20181113-introduction

Visualization Goals

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

Some analysis problems are ill-specified. People are unclear what questions to ask upfront or how to answer them. Visualization is used to **augment human capabilities**.

Visualization has 3 goals: to **explore**(e.g. data exploration), to **analyse**(verify hypotheses), to **present**(communication of results).

Scientific Visualization

It is the visualization of spatial data.

Data sources are:

- Medical data: MedVis-VolVis
- Flow data: FlowVis
- GIS data
- Microscopic data (molecular physics)
- Macroscopic data (astrononomy)

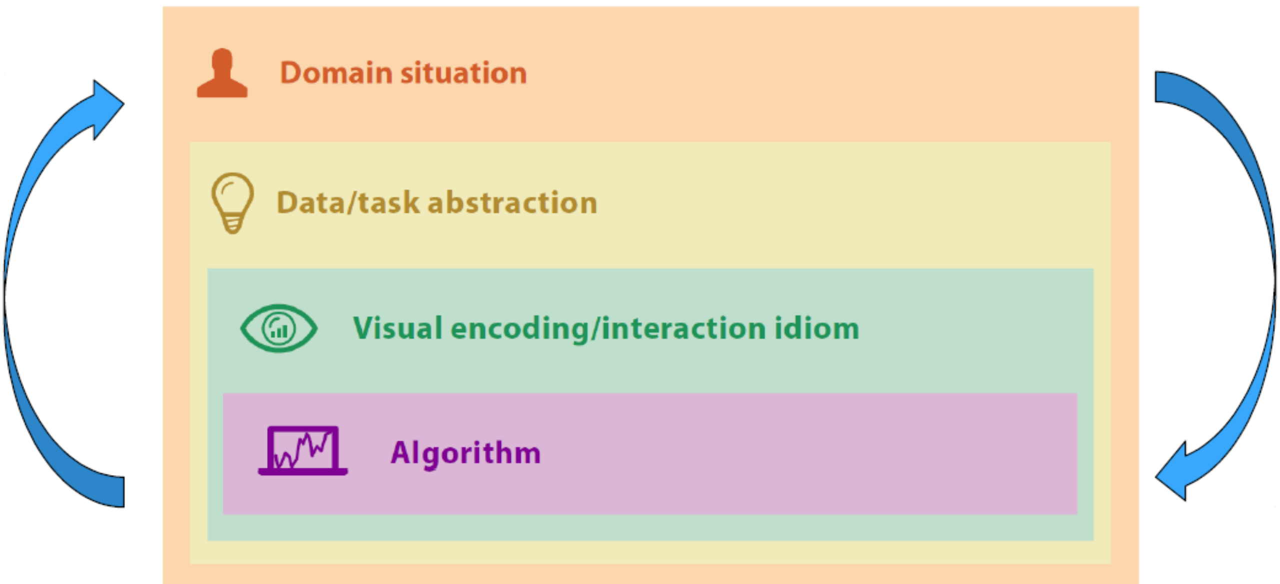
SciVis vs InfoVis

InfoVis	SciVis (VolVis/MedVis/FlowVis...)
Abstract Data No spatial reference.	Spatial Data Spatial reference.
N-dimensional. Heterogeneous.	Mostly 2 or 3-dimensional
Numerical, text, images, multimedia.	Scientific, engineering, biomedical.

Visual Analytics
Spatial and *Abstract* data.
N-dimensional space – hetorogeneous
Combination with pattern recognition, statistical data analysis machine learning

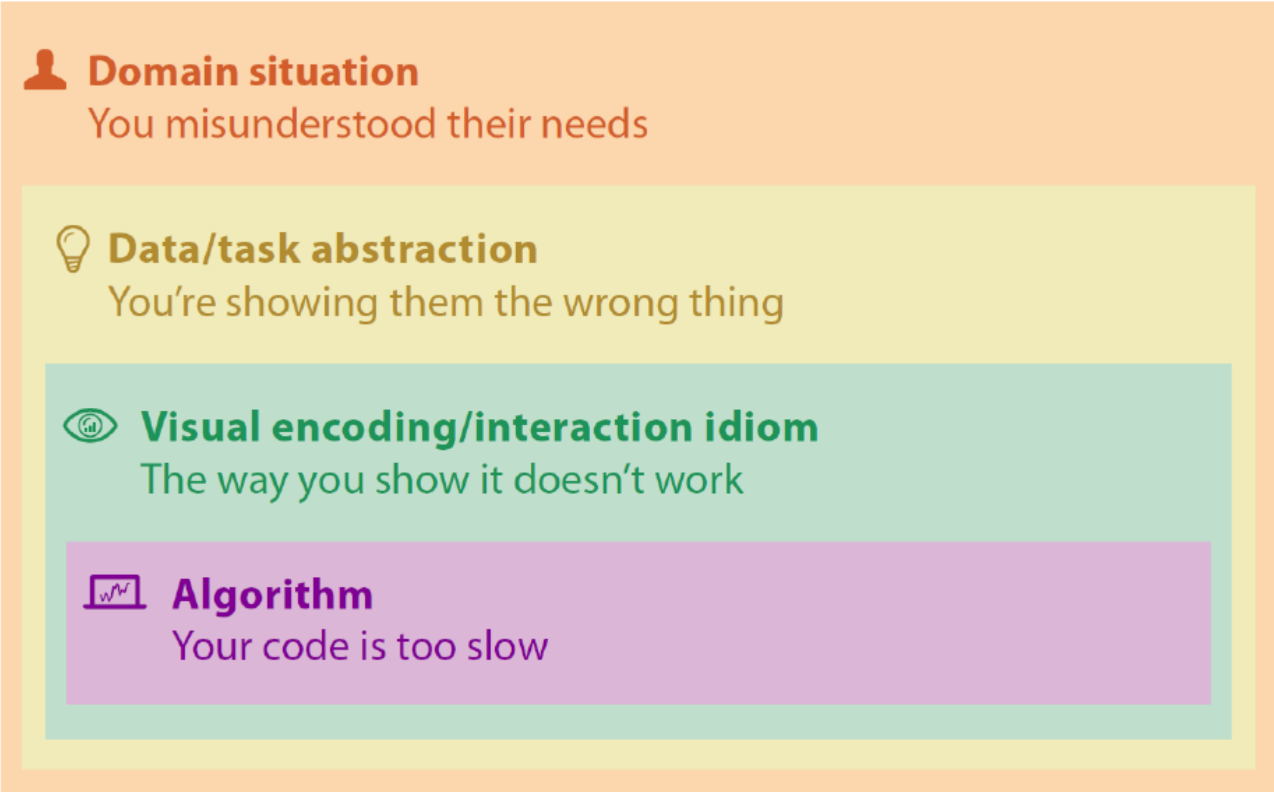
Visualization Design

Design Framework

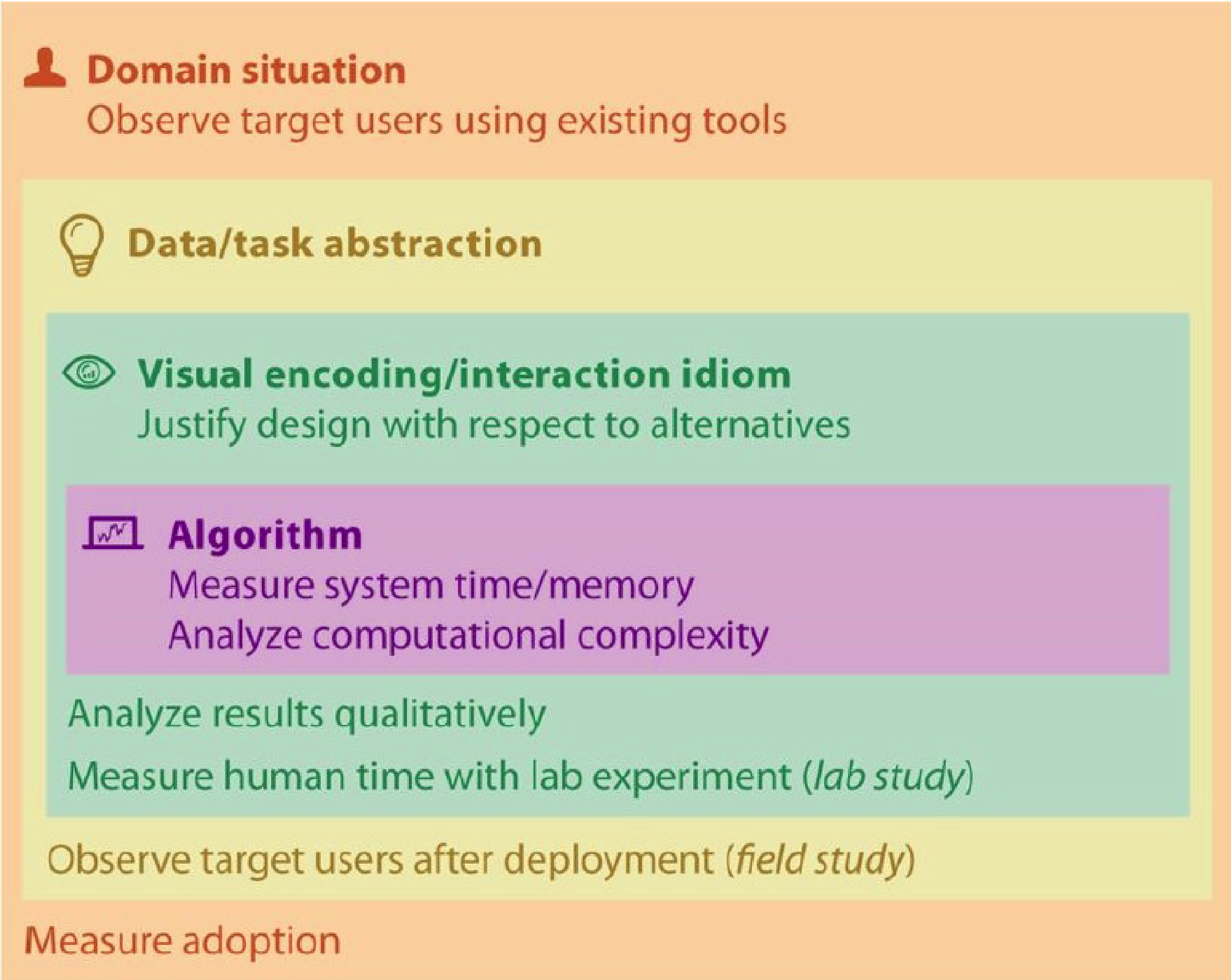


An iterative/refinement process.

Validation


Domain situation
You misunderstood their needs
Data/task abstraction
You're showing them the wrong thing
Visual encoding/interaction idiom
The way you show it doesn't work
Algorithm
Your code is too slow

Design Goal



Analysis framework

Anslysis framework contain three steps:

- **data abstraction** (What is shown?)
- **task abstraction** (Why is the user looking at it?)
- **visual encoding and interaction** (How is it shown?)

Data Abstraction: What?

It is the focus of visualization, everything is centred around the data. It is the driving factor (besides user and tasks) in choice and attribution of the visualization technique. Some important questions:

- Where do the data "live"(data space)
- Type of the data
- Which representation/visual encodings makes sense (secondary aspect)

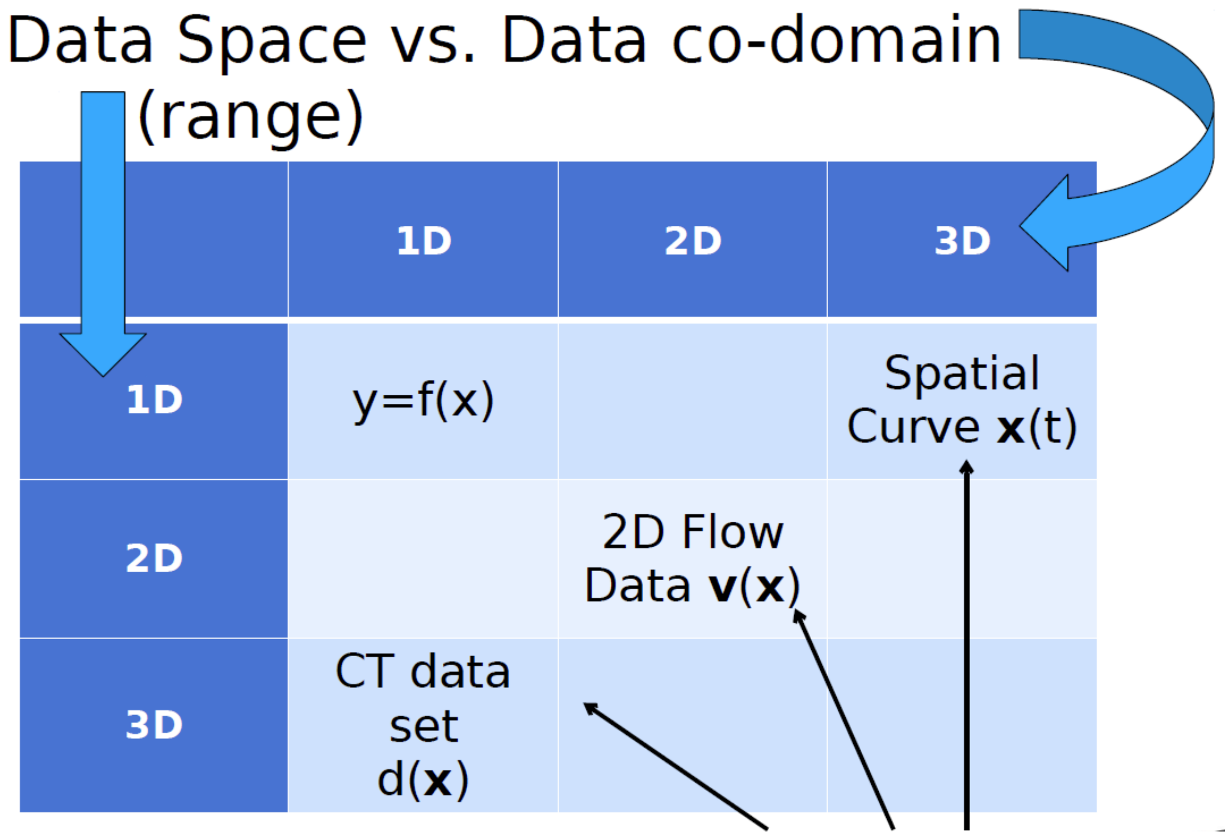
Data Space

- Inherent spatial domain (**SciVis**):
  - 2D/3D data space given

    - Examples: medical data, flow simulation data, GIS-data
- No inherent spatial reference (**InfoVis**)
    - abstract data, spatial arrangement through visualization
    - Example: databases
- Aspects: dimensionality (data space), coordinates, region of influence (local, global), domain

Data Types

- Scalar: numerical value: (natural, whole, rational, real, complex numbers)
- Non numerical: (nominal, ordinal values, ...)
- Multidimensional values: (n-dim, vectors, nxn dim, tensor data)
- Multimodal values: vectors of data with varying type (e.g., row in a table)
- Aspects: dimensionality, co-domain(range)



data	description	example
N1→R1	series of values	bar chart, pie chart
R1→R1	function	graph
R2→R1	funct. over R2	2D-heightfields in 3D,contour lines in 2D

data	description	example
R2→R2	2D-vector field	hedgehog plot, LIC, streamlets, etc.
R3→R1	3D-Density values	iso-surfaces in 3D, volume rendering
(N1→)Rn	tuple quantities	parallel coordinates, (multi-attribute data) glyphs, icons, etc.

2018/11/20 InfoVis basics

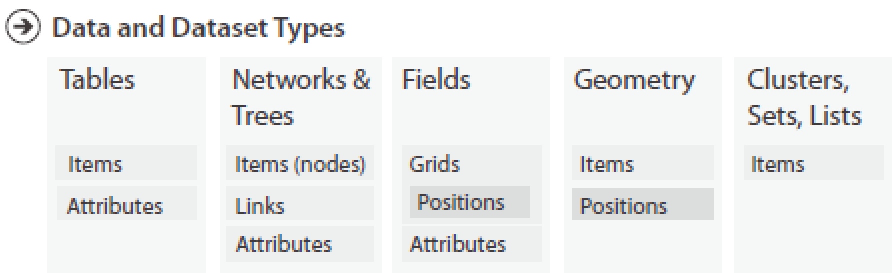
Data Abstraction: What?

Data Types

Five basic data types: items, attributes, links, positions, and grids.

- attribute: specific property that can be measured, observed, or logged
- item: an individual entity that is discrete, such as a row in a simple table or a node in a network
- link: a relationship between items, typically within a network.
- grid: strategy for sampling continuous data in terms of both geometric and topological relationships between its cells.
- position: spatial data

Dataset Types



Dataset is any collection of information that is the target of analysis. The four basic dataset types are tables, networks, fields, and geometry.

- networks:

  well suited for specifying that there is some kind of relationship between two or more items

- field:

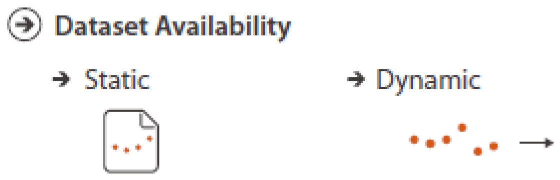
  contains attribute values associated with cells.

  Each cell in a field contains measurements or calculations from a continuous domain.

  the table and network datatypes are an example of discrete data where a finite number of individual items exist, and interpolation between them is not a meaningful concept.
- geometry:

  specifies information about the shape of items with explicit spatial positions

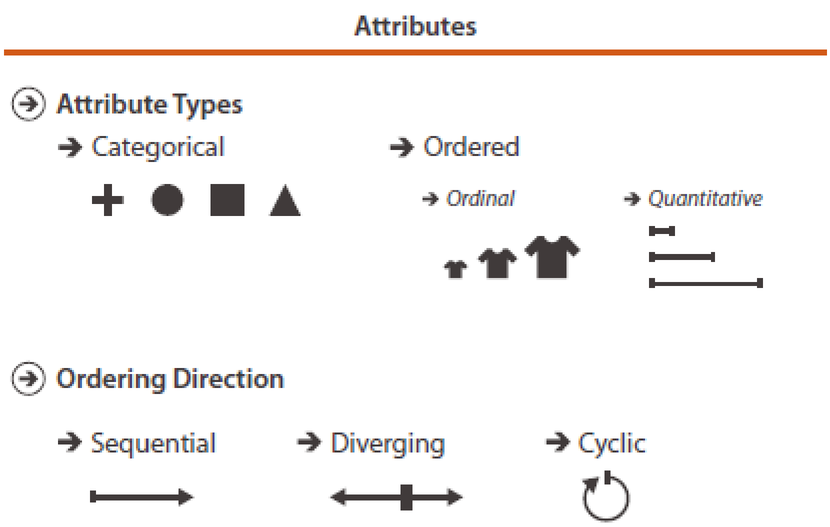
Dataset Availability



static file/offline: entire dataset is available all at once

dynamic streams/online: dataset information trickles in over the course of the vis session

Attribute Types



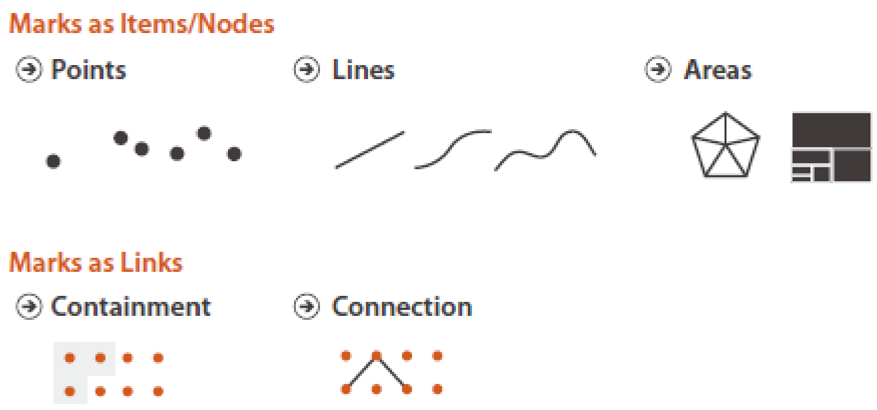
- categorical: does not have an implicit ordering
- ordered: have an implicit ordering
  - ordinal序数: a well-defined ordering e.g. size, ranking
  - quantitative: supports arithmetic comparison e.g. height, weight, temperature, stock price
  - ordering direction:
    - sequential: a homogeneous range from a minimum to a maximum value

    - diverging: two sequences pointing in opposite directions that meet at a common zero point e.g. sea level, elevation
    - cyclic: values wrap around back to a starting point rather than continuing to increase indefinite e.g. day of the week, month of the year

Visual Encoding: How?

Mark: a basic graphical element in an image

e.g. a point in 0D, a line in 1D, an area in 2D, a volume in 3D



In network dataset:

- node mark: represent an item
- link mark: represent a relationship between items
    - connection mark: a pairwise relationship between two items, using a line
    - containment mark: shows hierarchical relationships using areas

Channel: a way to control the appearance of marks

- **identity** channel: what something is or where it is
- **magnitude** channel: how much of something there is.

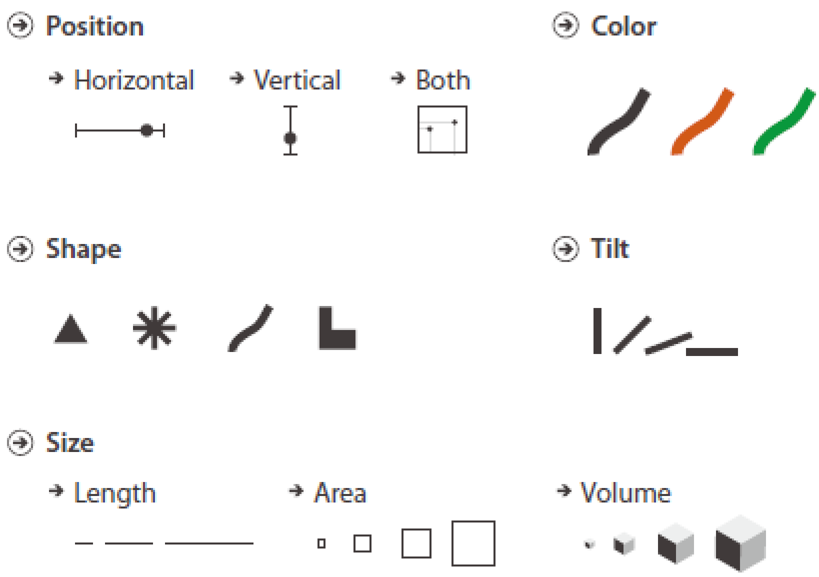


Figure 5.3. Visual channels control the appearance of marks.

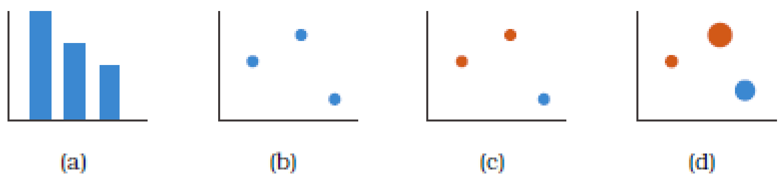


Figure 5.4. Using marks and channels. (a) Bar charts encode two attributes using a line mark with the vertical spatial position channel for the quantitative attribute, and the horizontal spatial position channel for the categorical attribute. (b) Scatterplots encode two quantitative attributes using point marks and both vertical and horizontal spatial position. (c) A third categorical attribute is encoded by adding color to the scatterplot. (d) Adding the visual channel of size encodes a fourth quantitative attribute as well.

Visualization principles

- **Expressiveness** principle:

  The visual encoding should express all of, and only, the information in the dataset attributes.

- **Effectiveness** principle:

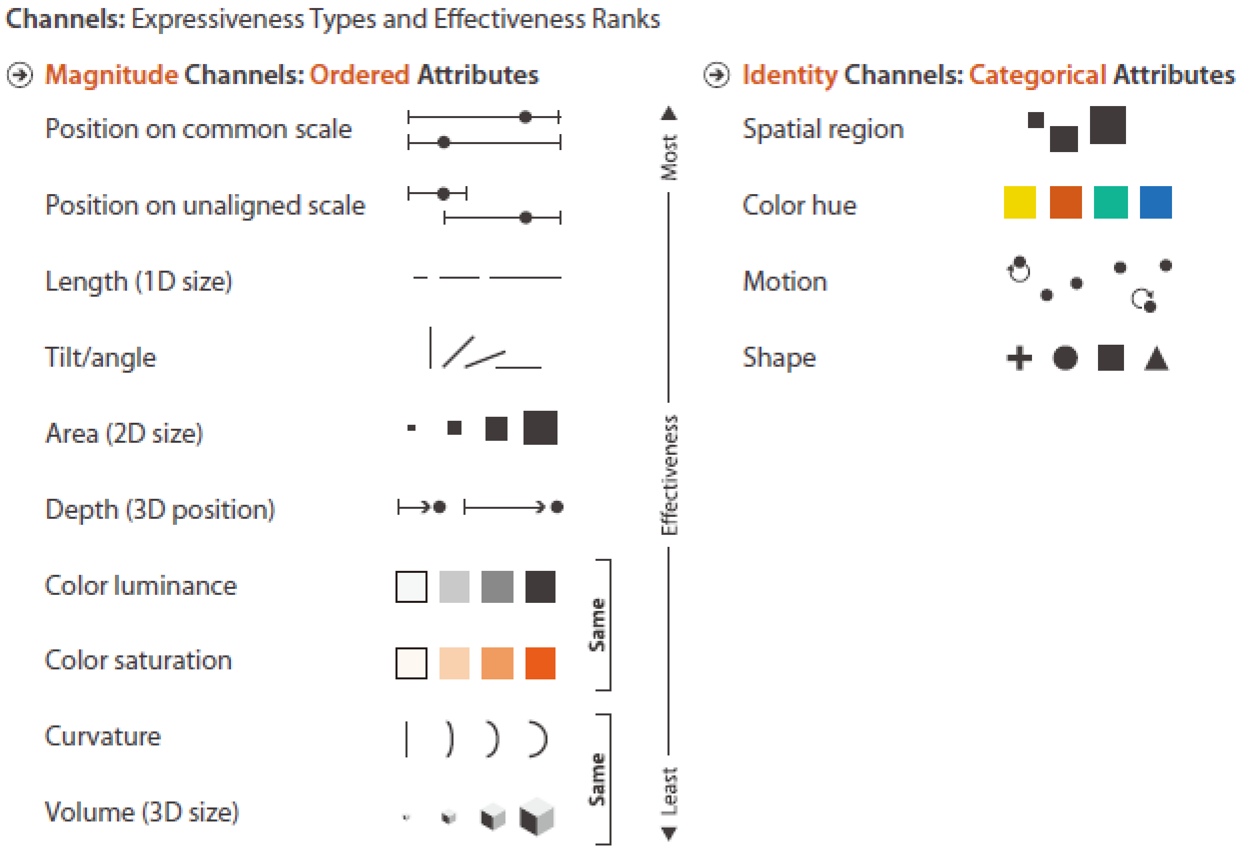
  Effectiveness principle dictates that the importance of the attribute should match the salience of the channel.

  In other words, the most important attributes should be encoded with the most effective channels in order to be most noticeable.

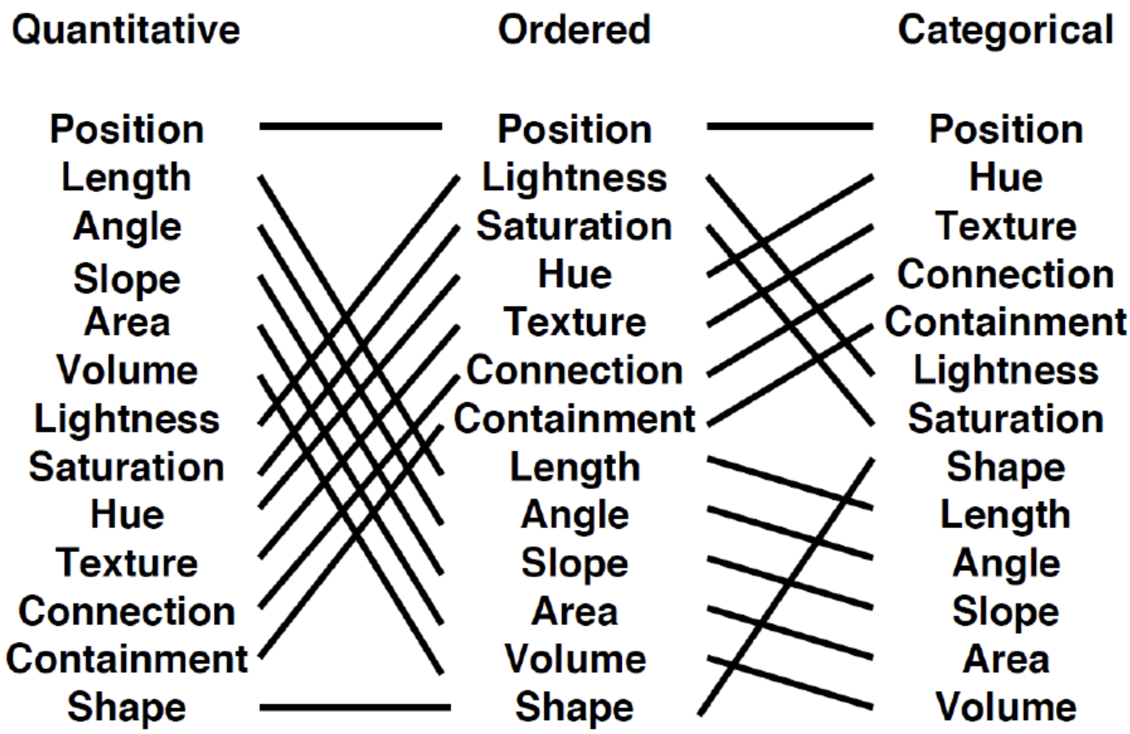
- Channel Ranking:

  Ordered attributes should be shown with the magnitude channels.

  Categorical attributes should be shown with the identity channels.



Visual Channel Rankings (Mackinlay)



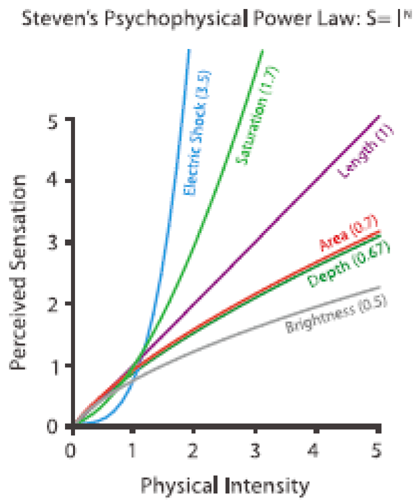
Mackinlay, Automating the Design of Graphical Presentations of Relational Information, ACM TOG 5:2, 1986]

Channel Effectiveness

Criteria: accuracy, discriminability, separability, popout

Accuracy: how close is human perceptual judgement to some objective measurement of the stimulus.

Some sensations are perceptually magnified compared with their objective intensity (when n > 1) and some compressed (when n < 1). Length is completely accurate.



Discriminability: if you encode data using a particular visual channel, are the differences between items perceptible to the human as intended? The characterization of visual channel thus should quantify the number of bins that are available for use within a visual channel, where each bin is a distinguishable step or level from the other.

Separability: If visual channels as completely independent from each other

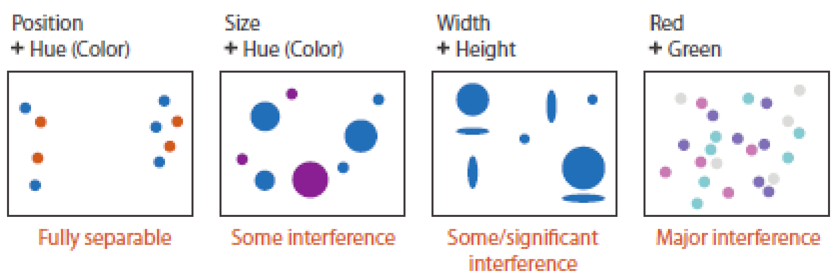


Figure 5.10. Pairs of visual channels fall along a continuum from fully separable to intrinsically integral. Color and location are separable channels well suited to encode different data attributes for two different groupings that can be selectively attended to. However, size interacts with hue, which is harder to perceive for small objects. The horizontal size and and vertical size channels are automatically fused into an integrated perception of area, yielding three groups. Attempts to code separate information along the red and green axes of the RGB color space fail, because we simply perceive four different hues. After [Ware 13, Figure 5.23].

Popout: a distinct item stands out from many others immediately.

Color

Color Spaces

The color space of what colors the human visual system can detect is three dimensional.

- Devide-oriented: corresponds to physical realisation(printer, screen)
  - RGB(triples of red, green, and blue)

    Advantage: computationally convenient

    Disadvantage: very poor match for the mechanics of how we see. The red, green, and blue axes of the RGB color space are not useful as separable channels
- Intuitive: Natural (Semantic properties of color)
  - HSL(hue–saturation–lightness)

    hue: pure colors that are not mixed with white or black (**identity** channel)

    saturation/chroma: the amount of white mixed with that pure color (**magnitude** channel)

    lightness: the amount of black mixed with a color
  - HSV(hue–saturation-value) value: stands for grayscale value and is linearly related to L
  - Disadvantage: It doesn't truely reflect how people perceive color. The lightness is wildly different from how we perceive L luminance.

- Perceptually uniform: Euclidean distance corresponds to perceptual difference
  - L∗a∗b∗

    L∗/**Luminance**: single black and white luminance channel, a nonlinear transformation of the luminance perceived by the human eye. (**magnitude** channel)

    a∗ and b∗: two color axes.

Colormap

A colormap specifies a mapping between colors and data values.

- categorical/quantitative: encode categories good choice of discriminable colors: opponent colors

- ordered: encode ordinal or quantitative attributes
    - sequential
    - diverging: have two hues at the endpoints and a neutral color as a midpoint

      Problems: how many hues used in continuous colormaps?

      Rainbow colormap:

      Advantage:Easy to tell difference between subranges

      Disadvantage: **perceptually unordered, perceptually nonlinear, perceptual borders that are not there**

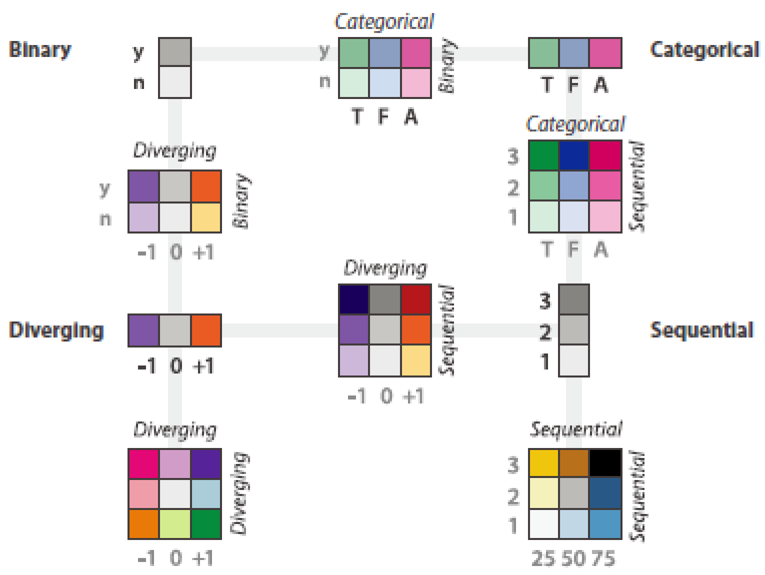
        - hue is used to indicate color, despite being an identity channel without an implicit perceptual ordering
        - the scale is not perceptually linear: steps of the same size at different points in the colormap range are not perceived equally by our eyes
        - fine detail cannot be perceived with the hue channel

      Alternatives:

        - **monotonically increasing luminance colormap**

          multiple hues are ordered according to their luminance from lowest to highest

        - segmented rainbows



Visualization Advice and Guidelines

Dangers of depth

Visualization of 3D spaces comes with some difficulties:

- Occlusion
- Interaction complexity
- Perspective distortion: interferes with all size channel encodings
- text legibility

Resolution beats Immersion

Eyes beat Memory

It is hard to compare visible items to memory of what you saw.

Animations:

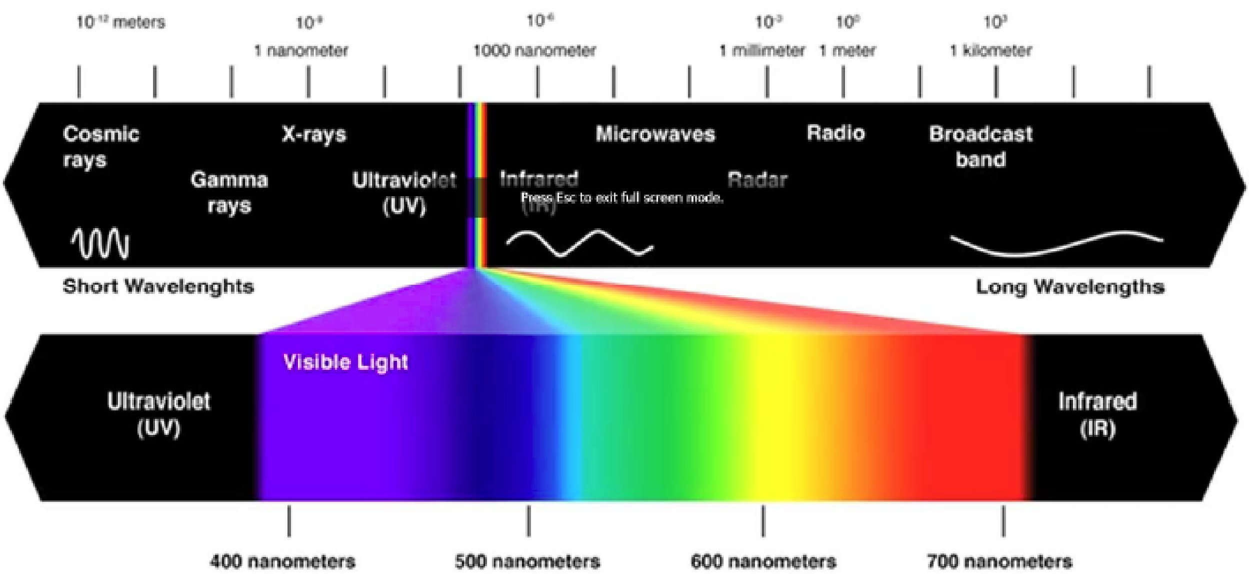
- Great for choreagraphed storytelling (localized action)
- Great for transition between two states/ data sets
- Give control to the user on the animation(re-play, stop, pause)
- Poor for many states with changes everywhere(**Change blindness**)

2018/11/23 Human Perception

See colors

Light Specturm

People observed spectrum mapped from LA to NY.



Eye biology

There are two types of photoreceptors in the human retina, rods and cones.

Cones: Chromatic perception

Rods: Achromatic perception, Low-light vision

At night, because cones shut off, people see all things in grey.

During the day, rods shut off.

Represent Colors

color blindness: People usually have three types of cones cells, some people only have two.

Color spaces:

- **RGB space**:

  Advantages: Extremely simple, Measured per device, Direct connection to hardware

  Disadvantages: For artists not easy to produce wanted colors, Not device independent

  Device dependent: the description of color information is related to the characteristics of a particular device

  Device independent: the description of color information is not dependent from the characteristics of a particular device
- HLS(Hue, Lightness, Saturation): converted from RGB

  Advantages: easy to mix colors
- XYZ

  Advantages:
  - XYZ covers all colors with positive coordinates
  - Easy conversion: Conversion to RGB is linear
  - Device independent

Disadvantages: a rather mathematical construction

Dependent: RGB, HVS

Independent: XYZ, Lab, luv

MacAdam ellipse: a region on a chromaticity diagram which contains all colors which are indistinguishable

2018/11/27 Interaction Graphs: High-Dimensional data

2018/12/04 Finalize InfoVis

Data Reduction

- Filter: Pro: straightforward/intuitive Cons: out of sight, out of mind
- Aggregation: Pro: inform about whole set Cons: details lost

  e..g Boxplot

- Dynamic query: Selection/filtering by pointing (not typing)

  Pro: Immediate and continuous display of results, Promote exploration

Dimensionality Reduction

- Linear: Resulting attributes are linear combination of existing attributes (interpretable)
  - Principal Component Analysis (PCA)
  - Linear Discriminant Analysis (LDA)
- Non Linear: Resulting attributes do not have straightforward relation to original attributes
  - Multi-Dimensional Scaling (MDS) -preserve distances
  - t-Distributed Stochastic Neighbor Embedding (t-SNE) –preserve neighbourhods

Static trees Visualization

- Node-link diagram

  Task: understand topology, follow paths

  Scale: 1000-10000 nodes

  Pro: Intuitive, Good at exposing structure of information Cons: A lot of empty space

- Enclosure: Space-filling diagram
  - Treemap

    Task: query attributes at leaf nodes

    Scale: million nodes and links

    Pro: Very good usage of available space –show attributes Cons: Difficulty distinction of the hierarchy (implicit)

- Combined representations

Static Graphs Visualization

- Node-link
  - Force-directed algorithms
- Adjacency Matrix
- Combination

Compound Graphs: Graphs with a hierarchy

2018/12/11 Volume Viusalization

2018/12/18 Flow Field Visualization

Flow Visualization

- Gaseous flow: development of cars, aircraft, spacecraft design of machines -turbines, combustion engines
- Liquid Flow: naval applications -ship design civil engineering -harbor design, coastal protection
- Medicine: blood flow

Vector-Field Visualization

Data set is given by a vector component and its magnitude.

- Glyphs

  Pro: Show full local information and it is intuitive

  Cons: Just shows local information, Cluttering extended to 3D (not really useful)

- Particle Tracing

  Visualizing the flow directions by releasing particles and calculating the trajectories based on the vector field

  - pathline: the path it takes through space as a function of time
  - streamline: a line which is everywhere parallel to the local velocity vector
  - streakline: concentrate on fluid particles that have gone through a fixed station or point. At some instant of time the position of all these particles are marked and a line is drawn through them.

  Pro: Shows global information

  Cons: Error accumulation, Seeding –Region of Interest, Cluttering

- Texture Based Methods

  - Spot Noise
  - Line Integral Convolution
  - Texture Splats

  Pro: Shows detailed global information, Continuous, No Seeding Cons: Computational cost, Cluttering -3D problematic

2019/01/09 Stereo and VR