

FIT3179 Data Visualisation

Week 01: Introduction to Data Visualisation

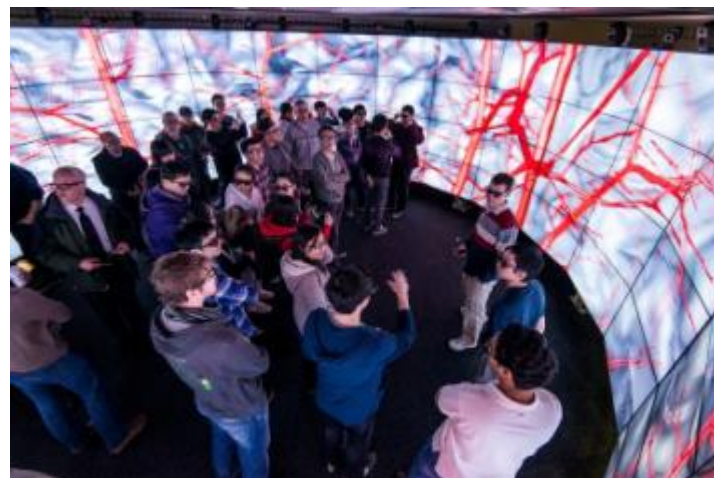
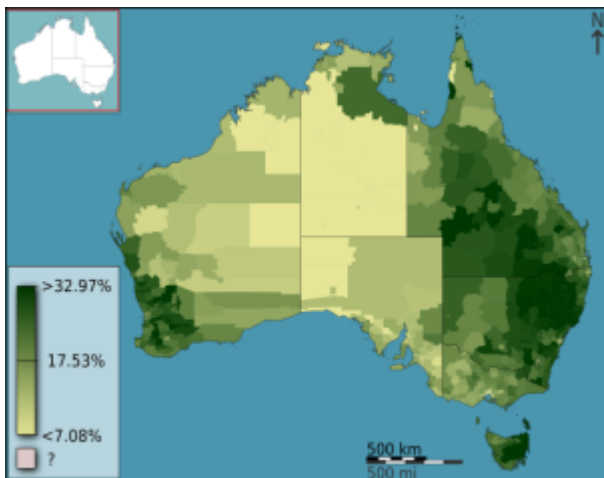


A definition of a Data Viz.

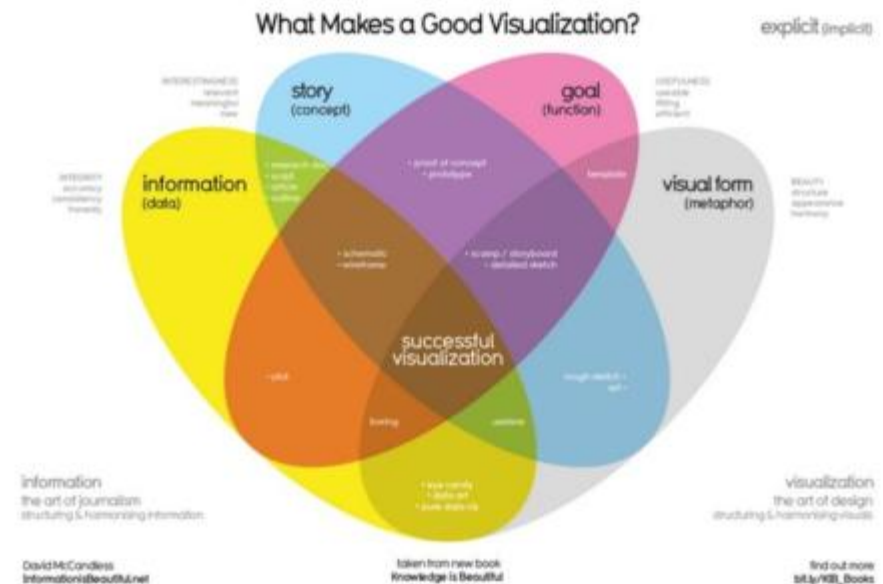
Can we define Data Visualisation?

There isn't a universal definition! For a starting place though: **any visual representation of information**

The process of using Data Visualisation in combination with interactive analysis tools is called *Visual Analytics*



- It is a conjunction of a number of areas:
 - Knowledge Discovery
 - Cognitive Science
 - Graphic Design
 - Interactive Computer Graphics
 - *Data Science*
- You may have experience in some or all of these already.
.. we just need to put them together



- A visualisation **transforms data into information** (then *understanding* and *insight*) and makes it useful to people
- Clichés...
 - “Seeing is believing”
 - “A picture is worth a thousand words”



- Making a Visualisation is often thought of as process of making a graphic or an image
- But really it is a cognitive process
 - Form a mental image of something
 - Internalize an understanding
 - “The purpose of visualization is insight, not pictures”
 - Insight: discovery, decision making, explanation



The History of Visualisations

- Humans acquire more information through vision than any of the other senses combined
- Human Vision
 - Highest bandwidth sense
 - Fast, parallel
 - Pattern recognition
 - Pre-attentive
 - Extends memory and cognitive capacity
 - People think visually
- Creating an image is a really effective way to transmit information

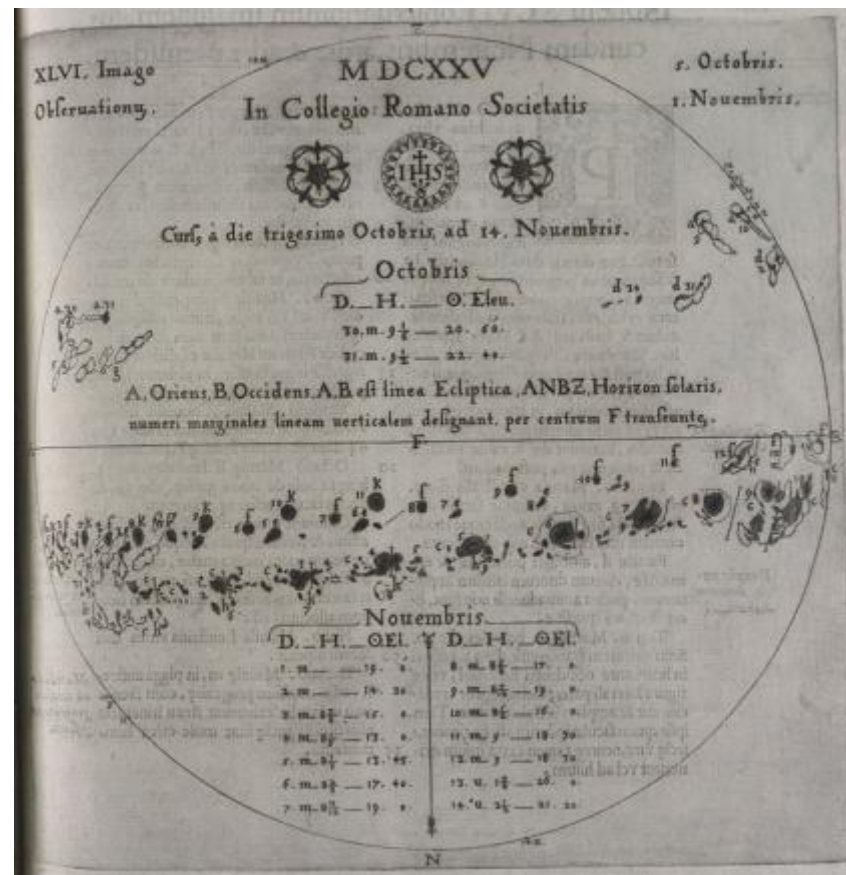
We can argue that the first viz. is the first image!

- **Narrow View:**
 - Started with Computer Graphics
 - 1987 first journal on Computer Graphics and Visualisation
- **Wider View:**
 - Cave Paintings, Hieroglyphics, Maps, Astrological Charts... Info Graphics
- **Which view? Depends on your philosophies**

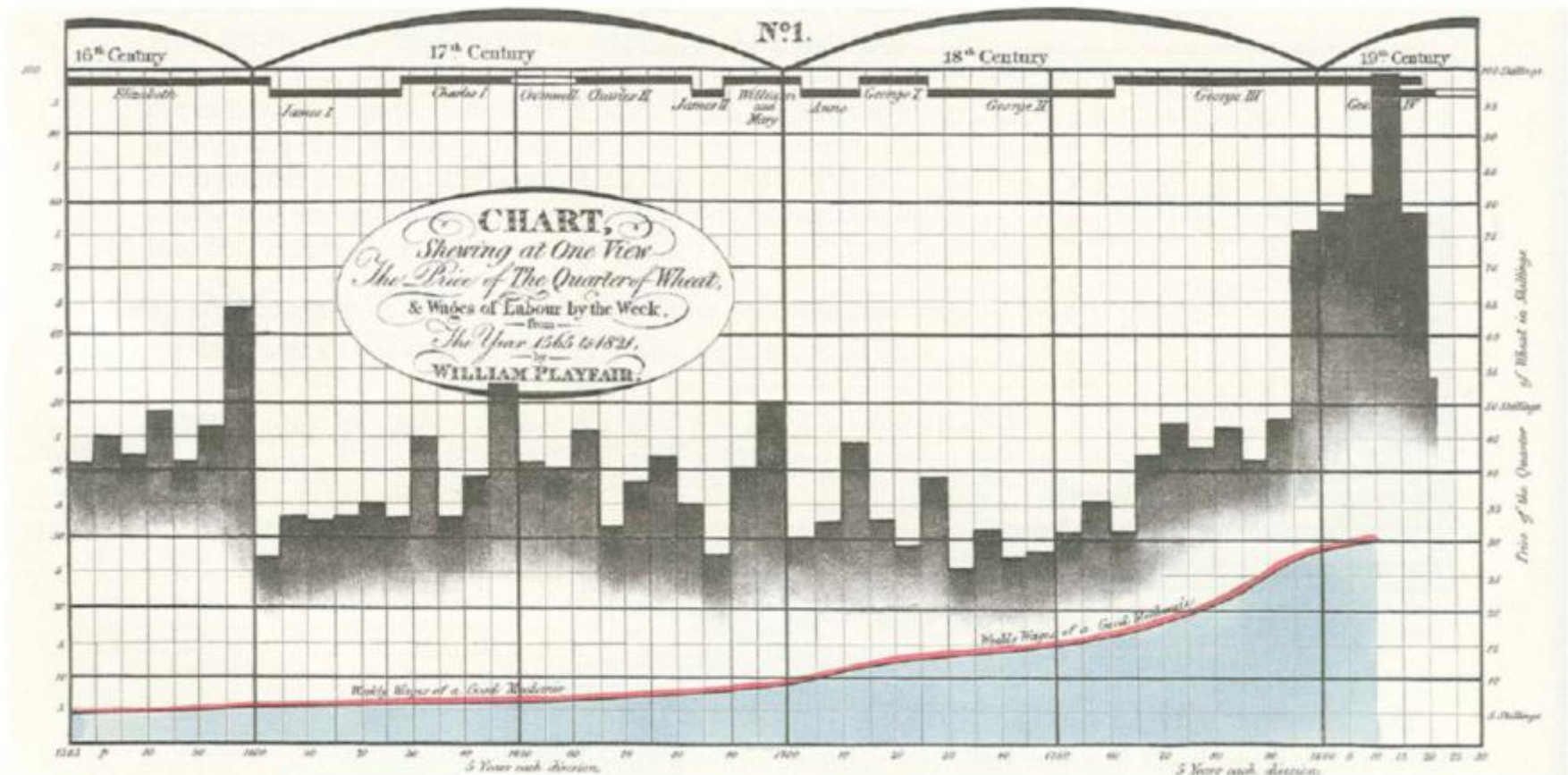


Rock art at Narwala Garbarnmang in Arnhem Land (from D. Bruno et al, 2011)

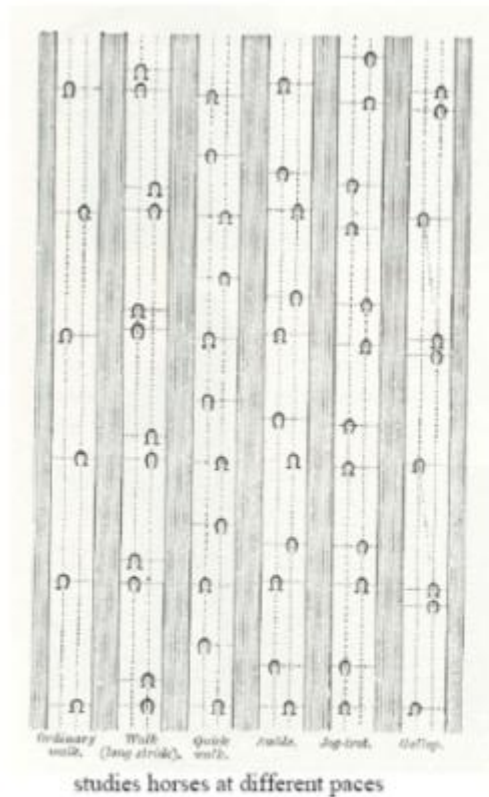
Astronomical Charts (1600's)



Graphs (William Playfair, 1700's)



Time Series (Etienne-Jules Marey, 1800's)

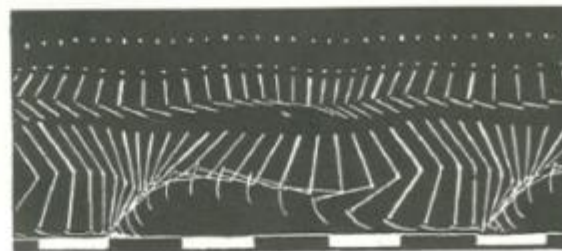


the undulations of the dorsal fin of a descending sea-horse,



the advances of the gecko

Marey's man in black velvet, photographed in stick-figure images, became the time-series forerunner of Marcel Duchamp's *Nude Descending a Staircase*.

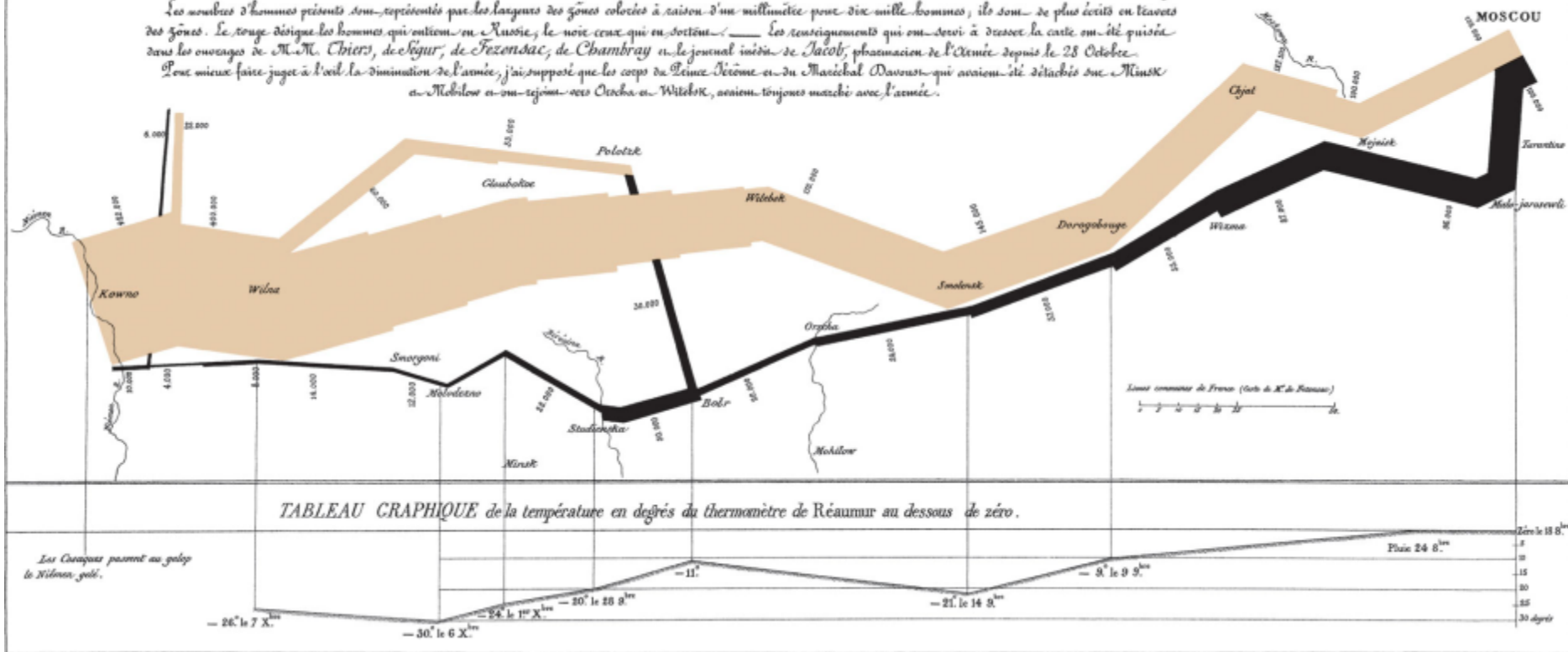


Napoleon's March by Charles Minard (1861)

Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite. Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en traits des zones. Le rouge désigne les hommes qui ont été en Russie; le noir ceux qui ont servi à dresser la carte ou à être prisés dans les ouvrages de M. M. Thiers, de Foy, de Fezensac, de Chambray et le journal inédit de Jacob; phrasologie de l'Armée depuis le 28 Octobre. Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davoust qui avaient été détachés sur Minsk et Mohilew et qui se rejoindront à Orel et Wilna, avaient toujours marché avec l'armée.



Aug. par Regnier, à Paris, 57° Mars 57° 0" à Paris.

Imp. Lab. Regnier et Devot.

London Underground Map based on a design by Harry Beck (1931)





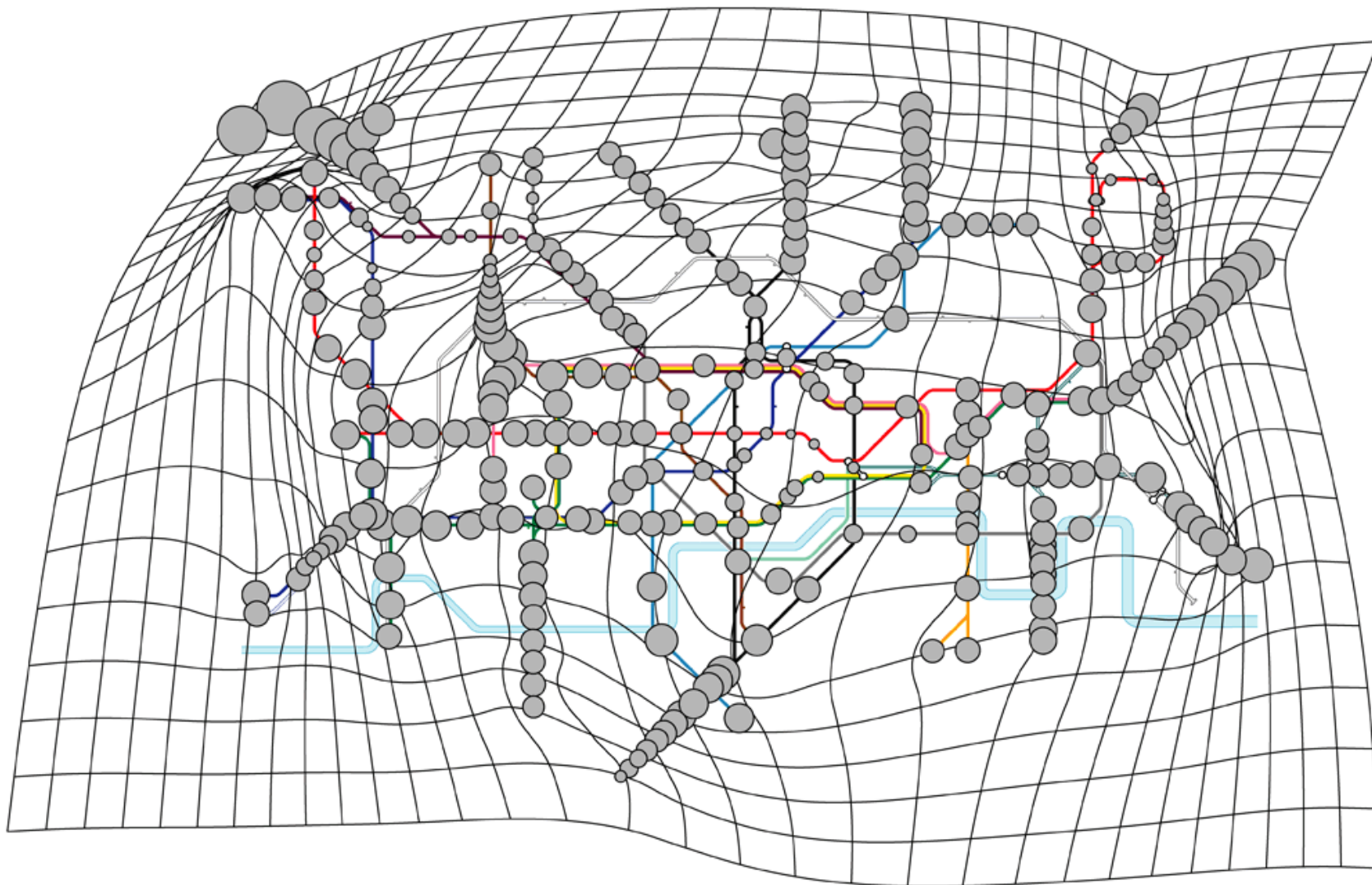
Source: http://ni.chol.as/media/geoff-files/sillymaps/large_geographical_map.gif



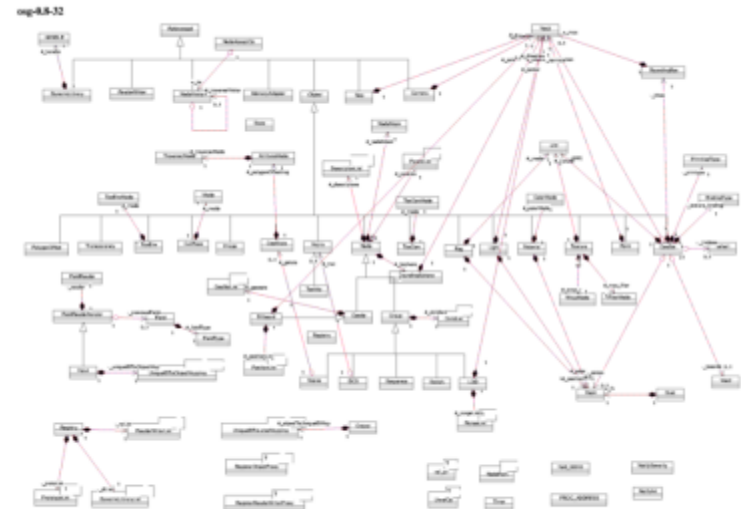
Source: <http://ni.chol.as/media/geoff-files/sillymaps/tubegeo.jpg>



Source: Jenny, B. 2006. Geometric distortion of schematic network maps.



Visualisation of Scientific and Engineering Data



Basic steps in Building a Viz.

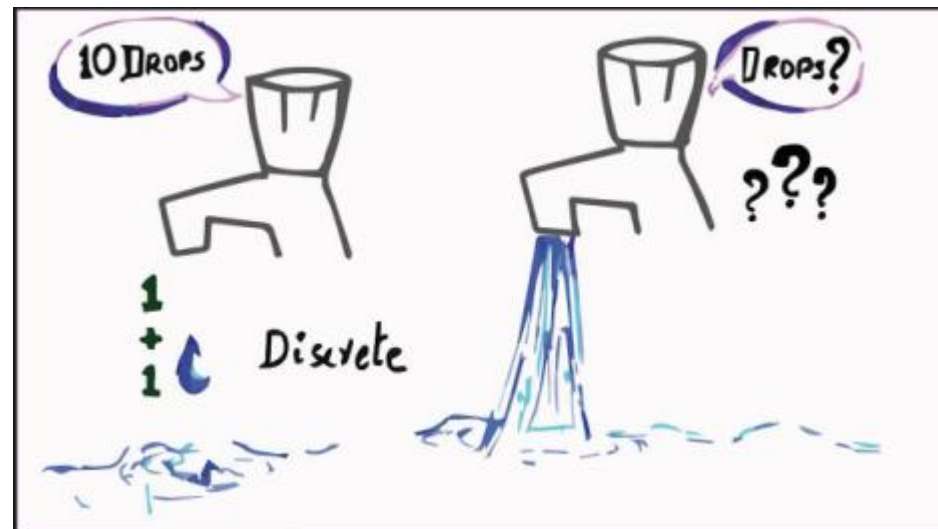
So maybe all this talk has made you interested in building your own viz. What would you need to do?

- Get data
 - Evaluate data in some way
 - Consider Interaction principles
 - Compose data into useful sets (build the viz)
 - Reduce data clutter

We'll cover these steps in detail during the semester!

Data input can be:

- Static (AKA *discrete*) information (e.g. one point in time)
- Dynamic (AKA *continuous*) information – what reality is
 - Hard for use to see continuous info from discrete data
 - Visualisation helps us see the continuous

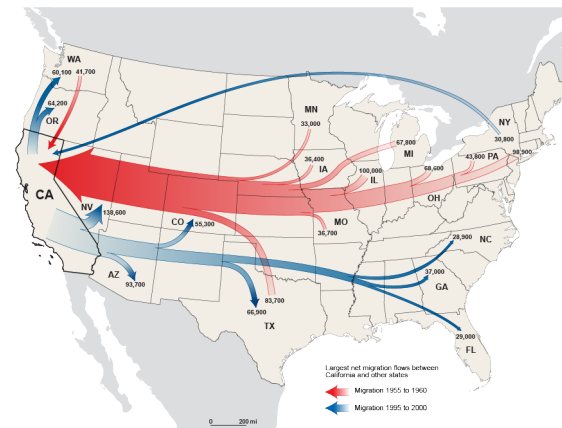


Data can belong to different dataset types:

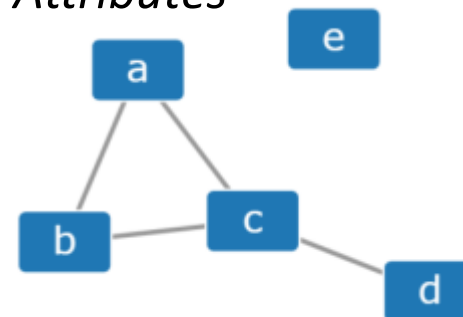
- Tables: *Items, Attributes*
- Geometry: *Items, Positions*

Table 1: Bilateral Remittance Estimates for 2011 using Migrant Stocks, Host Country Income, and Origin Country Income (millions of USD)

	FR	BY	GR	ES	BY	SE	PK
Remittance-receiving country (sender)							
Remittance-sending country (receiver)	France	Papua New Guinea	Paraguay	Pers	Philippines	Poland	Portugal
1 Afghanistan	0	0	0	0	0	0	0
2 Albania	0	0	0	0	0	0	0
3 Algeria	0	0	0	0	0	0	0
4 American Samoa	0	0	0	0	0	0	0
5 Andorra	0	0	0	0	0	0	15
6 Angola	0	0	0	0	0	0	30
7 Antigua and Barbuda	0	0	0	0	0	0	0
8 Argentina	0	0	0	524	188	29	0
9 Armenia	0	0	0	0	0	0	0
10 Austria	0	0	0	0	0	0	0
11 Australia	1	7	1	21	806	155	39
12 Austria	0	0	0	3	64	188	2
13 Azerbaijan	0	0	0	0	0	0	0
14 Bahamas, The	0	0	0	0	0	0	0
15 Bahrain	0	0	0	0	130	0	0
16 Bangladesh	0	0	0	0	0	0	0
17 Barbados	0	0	0	0	0	0	0
18 Belarus	0	0	0	0	0	534	0
19 Belgium	0	0	0	1	6	44	105
20							48



- Networks and Trees: *Items (nodes), Links, Attributes*



Data attributes can allow sorting/grouping in different ways

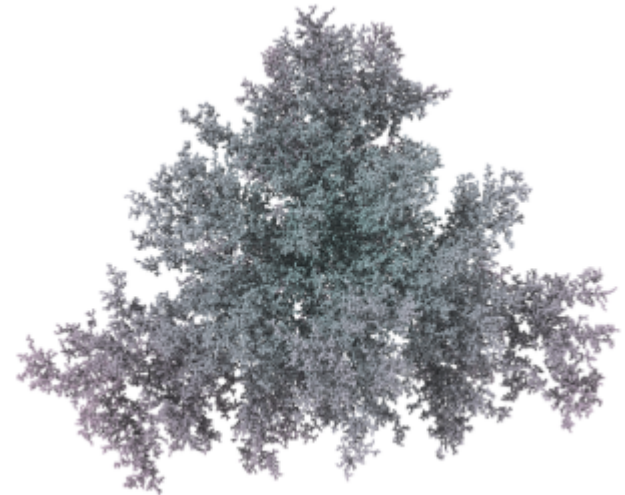
- Categorical: *departments, gender, state...*
- Ordered: *age bracket, day of the week...*
- Quantitative: *\$, voltage, counts of things, latitude/longitude...*



Generally it's better to have an image than a number

Humans are not good at interpreting numbers

- Better to combine multi-dimensional information into a single, easily understandable form
- Easier to extract and emphasise important info



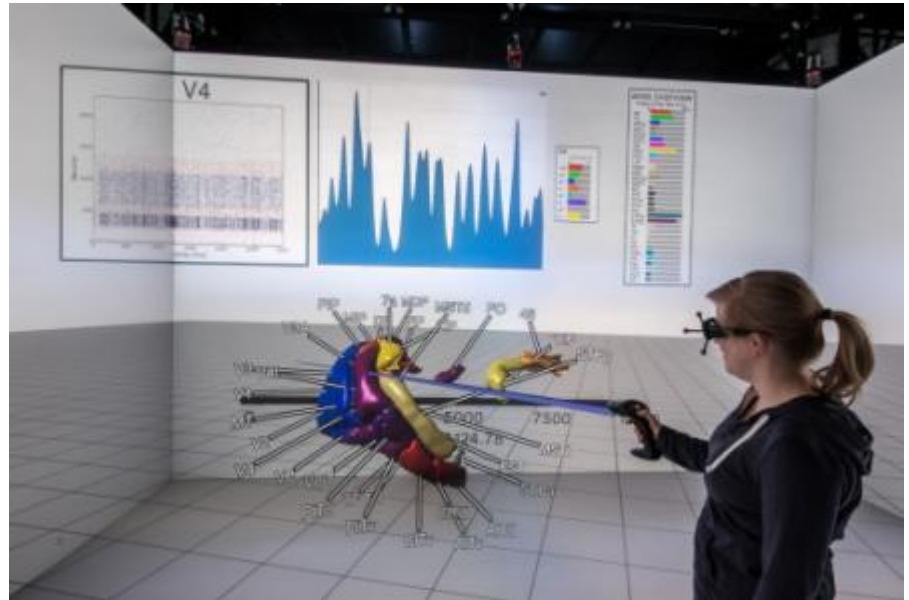
- In building a viz., there are three basic limitations to visualisations
 - Computer limits (for interactive vis)
 - Human perceptual & cognitive limits
 - http://en.wikipedia.org/wiki/Misleading_graph
 - <http://www.abc.net.au/news/2013-05-08/ericho--read-between-the-lines/4674322>
 - Display limits (run out of pixels to show fine changes)
- The designer always needs to *trade-off* between showing as much data as possible and reducing clutter




Modern Interactive Visualisations

To resolve the inherent tension between ‘more data’ vs ‘less clutter’, computers allow us to provide *interactivity*

- Allows us to show multiple different perspectives on the data
- Larger data sets maybe easier to work with now too (zooming, expanding, clicking down into data)
- The user can see what they need without distraction




[Contact Us](#)

[Discover Spotfire](#)
[Services](#)
[Resources](#)
[Community](#)
[About Us](#)

[Demos](#)
[Free Trial | Buy Spotfire](#)

FEATURED DEMO

Data Exploration features in TIBCO Spotfire 5.5



See the new Data Exploration features which are available in the TIBCO Spotfire 5.5 release.

Video Only

FEATURED DEMO

Production Optimization of Natural Gas



This analysis uses TIBCO Spotfire and S+ to analyze natural gas well data and predict the expected ultimate recovery (EUR) of potential well sites.

Video Available



Retail Cross-Sell Revenue Optimization Analysis

This is a demonstration of how TIBCO Spotfire can enable a retailer to design and run a real-time targeted cross-sell campaign to maximize revenue lift.

Video Only



Accounts Receivables Analysis

Do detailed analysis of account receivables aging in Spotfire!

RESET CRITERIA

Application

- ☐ Business Planning 8
- ☐ Chemistry 4
- ☐ Clinical Development 5
- ☐ Cross-Sell and Up-Sell 2
- ☐ Exploration and Production 6
- ☐ Finance 13
- ☐ Intelligence and Defense 4
- ☐ Manufacturing Operations 3
- ☐ Miscellaneous 9
- ☐ Pharmaceutical Research 2
- ☐ Portfolio Management 3
- ☐ Product Development 4
- ☐ Regulatory Compliance 5
- ☐ Risk Management 2
- ☐ Sales and Marketing 12
- ☐ Sports and Pop Culture 2
- ☐ Warranty 1
- ☐ Workforce Management 5

OneBayArea

OneBayArea / Maps

Traveling from:

37.763900,-122.413000

Locate

or + select the center of the map

by:



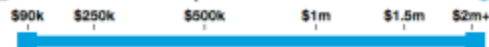
when:



☒ Show travel times between:



☒ Show median home prices between:



 For real-time travel information

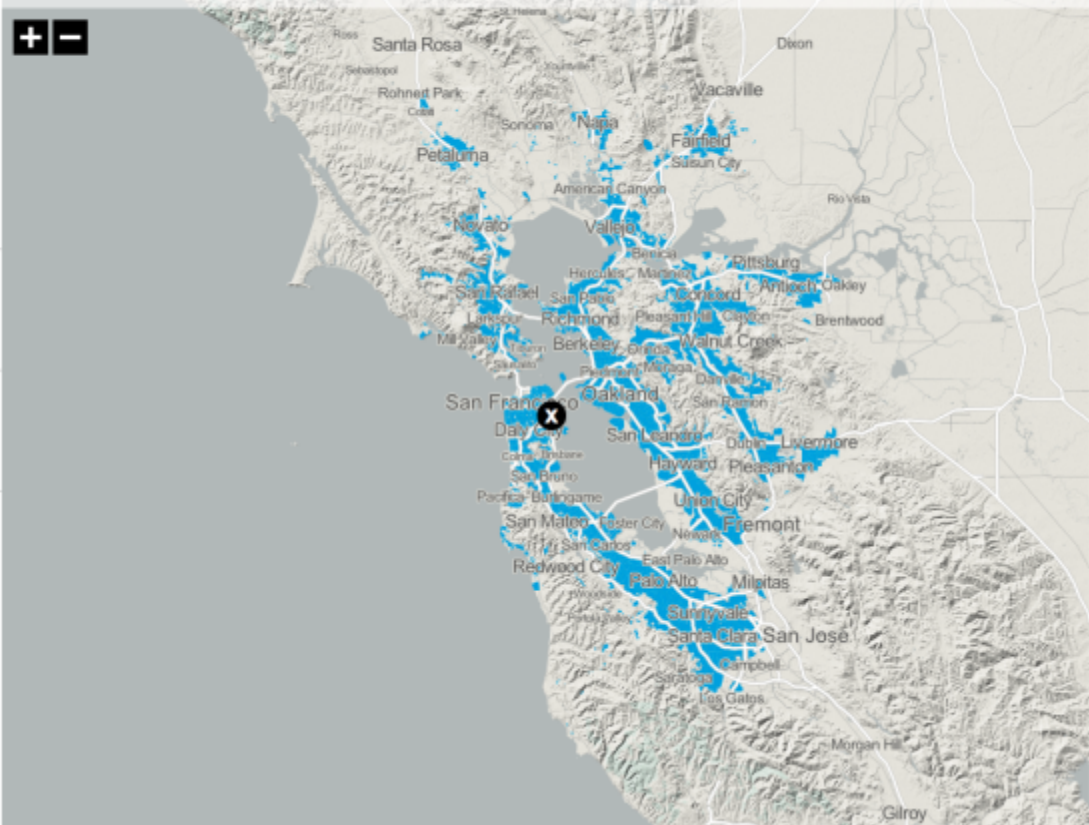
 Give us feedback!

 Follow OneBayArea on Twitter

 OneBayArea on Facebook



Blue indicates Bay Area places accessible from **X** by single-occupant auto in 1 hour during the **AM commute** with home prices between **\$90k and \$2m+**



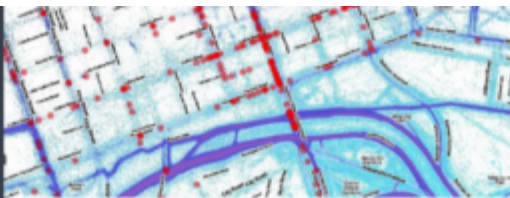
Source: http://maps.onebayarea.org/travel_housing/

Monash CityX (formerly *City Science*)



Melbourne Age Map

Australia has an ageing population. Millennials are changing the face of our cities. In this project, we visualize the spatial distribution of population by age in Melbourne.



Melbourne Bike Crash Map

Bicycle usage is growing in Melbourne. However, cycling safety still remains an issue. This project visualizes 5 years' worth of bicycle crashes in Melbourne.



Melbourne Housing Density Map

With the rapidly growing population of Melbourne, housing has become a major topic of debate. In this project, we visualize the distribution of dwelling density in Melbourne.



Science Isn't Broken

Hack Your Way To Scientific Glory

You're a social scientist with a hunch: **The U.S. economy is affected by whether Republicans or Democrats are in office.** Try to show that a connection exists, using real data going back to 1948. For your results to be publishable in an academic journal, you'll need to prove that they are "statistically significant" by achieving a low enough p-value.

1 CHOOSE A POLITICAL PARTY

Republicans

Democrats

2 DEFINE TERMS

Which politicians do you want to include?

- ☐ Presidents
- ☐ Governors
- ☒ Senators
- ☐ Representatives

How do you want to measure economic performance?

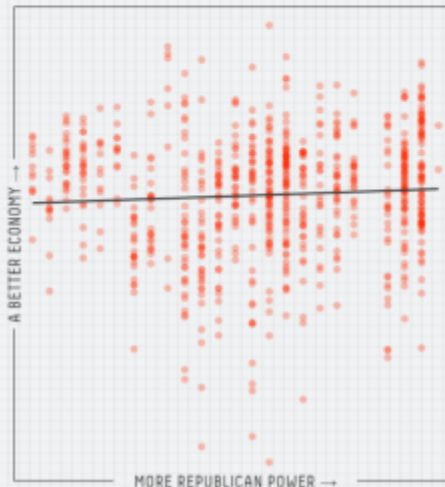
- ☐ Employment
- ☒ Inflation
- ☐ GDP
- ☒ Stock prices

Other options

- ☒ Factor in power
Weight more powerful positions more heavily
- ☐ Exclude recessions
Don't include economic recessions

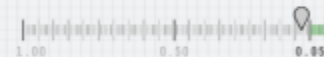
3 IS THERE A RELATIONSHIP?

Given how you've defined your terms, does the economy do better, worse or about the same when more Republicans are in power? Each dot below represents one month of data.



4 IS YOUR RESULT SIGNIFICANT?

If there were no connection between the economy and politics, what is the probability that you'd get results at least as strong as yours? That probability is your p-value, and by convention, you need a p-value of 0.05 or less to get published.



Result: Almost

Your **0.08** p-value is close to the 0.05 threshold. Try tweaking your variables to see if you can push it over the line!

If you're interested in reading real (and more rigorous) studies on the connection between politics and the economy, see the work of Larry Bartels and Alan Blinder and Mark Watson.

Data from The @unitedstates Project, National Governors Association, Bureau of Labor Statistics, Federal Reserve Bank of St. Louis and Yahoo Finance.

Research and Readings

- <http://flowingdata.com>
- <http://www.visualcomplexity.com/vc/>
- <http://www.informationisbeautiful.net/>
- <http://datavisualization.ch/>
- <http://www.visualizing.org/>
- <http://www.smashingmagazine.com/2009/09/11/25-useful-data-visualization-and-infographics-resources/>