Edinburgh Napier UNIVERSITY

Data Analytics
SET10109

Introduction to Visualisation

David Hunter
(d.hunter@napier.ac.uk)

Edinburgh Napier UNIVERSITY

Admin

- Contact details:

David Hunter
room C49
d.hunter@napier.ac.uk

- Feedback is most welcome!!
- We will have a break between lectures ☺
- Lecture and coursework material based on course written by Natalie Kerracher

Edinburgh Napier UNIVERSITY

General Overview

General overview of what we will cover in the data visualisation lectures (7 lectures):

1. Introduction to Visualisation

2-3. Understanding Data: data types, visual encodings; mapping data types to visual encodings

4-5 Main design principles; cognitive/perceptual factors

6. Focus on visual techniques for different types of data (tabular data, trees and networks, multivariate data, timeseries data); interaction techniques.

7. Evaluation of visualisation techniques.

Visualisation will be covered this week (**week 5**) and **weeks 10-12** (November)

Edinburgh Napier UNIVERSITY

Main Visualisation Texts

Both are available electronically

Daniel Keim, Jörn Kohlhammer, Geoffrey Ellis and Florian Mansmann (Eds.), Mastering the Information Age – Solving Problems with Visual Analytics, Eurographics Association, 2010.

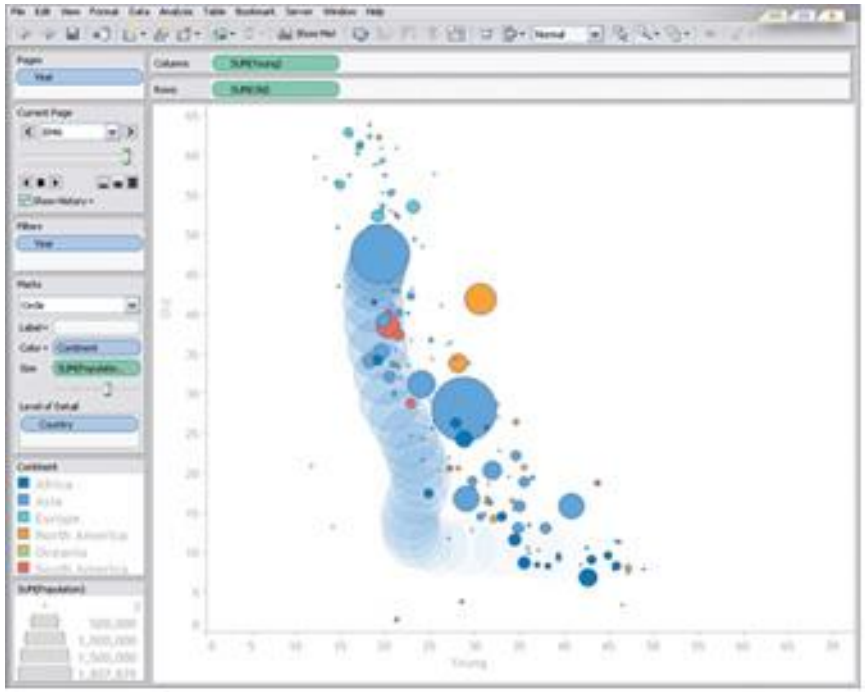
Electronic version available to download at: http://www.vismaster.eu/book/

Tamara Munzner (to appear 2014). Visualization Analysis and Design: Abstractions, Principles, and Methods

Pre-publication draft available at: http://www.cs.ubc.ca/~tmm/courses/533-11/book/

Edinburgh Napier UNIVERSITY

Tools



- We will be using Tableau to create interactive visualisations: http://www.tableausoftware.com/academic
- Will be discussed further in tomorrow's lab.

Image: http://www.tableausoftware.com/public/about-tableau-products

Edinburgh Napier UNIVERSITY

Overview of this lecture: Introduction to visualisation

- What is visualisation?
- Motivations: when and why to use a visual approach
- Visualisation for presenting data vs. visualisation for analysis
- A very brief history of visualisation

Edinburgh Napier UNIVERSITY

Suggested readings for this lecture

Munzner, chapter 1

Dix, A. (2013). Introduction to information visualisation. In *Information Retrieval Meets Information Visualization* (pp. 1-27). Springer Berlin Heidelberg. http://link.springer.com/book/10.1007/978-3-642-36415-0/page/1

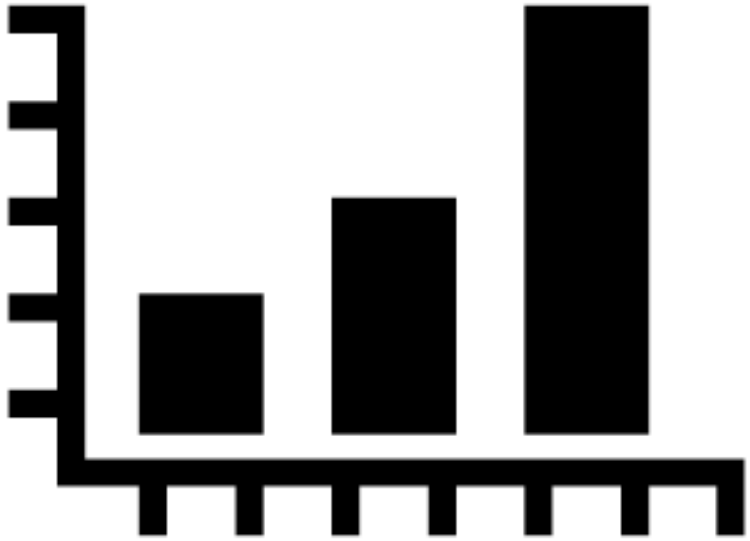
Fekete, J. D., Van Wijk, J. J., Stasko, J. T., & North, C. (2008). The value of information visualization. In *Information Visualization* (pp. 1-18). Springer Berlin Heidelberg. http://link.springer.com/chapter/10.1007/978-3-540-70956-5_1

Edinburgh Napier UNIVERSITY

WHAT IS VISUALISATION AND WHY SHOULD WE USE IT?

Edinburgh Napier UNIVERSITY

Visualisation: more than just barcharts



Humans are VISUAL creatures

Edinburgh Napier UNIVERSITY

touch

echolocation/sonar

smell

sight

"The visual cortex accounts for around 50% of our brain…it makes sense to use it."
(Dix, 2013)

Applications of visualisation

Edinburgh Napier UNIVERSITY

Presentation vs. Analysis

- Communicate data and ideas
- Explain and inform
- Influence and persuade
- Provide evidence and support

Goals:

- Clarify
- Focus
- Highlight
- Simplify

- Understand
- Compare
- Evaluate
- Judge
- Decide

Goals:

- Show many variables
- Illustrate overview + detail
- Facilitate comparison

Based on Professor John Stasko's seminar: The Value of Visualization for Exploring and Understanding Data, given at Edinburgh Napier University, 16th July 2013

Edinburgh Napier UNIVERSITY

VISUALISATION FOR PRESENTATION

Visualisation for *presentation*: why?

Edinburgh Napier UNIVERSITY

- Compact way to represent data
  - *"A picture is worth a thousand words…"*
- Understandable to a wide audience
  - not just statisticians: public, colleagues, decision makers
- Engaging, high impact
  - Simply presenting data visually can have a profound impact
- Convincing
  - you and others of the value of your data
  - some people just don't trust data until

    "they see it with their own eyes"
- Attractive
  - People prefer pretty pictures, even when pictures aren't actually giving them a lot
- Often static

thedetail.tv PSNI district and areas covered	Number of licence holders	Total number of weapons
F: Cookstown, Omagh, Fermanagh and Dungannon & South Tyrone	13,049	31,321
E: Craigavon, Newry & Mourne, Banbridge and Armagh	13,053	30,368
H: Coleraine, Ballymoney, Moyle, Ballymena and Larne.	8,842	21,677
C: Castlereagh, North Down, Ards and Down	7,706	21,586
G: Strabane, Foyle, Limavady and Magherafelt	7,997	19,737
D: Antrim, Carrickfergus, Lisburn and Newtownabbey	7,393	17,723
B: South and East Belfast	779	2,023
A: North and West Belfast	686	1,838

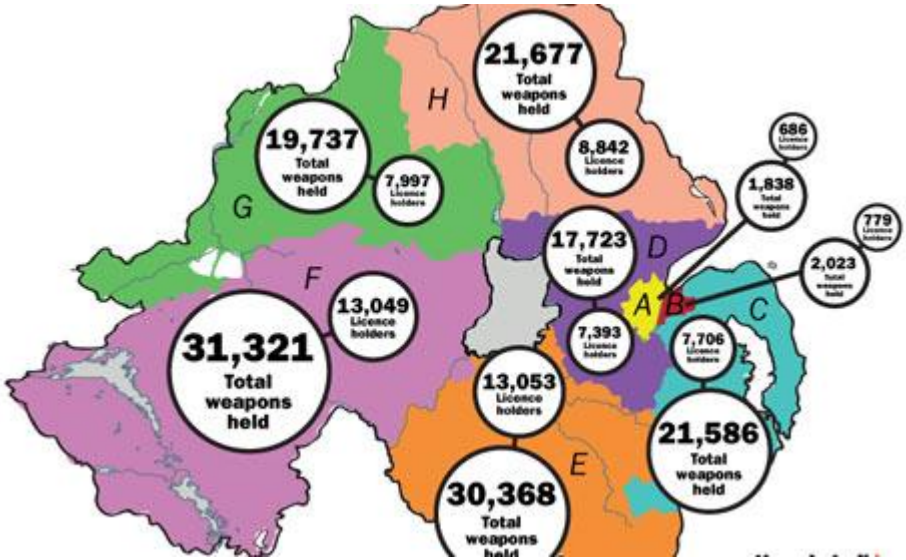


Image: Chris Scott - http://www.guardian.co.uk/news/datablog/2012/aug/23/arms-licence-northern-ireland

A good example of visual presentation…

Hans Rosling's 200 Countries, 200 Years – The Joy of Stats (4 minutes)
http://www.youtube.com/watch?v=jbkSRLYSojo

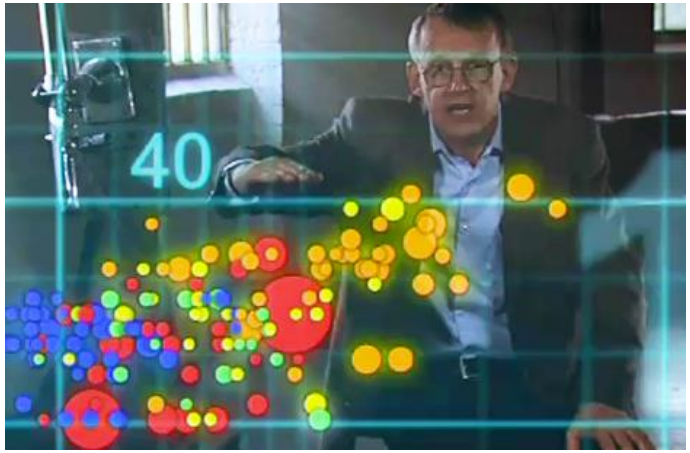
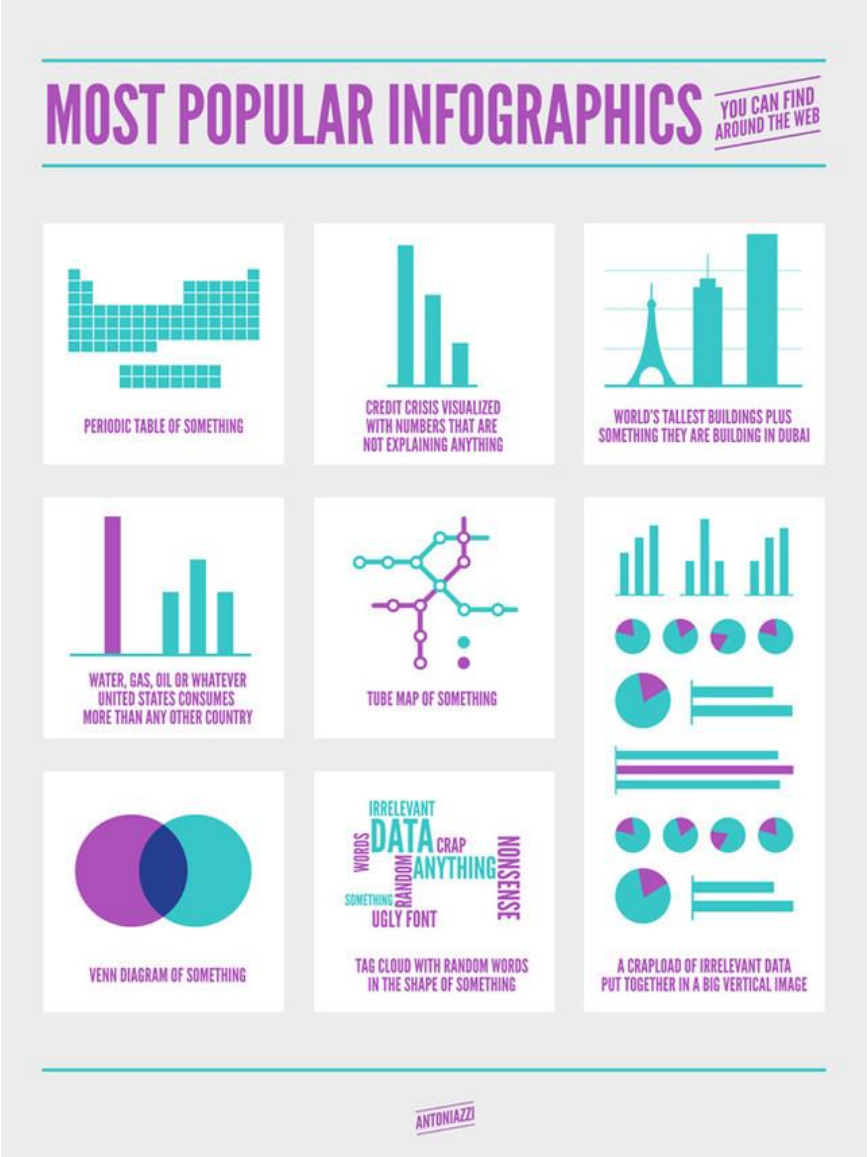


Image: http://www.stashmedia.tv/?p=6666

"Hans Rosling's famous lectures combine enormous quantities of public data with a sport's commentator's style to reveal the story of the world's past, present and future development. Now he explores stats in a way he has never done before - using augmented reality animation. In this spectacular section of 'The Joy of Stats' he tells the story of the world in 200 countries over 200 years using 120,000 numbers - in just four minutes. Plotting life expectancy against income for every country since 1810, Hans shows how the world we live in is radically different from the world most of us imagine"

Visual presentation: beware…



Most Popular Infographics You Can Find Around The Web, Alberto Antoniazzi http://www.albertoantoniazzi.com/

"*There are three kinds of lies: lies, damned lies, and statistics*." (Twain/Disraeli)

Be careful when representing data visually!

- For *presentation*, viewer must be able to interpret the visual easily and rapidly
- Avoid meaningless ,confusing or misleading representation of data (unintentional or otherwise)

"*the visualization design space is huge and full of tradeoffs; most visualization designs are ineffective."* (Munzner, 2014, p3)

- We will cover good and not-so-good approaches to visual representation in subsequent lectures

- Some good bad infographic examples… http://www.theguardian.com/news/datablog/gallery/2013/aug/01/16-useless-infographics

Edinburgh Napier UNIVERSITY

VISUALISATION FOR ANALYSIS

Edinburgh Napier UNIVERSITY

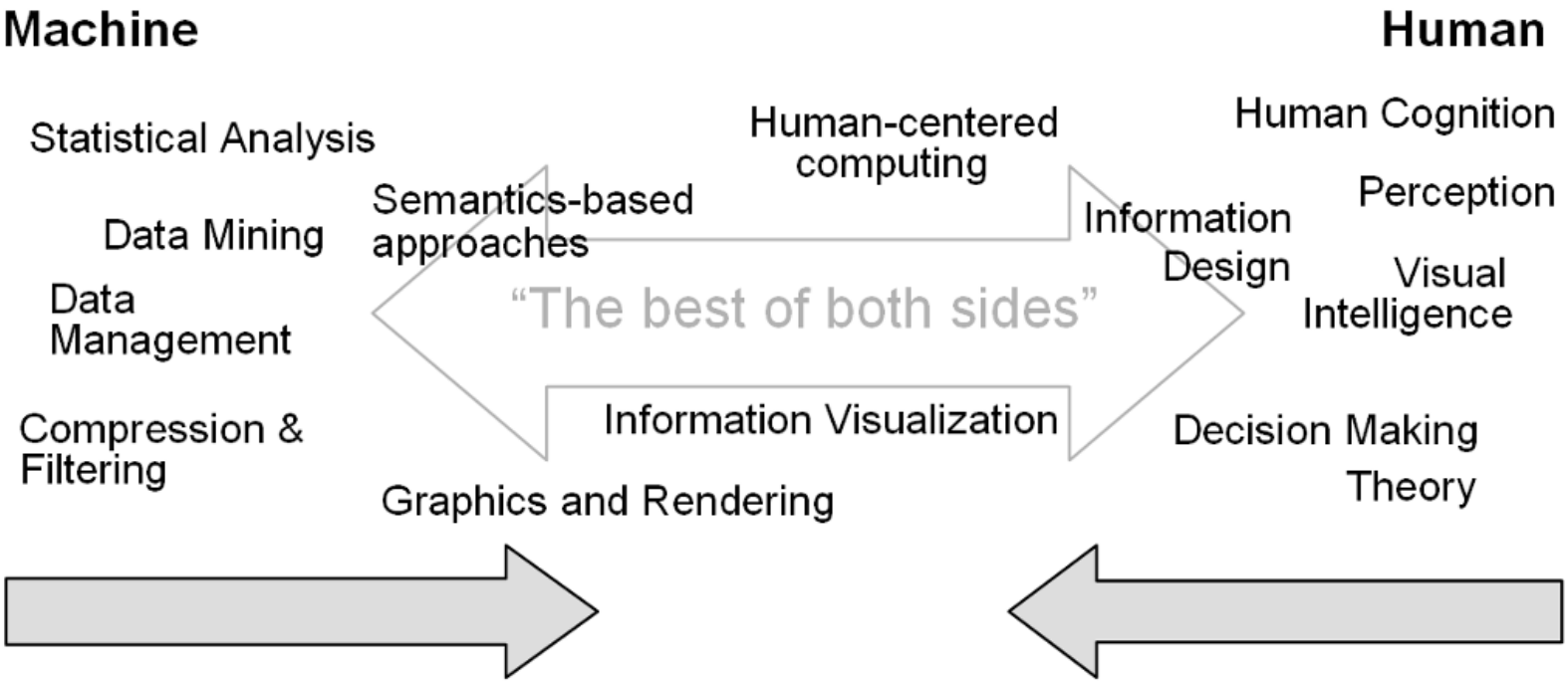
Definitions

- Data Visualisation
- Information Visualisation (InfoVis)
- Scientific Visualisation (SciVis)
- Geographic Visualisation (GeoVis)
- Knowledge Visualisation

Overlap:

  - Say inIf your data is scientific, which category?
  - formation is data with meaning, but what if someone decides it has no meaning to them?

Visual Analytics



"Visual analytics combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets.

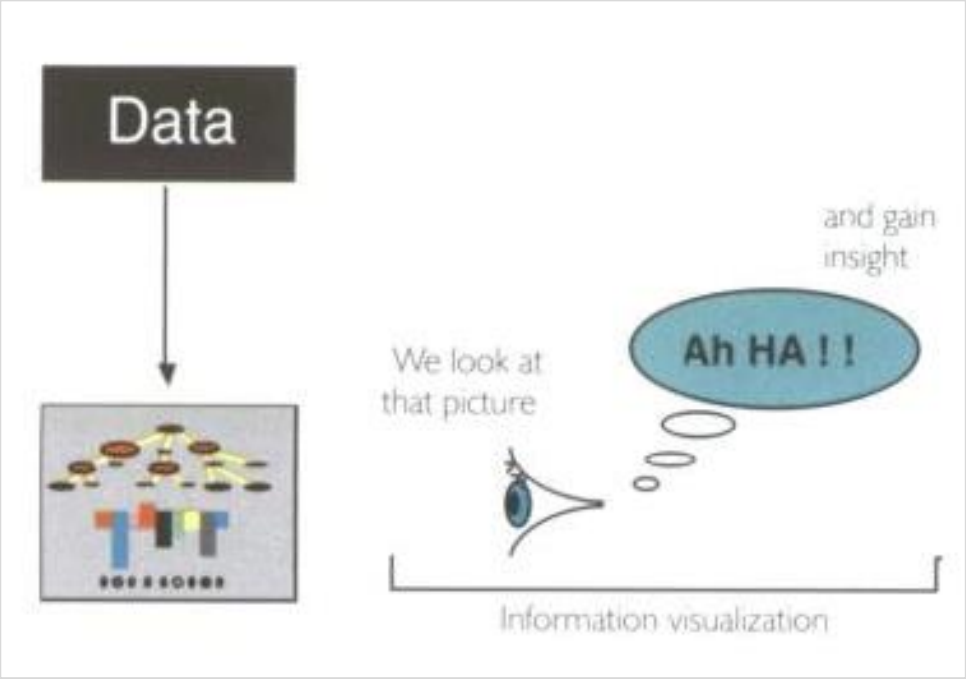
The goal of visual analytics is the creation of tools and techniques to enable people to:

- Synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data.
- Detect the expected and discover the unexpected.
- Provide timely, defensible, and understandable assessments.
- Communicate assessment effectively for action." (Keim et al., 2009, p157)

Information Visualisation: a definition

'The use of computer-supported, interactive, visual representations of abstract data to amplify cognition.'

(Card, Mackinlay, & Shneiderman, 1999, p7)

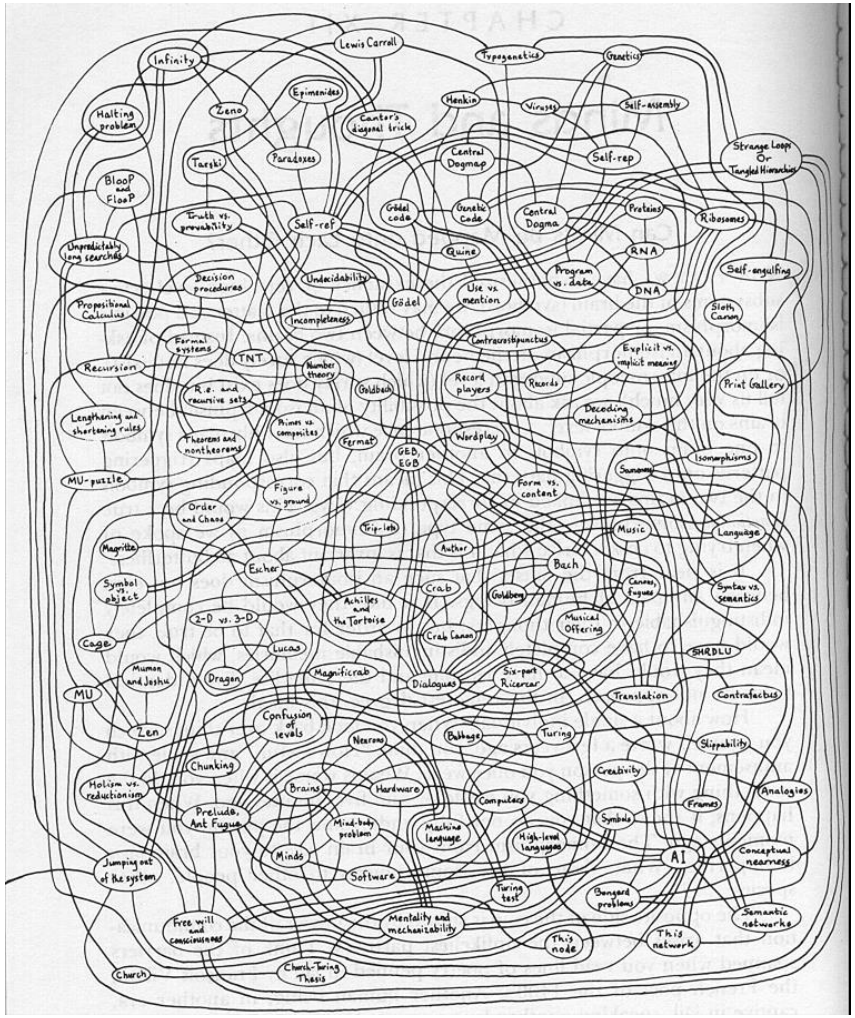


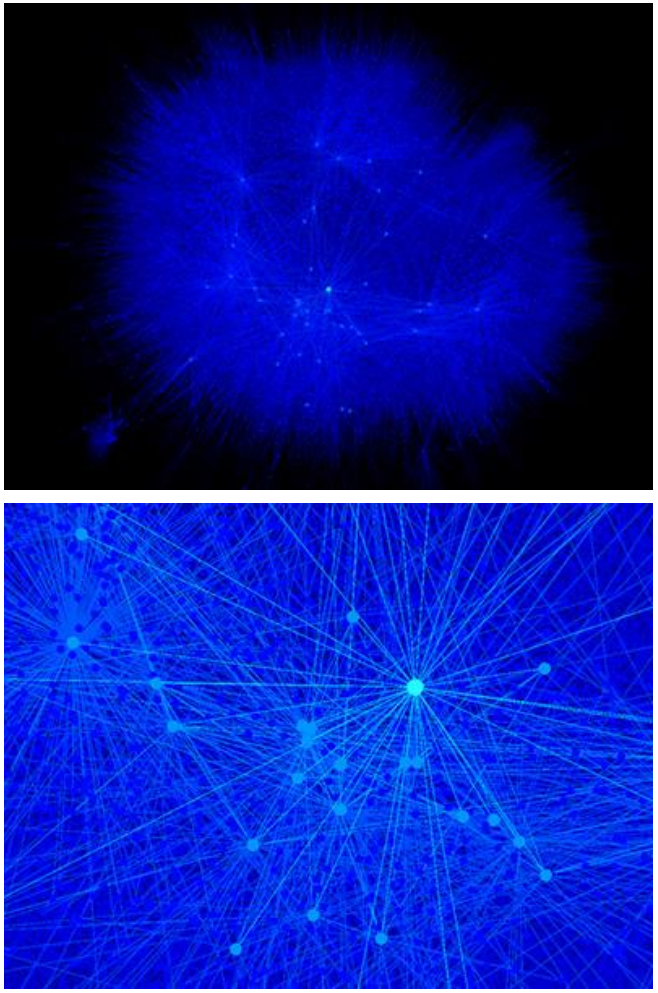
Spence (2007), p5.

'The use of ***computer-supported****,*

interactive, visual representations of abstract data to

Edinburgh Napier UNIVERSITY





Hand drawn semantic network

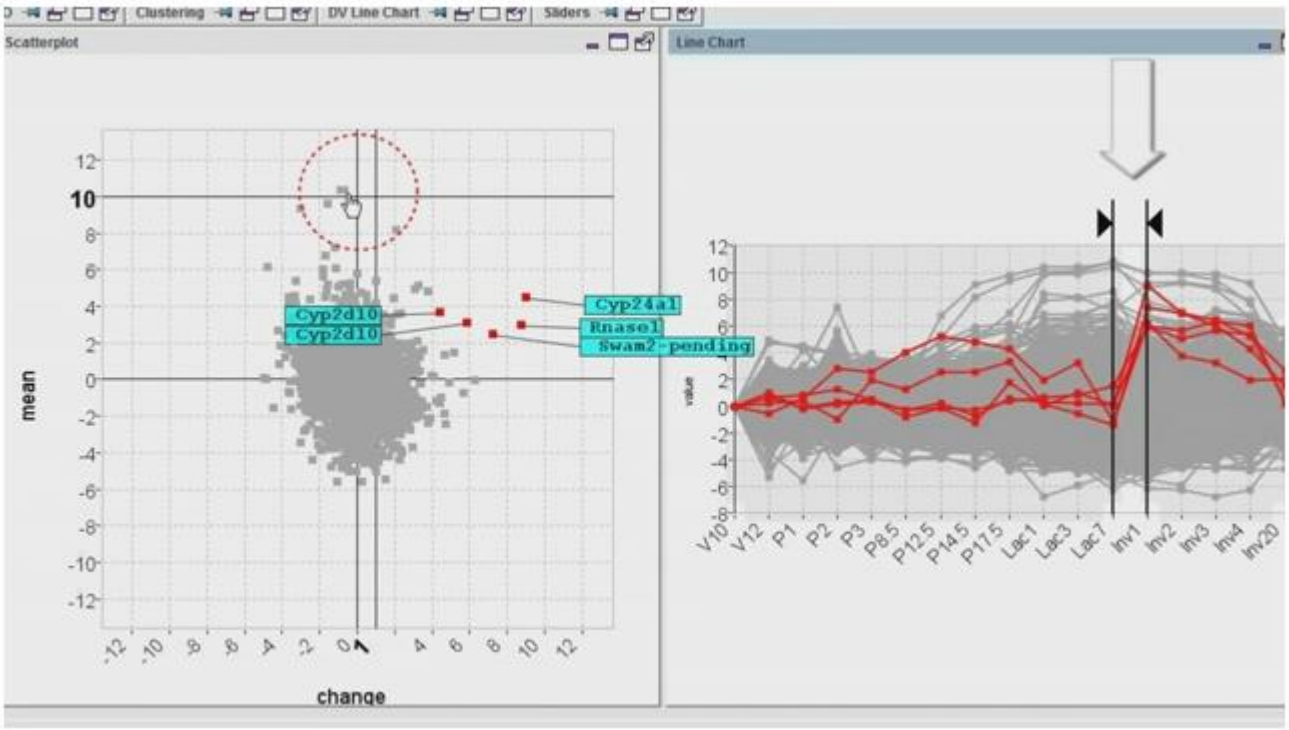
Original image from Hofstadter. Godel, Escher, Bach: an Eternal Gold Braid. Basic Books 1979

Computer drawn friendship network

of 47,471 people connected by 432,430 friendship relations (Jeffrey Heer).
Images from http://www.visualcomplexity.com/vc/project.cfm?id=97

Edinburgh Napier UNIVERSITY

'The use of computer-supported, ***interactive****, visual representations of abstract data to amplify cognition.'*



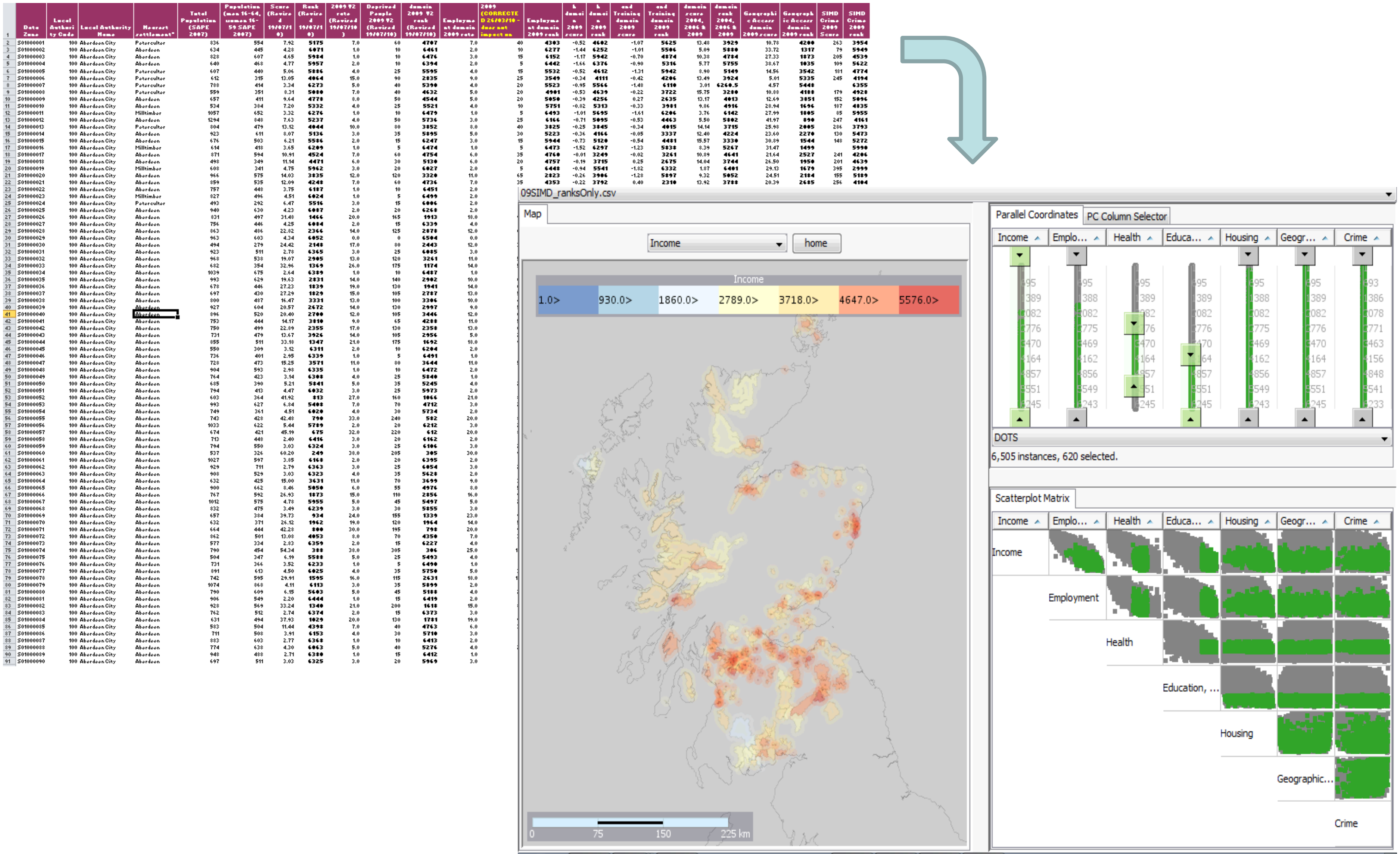
Introduction to MaTSE (Microarray Time-series Explorer) video:
http://www.iidi.napier.ac.uk/c/grants/grantid/11836532
(see also http://www.matse.org.uk for details of the project)

‘The use of computer-supported, interactive,

visual representations *of*

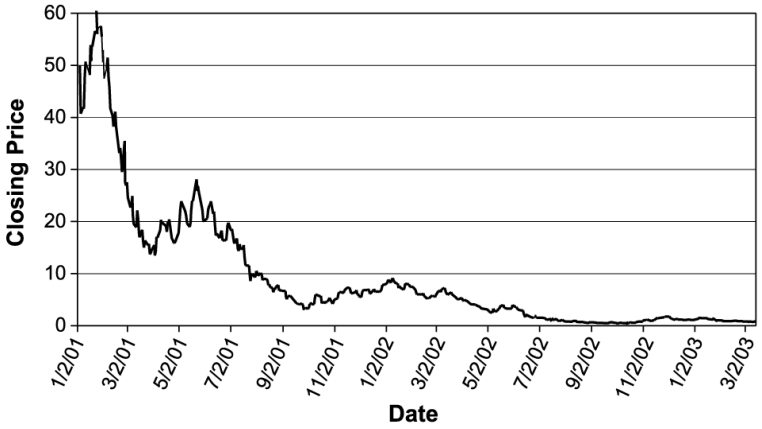
abstract data to amplify cognition.’

Edinburgh Napier
UNIVERSITY

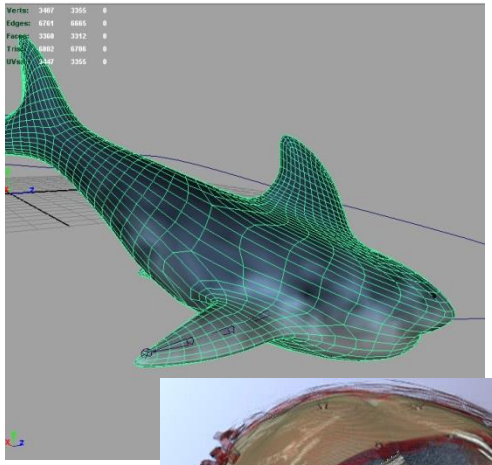


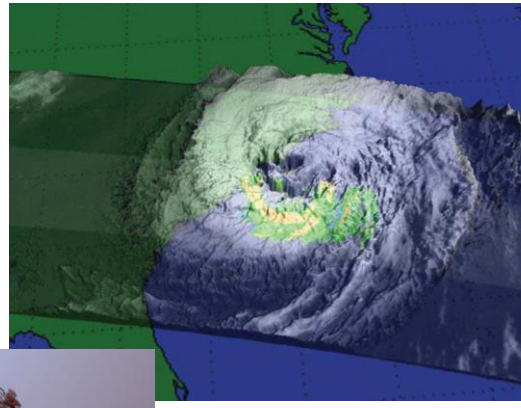
'The use of computer-supported, interactive, visual representations of ***abstract data*** *to amplify cognition.'*

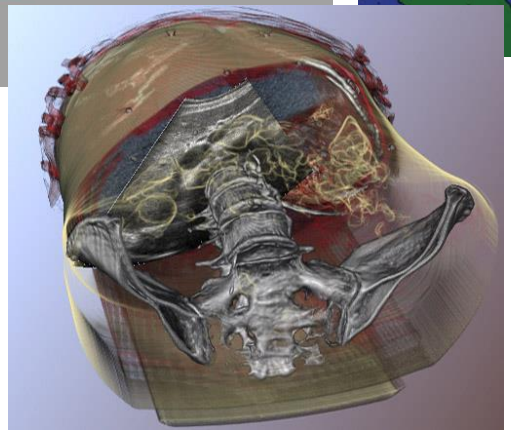
Edinburgh Napier UNIVERSITY



Note: I2 stock price from Jan 2001 to March 2003







Abstract data

- no inherent mapping in space
- e.g. stock prices, names and addresses, survey results etc.

Physical data

- Has a natural spatial mapping
- e.g. geographic data, medical visualisations of the human body etc.
- Dealt with by a different field: *Scientifc Visualisation*

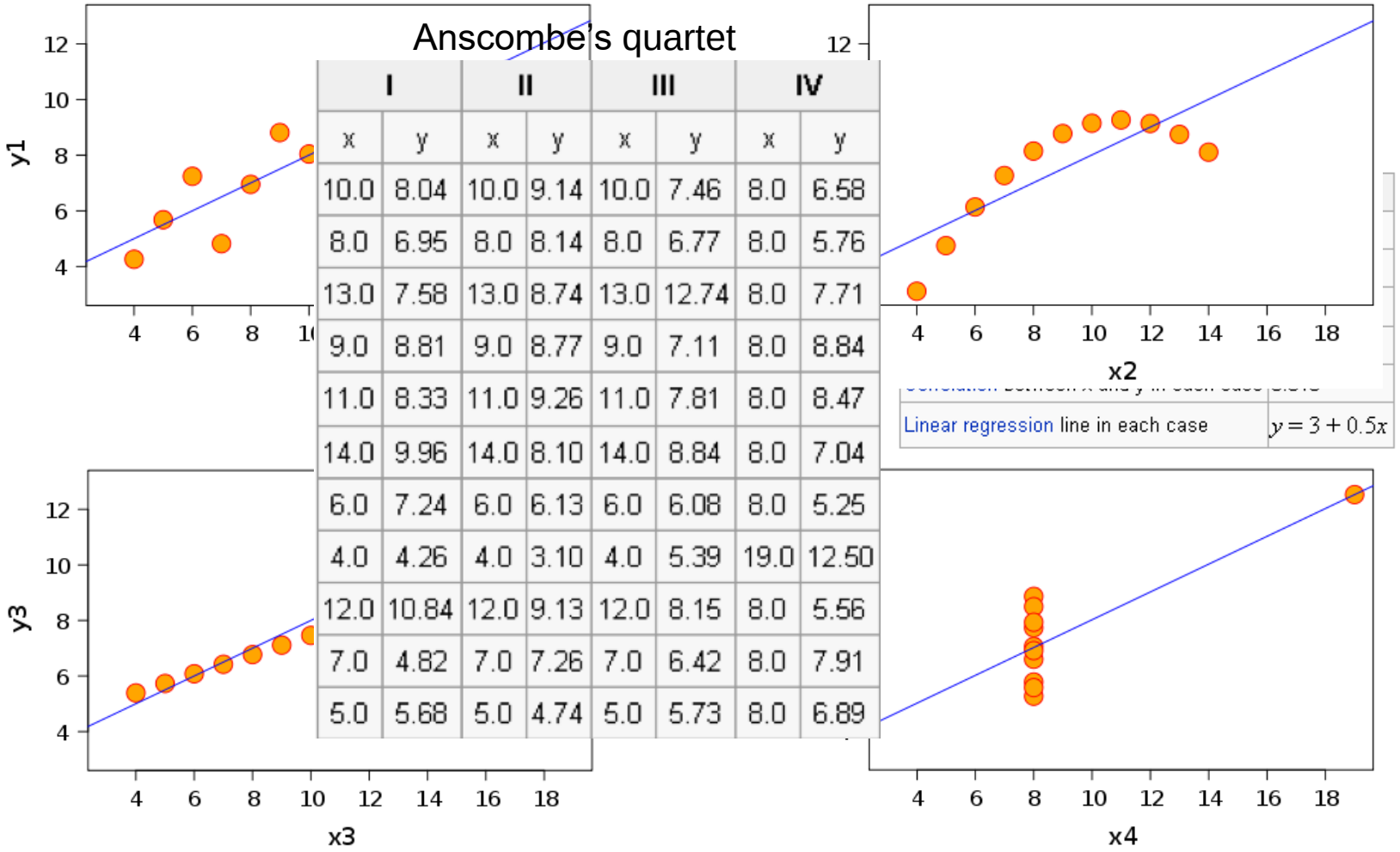
'The use of computer-supported, interactive, visual representations of abstract data to

amplify cognition.'

Edinburgh Napier UNIVERSITY

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



Edinburgh Napier UNIVERSITY

VISUALISATION FOR ANALYSIS: USES

Edinburgh Napier UNIVERSITY

Visualisation for analysis: uses

- To explore and gain understanding of a dataset (Fekete et al., 2008):
  - Where the data set is unfamiliar
  - When you don't have a specific goal in looking at the data
  - When information is easier to recognise than describe ("I'll know it when I see it")
  - As a starting point
    - When you want to examine the data to learn more about it (browsing to make discoveries/gain insight)
    - To narrow down the interesting parts of the data for closer inspection (when browsing a large data set)
    - When the exploratory process itself may influence future (more meaningful) questions and tasks
- Finding patterns and oddities that stay hidden from statistical analysis

Finding patterns

Difficult in a spreadsheet

	A	B	C	D	E	F	G	H	I
1	Model	MPG	Cylinders	Displacem	Horsepow	Weight	Accelerati	Year	Origin
2	hi 1200d	9	8	304	193	4732	18.5	70	US
3	ford f250	10	8	360	215	4615	14	70	US
4	chevy c20	10	8	307	200	4376	15	70	US
5	dodge d200	11	8	318	210	4382	13.5	70	US
6	mercury marquis	11	8	429	208	4633	11	72	US
7	chevrolet impala	11	8	400	150	4997	14	73	US
8	oldsmobile omega	11	8	350	180	3664	11	73	US
9	dodge monaco (sw)	12	8	383	180	4955	11.5	71	US
10	oldsmobile delta 88 ro	12	8	350	160	4456	13.5	72	US
11	mercury marquis broug	12	8	429	198	4952	11.5	73	US
12	buick electra 225 custo	12	8	455	225	4951	11	73	US
13	ford country	12	8	400	167	4906	12.5	73	US
14	oldsmobile vista cruise	12	8	350	180	4499	12.5	73	US
15	ford country squire (sw	13	8	400	170	4746	12	71	US
16	pontiac safari (sw)	13	8	400	175	5140	12	71	US
17	chevrolet impala	13	8	350	165	4274	12	72	US
18	buick lesabre custom	13	8	350	155	4502	13.5	72	US
19	chrysler newport royal	13	8	400	190	4422	12.5	72	US
20	chevrolet chevelle con	13	8	307	130	4098	14	72	US
21	ford gran torino (sw)	13	8	302	140	4294	16	72	US
22	buick century 350	13	8	350	175	4100	13	73	US
23	chevrolet malibu	13	8	350	145	3988	13	73	US
24	chevrolet caprice classi	13	8	400	150	4464	12	73	US
25	ford ltd	13	8	351	158	4363	13	73	US
26	chrysler new yorker br	13	8	440	215	4735	11	73	US
27	amc ambassador broug	13	8	360	175	3821	11	73	US
28	plymouth custom subu	13	8	360	170	4654	13	73	US
29	buick century luxus (sw	13	8	350	150	4699	14.5	74	US
30	ford mustang ii	13	8	302	129	3169	12	75	US
31	plymouth volare premi	13	8	318	150	3940	13.2	76	US
32	chevy c10	13	8	350	145	4055	12	76	US
33	ford f108	13	8	302	130	3870	15	76	US
34	dodge d100	13	8	318	150	3755	14	76	US
35	chevrolet impala	14	8	454	220	4354	9	70	US
36	plymouth fury iii	14	8	440	215	4312	8.5	70	US
37	pontiac catalina	14	8	455	225	4425	10	70	US
38	plymouth 'cuda 340	14	8	340	160	3609	8	70	US

cars

Finding patterns

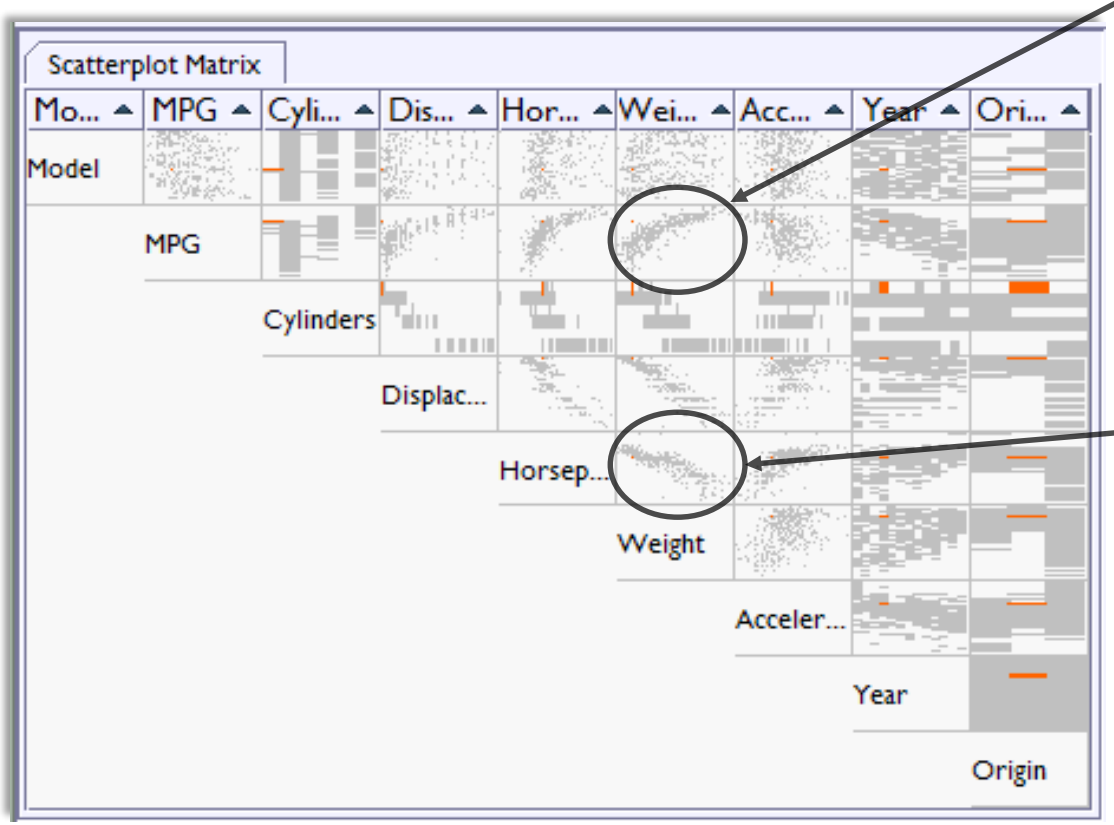
Stats are powerful, but (as we saw in Anscombe's quartet) don't always give the whole picture

Correlations

		price	mpg	weight
price	Pearson Correlation	1	-.478	.539
	Sig. (2-tailed)		.000	.000
	N	74	71	74
mpg	Pearson Correlation	-.478	1	-.807
	Sig. (2-tailed)	.000		.000
	N	71	71	71
weight	Pearson Correlation	.539	-.807	1
	Sig. (2-tailed)	.000	.000	
	N	74	71	74

Finding patterns

- One thing that humans are extremely good at is pattern recognition
- Representing our data visually allows us to exploit our perceptual capabilities.



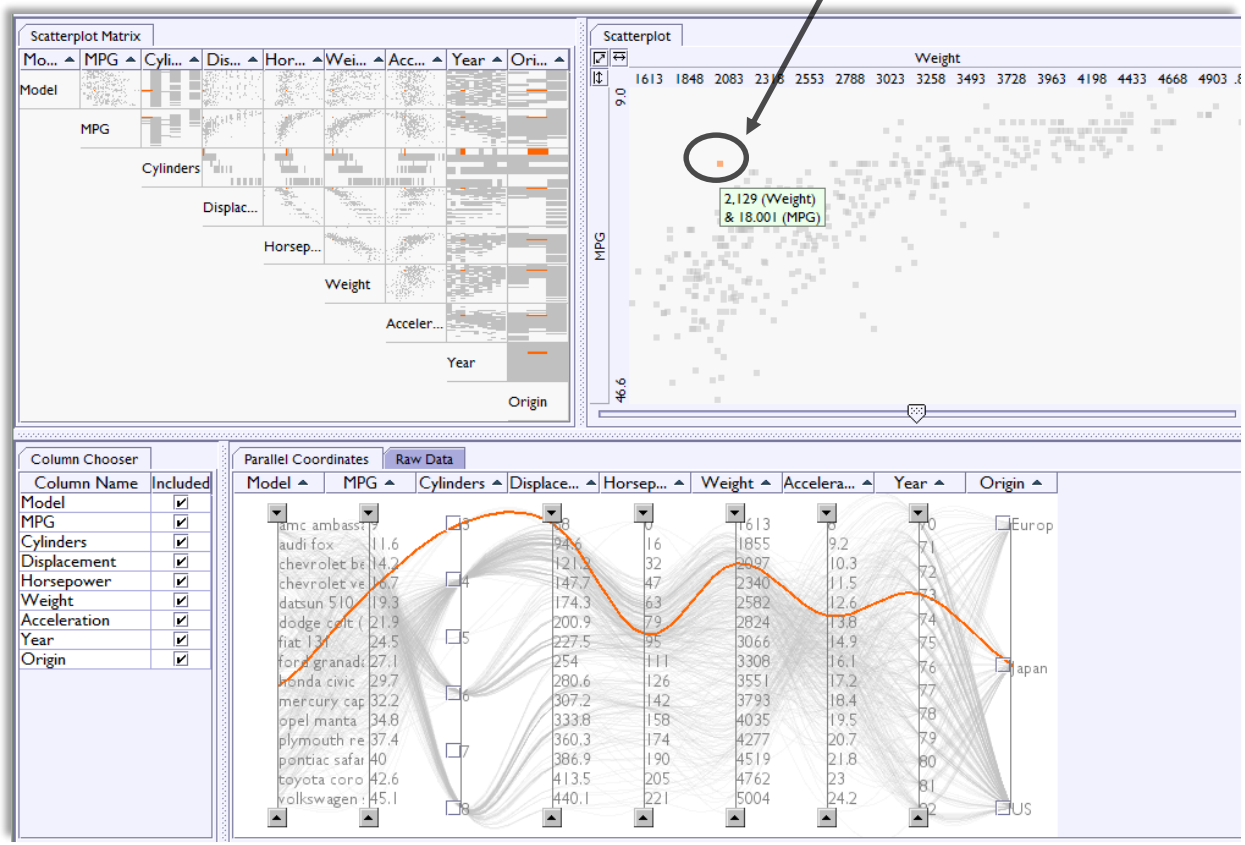
MPG and Weight negatively correlated

Horsepower and Weight positively correlated

Edinburgh Napier UNIVERSITY

Finding outliers / errors

Low weight but rubbish fuel economy



Edinburgh Napier UNIVERSITY

When not to use a visual approach…

Visualisation is probably *not* the best approach if you have a specific question about your data or you want to find individual data values.

- In this case, define the query and execute it against a database/search engine (Fekete et al., 2008)

Edinburgh Napier UNIVERSITY

Recap: Motivations for using Visualisation

- In presentation/communication of data
  - Accessibility: opens up data to a wider audience (not just statisticians)
  - Aesthetics: visualisation more attractive than statistics; engaging
  - Convincing: helps persuade yourself + others of the value of the data
  - Trust: visualisation more trusted by some than statistics

- In data analysis
  - As a starting point for exploration: understanding what's in the data, generating further questions to ask
  - Finding things that you can't with statistics alone (e.g. patterns and oddities)

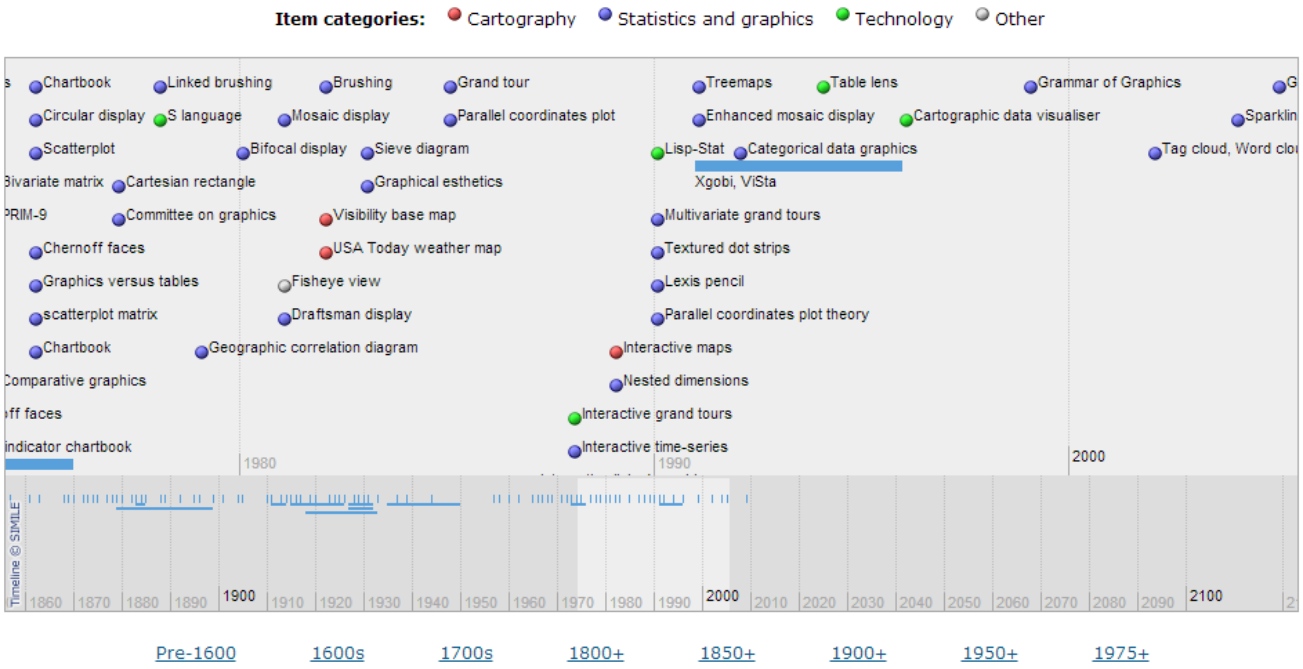
- In both cases:
  - Reduces cognitive effort by offloading some of the hard work onto the perceptual system

Edinburgh Napier
UNIVERSITY

VISUALISATION: BEGINNINGS

A history of visualisation, visualised

- For those interested in the development of the field, see Friendly and Denis' interactive timeline of milestones in the history of visualisation at http://datavis.ca/milestones/



Externalising cognition

- Humans have used mark making as a way of externalising cognition since ancient times
- Calculations are (for most of us) much easier with pen and paper than without
- Visualisation used as a “temporary working area” – a memory aid

A table (3000BC)

Clay tablet with a record of Beer quantities

Late Uruk Period, 3000BC.

.

Image: Francis T. Marchese

John Snow, cholera outbreak plot (1850s)

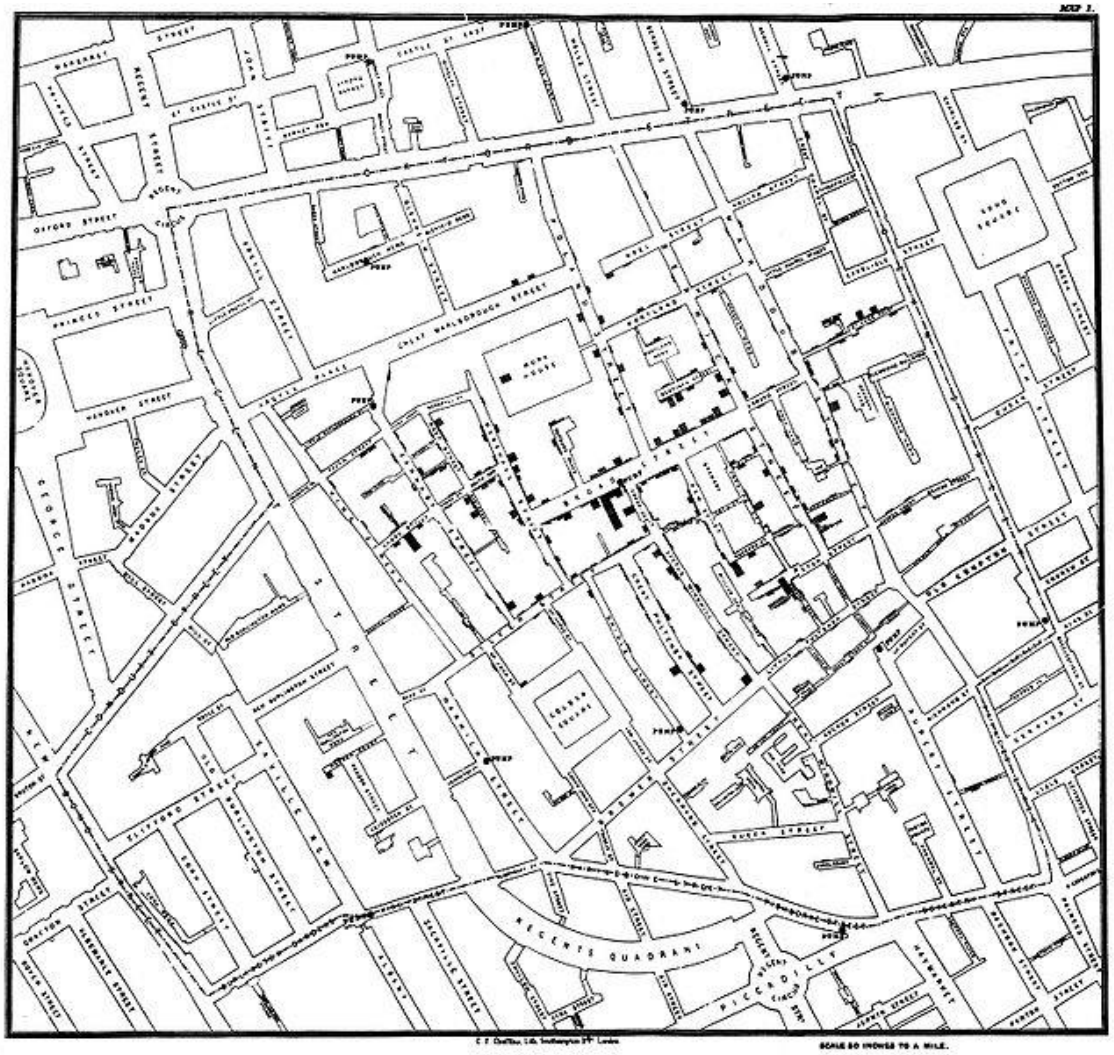
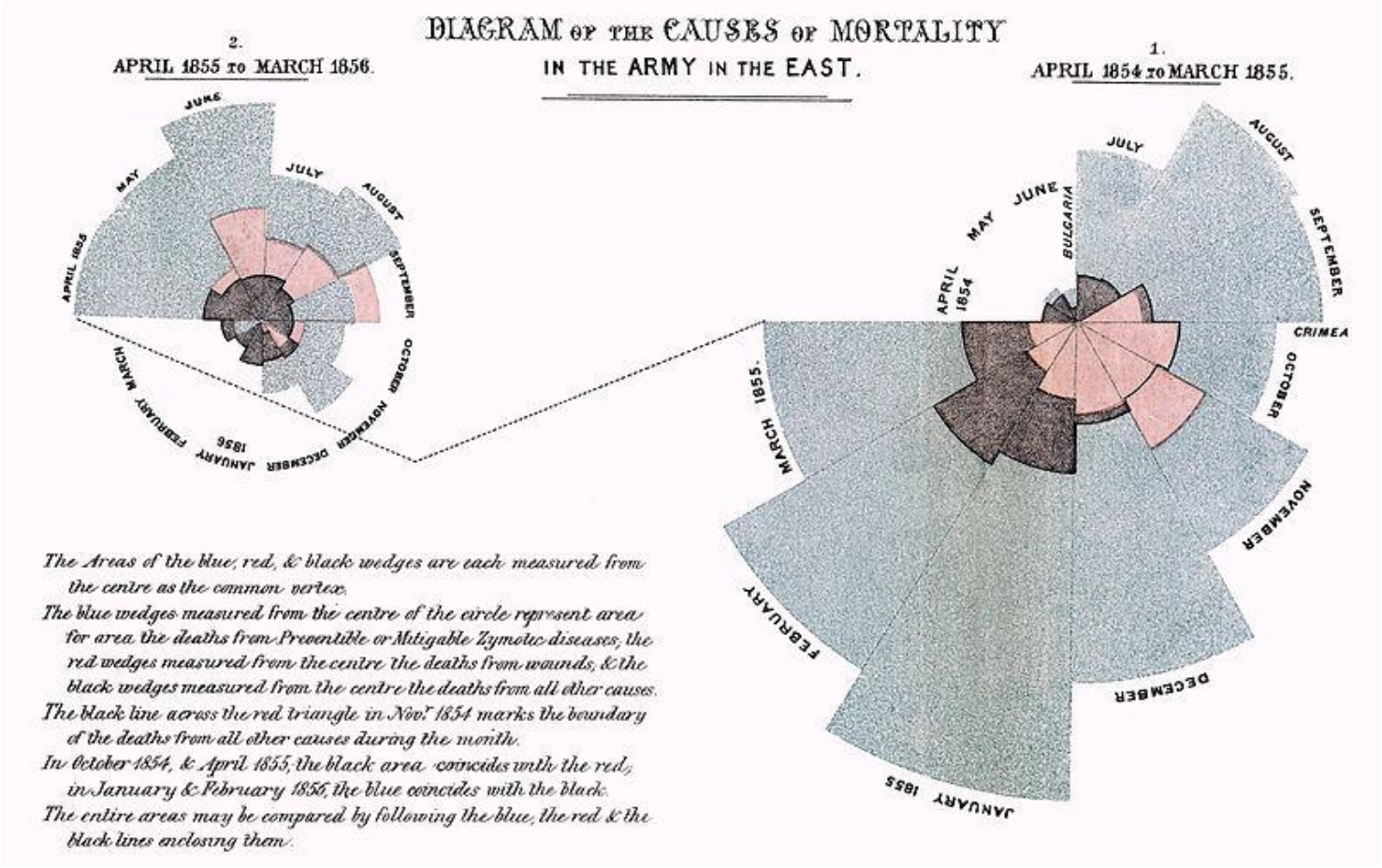


Image: http://en.wikipedia.org/wiki/File:Snow-cholera-map-1.jpg

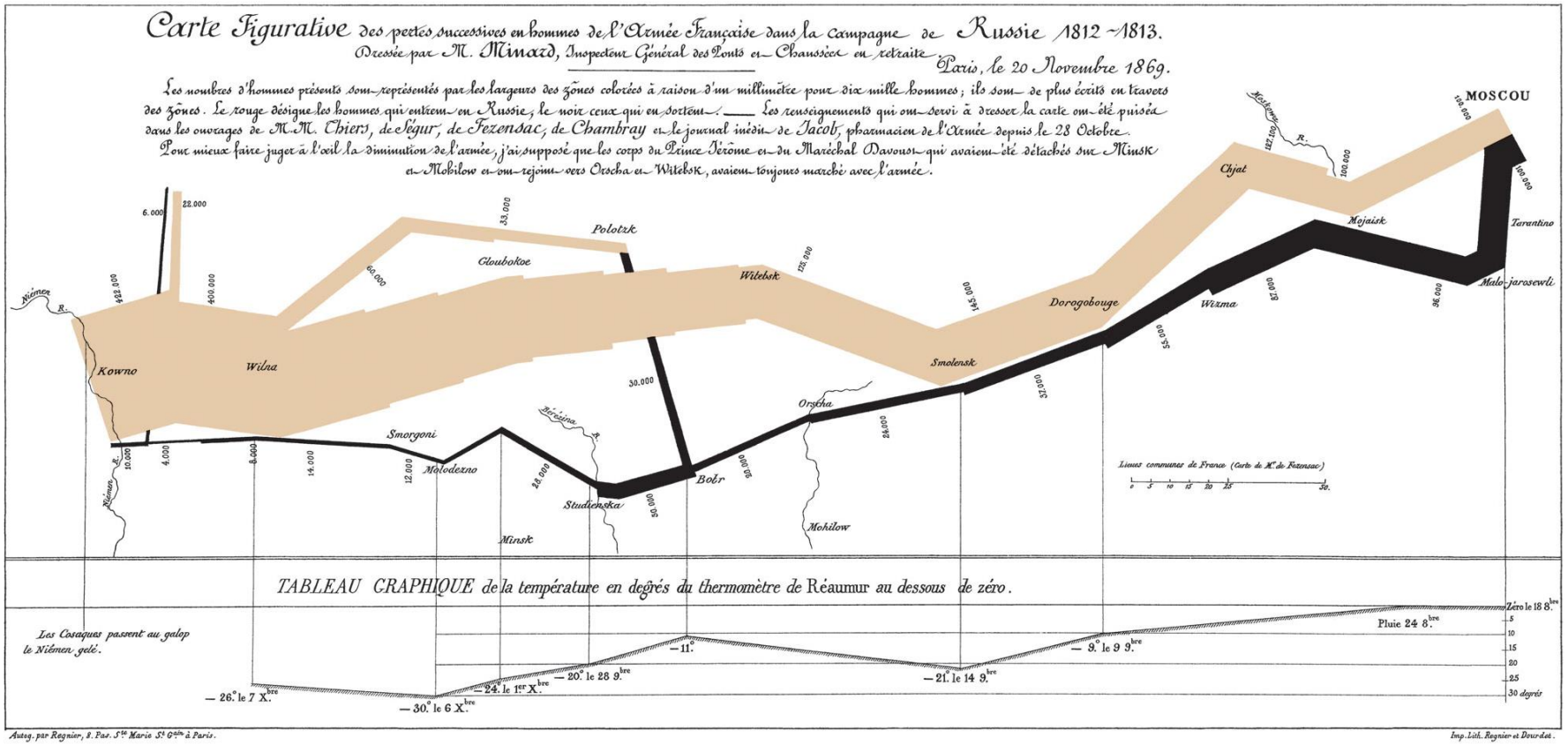
Edinburgh Napier UNIVERSITY

Florence Nightingale's rose diagram (1858)



Source: http://commons.wikimedia.org/wiki/File:Nightingale-mortality.jpg

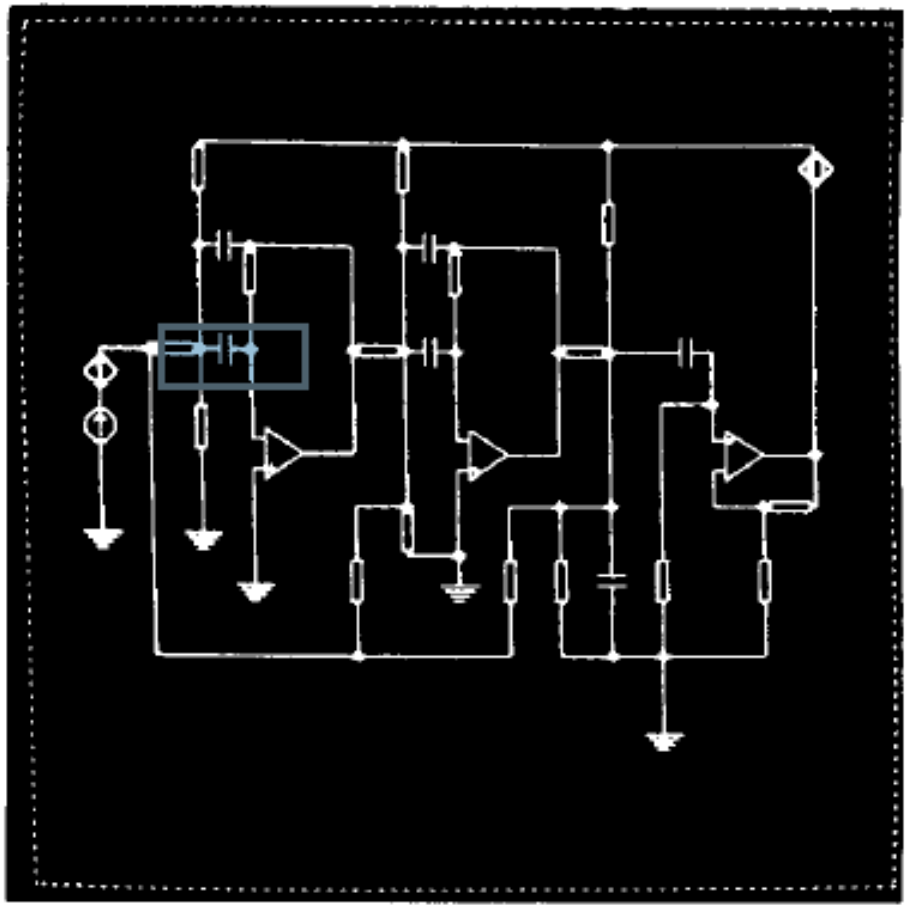
Minard's map (1869)



http://upload.wikimedia.org/wikipedia/commons/2/29/Minard.png

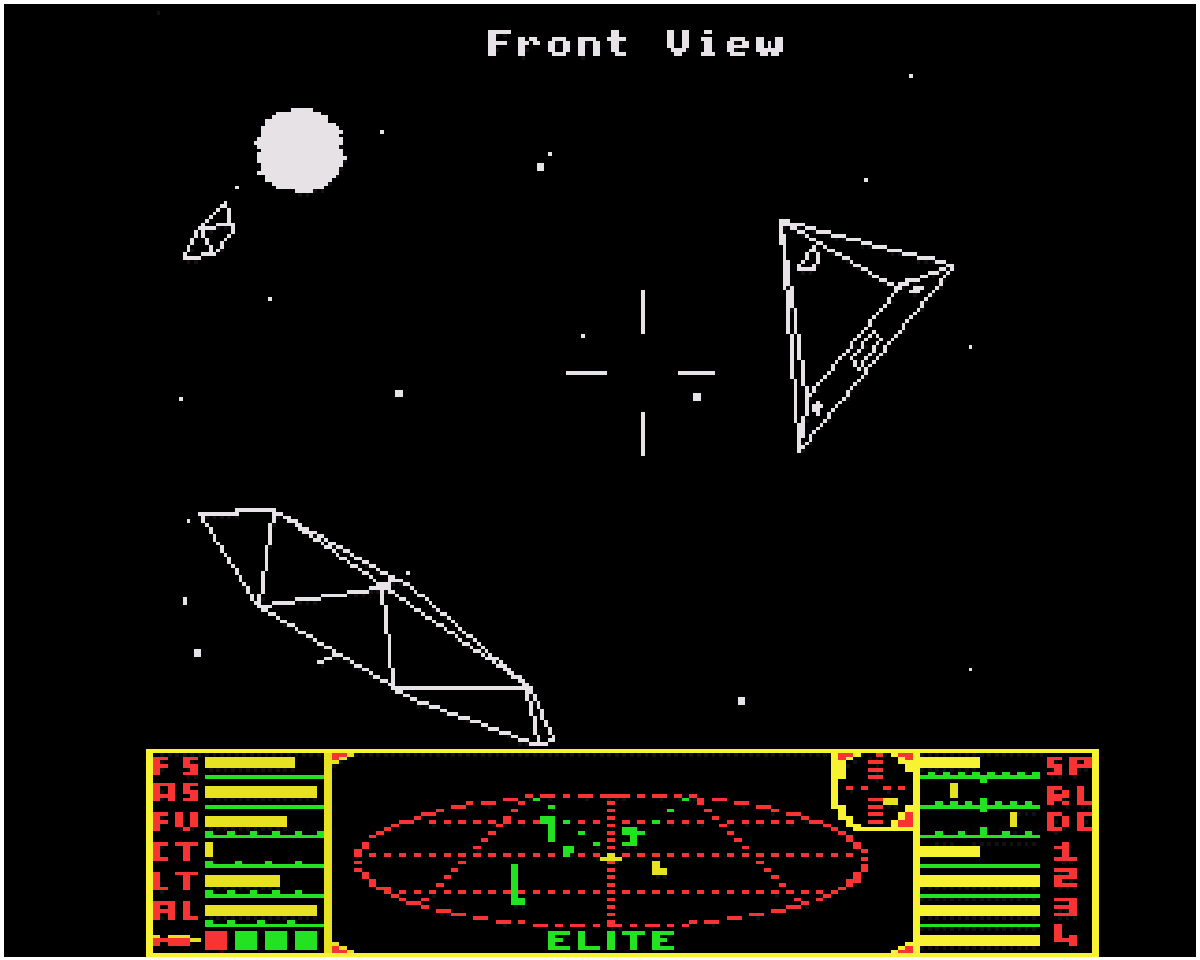
Edinburgh Napier UNIVERSITY

Interaction (late 60s)



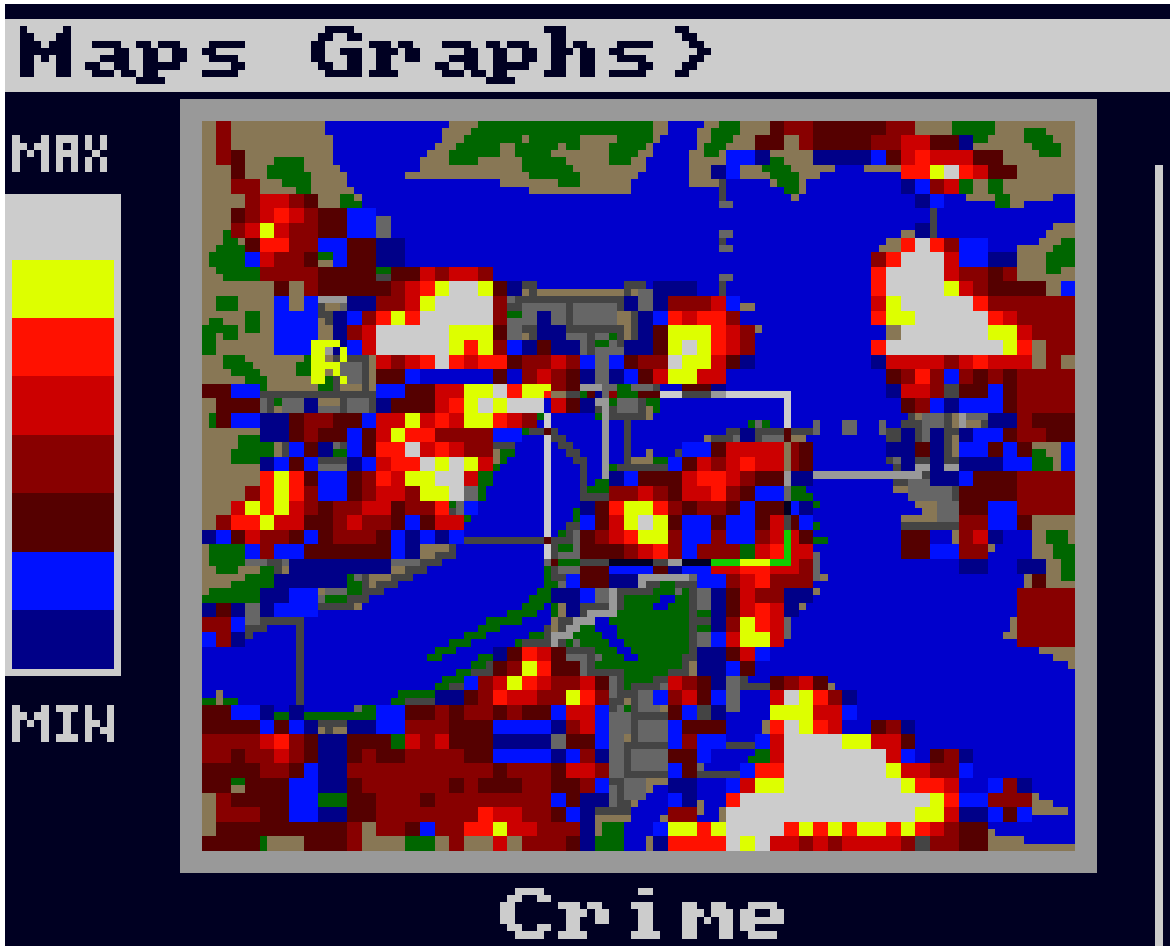
Bob Spence, circuit diagram, 1977

Visualisation in computer games (1980s)



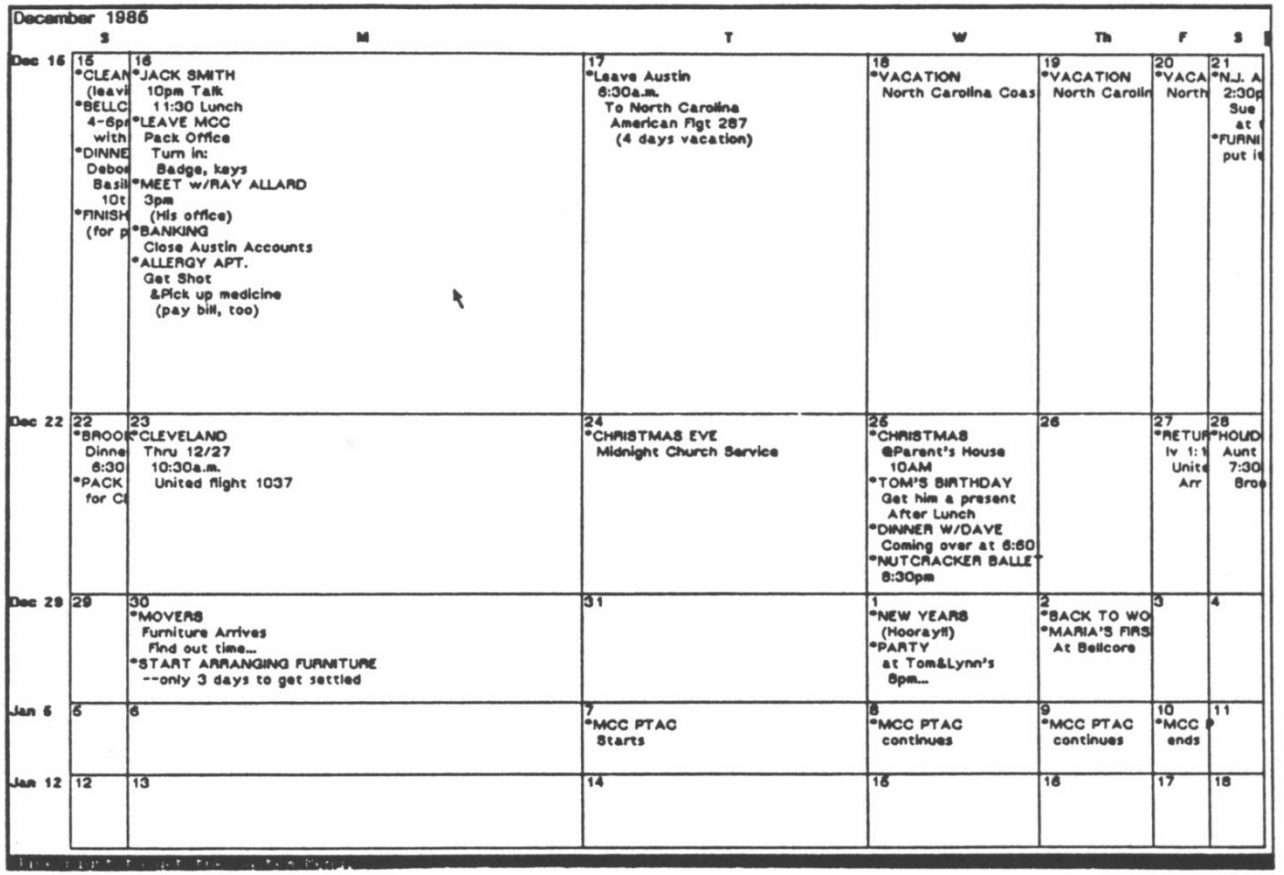
Radar plot navigation. David Braben, Elite, 1984

Bar chart of character attributes. Renegade Games, 1988



Heat map geo-visualisation plot. Sim City, Maxis, 1989

Focus + context interaction (1980s)



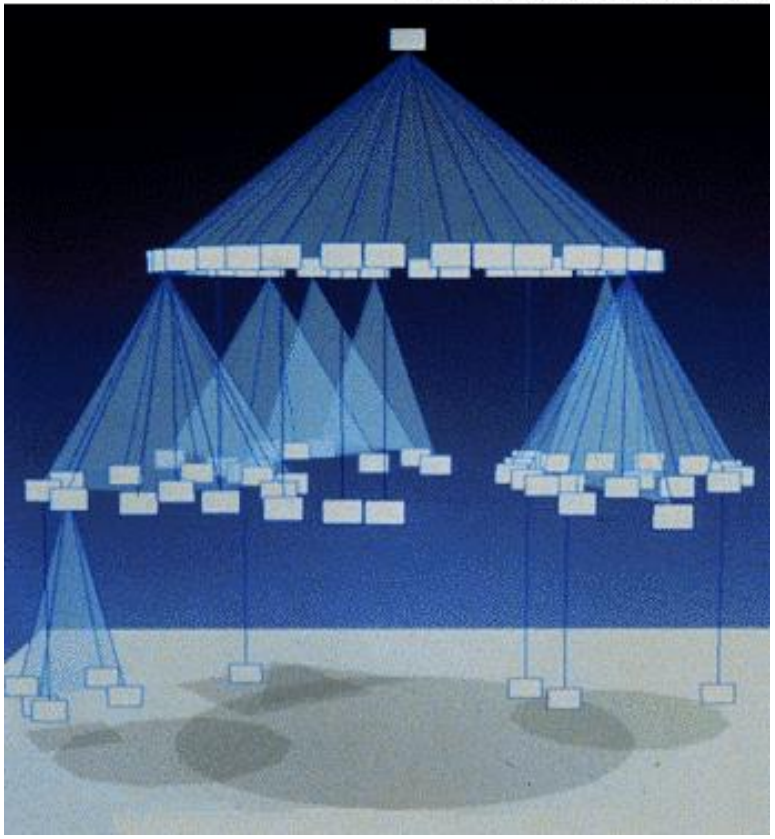
December 1986

	S	M	T	W	Th	F	S
Dec 15	15 *CLEAN (leavi *BELLC 4-6p with *DINNE Debo Basil 10t *FINISH (for p	16 *JACK SMITH 10pm Talk 11:30 Lunch *LEAVE MCC Pack Office Turn in: Badge, keys *MEET w/RAY ALLARD 3pm (His office) *BANKING Close Austin Accounts *ALLERGY APT. Get Shot &Pick up medicine (pay bill, too)	17 *Leave Austin 6:30a.m. To North Carolina American Flgt 287 (4 days vacation)	18 *VACATION North Carolina Coas	19 *VACATION North Carolin	20 *VACA North	21 *N.J. A 2:30p Sue at (*FURNI put it
Dec 22	22 *BROO Dinne 6:30 *PACK for C	23 *CLEVELAND Thru 12/27 10:30a.m. United flight 1037	24 *CHRISTMAS EVE Midnight Church Service	25 *CHRISTMAS @Parent's House 10AM *TOM'S BIRTHDAY Get him a present After Lunch *DINNER W/DAVE Coming over at 6:60 *NUTCRACKER BALLET 8:30pm	26	27 *RETUR lv 1: Unite Arr	28 *HOUD Aunt 7:30 Broo
Dec 29	29	30 *MOVERS Furniture Arrives Find out time... *START ARRANGING FURNITURE --only 3 days to get settled	31	1 *NEW YEARS (Hooray!!) *PARTY at Tom&Lynn's 8pm...	2 *BACK TO WO *MARIA'S FIRS At Bellcore	3	4
Jan 6	5	6	7 *MCC PTAC Starts	8 *MCC PTAC continues	9 *MCC PTAC continues	10 *MCC ends	11
Jan 12	12	13	14	15	16	17	18

George Furnas, Fisheye Calendar, 1986

Audiomaster, AEGIS, 1990

3D in InfoVis (1990s)



Cone Trees, File Directories, Robertson, Card & Mackinlay, 1990

Edinburgh Napier UNIVERSITY

References

- Card, Stuart K., Jock D. Mackinlay, and Ben Schneiderman, eds. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.
- Dix, A. (2013). Introduction to information visualisation. In *Information Retrieval Meets Information Visualization* (pp. 1-27). Springer Berlin Heidelberg. http://link.springer.com/book/10.1007/978-3-642-36415-0/page/1
- Jean-Daniel Fekete, Jarke J. Wijk, John T. Stasko, and Chris North. 2008. The Value of Information Visualization. In *Information Visualization*, Andreas Kerren, John T. Stasko, Jean-Daniel Fekete, and Chris North (Eds.). Lecture Notes In Computer Science, Vol. 4950. Springer-Verlag, Berlin, Heidelberg 1-18. DOI=10.1007/978-3-540-70956-5_1 http://dx.doi.org/10.1007/978-3-540-70956-5_1
- Keim, D. A., Mansmann, F., & Thomas, J. (2009). Visual analytics: how much visualization and how much analytics?. *ACM SIGKDD Explorations Newsletter*,*11*(2), 5-8. http://dl.acm.org/citation.cfm?id=1809403