



UNIVERSITY OF
CALGARY

Working with Data and Visualization

DATA 601: Fall 2019

Visual Thinking for Data Science

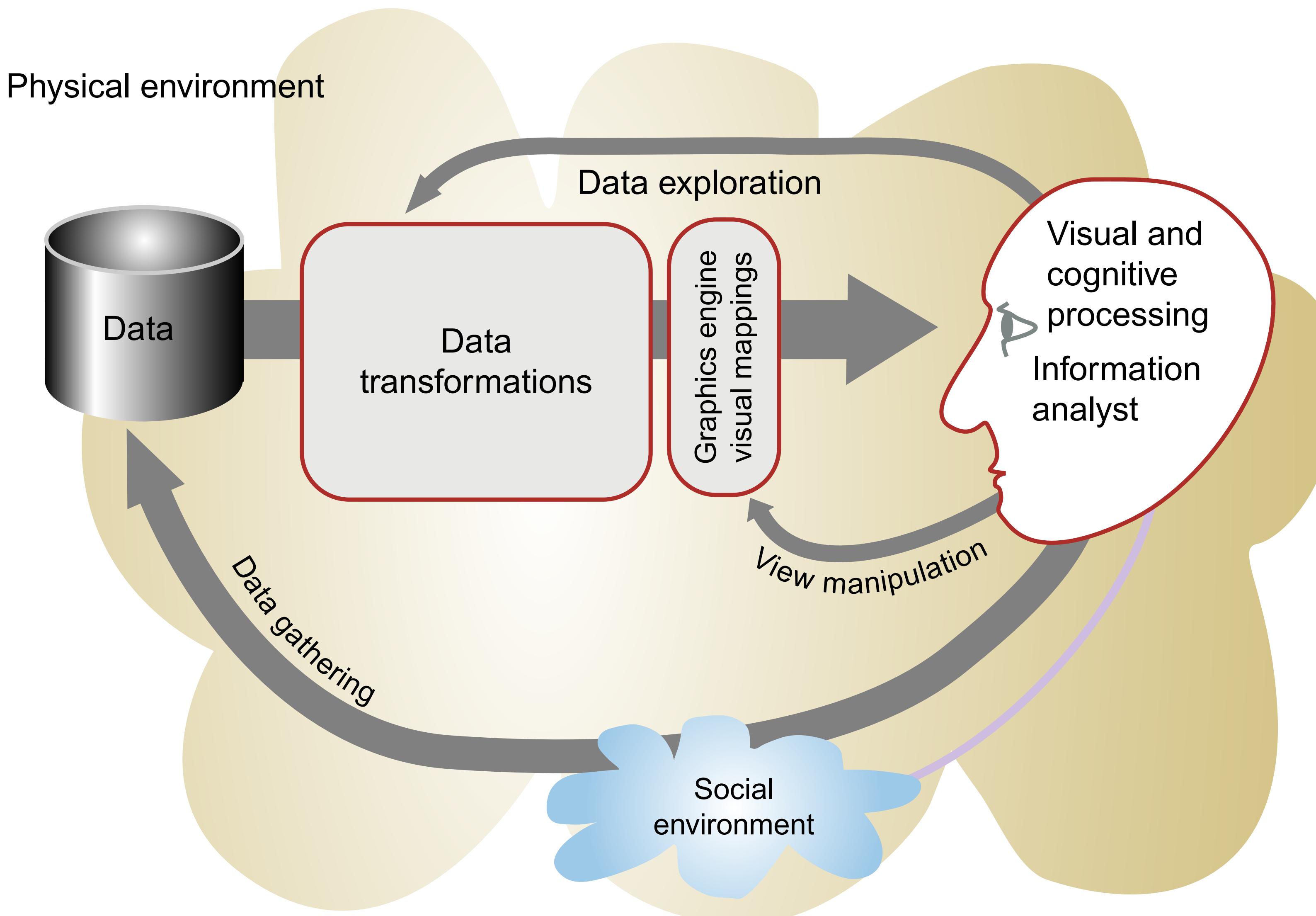
Usman R. Alim

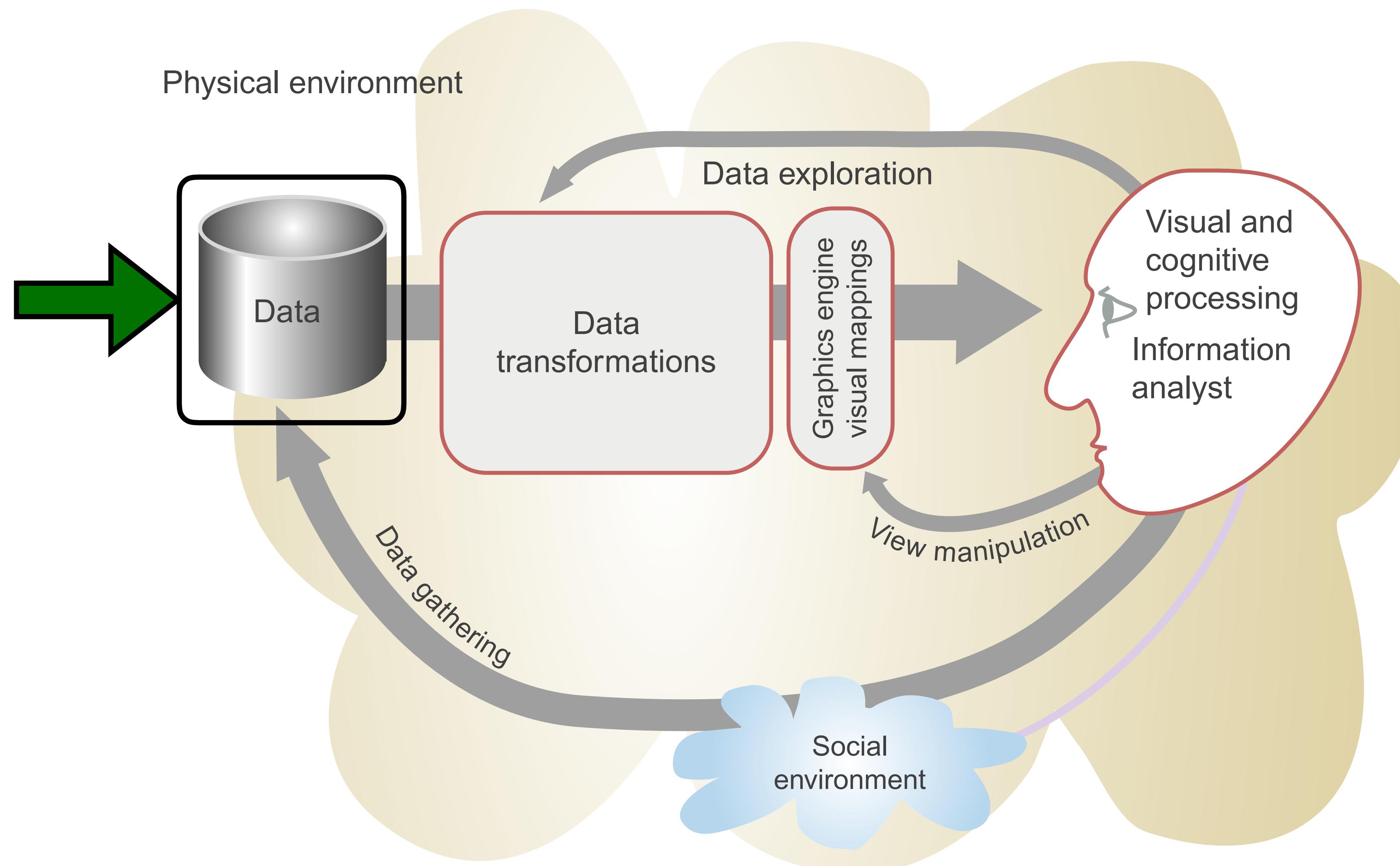
Associate Professor, Department of Computer Science
ualim@ucalgary.ca www.cpsc.ucalgary.ca/~ualim

Definitions

- A picture is worth a thousand words. Visualization aims at making sense of complex data via computer generated images.
- Telea:
The purpose of visualization is to get **insight**, by means of **interactive** graphics, into various aspects related to some process we are interested in...
- Munzner [2011]:
Computer-based visualization systems provide **visual representations** of datasets intended to help people carry out some **task** more **effectively**.

The Visualization Process

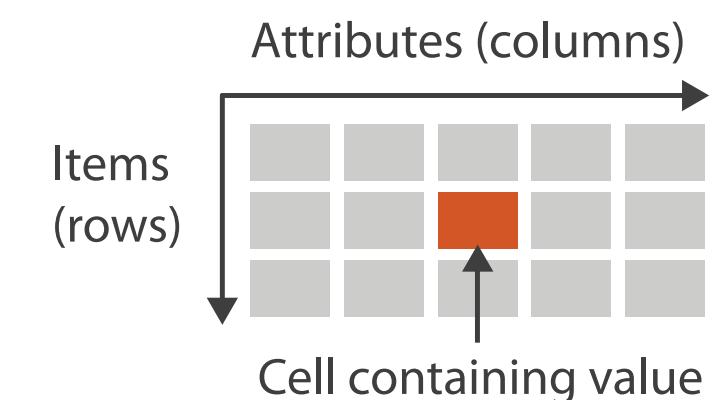




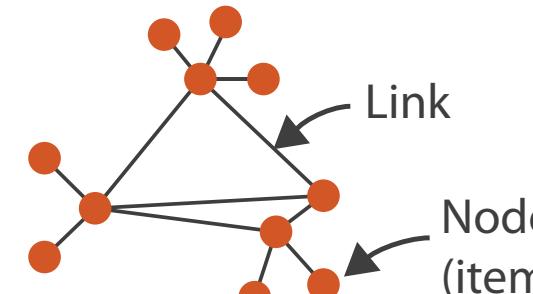
Flavours of Data

→ Dataset Types

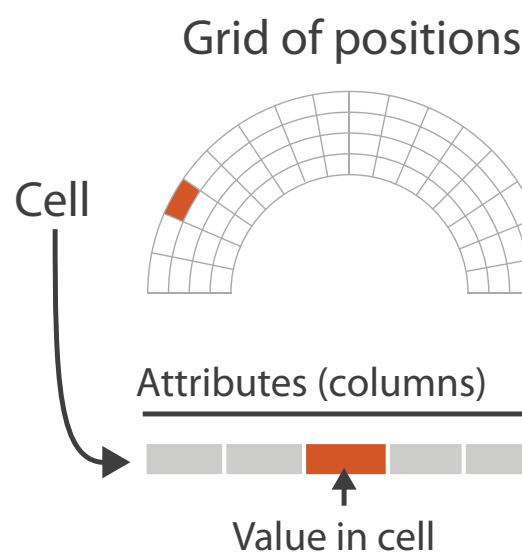
→ Tables



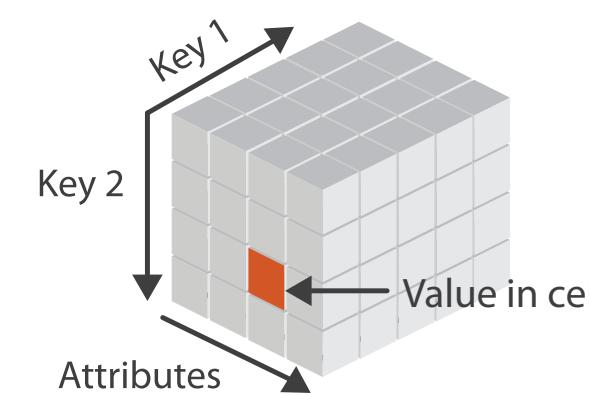
→ Networks



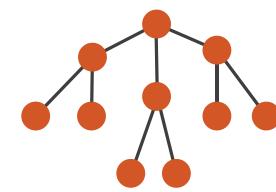
→ Fields (Continuous)



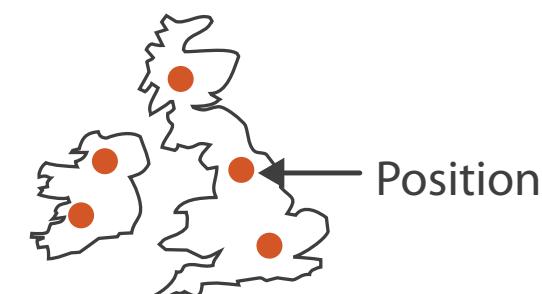
→ Multidimensional Table



→ Trees



→ Geometry (Spatial)



→ Attribute Types

→ Categorical



→ Ordered

→ Ordinal



→ Quantitative



→ Ordering Direction

→ Sequential



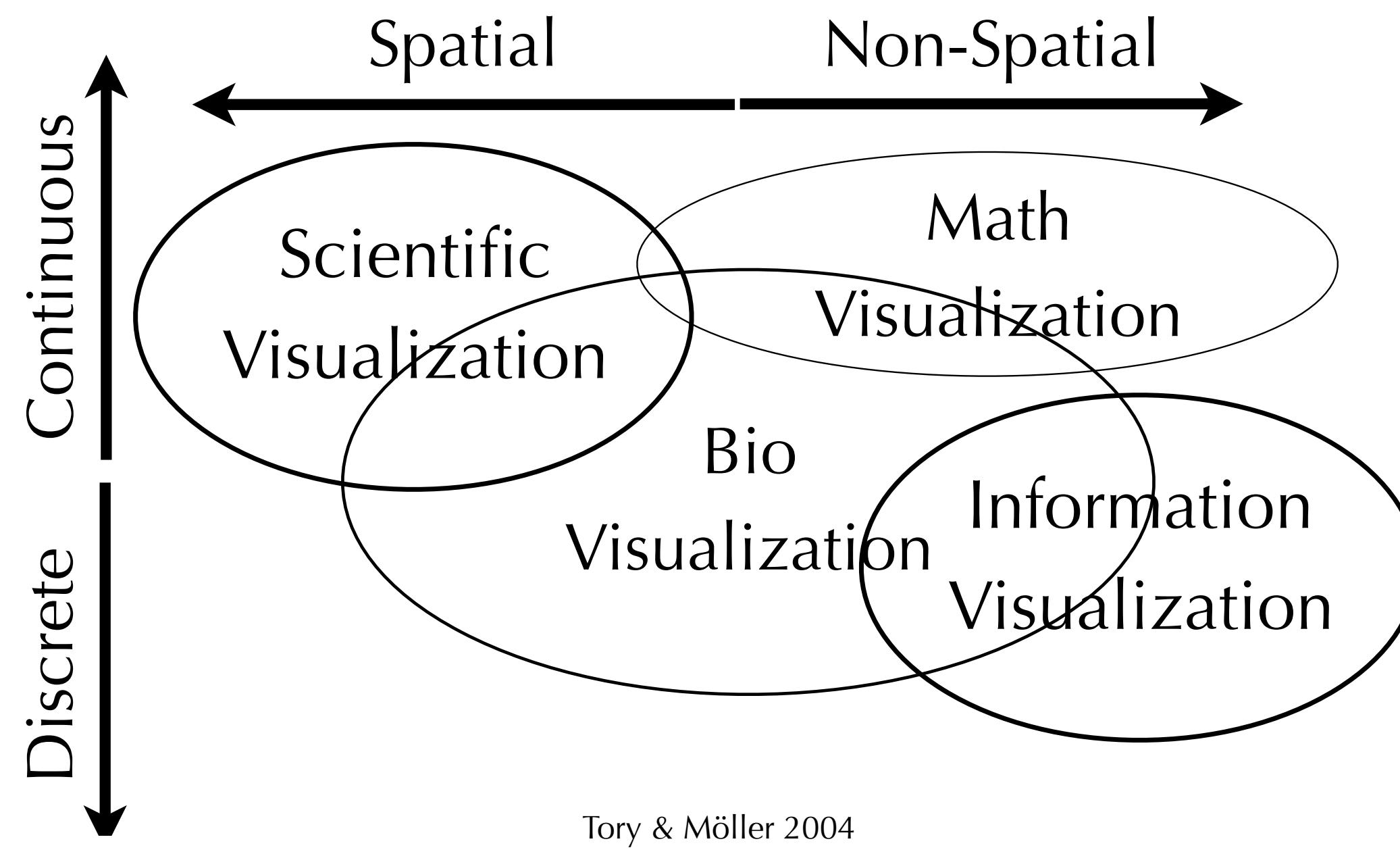
→ Diverging



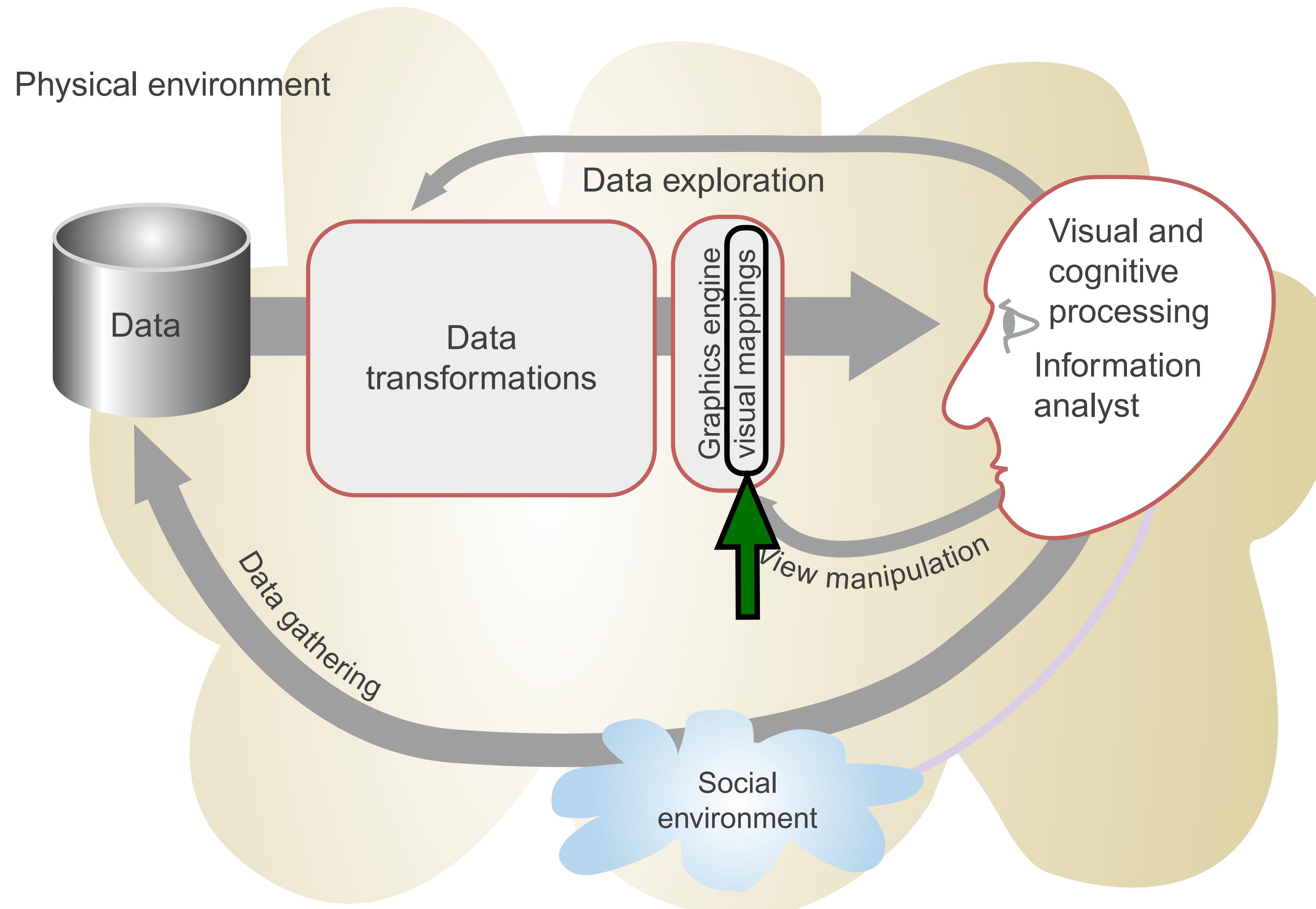
→ Cyclic



Flavours of Visualization

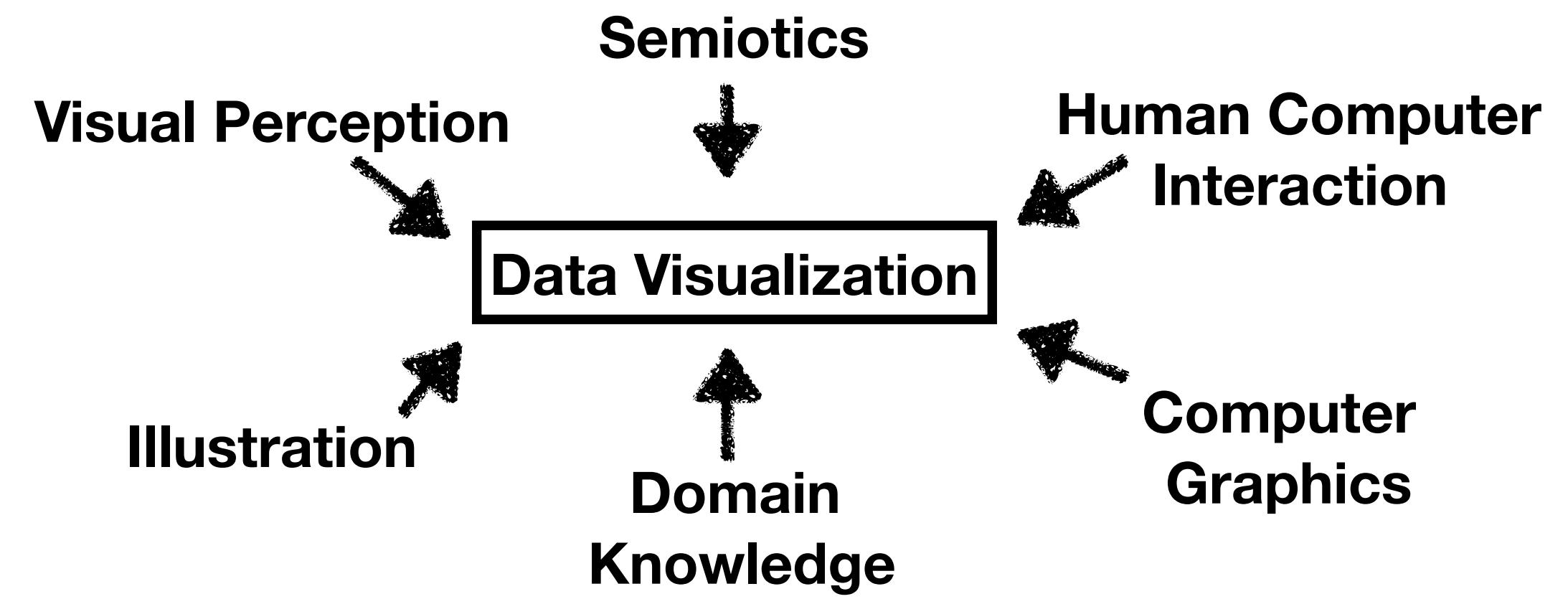


- **SciVis:**
Spatial layout of the visualization is a given (usually 3D).
- **InfoVis:**
Spatial layout of the visualization is a design choice (usually 2D).
- **Visual Analytics:**
combines automatic and visual analysis methods with a tight coupling through human interaction.



Visual Representations of Data

- **Semiotics:**
The study of symbols and how they convey meaning.
- **Visual Perception:**
The ability to perceive our surroundings through light entering the eye.
- **Illustration:**
A decoration, interpretation or visual explanation of a text, concept or process.
- **Domain Knowledge:**
Knowledge of the discipline or problem where the visualization is to be used.
- **Computer Graphics:**
A sub-field of Computer Science which deals with the synthesis and manipulation of visual content.
- **Human Computer Interaction:**
studies the design of computer technology focusing on the interfaces between users and computers.



The many facets of Data Vis.

Nature and Nurture



is

Some visuals are hard to interpret while others are very intuitive.

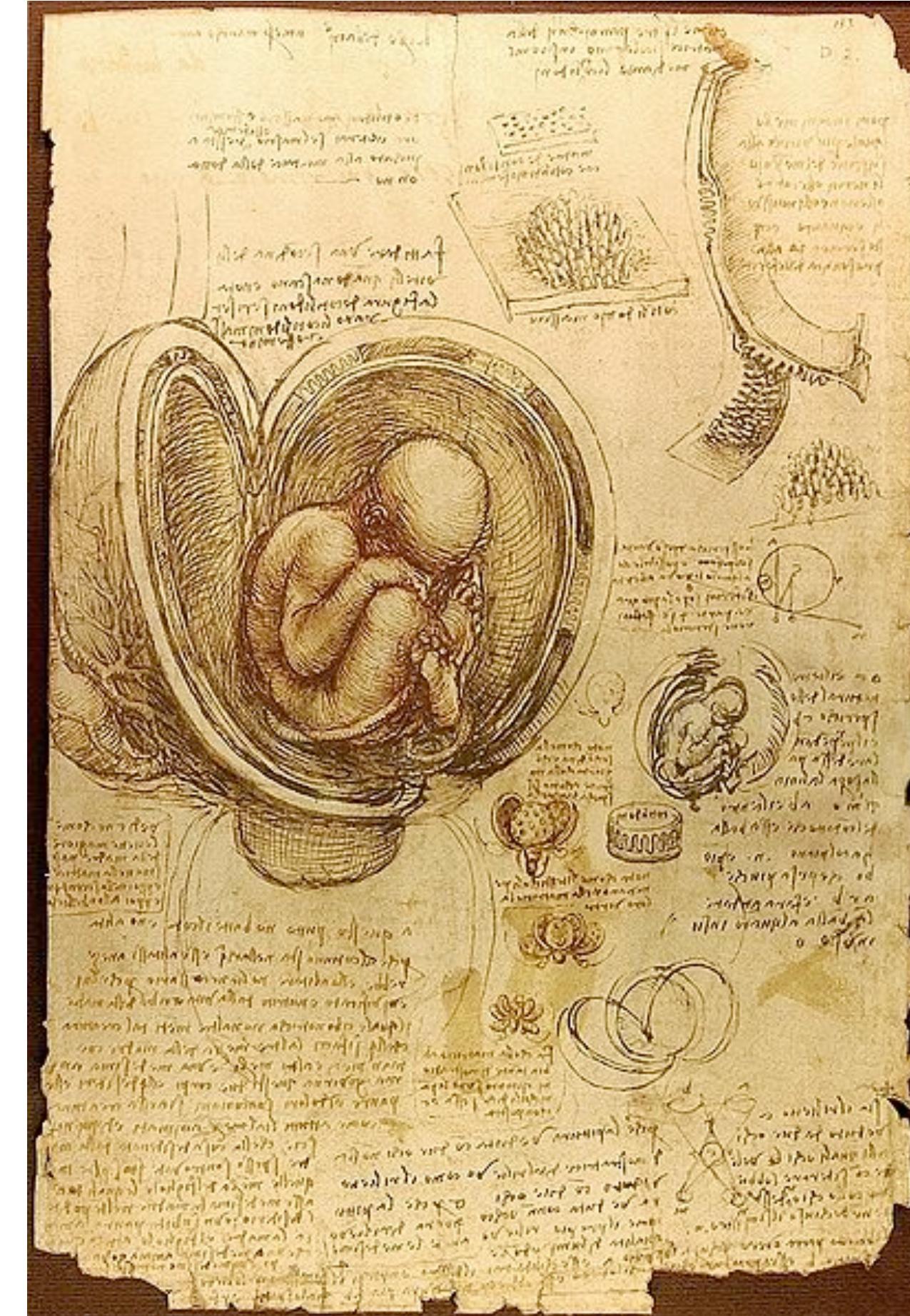
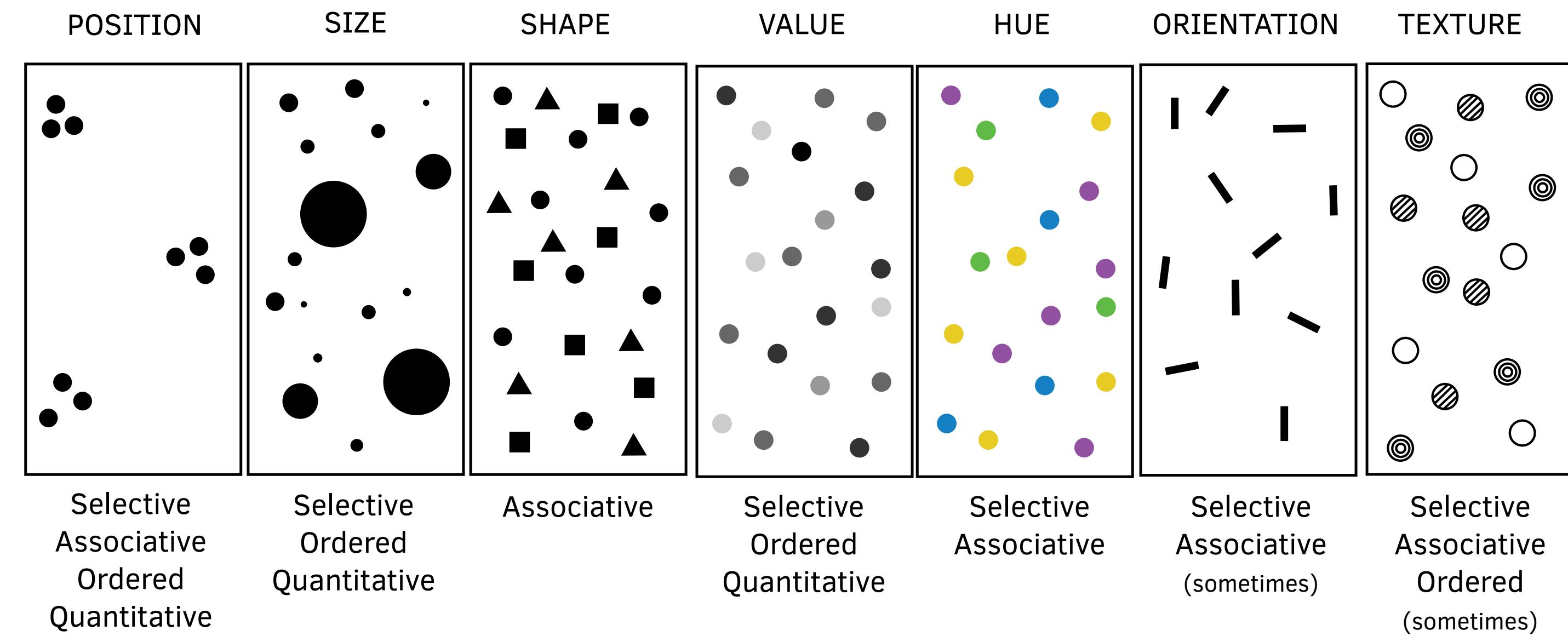


Illustration by Leonardo da Vinci, ca. 1500

Bertin's Visual Variables (1967)



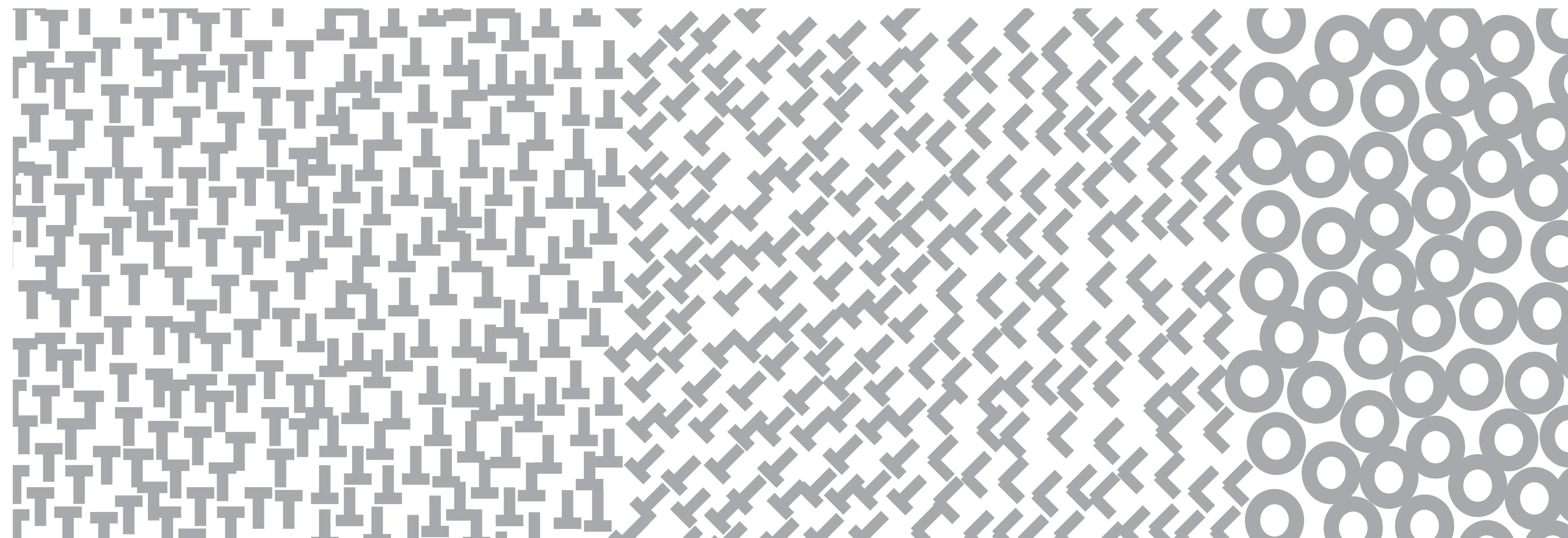
Born in the field of Cartography, *visual variables* are fundamental visual units used to encode data.

- **Selective:**
A selective variable allows us to immediately isolate a group of symbols based on change in the variable.
- **Associative:**
An associative variable allows grouping across changes in the variable.
- **Ordered:**
Variable has an immediately recognizable order.
- **Quantitative:**
Variable allows an estimation of the numerical difference between symbols.

Ware's Sensory Symbols

- *Sensory symbols* derive their expressive power from their ability to use the human (visual) perceptual system.
- *Understanding without training* - no specialized training is required to perceive the meaning.
- *Sensory immediacy* - the processing of certain kinds of sensory information is hardwired and fast.
- *Cross-cultural validity* - a sensory code will be understood across cultural boundaries.

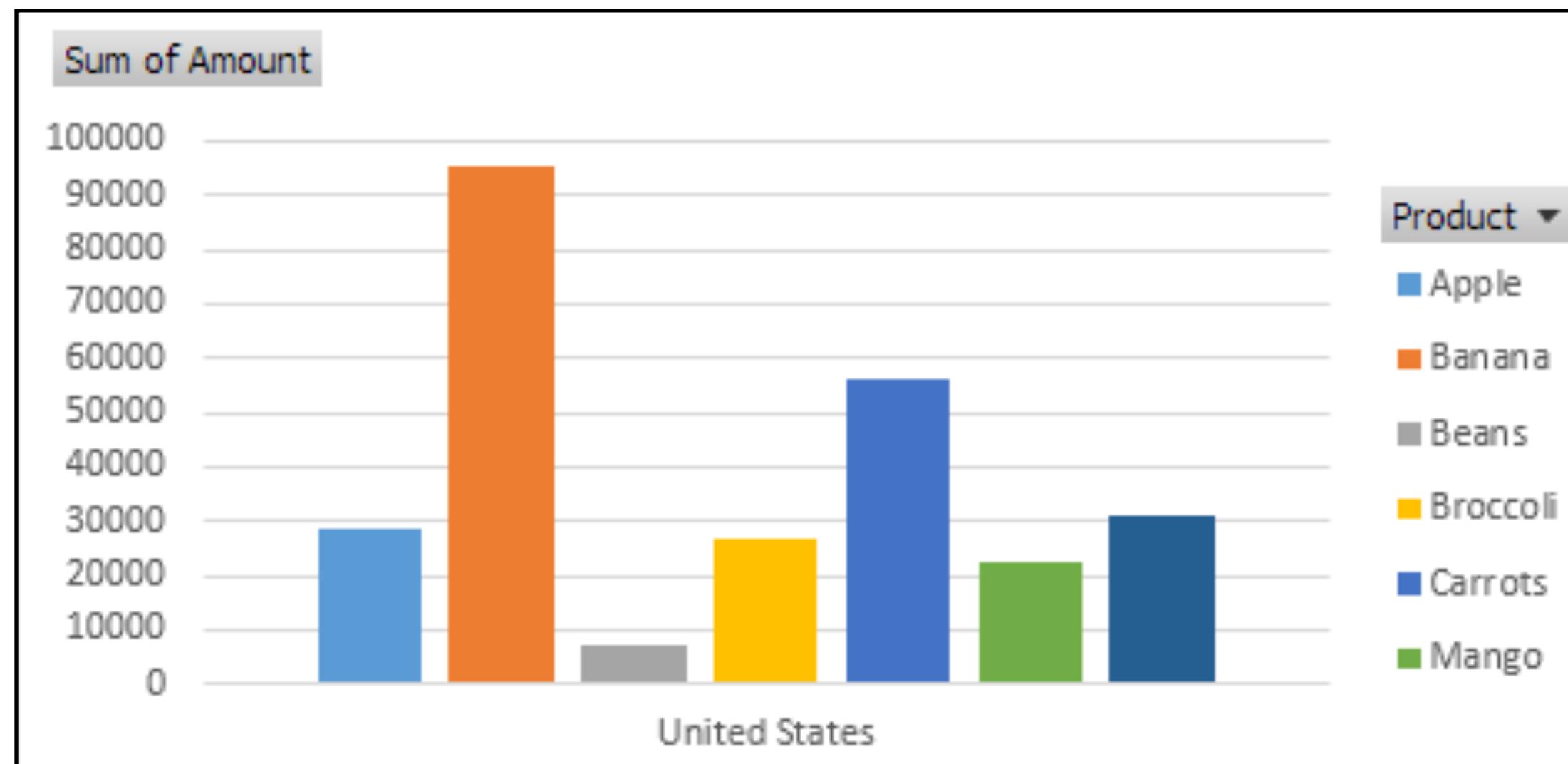
Sensory Immediacy



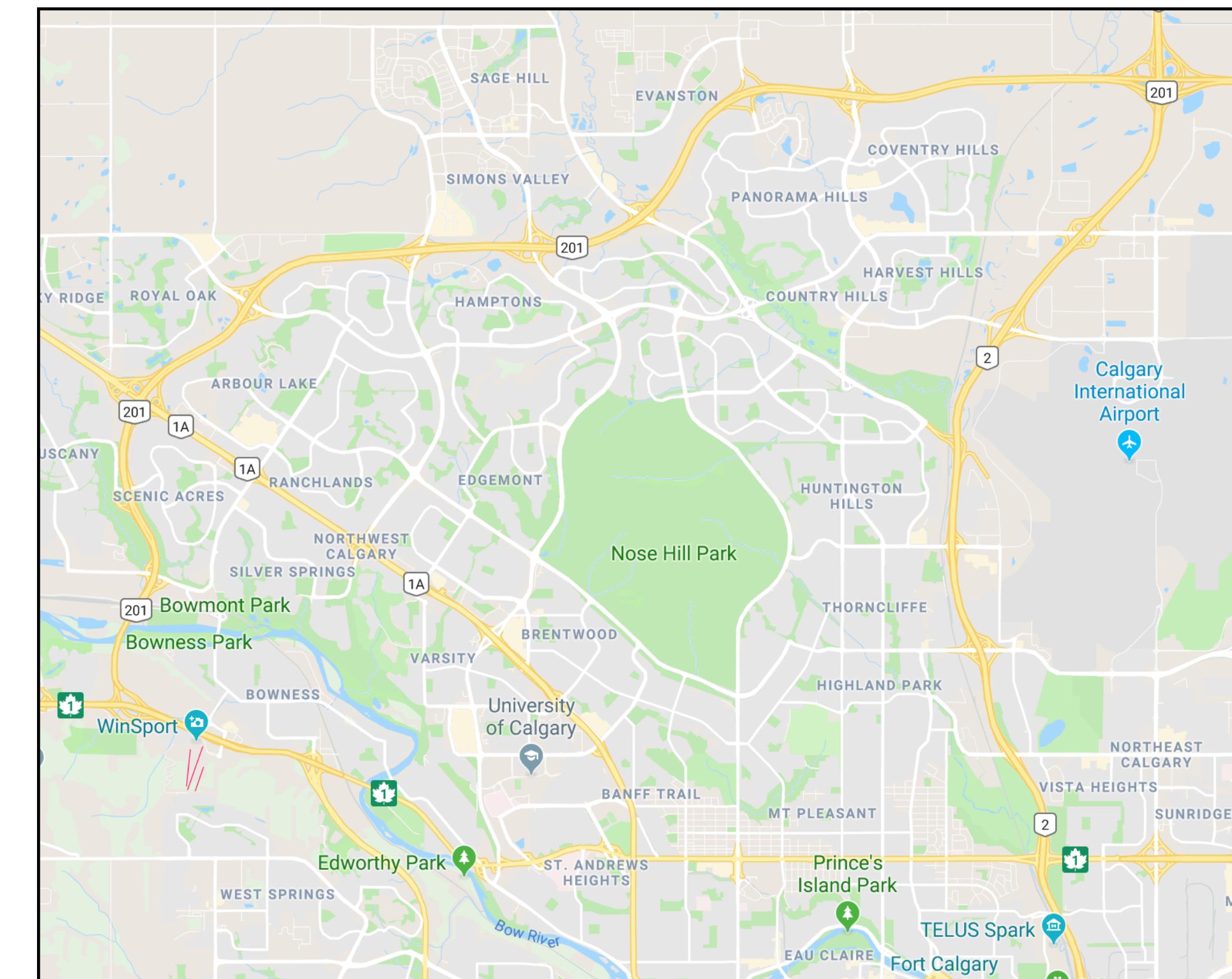
Five regions of texture

Visual Queries

Visual queries rely on the processing power of our visual system to help us quickly perform cognitive tasks.



Which product sold the most?
How do the sales compare to carrots?



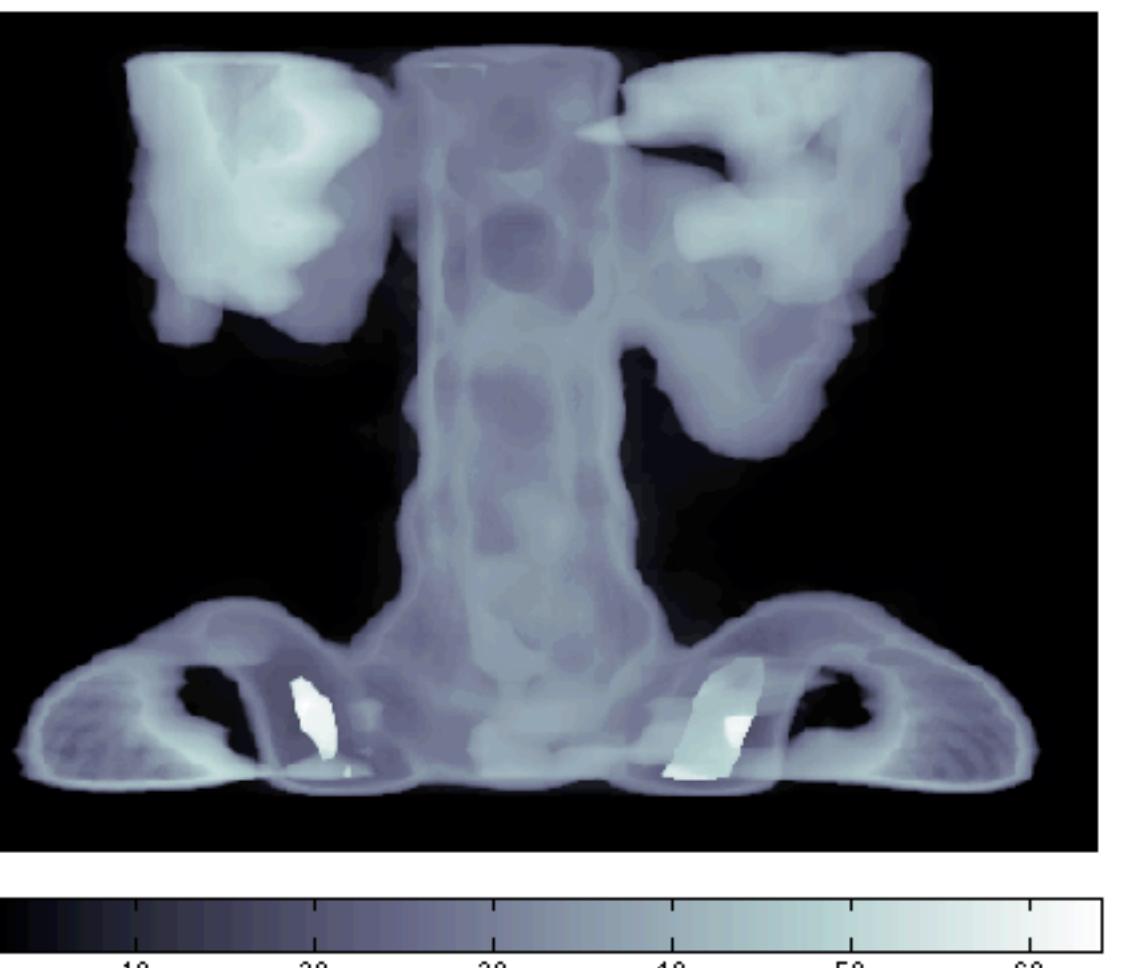
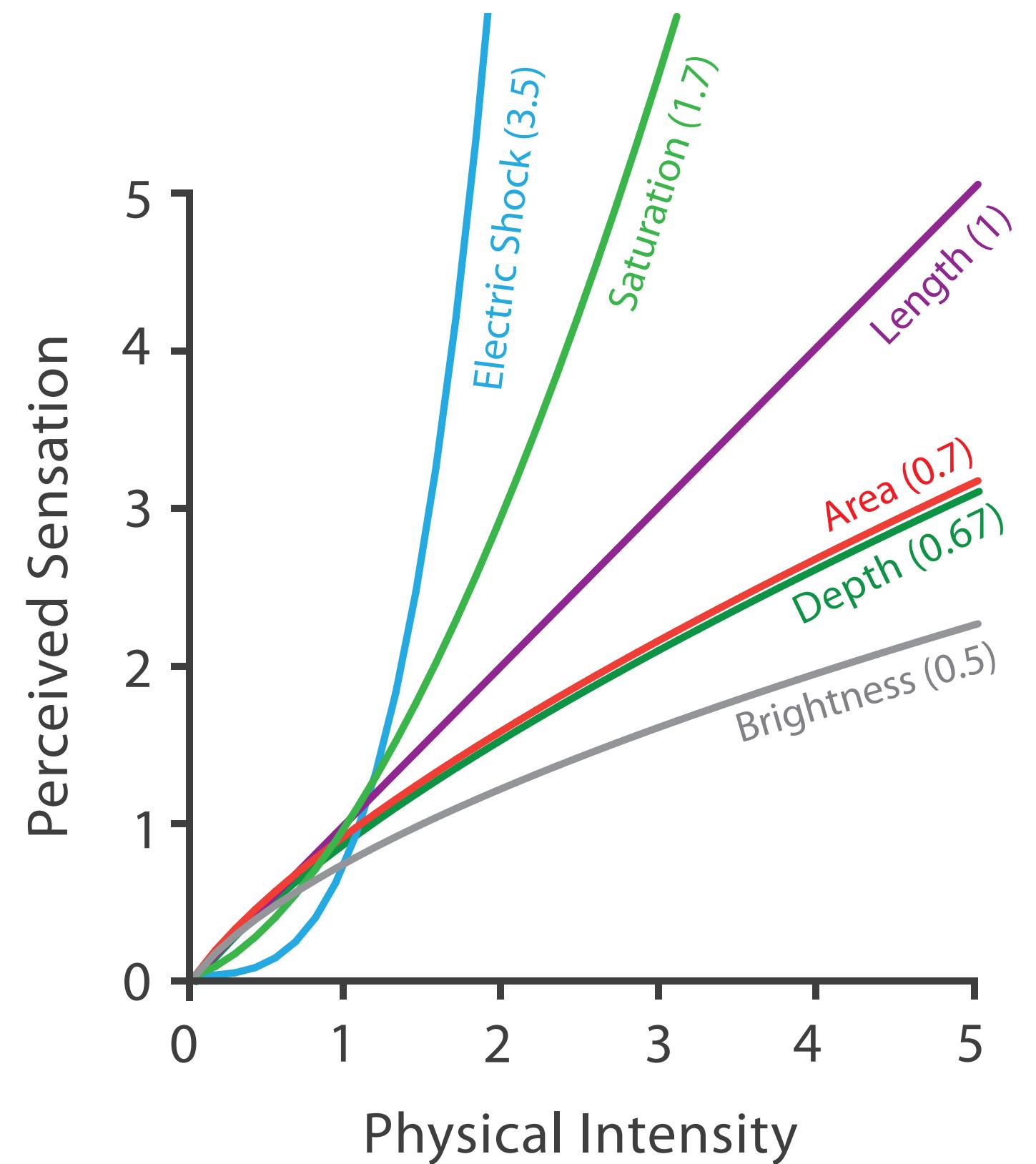
How to go from UCalgary to the airport?

Effective Visualizations

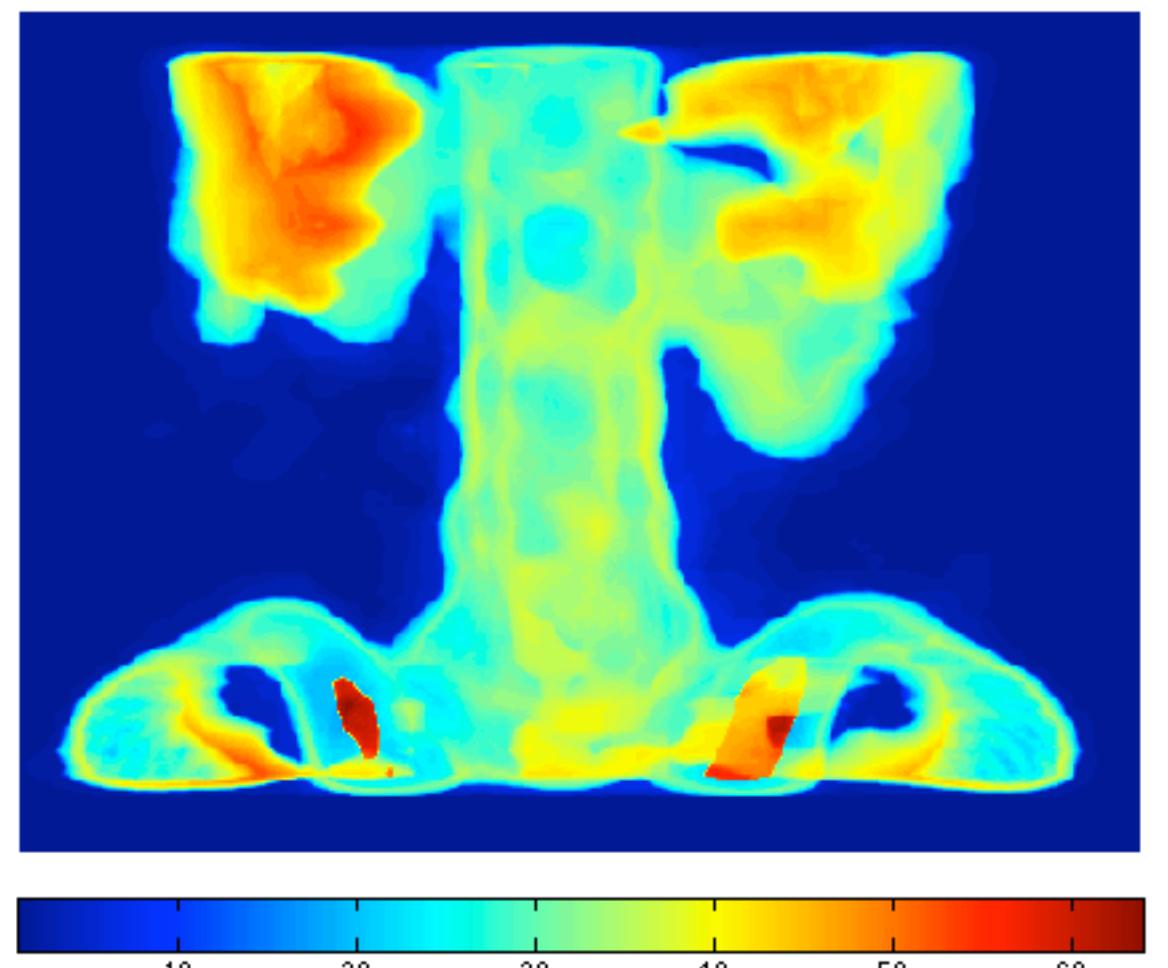
- Effective visualizations require effective *visual encodings* that help us carry out analysis *tasks* through effective *visual queries*.

Perception is Complex

Steven's Psychophysical Power Law: $S = I^n$



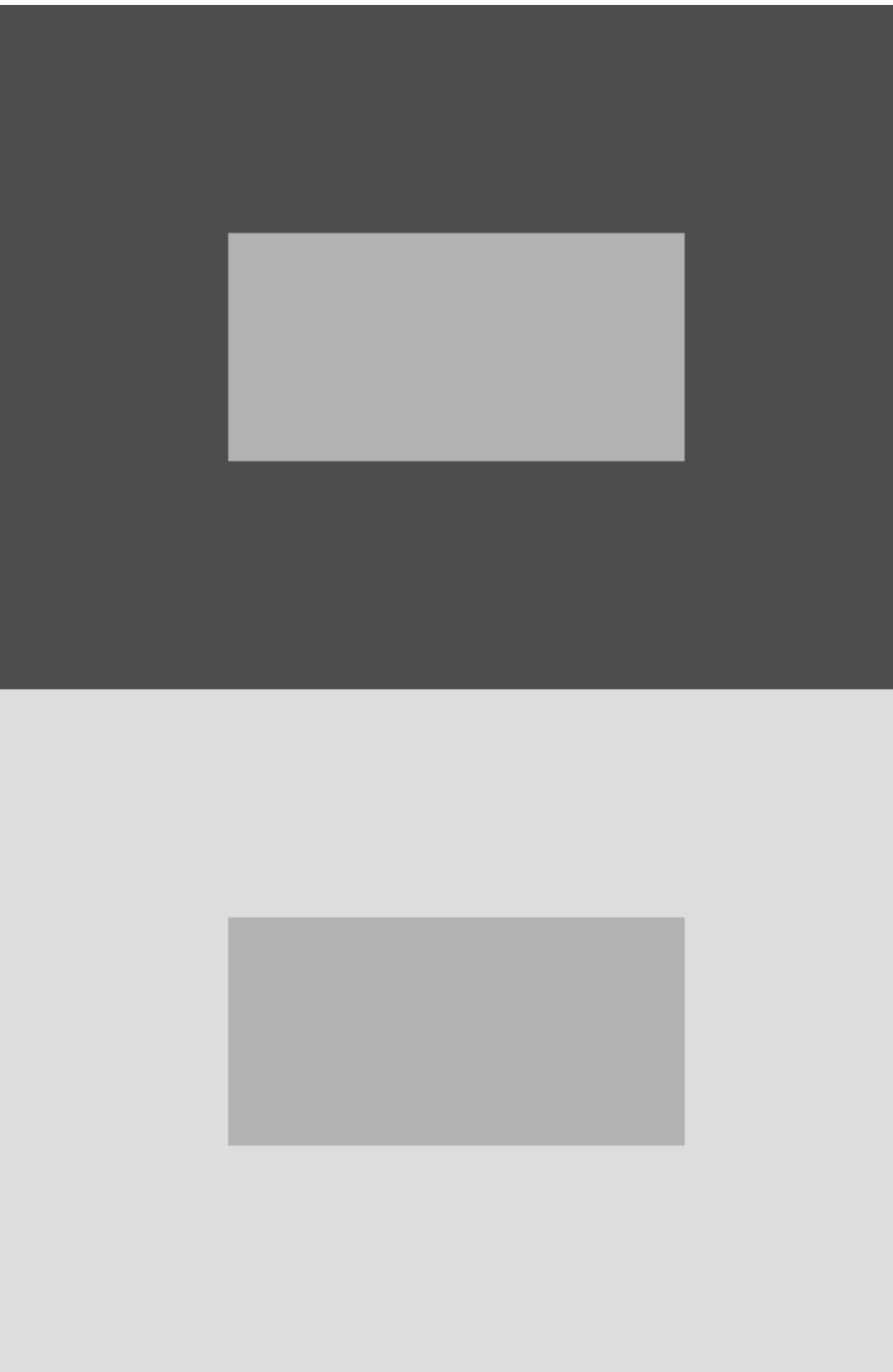
A luminance based colormap is more effective compared to the rainbow colormap.



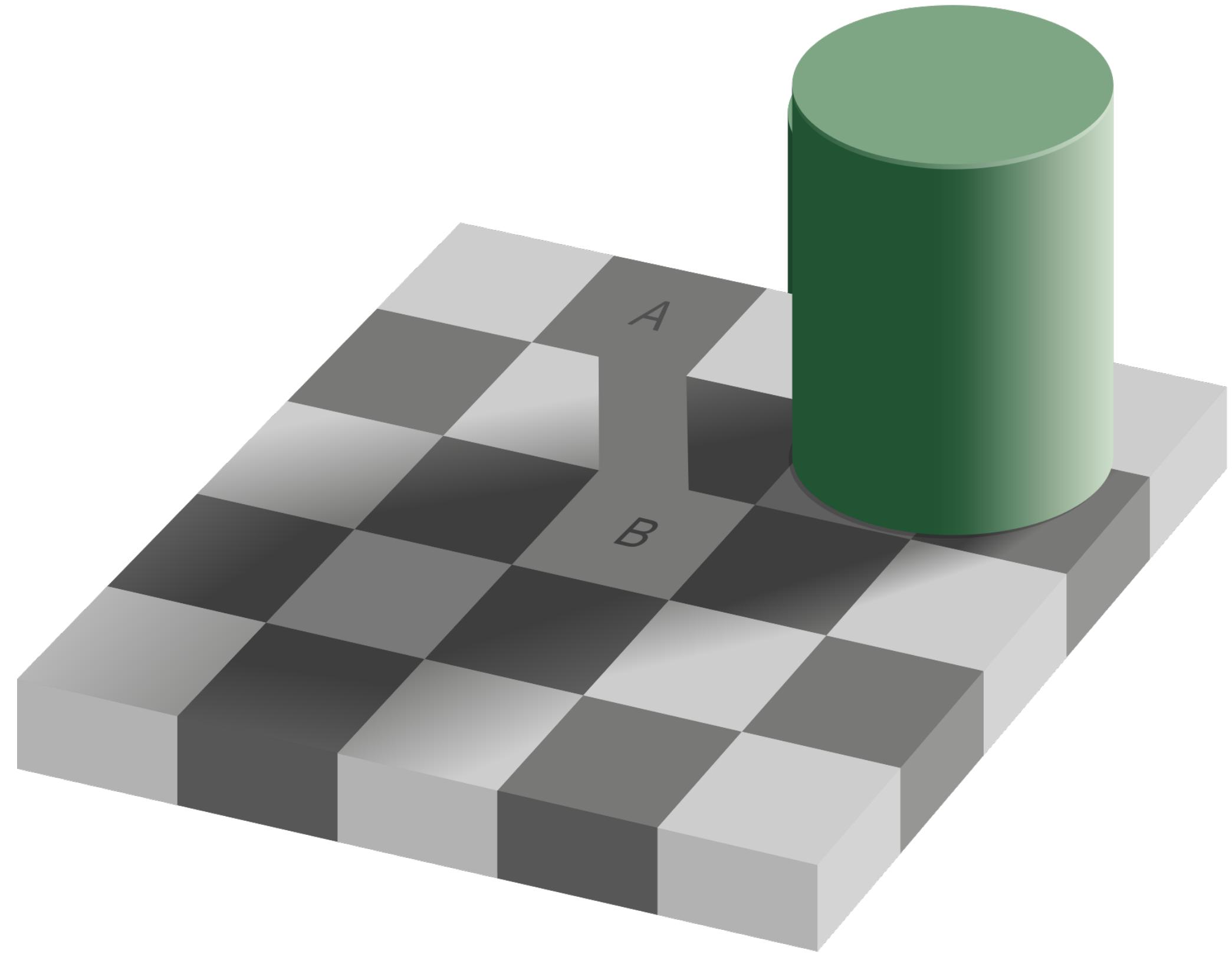
The 'rainbow' colourmap is notoriously non-linear.

Simultaneous Contrast

Colors of different objects affect each other.

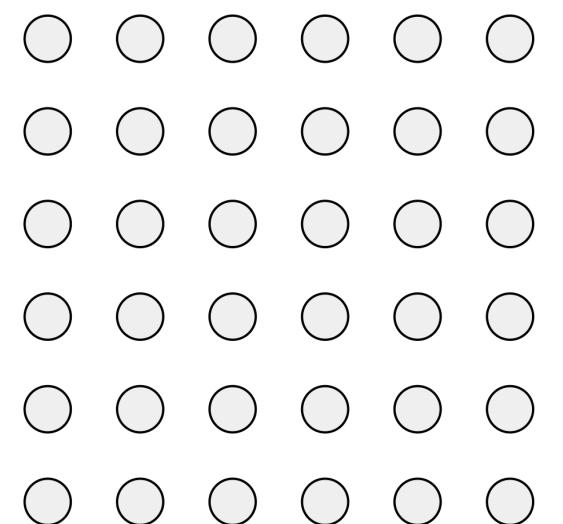


The inner rectangles have the same luminance

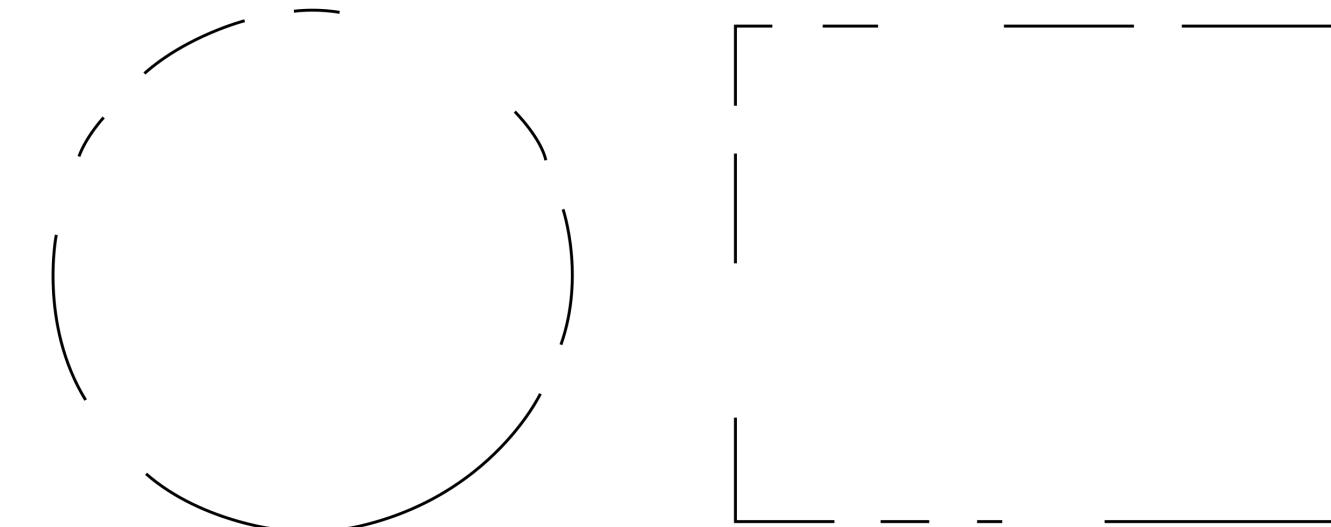
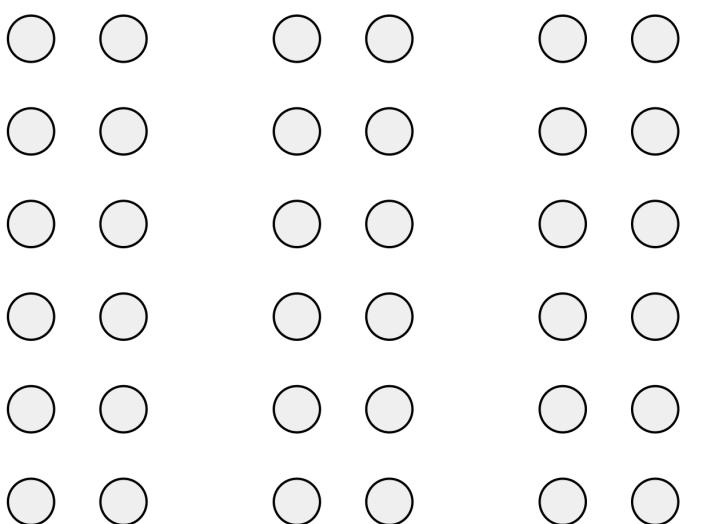


A and B have the same luminance

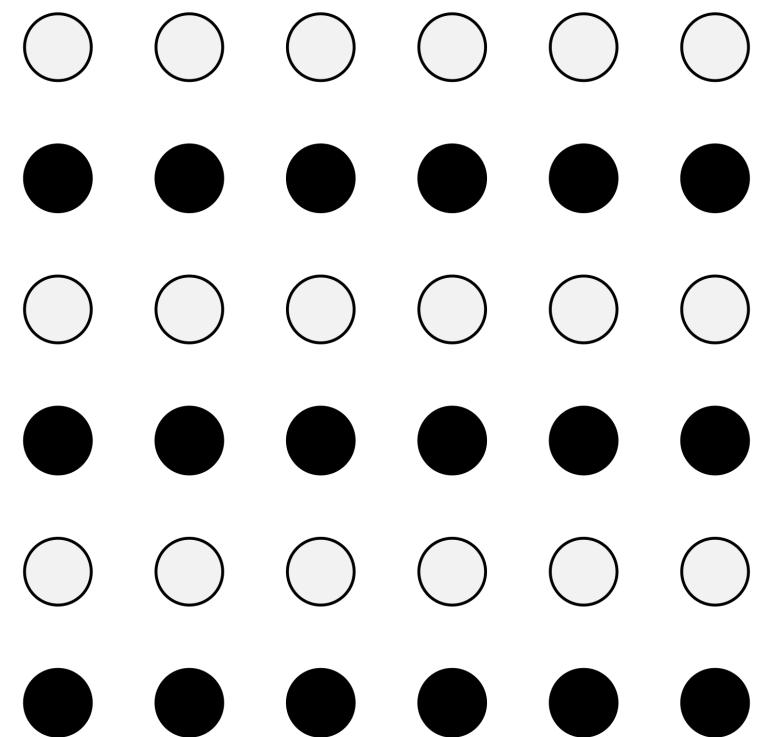
Gestalt Principles



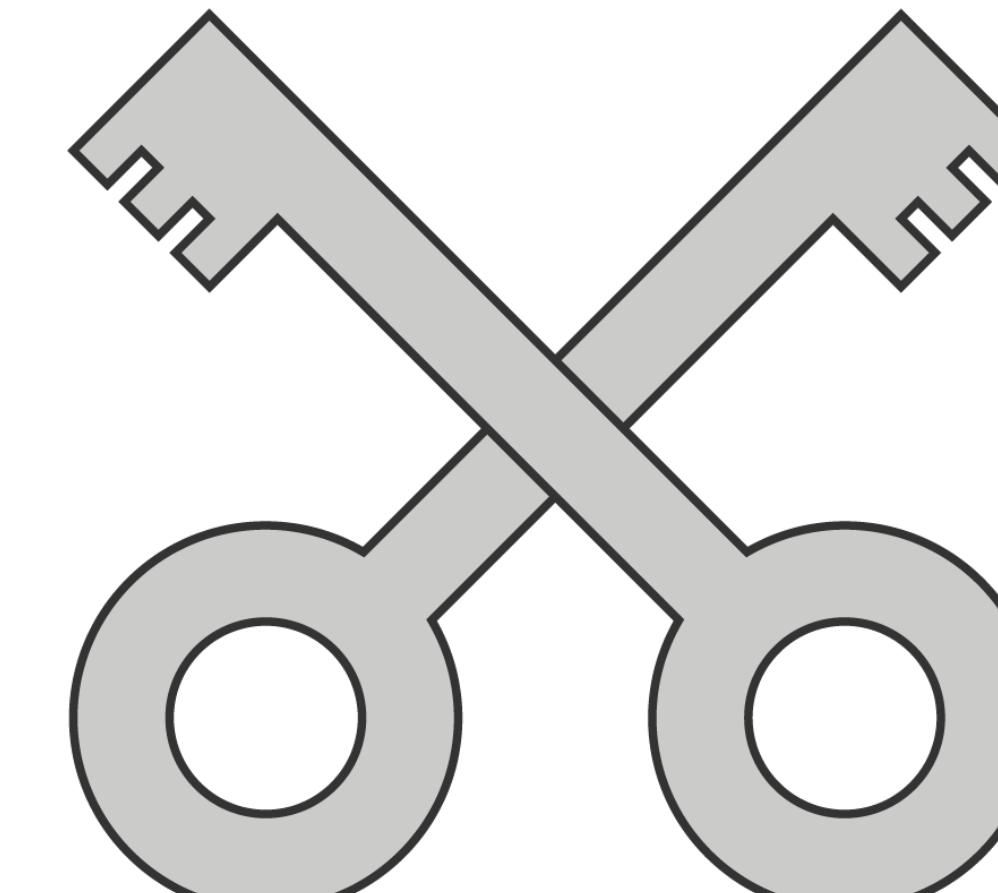
Proximity



Closure



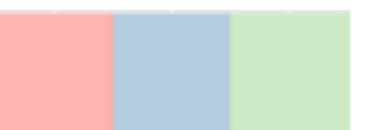
Similarity



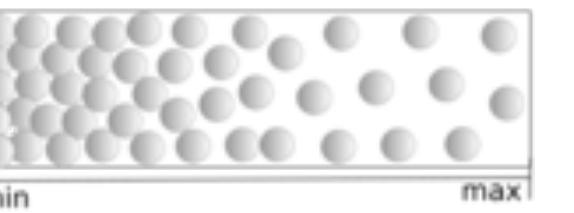
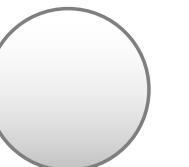
Continuity

Example: Geological Data

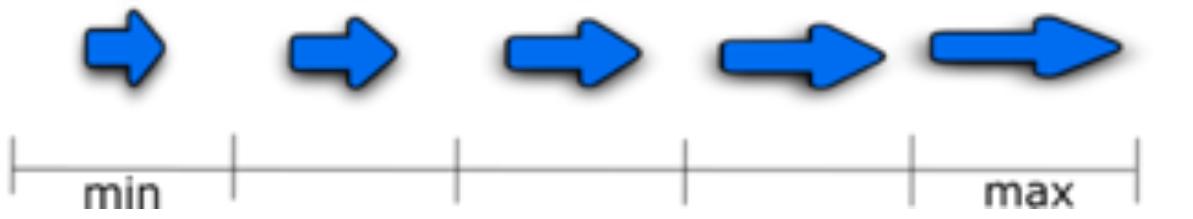
Rock type: Categorical



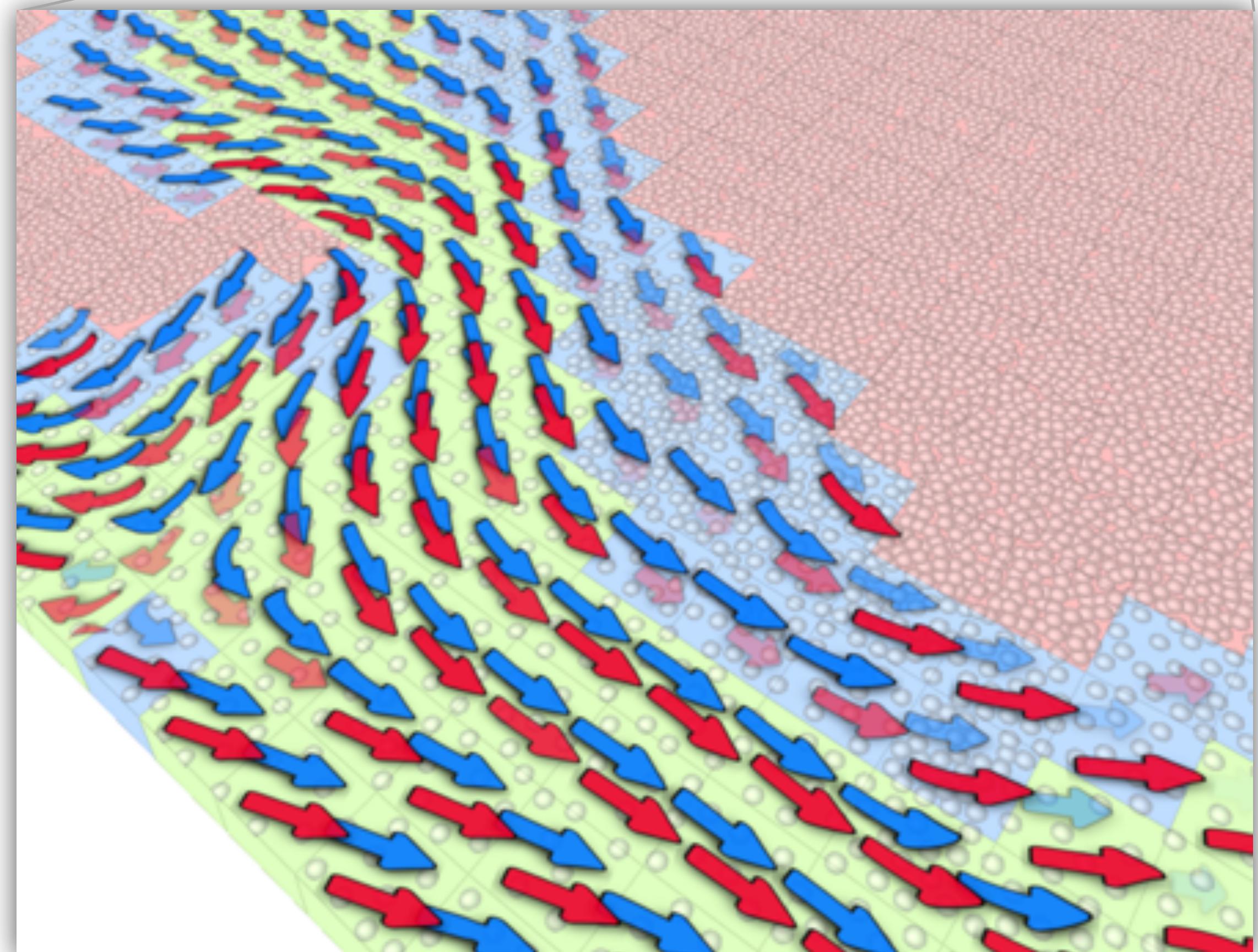
Porosity: Scalar



Water flow: Vector



Oil flow: Vector



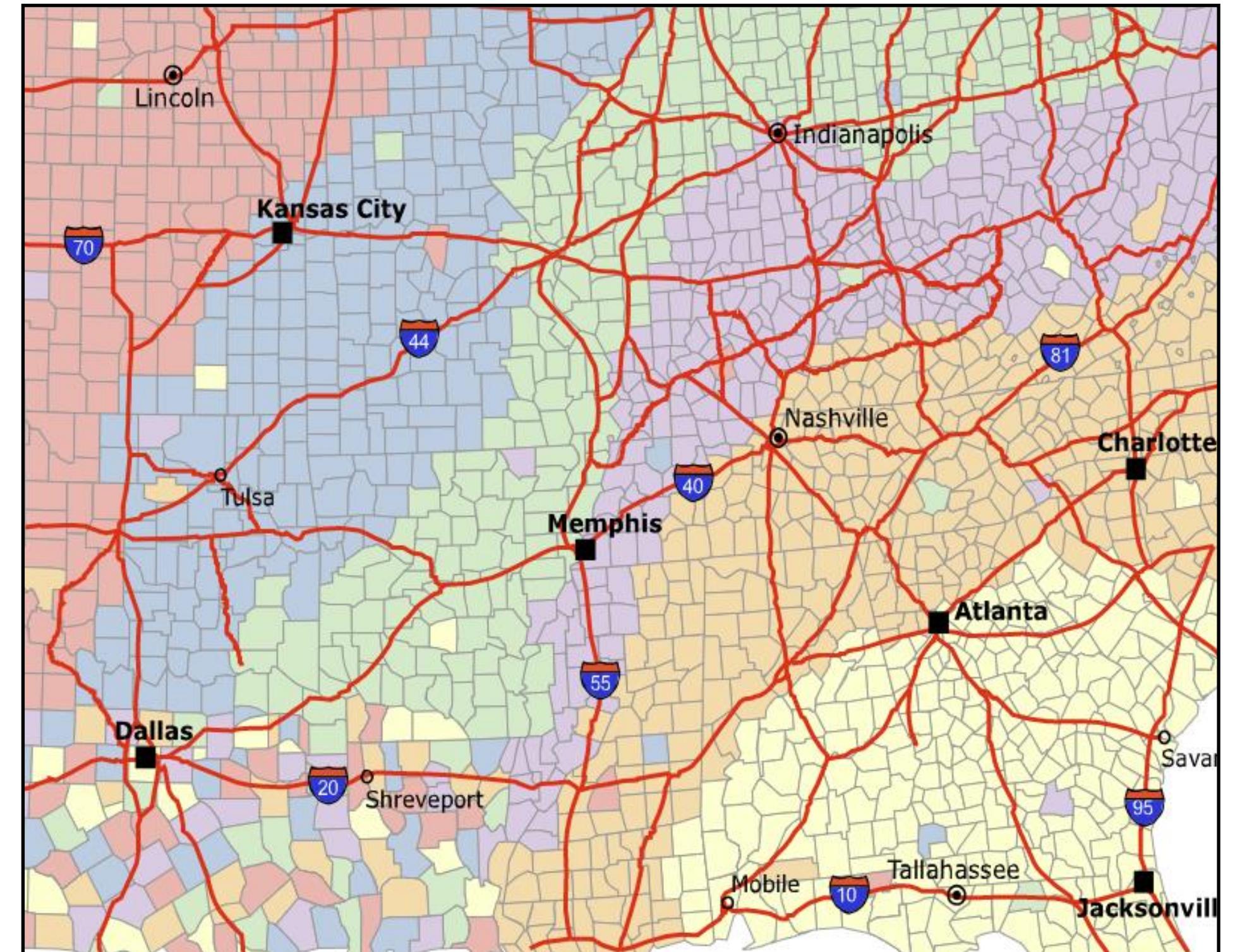
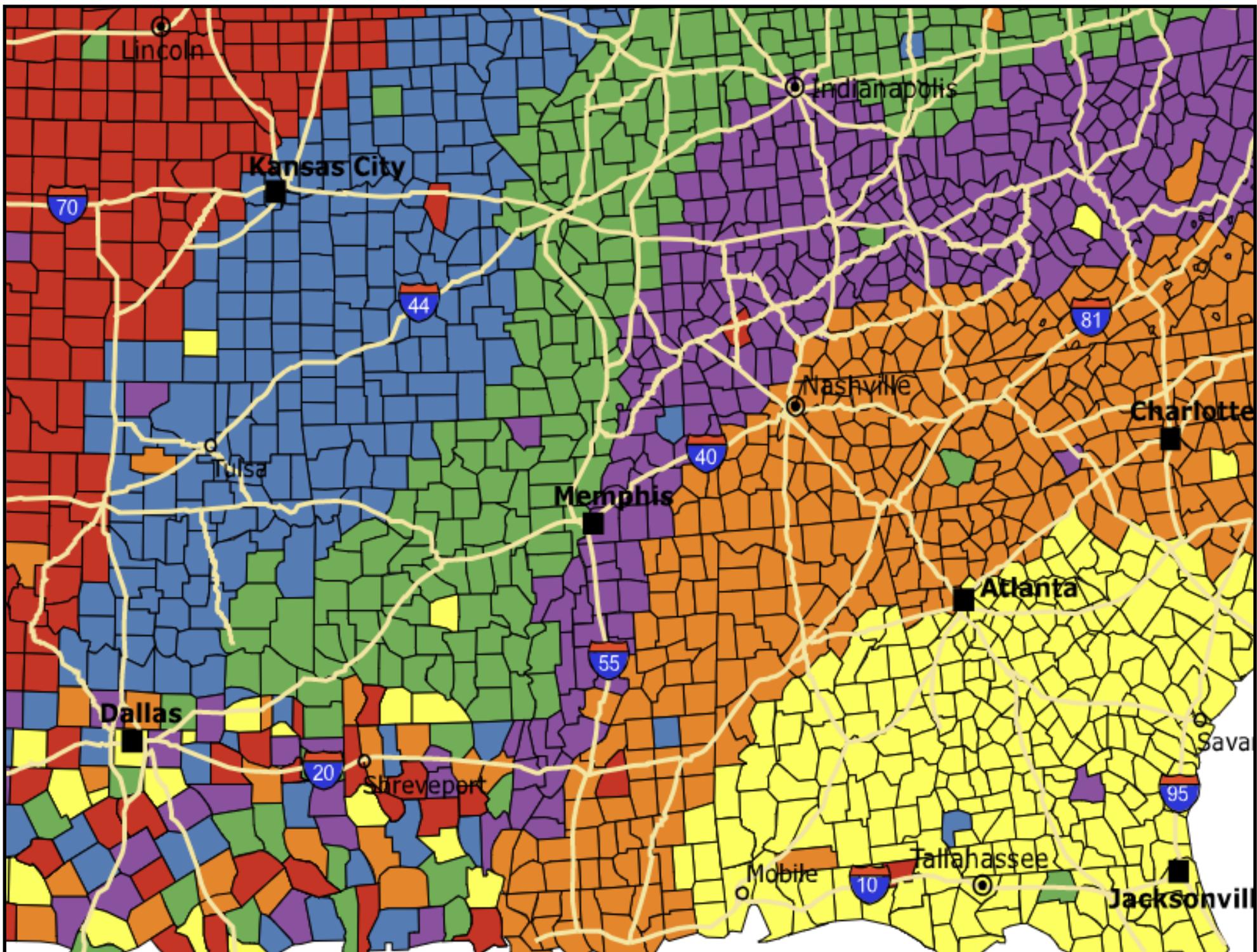
Some Guidelines

Ware's Design Guidelines

- *Design graphic representations of data by taking into account human sensory capabilities in such a way that important data elements and data patterns can be quickly perceived.*
- *Important data should be represented by graphical elements that are more visually distinct than those representing less important information.*
- *Where two or more tools can perform the same task, choose the one that allows for the most valuable work to be done per unit time.*

Some Guidelines

Effective Colour Use



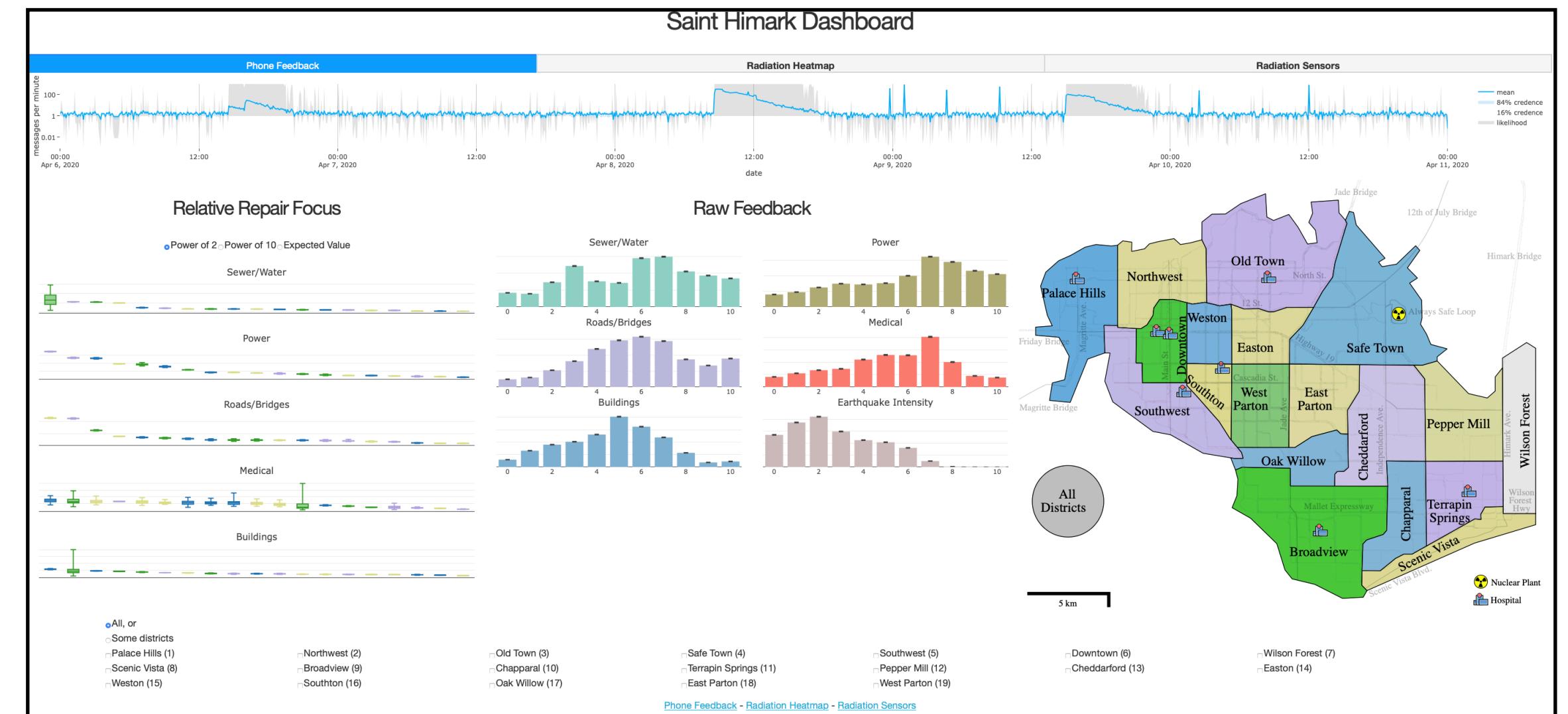
For colour advice, consult [ColorBrewer](#)

Some Guidelines

Schneiderman's Visual Information Seeking Mantra

- *Overview first, zoom and filter, then details-on-demand.*

Example:

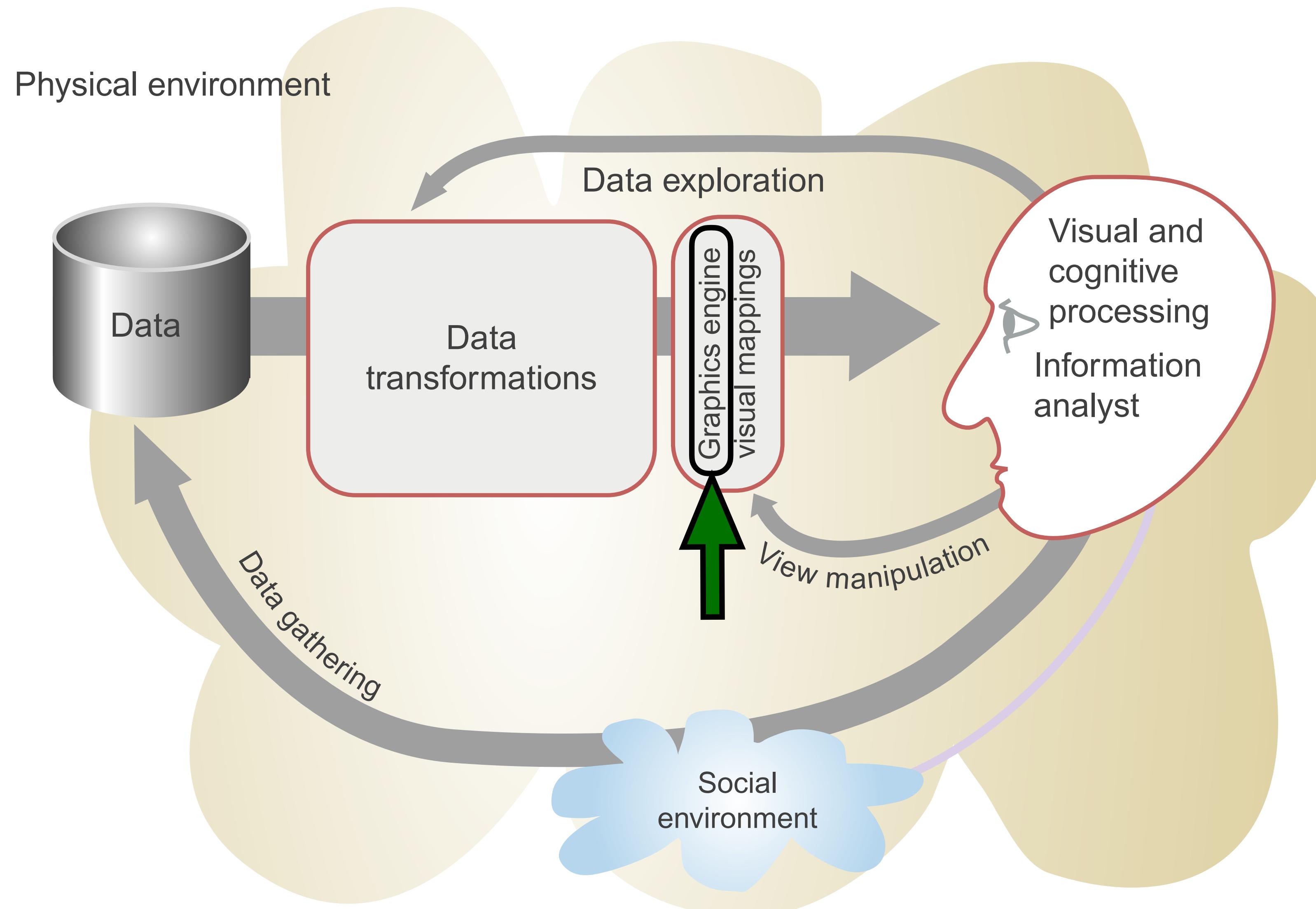


UofC-Bayes: Dashboard by H. Hornbeck and U. Alim, VAST 2019 Challenge

Some Guidelines

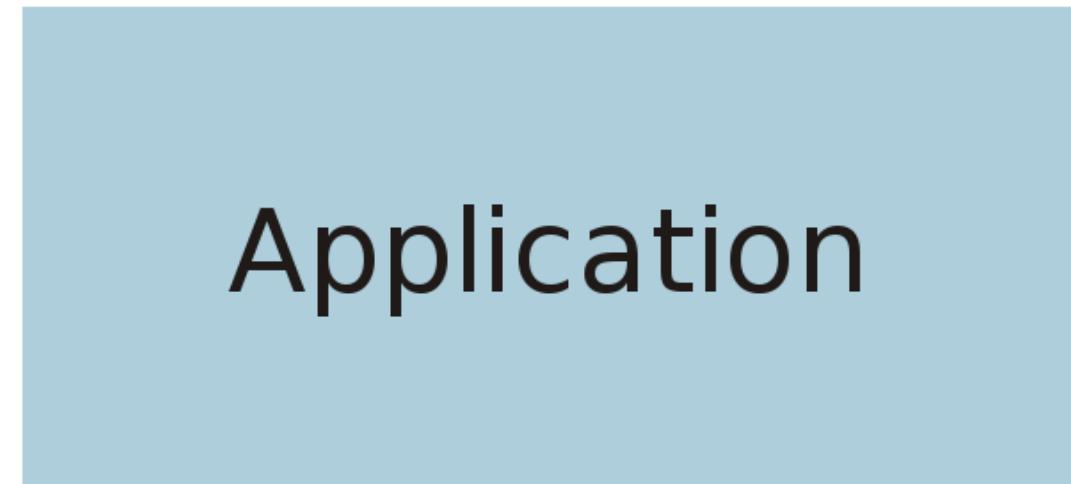
Catalogue of Charts

- If you are not sure about which chart type to use, perhaps consult a chart chooser:
 - Data Visualization Catalogue
 - Chart Suggestions
 - The Chartmaker Directory

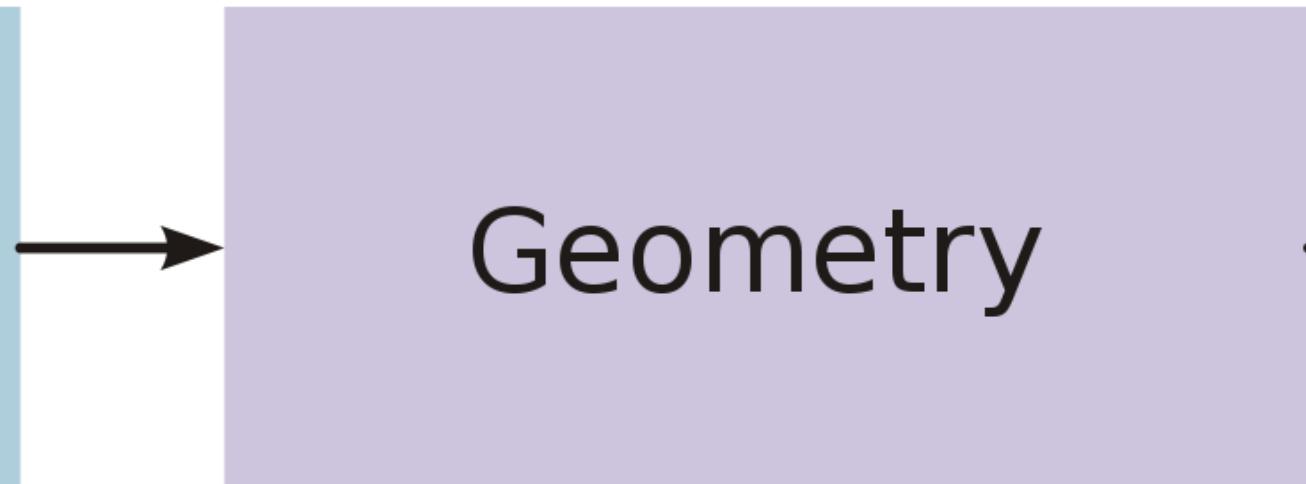


How Computers Draw

High level scene description:
- 3D environment (games)
- Chart data (visualization)
- ...



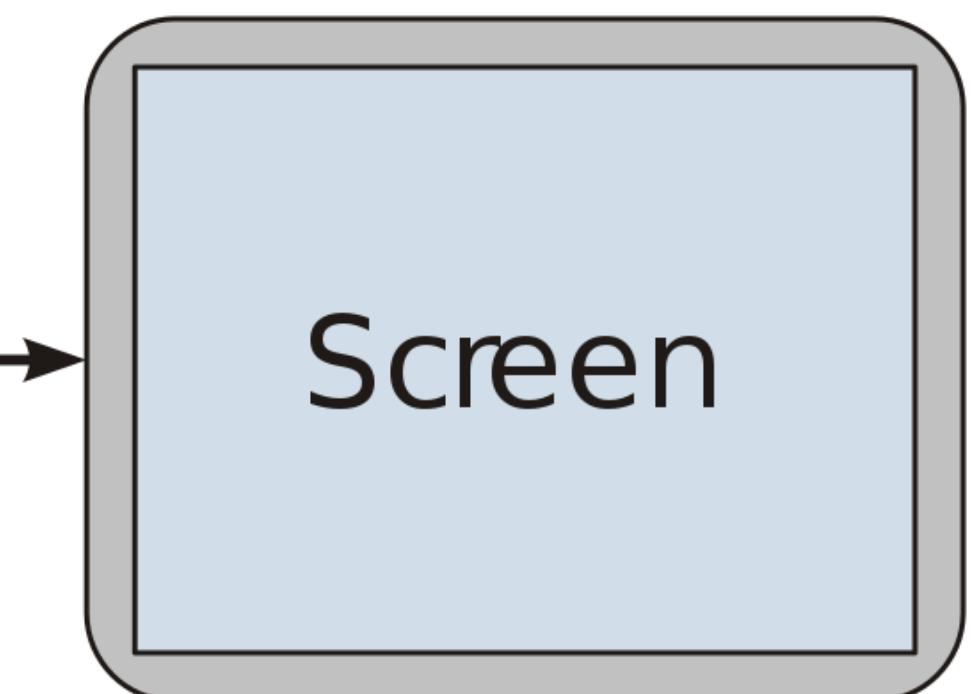
Scene is broken down into geometric primitives:
- points, lines, triangles
- shapes, paths
- ...



Geometric primitives are converted to drawable representations:
- vector -> raster for pixel images
- PDF, PS for print
- ...



Final image on output device:
- monitor, phone, tablet
- HMD
- paper



matplotlib

three.js

Bokeh

ggplot2

VTK

plotly

d3

unity



OpenGL

WebGL



Hardware

Tight coupling between drawing and interaction

A Typical Graphics Pipeline

Takeaways

- Data visualization is an involved process. Please pay attention to the essential ingredients: *data semantics, sensory visual encodings, analysis tasks, human in the loop.*
- From an implementation perspective, we'll largely focus on high-level APIs. Many of these APIs let us customize visualizations in a variety of ways.
- Interaction is an integral component of the process. GPUs are now widely used behind-the-scenes.