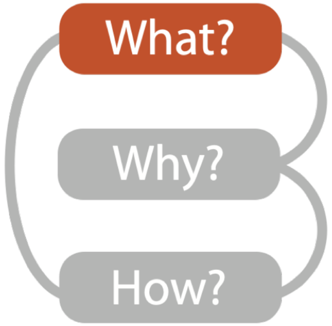
STA313 – Data Visualization | Lecture 2

What? Data Abstraction

Why? Task Abstraction

Prof. Fanny Chevalier

Acknowledgements:
The materials in this course is heavily based on the book and lecture materials from Tamara Munzner, as well as teaching materials and resources from John Stasko, Sheelagh Carpendale, Jeffrey Heer, Alex Lex, Miriah Meyer, Benjamin Bach, Charles Perin, Enrico Bertini, Evan Peck, Arvind Satyanarayan, Alberto Cairo, and Petra Isenberg.



Data **attribute** (*aka* data dimension, column)

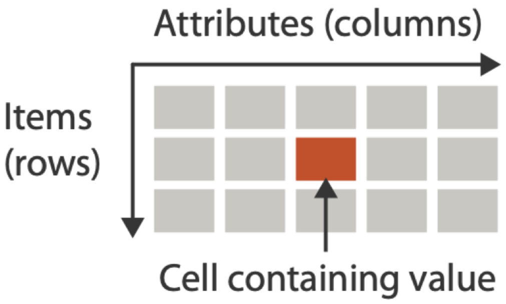
	A	B	C	D	E	F
1	Customer	City	Region	Product	Quantity	Total Sales
2	Orange	Big Town	West	Milk Chocolate	125	225
3	Red	Big Town	West	Dark Chocolate	210	441
4	Pink	Medium Town	East	Milk Chocolate	145	261
5	Grey	Big Town	West	Chocolate Hazelnut	21	63
6	Blue	Small Town	South	Dark Chocolate	50	105
7	Dark	Big Town	West	Chocolate Hazelnut	65	195
8	White	Big Town	West	Milk Chocolate	40	72
9	Green	Village	[illegible]	Chocolate Hazelnut	122	366
10	Yellow	Medium Town	East	Dark Chocolate	60	126
11	Silver	Medium Town	East	Extra Dark Chocolate	30	75
12	Gold	Medium Town	East	Chocolate Hazelnut	56	168

Data **item** (*aka* observation, entry, row)

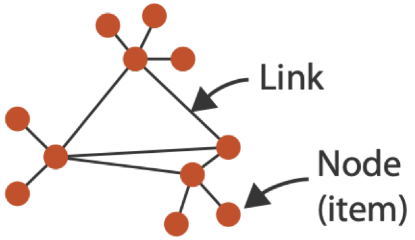
value (*aka* measurement, cell)

Three major datatypes

Tables

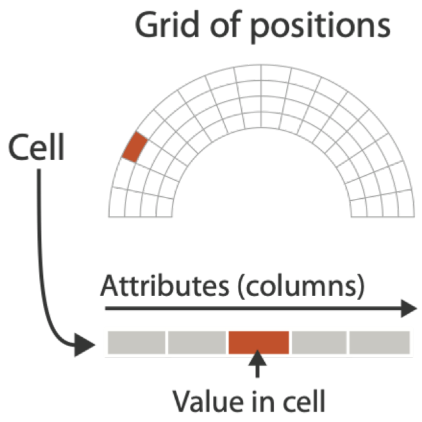


Networks



Spatial

Fields (continuous)



Geometry (spatial)

Attribute Types

Not Ordered

Categorical

Ordered

Ordinal

Quantitative

Ordering Direction

Sequential

Diverging

Circular

Attribute Types

Not Ordered

Categorical

$=, \neq$

Ordered

Ordinal

$=, \neq, <, >$

Quantitative

$=, \neq, <, >, +, -, \div$

Attribute Types

Not Ordered

Ordered

Categorical

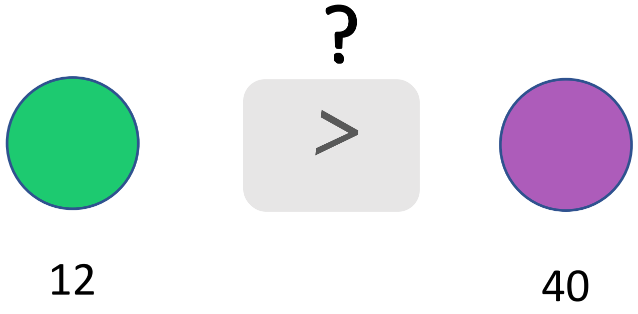
$= , \neq$

Ordinal

$= , \neq , <, >$

Quantitative

$= , \neq , <, >, +, -, \div$



Attribute Types

Not Ordered

Categorical

$= , \neq$

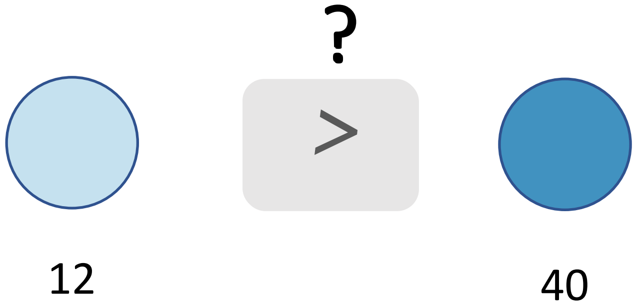
Ordered

Ordinal

$= , \neq , <, >$

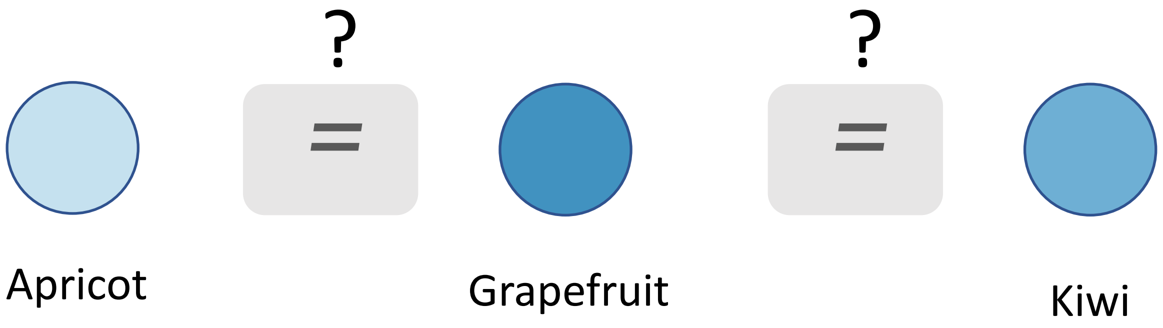
Quantitative

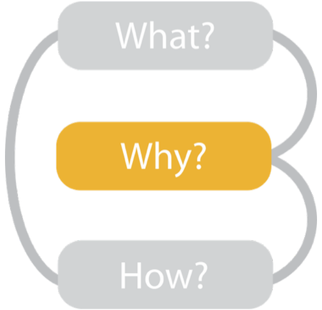
$= , \neq , <, >, +, -, \div$



Attribute Types

Not Ordered	Ordered	
Categorical	Ordinal	Quantitative
$=, \neq$	$=, \neq, <, >$	$=, \neq, <, >, +, -, \div$



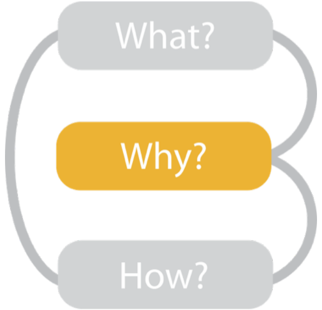


« Computer-based visualization systems provide ***visual representations*** *of* ***datasets*** *designed to* ***help people carry out tasks more effectively****. »*

– Munzner 2014

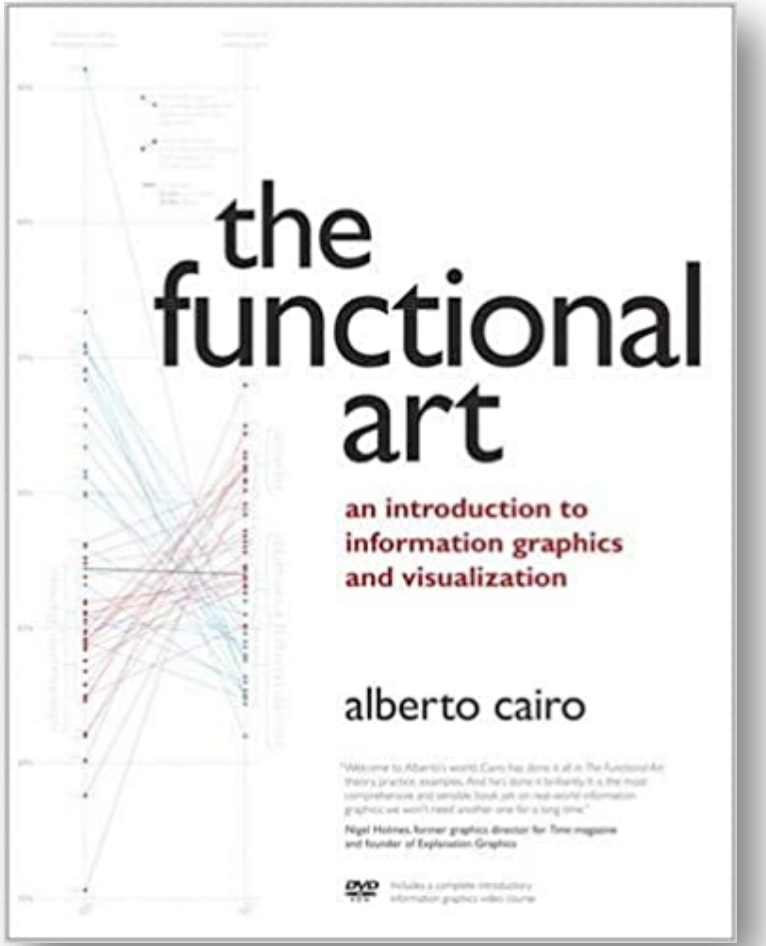
« The purpose of visualization is ***insight****, not pictures.»*

– Ben Shneiderman



« *Visualization should be seen as a* ***technology****. They are extenstions of oursleves […] They are* ***means to reach goals****.»*

– Alberto Cairo



Low-Level Components of Analytic Activity in Information Visualization

Robert Amar, James Eagan, and John Stasko

College of Computing/GVU Center
Georgia Institute of Technology

Abstract

Existing system-level taxonomies of visualization tasks are geared more towards the design of particular representations than the facilitation of user analytic activity. We present a set of ten low-level analysis tasks that largely capture people's activities while employing information visualization tools for understanding data. To help develop these tasks, we collected nearly 200 sample questions from students about how they would analyze five particular data sets from different domains. The questions, while not being totally comprehensive, illustrated the sheer variety of analytic questions typically posed by users when employing information visualization systems. We hope that the presented set of tasks is useful for information visualization system designers as a kind of common substrate to discuss the relative analytic capabilities of the systems. Further, the tasks may provide a form of checklist for system designers.

CR Categories and Subject Descriptors: H.5.0 [Information Interfaces and Presentation]: General; J.0 [Computer Applications]: General

Additional Keywords: Analytic activity, taxonomy, knowledge discovery, design, evaluation.

1 Introduction

Information visualization research, especially that dealing with the automatic generation of information presentations [10,15], has produced several taxonomies of system tasks that map visualization operations to user cognitive processes. In one sense, these taxonomies might be considered low-level task taxonomies or hierarchies, since they form part of a compositional language upon which automatic generation systems build higher-order externalizations of data.

However, considering these taxonomies as a basis upon which to build models of analytic activity is made difficult by their origins. While their elements can be algorithmically composed into presentations, the composition process itself is ad-hoc, relying on a designer's own insight and expressive capability within a particular tool. These taxonomies reflect this system-oriented approach, rather than providing ways to think about all

With the aim of generating an actionable means for supporting analytic activity, we wish to rethink some of the lower-level task taxonomies that focus on a generated presentation as an end result. In general, information visualization can benefit from understanding the tasks that users accomplish while doing actual analytic activity. Such understanding achieves two goals: first, it aids designers in creating novel presentations that amplify users' analytic abilities; second, it provides a common vocabulary for evaluating the abilities and affordances of information visualization systems with respect to user tasks.

We argue that a stronger focus on user tasks and analytic activities in information visualization is necessary as current tools do not seem to support analytic activity consistently. A 2004 study by Saraiya and North found that insights generated from tools used to visualize gene expression data were not generally valuable according to domain experts [11]. Systems such as IN-SPIRE [7] support analytic activities within the domain of document search but may not generalize across domains. Current tools may not even support representational activity very well; consider, for example, the Kobsa study showing only 68-75% accuracy on relatively simple tasks during commercial tool evaluation [8].

1.2 The Nature of Analytic Activity

User analysis questions and tasks as part of analytic activity typically range from broader, "high-level" goals to much more specific, "low-level" inquiries. For example, a person studying the history of motion picture films may have "high-level", uncertainty-tinged knowledge goals such as understanding trends in popularity over time or determining how to predict which movies will win Academy Awards. In the process of acquiring this knowledge, the person may generate more specific, low-level queries such as identifying the Academy Award-winning pictures of the past ten years and determining whether or not movie length correlates to the film's popularity.

It is this latter set of questions, more specific and focused in nature, on which we focus on in this article. In particular, we are interested in generating a relatively small set of question types that encompasses the set of user inquiries made while working with information visualization systems. While it seems unlikely

- Retrieve Value
- Filter
- Compute Derived Value
- Find Extremum
- Sort
- Determine Range
- Characterize Distribution
- Find Anomalies
- Cluster
- Correlate

A Task Taxonomy of Network Evolution Analysis

Jae-wook Ahn, Catherine Plaisant, and Ben Shneiderman

Abstract—Visualization is a useful tool for understanding the nature of networks. The recent growth of social media requires more powerful visualization techniques beyond static network diagrams. One of the most important challenges is the visualization of temporal network evolution. In order to provide strong temporal visualization methods, we need to understand what tasks users accomplish. This study provides a taxonomy of the temporal network visualization tasks. We identify (1) the entities, (2) the properties to be visualized, and (3) the hierarchy of temporal features, which were extracted by surveying existing temporal network visualization systems. By building and examining the task taxonomy, we report which tasks have been covered so far and suggest additions for designing the future visualizations. We also present example visualizations constructed using the task taxonomy for a social networking site in order to validate the quality of the taxonomy.

Index Terms—Network visualization, temporal evolution, task taxonomy, design space

1 Introduction

Network visualization is a crucial tool for understanding the nature of social networks. It can show the players of the networks and their relationships visually and can let the users explore and achieve important information for the tasks such as uncovering influential actors, finding helpful bridging people, and identifying destructive spammers. Existing social network analysis (SNA) software packages such as UCINET and Pajek support network visualization features. The recent growth of social media [5] requires more powerful social network analysis and visualization techniques beyond the conventional static network diagrams. One of the most important challenges is the visualization of temporal network evolution.

Time series visualization in general helps us to discover relations and patterns [3]; learn from the past to predict, plan, and build the future [1]. Therefore, various attempts have been made to provide effective tools for time series analysis. TimeSearcher [6, 15] provided an interactive pattern search in time series. Hochheiser and Shneiderman [18] introduced timeboxes which supported direct manipulation for specifying query constraints on time series data sets. Aris et al. [3] explored different strategies time intervals, specifically focusing on unevenly-spaced time series. Lifelines [33] and Lifeflow [35] provided methods to understand temporal categorical patterns. In social network analysis, the importance of temporal network analysis and longitudinal network models has been pointed out too [34]. However, this domain has been relatively less explored by social network researchers.

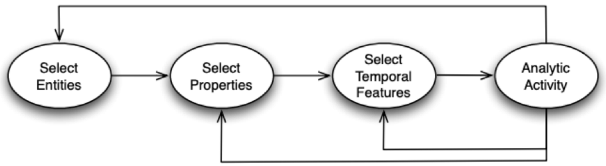
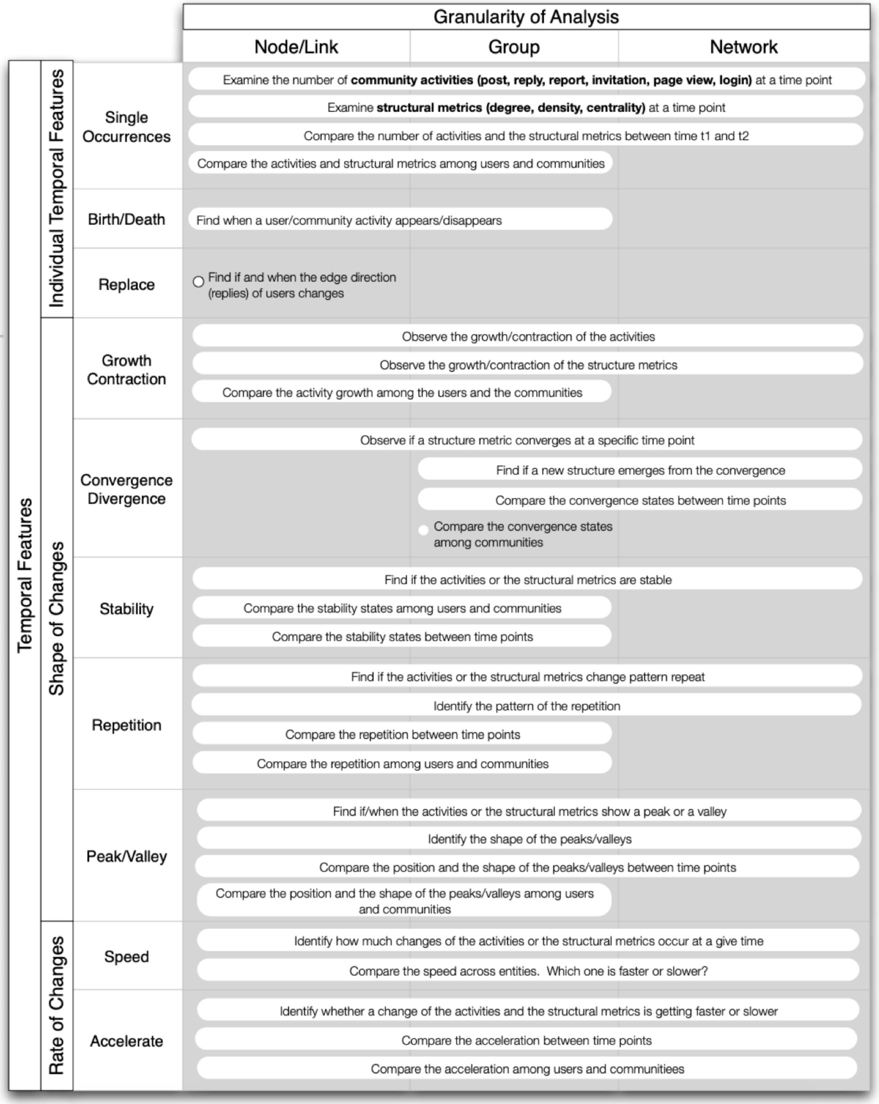
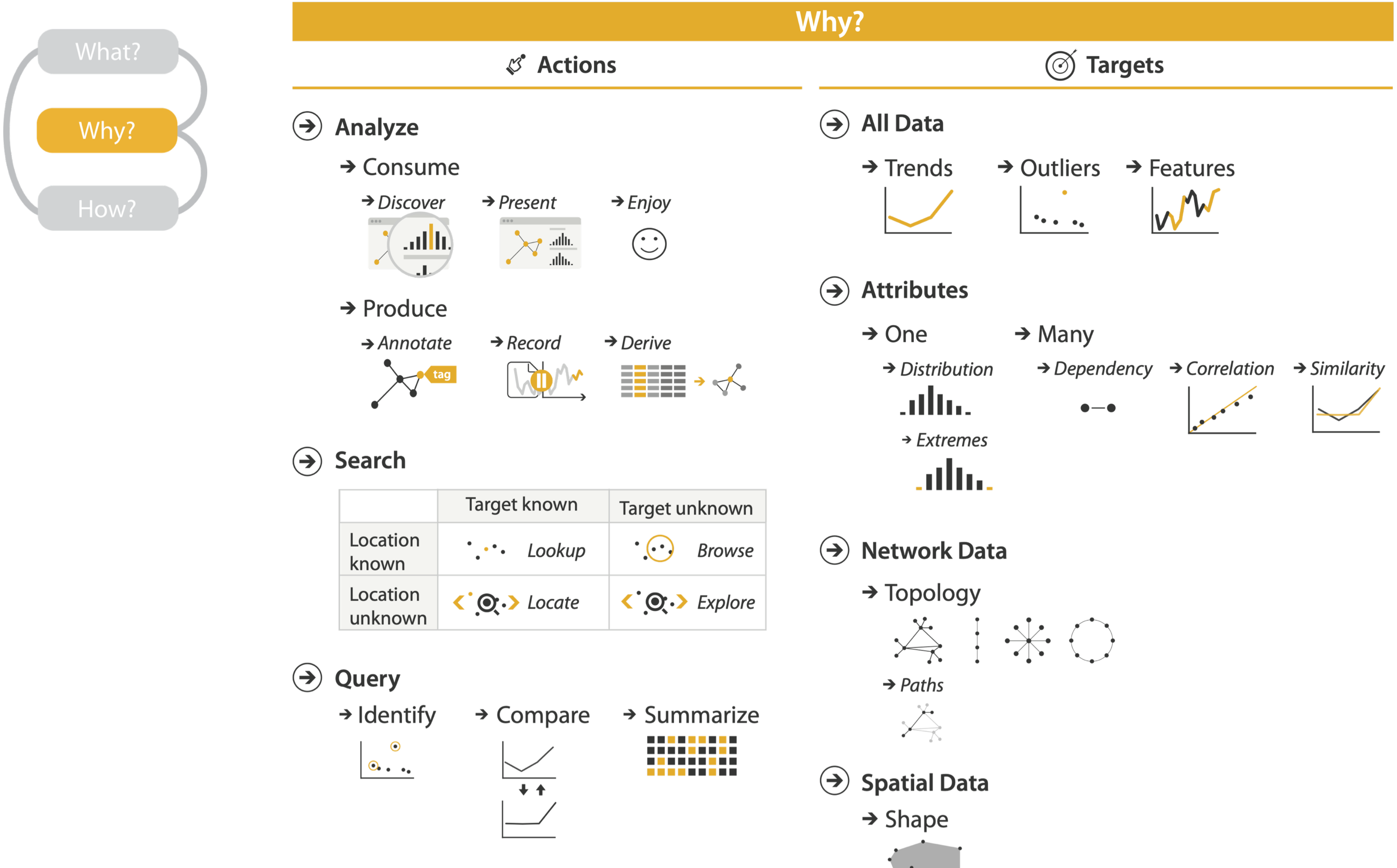


Fig. 1. Temporal network analysis process

study examples, we can find out which tasks have been covered so far and suggest additions for designing future visualizations. At the same time, we apply the task taxonomy to a social networking site called Nation of Neighbors and provide network evolution analysis tasks on a design space.

The following section introduces three dimensions for defining the temporal network visualization. A list of tasks identified by combining the dimensions and their taxonomy is provided in Section 3. In Section 4, we propose the network evolution analysis tasks and show examples of the usage of the task taxonomy in a Nation of Neighbors social network analysis project. The last section concludes the paper and reveals our future plans.

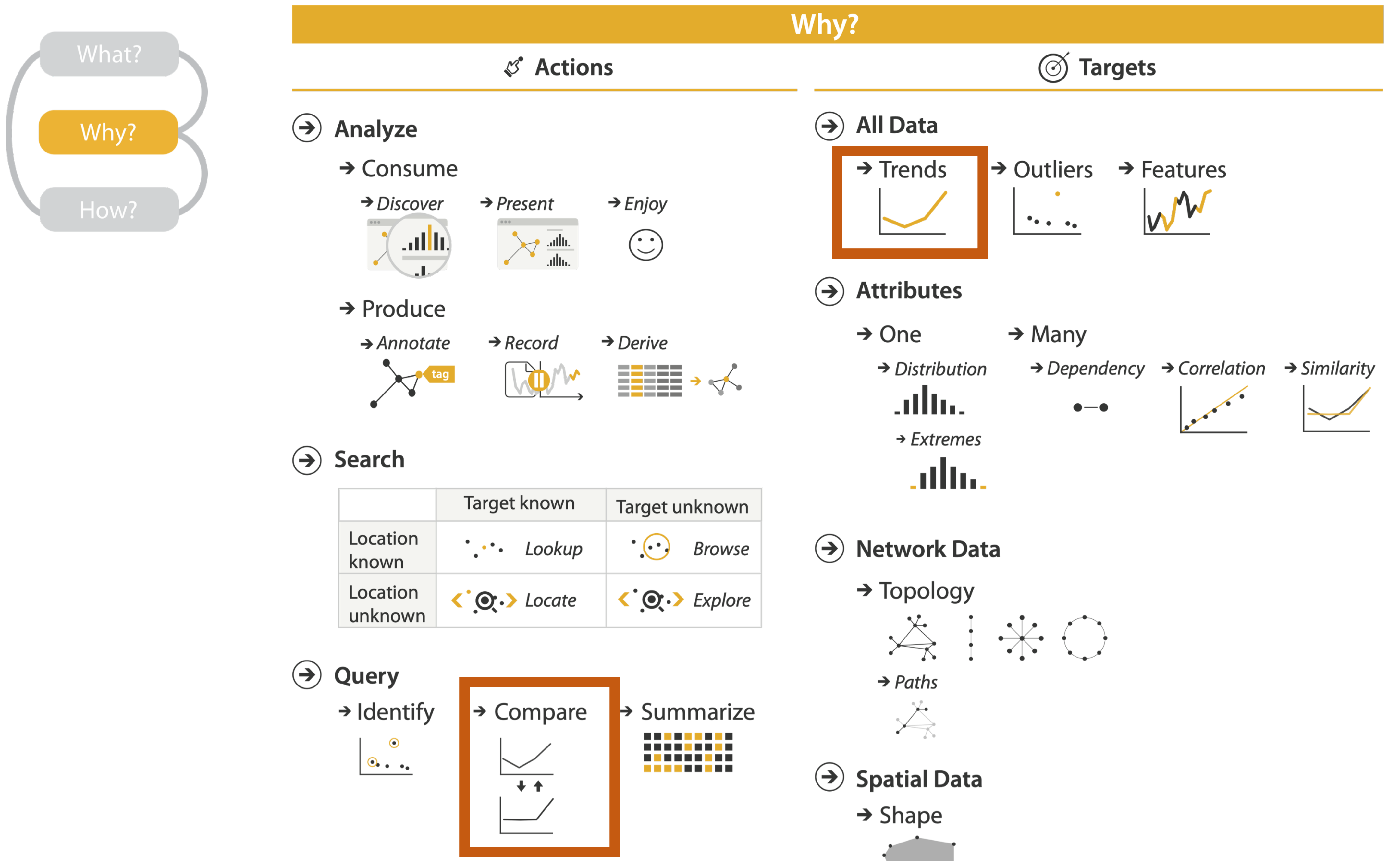




Source: Visualization Analysis & Design, by Tamara Munzner

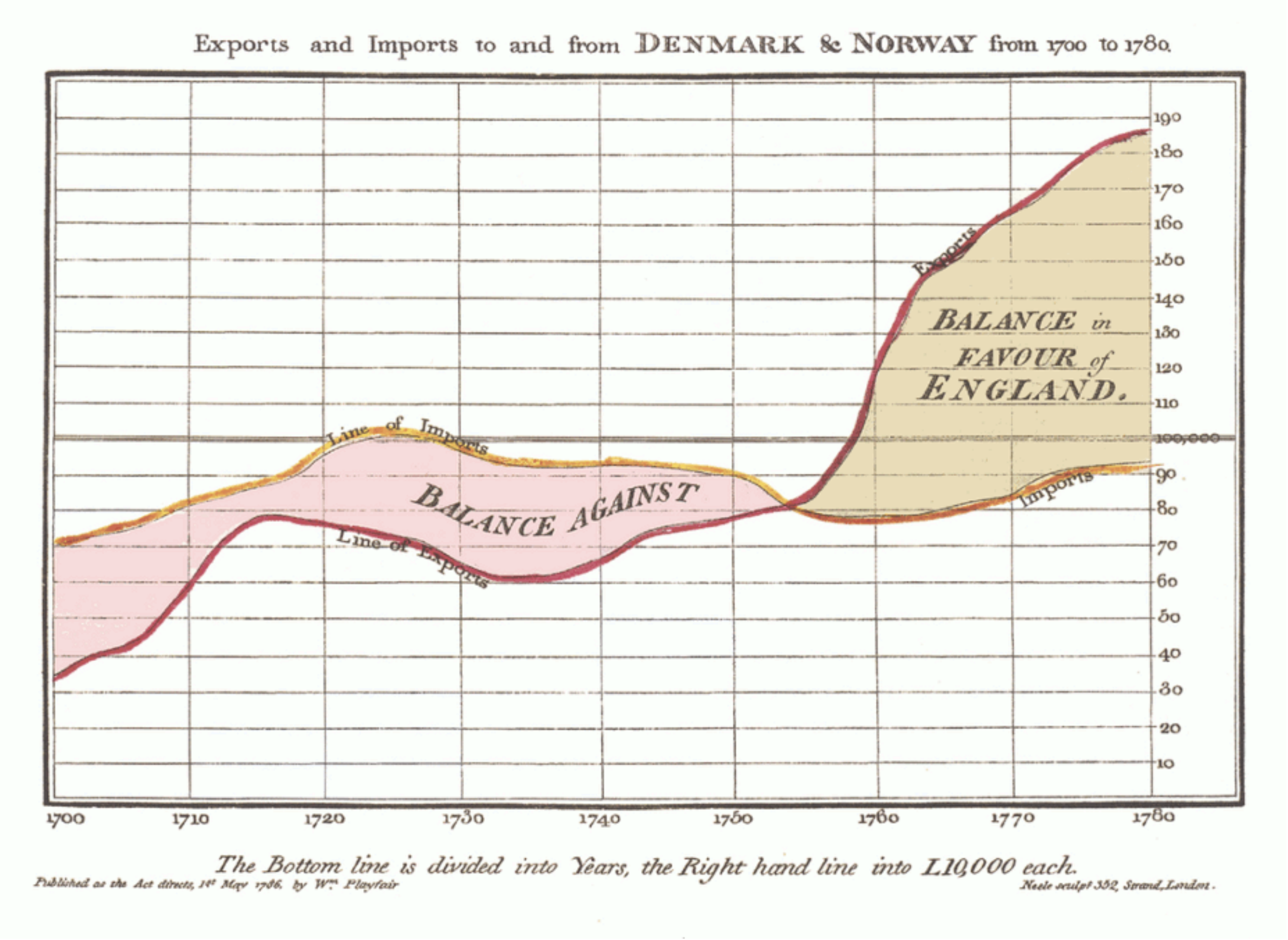
	Target known	Target unknown
Location known	Lookup	Browse
Location unknown	Locate	Explore

Source: Visualization Analysis & Design, by Tamara Munzner



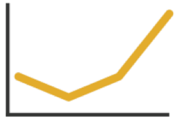
Source: Visualization Analysis & Design, by Tamara Munzner

year	import	export
1700	70,000	35,000
1710	80,000	60,000
1720	97,000	78,000
1730	97,000	65,000
1740	91,000	65,000
1750	90,000	79,000
1760	82,000	82,000
1770	79,000	120,000
1780	82,000	162,000
1790	92,000	185,000



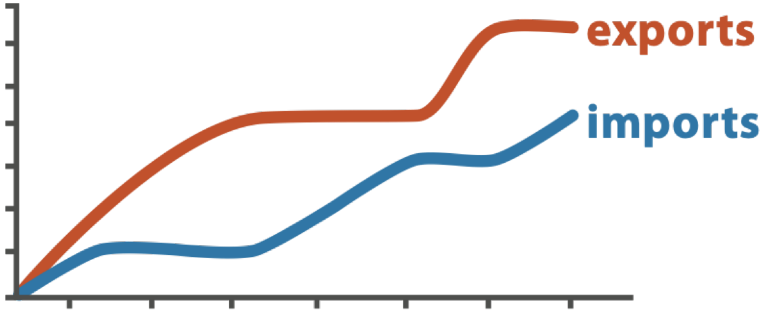
➔ *Derive*

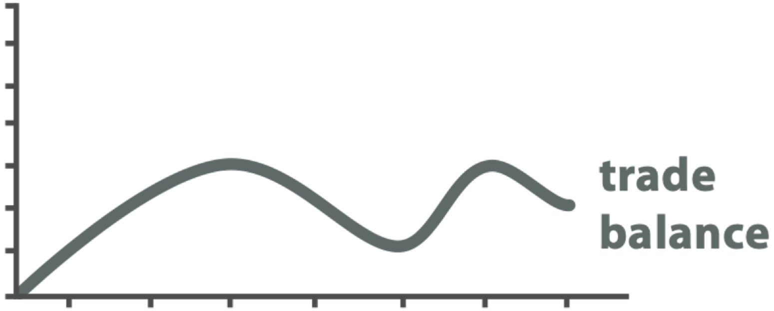
➔ Trends



Don’t necessarily draw what you’re given

→ *Derive*





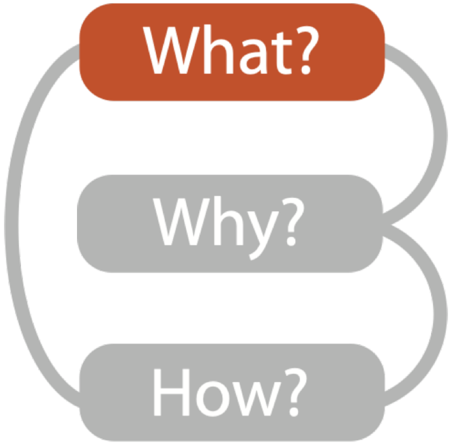
trade balance = exports − imports

Original Data

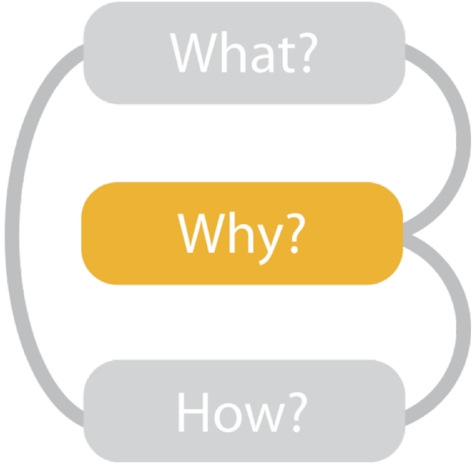
Derived Data

Good visualization design requires to know

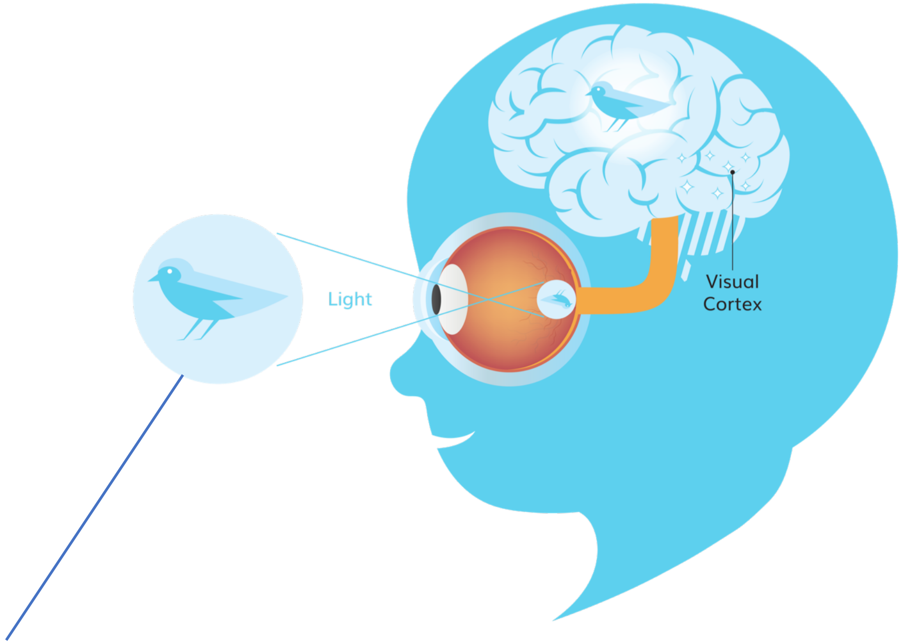
Data Abstraction



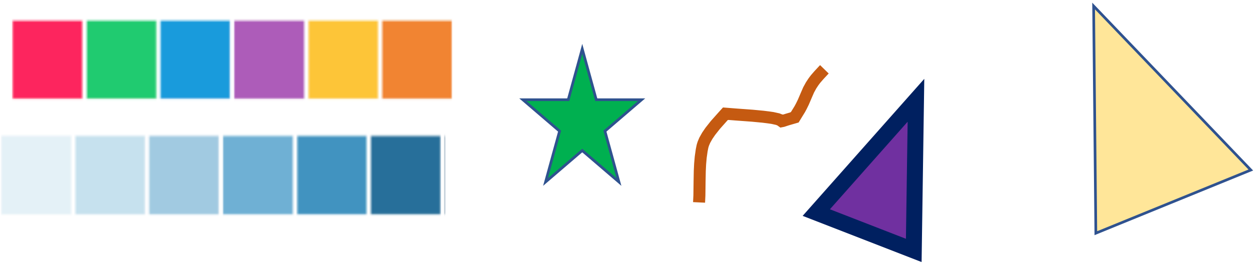
Task Abstraction



Perception & Cognition



Visual vocabulary



Next up:

Marks and Channels

Acknowledgements:
The materials in this course is heavily based on the book and lecture materials from Tamara Munzner, as well as teaching materials and resources from John Stasko, Sheelagh Carpendale, Jeffrey Heer, Alex Lex, Miriah Meyer, Benjamin Bach, Charles Perin, Enrico Bertini, Evan Peck, Arvind Satyanarayan, Alberto Cairo, and Petra Isenberg.