

Lesson 1

Emanuel Mejía

Contents

El número de reservaciones de pasajeros internacionales (en miles) por mes en una aerolínea (Pan Am) en Estados Unidos fue obtenido de la Administración Federal de Aviación para el periodo 1949–1960 (Brown, 1963). La empresa utilizó estos datos para predecir la demanda futura antes de ordenar nuevas aeronaves y capacitar a la tripulación. Los datos están disponibles como una serie temporal en R y ejemplifican varios conceptos importantes que surgen en un análisis exploratorio de series temporales.

```
##      Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
## 1949 112 118 132 129 121 135 148 148 136 119 104 118
## 1950 115 126 141 135 125 149 170 170 158 133 114 140
## 1951 145 150 178 163 172 178 199 199 184 162 146 166
## 1952 171 180 193 181 183 218 230 242 209 191 172 194
## 1953 196 196 236 235 229 243 264 272 237 211 180 201
## 1954 204 188 235 227 234 264 302 293 259 229 203 229
## 1955 242 233 267 269 270 315 364 347 312 274 237 278
## 1956 284 277 317 313 318 374 413 405 355 306 271 306
## 1957 315 301 356 348 355 422 465 467 404 347 305 336
## 1958 340 318 362 348 363 435 491 505 404 359 310 337
## 1959 360 342 406 396 420 472 548 559 463 407 362 405
## 1960 417 391 419 461 472 535 622 606 508 461 390 432
```

```
## [1] "ts"
```

En este caso, el objeto es de la clase *ts*, que es una abreviatura de “time series”. Estos objetos cuentan con varios métodos disponibles, que incluyen las funciones *start*, *end* y *frequency*.

```
start(pas)
```

```
## [1] 1949    1
```

```
end(pas)
```

```
## [1] 1960   12
```

```
frequency(pas)
```

```
## [1] 12
```

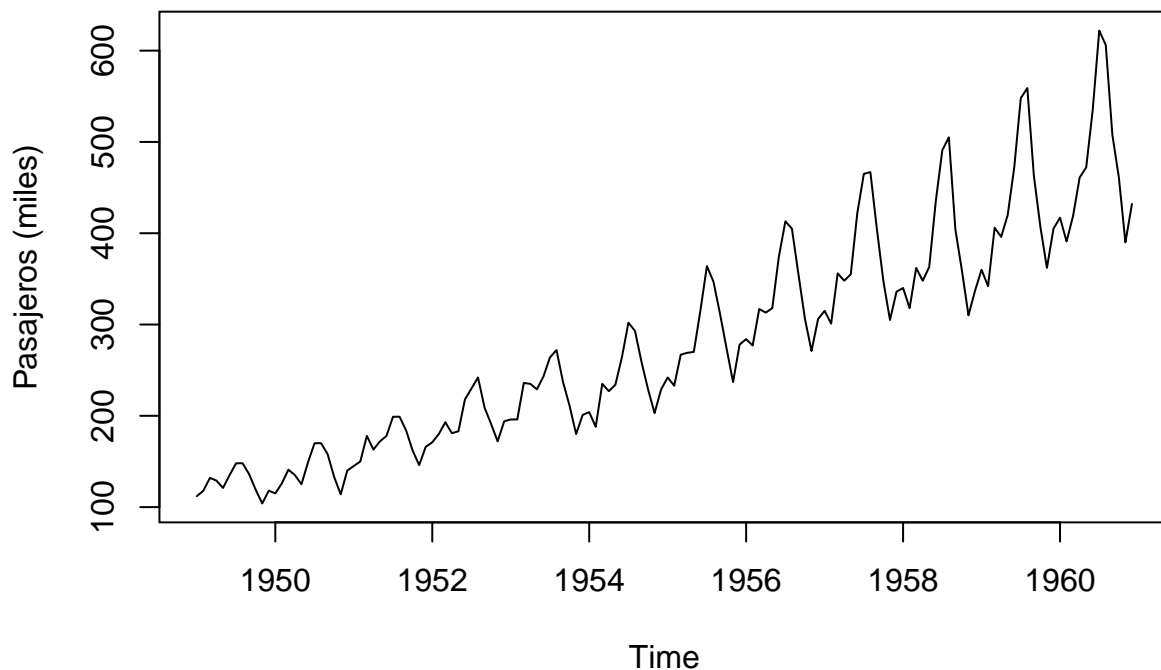
Funciones genéricas en *R*, como *plot* o *summary*, intentarán ofrecer la salida más apropiada para cualquier objeto de entrada.

```
summary(pas)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    104.0   180.0   265.5   280.3   360.5   622.0
```

Uno de los pasos más importantes en un análisis preliminar de series temporales es graficar los datos; es decir, crear un gráfico temporal. Para un objeto de clase `ts`, esto se logra con la función genérica `plot`:

```
plot(pas, ylab = 'Pasajeros (miles)')
```



Hay varias características que son comunes a muchas series temporales. Por ejemplo, puede notarse que el número de pasajeros que viajan en la aerolínea está aumentando con el tiempo. A este cambio sistemático en una serie temporal que no parece ser periódico se le conoce como *tendencia*.

Normalmente se asume que en ausencia de alguna perturbación en el sistema, una tendencia probablemente cambiará de forma relativamente lenta. Aunque suponer que las tendencias continuarán de forma lineal durante muchos años suele ser poco realista.

Un patrón que se repite dentro de cualquier periodo fijo (por ejemplo cada año) se le conoce como variación *estacional*. En la serie temporal de pasajeros aéreos es evidente la variación estacional. Las reservaciones alcanzaban su punto más alto durante los meses de verano y eran más bajas durante el invierno.

Comprender las posibles causas de las características que se observan en el gráfico nos ayuda a formular un modelo de serie temporal adecuado. En este caso, las posibles causas de la tendencia creciente incluyen el aumento en la prosperidad tras la Segunda Guerra Mundial, una mayor disponibilidad de aeronaves, vuelos más baratos debido a la competencia entre aerolíneas y el crecimiento de la población. La variación estacional coincide con los periodos vacacionales.

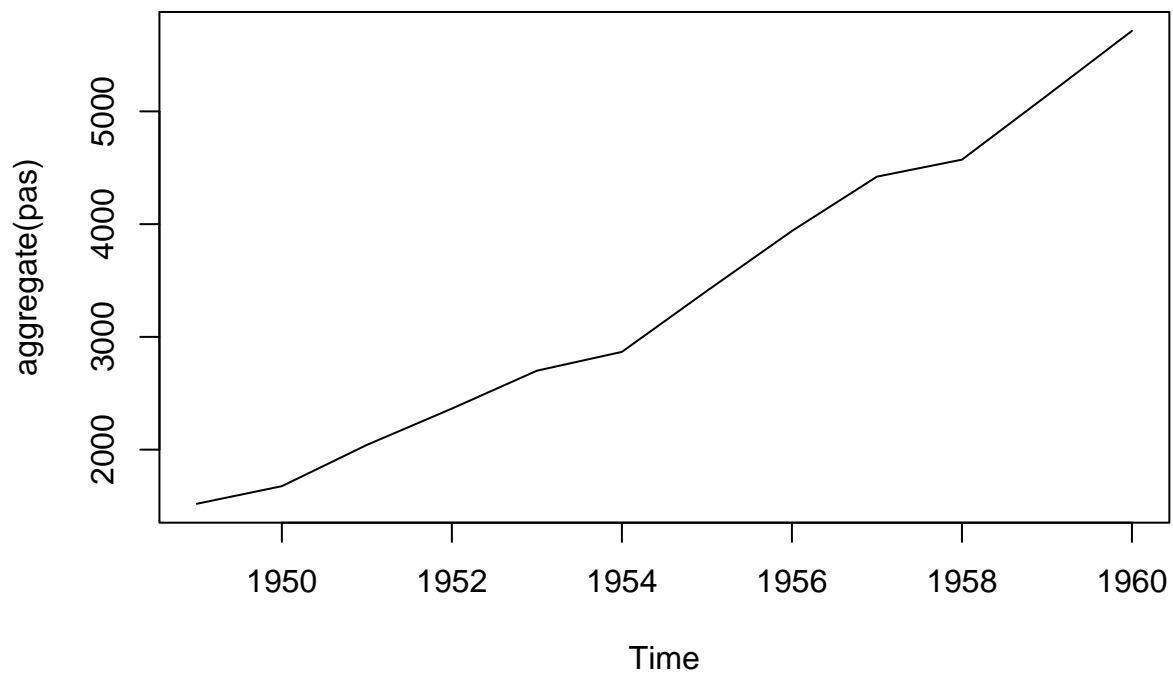
Un gráfico de series temporales no solo resalta los patrones y características de los datos, sino que también puede poner en evidencia valores atípicos (outliers) y errores.

Dichos errores deben tratarse de forma distinta durante el análisis y no deben incluirse como observaciones al ajustar un modelo a los datos. Los valores atípicos que no pueden atribuirse a errores de codificación deben revisarse cuidadosamente. Si son correctos es probable que sean de especial interés y no deben ser excluidos del análisis. Sin embargo, puede ser apropiado considerar métodos robustos para el ajuste de modelos, los cuales reducen la influencia de los valores atípicos.

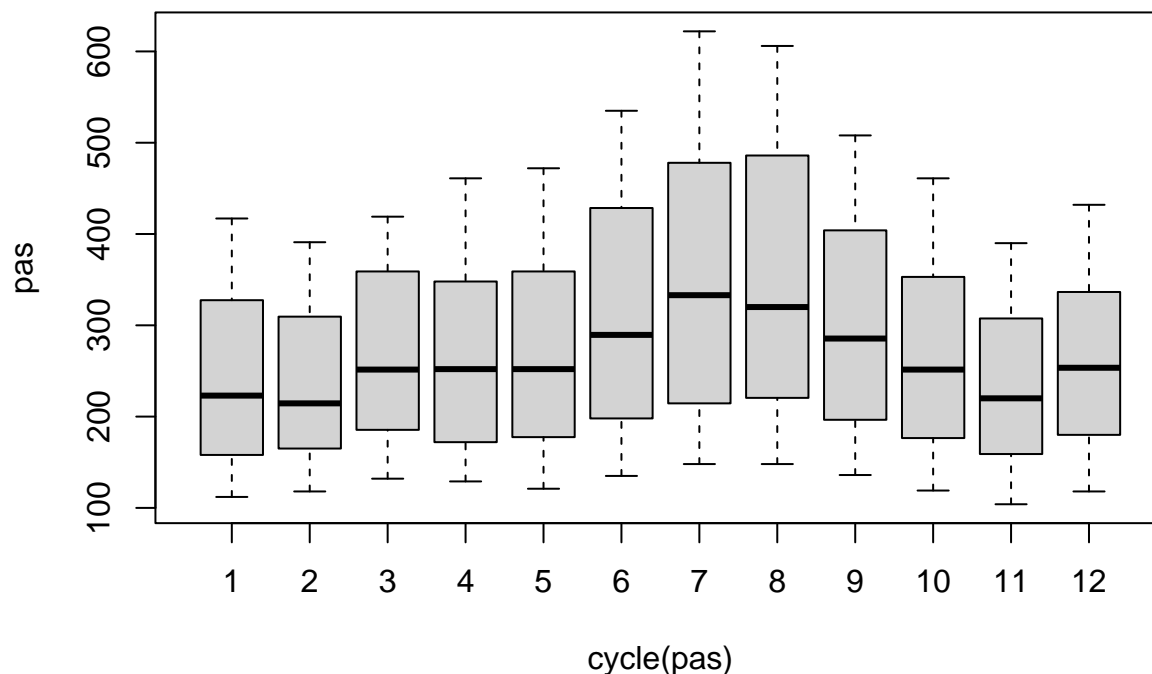
Para obtener una visión más clara de la tendencia, se puede eliminar el efecto estacional agregando los datos a nivel anual.

Mientras que un resumen de los valores para cada estación del año puede visualizarse mediante un diagrama de caja (boxplot), extrayendo la estación correspondiente a cada observación.

```
plot(aggregate(pas))
```



```
boxplot(pas~cycle(pas))
```



Las tendencias y patrones estacionales en el ejemplo anterior era evidente a partir de los gráficos. Además, se podían proponer explicaciones razonables para las posibles causas de estas características. En cambio, con datos financieros —por ejemplo, los tipos de cambio— es menos probable observar patrones tan marcados, y generalmente se requieren métodos de análisis diferentes.

Una serie financiera puede, en ocasiones, mostrar un cambio drástico con una causa clara, como una guerra o un desastre natural. Sin embargo, los cambios de un día a otro son más difíciles de explicar, ya que las causas subyacentes son complejas e imposibles de aislar. Por lo tanto, a menudo será poco realista suponer que exista algún componente determinista en el modelo de serie temporal.

Los tipos de cambio de la libra esterlina británica frente al dólar neozelandés para el periodo de enero de 1991 a marzo de 2000 se muestran en el siguiente ejemplo. Los datos son valores promedio tomados en periodos trimestrales de tres meses, siendo el primer trimestre de enero a marzo y el último de octubre a diciembre.

```
Z <- read.table("../data/pounds_nz.dat", header = T)
Z.ts <- ts(Z, st = 1991, fr = 4)

plot(Z.ts, xlab = "time/ years", ylab = "Quarterly exchange rate in $NZ/ pound")
```



En la serie temporal son evidentes tendencias de corto plazo: tras un aumento inicial que culmina en 1992, se observa una tendencia negativa que lleva a un mínimo alrededor de 1996, seguida por una tendencia positiva en la segunda mitad de la serie.

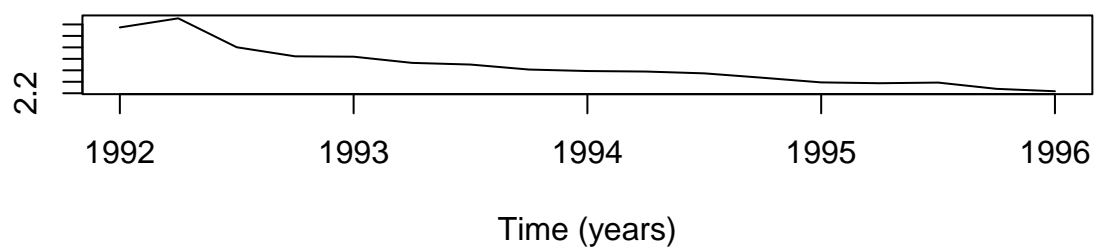
La tendencia parece cambiar de dirección en momentos impredecibles, en lugar de mostrar un patrón relativamente constante. A estas tendencias se les denomina tendencias estocásticas para enfatizar su aleatoriedad y distinguirlas de las tendencias más deterministas vistas en los ejemplos anteriores.

Dos tendencias locales se destacan cuando la serie se divide en dos subseries basadas en los periodos 1992–1996 y 1996–1998. La función `window` puede utilizarse para extraer estas subseries:

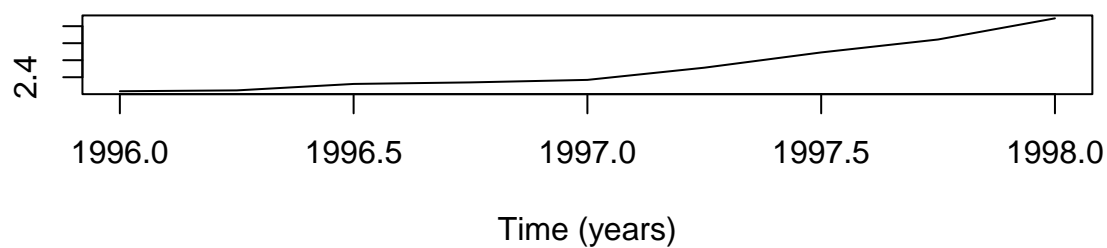
```
Z.92.96 <- window(Z.ts, start = c(1992, 1), end = c(1996, 1))
Z.96.98 <- window(Z.ts, start = c(1996, 1), end = c(1998, 1))

layout (1:2)
plot(Z.92.96, ylab = "Exchange rate in $NZ/pound", xlab = "Time (years)")
plot(Z.96.98, ylab = "Exchange rate in $NZ/pound", xlab = "Time (years)")
```

Exchange rate in \$NZ/pound

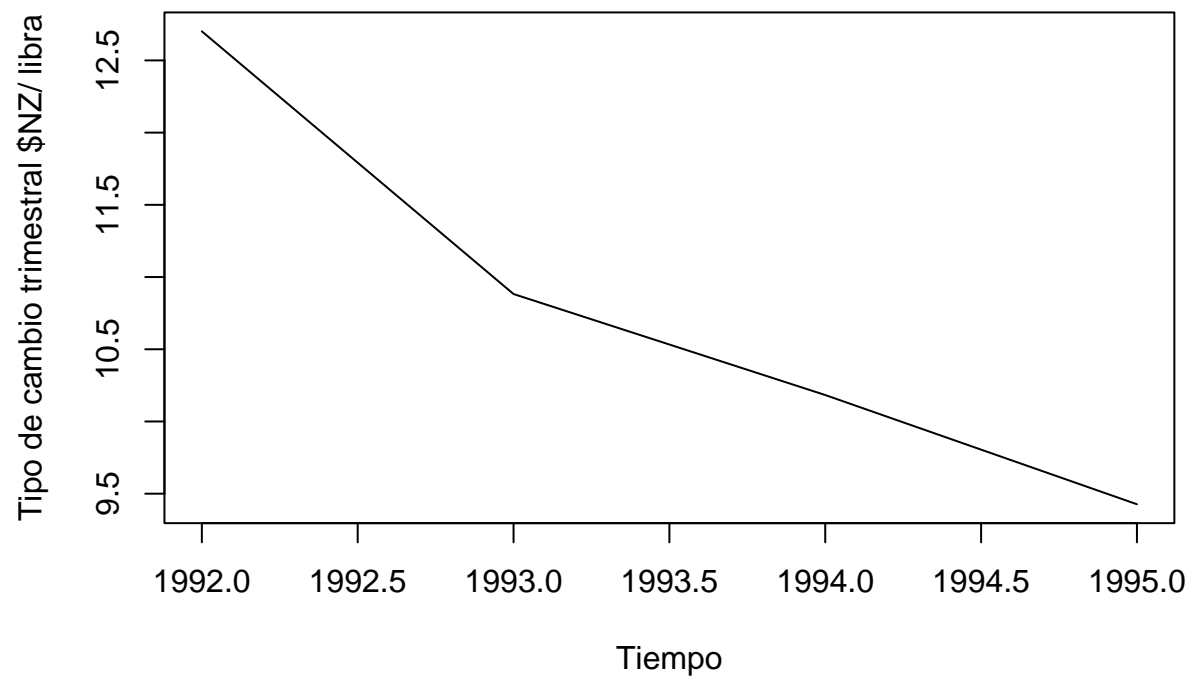


Exchange rate in \$NZ/pound

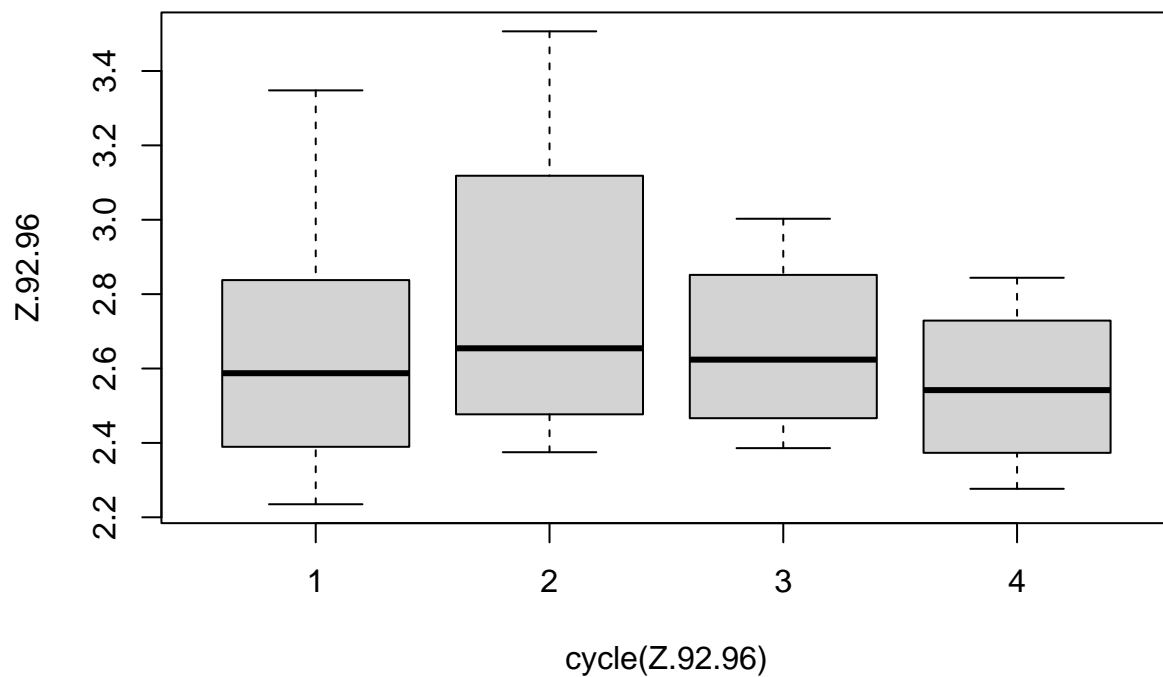


Al ser objetos ts a su vez podríamos utilizar los gráficos anteriores

```
plot(aggregate(Z.92.96),  
     ylab = 'Tipo de cambio trimestral $NZ/ libra',  
     xlab = 'Tiempo')
```



```
boxplot(Z.92.96~cycle(Z.92.96))
```



Imaginemos que nos encontramos al inicio de 1992.

Puede ser muy tentador predecir con base en la continuación de la tendencia descendente para los años siguientes.

Sin embargo, esta habría sido una muy mala predicción, muestra que los datos comenzaron a seguir una tendencia creciente.

Sin información adicional, no es recomendable extrapolar la tendencia mostrada.

Aunque veremos modelos que pueden reducir el riesgo y pruebas para evaluar si una tendencia es estocástica para reducir el riesgo de realizar predicciones incorrectas esto ilustra el peligro potencial de hacer extrapolaciones inapropiadas de tendencias estocásticas cuando no se comprendan adecuadamente las posibles causas subyacentes.