

# COMS21202: Symbols, Patterns and Signals

## Data Visualisation

[based on Dima Damen lecture notes]

Rui Ponte Costa

[rui.costa@bristol.ac.uk](mailto:rui.costa@bristol.ac.uk)

Department of Computer Science, University of Bristol  
Bristol BS8 1UB, UK

February 10, 2020

# Data Visualisation

- ▶ Information (or data) visualisation, as a research discipline, has emerged over the last 20 years
- ▶ Driven by the volumes of data and the accessibility of *big data*
- ▶ Characterised by large quantities of data - not necessarily numbers

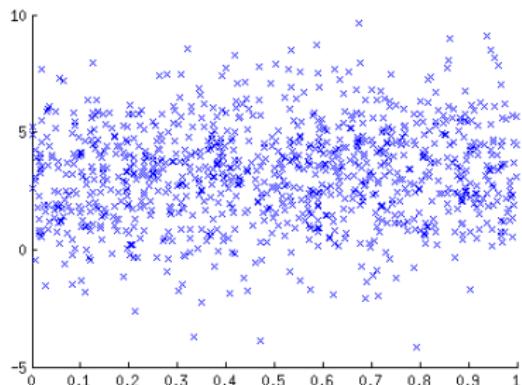
# Data visualisation is important!

Visualising the spread of coronavirus (Feb 2020, Johns Hopkins CSSE)



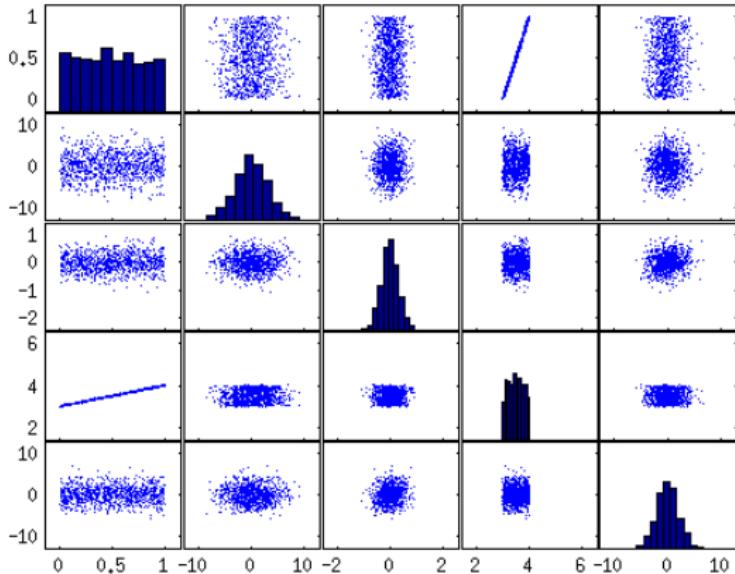
# Data Visualisation - Simple Graphs

## 1. Scatter Plot



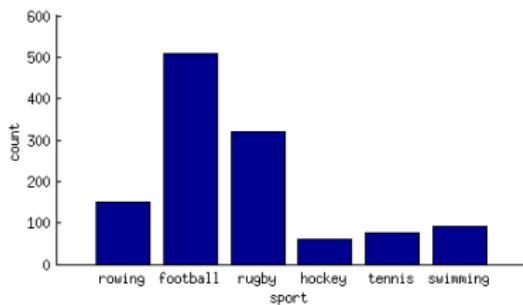
# Data Visualisation - Simple Graphs

1. Scatter Plot
2. Scatter Plot Matrix



# Data Visualisation - Simple Graphs

1. Scatter Plot
2. Scatter Plot Matrix
3. Histogram
  - ▶ Discrete Variable  
**(bar chart)**



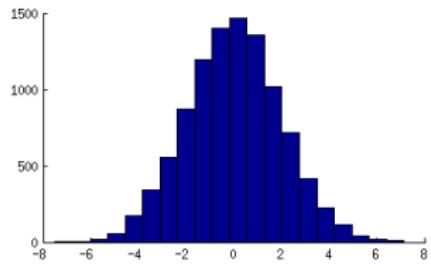
# Data Visualisation - Simple Graphs

1. Scatter Plot
2. Scatter Plot Matrix
3. Histogram
  - ▶ Discrete Variable
  - ▶ Continuous Variable

$\Delta v$ : width of bin

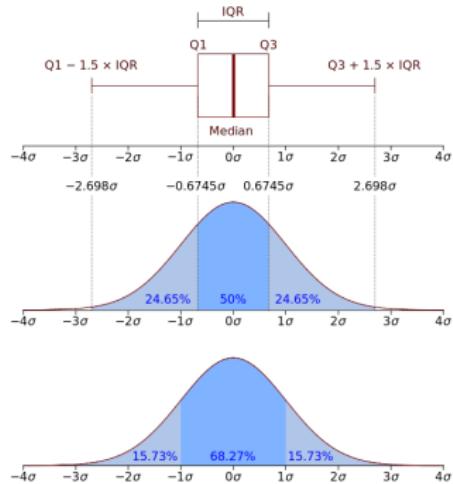
$c$ : bin number;  $0 \leq c \leq N$

$$min_x + c\Delta v \leq v < min_x + (c + 1)\Delta v$$

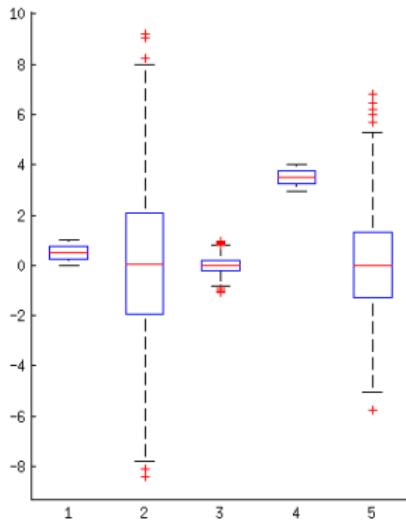


# Data Visualisation - Simple Graphs

1. Scatter Plot
2. Scatter Plot Matrix
3. Histogram
4. Box plot

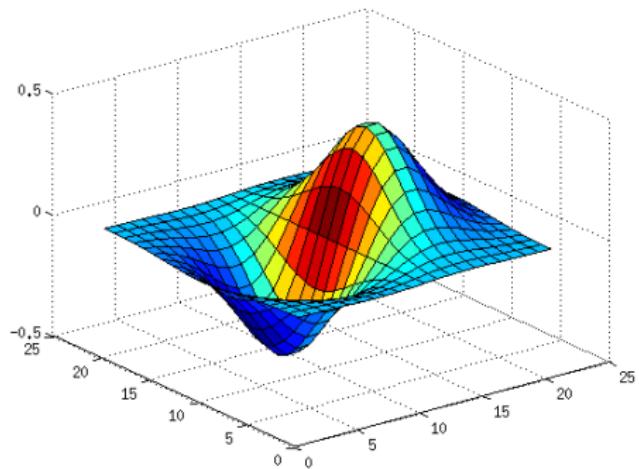


source:Wikipedia(2015)



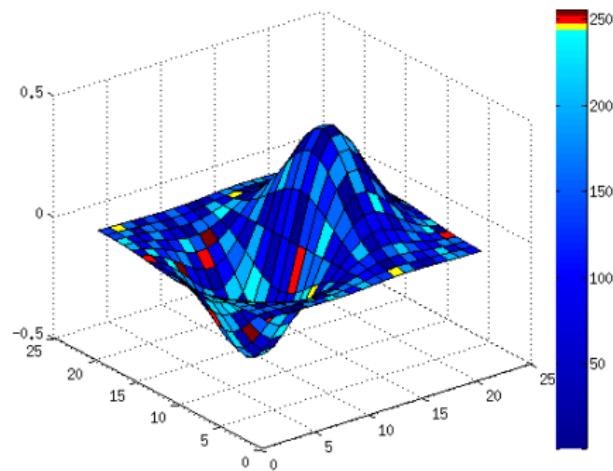
# Data Visualisation - Simple Graphs

1. Scatter Plot
2. Scatter Plot Matrix
3. Histogram
4. Box plot
5. Surface  
▶ 3D Data



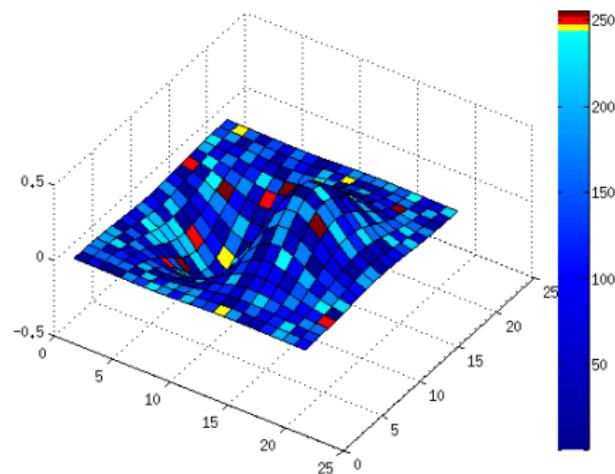
# Data Visualisation - Simple Graphs

1. Scatter Plot
2. Scatter Plot Matrix
3. Histogram
4. Box plot
5. Surface
  - ▶ 3D Data
  - ▶ 4D Data



# Data Visualisation - Simple Graphs

1. Scatter Plot
2. Scatter Plot Matrix
3. Histogram
4. Box plot
5. Surface
  - ▶ 3D Data
  - ▶ 4D Data



# Data Visualisation

- ▶ Is that all?
- ▶ **Problem:** How to understand massive datasets?
- ▶ **Solution:** Convert information into a graphical representation to take advantage of human perception
- ▶ Data visualisation – what can you vary?
  - ▶ Colour / Colour Maps
  - ▶ Size / contour width
  - ▶ Shape / line stroke
  - ▶ Location (2 dimensions)
  - ▶ Transparency
- ▶ Information visualisation: “The use of computer-supported, interactive, visual representations of abstract data to simplify cognition.” (Card, Mackinlay, Shneiderman, 1999)
- ▶ But it existed **before** computers!

# Historical Note - Ex1

## Napoleon's (disastrous) Russian invasion (1812)

*Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.*  
Dessiné par M. Minard, Inspecteur Général des Ponts et Chaussées en état-major.

Les nombreux hommes perdus sont représentés par les longueurs des lignes colorées à raison d'un millimètre pour dix mille hommes; ils sont le plus écrits en lettres des lignes. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été pris dans les messages de M. M. Cibot, de Ségur, de Tocqueville, de Chambry ou le journal intime de Jacob, pharmacien de l'Armée depuis le 23 Octobre. Pour mieux faire juger à l'œil la diminution de l'armée, j'ai superposé les corps de l'Armée à la Marche Davorin, qui aboutit sur Moscou. — Malgré ce rapprochement avec Cossack ou Witold, aucun voyage n'a été avec l'armée.

Paris, le 20 Novembre 1869.

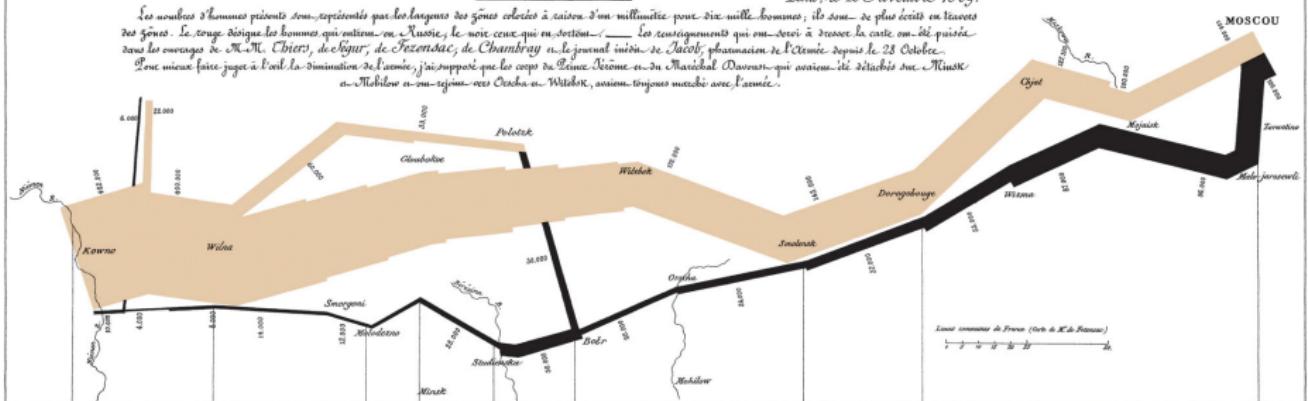
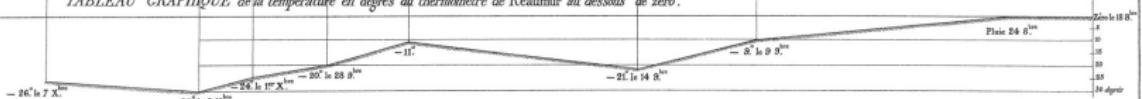


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les Campagnes passent au gelé  
le Nôtre, juillet.



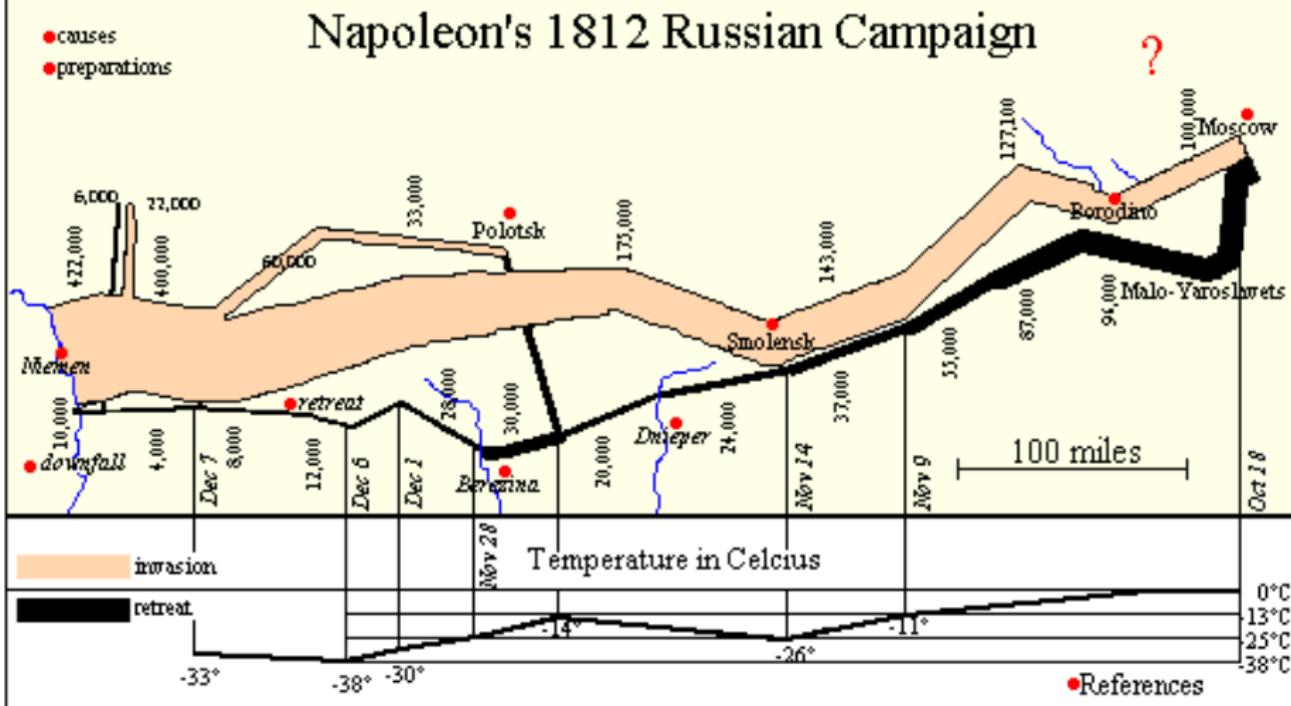
Arch. du Régiment, 8. Rue Félix-Marié 25<sup>e</sup> arr. de Paris.

Imp. L. Regnier à Bourges.

Further info <http://www.datavis.ca/gallery/re-minard.php>

# Historical Note - Ex1

Napoleon's (disastrous) Russian invasion (1812)



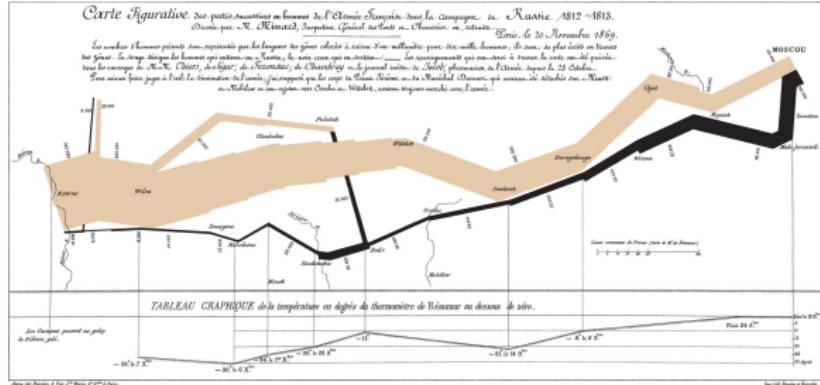
Further info <http://www.datavis.ca/gallery/re-minard.php>

Rui Ponte Costa

rui.costa@bristol.ac.uk

COMS21202: Data Visualisation

# Historical Note - Ex1



- ▶ Charles Minard
- ▶ Six dimensions of data
  - ▶ Number of Napoleon's troops
  - ▶ Direction
  - ▶ Distance
  - ▶ Temperature
  - ▶ Location: Longitude and Latitude
  - ▶ Dates
- ▶ Acclaimed best statistical graphic ever drawn

# Information Visualisation

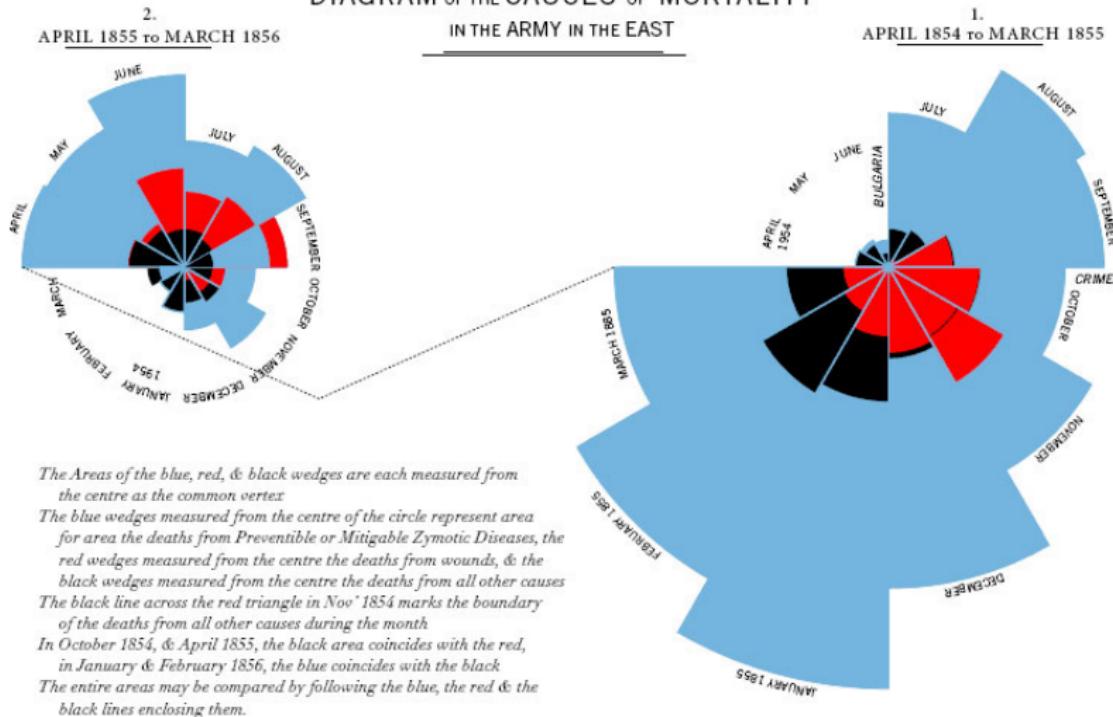
- ▶ Data is almost always multi-variate / multi-dimensional
- ▶ Visualisation should tell the truth about the data (or nearly)

$$\text{Lie Factor (LF)} = \frac{\text{size of effect in visualisation}}{\text{size of effect in data}}$$

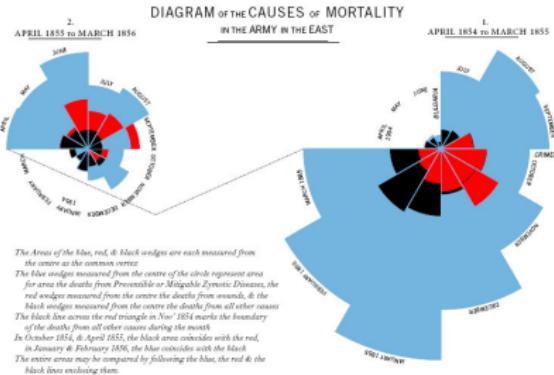
- ▶  $LF > 1 \rightarrow$  over stating
- ▶  $LF < 1 \rightarrow$  under stating
- ▶  $2 \leq LF \leq 5$  is common

# Historical Note - Ex2

DIAGRAM OF THE CAUSES OF MORTALITY  
IN THE ARMY IN THE EAST



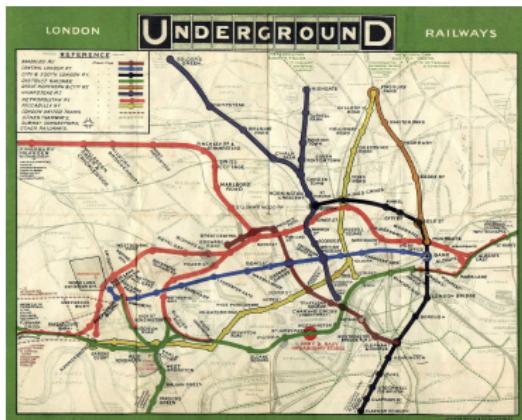
## Historical Note - Ex2



- ▶ Florence Nightingale
  - ▶ Four dimensional data (date, disease1, disease2, disease3)
  - ▶ Histogram-style (area instead of height)
  - ▶ colours for the various dimensions (ordered max to min)

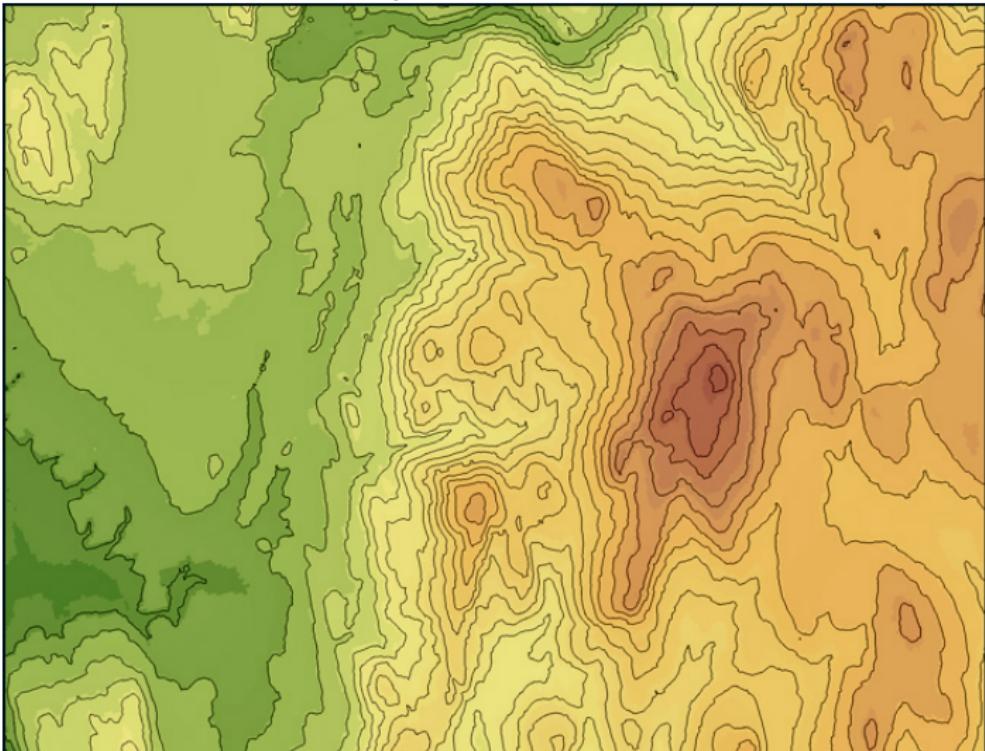
# Historical Note - Ex3

- ▶ A clean/structured representation is often more important [Harry Beck (1933)]

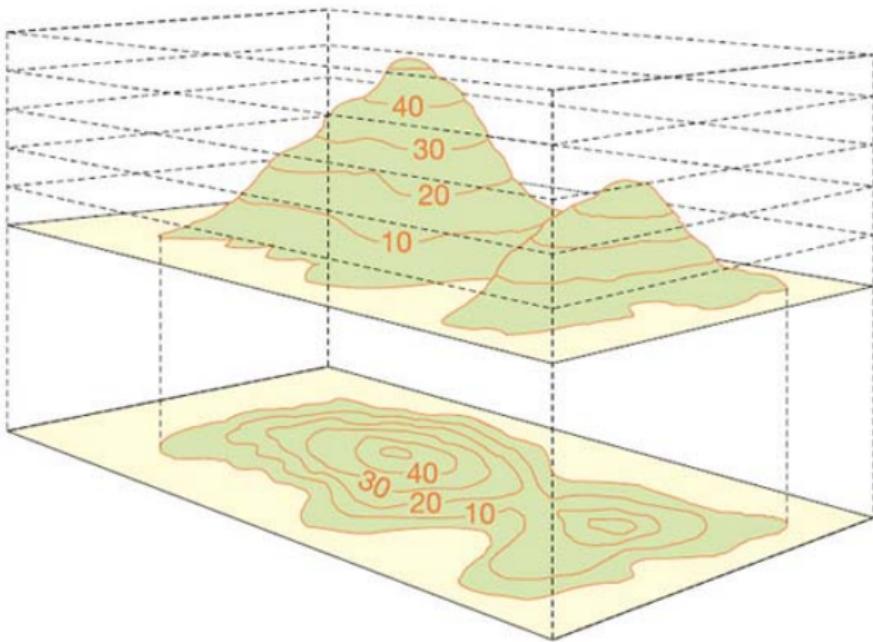


# Contour Maps

Colour map as a third dimension:



# Contour Maps



source: [ordnancesurvey.co.uk](http://ordnancesurvey.co.uk)

# Flow Charts

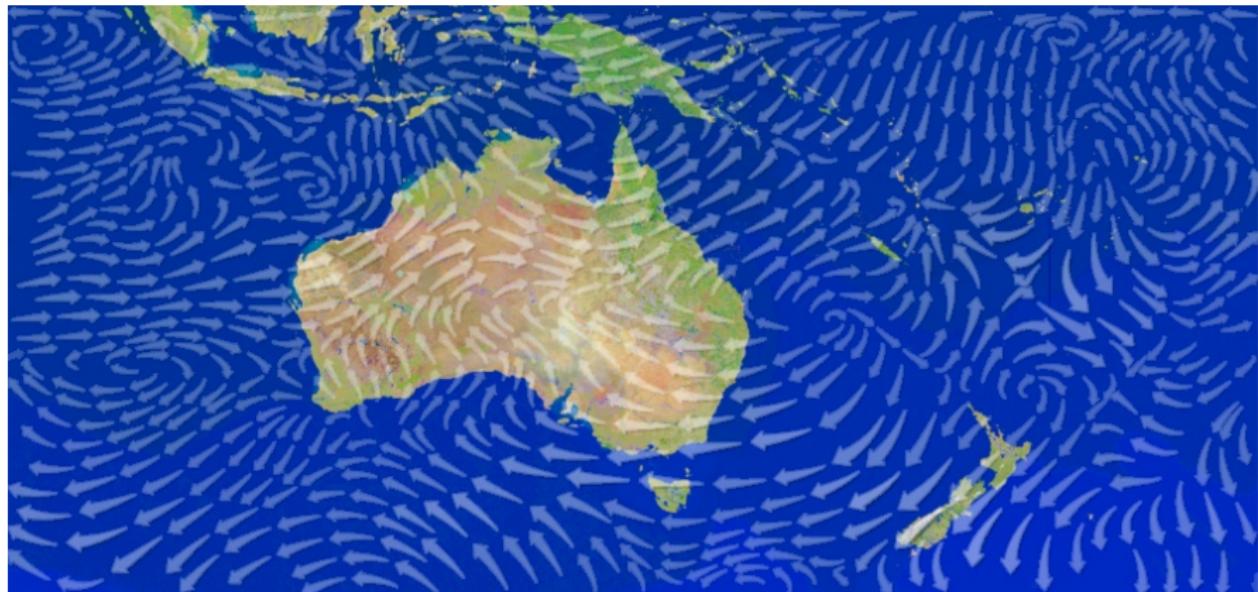
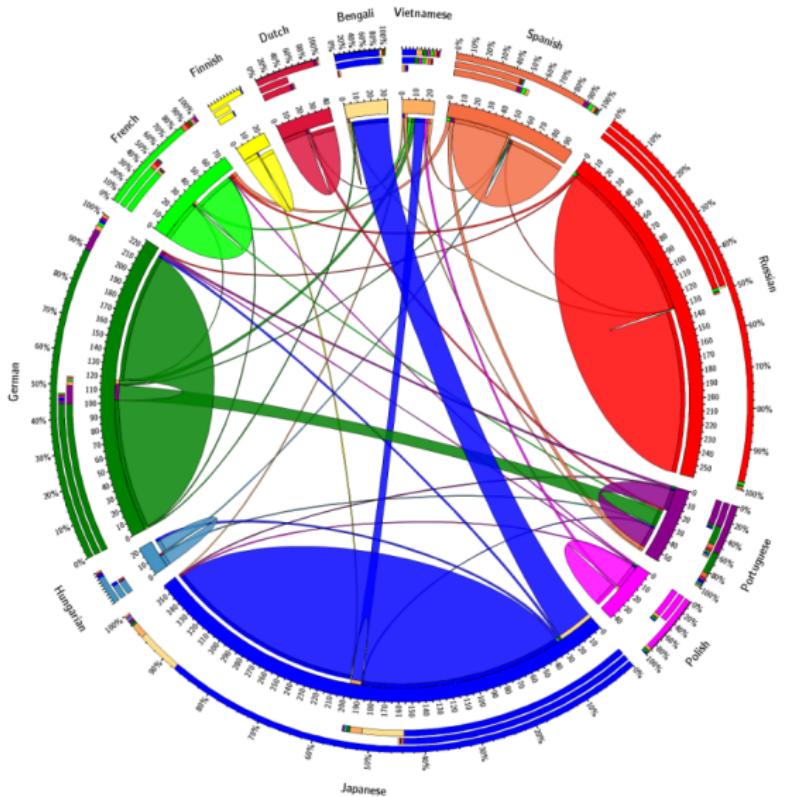


Image courtesy of Greg Turk

# Circular Graphs



Internet links between languages. AlNoamany et al (2013). Who and What Links to the Internet Archive

Rui Ponte Costa

rui.costa@bristol.ac.uk

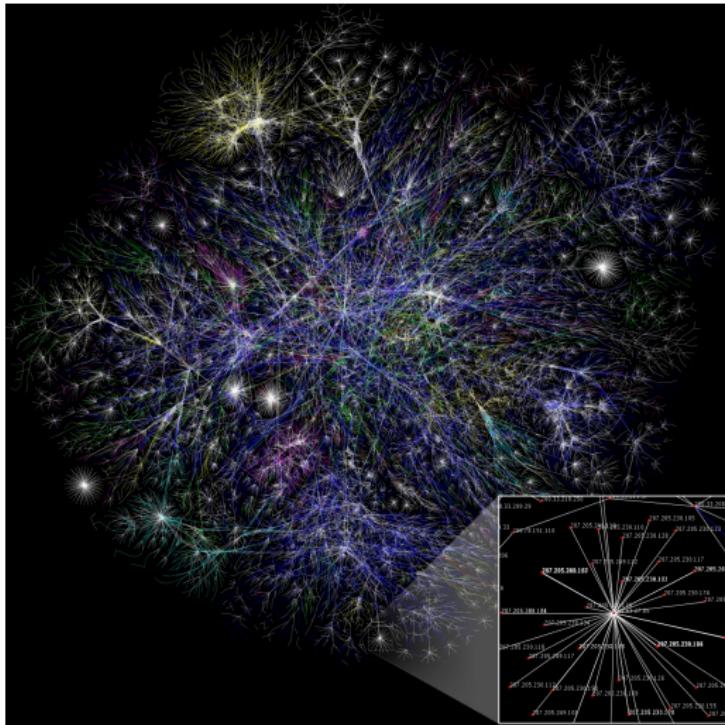
COMS21202: Data Visualisation

## Word clouds



Source: Wikipedia(2015)

# Graphs and Hierarchies



Complex networks: Universe or internet? [Wikipedia (2015)]

# Data visualisation as (informative) art

- ▶ Empires decline: [vimeo.com/11506746](https://vimeo.com/11506746) [Cruz and Machado, SIGGRAPH 2010]
- ▶ Lisbon traffic: [vimeo.com/10218235](https://vimeo.com/10218235) [Cruz and Machado, 2009]<sup>1</sup>

---

<sup>1</sup>More at [cdv.dei.uc.pt](http://cdv.dei.uc.pt)

# Further Reading

- ▶ **Visualization Handbook**  
Hansen and Johnson (2004)