



INTRODUCTION TO DATA VISUALIZATION

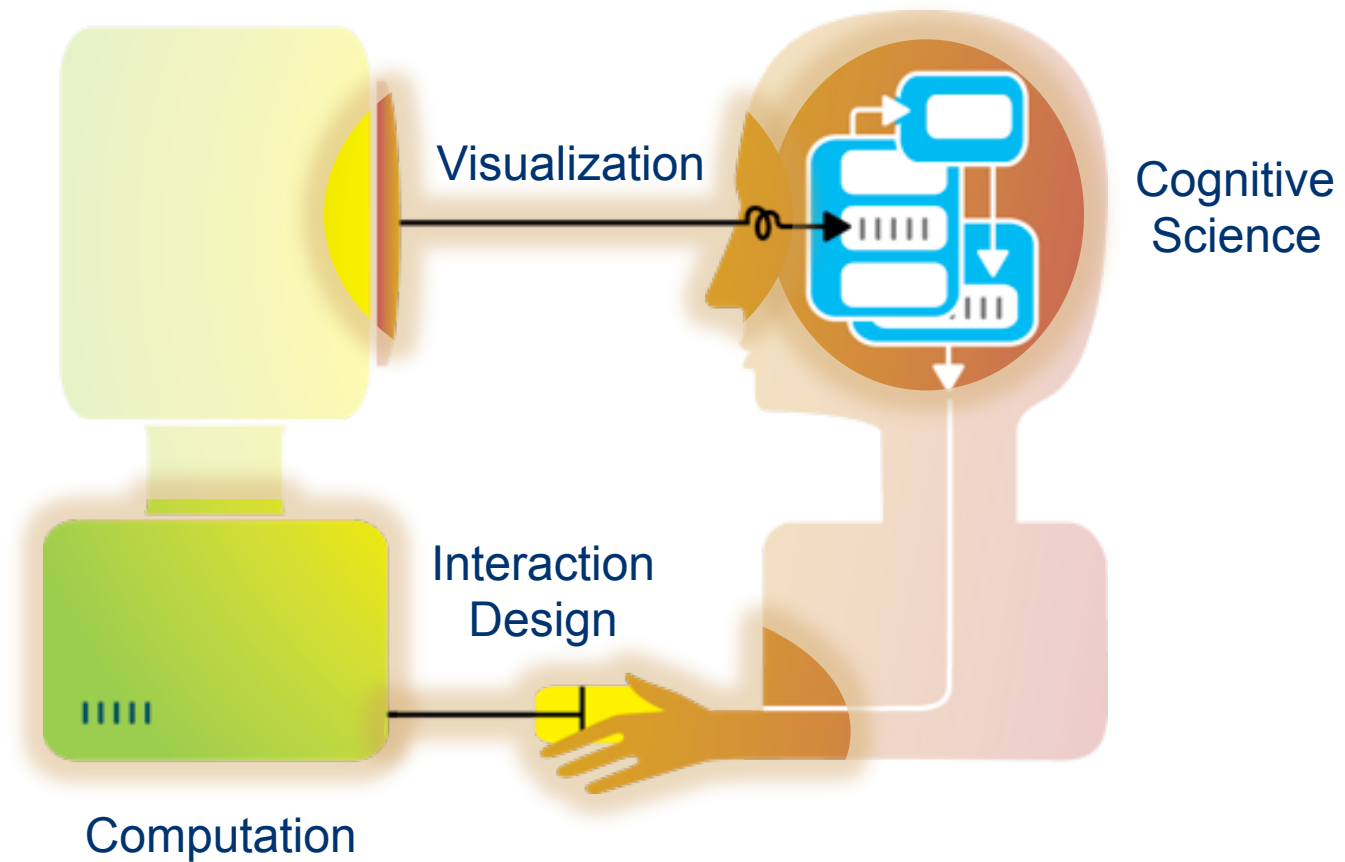
R. Jordan Crouser

Introductions & Background

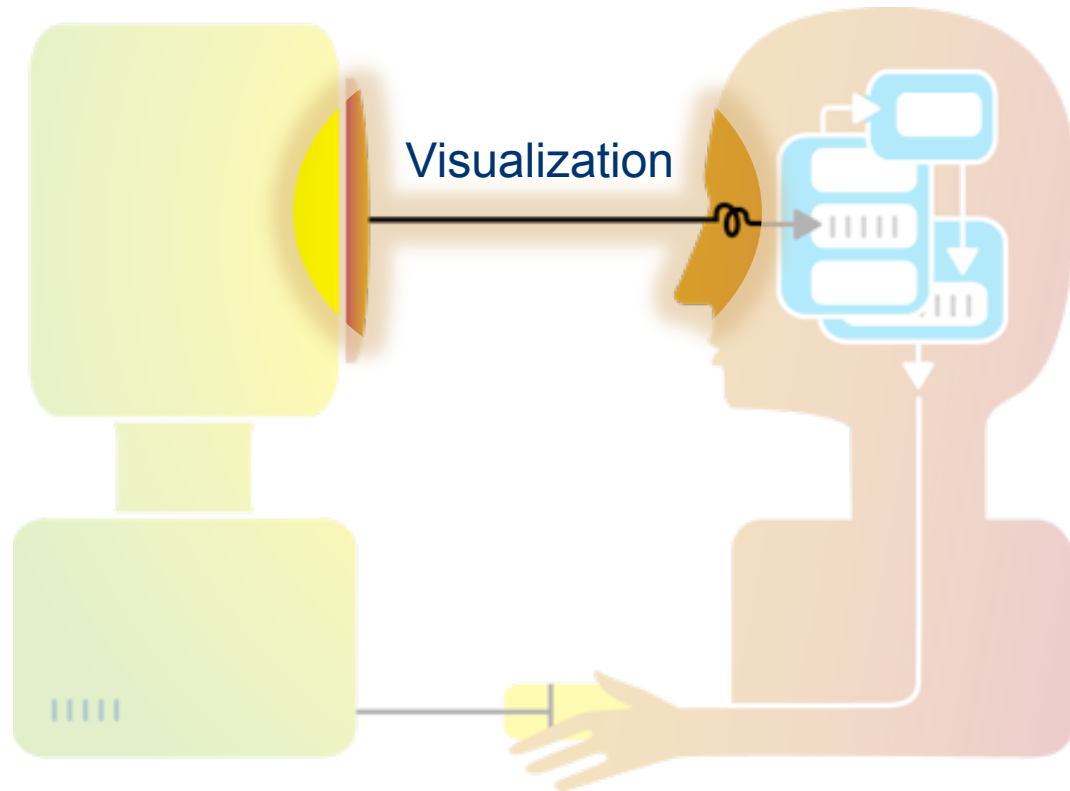


Jordan
(data scientist)

My research (generally)



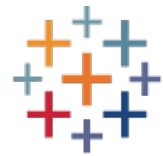
About this lecture / demo



Tons of useful tools...



Crossfilter



+ a b l e a u

Fast Multidimensional Filtering for Coordinated Views



visual.ly



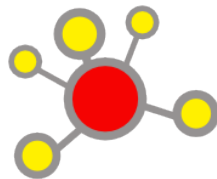
Data-Driven Documents



node



Gephi



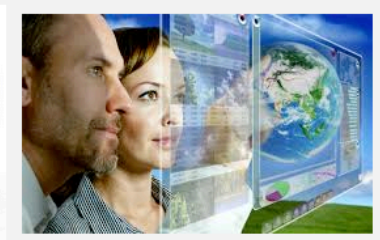
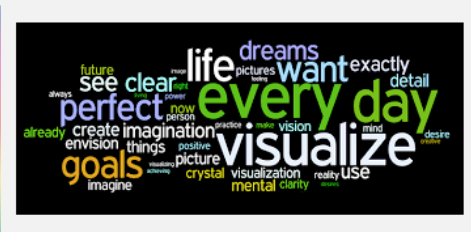
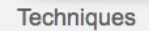
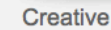
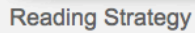
Leaflet



plotly

Google visualization   Jordan   

SafeSearch



Part 1: what is visualization??

Google information visualization

Jordan

All News **Images** Books Videos More Search tools

SafeSearch

Examples Design Tools 3D Data Visualization Map Data

Examples: A collection of various data visualizations, including a large circular network diagram, a small circular diagram, and a large circular diagram with many nodes and edges.

Design: A collection of various data visualizations, including a large circular network diagram, a small circular diagram, and a large circular diagram with many nodes and edges.

Tools: A collection of various data visualizations, including a large circular network diagram, a small circular diagram, and a large circular diagram with many nodes and edges.

3D: A collection of various data visualizations, including a large circular network diagram, a small circular diagram, and a large circular diagram with many nodes and edges.

Data Visualization Map: A collection of various data visualizations, including a large circular network diagram, a small circular diagram, and a large circular diagram with many nodes and edges.

Data: A collection of various data visualizations, including a large circular network diagram, a small circular diagram, and a large circular diagram with many nodes and edges.

ALL BLUES: A collection of various data visualizations, including a large circular network diagram, a small circular diagram, and a large circular diagram with many nodes and edges.

La natura della ricerca in design: A collection of various data visualizations, including a large circular network diagram, a small circular diagram, and a large circular diagram with many nodes and edges.

The World of Music: A collection of various data visualizations, including a large circular network diagram, a small circular diagram, and a large circular diagram with many nodes and edges.

INFORMATION VISUALIZATION AND COMPUTATIONAL DESIGN: A collection of various data visualizations, including a large circular network diagram, a small circular diagram, and a large circular diagram with many nodes and edges.

MSN: A collection of various data visualizations, including a large circular network diagram, a small circular diagram, and a large circular diagram with many nodes and edges.

2008: A collection of various data visualizations, including a large circular network diagram, a small circular diagram, and a large circular diagram with many nodes and edges.

Cosmopolitan: A collection of various data visualizations, including a large circular network diagram, a small circular diagram, and a large circular diagram with many nodes and edges.

Perhaps a more helpful question

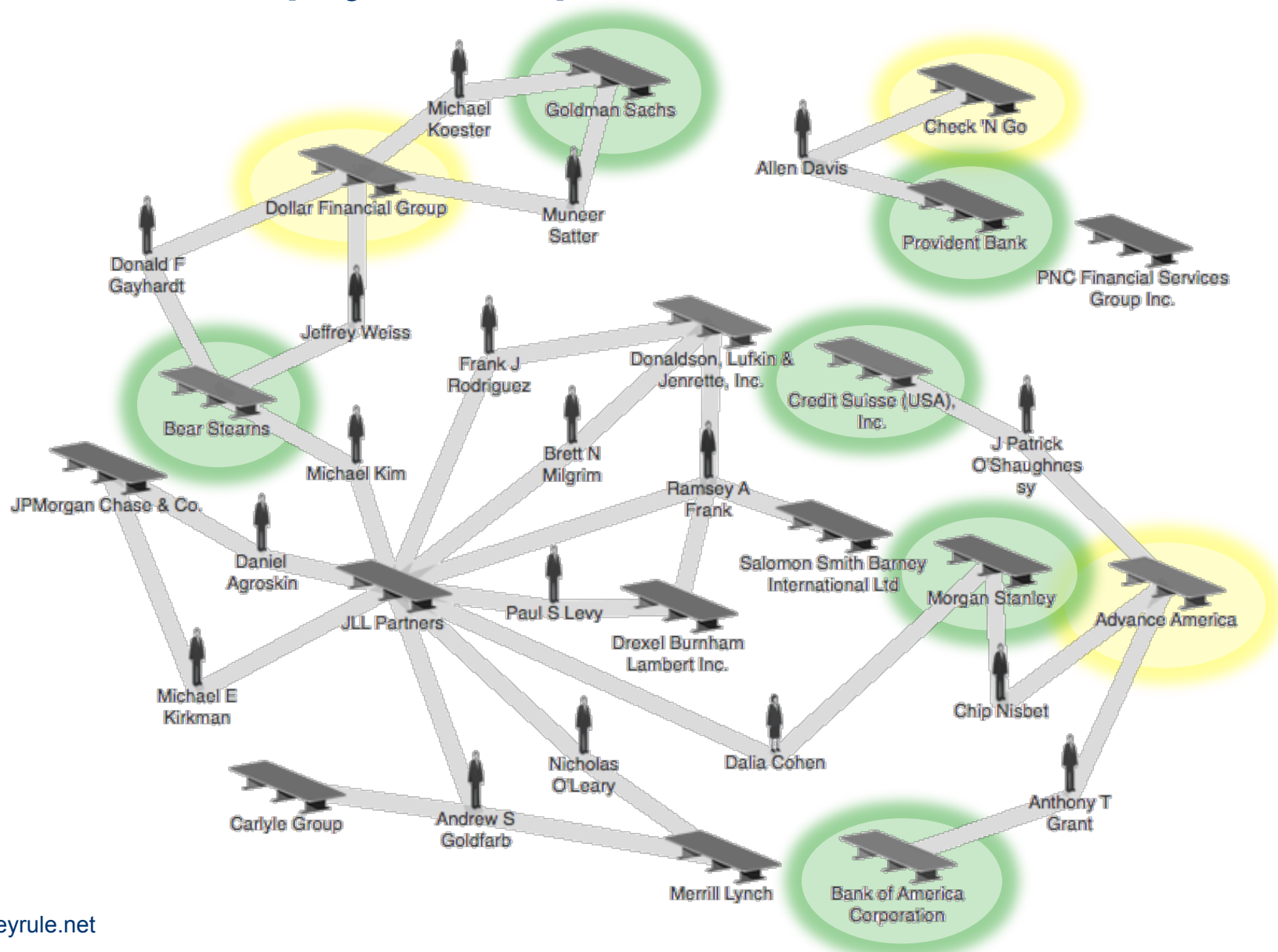
What are some ways
a “visualization” can be **useful**?

Does it help you spot trends?



More info here: http://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak

Does it help you explore?

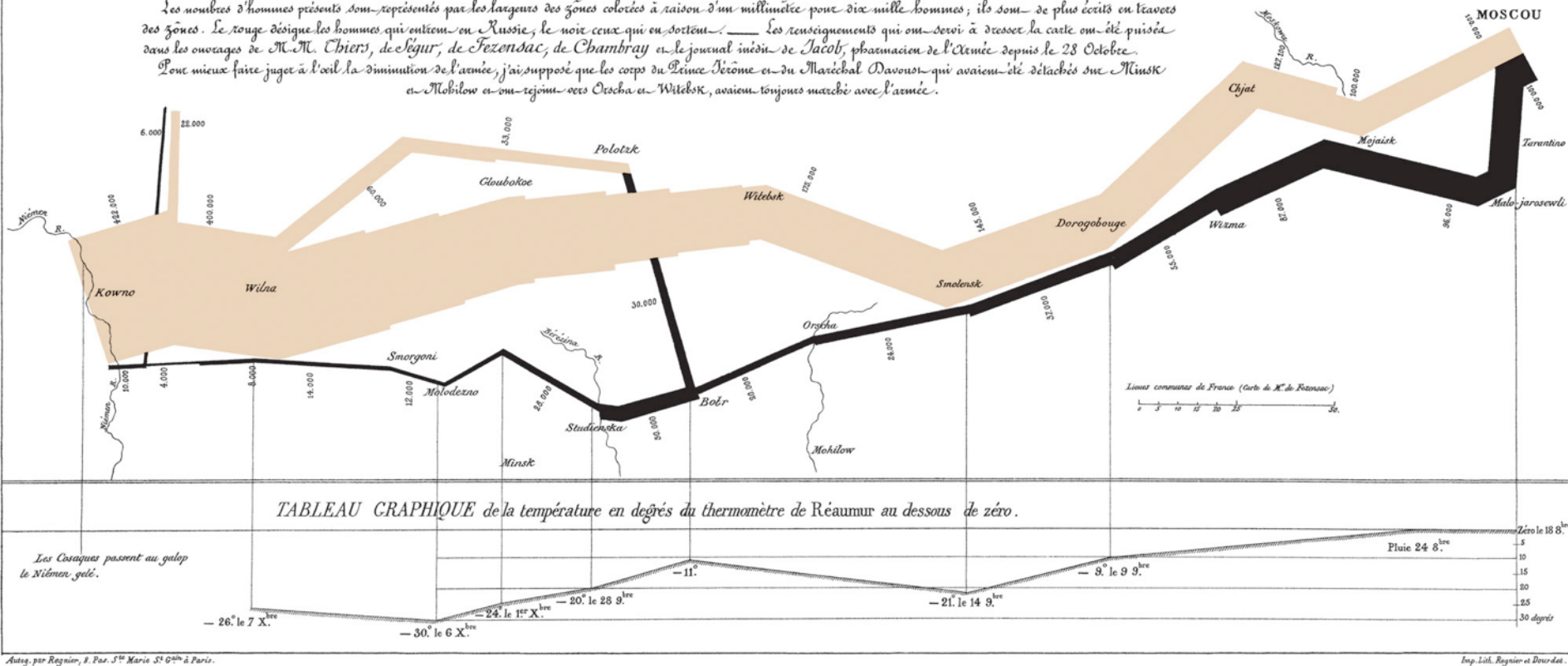


Does it tell a story?

Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en traits des zones. Le rouge désigne les hommes qui ont été en Russie; le noir ceux qui en sont sortis. Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Thiers, de Ségur, de Fozendac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre. Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk et Mohilew et qui rejoignent vers Otscha et Witebsk, avaient toujours marché avec l'armée.



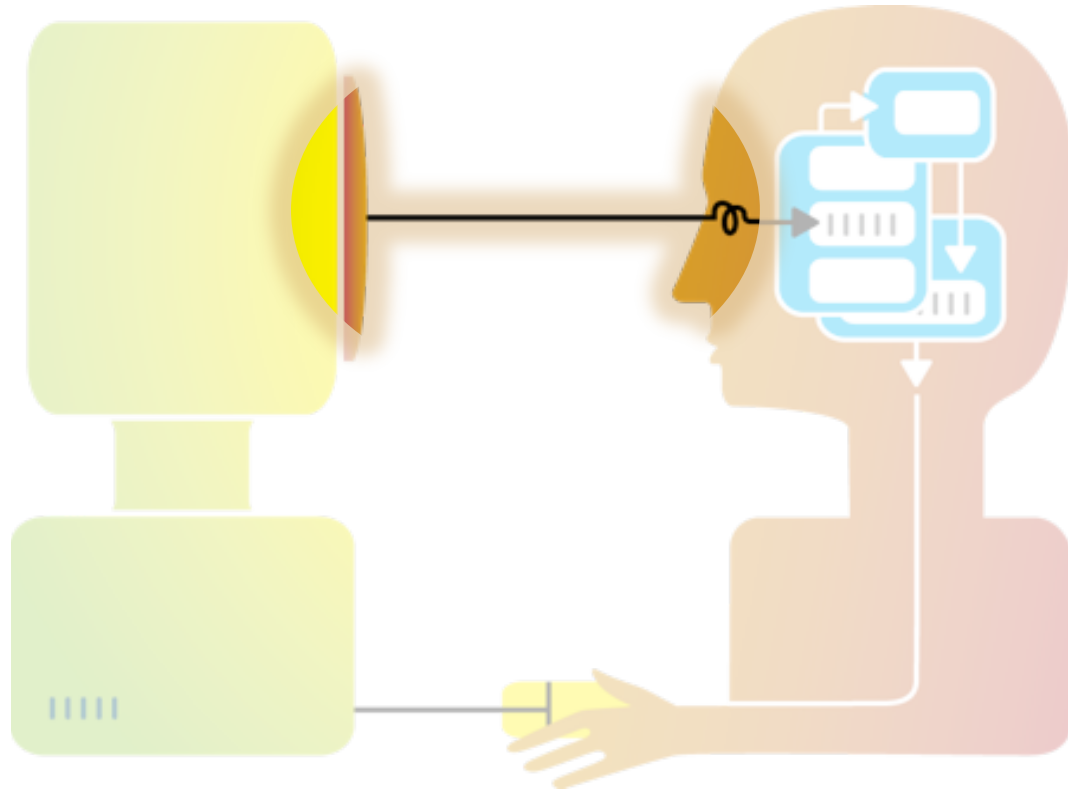
"Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812-1813."
 Drawn by Mr. Charles Minard, Inspector General of Bridges and Roads in retirement. Paris, 20 November 1869.

Visualization (def.)

Visual
representations
of data that
reinforce human
cognition

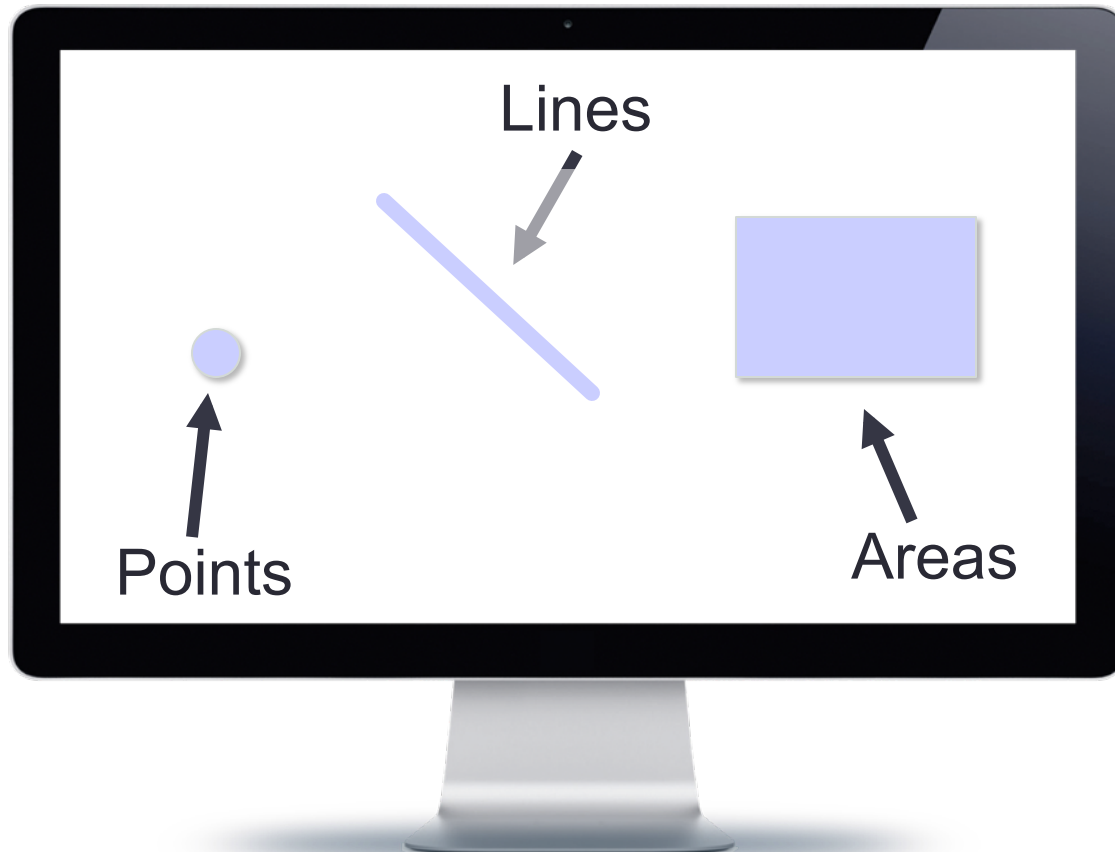


Part 2: how do we **build** them?



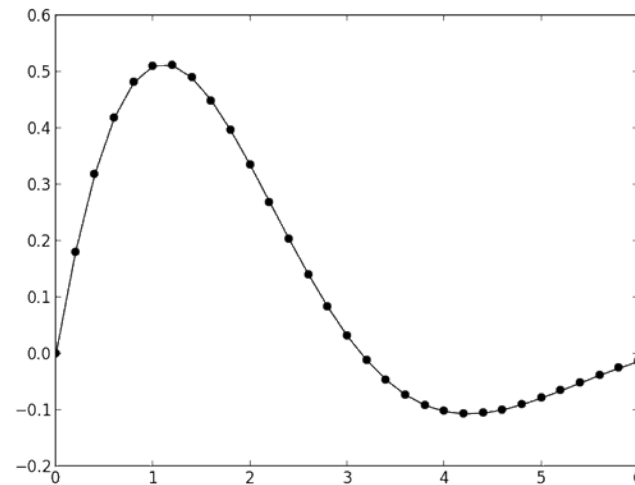
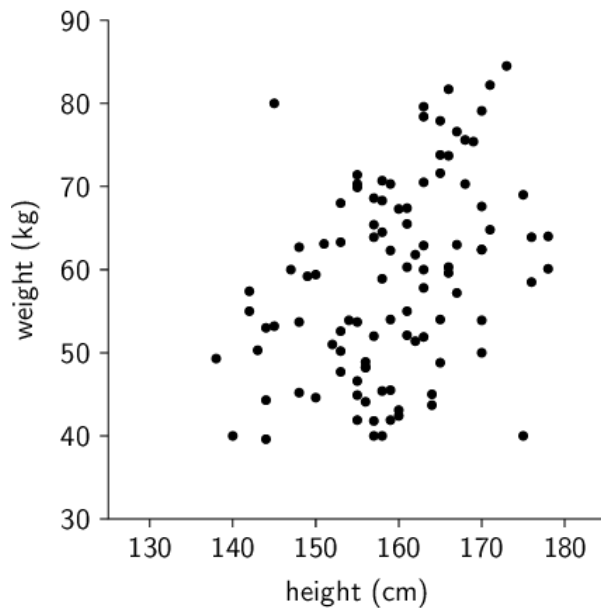
Graphical primitives

The images we draw are composed of marks: like ink



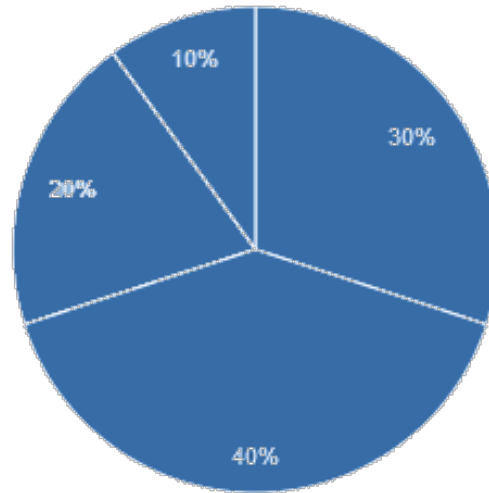
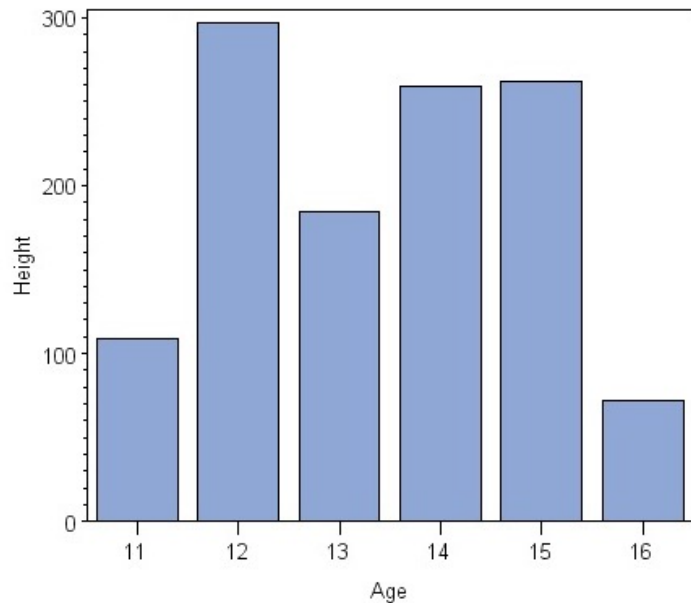
Visual dimension: position

- Encode information using **where** the mark is drawn
- Some examples:



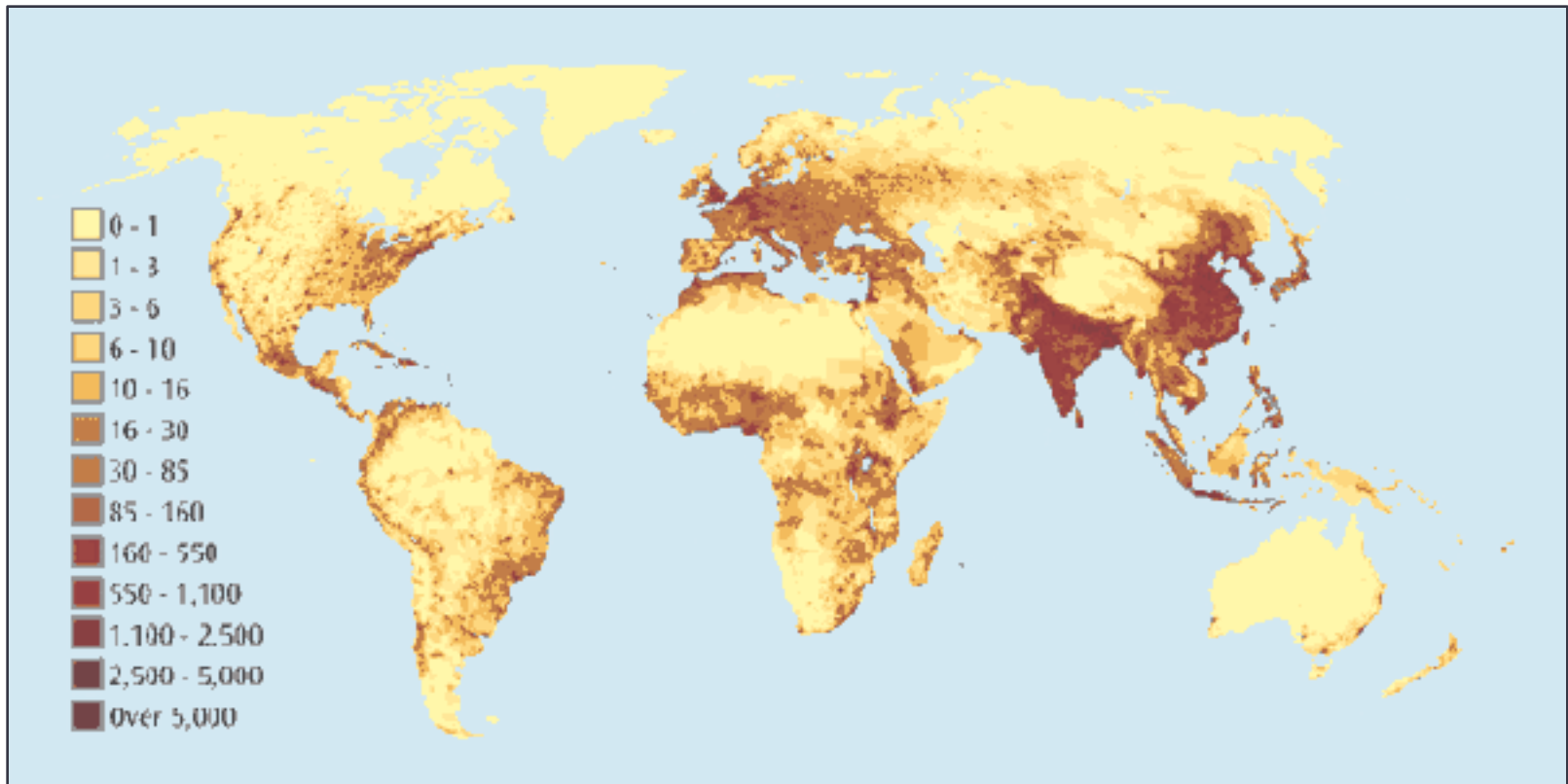
Visual dimension: size

- Encode information using **how big** the mark is drawn
- Examples:



Visual dimension: value

- Encode information using **how dark** the mark is drawn
- Example:



Visual dimension: color

- Encode information using the **hue** of the mark
- Examples:

Benefits

About 1 out of 10 women improved their symptoms using this medicine.



Side Effects

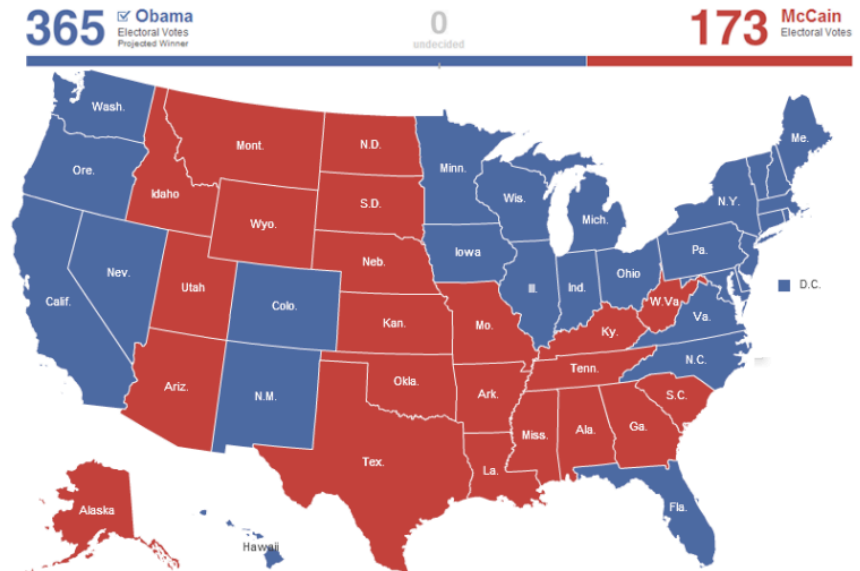
About 2 out of 10 women had dry mouth using this medicine.



About 1 out of 10 women had constipation using this medicine.

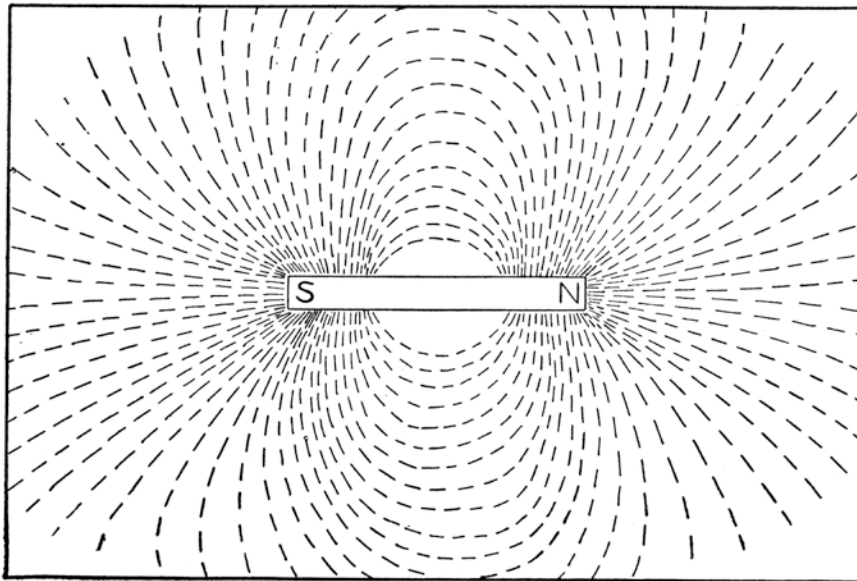


Less than 1 out of 10 women had an upset stomach using this medicine.



Visual dimension: orientation

- Encode information using how the mark is **rotated**
- Examples:

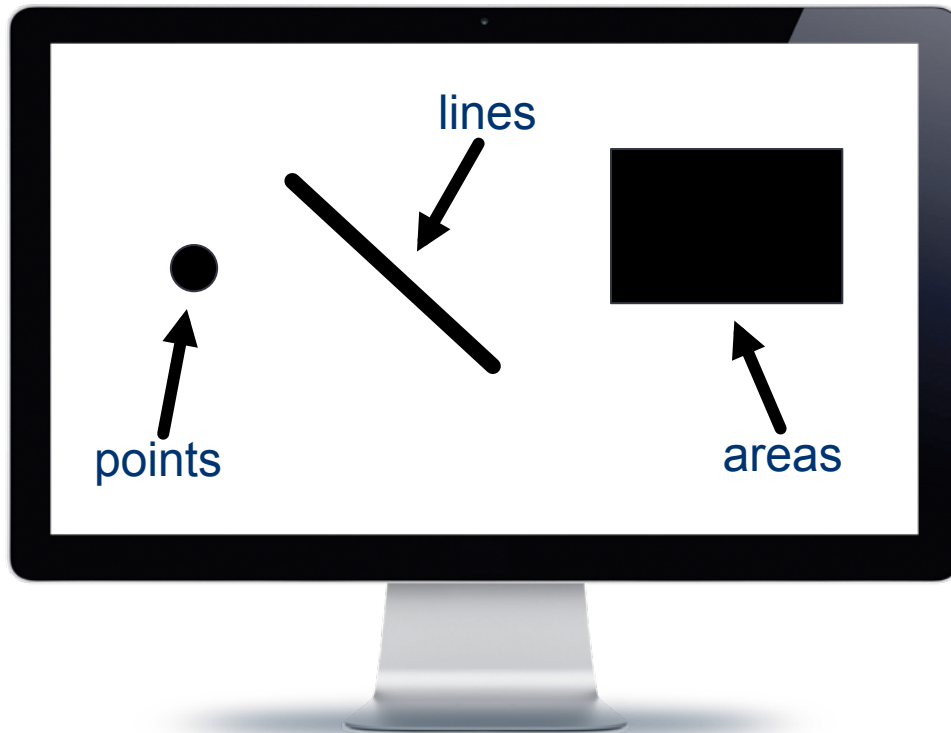


Visual dimension: shape

- Encode information using how the mark is **shaped**
- Examples:



Quick recap: what we have to work with



graphical primitives

dimensions

| | |
|--|-------------|
| | Position |
| | Size |
| | Value |
| | Color |
| | Orientation |
| | Shape |

Part 3: data visualization in python



Part 3: data visualization in python

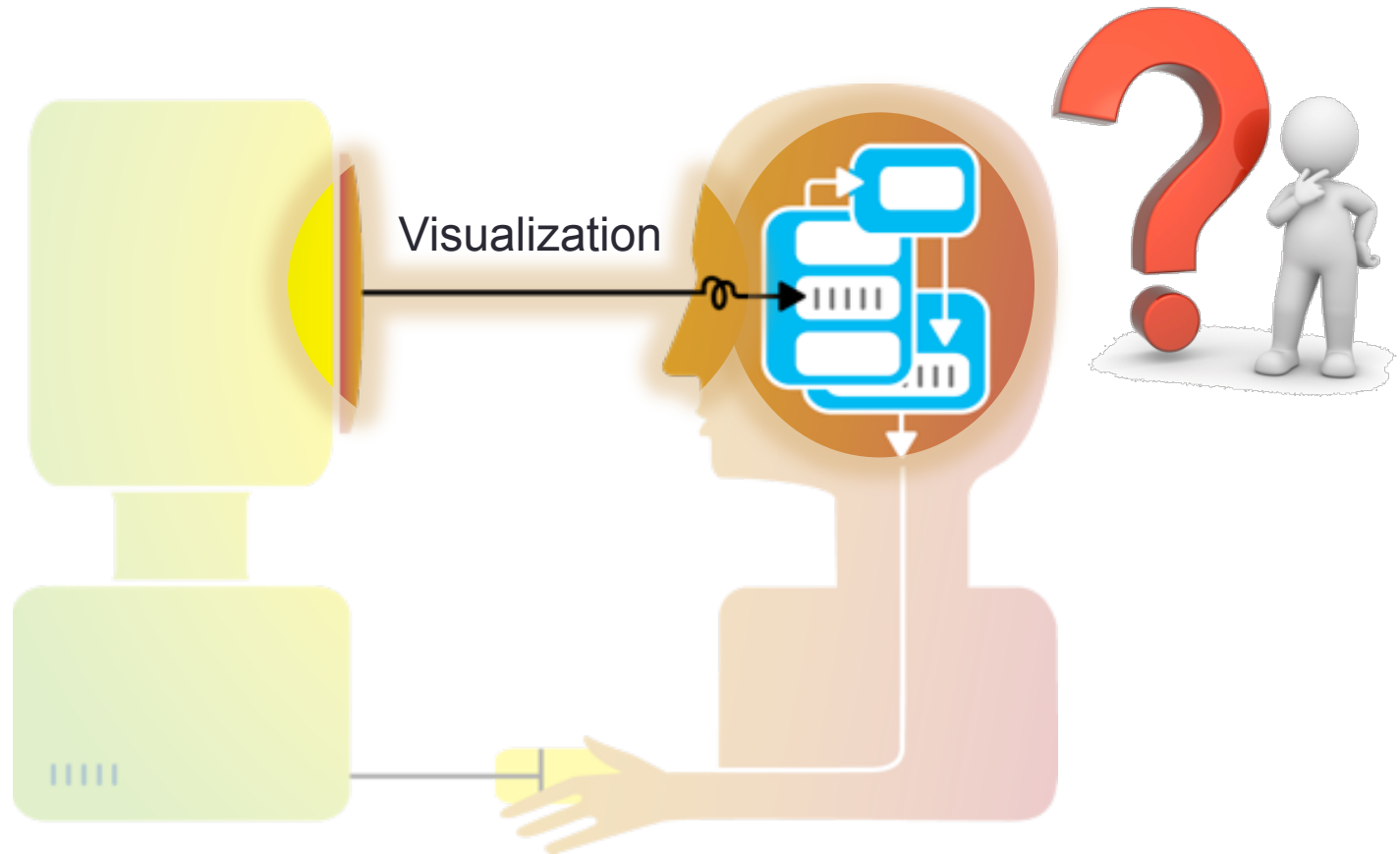


Bottom line: it's just like using any other functions!

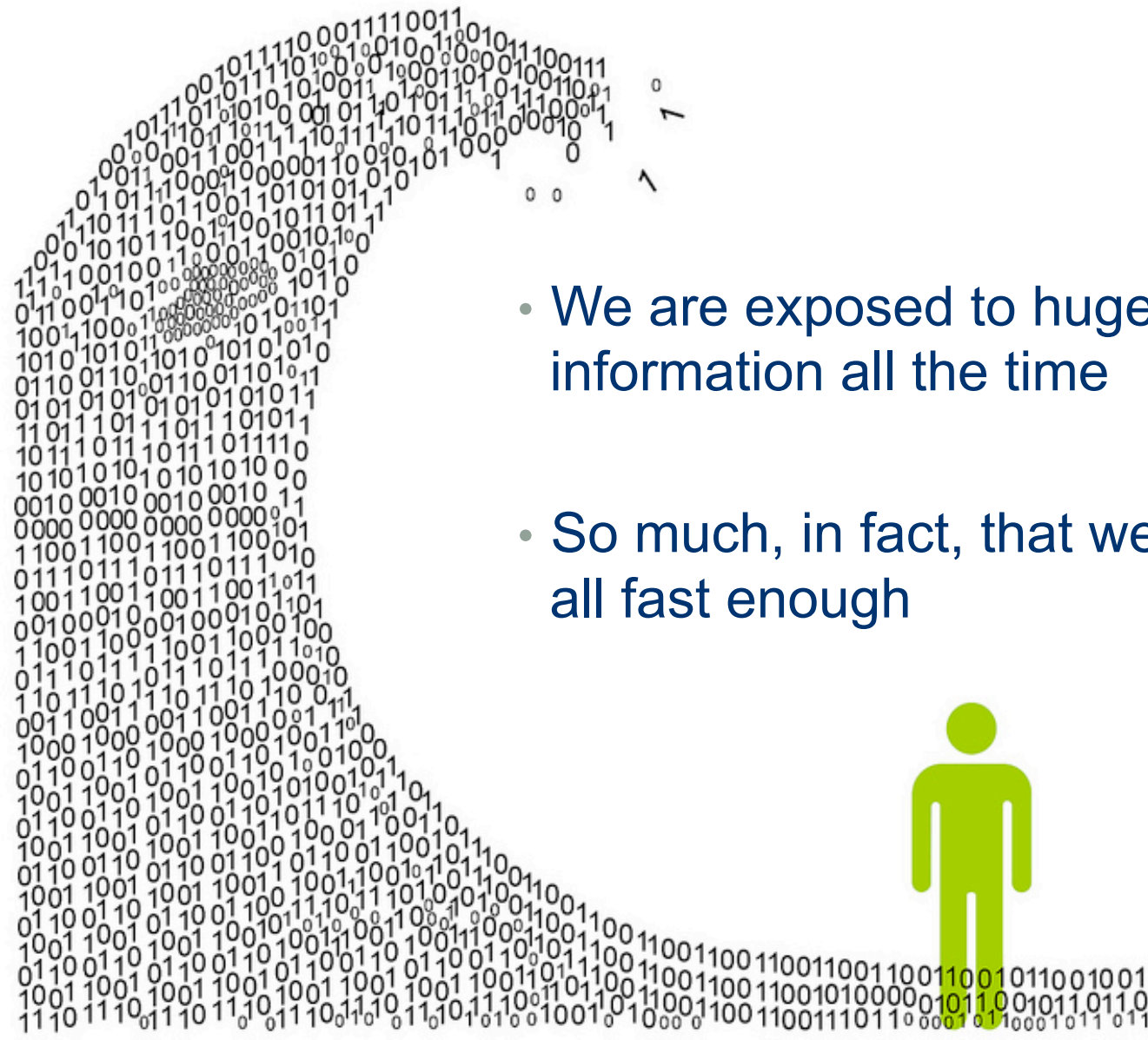
Part 3: why **visualization** works (or not)



Visualization helps shape *mental models*



Information overload



- We are exposed to huge amounts of information all the time
- So much, in fact, that we can't process it all fast enough



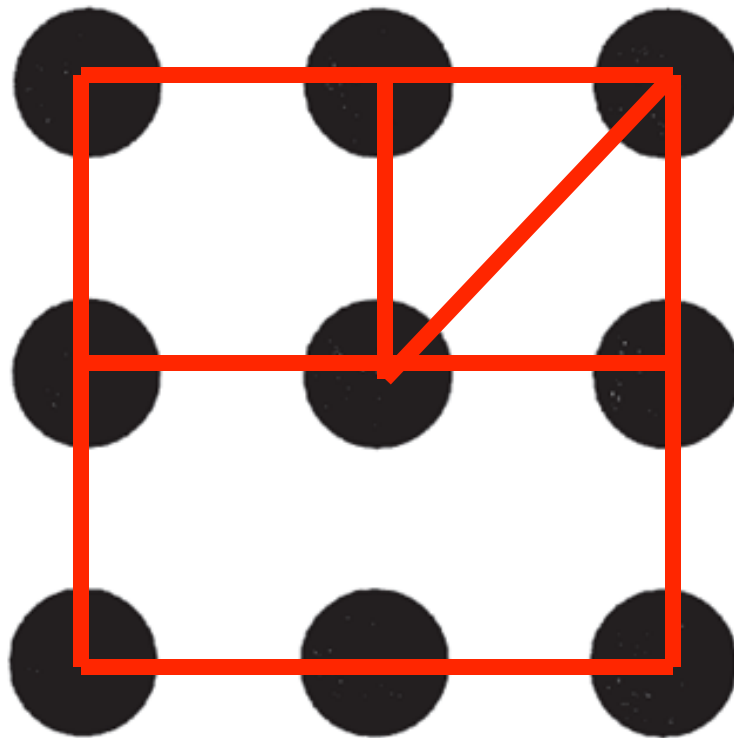
Mental models

To cope, we construct **mental models**:
abstracted, simplified versions of the world
that are more manageable



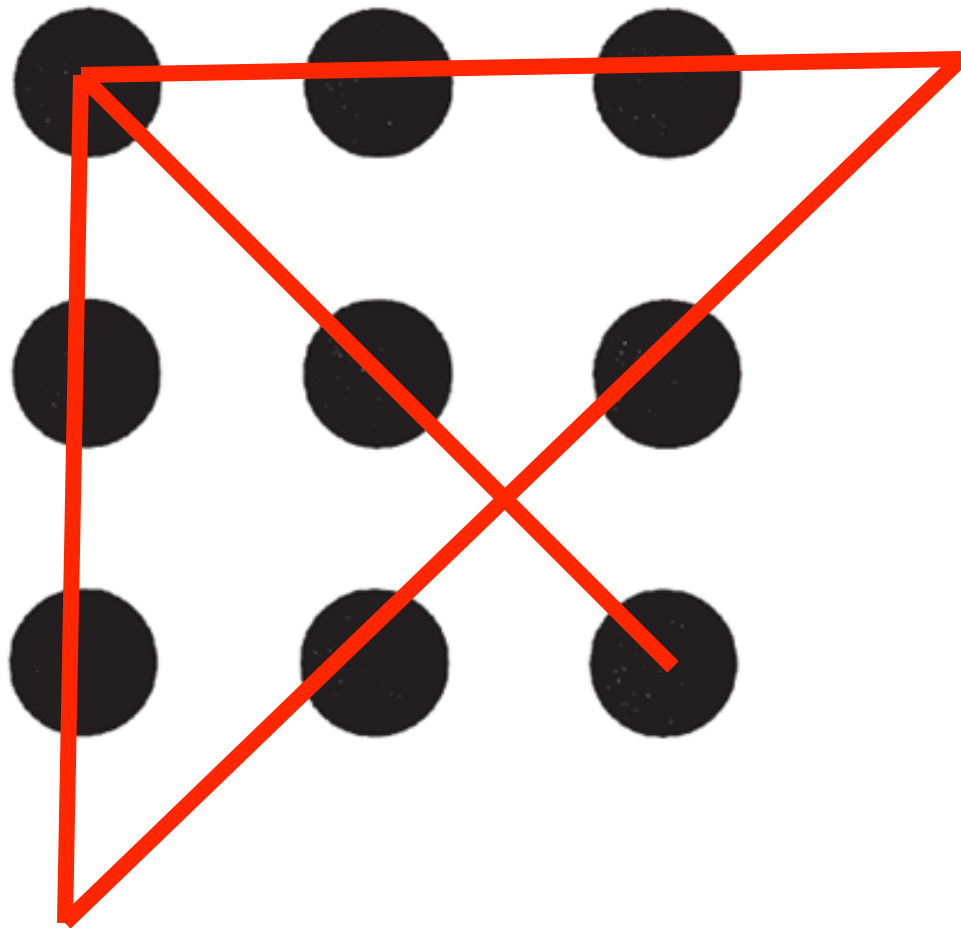
The 9-dot problem

Task 1: Connect all 9 dots using only straight lines



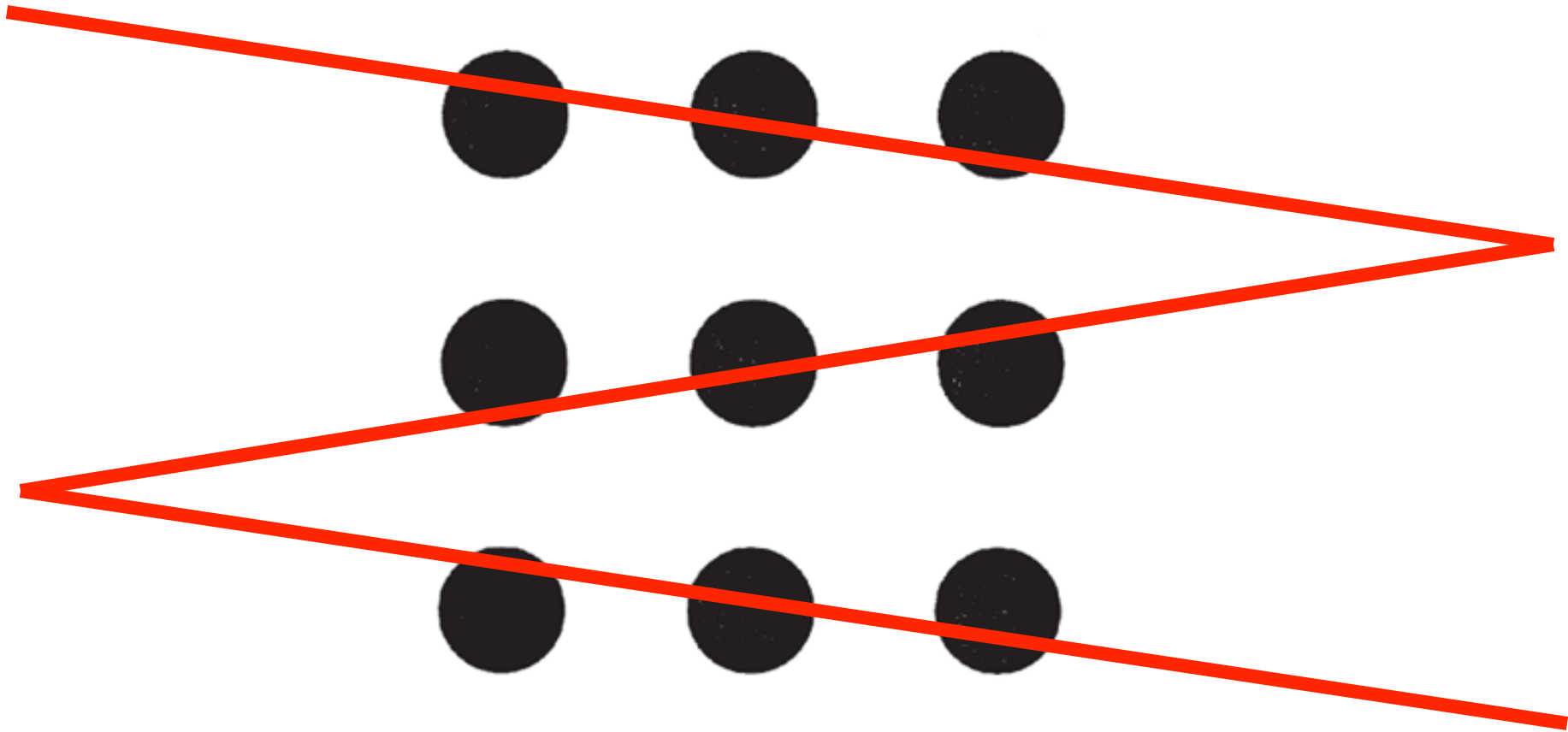
The 9-dot Problem

Task 2: Connect all 9 dots using 4 straight lines



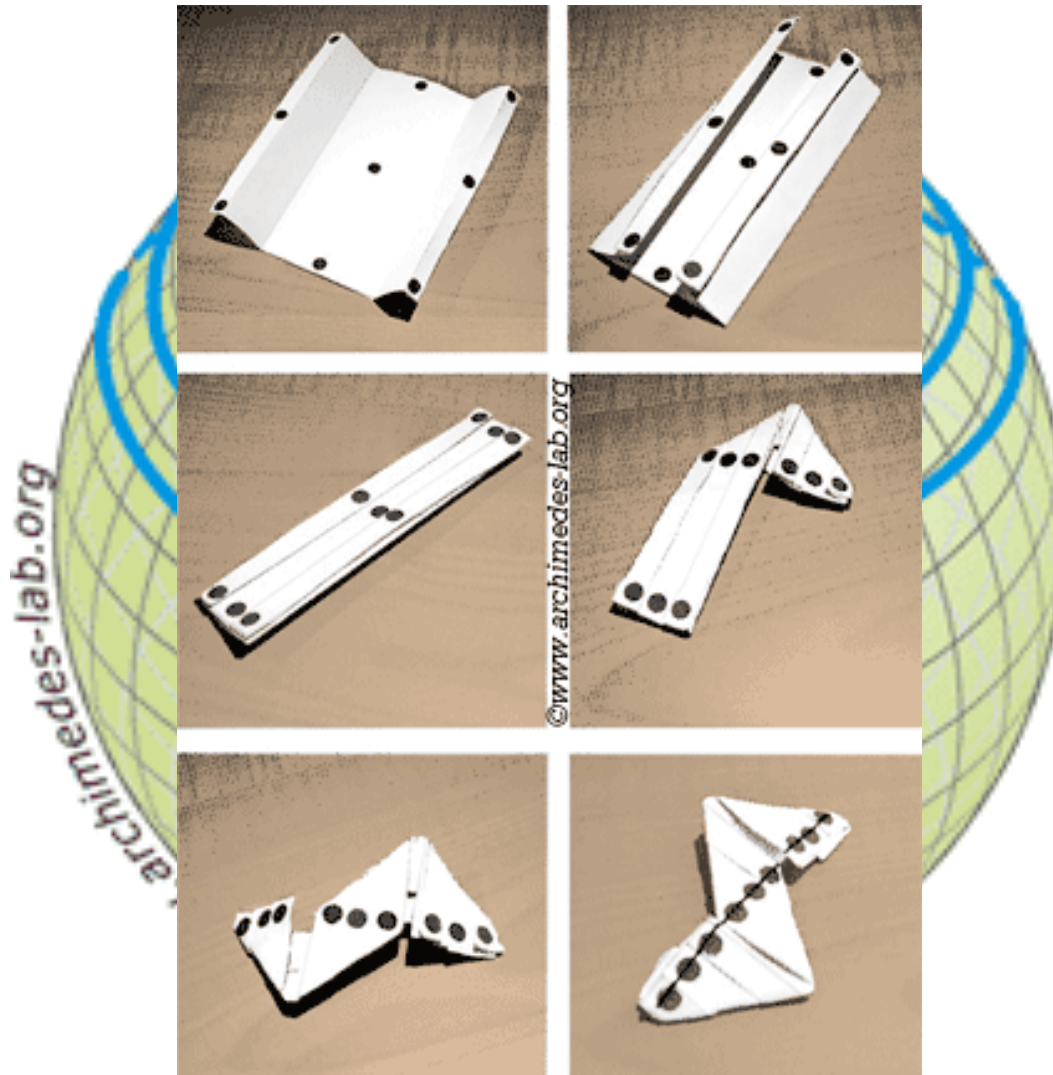
The 9-dot Problem

Task 3: Connect all 9 dots using 3 straight lines



The 9-dot Problem

Task 4: Connect all 9 dots using 1 straight line



Mental Models: a Sketch



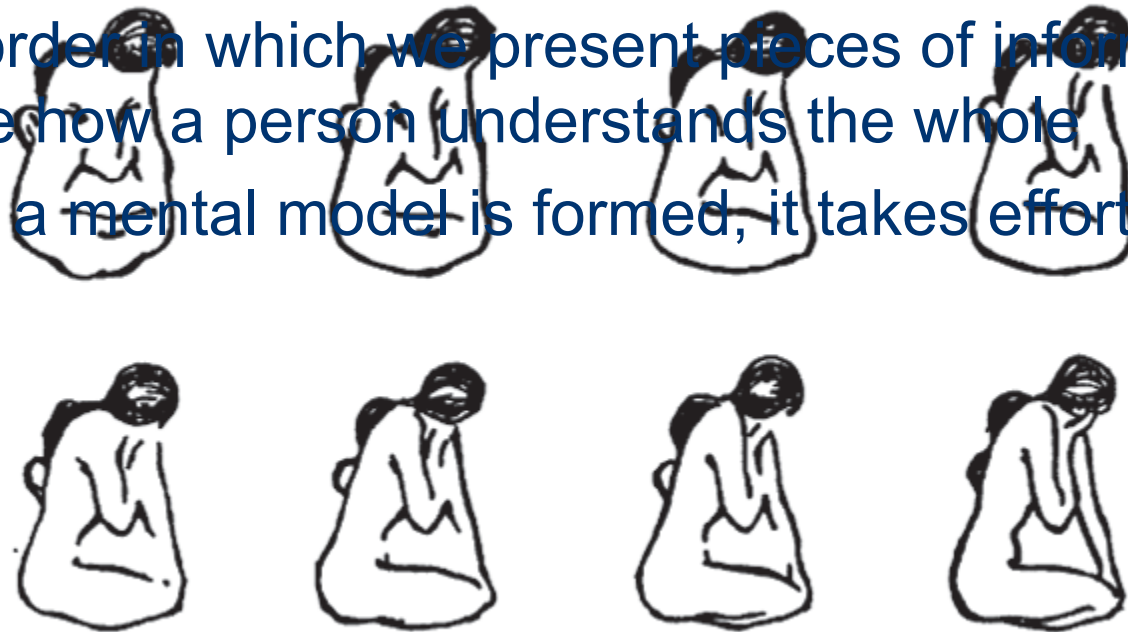
1. We tend to see what we expect to see

- Mental models are built from prior experience
- We expect new input to “fit” the existing model
- Updates are **expensive**: given input that almost fits, we’ll distort information to avoid re-fitting the model
- **Expectation** is at least as strong as perception



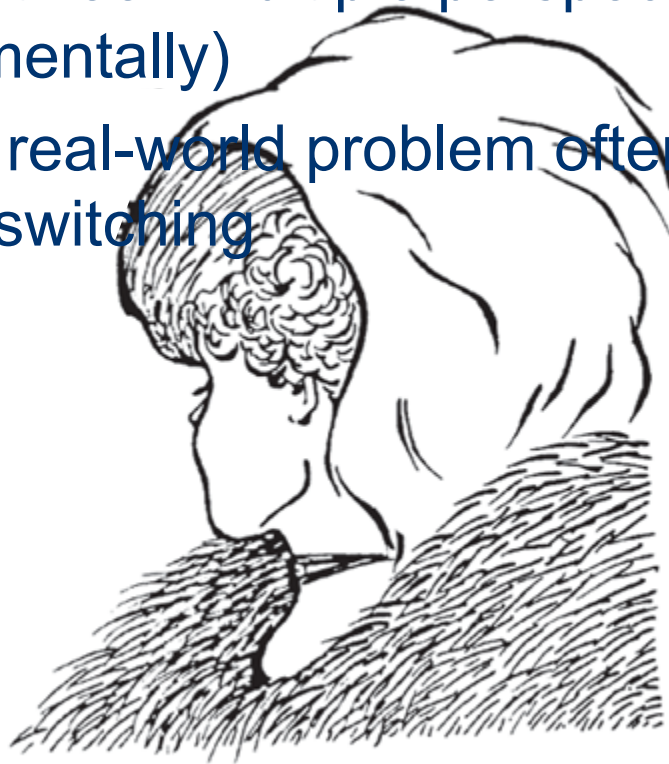
2. Mental models form quickly, and update slowly:

- “First impressions matter”
- Early information can have the highest impact
- The order in which we present pieces of information can shape how a person understands the whole
- Once a mental model is formed, it takes effort to alter it



3. New information gets incorporated into the existing model

- Integrating competing perspectives is challenging
- Switching between multiple perspectives is also difficult (visually or mentally)
- **Tricky part:** real-world problem often require such perspective switching



4. Initial exposure interferes with accurate perception



Blur size

128px

64px

32px

16px

8px

None

4. Initial exposure interferes with accurate perception

- Longer exposure to ambiguous data makes people **more confident** in their initial model
- This is true even if new data presents strong evidence that their model is **wrong**!
- **Important**: need to be intentional when we design, because incremental information can be **misleading**

The good, the bad, and the ugly...

The good:

- Well-tuned mental models let us process information quickly
- Frees up more processing power to synthesize information

The bad:

- People (esp. experts) tend not to notice information that contradicts their mental model
- A “fresh pair of eyes” can be beneficial

The ugly:

- Mental models are unavoidable: everyone has them, and they're all different
- **Key:** be aware of how mental models form, how they shape perception, and how to support (or challenge) them

Takeaways

- Visualization is about more than just **aesthetics**
- Tools to help us build a **mental model** of the data without actually having to look at all of it
- There are compelling **cognitive reasons** why some visualization techniques are helpful and others aren't

Want to learn more?
Take SDS136 with
Prof. McNamara next fall!

Thanks!



jcrouser@smith.edu

[@jordancrouser](#)

www.science.smith.edu/~jcrouser



HUMAN COMPUTATION
& VISUALIZATION
LABORATORY