Data+Dataset Types/Semantics Tasks

Cmpt 767 - Visualization

Steven Bergner
sbergner@sfu.ca

[Munzner/Möller]

Reading

- Munzner, “Visualization Analysis and Design”:
  - Chapter 2+3 (Why+What+How)
- Shneiderman, “The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations,” IEEE Symposium on Visual Languages, 1996
- Heer+Shneiderman, “Interactive Dynamics for Visual Analysis,” Communications of the ACM 2012.
- Amar et al., “Low-level components of analytic activity in information visualization,” InfoVis 2005.

Data/set types+semantics
Tasks

- What — Data abstraction
  - Data types
    - categorical, ordinal, quantitative
  - Dataset types
    - Tables
    - Networks/graph (trees)
    - Text / logs
    - Fields
    - Static file vs. dynamic stream
  - Attribute + dataset semantics
    - Spatial vs. non-spatial
    - Temporal vs. non-temporal
    - Keys vs. values
    - Continuous vs. discrete
    - Topology vs. geometry
  - Derived attributes / spaces
- Why+How —Task abstraction

Data types

32	7/16/07	2-High	Medium Box	0.65	7/18/07

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
32	7/16/07	2-High	Medium Box	0.65	7/18/07

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

Semantics vs. type

- Semantics: real-world meaning of data
- Type: abstract classification with implications on
  - mathematical operations
  - data structure (how to store)
- Given semantics -- type will follow

Basic variable types

- Physical type
  - Characterized by storage format & machine ops
  - e.g: bool, short, int, float, double, string, …
- Abstract type
  - Provide descriptions of the data
  - Characterized by methods / attributes
  - May be organized into a hierarchy

Data types

- categorical (nominal)
  apples, oranges, bananas
- ordered
  - ordinal
    e.g. rankings: small, medium, large
  - quantitative
    real numbers
  - sequential (interval)
  - diverging (ratios)
    well defined zero point

Quantitative

- Q - Interval (location of zero arbitrary)
  – Dates: Jan 19; Location: (Lat, Long)
  – Only differences (i.e., intervals) can be compared
- Q - Ratio (zero fixed)
  – Measurements: Length, Mass, Temp, ...
  – Origin is meaningful, can measure ratios & proportions
  – Weight A is twice as heavy as weight B
  – doesn't work for dates!

Hierarchies

- possible for any data type
- sometimes strong implicit hierarchies
- e.g. geography:
  – postal code
  – city district
  – city
  – state
  – country
  – continent

Example - Time

- has a strong (implicit) hierarchy:
  – minute
  – hour
  – day
  – week
  – month
- can be seen as ordinal (entries in a diary)
- can be seen as quantitative (timings in a race)
- interval vs. ratio --
  time-stamp vs. duration

Dataset types

Dataset types

- Tables
- Networks/graph (trees)
- Text / logs
- Fields
- Static file vs. dynamic stream

Tables

- each data **item** in a new row
- each column contains an **attribute**

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

Record / Item

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

Field / Dimension / Attribute

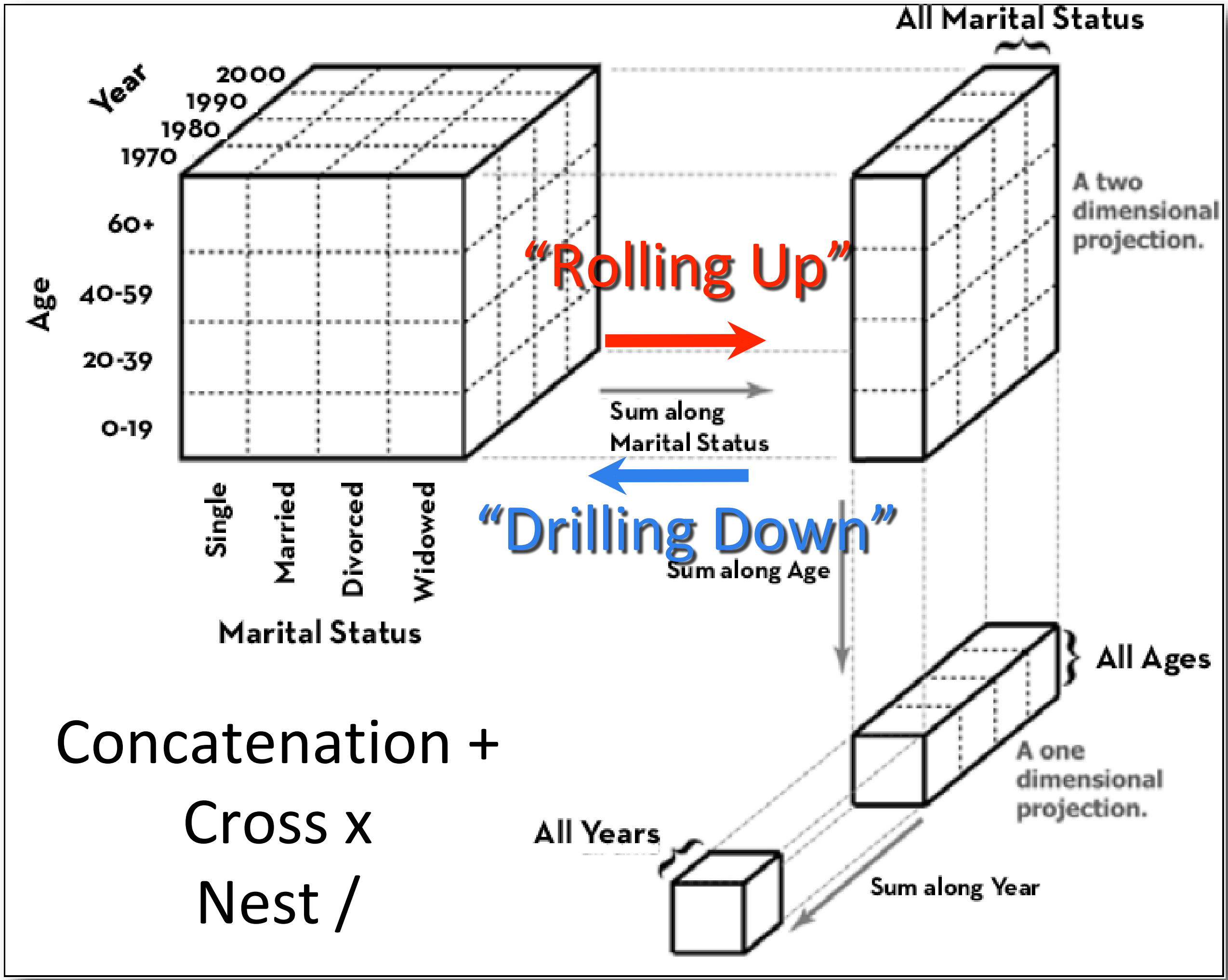
A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Specif		0.6	6/6/05
70	12/18/06	5-Low		0.59	12/23/06
70	12/18/06	5-Low		0.82	12/23/06
96	4/17/05	2-High		0.55	4/19/05
97	1/29/06	3-Medium		0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

1 = Quantitative
2 = Nominal
3 = Ordinal

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Speci[illegible]	[illegible]	0.6	6/6/05
70	12/18/06	5-Low	[illegible]	0.59	12/23/06
70	12/18/06	5-Low	[illegible]	0.82	12/23/06
96	4/17/05	2-High	[illegible]	0.55	4/19/05
97	1/29/06	3-Medium	[illegible]	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08
194	4/5/08	3-Medium	Wrap Bag	0.84	4/7/08

1 = Quantitative
2 = Nominal
3 = Ordinal

Relational Data Cubes



Concatenation +
Cross x
Nest /

Pivot Tables

	A	B	C	D	E
1	**Region**	East			
2					
3	**Sum of Sales**		**Product**		
4	**Month**	**Salesperson**	Dairy	Meat	Grand Total
5	May	Buchanan		$17,578	$17,578
6		Davolio	$22,977		$22,977
7	May Total		$22,977	$17,578	$40,555
8	Jun	Buchanan	$10,017	$7,711	$17,728
9		Davolio	$6,805	$5,575	$12,380
10	Jun Total		$16,822	$13,286	$30,108
11	Grand Total		$39,799	$30,864	$70,663

Networks / graphs

- item = **node**
- **link** between two items = **edge**
- e.g. social network: people + friendships
- both links+nodes can have attributes
- graphs can be represented by 2 tables

Types of graphs

- undirected graph vs. directed graph
  – edges do not/do have a direction
- DAG -- directed acyclic graphs
- connected graphs
- planar graphs
- trees -- connected graph with no cycles

Text + logs

- text document: ordered set of words
- document collection
- bag of words
- log files: designed for machine readability

Fields

- really continuous dataset
- specified through grids (connectivity)
- often connected to spatial data

Files vs. streams

- standard: static files
- challenge today: dynamic streams

Attribute + dataset semantics

Attribute+dataset semantics

- Spatial vs. non-spatial
- Temporal vs. non-temporal
- Keys vs. values
- Continuous vs. discrete

- Spatial vs. abstract
- Timevarying vs. static
- Topology vs. geometry

Spatial vs. abstract

- implications on visual encoding
- spatial
  - geographic information
  - physical simulation
  - medical data (MRI, CT scan etc.)
  - strong constraints on visual layout
- non-spatial / abstract
  - network data
  - financial transactions
  - up to the visualization expert to choose a visual layout

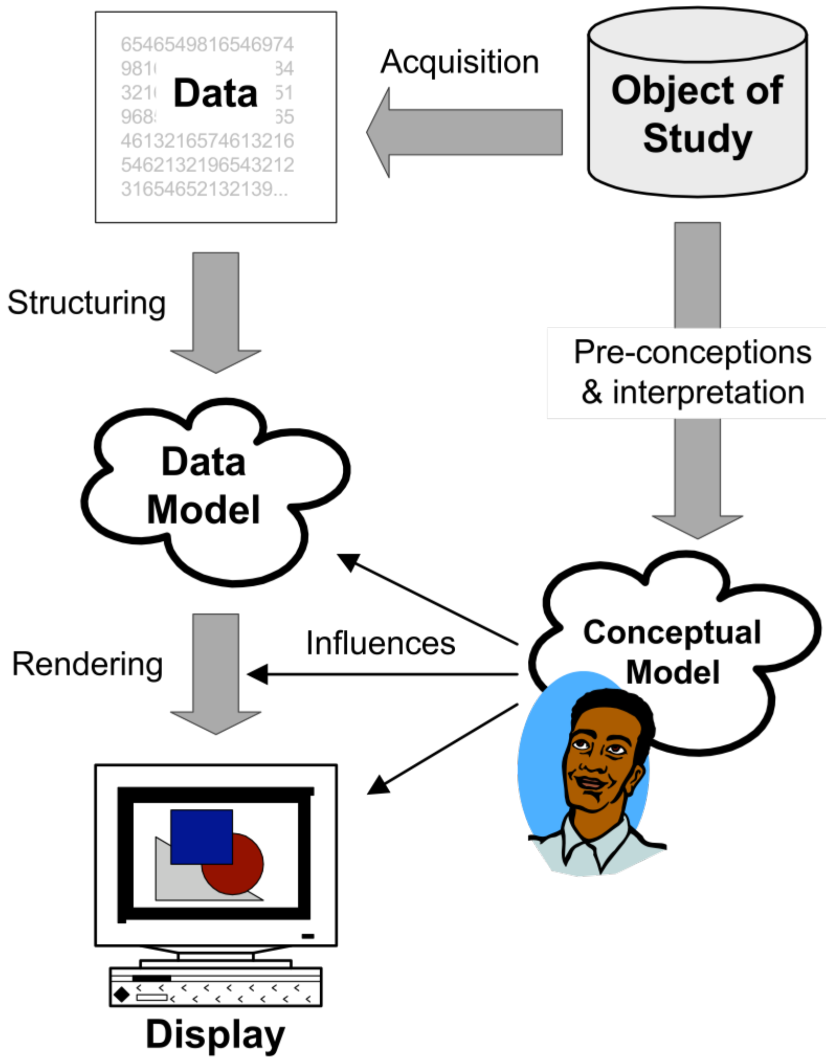
Temporal / time-varying vs. Non-temporal / static

- time has a strong meaning to us as humans
- special consideration for visual encoding
- time has a hierarchy
- time periods/cycles very important
- time-varying: time as quantitative
- time-series: time as ordinal

Continuous vs. discrete

- data is almost always discrete -- we need to store it in discrete memory cells
- it's really how we think about the data
- categorical is always discrete
- quantitative is continuous
- care must be taken when making discrete measurements continuous

Data vs. Conceptual Models



Data vs. Conceptual Models

- Data Model: Low-level description of the data
  – Set with operations, e.g., floats with +, -, /, *
- Conceptual Model: Mental construction
  – Includes semantics, supports reasoning

Physical Type	Conceptual
1D floats	temperature
3D vector of floats	space

Example

- From data model...
  – 32.5, 54.0, -17.3, … (floats)
- using conceptual model...
  – Temperature
- to data type
  – Continuous to 4 significant digits (Q)
  – Hot, warm, cold (O)
  – Burned vs. Not burned (N)

Keys vs. values

- databases: key vs. value
- statistics: independent vs. dependent variable or attribute or dimension
- computational science: inputs vs. outputs
- implies a mapping
  keys → values
- keys used to look up values in a table
- common keys: space + time

High-dimensional vs. multi-dimensional

- stats: high-dim (keys or values)
- physics: multi-dim (mostly about keys)
- implications:
  - multi-dim: two to tens of dimensions
  - high-dim: no constraint, can be thousands of dimensions

Multi-variate (about values)

- Number of values per key
  - 1: Univariate
  - 2: Bivariate
  - 3: Trivariate
  - >3: Hypervariate / Multi-variate

Spatial dimensions (keys)

- 1D: refers to a single ‘length’ scale (e.g. height)
- 2D: geographical information
- 3D: medical / physics
- time-varying:
  - 1D+time
  - 2D+time
  - 3D+time

Spatial values

- **Scalar data**
  - mapping $f: R^n \rightarrow R$, $(x_1, \ldots, x_n) \rightarrow y$
  - n independent variables (keys) x_i (1D, 2D, or 3D, +time)
  - value y is just univariate
- Example:
  - MRI data
  - 2D grey-scale image data

Spatial values

- **Vector data**
  – mapping $f: R^n \rightarrow R^m$, $(x_1,\ldots,x_n) \rightarrow (y_1,\ldots,y_m)$
  – representing direction and magnitude
  – usually $m=n$
  – Exceptions, e.g., due to projection
- Example:
  – weather map (wind direction)
  – flow around airplane wings

Spatial values

- **Tensor data**
  - mapping $f:R^n \rightarrow R^m$, $(x_1,\ldots,x_n) \rightarrow y_{i1,i2,\ldots,ik}$
  - tensor of level k
  - a tensor of level 1 is a vector
  - a tensor of level 2 is a matrix, …
- Example:
  - diffusion-tensor MRI
  - stress-tensor, etc.

Topology vs. geometry

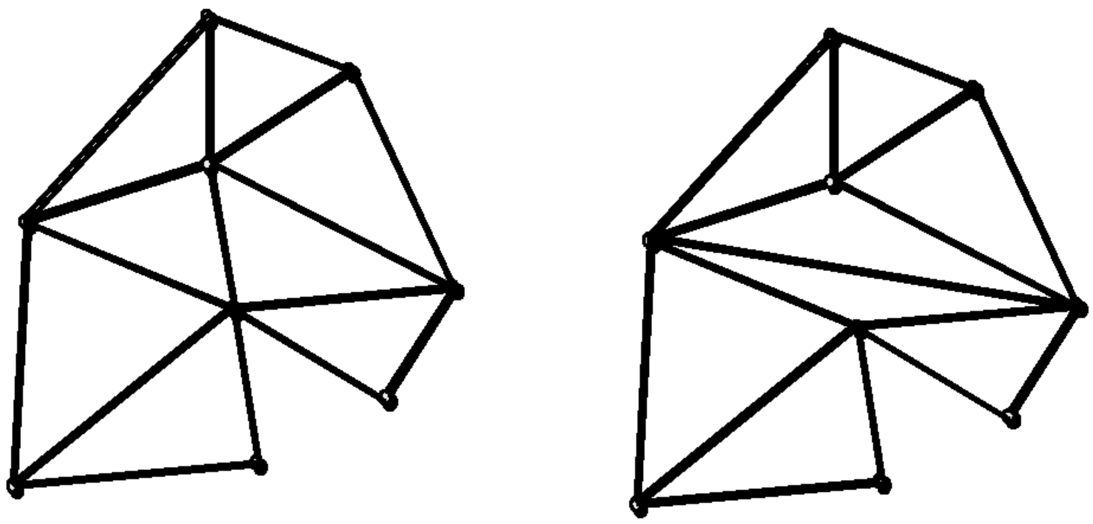
- **Topology** specifies the structure (**connectivity**) of the data
- **Geometry** specifies the **position** of the data

Topology vs. geometry

- In topology, qualitative questions about geometrical structures are the main concern
  - Does it have any holes in it?
  - Is it all connected together?
  - Can it be separated into parts?
- Underground map does not tell you how far one station is from the other, but rather how the lines are connected (topological map)

Topology

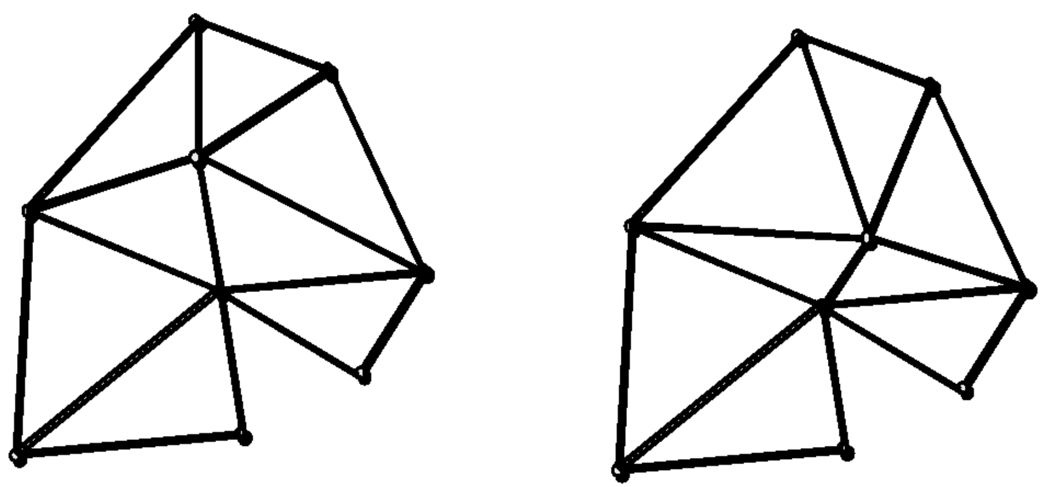
- Properties of datasets that remain unchanged even under different spatial layouts



Same geometry (vertex positions), different topology (connectivity)

Topological equivalence

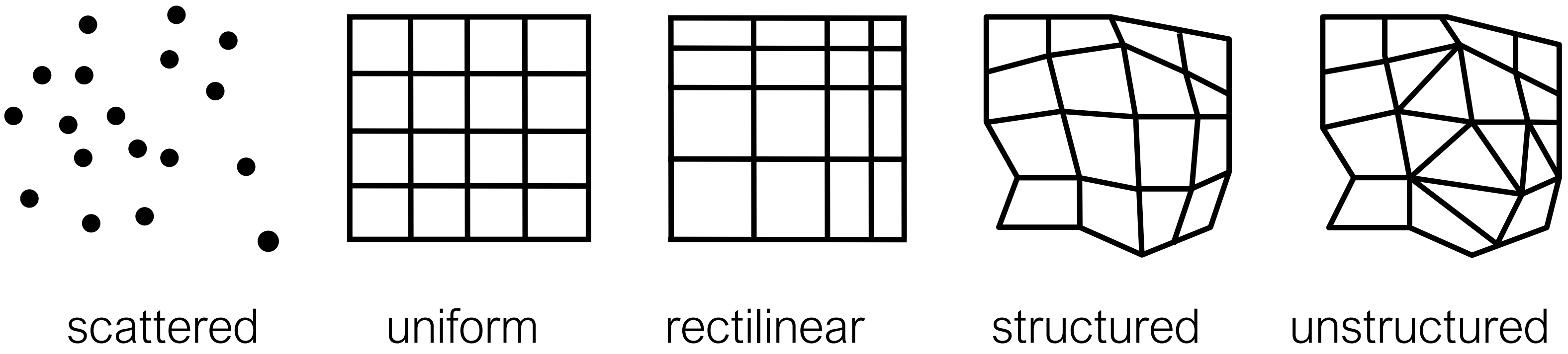
- Things that can be transformed into each other by stretching and squeezing, without tearing or sticking together bits which were previously separated



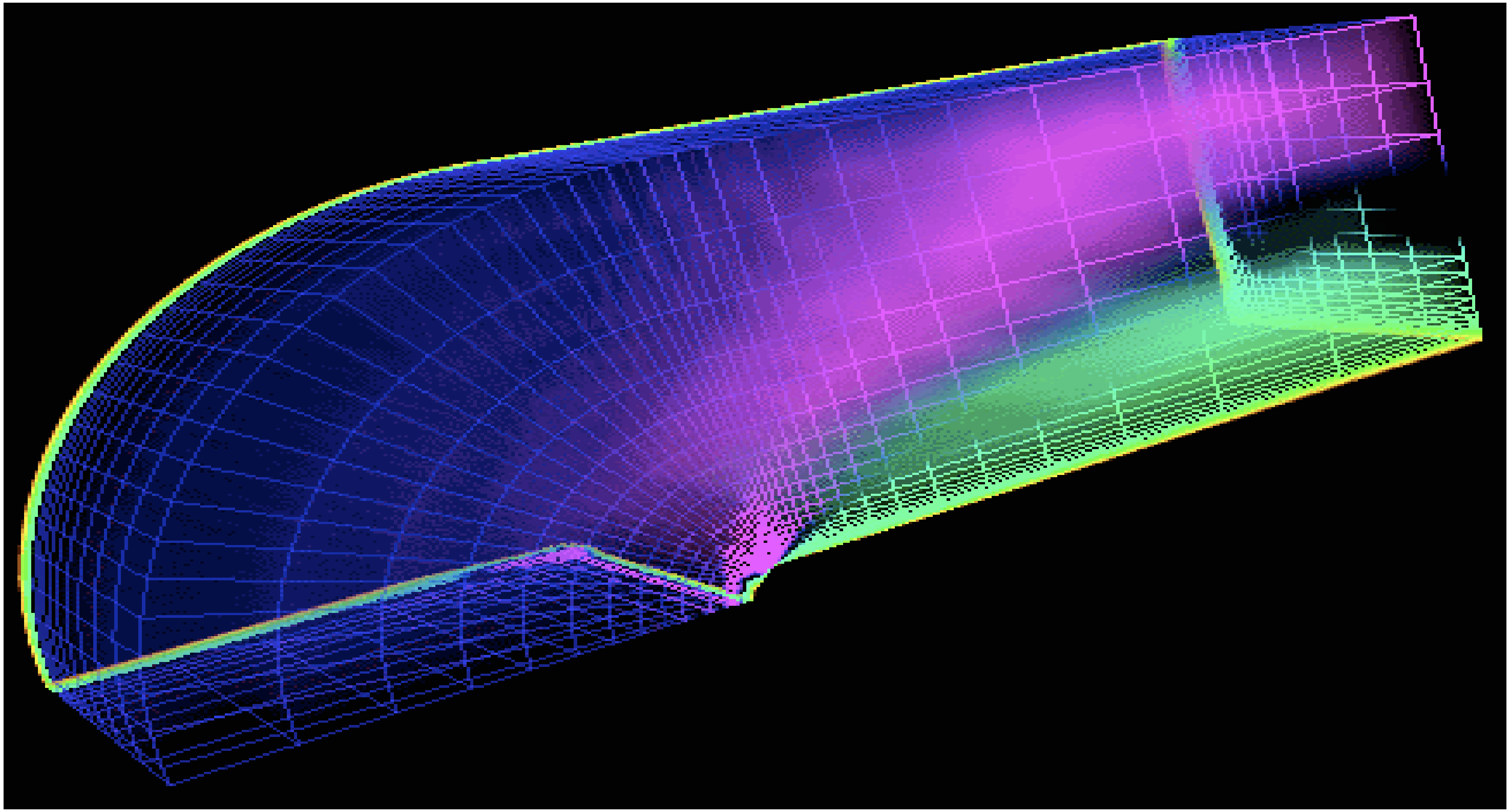
topologically equivalent

Grid types

- Grids differ substantially in the cells (basic building blocks) they are constructed from and in the way the topological information is given

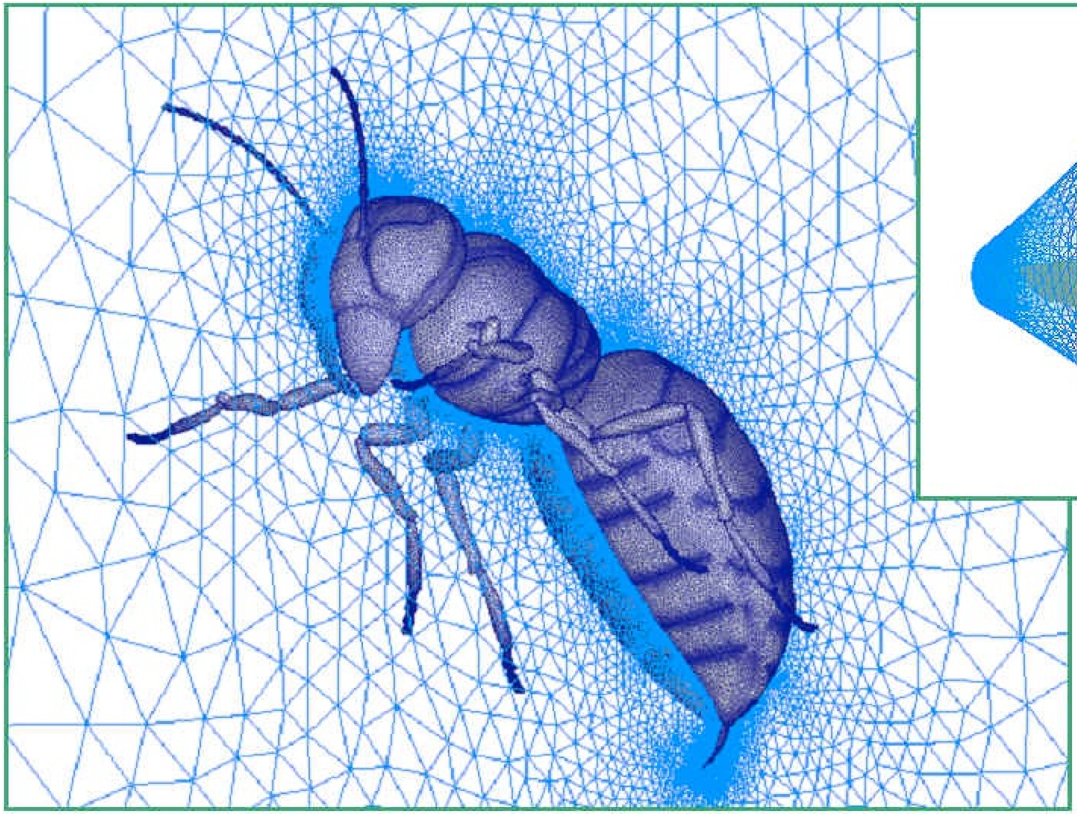


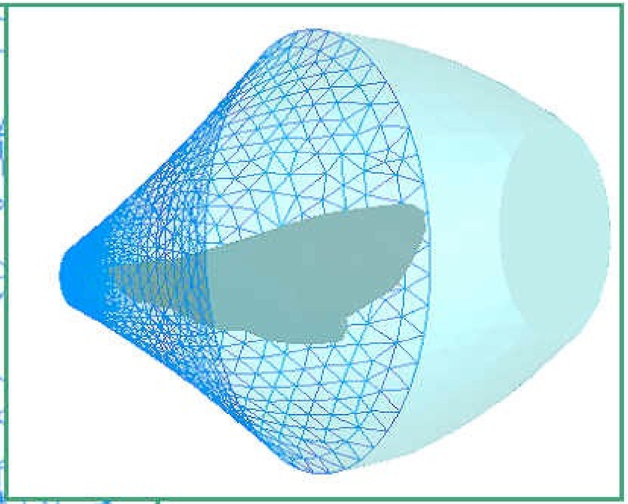
Curvilinear grids



Unstructured grids

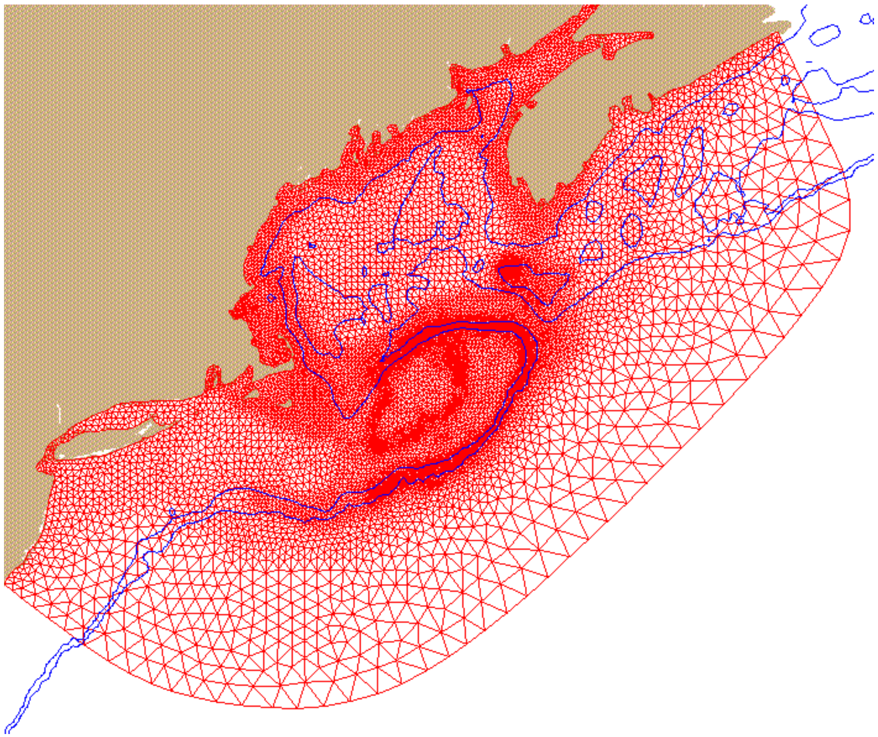
- Can be adapted to local features

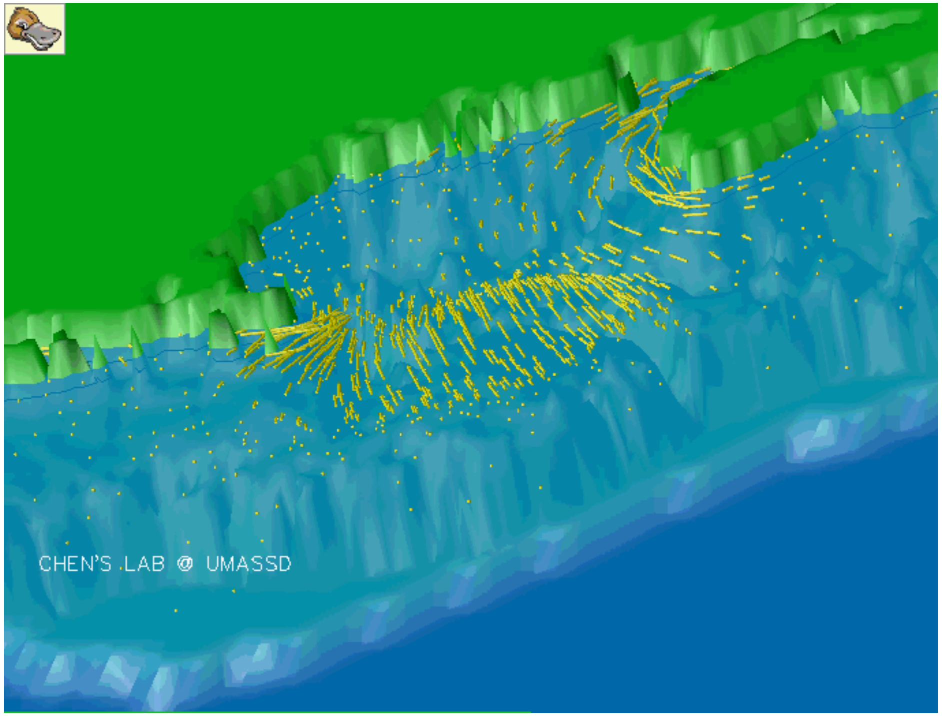




Unstructured grids

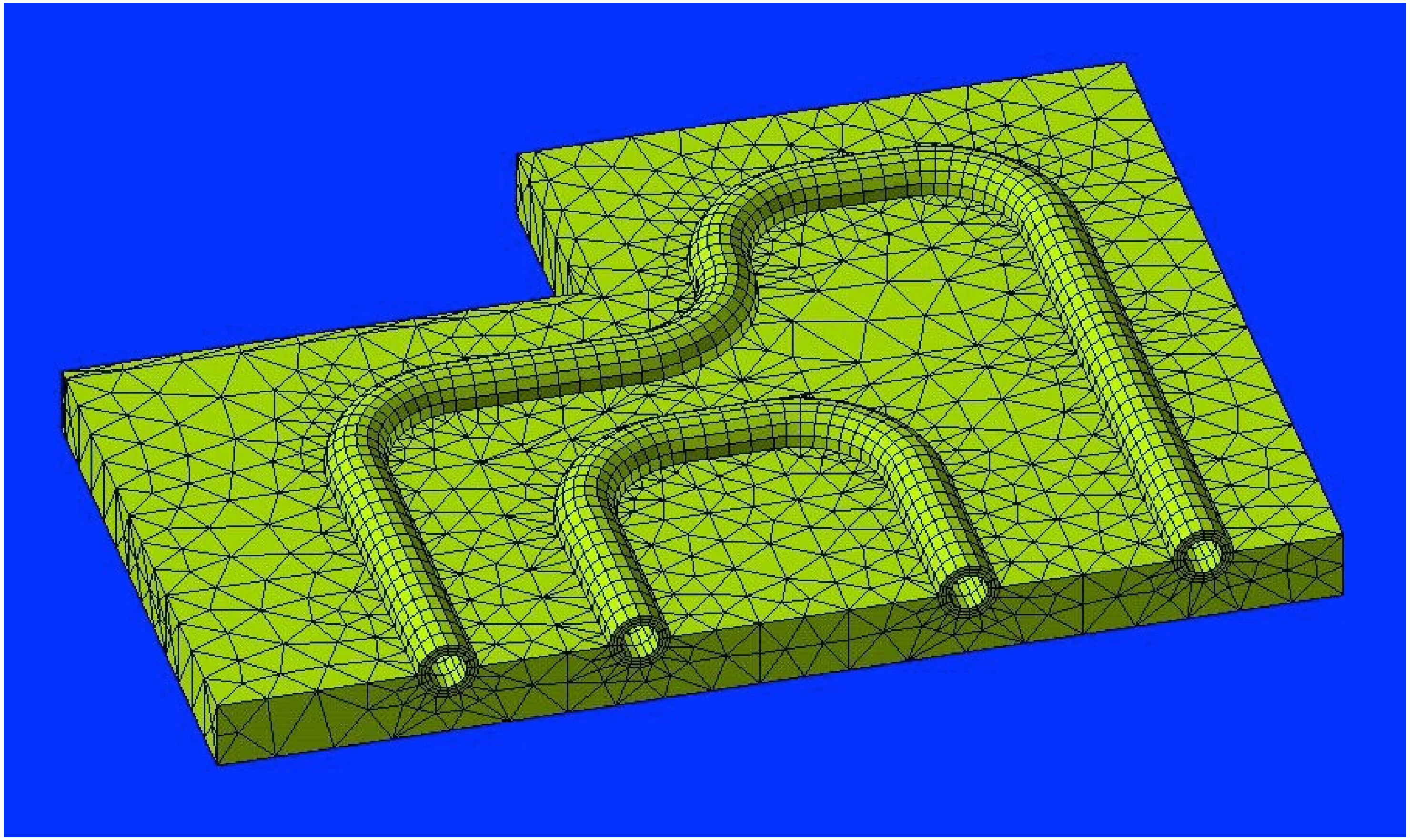
- Can be adapted to local features





Hybrid grids

- Combination of different grid types

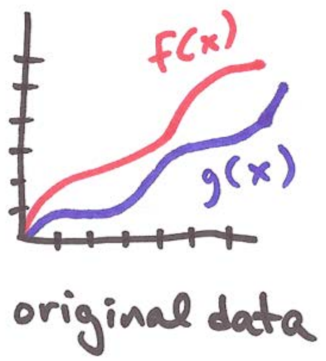


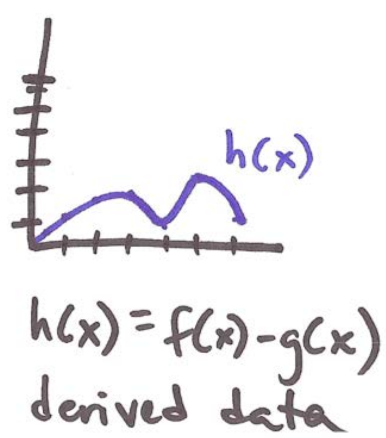
Data/set types+semantics
Tasks

- Data abstraction
  - Data types
  - Dataset types
  - Attribute + dataset semantics
  - **Derived attributes / spaces**
- Task abstraction

Derived attributes

- the norm, not the exception
- necessary for some of the tasks
- simple transformations
- statistical summaries of (lots of) data
- e.g. stagnostics





Data/set types+semantics Tasks

- What — Data abstraction
  - Data types
    - categorical, ordinal, quantitative
  - Dataset types
    - Tables
    - Networks/graph (trees)
    - Text / logs
    - Fields
    - Static file vs. dynamic stream
  - Attribute + dataset semantics
    - Spatial vs. non-spatial
    - Temporal vs. non-temporal
    - Keys vs. values
    - Continuous vs. discrete
    - Topology vs. geometry
  - Derived attributes / spaces
- Why+How — Task abstraction

Data/set types+semantics Tasks

- What — Data abstraction
- Why + How — Task abstraction
  - Shneiderman's Mantra
  - Heer+Shneiderman: Visual Analysis tasks
  - Empirical Study: Amar+Eagan+Stasko
  - Taxonomy: Brehmer + Munzner
    - Why
      - consume / produce
      - search
      - query
    - How
      - introduce
      - encode
      - facet
      - reduce

Task Abstraction

Table 1. Questions for the analysis of conserved syntenic data, with the scale and relationship addressed by each. The scales are: g, genome; c, chromosome; b, block; and f, feature. The relationships are: p, proximity/location; z, size; o, orientation; and s, similarity.

	question	*scale*				*relationship*			
		g	*c*	*b*	*f*	*p*	*z*	*o*	*s*
1	Which chromosomes share conserved blocks?	**X**				**X**			
2	For one chromosome, how many other chromosomes does it share blocks with?	**X**	**X**			**X**			
3	What is the density of coverage and where are the gaps on: chromosomes? blocks?	**X**	**X**	**X**		**X**			
4	Where are the blocks: on chromosomes? around a specific location on a chromosome?	**X**	**X**			**X**			
5	What are the sizes and locations of other genomic features near a block?		**X**			**X**	**X**		
6	How large are the blocks?		**X**				**X**		
7	Do neighboring blocks go to the same: chromosomes? relative location on a chromosome?	**X**	**X**			**X**			
8	Are the orientations matched or inverted for: block pairs? feature pairs?		**X**	**X**				**X**	
9	Do the orientations match for pairs of: neighboring blocks? features within a block?		**X**	**X**				**X**	
10	Are similarity scores alike: with respect to neighboring blocks? within a block?		**X**	**X**					**X**
11	Are the paired features within a block contiguous?			**X**		**X**			
12	How large is a feature relative to other genes within a block?			**X**			**X**		
13	What are the sizes, locations, and names of features within a block?			**X**		**X**	**X**		
14	What are the differences between individual nucleotides of feature pairs?				**X**				**X**

[Meyer et al., MizBee: A Multiscale Synteny Browser, 2009]

Shneiderman's Mantra

Task Abstraction

- **Overview**: Gain an overview of the entire collection
- **Zoom**: Zoom in on items of interest
- **Filter**: filter out uninteresting items
- **Details-on-demand**: Select an item or group and get details when needed
- **Relate**: View relationships among items
- **History**: Keep a history of actions to support undo, replay, and progressive refinement
- **Extract**: Allow extraction of sub-collections and of the query parameters

[Shneiderman, 1996]

Shneiderman’s Visual Information Seeking Mantra

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

Heer+Shneiderman: Visual analysis tasks

Taxonomy of interactive dynamics for visual analysis.

Data and View Specification	**Visualize** data by choosing visual encodings.
	Filter out data to focus on relevant items.
	Sort items to expose patterns.
	Derive values or models from source data.
View Manipulation	**Select** items to highlight, filter, or manipulate them.
	Navigate to examine high-level patterns and low-level detail.
	Coordinate views for linked, multidimensional exploration.
	Organize multiple windows and workspaces.
Process and Provenance	**Record** analysis histories for revisitation, review, and sharing.
	Annotate patterns to document findings.
	Share views and annotations to enable collaboration.
	Guide users through analysis tasks or stories.

Empirical Study: Amar+Eagan+Stasko

Task Abstraction

4 AN ANALYTIC TASK TAXONOMY

The ten tasks from the affinity diagramming analysis are:

- Retrieve Value
- Filter
- Compute Derived Value
- Find Extremum
- Sort
- Determine Range
- Characterize Distribution
- Find Anomalies
- Cluster
- Correlate

[Amar, Eagan, & Stasko, 2005]

Examples:
- Order the cars by weight.
- Rank the cereals by calories.

[Amar, Eagan, & Stasko, 2005]

1) **Filter**: Find data that satisfies conditions
2) **Find Extremum**: Find data with extreme values
3) **Sort**: Rank data according to some metric
4) **Determine Range**: Find span of data values
5) **Find Anomalies**: Find data with unexpected / extreme values

Examples:
- What Kellogg's cereals have high fiber?
- What comedies have won awards?
- Which funds underperformed the SP-500?

[Amar, Eagan, & Stasko, 2005]

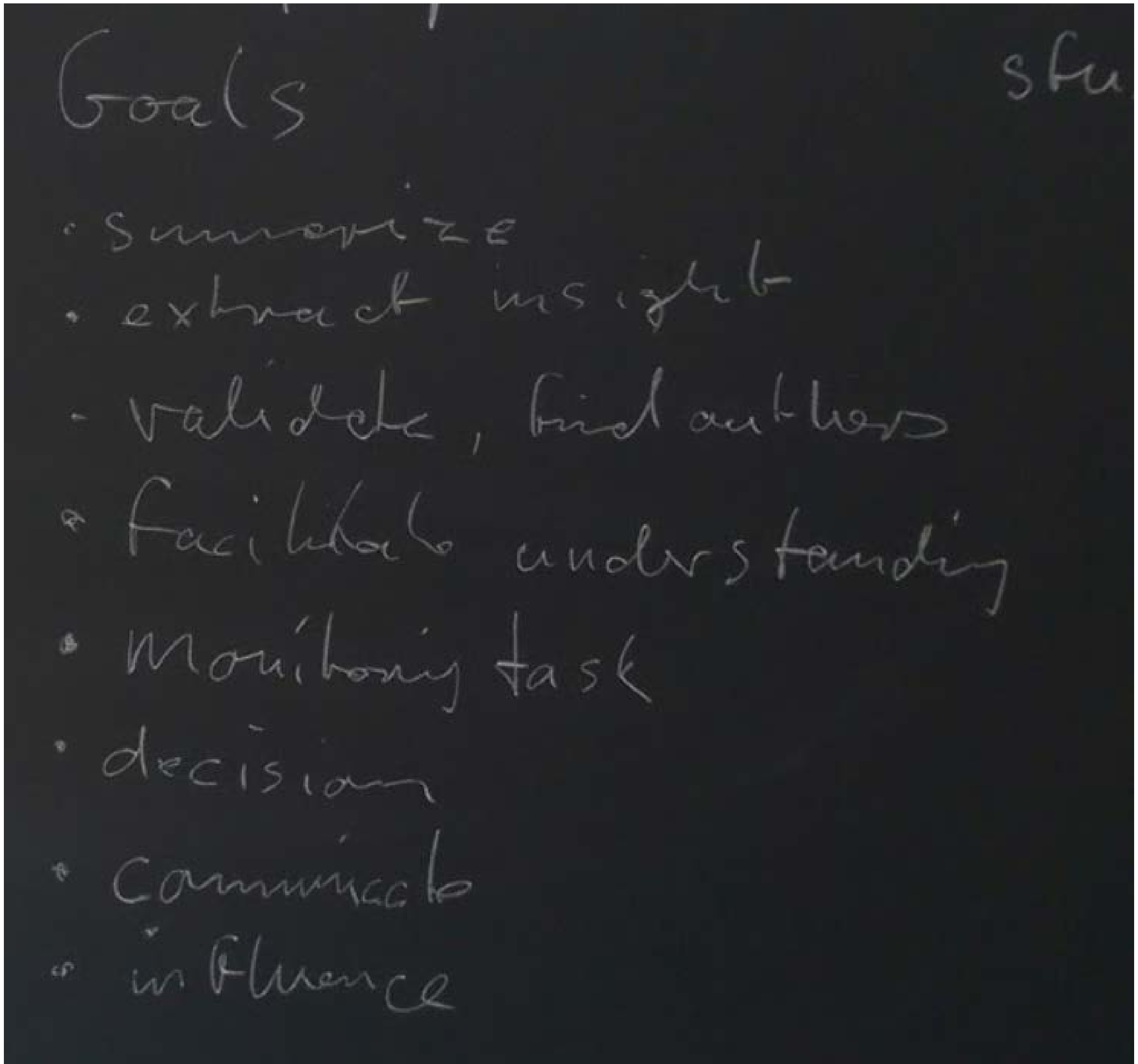
1) **Filter**: Find data that satisfies conditions
2) **Find Extremum**: Find data with extreme values
3) **Sort**: Rank data according to some metric
4) **Determine Range**: Find span of data values
5) **Find Anomalies**: Find data with unexpected / extreme values

Examples:
- Are there exceptions to the relationship between horsepower and acceleration?
- Are there any outliers in protein?

[Amar, Eagan, & Stasko, 2005]

1) **Filter**: Find data that satisfies conditions
2) **Find Extremum**: Find data with extreme values
3) **Sort**: Rank data according to some metric
4) **Determine Range**: Find span of data values
5) **Find Anomalies**: Find data with unexpected / extreme values

Task examples from you



Taxonomy: **Why?**
Brehmer + Munzner

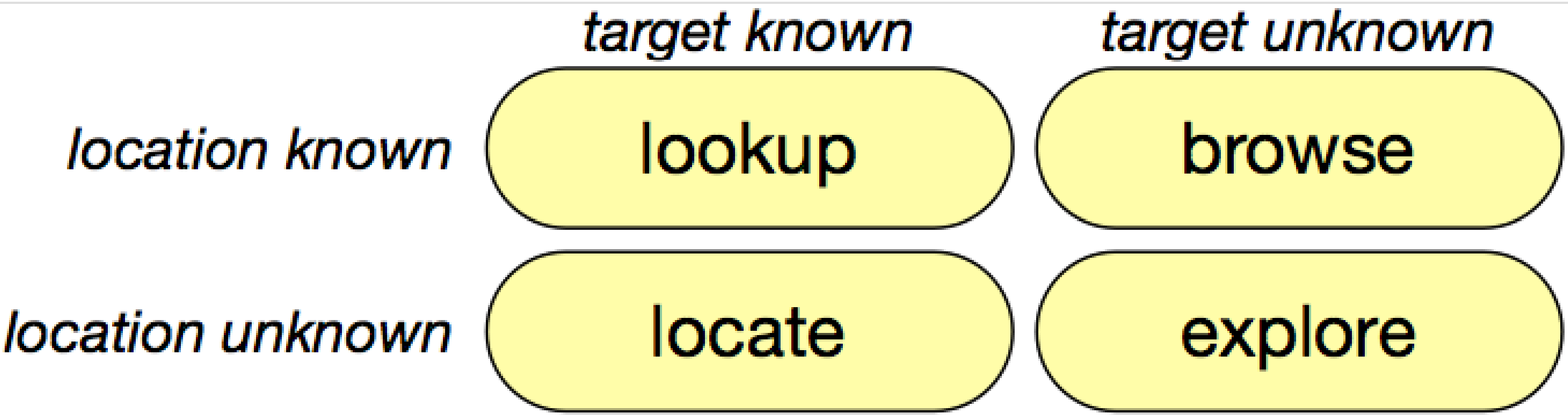
Why (Tasks!)

- Munzner has a hierarchy of
  - consume / produce
  - search
  - query

Consume vs. produce

- Produce
  - help the user produce vis!
- Consume (most common)
  - present
    - not just static (e.g. interactive graphics in newspapers / NY Times)
  - discover
    - generation / verification of hypothesis
  - enjoy
    - “casual” vis
    - e.g. Name Voyager (http://www.babynamewizard.com/voyager)

Search


target known
target unknown
location known
lookup
browse
location unknown
locate
explore

Query

- identify
  – refers to a single target
- compare
  – refers to two or multiple targets
- summarize
  – refers to whole set of targets

Taxonomy: **How?**
Brehmer + Munzner

How (what interactions enable the tasks)?

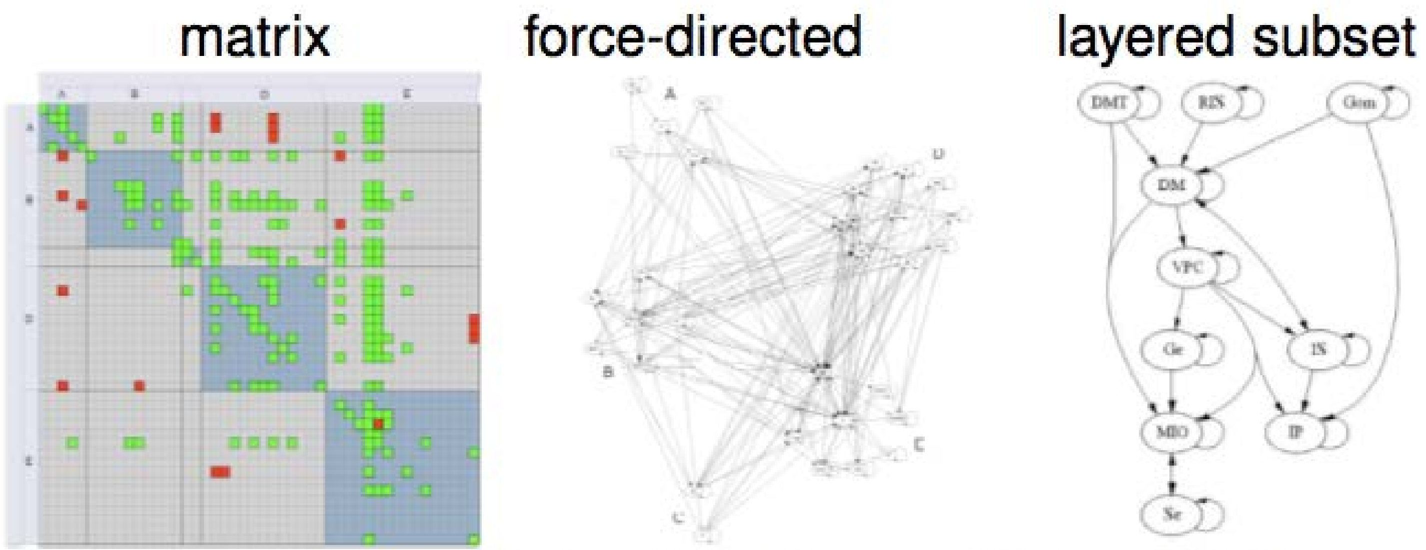
- Munzner considers these categories of how to manipulate visualizations:
  - introduce
  - encode
  - facet
  - reduce

Introduce

- import
  - new data items to be loaded
- derive
- annotate
  - with text label etc. (classification)
  - acts as a new attribute
- record
  - screenshots, bookmarks, parameter settings, logs, etc.
  - graphical / use history
  - analytical provenance!

Encode

- through channels and marks
- e.g. color, shapes, size, position etc.
- e.g. different visual encodings of a graph:

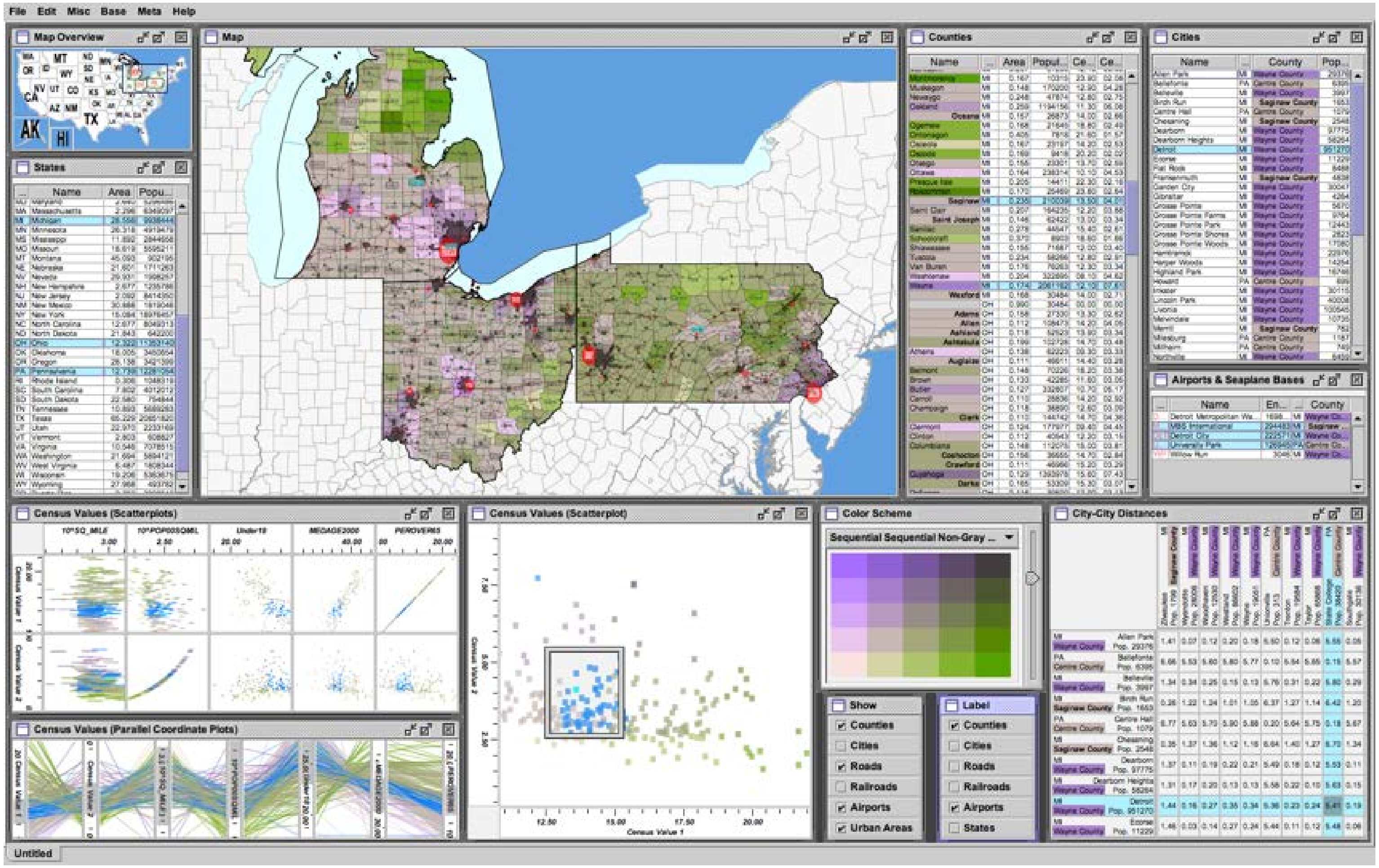


[van Ham, Using Multilevel Call Matrices in Large Software Projects. InfoVis03
http://www.win.tue.nl/~fvham/DL/callmatrix.pdf]

Facet

- how to use views:
  - partition (side-by-side, simultaneously)
  - superimpose (multiple layers)
  - change (layout, encoding --> interaction)
  - select (demarcation, highlighting)
  - coordinate (brushing+linking, linking views)

Partition


File Edit Misc Base Meta Help
Map Overview
Map
Counties
Cities
States
Airports & Seaplane Bases
Census Values (Scatterplots)
Census Values (Scatterplot)
Color Scheme
Sequential Sequential Non-Gray ...
City-City Distances
Census Values (Parallel Coordinate Plots)
Show
Counties
Cities
Roads
Railroads
Airports
Urban Areas
Label
Counties
Cities
Roads
Railroads
Airports
States
Untitled

Linking (coordinate)

Reduce

- reduce (increase) number of elements shown
  - filter
  - aggregate
  - navigate (alter viewpoint, e.g. zooming, detail-on-demand)
  - embed (focus+context)

Embed — fisheye lens