After the next 4 sessions on time series analysis, you should

- Grasp the wider concept of data where time is a factor
- Grasp such basic terms concepts as trend, seasonality, as well as cyclical and residual contributions in time series
- Have the ability to decompose a time series to such effects
- Be able to perform simple forecasting on a time series
- Be able to synthesize the results of time series analysis concisely
- Predict future values using simple models of time series data
Time Series Analysis: Session Learning Objectives

After the first session, you should be able to
1. Give instances of data collected over time
2. Outline the main objectives of time series analysis
3. Understand the nature of the data
4. Understand how crucial it is to graph/examine data
5. Infer key aspects arising out of studying time series
6. Report key findings from a graph of time series data

Definitions and Notation
- A time series is a collection of observations made sequentially through time
- Aside: To be pedantic, a time ‘series’ is actually a time ‘sequence’ as the values are sampled discretely
- Such observations may be denoted by

\[ x_1, x_2, x_3, \ldots, x_t, \ldots, x_T \]

(where \( x_t \) is an observation at time \( t \))
- Intervals between observations can be at any granularity (hours within days, weeks, months, etc).
Time Series Analysis: Objectives

Goals of time series analysis:

1. **Description**: Identify patterns in correlated data trends or seasonal variation
2. **Explanation**: to understand and model the data
3. **Forecasting**: to predict short-term trends from historical patterns
4. **Intervention analysis**: to gauge how single events change the series
5. **Quality control**: to assess how much deviations of a specified size indicate a problem.

We analyse time series to understand any underlying structures and functions producing the observed data

Understanding the mechanisms of a time series allows a mathematical model to be developed

Such models (hopefully) explain the data so prediction, monitoring, or control can occur
Types of Time Series

We differentiate between two types of Time Series:

- Observations made continually in time give rise to a **Continuous Time Series**, e.g.
  - Continuously measured temperature data at a Met station
  - Taking air pollution levels for limit (levels are continuous)

- More often, observations taken only at specific time points, giving rise to a **Discrete Time Series**, e.g.
  - Annual number of road accidents (intrinsically discrete)
  - Maximum daily atmospheric pressure levels (continuous but sampled at discrete intervals)
  - Whether or not it rained on particular days (binary)
  - Closing stock market prices (discrete)
Both Continuous & Discrete types of observations can be equally spaced, unequally spaced, or have missing data.

- **Discrete Time Series**
  - Noise also has a big role in data & must be dealt with (more later!).
  - Discrete measurements can be taken at any time interval.
  - However, they are mostly taken at evenly spaced intervals.

- **Continuous Time Series**
  - Such measurements can be spaced randomly in time,
  - Earthquakes measurements are such as a result of an instrument constantly recording.
When Time Series Turn Nasty...

Complexities in Time Series

Time series can be very complex due to each observation being dependent upon the previous observation(s).

- **Autocorrelation**
  - Random error is also influential between data.
  - Such influences are called autocorrelation.
  - These are dependent relationships between successive observations of the same variable.
  - A challenge of time series analysis is to extract the autocorrelation elements of the data.
  - This is to see the trend or model its underlying mechanisms.

Complexities in Time Series (cont’d)

Time series reflect stochasticity in most measurements over time.

- **Stochasticity**
  - Thus, data may be skewed, with non-constant mean/variation not constant.
  - Also maybe non-normally distributed, and not randomly sampled or independent.
  - Another non-normal aspect of observations is that often aren’t evenly spaced in time.
  - Instrument failure? Even variable no. of days in a month?
Here’s the plan for this lecture:

- To clarify these (somewhat vague) concepts we begin with some simple examples
- These show the importance of graphing data to get an insight into the distribution over time
- They also show how initial readings of the (raw) data can be misleading
- We demonstrate how the choice of simple factors can adversely affect the conclusions drawn from the data
- We conclude by showing two nice examples of data presented on graphs

Example 2.1

Things may not be as they first seem...
or...leaping to erroneous conclusions from raw data.

<table>
<thead>
<tr>
<th>Year</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>667</td>
<td>631</td>
<td>675</td>
<td>699</td>
</tr>
<tr>
<td>2</td>
<td>739</td>
<td>695</td>
<td>751</td>
<td>779</td>
</tr>
<tr>
<td>3</td>
<td>823</td>
<td>795</td>
<td>835</td>
<td>875</td>
</tr>
<tr>
<td>4</td>
<td>931</td>
<td>855</td>
<td>939</td>
<td>967</td>
</tr>
</tbody>
</table>

**Table 2.1:** Data is Quarterly company profits (in USD Ms)

- Objective: To study changes in profit figures over consecutive quarters
- Naïve Reading: 4th quarter is profits are always higher than those from the 1st quarter
Example 2.1(2/2): Take a look again...

**Figure 2.1:** Quarterly Company Profits
- False impression from data is largely due to increase over time
- Occurs often in time series, graphing helps understanding the data
- Q1 Figures always seem to dip after Q4 of previous year, however

**Example 2.2**

**Jumping to conclusions from Summary Statistics**

**Objective:** to emphasize the need to graph distributions to better understand the data distribution

<table>
<thead>
<tr>
<th></th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>20.81</td>
<td>20.81</td>
<td>20.81</td>
</tr>
<tr>
<td>Std Dev</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
</tr>
</tbody>
</table>

**Table 2.2:** Data from Petruccelli, J; MSOR Connections V7(1), 2007

Data* (interval-scale):
- Tensile strengths of string tested on a piece sampled every 5 min. from 1 spool during production.
- 100 samples each day from 3 days (simulated data)
Example 2.2(2/2): Take a look again...

(a) Day 1

(b) Day 2

(c) Day 3

Figure 2.2: Some Pieces of String
Underlying distributions are obviously very different!

Example 2.3: More False Conclusions from Stats...

<table>
<thead>
<tr>
<th>Time (days)</th>
<th>Date 1</th>
<th>Date 2</th>
<th>Date 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>8.04</td>
<td>9.14</td>
<td>7.46</td>
</tr>
<tr>
<td>8</td>
<td>6.95</td>
<td>8.14</td>
<td>6.77</td>
</tr>
<tr>
<td>13</td>
<td>7.58</td>
<td>8.74</td>
<td>12.74</td>
</tr>
<tr>
<td>9</td>
<td>8.81</td>
<td>8.77</td>
<td>7.11</td>
</tr>
<tr>
<td>11</td>
<td>8.33</td>
<td>9.26</td>
<td>7.81</td>
</tr>
<tr>
<td>14</td>
<td>9.96</td>
<td>8.1</td>
<td>8.84</td>
</tr>
<tr>
<td>6</td>
<td>7.24</td>
<td>6.13</td>
<td>6.08</td>
</tr>
<tr>
<td>4</td>
<td>4.26</td>
<td>3.1</td>
<td>5.39</td>
</tr>
<tr>
<td>12</td>
<td>10.84</td>
<td>9.13</td>
<td>8.15</td>
</tr>
<tr>
<td>7</td>
<td>4.82</td>
<td>7.26</td>
<td>6.42</td>
</tr>
<tr>
<td>Mean</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>Std Dev</td>
<td>2.03</td>
<td>2.03</td>
<td>2.03</td>
</tr>
</tbody>
</table>

Table 2.3: From Petruccelli, J; MSOR Connections Vol 7 No 2, 2007

*Note Again: Graphing is very important!*

Notes
Ex 2.3: More False Conclusions from Stats...(2/2)

Again the underlying distributions differ substantially!
- Series 1: Roughly upward trend?
- Series 2: Polynomial fit, starting to trend downwards?
- Series 3: Linear fit with an outlier/error?

Example 2.4: Zambian Rainfall Data

Problem:
- Trend in Southern Zambia for farmers to move into cities
- Maybe believe that climate change affects farm production.
- A local NGO promoting Conservation Farming thinks the issue is bad farming.

Study commissioned to investigate the problem:
- Examined the ‘Start of the Rains’ (defined as >20mm of rainfall in 3 days, after 15 November)
- Another was the number and effect of dry weather events following on from this event
Example 2.4: Zambian Rainfall Data (/2)

**Figure 2.4:** Rainfall Data Graph Over Years
- Question of change in rainy season: Examine the data!

Example 2.4: Zambian Rainfall Data (/3)

**Figure 2.5:** As Figure 2.4 Except Centred Around Day 150
- Answer:
  - Maybe fluctuations around July 1+ (i.e. day 150)
  - Seem to become more extreme and persistent, for period 1960-Now vis à vis before.
Example 2.4: Zambian Rainfall Data

(a) Rainy Days Data  
(b) Start of Rains Data

Closer Examination

- Fig 2.6(a) has small trend (downward for No. of Rainy days/season, upward for average rainfall per rainy day)
- Fig 2.6(b): some increase in no. of 10-day dry spells in 30 days after ‘Start of the Rains’ (maybe need for replanting?)
- Neither of these was found to be statistically significant, however.

Lessons Summarised

- The level of summary of data before analysis depends on study objective(s)
- Type of analysis depends on objectives - often descriptive analysis is ok
- Different data levels needed depending on whether the problem is being looked at the international level, national level or local level
- Imperative however that quality data be made accessible to ensure that conclusions arising from the analysis are correct.
- This means that data should be (as far as possible) clean, whole and free from outliers
- In examining data, remember human brain is preprogramed to find patterns in data/graphs.
This shows the measurement of interest against time of observation.

No matter what you decide is the appropriate way to analyse your data, the time factor must not be ignored.

From examples, very important to start the exploration of a time series with a graphical representation of the data.

However, there are a number of points to be kept in mind when drawing such a plot, as discussed below.

Both figures depict the same subject
- lower is measured at a smaller interval
- but top is measured at a longer interval missing heartbeats odd (and maybe significant) peak.

So sampling interval choice is quite important:
- too fine can be costly (or maybe too invasive for some tests)
- too coarse might miss out essential characteristics.
**Time Plots (2): A Subtlety: Choice of Aspect Ratio**

- **Figure 2.8:** Effects of Different Aspect Ratios
  - Nb: different aspect ratios emphasize different characteristics of the series:
    - the top one brings out the differences in the peaks
    - while the lower one highlights the way the peaks rise and fall.

**Time Plots (3): To join or not to join**

- **Figure 2.9:** A Figure Showing the Subtleties of Not Joining Data
  - Pluses and minuses of joining:
    - **Advantage:** usually easier to digest data, draw conclusions
    - **Disadvantage:** gives impression of continuity; definitely a risk when missing values exist.
Finally, Two Elegant Graphs

- Could just present as 12 graphs for monthly data
- Would not demonstrate as clearly key features:
  - ‘Christmas Baby’ Effect
  - ‘Few Spring/Summer Babies’
  - ‘Sept 15th???

This is a very famous graph from Charles Minard
- Shows in 2D six types of data: the number of Napoleon’s troops; distance; temperature; the latitude longitude; direction of travel; and location relative to specific dates.
Finally: Two Very Elegant Graphs (/2)

This is a very famous graph from Charles Minard

Shows in 2D six types of data: the number of Napoleon's troops; distance; temperature; the latitude longitude; direction of travel; and location relative to specific dates.
Finally: Two Very Elegant Graphs (/2)

This is a very famous graph from Charles Minard

Shows in 2D six types of data: the number of Napoleon’s troops; distance; temperature; the latitude longitude; direction of travel; and location relative to specific dates.