Data Visualisation

(CMP020L013S)

Week 2: Application & Evaluation

Mohammad Javaheri

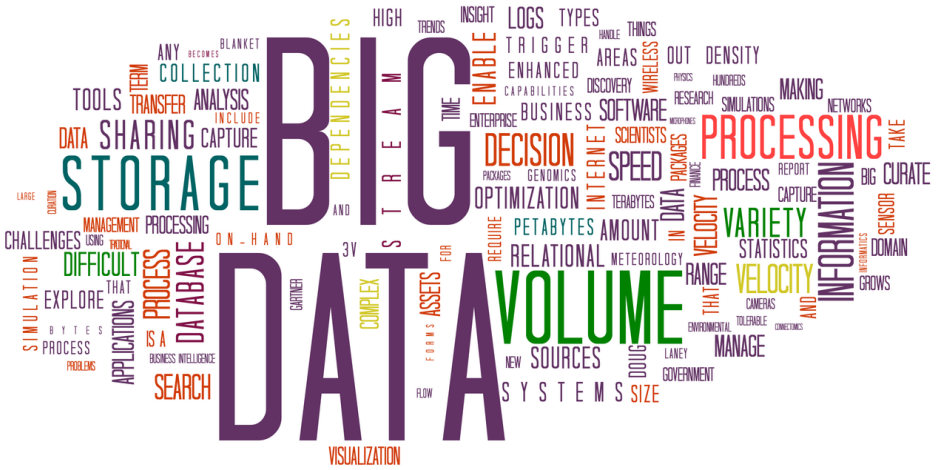
(Dr Mohammad Ali Javaheri Javid)

Agenda

- Review
- Types of Data Visualization
- Evaluating visualisation
  - Good and bad visualisations

Big Data and The Information Problem

- We are Data Rich but Information Poor (DRIP syndrome)

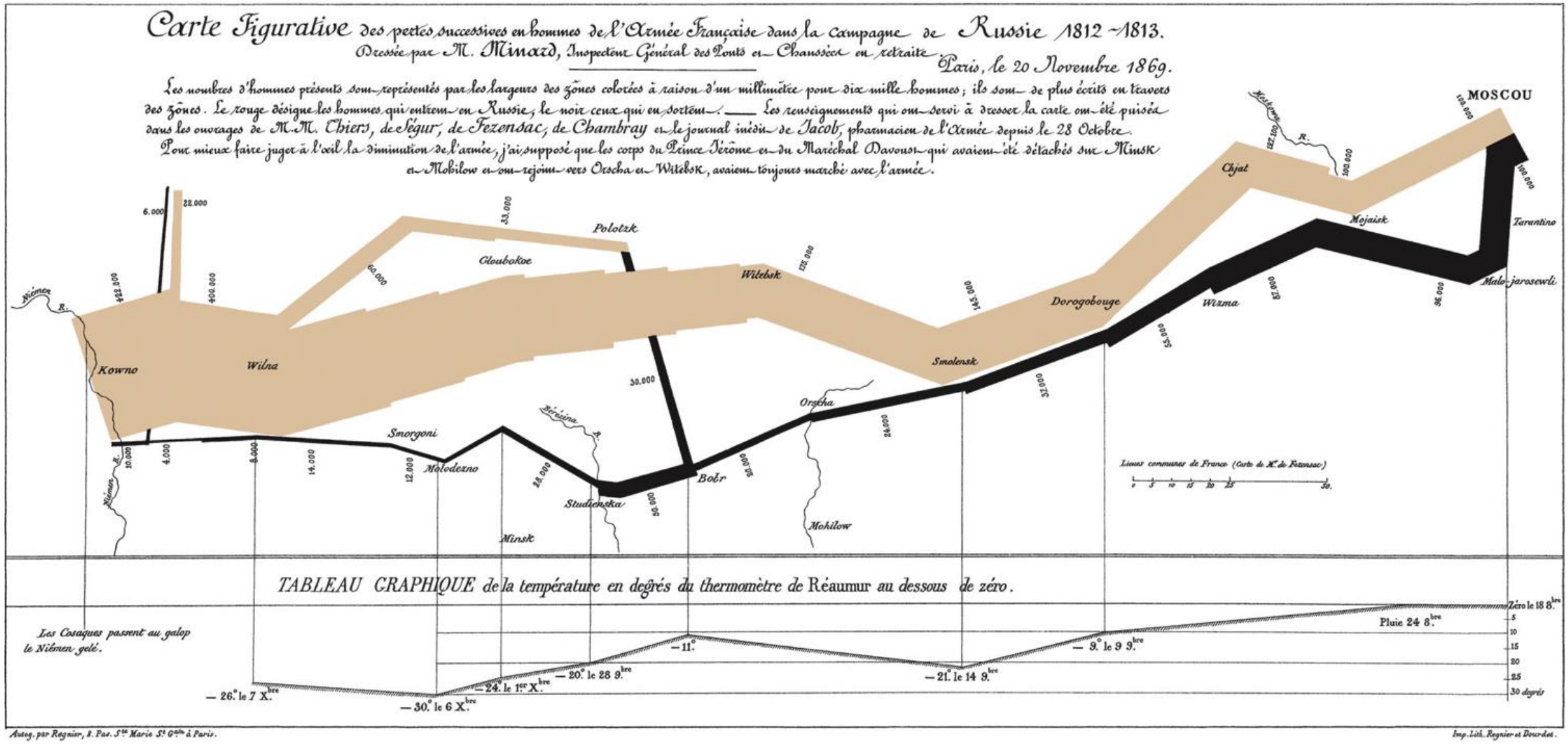


Is data enough?

Data vs Information

- Data != Information
- Data Vis: data → information
  - Using power of human visual processing
  - Making visible the patterns and structures in the data
  - Using graphs, tables, diagrams, maps…

Charles Joseph Minard 1869 Napoleon's March

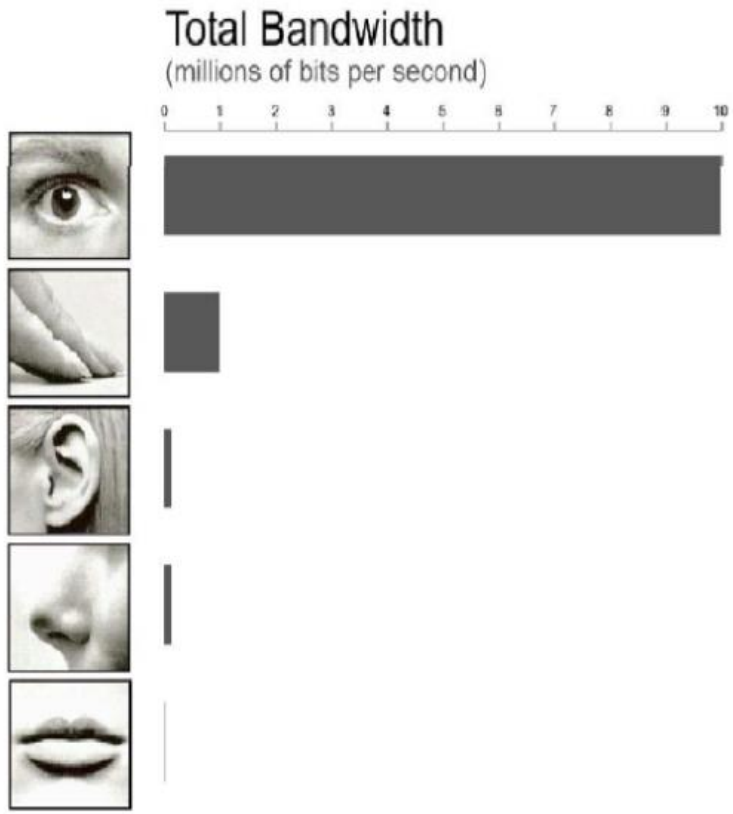


Defining Visualisation

- Visualisation is the process that transforms (abstract) data into interactive graphical representations for the purpose of exploration, confirmation, or presentation.
- **Tool to enable a user insight into Data**

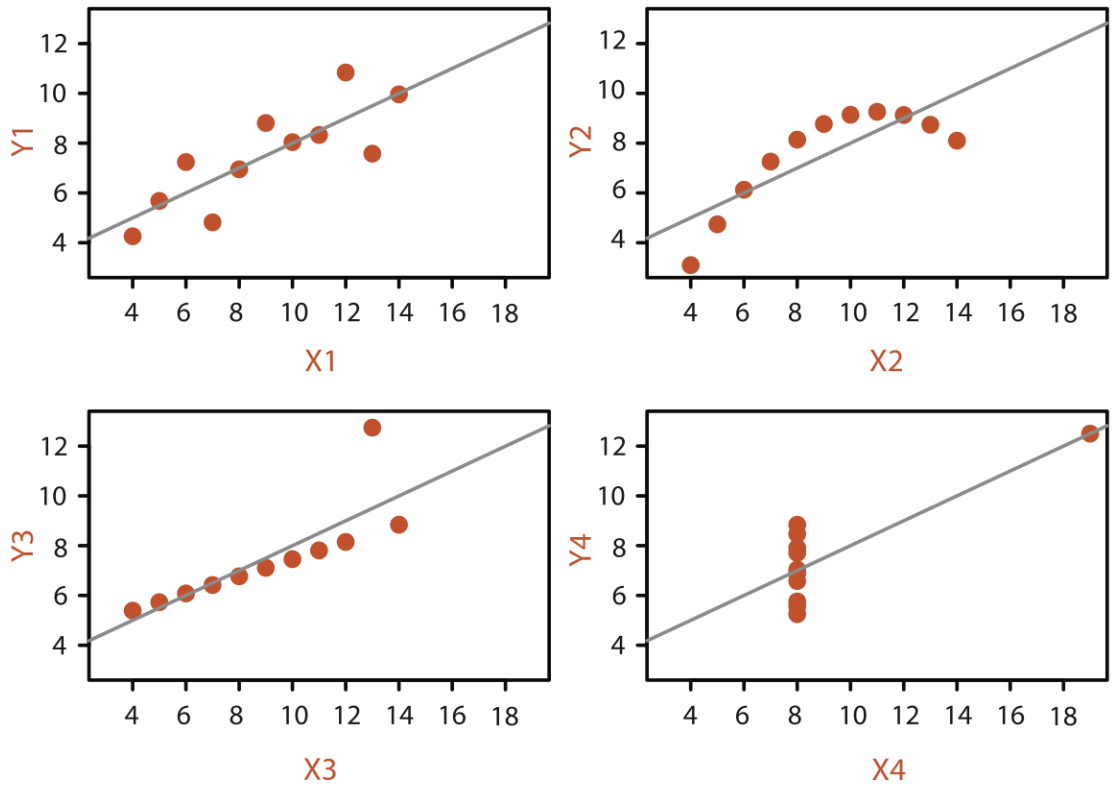
Why depend on vision?

- Figures are richer; provide more information with less clutter and in less space.
- Figures provide the *gestalt* effect: they give an overview; make a structure more visible.
- Figures are more accessible, easier to understand, faster to grasp, more comprehensible, more memorable, more fun, and less formal.



Why represent all the data?

- summaries lose information, details matter
  - confirm expected and find unexpected patterns
  - assess the validity of statistical model



Identical statistics	
x mean	9
x variance	10
y mean	7.5
y variance	3.75
x/y correlation	0.816

Uses for Visualisation

A: Support reasoning about information (analysis)

- Finding relationships
- Discover structure
- Quantifying values and influences
- Should be part of a query/analyze cycle

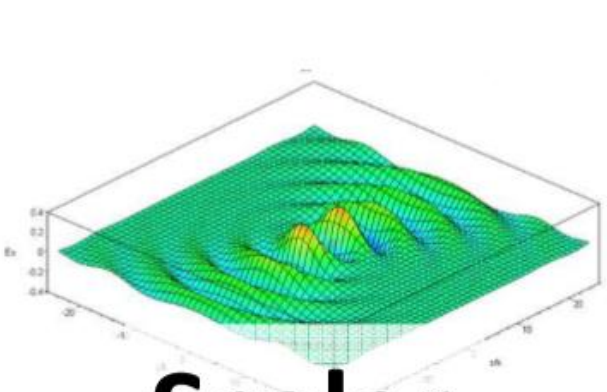
B: Inform and persuade others (communication)

- Capture attention, engage
- Tell a story visually
- Focus on certain aspects, and omit others

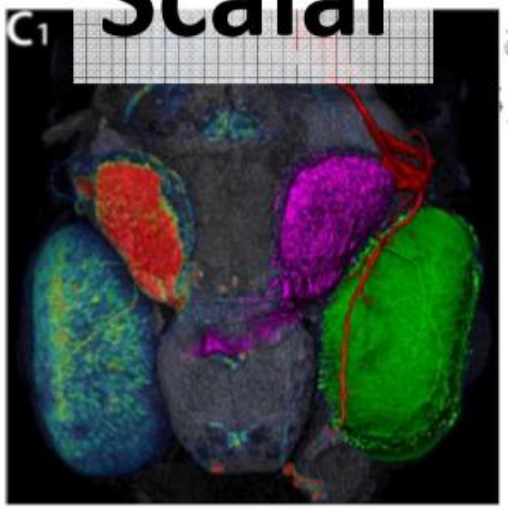
Types of Data Visualization

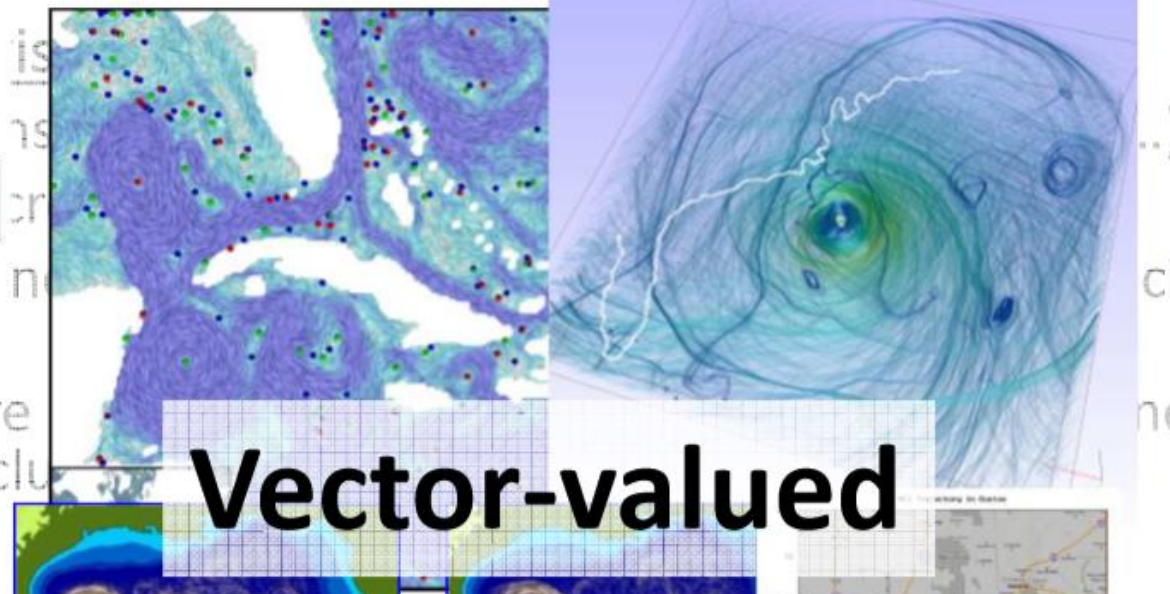
- Scientific Visualization
  - Structural Data
    - Medical, ..
- Information Visualization
  - No inherent structure
    - News, stock market, top grossing movies, facebook connections
- Visual Analytics
  - Use visualization to understand and synthesize large amounts of multimodal data
    - audio, video, text, images, networks of people

Information vs Scientific Visualisation

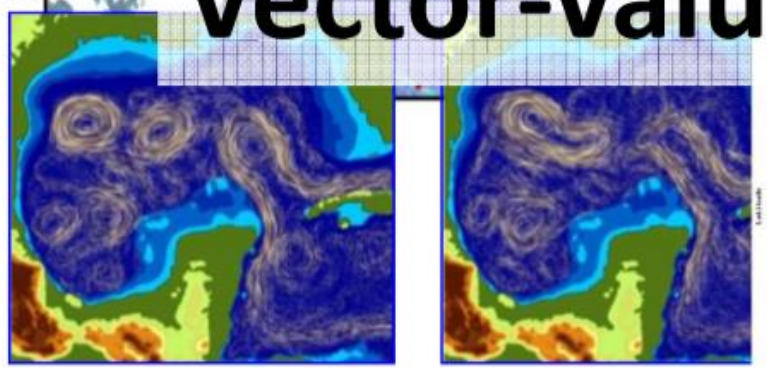


Scalar

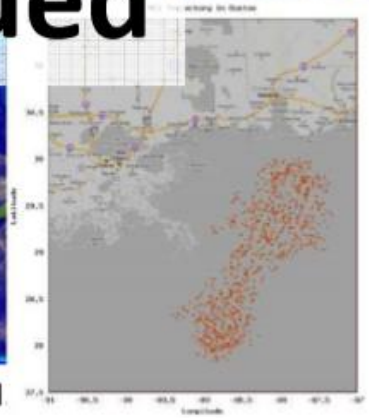


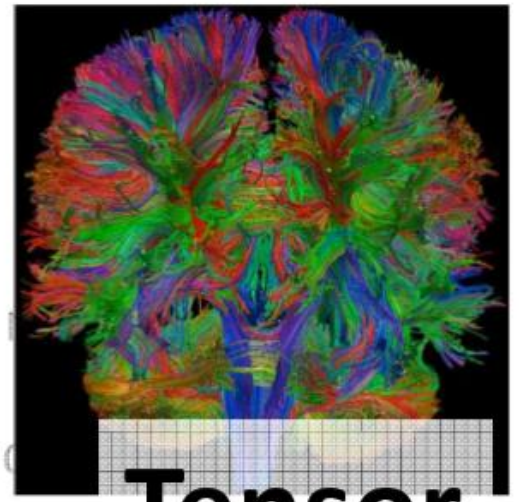


Vector-valued

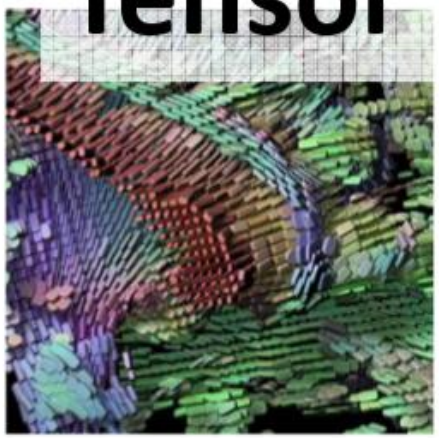


(Bruno Jobard, Gordon Erlebacher, and M



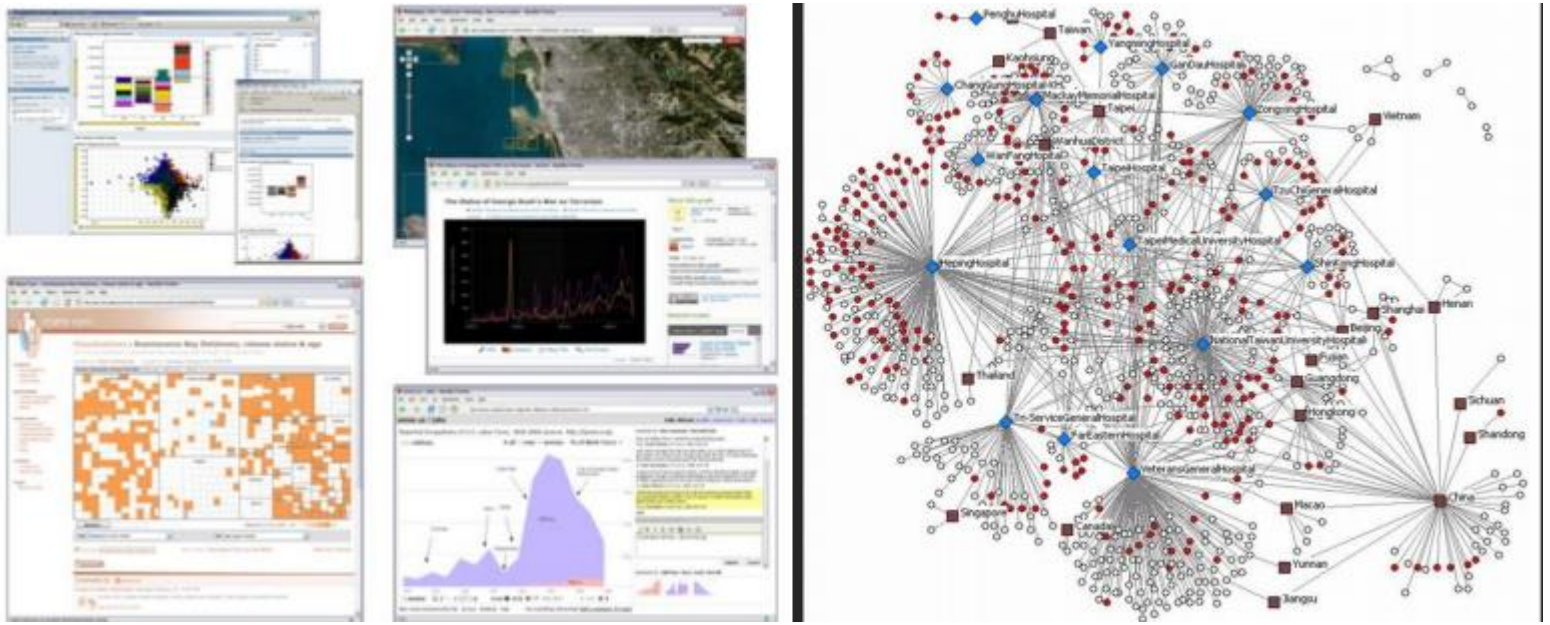


Tensor



Visual Analytics

- Integration of interactive visualization with analysis techniques to answer a growing range of questions in science, business, and analysis.
- Making sense of multimodal data-audio clips, video, photographs, transcripts, …



Use Cases of Data Visualisation

- Exploratory
  - helps discover trends and patterns interactive
  - quick and dirty
  - early stages of research
- Explanatory
  - conveys a clear message
  - makes a point or answers a question
  - polished, publication/presentation ready
  - final stages of research

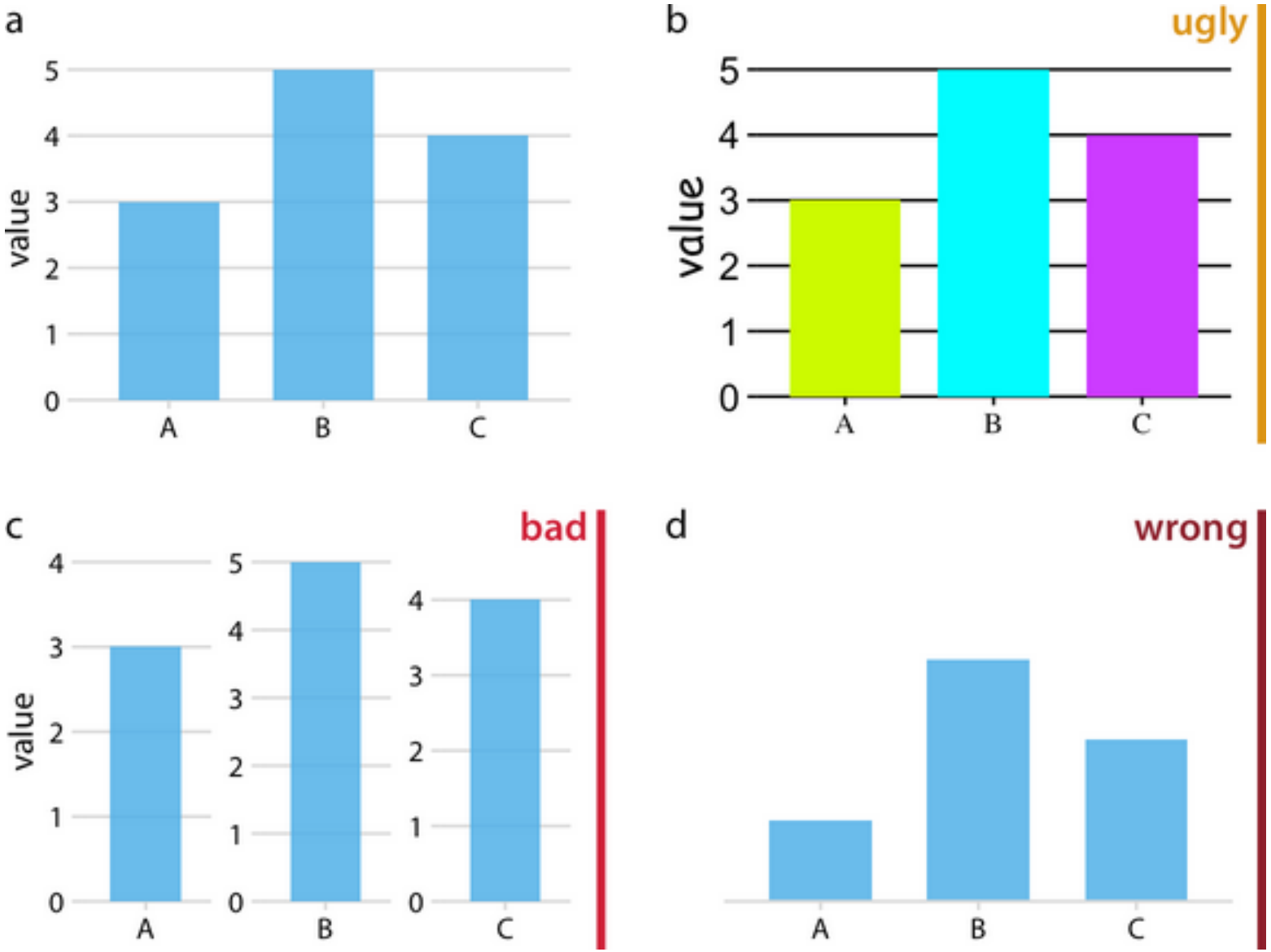
Value of Data Visualisation

- Useful in many areas of work and study
- Practical data manipulation skills, enhances experience of research
- Visualisation is a powerful means of communication, and making a difference
- **The Purpose**
  - Understanding
  - Reveal things that are
    - Important
    - Meaningful
    - Useful

Good Data Visualization

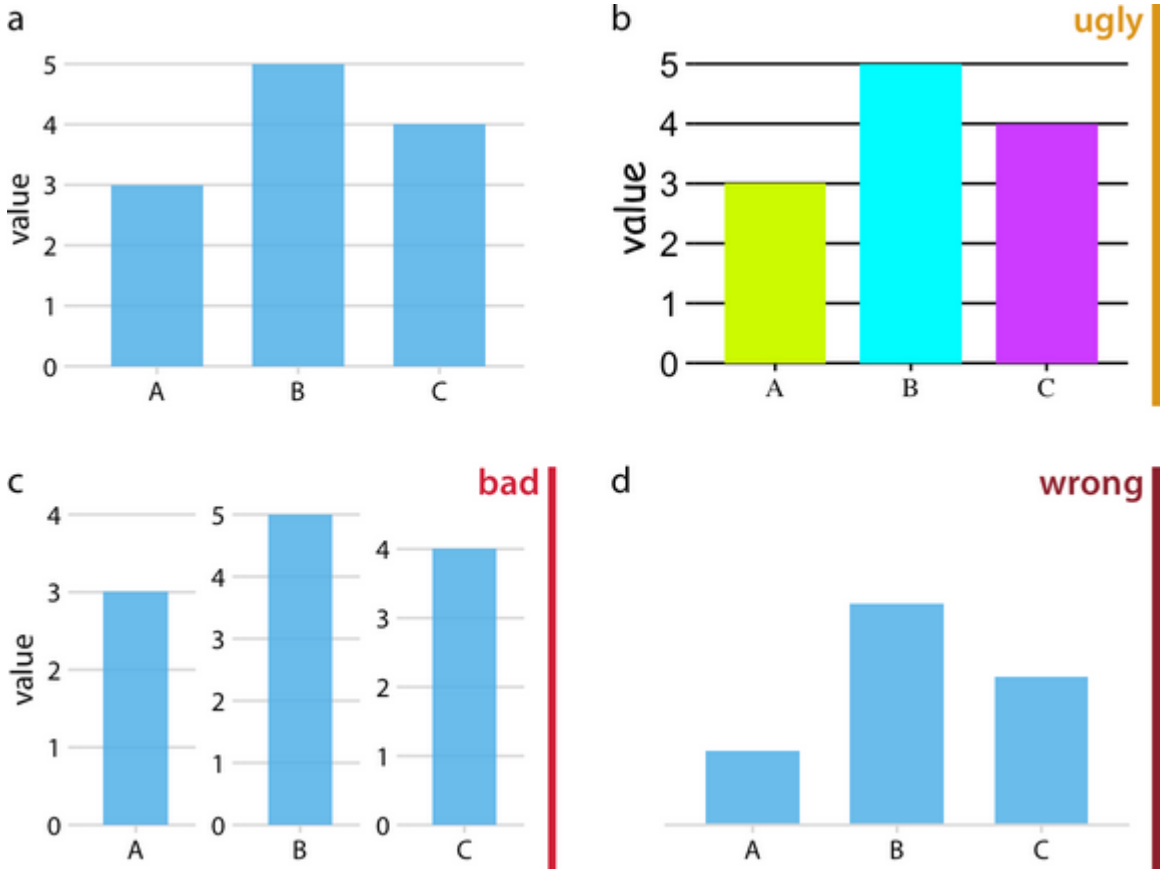
- Makes data accessible
- Combines strengths of humans and computers
- Enables insight
- Communicates effectively
- Visualisation is really about external cognition, that is, *how resources outside the mind can be used to boost the cognitive capabilities of the mind*.

A Case for Ugly Visualisations

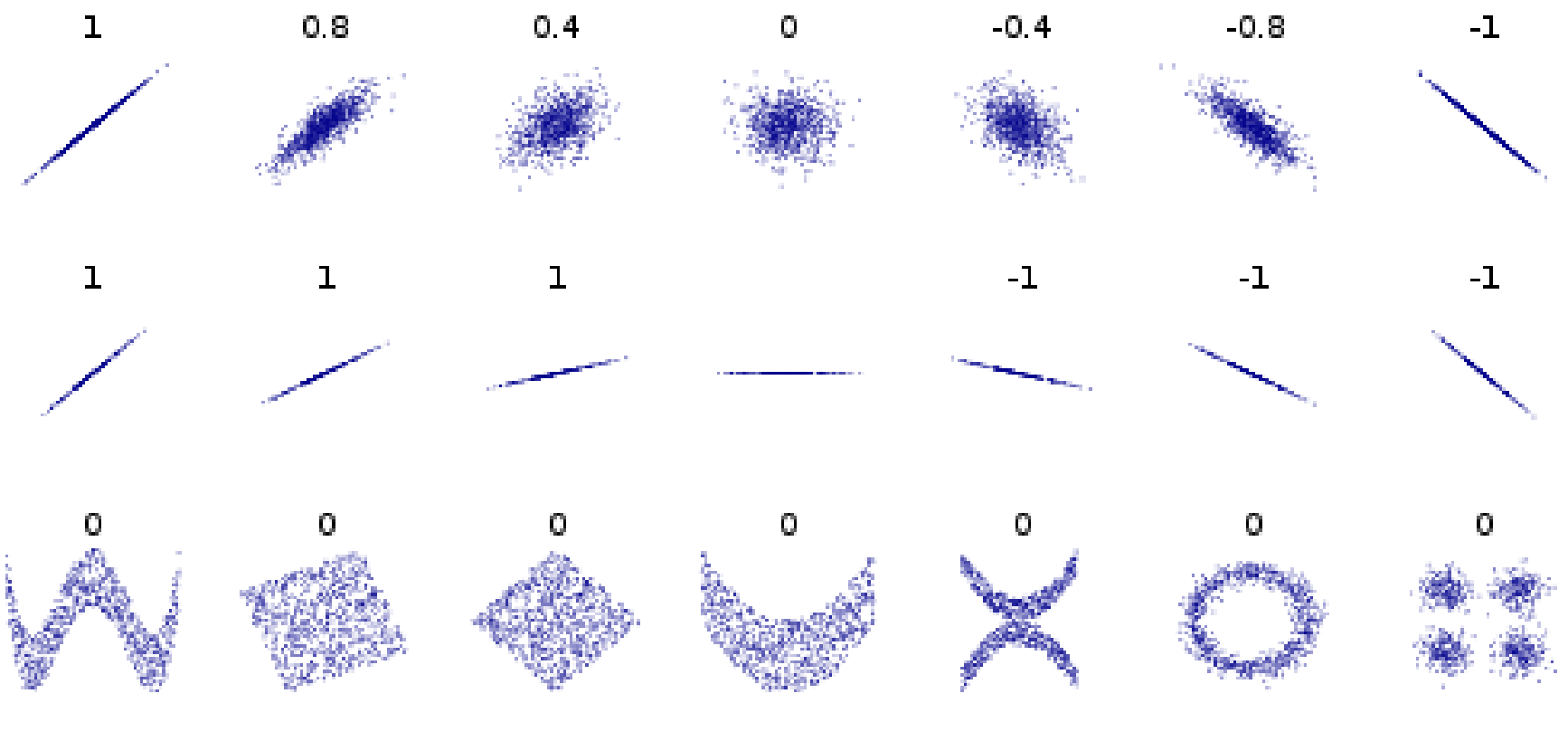


A Case for Ugly Visualisations

- Ugly
  - A figure that has aesthetic problems but otherwise is clear and informative.
- Bad
  - A figure that has problems related to perception; it may be unclear, confusing, overly complicated, or deceiving.
- Wrong
  - A figure that has problems related to mathematics; it is objectively incorrect.



Pearson Correlation



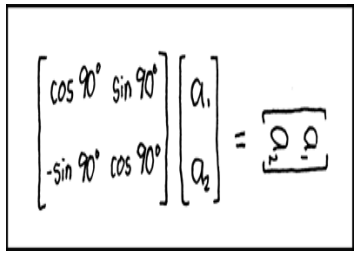
What makes a good Visualisation?

- Use the appropriate visual element for the relationship and data being analysed (Jock Mackinlay)
- Effectiveness
- A visualisation is more effective than another visualisation if the information conveyed by one visualisation is more readily perceived than the information in the other visualisation.

Data Scientist's Workflow

Sandbox

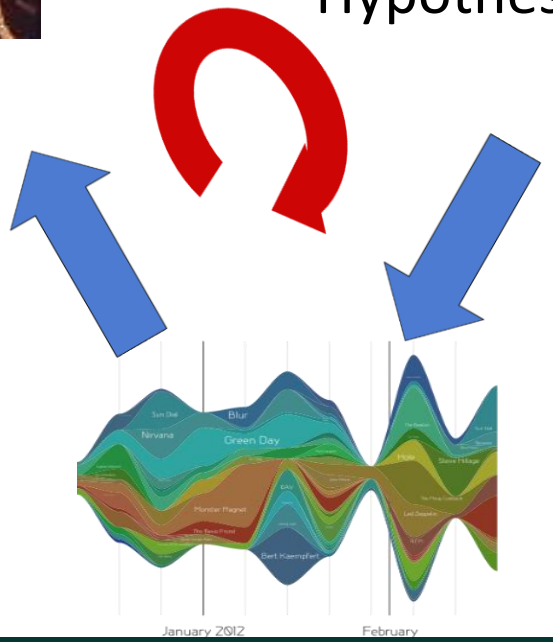
Digging Around in Data



Hypothesize Model

Production

Large Scale Exploitation



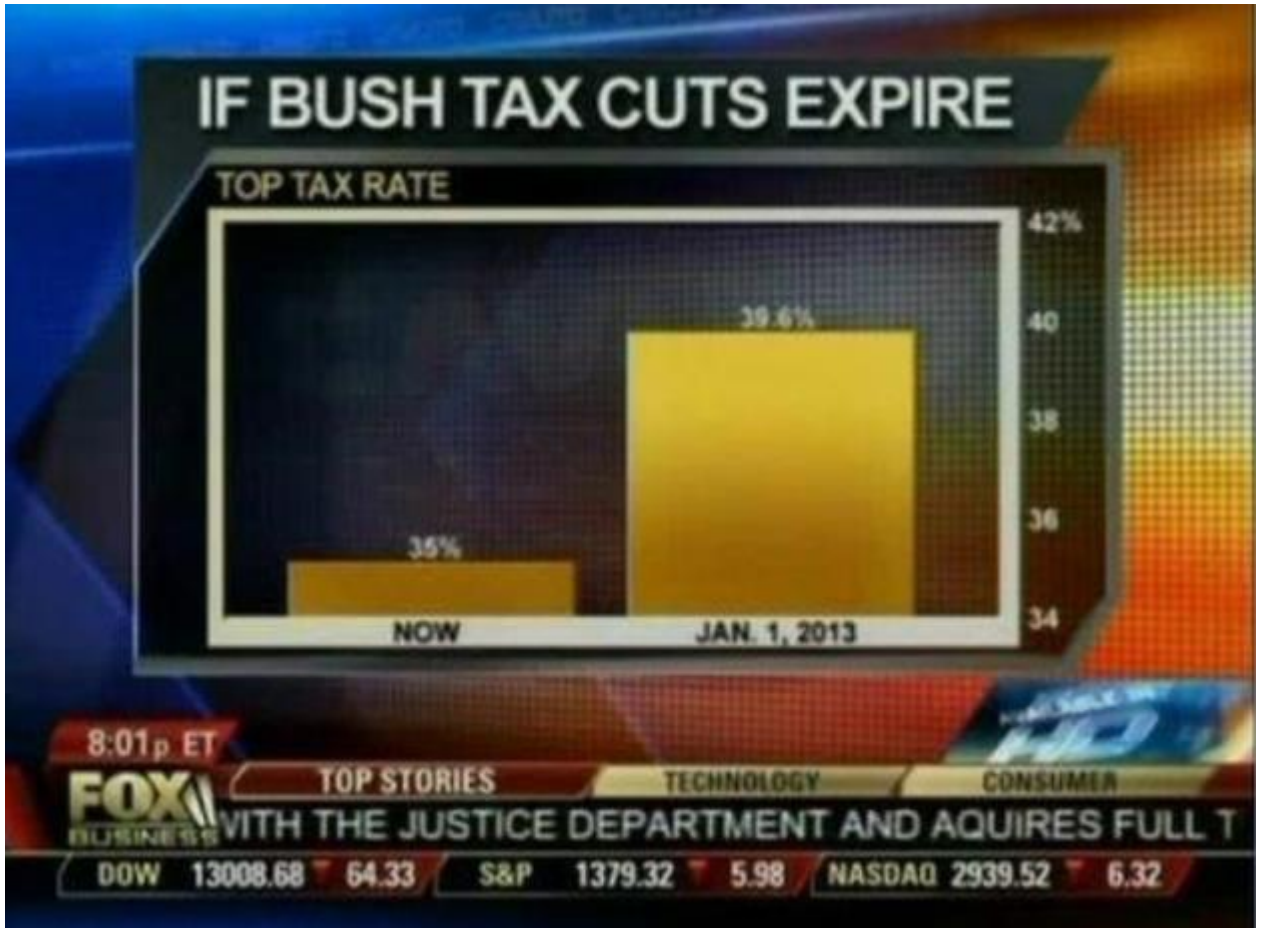
Evaluate
Interpret

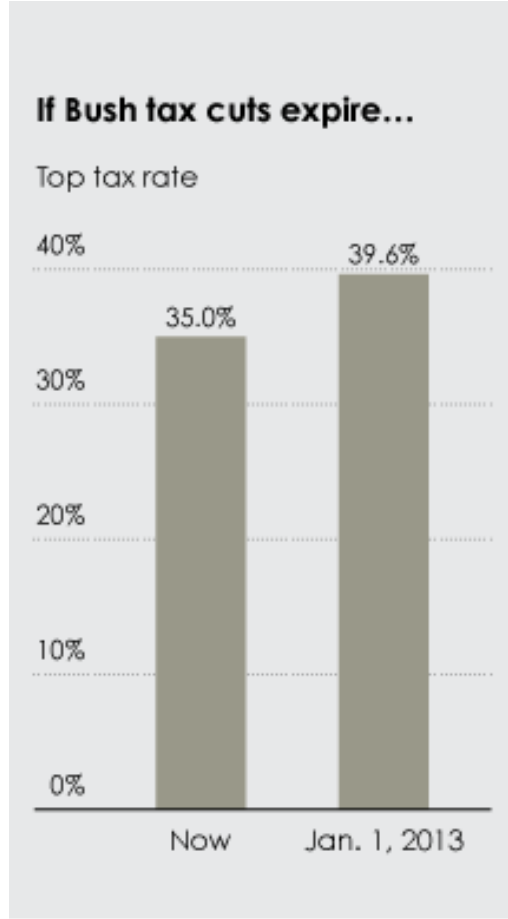
Top 3 tips for **BAD** Data Visualisation

- i.e. what NOT to do!

1) Cram everything you can into the chart – readability is overrated
2) Choose the scale to hide the inconvenient truth
3) Emphasise the trivial and ignore the important

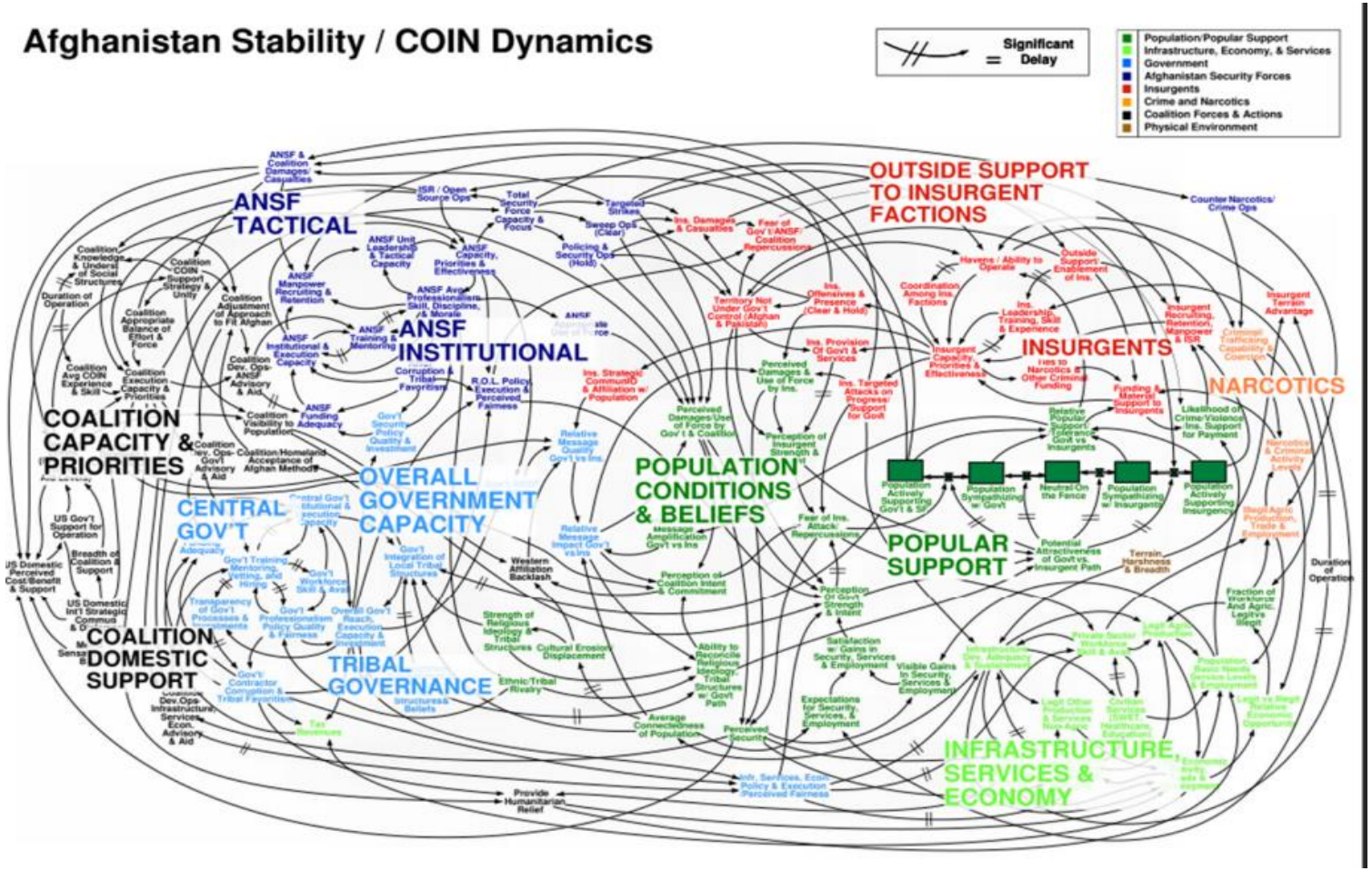
Example: Misleading Bar Chart



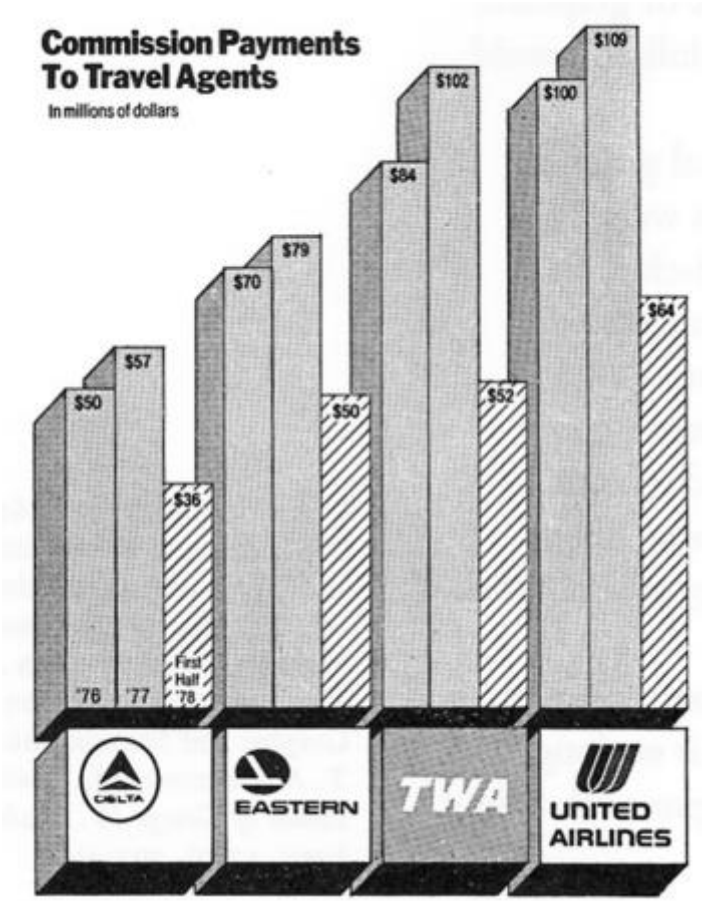


SOURCE: http://flowingdata.com/category/statistics/mistaken-data/

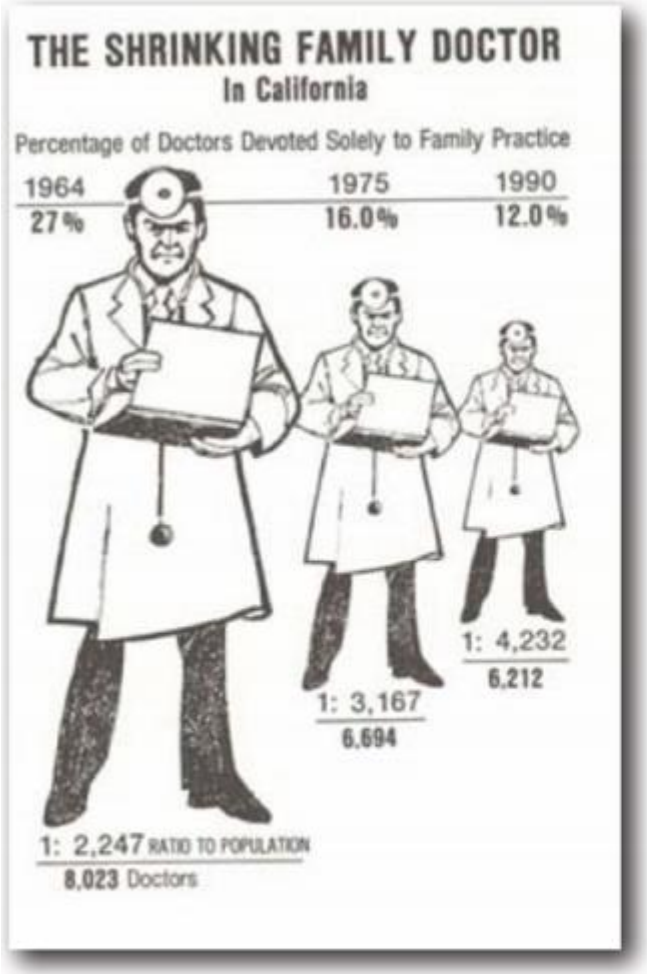
Visualization to Educate?

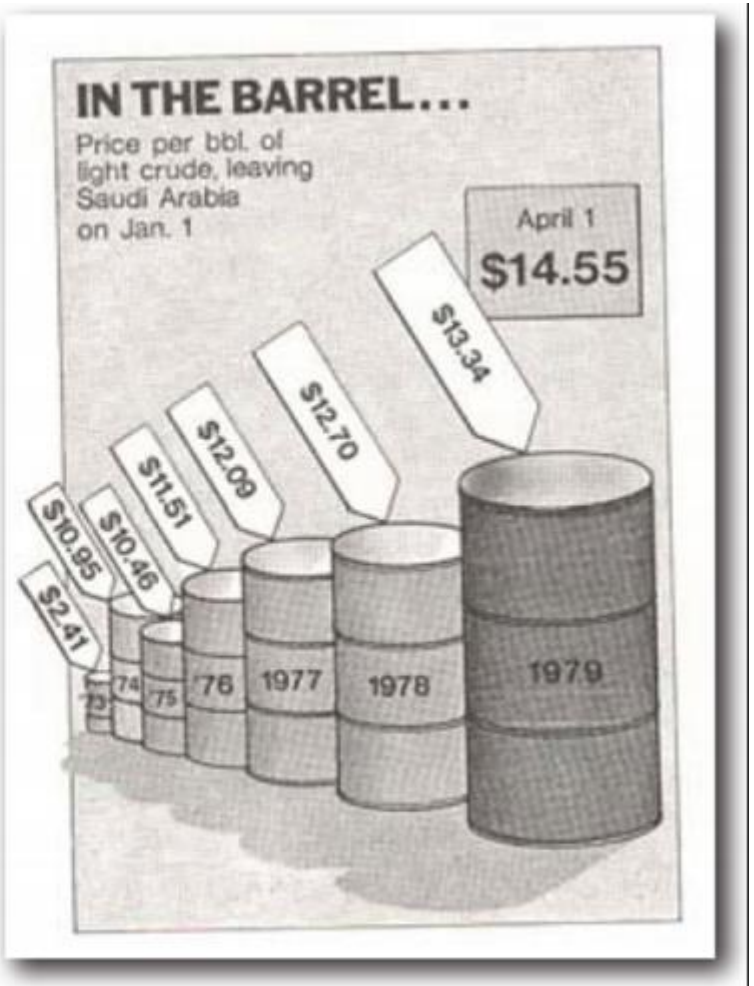


https://www.theguardian.com/news/datablog/2010/apr/29/mcchrystal-afghanistan-powerpoint-slide



Comparing six months of 1978 with 1976, 1977





Inconsistent Visual area and numeric measure

Further Material

- Look at some of these visualisation examples and find out /summarise what is the root of their problems
- http://www.perceptualedge.com/examples.php