Visualising Time Series Data

NANYANG THE INNOVATIVE POLYTECHNIC School of Information Technology

Learning Outcomes

By the end of this class, you should be able to

- Identify the pattern of time-series data
- Identify and explain the difference of discrete and continuous time-series data
- Explain and apply the techniques and best practices used

Introduction

- Time series analysis is about visualising the quantitative relationship through time:
  - Business analysts hope to see profits increase over time
  - Government expects to see changes in relation to influential events

- Time gives us a context for understanding data so that *present can be understood* and *future can be predicted* in light of the past.

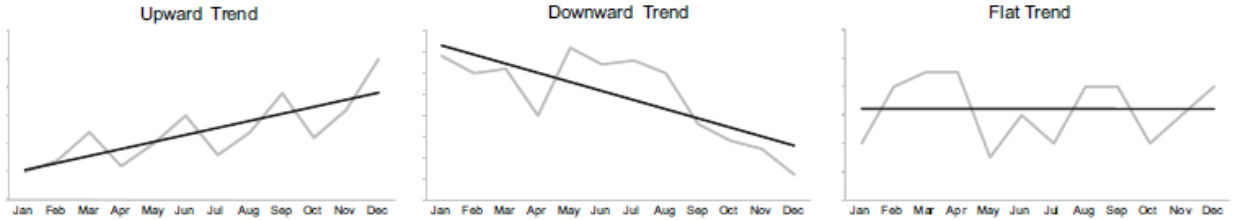
Time-Series Patterns

Six basic patterns:

- Trend
- Variability
- Rate of change
- Co-variation
- Cycles
- Exceptions

Trend

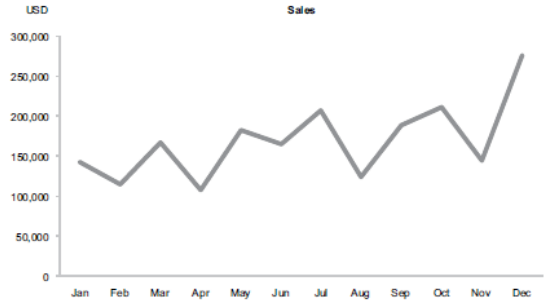
- A trend is the overall tendency of a series of values to increase, decrease or remain relatively stable during a particular period of time (like 12-months)
- Any period of time can be the basis for determining a trend but the *comparison* has to be within the same period
- Using lines of best fit based on statistical models

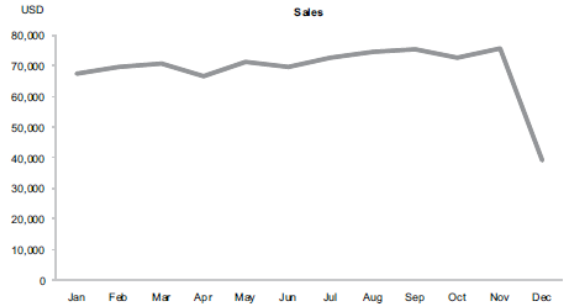


Variability

- Variability is the average degree of change from one point in time to the next throughout a span of time.

Which one has high variability?





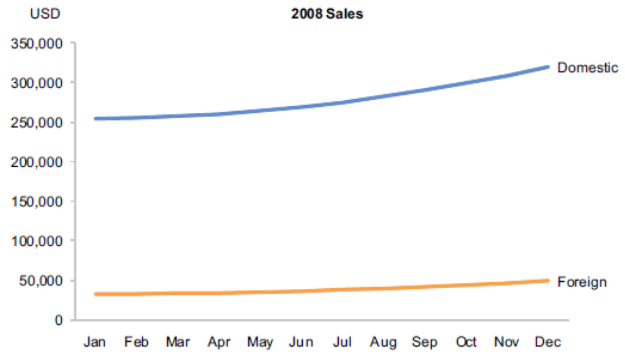
Consider month-month variability

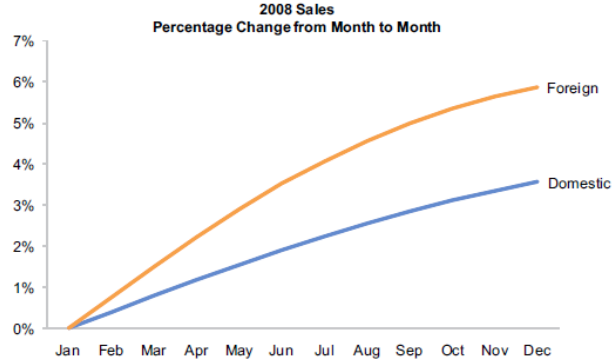
Rate of Change

Domestic sales is larger than foreign sales but if percentage is considered, foreign sales is increased at a faster rate though they are increasing by smaller amount.

- This is useful when comparing multiple series of values.
- Percentage difference is considered (not absolute value)

10k over 50k is 20% increase but 10k over 200k is just 5% increase



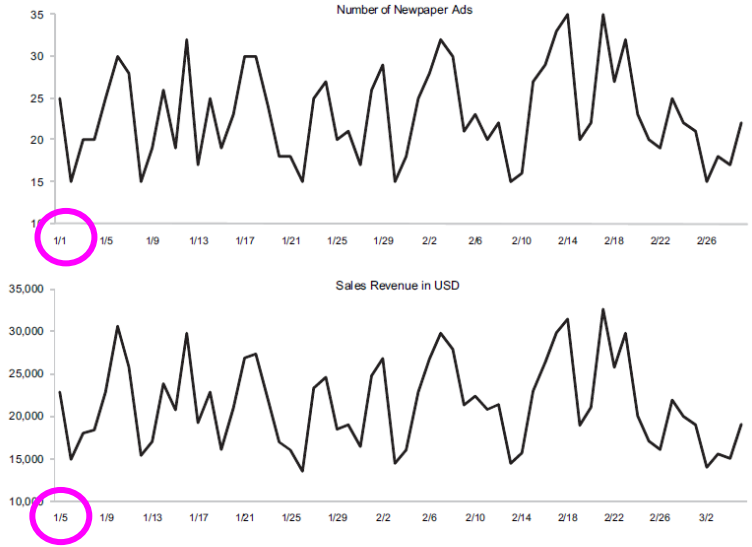


What conclusion can you tell from the graph?

Co-variation

- When 2 time series relate to one another so that changes in one are reflected as changes in the other, either immediately or later, is called co-variation.

Co-variation between newspapers ads (leading indicator) and orders (lagging indicator)



4 days later →

Cycles

- Cycles are patterns that repeat at regular intervals, such as daily, weekly, monthly, quarterly or seasonally.

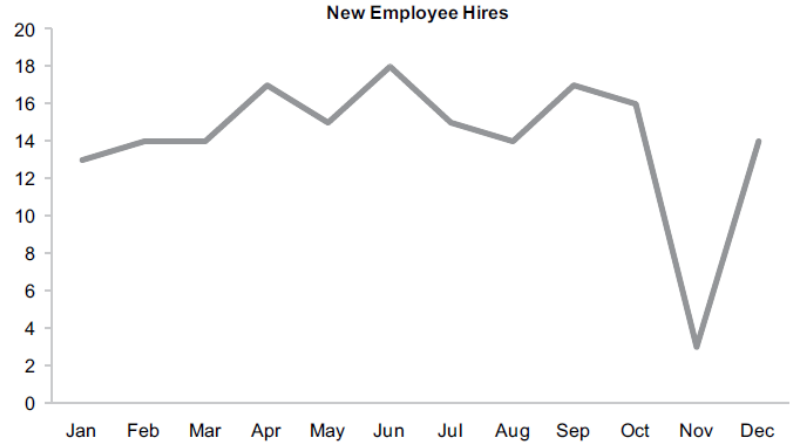
Any idea why such pattern for Sales data?

Customer buying preference ??

- *Hockey stick* pattern – upward bend near the end.

Exceptions

- Exceptions – values fall outside the norm.
- Every industry or work function has particular and unique time-series patterns that are of special interest.
- Eg: Floral & hamper company



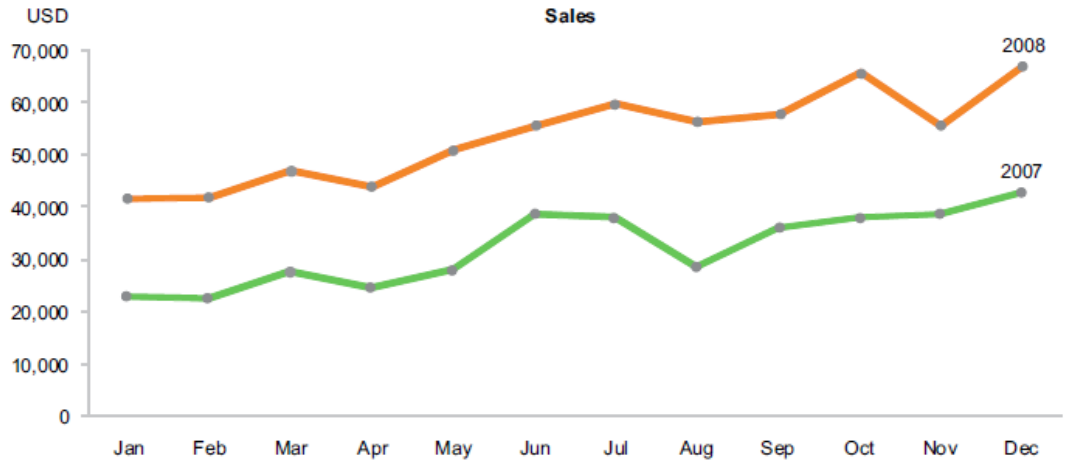
Time-Series Displays

Most time-series analysis can and should be accomplished using line chart. However sometimes other charts are suitable too:

- Bar charts
- Dot plot
- Radar charts
- Heatmaps
- Scatterplots *(esp. in animation to analyse the correlation changes)*

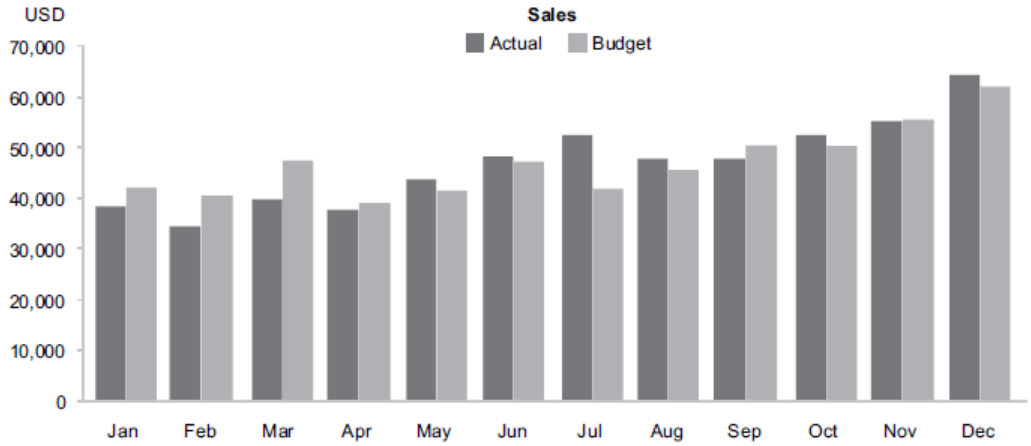
Line Chart

- For analysing Patterns and Exceptions



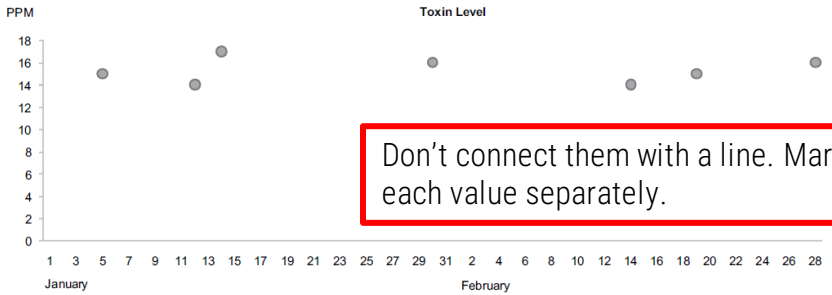
Bar Chart

- For Emphasizing and Comparing Individual Values

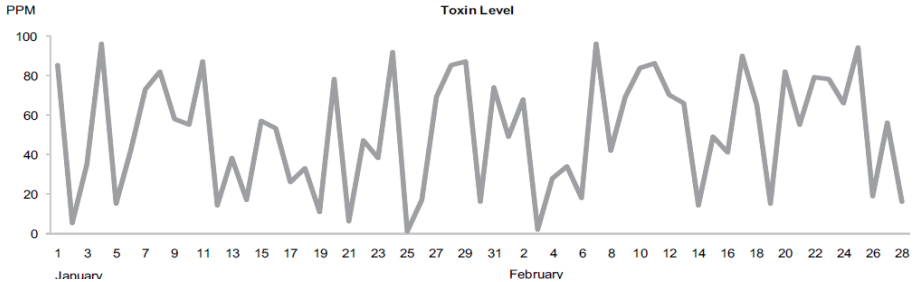


Dot Plots

- For analysing Irregular Intervals
- Eg: data for certain days

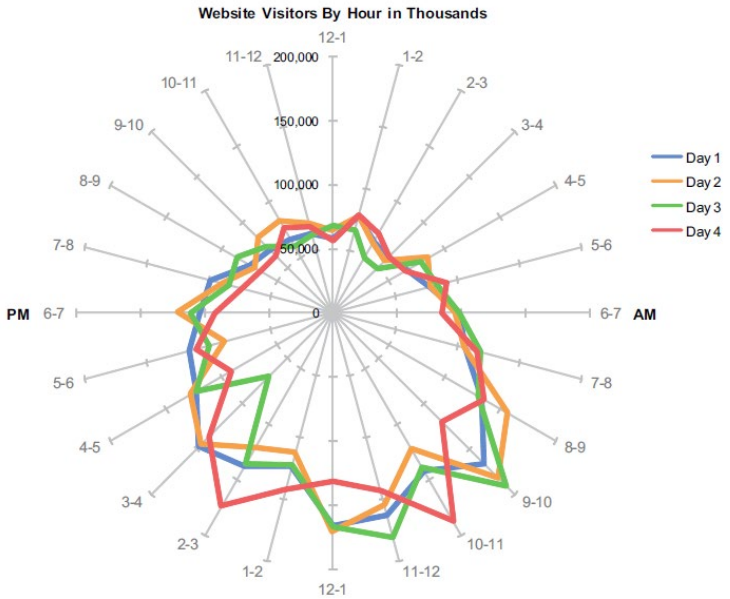


Don't connect them with a line. Mark each value separately.



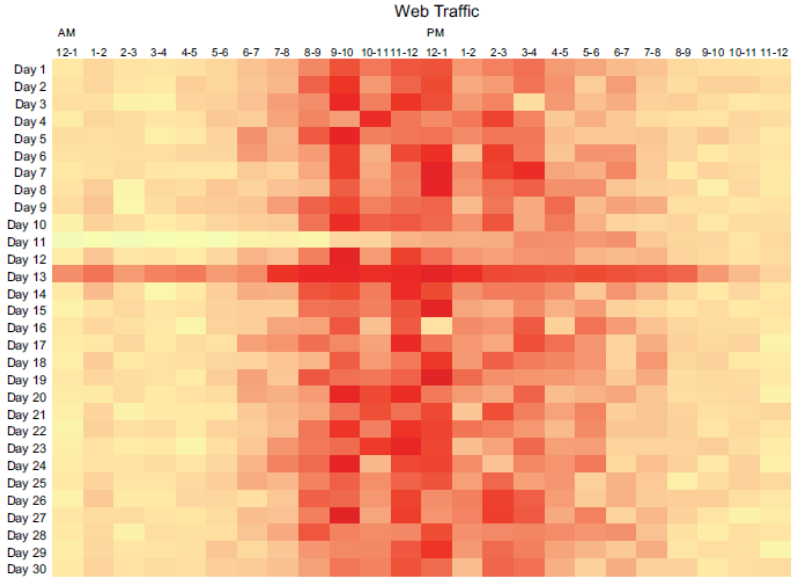
Radar Chart [Spider / Star Chart]

- For comparing cycles and spot exception
- Eg: 2 dimension- hr and day



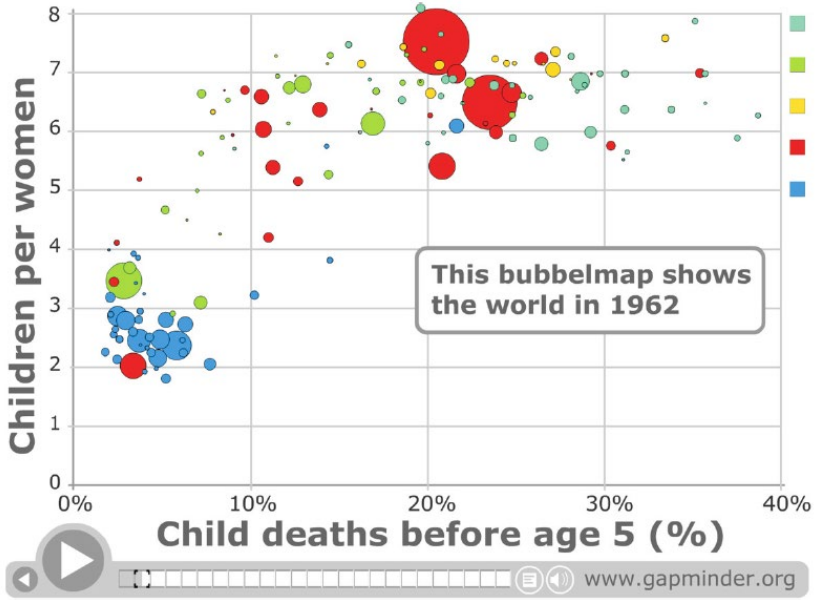
Heatmaps

- For analysing high volume cyclical Patterns and Exception



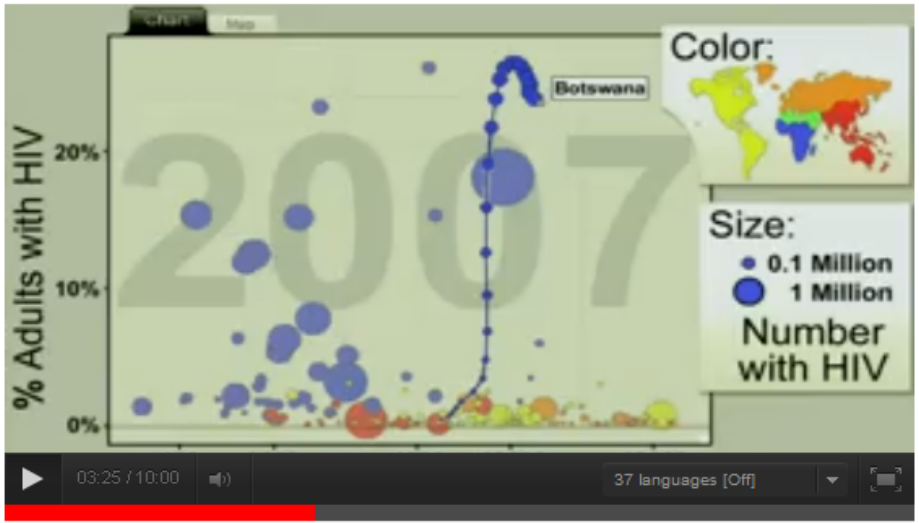
Scatterplots

- Animated for analysing correlation changes



Hans Rosling: HIV -- new facts and stunning data visuals

FILMED FEB 2009 • POSTED MAY 2009 • TED2009



556,689 Views

Like 2k

Hans Rosling unveils new data visuals that untangle the complex risk factors of one of the world's deadliest (and most misunderstood) diseases: HIV. He argues that preventing transmissions -- not drug treatments -- is the key to ending the epidemic.

In Hans Rosling's hands, data sings. Global trends in health and economics come to vivid life. And the big picture of global development—with some surprisingly good news—snaps into sharp focus. Full bio »

RELATED PLAYLISTS NEW View more »

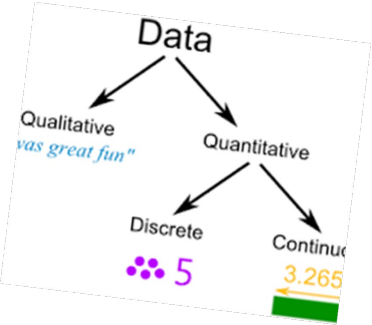
Hans Rosling: 5 talks on global issues

*Curated by **Hans Rosling***

Data visionary Hans Rosling's favorite talks keep an eye toward health, economics and population growth -- from business in Africa to youth...

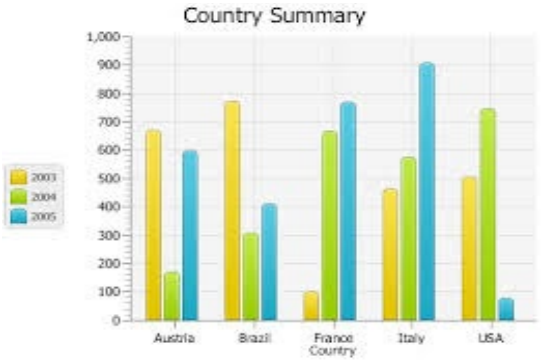
Discrete and Continuous Time-Series Data

- Discrete – values are from specific points or blocks of time and there is a finite number of possible values.
  - E.g. Percentage of people who pass a test each year – this value don't change afterward
- Continuous – value that can be measured at any time of day during any interval and it is constantly changing.
  - E.g. Environment or body temperature

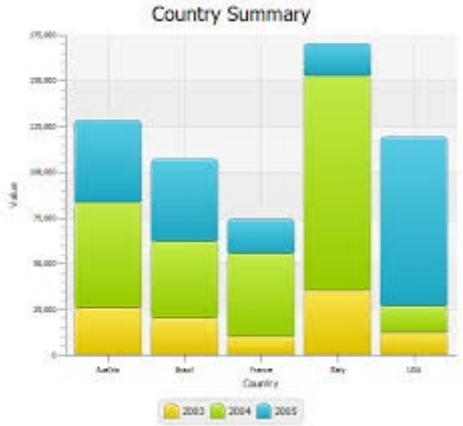


Discrete Time-Series Data

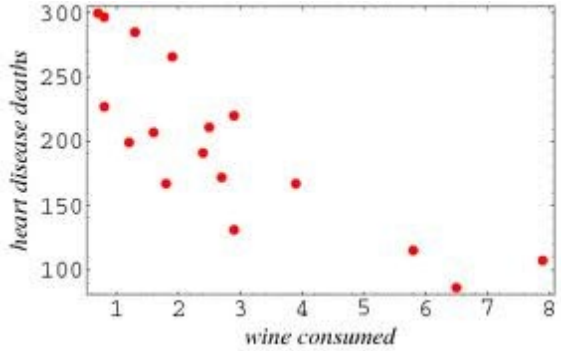
Bar chart



Stacked Bar chart



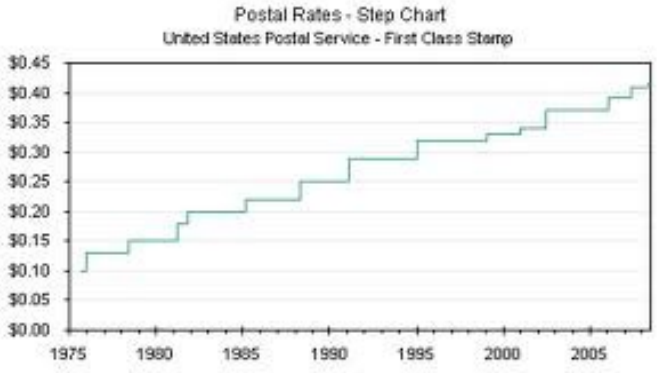
Dot plot



Continuous Time Series Data

Line chart

Step chart

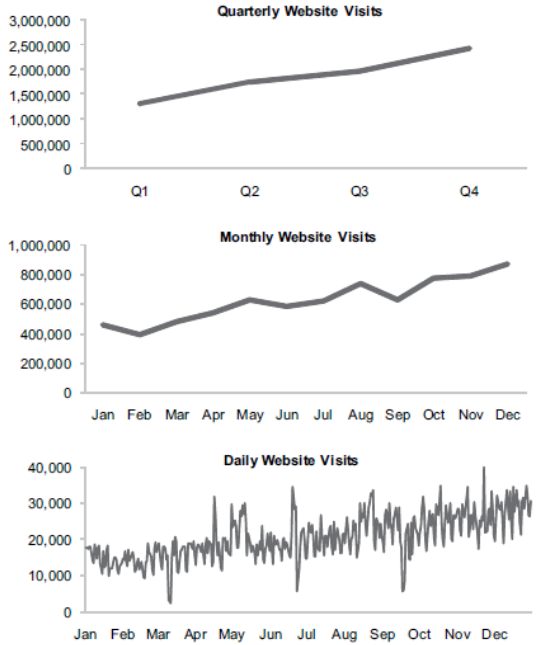


Techniques and Best Practices

- Aggregating to Various Time Intervals
- Viewing Time Periods in Context
- Using Trend Line to enhance Perception of high-level Patterns
- Omitting Missing Values from a display
- Using Logarithmic Scales and Percentages to Compare Rates of Change
- Overlapping time scales to Compare Cyclical Patterns
- Stacking Line Charts to Compare Multiple Variables

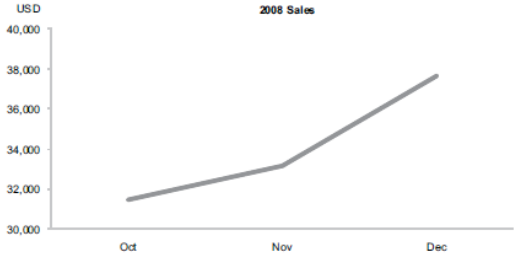
Aggregating

- Time-series data can look quite different if you change the level of aggregation.
- There is a need to quickly and easily switch between various intervals of time to examine the data and identify the insights.



Context

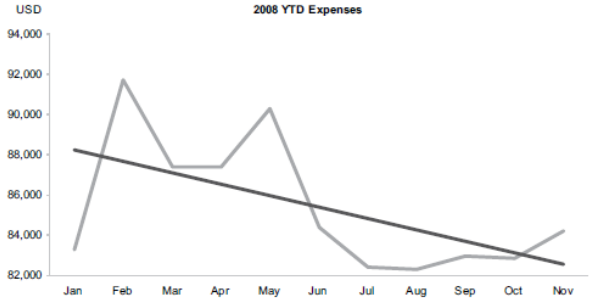
- Views of various time spans might lead to insights that are not available if we stick to one time span.
- Important to examine data in the context of a longer period.

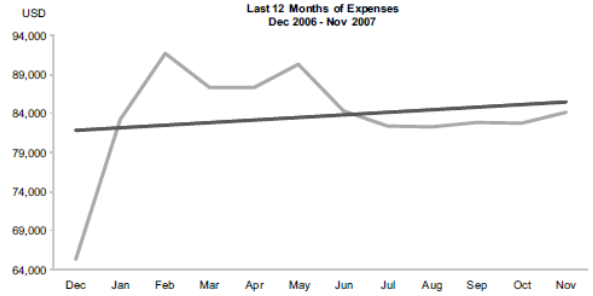


Based on this graph, we might conclude that sales are trending upward. Now look at the same three months of data, this time in the context of the entire year.

Trend line

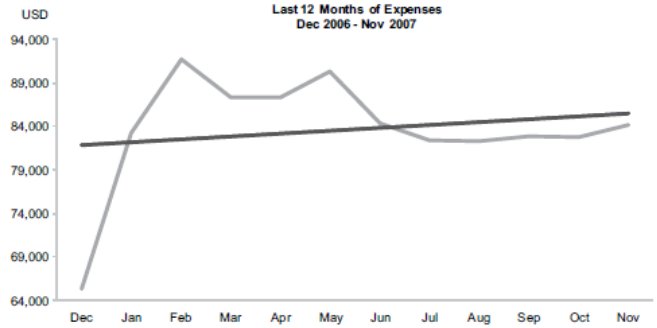
- Take note of the time period used for the trend line
- Be sure to examine values that fall outside the specified time period and not isolated a section that would trend quite differently if the period is altered

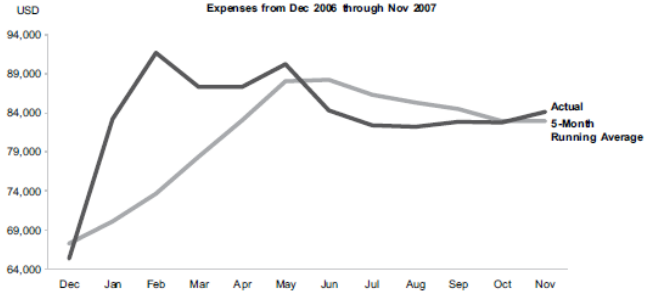




Trend line

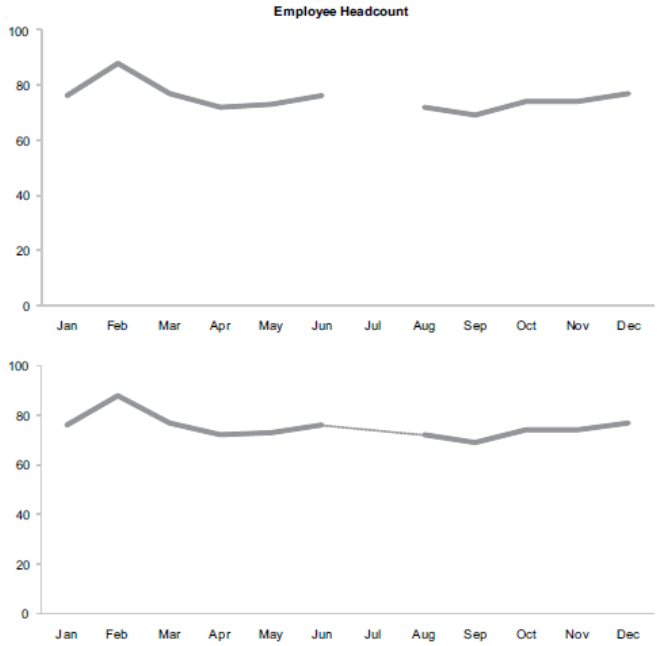
- Usually a straight line is used and based on linear regression
- However it may not be the best fit so may want to consider – running average





Omitting

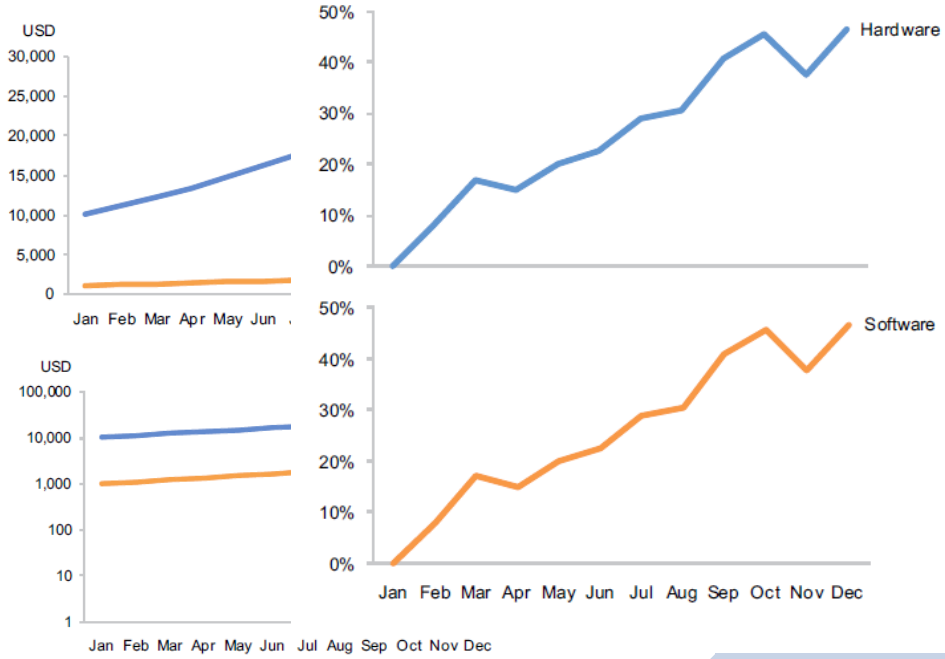
- If missing values are found, we can estimate the value or omit it from displaying.
- It is not right to treat missing value as zero.



Logarithmic

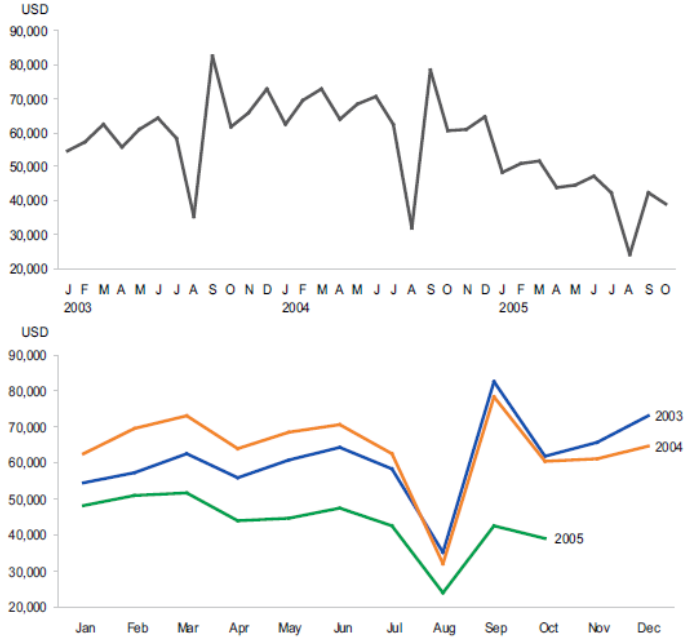
- When comparing rate of changes, important to use percentage and logarithm scale
- Both increased at 10% but 10% of $1,000 is $100 and 10% of $10,000 is $1,000 so increase of $1,000 is steeper

equal rate of change appear as equal slope



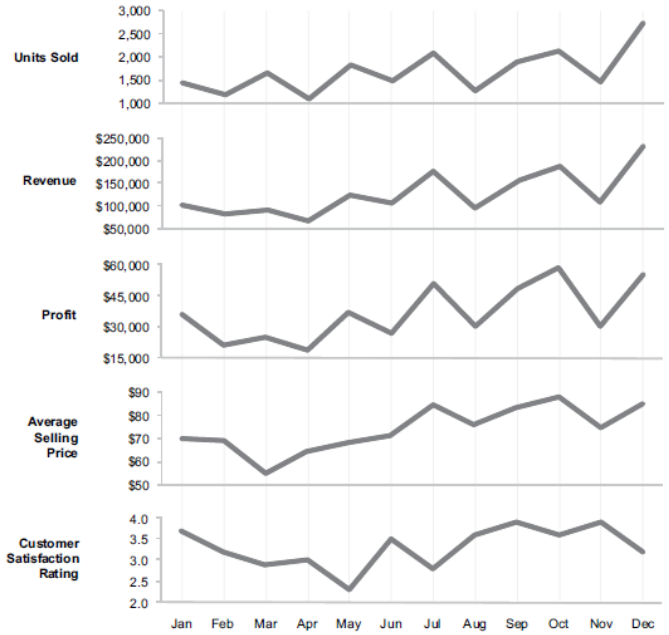
Overlapping

- It is easier to detect cyclical patterns by overlapping the multiple cycles on the same chart



Stacking

- If there is a need to compare multiple variables in the same time period but they are expressed using different units of measure, stacking line charts can be used.



Summary

- Time-series data pattern
- How to display time-series data
- The difference between discrete and continuous time-series data
- Techniques and best practices to consider for time-series data

Reference

- Now You See It : Simple Visualisation Techniques for Quantitative Analysis/ Stephen Few, Analytics Press, c2009

  P.143 – P.188 Chapter 7 Time-Series Analysis

  - All the images are extracted from the book unless it is explicitly stated

- Visualize This / Nathan Yau, Wiley, c2011

Reference

Most materials (image, figure, content) are from Interactive Data Visualisation by Matthew Ward, Georges Grinstein and Daniel Keim © A K Peters, Ltd.

- Chapter 1
- Chapter 4
- Chapter 12