DS4200: Information Presentation and Visualization

Data abstraction

Xiaoyi Yang
Khoury College of Computer Sciences
Northeastern University

Goals for Today

- Data type and abstraction
- Nested model of data visualization
- Quick introduction to Altair

Data analysis

Even though this course is about data visualization and information presentation...

These are just tools for data analysis

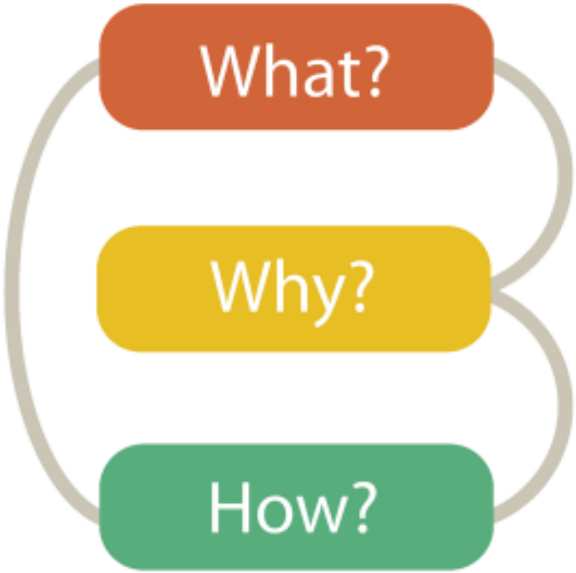
Data analysis

Even though this course is about data visualization and information presentation...

These are just tools for data analysis

Recall last time we discuss the components for a data visualization...

Data analysis

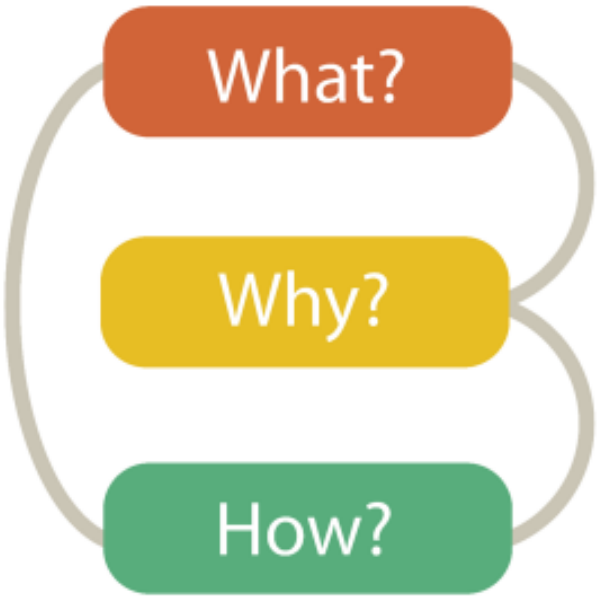


What data is shown?

Why is the user analyzing / viewing it?

How is the data presented?

Data analysis

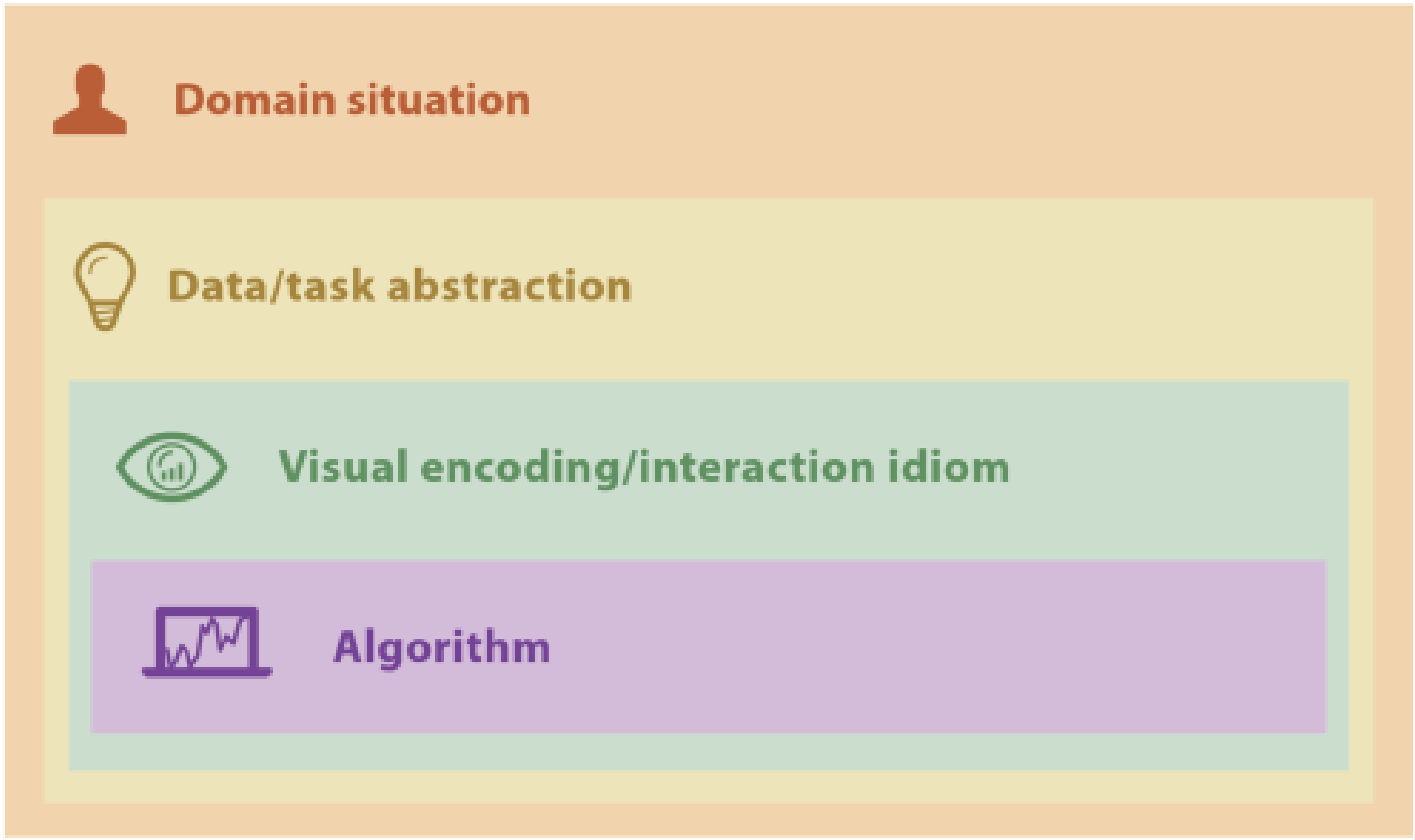


What data is shown?
DATA ABSTRACTION

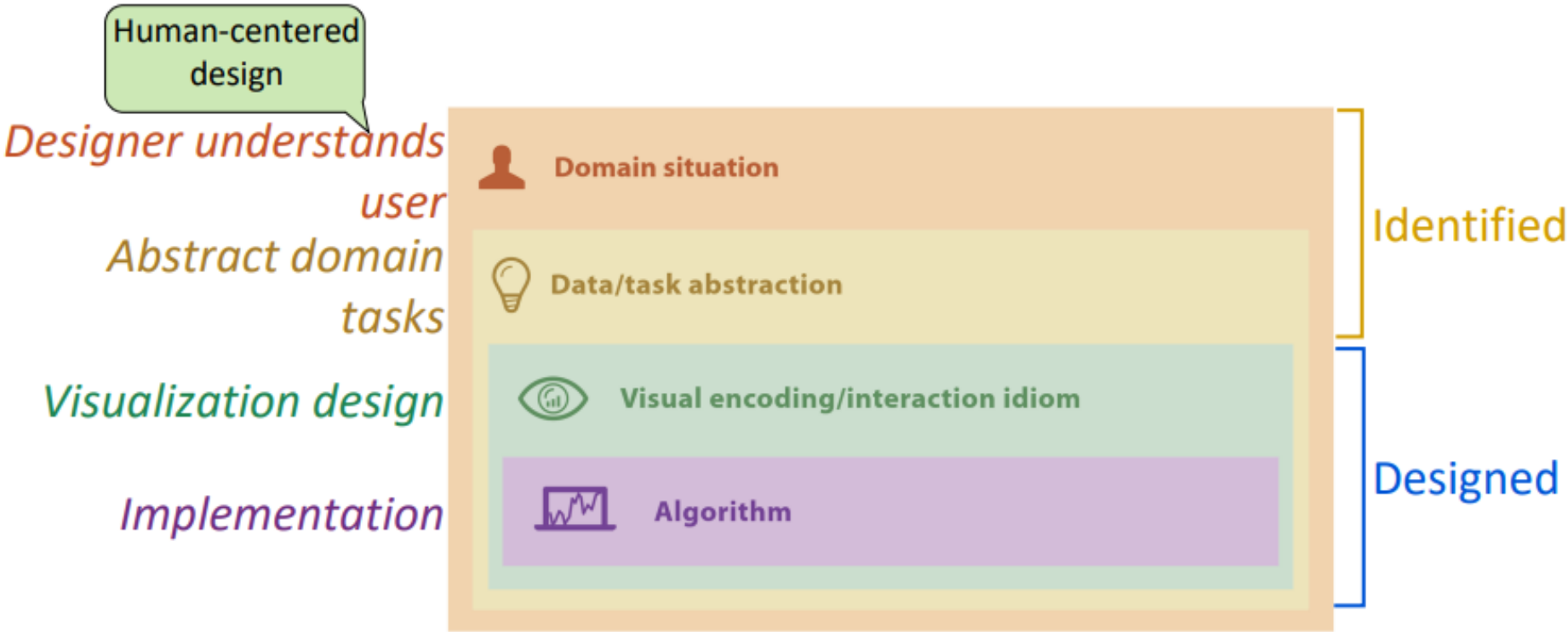
Why is the user analyzing / viewing it?
TASK ABSTRACTION

How is the data presented?
VISUAL ENCODING

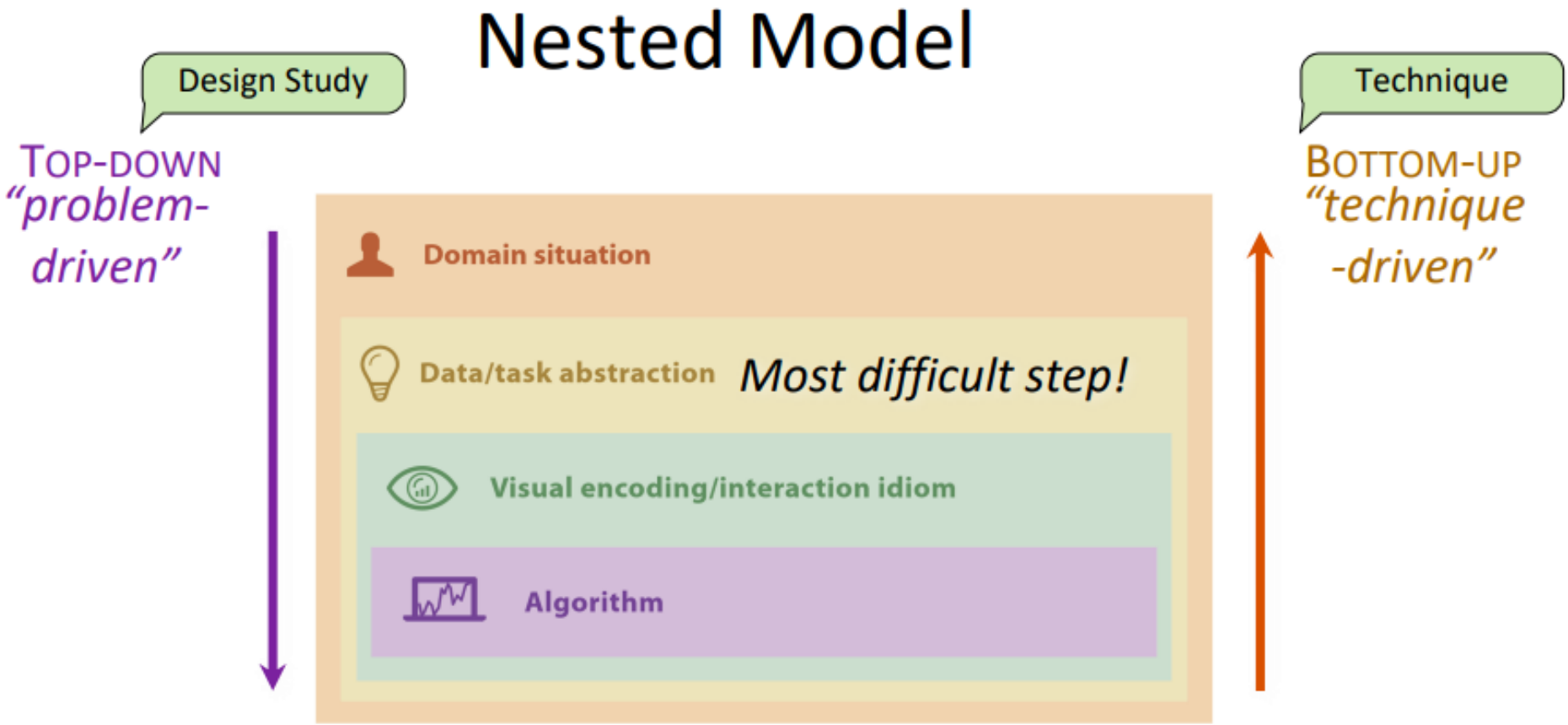
Nested model of data visualization



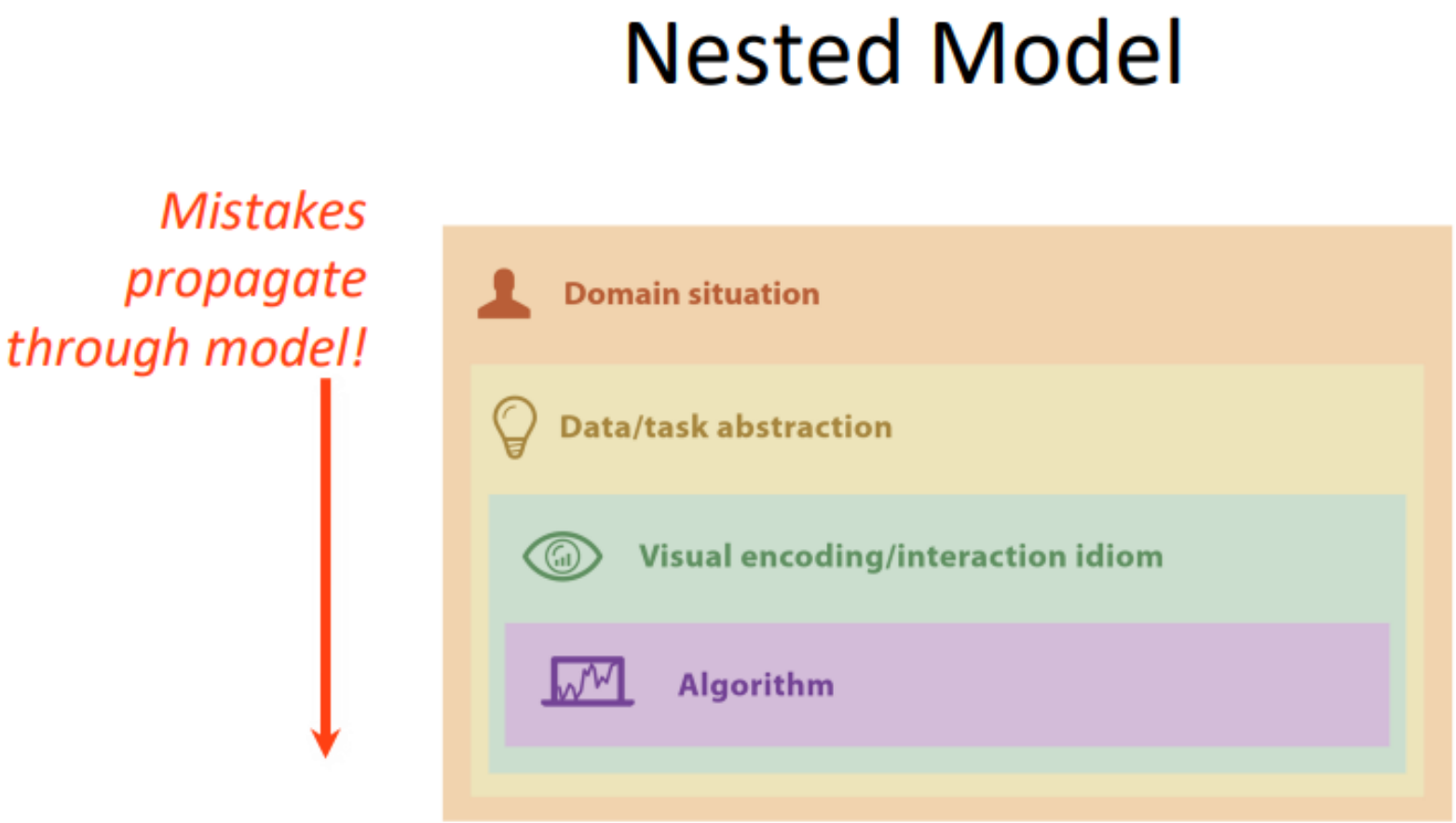
Nested model of data visualization



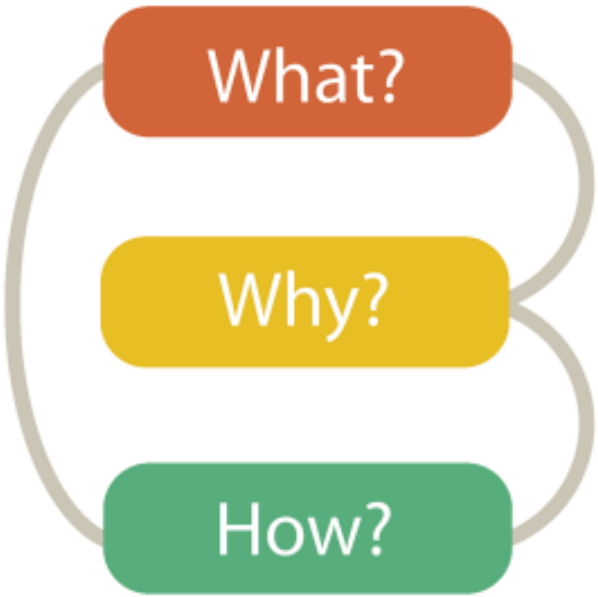
Nested model of data visualization



Nested model of data visualization



Data analysis



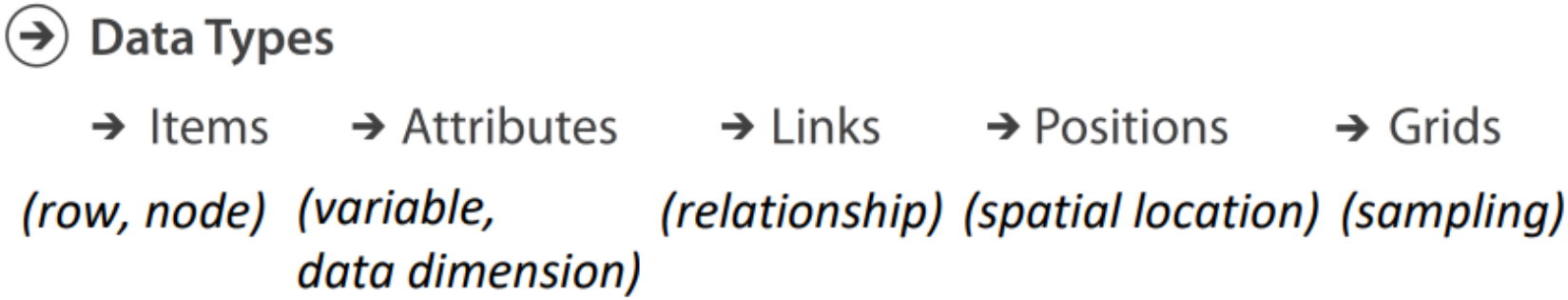
What data is shown?
DATA ABSTRACTION

Why is the user analyzing / viewing it?
TASK ABSTRACTION

How is the data presented?
VISUAL ENCODING

Data Types

TYPE = structural or mathematical interpretation of the data

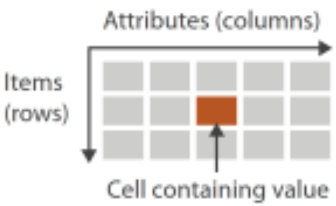


Data Types

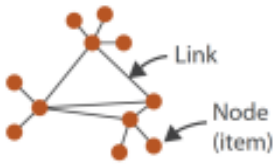
DATASET = collection of information that is the target of analysis

Dataset Types

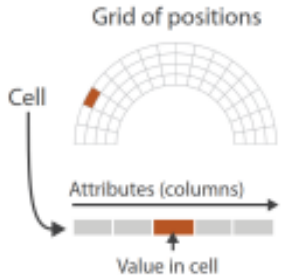
→ Tables



→ Networks

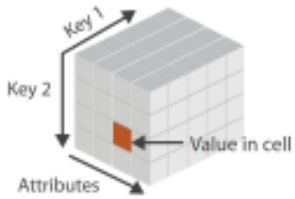


→ Fields (Continuous)



→ Geometry (Spatial)

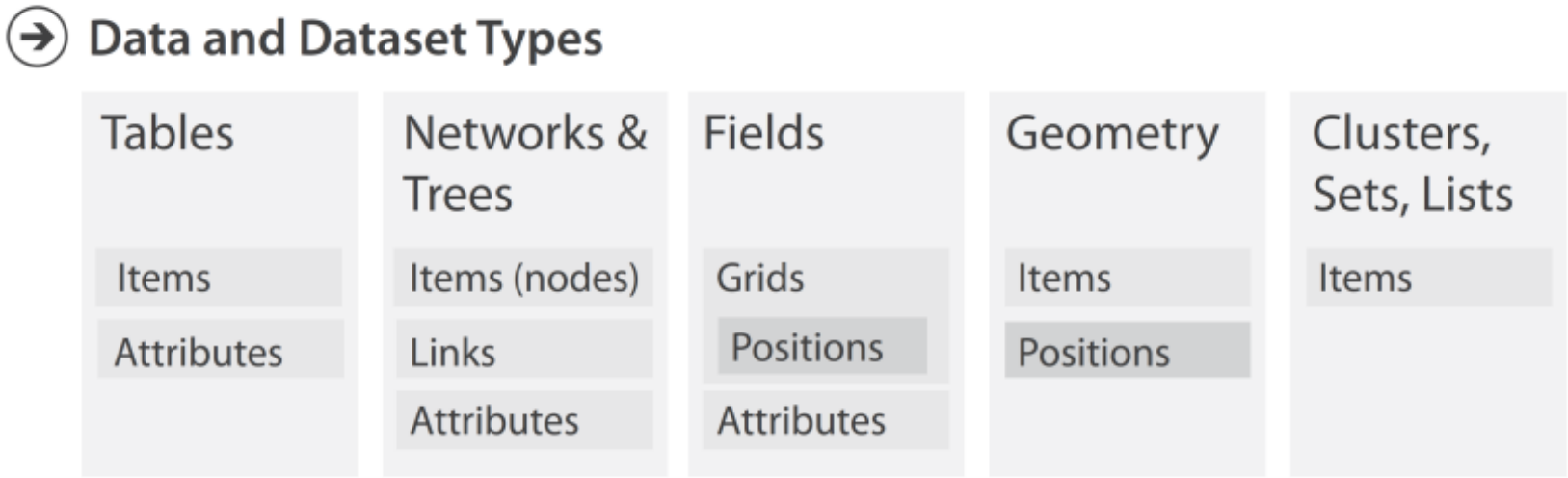
→ *Multidimensional Table*



→ *Trees*

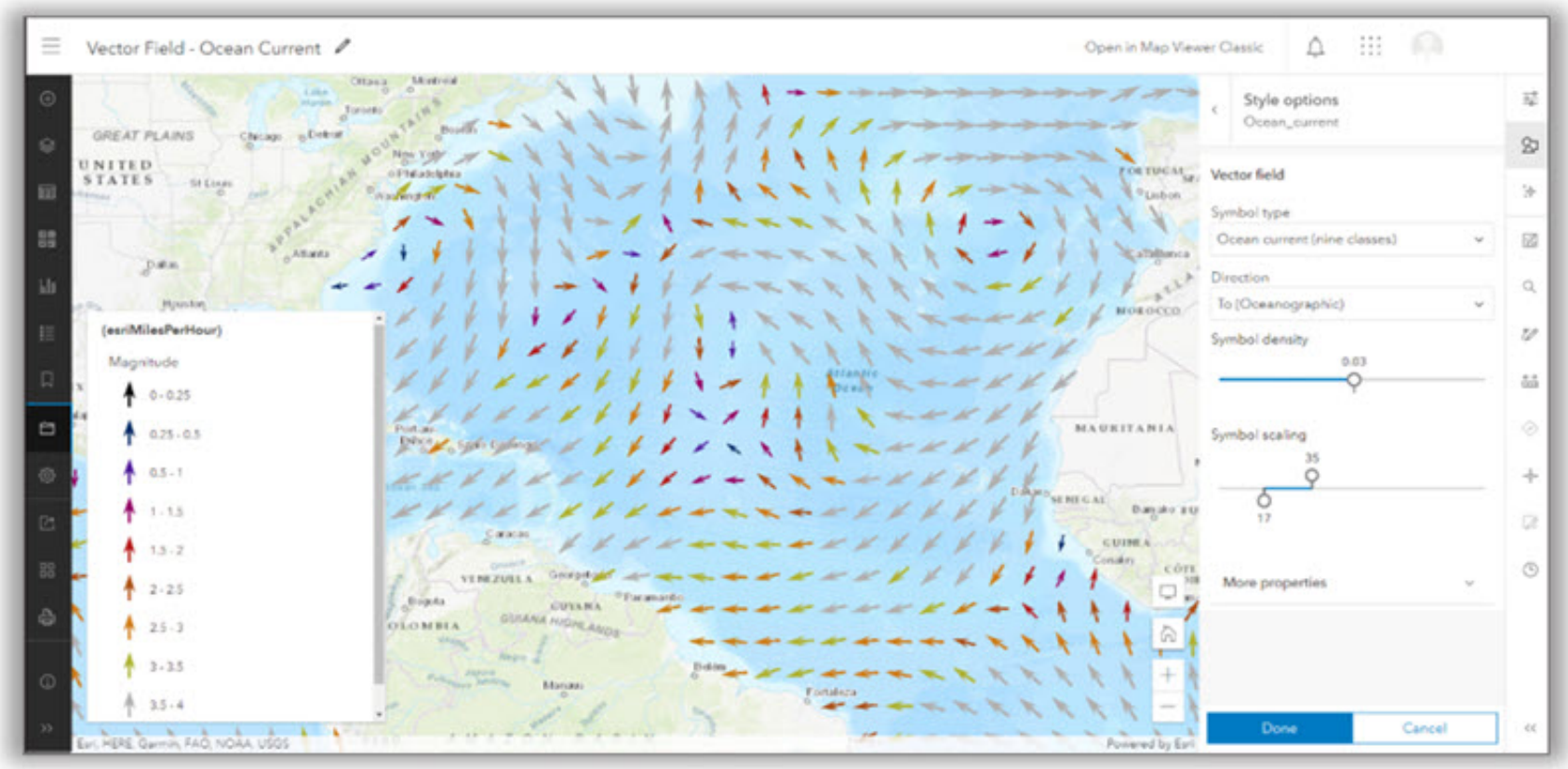
Data Types

DATASET = collection of information that is the target of analysis



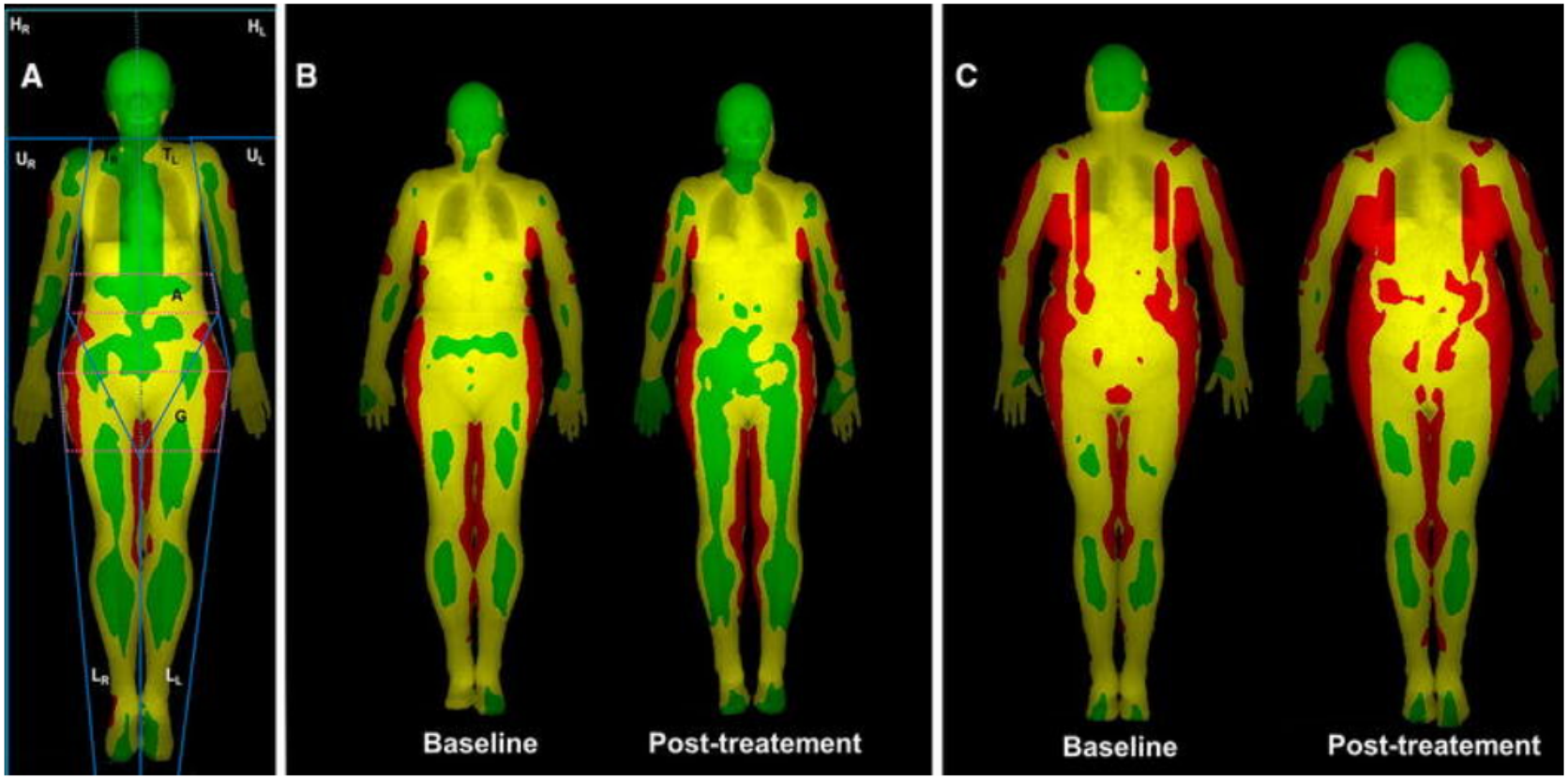
Fields

Field data type: Each cell in a field contains measurements or calculations from a continuous domain.



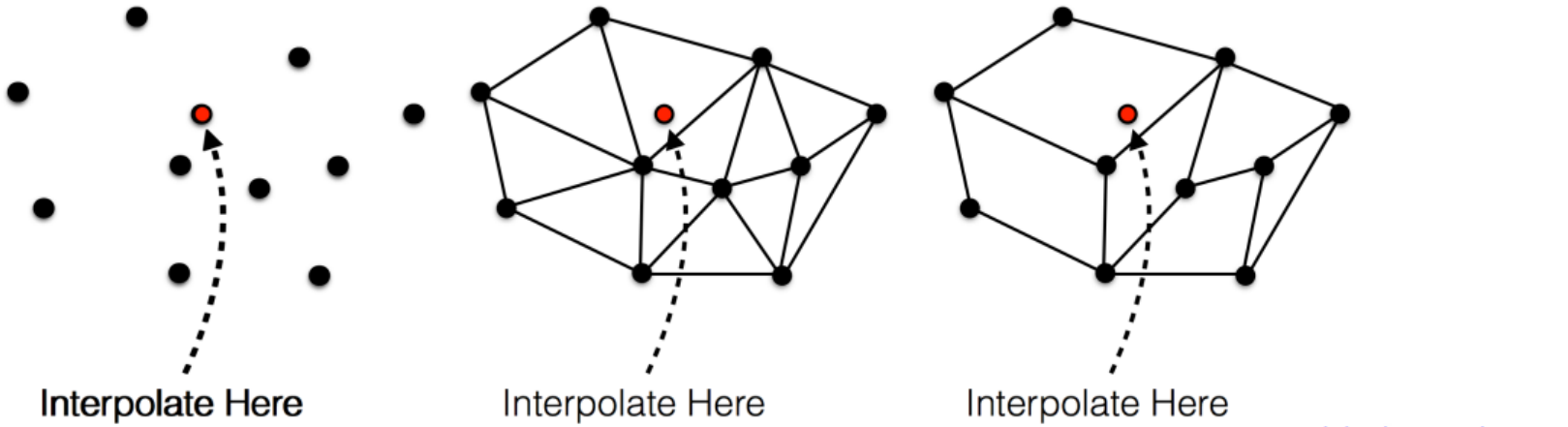
Fields

Field data type: Each cell in a field contains measurements or calculations from a continuous domain.



Two problems in the Field data

- Sampling, or the choice of where attributes are measured
- interpolation, or how to model the attributes in the rest of place



Slides by Miriah Meyer

Introduction to Altair

Altair is a declarative statistical visualization library for Python.

- Declarative Syntax
- Interactivity
- Data Transformation

Here is the reference: Altair curriculum

In-class activities

Let's redo your Titanic data visualization from the last lecture with Altair.

Requirement:

- First, re-create the same static data visualization.
- Add the Tooltip for interaction.
- Output the HTML file.