CSE3020 - Data Visualization

Dr. K.P. Vijayakumar, VIT Chennai

Syllabus - Data Visualization

- **Module 1 : Introduction to Data Visualization**
  - Overview of data visualization - Data Abstraction - Task Abstraction - Analysis: Four Levels for Validation
- **Module 2 :Visualization Techniques**
  - Scalar and point techniques- Color maps Contouring Height Plots - Vector visualization techniques- Vector properties Vector Glyphs Vector Color Coding Stream Objects
- **Module 3 :Visual Analytics**
  - Visual Variables- Networks and Trees - Heat Map, Map Color and Other Channels- Manipulate View

Syllabus - Data Visualization

- **Module 4 :Visual Analytics**
  - Arrange Tables, Geo Spatial data Reduce Items and Attributes
- **Module 5 : Visualization Tools and Techniques**
  - Introduction to data visualization Tools - Tableau – Visualization using R
- **Module 6 :Diverse Types Of Visual Analysis**
  - Time- Series data visualization Text data visualization Multivariate data visualization and case studies
- **Module 7 :Visualization Dashboard Creations**
  - Dashboard creation using visualization tools for the use cases: Finance-marketing-insurance healthcare etc.,
- **Module 8 : Recent Trends**
  - Industry Expert Talk

What is Data Visualization?

- Graphical representation of data
- Effective way to communicate with data
- Visual elements
  - Charts
    - Pie charts, Bar charts, Histograms, Gantt charts, Box-and-whisker plots, Waterfall charts, Area charts, Scatter plots
  - Graphs
  - Maps
- Data Visualization Tools
  - To see and understand
    - Trends
    - outliers
    - Patterns

Image Source: Boostlabs.com

What is Data Visualization?

- **Data Visualization Tools**
  - Tableau, Qlikview, Fusionchart
  - Highcharts,Plotly, Sisense, Infogram
  - ChartBlocks, Google Charts, Datawrapper
  - D3.Js, Chart.js
  - Polymaps

What is Data Visualization?

- **Why is Data Visualization important?**
  - makes it easier to understand and draw insights from
  - Takes raw data, model it and deliver the data
- **Benefits of Data Visualization**
  - Correlations in Relationship
  - Trends Over Time
    - Predictions – Past and Present
  - Frequency
  - Examining the Market
  - Risk and Reward
    - pinpoint areas that may or may not require action
  - Reacting to the Market
    - helps to avoid making mistakes

Source: https://analytiks.co/importance-of-data-visualization/

Visualization (Vis)

- Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively
- Vis systems are appropriate for use when your goal is to augment human capabilities, rather than completely replace the human in the loop.

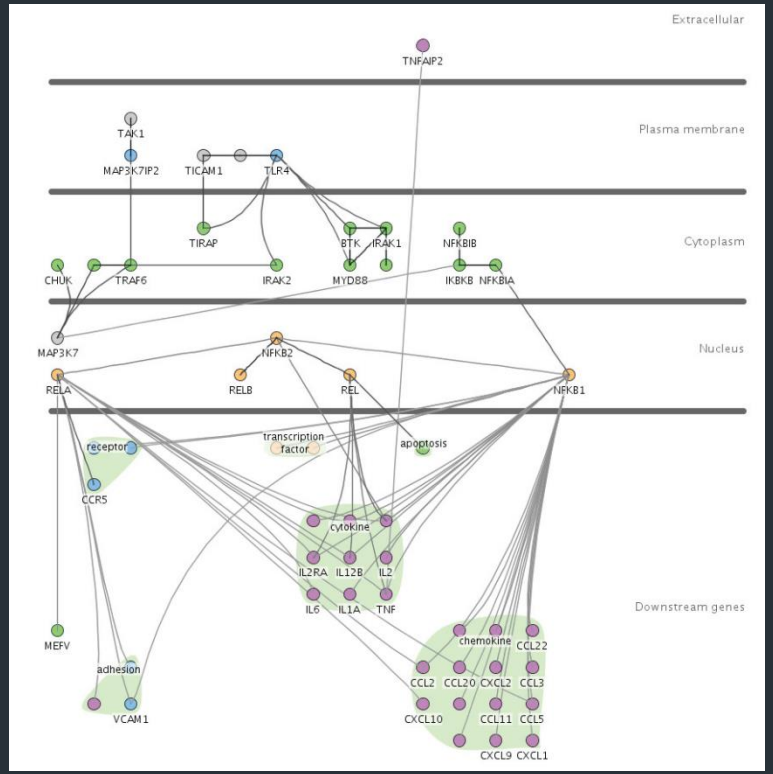
Source: Tamara Munzer, Visualization Analysis and Design

Why have a human in loop?

- Vis allows people to analyze data when they don’t know exactly what questions they need to ask in advance
- Example : Descriptive analysis

Source: Tamara Munzer, Visualization Analysis and Design

Why have a computer in loop?



- By enlisting computation, you can build tools that allow people to explore or present large datasets that would be completely infeasible to draw by hand, thus opening up the possibility of seeing how datasets change over time
- Ex: A small network of 57 nodes and 74 edges

Source: Tamara Munzer, Visualization Analysis and Design

Why Show the Data in Detail?

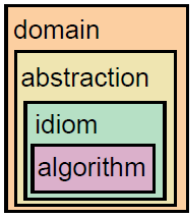
- Vis tools help people in situations where seeing the dataset structure in detail is better than seeing only a brief summary of it.
- Summaries lose information, detail matter
  - confirm expected ones and find unexpected ones
  - Assess the validity of the model

Source: Tamara Munzer, Visualization Analysis and Design

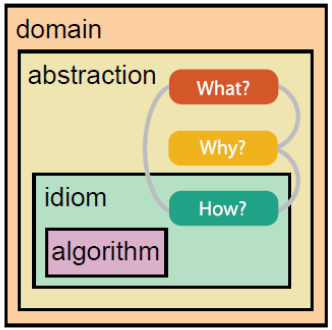
Analysis

Analysis framework: Four levels, three questions

- *domain* situation
  - who are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**
    - often don't just draw what you're given: transform to new form
  - **why** is the user looking at it? **task abstraction**
- *idiom*
  - **how** is it shown?
    - **visual encoding idiom**: how to draw
    - **interaction idiom**: how to manipulate
- *algorithm*
  - efficient computation



[A Nested Model of Visualization Design and Validation. *Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).*]

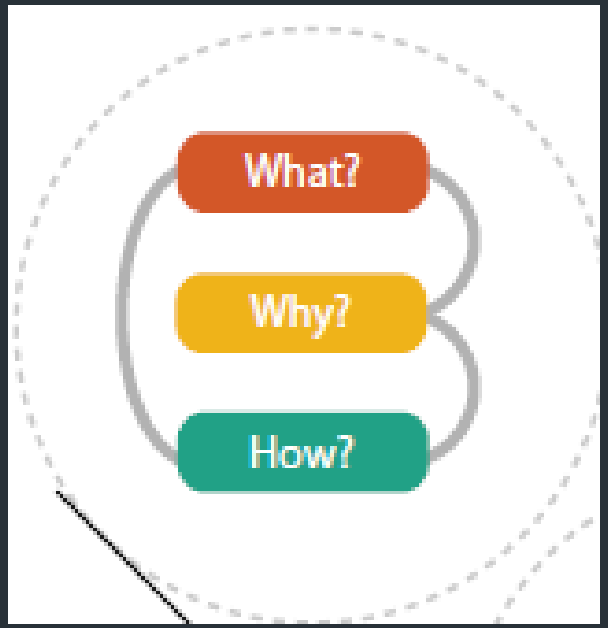


[A Multi-Level Typology of Abstract Visualization Tasks *Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).*]

16

Source: Tamara Munzer, Visualization Analysis and Design

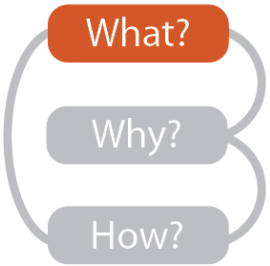
Analysis



- What data the user sees,
- Why the user intends to use a vis tool,
- How the visual encoding and interaction idioms (*approach to create and manipulation visual representation- chart, graph etc*) are constructed in terms of design choices

Source: Tamara Munzer, Visualization Analysis and Design

Data Abstraction



What?

Datasets

- Data Types
  - Items
  - Attributes
  - Links
  - Positions
  - Grids
- Data and Dataset Types

Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

- Dataset Types
  - Tables
  - Networks
  - Fields (Continuous)
  - *Multidimensional Table*
  - *Trees*
- Geometry (Spatial)
- Dataset Availability
  - Static
  - Dynamic

Attributes

- Attribute Types
  - Categorical
  - Ordered
    - *Ordinal*
    - *Quantitative*
- Ordering Direction
  - Sequential
  - Diverging
  - Cyclic

Source: Tamara Munzer, Visualization Analysis and Design

Data Abstraction

- Why do Data Semantics and Types matters?
- Data Semantics
  - Real world meaning
  - Ex: Ram Male 50 50000 Clerk
- Types
  - structural or mathematical interpretation
  - What kind of things?
    - Item, link or attribute
  - Combined into large structure?
    - Table, Tree or Field

Source: Tamara Munzer, Visualization Analysis and Design

Data Abstraction

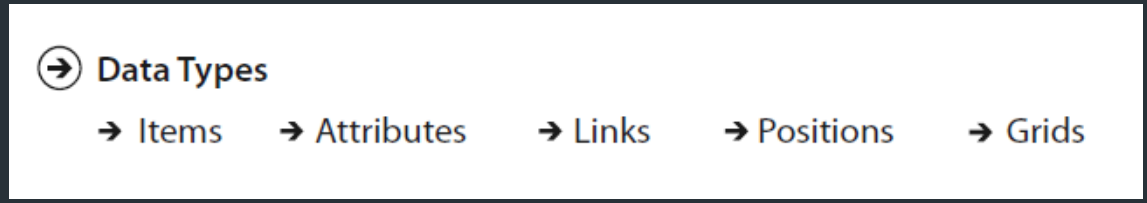
- Types
  - Quantitative data
  - Categorical data
  - A full table with column titles that provide the intended semantics of the attributes

Name	Gender	Age	Salary	Designation
Ram	Male	50	50000	Manager
Sam	Male	40	40000	Clerk
Kumar	Male	30	25000	Clerk

Source: Tamara Munzer, Visualization Analysis and Design

Data Abstraction

- Data Types
  - **Five basic data types**



  - **Items – individual entity**
    - Ex: Person- Student or Employee, city
  - **Attribute- Property of an item**
    - Ex: Salary, Price, No. of Students
  - **Links – relationship between items**
    - Network

Source: Tamara Munzer, Visualization Analysis and Design

Data Abstraction

- Data Types
  - **Positions**
    - Spatial data
      - Latitude and longitude
      - 2D or 3D space
  - **Grids**
    - strategy for sampling continuous data in terms of both geometric and topological relationships between its cells

Source: Tamara Munzer, Visualization Analysis and Design

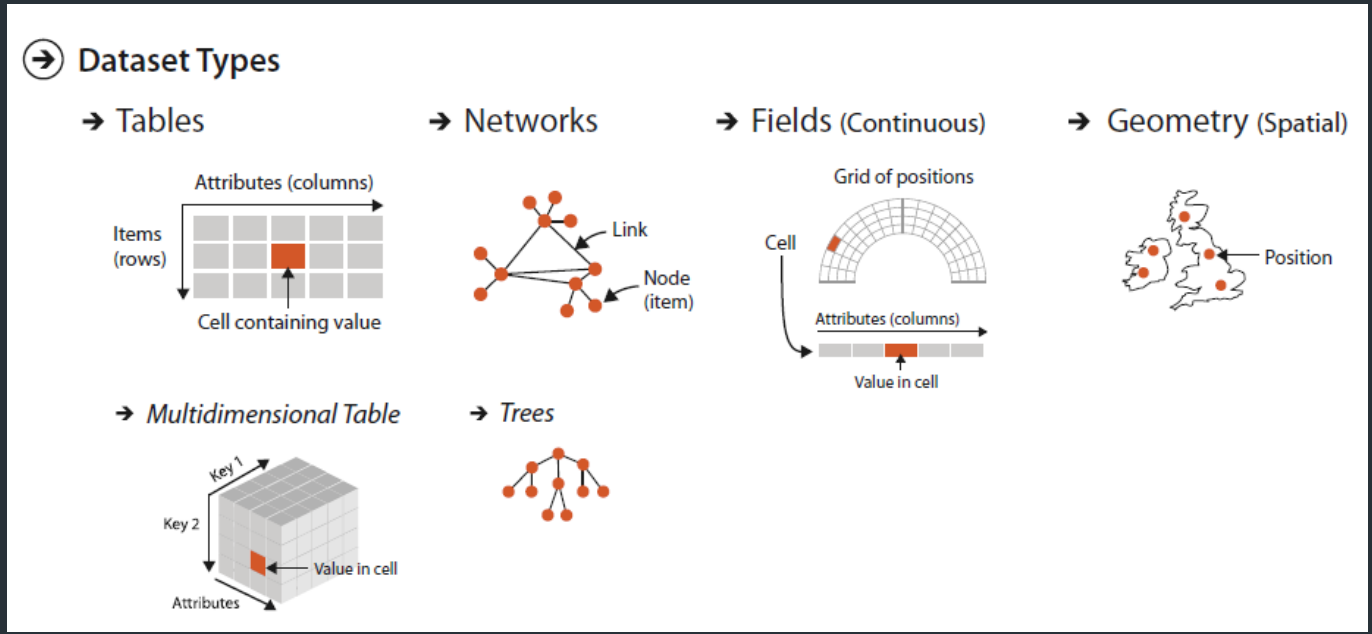
Data Abstraction

- Dataset Types
  - A dataset is any collection of information that is the target of analysis.
  - The four basic dataset types are tables, networks, fields, and geometry.
  - Other ways to group items together include clusters, sets, and lists.

Source: Tamara Munzer, Visualization Analysis and Design

Data Abstraction

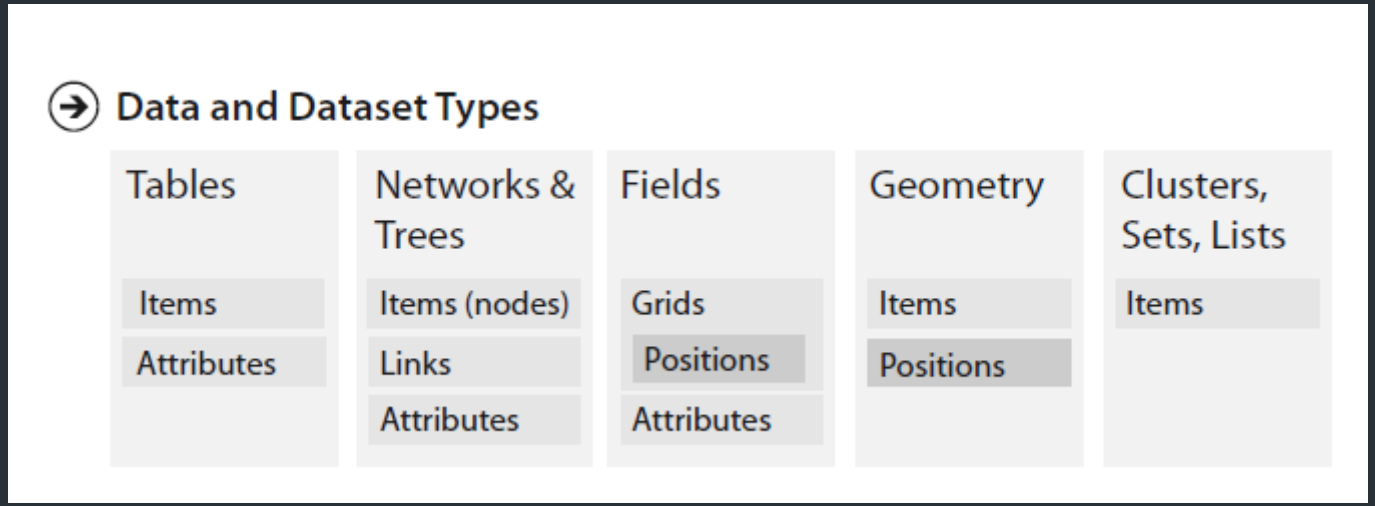
- Dataset Types



Source: Tamara Munzer, Visualization Analysis and Design

Data Abstraction

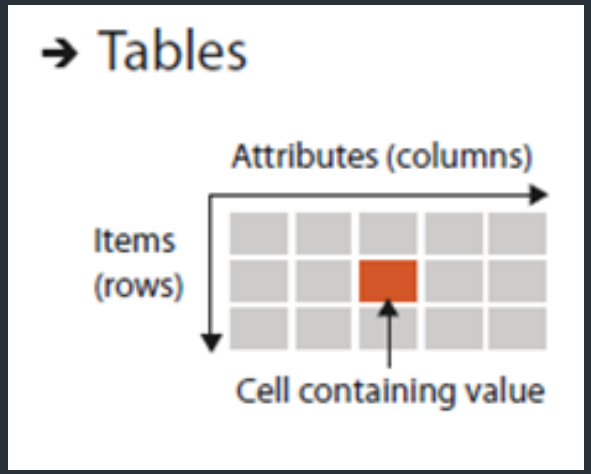
- Data and Dataset Types



Source: Tamara Munzer, Visualization Analysis and Design

Data Abstraction

- Tables
  - Rows and Columns
  - For a simple **flat table:**
    - **Each row** represents an **item of data**
    - **Each column is an attribute of** the dataset

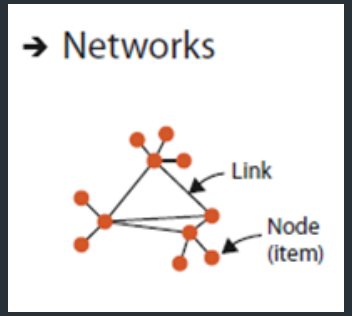


Name	Gender	Age	Salary	Designation
Ram	Male	50	50000	Manager
Sam	Male	40	40000	Clerk
Kumar	Male	30	25000	Clerk

Source: Tamara Munzer, Visualization Analysis and Design

Data Abstraction

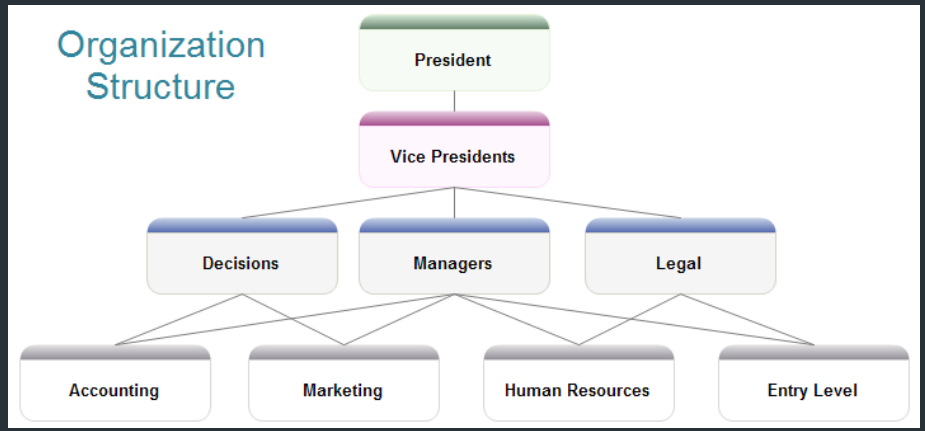
- Networks and Trees
  - well suited for specifying that there is some kind of relationship between two or more items
  - Items - > Nodes
  - Relationship between items - > link
  - Ex: Social Network
    - **nodes** are people
    - **links** mean friendship



Source: Tamara Munzer, Visualization Analysis and Design

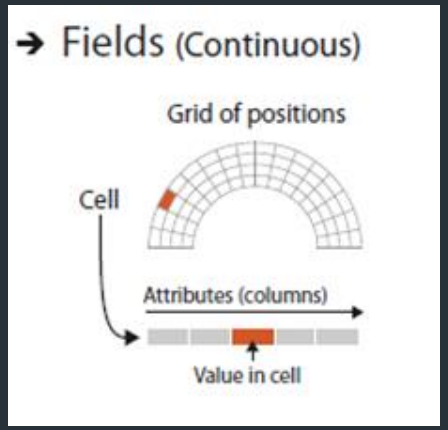
Data Abstraction

- Networks and Trees
  - Networks with hierarchical structure are more specifically called **trees. In contrast to a general network, trees do not have cycles**
  - Ex: organizational chart of a company



Source: Tamara Munzer, Visualization Analysis and Design

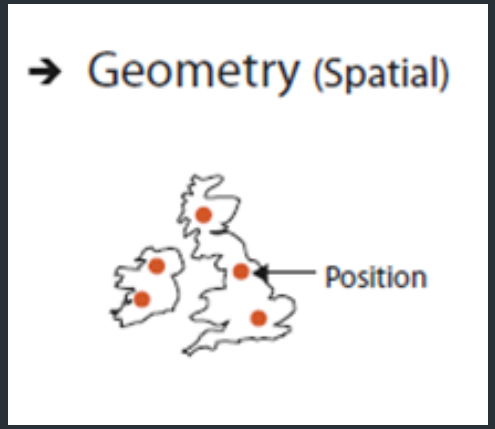
Data Abstraction



- Fields
  - contains measurements or calculations from a continuous domain
  - Spatial Field
  - Grid Types
  - Ex:
    - field dataset representing a medical scan of a human body
    - ECG Report
    - Temperature, Pressure, Speed, Force, Density

Source: Tamara Munzer, Visualization Analysis and Design

Data Abstraction

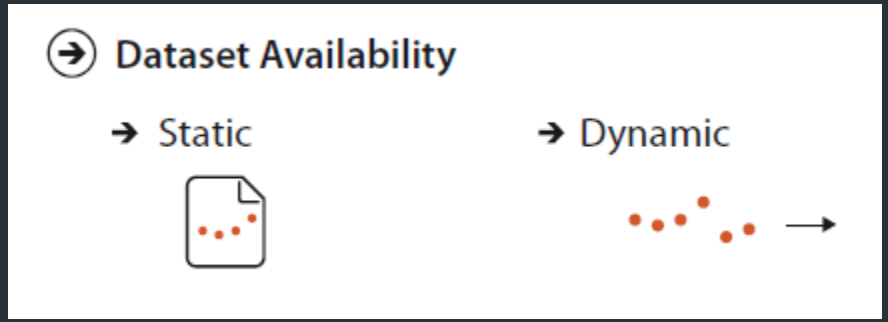


- Geometry
  - Shape of items with explicit spatial positions
  - The items could be points, or one-dimensional lines or curves, or 2D surfaces or regions, or 3D volume
  - Ex:
    - field dataset representing a medical scan of a human body
    - ECG Report
    - Temperature, Pressure, Speed, Force, Density

Source: Tamara Munzer, Visualization Analysis and Design

Data Abstraction

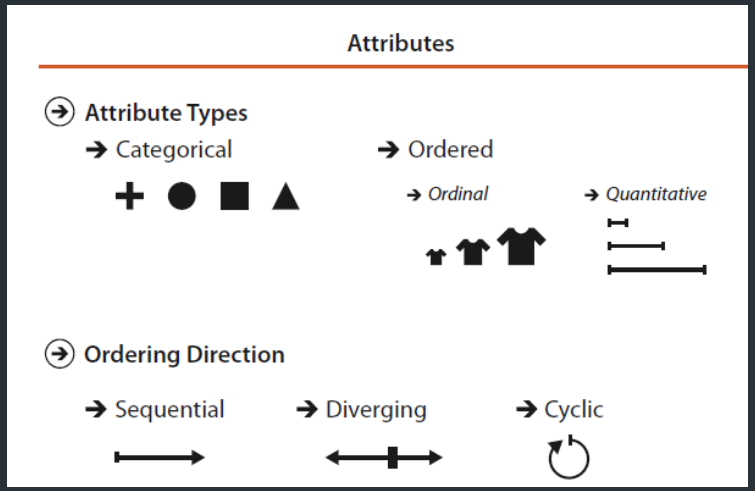
- Data Availability
  - *static* or *dynamic*
  - *Dynamic - Ex*: change or deletion of data



Source: Tamara Munzer, Visualization Analysis and Design

Data Abstraction

- Attribute Types
  - *Categorical – fruit, name, city names, etc.,*
  - *Ordered - Ordinal and Quantitative*
  - *Ordinal*
    - *Shirt size*
  - *Quantitative*
    - *Height, weight, price, etc.,*



Source: Tamara Munzer, Visualization Analysis and Design

Data Abstraction

- Attribute Types with Operations

Attribute Types (N,O,Q)

N - Nominal (same as Categorical)

- Fruits: Apples, oranges, ...

O - Ordered

- Quality of meat: Grade A, AA, AAA

Q - Interval (Location of zero arbitrary)

- Dates: Jan, 19, 2006; Location: (LAT 33.98, LONG -118.45)
- Only differences (i.e. intervals) may be compared

Q - Ratio (zero fixed)

- Physical measurement: Length, Mass, Temp, ...
- Counts and amounts

Attribute Types (N,O,Q)

N - Nominal (labels, categories)

- Operations: =, ≠

O - Ordinal (ordered)

- Operations: =, ≠, >, <

Q - Interval (location of zero arbitrary)

- Operations: =, ≠, >, <, +, −
- Can measure distances or spans

Q - Ratio (zero fixed)

- Operations: =, ≠, >, <, +, −, ×, ÷
- Can measure ratios or proportions

Source: Tamara Munzer, Visualization Analysis and Design

Data Abstraction

Key

EmpID	Name	Age	Salary	Designation
1	Ram	50	50000	Manager
2	Sam	40	40000	Clerk
3	Kumar	30	25000	Clerk

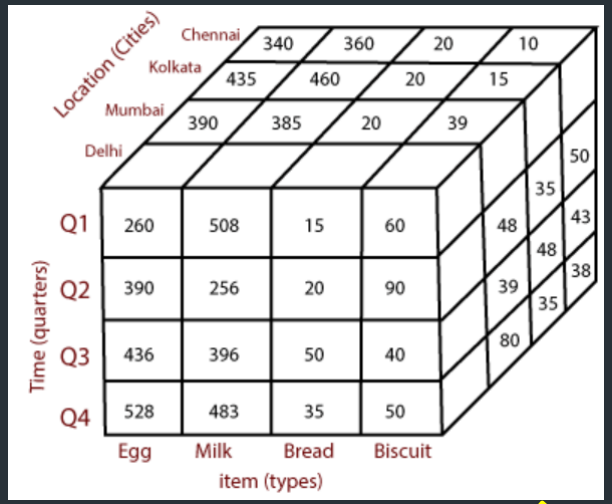
A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	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	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	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

- Key vs Value Semantics
  - Key
    - act as an index to look up the value
  - Flat Tables
  - Multidimensional Tables
  - Fields - Scalar, Vector, Tensor
- Temporal Semantics
  - Time Varying Data

Source: Tamara Munzer, Visualization Analysis and Design

Data Abstraction

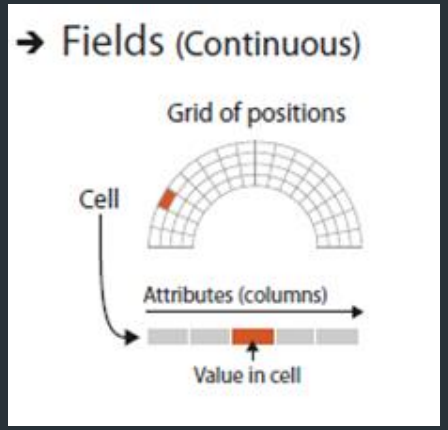
- Multidimensional Tables



	Location="Chennai"				Location="Kolkata"				Location="Mumbai"				Location="Delhi"			
	item				item				item				item			
Time	Egg	Milk	Bread	Biscuit	Egg	Milk	Bread	Biscuit	Egg	Milk	Bread	Biscuit	Egg	Milk	Bread	Biscuit
Q1	340	360	20	10	435	460	20	15	390	385	20	39	260	508	15	60
Q2	490	490	16	50	389	385	45	35	463	366	25	48	390	256	20	90
Q3	680	583	46	43	684	490	39	48	568	594	36	39	436	396	50	40
Q4	535	694	39	38	335	365	83	35	338	484	48	80	528	483	35	50

Source: Tamara Munzer, Visualization Analysis and Design , Javatpoint

Data Abstraction



- Fields
  - Represents continuous data
  - Scalar field
    - One attribute per cell
  - Vector field
    - Two or more attributes per cell
  - Tensor field
    - many attributes per cell

Source: Tamara Munzer, Visualization Analysis and Design

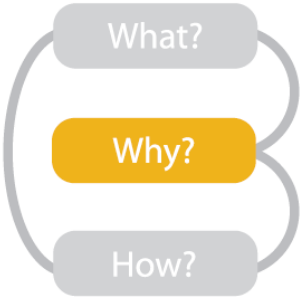
Data Abstraction

- Temporal Semantics
  - simply any kind of information that relates to time
  - Time Varying Data

City	Date	Temperature (celsius)	Humidity %	Wind (km/hr)
Chennai	05.08.2021	31	62	10
Chennai	06.08.2021	27	83	7
Chennai	07.08.2021	27	83	13

Source: Tamara Munzer, Visualization Analysis and Design

Task Abstraction



- {action, target} pairs
  - *discover distribution*
  - *compare trends*
  - *locate outliers*
  - *browse topology*

Why?

Actions

Analyze

→ Consume

→ *Discover* → *Present* → *Enjoy*

→ Produce

→ *Annotate* → *Record* → *Derive*

Search

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

Query

→ Identify → Compare → Summarize

Targets

All Data

→ Trends → Outliers → Features

Attributes

→ One → Many

→ *Distribution* → *Dependency* → *Correlation* → *Similarity*

→ *Extremes*

Network Data

→ Topology

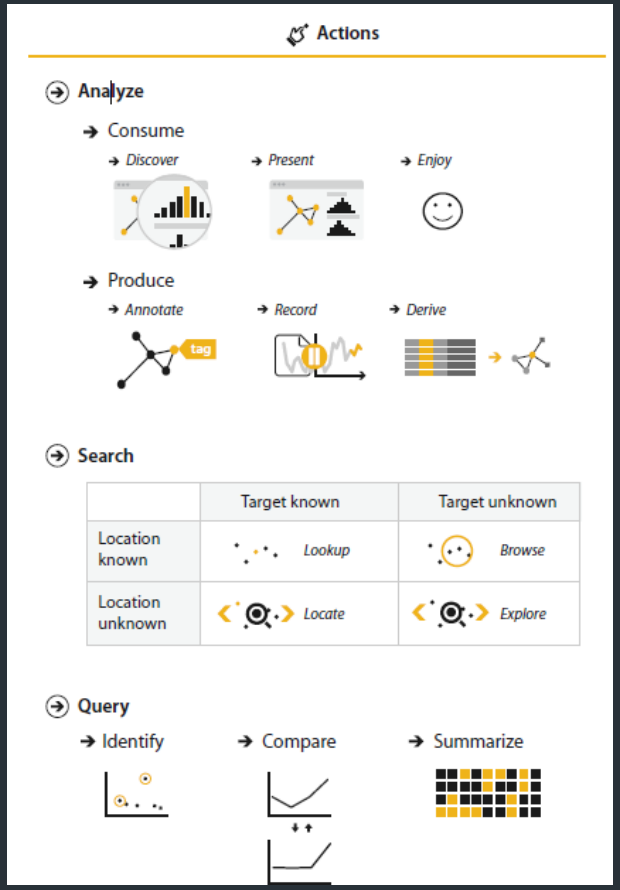
→ *Paths*

Spatial Data

→ Shape

Source: Tamara Munzer, Visualization Analysis and Design

Task Abstraction



- Actions
  - Analyze
    - high-level choices describe how the visualization(vis) is used to *analyze*
    - *Existing data or additional data*
  - Search
    - mid-level choices cover what kind of *search* is involved
    - target and location are known or not
  - Query
    - low-level choices pertain to the kind of *query*
    - to identify one target, compare some targets, or summarize all of the targets
- defines the user goals

Source: Tamara Munzer, Visualization Analysis and Design

Task Abstraction

- Analyze
  - Consume
    - Discover
      - goal - finding completely new things
      - **generate** a new **hypothesis**
      - **verify**—or disconfirm—an existing hypothesis
    - Present
      - Presentation – graphics, diagram, image
      - Size of the audience
      - Live, prerecorded

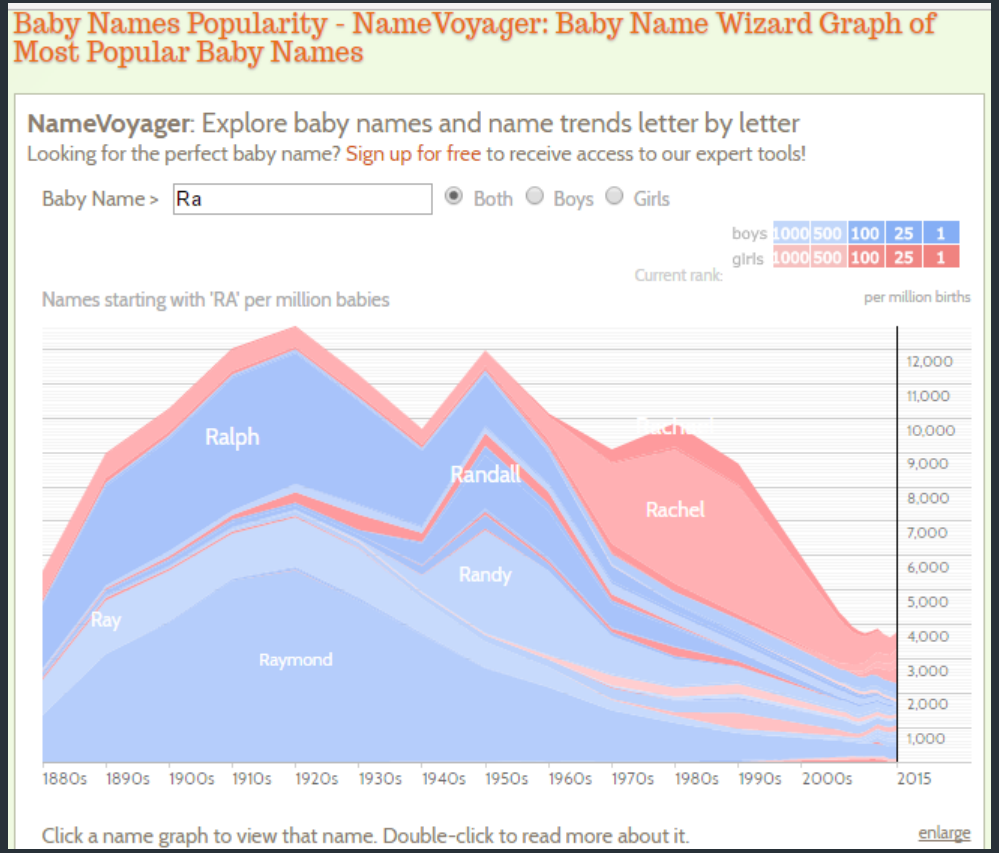
Source: Tamara Munzer, Visualization Analysis and Design

Task Abstraction

- Analyze
  - Enjoy
    - Infographic
    - User goal does not match with designer goal
    - Ex: Baby Name Voyager – Popular name in US since 1900

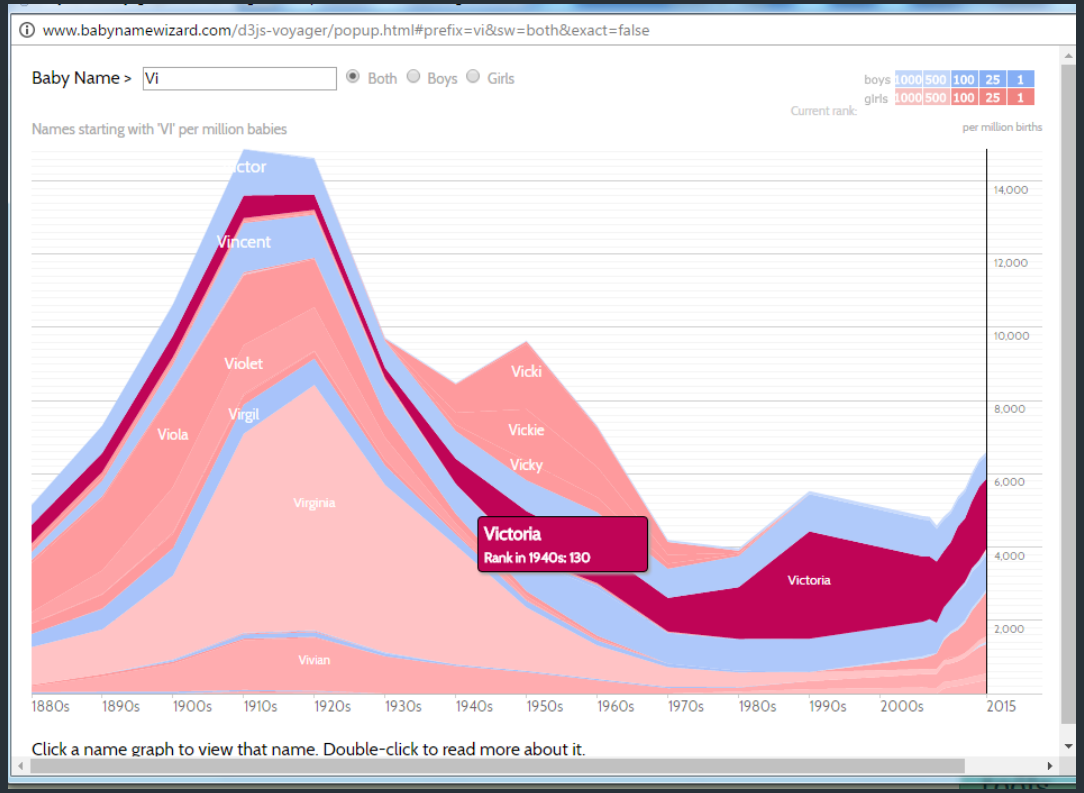
Source: Tamara Munzer, Visualization Analysis and Design

Task Abstraction



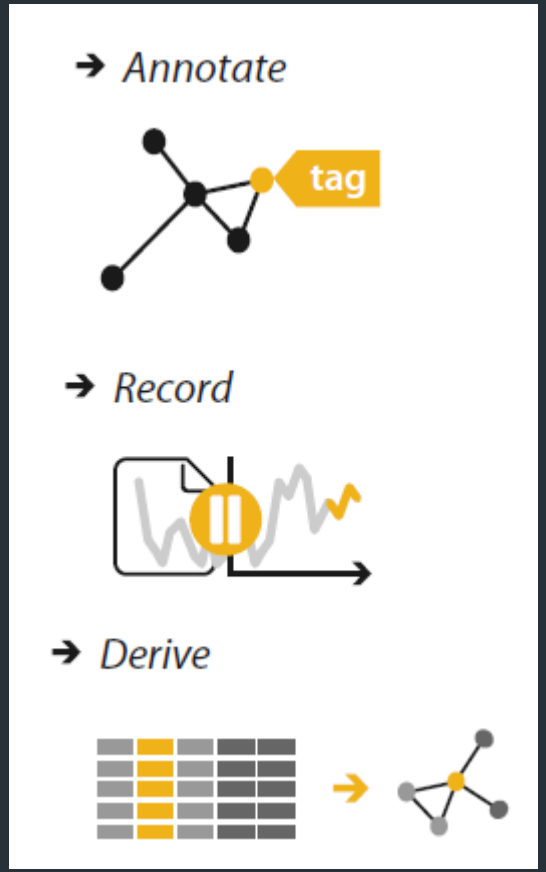
Source: Tamara Munzer, Visualization Analysis and Design

Task Abstraction



Source: Tamara Munzer, Visualization Analysis and Design

Task Abstraction

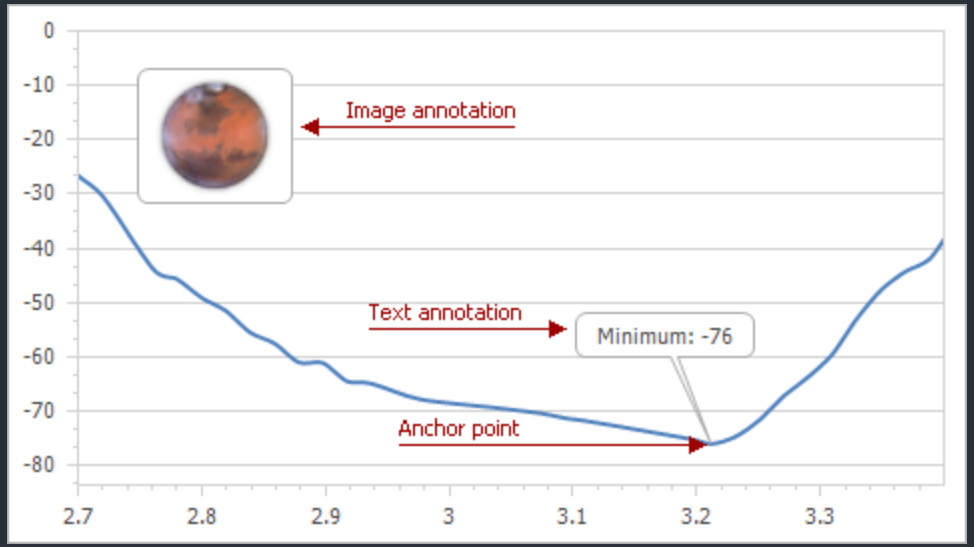


- Analyze
  - Produce
  - produce output to a next instance
  - **Types of Produce**
    - Annotate - Graphical or Textual annotation
    - Record
      - To assemble a graphical history
      - Artifacts- screen shots, book marked elements, parameter settings, interaction logs or annotation
    - Derive
      - produce new data elements based on existing data elements
      - Temperature - Weather, Shower, Bread Toast

Source: Tamara Munzer, Visualization Analysis and Design

Task Abstraction

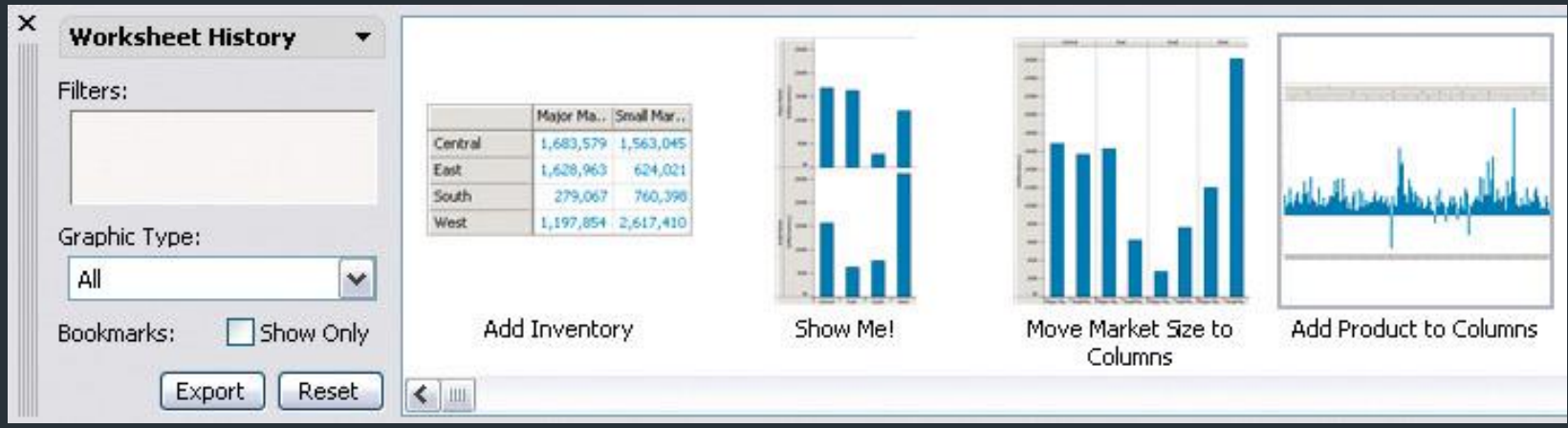
- Analyze
  - Annotate



Source: Tamara Munzer, Visualization Analysis and Design,
Image Source: DevExpress

Task Abstraction

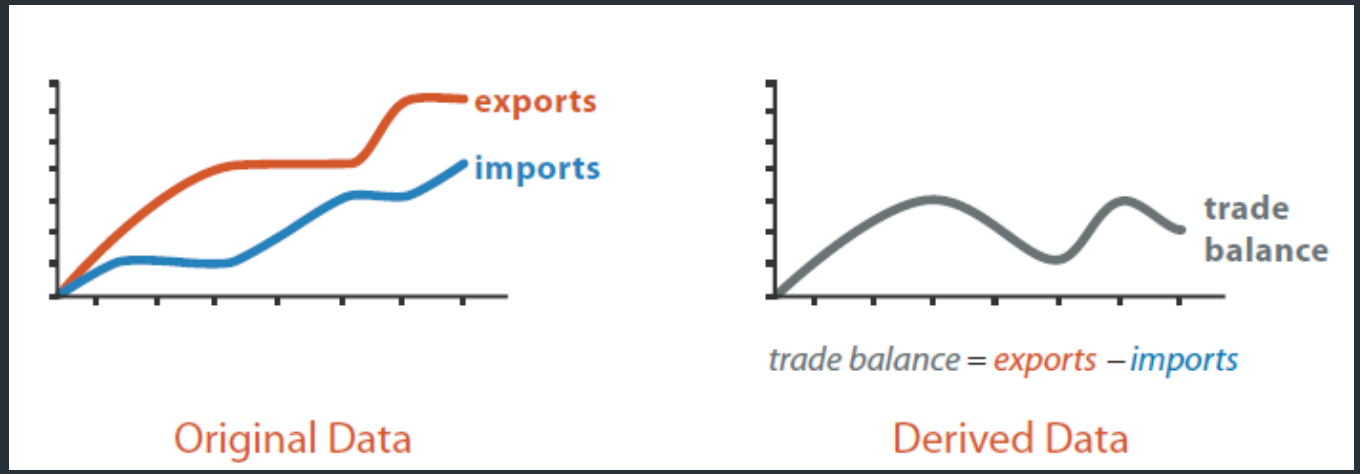
- Analyze
  - Record



Source: Tamara Munzer, Visualization Analysis and Design,
Image Source: DevExpress

Task Abstraction

- Analyze
  - Derieve



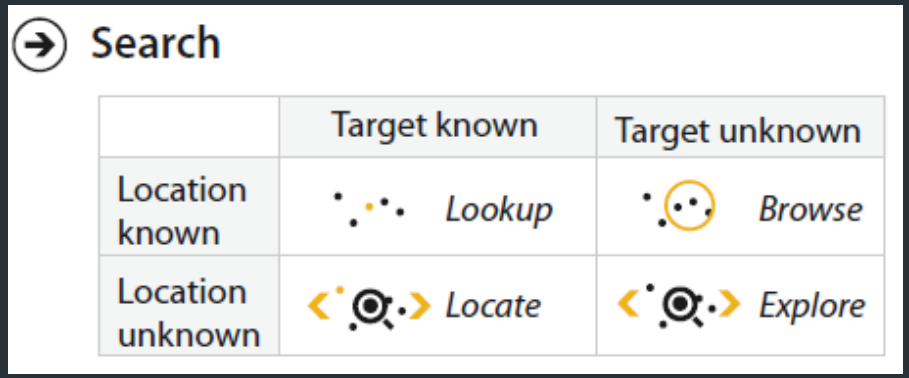
Source: Tamara Munzer, Visualization Analysis and Design,
Image Source: DevExpress

Task Abstraction

- Analyze
  - Search
    - Lookup, Locate, Browse, Explore

Target Location

What?	Where?	
Yes	Yes	Lookup
Yes	No	Locate
No	Yes	Browse
No	No	Explore



Source: Tamara Munzer, Visualization Analysis and Design

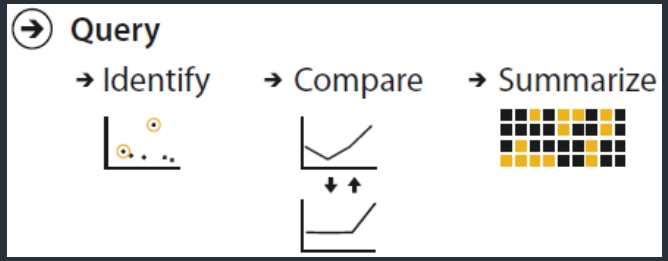
Task Abstraction

- Analyze
  - Search
    - Lookup
      - If users already know both what they're looking for and where it is, then the search type is simply **lookup**
    - Locate
      - To find a known target at an unknown location, the search type is **locate: that is, find out where the specific object is**
    - Browse
      - When users don't know exactly what they're looking for, but they do have a location in mind of where to look for it, the search type is **browse**
    - Explore
      - When users are not even sure of the location, the search type is **explore**

Source: Tamara Munzer, Visualization Analysis and Design

Task Abstraction

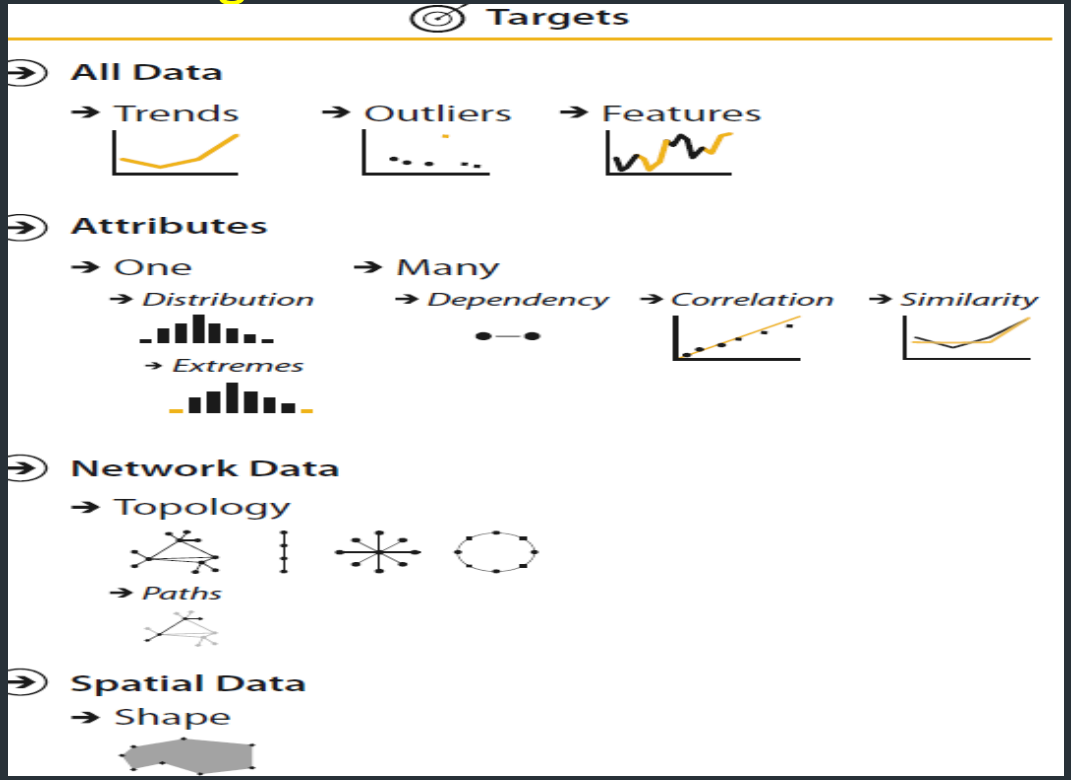
- Analyze
  - Query
    - **Identify** refers to a single target
      - Election Map
    - **Compare** refers to multiple targets, and
    - **Summarize** refers to the full set of possible targets.



Source: Tamara Munzer, Visualization Analysis and Design

Task Abstraction

- Targets

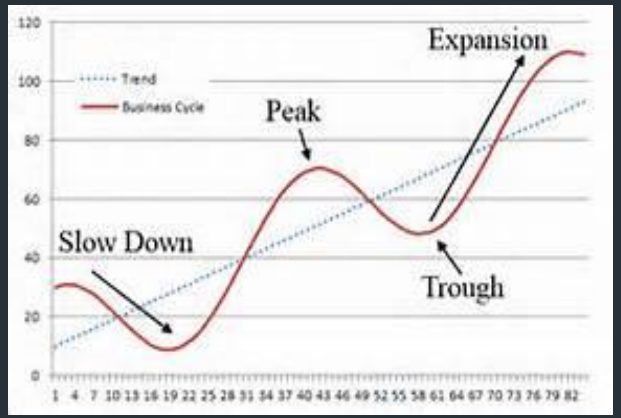


Source: Tamara Munzer, Visualization Analysis and Design

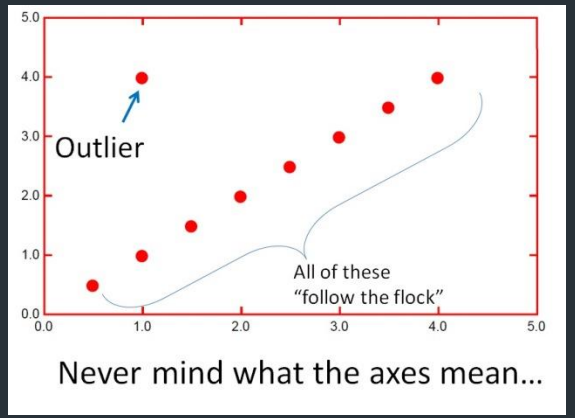
Task Abstraction

- Targets

Trends



Outlier

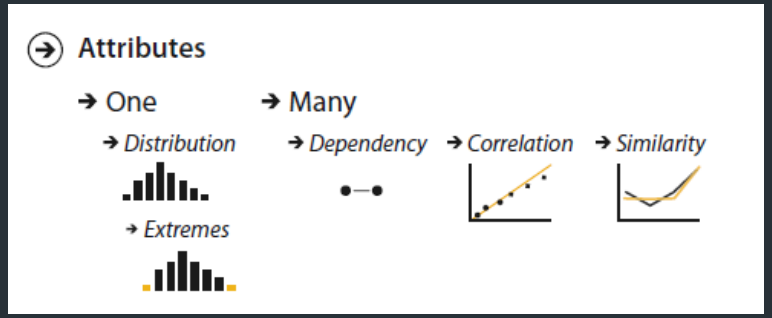


Features

Source: Tamara Munzer, Visualization Analysis and Design

Task Abstraction

- Targets
  - Attributes



Attributes are specific properties that are visually encoded.

High-level scope is the **distribution** of all values for an attribute.

The lowest-level target for an attribute is to find an individual value.

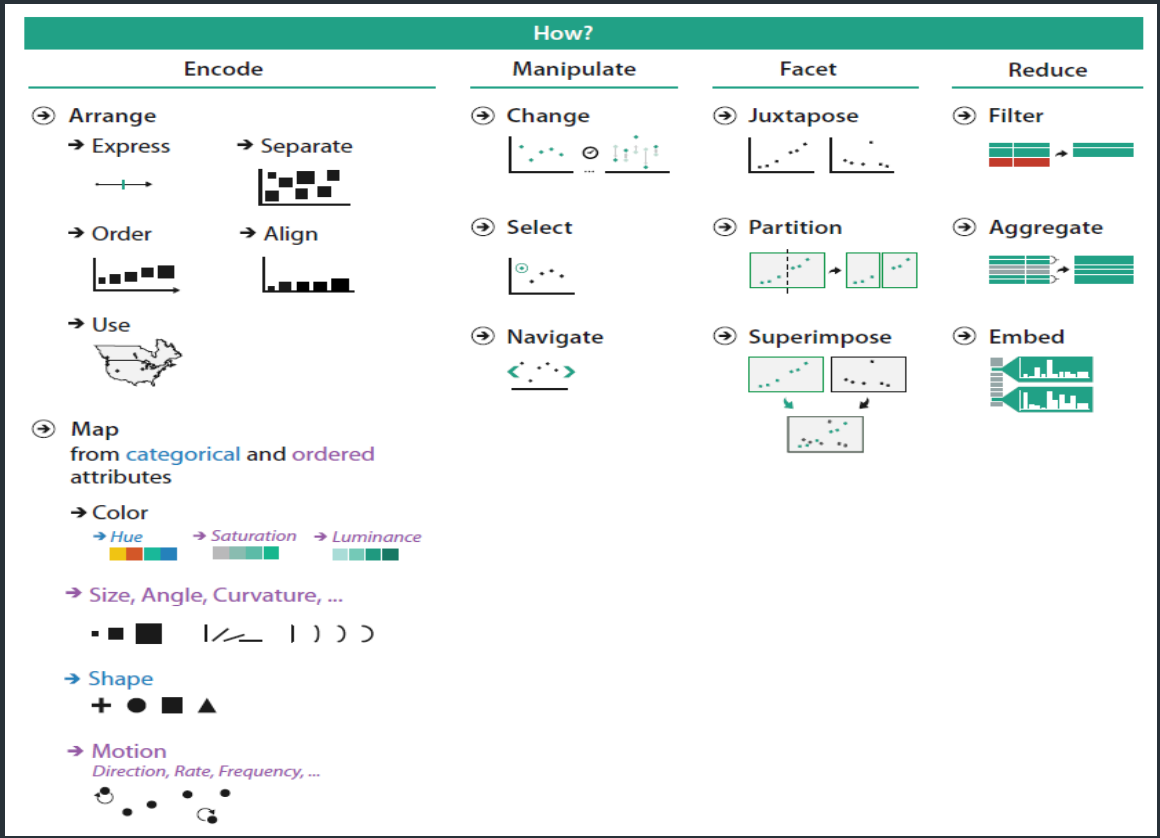
A first attribute can have a **dependency on a second, where the values for** the first directly depend on those of the second.

There is a **correlation** between one attribute and another if there is a tendency for the values of second to be tied to those of the first.

The **similarity between** two attributes can be defined as a quantitative measurement calculated on all of their values, allowing attributes to be ranked with respect to how similar, or different, they are from each other.

Source: Tamara Munzer, Visualization Analysis and Design

Task Abstraction



Source: Tamara Munzer, Visualization Analysis and Design

Visual Encoding

- How? Visual Encoding
  - **The way in which data is mapped into visual structures**
  - **Defining your goal**
  - **Examples of goals for visualizations**
    - to monitor systems,
    - find bargains,
    - compare company performances,
    - select suitable solutions,
    - track populations,
    - tell stories,
    - find specific data points,
    - find outliers,
    - show trends or simply give an overview of the data.

Source: Tamara Munzer, Visualization Analysis and Design

Visual Encoding

- Visual Encoding
  - **To begin defining the goal of your visualization, ask yourself**
    - What values or data dimensions are relevant in this context?
    - Which of these dimensions matter
      - **Most or Least**?
    - What are the key relationships that need to be communicated?
    - What properties or values would make some individual data points more interesting than the rest?

Source: Tamara Munzer, Visualization Analysis and Design

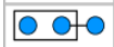
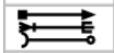
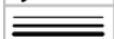
Choose Appropriate Visual Encoding

- Visual Property
  - ***Naturally ordered or Distinct values***
- **Natural Ordering**
  - For example, position has a natural ordering; shape doesn't.
  - Length has a natural ordering; texture doesn't (but pattern density does).
  - Line thickness or weight has a natural ordering; ; line style (solid, dotted, dashed) doesn't.
- **Distinct Values**
  - The second main factor to consider when choosing a visual property is how many *distinct values* it has that your reader will be able to perceive, differentiate, and possibly remember.

Source: Designing Data Visualisation, Noah Iliinsky

Choose Appropriate Visual Encoding

- Visual Property

Example	Encoding	Ordered	Useful values	Quantitative	Ordinal	Categorical	Relational
	position, placement	yes	infinite	Good	Good	Good	Good
1, 2, 3; A, B, C	text labels	optional alpha or num	infinite	Good	Good	Good	Good
	length	yes	many	Good	Good		
	size, area	yes	many	Good	Good		
	angle	yes	medium	Good	Good		
	pattern density	yes	few	Good	Good		
	weight, boldness	yes	few		Good		
	saturation, brightness	yes	few		Good		
	color	no	few (<20)			Good	
	shape, icon	no	medium			Good	
	pattern texture	no	medium			Good	
	enclosure, connection	no	infinite			Good	Good
	line pattern	no	few				Good
	line endings	no	few				Good
	line weight	yes	few		Good		

Source: Designing Data Visualisation, Noah Iliinsky

Choose Appropriate Visual Encoding

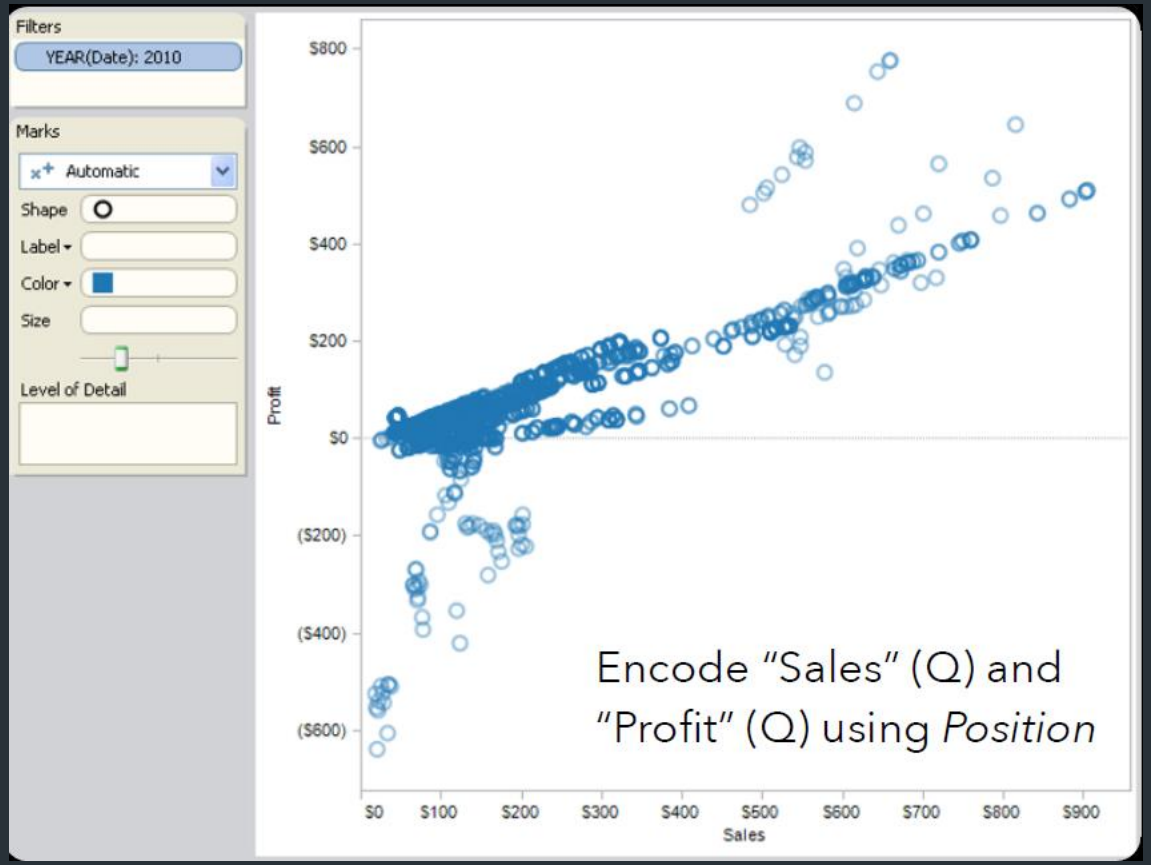
Example: Coffee Sales

Sales figures for a fictional coffee chain:

Sales	Q-Ratio
Profit	Q-Ratio
Marketing	Q-Ratio
Product Type	N {Coffee, Espresso, Herbal Tea, Tea}
Market	N {Central, East, South, West}

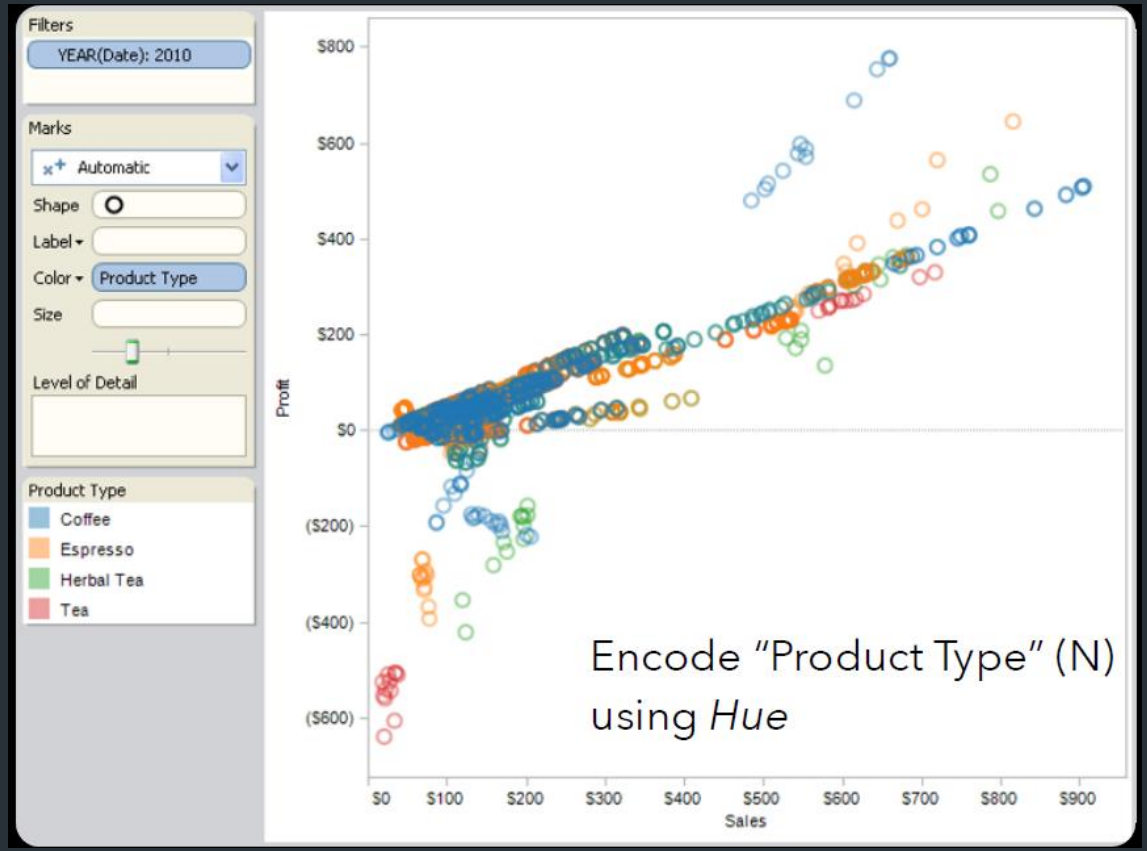
Source: Designing Data Visualisation, Noah Iliinsky

Choose Appropriate Visual Encoding



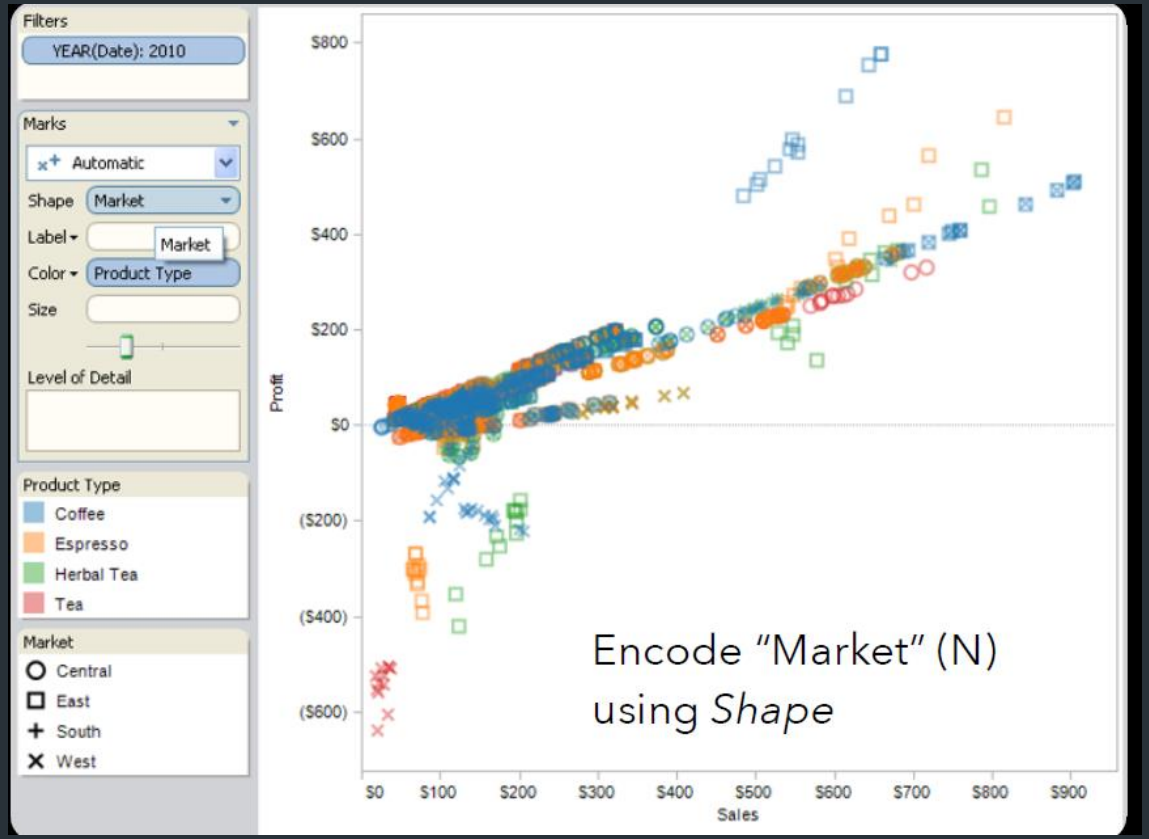
Source: Designing Data Visualisation, Noah Iliinsky

Choose Appropriate Visual Encoding



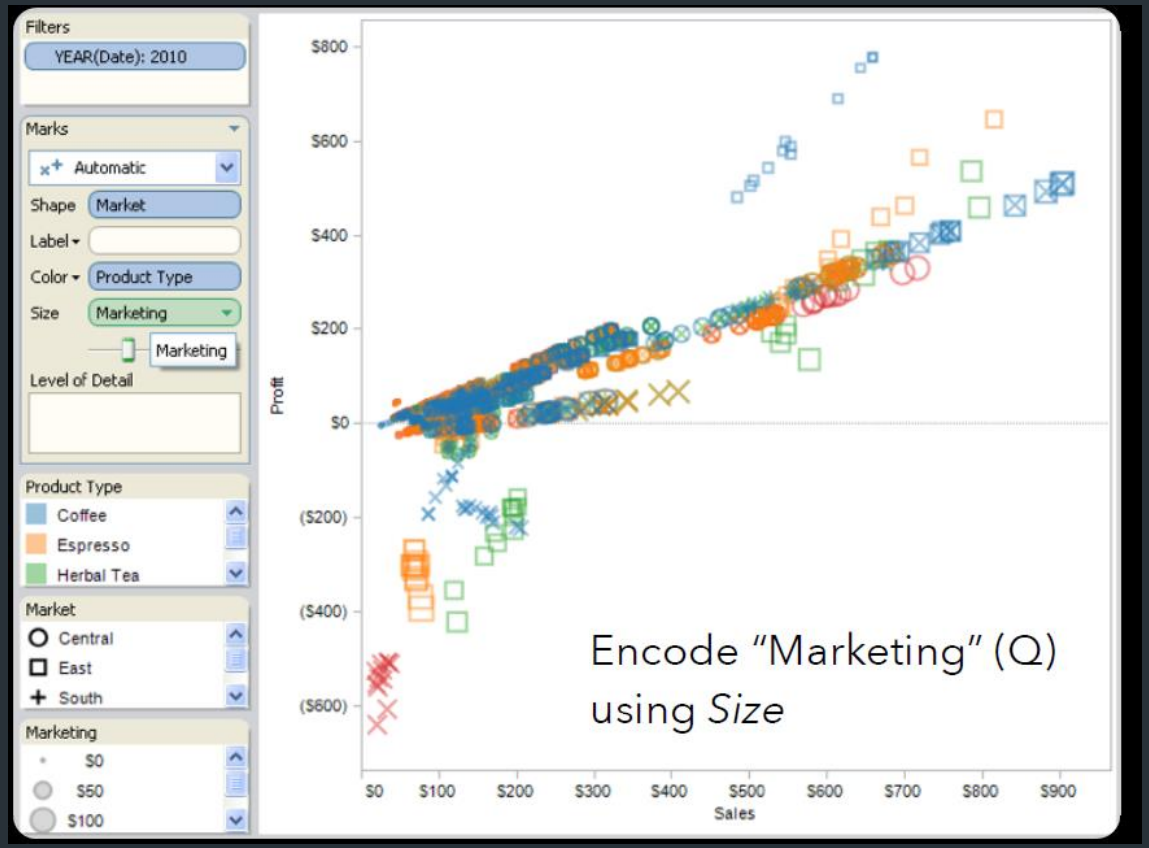
Source: Designing Data Visualisation, Noah Iliinsky

Choose Appropriate Visual Encoding



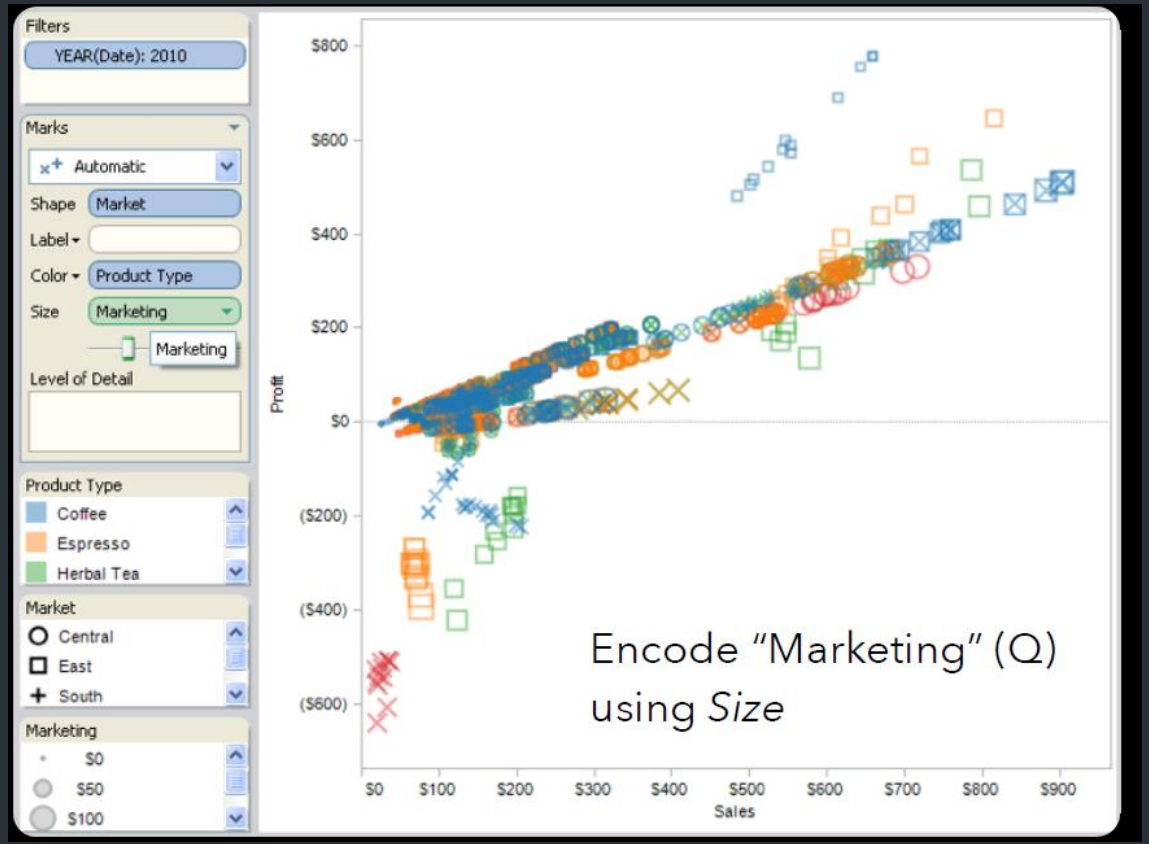
Source: Designing Data Visualisation, Noah Iliinsky

Choose Appropriate Visual Encoding



Source: Designing Data Visualisation, Noah Iliinsky

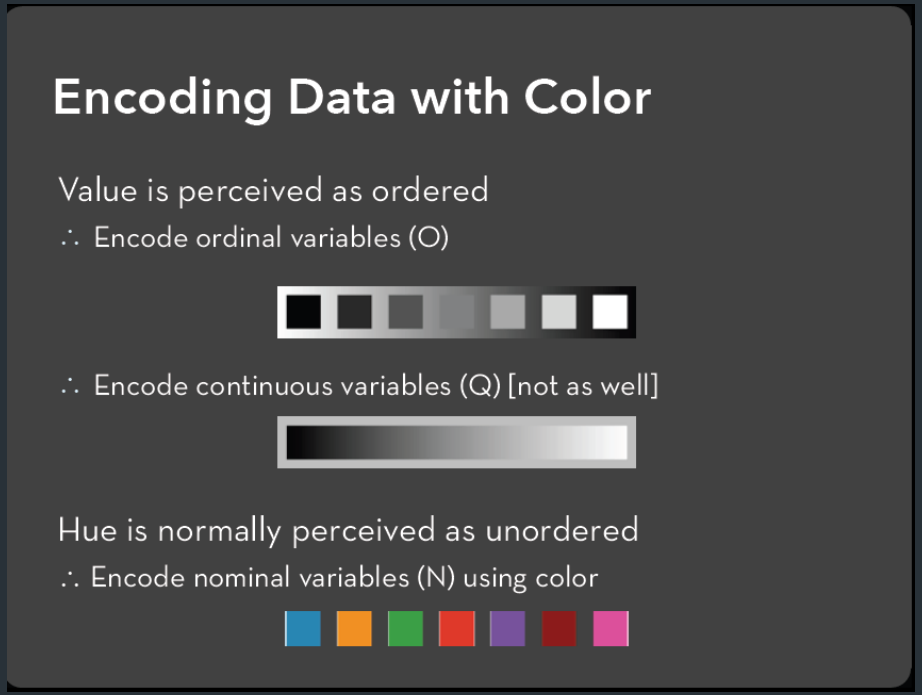
Choose Appropriate Visual Encoding



Source: Designing Data Visualisation, Noah Iliinsky

Choose Appropriate Visual Encoding

- Encoding Data with Color



Source: Designing Data Visualisation, Noah Iliinsky

Choose Appropriate Visual Encoding

- Encoding Data with Color

Quantitative Color Encoding

Sequential color scale

Constrain hue, vary luminance/saturation

Map higher values to darker colors

Diverging color scale

Useful when data has meaningful "midpoint"

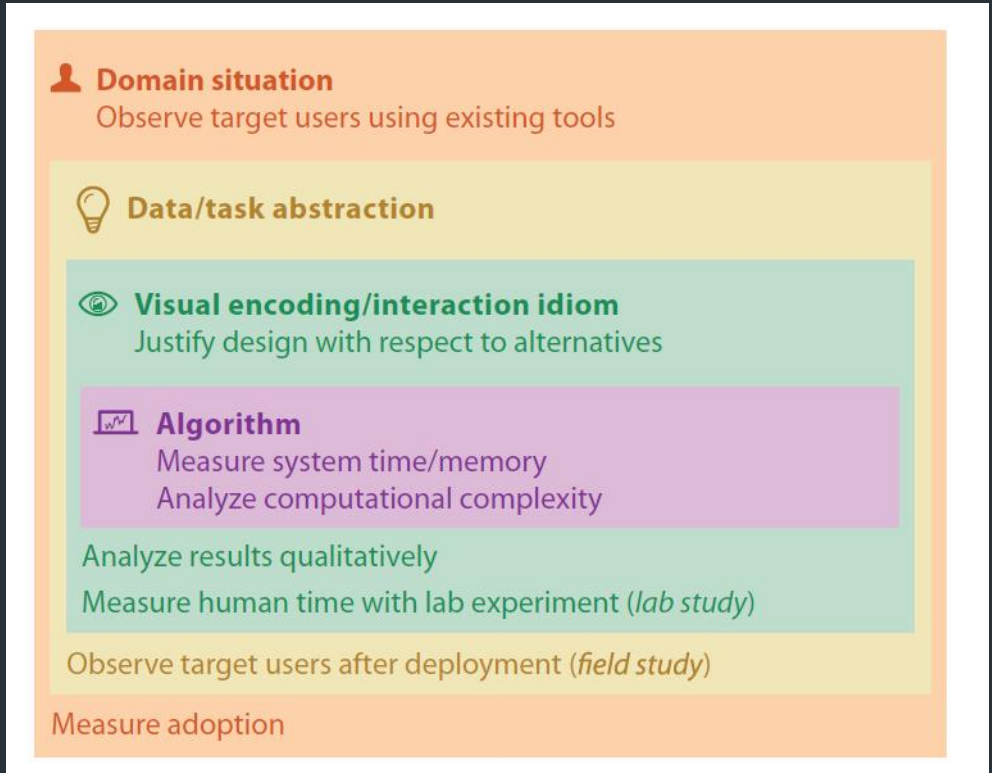
Use neutral color (e.g., grey) for midpoint

Use saturated colors for endpoints

Source: Designing Data Visualisation, Noah Iliinsky

Four Levels for Validation

- Four levels of Design



Source: Designing Data Visualisation, Noah Iliinsky

Four Levels for Validation

- Four levels of Design have different threads to validity at each level

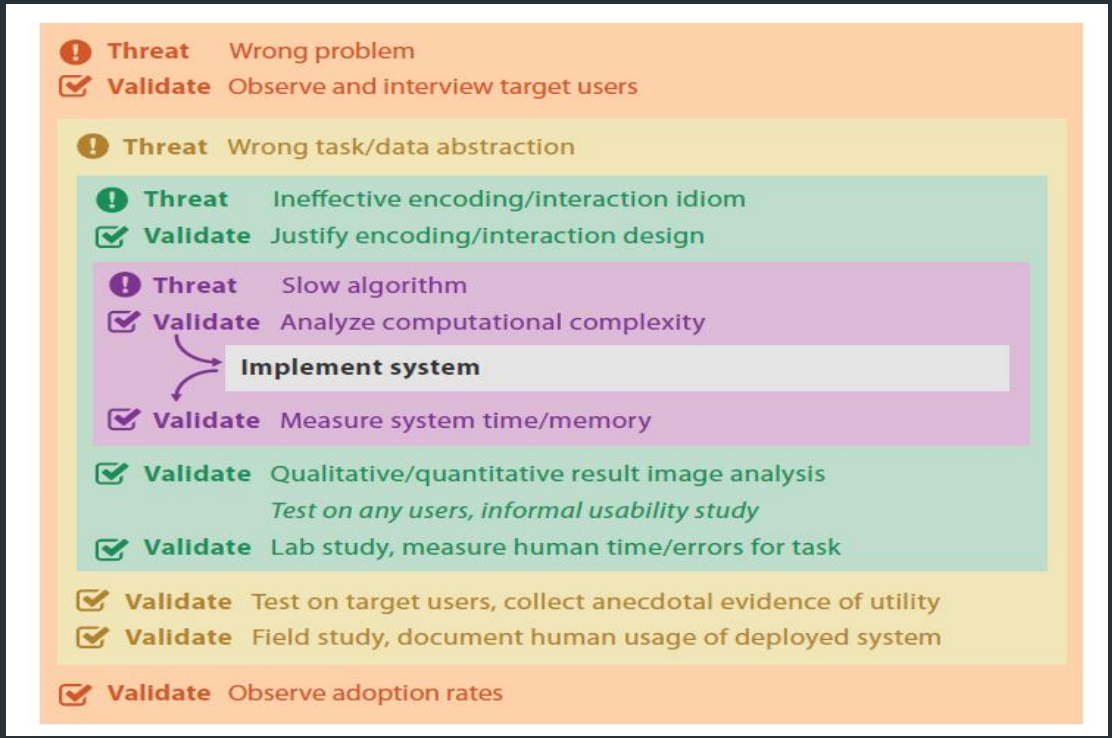
- Wrong problem: You misunderstood their needs.
- Wrong abstraction: You're showing them the wrong thing.
- Wrong idiom: The way you show it doesn't work.
- Wrong algorithm: Your code is too slow.

They – Target users
You – Designer

Source: Designing Data Visualisation, Noah Iliinsky

Four Levels for Validation

- Validation Approaches



Source: Designing Data Visualisation, Noah Iliinsky