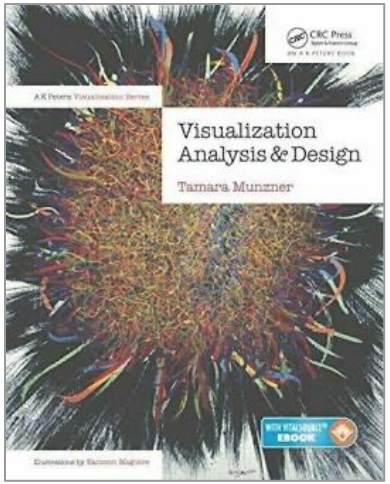
Tamara Munzner

Professor of Comp Sci Dept at the University of British Columbia, Canada



Website

Talks

On YouTube

on Safari

> The essence of abstraction is preserving information that is relevant in a given context, and forgetting information that is irrelevant in that context.
>
> – John V. Guttag[1]

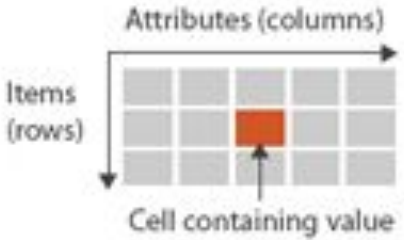
What: Data Abstraction.

Many aspects of vis design are driven by the kind of data that you have at your disposal.

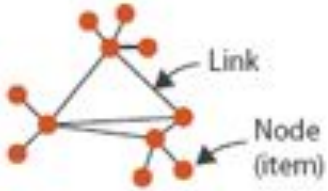
Chapter 2. Summary.

Dataset Types

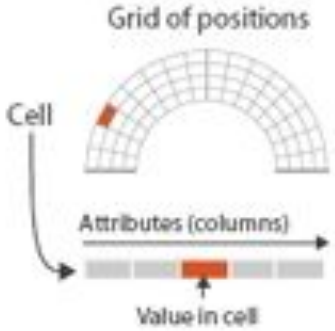
Tables



Networks

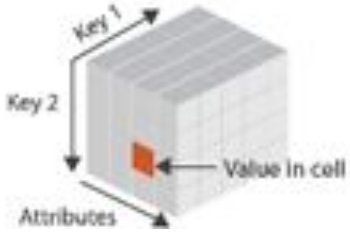


Fields (Continuous)



Geometry (Spatial)

Multidimensional Table



Trees

Continuous fields have grids based on spatial positions where cells contain attributes.

https://learning.oreilly.com/library/view/visualization-analysis-and/9781466508910/K14708_C002.xhtml

Data Types

- Items
- Attributes
- Links
- Positions
- Grids

An item is an individual entity that is discrete, such as a row in a simple table or a node in a network: people, stocks, coffee shops, genes, or cities.
An attribute is some specific property that can be measured, observed, or logged: salary, price, number of sales, protein expression levels, or temperature.
A link is a relationship between items, typically within a network.
A position is spatial data, providing a location in two-dimensional (**2D**) or three-dimensional (**3D**) space: a latitude–longitude pair describing a location on the Earth's surface or three numbers specifying a location within the region of space measured by a medical scanner.
A grid specifies the strategy for sampling continuous data in terms of both geometric and topological relationships between its cells.

https://learning.oreilly.com/library/view/visualization-analysis-and/9781466508910/K14708_C002.xhtml

The datasets are made up of five core data types: items, attributes, links, positions, and grids.

https://learning.oreilly.com/library/view/visualization-analysis-and/9781466508910/K14708_C002.xhtml

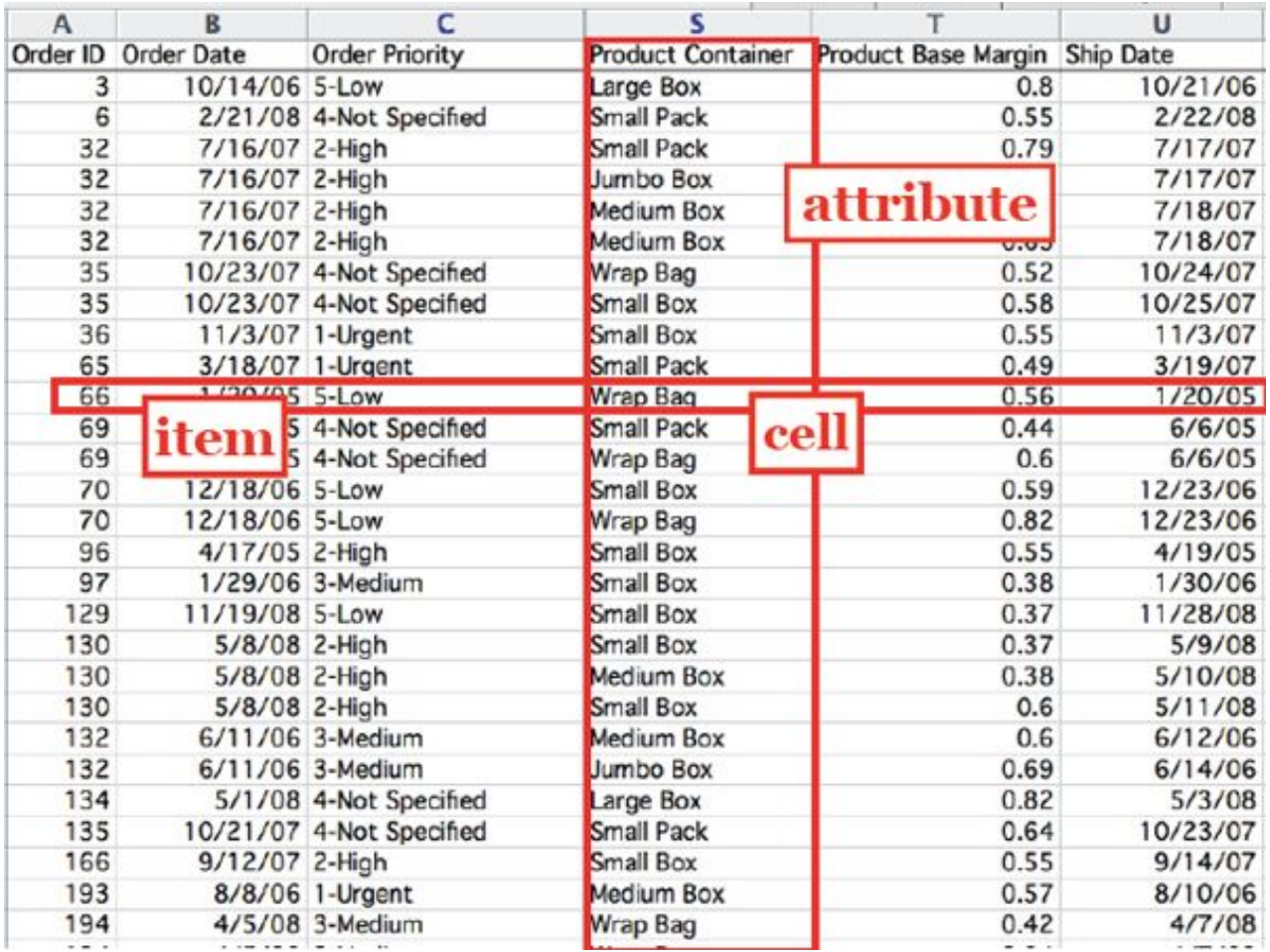
Dataset Availability

- Static

- Dynamic

For any of these dataset types, the full dataset could be **available immediately** in the form of a static file, or it might be **dynamic data processed gradually** in the form of a stream.

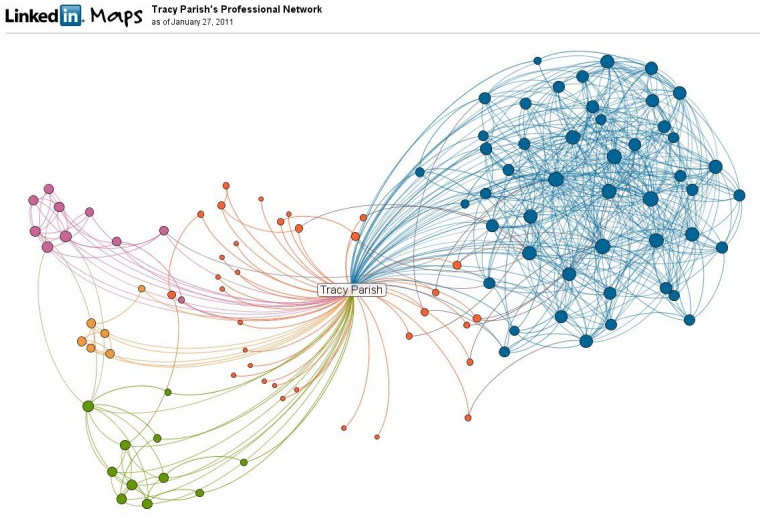
https://learning.oreilly.com/library/view/visualization-analysis-and/9781466508910/K14708_C002.xhtml



https://learning.oreilly.com/library/view/visualization-analysis-and/9781466508910/K14708_C002.xhtml

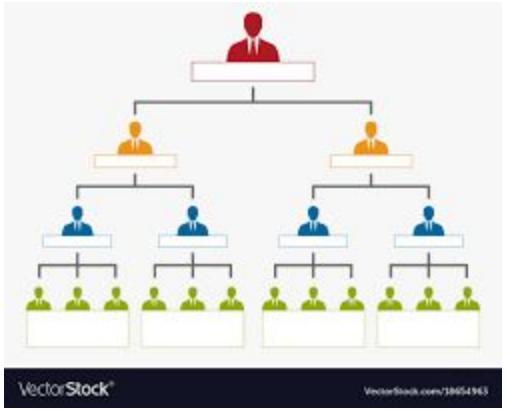
A Network (graph) and a Tree.

Tracy's professional network on LinkedIn.



A network is a collection of **nodes** and **links**. In this example, people are nodes and their friendships are links.

Corporate organization hierarchy.



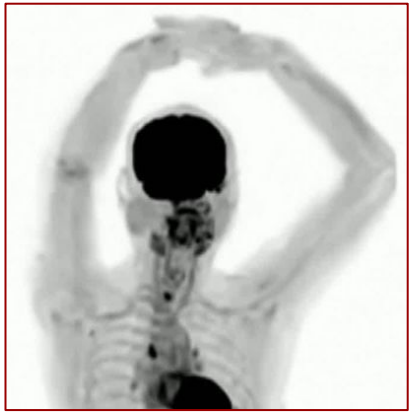
Networks with hierarchical structure are also known as trees.

A Field Dataset Example

A medical scan of a human body containing measurements indicating the density of tissue at many sample points, spread regularly throughout the volume of 3D space.

A field of sampled values (a.k.a a spatial field):

- Represent continuous data;
- Have grids based on spatial positions where cells contain attributes;



Attributes

Attribute Types

- Categorical

- Ordered
  - *Ordinal*

  - *Quantitative*

Ordering Direction

- Sequential

- Diverging

- Cyclic

https://learning.oreilly.com/library/view/visualization-analysis-and/9781466508910/K14708_C002.xhtml

Categories:

1. Can only distinguish whether two things are the same or different.
2. Does not have an implicit ordering.
3. Ordering is possible based on arbitrary, external information, e.g. fruits can be ordered by its name or price, but only if this information is available.

https://learning.oreilly.com/library/view/visualization-analysis-and/9781466508910/K14708_C002.xhtml

Categorical (Nominal) Data **Examples**

Movie genres: Action, Comedy, Fantasy

File Types: txt, pdf, mp4

Favorite Fruit: apples, oranges

City Names: NY, LA, Miami Beach

https://learning.oreilly.com/library/view/visualization-analysis-and/9781466508910/K14708_C002.xhtml

Ordered Attributes

Ordinal

Categories are ordered, but the difference between them is not meaningful.

Examples:

Shirt Size: Large, Medium, Small

Ranking: Never, Sometimes, Often

Quantitative

Numeric values: int, floats.

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Specified	Small Box	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Small Box	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

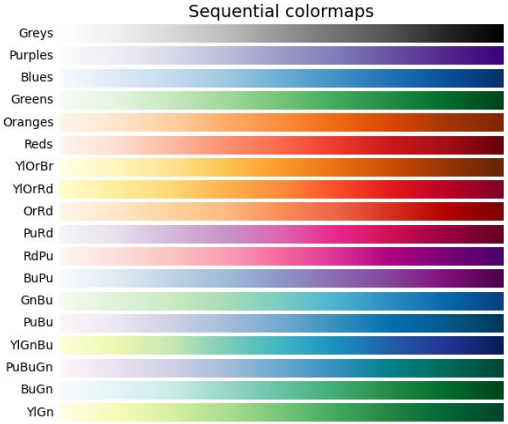
quantitative
ordinal
categorical

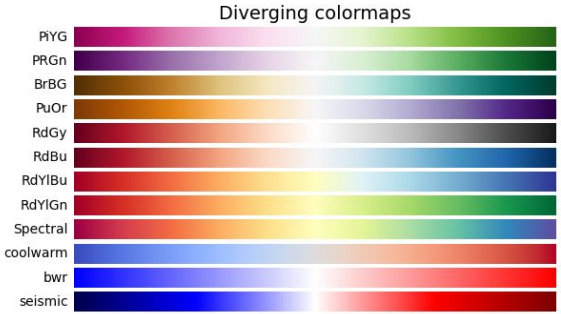
https://learning.oreilly.com/library/view/visualization-analysis-and/9781466508910/K14708_C002.xhtml

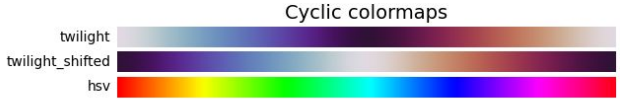
Ordering Direction

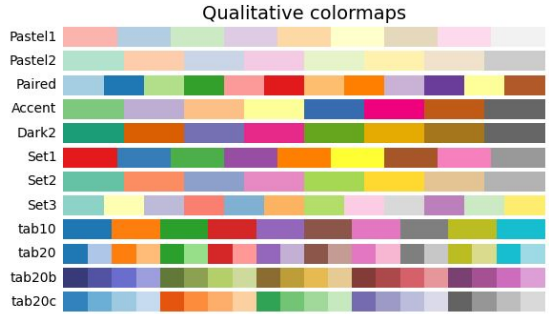
Sequential	**Diverging**	**Cyclic**
Homogeneous range from a min to a max value.	Can be deconstructed into two sequences pointing in opposite directions that meet at **a common zero point**.	The values wrap around **back to a starting point** rather than continuing to increase indefinitely.
e.g., a mountain height from 0 (sea level) to a max point of Mount Everest.	e.g., the values go up for mountains and down for undersea valleys.	e.g., the hour of the day, the day of the week, and the month of the year.

Example: MatPlotLib Colormaps









https://matplotlib.org/stable/tutorials/colors/colormaps.html