Information Visualisation
(2) Data, Tasks and Validation

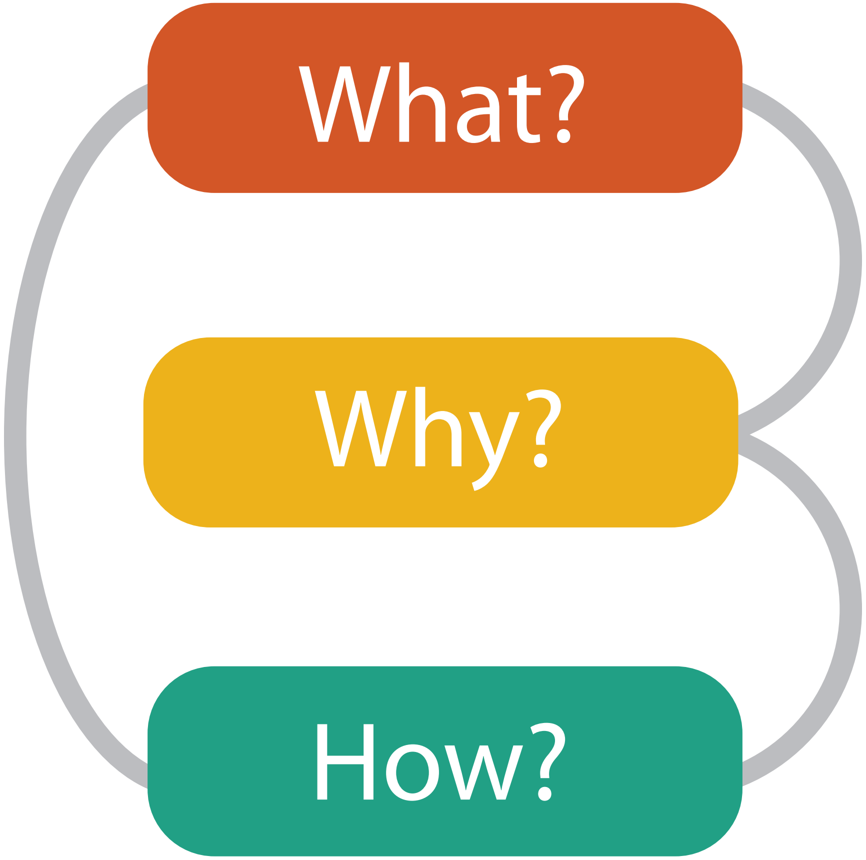
Prof. Bruno Dumas

UNIVERSITÉ DE NAMUR

Course Outline

- Introduction to Information Visualisation
- Data, Task and Validation
- Marks and Channels, Color Mapping
- Tables, Spatial data, Networks and Trees
- Manipulating View, Facetting, Focus + Context
- Reduce Items and Attributes + Some Cases Analysis

Where We Left Last Week…



Today's Program

- Data
  - Datasets
  - Attributes
- Tasks attributions
  - Why and who
  - Actions
  - Targets
- Analysis
- Threats and validation

1. Data

- Datasets
  - Data types
  - Data and dataset types
  - Dataset types
  - Dataset availability
- Attributes
  - Attribute types
  - Ordering Direction

Datasets: Data Types

➔ Data Types

➔ Items ➔ Attributes ➔ Links ➔ Positions ➔ Grids

- Item: individual discrete entity
  - Row in a table, node in a tree…
- Attribute: specific property that can be measured, logged
  - Aka variables or dimensions
- Link: relationship between items
- Position: spatial data (in 2D or 3D)
- Grid: strategy for sampling continuous data

Datasets: Data and Dataset Types

- Dataset: any collection of information that is the target of analysis
- Four basic dataset types: tables, networks, fields, geometry
  - + subtypes (clusters, sets, lists)
  - x Data types as seen before = ...

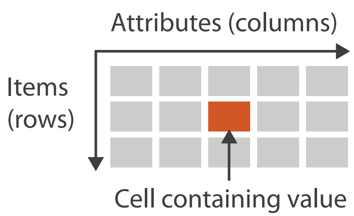
➔ Data and Dataset Types

Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

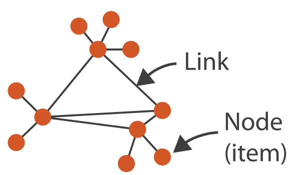
Datasets: Dataset Types

Dataset Types

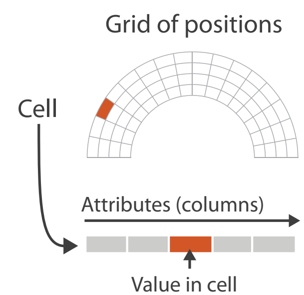
- Tables



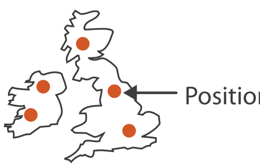
- Networks



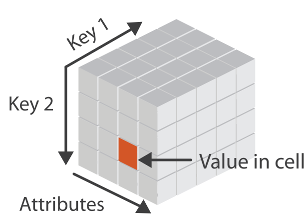
- Fields (Continuous)



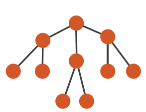
- Geometry (Spatial)



- *Multidimensional Table*



- *Trees*

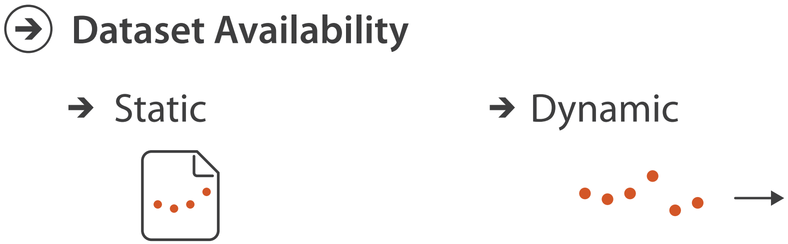


Datasets: Dataset Types

- Tables: rows and columns; can have multiple dimensions
- Networks (graphs) and trees: has nodes and links
  - Trees have hierarchies while networks don't
- Fields: cells that contain measurement from continuous domain
  - E.g. vector fields, medical data from a scanner, simulations...
  - Sampling !
  - Important in scientific visualisation (where the spatial position is *given* with the dataset) vs. Information visualisation (where use is space is *chosen* by designer)
- Geometry: items with explicit spatial positions

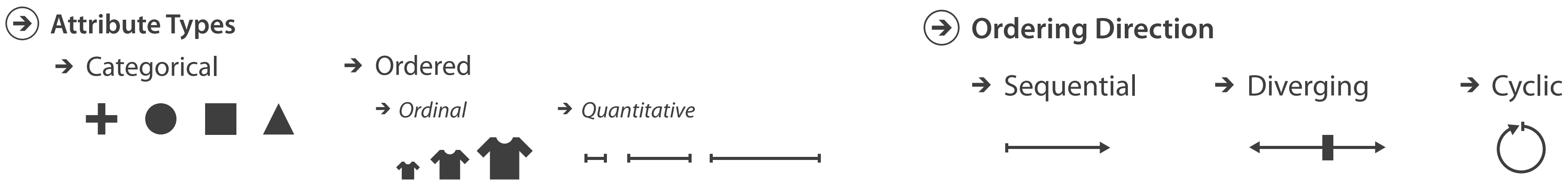
Datasets: Dataset Availability

- Datasets can be static as well as dynamic
- Static: data available all at once
  - Example: reading from a CSV file
- Dynamic: data changes over time
  - Example: database updated with new values while the visualisation is running



Attributes: Attribute Types

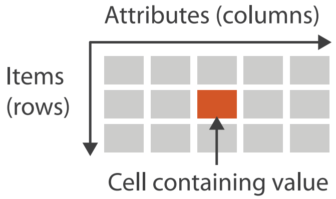
- Categorical vs. ordered
  - Categorical: no implicit ordering (fruits, names...)
  - Ordered: have an implicit ordering
- Ordered: ordinal vs. quantitative
  - Ordinal: no mathematics possible, but still ordered (e.g. shirt sizes)
  - Quantitative: measurements
  - Ordered data can be sequential or diverging; can also be cyclic



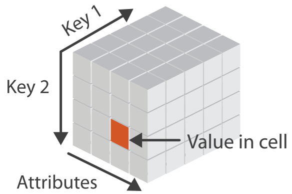
Data Semantics

- Type of attribute is one thing, its semantics is another
- Key semantics vs. value semantics
  - E.g. keys and values in an Excel sheet or a database
- Temporal semantics

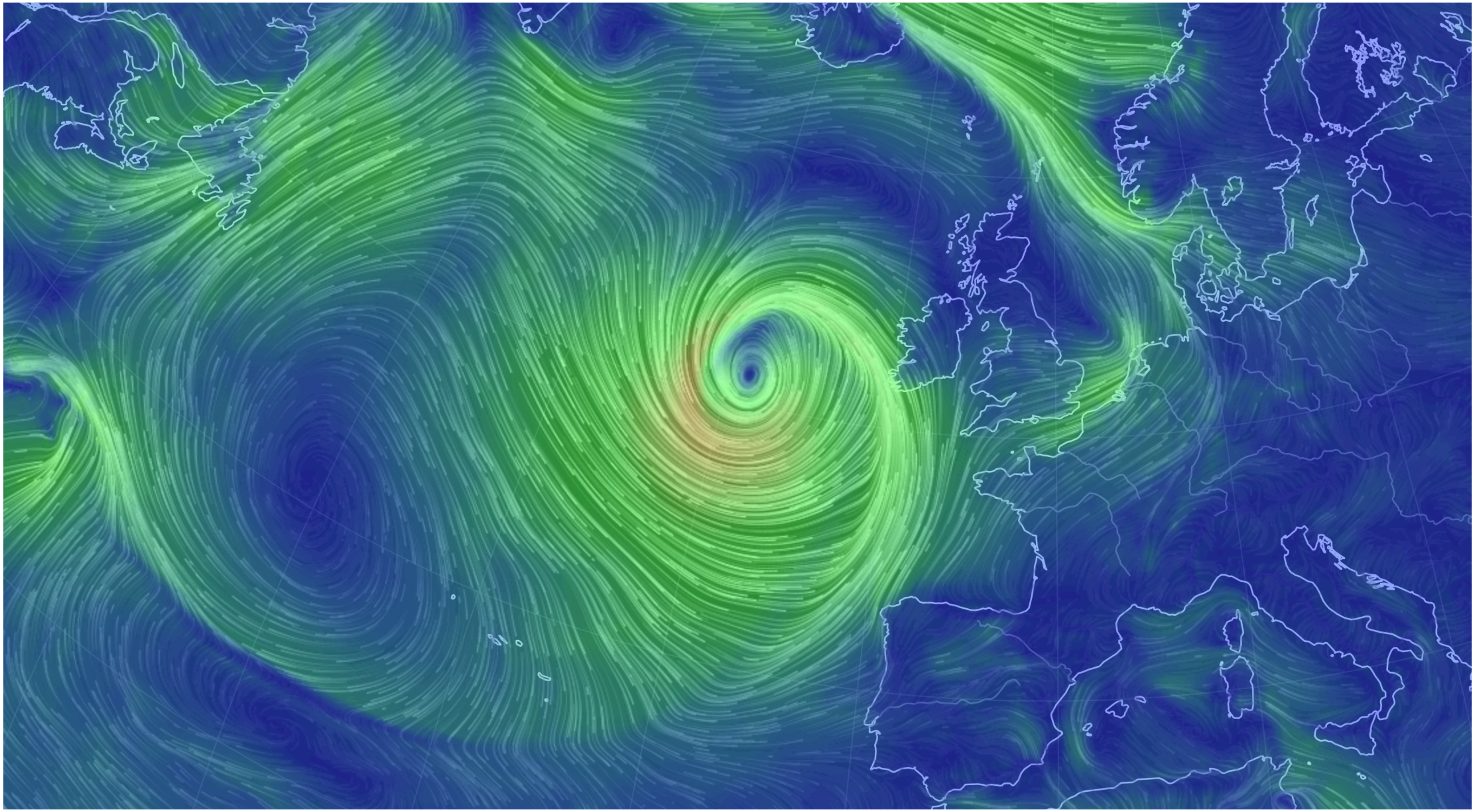
➔ Tables



➔ *Multidimensional Table*



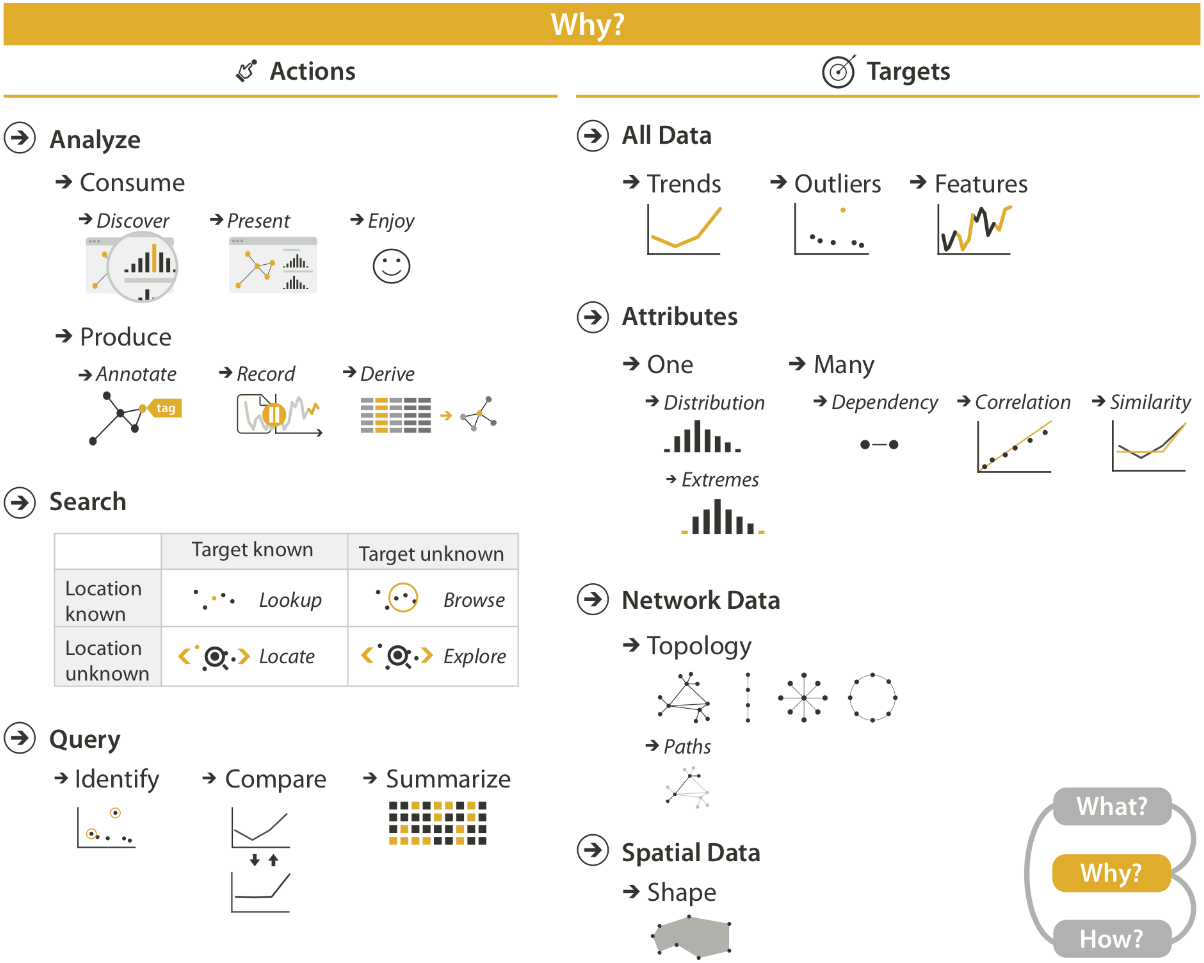
Test Time !



https://earth.nullschool.net/#current/wind/surface/level/orthographic=-0.61,47.21,1137/loc=26.765,35.377

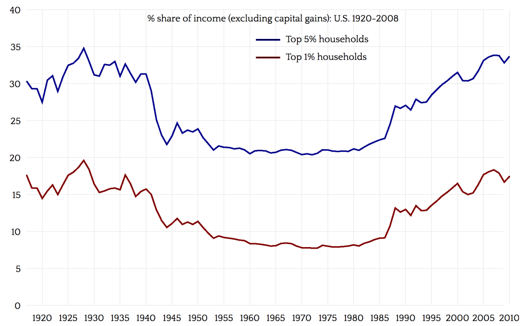
2. Task Abstraction

- Why analyse?
- Who?
- Actions (verbs):
  - Analyse
  - Produce
  - Search
  - Query
- Targets (nouns):
  - All data
  - Attributes
  - Network data
  - Spatial data



Why Analyse?

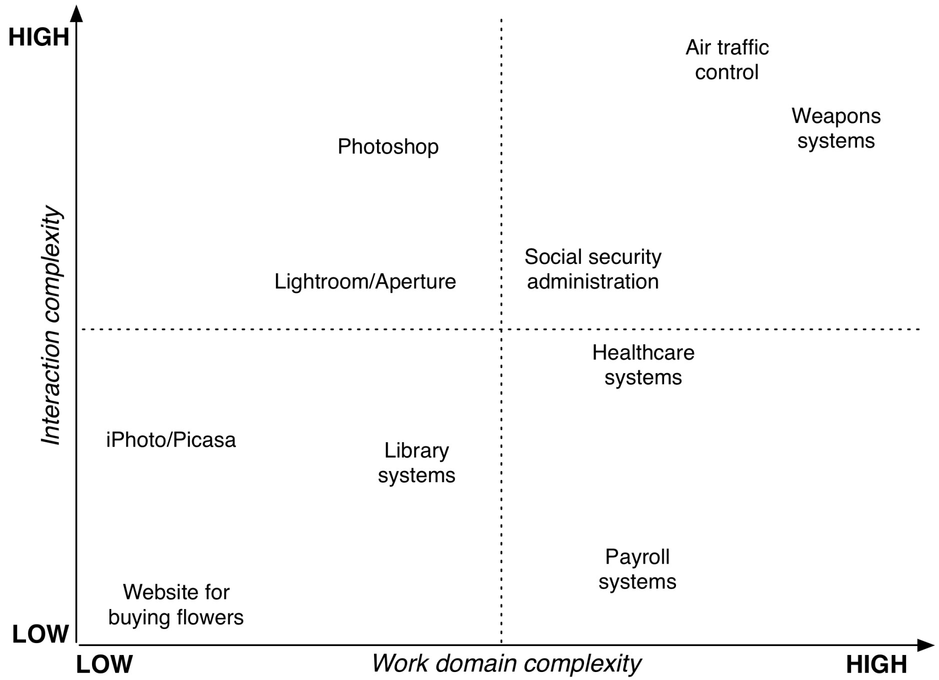
- From one domain-specific task…
  - E.g. Compare a group of patients who take some specific medication to another
- … To the generic underlying task
  - -> « compare values between two groups »
- … To the good (or: a good) vis idiom



- Goal of the framework: describe why people use visualisation
  - Helps abstracting from domain-specific vocabulary
  - Helps identify the correct idioms to use
  - Helps understand whether and how to transform the user's original data

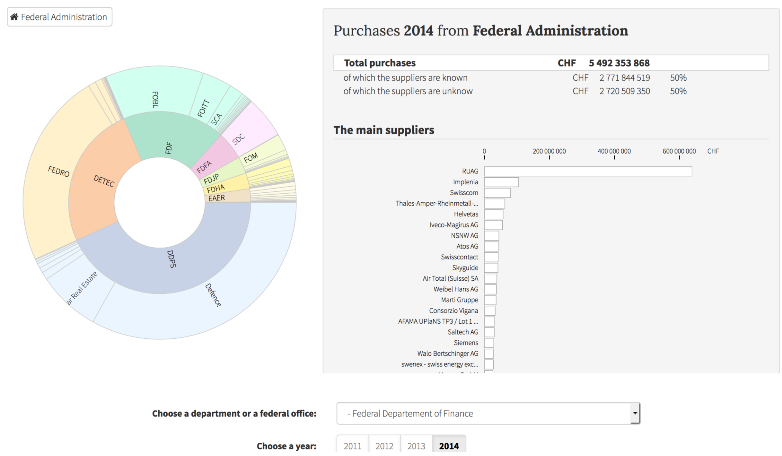
Who Is the Audience?

- Domain expert? Average web user? Namur resident? You?
  - Cf. *Interaction Homme-Machine* course (for those that followed it)



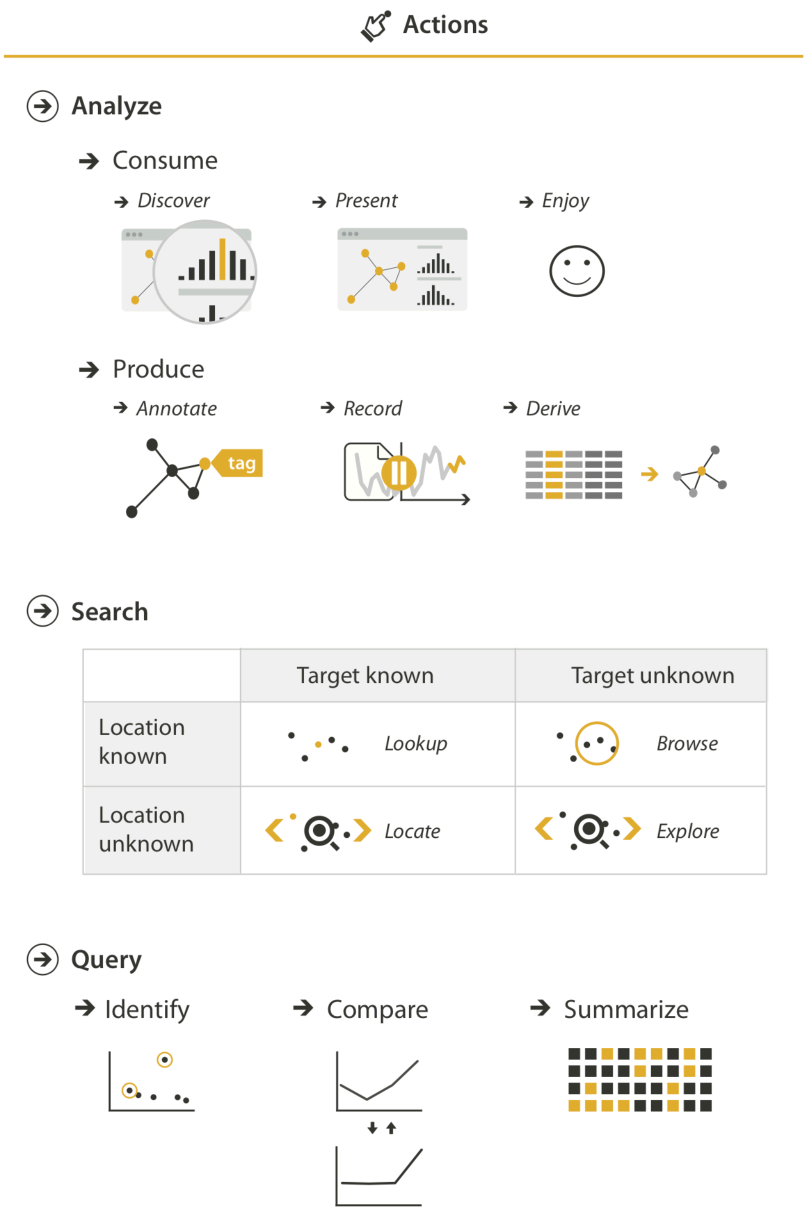
Who Makes the Choices?

- Part of the analysis: who will be the main user of the visualisation?
- Corollary of that question: who refines the design?
  - Filters data?
  - Selects representation?
  - Focuses on a specific part of the data?
  - Offline choices (designer) vs.
    Online choices (user)
  - ... Vs. both



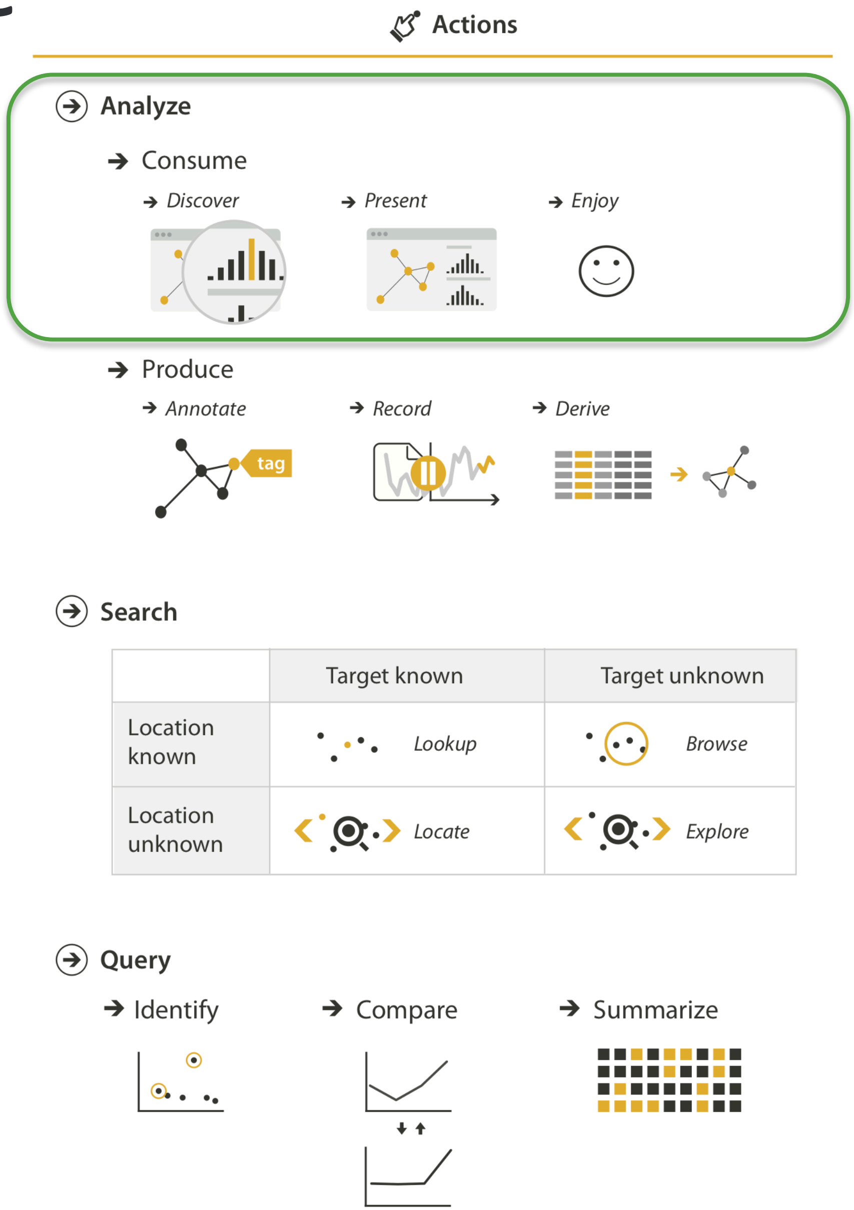
Actions (Verbs)

- Four type of actions:
  - Analyse
  - Produce
  - Search
  - Query



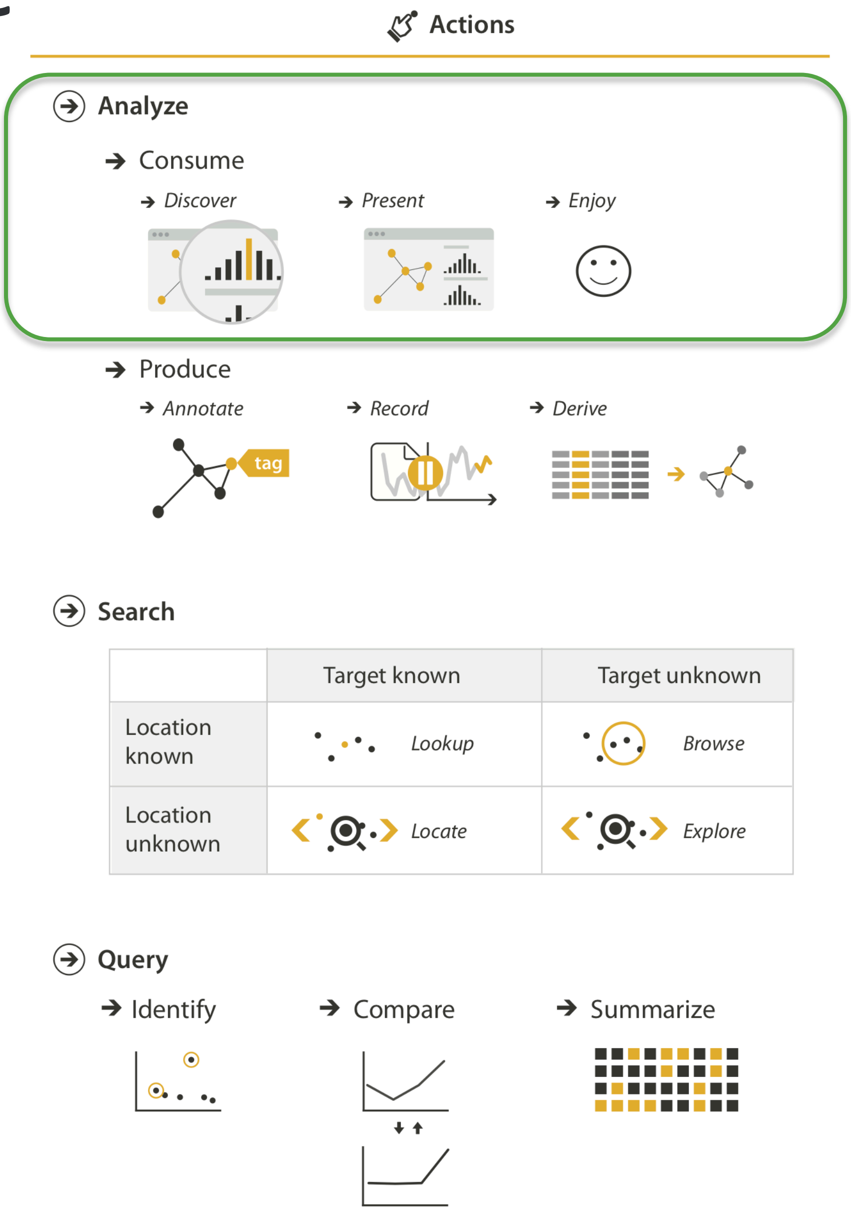
Actions: Analyse/Consume

- (Common) use case: *consume* existing information
- Discover: using vis to find new, not previously known knowledge
  - Generate new hypotheses
  - Verifying existing hypotheses
  - Frequently involves sophisticated interactive idioms
  - The designer doesn't know in advance what the user will look for!

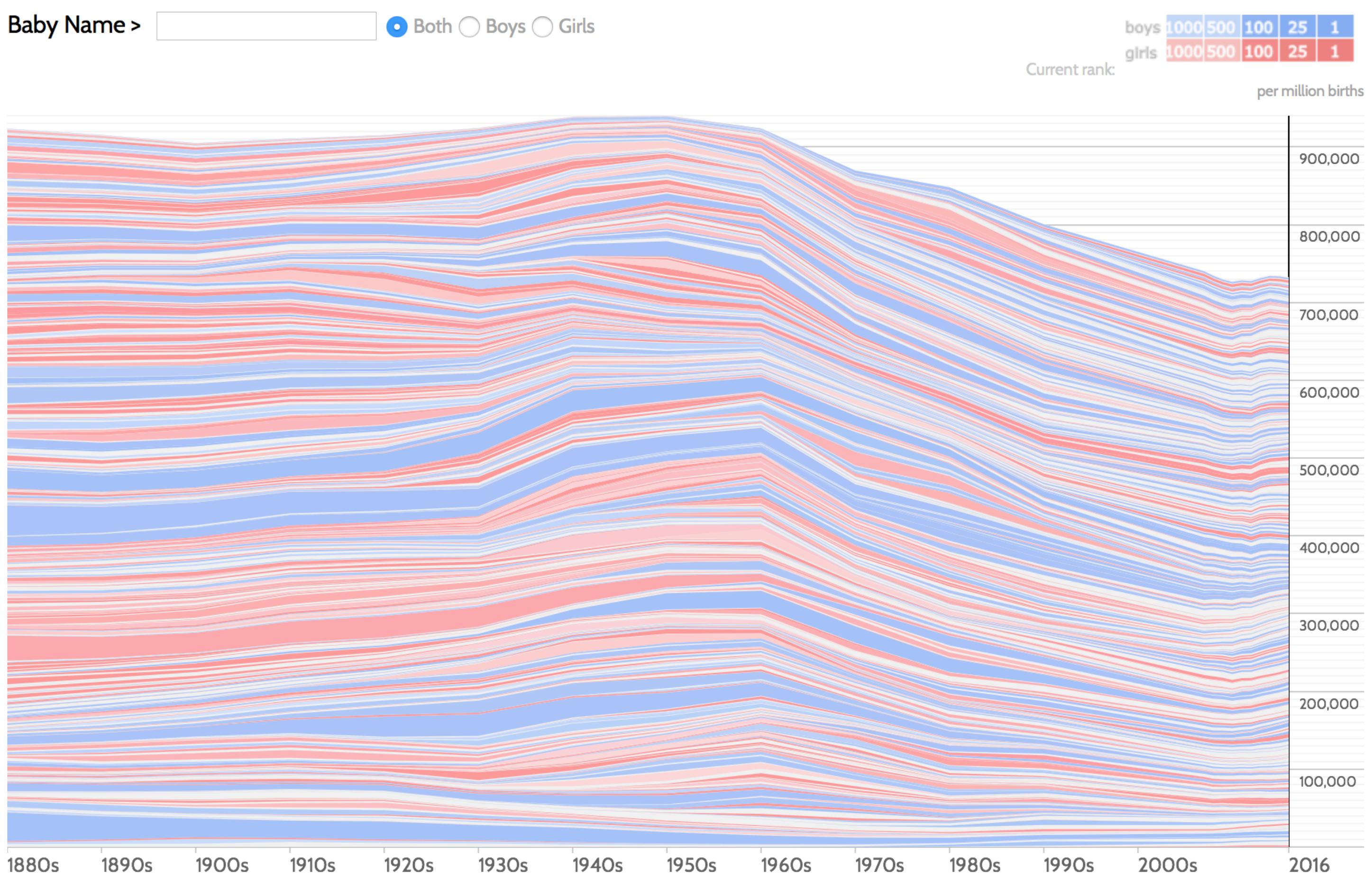


Actions: Analyse/Consume

- Present: using vis to communicate information
  - Communicate known hypotheses
  - Telling a story through data
  - Decision making, planning, teaching
  - Typical example: static diagram in a newspaper
- Enjoy: casually encountering/ browsing a vis for enjoyment
  - Not driven by a specific need
  - Casual or intensive use
  - Fleeting, maybe ill-defined goals !



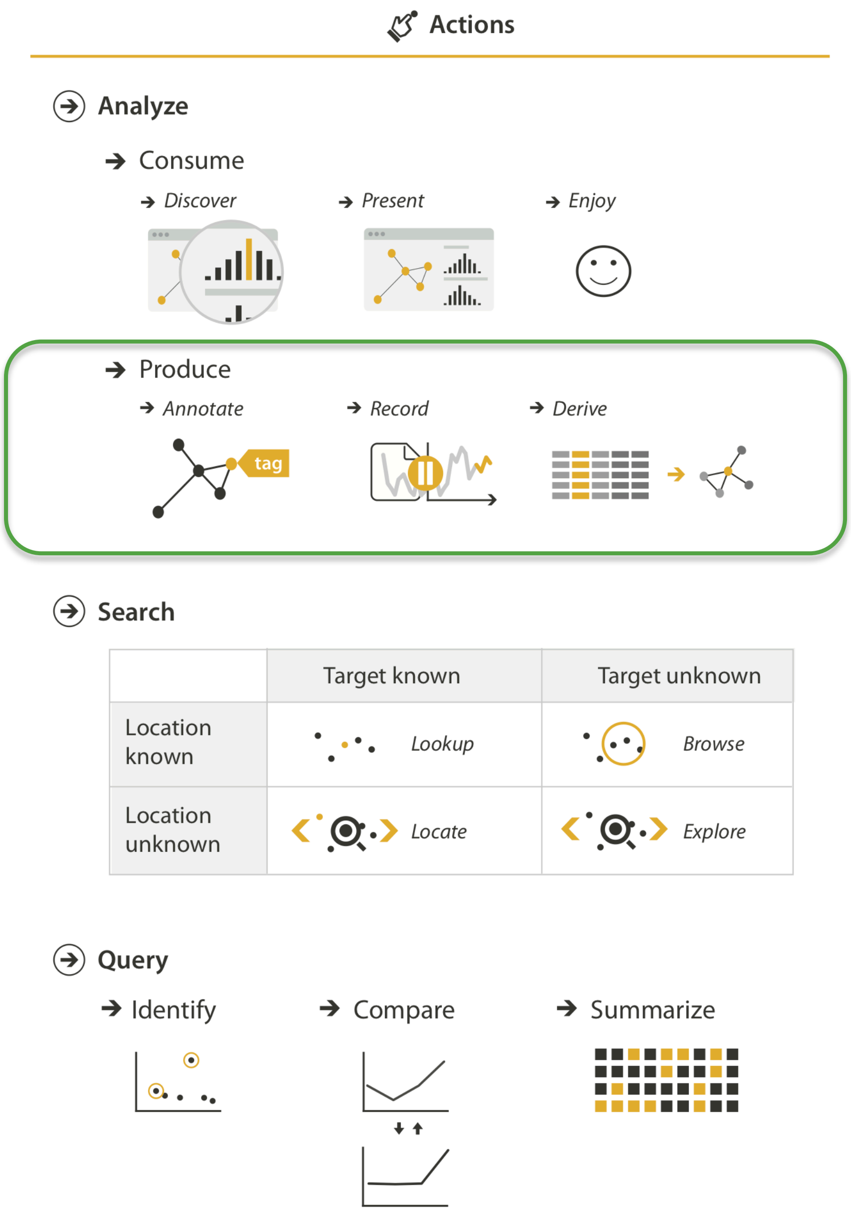
Enjoyment Example



http://www.babynamewizard.com/voyager#prefix=&sw=both&exact=false

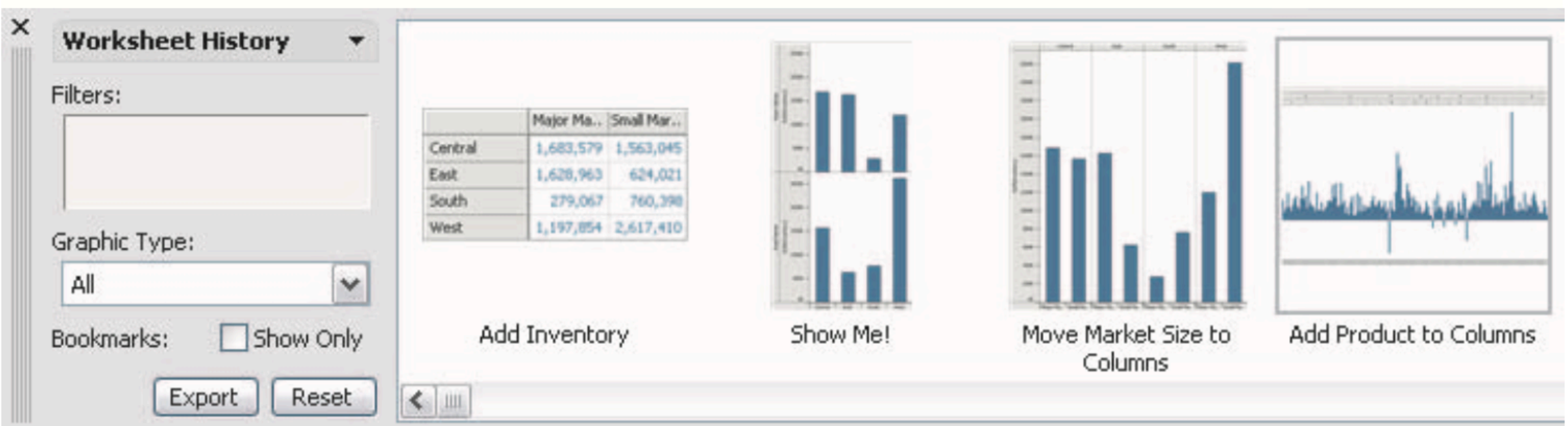
Actions: Produce

- Use case: generate new material
- Annotate: addition of graphical or textual annotations
  - Effectively new attributes on the data
- Record: capture visualisation elements as persistent artifacts
  - One artifact may feed the next
  - Example: graphical history (cf. next slide)



Record Example

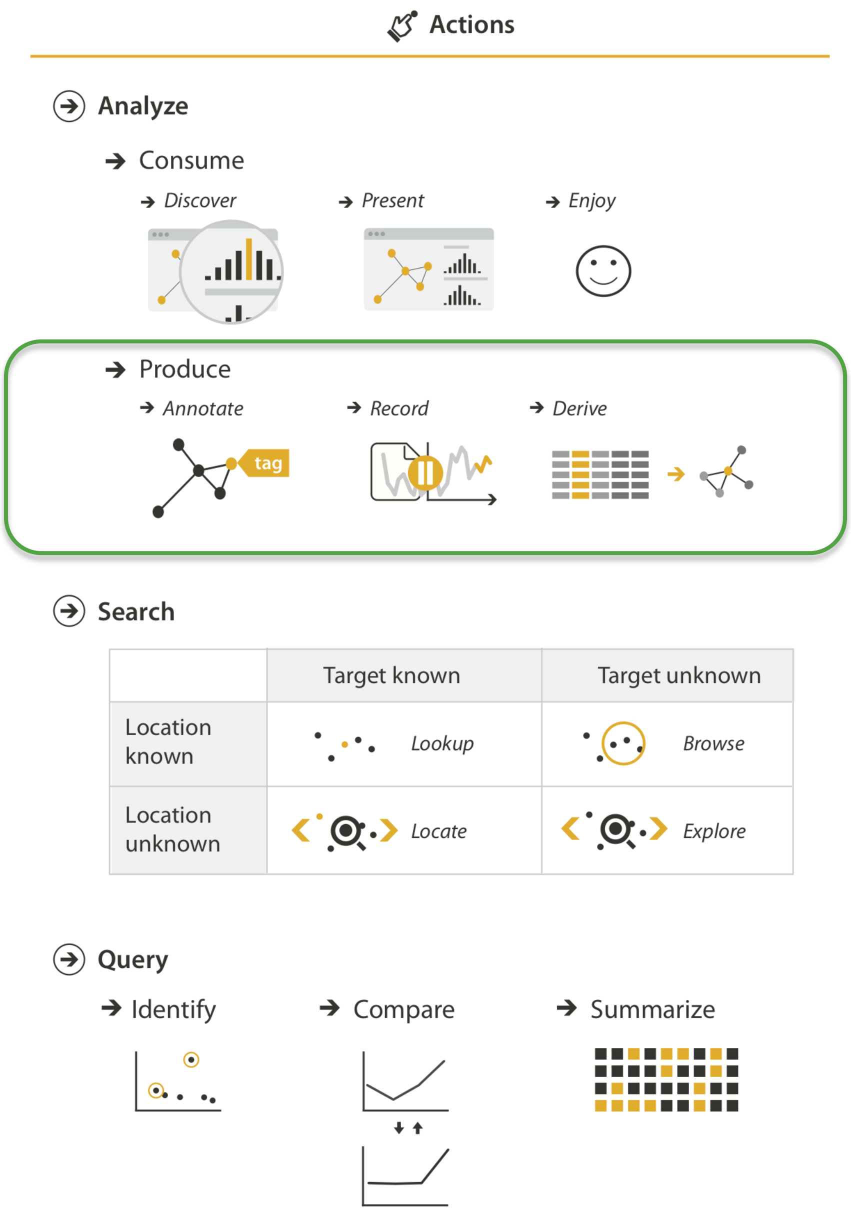
- Example generated using the Tableau tool
- Output of each task includes a static snapshot
- Sort of « meta-visualisation » showing the user's entire session



Heer, J., Mackinlay, J., Stolte, C., & Agrawala, M. (2008). Graphical histories for visualization: Supporting analysis, communication, and evaluation. IEEE transactions on visualization and computer graphics, 14(6).

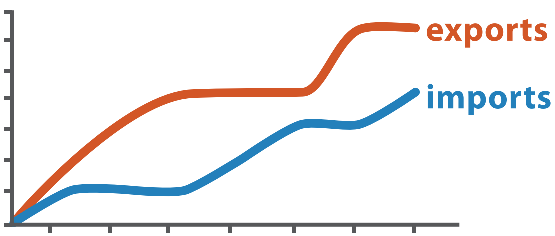
Actions: Produce

- Derive: produce new data elements based on existing data
- Transformation of dataset
  - Critical part of the vis design process
  - Often needed to create a visual encoding solving a current problem
- Example: deriving quantitative temperature data in two « warm » and « cold » sets

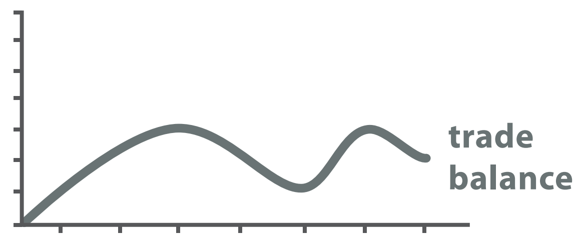


Derivation Example

- Left-hand vis shows two raw, underived data
- Right-hand vis shows the derived data
- Notice the change in semantics?



Original Data

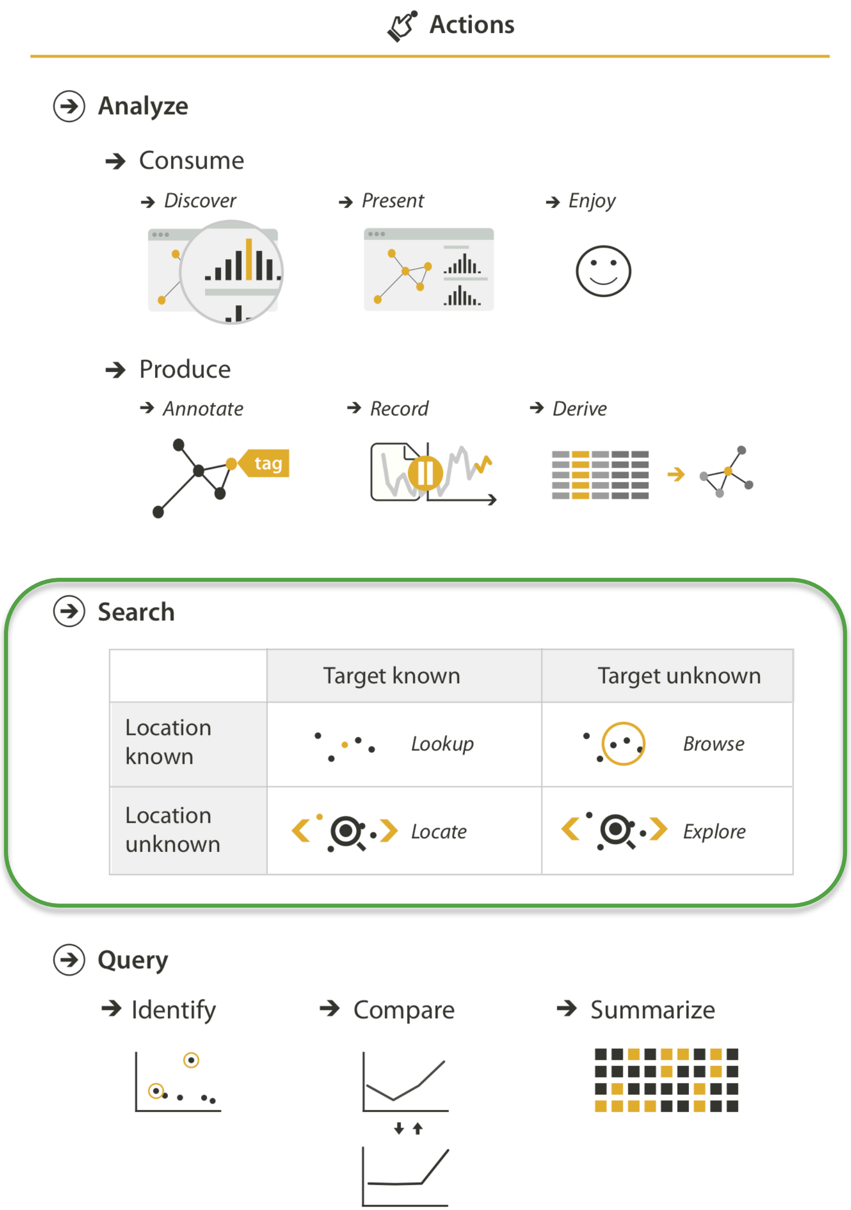


trade balance = exports − imports

Derived Data

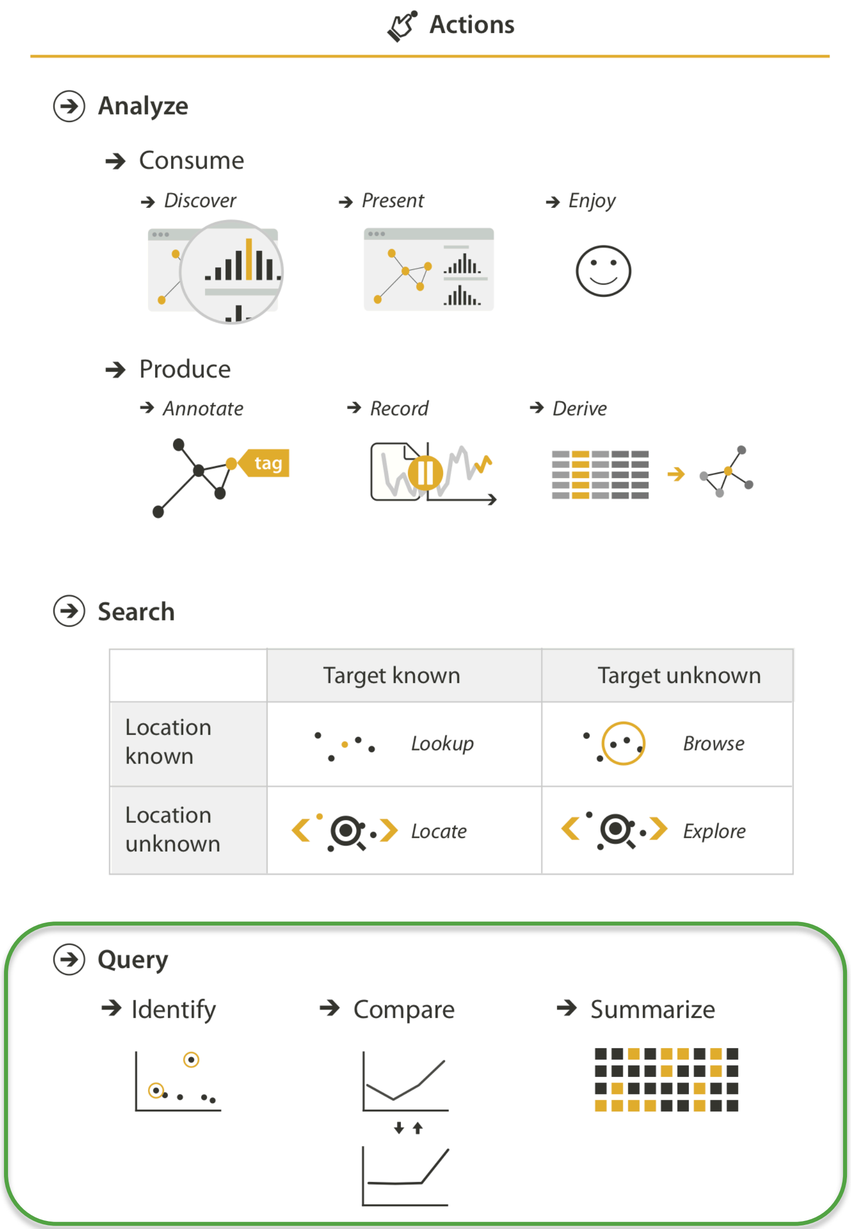
Actions: Search

- Find elements of interest as mid-level goal for the user
- Four alternatives, depending on whether the location and target are known/unknown
  - Lookup
  - Locate
  - Browse
  - Explore



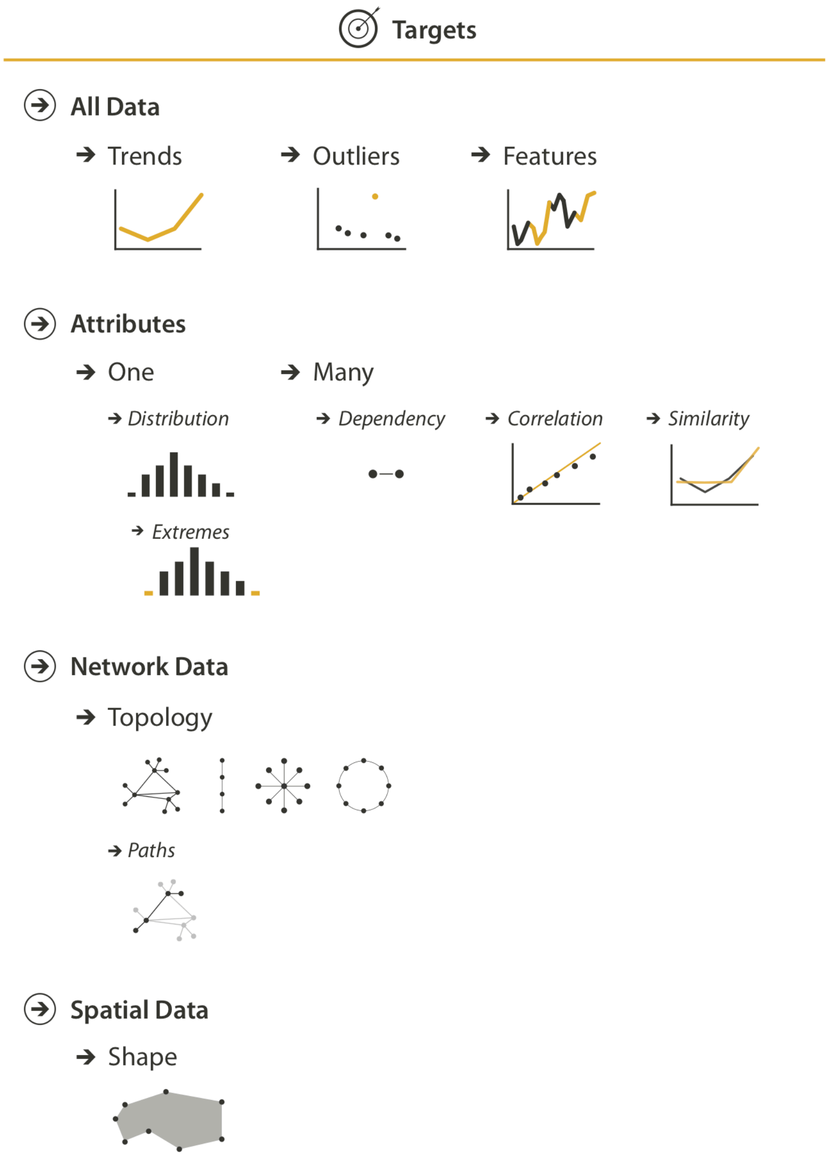
Actions: Query

- Querying targets to get more information
- Querying one target: identify
- Querying multiple targets: compare
- Querying all targets: summarize



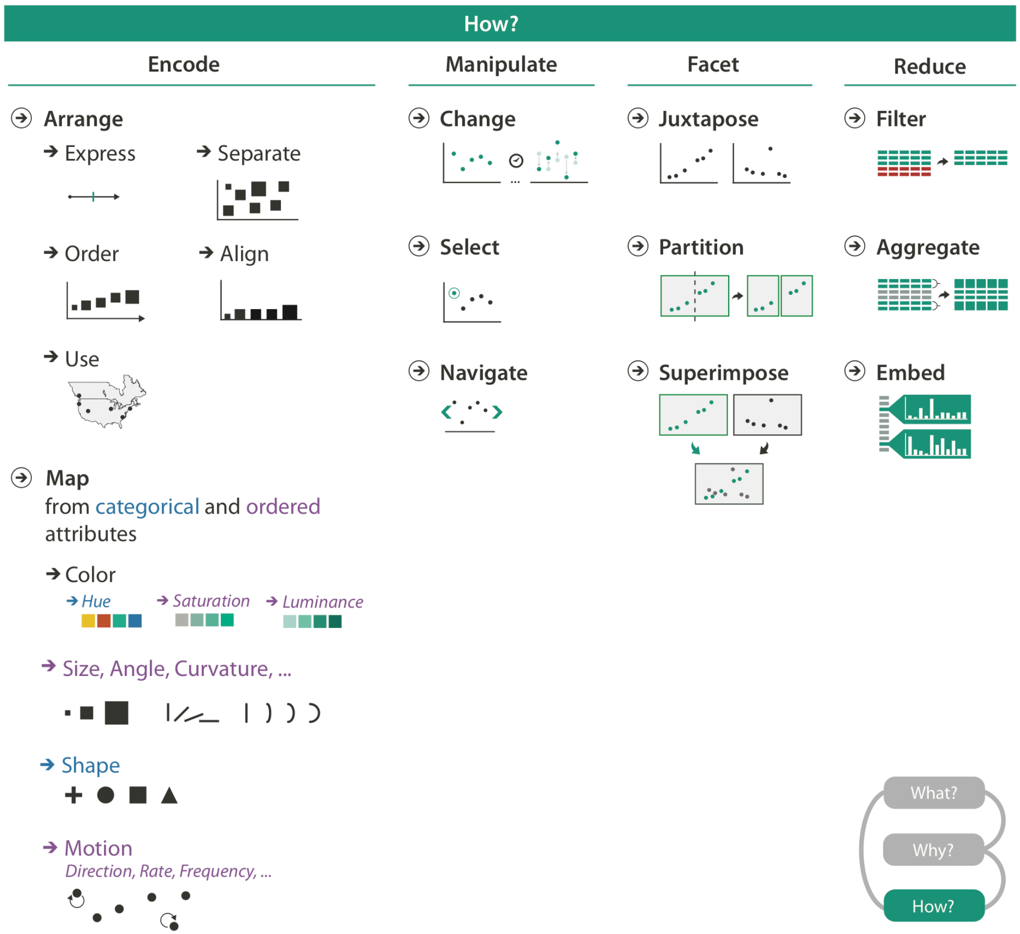
Targets (Nouns)

- On all data, three prominent:
  - Trends: increases, decreases, peaks, troughs, plateaus...
  - Outliers
  - Features
- Attributes: visually encoded specific properties
- Network data: topology and paths
- Spatial data: shape

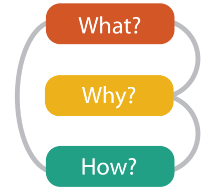


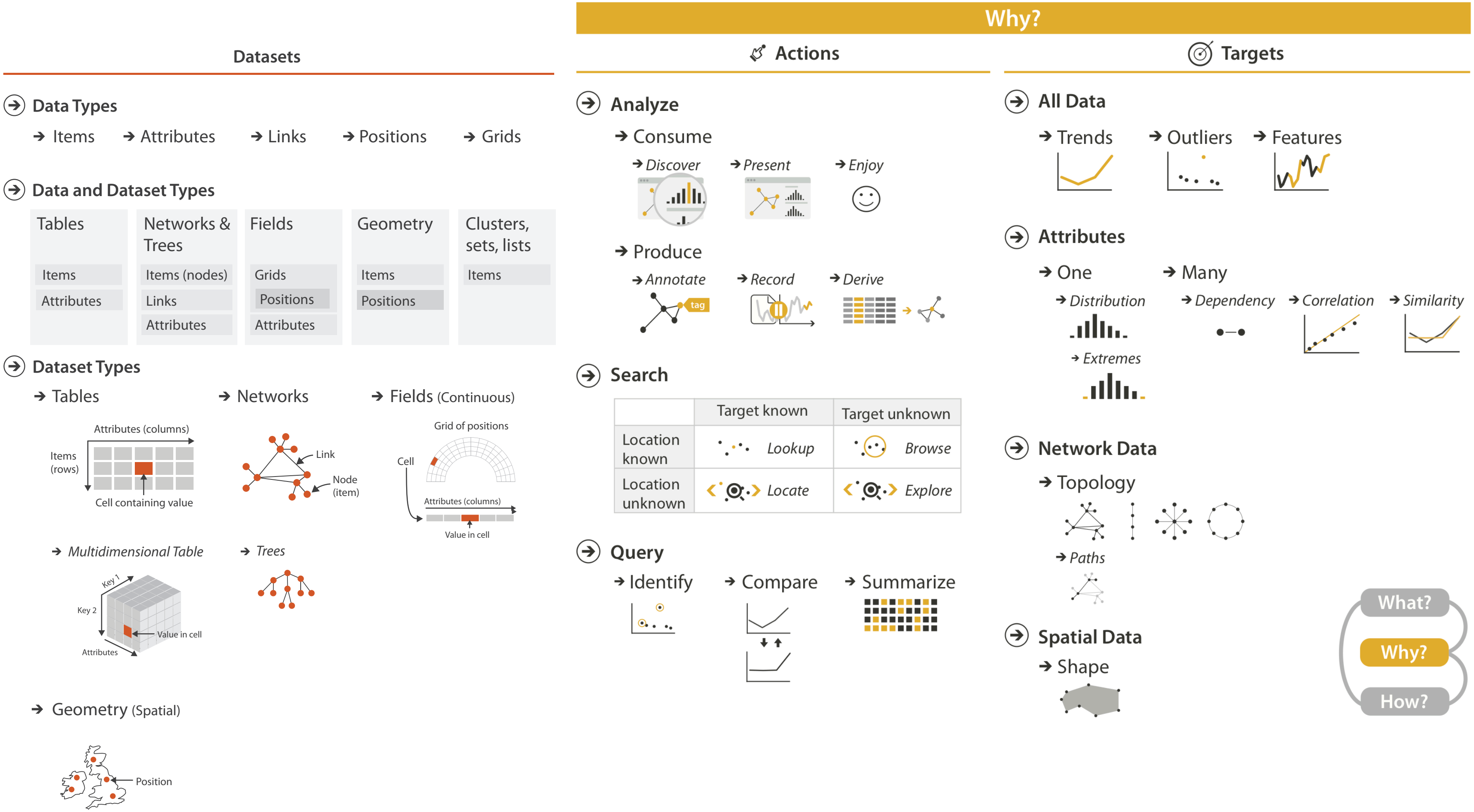
How? (Sneak Peek, More on It Later)

What?
Why?
How?



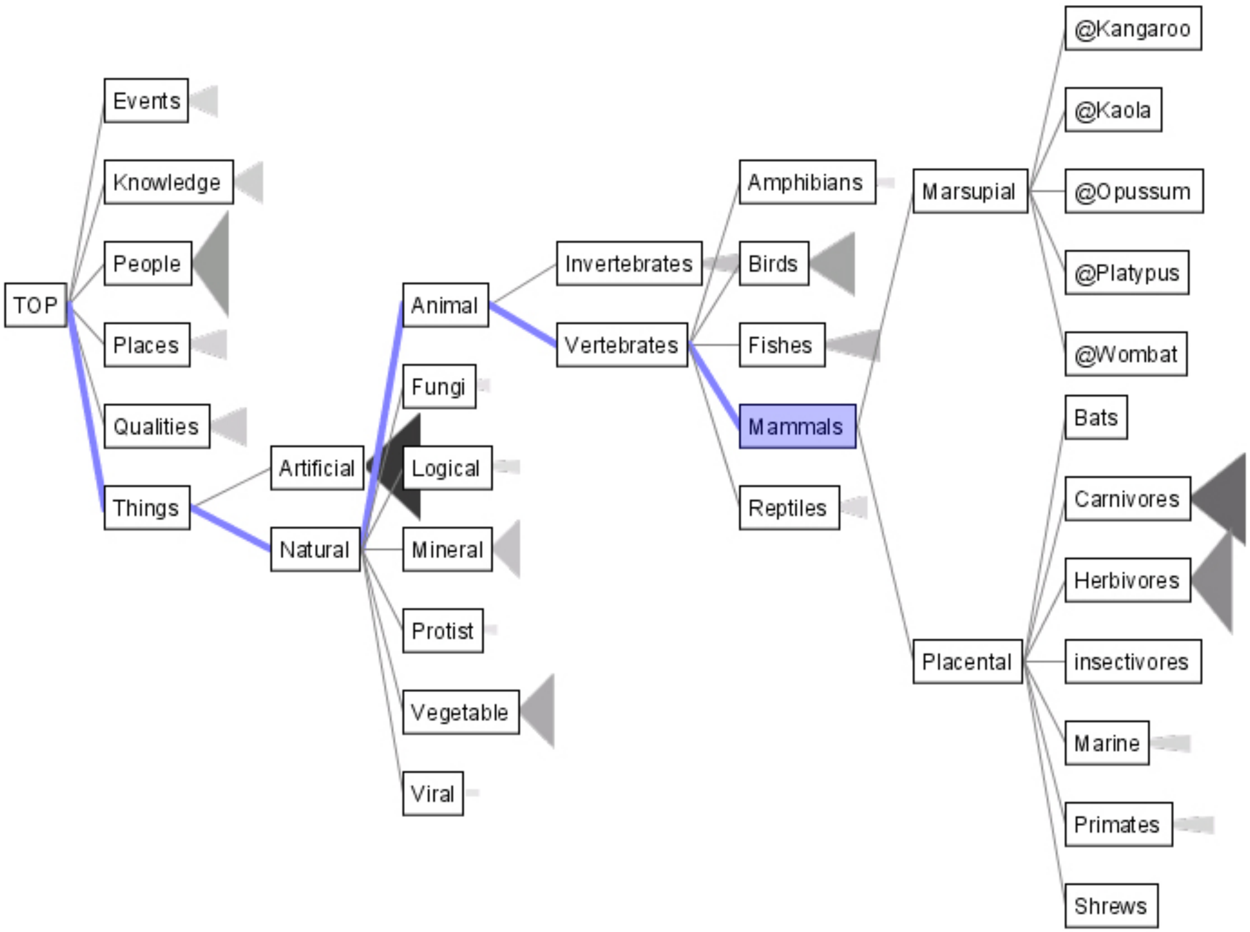
A Few Examples





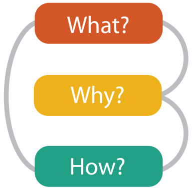
Example 1: SpaceTree

What?
Why?
How?



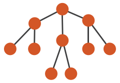
Plaisant, C., Grosjean, J., & Bederson, B. B. (2002). Spacetree: Supporting exploration in large node link tree, design evolution and empirical evaluation. In Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on (pp. 57-64). IEEE.

Example 1: SpaceTree

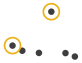


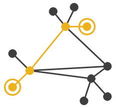
What?

- **Tree**

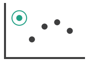
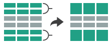


Why?

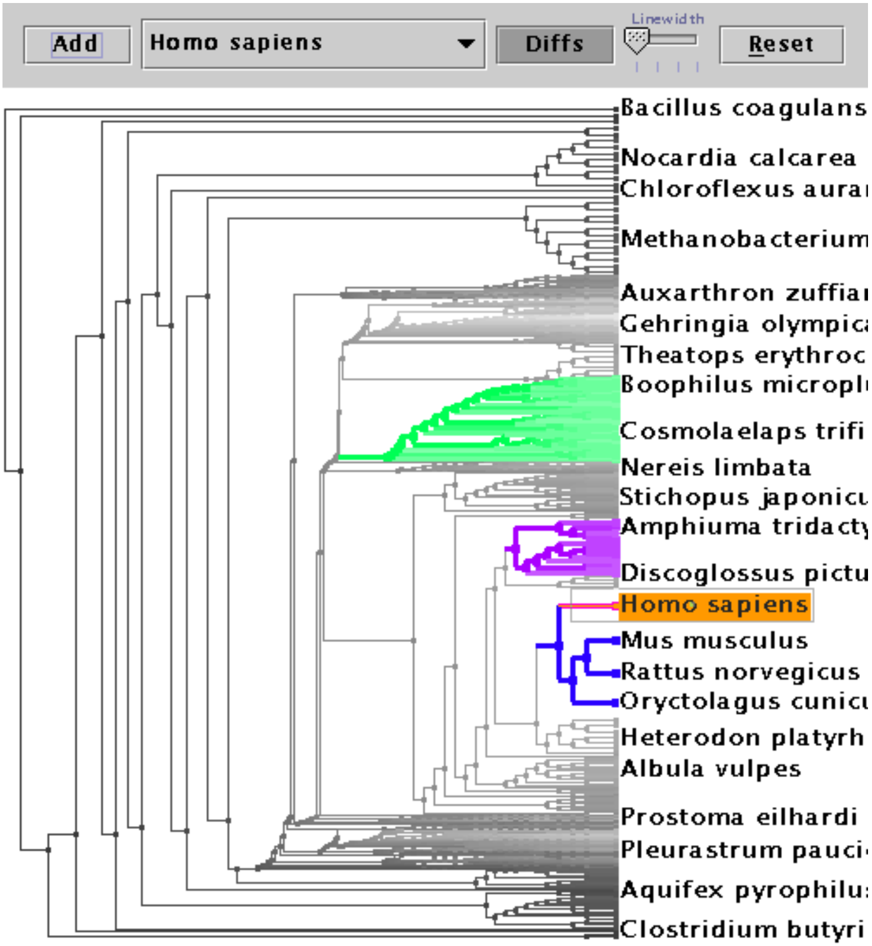
- **Actions**
  - Present 
  - Locate 
  - Identify 
- **Targets**
  - Path between two nodes



How?

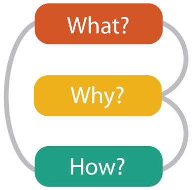
- **SpaceTree**
  - Encode 
  - Navigate 
  - Select 
  - Filter 
  - Aggregate 

Example 2: TreeJuxtaposer



Munzner, T., Guimbretière, F., Tasiran, S., Zhang, L., & Zhou, Y. (2003, July). TreeJuxtaposer: scalable tree comparison using Focus+ Context with guaranteed visibility. In ACM Transactions on Graphics (TOG) (Vol. 22, No. 3, pp. 453-462). ACM.

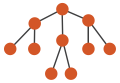
Example 2: TreeJuxtaposer



What?	Why?	How?

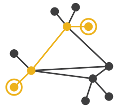
What?

- **Tree**



Why?

- **Actions**
  - Present
  - Locate
  - Identify
- **Targets**
  - Path between two nodes



How?

- **SpaceTree**
  - Encode
  - Navigate
  - Select
  - Filter
  - Aggregate
- **TreeJuxtaposer**
  - Encode
  - Navigate
  - Select
  - Arrange

Example 3: Strahler Number

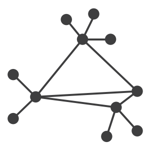
What?
Why?
How?

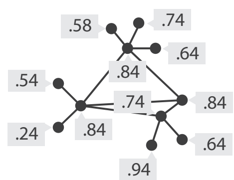
Auber, D. (2002, September). Using Strahler numbers for real time visual exploration of huge graphs. In International Conference on Computer Vision and Graphics (Vol. 1, p. 3).

Example 3: Strahler Number

What?
Why?
How?

Task 1





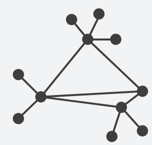
In
Tree

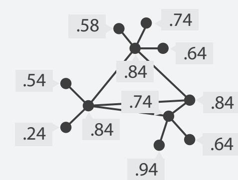
Out
Quantitative attribute on nodes

What?
- In Tree
- Out Quantitative attribute on nodes

Why?
- Derive

Task 2

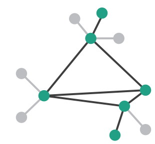




In
Tree

+

In
Quantitative attribute on nodes



Out
Filtered tree
Removed unimportant parts

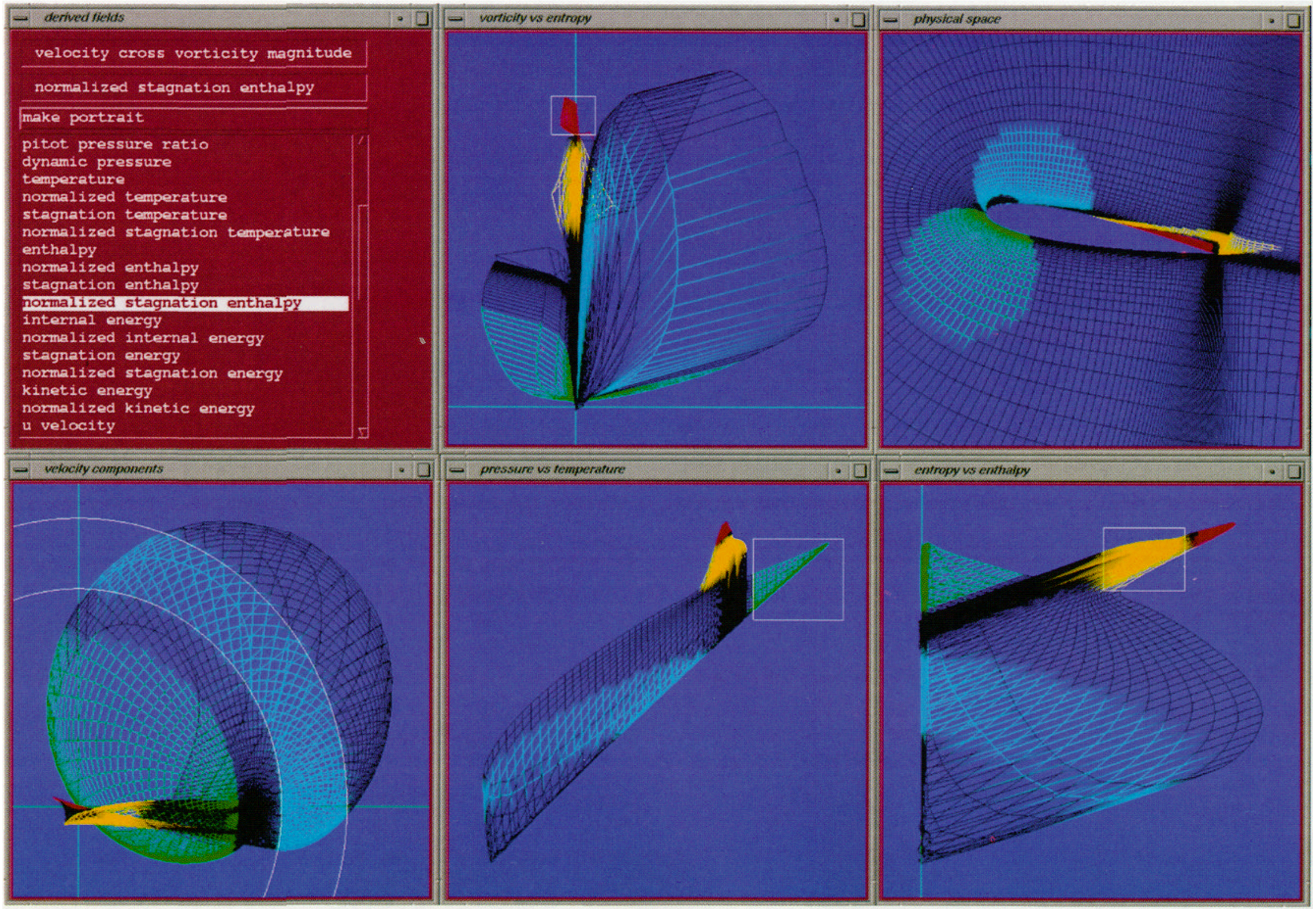
What?
- In Tree
- In Quantitative attribute on nodes
- Out Filtered tree

Why?
- Summarize
- Topology

How?
- Reduce
- Filter

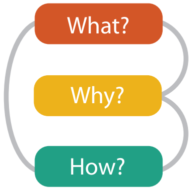
Example 4: Fluid Dynamics

What?
Why?
How?

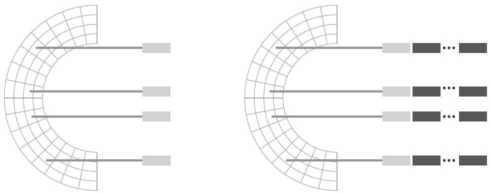


Henze, C. (1998, October). Feature detection in linked derived spaces. In Proceedings of the conference on Visualization'98 (pp. 87-94). IEEE Computer Society Press.

Example 4: Fluid Dynamics



Task 1



In
Spatial field → Out
Many quantitative attributes

Task 2

In
Spatial field + In
Many quantitative attributes → Out
Juxtaposed attribute plots with linked coloring

Task 1

What?
- **In Spatial field**
- **Out Many quantitative attributes**

Why?
- Derive

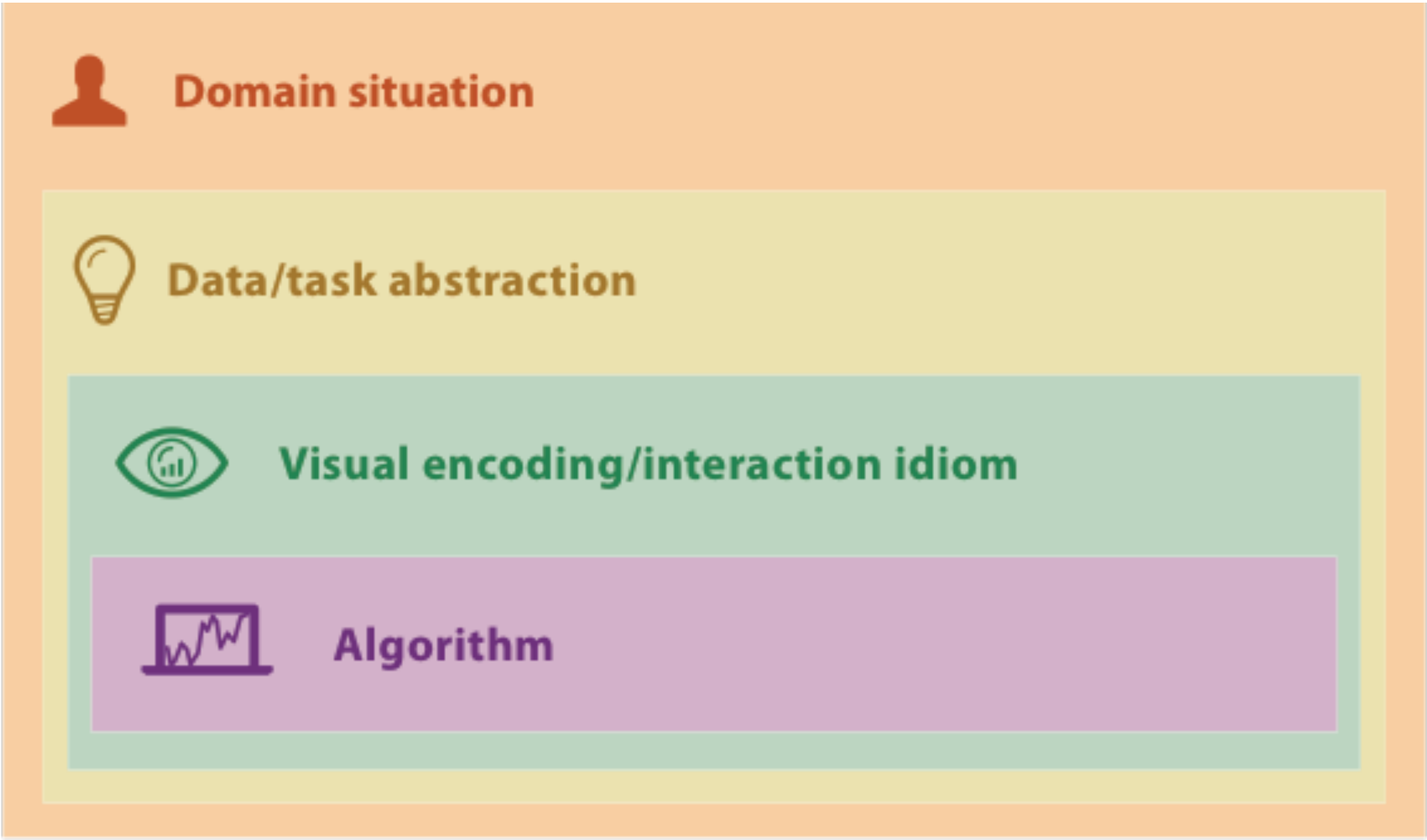
Task 2

What?
- **In Spatial field**
- **In Many quantitative attributes**
- **Out Juxtaposed attribute plots with linked coloring**

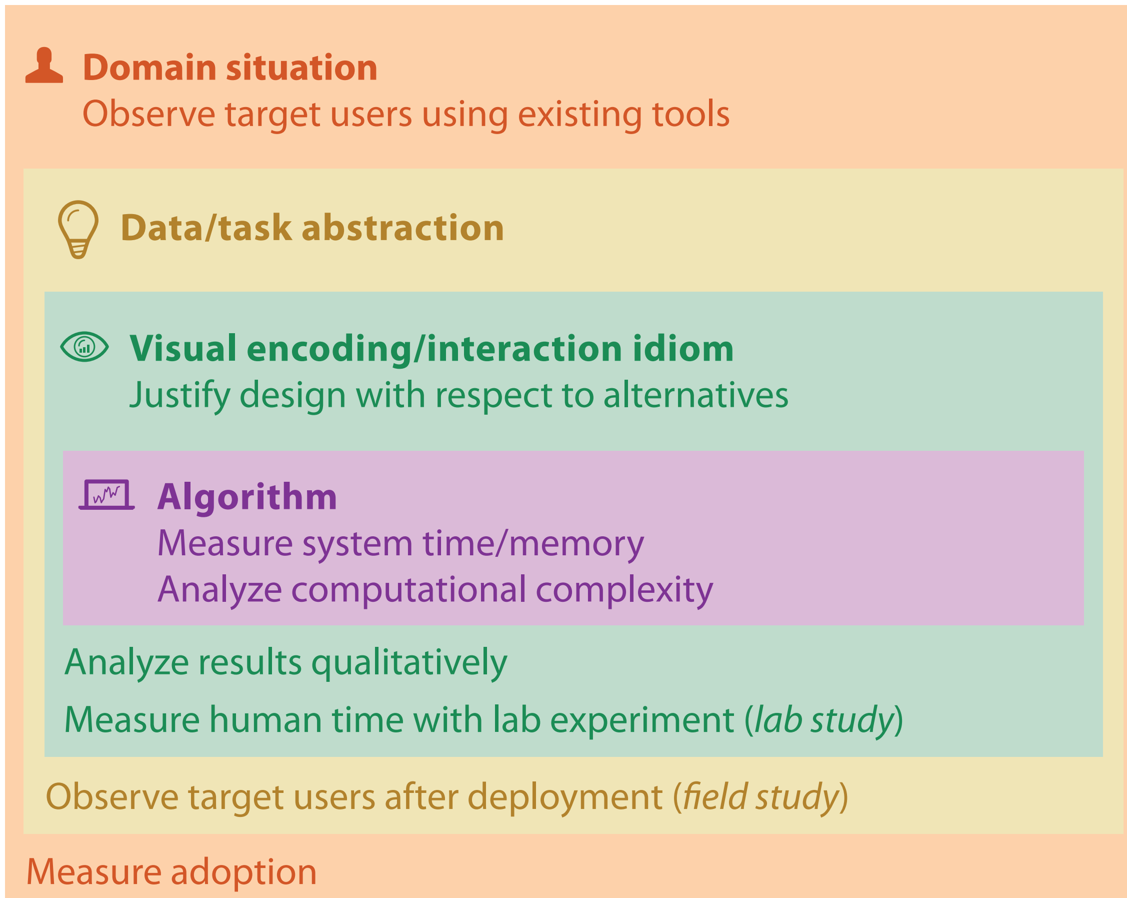
Why?
- **Actions**
  - → Discover
  - → Explore
  - → Browse
  - → Identify
  - → Compare
- **Targets**
  - → Features

How?
- **Map**
  - → Hue
- **Facet**
  - → Juxtapose
  - → Partition
- **Arrange**
  - → Express
- **Manipulate**
  - → Select
  - → Navigate

Analysis: Four Levels of Design



Analysis: Four Levels of Design



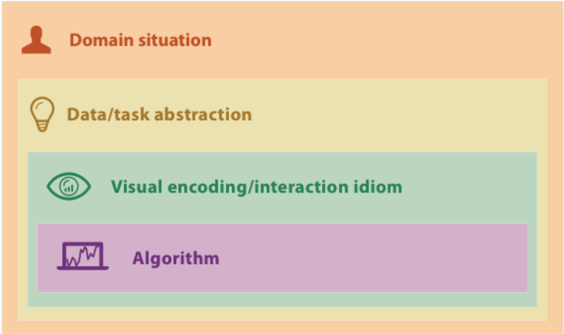
Domain

What? Why?

How?

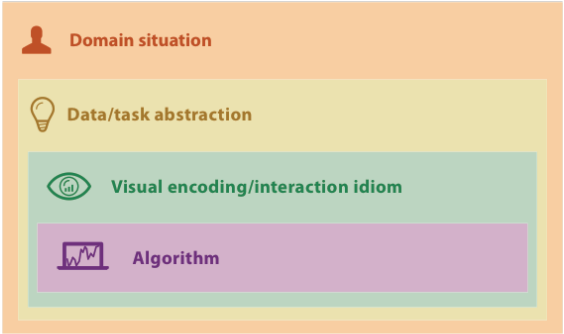
Technological aspects

Four Levels: Domain Situation

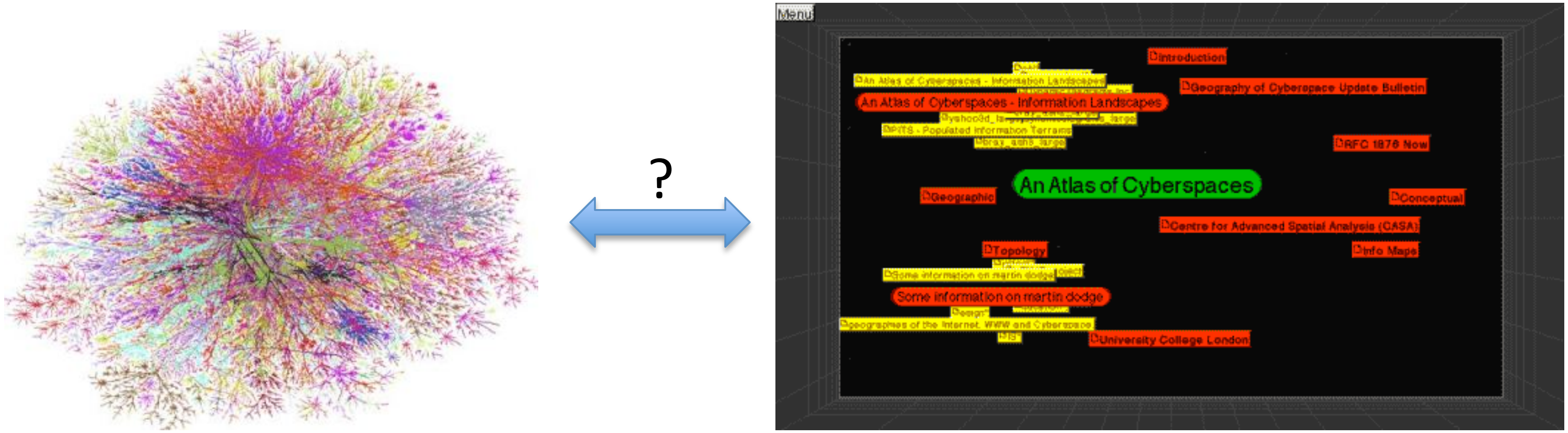


- Group of target users
  - From handful of users to anyone on the web
- + Their domain of interest (~ work domain)
  - Biology, security, e-commerce...
- + Their questions, their data…
  - Frequently, a workflow may already exist
- User experience (UX) approach needed
  - Engage with your target users
  - Observe them, their work context, talk to them...
  - Don't ask them directly what they need: users seldom are able to articulate their needs clearly
  - Eliciting design requirements is not easy !
    - -> *Interaction Homme-Machine* class, University of Namur, CS faculty

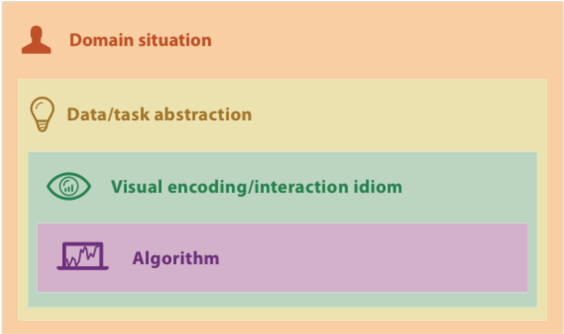
Four Levels: Data/Task Abstraction



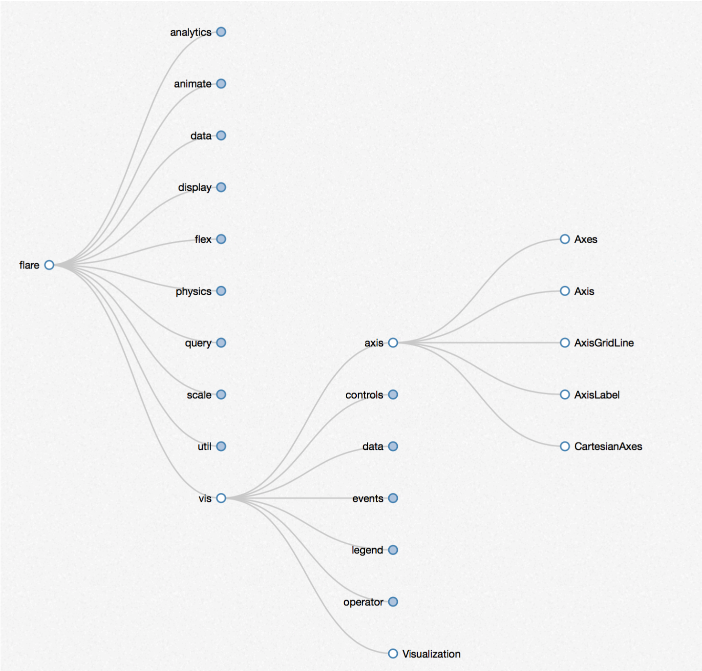
- From domain-specific to task-specific
- Use task actions and targets, as seen above to go from domain dependent to domain-independant
- Data: transform it, manipulate it, make it suit your vis idioms needs…
- Data types are not set in stones!



Four Levels: Visual Encoding

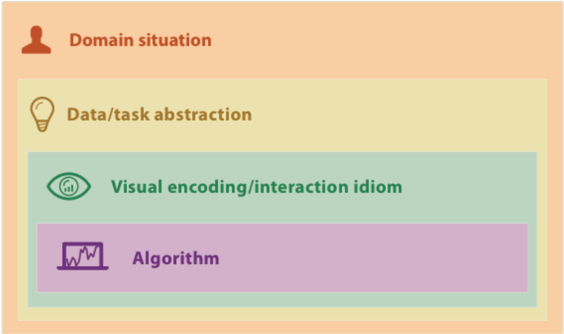


- From data and abstract task to the specific vis idiom (idiom = any possible approach)
  - Visual encoding (representation)
  - Maybe multiple idioms in one tool
  - Interaction
- Design choices!
  - Your decisions among the millions of possibilities within the design space
  - Taking into account human capabilities removes bad choices



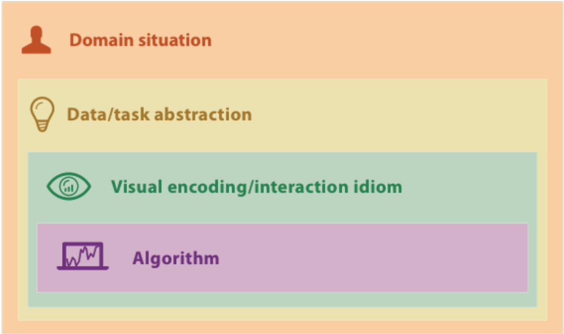
http://mbostock.github.io/d3/talk/20111018/tree.html

Four Levels: Algorithm



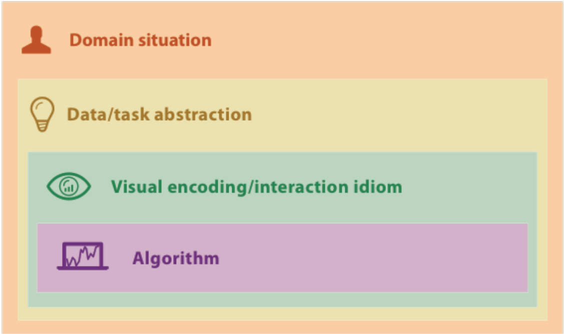
- Actual implementation of the chosen visual encoding + interaction choice(s)
- Also a design space
  - From raw code (C, C++, Java, Python...)
  - ... To existing frameworks to help you code (D3, prefuse...)
  - ... To embedded tools (R, Mathematica, Excel...)
  - ... To infovis-specific systems (Tableau, SAS...)
  - ... To paper (pen, glue, tape...)
- Efficiency vs. Comfort vs. Implementation speed vs. …

Angles of Attack



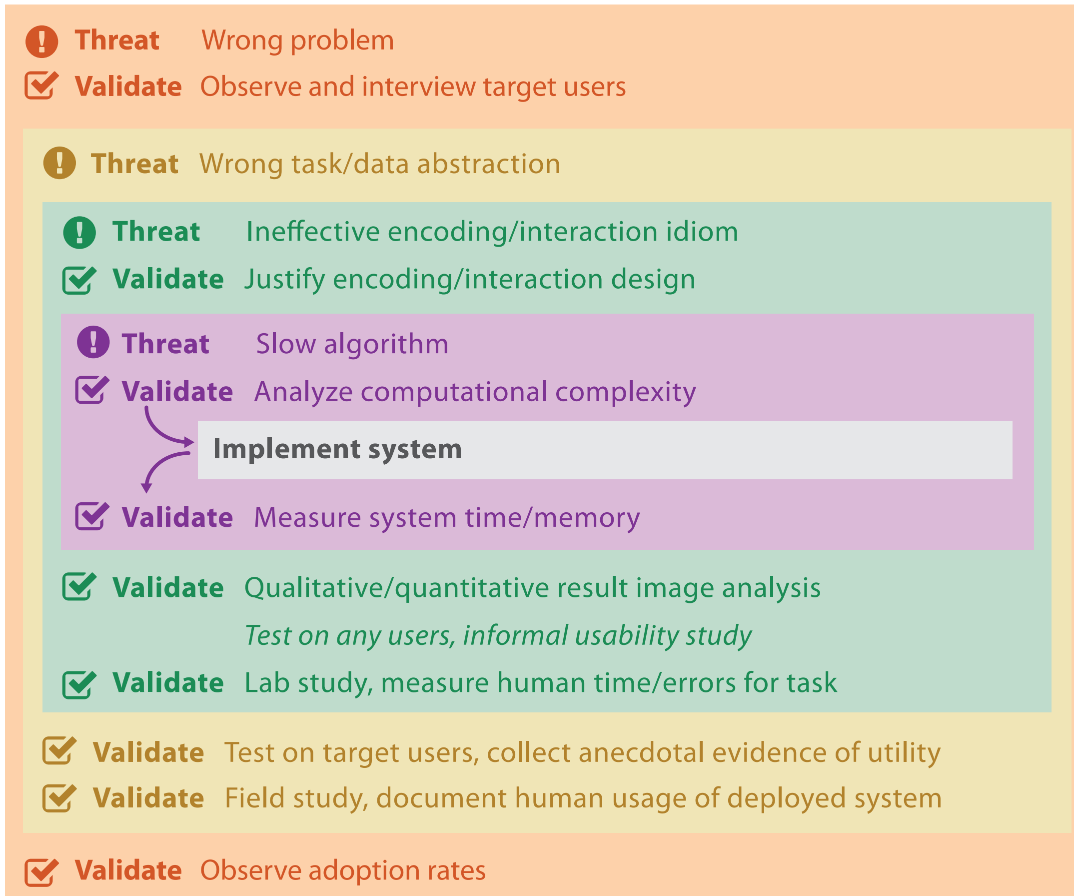
- Problem-driven vs. Technique-driven
- Problem-driven: starting from the task
  - Aka top-down
  - What's the user problem? How can I solve it?
- Technique-driven: starting from the tool
  - Aka bottom-up
  - I want to experiment this new visual encoding! What can my favorite tool do?
- Data/task abstraction frequently the hardest to get right!
- Iterative refinement

Threats to validity



- Four levels:
- Wrong problem: the user's needs were misunderstood
- Wrong abstraction: you're showing users the wrong thing
- Wrong idiom: The way you show it doesn't work
- Wrong algorithm: your code breaks/is too slow

Threats vs. Validation



For Next Week…

- Perception, marks, channels, colour