Time Series

Emerson Amirhosein Azarbakht

Time Series (TS)

Used in

- control of inventory, based on demand trends
- airline's decision to buy airplanes bc of passenger trends and decision to increase/maintain market share
- climate change decisions based on temperature change trends
- business/sales forecasting
- everyday operational decisions
- long-term effects of proposed water management policies by simulating daily rainfall and sea state time series
- understanding fluctuations in monthly sales
- basis for signal processing in telecommunications <?>
- disease incidence tracking, yearly rates
- census analysis
- tracking monthly unemployment rate; as an economic indicator used by decision makers

Used to

- to understand the past, and predict the future
- forcasting (predicting inference, a subset of statistical inference). assumes that present trends continue. This assumption cannot be checked empirically, but, when we identify the likely causes for a trend, we can justify the forecasting(extrapolating it) for a few time-steps at least
- anomaly detection
- clustering
- classification (assigning a time series pattern to a specific category: e.g. gesture recognition of hand movements in sign language videos)
- query by content ~ Content-based image retrieval

Data: a variable measured sequentially in time, or at a fixed [sampling] interval

serial dependence problem: observations close together in time tend to be correlated (serially dependent)

TS tries to explain this correlation (serial dependence) autocorrelation analysis examines this serial dependence <?>

conditions (assumptions of TS)

- stationary process ?
- Ergodic process ?

```
plot(AirPassengers)
start(AirPassengers)
```

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

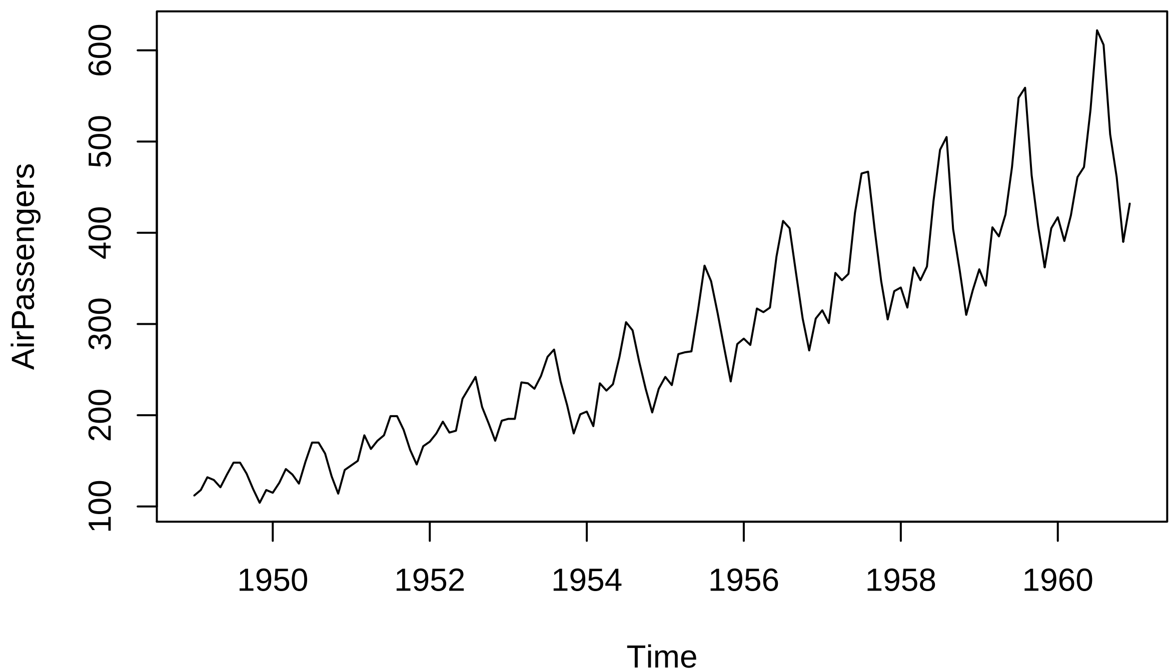
```r
end(AirPassengers)
```

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

```r
frequency(AirPassengers)
```

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

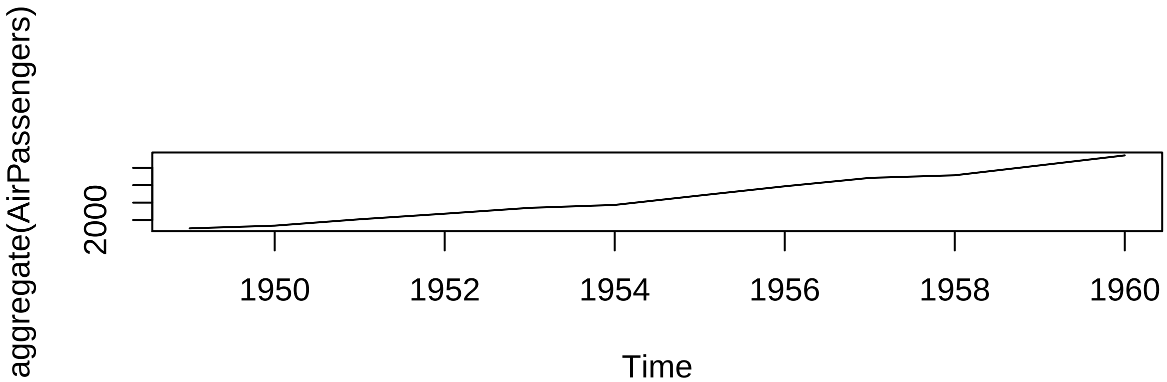
```r
plot(AirPassengers)
```

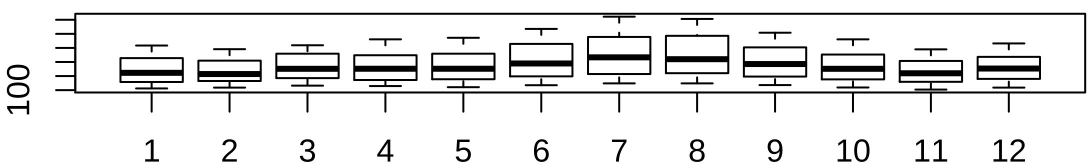



```r
summary(AirPassengers)
```

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

```r
layout(1:2)
# takes an input matrix for the location of each plot in the graphics window
plot(aggregate(AirPassengers))
boxplot(AirPassengers ~ cycle(AirPassengers))
```





plotting shows *patterns*, and *features* of the data + *outliers* and *erroneous* values

patterns

1. trend = a non-periodic systematic change in a TS
   - can be modeled simply by a linear increase or decrease. (only if it's non-stochastic)
   - stochastic trend: seems to change direction at unpredictable times rather than displaying a consistent pattern (e.g. like the air passenger series)
2. seasonal variation = a repeating pattern within a fixed period (e.g. each year)
3. cycles = a non-fixed-period cycle (without a fixed period). example: El-Nino

```
# monthly unemployment rate for the US state of Maine from January 1996 until August 2006
Maine.month <- read.table("http://staff.elena.aut.ac.nz/Paul-Cowpertwait/ts/Maine.dat", header = TRUE)
# header TRUE means treat first row as column names
attach(Maine.month)
str(Maine.month)
```

```
## 'data.frame':    128 obs. of  1 variable:
##  $ unemploy: num  6.7 6.7 6.4 5.9 5.2 4.8 4.8 4 4.2 4.4 ...
```

```
head(Maine.month)
```

```
##   unemploy
## 1      6.7
## 2      6.7
## 3      6.4
## 4      5.9
## 5      5.2
## 6      4.8
```

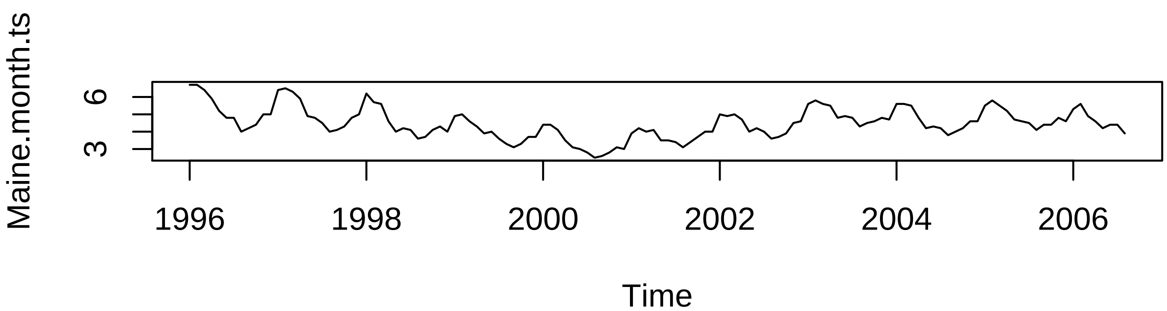
```r
class(Maine.month)
```

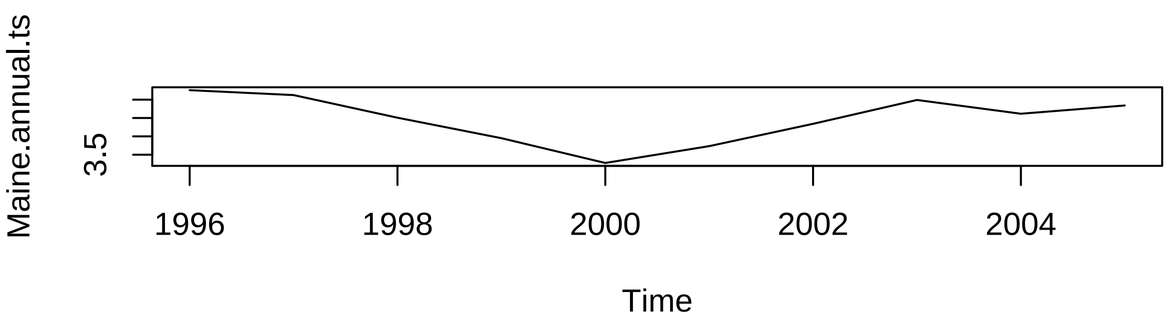
```
## [1] "data.frame"
```

```r
# it's a data.frame, not a ts object. So, we need to convert it to ts
Maine.month.ts <- ts(unemploy, start = c(1996, 1), freq = 12)
Maine.month.ts
```

```
##      Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
## 1996 6.7 6.7 6.4 5.9 5.2 4.8 4.8 4.0 4.2 4.4 5.0 5.0
## 1997 6.4 6.5 6.3 5.9 4.9 4.8 4.5 4.0 4.1 4.3 4.8 5.0
## 1998 6.2 5.7 5.6 4.6 4.0 4.2 4.1 3.6 3.7 4.1 4.3 4.0
## 1999 4.9 5.0 4.6 4.3 3.9 4.0 3.6 3.3 3.1 3.3 3.7 3.7
## 2000 4.4 4.4 4.1 3.5 3.1 3.0 2.8 2.5 2.6 2.8 3.1 3.0
## 2001 3.9 4.2 4.0 4.1 3.5 3.5 3.4 3.1 3.4 3.7 4.0 4.0
## 2002 5.0 4.9 5.0 4.7 4.0 4.2 4.0 3.6 3.7 3.9 4.5 4.6
## 2003 5.6 5.8 5.6 5.5 4.8 4.9 4.8 4.3 4.5 4.6 4.8 4.7
## 2004 5.6 5.6 5.5 4.8 4.2 4.3 4.2 3.8 4.0 4.2 4.6 4.6
## 2005 5.5 5.8 5.5 5.2 4.7 4.6 4.5 4.1 4.4 4.4 4.8 4.6
## 2006 5.3 5.6 4.9 4.6 4.2 4.4 4.4 3.9
```

```r
Maine.annual.ts <- aggregate(Maine.month.ts)/12
layout(1:2)
plot(Maine.month.ts)
plot(Maine.annual.ts)
```





```r
Maine.Feb <- window(Maine.month.ts, start = c(1996,2), freq = TRUE)
Maine.Aug <- window(Maine.month.ts, start = c(1996,8), freq = TRUE)
Feb.ratio <- mean(Maine.Feb) / mean(Maine.month.ts)
Aug.ratio <- mean(Maine.Aug) / mean(Maine.month.ts)
```

========

```r
# monthly unemployment rate for all of the United States from January 1996 until October 2006
US.month <- read.table("http://staff.elena.aut.ac.nz/Paul-Cowpertwait/ts/USunemp.dat", header = T)
attach(US.month)
US.month.ts <- ts(USun, start=c(1996,1), end=c(2006,10), freq = 12)
plot(US.month.ts, ylab = "unemployed (%)")
```

natoionwide unemployment rate

========

```r
ChocolateBeerElectricity <- read.table("http://staff.elena.aut.ac.nz/Paul-Cowpertwait/ts/cbe.dat", heade
class(ChocolateBeerElectricity)
```

Multiple TS

```
## [1] "data.frame"
```

```r
str(ChocolateBeerElectricity)
```

```
## 'data.frame':    396 obs. of  3 variables:
```

```
##  $ choc: int  1451 2037 2477 2785 2994 2681 3098 2708 2517 2445 ...
##  $ beer: num  96.3 84.4 91.2 81.9 80.5 70.4 74.8 75.9 86.3 98.7 ...
##  $ elec: int  1497 1463 1648 1595 1777 1824 1994 1835 1787 1699 ...
```

```r
Chocolate.ts <- ts(ChocolateBeerElectricity[,1], start = 1958, frequency = 12)
Beer.ts <- ts(ChocolateBeerElectricity[,2], start = 1958, frequency = 12)
Electricity.ts <- ts(ChocolateBeerElectricity[,3], start = 1958, frequency = 12)

plot(cbind(Chocolate.ts, Beer.ts, Electricity.ts))
```