Visualization Analysis & Design   Tamara Munzner

CRC Press, 2014

What is Visualization and Why
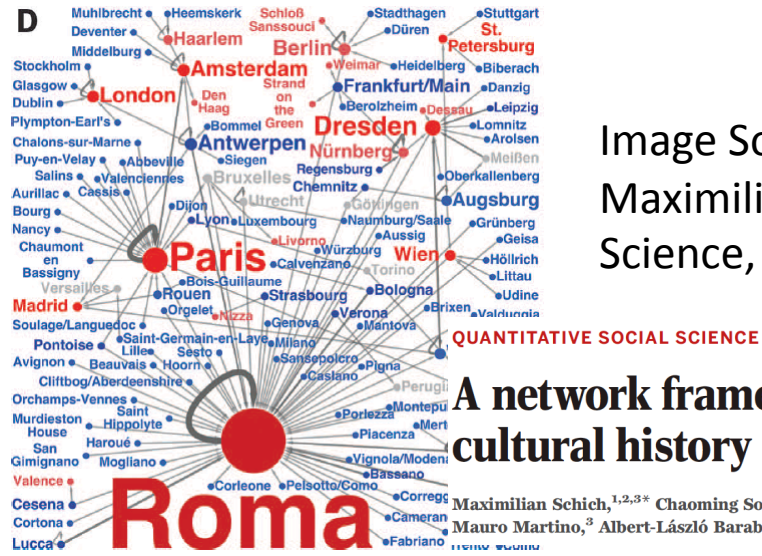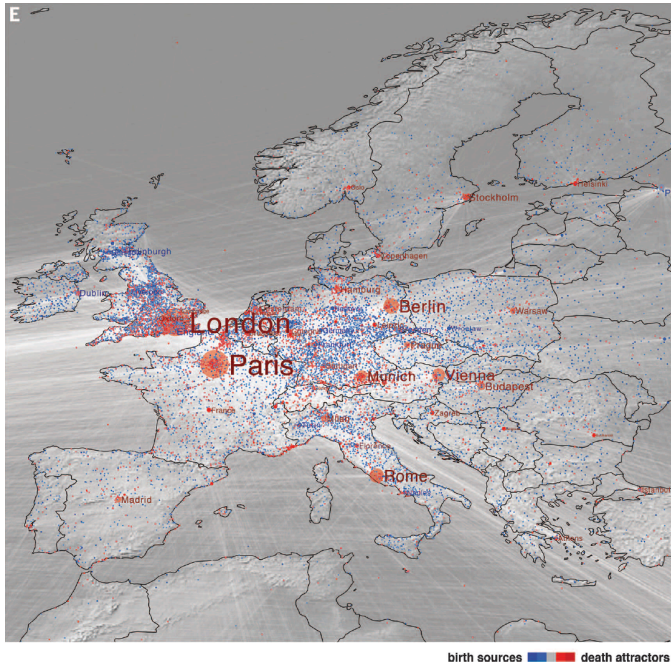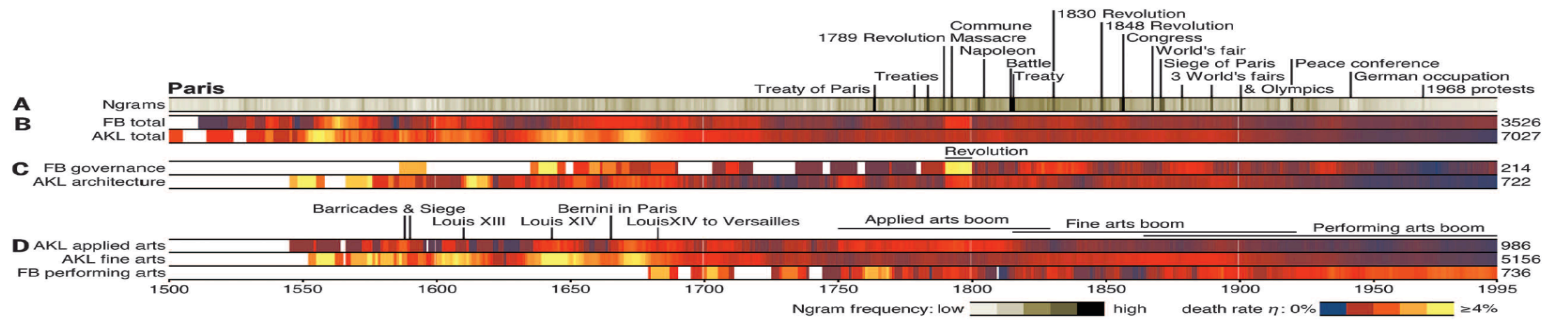

Image Source:
Maximilian Schich eta al.
Science, March 2014

What is Visualization

Visual representation of data sets designed to help people carry out tasks more effectively

- Augmenting people's ability rather than replacing
- The design space is huge, with many trade offs
- Validating design is both necessary and challenging
- Need to address three resource limitations
  - Computers
  - Human
  - Displays

We analyzing visualization by answering the following questions:
    Why - people need it
   What - data are shown
    How - the idiom is designed

Why human in the loop

- When people do not know exactly what questions to ask in advance
  - They do not know how to approach the problem
- Putting human in the loop – enhance their ability rather than replace
- Three uses of visualization:
  - Transitional use: help designers with future solutions that are completely computational
  - Long term use: exploratory data analysis
  - Presentation use: explain something that is already known

Why computer in the loop

- People are unlikely to move beyond tiny datasets

- Complex drawings cannot be done by human
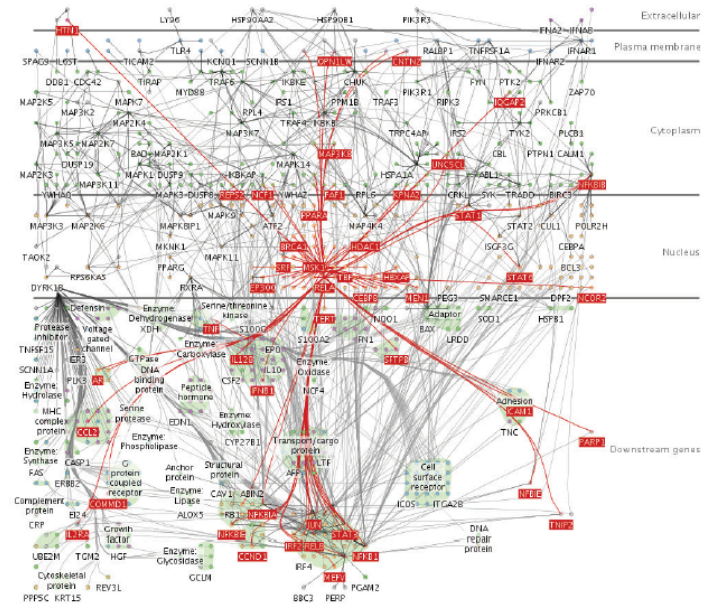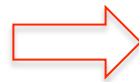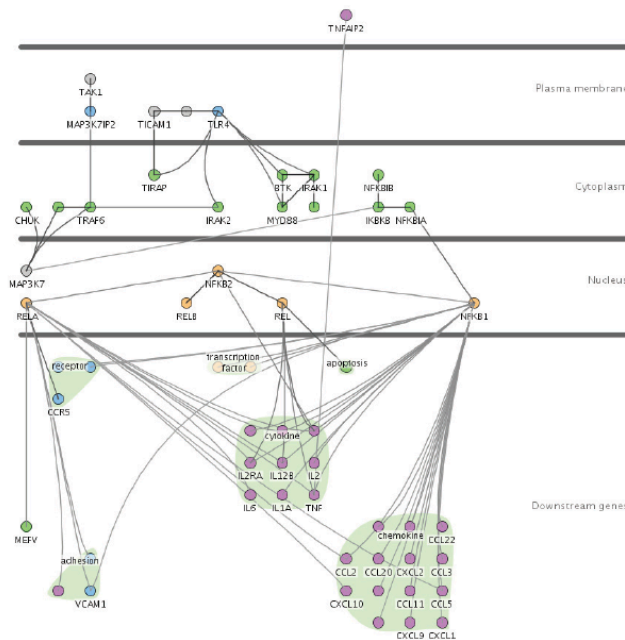
# Why use external representations

- Augment human's capacity by surpassing our internal cognition and memory limitations
- External representation (diagram)= external memory
  - Organize information in spatial locations
  - Accelerate search and cognition
  - Grouping relevant information in nearby locations

Why depend on vision

- The visual system provides a high bandwidth channel

- Process information in parallel

- Popout occurs even when the number of objects is large

- Sound is not as effective (sequential stream)

- Technological limitations rule out other senses
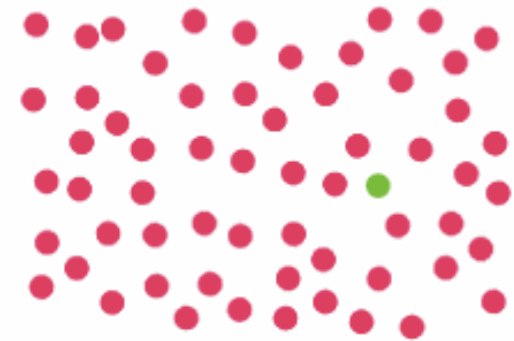
The green dot pops out
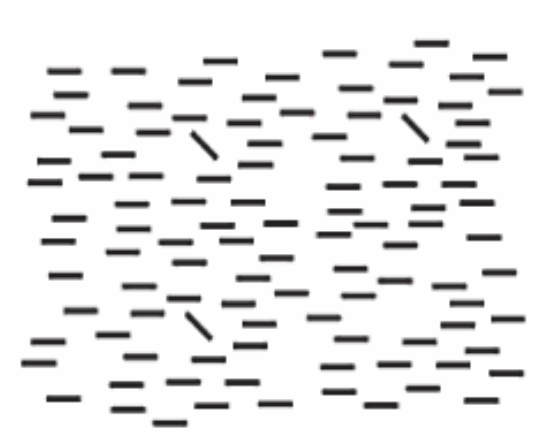

The oblique lines pop out

Why show data in detail

- Exploring data to find patterns
  - Confirm expected ones and discover unexpected ones
  - Assess the validity of statistical models
    - Identical statistics does not mean similar data

Anscombe's Quartet: Raw Data

|		1		2		3		4	
		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
Mean		9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance		10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation		0.816		0.816		0.816		0.816	

Why show data in detail

- Exploring data to find patterns
  - Confirm expected ones and discover unexpected ones
  - Assess the validity of statistical models
   - Identical statistics does not mean similar data
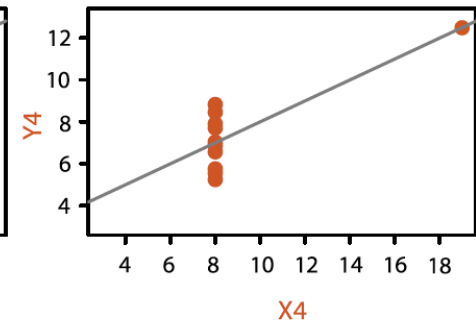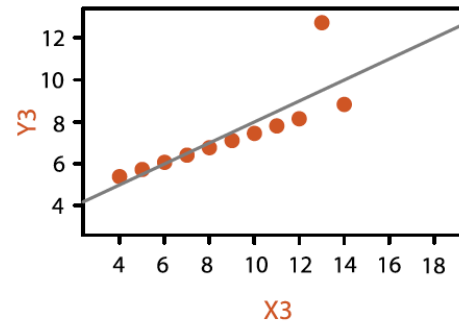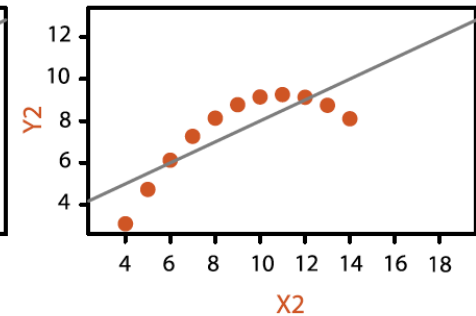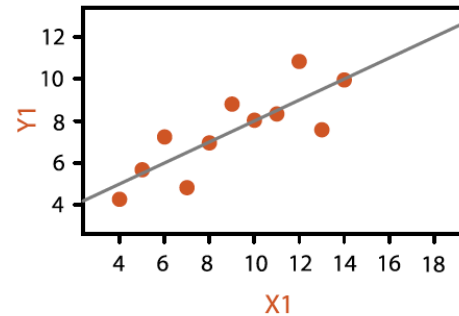
Why interactivity

- Crucial for handling data complexity
- Overcome the limitations of people and display – not possible to show all the data at once
- Nee to show multiple aspects of the data
- Change displays to support many queries

Why vis idiom design space is huge

- There exist many ways to create visual encoding
  - Become even bigger when considering interaction
- Idioms: scatter plots, line charts, bar charts , etc., or the linking between them
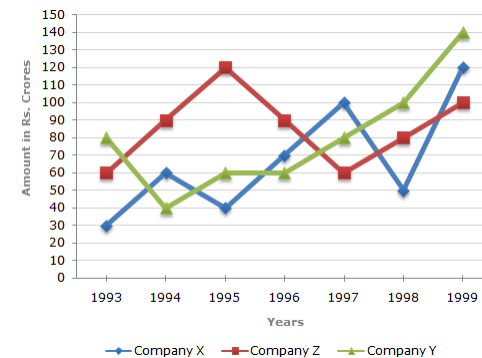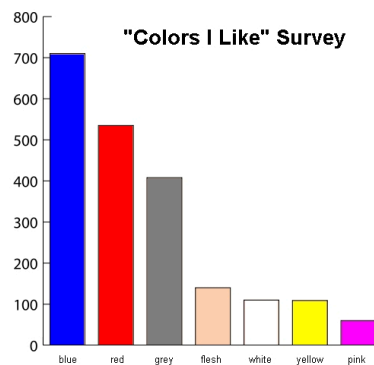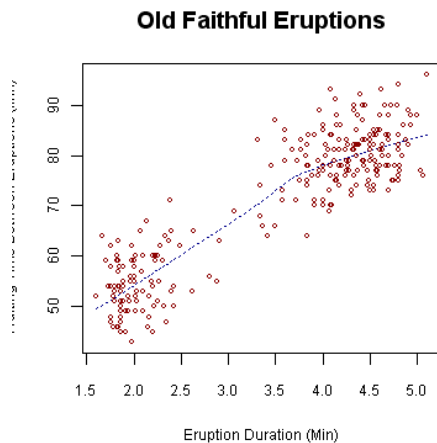
Why focus on tasks and effectiveness

- Tasks
  - a tool good for one tasks can be bad for another
  - Abstract tasks allow one to consider the similar or different needs across multiple fields
- Effectiveness
  - Leads to concerns of correctness, accuracy, and truth
  - A vast majority of design is ineffective for any special context
    - Optimize vs. Satisfy
  - Use a big consideration space, i.e., consider multiple alternatives

Why are most designs ineffective?

- Poor match with the properties of human perception and cognitive system
- Bad match for the intended task
- We should satisfy instead of optimizing
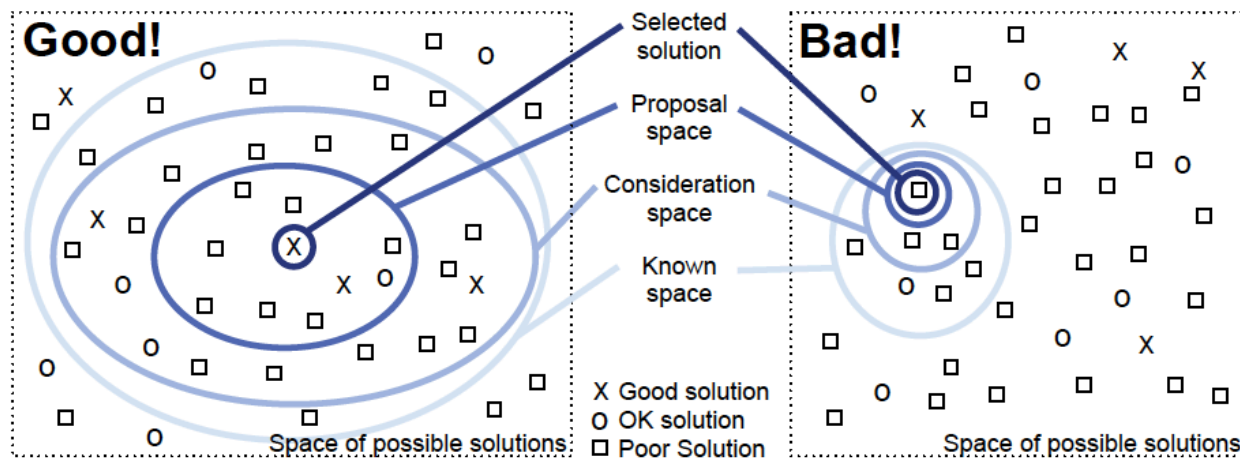
Why is validating difficult

- There are so many questions that you can ask
  - What does better mean?
  - What does effective mean?
  - How to measure insight/engagement?
  - Automatically or manually?
  - Who is the user?
  - What benchmark data and tasks to use?
  - How to measure the quality of images?
  - What is the algorithm's scalability wrt. data and image sizes?
  - ...

Why resource limitations

- Limitations: Computation, Human Cognition, Display
- Especially for large data sets (scalability issues)
- Computation: memory and compute time
- Human: memory and attention
- Display: run out of pixels
  - Information density

Why analyzing existing techniques

- Impose a structure on the enormous design space
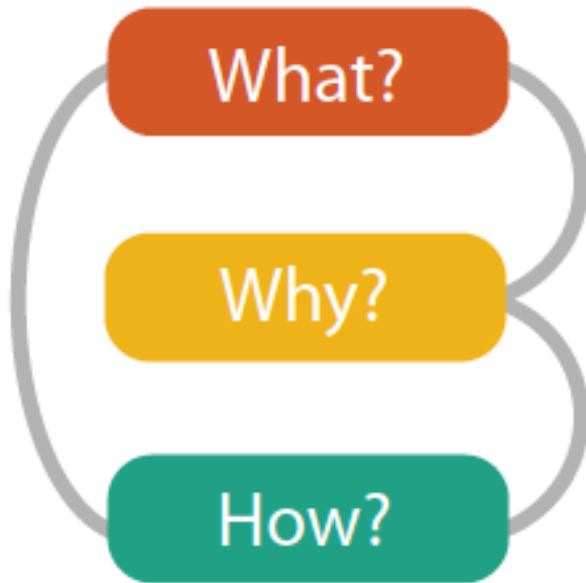
What data the user sees

Why the user intends to use vis tools

How the visual encoding and interaction idoms are constructed

Why analyzing existing techniques

- Complex vis tools are often a sequence of what-why-how instances chained together (input/output dependencies)