

DATA 230 Spring 2023

Week 2

Lecture 2:

Basic Concepts in Data Visualization

Dr. Seungjoon (Joon) Lee

DATA 230: Data Visualization

Human in the Loop

1. Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.
2. Don't need visualization when fully automatic solution exists and is trusted.
3. However, many analysis problems ill-specified.
 - don't know exactly what questions to ask in advance.
4. Start with a human in the loop.

Computer in the Loop

- Draw a visualization!
- Limitations:

Why Depend on Vision?

Biggest bandwidth

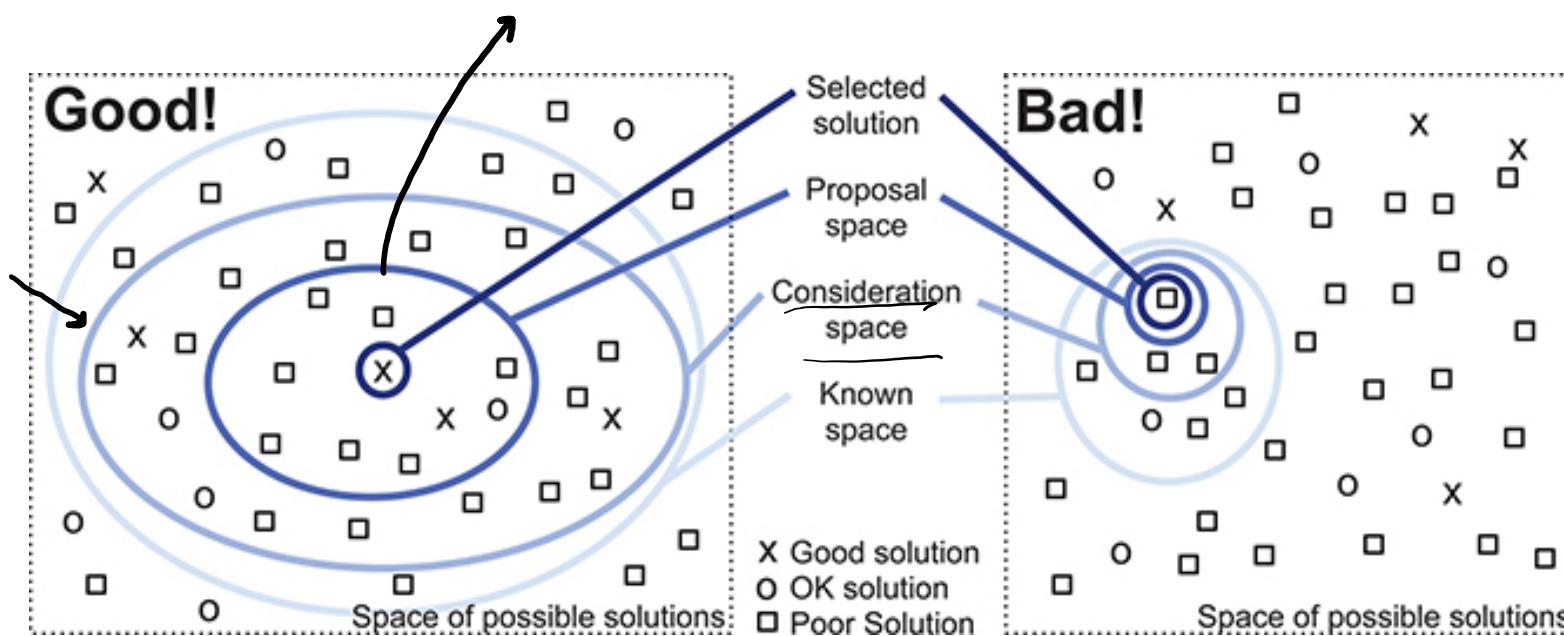
- Eyes (vision): 10,000,000 bps (10 MB/sec).
- Skin (touch): 1,000,000 bps.
- Ears (hearing): 100,000 bps.
- Nose (smell): 100,000 bps.
- Mouth (taste): 1,000 bps.

Accurate.

Technological Limitation.

Why Focus on Effectiveness

- It is not just about making pretty pictures.
- Making pictures to support user tasks (or hypothesis).
- Effectiveness vs ineffectiveness



Nested Model



Domain situation



Data/task abstraction



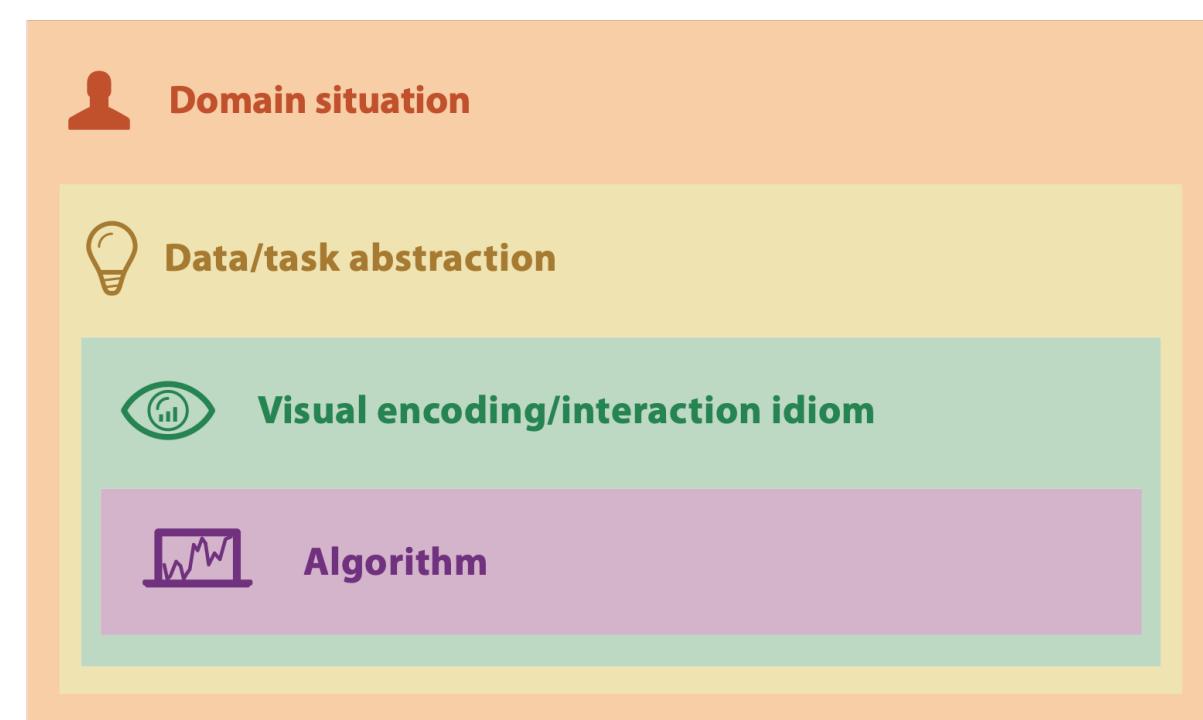
Visual encoding/interaction idiom



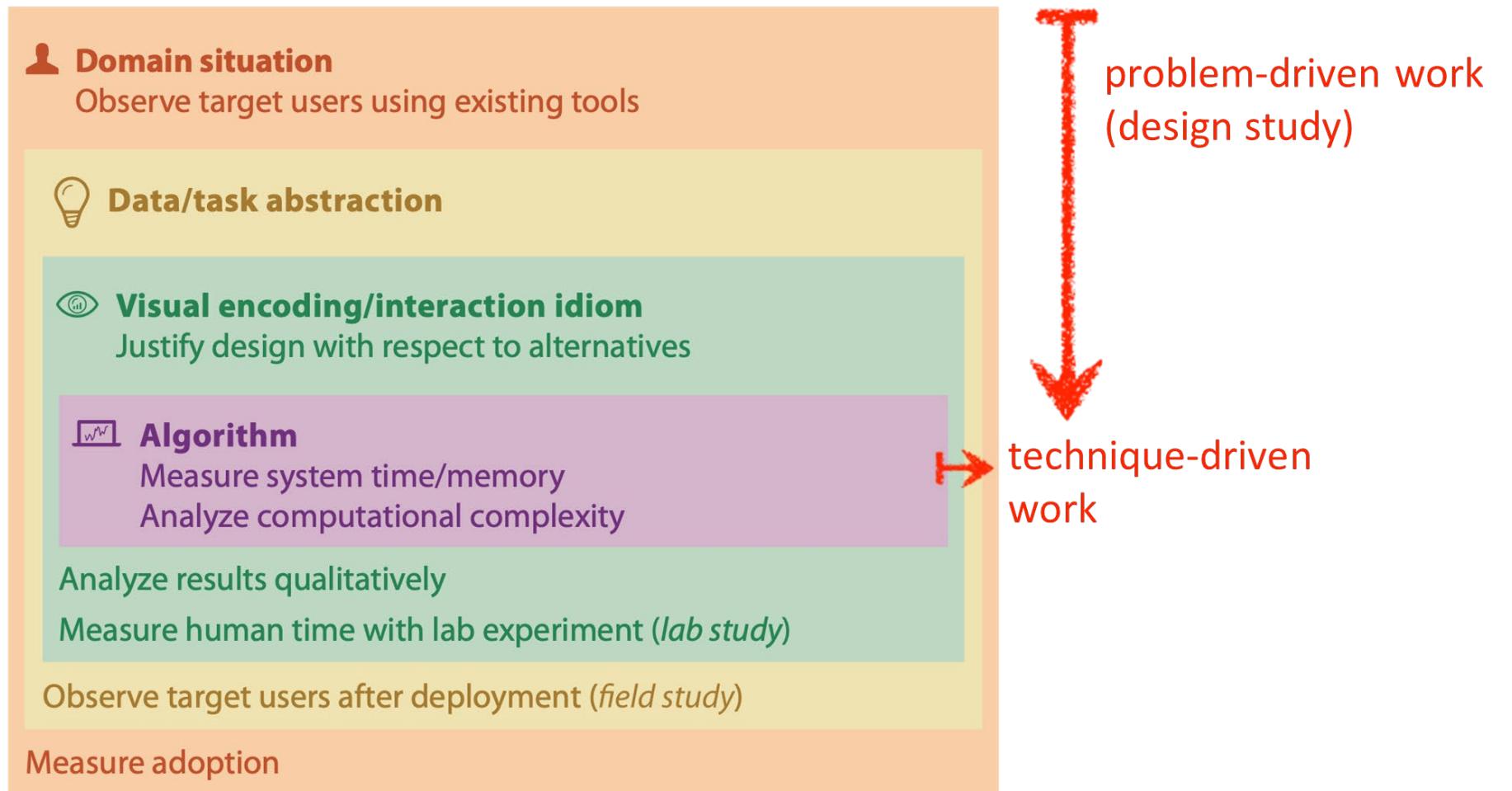
Algorithm

Nested Model

1. Domain situation: who are the target users?
2. Abstraction: translate from specifics of domain to vocabulary of visualization.
 - **what** is shown? **data abstraction**
 - **why** is the user looking at it? **task abstraction.**
3. Idiom: **how** is it shown?
 - visual encoding idiom: how to draw?
 - interaction idiom: how to manipulate?
4. Algorithm: efficient computation



Validation



Data Abstraction

Categorize and Characterize data types

- **10, 15, 20, 45, 75**: what does this sequence of five numbers mean?
- **Basil, 7, S, Pear**: what about this data?

Insights gained from such analysis can help to design a visualization system!

Data Semantics and Types

- Semantic:
- Type:

Name	Age	Shirt Size	Favorite Fruit
Amy	8	S	Apple
Basil	7	S	Pear
Clara	9	M	Durian
Desmond	13	L	Elderberry
Ernest	12	L	Peach
Fanny	10	S	Lychee
George	9	M	Orange
Hector	8	L	Loquat
Ida	10	M	Pear
Amy	12	M	Orange

Data, Dataset, and Attributes

- Data type: item, link, attribute
- Dataset type: table, tree, field, ...
- Attribute (feature/variable/dimension) type: numerical, quantity, category, ...

Name	Age	Shirt Size	Favorite Fruit
Amy	8	S	Apple
Basil	7	S	Pear
Clara	9	M	Durian
Desmond	13	L	Elderberry
Ernest	12	L	Peach
Fanny	10	S	Lychee
George	9	M	Orange
Hector	8	L	Loquat
Ida	10	M	Pear
Amy	12	M	Orange

Items and Features

1. Item: individual entity, discrete

2. Attribute: property that is measured, observed, logged...

Name	Age	Shirt Size	Favorite Fruit
Amy	8	S	Apple
Basil	7	S	Pear
Clara	9	M	Durian
Desmond	13	L	Elderberry
Ernest	12	L	Peach
Fanny	10	S	Lychee
George	9	M	Orange
Hector	8	L	Loquat
Ida	10	M	Pear
Amy	12	M	Orange

Data Type

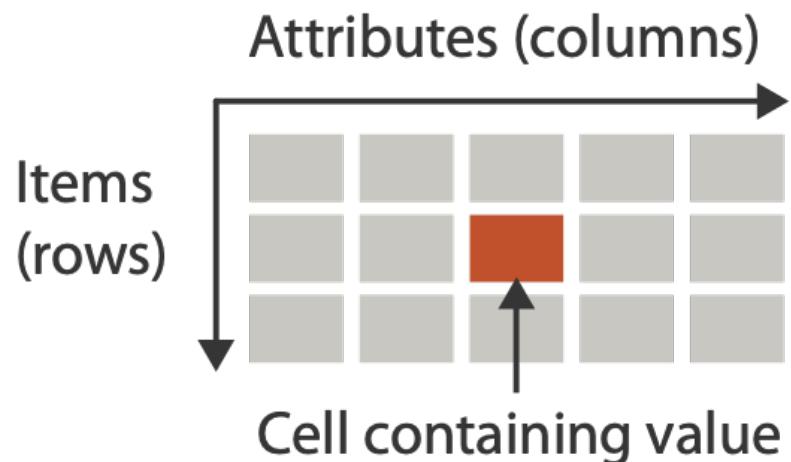
1. Item: individual entity (discrete).
2. Attribute (feature/variable/dimension): specific property.
3. Link: relationship between items (generally in a network), marriage relationship
4. Grid: geometric/topological relationship (sampling data) → temperature in the classroom.
5. Position: for spatial data (latitude, longitude), x-y-z dimension. MRI scan.

Dataset Type

1. Table = item (rows) + attribute (columns)
2. Network (or tree)= item (nodes) + link (links) (+ attribute/feature on node).
3. Fields = grid + position
4. Geometry = item + position
5. Cluster (Set/List) = item

Dataset Type - Table

Row = item, Column = attribute (variable)



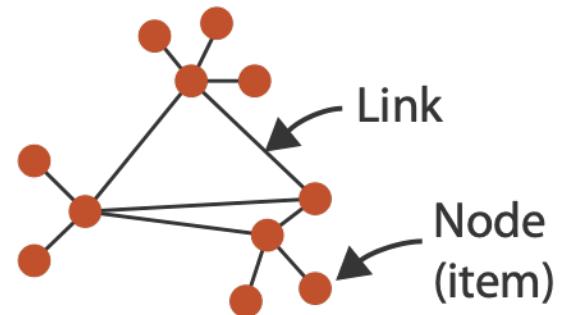
Name	Age	Shirt Size	Favorite Fruit
Amy	8	S	Apple
Basil	7	S	Pear
Clara	9	M	Durian
Desmond	13	L	Elderberry
Ernest	12	L	Peach
Fanny	10	S	Lychee
George	9	M	Orange
Hector	8	L	Loquat
Ida	10	M	Pear
Amy	12	M	Orange

Dataset Type – Network and Tree

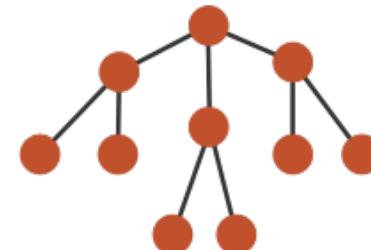
Node and link can have variable separately.

Example) Social Network on Facebook

Network (graph)

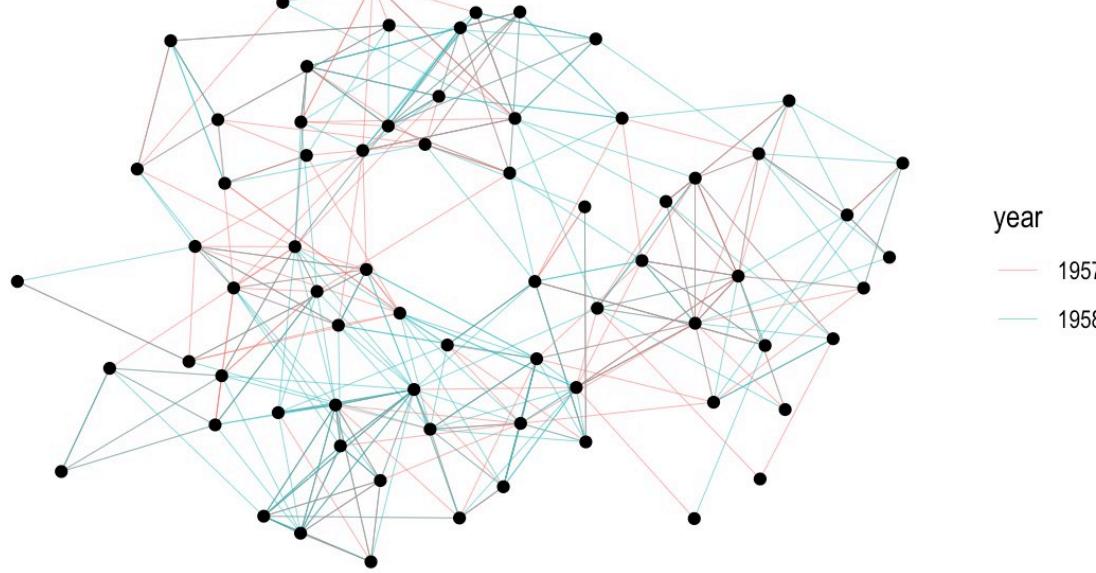


Tree

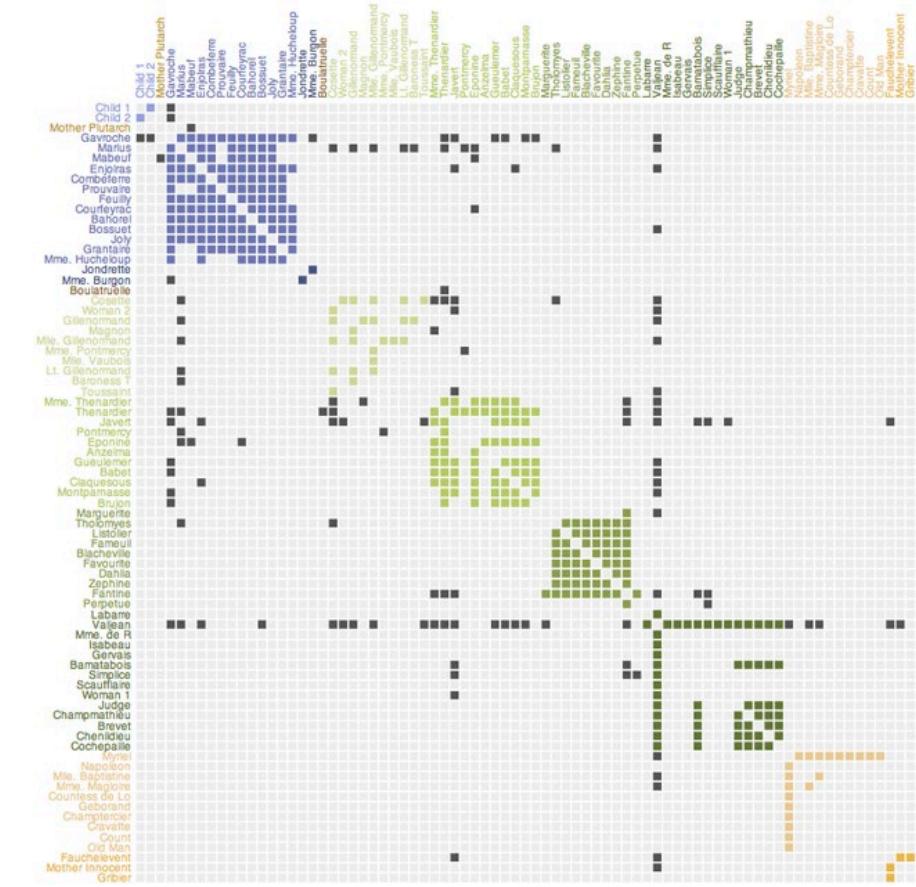


Dataset Type – Network and Tree

Visualization



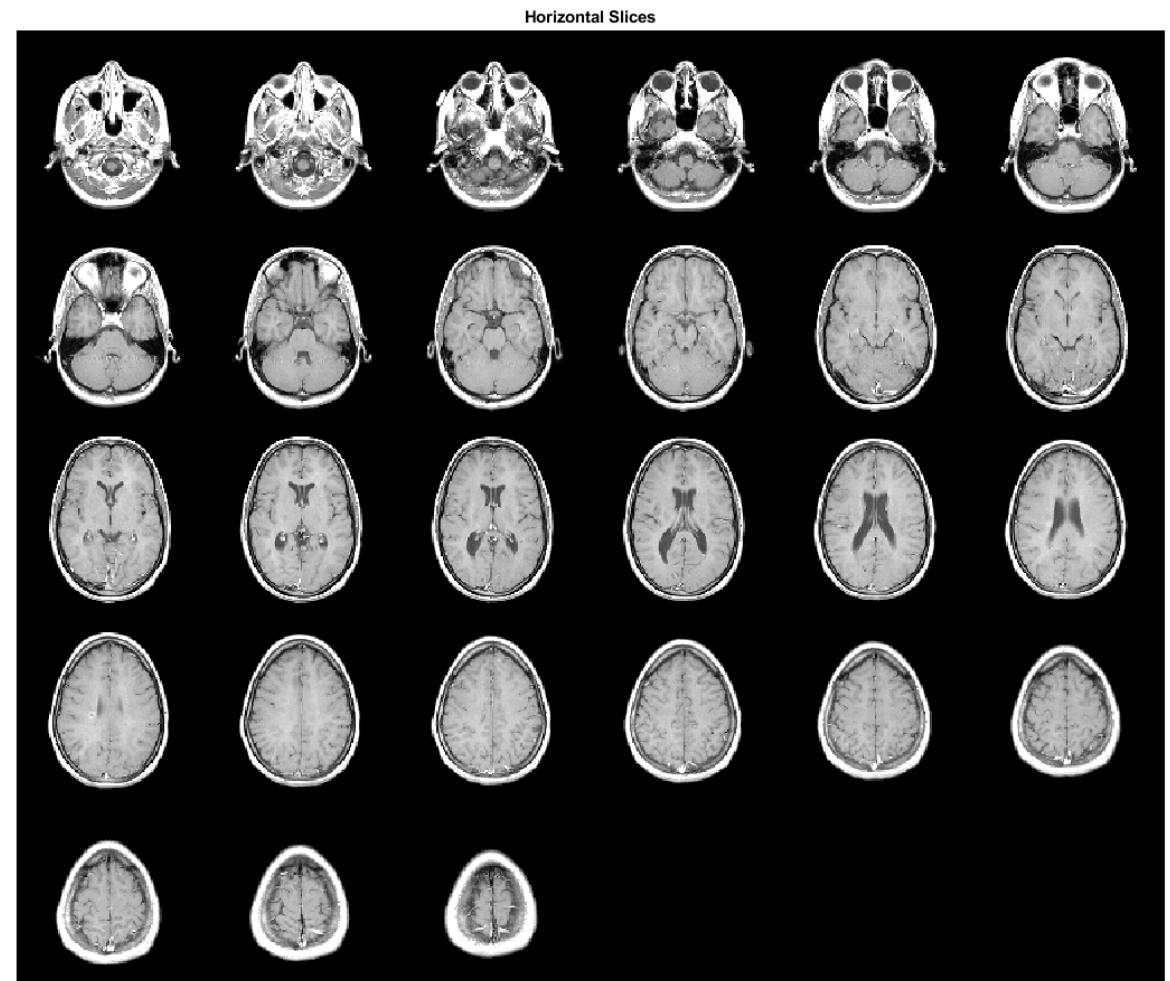
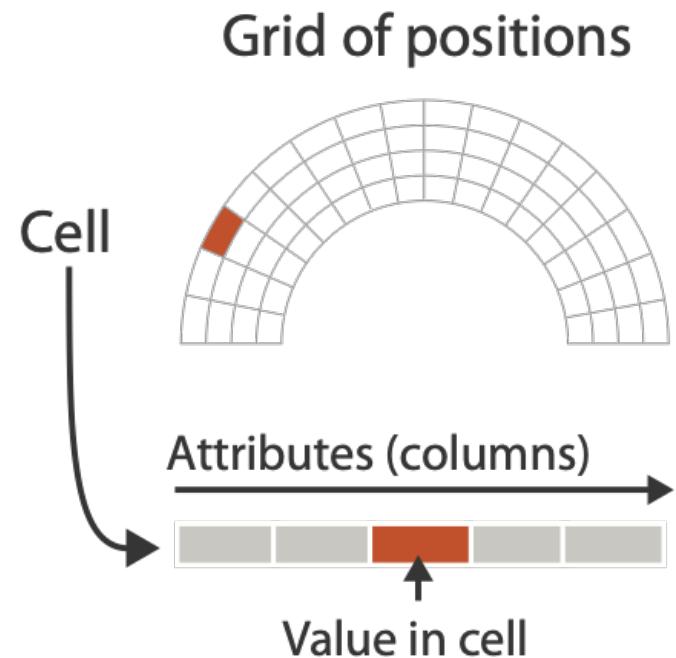
Node-link diagram



Adjacency matrix

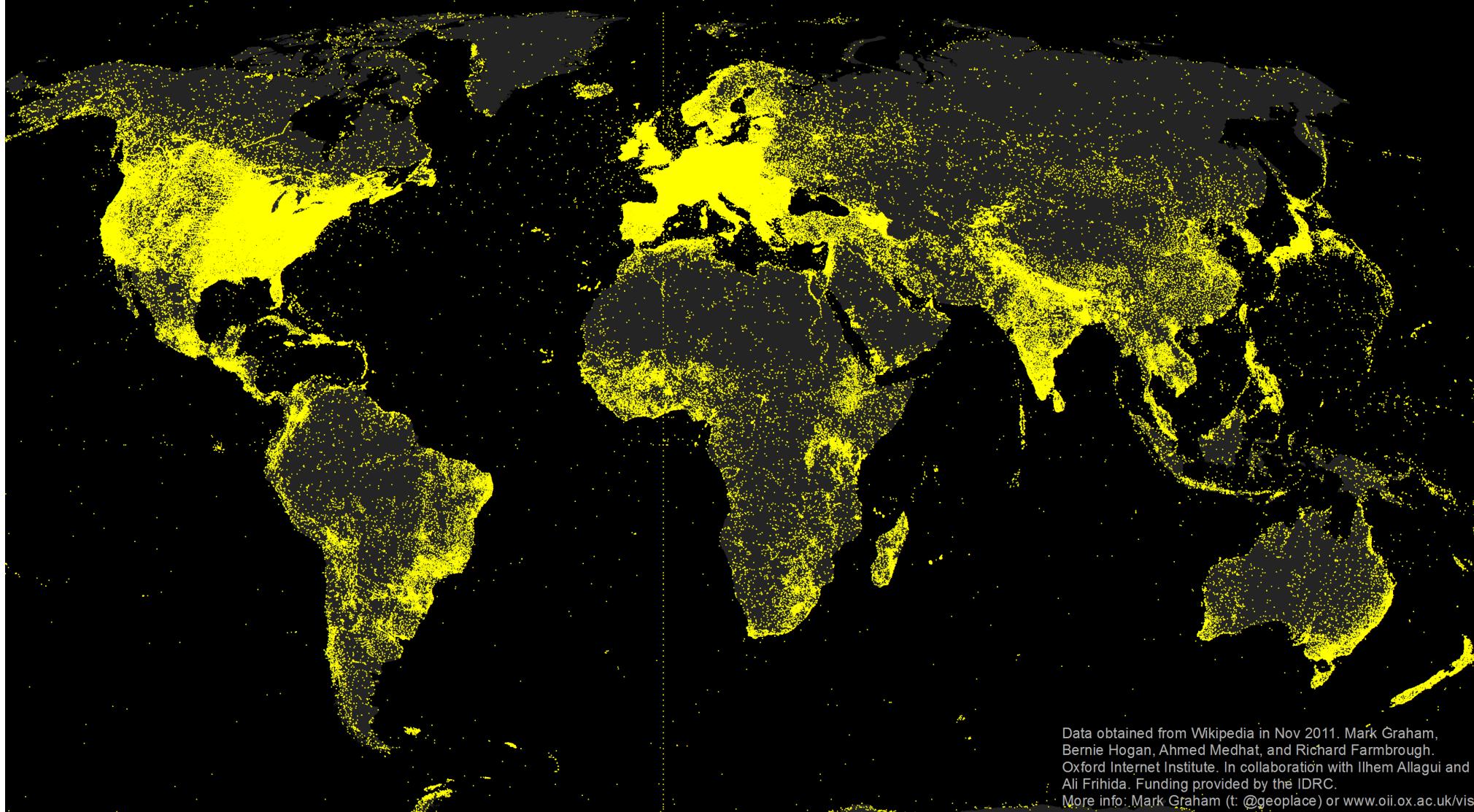
Dataset Type - Field

Sampling on grid from a continuous domain.



Dataset Type – Position (Spatial Data)

Geotagged Articles in English Wikipedia

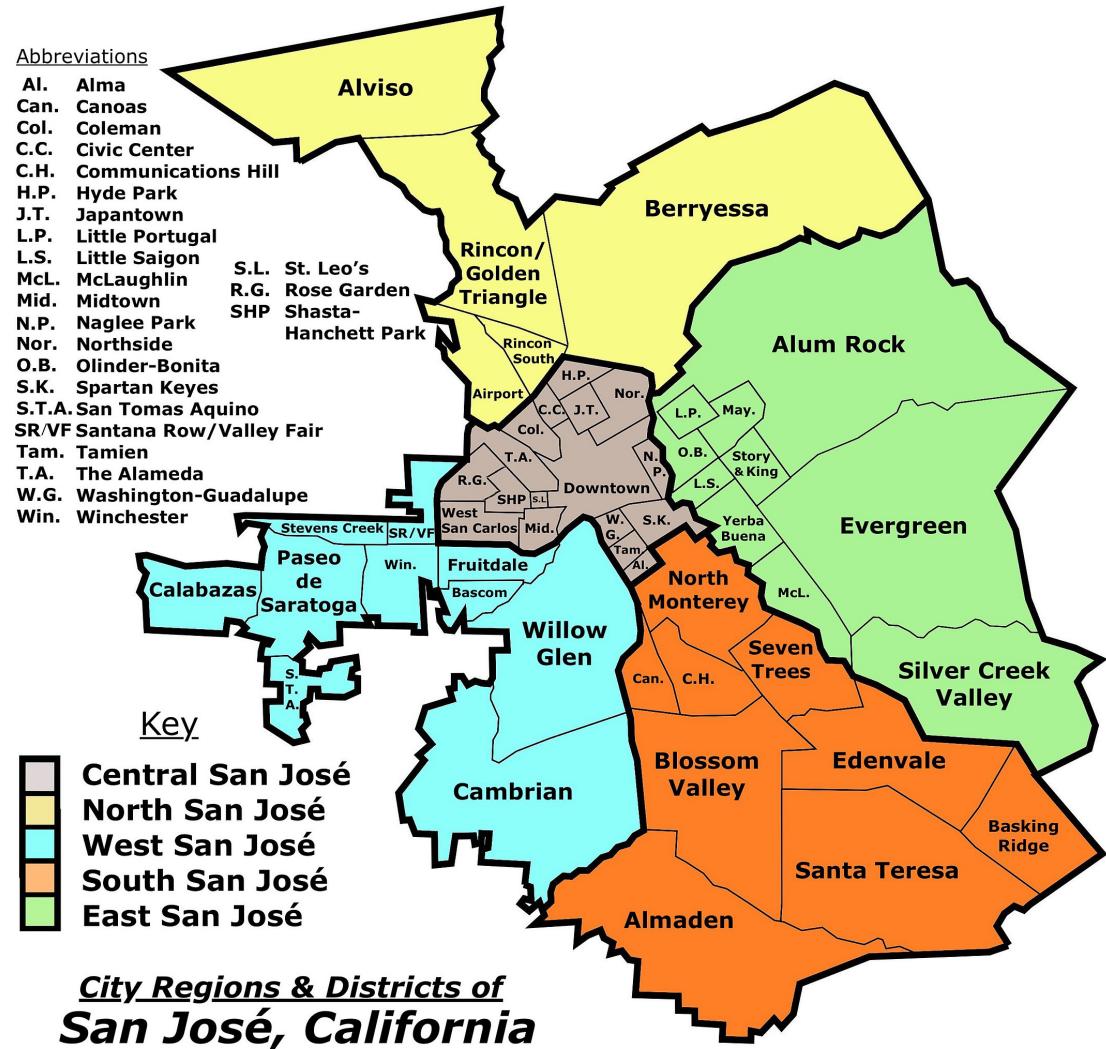


Attribute (Variable/Feature) Types

1. Categorical (nominal):
2. Ordinal:
3. Quantitative:

Categorical (Nominal)

Name	Age	Shirt Size	Favorite Fruit
Amy	8	S	Apple
Basil	7	S	Pear
Clara	9	M	Durian
Desmond	13	L	Elderberry
Ernest	12	L	Peach
Fanny	10	S	Lychee
George	9	M	Orange
Hector	8	L	Loquat
Ida	10	M	Pear
Amy	12	M	Orange



Ordinal



Name	Age	Shirt Size	Favorite Fruit
Amy	8	S	Apple
Basil	7	S	Pear
Clara	9	M	Durian
Desmond	13	L	Elderberry
Ernest	12	L	Peach
Fanny	10	S	Lychee
George	9	M	Orange
Hector	8	L	Loquat
Ida	10	M	Pear
Amy	12	M	Orange

Quantitative

Quantitative: order and arithmetic (+ plus/minus).

Name	Age	Shirt Size	Favorite Fruit
Amy	8	S	Apple
Basil	7	S	Pear
Clara	9	M	Durian
Desmond	13	L	Elderberry
Ernest	12	L	Peach
Fanny	10	S	Lychee
George	9	M	Orange
Hector	8	L	Loquat
Ida	10	M	Pear
Amy	12	M	Orange

Data Abstraction

1. Translate from domain-specific language to generic visualization language.
2. Identify dataset type(s), attribute (feature) types.
3. Identify cardinality.
 - How many items in the dataset?
 - What is cardinality of each attribute?
 - number of levels for categorical data
 - range for quantitative data
4. Consider whether to transform data: guided by understanding of task.

Data Model and Conceptual Model

Data Model

Mathematical abstraction:

Sets with operations.

Variable data type in programming.

Conceptual Model

Mental construction.

Supports reasoning.

Based on understanding task.

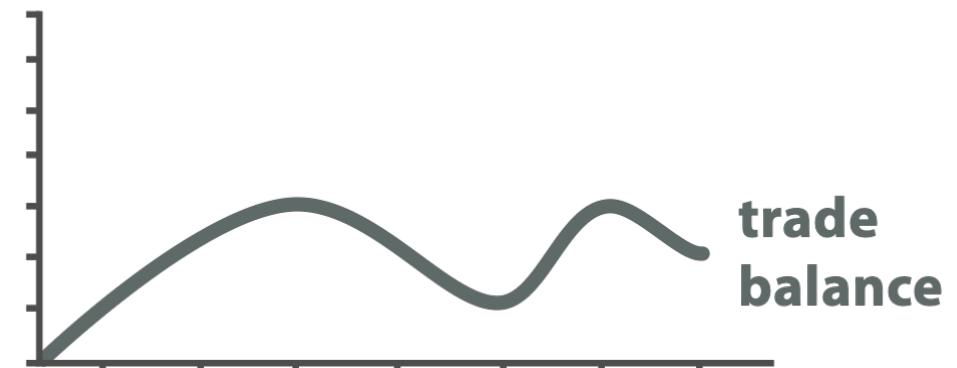
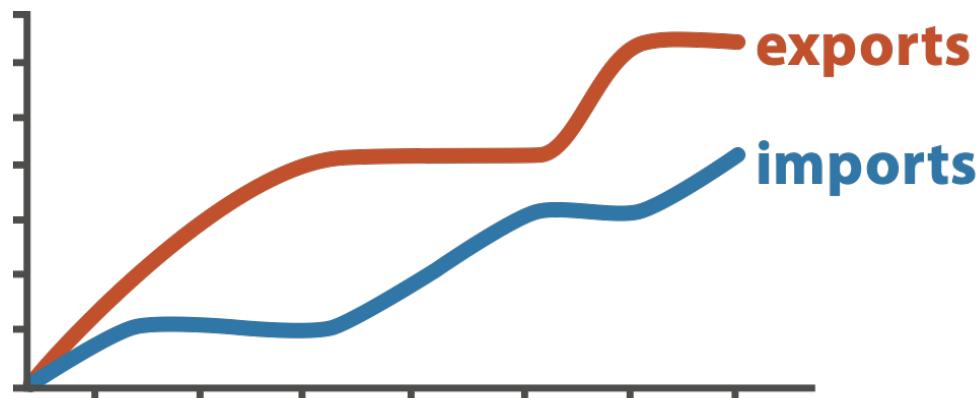
Data Abstraction and Task

- Data model: floats, 32.52, 54.06, -14.35, ...
- Conceptual model: temperature.
- Multiple possible data abstractions:

Data Abstraction

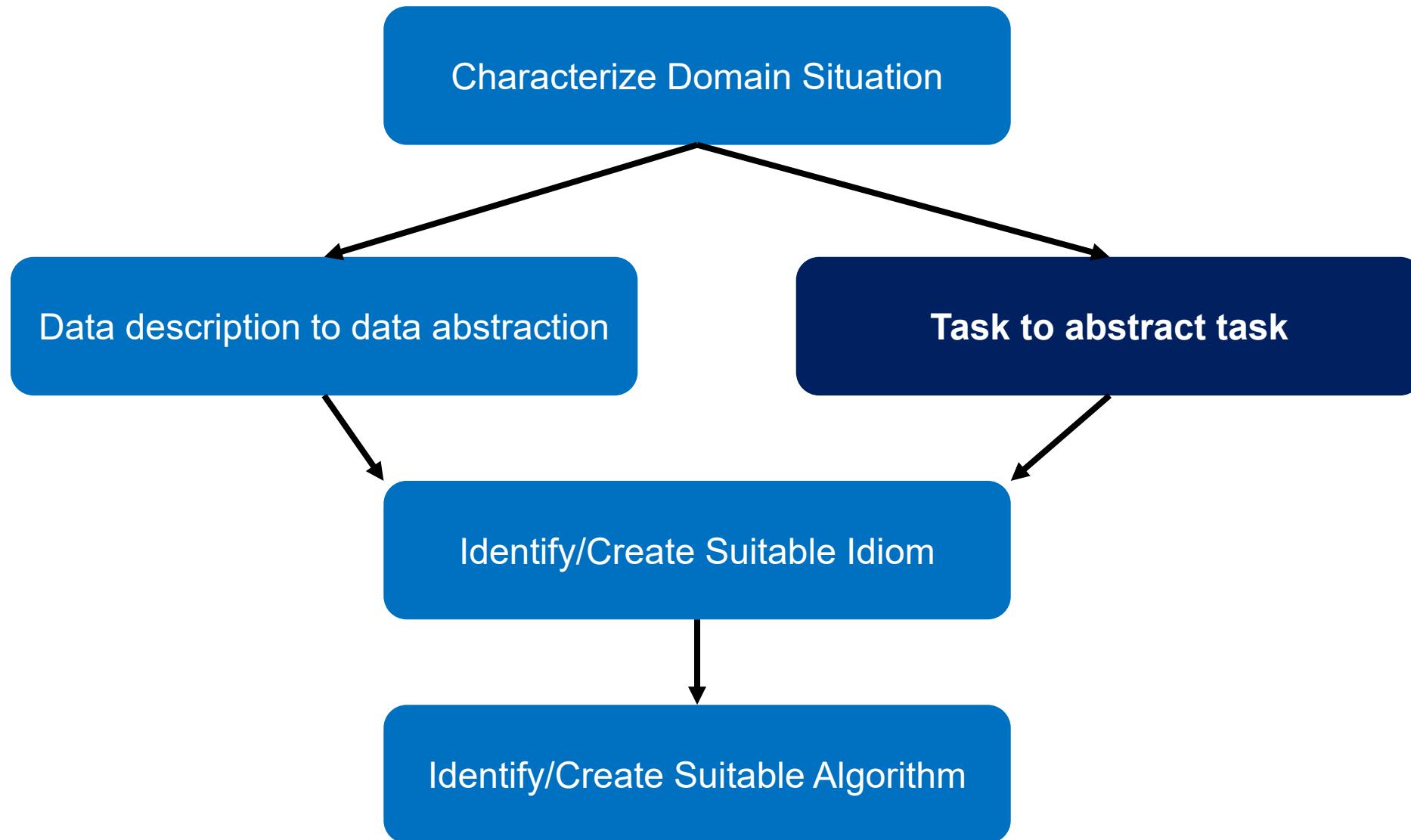
Derived attribute: compute from originals.

- simple change of type.
- acquire additional data.
- complex transformation.



$$\text{trade balance} = \text{exports} - \text{imports}$$

Design Process



Task Abstraction: Actions and Targets

- Business manager

“I want to see if new marketing strategy was successfully resulted in selling more products in the jewelry.”

- Biologist

“I want to see if the results for the tissue samples treated with Joon-55”

Domain Question

Why are there so many failed requests today?



Abstract task

Identify extremums.
Analyze outliers.

Task and Data Abstraction

- Data model: floats, 32.52, 54.06, -14.35, ...
- Conceptual model: temperature.
- Multiple possible data abstractions.
continuous to 2 significant figures: **quantitative**.

hot, warm, cold: **ordinal**.

above freezing, below freezing: **categorical**.

Action: Analyze (High-level)

Analyze: Consume existing data or produce additional data

Consume:

Consume information generated already.

This is the most common “why”

Produce:

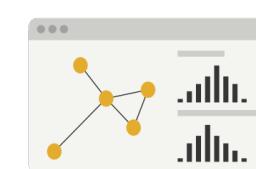
Produce new materials!

→ Consume

→ Discover



→ Present



→ Enjoy



→ Produce

→ Annotate



→ Record



→ Derive



Action: Analyze (High-level)

→ Consume

→ Discover



→ Present



→ Enjoy



→ Produce

→ Annotate



→ Record



→ Derive



Action: Search (Mid-level)

- **Lookup:** word in dictionary.
- **Locate:** node in network.
- **Browse:** books in bookstore.
- **Explore:** Find neighborhood in new city.

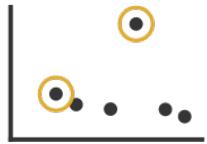
	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

Action: Query (Low-level)

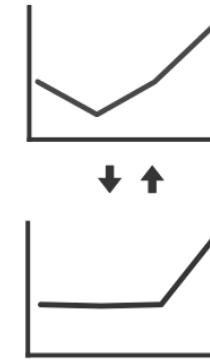
Returns the characteristics.

- **One:** identify
- **Some:** compare.
- **All:** summarize.

→ Identify



→ Compare



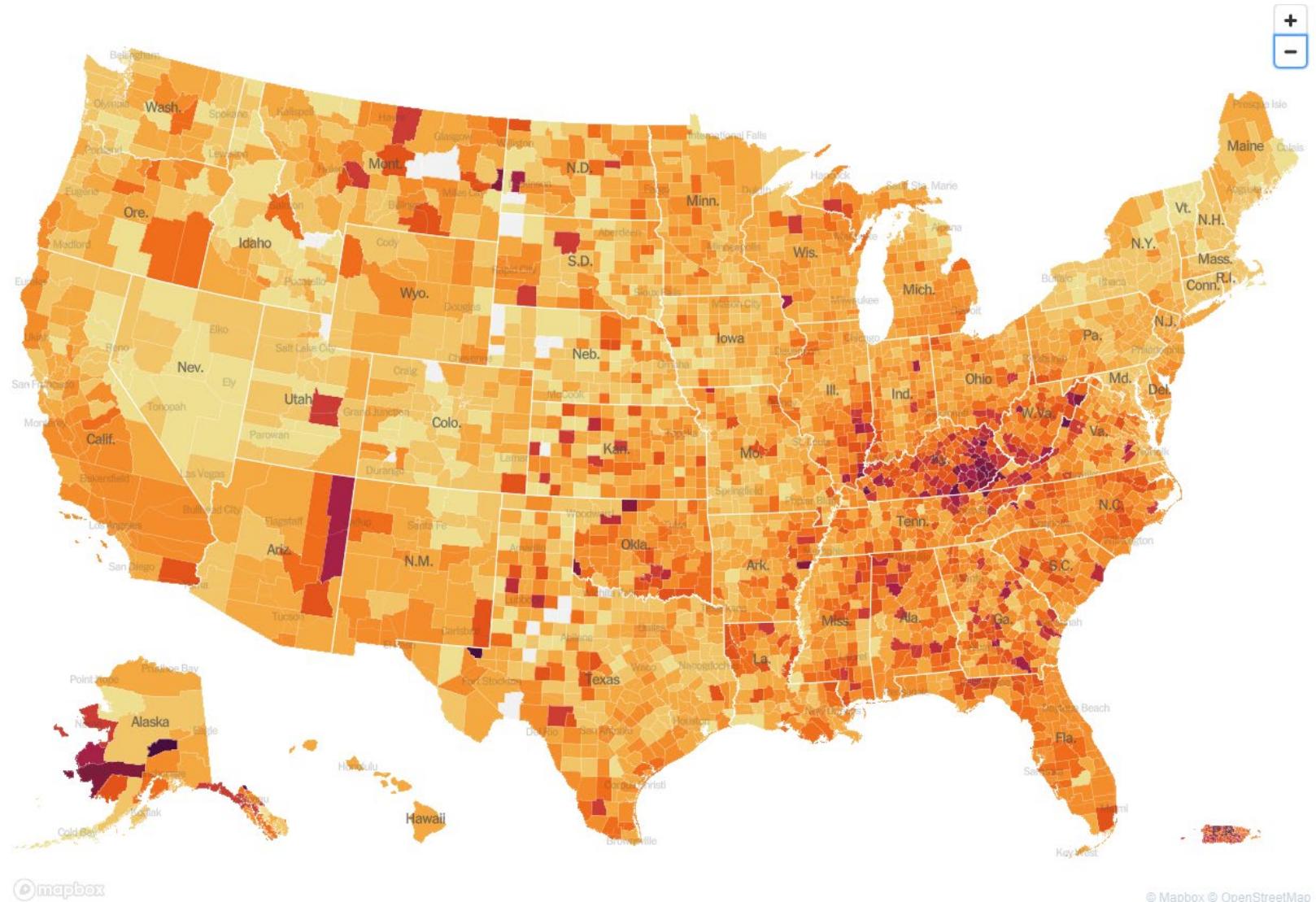
→ Summarize



Action: Query

Returns the characteristics.

- **One:** identify
- **Some:** compare.
- **All:** summarize.



Target

→ All Data

→ Trends



→ Outliers



→ Features



→ Attributes

→ One



→ Many

→ Distribution

→ Dependency

→ Correlation

→ Similarity



→ Extremes



Task Abstraction

- These {action, target} pairs are good starting point for vocabulary.
- Interplay: task and data abstraction.
 - need to use data abstraction within task abstraction.
but task abstraction can lead you to transform the data.
 - iterate back and forth.
first pass data, first pass task, second pass data, ...