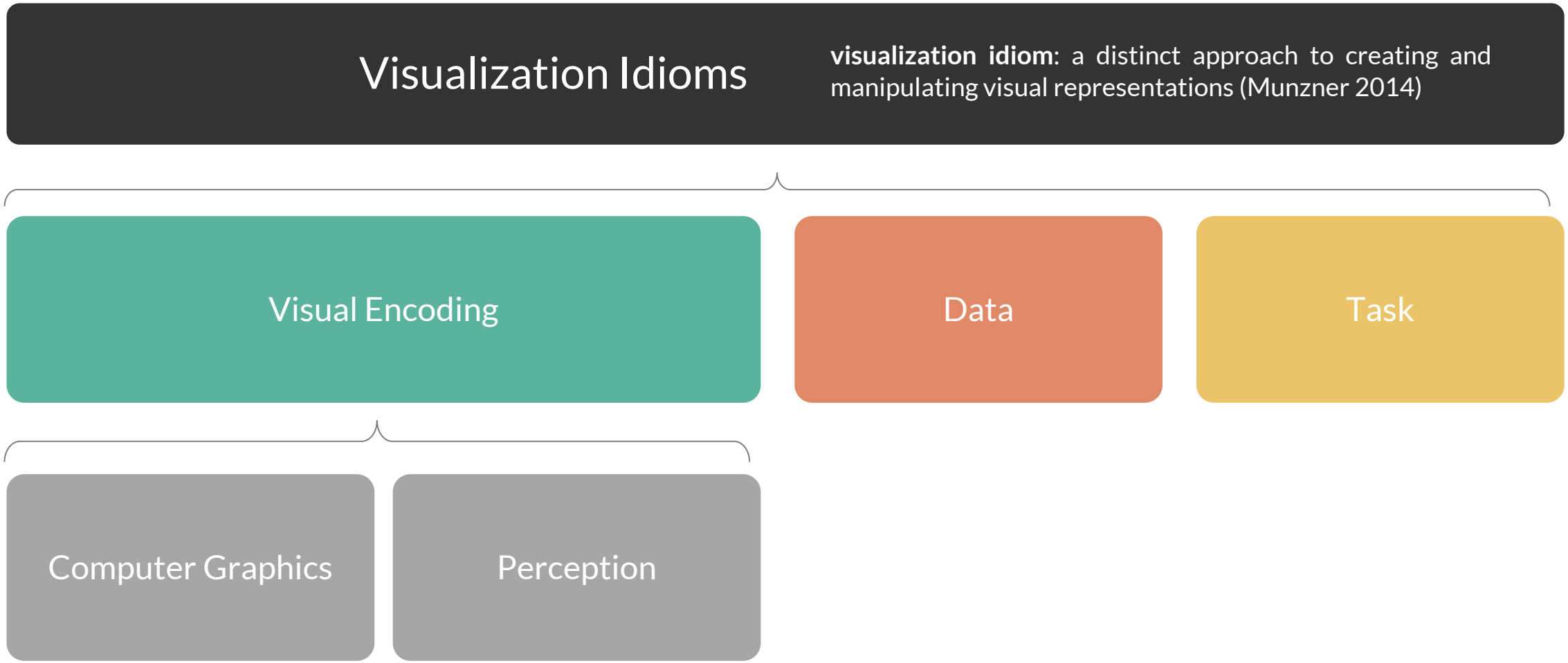
Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

# Data and Task Abstractions

05/10/2023

# Foundations of Visualization

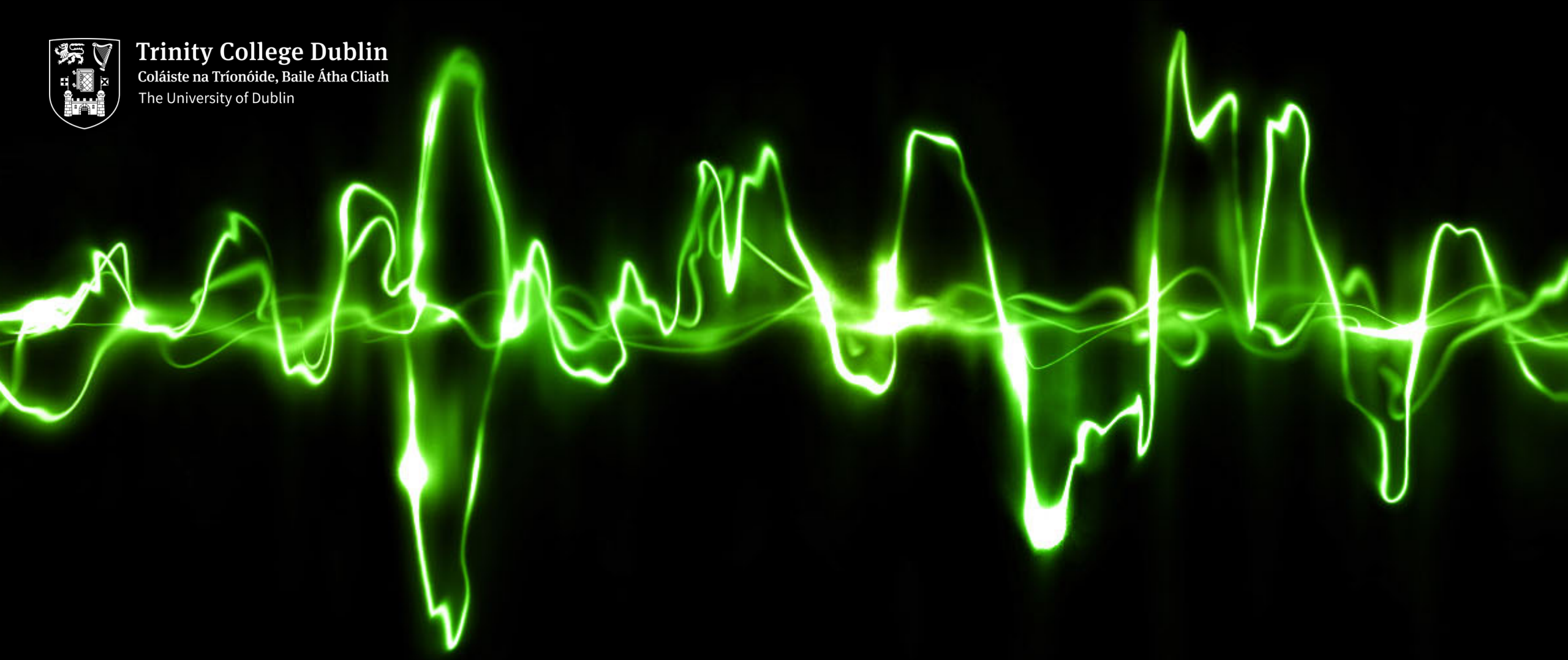


# Information vs meaning


The dangers of depiction without regard for meaning:

"Consequences of communication gap" [sic] - Original Source Unknown

https://www.youtube.com/watch?v=IOKeTZwgx-k



Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

# Data Abstraction

# Data/Dataset Types

Images and Classification from Munzner [2014]

Common data types in visualization :

**Attributes**: measurable property value or characteristic

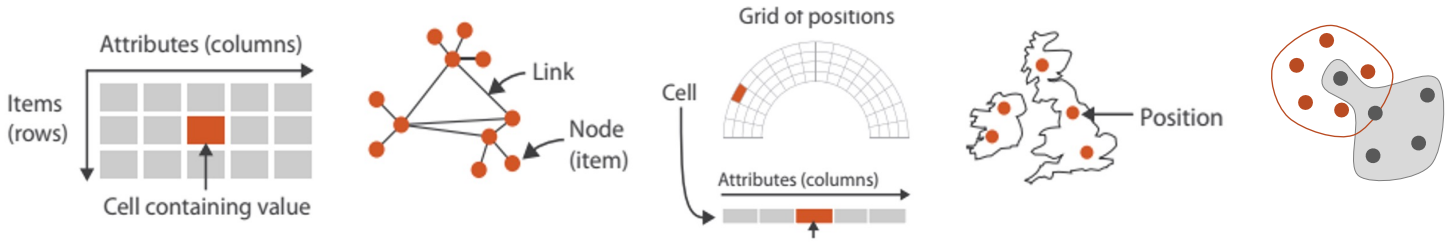
**Item:** discrete individual entity e.g., a row in a table (could be comprised of one or more attributes)

**Link:** relationship between items

**Position:** spatial information

**Grid Points:** discrete sampling of continuous data

Higher level data are comprised of atomic data types. The following are some commonly occurring dataset types:




| Tables | Networks | Fields | Geometry | Sets, Lists, Clusters |
|---|---|---|---|---|
| Items | Items | Grid points | Items | Items |
| Attributes | Attributes | Positions | Positions | |
| | Links | Attributes | | |

# Dataset Type: **Tables**



A dataset made up of rows and columns

For a standard tables:

- Row = Item of data
- Column = Attribute of dataset
- Cell defined by combination or row and column

- A **key** attribute (a.k.a. independent attribute or dimension) is used to look up a *value* (a.k.a. dependent attribute, measure) of items. The key should be unique to an item

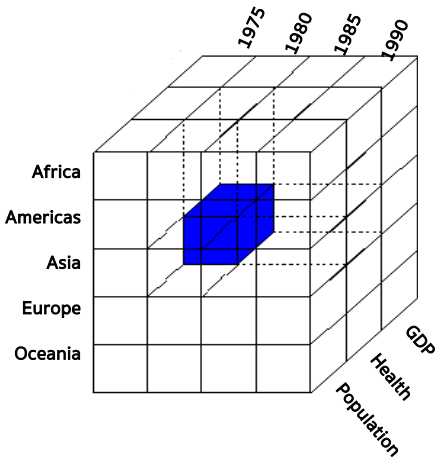
Attribute ▼

Item ▶

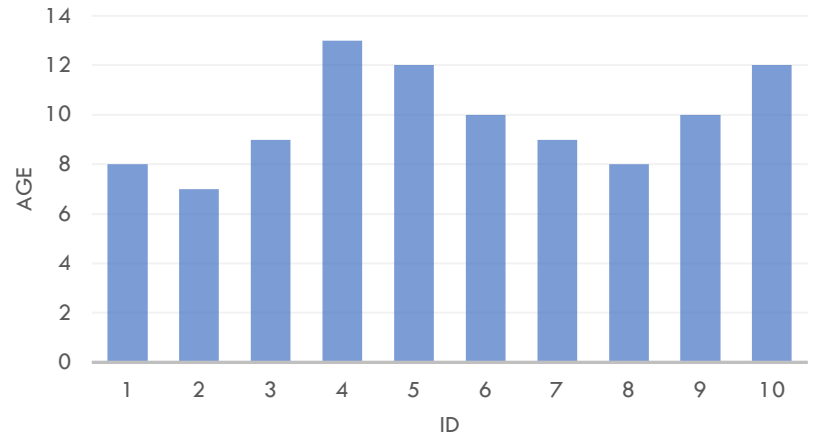
| ID | Name | Age | Shirt Size | Favorite Fruit |
|---|---|---|---|---|
| 1 | Amy | 8 | S | Apple |
| 2 | Basil | 7 | S | Pear |
| 3 | Clara | 9 | M | Durian |
| 4 | Desmond | 13 | L | Elderberry |
| 5 | Ernest | 12 | L | Peach |
| 6 | Fanny | 10 | S | Lychee |
| 7 | George | 9 | M | Orange |
| 8 | Hector | 8 | L | Loquat |
| 9 | Ida | 10 | M | Pear |
| 10 | Amy | 12 | M | Orange |

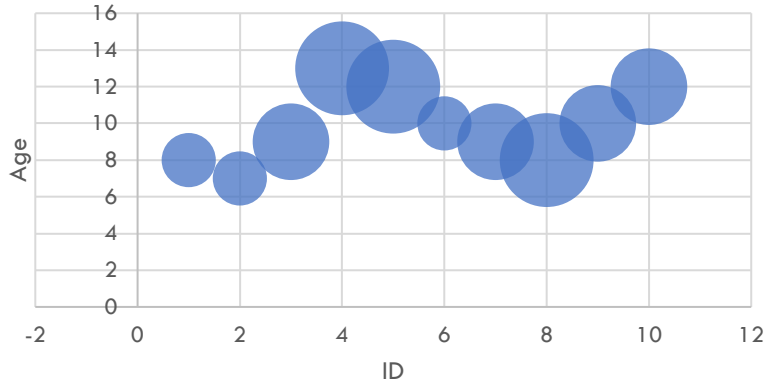
Cell

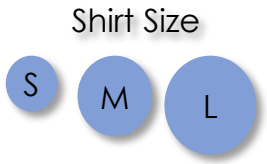
**multi-dimensional table**: requires multiple keys for indexing cells ▶



# Dataset Type: **Tables**




Attribute ▼

Item ▶

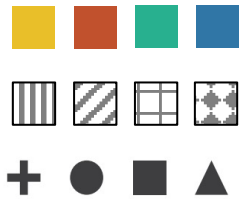
| ID | Name | Age | Shirt Size | Favorite Fruit |
|---|---|---|---|---|
| 1 | Amy | 8 | S | Apple |
| 2 | Basil | 7 | S | Pear |
| 3 | Clara | 9 | M | Durian |
| 4 | Desmond | 13 | L | Elderberry |
| 5 | Ernest | 12 | L | Peach |
| 6 | Fanny | 10 | S | Lychee |
| 7 | George | 9 | M | Orange |
| 8 | Hector | 8 | L | Loquat |
| 9 | Ida | 10 | M | Pear |
| 10 | Amy | 12 | M | Orange |

Cell

◀ Visualization goal is to visually encode the values of cells in the table. This can be done in many different ways. Note that the table itself is a form of visualization.

# Attribute Types

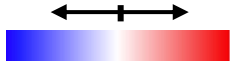
All other things being equal, the attribute type may define how best to encode the data. However, this may be impacted by other facts such as the higher level structure or task (see later).



CATEGORICAL / Nominal / Qualitative Data:

- No implicit ordering (could enforce some ordering)
- Often has grouping, sometimes hierarchical structure
- Applicable Operations: =, ≠

Sequential



Diverging

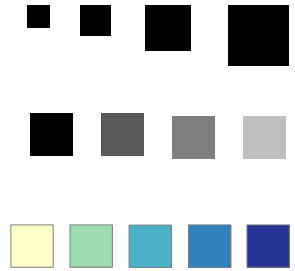
Cyclic

QUANTITATIVE / Numerical Data:

- *Sequential:* from min to max e.g. populations of countries
- *Diverging:* two or multiple sequences that meet in middle e.g. Elevation : above sea level & below sea level
- *Cyclic:* values wrap around e.g. days of the week

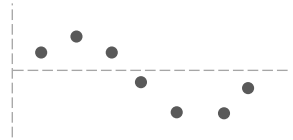
**Discrete / Ordinal:** discrete elements that have some implicit ordering e.g. shirt size (small, medium, large), top-10 list of movies

- Applicable Operations: =, ≠, >, <



**Continuous:** comprises measurement of magnitude supporting arithmetic comparisons

- *Interval (arbitrary zero)* : Cannot compare directly. Only differences/distances can be compared. e.g. Dates; Location
  - Operations: =, ≠, >, <, +, -
- *Measurement* (there exists an absolute zero): Can measure ratios & proportions. e.g. Length, Mass
  - Operations: =, ≠, >, <, +, -, ×, ÷



# Dataset Type: **Networks And Trees**

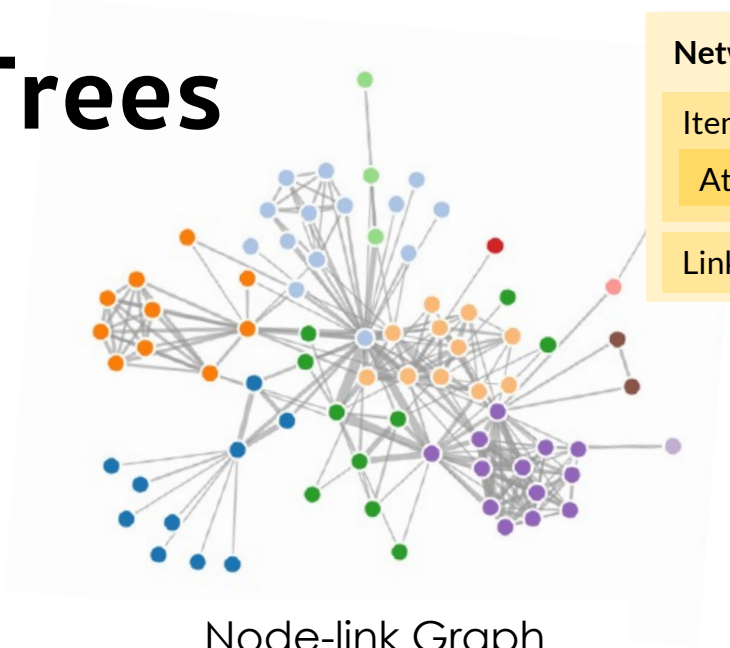
**Network**: a type of dataset for specifying relationships between two or more items. Typically comprised of:

- *Nodes*: an item in a network
  - associated with some attribute(s)
- *Link*: relation between items

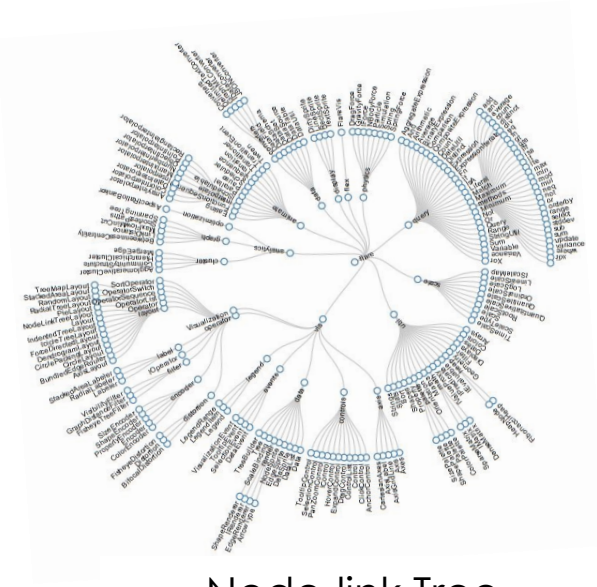
**Trees:** hierarchical network structure (each child has only one parent), generally without cycles

Both are typically encoded by positional arrangement of nodes, link marks for connection, categorical encoding of node identity/association. *May include other attributes*

More on this in a later lecture



Node-link Graph



Node-link Tree

**Networks**

Items

Attributes

Links

# Dataset Type: **Field**

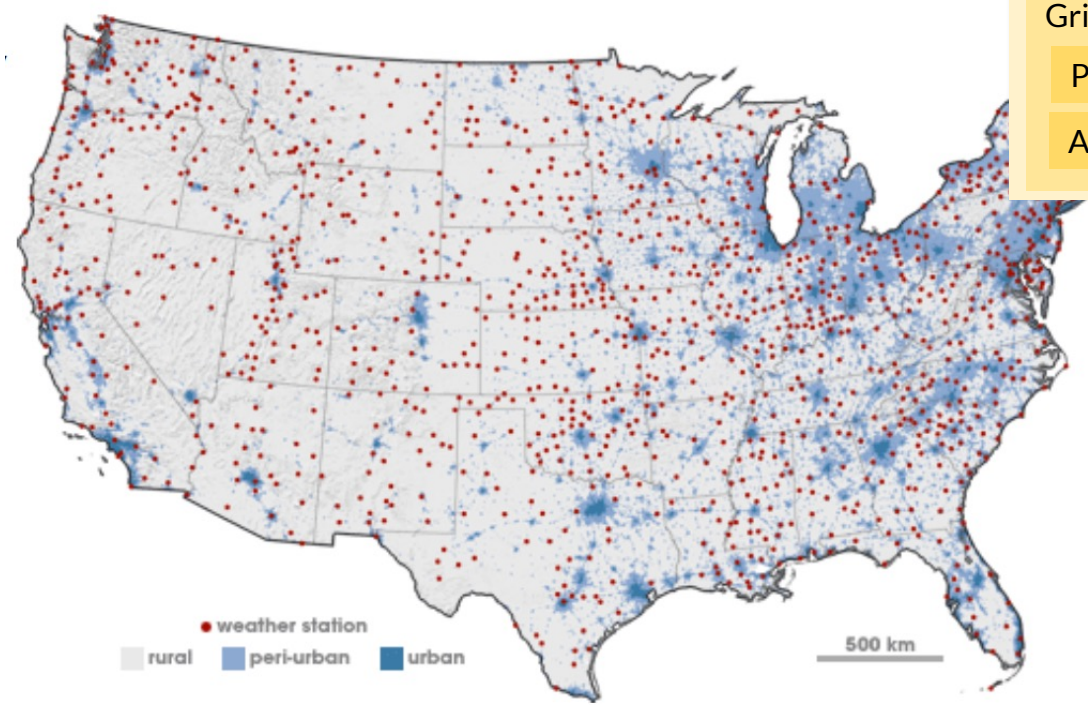
**Fields**

- Grid points
  - Positions
  - Attributes

Dataset comprised of cells that contain measurements from a continuous domain

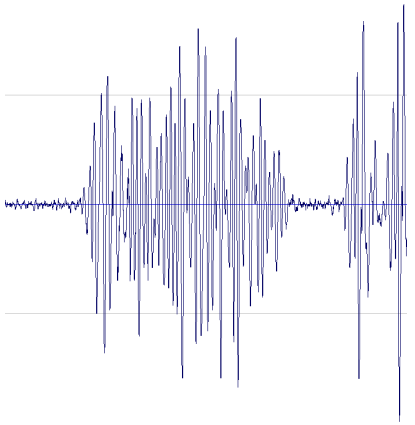
Data typically needs to be discretized for visualization/analysis:

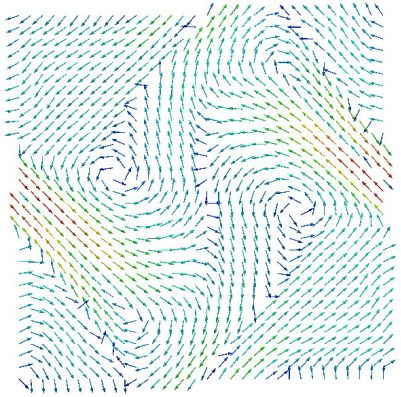
- *Sampling* determines how frequently measurements are taken
- *Interpolation* is a means of approximating values in between sampled points

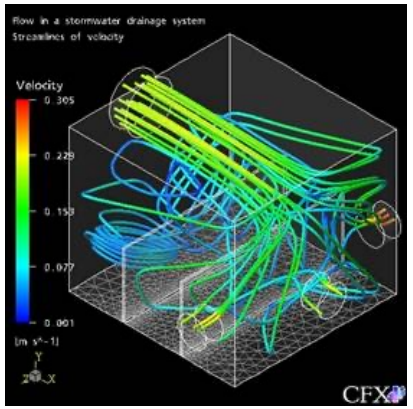



Spatial field data constrains choice in use of visualization space

- position & size channels are reserved, use other channels for expressing other attributes
- many examples in scientific visualization








More on this in a later lecture

# Dataset Type: **Geometry**

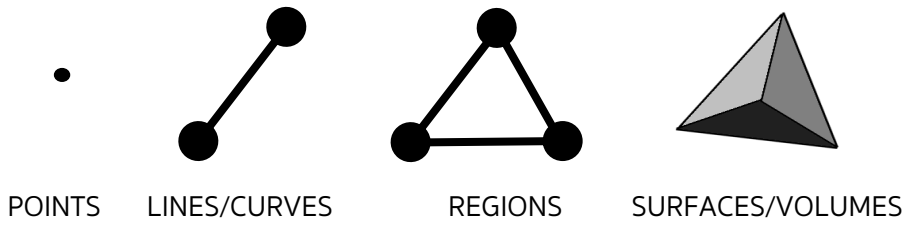
**Geometry**

Items

Positions

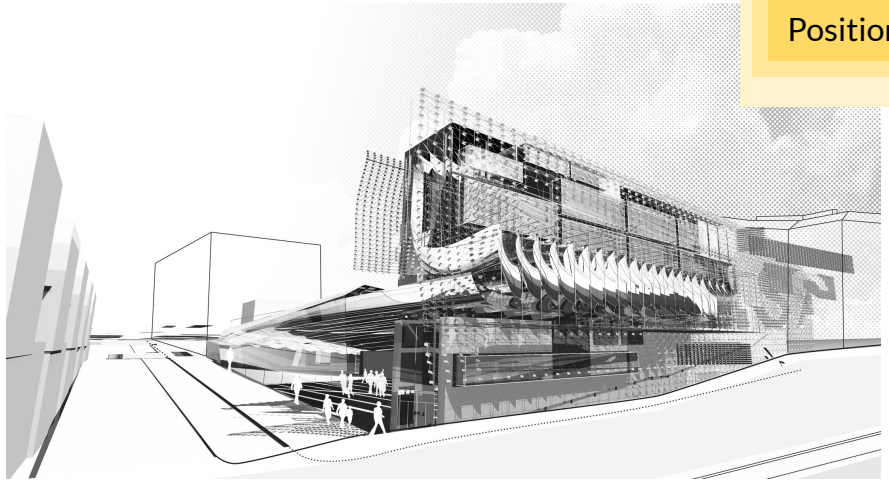
**Dataset specifying information about the shape of items with explicit spatial properties.**

- **Items** must include **position**, but this can be used to build up higher level geometry

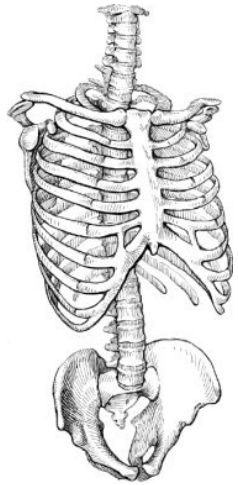



- Intrinsically **spatial**; does not necessarily have other attributes
- Primarily occurs in context of tasks that require spatial understanding (position, shape, curvature, area, etc..)

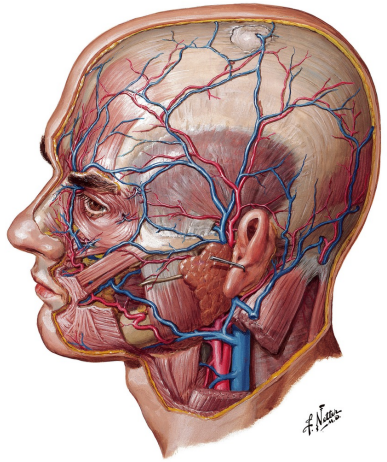
More on this in a later lecture



© Lostritto (2009)



© Isenberg et al (2004)



© Frank Netter

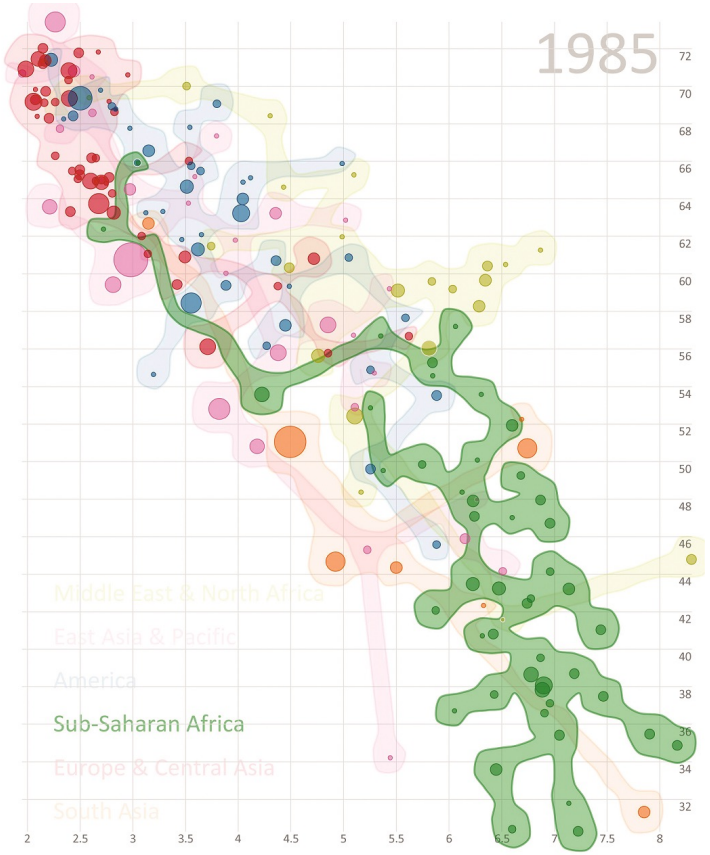
# Other Data Set Types: **Sets, Lists, Clusters, etc**

**Sets, Lists, Clusters**

Items

**Set:** a group of *items*.

Encode containment, association



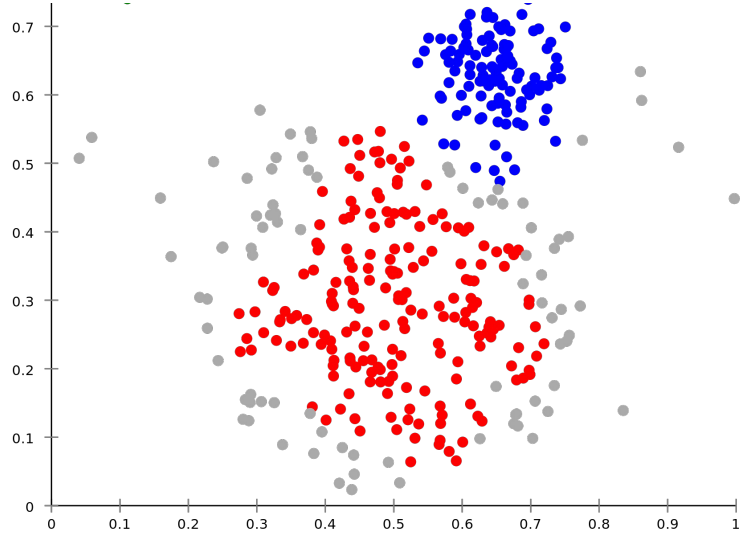

**List:** group of *items* with an ordering

Encode grouping + ordered arrangement

| Rank | School Name | Academic repu | E | Facult | Citatio | I | I |
|---|---|---|---|---|---|---|---|
| | Filter: <None> | | | | | | |
| 1. | Massachusetts Inst | | | | | | |
| 2. | University of Camb | | | | | | |
| 3. | Harvard University | 100 (1) | | | 100 (1) | | |
| 4. | UCL (University Co | | | | | | |
| 5. | University of Oxfor | | | | | | |
| 6. | Imperial College L | | | | | | |

**Cluster:** a grouping of items based on *attribute* similarity

Encode association



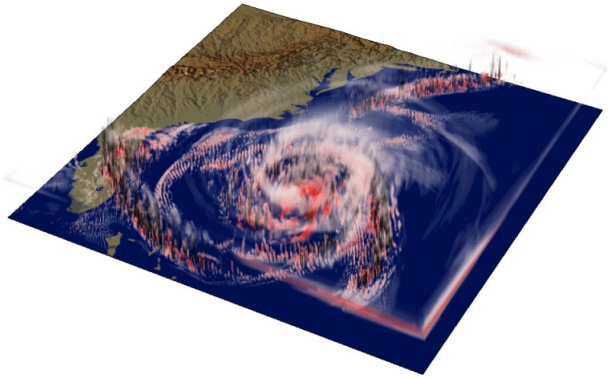
# Aside: **Dataset availability**

## Static data:

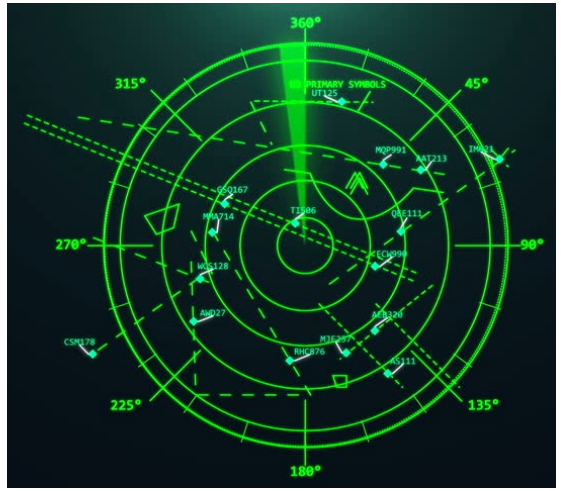
- aka **offline** data; available all at once

## Dynamic data:

- aka **online** data; available as a changing stream e.g. air traffic control, real-time computer vision or sensor processing applications



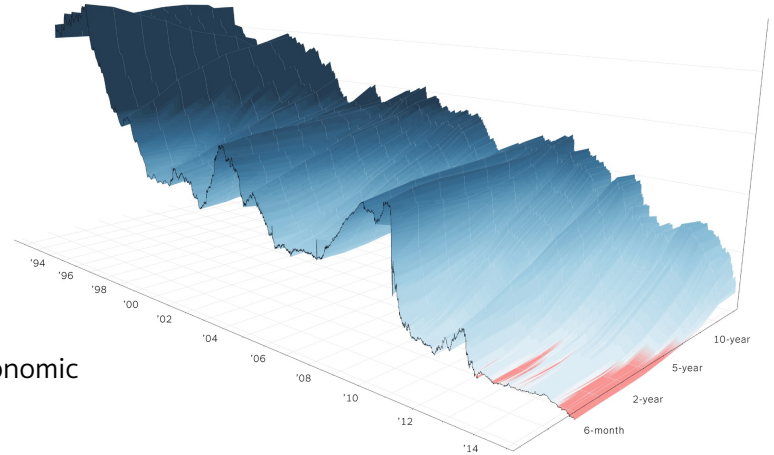
**Time-varying data**: weather simulation




**Real-time data**: Air-traffic control console

## **Also related : Time-varying data / time-series data**

- Could be either static or dynamic
- *Consequences for Encoding* : use animation or map time to position or another ordered visual encoding channel




**Time-series data:** economic yield curve


# Task abstractions

# Task Abstraction

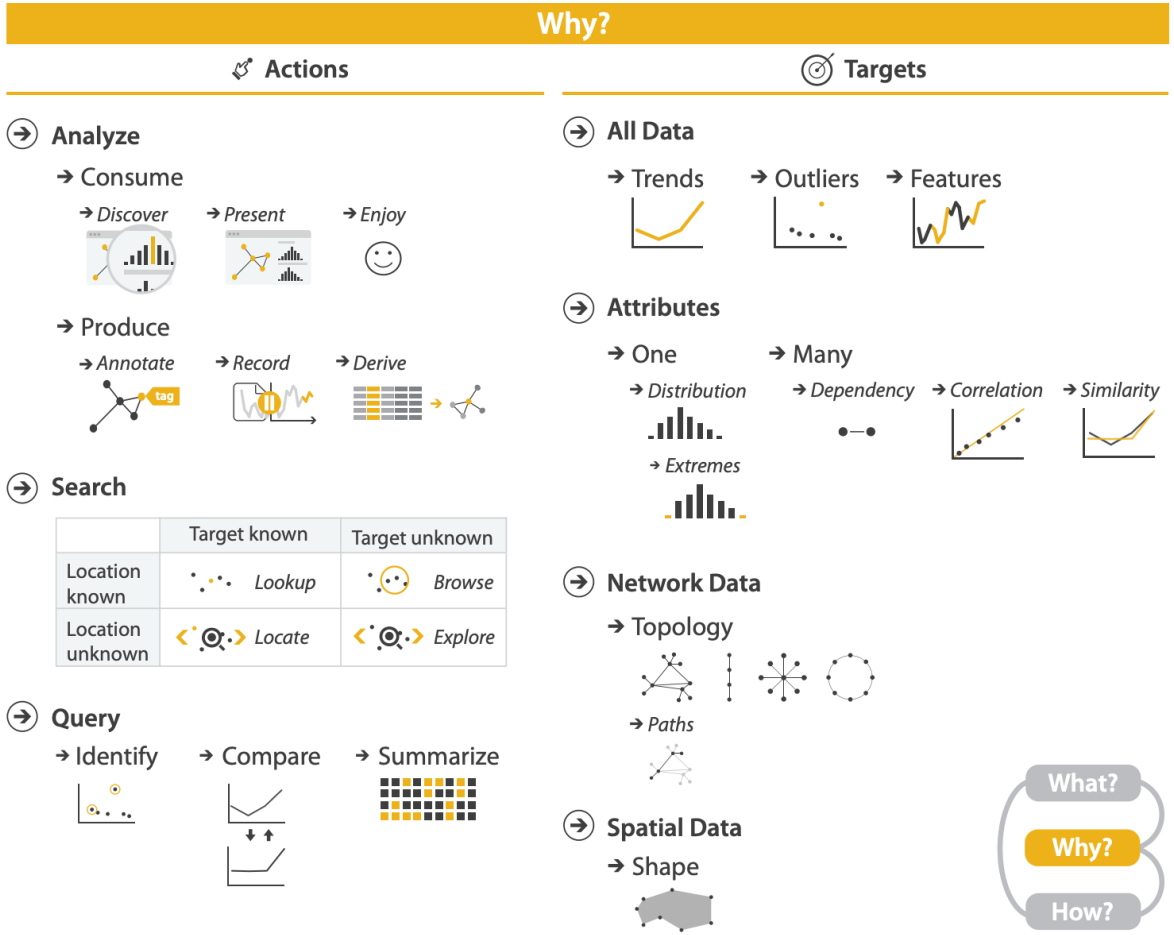
For effective visualization, we must take into account also the task(s) for which visualization is required

Define as action-target pair [Munzner 2014]

- Action: 'verbs', define user goals/process
- Target: 'nouns', define some aspect of data of interest to the user

N.B. Non-exclusive: often there will be multiple sub-tasks

From Munzner [2014]



# Action: **Consume**

**Exploit/utilize information that has already been generated to analyze the data**

- Discover: (Explore)

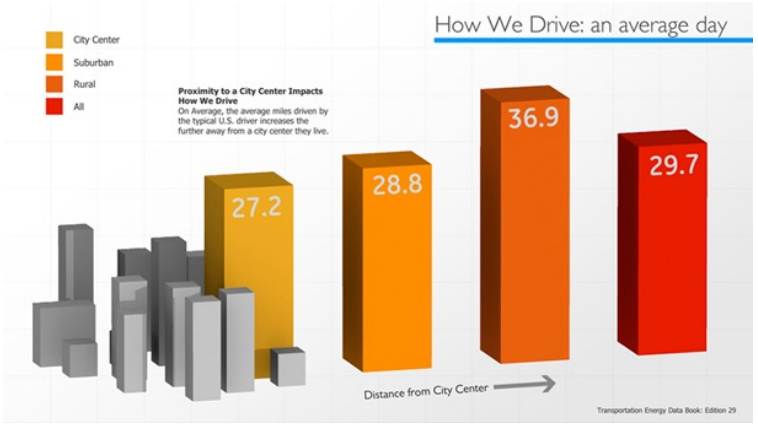
  - Find new knowledge
  - Generate new hypothesis

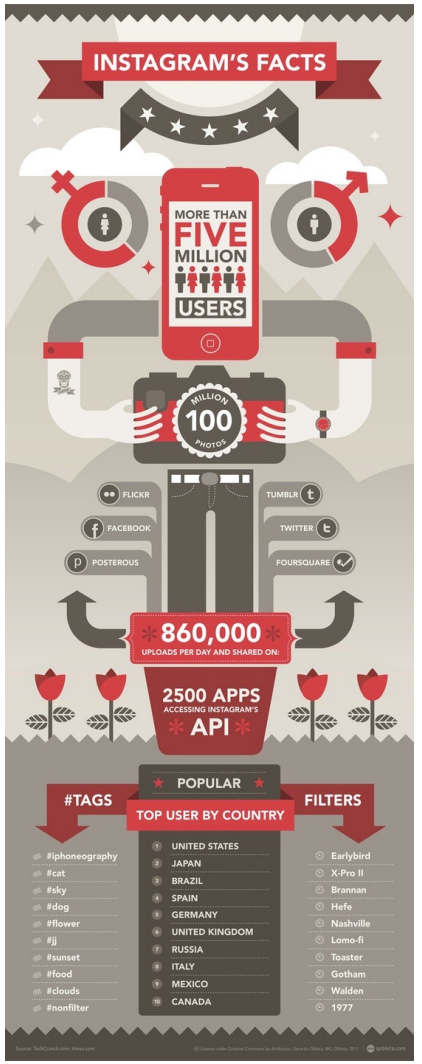
- Present: (Explain/Persuade)

  - Communicate information succinctly
  - Tell a story with data
  - Guide users through a cognitive process

- Enjoy (Casual/Social Visualization)

  - Curiosity driven
  - Encode for Engagement

# Action: **Produce**

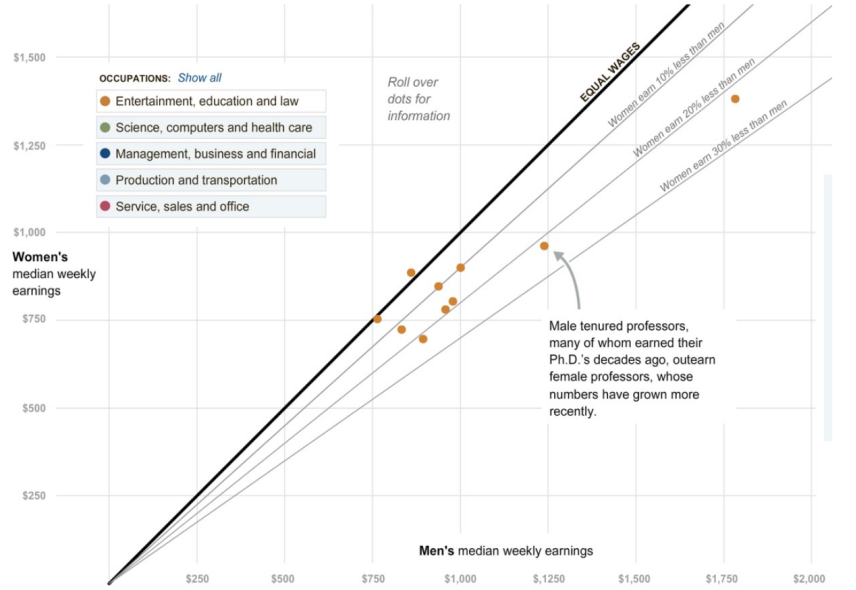
**Generate new material, often output that is used as input to next stage of analysis.**

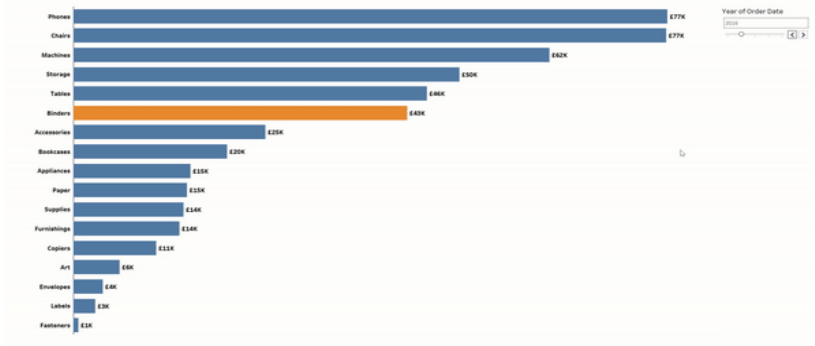
- Present explanatory outputs from exploratory discovery.

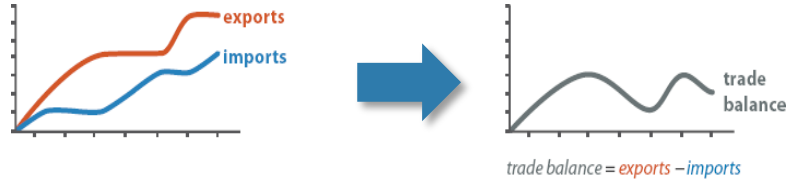

**Annotate** augment with graphical/textual details (often manually)

**Record** capture visualization elements as persistent artefacts

**Derive** produce new data elements based on existing ones.





# Action: **Search**

Images from Munzner (2014)

All viz tasks have search as an intermediate goal

Can be classified based on what user knows of the identity/location of target

- Lookup: user has some notion of location and target [ideal encoding: ordered]
- Locate: user knows target but needs to determine location [selective]
- Browse: user has some notion of location but is unclear what they will find [ordered, selective]
- Explore: user is not sure of either location or what they will find

| | Target known | Target unknown |
|---|---|---|
| Location known | *Lookup* | *Browse* |
| Location unknown | *Locate* | *Explore* |

# Action: **Query**

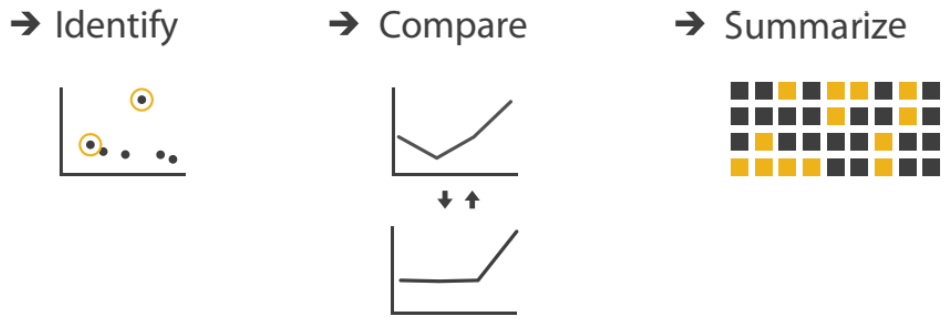
Images from Munzner (2014)

Once a search target is found, analyse it to draw insights

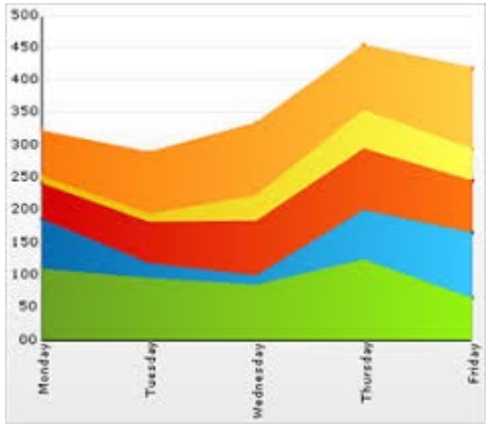
We may sometimes also query on-the-fly within the process of search

Classified by amount of search targets

- Identify: return characteristics of a single search target e.g., is this point an outlier?
- Compare: correlate/contrast multiple targets e.g. which value increases more rabidly?
- Summarize: obtain an overview of multiple possible targets e.g., what feature occurs most often?





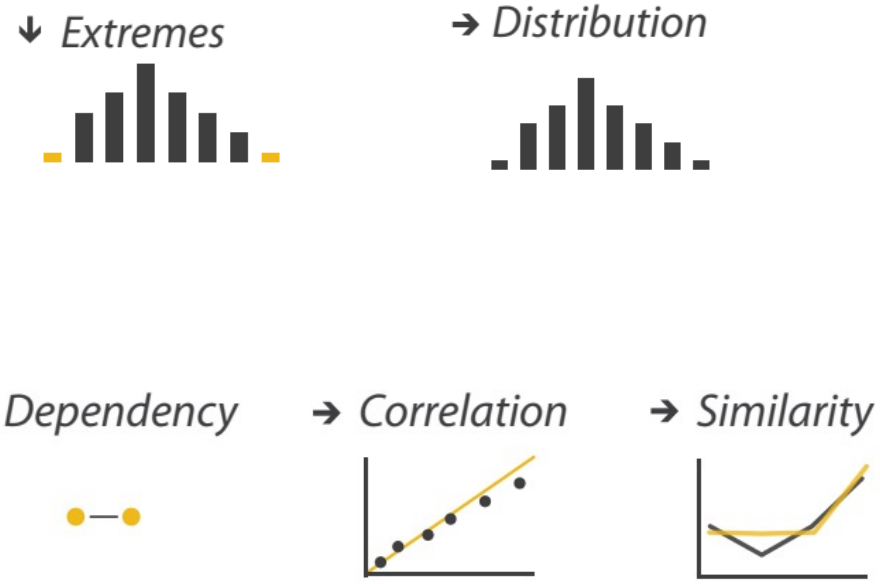


# Attribute Targets

Attribute targets:

- Based on single attribute:
  - Range: min/max of occurrence
  - Distribution: pattern of occurrence

- Based on multiple attributes:
  - Dependencies: value of attribute A depends on that of attribute B
  - Correlation: tendency for values to be tied
  - Similarity: quantitative measure ranking differences between attributes



# General Targets

Analysis actions require a target: some aspect of data of interest to user / relevant to task

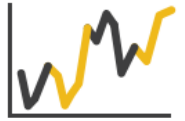
Broadly relevant Targets:

- Trends: high-level characterization of a pattern in the data [ideal encodings should have continuity: connection, associative]
- Outliers: elements of data that don't fit well with the prevailing patterns in the data [selective, discriminable]
- Features: specific structures of interest for a particular task [associative, discriminable]

➔ Trends

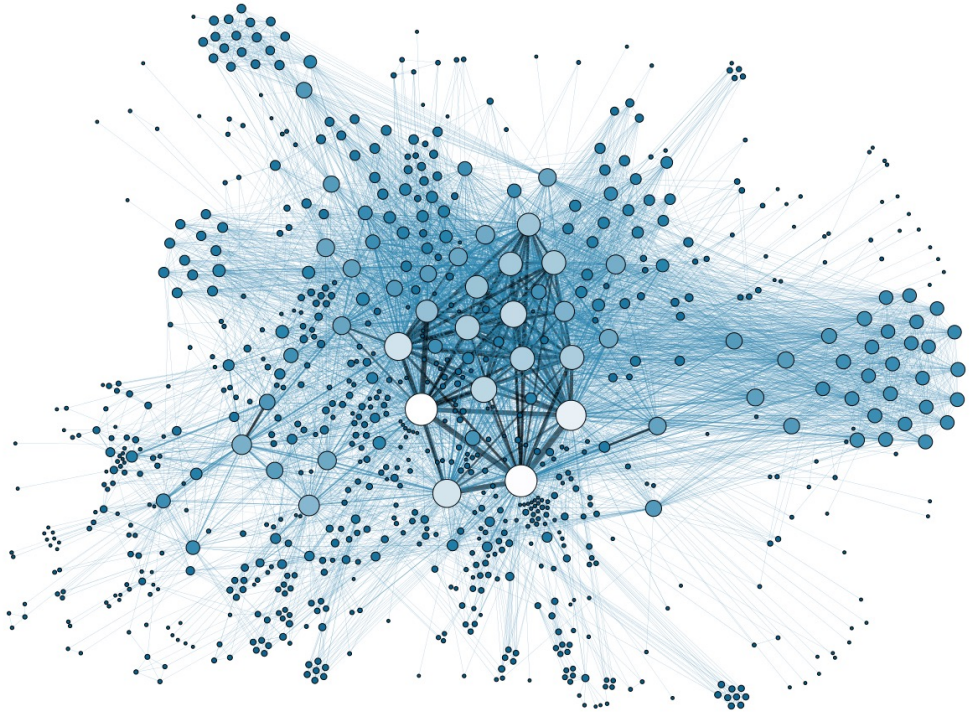
➔ Outliers

➔ Features

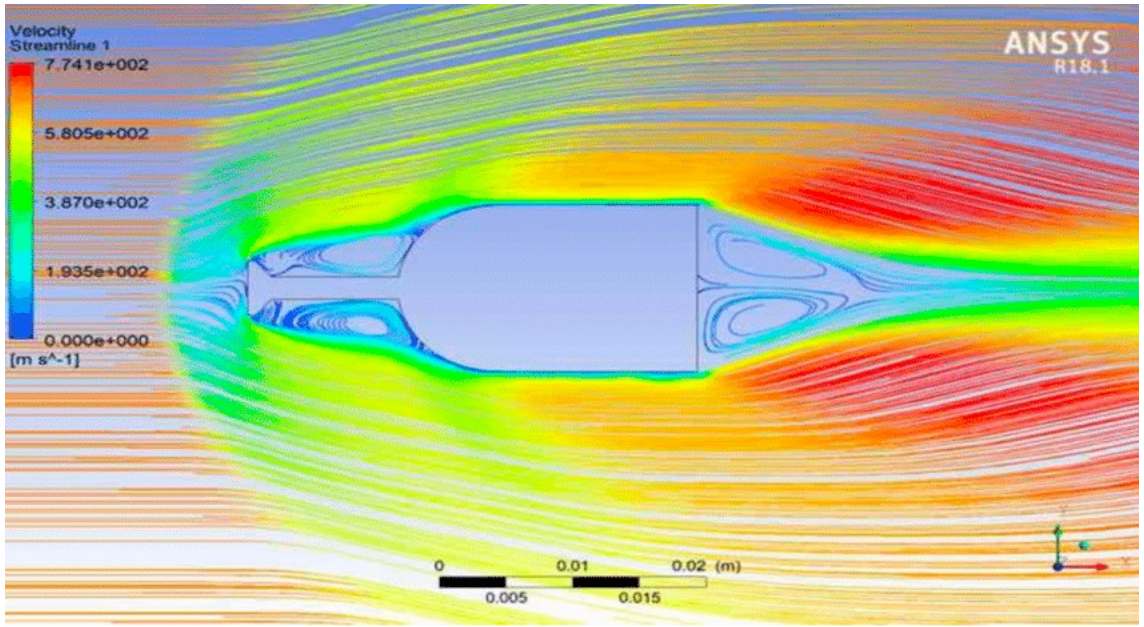


# Specialized Targets

Topology (e.g., paths OR links in a network),



Shape (of spatialized or geometric data)



# Other Classifications

**Important to note:** data/task classifications discussed in previous slides are based on Munzner [2014]. But many other classification schemes exist. None are exhaustive.

# Keller and Kelly [1994]

## Visualization Goals (data types)

- scalar (or scalar field);
- nominal;
- direction (or direction field);
- shape;
- position;
- spatially extended region or object (SERO).

Ref: Peter R. Keller and Mary M. Kelly. Visual Cues: Practical Data Visualization. Los Alamitos, CA: IEEE Computer Society Press, 1994.

## Visualization Tasks

- identify—establish characteristics by which an object is recognizable;
- locate—ascertain the position (absolute or relative);
- distinguish—recognize as distinct or different (identification is not needed);
- categorize—place into divisions or classes;
- cluster—group similar objects;
- rank—assign an order or position relative to other objects;
- compare—notice similarities and differences;
- associate—link or join in a relationship that may or may not be of the same type;
- correlate—establish a direct connection, such as causal or reciprocal.

# Schneiderman [1996]

Task Classification is based on the behavior of the analyst or visualization creator (rather than the reader)

## Data Types

- one-dimensional linear;
- two-dimensional map;
- three-dimensional world;
- temporal;
- multidimensional;
- tree;
- network.

Ref: Ben Shneiderman. “The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations.” In Proceedings of the 1996 IEEE Symposium on Visual Languages, pp. 336–343. Los Alamitos, CA: IEEE Computer Society, 1996.

## Tasks (Schneiderman 1996)

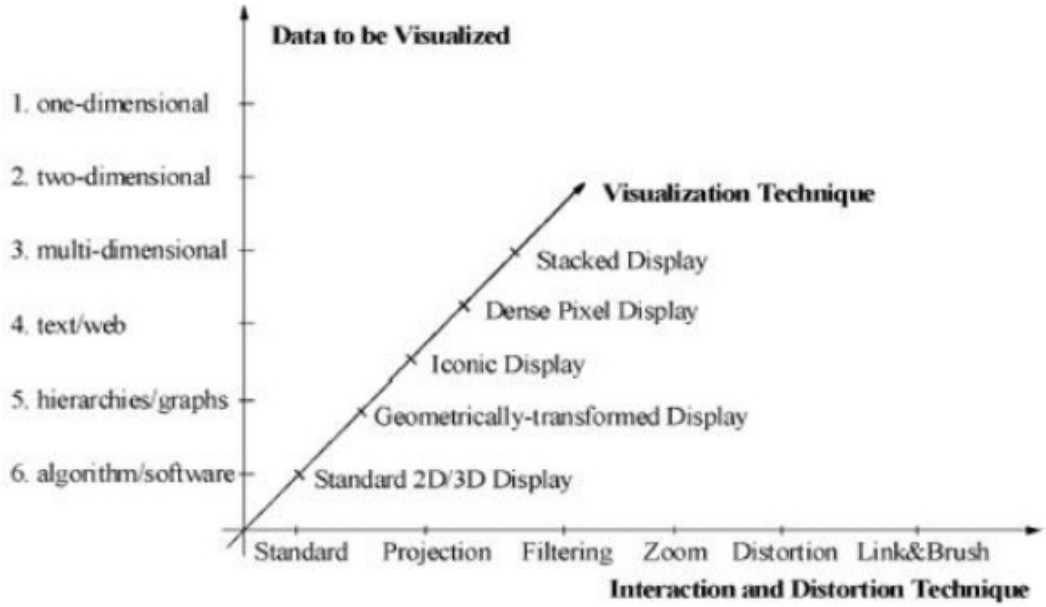
- **Zoom.** Zoom in items of interest to gain a more detailed view, e.g., holding down a mouse button to enlarge a region of the display.
- **Filter**. Filter out uninteresting items to allow the user to reduce the size of a search, e.g., dynamic queries that can be invoked via sliders.
- **Details-on-demand**. Select an item or group and get details when needed, e.g., a pop-up window can show more attributes of a specific object on the screen.
- **Relate**. View relationships among items, e.g., select a particular object that can then show all other objects related to it.
- **History**. Keep a history to allow undo, replay, and progressive refinement, such as allowing a mistake to be undone, or a series of steps to be replayed.
- **Extract.** Extract the items or data in a format that would facilitate other uses, i.e., saving to file, sending via e-mail, printing, or dragging into another application (statistical or presentation package)

# Keim [2002]

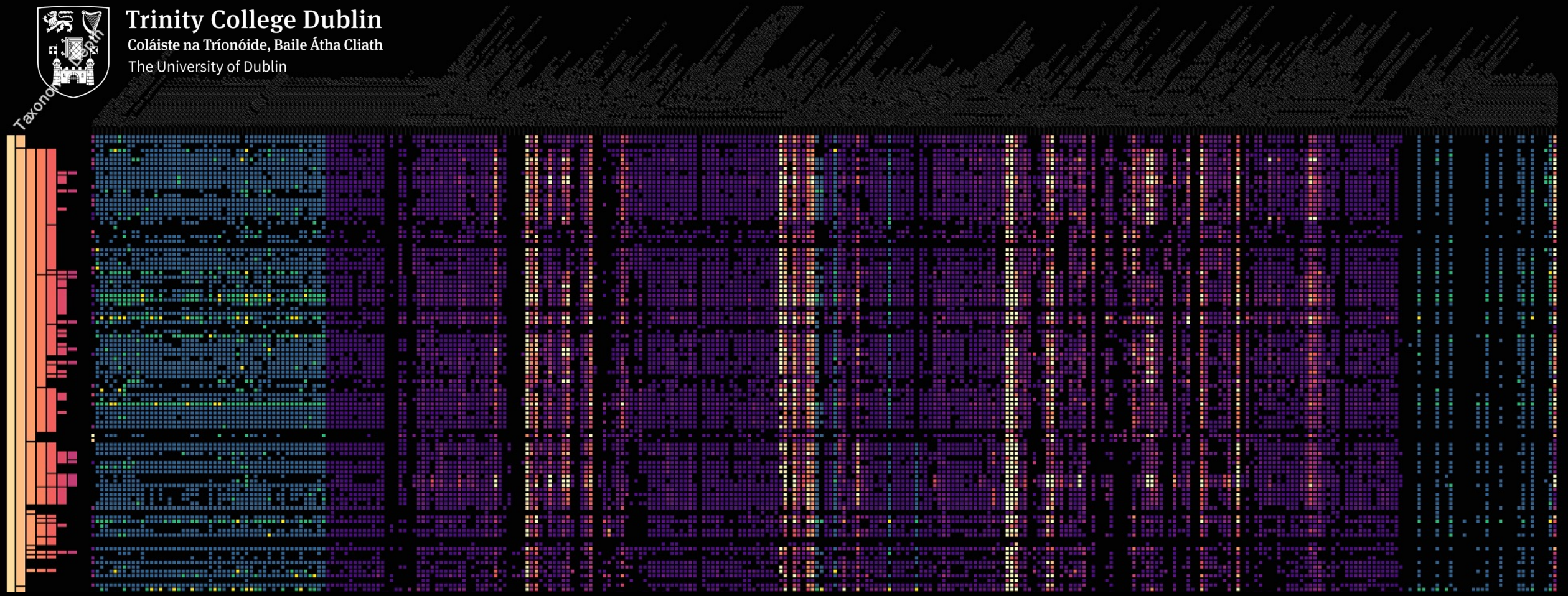
## Classification of Data Types

- One-dimensional data e.g. stock prices
- Two-dimensional data—e.g. floor plans
- Multidimensional data—e.g. relational tables
- Text and hypertext—e.g. web documents
- Hierarchies and graphs—e.g., telephone/network traffic
- Algorithm and software—e.g. execution traces

## Classification of visualization techniques

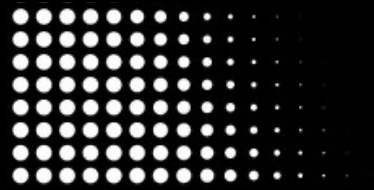



Ref. Daniel A. Keim. "Information Visualization and Visual Data Mining." IEEE Transactions on Visualization and Computer Graphics 8:1 (2002), 1–8.

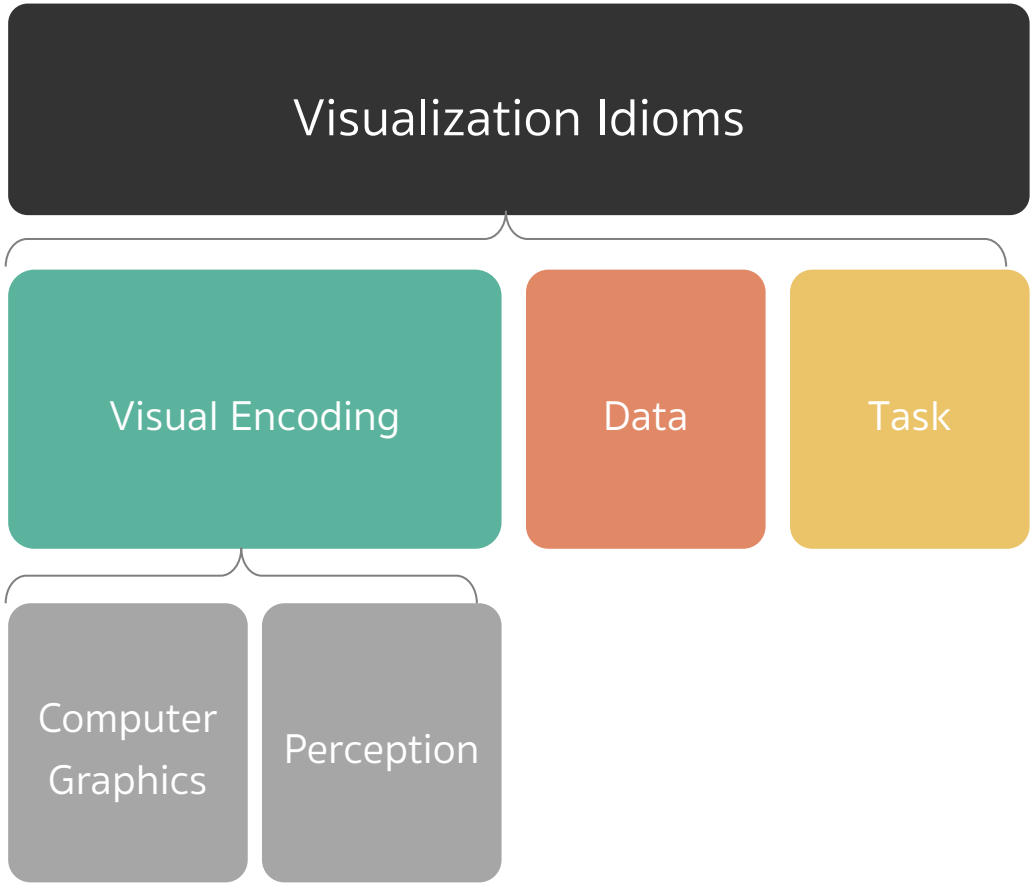



# Visualization Idioms

Some examples / A preview of more detailed discussion tomorrow



# Next : **Visualization Idioms, How?**




**How?**

**Encode**

- **Arrange**
  - Express
  - Separate
  - Order
  - Align
  - Use
- **Map** from categorical and ordered attributes
  - Color
    - *Hue*
    - *Saturation*
    - *Luminance*
  - Size, Angle, Curvature, ...
  - Shape
  - Motion
    *Direction, Rate, Frequency, ...*

**Manipulate**

- **Change**
- **Select**
- **Navigate**

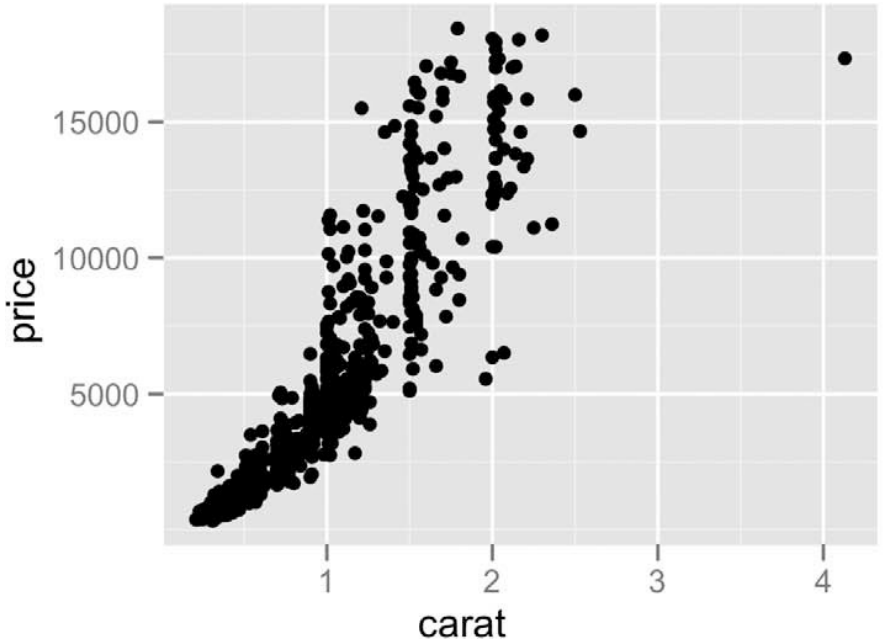
**Facet**

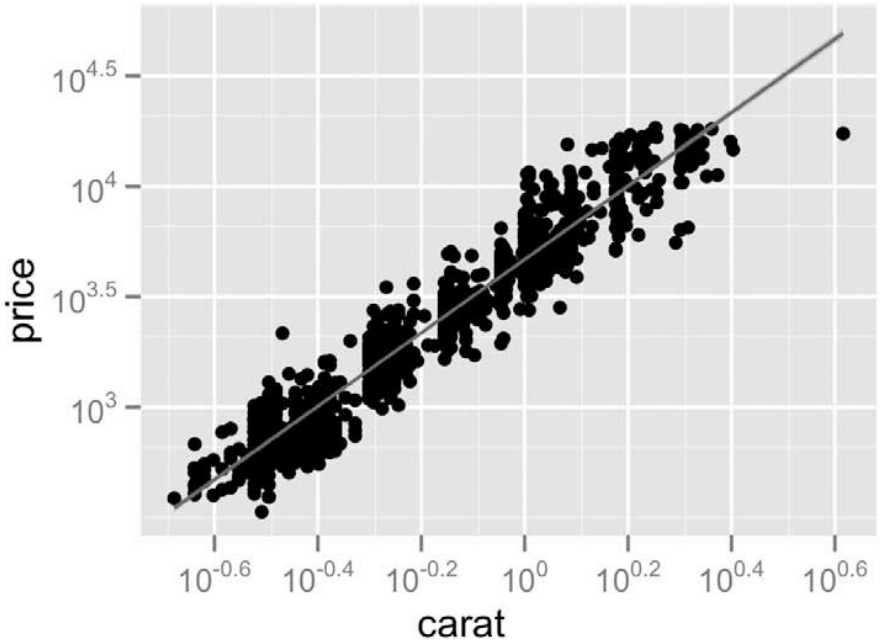
- **Juxtapose**
- **Partition**
- **Superimpose**

**Reduce**

- **Filter**
- **Aggregate**
- **Embed**

# Example: **Scatterplot**



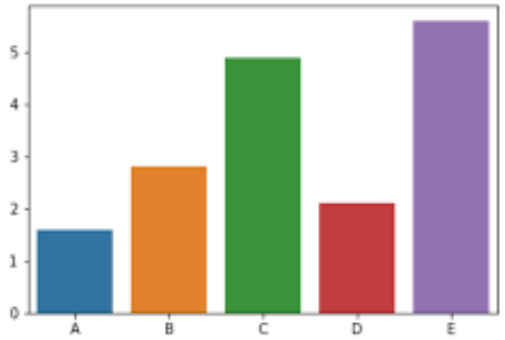



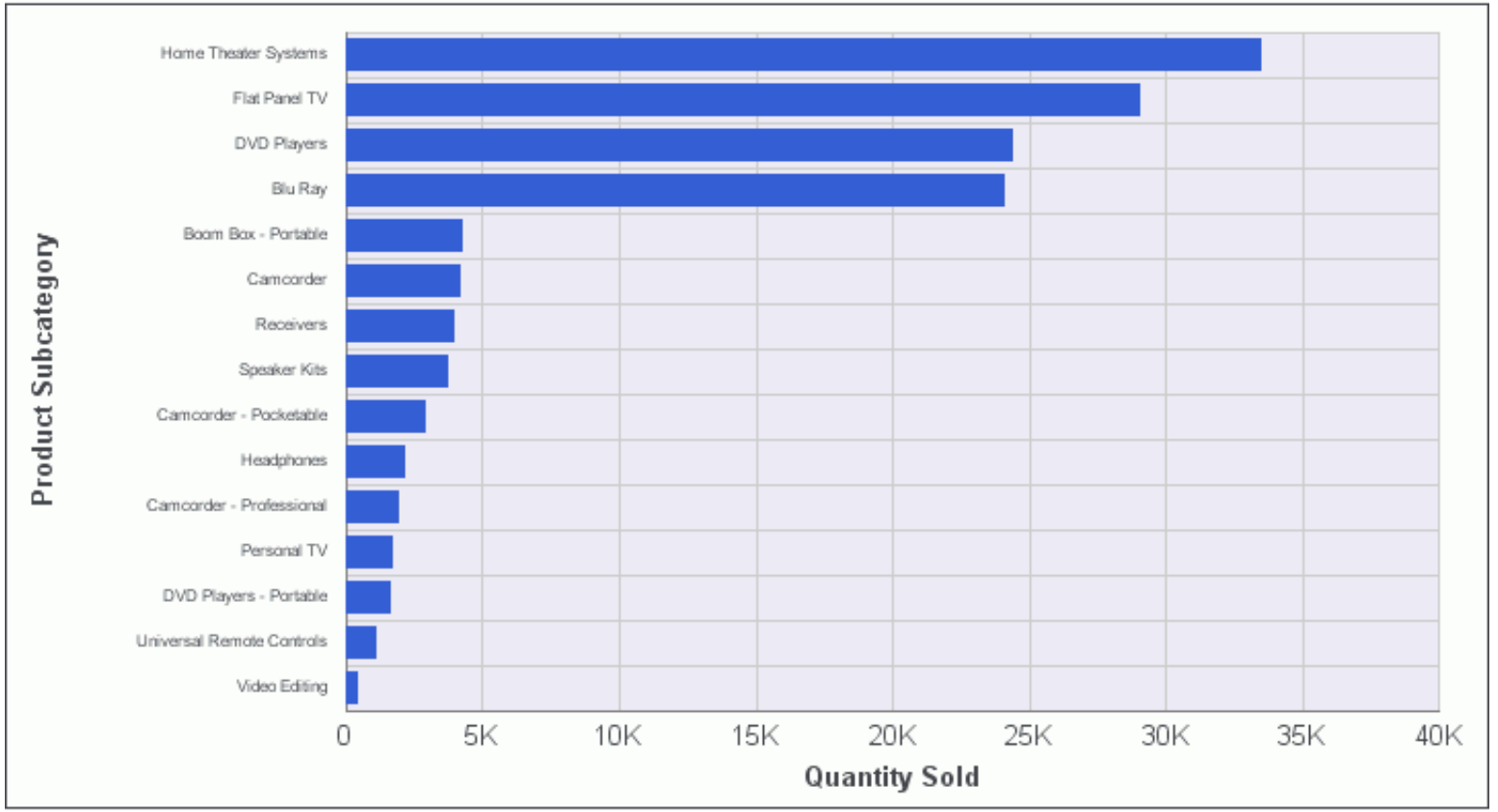



| Idiom | SCATTERPLOT |
|---|---|
| Data: WHAT? | Table Dataset, 2 quantitative continuous measurement attributes |
| Encoding: HOW? | Point marks **express** quant. attributes with horizontal and vertical position |
| Task: WHY? | **Discover** trends, outliers, distribution, correlation; **Locate** clusters |

***Summary design and analysis.*** Note that this doesn't mean all scatter plots are limited to these parameters

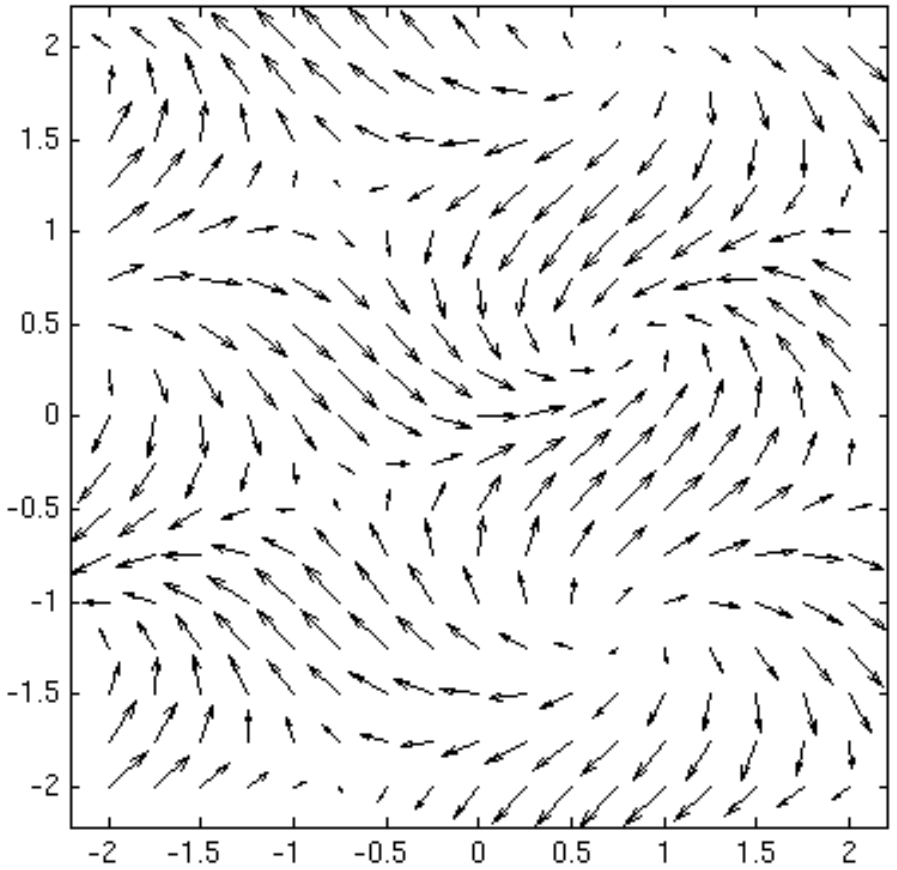
# Example: **Bar chart**



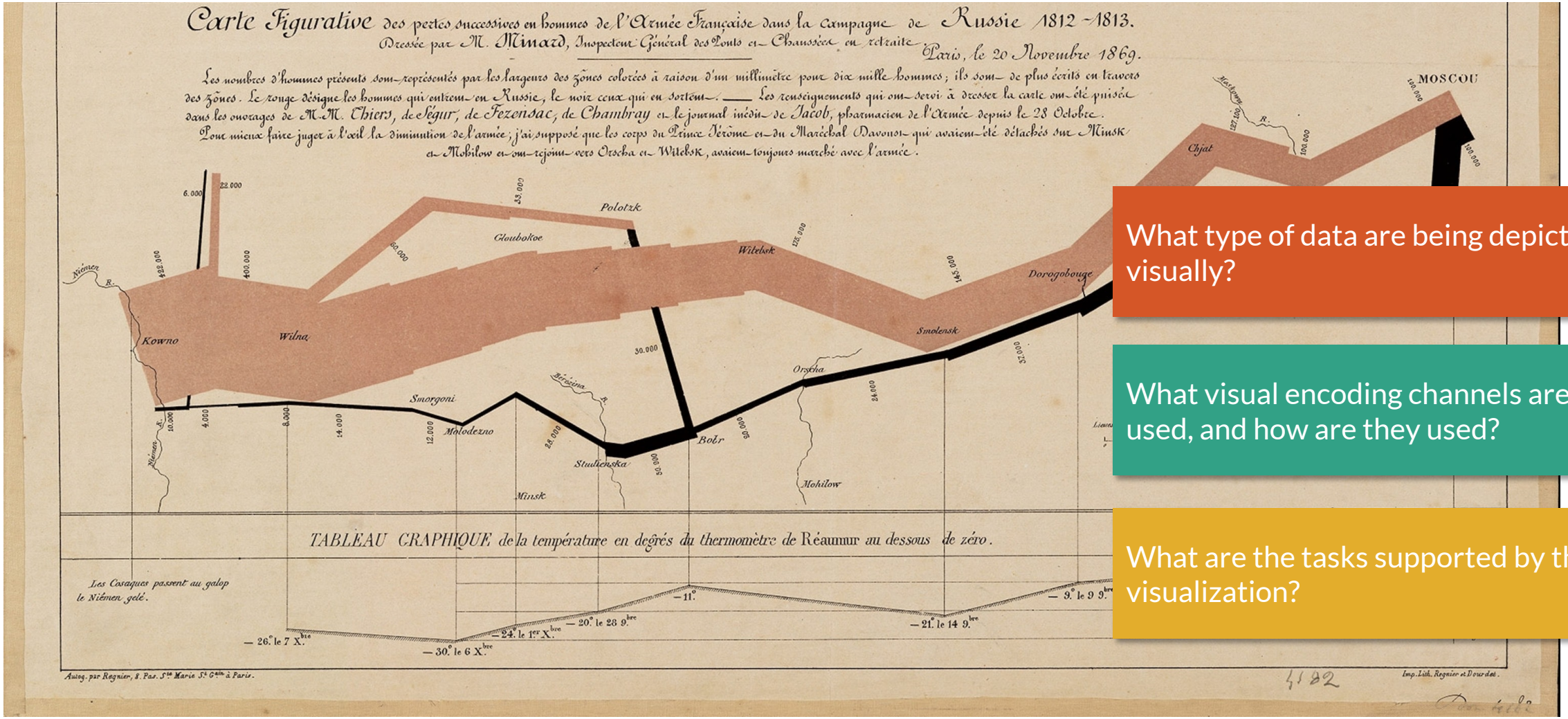

| Idiom | HORIZONTAL BAR CHART |
|---|---|
| Data: WHAT? | Table dataset. One quantitative + one categorical attribute |
| Encoding: HOW? | **Express** quantitative attribute with aligned position/length in primary direction, **separate** key attribute with position in second direction |
| Task: WHY? | Lookup and compare values |

# More Specialized Example: **Vector Field**




| Idiom | DISCRETIZED VECTOR FIELD |
|---|---|
| Data: WHAT? | Grid with high-dimensional quantitative attribute (here, a 2D-vector) sampled at each grid position |
| Encoding: HOW? | Arrow marks with angular encoding for direction, hue for magnitude. |
| Task: WHY? | Discover/present/identify trends, features. Derive shape or outliers. |

# Try this at home: **More complex example**




What type of data are being depicted visually?

What visual encoding channels are used, and how are they used?

What are the tasks supported by the visualization?

# Recommended Readings : This Lecture

## Data Abstraction

- Chapter 2 – "Data Abstraction" in **Visualization Analysis and Design**, Tamara Munzner 2014 [Available as e-book in Library Reading Rooms]
- *Alternative reference*: the following early draft Munzner's book is available free online. Chapter 3 discusses Data Abstractions:
  - Visualization Design and Analysis: Abstractions, Principles, and Methods (DRAFT). Tamara Munzner, 2012.

    [ https://web.cse.ohio-state.edu/~machiraju.1/teaching/CSE5544/ClassLectures/PDF-old/book.120803.pdf#page=60 ]

## Task Abstraction

- Chapter 3 – "Task Abstraction" in **Visualization Analysis and Design**, Tamara Munzner 2014 [Available as e-book in Library Reading Rooms]
- *Alternative reference*: The following paper co-authored by Munzner, freely available online, discusses some of the aspects of task abstraction covered in Munzner's book:
  - A Multi-Level Typology of Abstract Visualization Tasks. Brehmer and Munzner, Infoviz, 2013.

    [ https://www.cs.ubc.ca/labs/imager/tr/2013/MultiLevelTaskTypology/brehmer_infovis13.pdf ]

N.B. The actual 2014 book by Munzner is a much better to read, if you can get it,

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

# END