Data Visualisation

(CMP020L013S)

Week 1: Introduction

Mohammad Javaheri

(Dr Mohammad Ali Javaheri Javid)

Agenda

- Introduction to the module and logistics
- Data, information, history of visualisation
- The need for visualisations, definitions

Learning Outcomes

- Critically analyse the effectiveness of a given visualisation technique for a particular task.
- Critically evaluate the appropriate visualisation technique for specific problems.
- Develop visualisations to report on studies of both quantitative and qualitative evaluations.

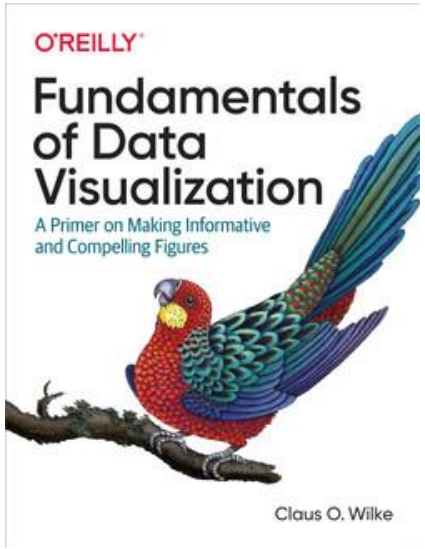
- Python skill are pre-requisite for this module

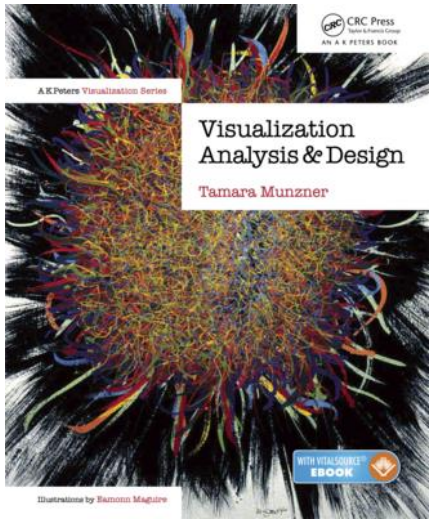
Module Logistics

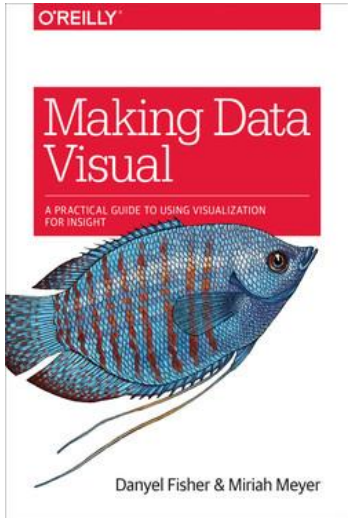
- Delivery approach
- Expectations, preparation, lab sessions
- Attendance monitoring

Module Material/Reading List

- Matplotlib, NumPy, and Seaborn etc.
- We will use only libraries which are available/can be imported on the uni machines

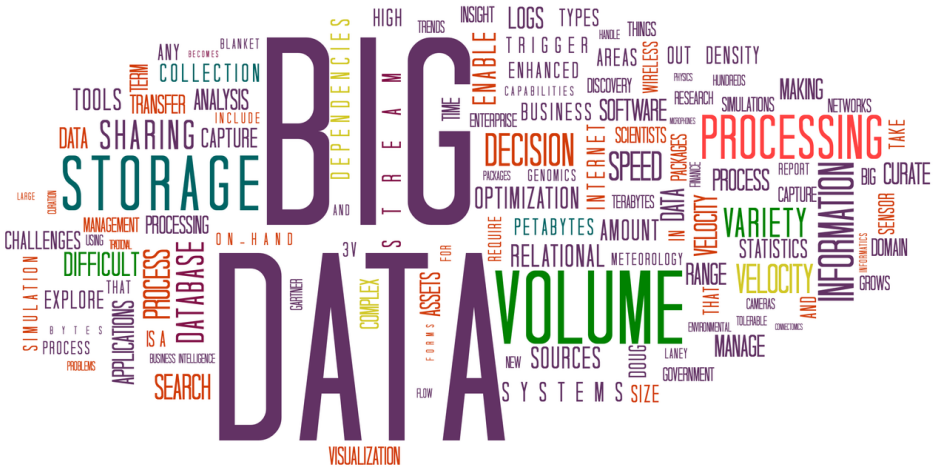






Big Data and The Information Problem

- We are Data Rich but Information Poor (DRIP syndrome)



Is data enough?

Data vs Information

- Data != Information
- Data Vis: data → information
  - Using power of human visual processing
  - Making visible the patterns and structures in the data
  - Using graphs, tables, diagrams, maps…

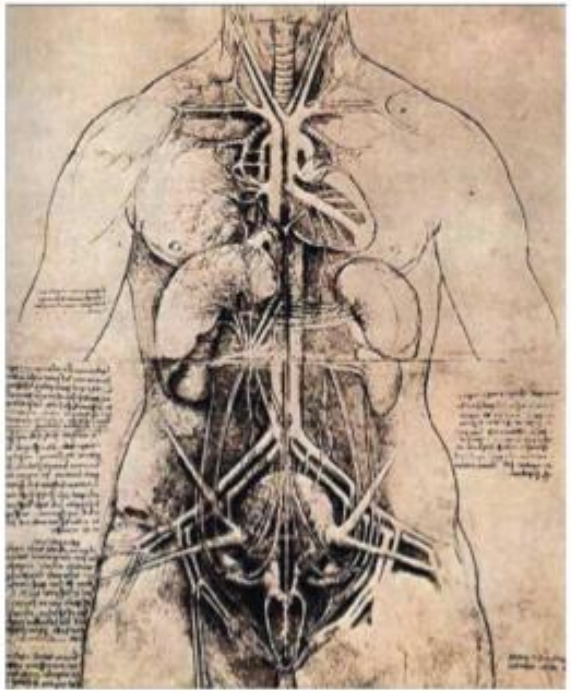
History of Visualisation

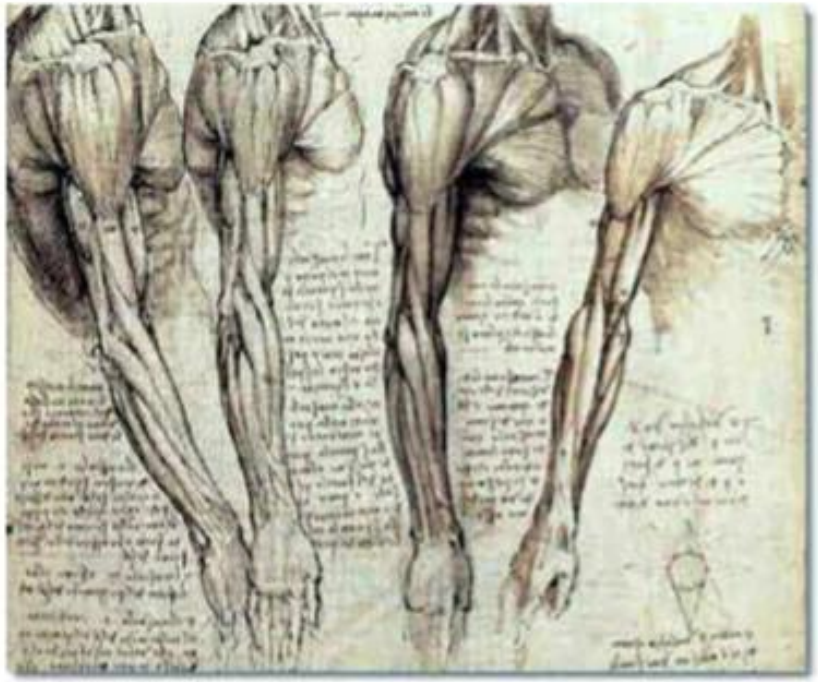
- Dr. John Snow
- Location of deaths in the 1854 London Cholera Epidemic
- X marks the locations of the water pumps

History of Visualization

- Visualization = rather old

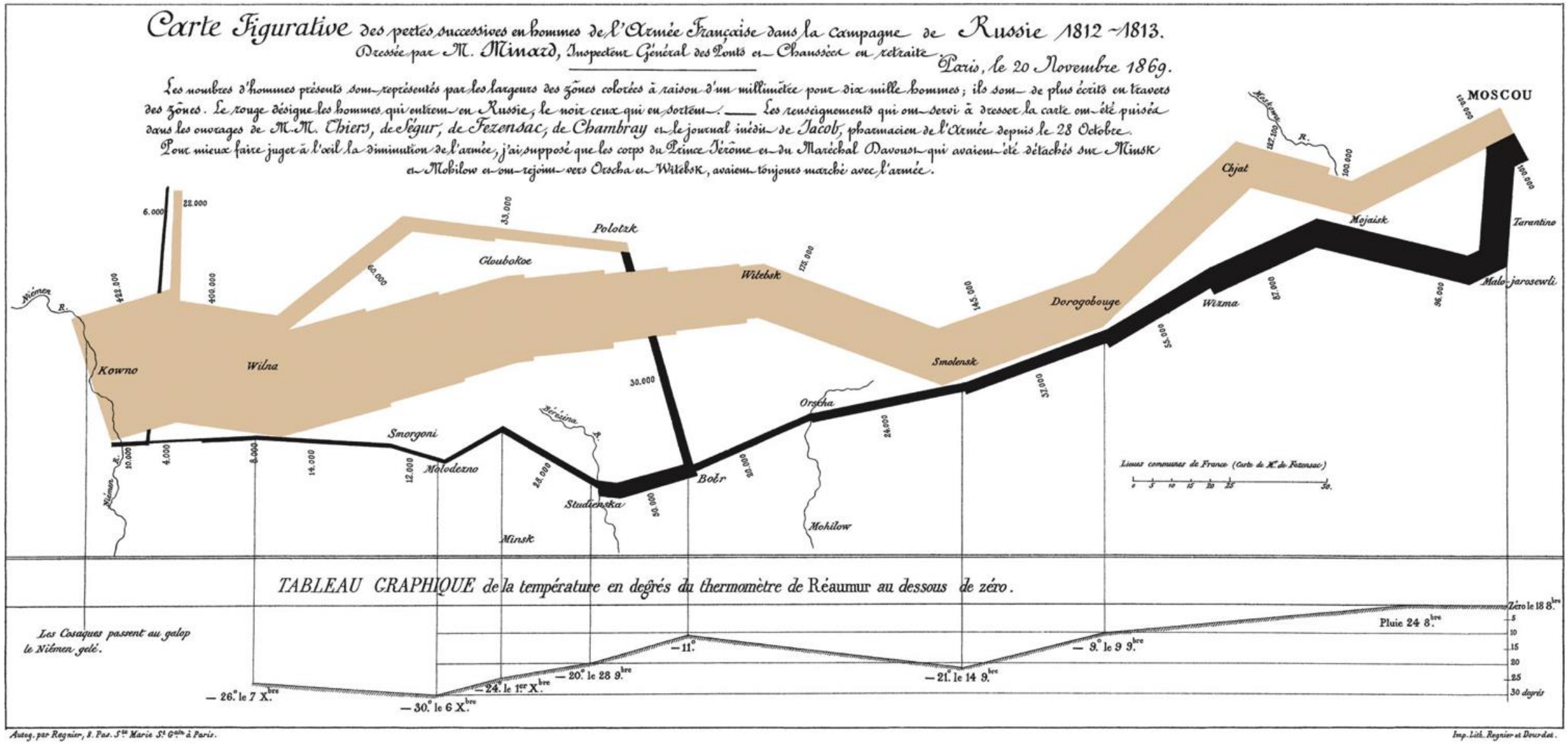
L. da Vinci (1452-1519)





- Often an intuitive step: graphical illustration

Charles Joseph Minard 1869 Napoleon's March



Defining Visualisation

- The National Science Foundation (U.S.):
  - The use of computer graphics for the analysis and presentation of computed or measured scientific data.
- Oxford Engl. Dict., 1989
  - To form a mental vision, image, or picture of (something not visible or present to the sight, or of an abstraction); to make visible to the mind or imagination
- Visualization transforms data into images that effectively and accurately represent information about the data.

Defining Visualisation

- "Transformation of the symbolic into the geometric"

  [McCormick et al. 1987]

- "... finding the artificial memory that best supports our natural means of perception." [Bertin 1967]

- "The use of computer-generated, interactive, visual representations of data to amplify cognition."

  [Card, Mackinlay, & Shneiderman 1999]

Defining Visualisation

- Visualisation is the process that transforms (abstract) data into interactive graphical representations for the purpose of exploration, confirmation, or presentation.
- **Tool to enable a user insight into Data**

Defining Visualisation

- Computer-based visualisation systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- Why we need Vis?
- Why have a human in the loop?

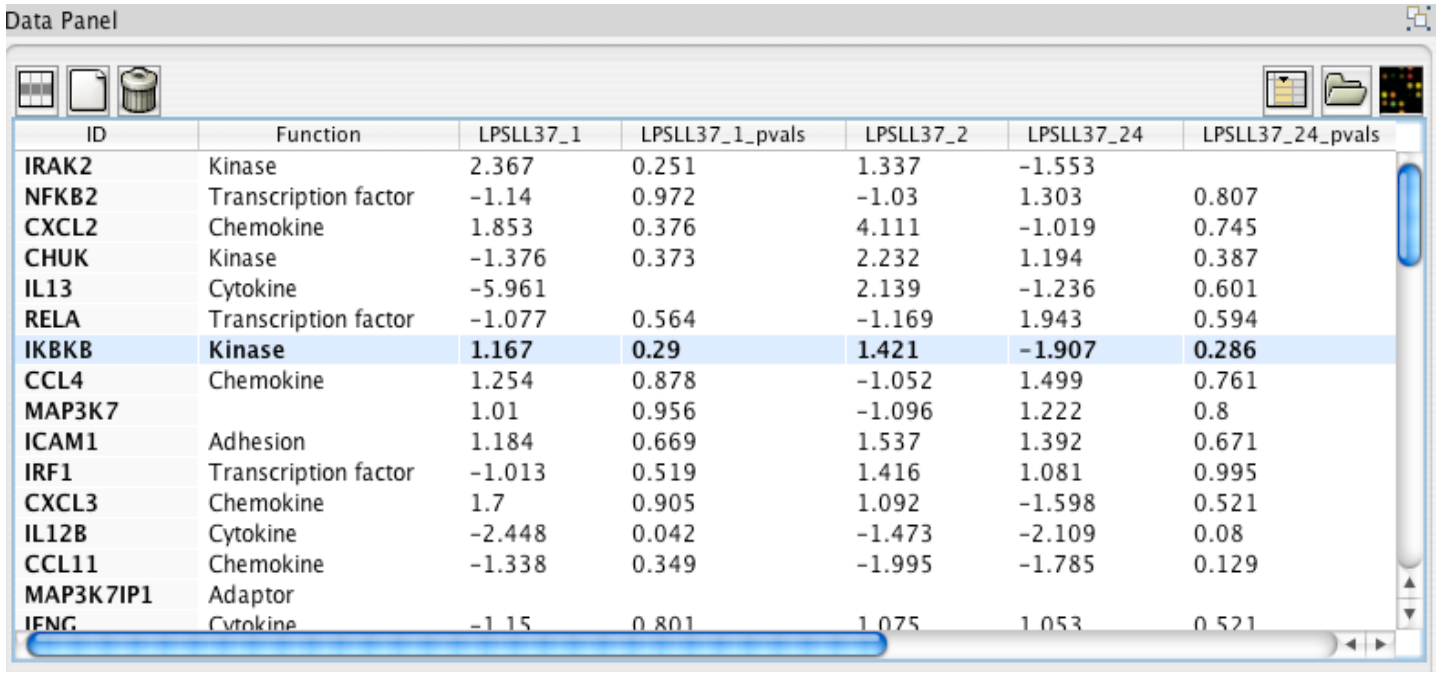
Defining Visualisation

- Computer-based visualisation systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

Why use an external representation?

- visual representations as external representation: replace cognition with perception



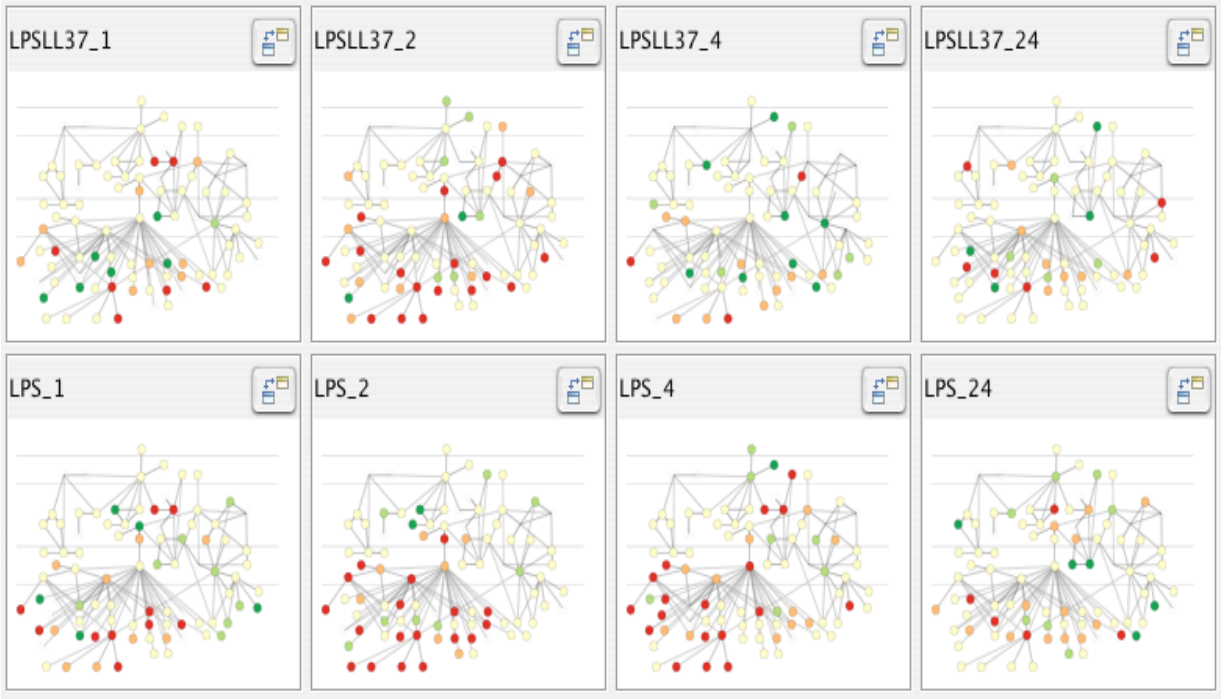
ID	Function	LPSLL37_1	LPSLL37_1_pvals	LPSLL37_2	LPSLL37_24	LPSLL37_24_pvals
IRAK2	Kinase	2.367	0.251	1.337	-1.553	
NFKB2	Transcription factor	-1.14	0.972	-1.03	1.303	0.807
CXCL2	Chemokine	1.853	0.376	4.111	-1.019	0.745
CHUK	Kinase	-1.376	0.373	2.232	1.194	0.387
IL13	Cytokine	-5.961		2.139	-1.236	0.601
RELA	Transcription factor	-1.077	0.564	-1.169	1.943	0.594
IKBKB	Kinase	1.167	0.29	1.421	-1.907	0.286
CCL4	Chemokine	1.254	0.878	-1.052	1.499	0.761
MAP3K7		1.01	0.956	-1.096	1.222	0.8
ICAM1	Adhesion	1.184	0.669	1.537	1.392	0.671
IRF1	Transcription factor	-1.013	0.519	1.416	1.081	0.995
CXCL3	Chemokine	1.7	0.905	1.092	-1.598	0.521
IL12B	Cytokine	-2.448	0.042	-1.473	-2.109	0.08
CCL11	Chemokine	-1.338	0.349	-1.995	-1.785	0.129
MAP3K7IP1	Adaptor					
IFNG	Cytokine	-1.15	0.801	1.075	1.053	0.521

[Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, and Kincaid. IEEE TVCG (Proc. InfoVis) 14(6):1253-1260, 2008.]

Why use an external representation?

Data Panel

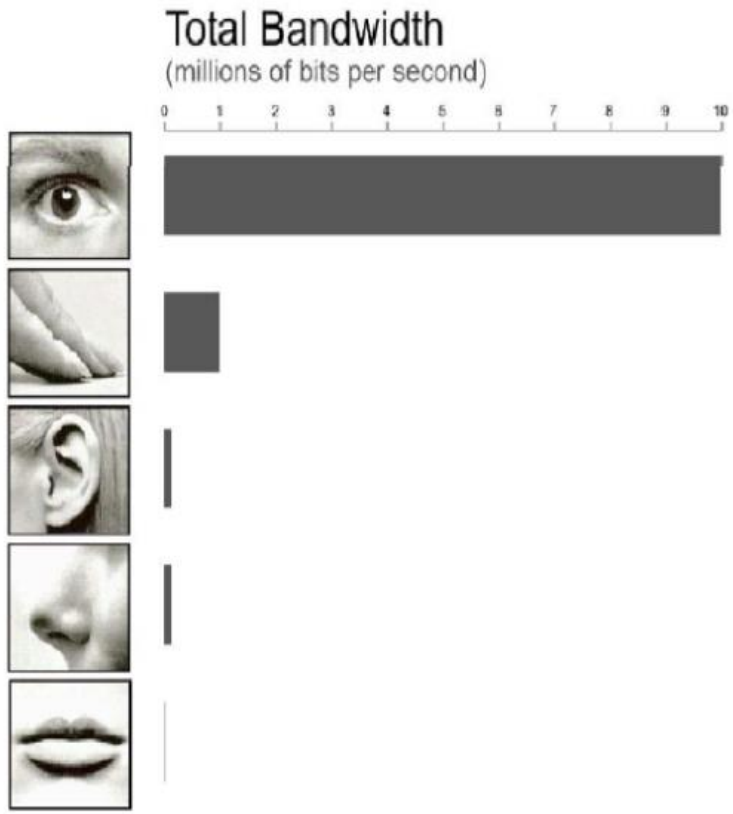
ID	Function	LPSLL37_1	LPSLL37_1_pvals	LPSLL37_2	LPSLL37_24	LPSLL37_24_pvals
IRAK2	Kinase	2.367	0.251	1.337	-1.553	
NFKB2	Transcription factor	-1.14	0.972	-1.03	1.303	0.807
CXCL2	Chemokine	1.853	0.376	4.111	-1.019	0.745
CHUK	Kinase	-1.376	0.373	2.232	1.194	0.387
IL13	Cytokine	-5.961		2.139	-1.236	0.601
RELA	Transcription factor	-1.077	0.564	-1.169	1.943	0.594
IKBKB	Kinase	1.167	0.29	1.421	-1.907	0.286
CCL4	Chemokine	1.254	0.878	-1.052	1.499	0.761
MAP3K7		1.01	0.956	-1.096	1.222	0.8
ICAM1	Adhesion	1.184	0.669	1.537	1.392	0.671
IRF1	Transcription factor	-1.013	0.519	1.416	1.081	0.995
CXCL3	Chemokine	1.7	0.905	1.092	-1.598	0.521
IL12B	Cytokine	-2.448	0.042	-1.473	-2.109	0.08
CCL11	Chemokine	-1.338	0.349	-1.995	-1.785	0.129
MAP3K7IP1	Adaptor					
IFNG	Cytokine	-1.15	0.801	1.075	1.053	0.521



[Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, and Kincaid. IEEE TVCG (Proc. InfoVis) 14(6):1253-1260, 2008.]

Why depend on vision?

- Figures are richer; provide more information with less clutter and in less space.
- Figures provide the *gestalt* effect: they give an overview; make a structure more visible.
- Figures are more accessible, easier to understand, faster to grasp, more comprehensible, more memorable, more fun, and less formal.

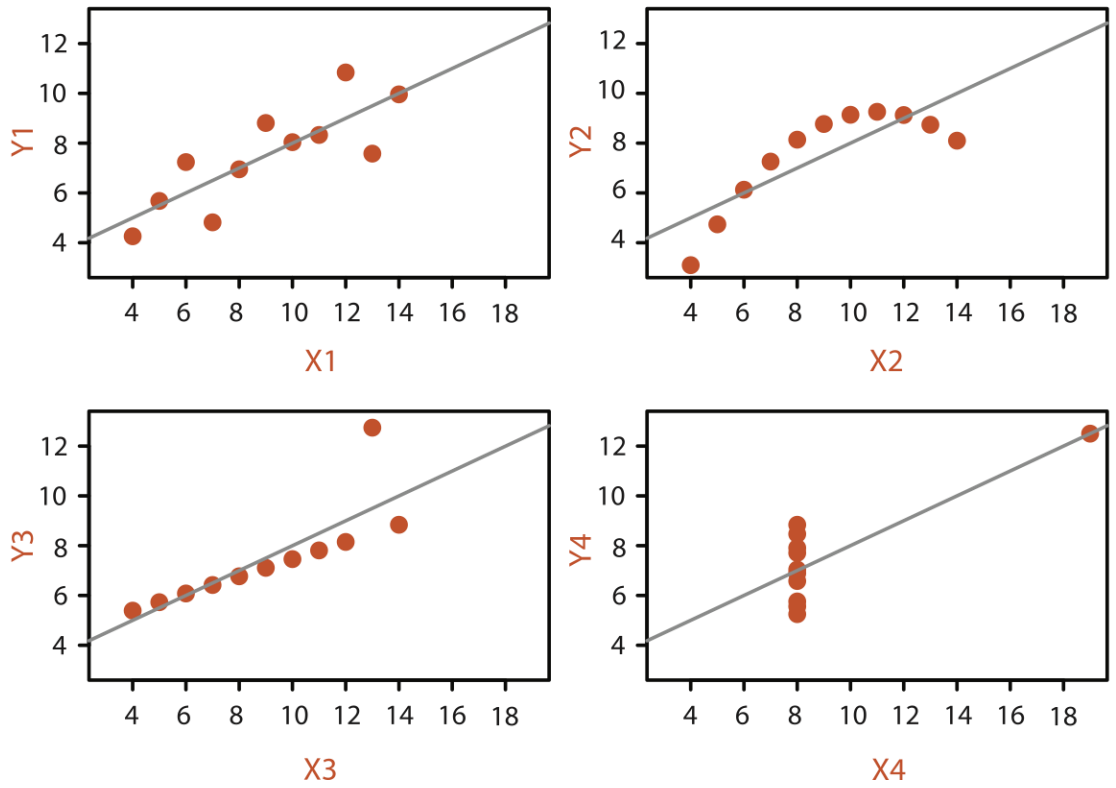


Why depend on vision?

- Human visual system is a high-bandwidth channel to the brain
  - overview possible due to background processing
    - subjective experience of seeing everything simultaneously
    - significant processing occurs in parallel and pre-attentively
- Sound: lower bandwidth and different semantics
  - overview not supported
    - subjective experience of sequential stream
- Touch/haptics: impoverished record/replay capacity
  - only very low-bandwidth communication thus far
- Taste/smell: no viable record/replay devices

Why represent all the data?

- summaries lose information, details matter
  - confirm expected and find unexpected patterns
  - assess the validity of the statistical model



Identical statistics	
x mean	9
x variance	10
y mean	7.5
y variance	3.75
x/y correlation	0.816

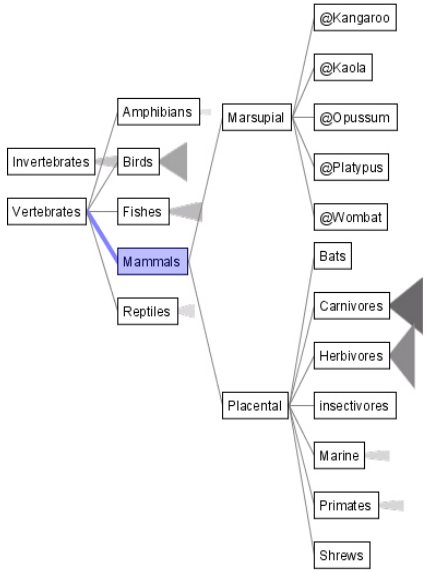
What resource limitations are we faced with?

- Vis designers must take into account three very different kinds of resource limitations:
- Display limits
  - pixels are precious and the most constrained resource
  - **information density**:
  - ratio of space used to encode info vs unused whitespace
    - tradeoff between clutter and wasting space
    - find a sweet spot between dense and sparse
- human limits
  - human time, human memory, human attention

Why analyse Visualition?

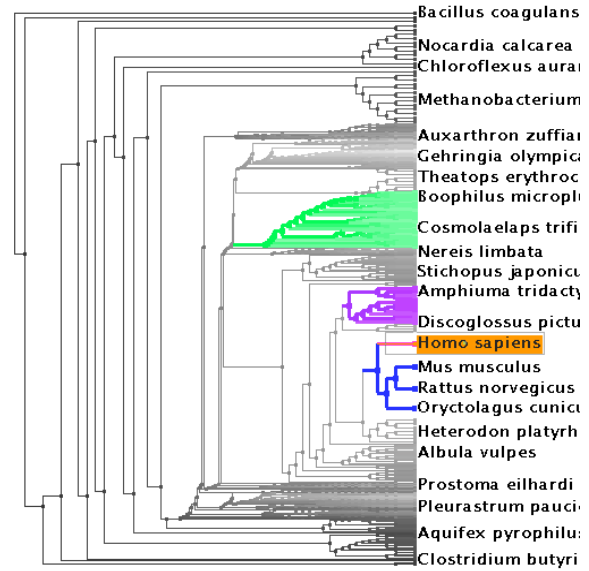
- Helps to think systematically about choices of visualisation
- Analysing existing visualisations as a stepping stone to designing new ones
- Vis design is full of trade-offs,
- most possibilities in the design space are ineffective for a particular task
- Validating the effectiveness of a design is both necessary and difficult

SpaceTree



[SpaceTree: Supporting Exploration in Large Node Link Tree, Design Evolution and Empirical Evaluation. Grosjean, Plaisant, and Bederson. Proc. InfoVis 2002, p 57–64.]

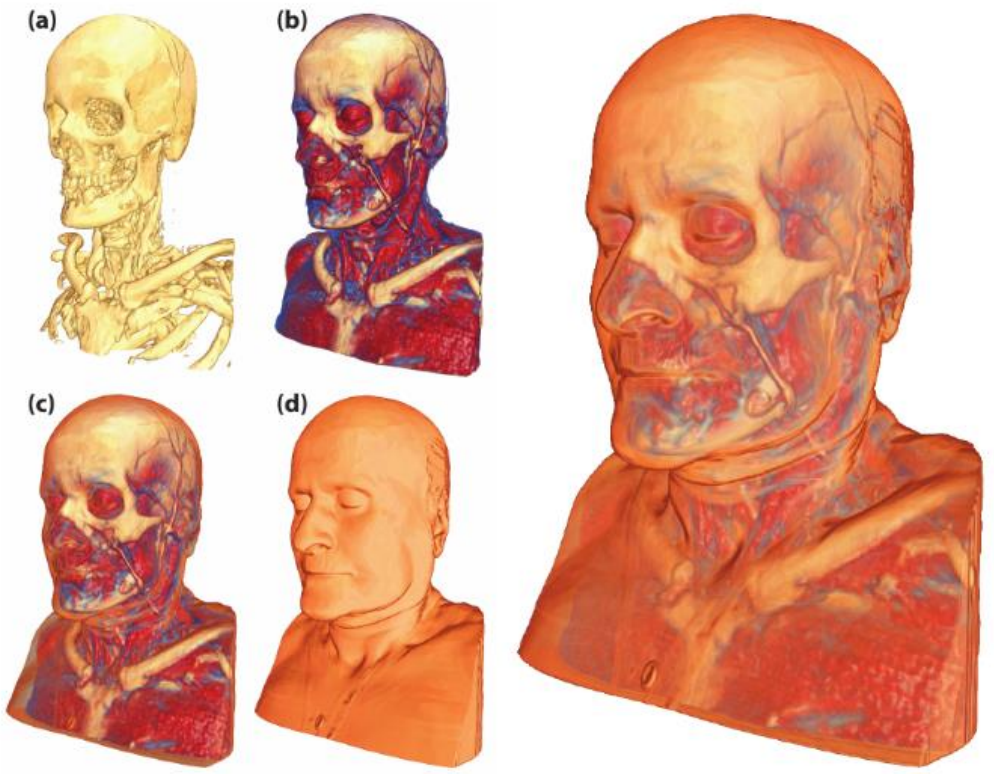
TreeJuxtaposer



[TreeJuxtaposer: Scalable Tree Comparison Using Focus+Context With Guaranteed Visibility. ACM Trans. on Graphics (Proc. SIGGRAPH) 22:453– 462, 2003.]

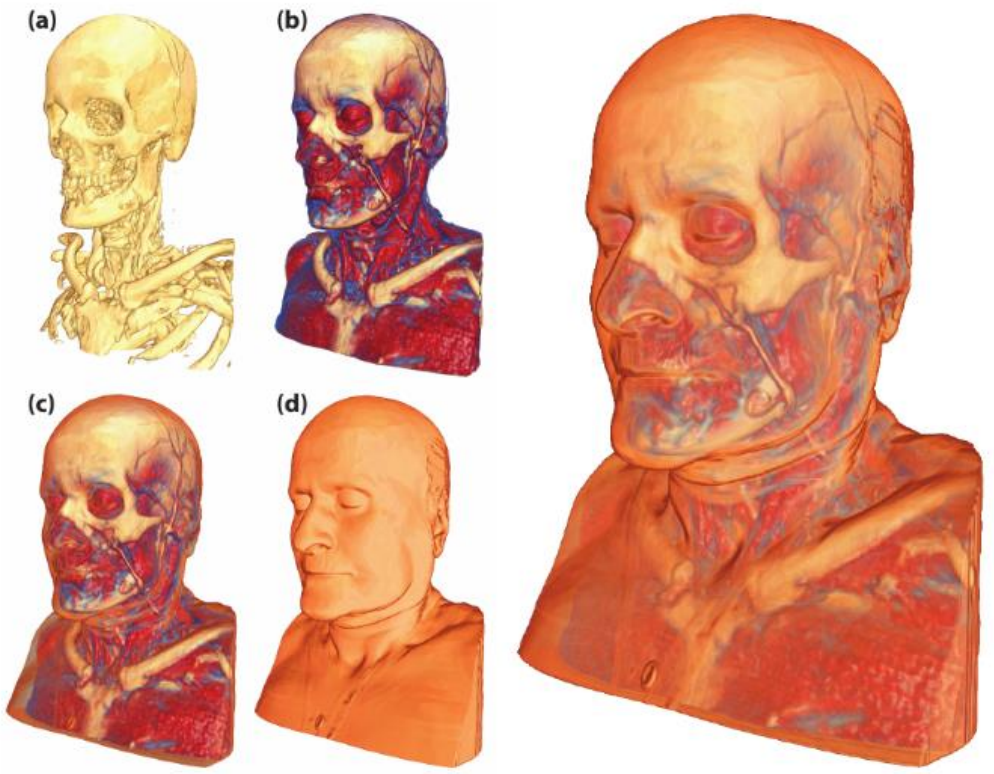
Why Use Computers?

- Scale
- Drawing by hand(Illustrator)
- Infeasible
- inflexible (updates!)
- How to draw an MRI scan?



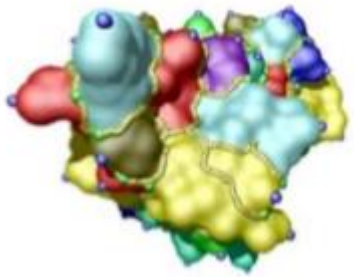
Why Use Computers?

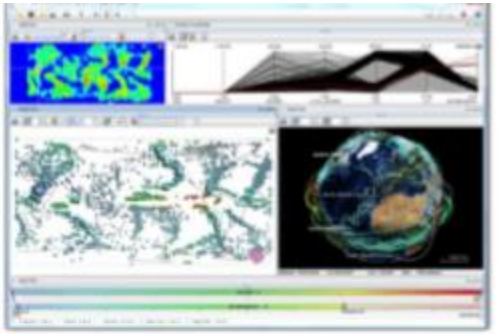
- Integration with algorithms
- Make visualization part of a data analysis pipeline
- Efficiency
  - Re-use charts / methods for different datasets
- Quality
  - Precise data driven rendering
- Storytelling
  - Use time

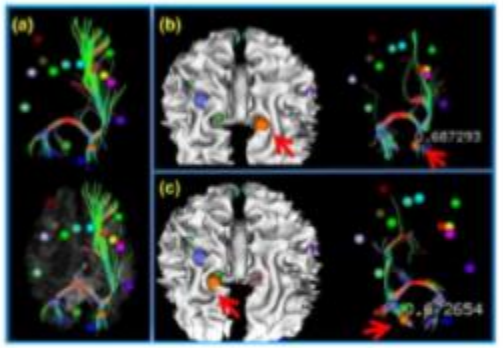


Goals of Visualisation

- Three goals for visualisation
- Explore
  - Nothing is known about the data
  - Vis is used for data exploration
- Analyse
  - There are hypotheses/assumptions
  - Vis is used for verification or falsification
- Present
  - Everything is known about the data
  - Vis is used to communicate the results







Further Applications of Visualisation

A: Support reasoning about information (analysis)

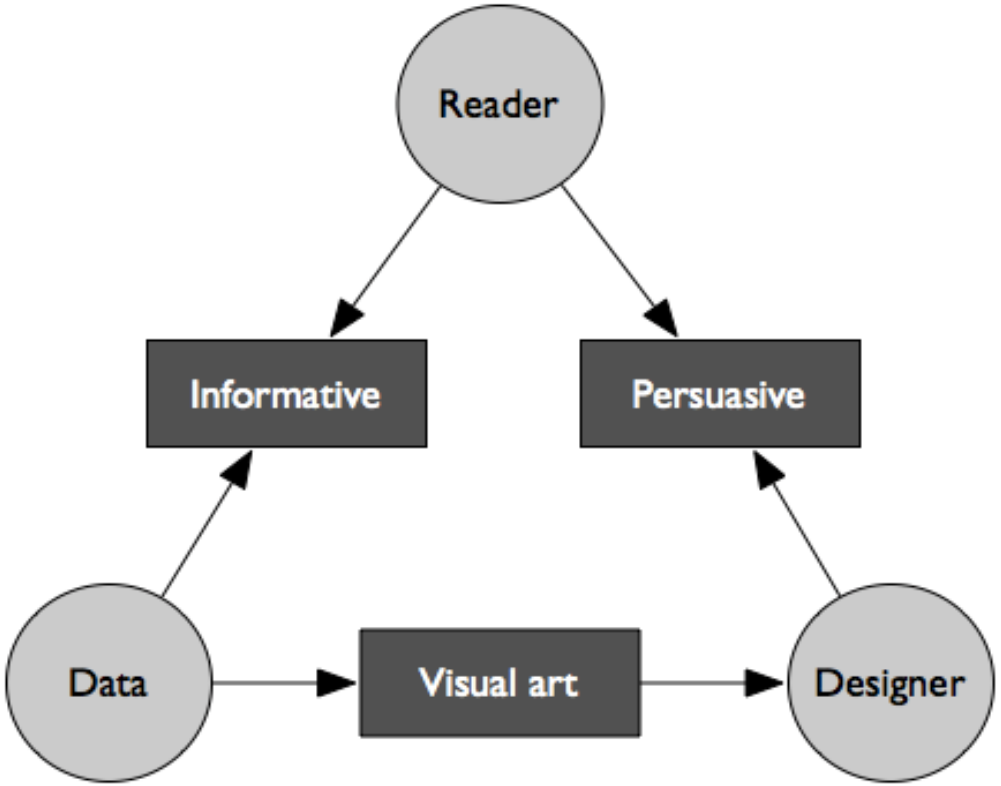
- Finding relationships
- Discover structure
- Quantifying values and influences
- Should be part of a query/analysis cycle

B: Inform and persuade others (communication)

- Capture attention, engage
- Tell a story visually
- Focus on certain aspects, and omit others

Data Presentation Trinity in Vis

- Designer-Reader-Data Trinity



"Designing Data Visualizations", Iliinsky and Steele, O'Reilly, 2011

Next Week

- Types of Data Visualisation
- Use Cases of Data Visualisation
- Value of Data Visualisation
- Good, bad and ugly visualiations

Further Material

- The famous Gapminder Video, Hans Rosling:
  200 Countries, 200 Years, 4 Minutes
- https://www.youtube.com/watch?feature=player_embedded&v=jbkSRLYSojo
- Florence Nightingale was one of the pioneers of data visualisation
- https://youtu.be/yhX0OR1_Vfc