Visualization for Machine Learning

Fernanda Viégas @viegasf
Martin Wattenberg @wattenberg
Google Brain

Taller de Nous Usos de la Informàtica

Universitat de Barcelona

What is data visualization?

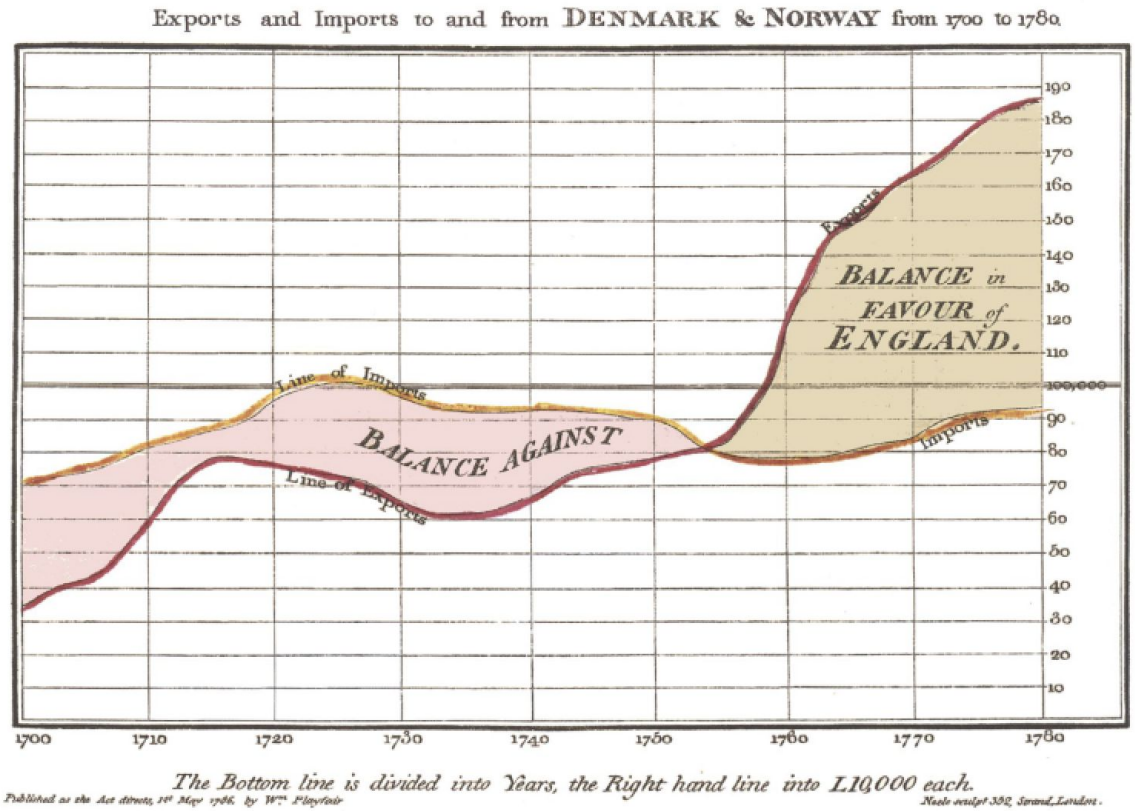
Transform data into visual encodings. What is it good for?

- Data exploration
- Scientific insight
- Communication
- Education

How to ensure it works well?

- Engage the visual system in smart ways
- Take advantage of pre-attentive processing

William Playfair (1786)

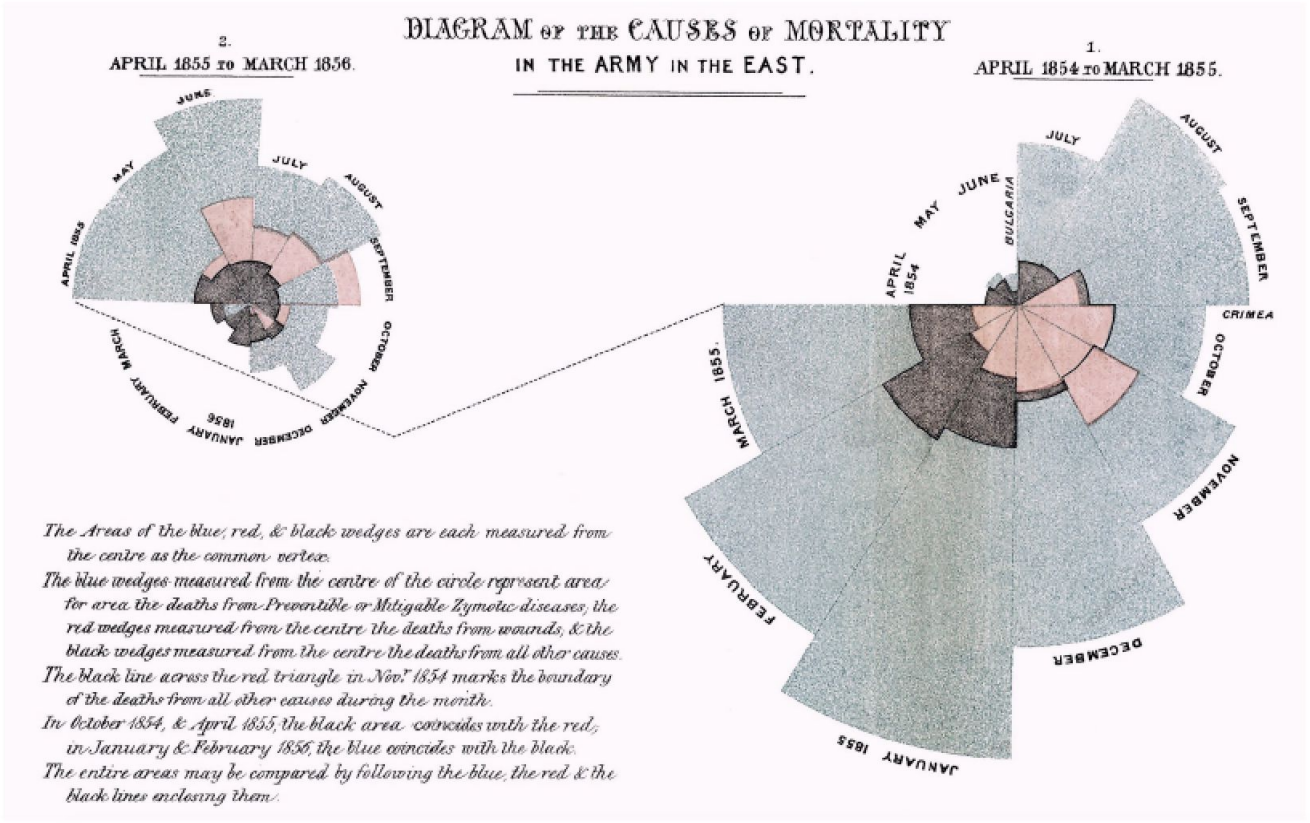


Line, bar, pie charts were all invented by the same person!

Aside from revolutionizing graphics, Playfair was an economist, engineer, and even a secret agent.

(Image: Wikipedia)

Florence Nightingale (1858)



These charts led to the adoption of better hygiene / sanitary practices in military medicine, saving millions of lives.

Arguably the most effective visualization ever!

This particular visualization technique would be frowned on today. Lesson: technique is less important than having the right data and right message.

(Image: Wikipedia)

What do they have in common?

Using special properties of the visual system to help us think.

Our visual system is like a GPU

- Incredibly good at a few special tasks
- With work, can be repurposed for more general situations

How do visualizations work?

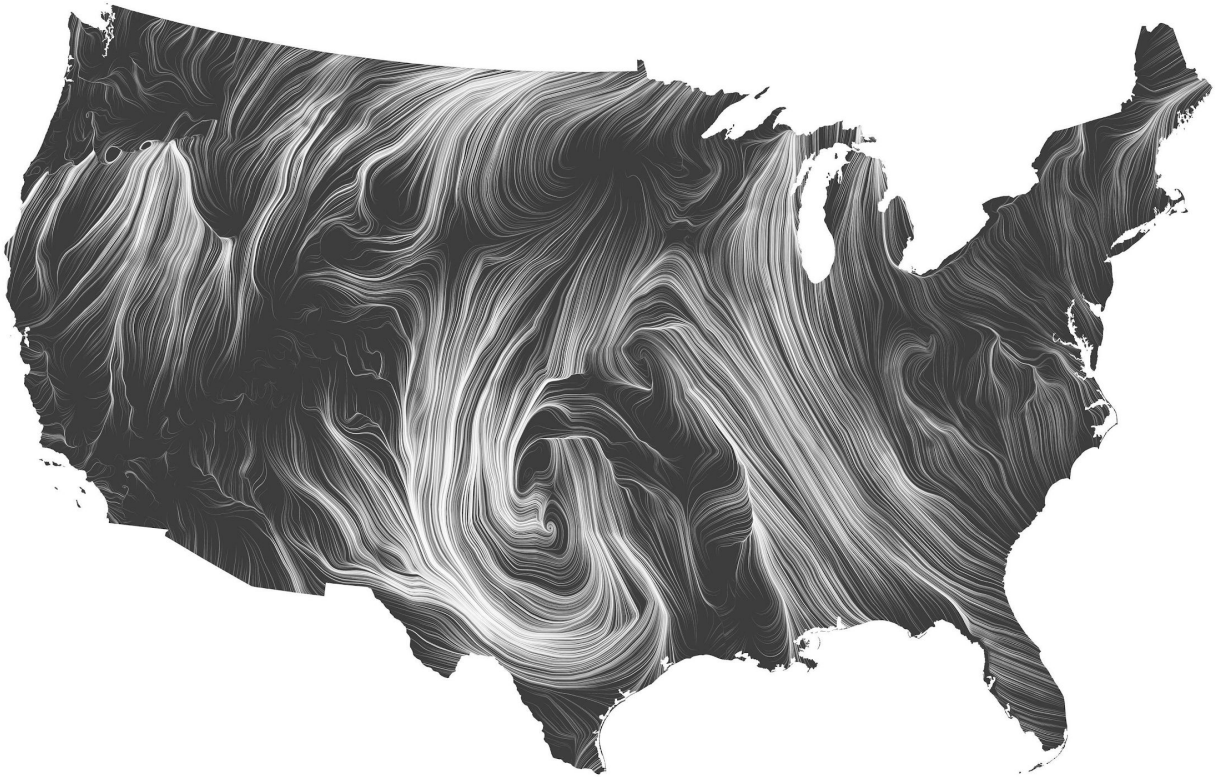
Find visual encodings that

- Guide viewer's attention
- Communicate data to the viewer
- Let viewer calculate with data

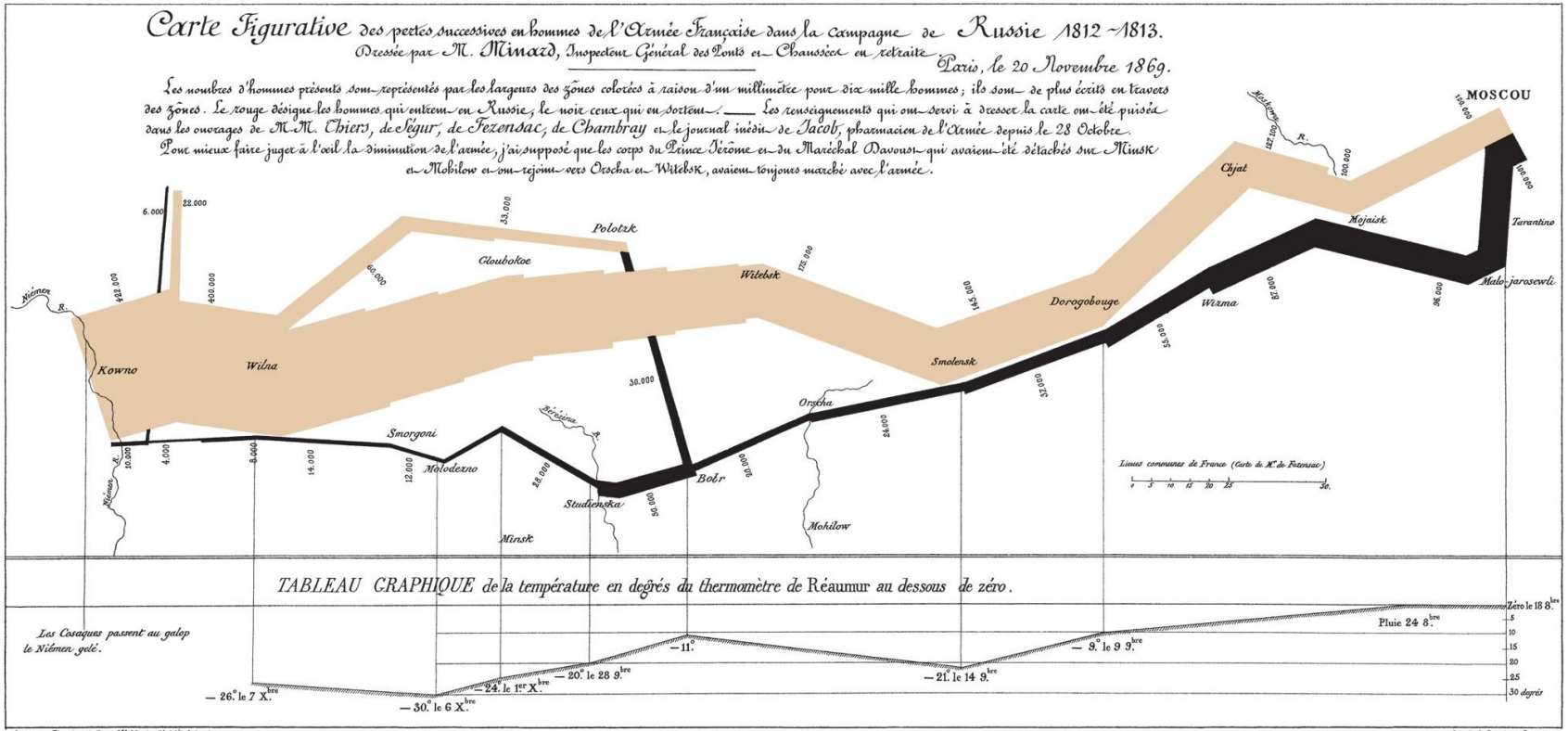
On computer

- Interactive exploration

Encodings: Examples



Encodings: Examples



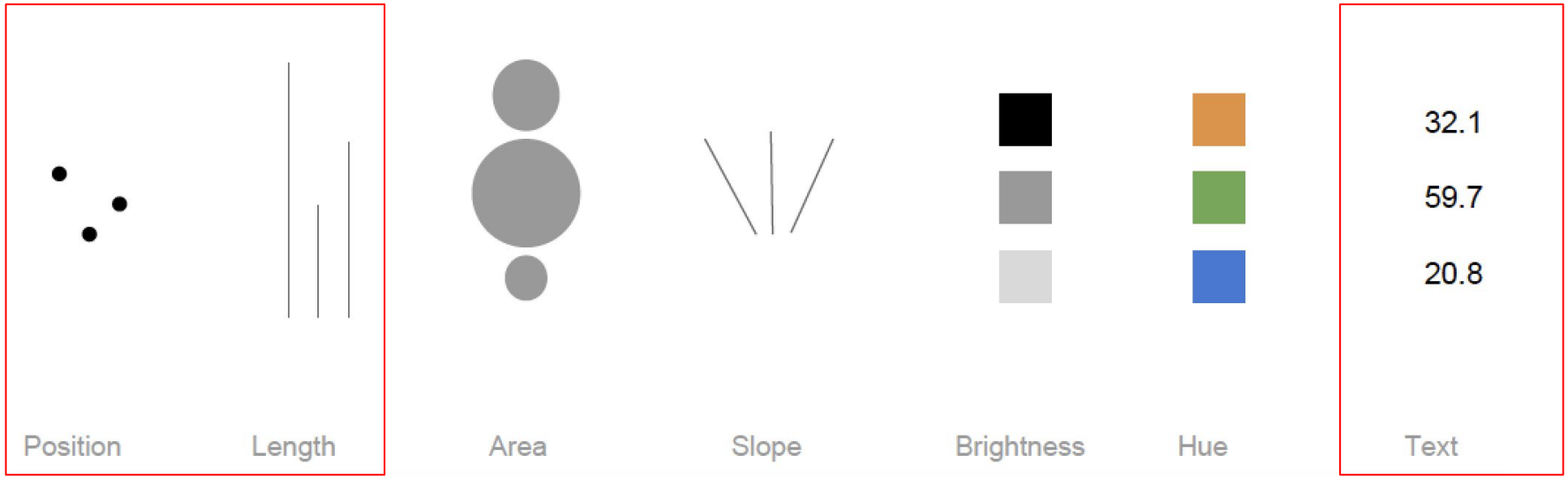
Encodings: Theory

From perceptual psychology: different encodings have different properties.

Encodings: Theory

From perceptual psychology: different encodings have different properties.

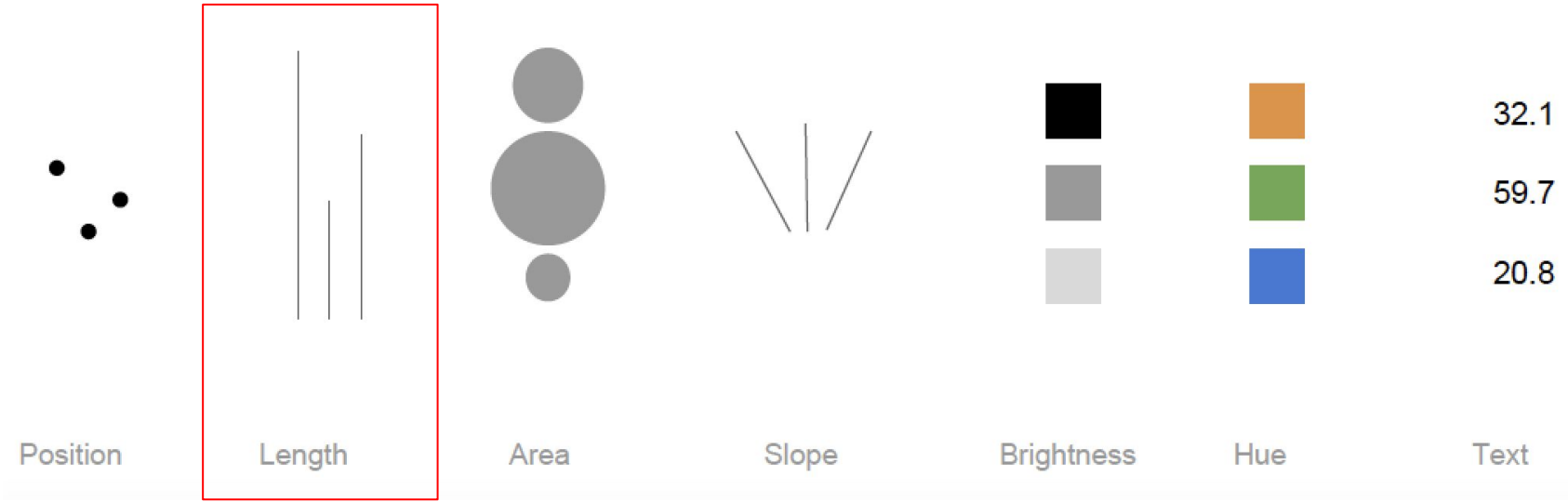
Communicating exact values:



Encodings: Theory

From perceptual psychology: different encodings have different properties.

Communicating ratios:

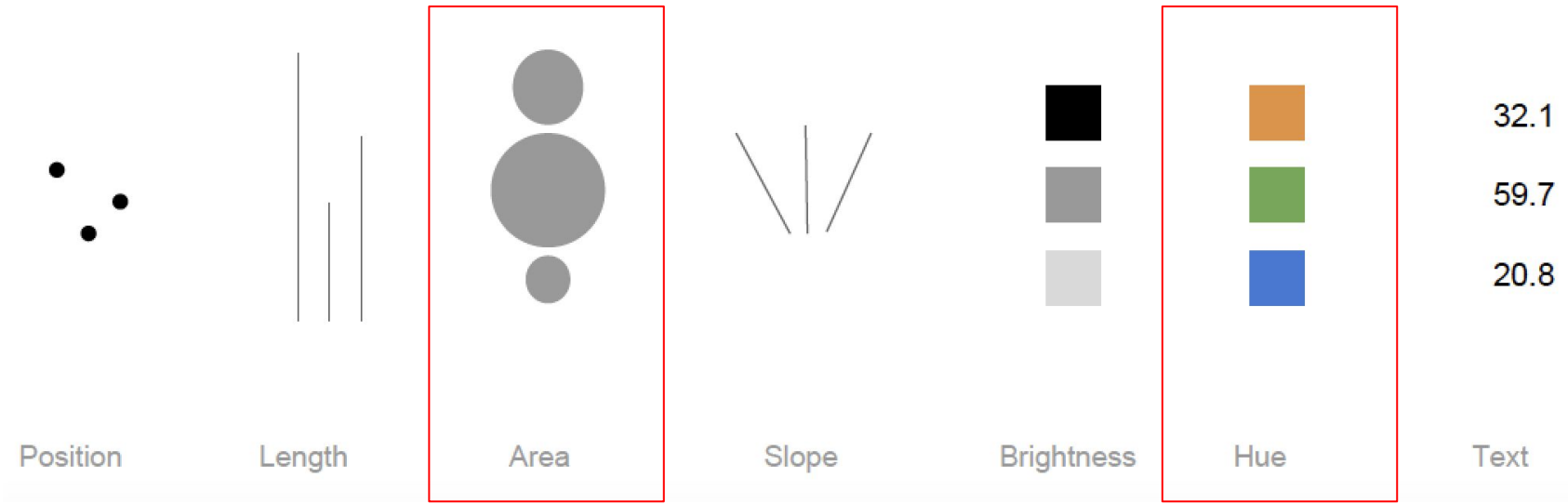


Encodings: Theory

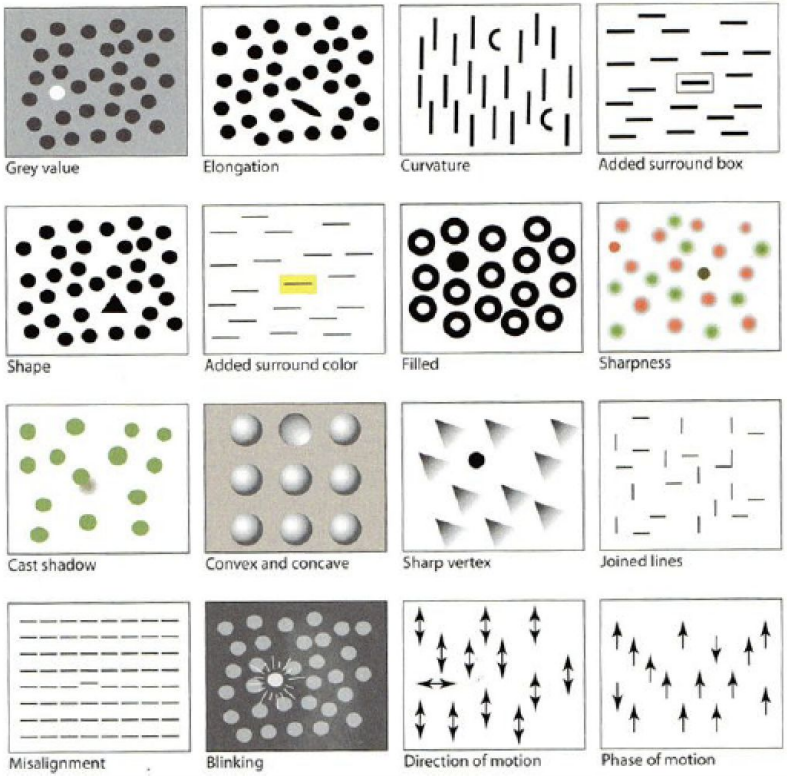
From perceptual psychology: different encodings have different properties.

Drawing attention:

When in doubt, use the "Color Brewer" site:
http://colorbrewer2.org



Encodings: Theory of attention



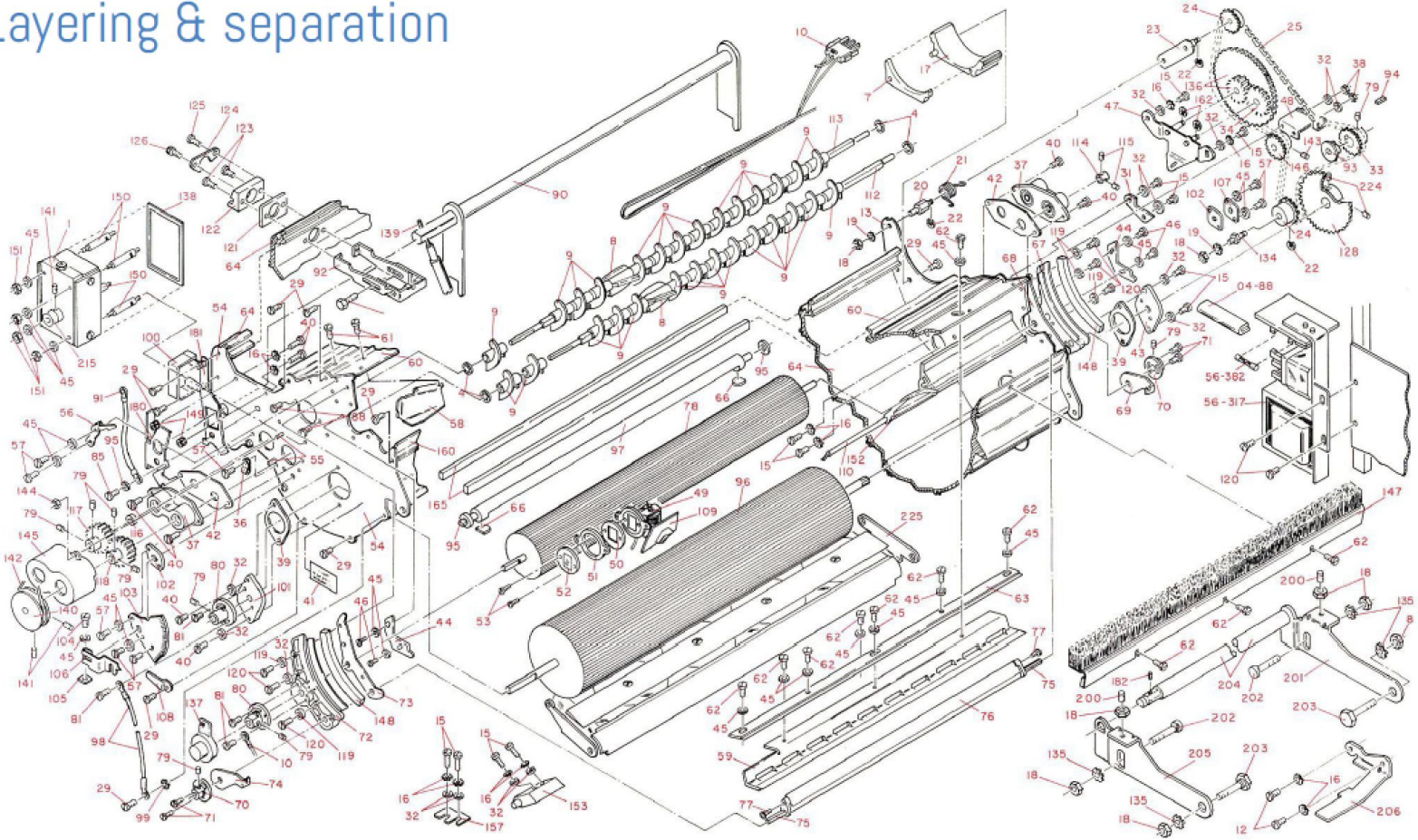
(Colin Ware, Visual Thinking for Design)

Pre-attentive processing / "popout"

Under the right circumstances, visual search can be parallel, rather than serial

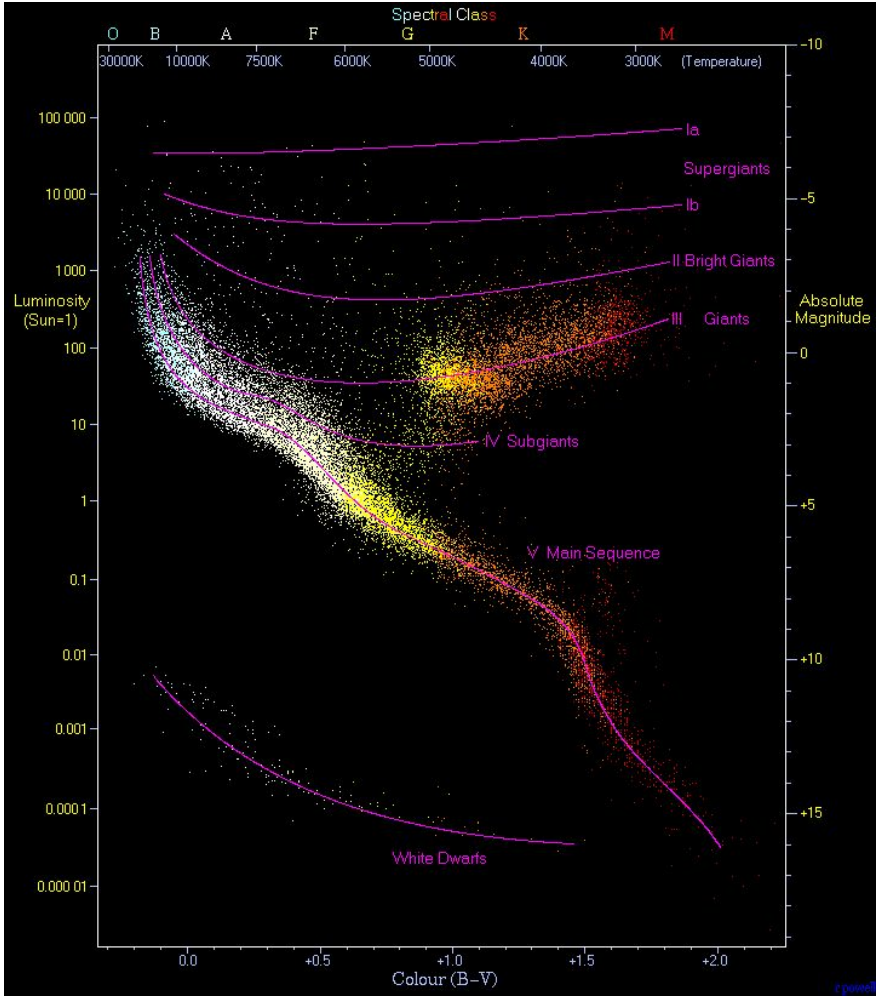
Time to find target does not increase as number of distractors increases

Layering & separation



after Tufte

Calculation



How do visualizations work - on computers?

Beyond static representations

- Interaction
- Conversation and collaboration

Shneiderman “mantra”:

- Overview first
- Zoom and filter
- Details on demand

Example: http://demographics.virginia.edu/DotMap/

Some common techniques: small multiples

Small multiples are a visualization concept introduced by Edward Tufte. He described them as:

"Illustrations of postage-stamp size are indexed by category or a label, sequenced over time like the frames of a movie, or ordered by a quantitative variable not used in the single image itself."

In other words, small multiples use the same basic graphic or chart to display difference slices of a data set.

Examples: Droughts footprint.

Some common techniques: data faceting.

Visualize data and slice it in all sorts of manners, which can help us begin to see how our dataset is laid out.

Examples: Data faceting.

Other Examples

https://youtu.be/jbkSRLYSojo

https://youtu.be/ezVk1ahRF78

http://johnguerra.co/viz/berkeleyFinalProjectsSummer2018/

http://rhythm-of-food.net/

http://roadstorome.moovellab.com/countries

https://guns.periscopic.com/?year=2013

http://chriswhong.github.io/nyctaxi/