INM433: Session 01

Introduction to Visual Analytics and Visualisation

INM433 Visual Analytics

Welcome to Visual Analytics!

- Taught by
  - Aidan Slingsby
  - Gennady and Natalia Andrienko
- Not much specialised Visual Analytics software
  - We'll use Andrienkos' **V-Analytics**
  - **Tableau/Mondrian/R/Excel** used in a "VA way"
- 100% coursework
  - Groupwork: literature review
  - Individual work: Analysis project

Visual Analytics

What is Visual Analytics?

- The science and practice of **analytical reasoning** by combining **computational analysis** with **interactive visual interfaces**.
  - Let computers do what computers are good at (summarising and searching large amounts of data)
  - Let humans do what humans are best at (interpreting, thinking, reasoning, applying expert knowledge)
- Analytical reasoning
  - Data→Information→Knowledge→Solution/decision

Human vs Computer

Human

- flexible and inventive, can deal with new situations and problems
- solve problems that are hard to formalise
- can cope with incomplete/ inconsistent information
- can recognise things that are hard to compute or formalise

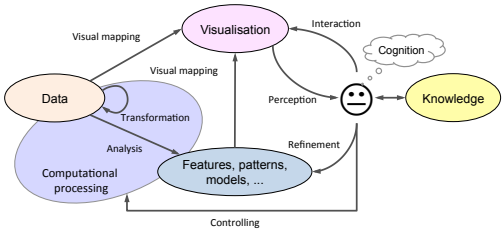
Computer

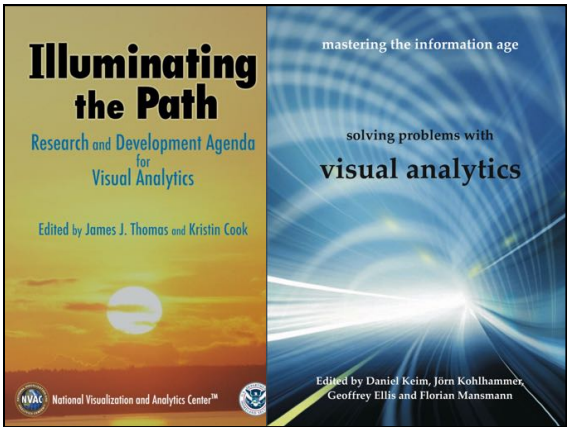
- large amounts of data
- fast search
- fast data processing
- link to other databases/ services
- high quality graphics

Why Visual Analytics?

- Incorporate human reasoning
  - Too much data to simply explore ourselves
    - need a computer to reduce/filter/generalise/identify
    - …but which parameters to use?
  - Want to be aware of potential problems in data
  - Don't know what we're looking for (yet)
  - Want to incorporate "expert knowledge"
  - Want to record and/or understand the process of reasoning

Divide the labour between human and computer





Examples from later in the module

- London Bike Hire scheme
  - Are there distinct bike hiring behaviours? What? Where? Can we manage the scheme better?
- Tweets
  - Can we detect events from tweets? Where? When?
  - Can be characterise neighbourhoods?
- Journeys
  - What are common journeys made

Example

- IBM's entry to Orange's **Data for Development** challenge
  - Participants were given 2.5 billion phone records (cell mast only) for the Ivory Coast
- Suggested bus routes based on mobile phone patterns of communication

http://www.bbc.co.uk/news/technology-22357748

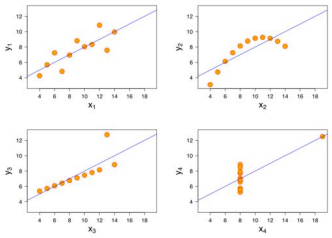
Role of visualisation

Role of visualisation

- **Today is about visualisation**
- Present data in a way that facilitates comparison
  - Compare hundreds of numbers - humans are good as seeing visual patterns
  - See distributions and uncertainties
  - Compare alternative outputs
- As an interactive interface to data:
  - Details on demand
  - Trigger/direct the computational analysis
  - Humans can act on their interpretations

Anscombe's Quartet (1973)

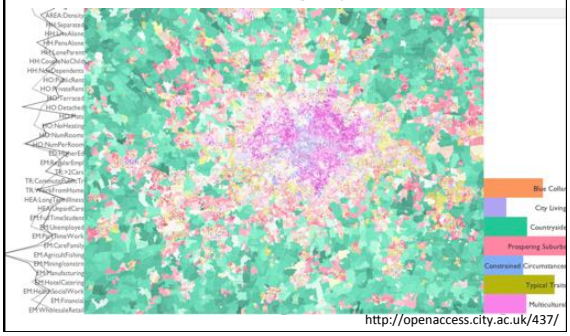
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	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



https://en.wikipedia.org/wiki/Anscombe%27s_quartet

(Static) example: transport

(Interactive) example: Geodemographics



http://openaccess.city.ac.uk/437/

(Interactive) example: Geodemographics: Purpose

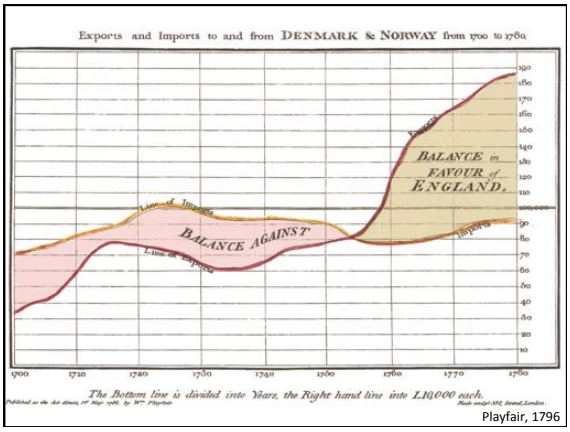
- To explore a population classifier
  - What variables drive the classifier?
  - What's the spatial distribution of the variables?
  - How certain are we that the classifier is a good description?
- To look at the impact on users
  - Does it affect the way that the classifier is used?

http://openaccess.city.ac.uk/437/

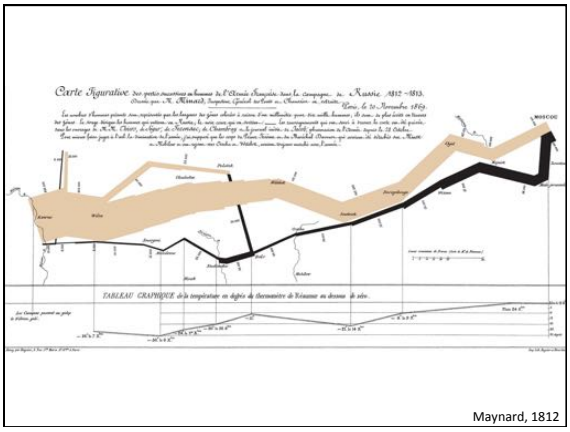
A historical perspective

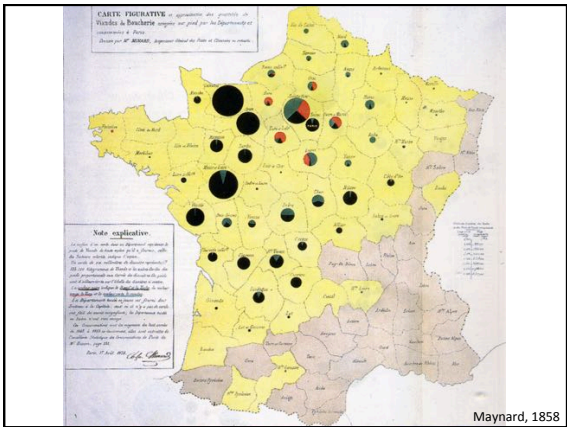
Statistical Graphics

- From the statistics domain
  - Scatter plots
  - histograms,
  - box plots
  - barcharts
  - Realisation that statistics may hide important things
- Exploratory data analysis (Turkey, 1977)
  - Interactive versions from 1970s onwards



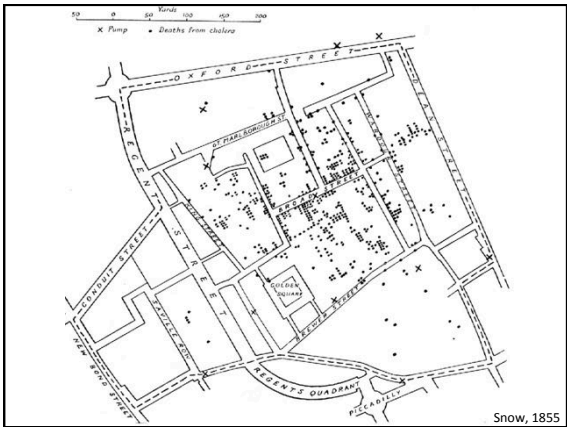
Playfair, 1796


Maynard, 1812


Maynard, 1858

Geographical Information Science (GIS)

- Geographical data storage, visualisation and analysis
  - Facility managements
  - Automated cartography
  - Layers and geographical linking
  - Data representation
  - Spatial analysis
  - (poor temporal handing)
- Recent web-based, mashup developments



Snow, 1855

Information Visualisation

- From computer graphics and human-computer interaction
- Emphasis in novelty in visual encoding and interaction
- Increasingly “spatial” and “temporal”

Visual Analytics

- Incorporation of computational analysis methods
  - particularly for large data
- Emphasis on reasoning and sensemaking in an application domain

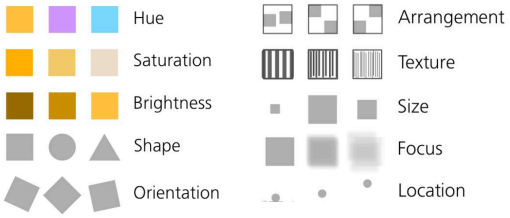
The visualisation display

The visualisation display

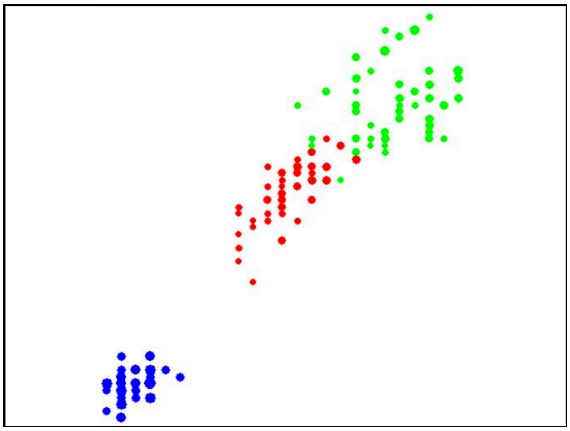
- It's helpful to think of the visualisation display as a **2D space** that contains **marks:**
  - Points, lines, areas, surfaces, volumes
- See:
  - Munzner (2014): *Visualization Analysis and Design*. A K Peters Visualization Series, CRC Press

Visual variables

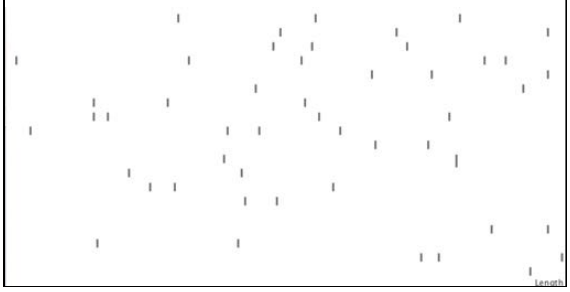
- First proposed by Bertin (1983)
  - also added to by others!



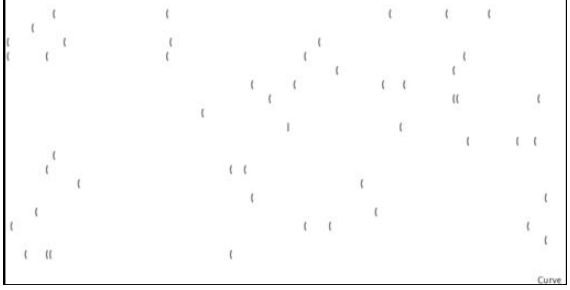
MacEachren, 1995



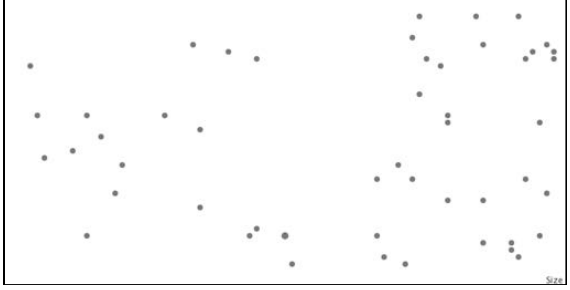
Pre-attentive: length



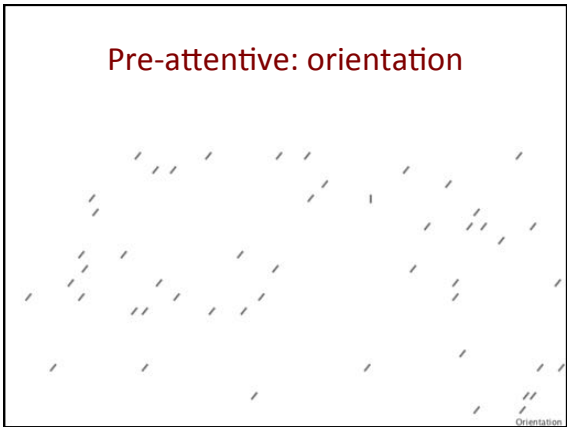
Pre-attentive: curve



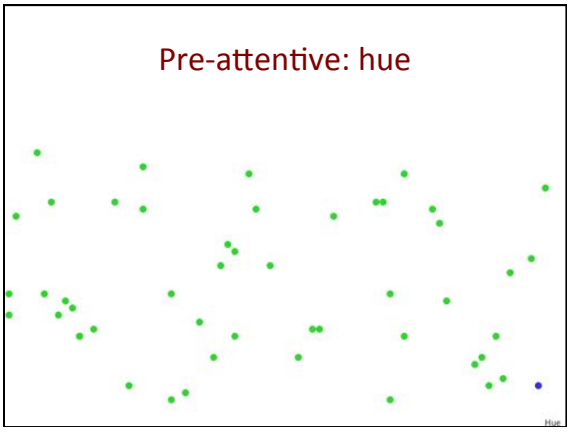
Pre-attentive: size



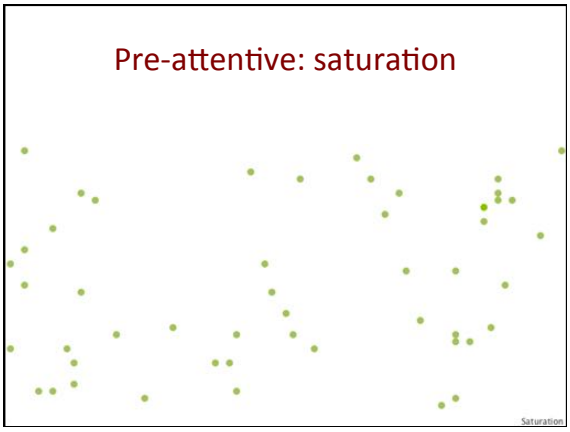
Pre-attentive: orientation



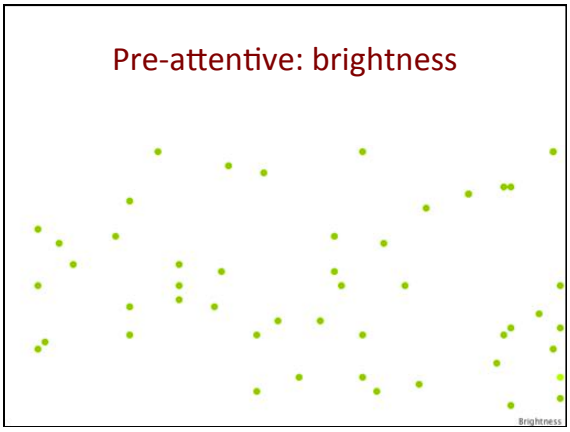
Pre-attentive: hue



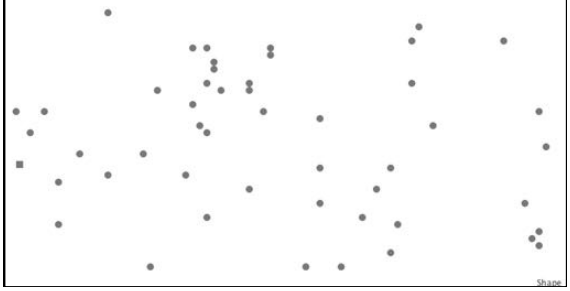
Pre-attentive: saturation



Pre-attentive: brightness



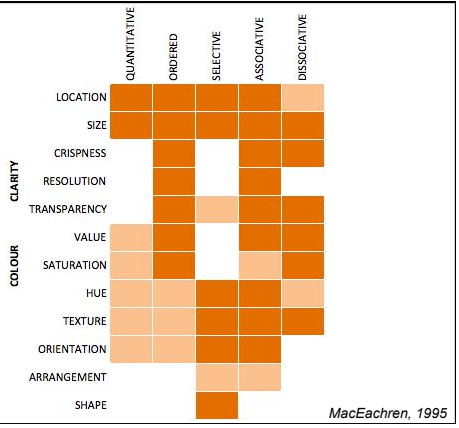
Pre-attentive: shape



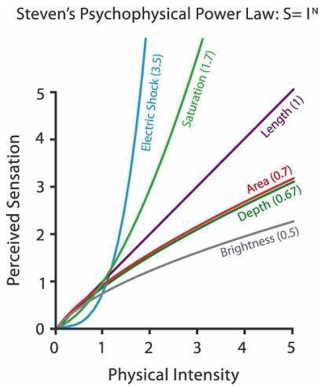
Visual variables not created equal

- Quantitative
  - Allowing **quantity** to be estimated
- Ordered
  - Allowing **order** to be determined
- Selective
  - Allowing **particular things** to be identified
- Associative
  - Allowing **groups of things** to be identified

Bertin, 1983

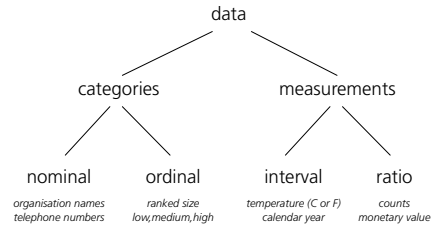


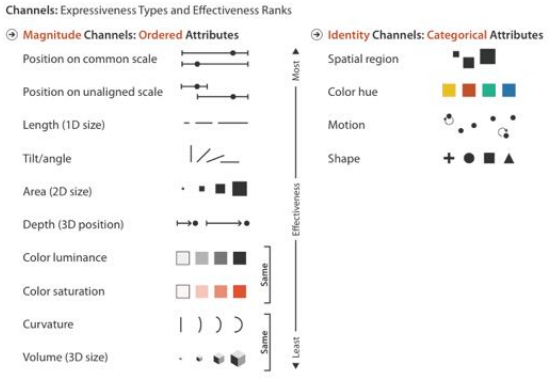
MacEachren, 1995



Stevens, S.S. (1957): On the psychophysical law. *Psychological Review* 64 (3): 153–181

Depends on your data





T. Munzner (2014): *Visualization Analysis and Design*. A K Peters Visualization Series, CRC Press

Use theory to inform design choices

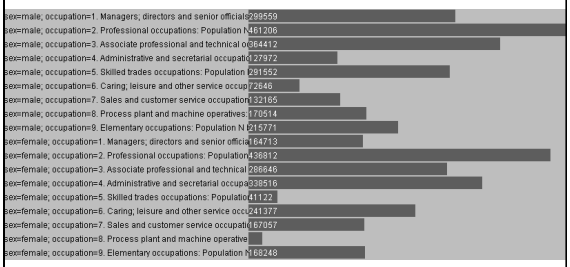
- For example:
  - Position or (aligned) length is the strongest
    - implies quantity/order, so take care with categories
  - Hue and shape good for unordered categories (<8ish)
  - Lightness has a much lower resolution than position
    - can be good for ordered categories.
  - Maps use position to show geography
    - but if geography isn't the main point, are you "wasting" this important visual variable?

Types of display

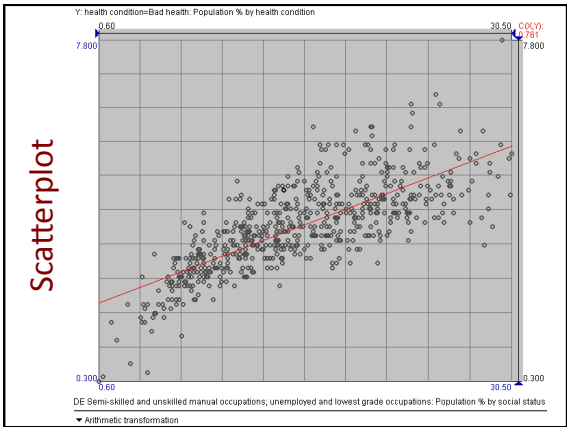
Some common types of display

	Display elements	Data components
Bar graph (bar chart) 	x- or y-position bar size (length)	references numeric attribute
Scatter plot 	marks (dots) x-position y-position	references numeric attribute numeric attribute
Line graph 	x-position y-position	ordered references, especially temporal numeric attribute
Gantt chart 	y-position x-position bar size (length)	temporal objects (events, processes ...) existence time (start, end) duration
Map 	marks (x,y)-position other variables	spatial references (refer to spatial objects) spatial references or spatial attributes attributes

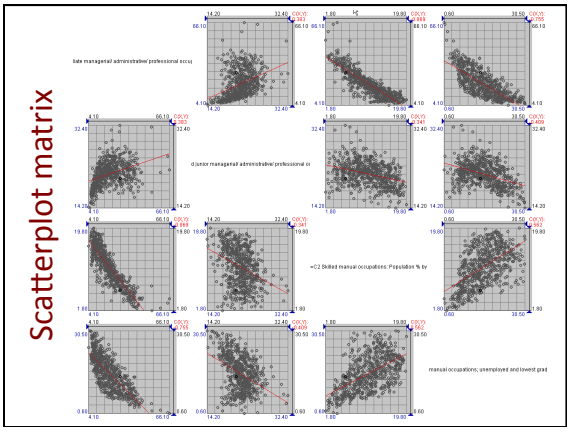
Barchart



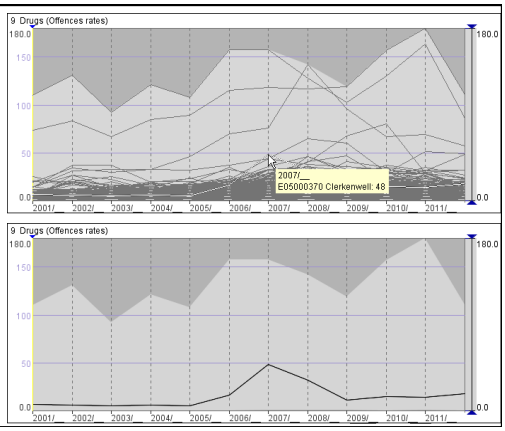
Scatterplot



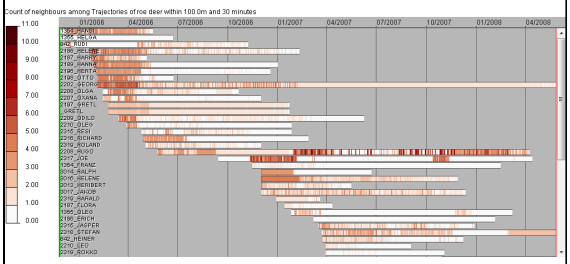
Scatterplot matrix



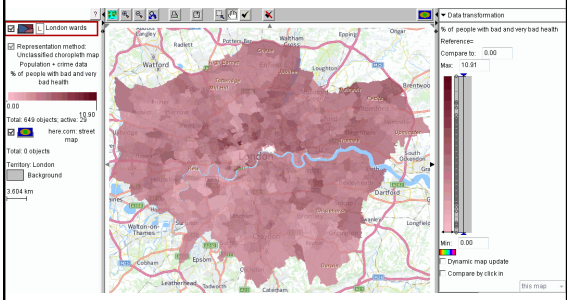
Line graph



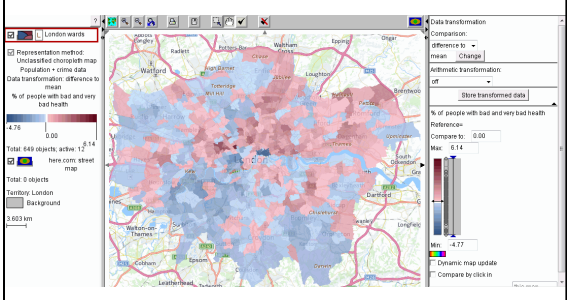
Gantt Chart



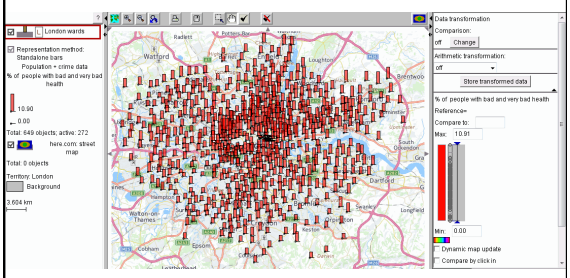
Choropleth map



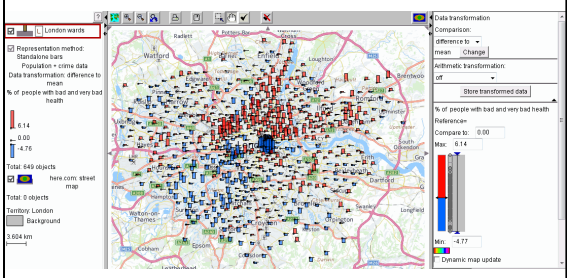
Choropleth map



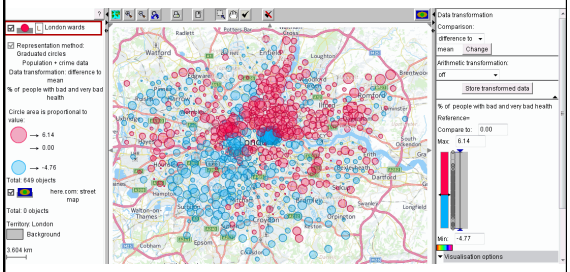
Proportional symbol map



Proportional symbol map

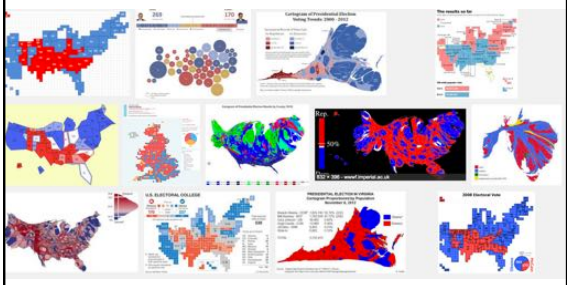


Proportional symbol map



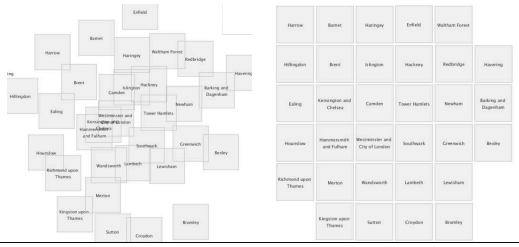
Area cartograms

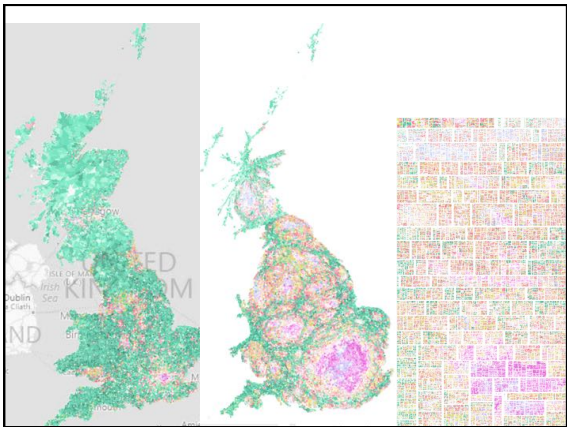
Some results of Google Image search

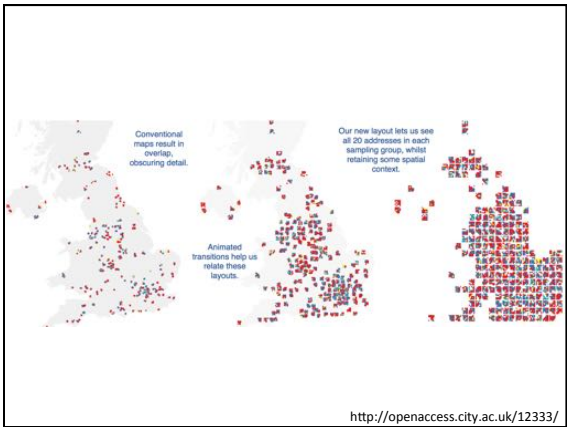


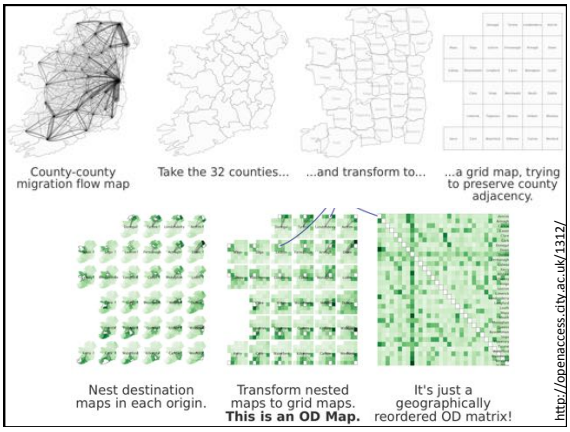
Distorting space

- giCentre often distort geographical space to give space and/or salience to particular places

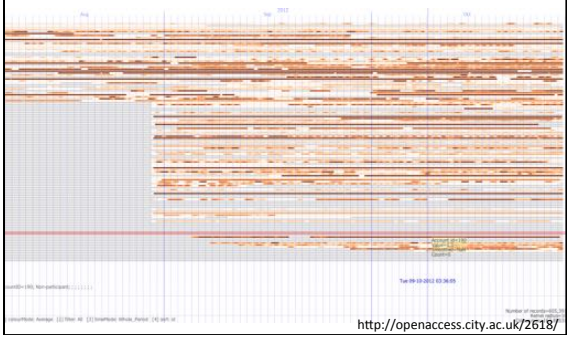








Example: Smart homes

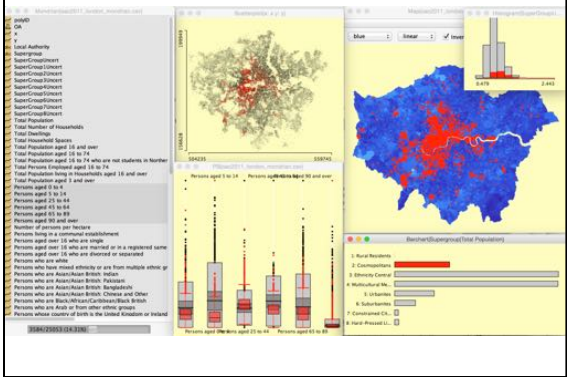


Example: GPS tracking of birds

Mondrian

Mondrian

- Interactive visual exploration software
  - Written by Martin Theus
  - http://www.theusrus.de/Mondrian/
  - Very fast, simple and effective
  - Reads CSV/TSV files



Mondrian

- Simple individual charts… but coordinated!
  - Barchart (& weighted)
  - Histogram (& weighted)
  - Scatterplot
  - Mosiac plot
  - Parallel coordinates
  - Scatter plot matrix (SPLOM)
  - Map

Mondrian: Data

- Standard single CSV or TSV file
  - Comma/tab-separated values
    - Rows = records; columns = attributes
  - Plus an additional `.map` file for choropleth maps
    - Polygons need to be specified in a particular format. Very fiddly. See example datasets and/or ask.

Mondrian: operation

- List of attributes
  - Icon identifies type –click to change
- Select one or more (SHIFT/CTRL for multiple)
  - Can weight by something (e.g. population)
- Use the plot menu to plot simple charts
  - Coordinated: selecting items will select corresponding items
- Calculate: min/max; can also connect to R
- Selection types: see options menu

Tableau

Tableau

- Interactive data visualisation focused on business intelligence
- Perhaps less suited for exploratory visual analysis

Tableau: data

- Tabular data
  - Many types: CSV, database connection, Excel, etc
- Concept of
  - Dimensions: categorical variables for pivoting
  - Measures: quantitative (numerical) variables for mapping to colour/size/etc
- May need to reshape data
  - http://kb.tableau.com/articles/knowledgebase/preparing-excel-files-analysis

Tableau: operation

- Column and row shelves:
  - Broadly, x-position and y-position
- Marks: the shapes that represent data
  - Map to "visual channels": size/colour/label
- Tableau will aggregate data according to what's in the shelves and marks
  - Drag the attribute for disaggregation to marks "detail".

Wrap up

Conclusions

- Visual Analytics is:
  - **analytical reasoning** by combining **computational analysis** with **interactive visual interfaces**
- Visualisation has an important role:
  - Well-designed visualisation present data effectively and facilitating comparison
  - Well-design interactions act as an interface to data, particularly in multiple-linked views
- Mondrian and Tableau enable visualisations/ interactions that can facilitate Visual Analytics

Intended learning outcomes

- Know what Visual Analytics is
- Know the role of interactive visualisation in Visual Analytics
  - Visual variables and when to use
  - Types of visualisation display and when to use
  - Types of interaction
  - Coordinated linked views
- (Practical) how to use Mondrian and Tableau

Reading

- Exploratory Analysis of Spatial and Temporal Data A Systematic Approach.
  - *Chapter 4.3 Visualization in a Nutshell*
  - *Download whole book!* http://0-dx.doi.org.wam.city.ac.uk/10.1007/3-540-31190-4
- **Munzner, T.** Visualization Analysis & Design, CRC Press
  - Chapters 2 and 5.