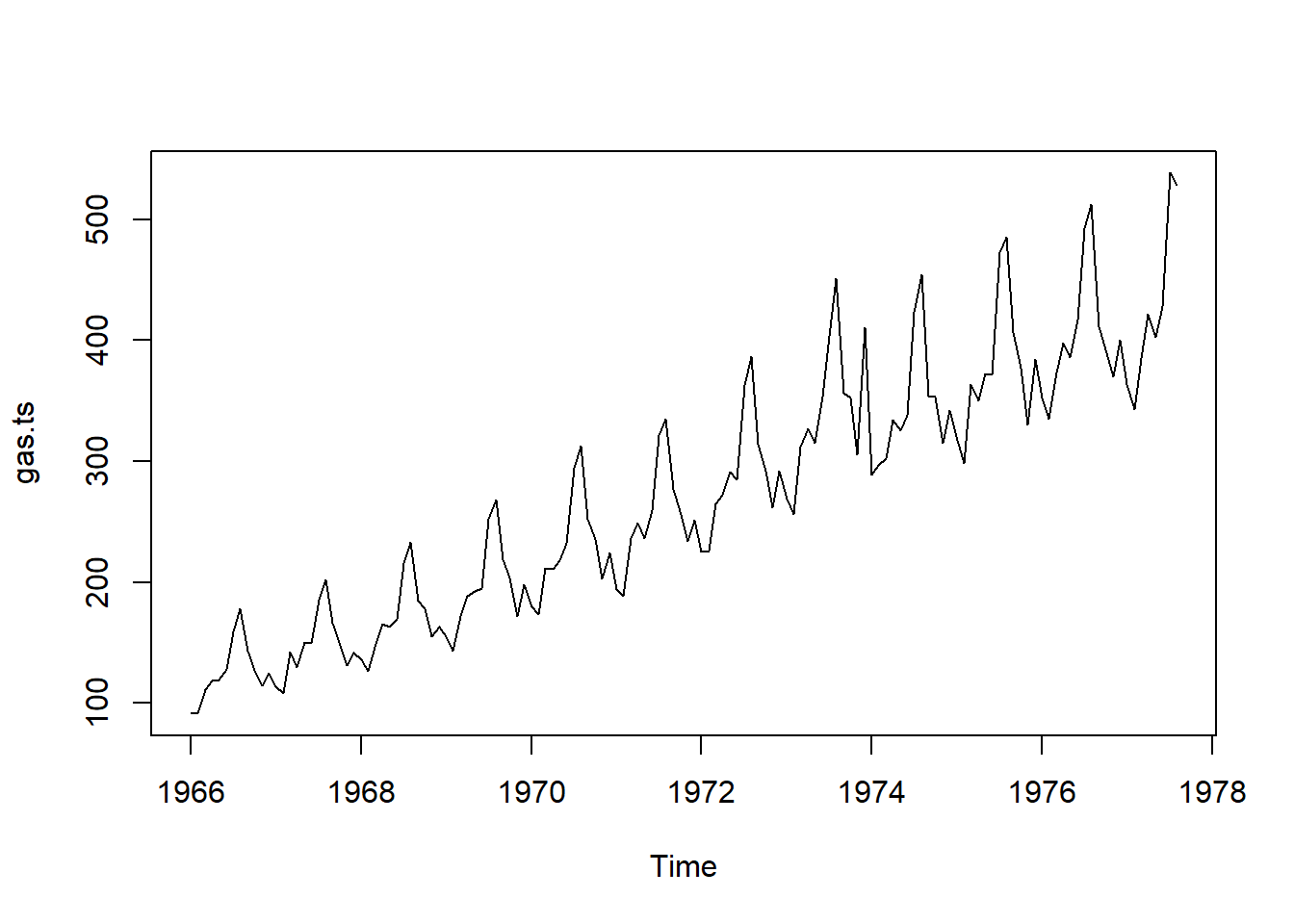
## 1974 353.727 353.413 315.272 341.902

## 1975 406.223 377.262 329.794 384.350

## 1976 411.514 392.380 369.671 400.243

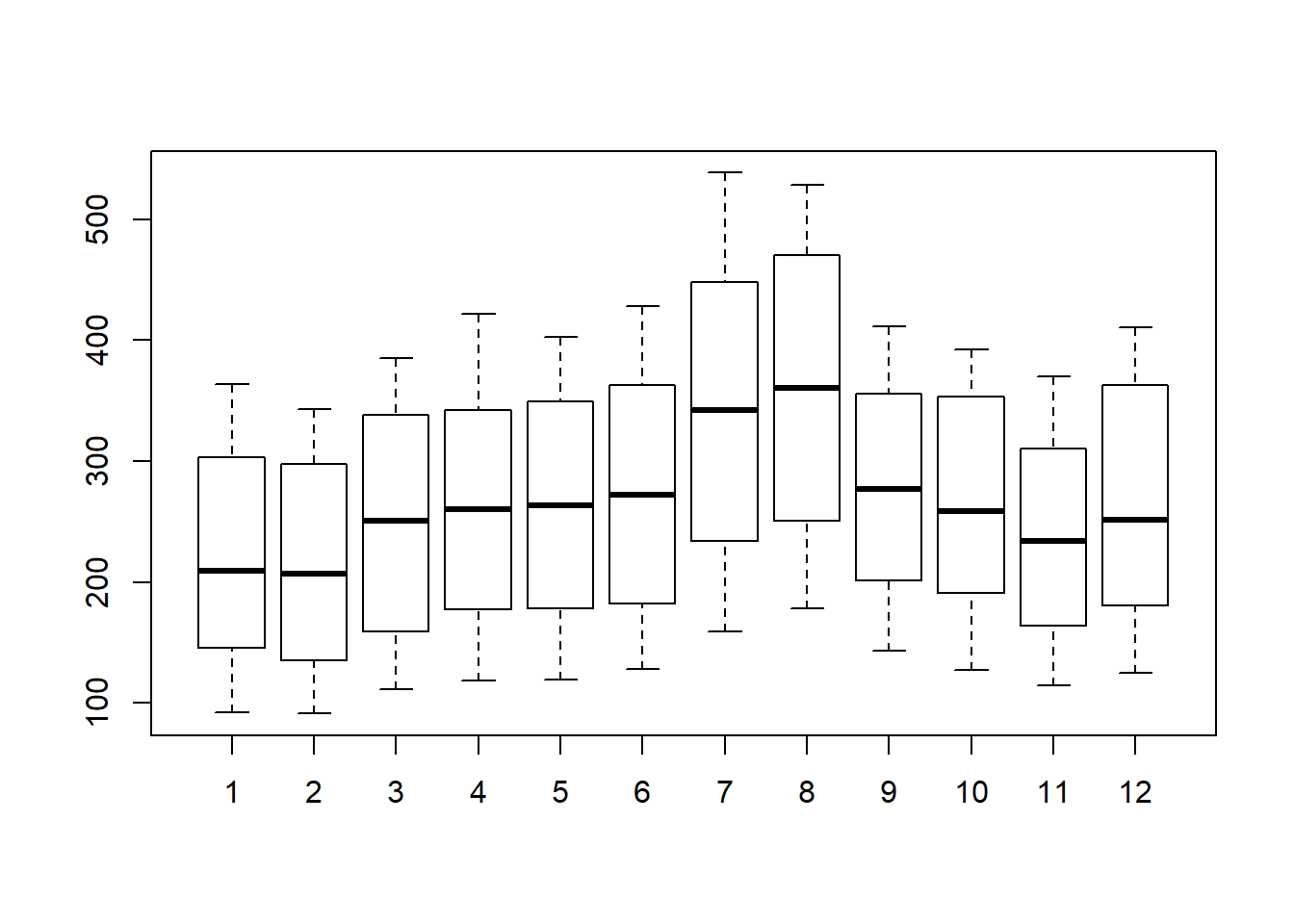
## 1977

plot(gas.ts)



Vemos ahora cómo el resultado depende de las características que hemos definido para la serie. Si queremos comparar la distribución del consumo de gasolina para cada mes, un gráfico útil es

boxplot(gas.ts ~ cycle(gas.ts))



El comando cycle determina la unidad de tiempo a la que pertenece cada observación de la serie:

cycle(gas.ts)

## Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

## 1966 1 2 3 4 5 6 7 8 9 10 11 12

## 1967 1 2 3 4 5 6 7 8 9 10 11 12

## 1968 1 2 3 4 5 6 7 8 9 10 11 12

## 1969 1 2 3 4 5 6 7 8 9 10 11 12

## 1970 1 2 3 4 5 6 7 8 9 10 11 12

## 1971 1 2 3 4 5 6 7 8 9 10 11 12

## 1972 1 2 3 4 5 6 7 8 9 10 11 12

## 1973 1 2 3 4 5 6 7 8 9 10 11 12

## 1974 1 2 3 4 5 6 7 8 9 10 11 12

## 1975 1 2 3 4 5 6 7 8 9 10 11 12

## 1976 1 2 3 4 5 6 7 8 9 10 11 12

## 1977 1 2 3 4 5 6 7 8

Por lo tanto, en el gráfico anterior se ha producido un diagrama de cajas para cada mes del año.

**Ejercicio:** El fichero jj.dat contiene los beneficios trimestrales de la empresa Johnson & Johnson entre 1960 y 1980:

jj = scan('http://verso.mat.uam.es/~joser.berrendero/datos/jj.dat')

Define la serie temporal y represéntala. ¿Cuál es el valor de la serie para el tercer trimestre de 1980? ¿Cuáles son las principales características (tendencia, estacionalidad) de esta serie?

## **Descomposición de una serie**

Es frecuente analizar las series temporales desde el punto de vista de sus componentes estructurales:

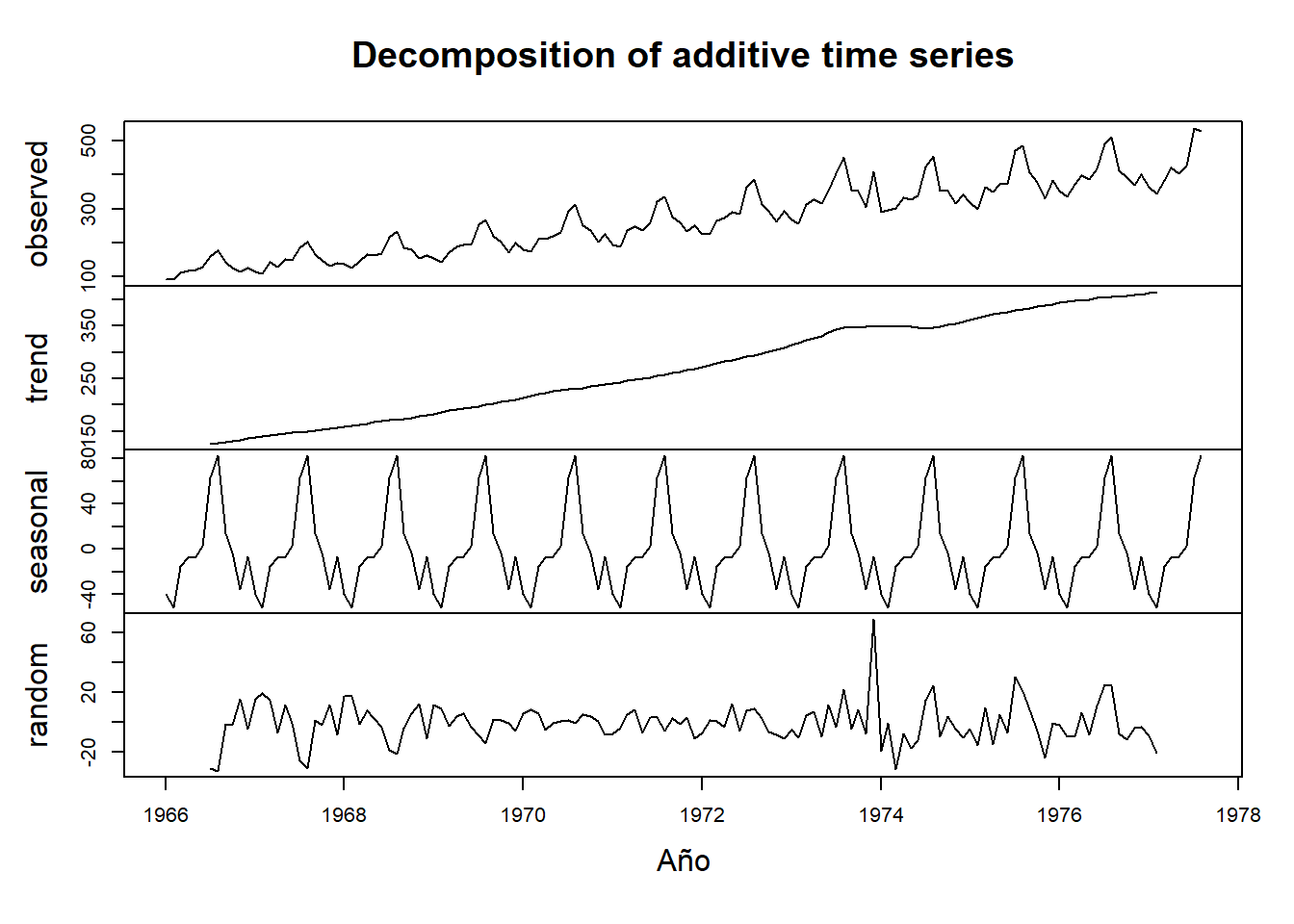
Serie observada = Tendencia + Efecto estacional + Residuos.

En este modelo, la serie observada es el resultado de sumar una tendencia que representa el comportamiento a largo plazo de la serie, un efecto estacional que describe sus fluctuaciones periódicas y un componente residual que describe las variaciones a corto plazo, normalmente impredecibles.

Con **R** es muy sencillo obtener una descomposición estructural de este tipo. Se usa el comando decompose:

gas.ts.desc = decompose(gas.ts)

plot(gas.ts.desc, xlab='Año')



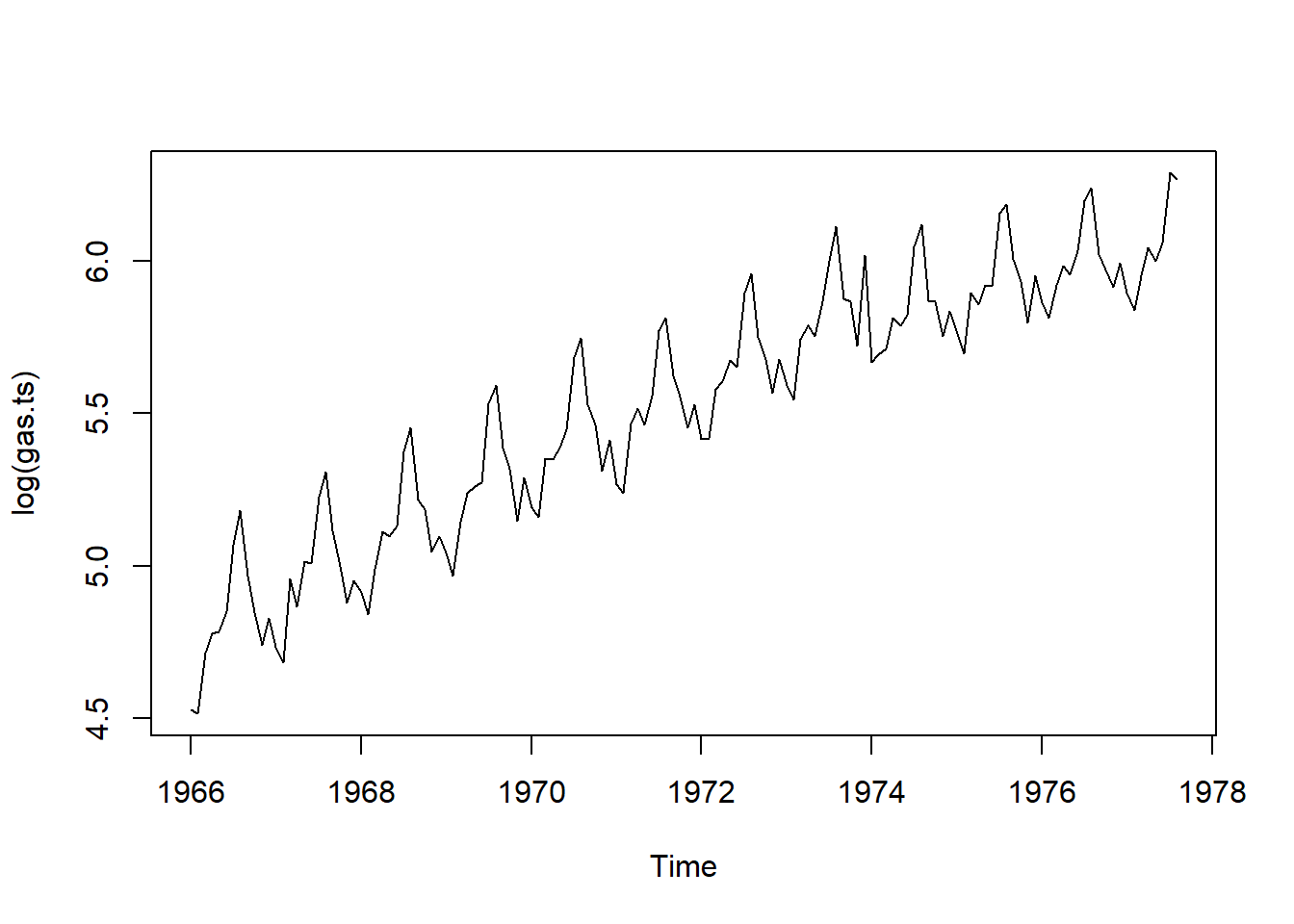
Esta descomposición se basa en métodos elementales: la tendencia se calcula con una media móvil, el efecto estacional se calcula promediando los valores de cada unidad de tiempo para todos los periodos (por ejemplo, todos los meses de enero si la serie es mensual) y luego centrando el resultado. Finalmente, los residuos se obtienen restando a la serie observada las dos componentes anteriores. La descomposicion solo es totalmente adecuada si se dispone de un número completo de periodos (por ejemplo, un múltiplo de 12 si la serie es mensual).

## **Transformaciones básicas de una serie**

En el gráfico de gas.ts se observa que la serie no es estacionaria. La serie presenta una tendencia aparentemente lineal y una estacionalidad muy marcada (el consumo aumenta los meses de verano). Además, la amplitud de las fluctuaciones aumenta con el tiempo por lo que la variabilidad tampoco es constante. Sin embargo, muchos modelos importantes de series temporales corresponden a series estacionarias (es decir, sin tendencia ni estacionalidad y con variabilidad constante). Antes de ajustar un modelo estacionario tenemos que transformar la serie original.

**Estabilización de la varianza:** Para estabilizar la variabilidad se suelen tomar logaritmos. Esta transformación funcionará bien cuando la variabilidad sea aproximadamente proporcional al nivel de la serie. Representamos la serie transformada mediante

plot(log(gas.ts))

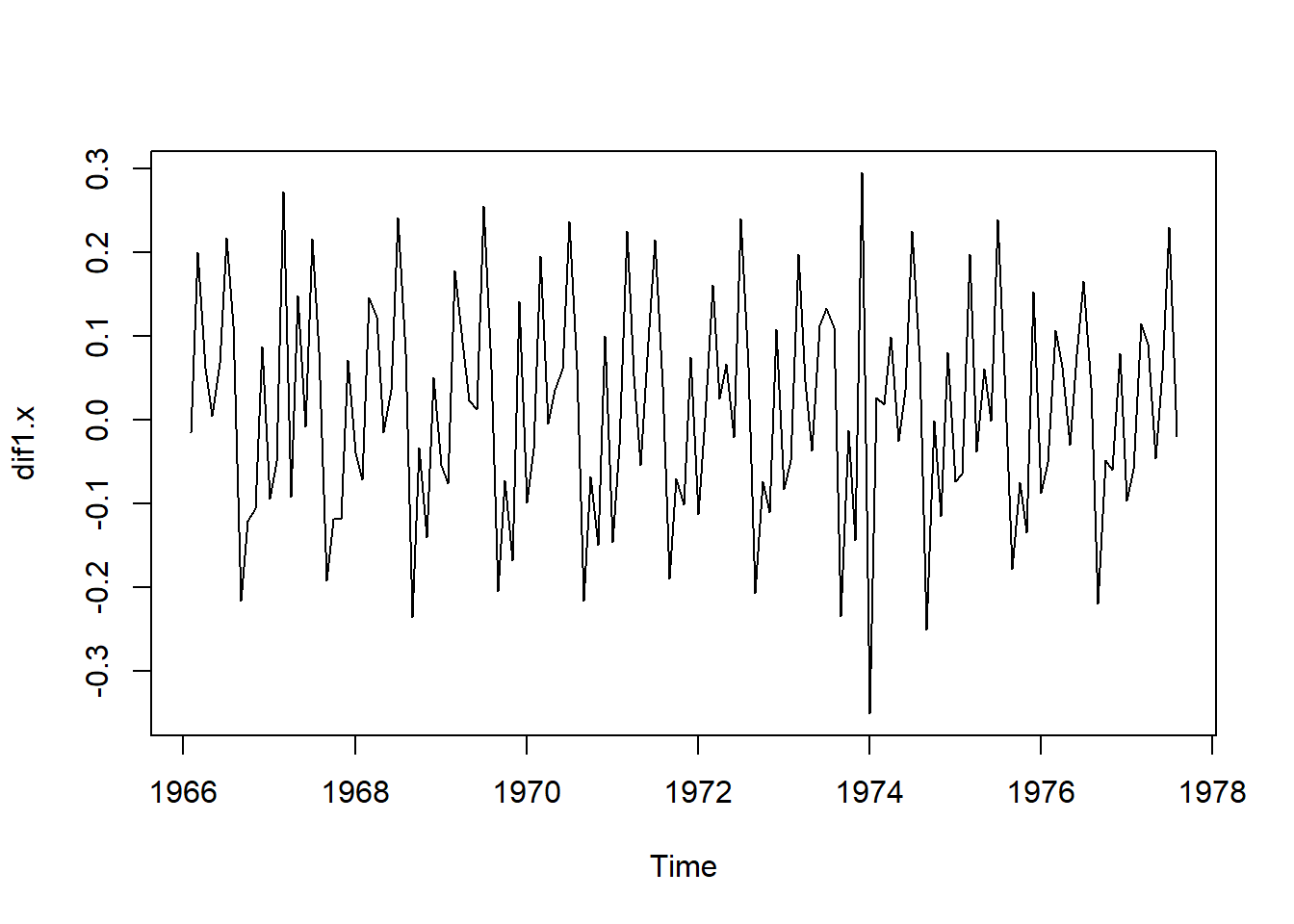


**Eliminación de tendencia:** Una forma sencilla de eliminar una tendencia aproximadamente lineal es diferenciar la serie, es decir, considerar la serie de diferencias entre una observación y la anterior en lugar de la serie original. Si *x**t* es una serie contenida en x, para calcular ∇*x**t*=*x**t*−*x**t*−1 con **R** se escribe:

x = log(gas.ts)

dif1.x = diff(x)

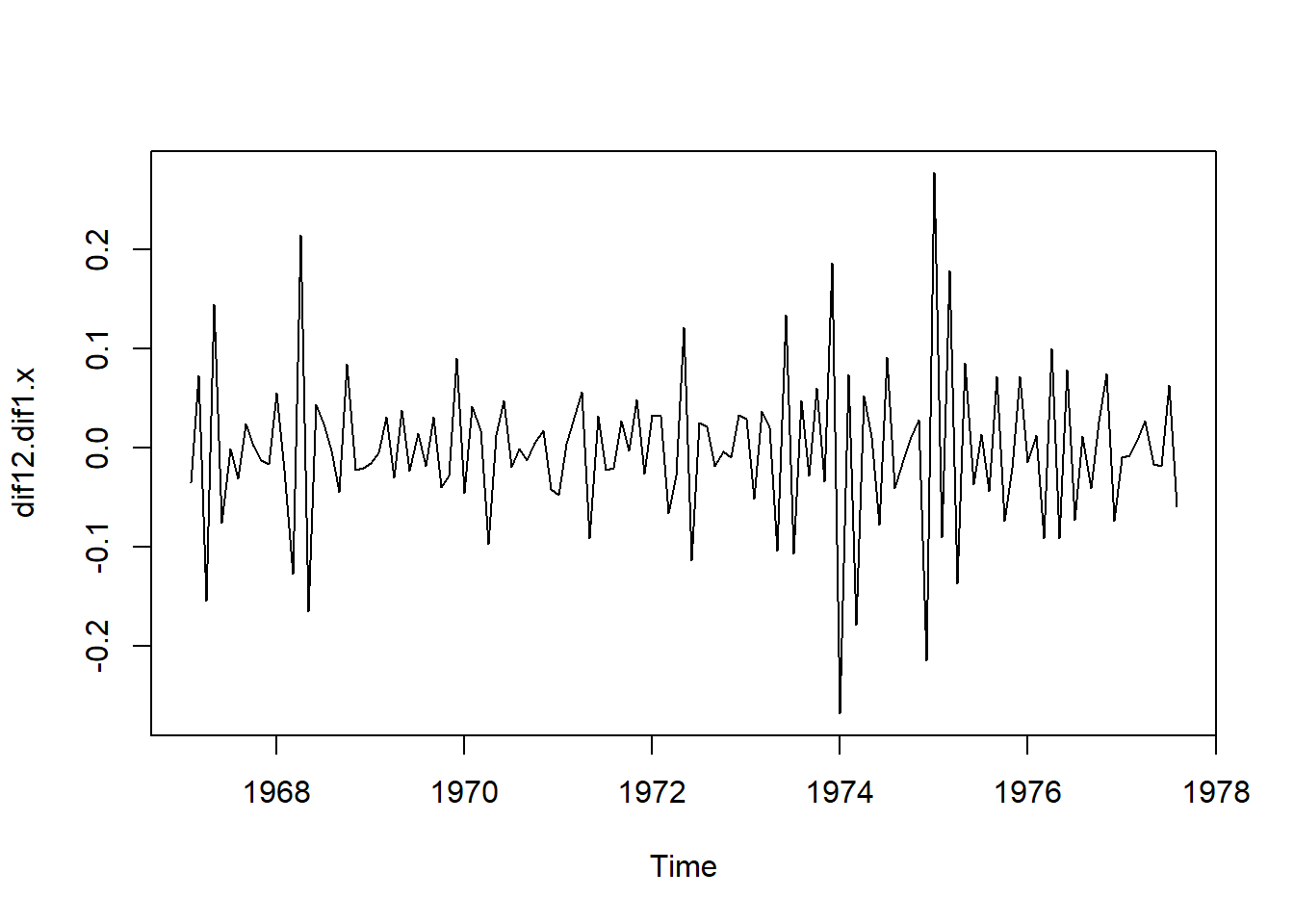
plot(dif1.x)



**Eliminación de estacionalidad:** Para eliminar la estacionalidad de una serie mensual se pueden tomar diferencias estacionales de orden 12. Si *x**t* es la serie que queremos desestacionalizar, se trata de calcular ∇12*x**t*=*x**t*−*x**t*−12:

dif12.dif1.x = diff(dif1.x, lag=12)

plot(dif12.dif1.x)



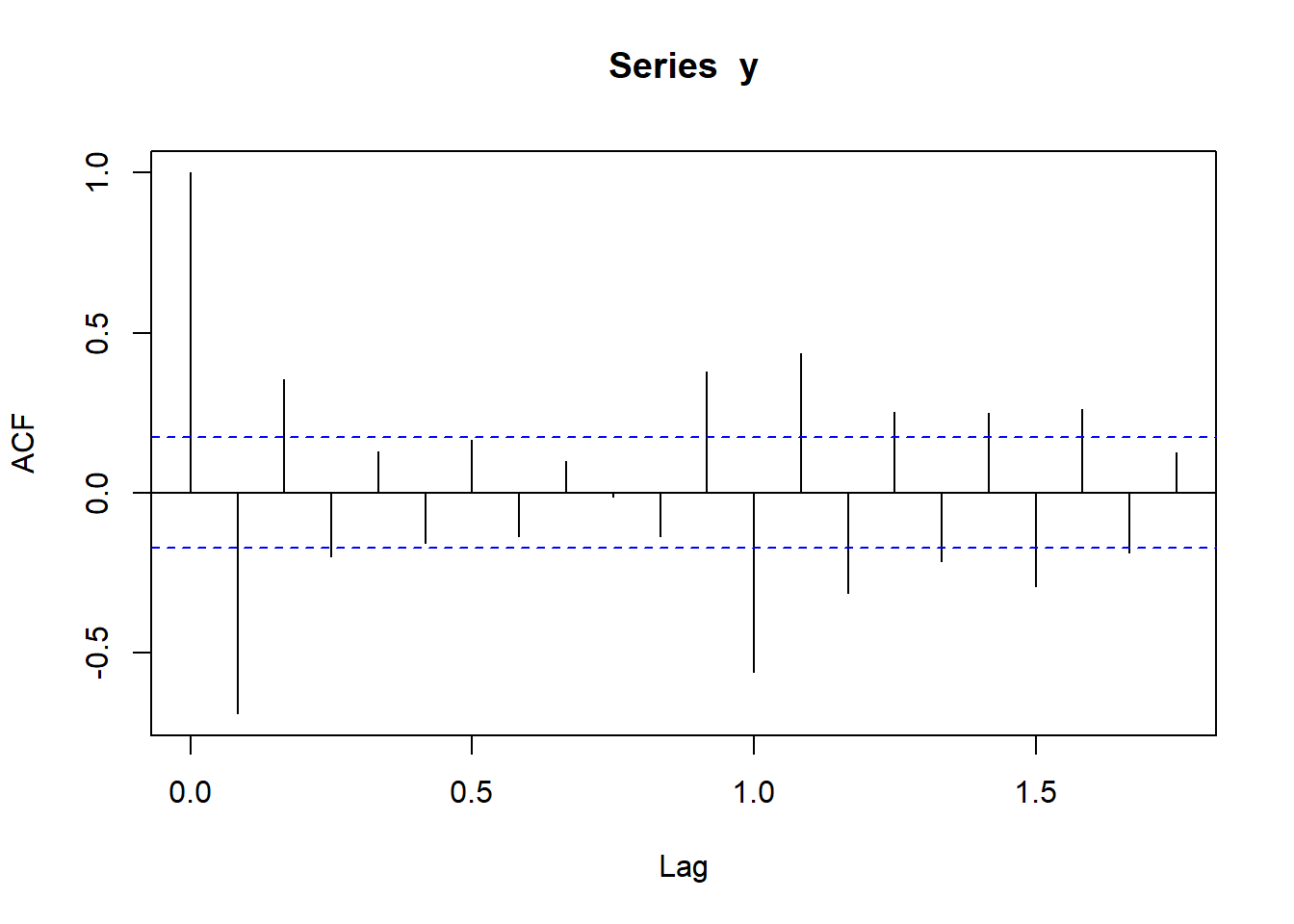
**Ejercicio:** Aplica las transformaciones adecuadas a la serie de beneficios de la empresa Johnson & Johnson para obtener una serie aproximadamente estacionaria.

## **Las funciones de autocovarianzas y de autocorrelaciones**

Transformamos la serie del consumo de gasolina de manera que un modelo estacionario sea apropiado para la serie transformada. El siguiente código se puede utilizar para representar el correlograma de la serie. El correlograma es una representación gráfica de las autocorrelaciones *ρ*(*k*), es decir, las correlaciones entre *x**t* y *x**t*+*k*, en función de *k*:

y = dif12.dif1.x

acf(y)



Siempre se tiene que *ρ*(0)=1. Las líneas discontinuas representan las bandas de confianza de *ρ*(*k*) de nivel 95% bajo la hipótesis de que la serie es un ruido blanco (incorrelada). En el ejemplo las autocorrelaciones más significativas son las correlaciones entre la observación de un mes y la del mes siguiente, y la observación de un mes con la del mismo mes del año siguiente.

Si lo que queremos obtener son los valores numéricos de las correlaciones estimadas escribimos

acf(y, plot=FALSE)$acf

## , , 1

##

## [,1]

## [1,] 1.00000000

## [2,] -0.69060340

## [3,] 0.35402186

## [4,] -0.19960669

## [5,] 0.12938794

## [6,] -0.15876250

## [7,] 0.16437320

## [8,] -0.13672993

## [9,] 0.09869406

## [10,] -0.01149061

## [11,] -0.13667518

## [12,] 0.37817656

## [13,] -0.56158980

## [14,] 0.43614496

## [15,] -0.31315303

## [16,] 0.25193117

## [17,] -0.21379059

## [18,] 0.24804411

## [19,] -0.29321039

## [20,] 0.26216941

## [21,] -0.18885352

## [22,] 0.12540136

**Ejercicio:** Representa el correlograma de la serie de beneficios de Johnson & Johnson transformada que se ha obtenido en el ejercicio anterior. ¿Podemos decir que la serie transformada es un ruido blanco?

**Ejercicio:** Escribe una función de **R** que tenga como argumento una serie temporal y como salida la estimación de la autocorrelación de primer orden *ρ*^(1).