Data Abstraction

Thursday, October 11, 2018 17:52

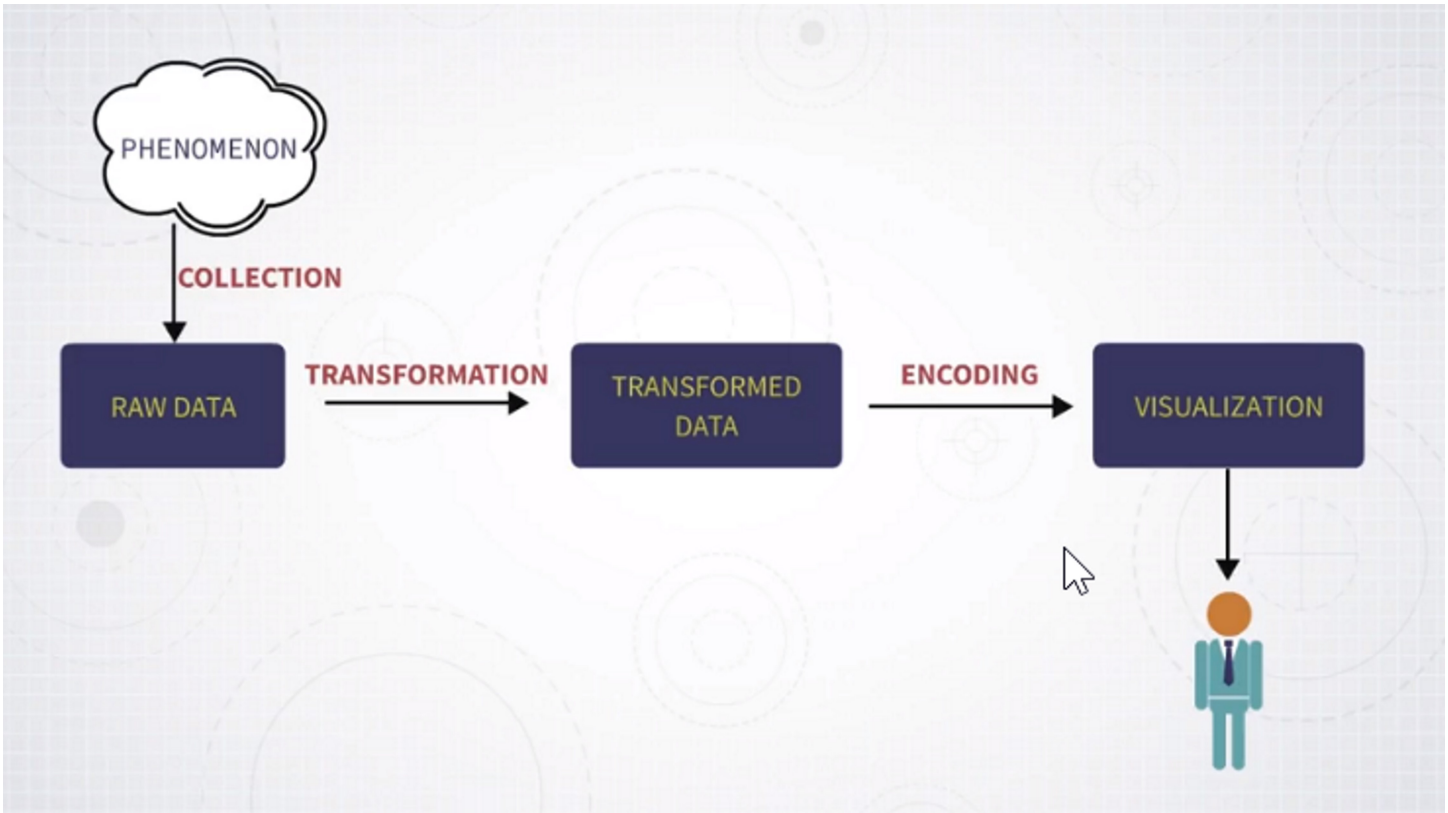
What is data?

Factual information -> measurements or statistics
Used as a basis for

- Reasoning
- Discussion
- Calculations

Data as a general concept refers to the fact that some existing information or knowledge is represented or coded in some form suitable for better usage or processing. (Wikipedia)

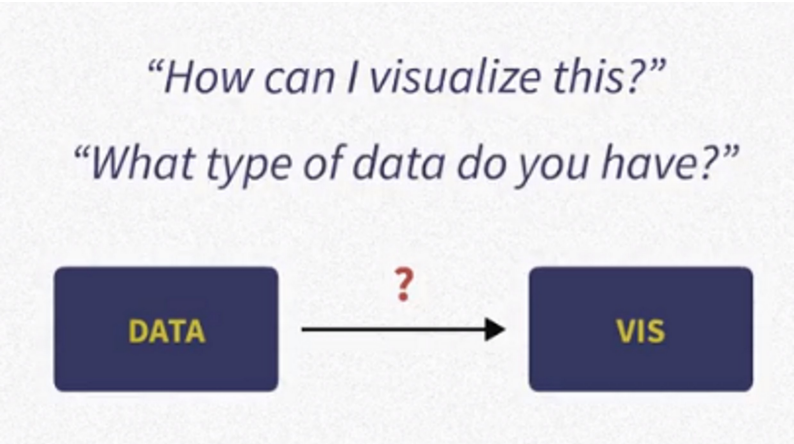
Role of data in visualization process:



TRANSFORM DATA!!!

1. Collection: data is collected in different formats (granularity) exclusively!
2. Transformation ->
3. Visual Encoding: how do I know what is the appropriate visual representation? -> Data abstraction

Data abstraction
The idea that you can describe data in ways that help you decide what encoding methods are available and appropriate for this information.



(… a way to recognize common structures in data coming from very different domains.)

Abstracting away from the domain
To identify characteristics that are useful
To decide what visual representations are available and appropriate

Dataset types: Tables and Networks
Data as a collections of:

- Items: objects/entities (you want to visualize)
- Attributes: properties of these objects

Tables: grid or collections of

- Rows: to represent item
- Column: to represent attributes -> Same type

Networks & trees: the defining characteristic of this collection of items is that they are connected by links:

- Nodes: items
- Links: item's connection

Nodes and links can have attributes associated.

Attribute types
There are three main types:

1. Categorical
    - Values that describe categories
    - There's no inherent order

2. Ordinal
    - Categories that have an order
    - "distance" among the categories is not known -> no meaning in applying

arithmetical operations
    - Example: Education: [Elementary, High School, Undergrad, Grad, ...]

3. Quantitative (numerical)
    - Values represent some measure quantity
    - "Distance" is known -> arithmetic operations IS POSSIBLE
    - Example: height and weight

Attribute semantics/characteristics
Semantics: the meaning of some attributes...

Predefined semantics:

First classification

1. Spatial : geographic : quantitative
2. Temporal : time : quantitative

Secondary classification

1. Sequential:
2. Diverging: it's possible to identify a zero value, that separates negative and positive values
3. Cyclic: e.g. months of the year
4. Hierarchical: e.g. category - subcategory

Data abstraction as a Visualization Design Tool
Data abstraction gives guidance in selecting appropriate visual encoding strategies.

Examples:
- Line chart is **inappropriate** for *unordered* attributes.
    - Inexistent patterns can be read from categorical variables.

- *Categorical* attributes can be sorted arbitrarily.

- Spatial attributes can be visualized with *spatial* metaphors.

- Divergent color scale more appropriate to represent a *divergent* quantitative attribute.

Data Profiling and Quality
Data familiarization and profiling -> You have to become familiar with your dataset

- Data come from different sources
- Meaning of data is often not evident
- Effort and interactions needed to understand meaning
    - by your own or talking with the person producing the data

Data quality and data "wrangling"
Sometimes there are missing/wrong values in the data.
Sometimes data comes in different types that are not useful for processing.

- Very important and time consuming