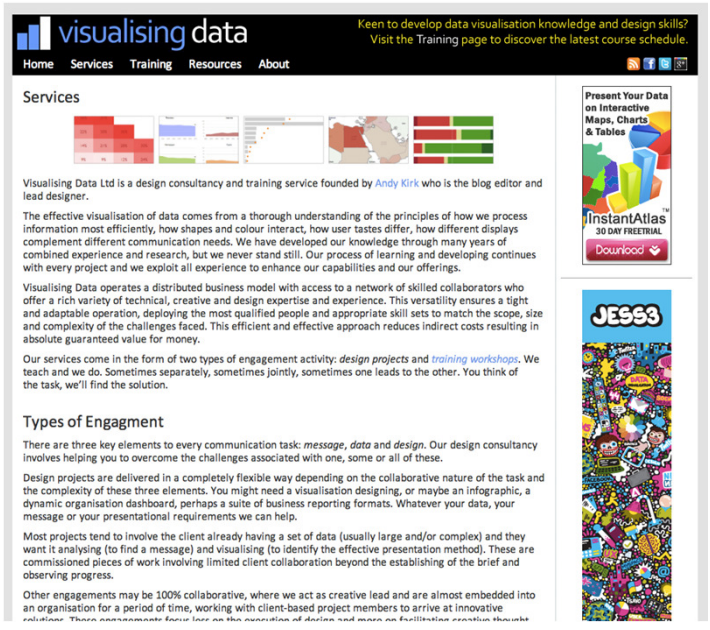
visualising data

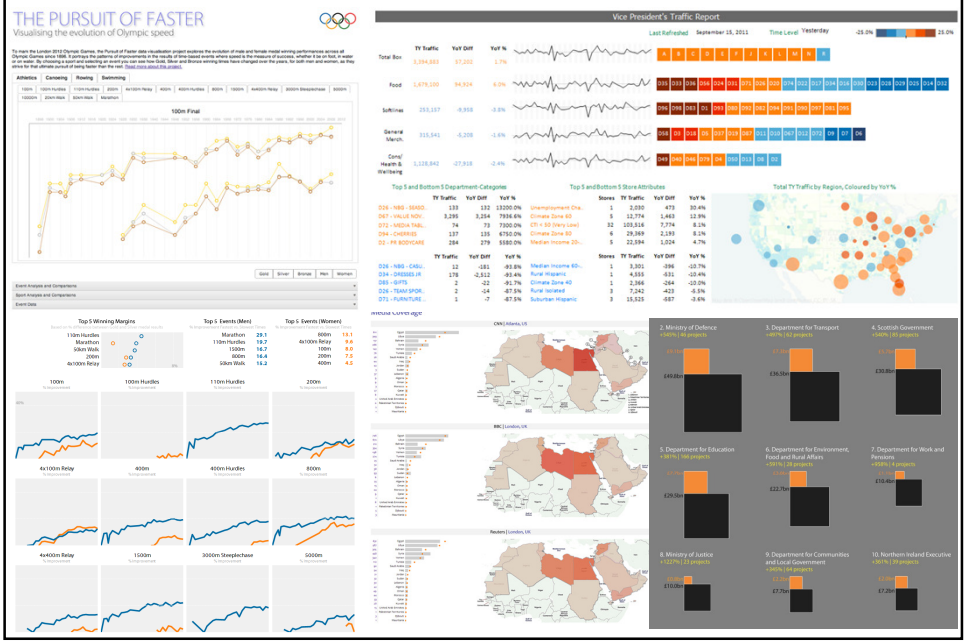
Visualisation's Duality: Finding Stories and Showing Stories

Andy Kirk

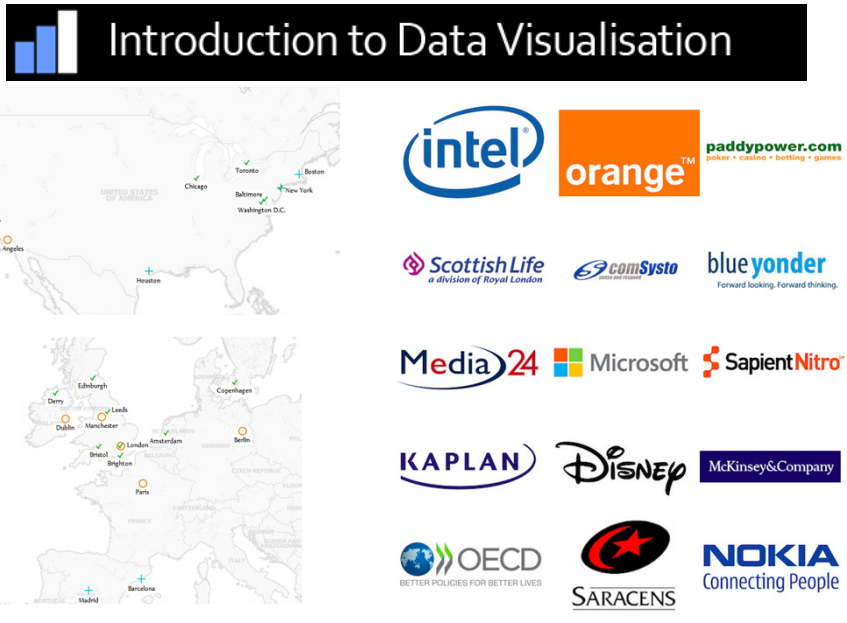
www.visualisingdata.com



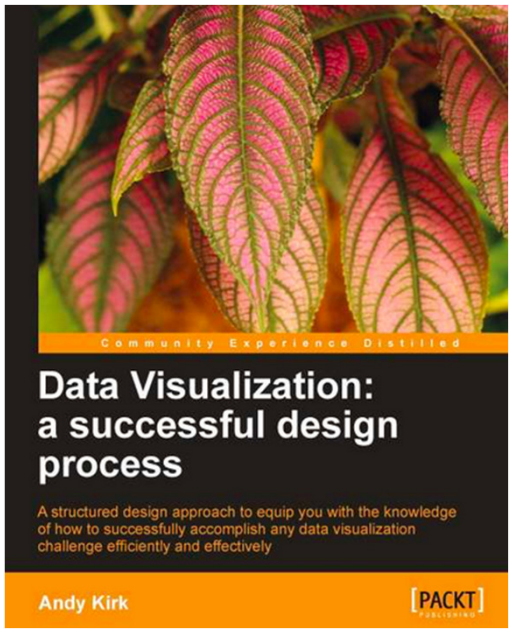
Design architect/consultant



Trainer



Author



What to show | How to show it

The real craft behind data visualisation design is being able to rationalise choices

1. Establish the visualisation's purpose and identify key factors

What is 'Purpose'?

Trigger

Its reason for existing
How well is it defined?

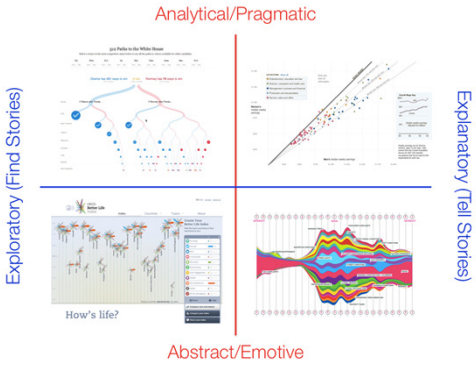
Client project (brief)
Internal project (brief)

↕

Self-initiated

Intent

The intended *tone* and *function*



Intent: Tone

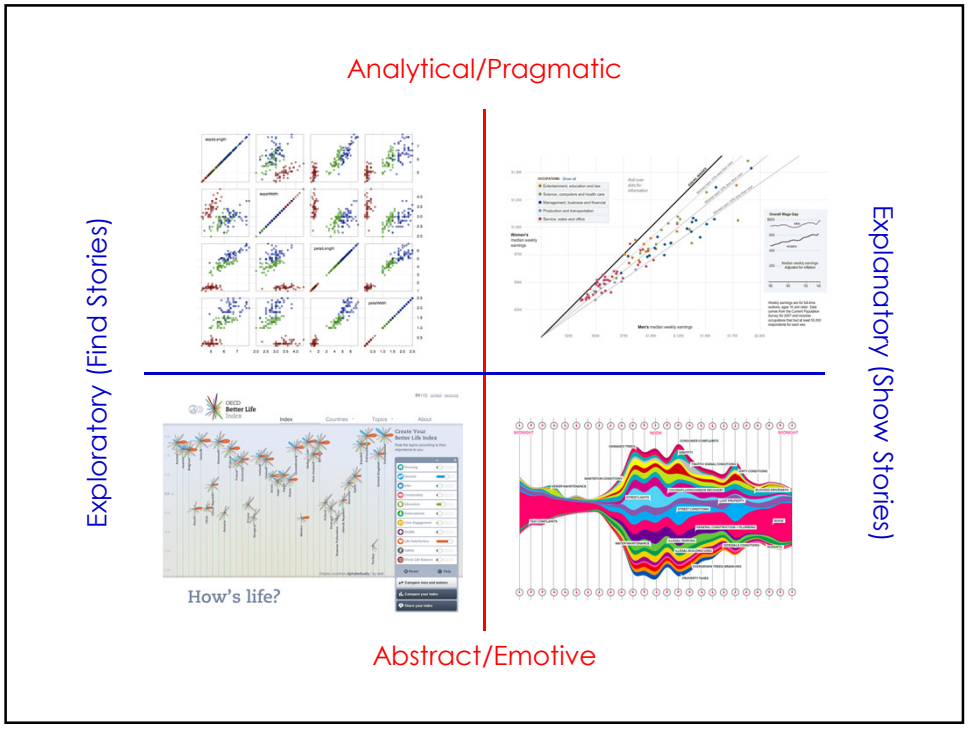
How important is accuracy compared to aesthetics?

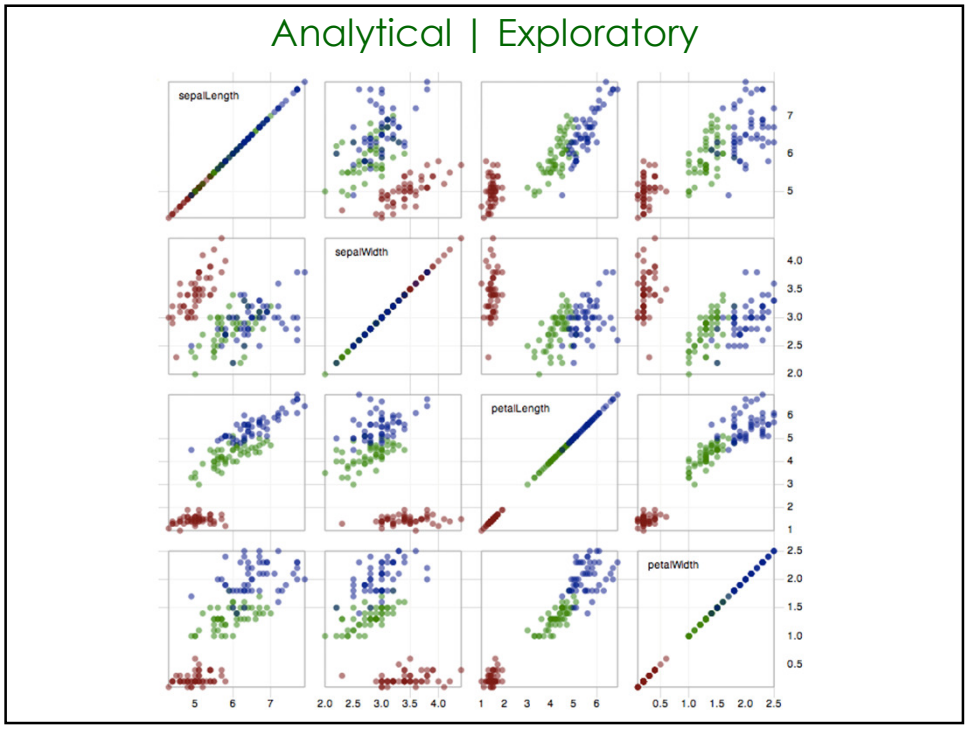
Read data vs Feel data
Precision vs Beauty
Pragmatism vs Emotion

Intent: Function

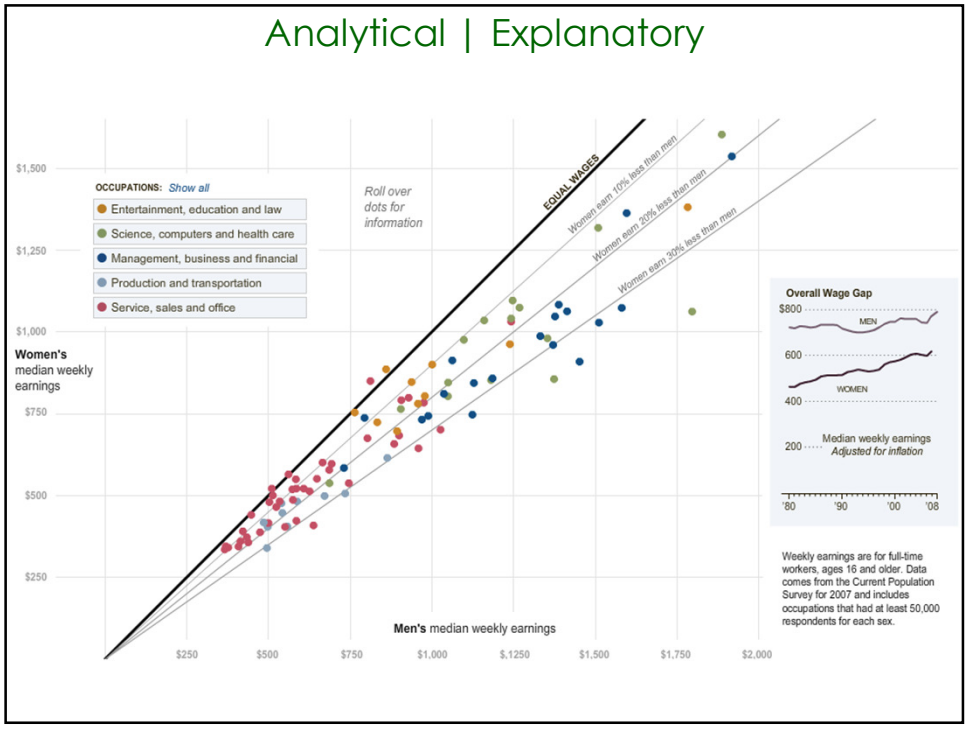
Who does the work to surface the insights?

Find or Show
Reader or Designer
Explore or Explain

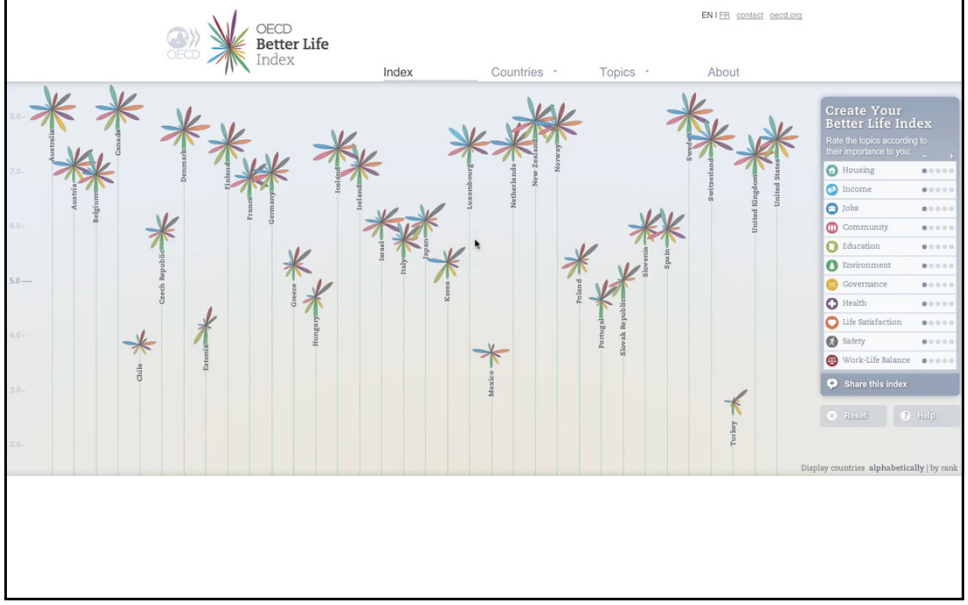




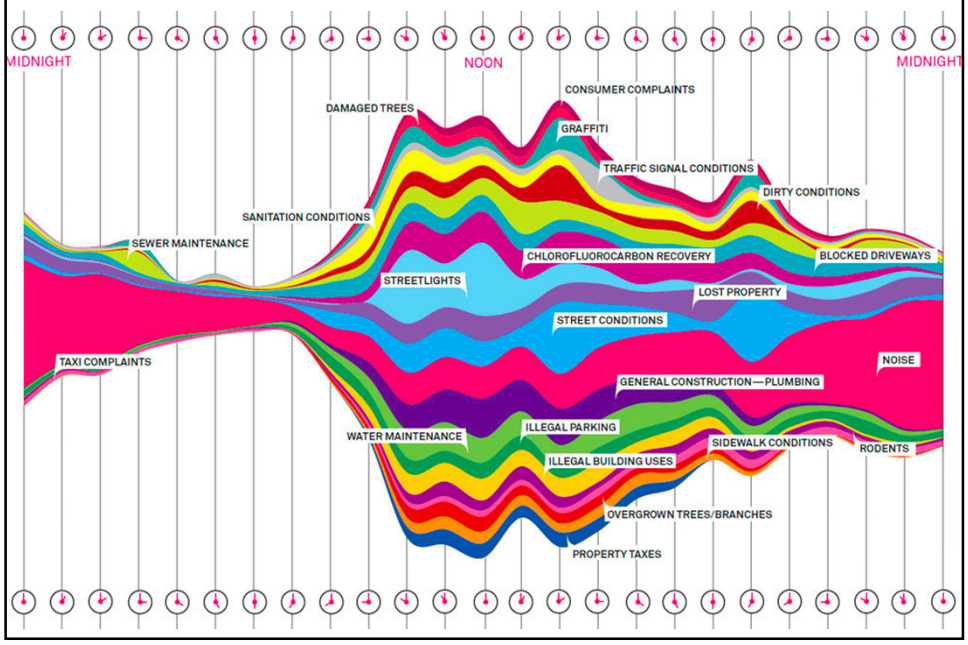
Analytical | Explanatory



Emotive | Exploratory



Emotive | Explanatory



Potential key factors

The brief? Open, strict, helpful, unhelpful, clarity
Pressures? Timescales, managerial, financial
Format? Static, interactive, video, tools
Setting? Issued, presented, instant, prolonged
Technical? Software, hardware, infrastructure
Audience size? One, group, organisation, outside
Audience type? Domain, captive, general
Resolution? Headlines, detail
Frequency? One-off, regular
Rules? Structure, layout, style, colour
People? Individual, team, the 8 hats…

2. Acquire and prepare your data

The hidden burden…

Acquisition
Examination
Transform for quality

The hidden cleverness…

Transform for analysis
Consolidation
Visual Analysis

Visual analysis

Using visualisation techniques to familiarise, learn about and discover insights from data

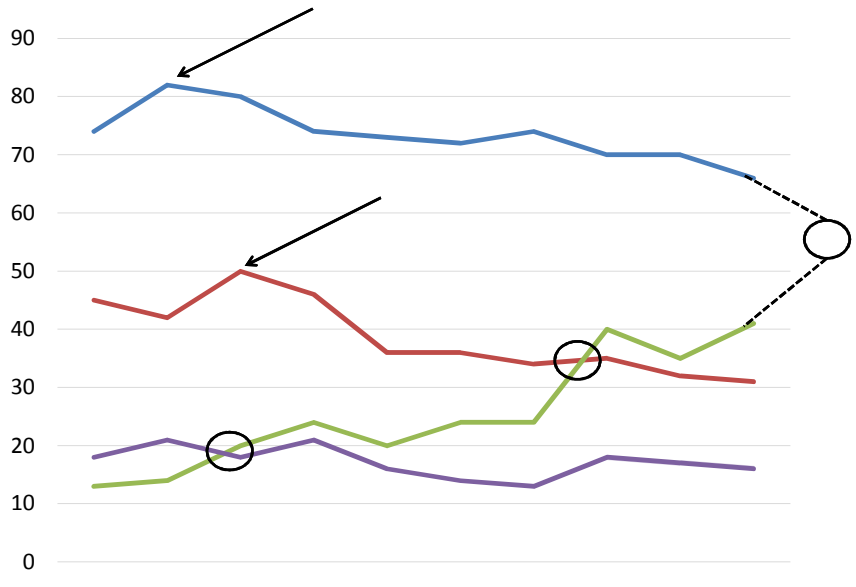
Requires curiosity and graphical literacy

Graphical literacy

Trends and patterns (or lack of)

- Up and down vs. flat?
- Linear vs. exponential
- Steady vs. fluctuating
- Seasonal vs. random
- Rate of change vs. steepness

Graphical literacy

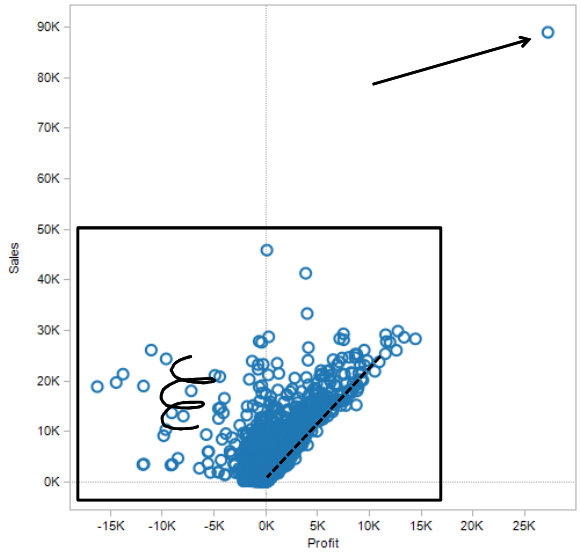


Graphical literacy

Relationships

- Outliers
- Intersections
- Correlations
- Connections
- Clusters
- Associations
- Gaps

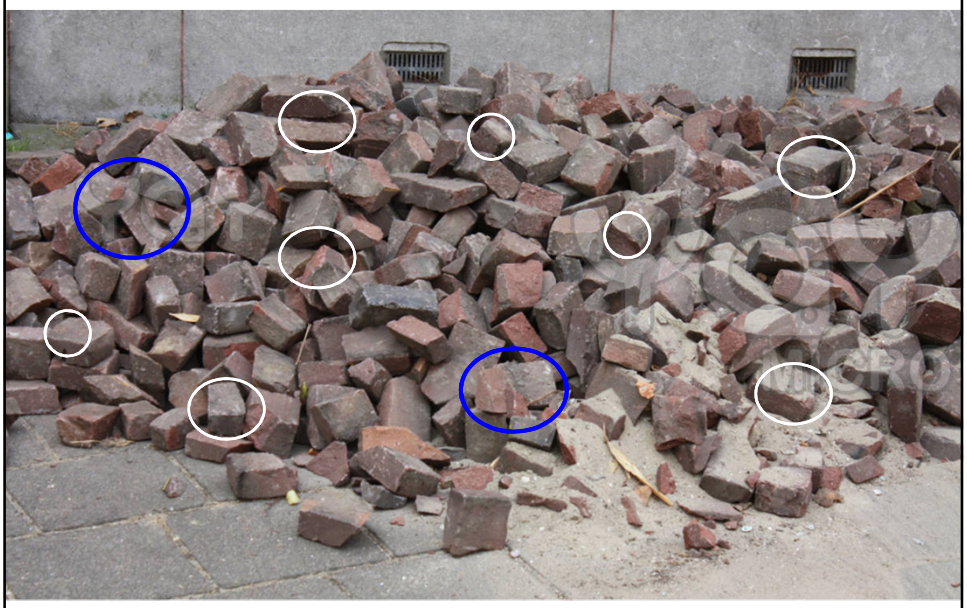
Graphical literacy



3. Establishing editorial focus by finding stories

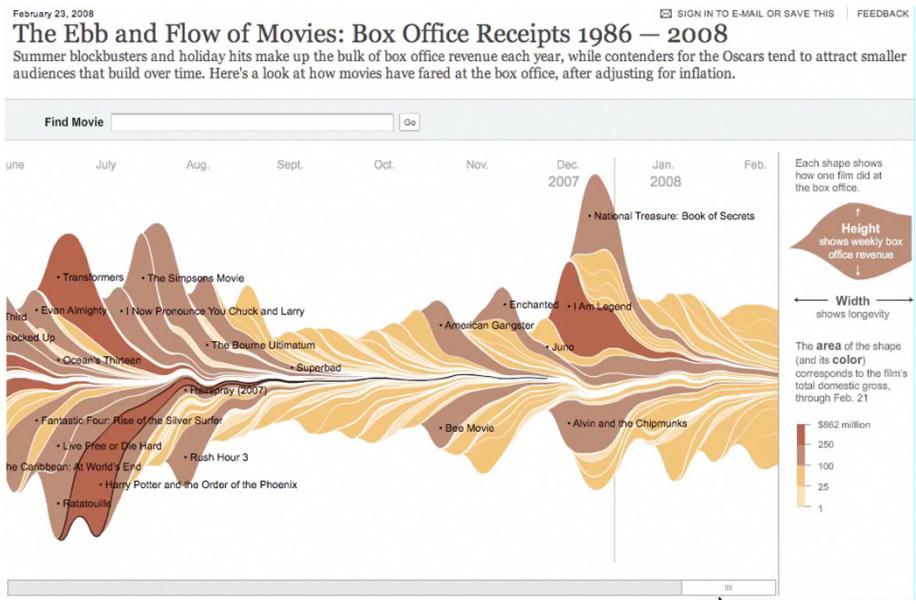
Good content reasoners and presenters are rare, designers are not.

Edward Tufte



What questions do you have about this data?

What questions do you want readers to be able to answer about this data?



We rejected them because they didn't do a good job of answering some of the most interesting questions... Different forms do better jobs at answering different questions.

Amanda Cox (on NYT Stream Graph)

4. Conceive your visualisation design specification

The 5 layers of a visualisation

1. Data representation

Which chart?

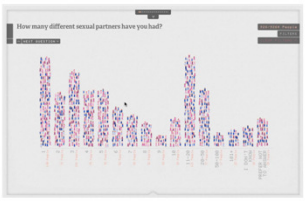
What are we trying to say
with what we are showing?

Data representation ingredients

1. Consistency with purpose
2. Choose the correct visualisation method
3. Effectiveness of visual analysis techniques
4. Consider physical properties of your data
5. Create the appropriate metaphor

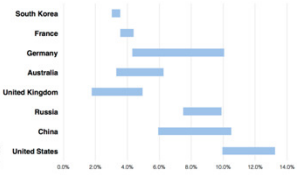
Comparing categories

Pixelated Bar Chart



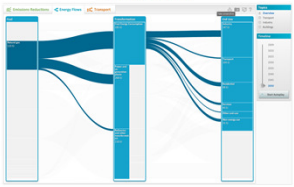
Portion/Comparative Chart

Floating Bar



Slopegraph/Bump Chart

Sankey Diagram

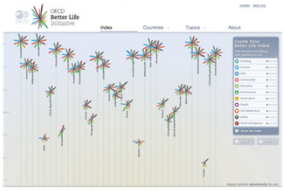


Bar Chart/Column Chart

Word Cloud

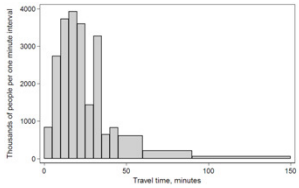
Dot Plot

Glyph Chart



Stacked Bar Chart

Histogram

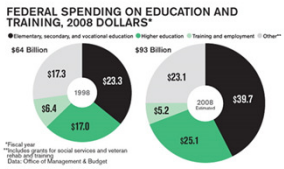


Assessing hierarchies & part-to-whole relationships

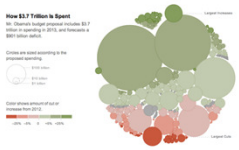
Pie Chart

Square Pie/Unit Chart

Doughnut Chart

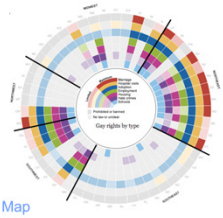


Circle Packing Diagram

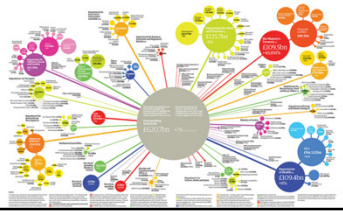


Polar Diagram/Coxcomb

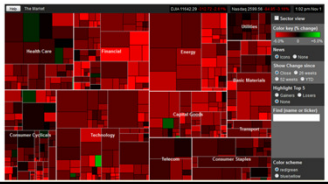
Radial Chart



Tree/Bubble Hierarchy

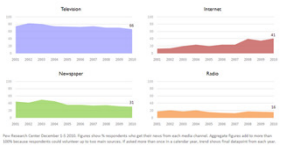


Tree Map



Showing changes over time

Area Chart

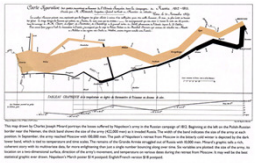


Symbol Time Chart

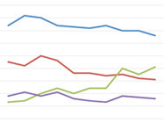
Spark Lines

Candlestick Chart

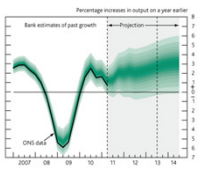
Flow Map



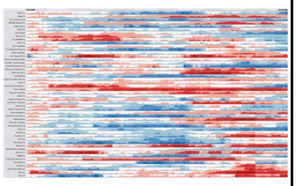
Multi-line Chart



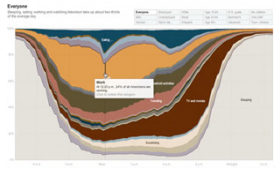
Fan Chart



Horizon Chart

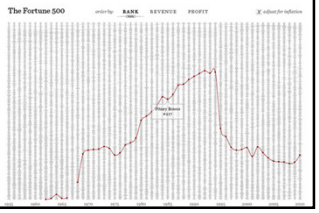


Stacked Area Chart



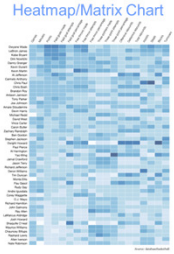
Stream Graph

Stacked Line Chart

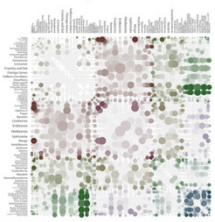


Charting connections and relationships

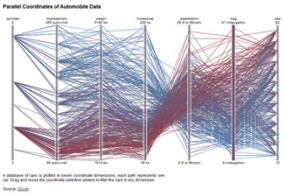
Heatmap/Matrix Chart



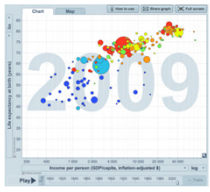
Heatmap/Matrix Chart



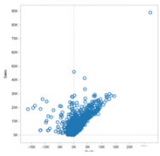
Parallel Sets/Coordinates



Bubble Plot

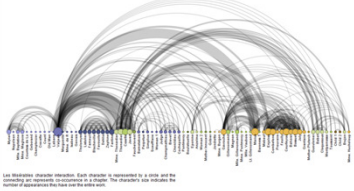


Scatter Plot

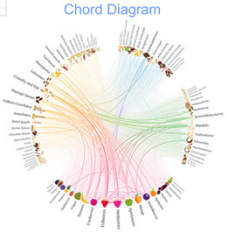


Force-Directed/Node-Link Network

Arc Diagram

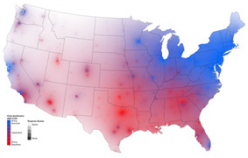


Chord Diagram

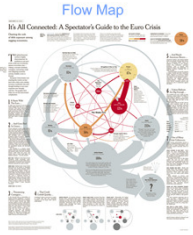


Mapping spatial data

Isarithmic/Contour/Topological Map

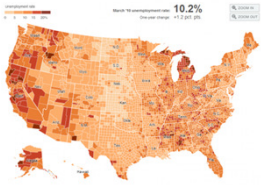


Flow Map

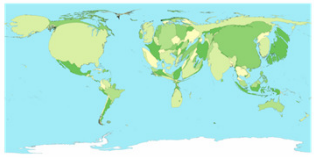


Connection Map

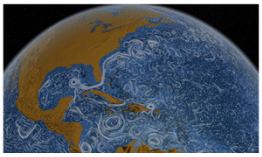
Choropleth Map



Cartogram

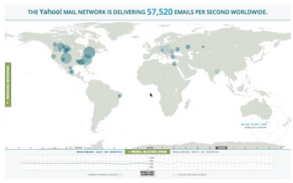


Particle Flow

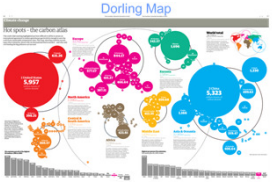


Dot/Point Map

Bubble Overlay Map



Dorling Map



The 5 layers of a visualisation

2. Colour

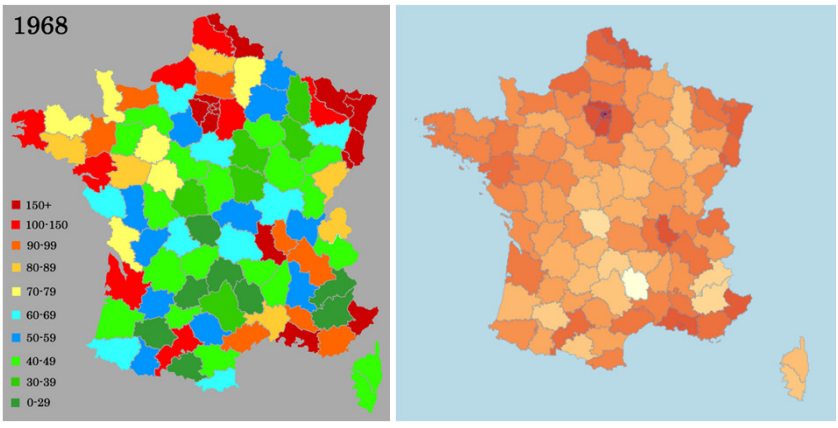
Colour used well can enhance and clarify a presentation. Colour used poorly will obscure, muddle and confuse.

Maureen Stone

Represent data values

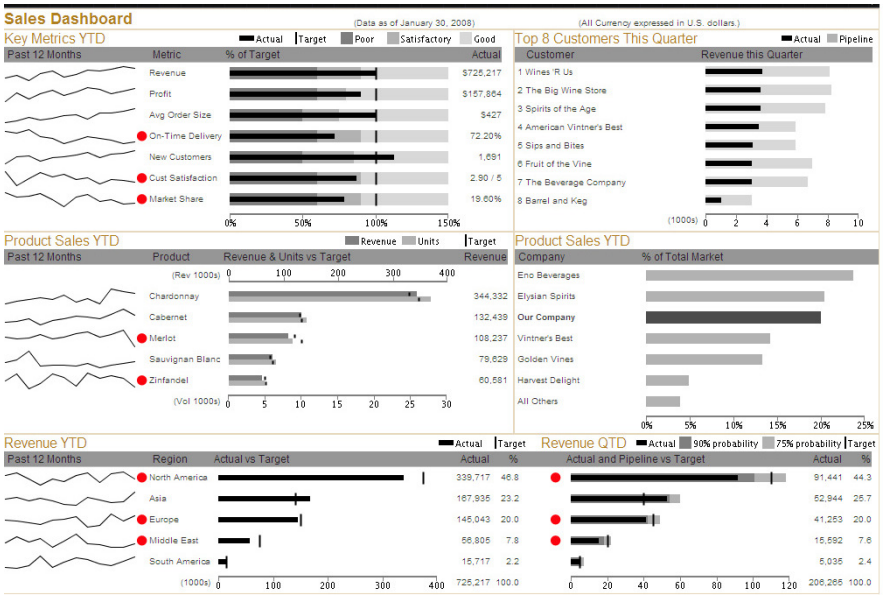
Colour (Hue)

Colour (Saturation)

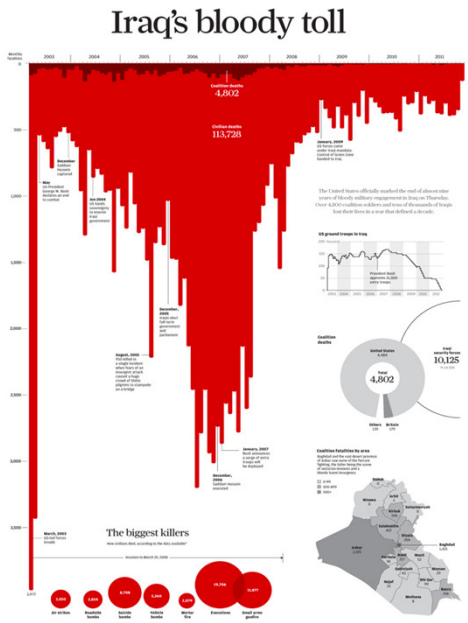


Distinguish between categorical items

Accentuate data



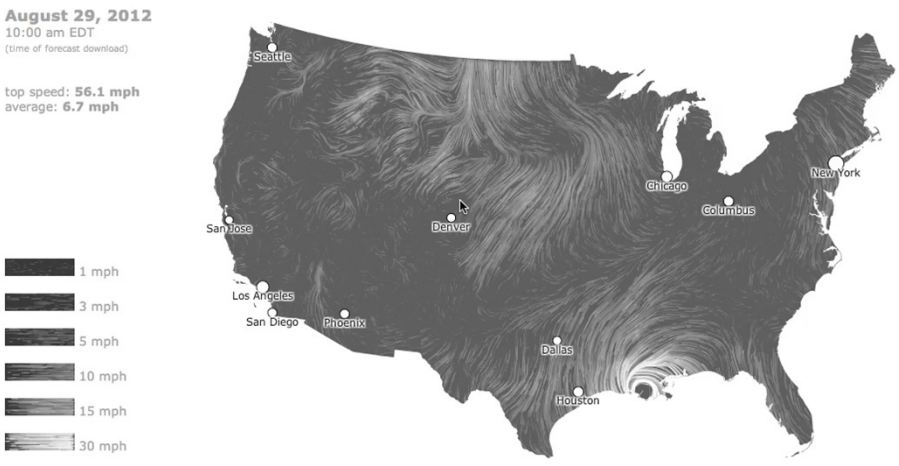
Exploit visual language



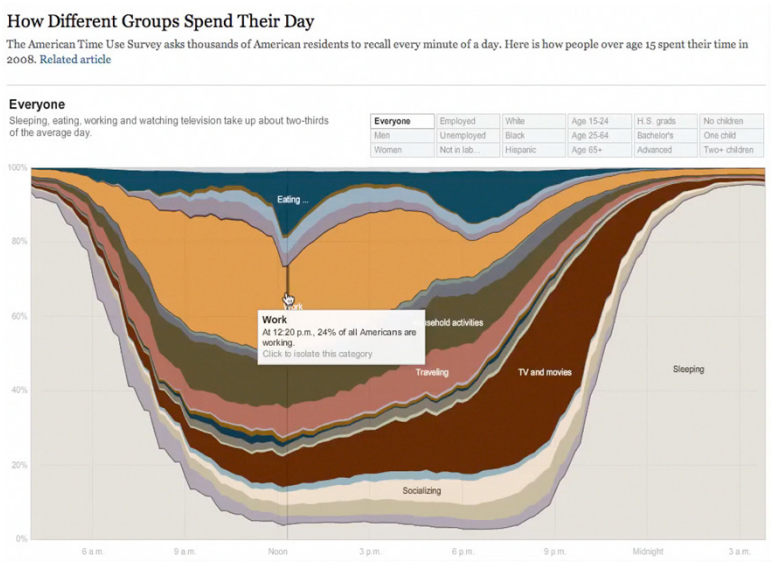
The 5 layers of a visualisation

3. Interactivity

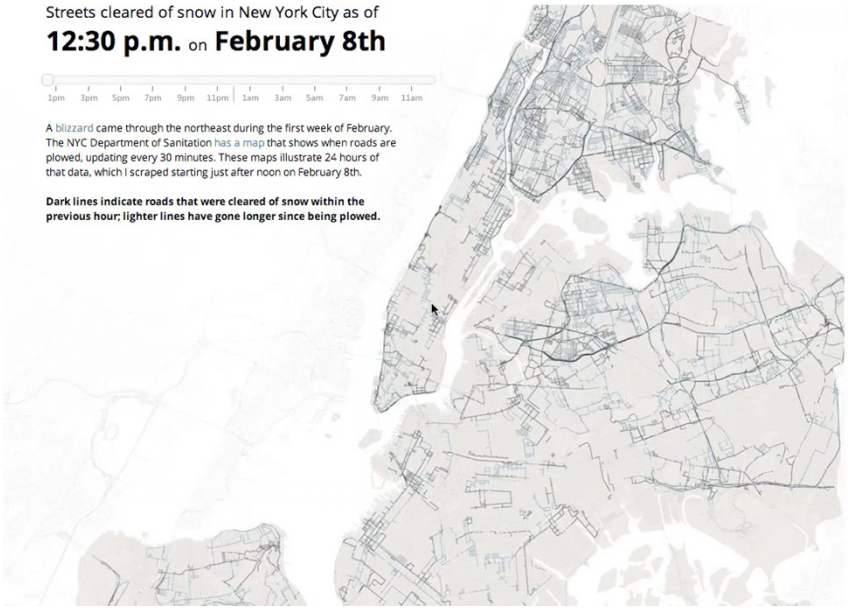
Immersive interactivity



Details on demand



Potential for animation



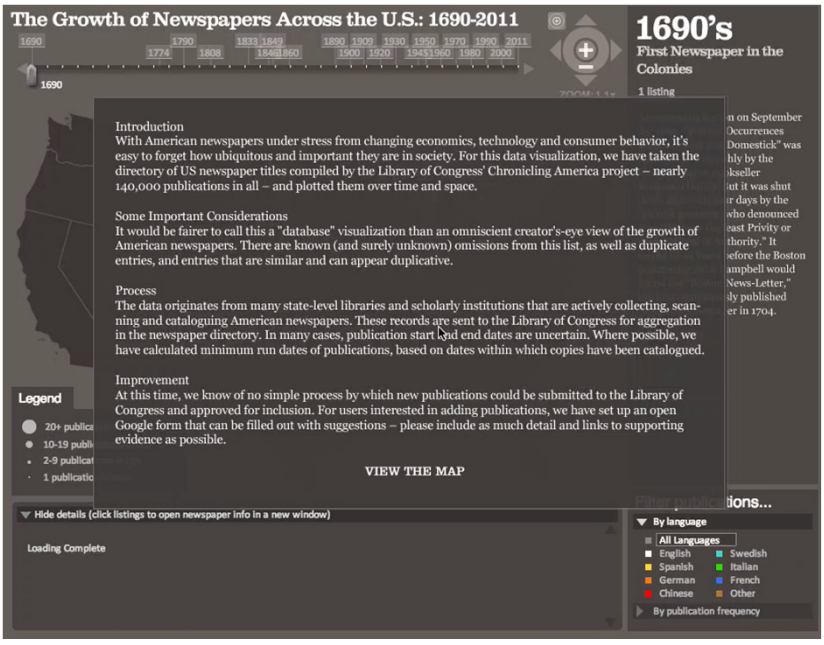
The 5 layers of a visualisation

4. Annotation

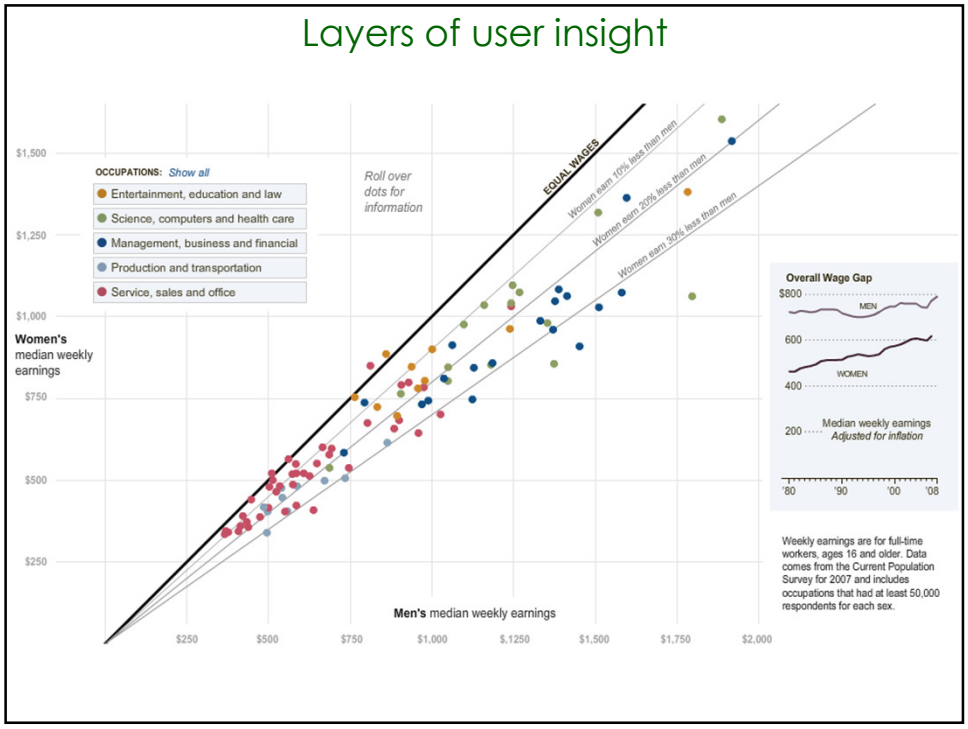
The annotation layer is the most important thing we do... otherwise it's a case of here it is, you go figure it out.

Amanda Cox, Graphics Editor, New York Times

Layers of user assistance



Layers of user insight

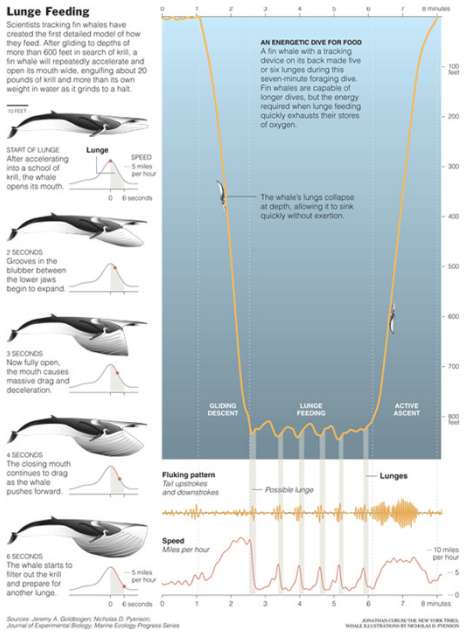


The 5 layers of a visualisation

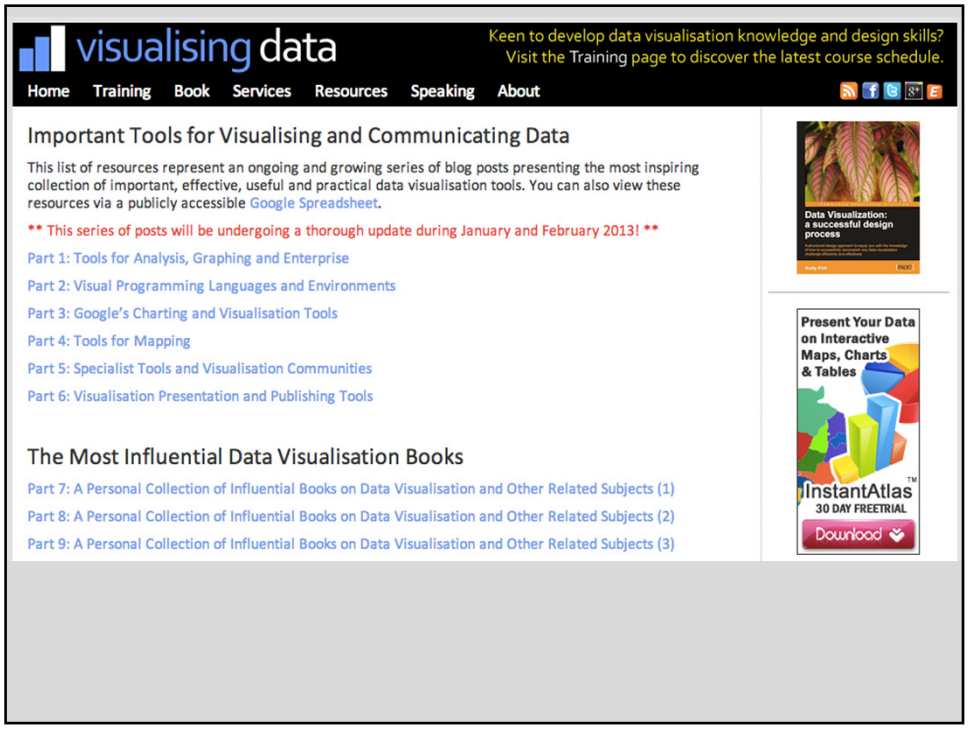
5. Arrangement

Consider the placement of every single visible element in a way that minimises thinking and maximises interpretation

Size, sequence, position, grouping, orientation…



5. Construct and launch your data visualisation solution



Microsoft Excel

Microsoft Excel is the most popular spreadsheet tool in the world with over 400 million users and therefore the most accessible tool for conducting analysis and presenting data in graphical format. The package receives a great deal of justified criticism within the visualisation field for the appalling default and range of bad practice graph designs it promotes, yet in the right hands it can be an incredibly powerful and effective visualisation tool.

Find out more information | Cost: Trial > under £100/$150 per license | Tags: Spreadsheet, Office, Graphing

Good examples and references: Peltier Tech | Excel Charts Blog | Chandoo

Status: Ongoing (July 7, 2011)

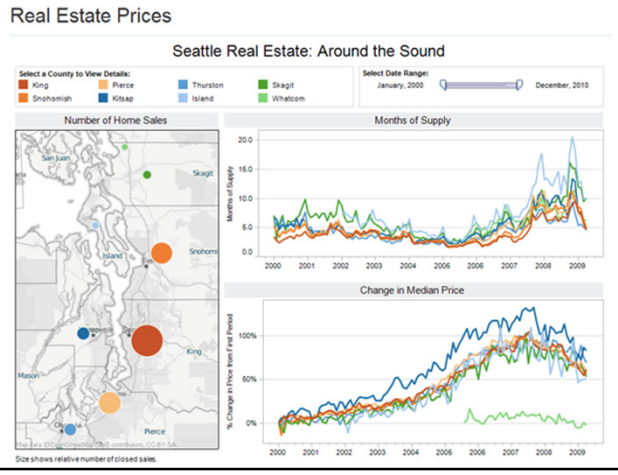
Tableau Desktop

Tableau Desktop is based on breakthrough technology from Stanford University that lets you drag & drop to analyse data rapidly and fluidly, connect to data in a few clicks, then visualise and create interactive dashboards in an instant. Tableau have based their product on years of research to build a system that supports people's natural ability to think visually providing a tool that lets you easily build beautiful, effective, rich data visualisations.

Find out more information | Cost: Trial > £600/$999 Personal, £1200/$1999 Professional | Tags: Statistical Analysis, Business Intelligence, Dashboard

Examples and references: The Information Lab | The Data Studio | Freakalytics

Status: Ongoing (July 7, 2011)



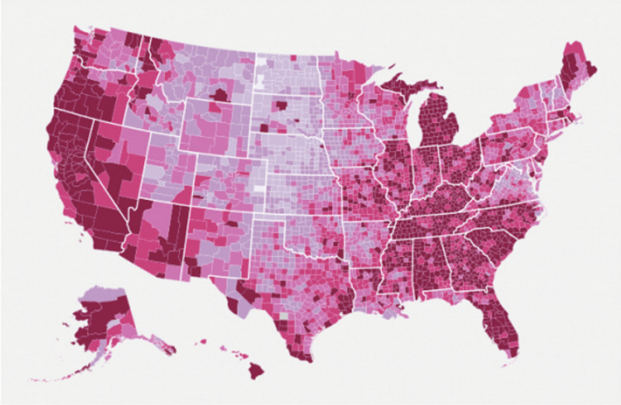
R

R is a highly extensible, open source language and environment for data handling, statistical computing and graphical techniques. One of R's key strengths is the ease with which well-designed publication-quality graphical plots can be produced. Importantly, from a good practice principles perspective, great care has been taken over the defaults for the minor design choices in graphics whilst allowing the user to retain full control beyond. Many people traditionally think of R as a statistics system but its power and potential as a visualisation tool is significant and its popularity within the field is growing constantly.

Find out more information | Cost: Free | Tags: Statistical Programming Language, Graphics, Open Source

Good examples and references: RStudio | Hadley Wickham |ggPlot2 Library | FlowingData

Status: Ongoing (July 7, 2011)



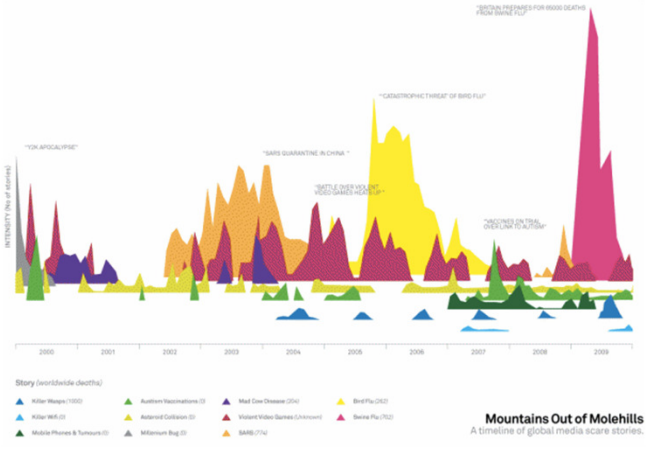
Adobe Illustrator

Like much of the Creative Suite software, Adobe Illustrator is the industry standard application for creating incredible electronic illustrations. Whilst most of us think of creating graphs using statistical software, like many of those listed above, its huge depth of features and capabilities make it the ideal accompaniment tool to take the graphic composition to the next level. Illustrator gives you the complete control you require to polish existing graphs, creating new ones or combine separate elements into an inspiring, customised and hand-crafted infographic design.

Find out more information | Cost: Trial > £375 /$599 per license | Tags: Illustration, Graphics, Graphing

Examples and references: Information is Beautiful | Flowing Data

Status: Ongoing (July 7, 2011)



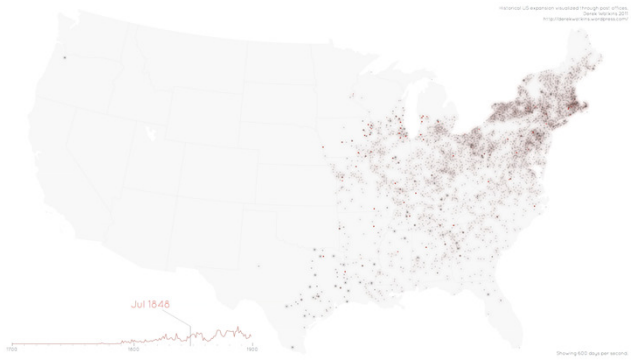
Processing

Processing is an open source programming language and environment for people who want to create images, animations, and interactions. It was founded by Ben Fry and Casey Reas in 2001 while both were students under John Maeda at the MIT Media Lab. In recent recent years it has evolved into one of the most significant and versatile visual programming languages in the visualisation field, particularly off the back of Fry's excellent Visualizing Data book about computational information design which extensively described the potential power of the Processing environment. Today, there are tens of thousands of students, artists, designers, researchers, and hobbyists who use Processing for learning, prototyping, and production.

Find out more information | Cost: Free | Tags: Visual Programming Language, Open Source, Java

Good examples and references: Processing Exhibition Gallery | Ben Fry | Jer Thorp's blprnt.com | 'giCentre' Visualisation Library

Status: Ongoing (July 7, 2011)



D3

D3 is a fresh and efficient visualisation framework created by one of the Protovis creators, Mike Bostock. Its key feature is that it allows you to bind data to a Document Object Model (DOM) and then apply smooth data-driven transformations and interactions to the document. D3's lightweight characteristics enables it to expose the full capabilities of underlying technologies such as CSS3, HTML5 and SVG. Accordingly D3 is extremely fast, supports large datasets and overall provides great flexibility in designing a contemporary web-based visualisation.

Find out more information | Cost: Free | Tags: JavaScript Library, Open Source, SVG

Good examples and references: D3 Gallery | Jan Willem Tulp

Status: Ongoing (July 7, 2011)