CS 571 - Data Visualization & Exploration

Data Abstraction

Instructor: Hamza Elhamdadi

UMassAmherst

Upcoming Dates

Feb 18 (Today): Group Activity 2 (due at 5pm)

Feb 28: Project Proposal Due

Feb 21: Homework 2 will be released (due Mar 7)

Visualization Idioms

Visualization Idioms

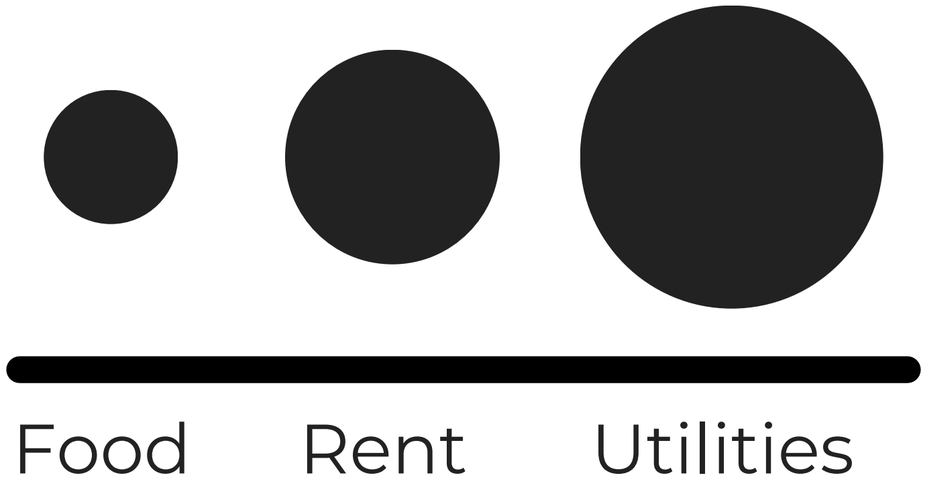
A vis idiom is a distinct approach for creating and manipulating visual representations

Category	Budget
Food	500
Rent	1000
Utilities	200

Visualization Idioms

A vis idiom is a distinct approach for creating and manipulating visual representations

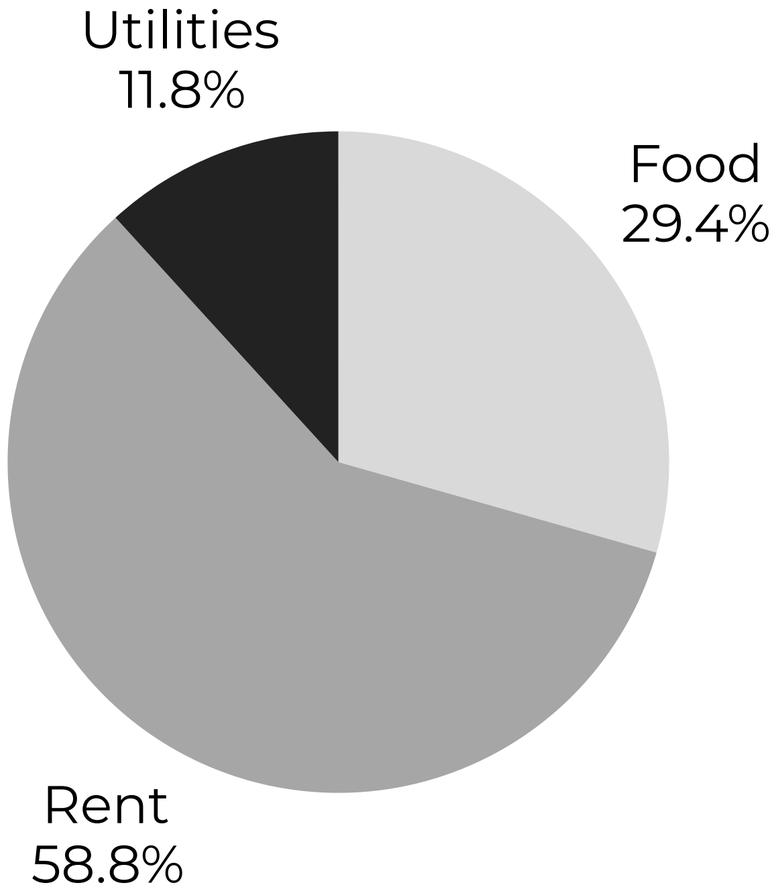
Category	Budget
Food	500
Rent	1000
Utilities	200



Visualization Idioms

A vis idiom is a distinct approach for creating and manipulating visual representations

Category	Budget
Food	500
Rent	1000
Utilities	200

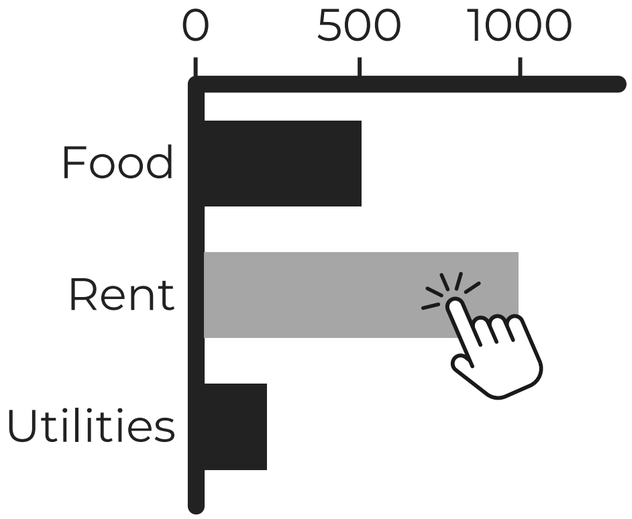


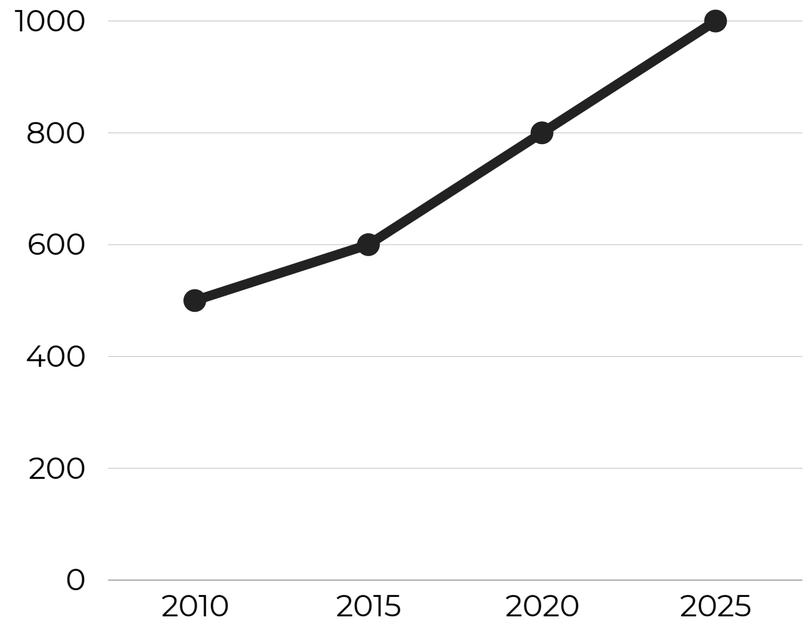
Visualization Idioms

A vis idiom is a distinct approach for creating and manipulating visual representations

There are many ways to visually represent the same data

The design space of possibilities gets even larger when we think about interactions





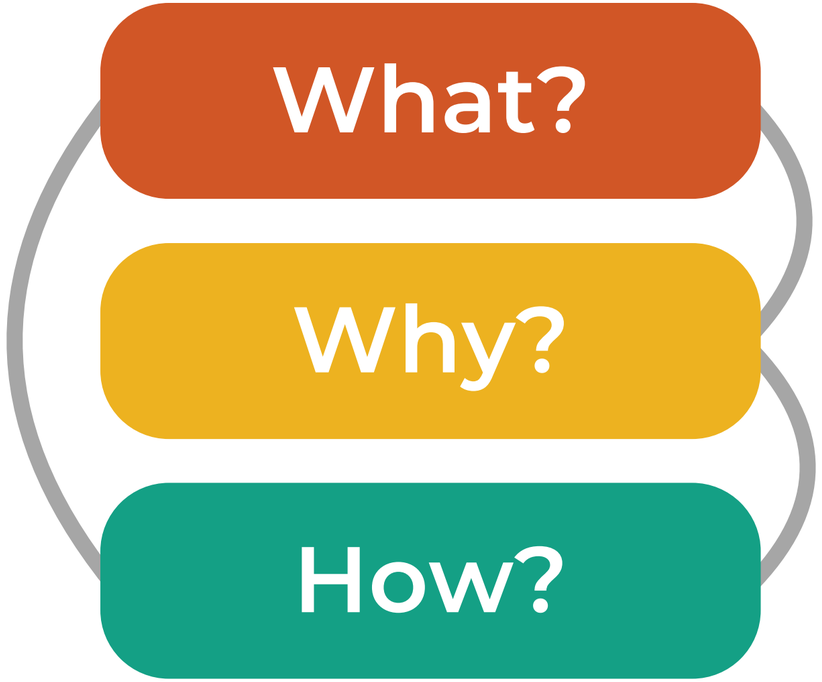
Analyzing Data Visualizations

Analyzing Data Visualizations

What is shown?

Why is the user looking at it?

How is it shown?



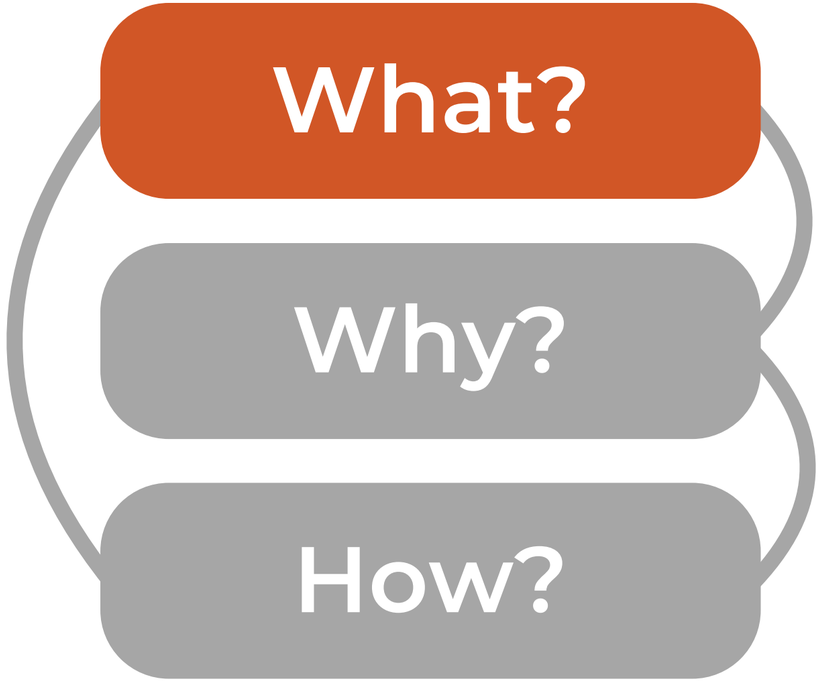
Abstract vocabulary allows us to avoid domain-specific terms

What-Why-How framework is a scaffold we can use to systematically think about the huge visualization design space

Data Abstraction

The **what** part pertaining to the data

Data abstraction is the translation of domain-specific terms into more general terms



Data Types

Attribute:

- a specific property that can be measured, observed, or logged
- e.g., salary, price, number of sales

Item:

- an individual entity that is discreet
- e.g., a row in a simple table, or a node in a network

Link:

- a relationship between items (usually in a network)

Position:

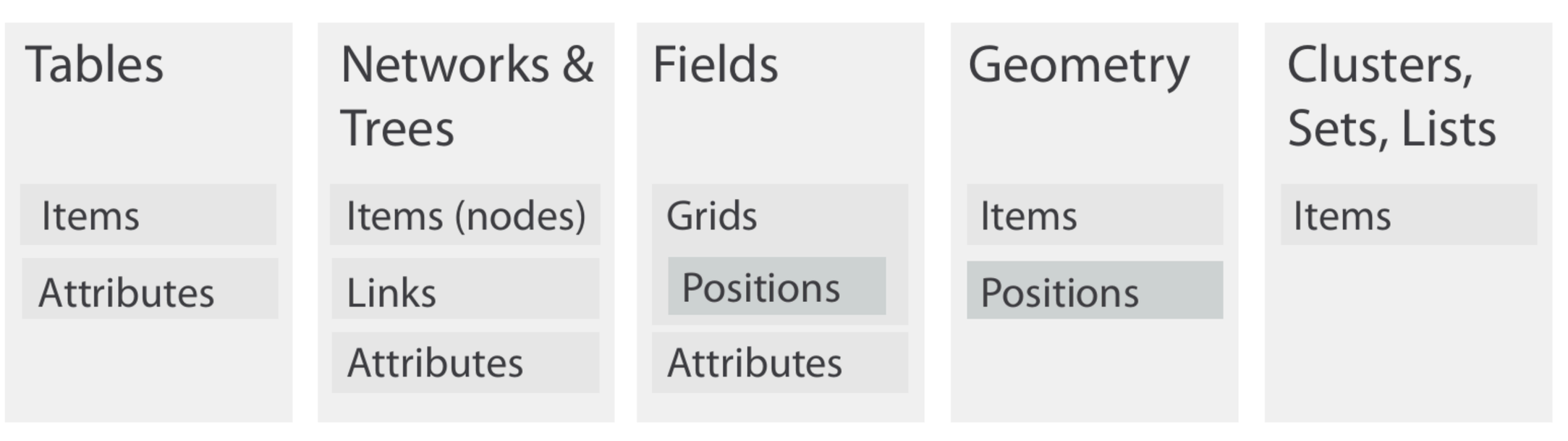
- spatial data that specifies a location in 2D or 3D space

Grid:

- specifies the geometric/topological relationship between positions

Dataset Types

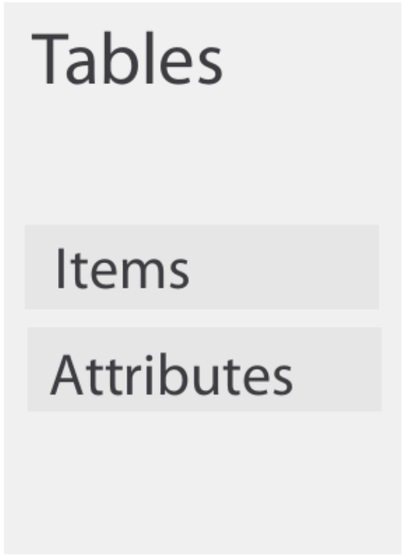
Dataset Types



Tables

Consist of items and attributes in rows and columns

Each cell in the table is fully specified by an item-attribute pair



Order Id	Order Date	Container	Ship Date	Priority
3	10/04/06	Large B	/06	Low
6	2/21/08	Small Pa	08	Not Specified
32	7/16/07	Small Pack	7/17/07	High
32	7/16/07	Jumbo Box	7/17/07	High
32	7/16/07	Medi	/18/07	High
32	7/16/07	Medium Box	7/18/07	High
25	10/23/07	Wrap Bag	10/24/07	Not Specified
25	10/23/07	Small Box	10/25/07	Not Specified

attribute

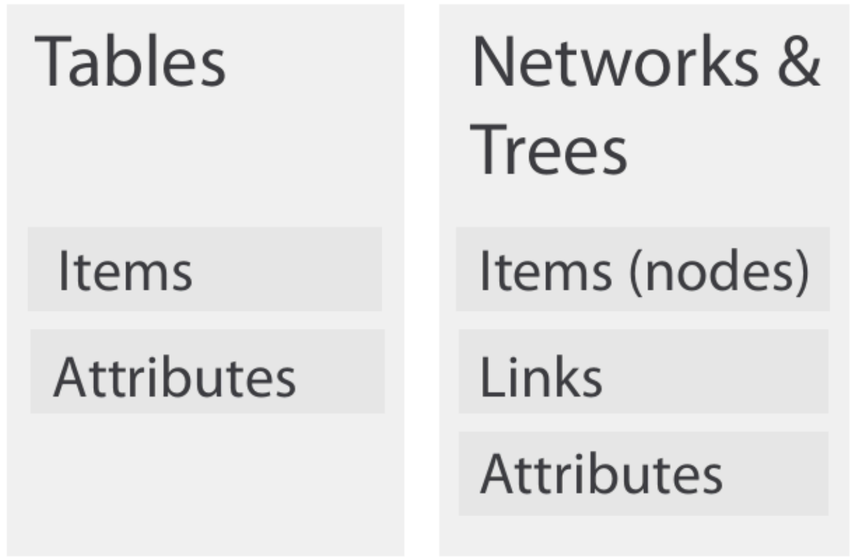
item

cell

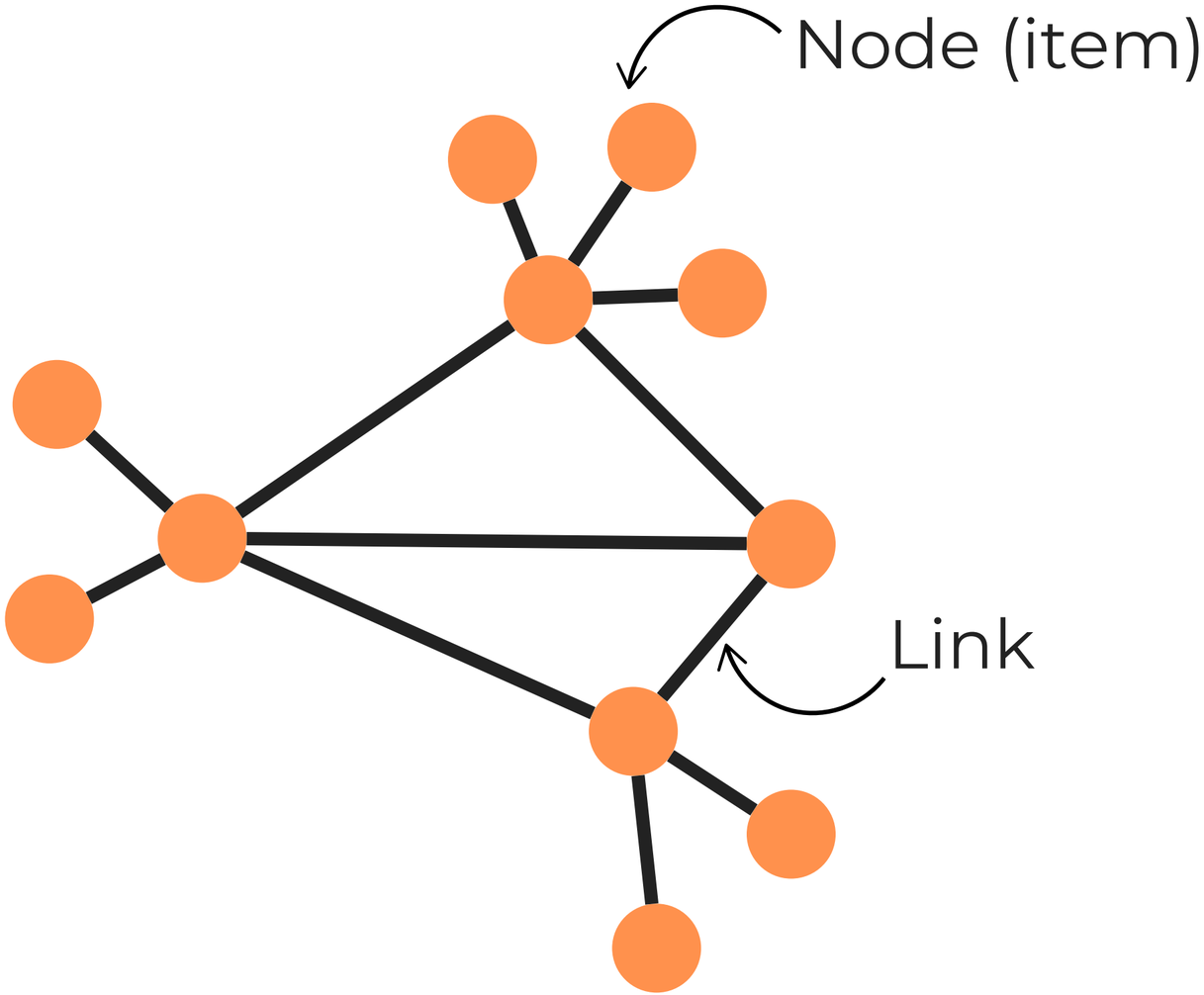
Networks & Trees

Consist of items (nodes), links (edges) and attributes

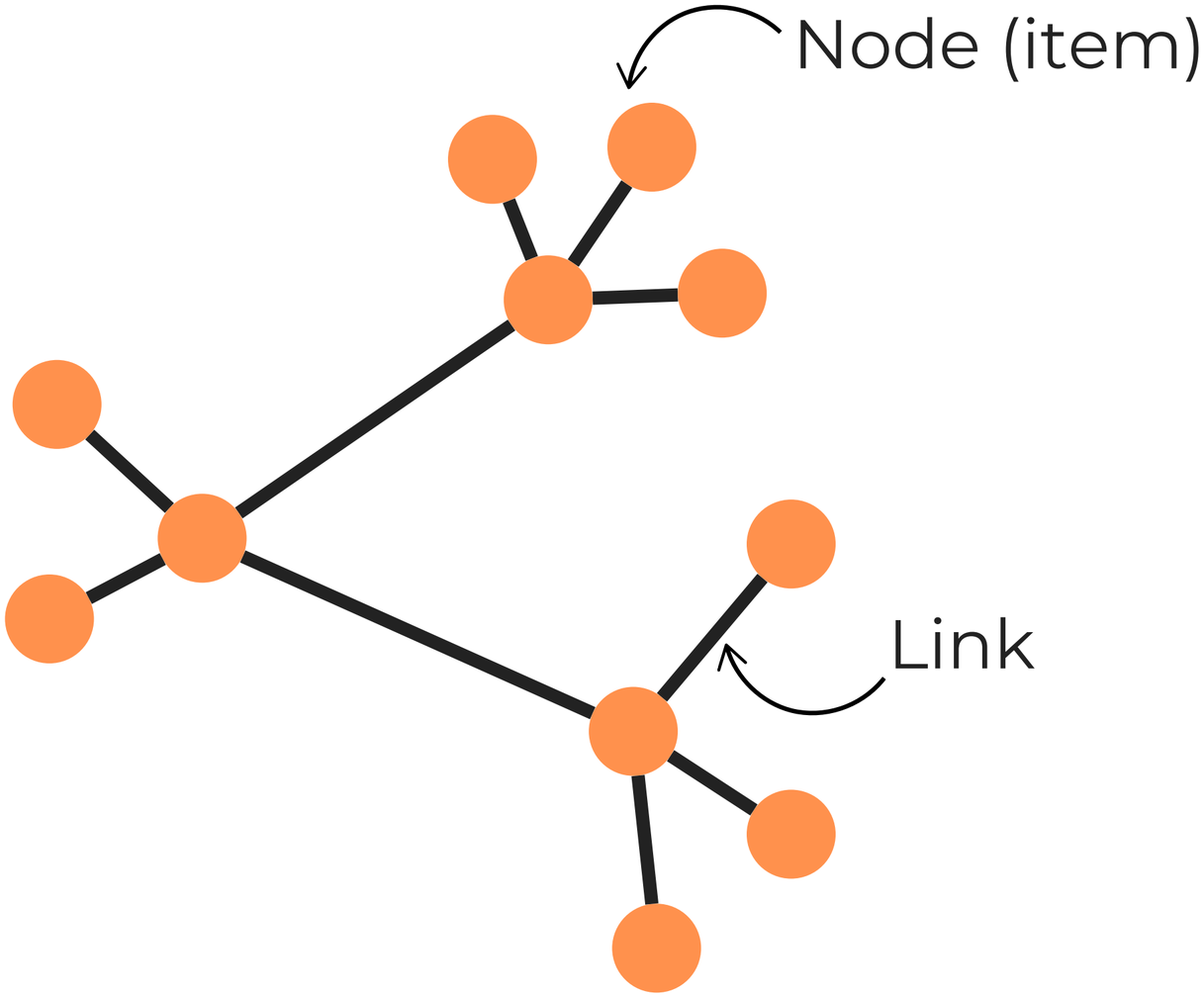
Useful for specifying relationships between two or more items



Networks



Trees

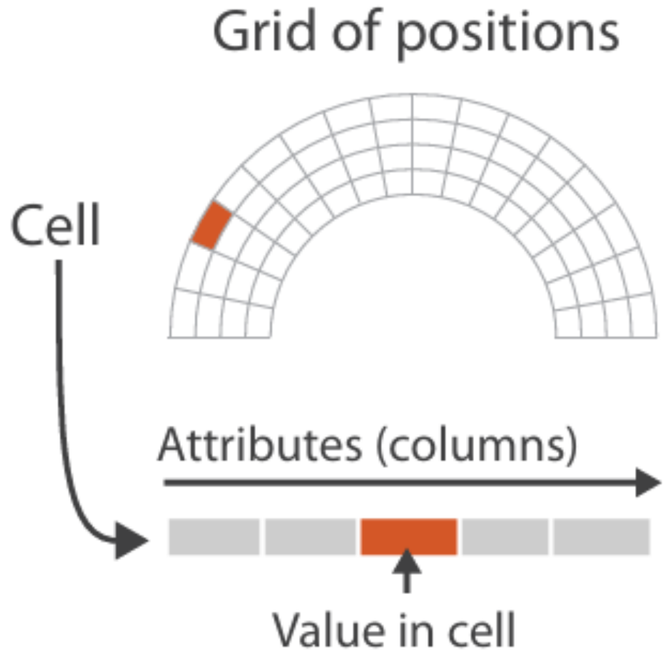


Fields

A field dataset contains attributes associated with cells

Each cell contains attributes associated with measurements from a continuous domain, defined by a grid

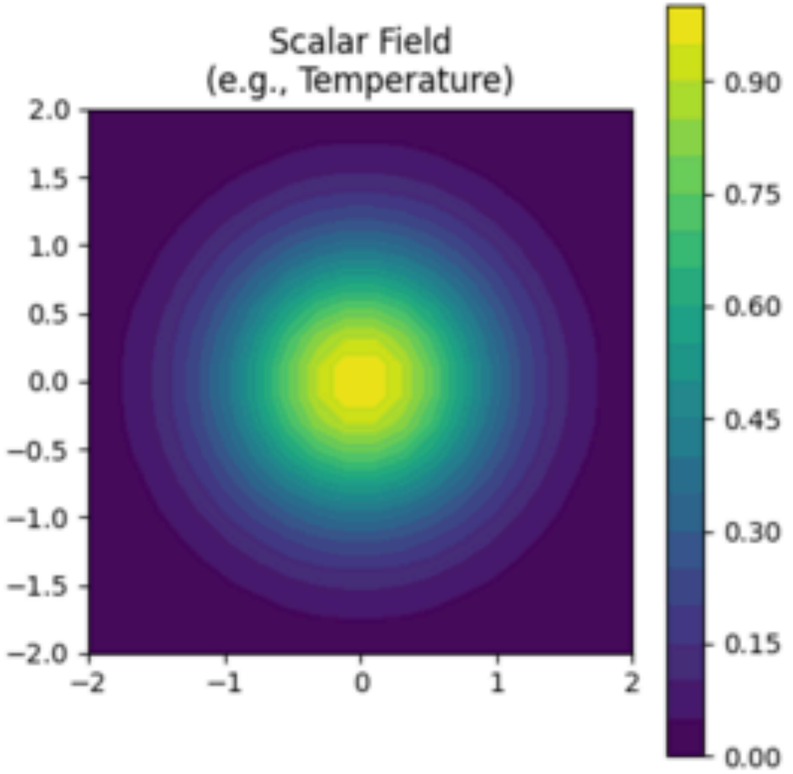
Tables	Networks & Trees	Fields
Items	Items (nodes)	Grids
Attributes	Links	Positions
	Attributes	Attributes



Fields

Scalar Fields:

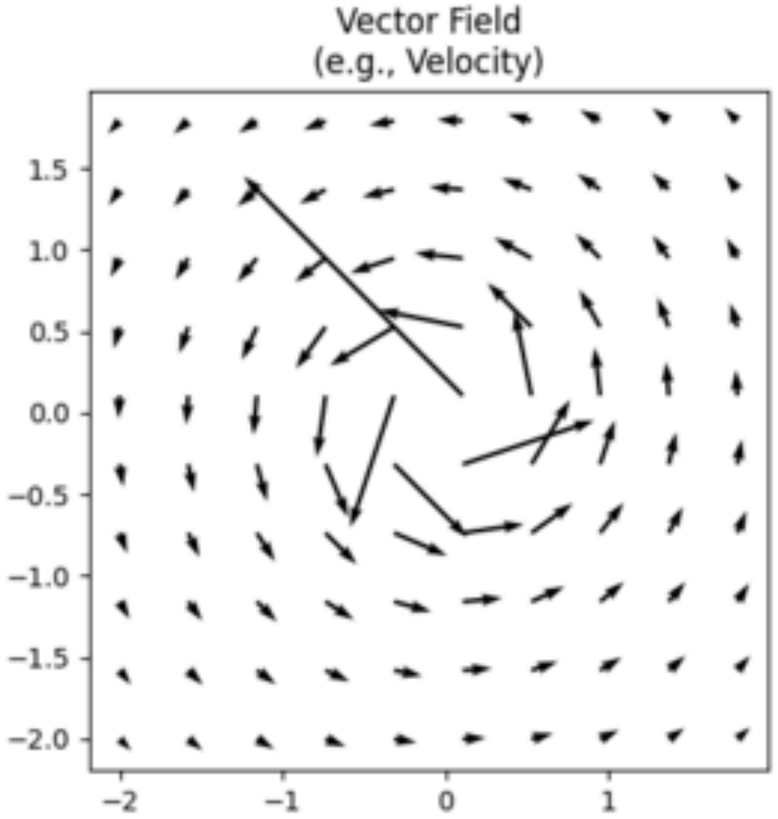
- univariate (one attribute at each position in the grid)



Fields

Vector Fields:

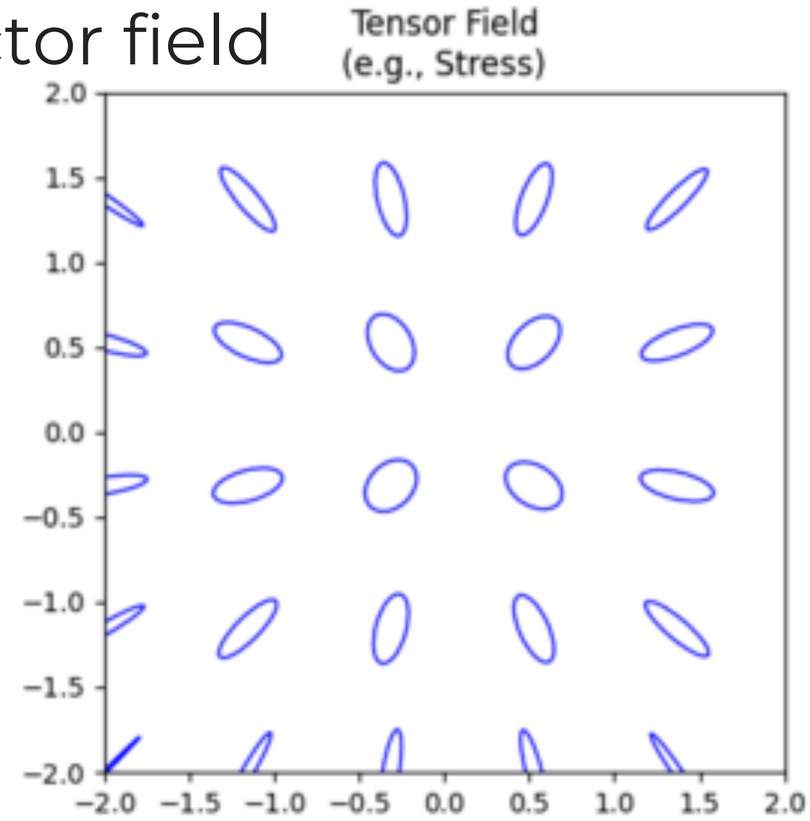
- multivariate (list of attributes at each position in the grid)



Fields

Tensor Fields:

- multivariate (array of attributes assigned to each position)
- generalized version of a tensor and vector field



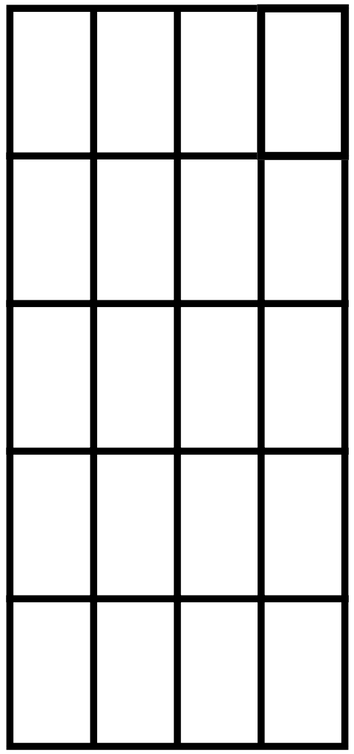
Grid Choices Impact How Continuous Data is Interpreted

Two key considerations:

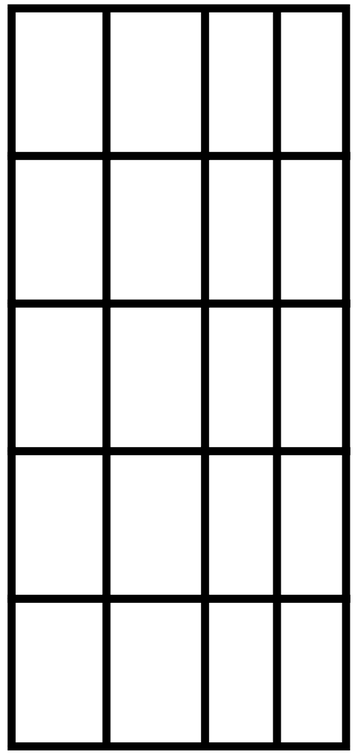
- **Sampling:**
  - how frequently to take measurements
- **Interpolation:**
  - how to show values between measurements without misleading

Grid Types

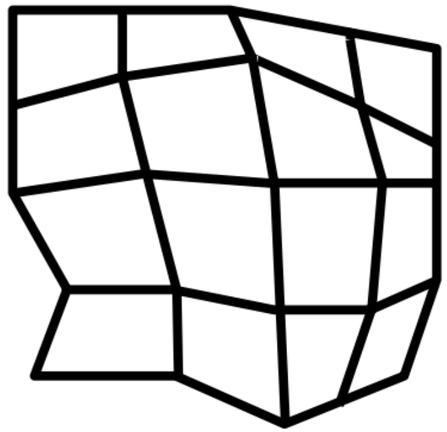
Uniform



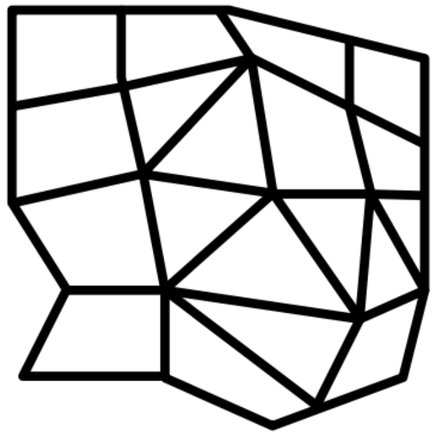
Rectilinear



Structured

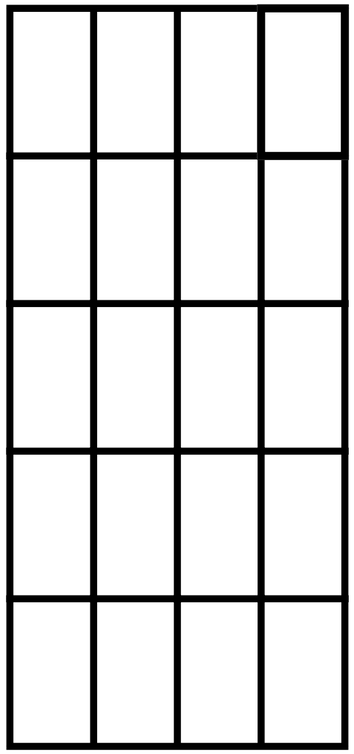


Unstructured



Grid Types

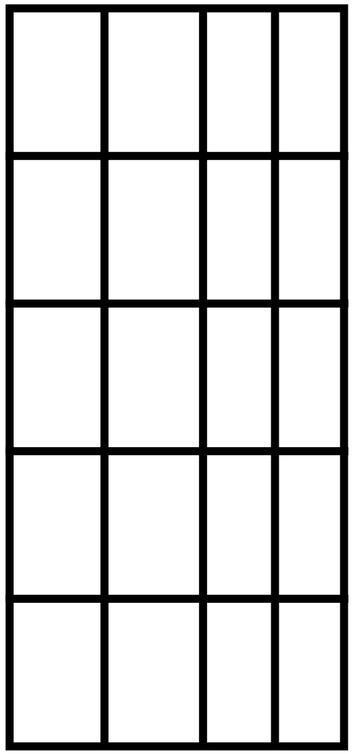
Uniform



Points are sampled uniformly (at regular intervals)

Grid Types

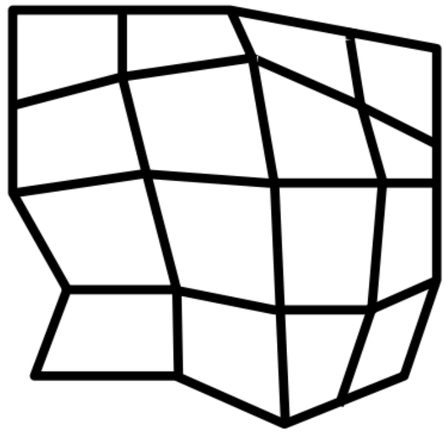
Rectilinear



Points are sampled nonuniformly

Grid Types

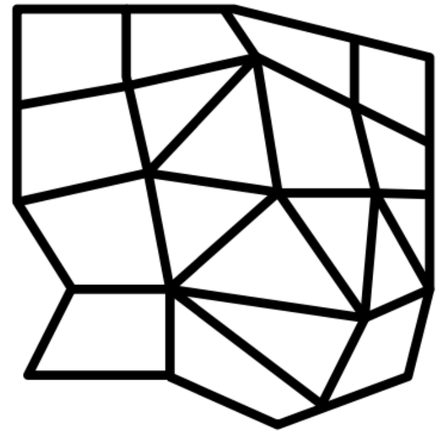
Structured



Allows curvilinear shapes, need to specify the geometry of the cells in the grid

Grid Types

Unstructured

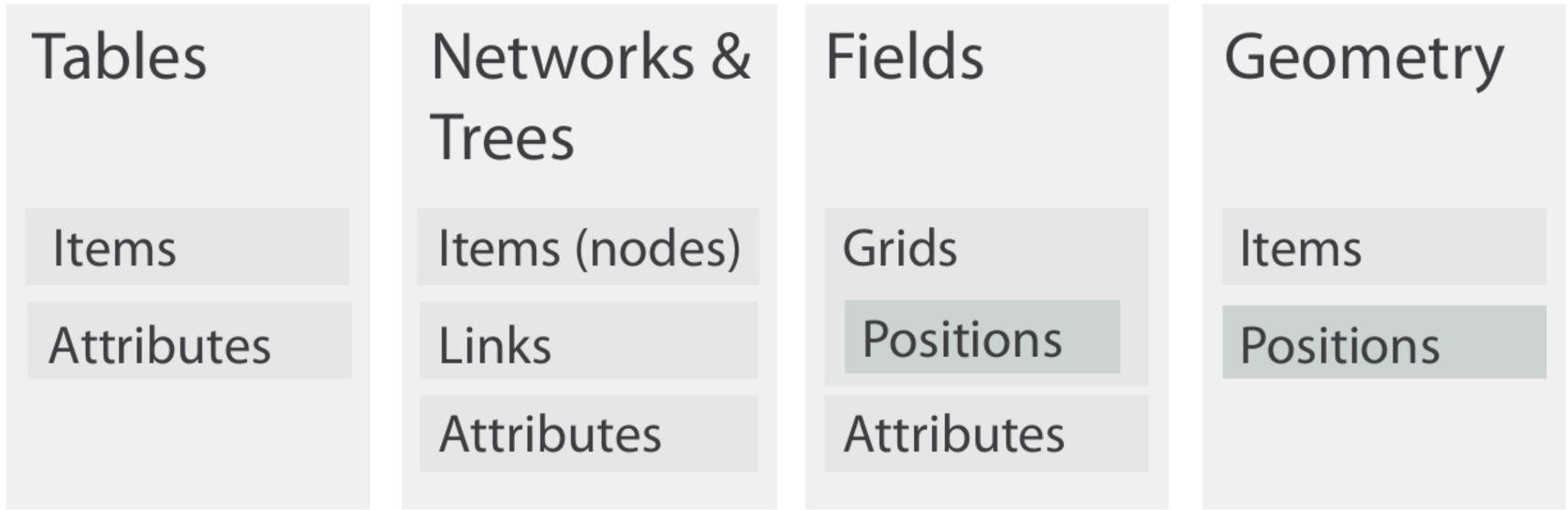


Complete geometrical freedom, but topological information about how cells connect to each other must be maintained

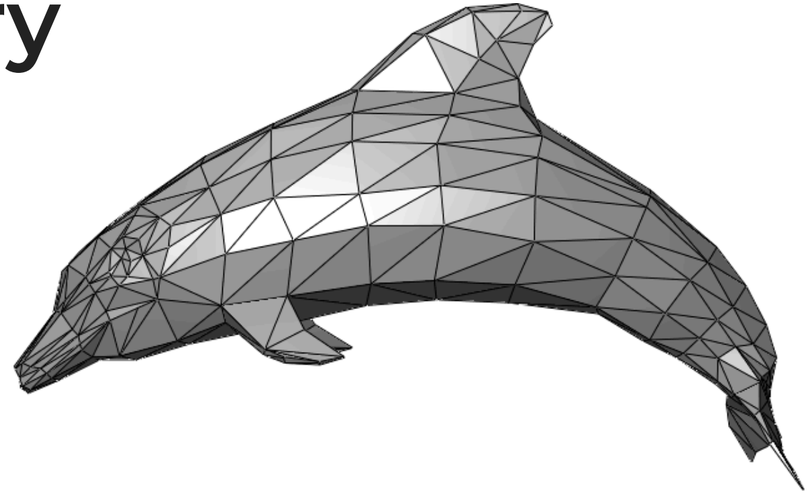
Geometry

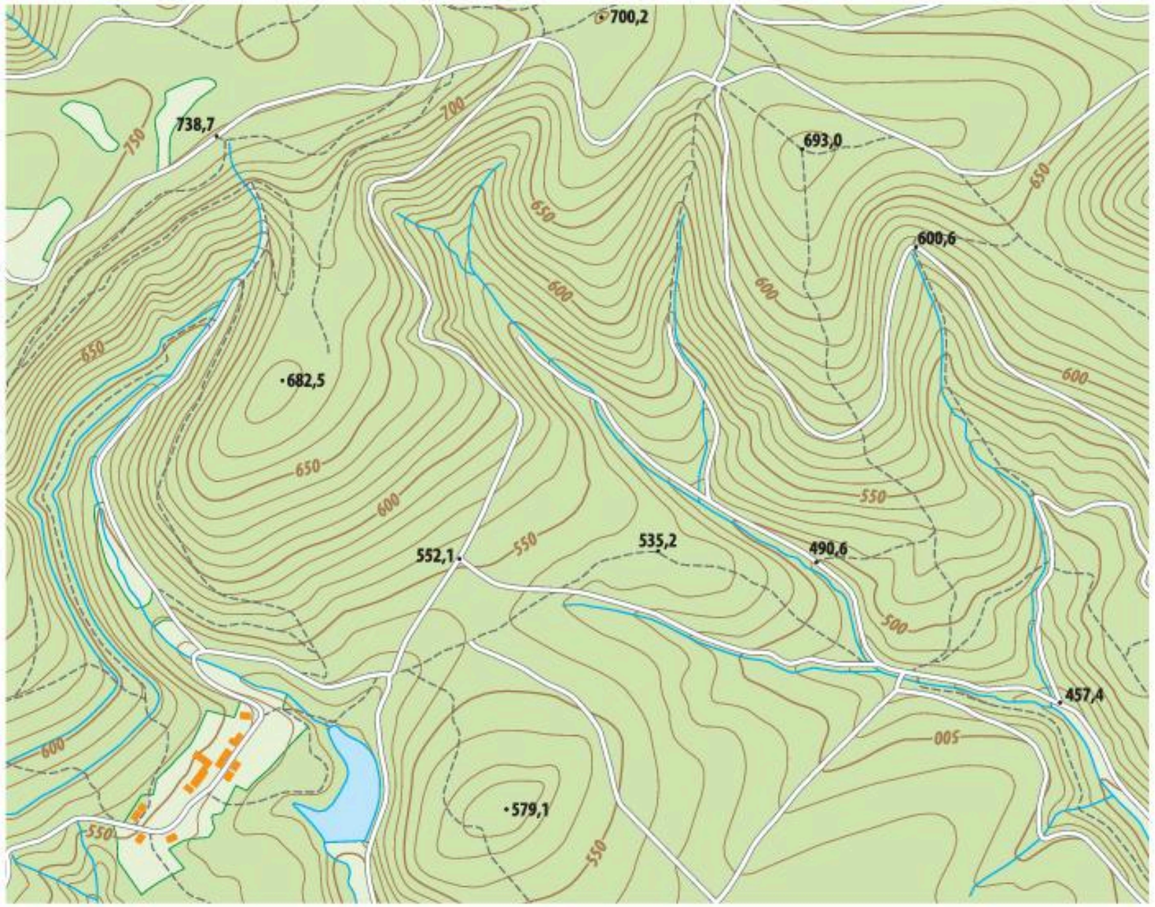
A geometry dataset specifies the spatial position of its items

Items in a geometry dataset do not necessarily have attributes



Geometry

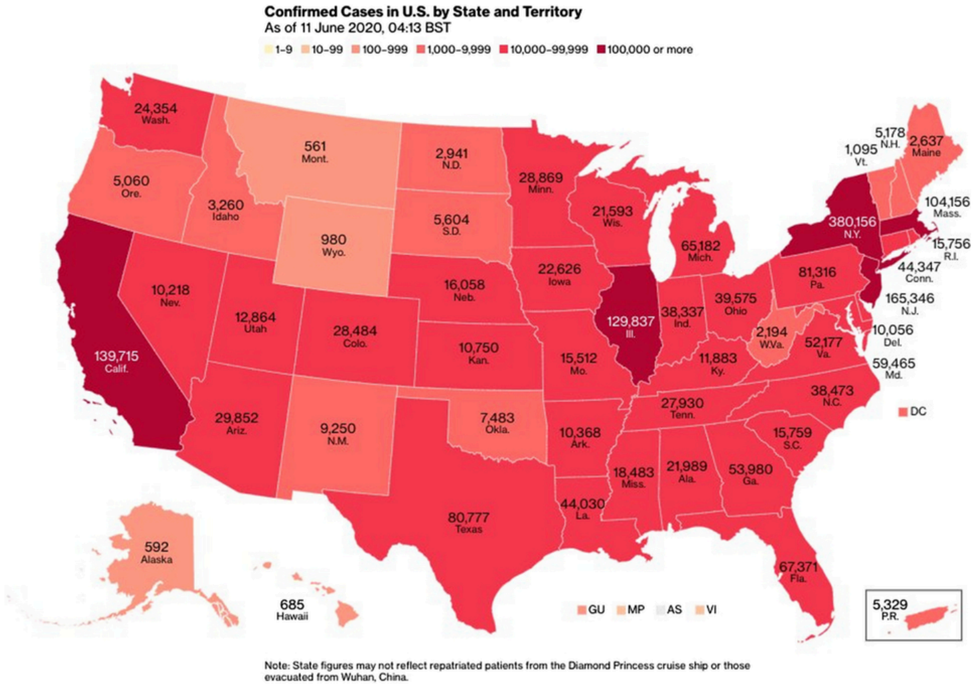




Geometry

From a Data Vis perspective, geometry datasets are not that interesting by themselves

Instead, they are useful as a backdrop on which to overlay other data

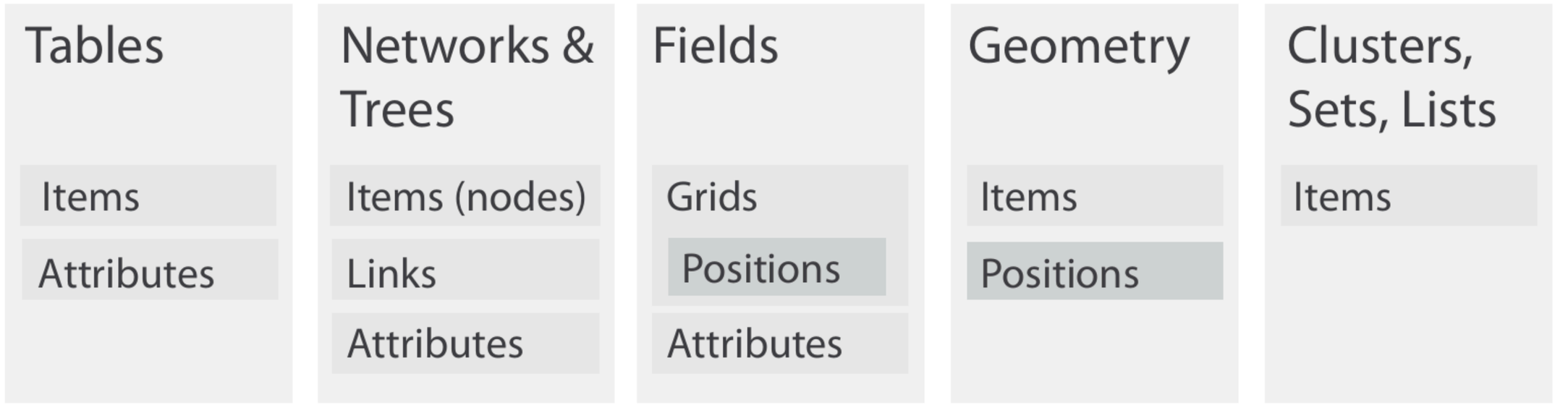


Clusters, Sets, and Lists

A set is an unordered set of items

A list is an ordered set of items

A cluster is a grouping based on attribute similarity



Attribute Types

Attribute Types

Categorical:

- no implicit ordering
- e.g., name, hometown, favorite color

Ordinal:

- has an implicit ordering
- cannot do full-fledged arithmetic
- e.g., shirt size, education level, tournament placing

Quantitative:

- is ordered, and supports arithmetic comparison
- e.g., height, temperature, stock price, number of drinks sold

Order Id	Order Date	Container	Ship Date	Priority
3	10/04/06	Large Box	10/21/06	Low
6	2/21/08	Small Pack	2/22/08	Not Specified
32	7/16/07	Small Pack	7/17/07	High
32	7/16/07	Jumbo Box	7/17/07	High
32	7/16/07	Medium Box	7/18/07	High
32	7/16/07	Medium Box	7/18/07	High
25	10/23/07	Wrap Bag	10/24/07	Not Specified
25	10/23/07	Small Box	10/25/07	Not Specified

Order Id	Order Date	Container	Ship Date	Priority
3	10/04/06	Large Box	10/21/06	Low
6	2/21/08	Small Pack	2/22/08	Not Specified
32	7/16/07	Small Pack	7/17/07	High
32	7/16/07	Jumbo Box	7/17/07	High
32	7/16/07	Medium Box	7/18/07	High
32	7/16/07	Medium Box	7/18/07	High
25	10/23/07	Wrap Bag	10/24/07	
25	10/23/07	Small Box	10/25/07	

categorical
ordinal
quantitative

Ordering Direction

Ordering Direction

Some attributes (ordinal and quantitative) have an order

The direction of their order can be one of three types:

- Sequential
- Diverging
- Cyclic

Ordering Direction

Sequential:

- goes from a minimum value to a maximum value
- e.g., mountain height dataset, bathymetric dataset

Diverging:

- can be deconstructed into two sequential orders that meet at a common zero point
- e.g., elevation dataset

Cyclic:

- the values wrap around back to a starting point
- e.g., month of the year, day of the week

Derived Attributes

We can derived new attributes from existing attributes by:

- simply **changing the attribute type**
  - temperature in degrees >> hot, warm, cold
- **acquiring additional information**
  - city name >> latitude, longitude pairs
- **using arithmetic, logical, or statistical operations**
  - computing the difference of two quantitative atttributes
  - computing the mean of a single attribute

We can also transform datasets into new ones of different type by creating derived data

Attribute Type vs. Attribute Semantic

Data Model: mathematical abstraction (data abstraction)

- focused on attribute types
- which operations are allowed (e.g., + - * /)

Conceptual Model: mental construction

- focused on semantics
- supports reasoning about idiom design choice

Attribute Type vs. Attribute Semantic (example)

Data Model:

- -32.52, 54.06, -17.35, … (float)

Conceptual Model:

- temperature

To new data abstraction:

- round to 2 significant figures (quantitative)
- convert to “hot”, “warm”, “cold” (ordinal)
- “above freezing”, “below freezing” (categorical)

FIN

Upcoming Dates

Feb 18 (Today): Group Activity 2 (due at 5pm)

Feb 28: Project Proposal Due

Feb 21: Homework 2 will be released (due Mar 7)