



공정모형 및 해석

유준

부경대학교 화학공학과



Statistics?

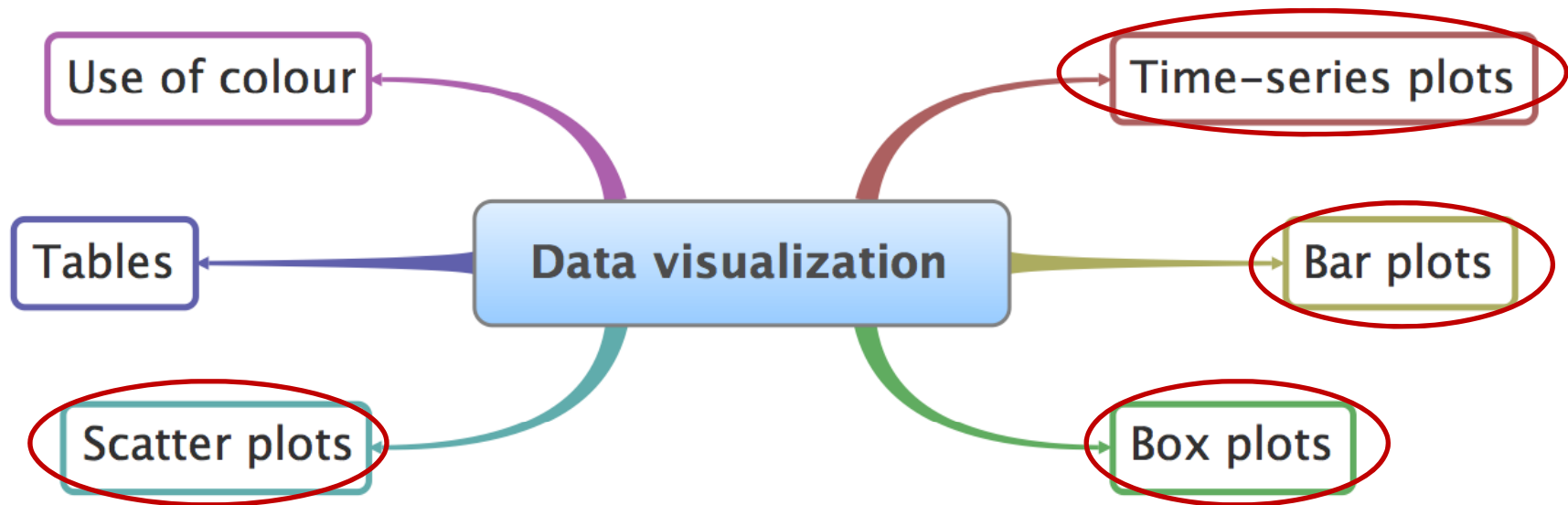
- ➔ Collection of tools for data analysis.
- ➔ Deals with the **collection**, **presentation**, **analysis**, and **use** of data to **make decisions**, **solve problems**, and **design products and processes**.
- ➔ Extracts information from data and combines all information to get knowledge.



Data visualization

➤ Plot your data – the first step of data analysis

➤ Tools



And many more (e.g., pareto chart, histogram)

Reading: Minitab manual p37-70 (or use Minitab “help”)

Usage examples

- Co-worker: Here are the yields from a batch process for the last 3 years (1256 data points), can you help me:
 - understand more about the time-trends in the data?
 - efficiently summarize the batch yields?
- Manager: effectively summarize the (a) number and (b) types of defects on 17 aluminum grades for the past 12 months
- Yourself: 24 different measurements vs time (5 readings per minute, over 300 minutes) for each batch of methadone we produce; how can we visualize these 36,000 data points?
- Life insurer: understand relationship between background education life expectancy, with data on 3,500 staff

Background

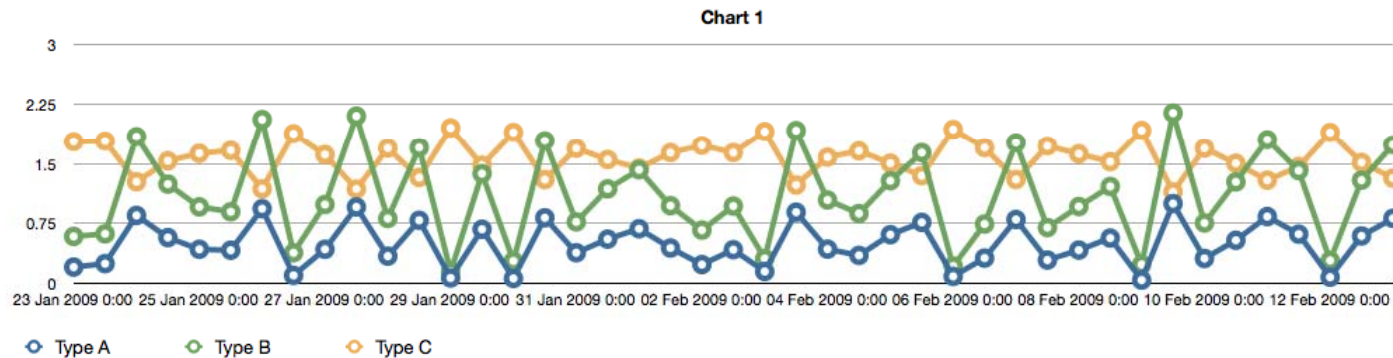
- This class might seem too easy, too obvious. It is!
 - The human eye and brain are excellent at pattern recognition, sorting through signal and noise.
 - We can easily cope with bad plots; but good plots save time and show a clearer, more honest picture.
 - Clichés: "Let the data speak for themselves", "Plot the data"
 - We will look at: how

Time-series plots

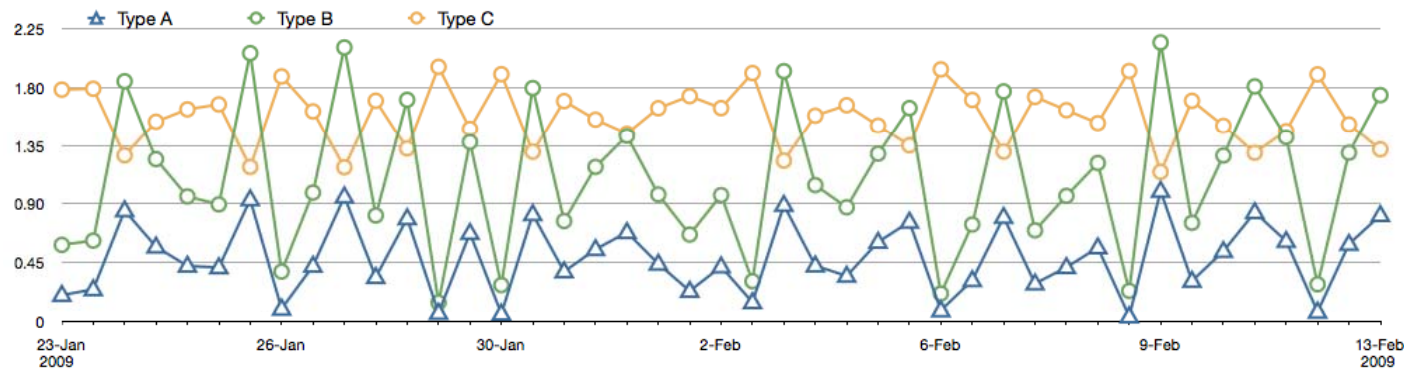
- It is a 2-dimensional plot:
 - (usually) horizontal x-axis: time or sequence order
 - other axis: the data values
- Univariate plot
- Our eyes can deal with high data density:
 - sinusoids
 - Spikes
 - Outliers
 - separate noise from signal

Time-series plots (cont.)

➡ Multiple lines (trajectories): had better not to cross and jumble

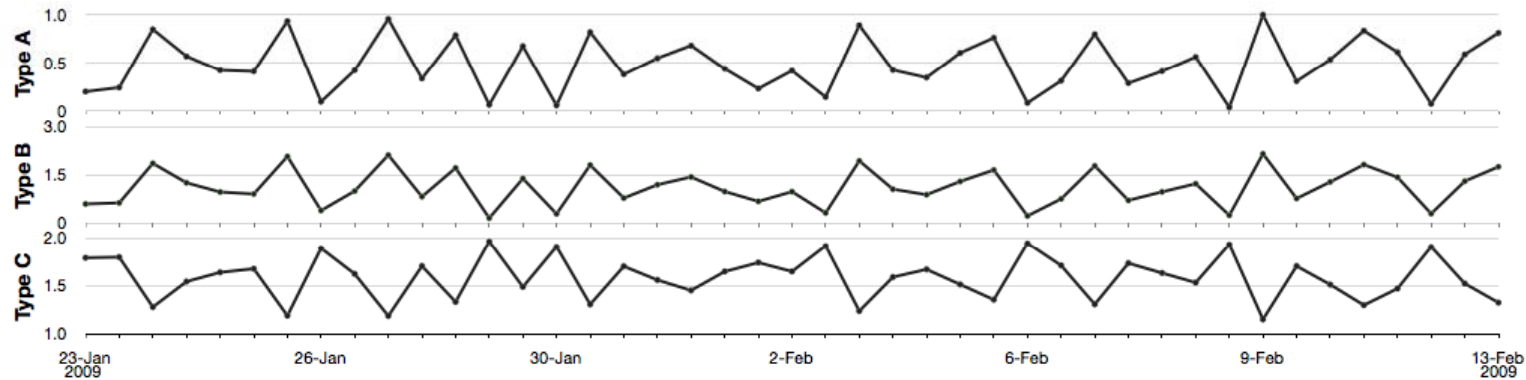


➡ Colours and markers help only slightly



Time-series plots (cont.)

➡ Use separate, parallel axes rather



These non-default settings can take a long time to set.

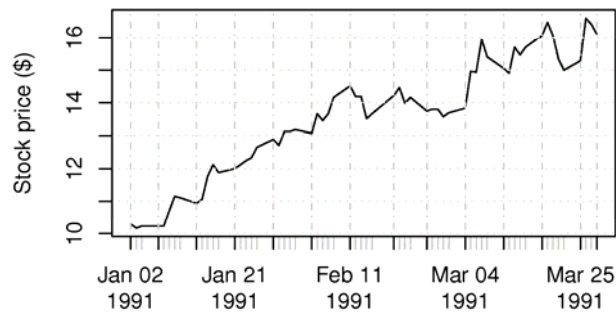
Think about if you were a plant engineer who reports this to a plant manager daily basis.

Time-series plots (cont.)

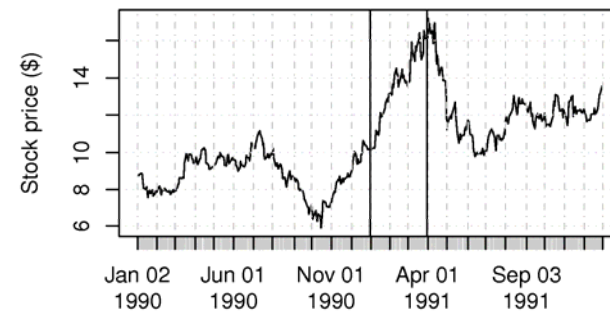
Example: [newmarket.mtw](#)

➡ Show reasonable amount of data for context

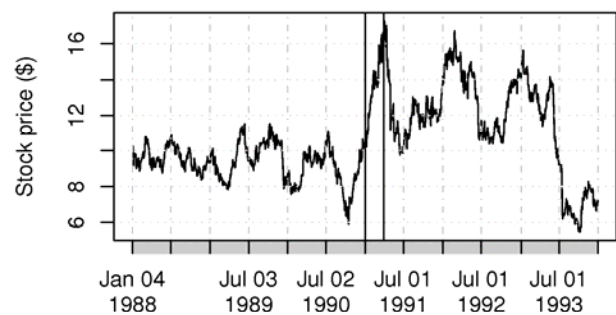
1. Got to buy some of this stock!



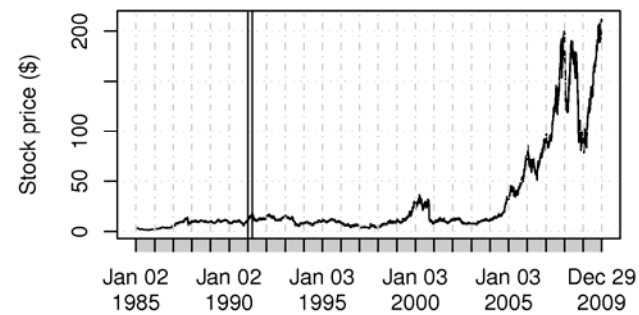
2. But, here is some more context



3. And, even further context

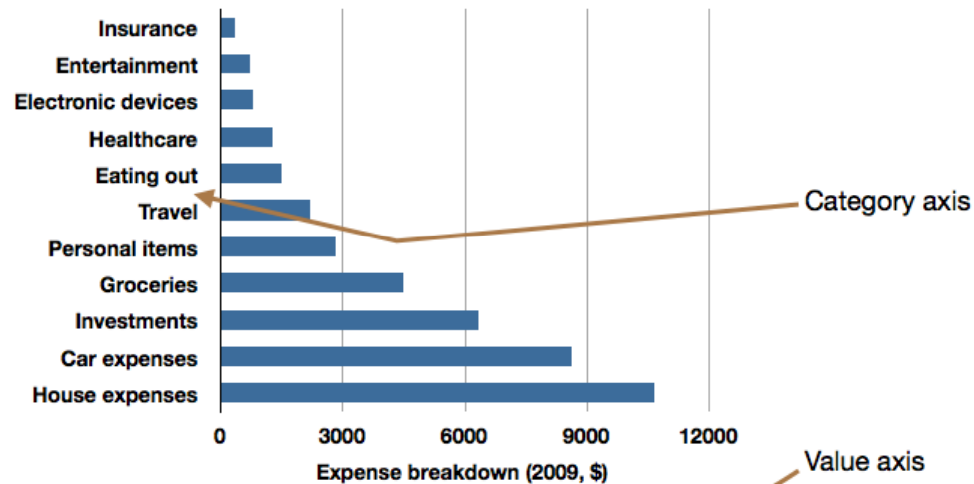


4. To finish: all available data



Bar plots

- A univariate plot on a two dimensional axis.
- Has a category axis and value axis

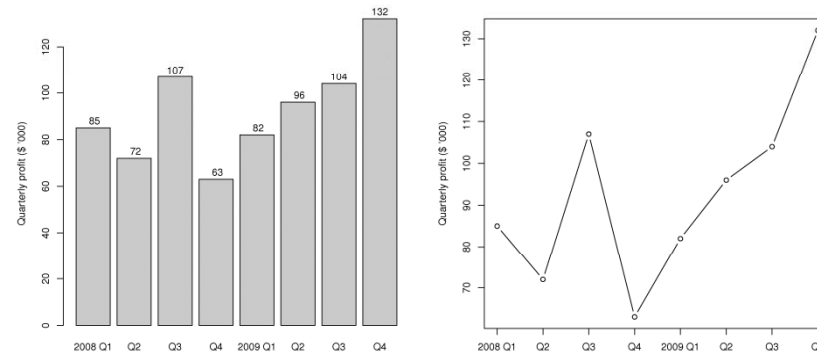


- Use a bar plot when:
 - many categories
 - interpretation does not change if category axis is reordered

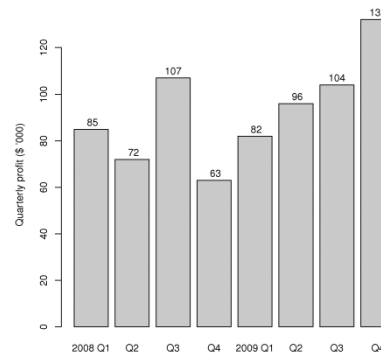
Bar plots (cont.)

➤ Further tips

- Rather use a time-series plot if the data have a sequence.



- Bar plots can be wasteful as each data point is repeated several times.



Box plots

- A graphical display of the "5-number summary" for one variable
 - ① minimum sample value (or $Q3 - 1.5(Q3 - Q1)$ in Minitab)
 - ② 25th percentile (1st quartile)
 - ③ 50th percentile (median)
 - ④ 75th percentile (3rd quartile)
 - ⑤ maximum sample value (or $Q3 + 1.5(Q3 - Q1)$ in Minitab)

Notes:

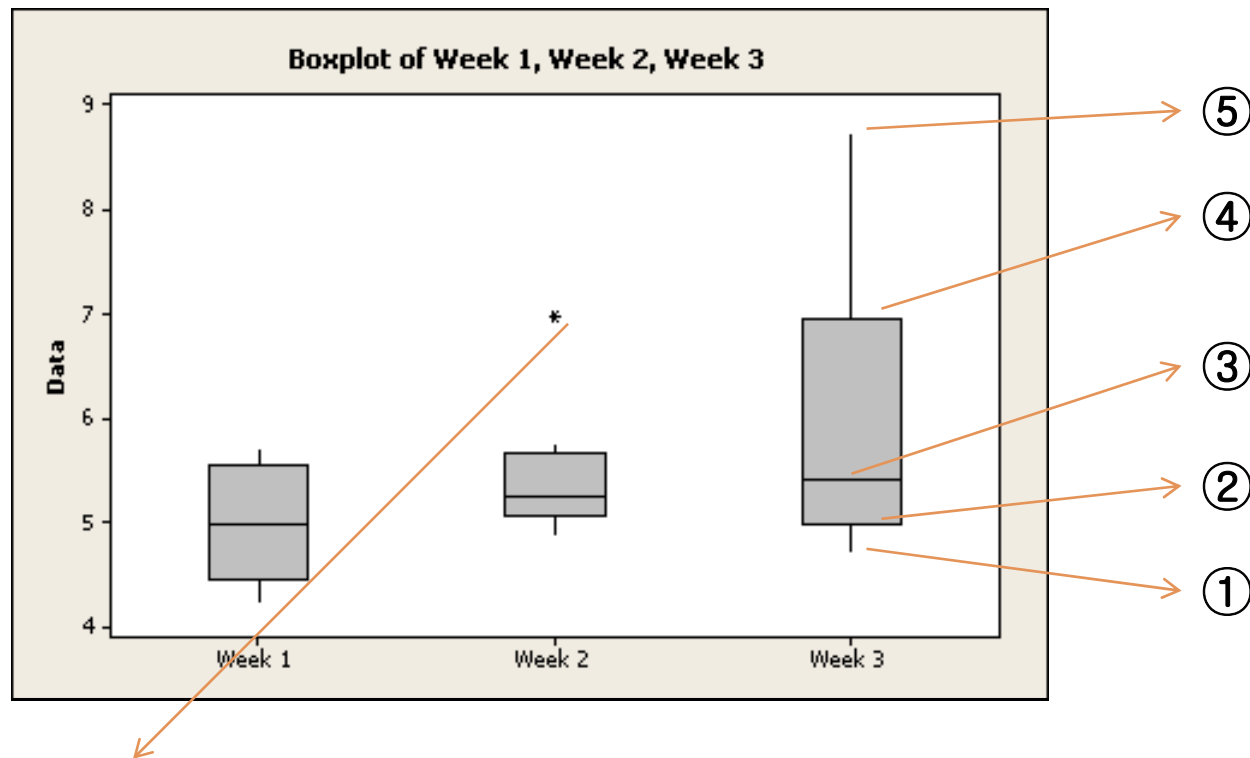
1. 25th percentile is the value below which 25 % of the observations in the sample are found
2. distance from 3rd to 1st quartile = interquartile range (IQR)

Box plots are effective for visualizing overall information of variables (without using basic statistics)

Box plots (cont.)

Example: pipe.mtw

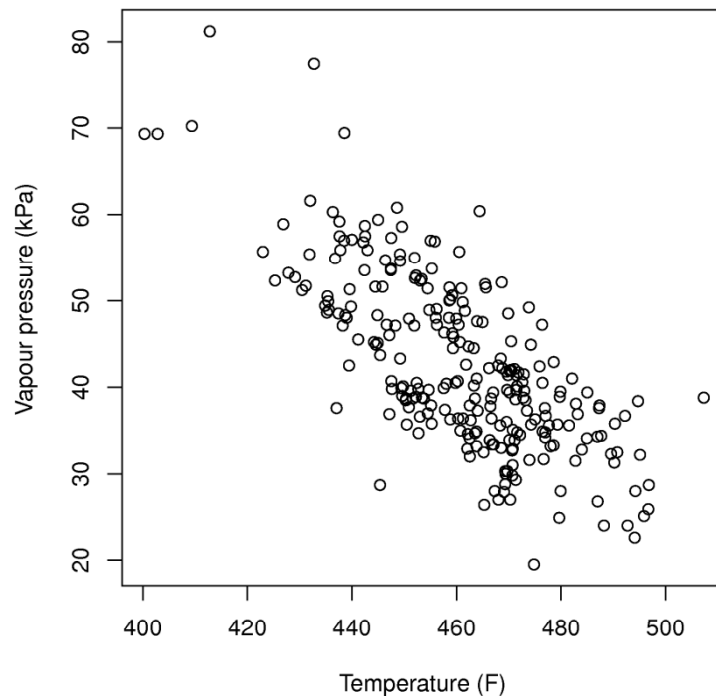
➔ Example



outliers shown as asterisk (*), most commonly defined as above ⑤ or below ①.

Scatter plots

- ➔ Used to help understand the relationship between **two variables**: a bivariate plot
- ➔ Collection of points in the 2 axes
- ➔ Each point is the intersection of the values on each axis



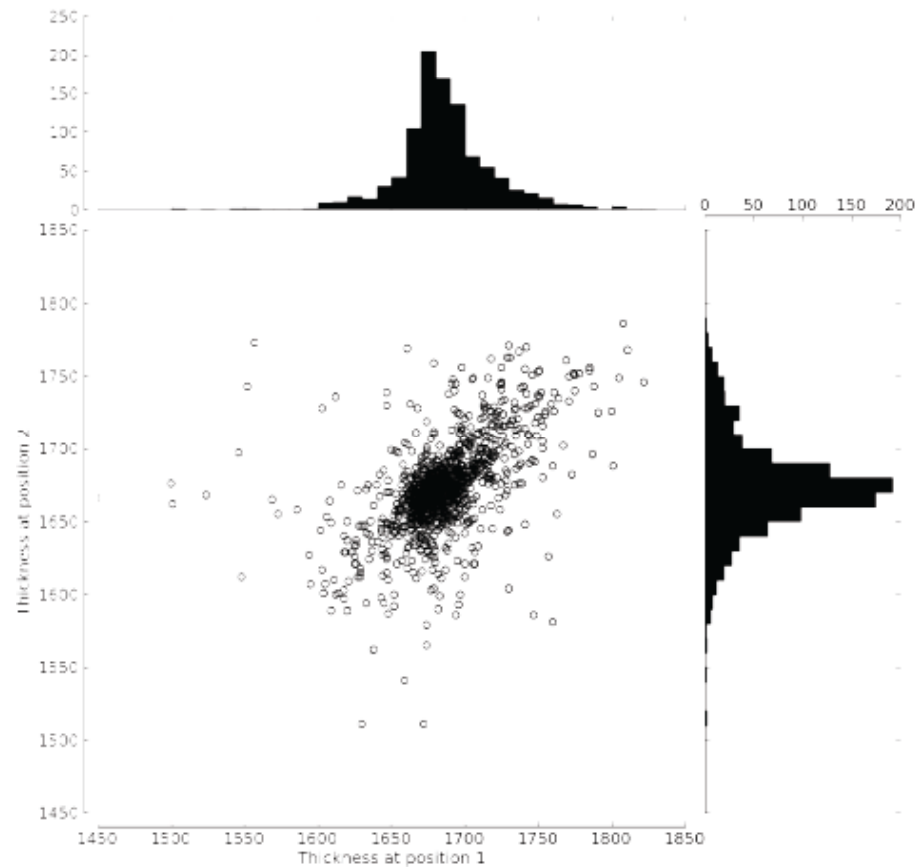
Intention of a scatter plot:
Asks the viewer to draw some relationship between the two variables

Example: [batteries.mtw](#)

Scatter plots (cont.)

➤ Variations

➤ Add box plots or histograms to aid interpretation:

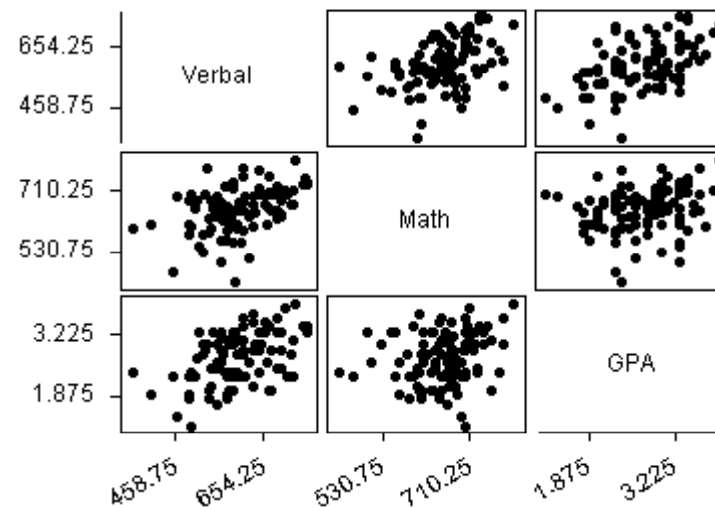


Scatter plots (cont.)

➤ Variations

➤ With **more than three variables**

ex. verbal vs. math, verbal vs. GPA, math vs. GPA

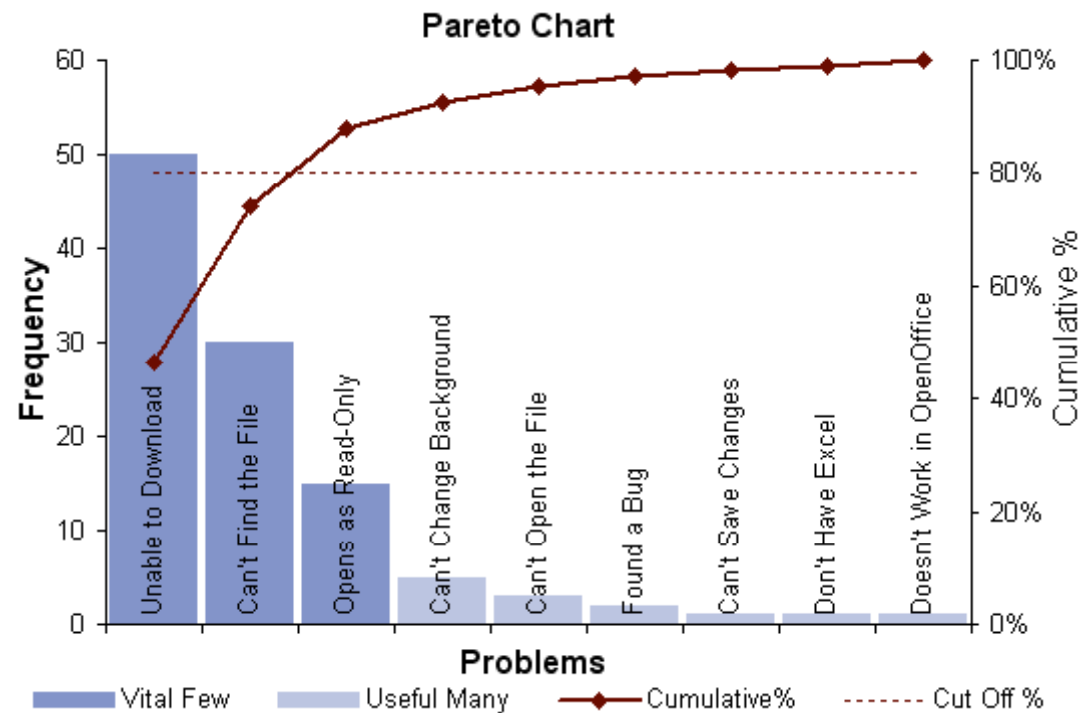


“Matrix plot” in Minitab

Pareto charts

➤ Named after economist Vilfredo Pareto

➤ The purpose of the Pareto chart is **to highlight the most important among a (typically large) set of factors.**



[FYI] Pareto principle (By Joseph Juran, an MBA guy)

➔ For many events, roughly 80% of the effects come from 20% of the causes.

➔ Also known as:

➔ 80-20 rule, the law of the vital few, and the principle of factor sparsity

➔ Example 1

Distribution of world GDP, 1989

Quintile of population	Income
Richest 20%	82.70%
Second 20%	11.75%
Third 20%	2.30%
Fourth 20%	1.85%
Poorest 20%	1.40%

➔ Example 2 (a common rule of thumb in business)

"80% of your sales come from 20% of your clients"

➔ Further reading: http://en.wikipedia.org/wiki/Pareto_principle

[FYI] Seven basic quality tools

➤ Seven “indispensable” tools of quality

1. Cause-and-effect diagram
2. Check sheet
3. Control charts
4. Histogram
5. Pareto chart
6. Scatter plot
7. Stratification (separating data gathered from various sources)

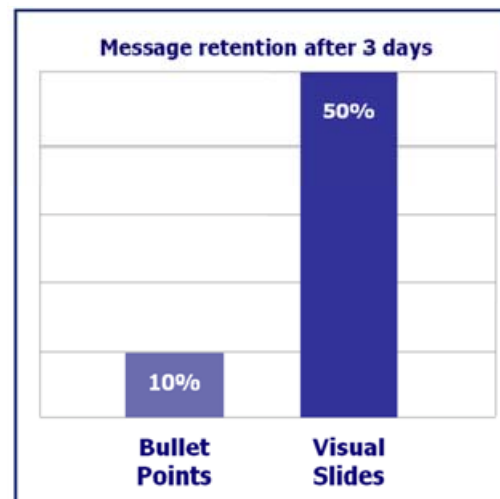
➤ Further reading: <http://www.asq.org/learn-about-quality/seven-basic-quality-tools/overview/overview.html>

[Tips] Presentation skills

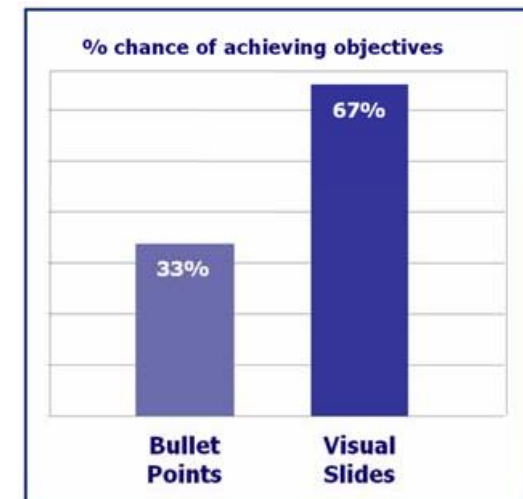
- ➔ Three essential presentation skills
 1. **Use visual aids where you can**
 2. Rehearse, rehearse, rehearse
 3. The audience will only remember three messages



Information take from presentations



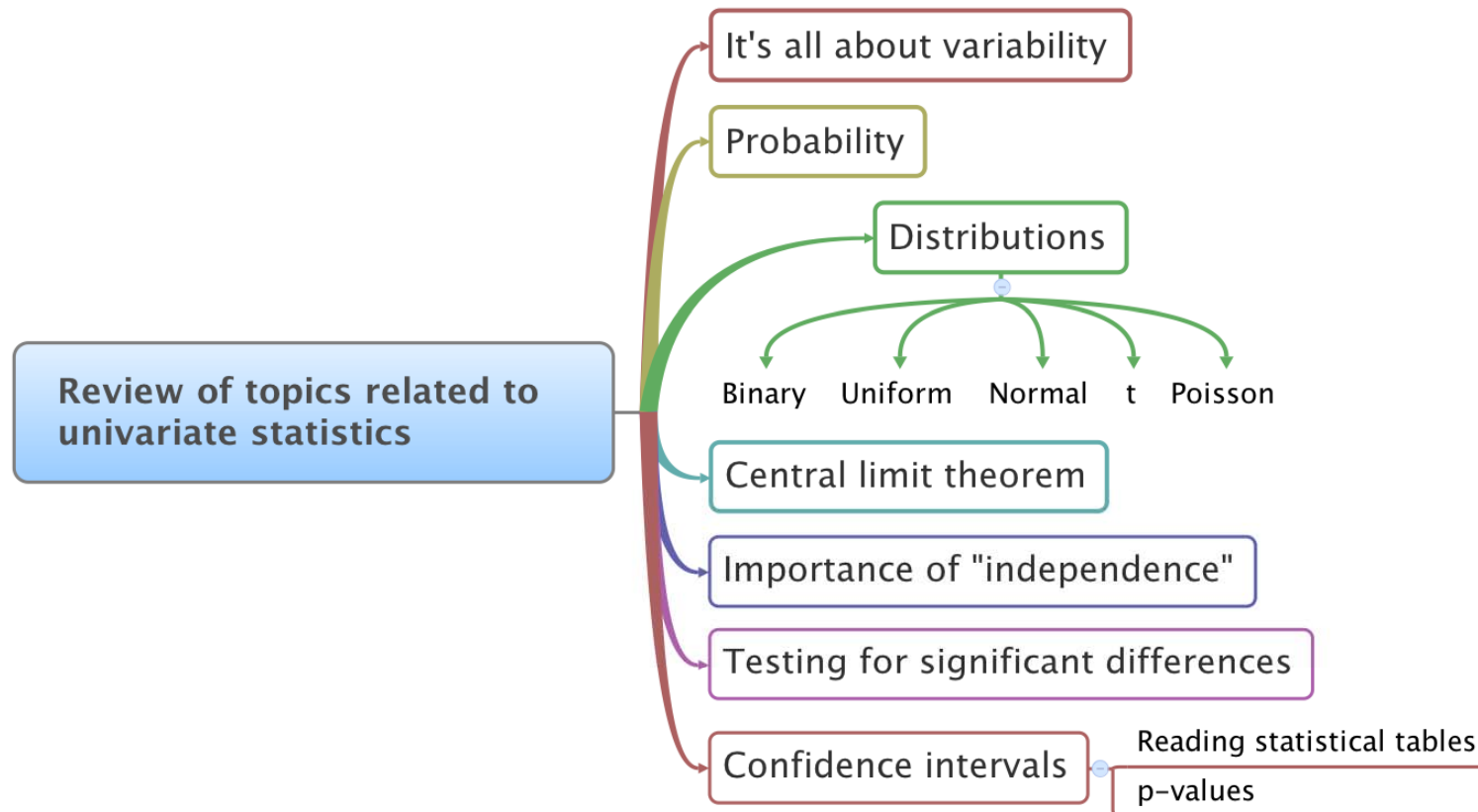
Visual slides last longer.



Visual slides achieve more.

Review of statistics and more

➔ In next class,



[숙제 #2]

- QC 7 tools(7가지 품질관리도구)에 관한 보고서를 작성하여 제출할 것.
 - (공백/도표/그림 제외) 5,000자 이내로 작성할 것
 - 7가지 중 3가지 이상에 대한 예제도 포함 시킬 것.
 - 제출기한: ~9/16