# Time Series Analysis

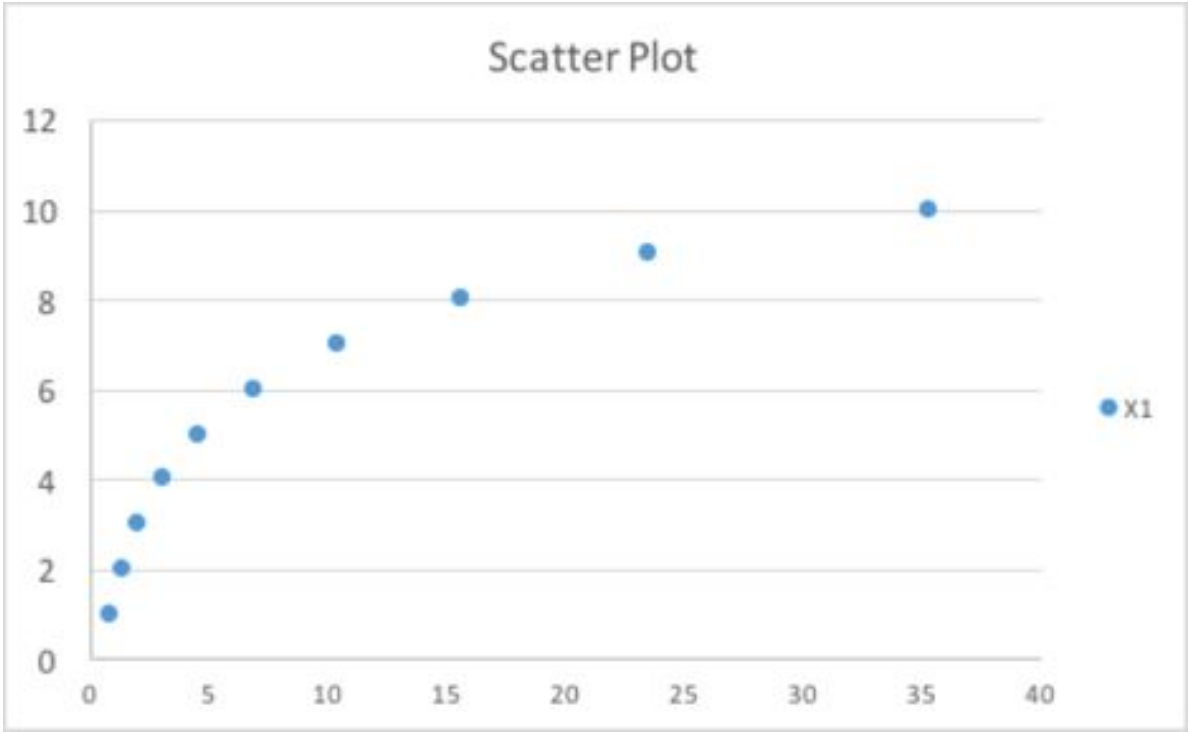
## Business Analytics

# Lesson Objectives

- Background on time series analysis
- Working with temporal data
  - Smoothing methods
  - Seasonality analysis

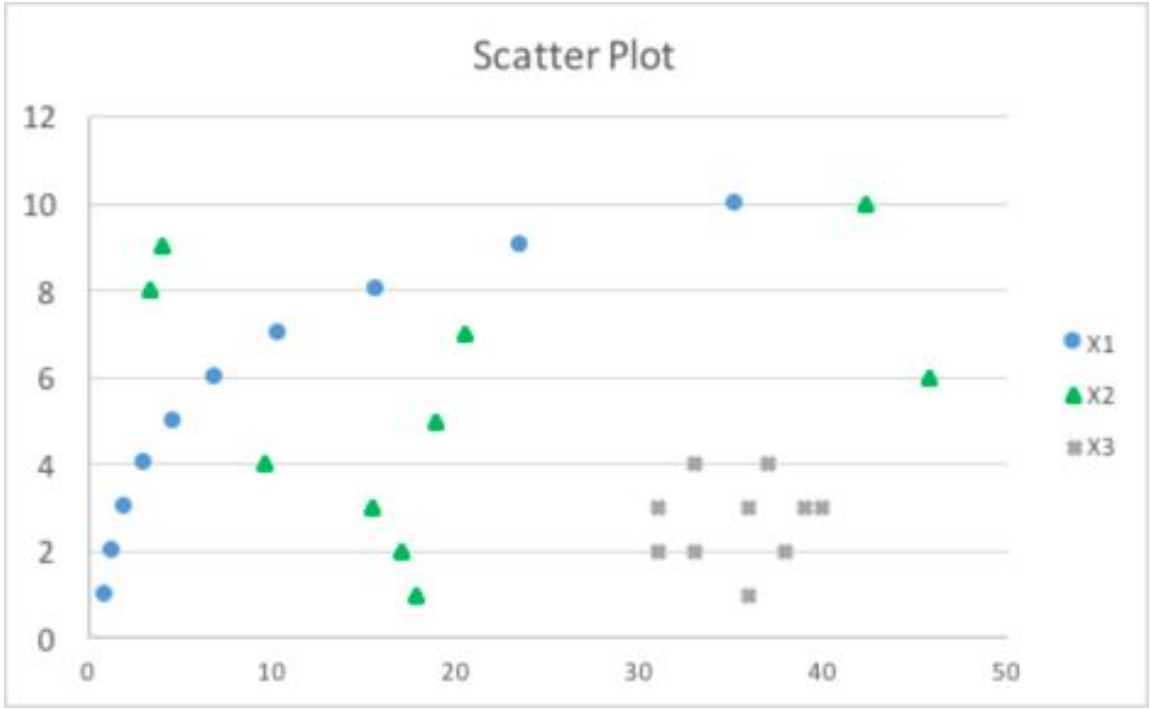
# Motivation

Univariate→

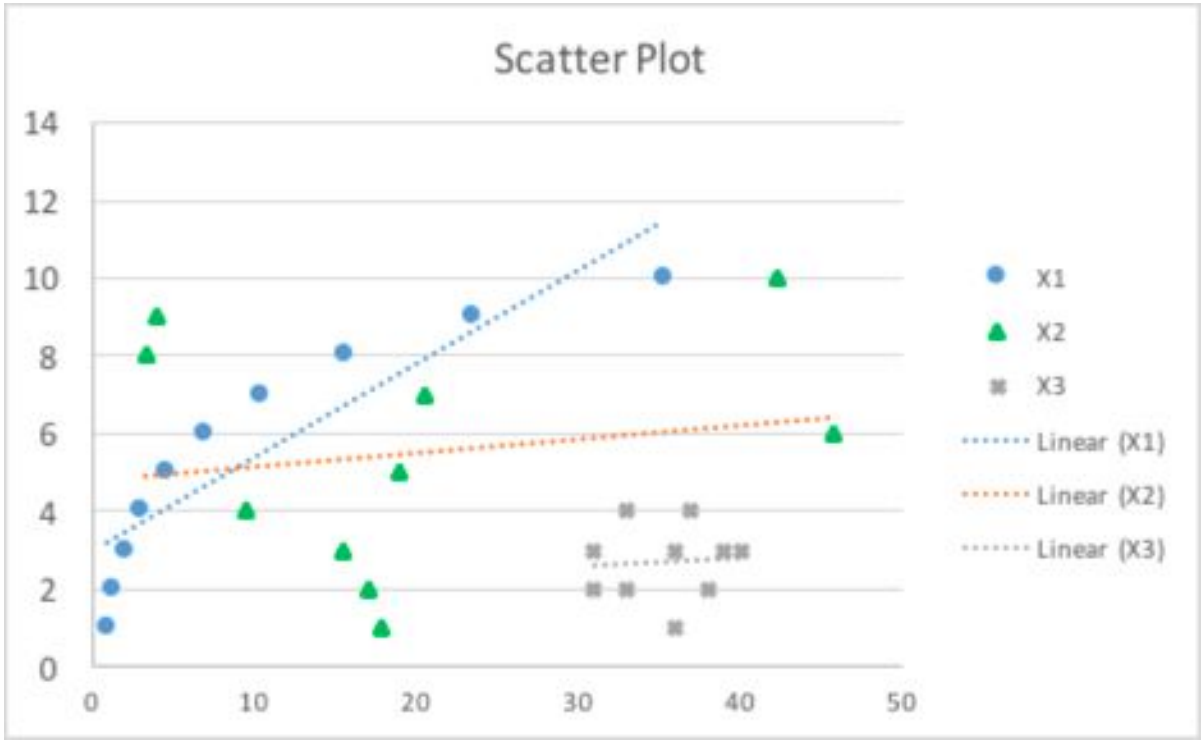


# Motivation

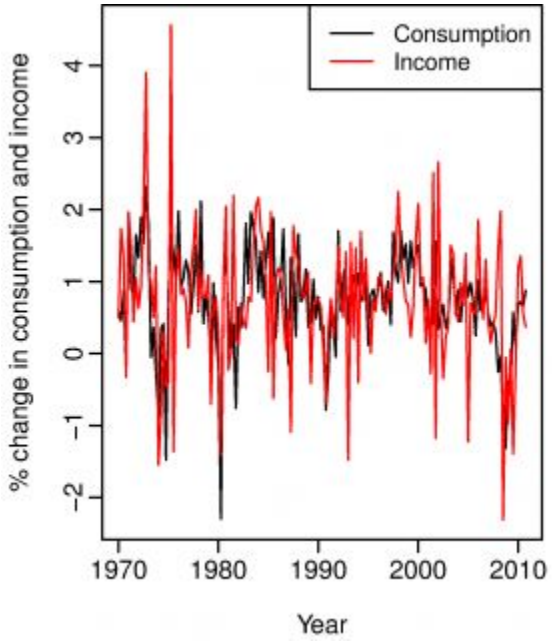
Multivariate →

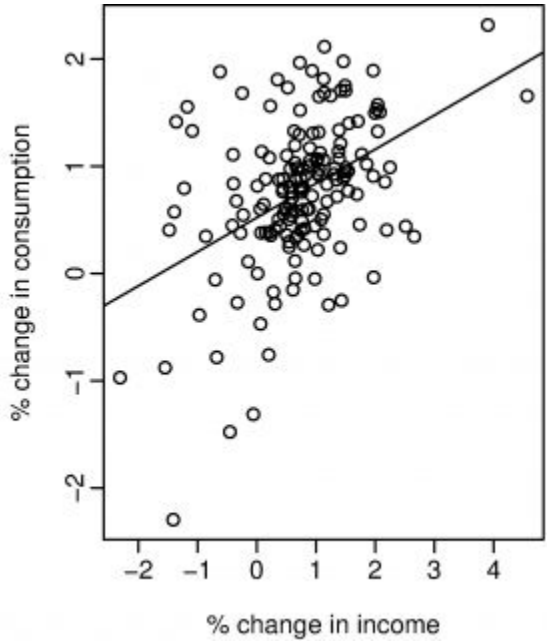


# Motivation


Scatter Plot
14
12
10
8
6
4
2
0
0
10
20
30
40
50
X1
X2
X3
Linear (X1)
Linear (X2)
Linear (X3)

# Time Series vs Cross-Sectional Data



# Time Series vs Cross-Sectional Data

Time Period: Jan 01, 2015 - Dec 31, 2015 Show: Historical Prices Frequency: Monthly Apply

Currency in USD. Download Data

| Date | Open | High | Low | Close | Adj Close* | Volume |
|---|---|---|---|---|---|---|
| Dec 01, 2015 | 2,082.93 | 2,104.27 | 1,993.26 | 2,043.94 | 2,043.94 | 83,649,260,000 |
| Oct 31, 2015 | 2,080.76 | 2,116.48 | 2,019.39 | 2,080.41 | 2,080.41 | 75,913,590,000 |
| Sep 30, 2015 | 1,919.65 | 2,094.32 | 1,893.70 | 2,079.36 | 2,079.36 | 85,844,900,000 |
| Aug 31, 2015 | 1,970.09 | 2,020.86 | 1,871.91 | 1,920.03 | 1,920.03 | 79,989,370,000 |
| Jul 31, 2015 | 2,104.49 | 2,112.66 | 1,867.01 | 1,972.18 | 1,972.18 | 84,626,790,000 |
| Jun 30, 2015 | 2,067.00 | 2,132.82 | 2,044.02 | 2,103.84 | 2,103.84 | 77,920,590,000 |
| May 31, 2015 | 2,108.64 | 2,129.87 | 2,056.32 | 2,063.11 | 2,063.11 | 73,213,980,000 |
| Apr 30, 2015 | 2,087.38 | 2,134.72 | 2,067.93 | 2,107.39 | 2,107.39 | 65,187,730,000 |
| Mar 31, 2015 | 2,067.63 | 2,125.92 | 2,048.38 | 2,085.51 | 2,085.51 | 69,440,940,000 |
| Mar 01, 2015 | 2,105.23 | 2,117.52 | 2,039.69 | 2,067.89 | 2,067.89 | 76,675,850,000 |
| Feb 01, 2015 | 1,996.67 | 2,119.59 | 1,980.90 | 2,104.50 | 2,104.50 | 68,775,560,000 |
| Jan 01, 2015 | 2,058.90 | 2,072.36 | 1,988.12 | 1,994.99 | 1,994.99 | 77,300,040,000 |

# Time Series vs Cross-Sectional Data

Time Period: Jan 01, 2015 - Dec 31, 2015 Show: Historical Prices Frequency: Monthly Apply

Currency in USD. ← Time series Download Data

| Date | Open | High | Low | Close | Adj Close* | Volume |
|---|---|---|---|---|---|---|
| Dec 01, 2015 | 2,082.93 | 2,104.27 | 1,993.26 | 2,043.94 | 2,043.94 | 83,649,260,000 |
| Oct 31, 2015 | 2,080.76 | 2,116.48 | 2,019.39 | 2,080.41 | 2,080.41 | 75,913,590,000 |
| Sep 30, 2015 | 1,919.65 | 2,094.32 | 1,893.70 | 2,079.36 | 2,079.36 | 85,844,900,000 |
| Aug 31, 2015 | 1,970.09 | 2,020.86 | 1,871.91 | 1,920.03 | 1,920.03 | 79,989,370,000 |
| Jul 31, 2015 | 2,104.49 | 2,112.66 | 1,867.01 | 1,972.18 | 1,972.18 | 84,626,790,000 |
| Jun 30, 2015 | 2,067.00 | 2,132.82 | 2,044.02 | 2,103.84 | 2,103.84 | 77,920,590,000 |
| May 31, 2015 | 2,108.64 | 2,129.87 | 2,056.32 | 2,063.11 | 2,063.11 | 73,213,980,000 |
| Apr 30, 2015 | 2,087.38 | 2,134.72 | 2,067.93 | 2,107.39 | 2,107.39 | 65,187,730,000 |
| Mar 31, 2015 | 2,067.63 | 2,125.92 | 2,048.38 | 2,085.51 | 2,085.51 | 69,440,940,000 |
| Mar 01, 2015 | 2,105.23 | 2,117.52 | 2,039.69 | 2,067.89 | 2,067.89 | 76,675,850,000 |
| Feb 01, 2015 | 1,996.67 | 2,119.59 | 1,980.90 | 2,104.50 | 2,104.50 | 68,775,560,000 |
| Jan 01, 2015 | 2,058.90 | 2,072.36 | 1,988.12 | 1,994.99 | 1,994.99 | 77,300,040,000 |

Cross Sectional →

NYU TA O

# Time Series

- A time series is a series of data points in time order. Time series data have a ***natural temporal ordering***. This makes time series analysis distinct from cross-sectional studies, in which there is no natural ordering of the observations
- Another name is temporal data, because information is spaced in time. Examples
  - Weather
  - Signal processing
  - Pattern recognition,
  - Finance
  - Website traffic

# Time Series Analysis

Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data.

Time series analysis is distinct from spatial data analysis where the observations typically relate to geographical locations

The biggest difference is that time series regression accounts for the autocorrelation between time events: the next value in a time series is affected by earlier values in the time series, while in normal regression, independence of serial errors are presumed, or at least minimized.

# Time Series Analysis

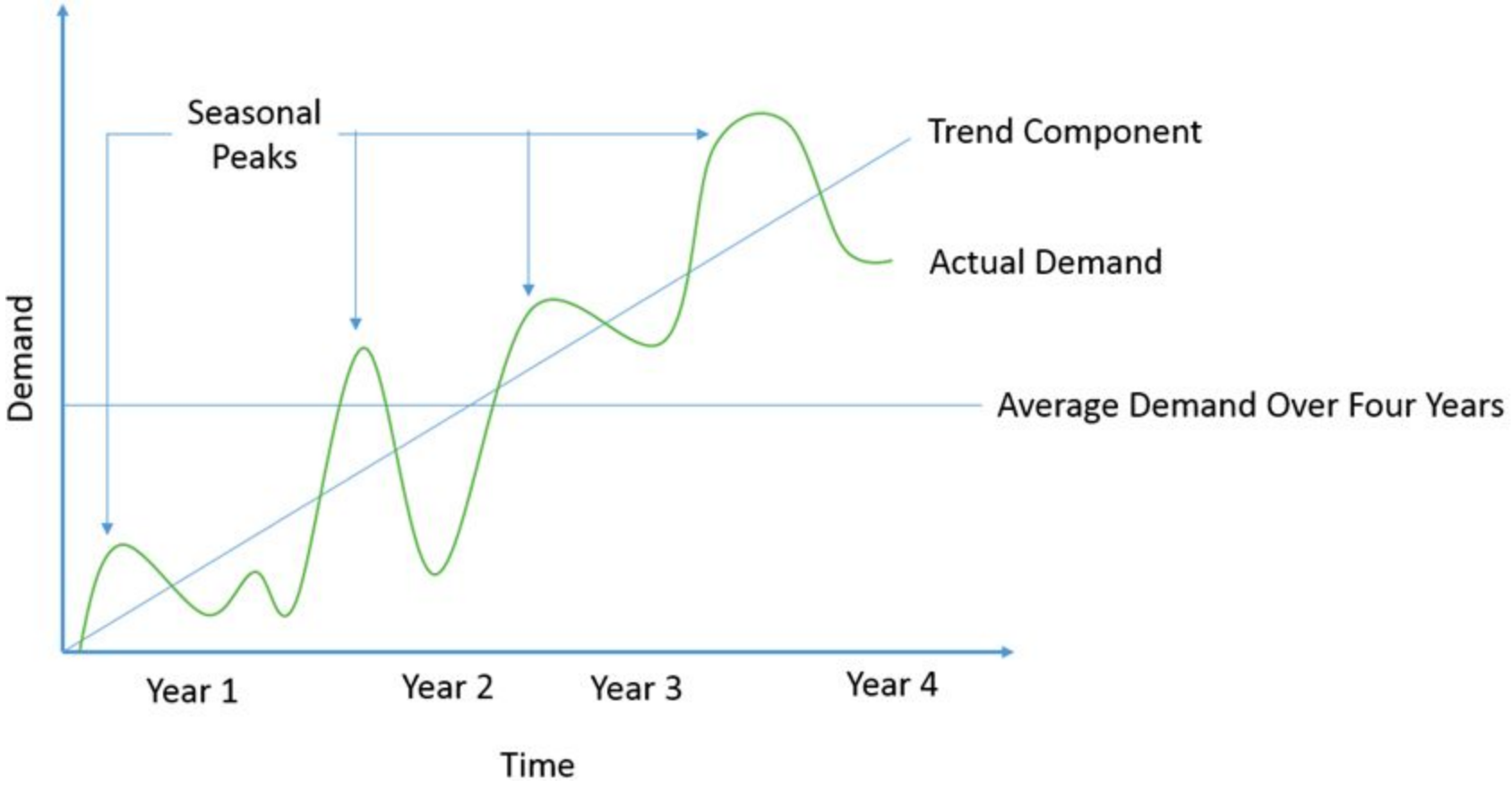
There are two main goals of time series analysis.

- Identify the nature of the phenomenon represented by the sequence of observations in the data.
- Use the data to forecast or predict future values of the time series variable.

Both of these goals require that we identify the pattern of observed time series data and more or less formally describe it.

# Time Series Analysis

- Trend
  - A trend exists when there is a long-term increase or decrease in the data. It does not have to be linear. Sometimes we will refer to a trend “changing direction” when it might go from an increasing trend to a decreasing trend.
- Seasonal
  - A seasonal pattern exists when a series is influenced by seasonal factors (e.g., the quarter of the year, the month, or day of the week). Seasonality is always of a fixed and known period.
- Cyclic
  - A cyclic pattern exists when data exhibit rises and falls that are not of fixed period. The duration of these fluctuations is usually of at least 2 years
- Random
  - Irregular component or residuals

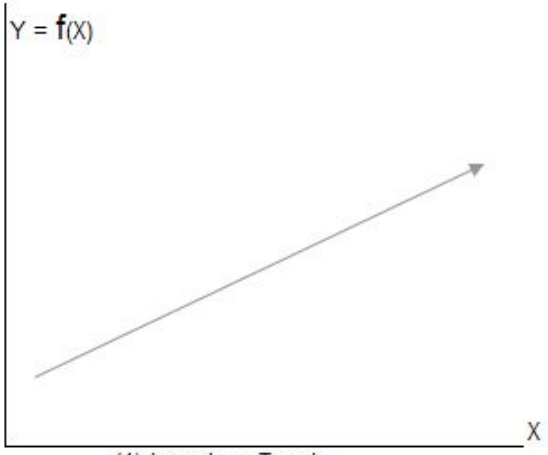

Seasonal
Peaks
Trend Component
Actual Demand
Average Demand Over Four Years
Demand
Year 1
Year 2
Year 3
Year 4
Time

# Cyclic vs Seasonal behavior

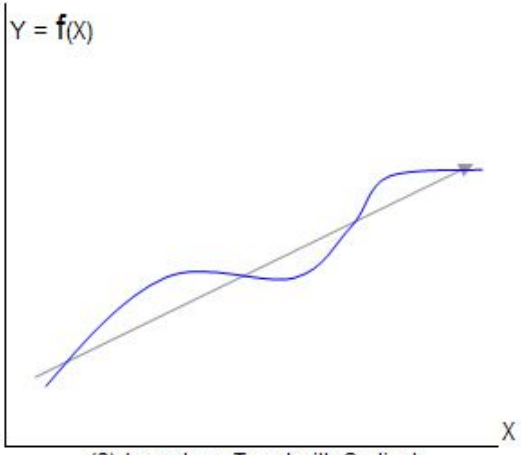
Many people confuse cyclic behaviour with seasonal behaviour, but they are really quite different.

- If the fluctuations are not of fixed period then they are cyclic
- If the period is unchanging and associated with some aspect of the calendar, then the pattern is seasonal.

In general, the average length of cycles is longer than the length of a seasonal pattern, and the magnitude of cycles tends to be more variable than the magnitude of seasonal patterns

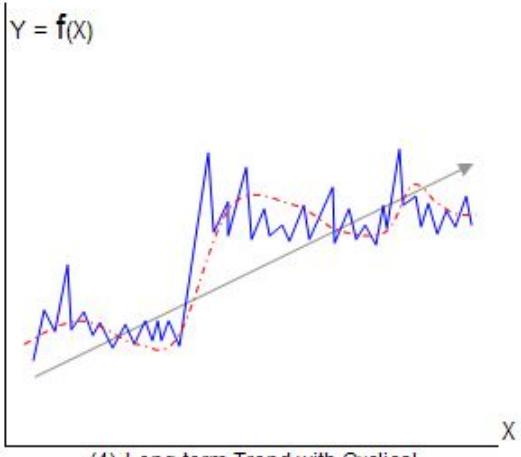



(1) Long-term Trend




(2) Long-term Trend with Cyclical Variations/Movements


(3) Long-term Trend with Cyclical and Seasonal Variations/Movements



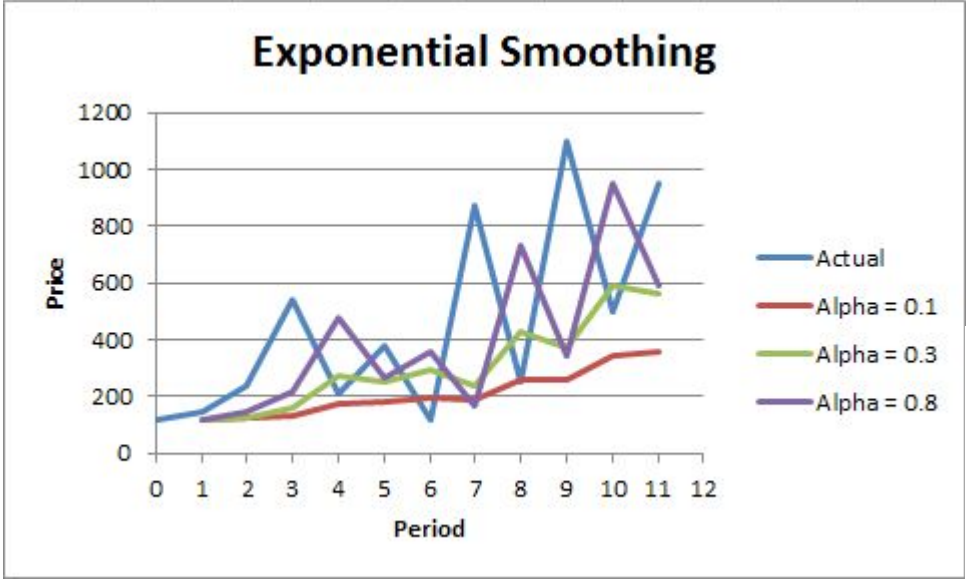

(4) Long-term Trend with Cyclical, Seasonal, and Random Variations/Movements

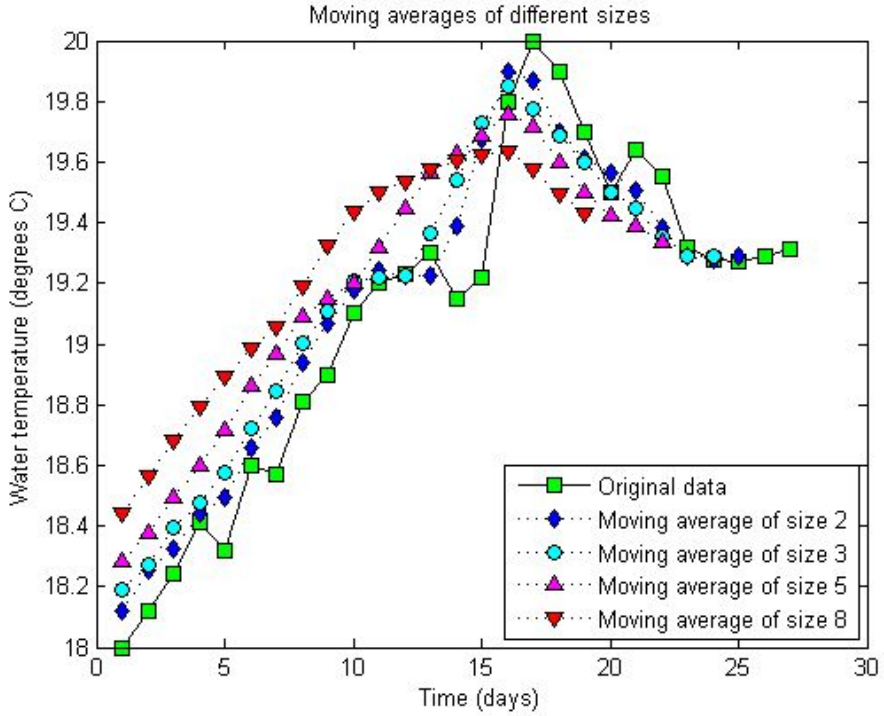
# Smoothing Methods

To smooth a data set is to create an approximating function that attempts to capture important patterns in the data, while leaving out noise or other fine-scale structures/rapid phenomena.

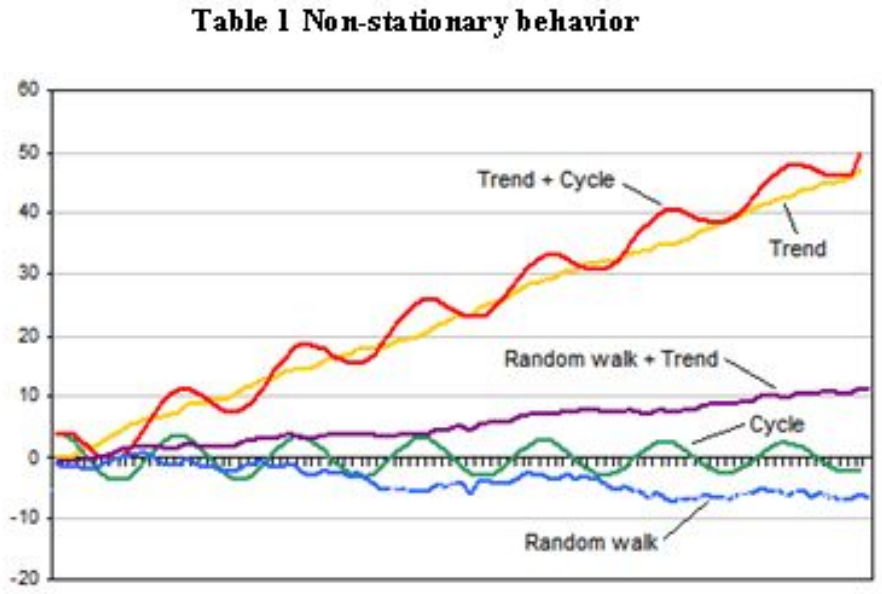
Two common techniques:

- Moving average a form of average which has been adjusted to allow for seasonal or cyclical components of a time series. Moving average smoothing is a smoothing technique used to make the long term trends of a time series clearer.  Variations include: simple, and cumulative, or weighted forms (described below).
- Exponential smoothing used to reduce irregularities (random fluctuations) in time series data, thus providing a clearer view of the true underlying behaviour of the series. It also provides an effective means of predicting future values of the time series (forecasting)


Exponential Smoothing
Price
Period
Actual
Alpha = 0.1
Alpha = 0.3
Alpha = 0.8


Moving averages of different sizes
Water temperature (degrees C)
Time (days)
Original data
Moving average of size 2
Moving average of size 3
Moving average of size 5
Moving average of size 8

# Stationary data




More details:
http://www.investopedia.com/articles/trading/07/stationary.asp

# R packages

**library(forecast)** - Forecast is a generic function for forecasting from time series or time series models. The function invokes particular methods which depend on the class of the first argument.

**R Markdown** - R Markdown is a variant of Markdown that has embedded R code chunks, to be used with knitr to make it easy to create reproducible web-based reports. http://kbroman.org/knitr_knutshell/pages/Rmarkdown.html