Topic 8

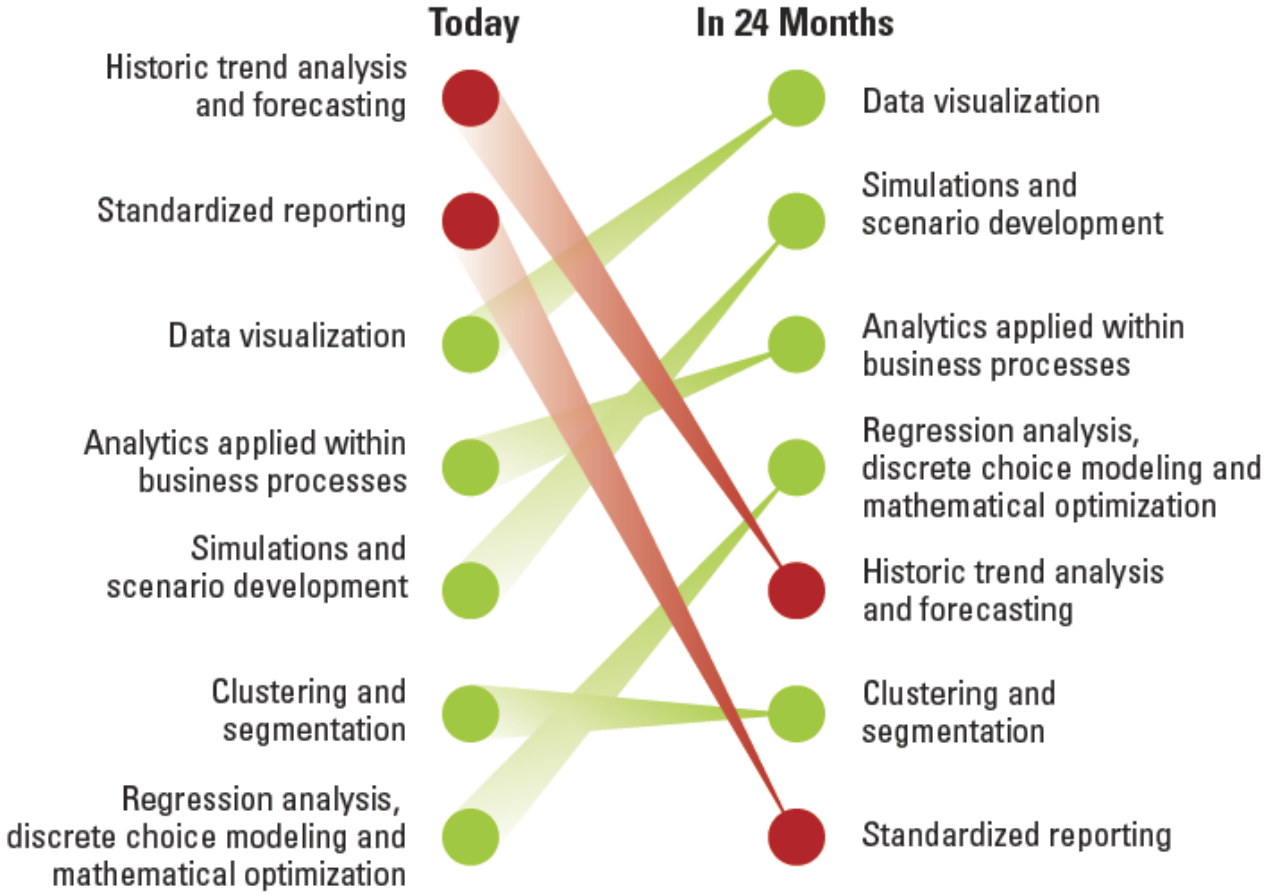
Data Visualisation and Design Principles

Most Valuable

Tableau, Qlikview, Spotfire and Microsoft BI lead the pack. Pentaho offers a solid functionality to business users to effectively visualise and analyse their enterprise data.

WHERE ARE DATA-DRIVEN MANAGERS HEADED?

Organizations expect that the ability to visualize data differently will be the most valuable technique in two years. Other techniques and activities that are currently delivering the most value today will still be done, but will be of less value.



MITSloan Management Review

Respondents were asked to identify the top three analytic techniques creating value for the organization, and predict which three would be creating the most value in 24 months.

Role of Data Visualisation

Data visualisation in Business Analytics serves two main purposes

Assist in analysis (as an exploratory tool)

- Origins in exploratory data analysis (EDA)
- Has gained prominence in recent years with the rise of Big Data and Real-time Web
- Visualisation techniques has evolved to include **maps and hierarchal data and features interactive and drill-down facilities**

Assist in communication of insight (as a way of telling a story)

- Communicate something new about the underlying patterns and relationships contained within data
- Used for decision making – thus must be honest and unambiguous
- Works best when someone understands not only the data, but also the principles of design as well as visual communication

Anscombe’s quartet

Anscombe’s Quartet shows the importance of visualising data.

It comprises four datasets that have identical statistics but looks very different when visualised.

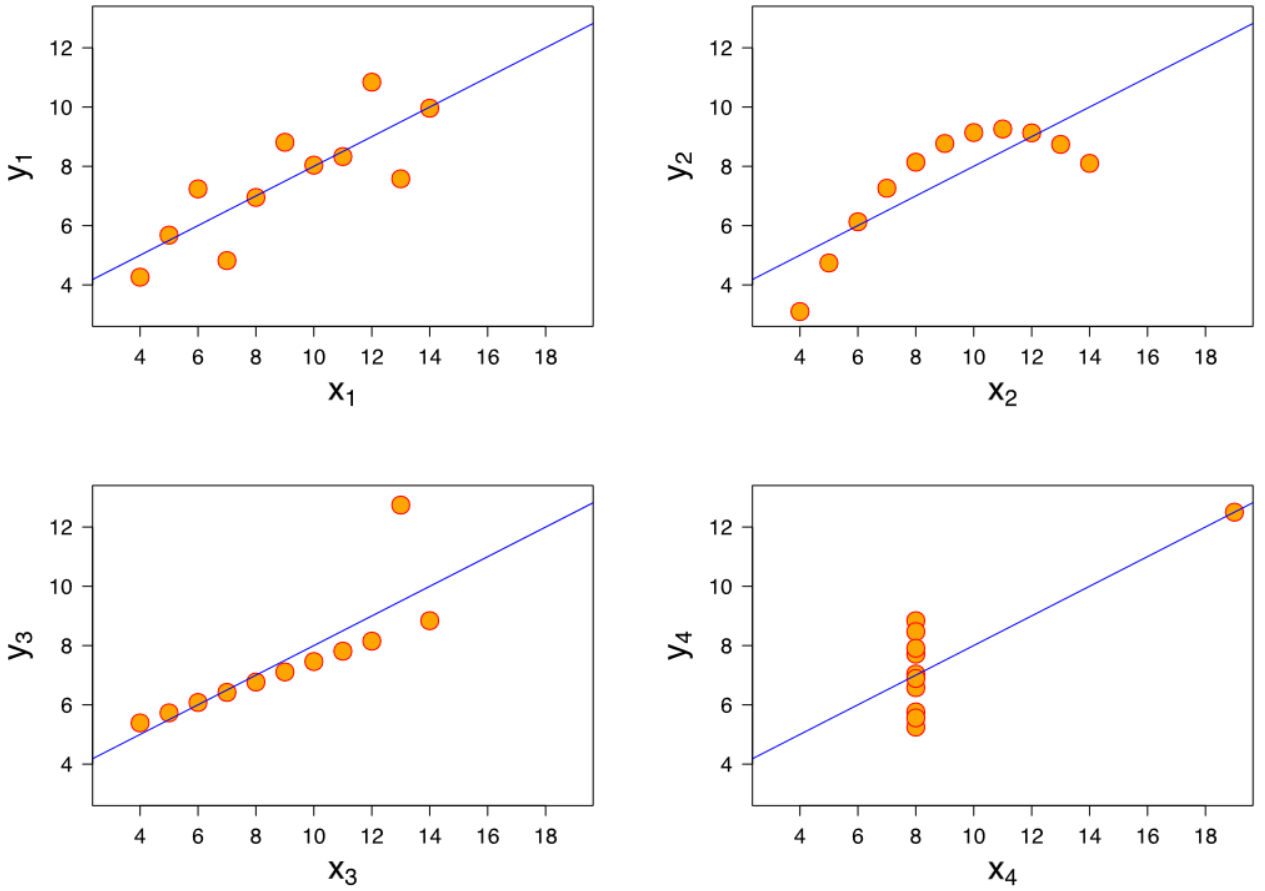
Anscombe's quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Anscombe's quartet …

Each of the values below is the same for each dataset

- Average X = 9
- Average Y = 7.5
- Regression Line :
- y = 3 + 0.5x
- Correlation Coefficient = 0.816



As an exploratory tool

As an exploratory tool

- Visualisation provides an ability to comprehend huge amounts of data
- Visual exploration of data to discover patterns and relationships that are not readily visible
- Visualisations often enable problems with data to become immediately apparent
- Use many visualisations for exploration
- Mostly reliant on the cognitive skills of the analyst
- Data quality issues?
- Will be spending most of the time preparing data

Seeking Analytical patterns

- We usually look for following analytical patterns when looking at

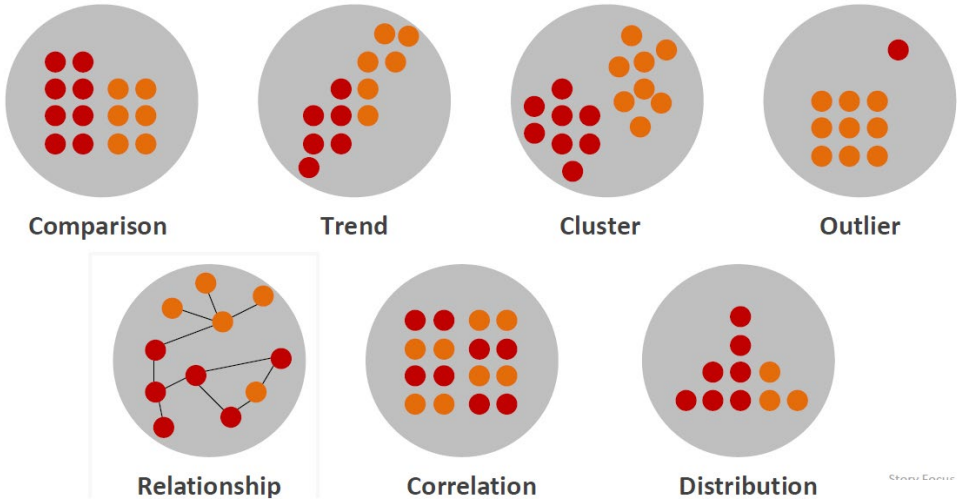
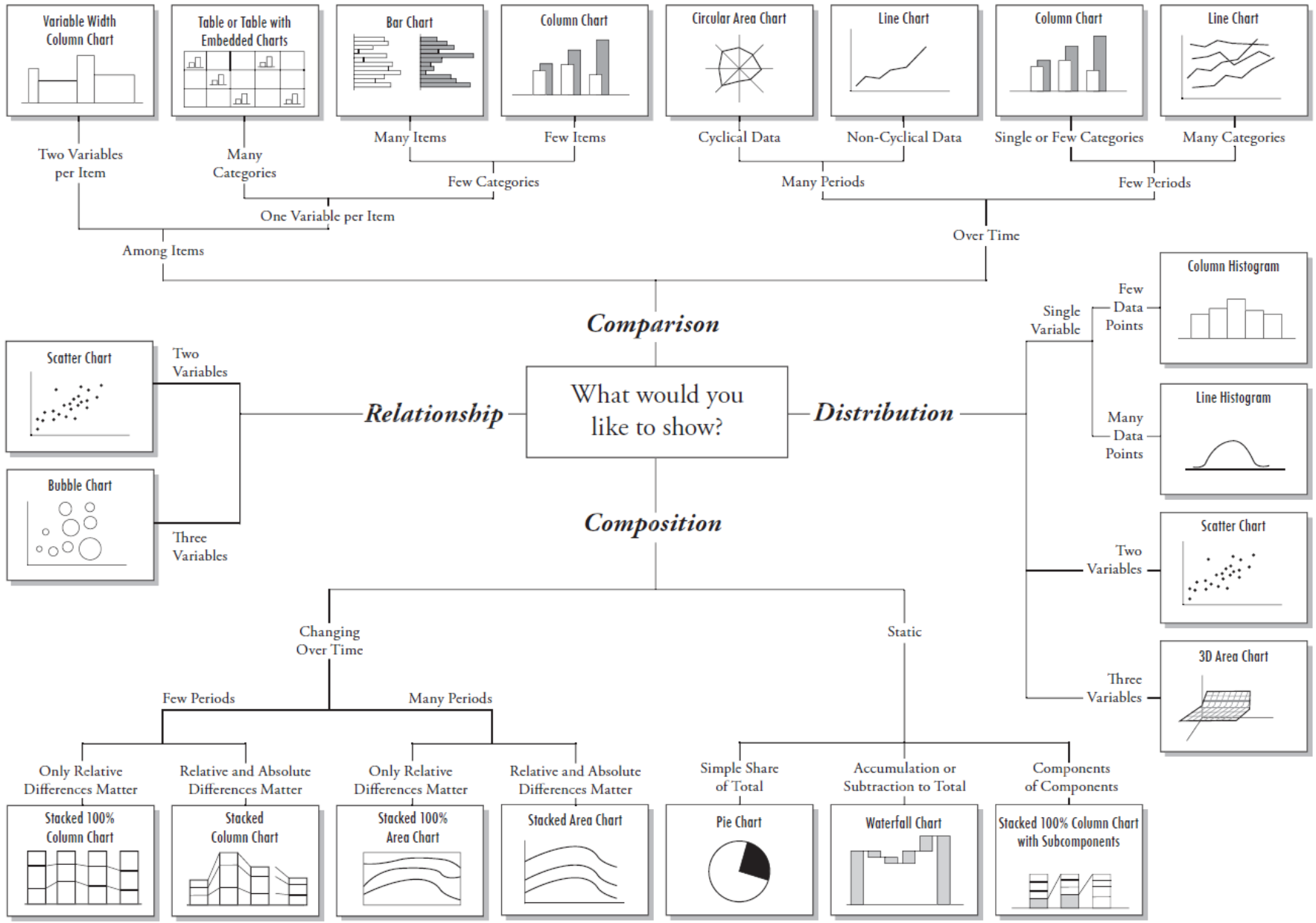


Chart Suggestions—A Thought-Starter



Common Visualisations (univariate and bivariate)

- Bar charts: comparing data across categories
- Pie charts: observe a portion of the total (a particular category)
- Line charts: observe trend over time
- Box/Dot plots: view distribution (shape) and unusual data elements
- Histogram: view distribution and unusual data elements (grouping)
- Multiple box plot: comparison to find the difference between groups
- Scatter plots: investigate relationship between variables

Complex Visualisations (multivariate)

- Maps : geo-locations
- Bubble charts
- Heatmaps / Treemaps / Choropleth
- Parallel Set / Mosaic plot
- Area charts
- Decision Trees
- Radar charts

As a way of communicating

Visualisation is used to tell the story or show the patterns that have been discovered in data.

Visualisation focuses on the message so it is clear and easy to understand.

Therefore most of the focus is on the design of the visualisation. It should be honest, unambiguous and an effective presentation of the data.

There are a lot of theories and principles of data visualisation – let us look at a couple of key ideas.

Design – Pre attentive attributes

These attributes immediately catch our eye when we look at a visualisation.

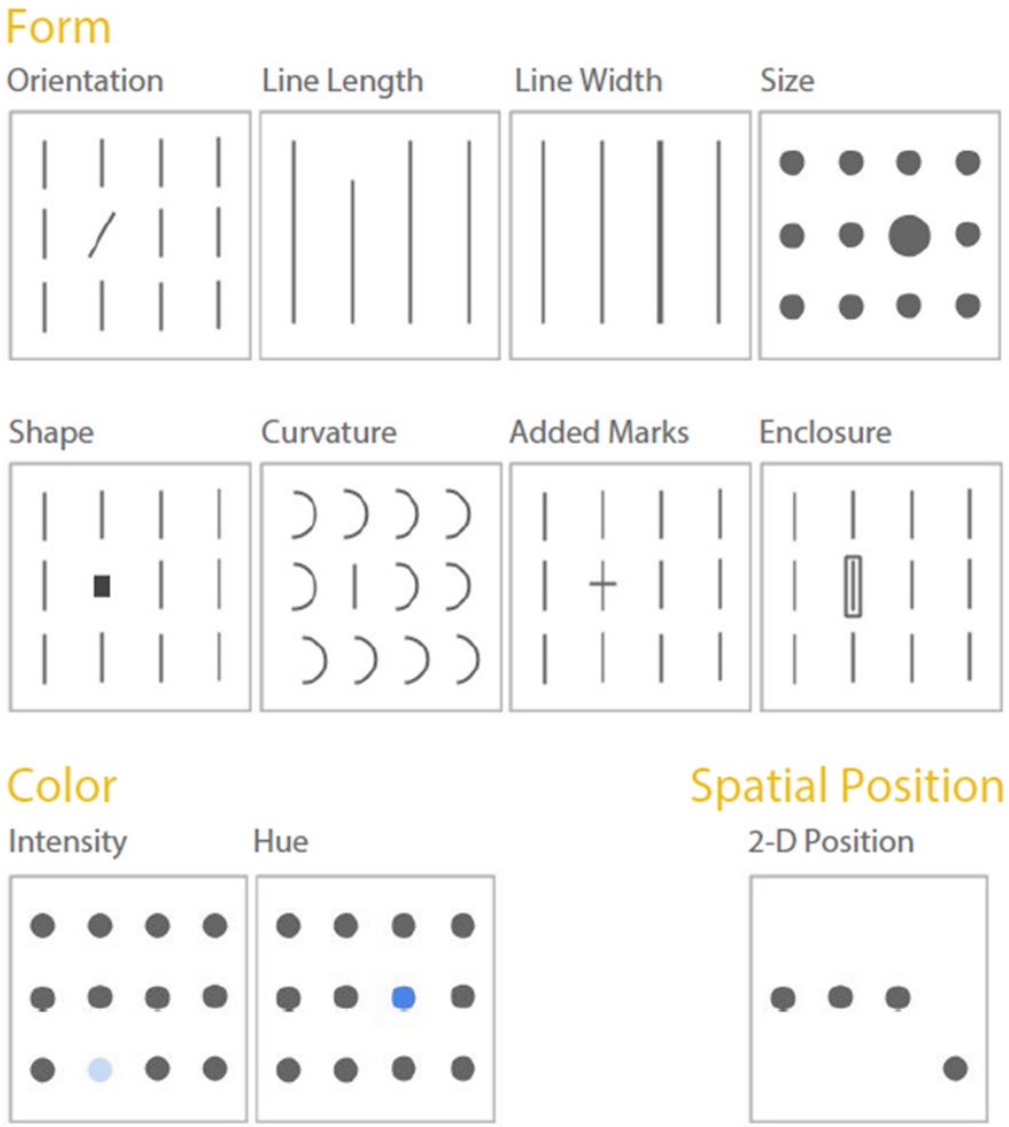
Only “Position” and “Length” can be used to perceive numerical data with precision.

Other attributes are good for perceiving categorical data.

For example, both the pie chart and the bar chart show the same data. But you can’t easily tell from the pie chart which is the biggest pie.

That’s more clearly visible in the bar chart as it calls on the pre-attentive attribute of length.

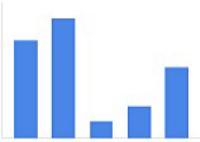
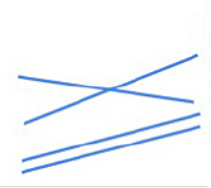
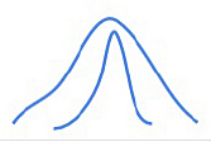
These attributes are the basis for comprehension of patterns in a graphs (Ware 2008).



Analytical patterns

We immediately identify the pre-attentive attribute in a visualisation. We then combine the pre-attentive attribute with analytical patterns in the visualisation (Ware 2008).

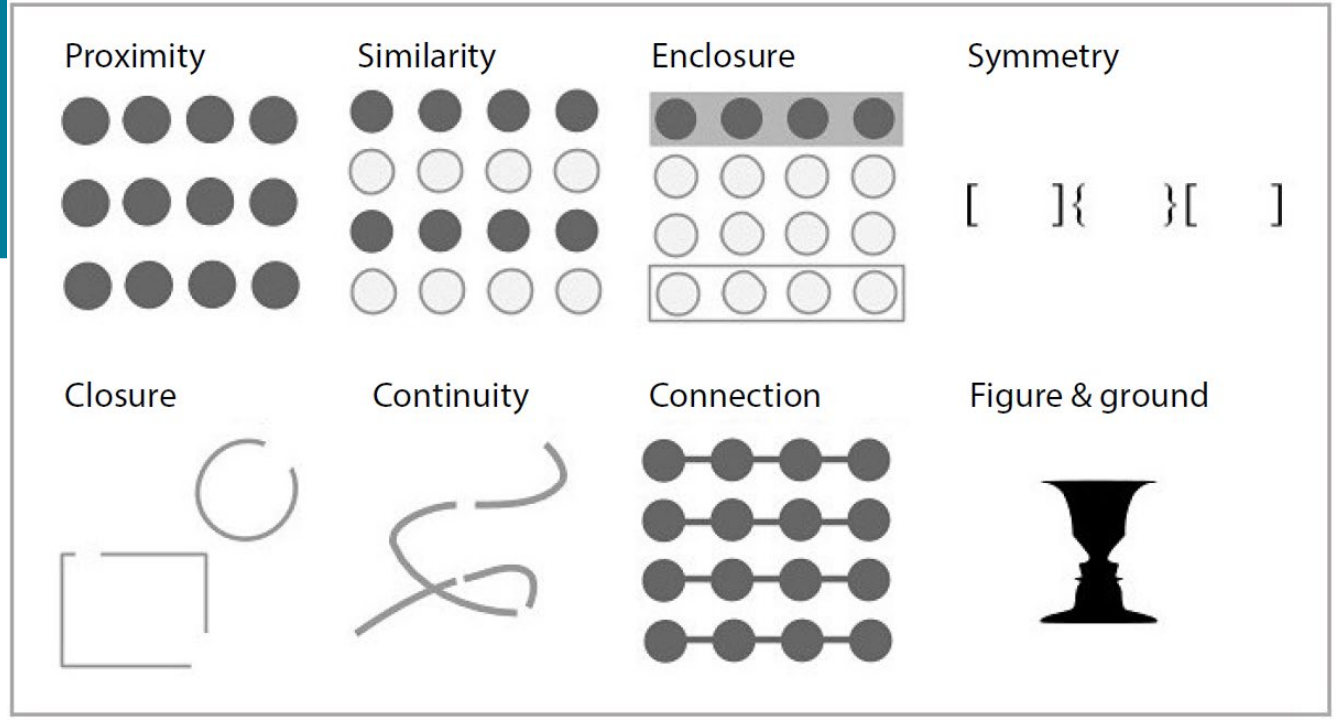
When designing visualisations, we would highlight one pattern over the other.

Pattern	Example	Pattern	Example
High, low and in between		Non-intersecting and intersecting	
Going up, going down and remaining flat		Symmetrical and skewed	
Steep and gradual		Wide and narrow	
Steady and fluctuating		Clusters and gaps	
Random and repeating		Tightly and loosely distributed	
Straight and curved		Normal and abnormal	

Gestalt Principles

Gestalt principles describe how our mind organises individual elements into groups

Highlight patterns that are important



- Proximity: we see three rows of dots instead of four columns because they are closer horizontally than vertically.
- Similarity: we see similar looking objects as part of the same group.
- Enclosure: we group the first four and last four dots as two rows instead of eight dots.
- Symmetry: We see three pairs of symmetrical brackets rather than six individual brackets.
- Closure: we automatically close the square and circle instead of seeing three disconnected paths.
- Continuity: we see one continuous path instead of three arbitrary ones.
- Connection: we group the connected dots as belonging to the same group.
- Figure & ground: We either notice the two faces, or the vase. Whichever we notice becomes the figure, and the other the ground.

Tufte’s principles of graphical display

Graphical displays should:

- show the data
- tell the truth (avoid distorting what the data has to say)
- focus on the content (help the viewer think about the information rather than design)
- encourage the eye to compare the data
- make large data sets coherent
- reveal data at several levels of details
- Closely integrate statistical and verbal descriptions

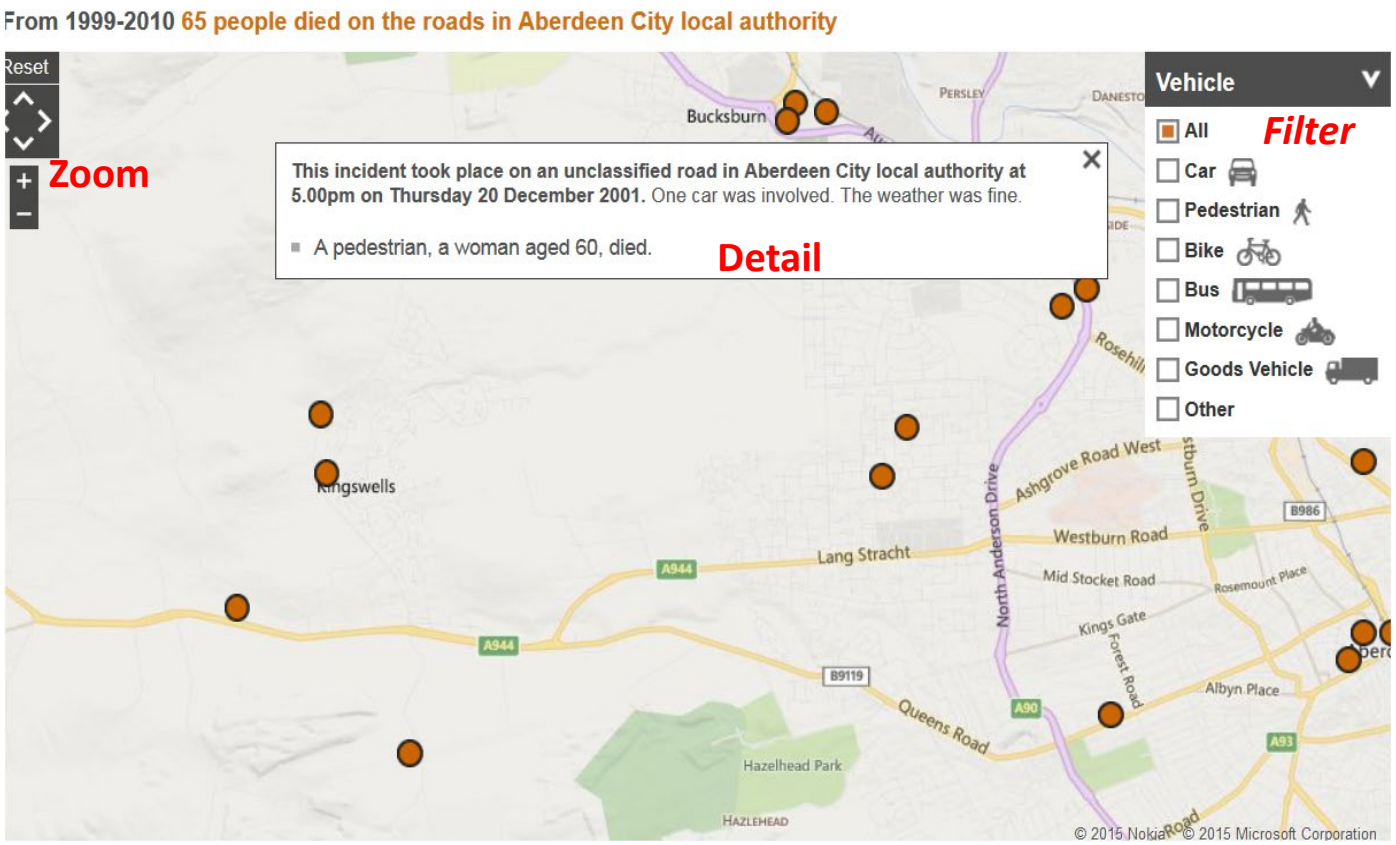
Task Taxonomy [Schneiderman Mantra]

Providing just Data is not enough!

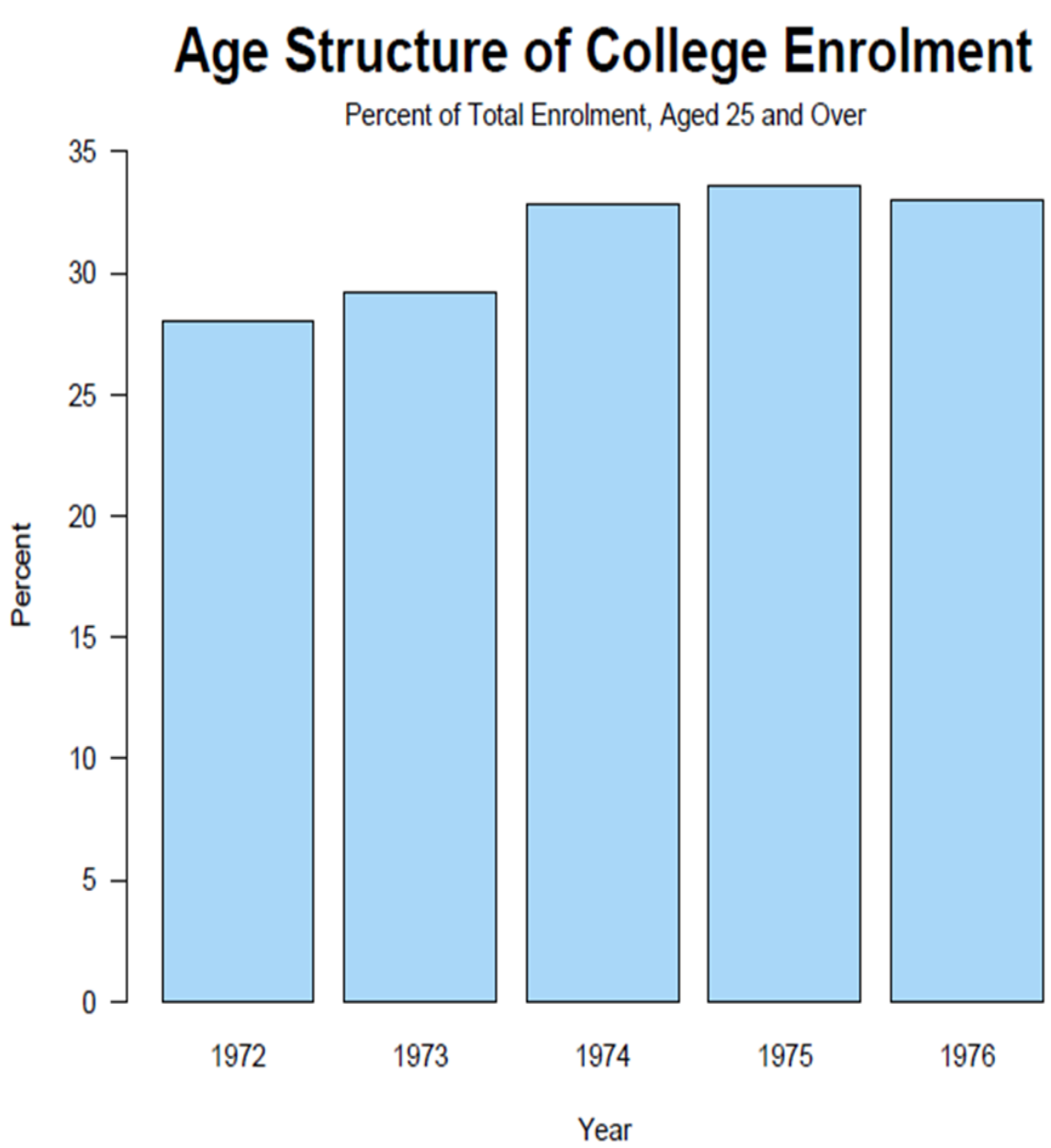
Need to relate to tasks or introduce *interaction*

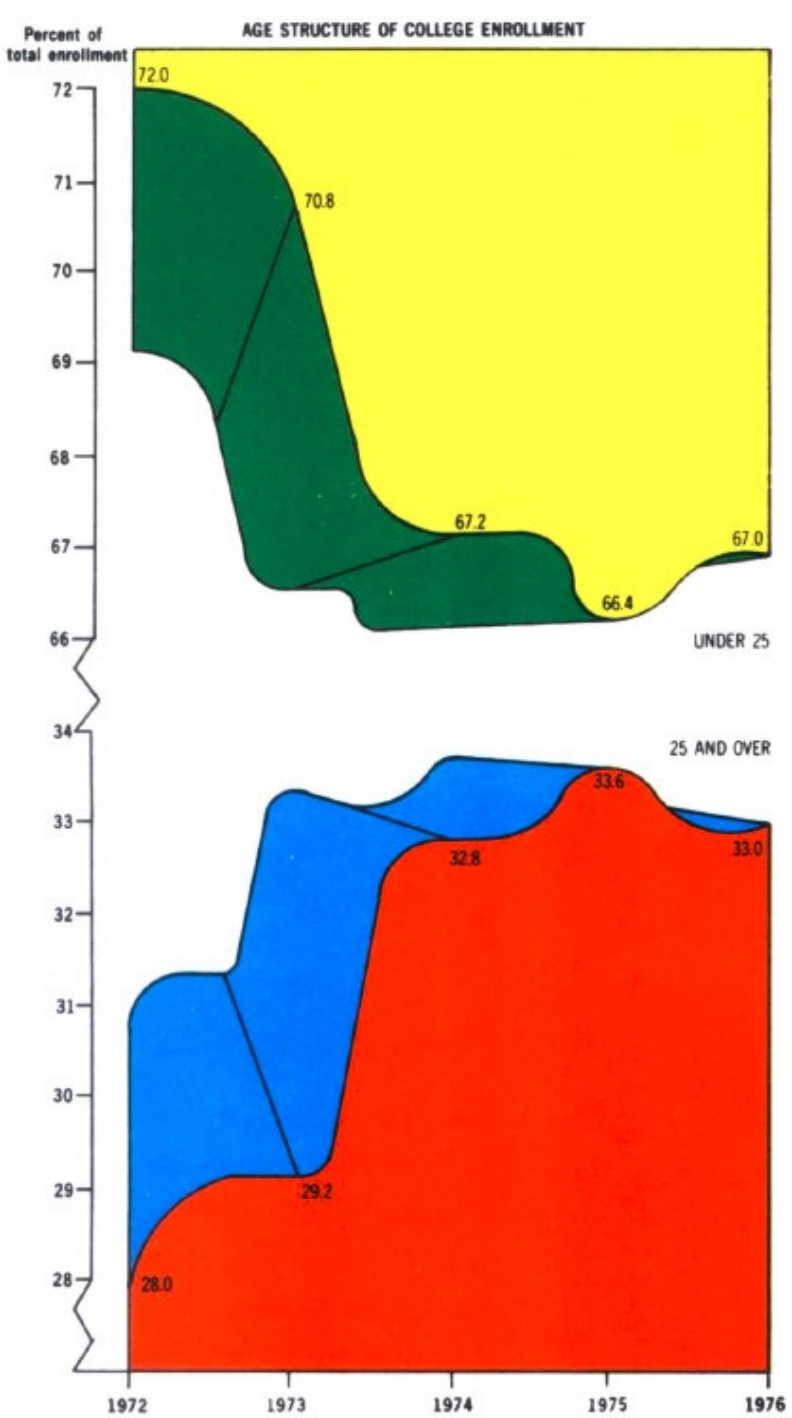
For example

- Overview first,
- Zoom and Filter, then
- Details on demand
- Relate (to view relationships between data points)
- History (let the user undo or redo any action taken while visualizing the data)
- Extract (allows users to visualize a part of the graph in order to focus only on the data that is necessary for immediate use) – e.g. Hovering



Cautionary notes





Cautionary notes…

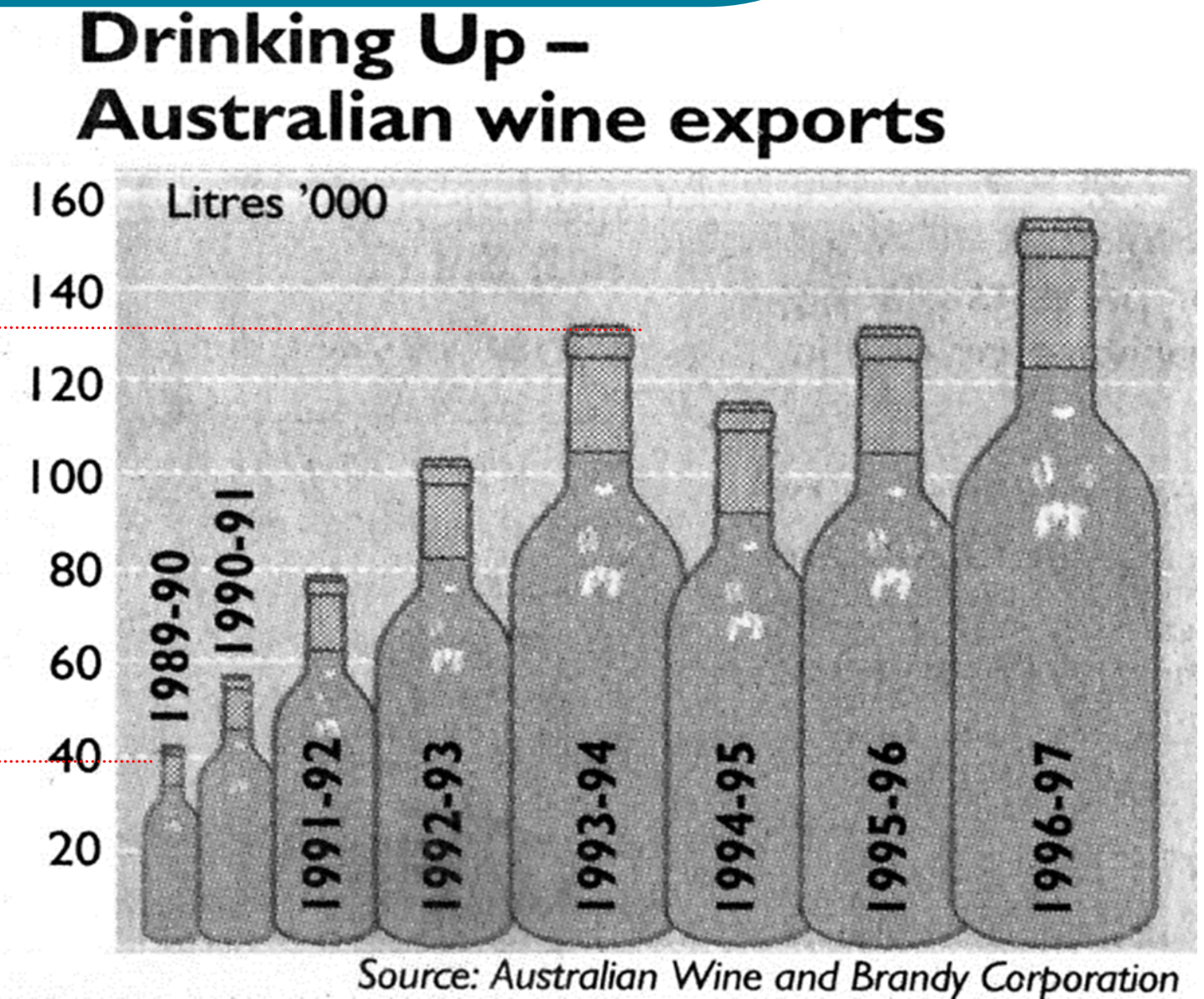
How much larger does the 1993-4 bottle look than the smallest bottle?

About 12 times?

Change from 1989-90 to 1993-94

≈130/40

= 3¼ times

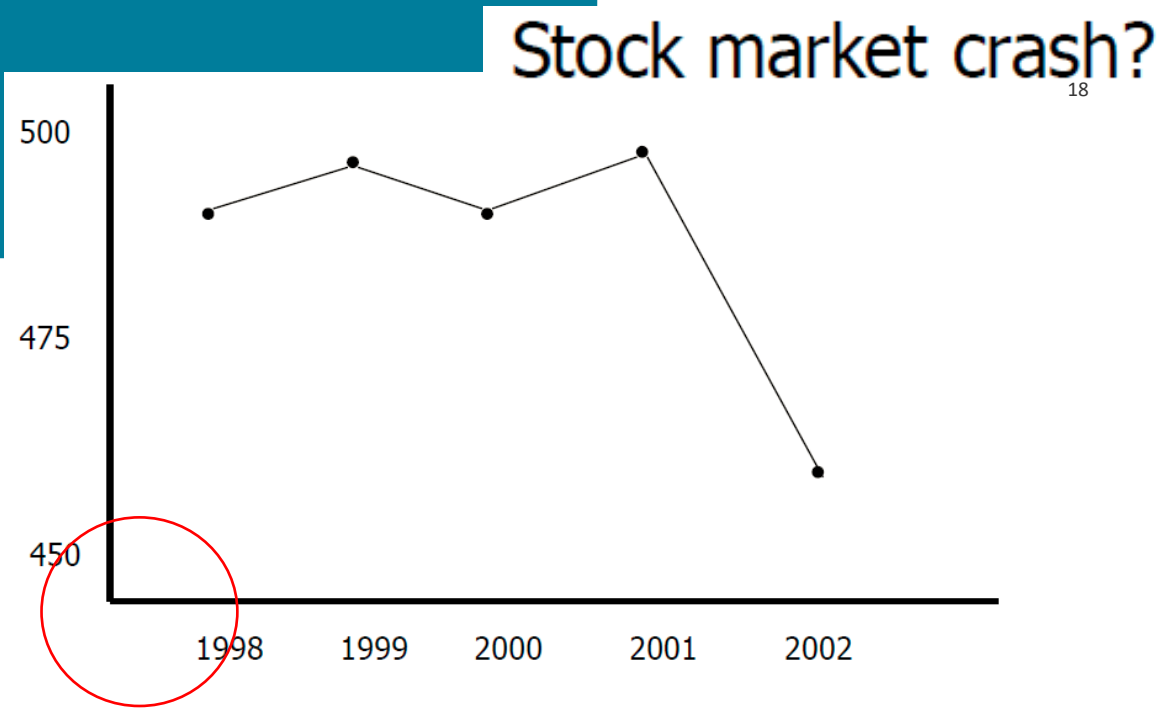


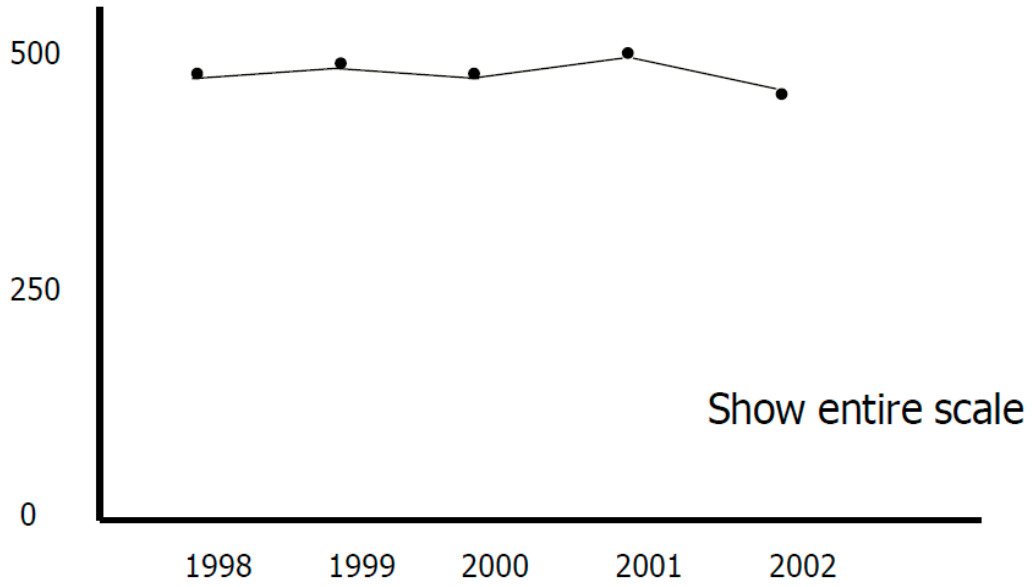
Cautionary notes…

Graphs should not present data out of context.

The graph appears to show an enormous reduction in share prices.

However, when you see the graph in context, the reduction is not so dramatic.

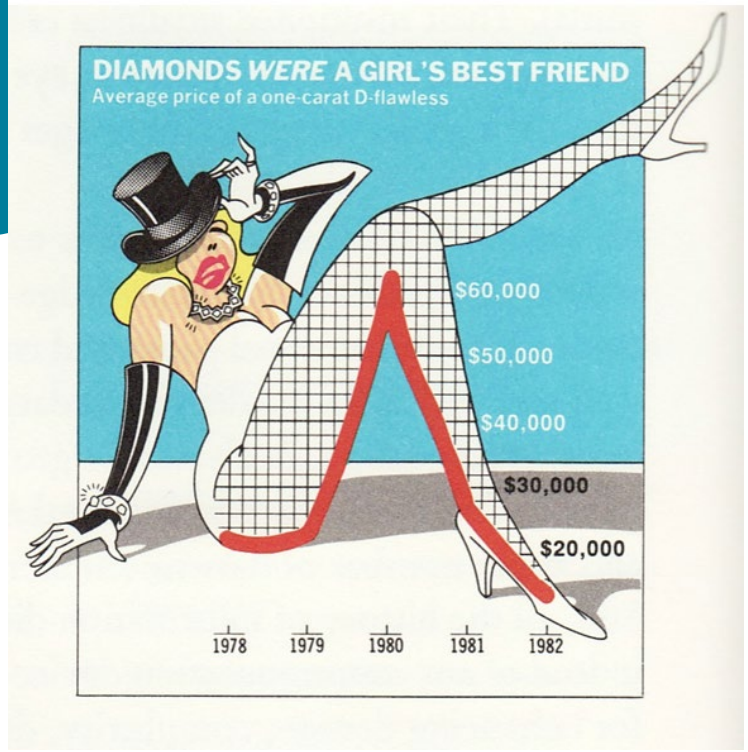




Cautionary notes...

Avoid chart junk (no data link decorations)

- Unintended Optical Art (vibration)
- The Grid
- The Duck: Self Promoting Graphics

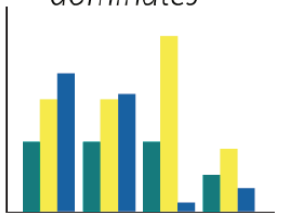


The choice of colour:

- Call attention to specific items
- Distinguish between classes of items
- Proceed with caution
- Less is more
- Representing magnitude is tricky

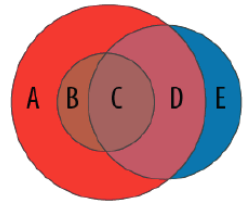
Examples

Poor use of colour



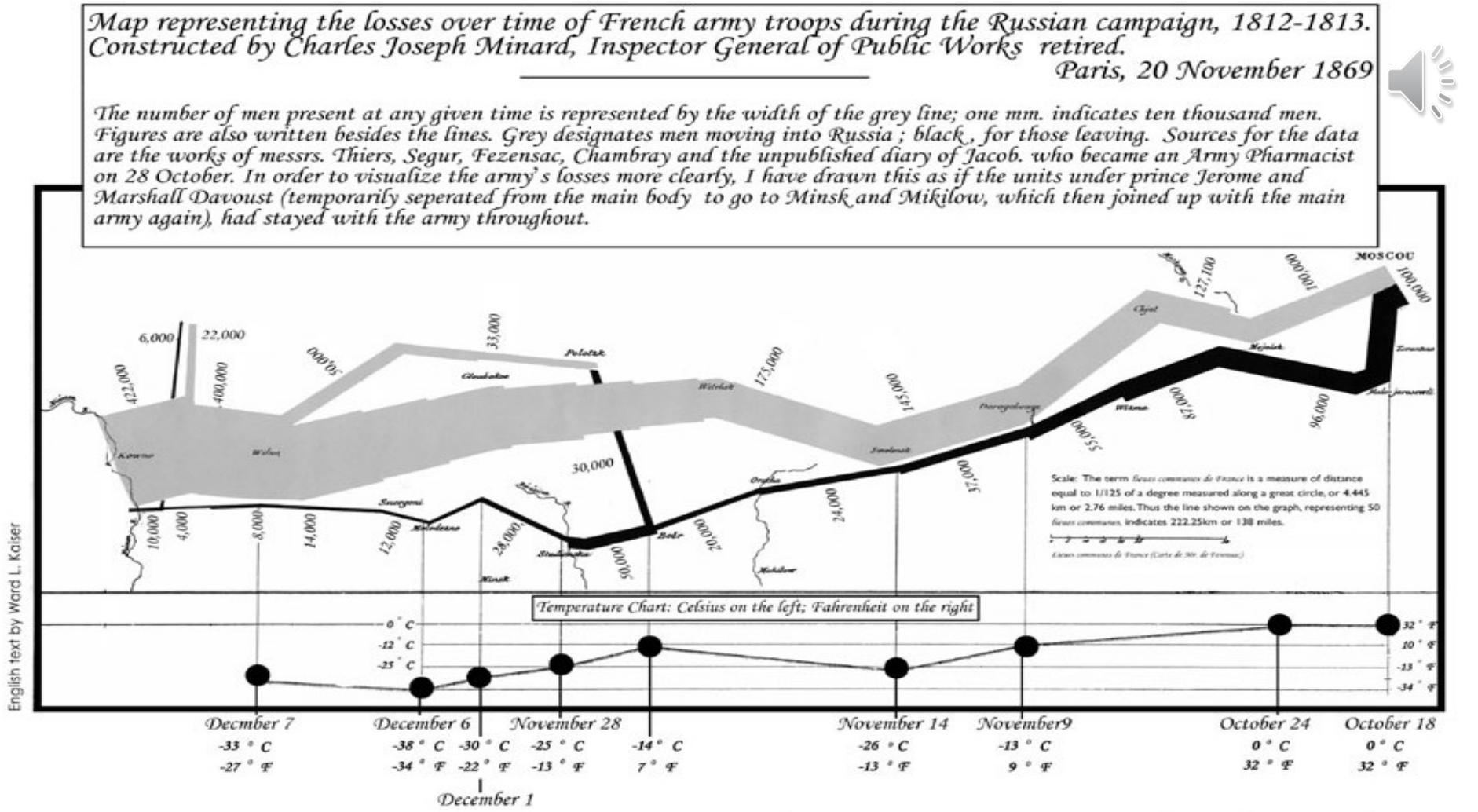
one color dominates

difficult to distinguish



murky

Classic Visualisation



Editor s note: dates & temperatures are only referenced for the retreat from Moscow

Figure 58. ***Minard s map of Napoleon s Russian campaign.***
This graphic has been translated from French to English and modified to most effectively display the temperature data.

Napoleon’s March - Review

Visual comparisons

– Width of the grey and black line gives us an immediate comparison of the size of the Napoleon's army at different times during the campaign.

Show Causality

– Map shows temperature records and some geographical locations. It illustrates that weather and terrain defeated Napoleon as much as his opponents.

Shows multivariate data

– Napoleon's March shows six variables : Army size, location (in two dimensions X and Y), directions, time, and temperature.

Use Direct Labelling

– Integrate words, numbers and images and do not have to go to legend to work it out.

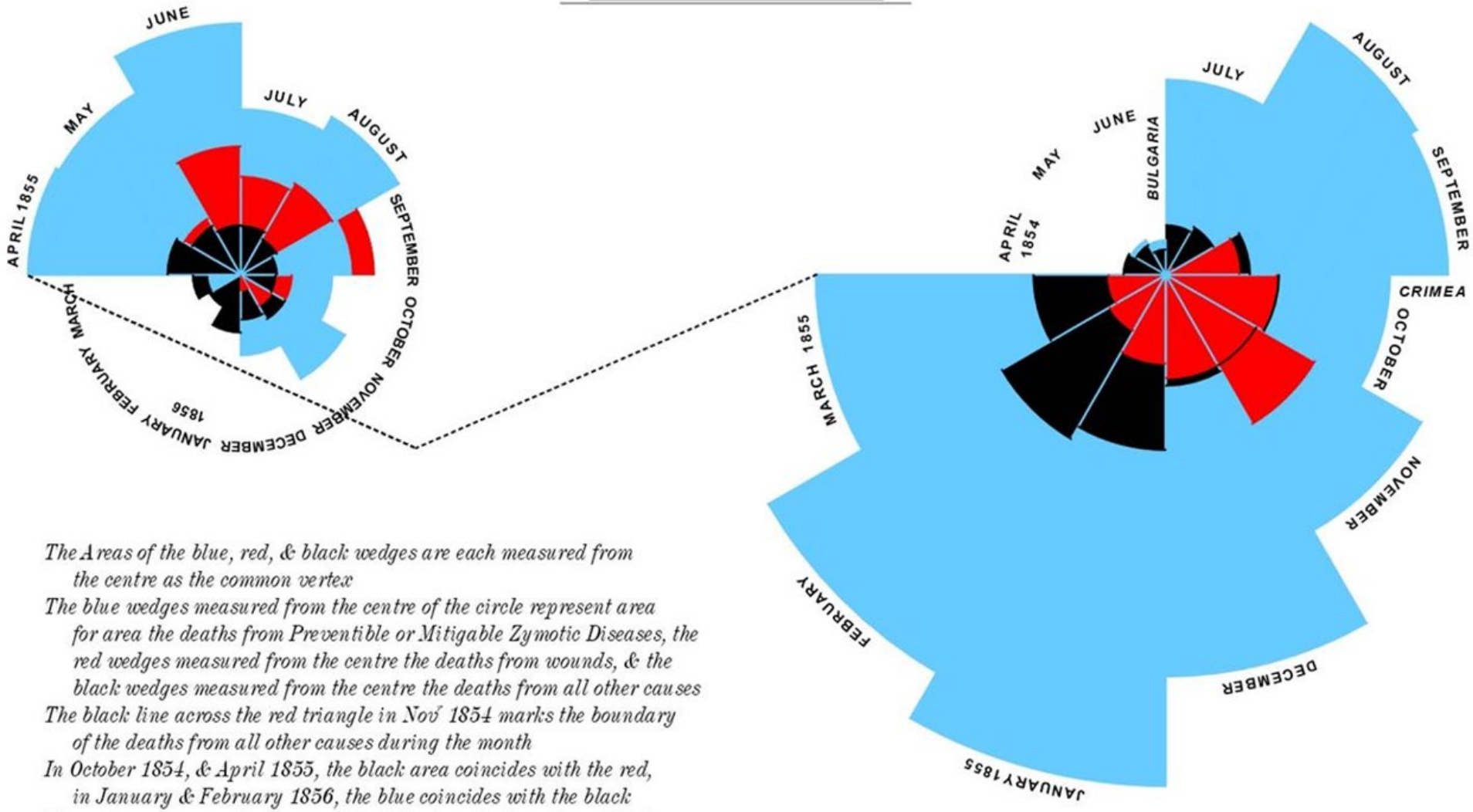
Design Content driven

– The whole design is driven by what is trying to visualise, that it is not a cookie cutter visualisation.

DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

2.
APRIL 1855 TO MARCH 1856.

1.
APRIL 1854 TO MARCH 1855.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex

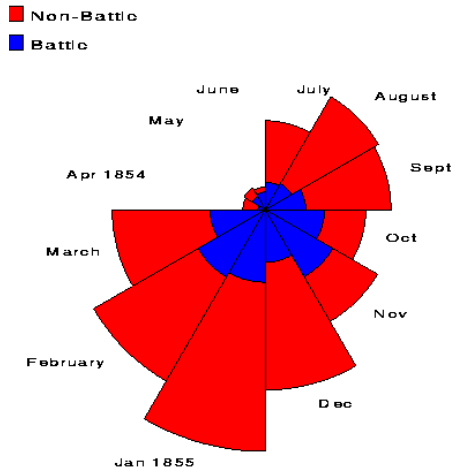
The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic Diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes

The black line across the red triangle in Nov 1854 marks the boundary of the deaths from all other causes during the month

In October 1854, & April 1855, the black area coincides with the red, in January & February 1856, the blue coincides with the black

The entire areas may be compared by following the blue, the red & the black lines enclosing them

Causes of Mortality in the Army in the East
April, 1854 to March 1855



rom: F. Nightingale, "Notes on Matters Affecting the Health, ciency and Hospital Administration of the British Army", 1858


50 0 50 100 150 200 Yards
× Pump • Deaths from cholera
OXFORD STREET
REGENT STREET
GT. MARLBOROUGH ST.
BROAD STREET
WARWICK STREET
DEAN STREET
KING STREET
CONDUIT STREET
SAVILLE ROW
GOLDEN SQUARE
BREWER STREET
NEW BOND STREET
REGENTS QUADRANT
PICCADILLY

Hans Rosling – Color Bubble Motion Chart

During last 200+ years people kept inventing all type of charts to be printed on paper or shown on screen, so most charts showing 2- or 3-dimensional data sets. Professor Hans Rosling led Gapminder.org to create the web-based, animated 6-dimensional Color Bubble Motion Chart.

•X coordinate of the Bubble = Income per person,

•Y coordinate of the Bubble = Life expectancy,

•Size of the Bubble = Population of the Country,

•Color of the Bubble = Continent of the Country,

•Name of the Bubble = Country,

•Year = animated 6th Dimension/Parameter as time-stamp of the Bubble.

Summary

Making sense of large quantities of disparate data is necessary not only for gaining a competitive advantage, but also for surviving in today's business environment. Raw data is often difficult to understand and interpret. Graphs and Charts provide a convenient way to analyse data as well as communicating the insight of the analysis to all levels of the organisation. In the labs, we will use pivot tables to design a dashboard with some interactivity (slice and dice).

References

Ware, Colin, Information Visualization: Perception for Design, Academic Press, 2000.

Tufte, Edward, Envisioning Information, Graphics Press, 1990.

Tufte, Edward, (2001). The Visual Display of Quantitative Information, 2nd Edition. Graphics.

Few, Stephen. (2009). Now You See It: Simple Visualization Techniques for Quantitative Analysis. Analytics Press.

Bertin, Jacques, Semiology of Graphics, 1967

Schneiderman, Ben, “The eyes have it: A task by data type taxonomy for information visualization”, Visual Languages, 1996