VISUALIZATION DESIGN

Hans-Jörg Schulz

1. INTRODUCTION

2. VISUAL ENCODING

3. BASIC CHART TYPES

Basics

4. INTERACTION

5. VISUALIZATION DESIGN

6. DATA PREPROCESSING

Visualization Building Blocks & Processes

7. RECAP 1st Half

8. MULTIVARIATE DATAVIS

9. TEMPORAL DATAVIS

10. GEOSPATIAL DATAVIS

11. GRAPH DATAVIS

Visualization Techniques

12. 3D DATAVIS

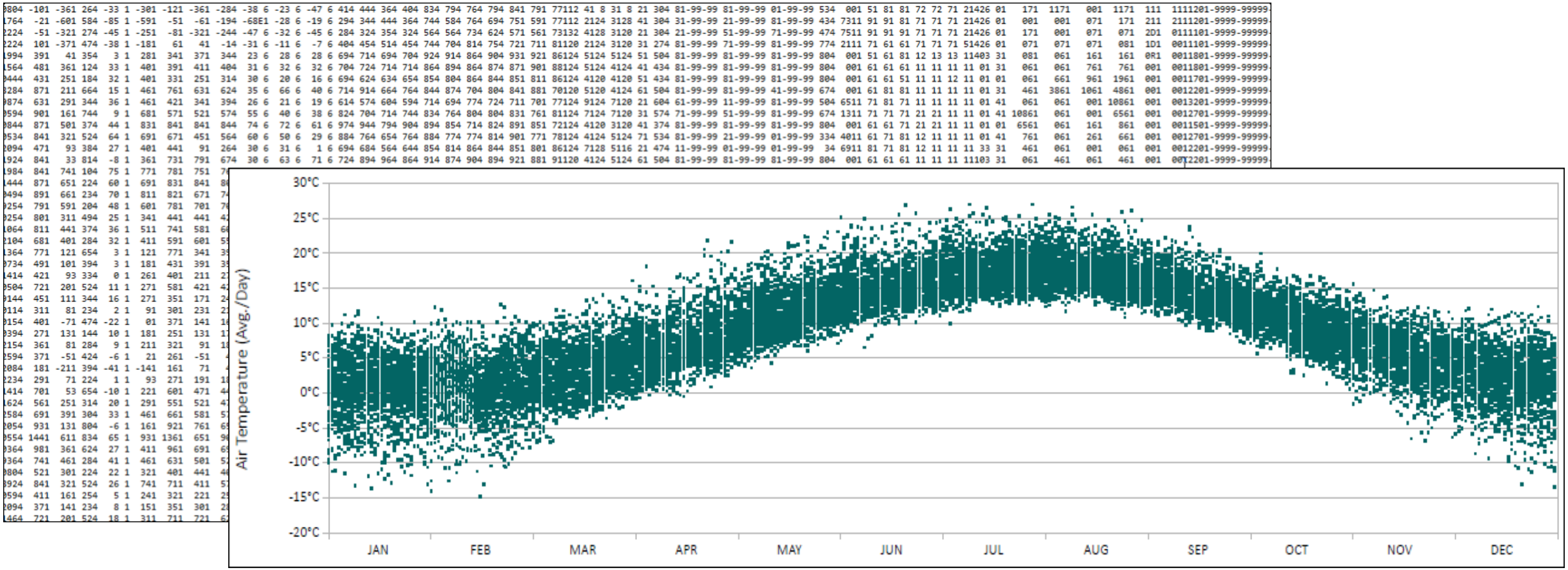
13. VISUAL ANALYTICS

Visualization Applications

14. RECAP 2nd Half

WHY TALK ABOUT VIS DESIGN?

How to get from a visualization problem to a visualization solution?



OVERVIEW

- **The Visualization Problem**
  - Data Abstraction
  - Task Abstraction

- **Design Processes for Data Visualization**
  - Design Activity Framework
  - 5 Design Sheet Methodology

- **Evaluating the Resulting Visualization Solution**
  - Munzner's Nested Model for Visualization Design

THE VISUALIZATION PROBLEM: DATA ABSTRACTION

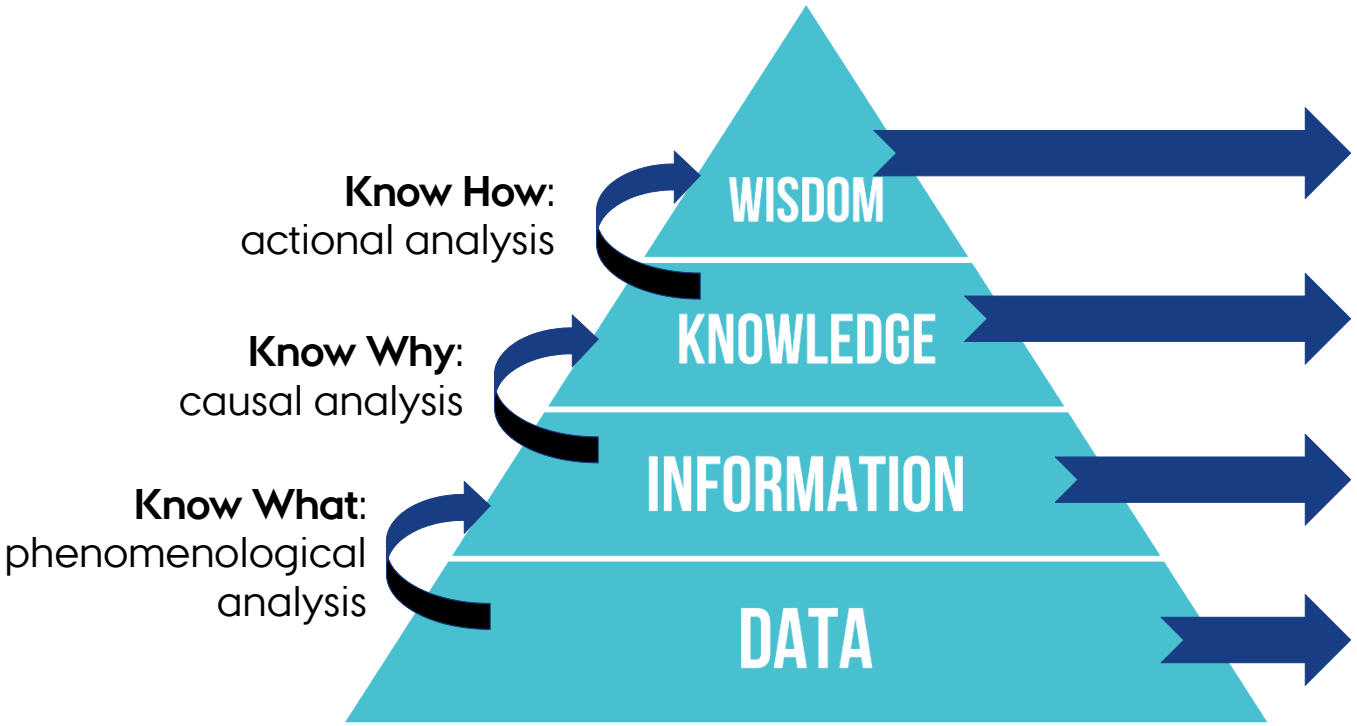
AARHUS
UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

HANS-JÖRG SCHULZ
ASSOCIATE PROFESSOR

WHAT IS DATA?

Further reading: Min Chen et al. 2009
"Data, Information, and Knowledge in Visualization"

The DIKW Hierarchy



- Prognoses (e.g., decision X will lead to outcome Y)
- Reasoning (e.g., if X leads to Y, then X' will lead to Y')
- Adaptation (e.g., situation Z => situation Z')

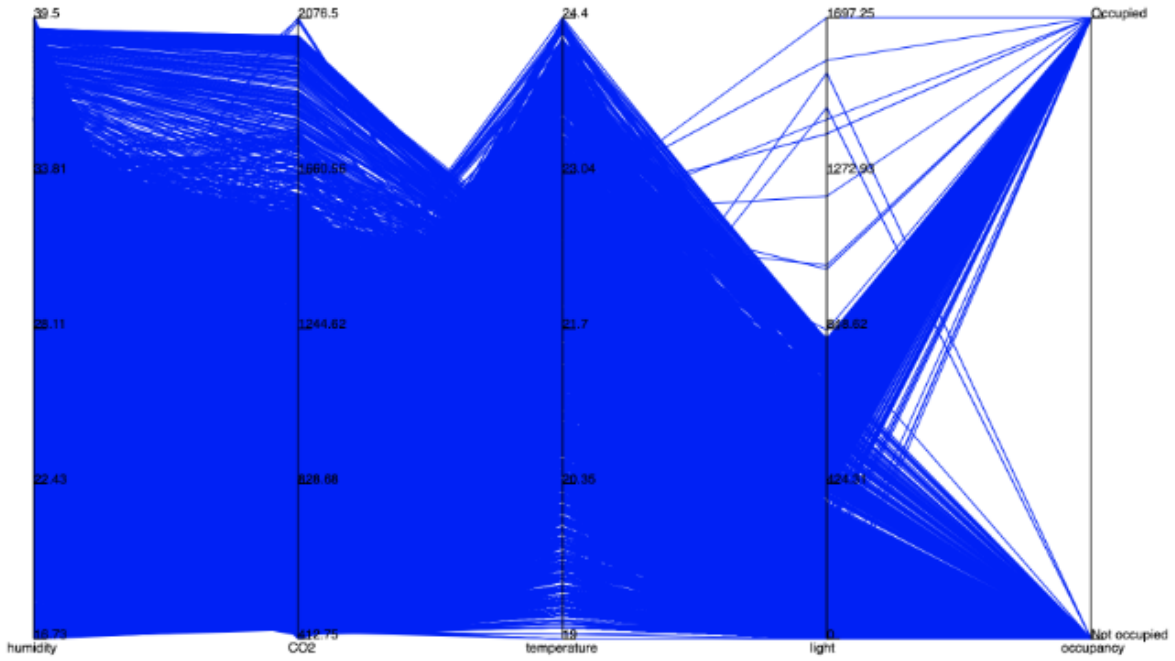
- Models (e.g., simulation models, statistical models)
- Parameters (e.g., controls, influencing factors)
- Limitations (e.g., boundary conditions)

- Distributions (e.g., trends, clusters, outliers)
- Relations (e.g., correlations, associations)
- Recognition (e.g., classification, matching)

- Facts (e.g., names and favorite colors)
- Measurements (e.g., weight and length)
- Observations (e.g., eye color and spoken languages)

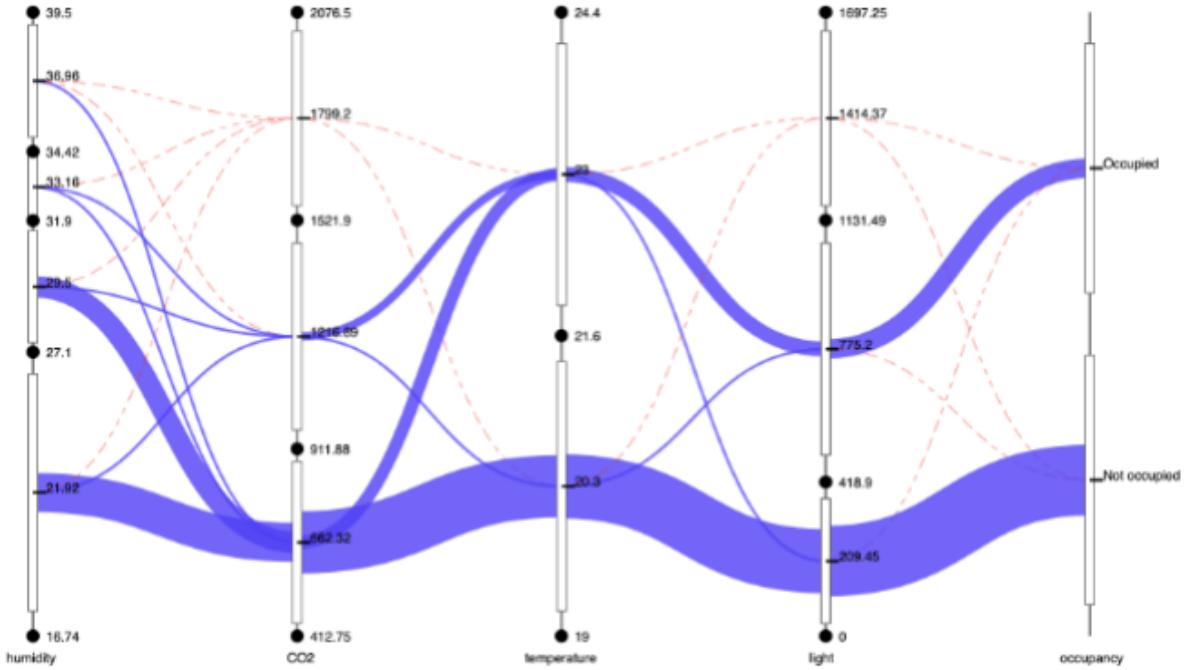
VISUALIZING DATA VS. INFORMATION

Data



Plain Parallel Coordinates Visualization

Information



Confluent Parallel Coordinates Visualization

Image Source: Cui et al. 2019

THE DATA SPACE

Source: [Schulz et al. 2017]

DATA CONTEXT

DATA DOMAIN
REFERENCE SPACE
INDEPENDENT VARIABLES

Frame of reference for any measured or observed data characteristic. Examples:

- Latitude / Longitude
- Time
- CPR number

Important Properties:

- Extent of Validity: point, local, global
- Topology, Grid type

DATA CONTENT

DATA VALUES
ATTRIBUTE SPACE
DEPENDENT VARIABLES

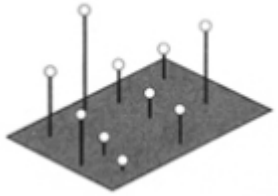
Gathered data at a particular reference point. Examples:

- Temperature
- CO2 concentration
- Wind speed & direction

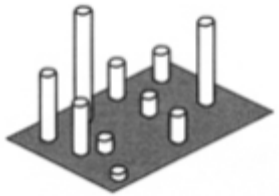
Important Properties:

- Data type: categorical, numerical,...
- Data structure: scalar, vector, tensor
- Value ranges

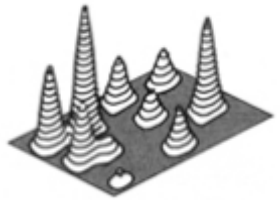
THE DATA CONTEXT – EXTENT



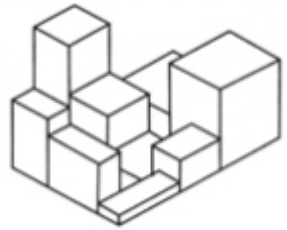
Point



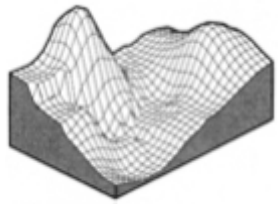
narrow, discrete



narrow, continuous

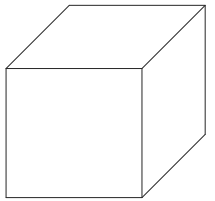


broad, discrete



broad, continuous

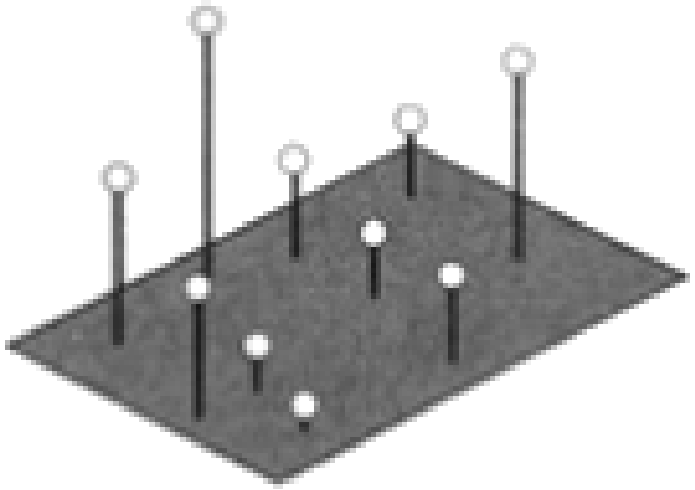
Local



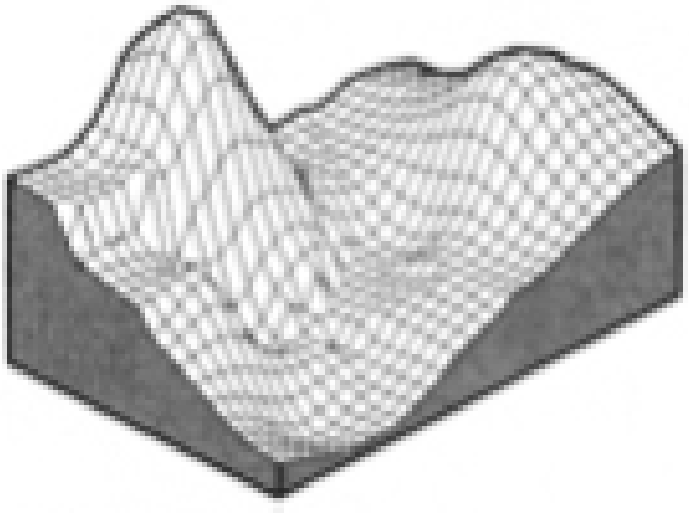
Global

Image source [MacEachren 1992]

THE DATA CONTEXT – INTERPOLATION



?



Inverse Distance Weighting – Shephard's Method

$$z(x,y) = \frac{\sum_{i=1}^{k} w_i(x,y) \times z_i}{\sum_{i=1}^{k} w_i(x,y)} \quad \text{with} \quad w_i(x,y) = \frac{1}{\sqrt[2]{(x_i - x) \times (y_i - y)}^{p}}$$

THE DATA CONTEXT – INTERPOLATION

Inverse Distance Weighting – Shephard's Method

$$z(x,y) = \frac{\sum_{i=1}^{k} w_i(x,y) \times z_i}{\sum_{i=1}^{k} w_i(x,y)} \quad \text{with} \quad w_i(x,y) = \frac{1}{\sqrt[2]{(x_i - x) \times (y_i - y)}^{p}}$$

Influence of the Power Parameter *p:*

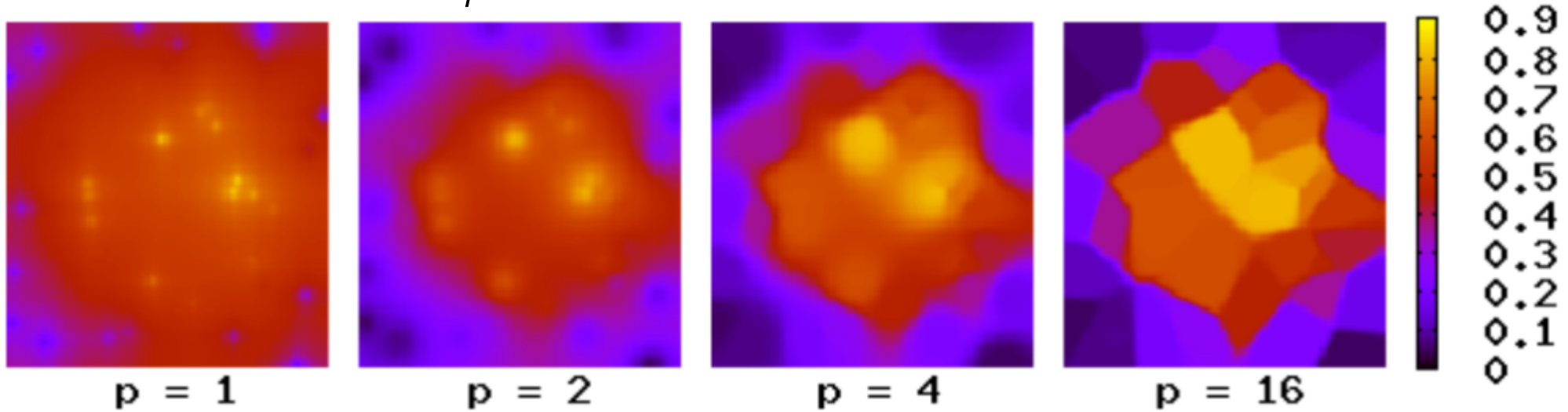
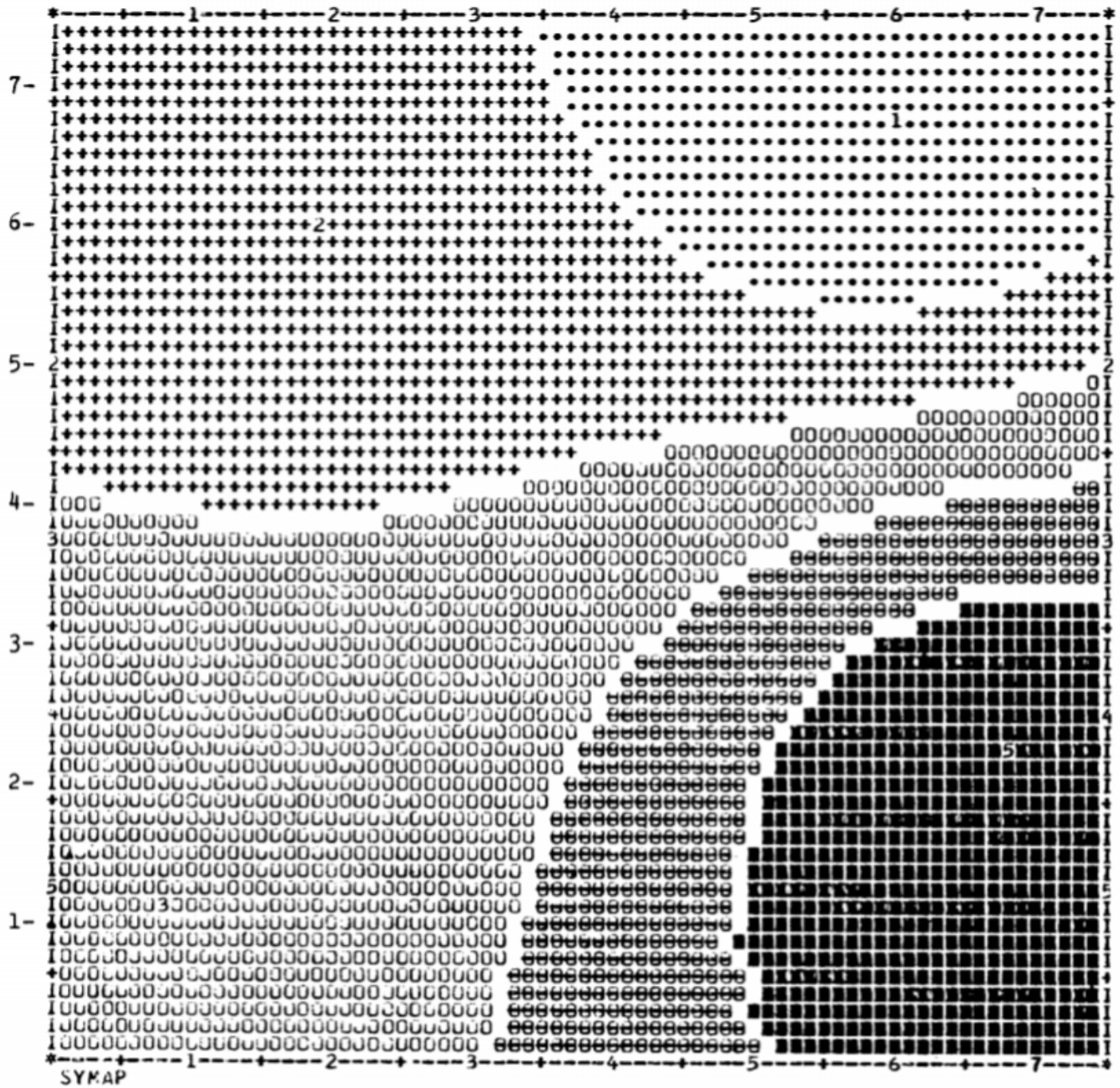


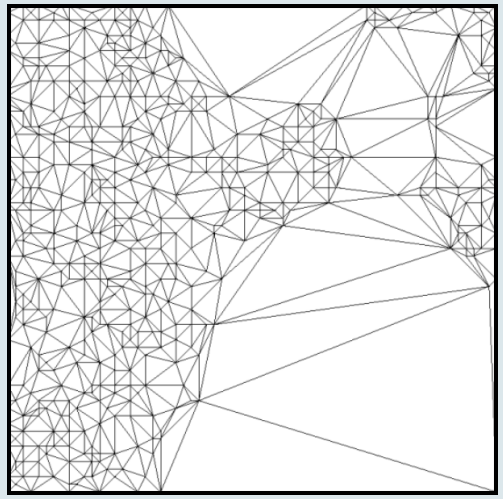
Image source: Wikipedia



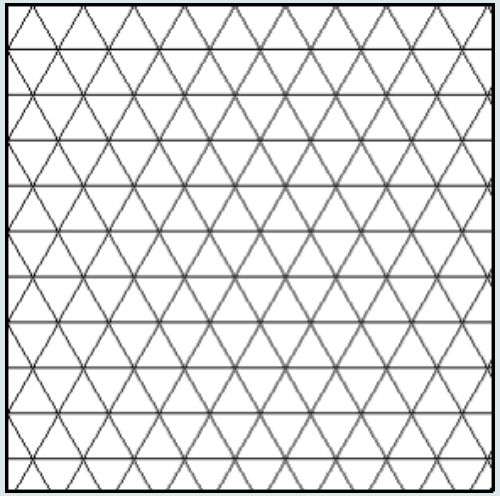
[Donald Shepard 1968]

THE DATA CONTEXT – TOPOLOGY

Grid-based Topology

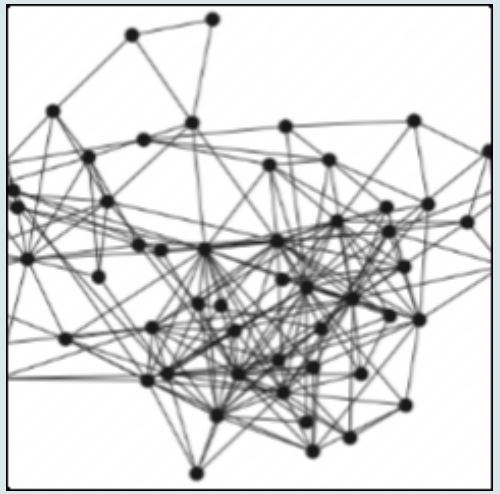


Irregular Grid

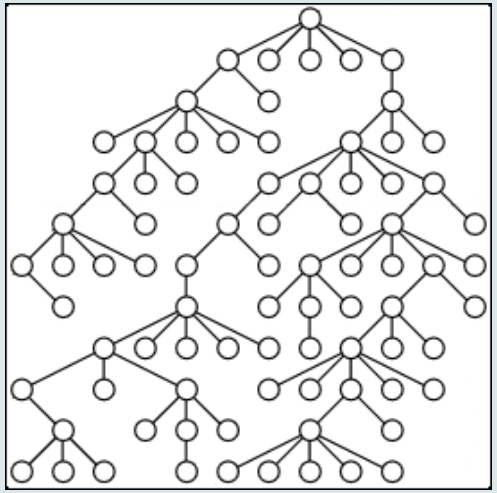


Regular Grid

Graph-based Topology



Network



Hierarchy

THE DATA CONTENT – TYPE

	Quantitative Data		Qualitative Data	
	Continuous Data	Discrete Data	Ordinal Data	Categorical Data
Interpolate	✓			
Difference	✓	✓		
Sort	✓	✓	✓	
Match	✓	✓	✓	✓

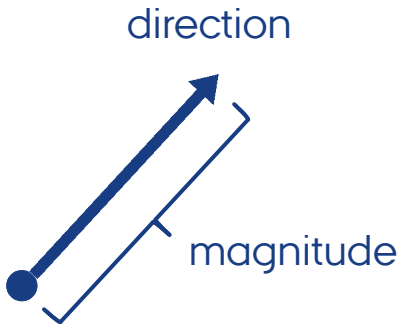
Adapted from [Stevens 1946] – DOI: 10.1126/science.103.2684.677

THE DATA CONTENT – STRUCTURE

Image source: https://www.tatvic.com/blog/calender-heatmap-with-google-analytics-data/

Scalar Data := only 1 magnitude

THE DATA CONTENT – STRUCTURE



Vector Data := 1 magnitude + 1 direction

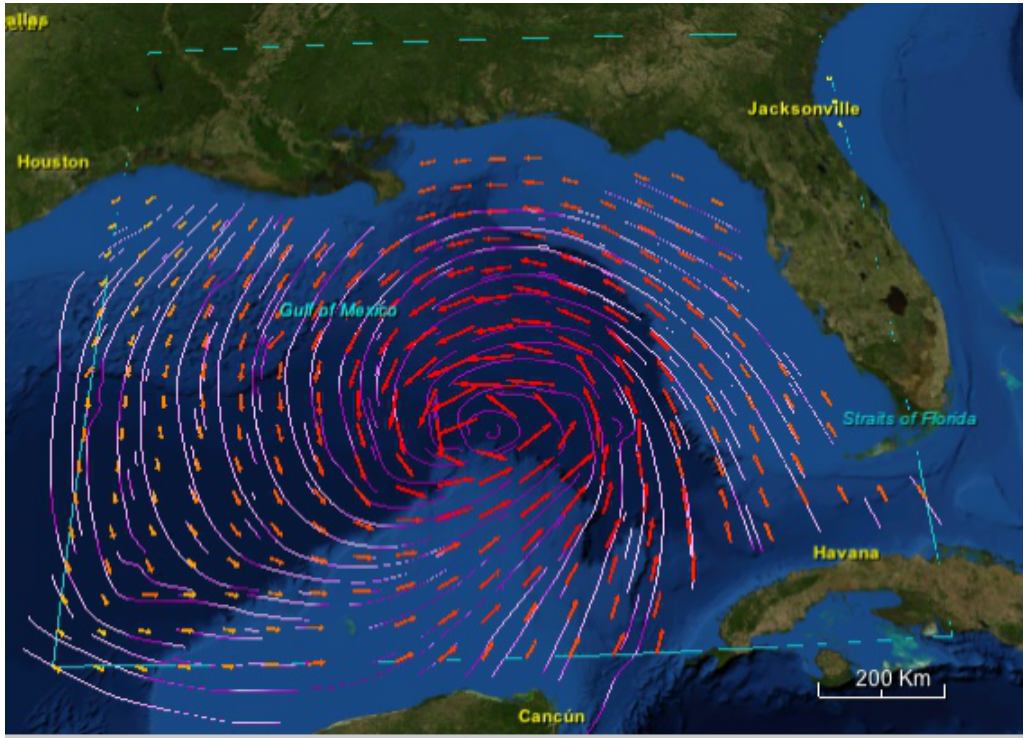
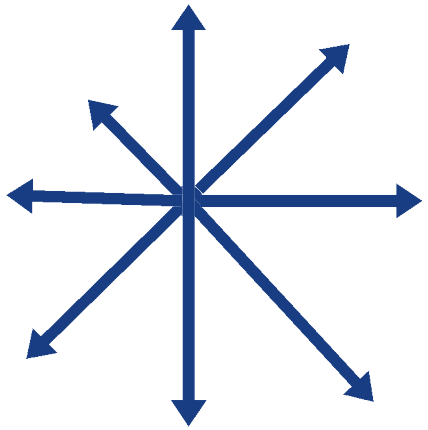


Image Source: https://people.eecs.ku.edu/~jrmiller/Courses/775/Description/Description.php

THE DATA CONTENT – STRUCTURE



Tensor Data := n magnitudes + n directions

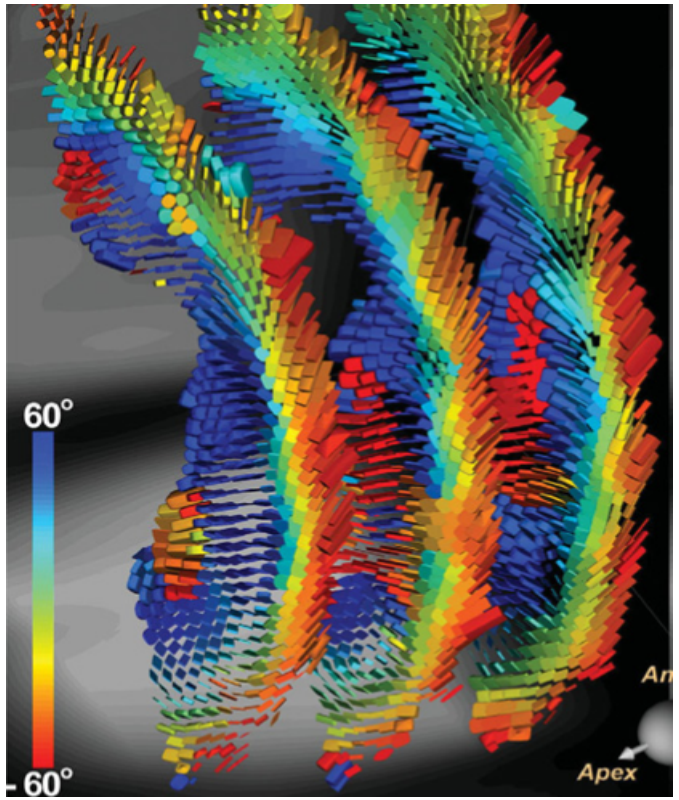
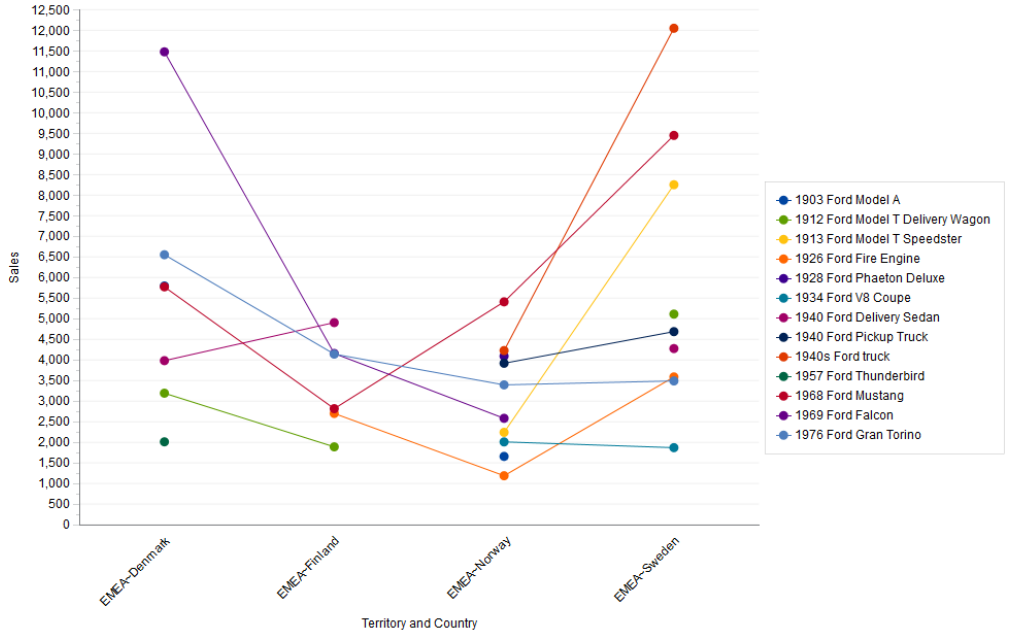


Image Source: [Nielles-Vallespin et al. 2013]

WHAT TO GAIN FROM DATA ABSTRACTION?

- domain-independent way to discuss data
- visualization requirements, generalization of problem & solution
- expressive visualization that show the data and only the data



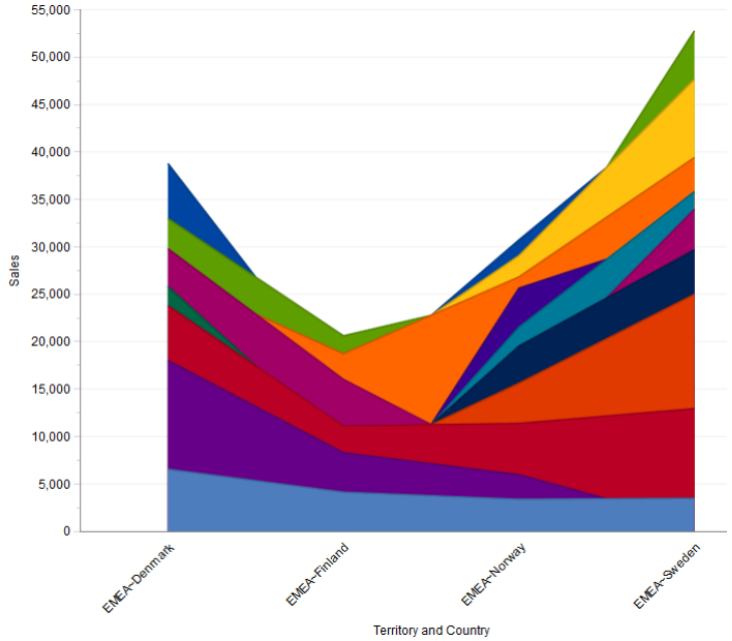


Image source: Pentaho 5.1 Online Documentation

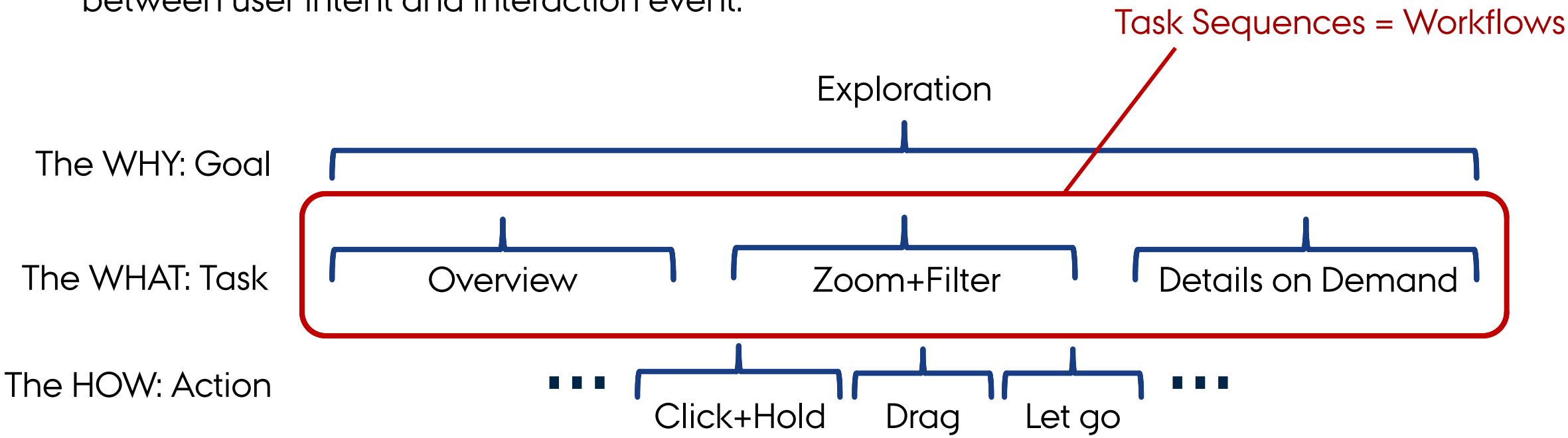
THE VISUALIZATION PROBLEM: TASK ABSTRACTION

AARHUS
UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

HANS-JÖRG SCHULZ
ASSOCIATE PROFESSOR

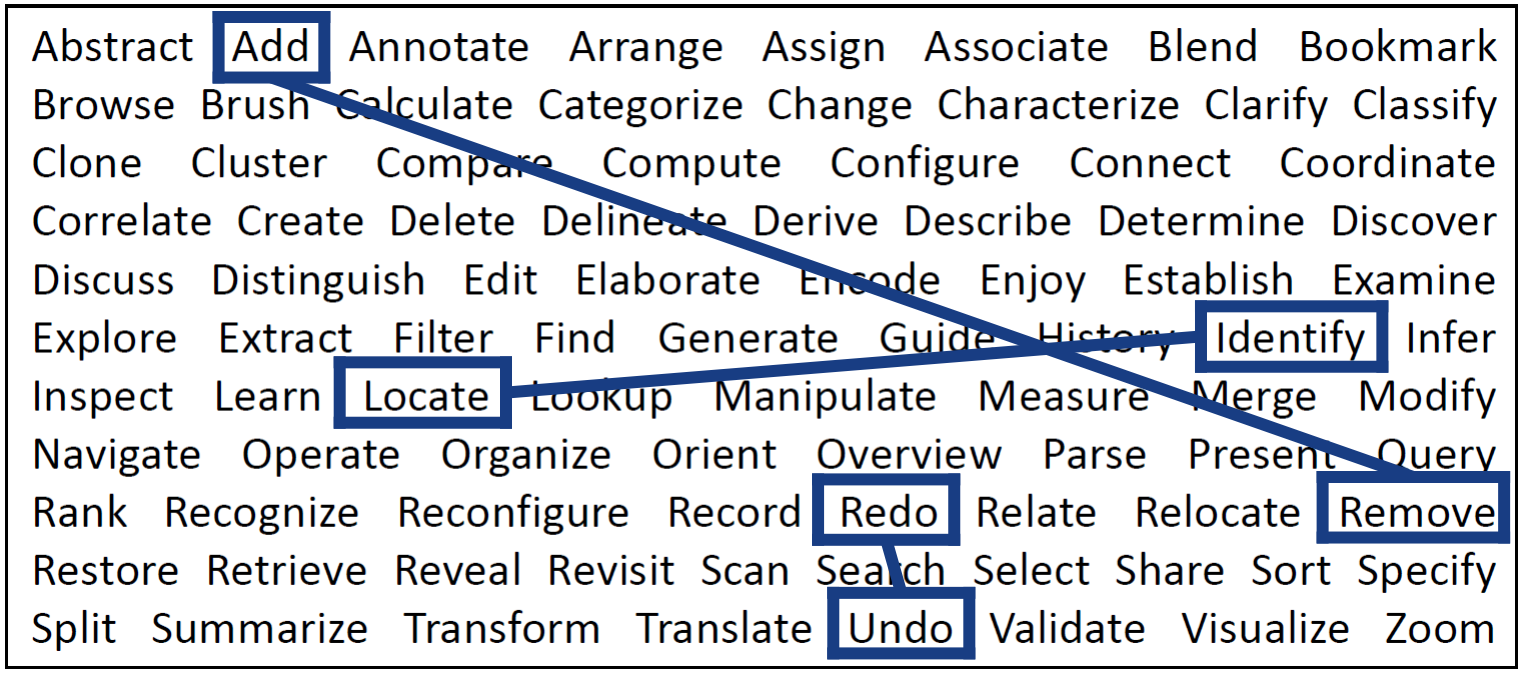
WHAT IS A TASK?

Task as a mid-level abstraction
between user intent and interaction event:



Adapted/Merged from Gotz+Zhou 2008 / Schulz et al. 2013 / Rind et al. 2016

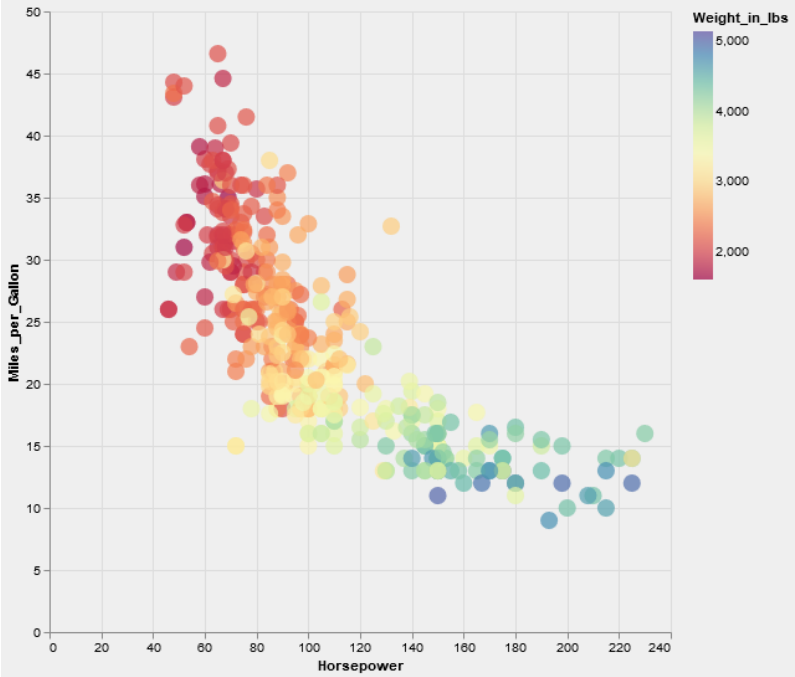
WHAT TASKS ARE THERE?



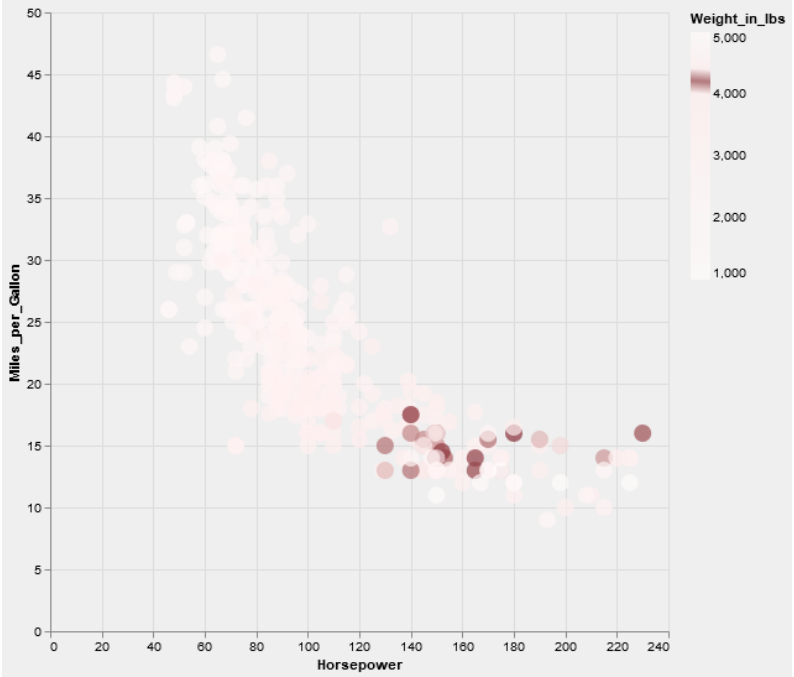
Distinction between **unipolar** and **bipolar** tasks, depending on whether a task has a “natural opposite”. [Sedig, Parsons 2013]

IDENTIFY TASK VS. LOCATE TASK

Data Identification:
Find the value of a particular data point



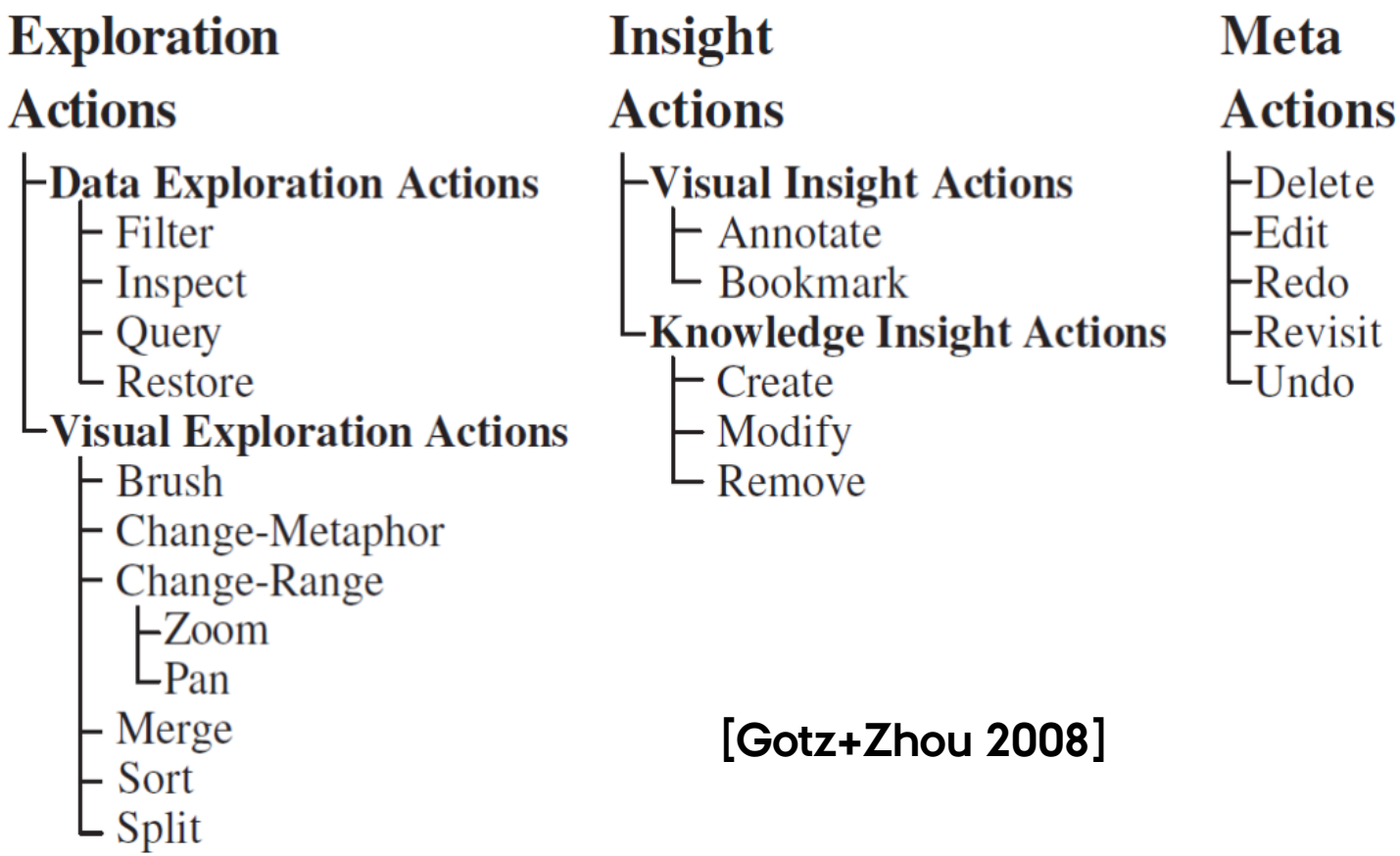
Data Localization:
Find data points of a particular value



TASK TAXONOMIES

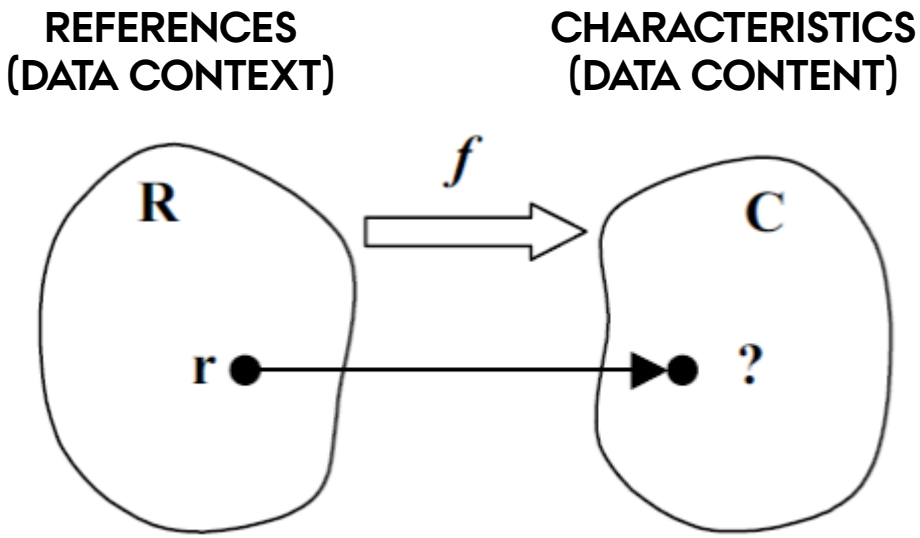
Taxonomy is the practice and science of classification of things or concepts, including the principles that underlie such classification.

-- From Wikipedia



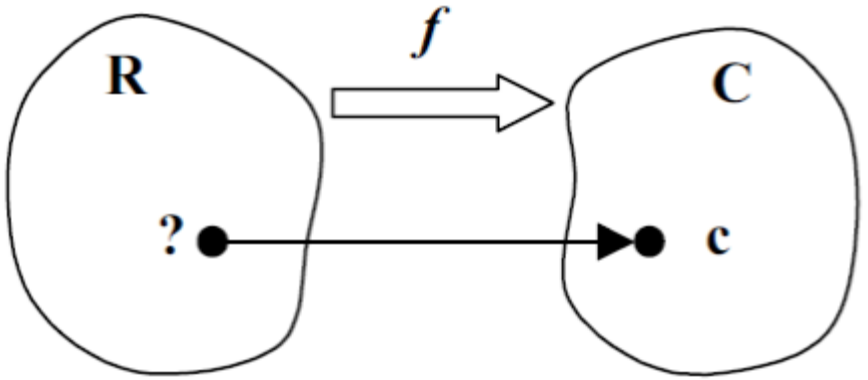
[Gotz+Zhou 2008]

THE TASK FRAMEWORK BY ANDRIENKO [2006]



Lookup Task:
On a given date what is the price of the Apple stock?

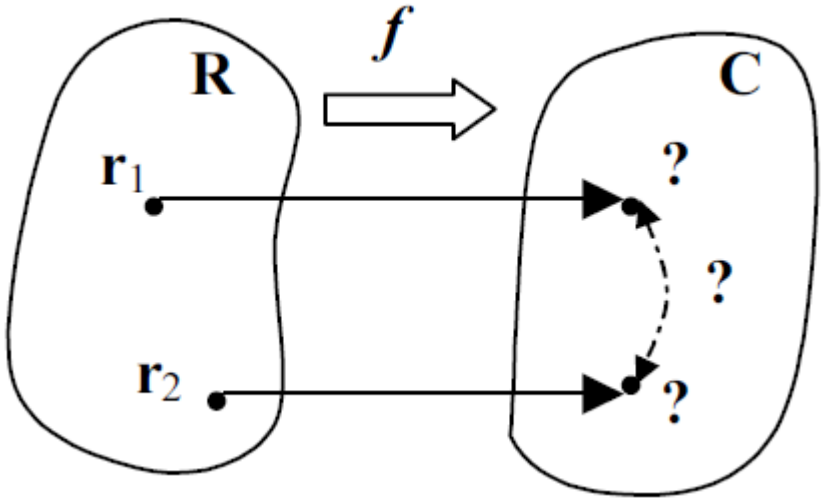
-> Identify Task



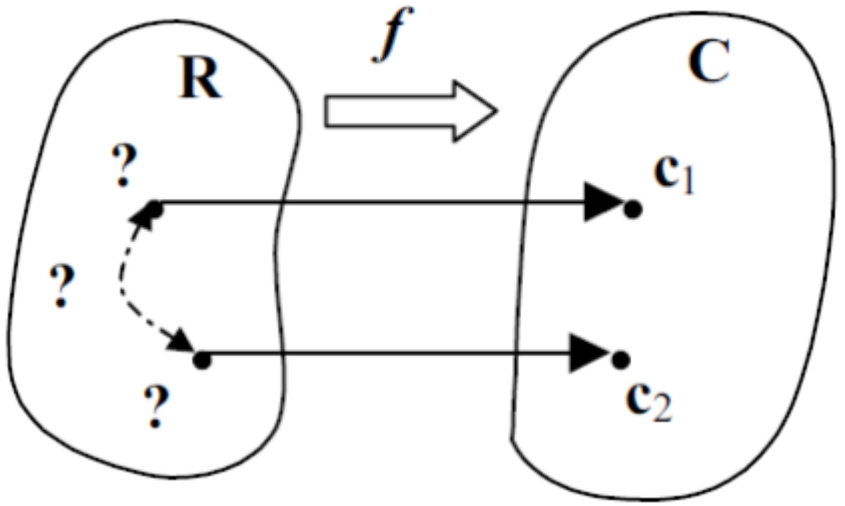
Inverse Lookup Task:
On which dates was the price of the Apple stock at 200$?

-> Locate Task

THE TASK FRAMEWORK BY ANDRIENKO [2006]

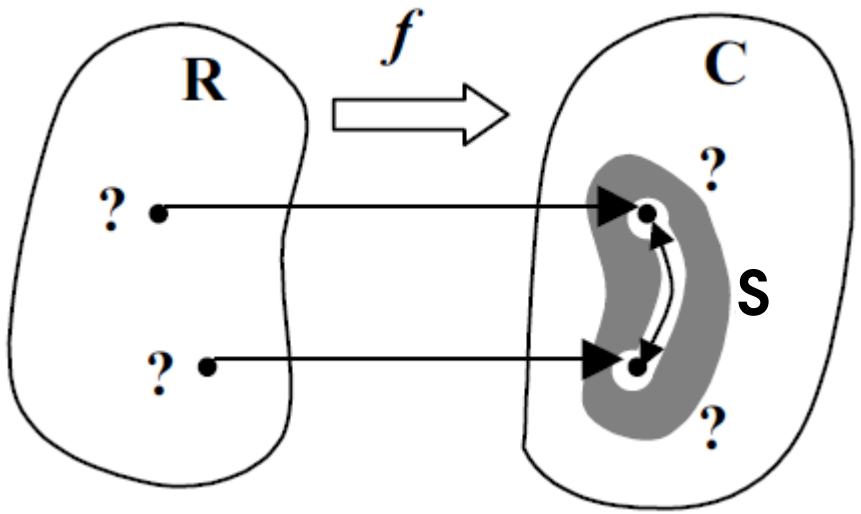


Comparison Task:
How did the Apple stock price change from 10 years ago to today?

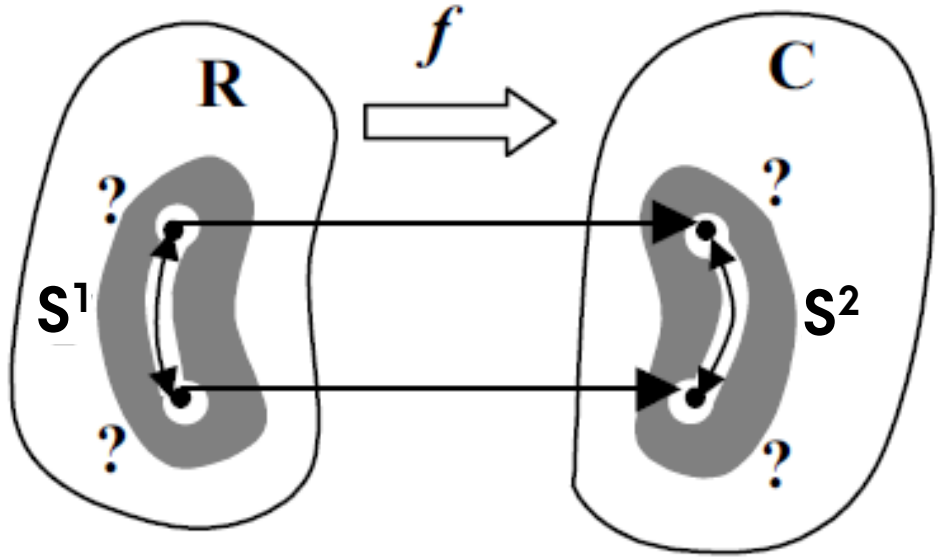


Inverse Comparison Task:
How much time went by from the Apple stock price hitting 100$ the first time and doubling that?

THE TASK FRAMEWORK BY ANDRIENKO [2006]



Relation-Seeking Task with 1 relation:
When should I have bought and sold Apple stock to triple my investment?



Relation-Seeking Task with 2 relations:
When did the Apple stock price increase by 10% from one day to the next?

THE TASK FRAMEWORK BY ANDRIENKO [2006]

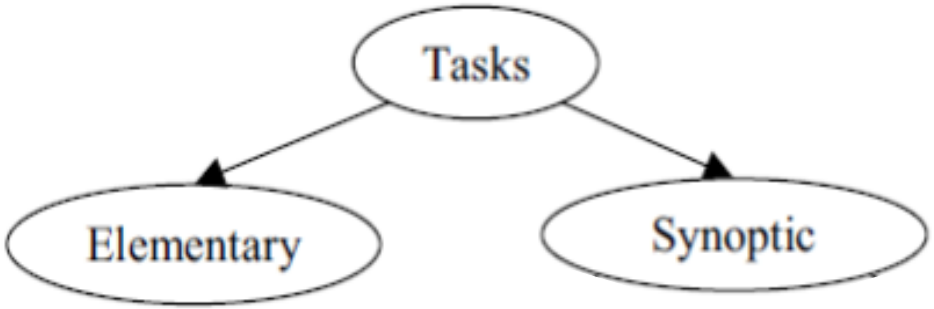
So far: **Elementary tasks** that ask/answer questions about individual data items and their relations

Now: **Synoptic tasks** that ask/answer questions about sets of data items and their relations

Examples of synoptic tasks:

What trend does the Apple stock price exhibit between 1996 and 2000? How does this trend compare to the one between 2000 and 2004?

How is the ownership of Apple stock distributed across the world? How did this distribution evolve over time?



THE TASK FRAMEWORK BY ANDRIENKO [2006]

Elementary lookup -> Synoptic lookup: “Pattern definition”

Elementary: *On a given date, what was the price of the Apple stock?*

Synoptic: *During a given time interval, what was the trend of the Apple stock price?*

Elementary inverse lookup -> Synoptic inverse lookup: “Pattern search”

Elementary: *For a given Apple stock price, on what date(s) was it attained?*

Synoptic: *In which time intervals did the Apple stock price increase by 300%?*

Elementary comparison -> Synoptic comparison: “Pattern comparison”

Elementary: *How do the stock prices at the beginning and end of a trading week differ?*

Synoptic: *What happens more often: the stock price increasing or decreasing over the course of a week?*

Elementary inverse comparison -> Synoptic inverse comparison: “Pattern comparison”

Elementary: *How long after the stock price hit 100$, it went up to 200$?*

Synoptic: *How do the periods of stock price growth relate to those of stock price decrease?*

THE TASK FRAMEWORK BY ANDRIENKO [2006]

Elementary relation seeking -> Synoptic relation seeking (1 relation):

Elementary: *When should I have bought and sold Apple stock to triple my investment?*

Synoptic: *Over how long of a time period does one on average need to hold Apple stock, in order to triple one's investment?*

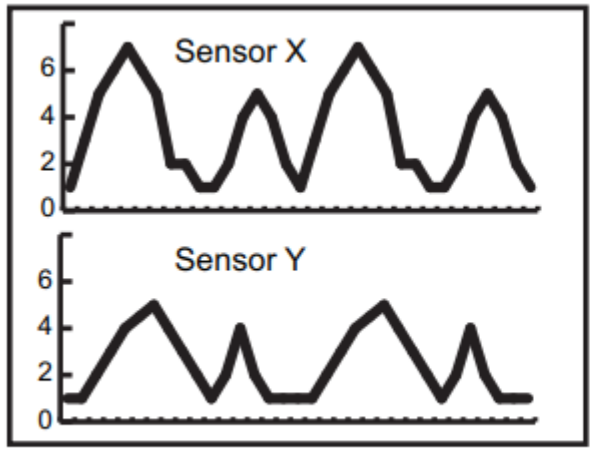
Elementary relation seeking -> Synoptic relation seeking (2 relations):

Elementary: *When did the Apple stock price increase by 10% from one day to the next?*

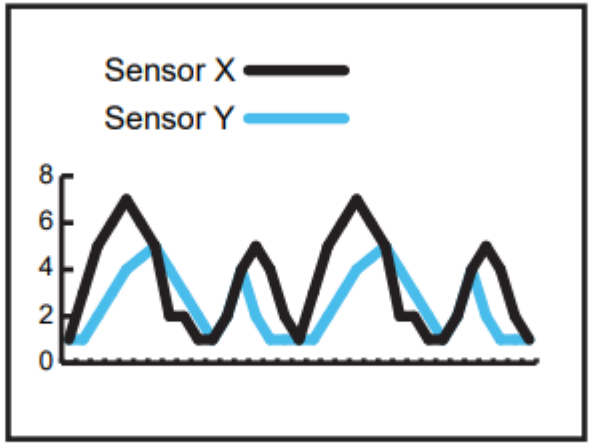
Synoptic: *How frequent are Apple stock price increases by 10% from one day to the next?*

WHAT TO GAIN FROM TASK ABSTRACTION?

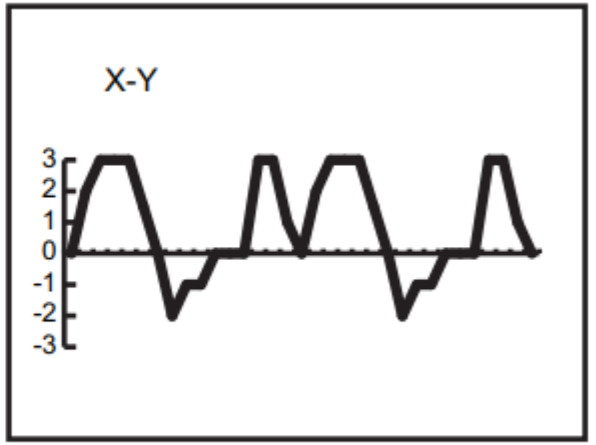
- domain-independent way to discuss analytic tasks
- requirements engineering, generalization of problem & solution
- effective visualization that support the task



(a) Juxtaposition



(b) Superposition



(c) Explicit Encoding: Difference

Image source: [Gleicher 2018]

VISUALIZATION DESIGN: DESIGN ACTIVITY FRAMEWORK

AARHUS
UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

HANS-JÖRG SCHULZ
ASSOCIATE PROFESSOR

DESIGN ACTIVITY FRAMEWORK (DAF)

[McKenna et al. 2014]

understand

Motivation: to gather and research the needs of the user

Methods: interviews, surveys, identification of usage scenarios, abstractions

Outcome: visualization requirements

ideate

Motivation: to generate good ideas for supporting the requirements

Methods: sketching, lo-fi mock-ups

Outcome: a set of externalized visualization ideas

make

Motivation: to concretize ideas into tangible prototypes

Methods: visual design, algorithm design

Outcome: a testable hi-fi prototype

deploy

Motivation: to bring a prototype into effective action in a real-world setting

Methods: software engineering

Outcome: usable visualization system

Understand

HOW TO GET VIS REQUIREMENTS?

Depending on your access to the data:

e.g., Box-Muller Transform for normal-distributed data

- **No access:**
  1) Use synthetic, realistic, or historic data from the earliest moment possible.
  2) Establish metadata – e.g., data format, expected data size,...
- **Access unusable:** Export snapshot / sample in usable format.
- **"Incrementally increasing" access:** Use a staged/iterative development process with minimal viable products to progressively gain access to data.
  -> Positive side effect: If the visualization project gets cancelled early, you have already an intermediary outcome to show for (and to bill for).

VISUALIZATION DESIGN: 5 DESIGN SHEET METHOD

ideate

HOW TO COME UP WITH VIS IDEAS?

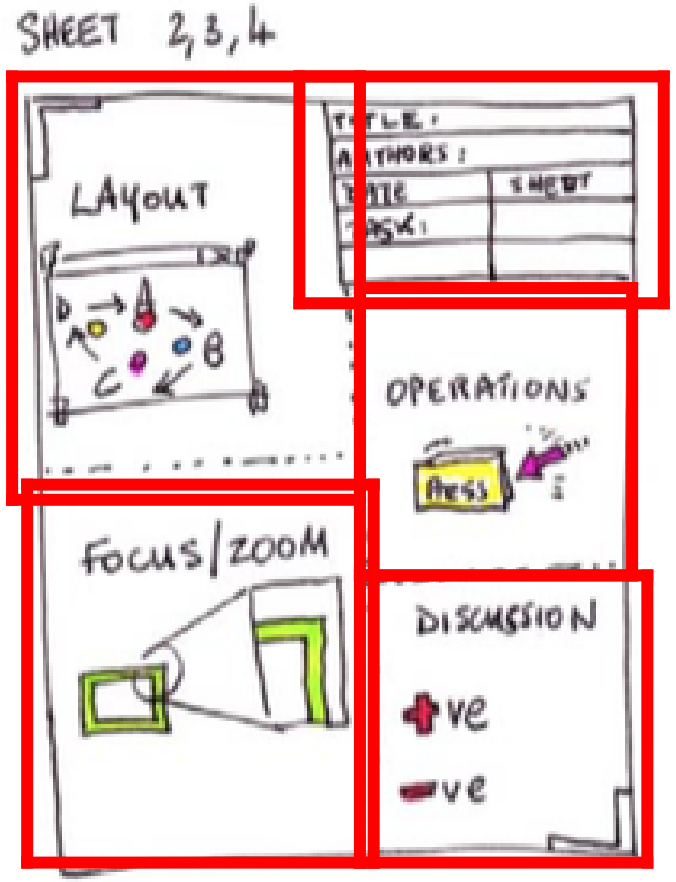
Sheet 1: The Brainstorm Sheet

- Ideas:15-20 small, possibly partial ideas
- Filter: what works, what doesn't work
- Categorize: group your ideas into similar clusters
- Combine & Refine: put together properties from dissimilar clusters to generate more possibilities, look for improvements
- Question: Which three ideas are the best ones to carry forward into the next design step?

The 5 Design Sheet Methodology (https://doi.org/10.1007/978-3-319-55627-7)

ideate

HOW TO COME UP WITH VIS IDEAS?



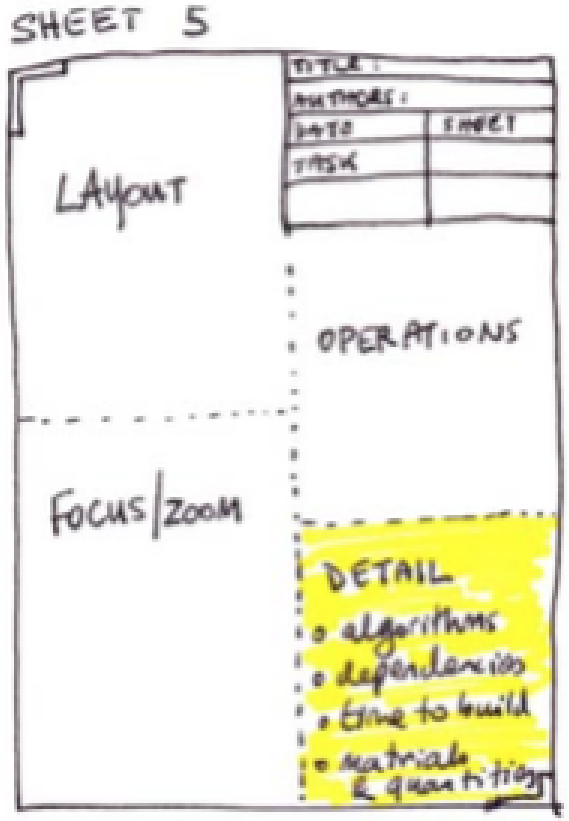
Sheets 2-4: The Alternative Designs

- Information: textual metadata
- Big Picture: principal layout, overview
- Functions: operations to be performed on them
- Focus: look & feel of the most important details, unique selling points, main difference to existing approaches
- Pros & Cons: balanced review

The 5 Design Sheet Methodology (https://doi.org/10.1007/978-3-319-55627-7)

ideate

HOW TO COME UP WITH VIS IDEAS?

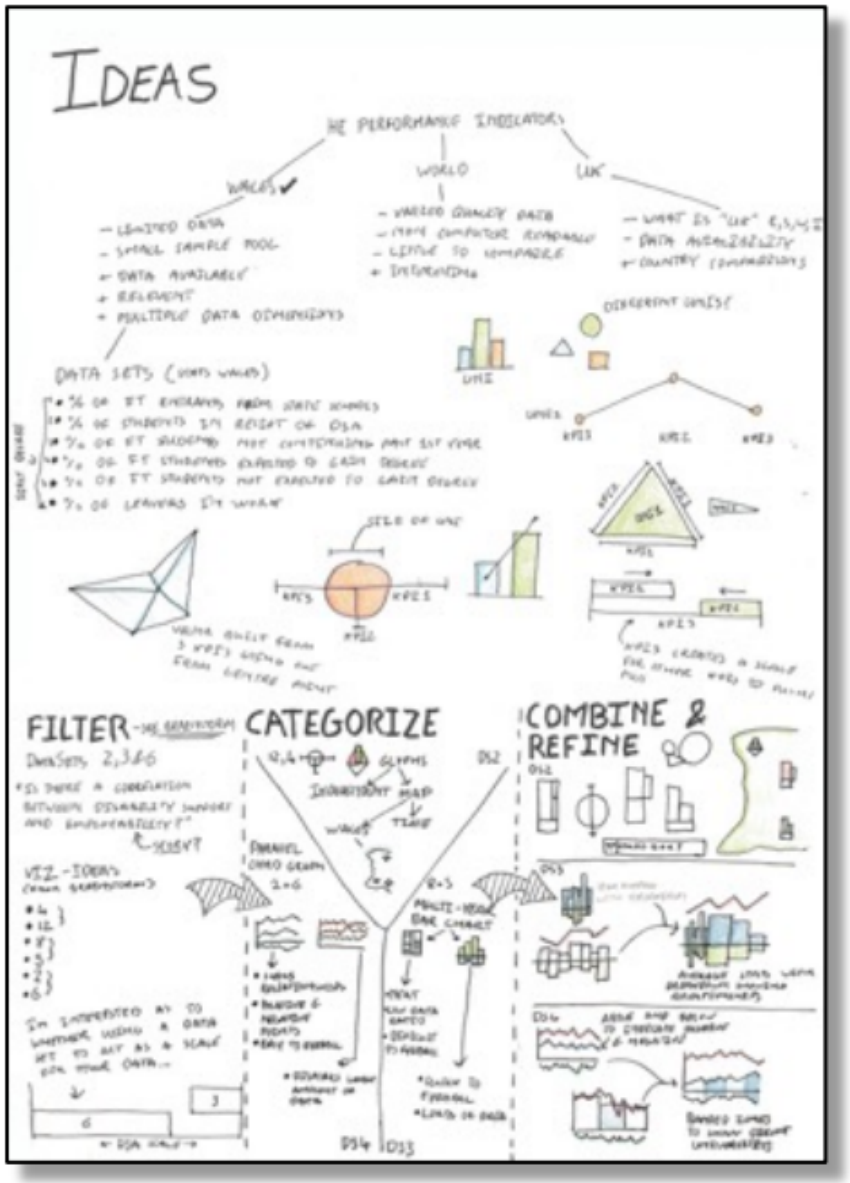


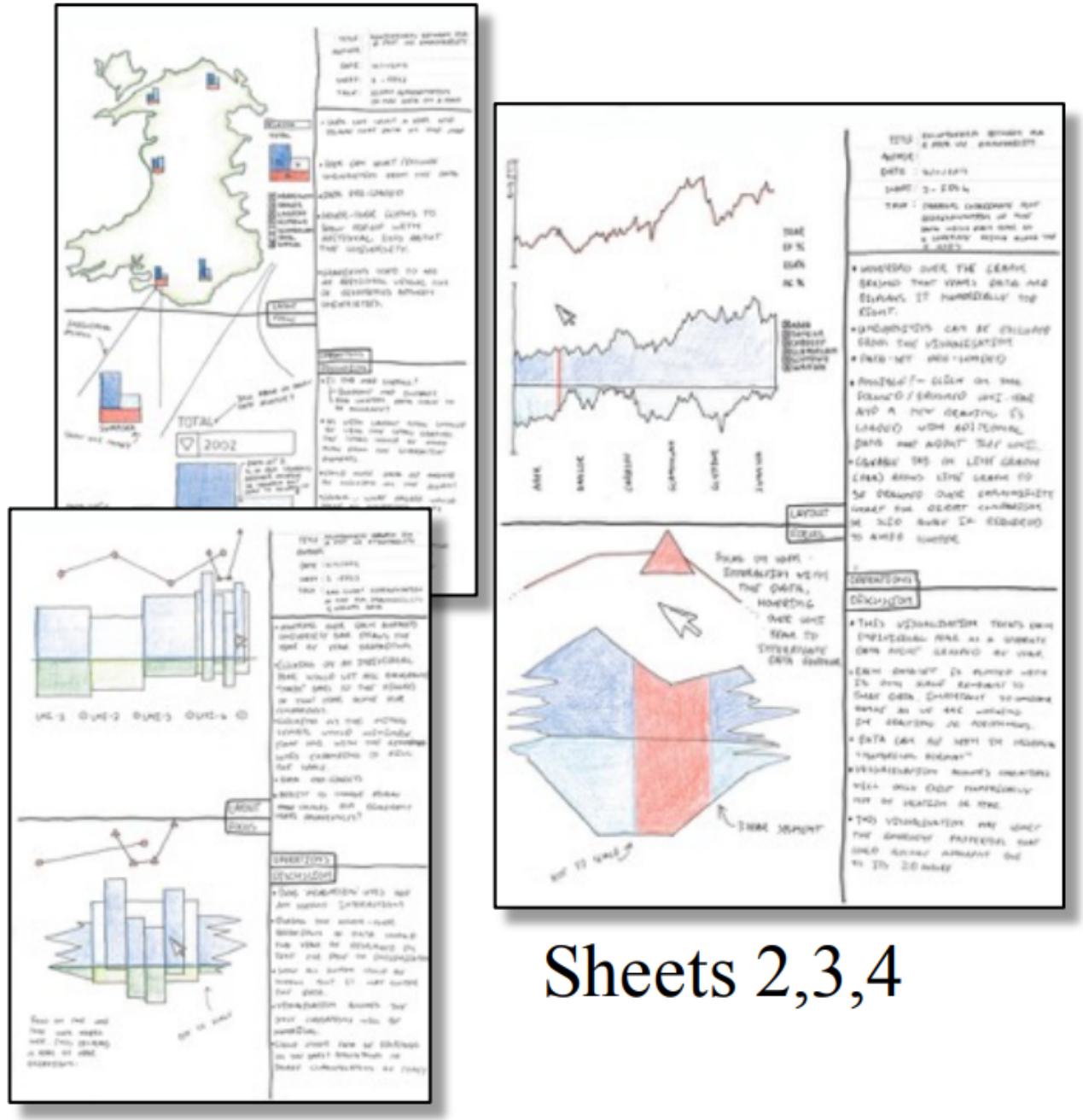
Sheet 5: The Design Realization

- Description of what algorithms are being used (e.g., perhaps citations or some critical maths)
- Any dependencies (e.g., software libraries or data formats that it must be compatible with)
- Estimates of cost, time, and person-months
- Specific requirements (e.g. hardware requirements, amount of pixels on a screen)

The 5 Design Sheet Methodology (https://doi.org/10.1007/978-3-319-55627-7)

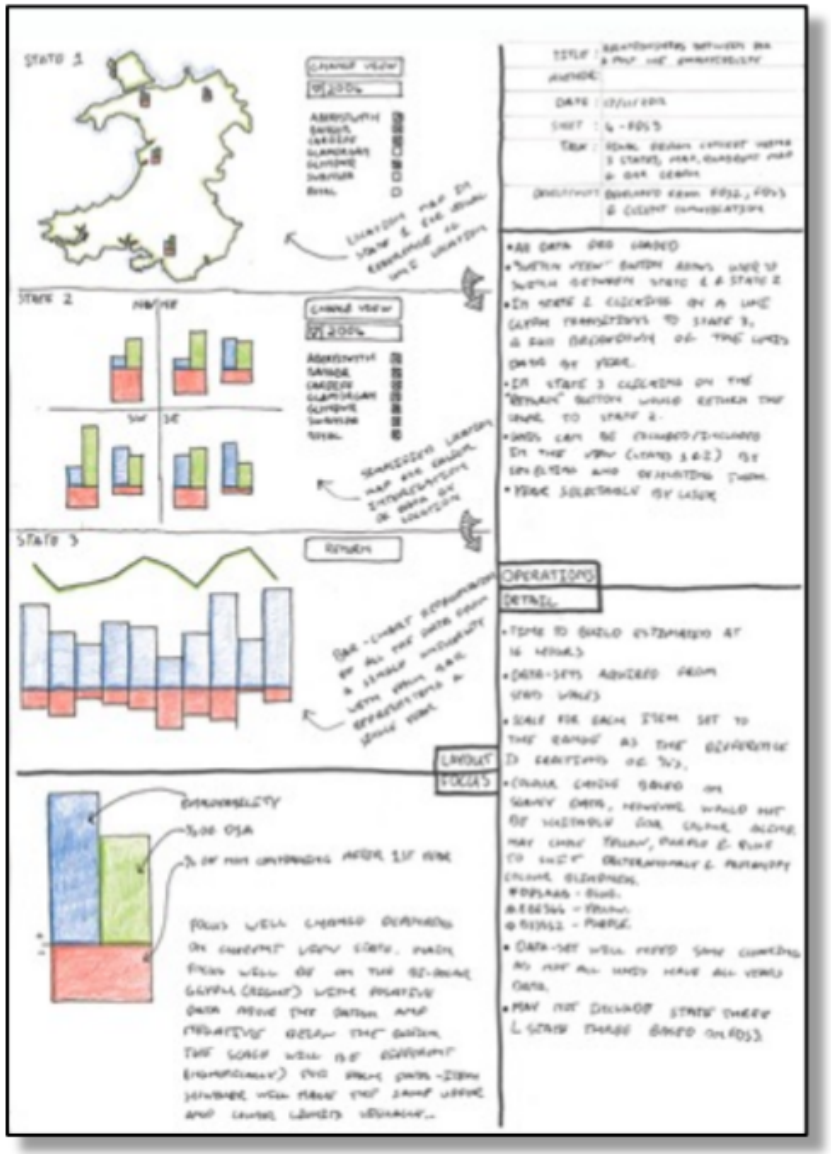
Dataset of degree entrants

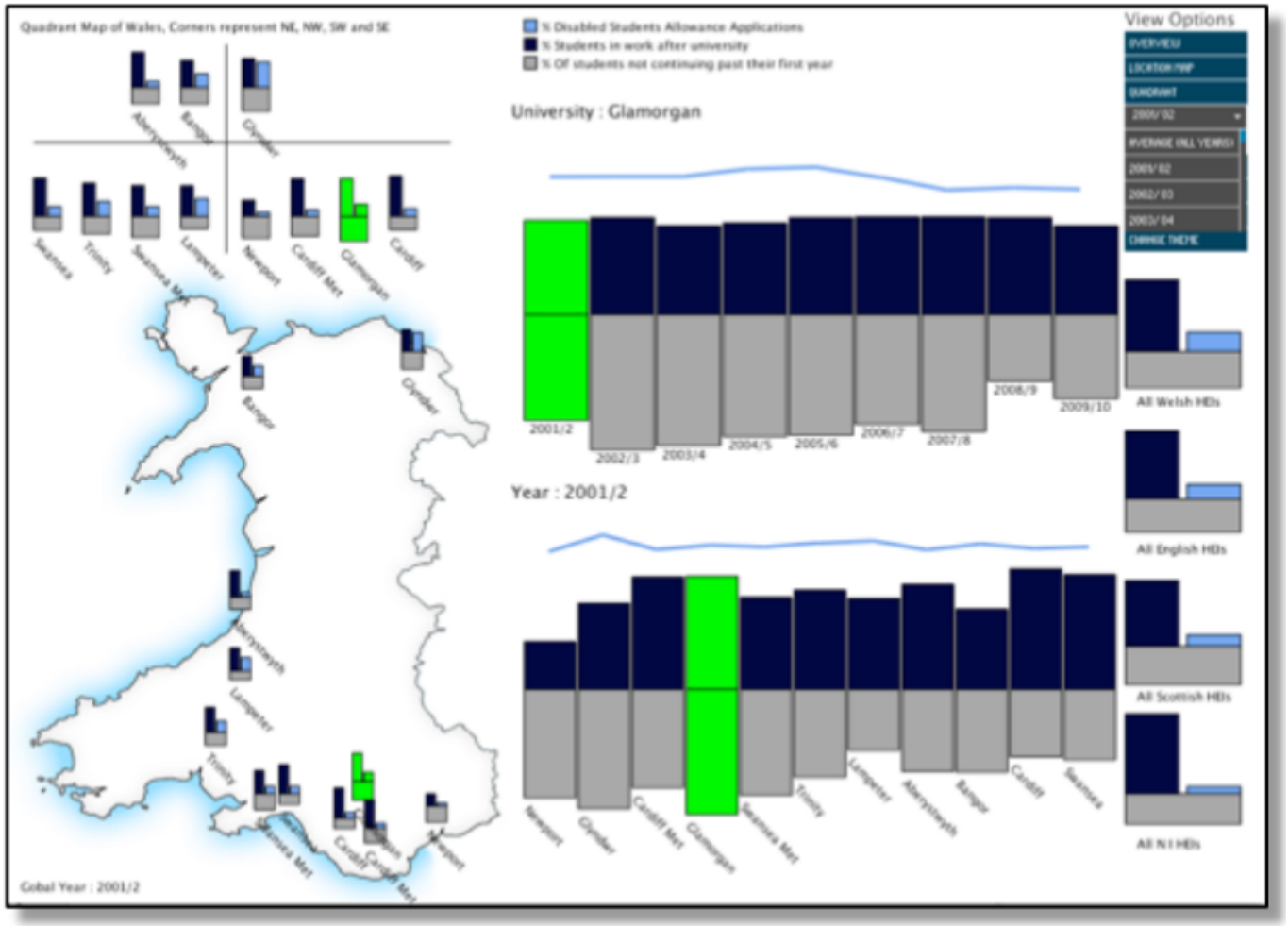




Sheets 2,3,4

Realization sheet





GUIDING DESIGN PRINCIPLES

- **Functional Design:**
  -> First create something useful, then make it more beautiful.

- **Deliberate Design:**
  -> Don’t include any visual feature without justification (purpose or reason).

- **Intuitive Design:**
  -> The visualization should not be more complex than the underlying data.

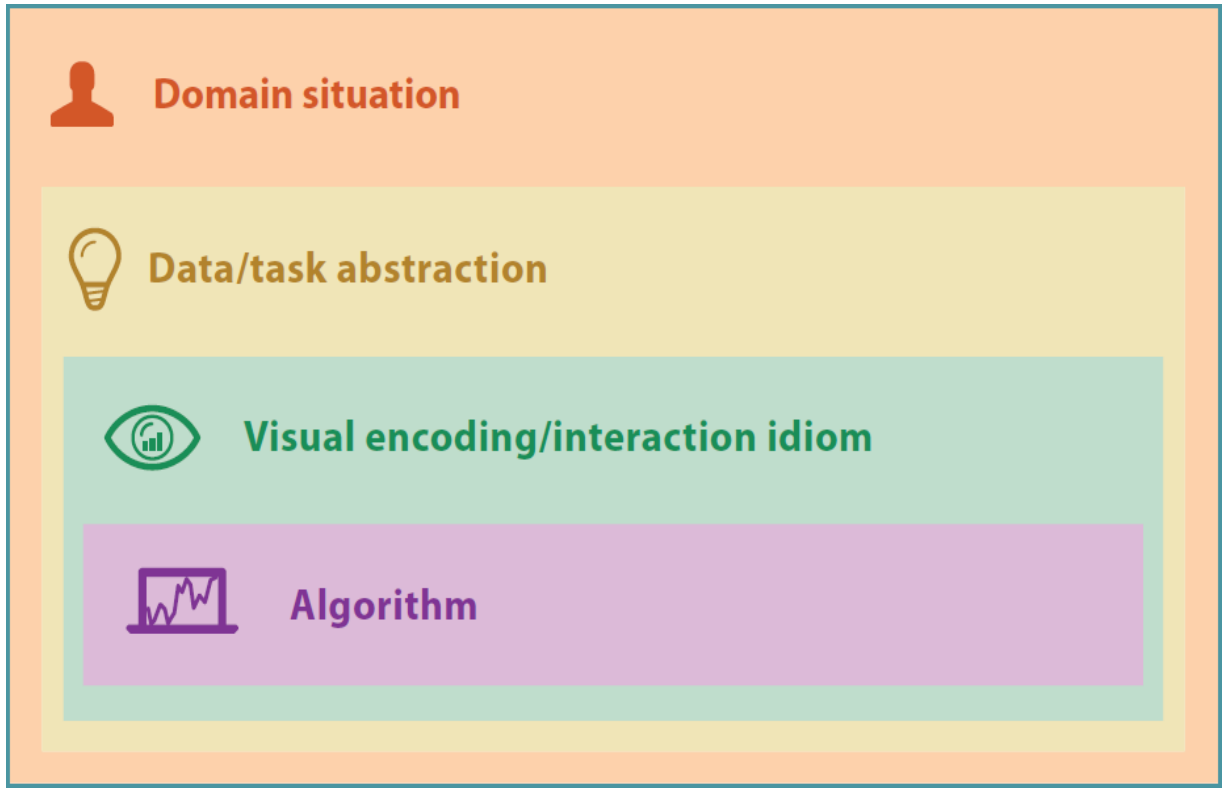
- **Ethical Design:**
  -> Do not visualize to deceive the reader/viewer.

EVALUATION OF VISUALIZATION: MUNZNER'S NESTED MODEL

AARHUS
UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

HANS-JÖRG SCHULZ
ASSOCIATE PROFESSOR

EVALUATION OF WHAT WITH WHAT



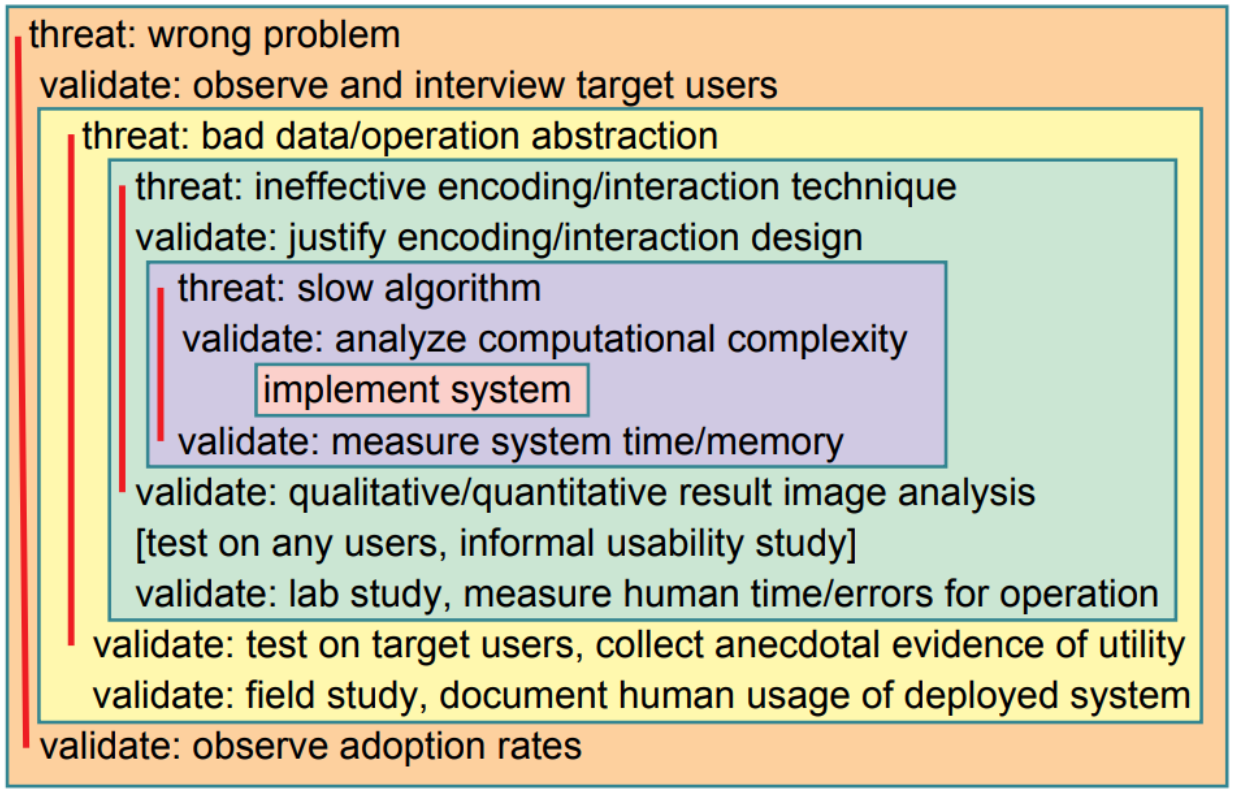


Image source: T.Munzner 2009

RESULT IMAGE ANALYSIS

For comparing the visualization outputs, rather than their runtimes,
use visualization metrics to quantify the quality of a visualization.

Pixel-based metrics for clutter quantification:

- overplotted% := percentage of pixels with more than 1 plotted point
- overcrowded% := percentage of plotted points that are in overplotted pixels

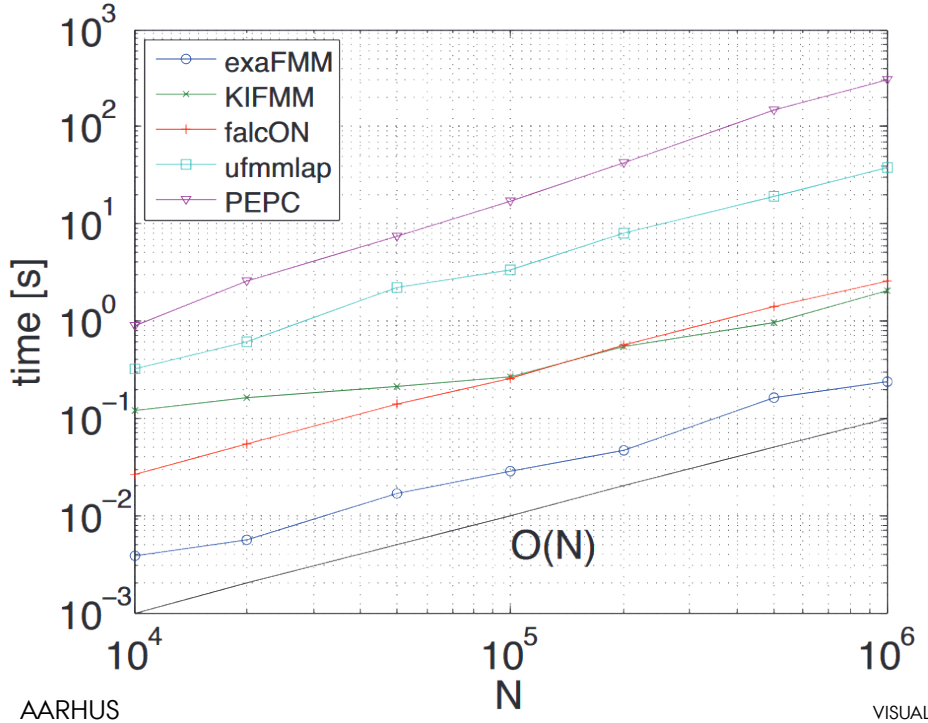
Graph-based metrics for layout optimization:

- Number of edge crossings
- Number of edge bends
- Minimum angle of crossing edges

BENCHMARKING

For comparing

- Different algorithms for the very same layout problem
- Different implementations of the very same algorithm
- Different # of CPUs/GPUs on which to run the very same implementation



One important condition:
Must be run on the same dataset!

Visual Analytics Benchmark Repository

A service of the SEMVAST Project
Managed by HCIL, University of Maryland

https://www.cs.umd.edu/hcil/varepository/benchmarks.php

LAB STUDY (AKA USER STUDY)

For comparing a novel visualization to a known prior for the same data / task characterization in terms of:

- Task completion time
- Error rates
- User satisfaction
  - User Experience Questionnaire: https://www.ueq-online.org
  - Computer System Usability Questionnaire: https://garyperlman.com/quest/quest.cgi
  - System Usability Scale: https://hell.meiert.org/core/pdf/sus.pdf

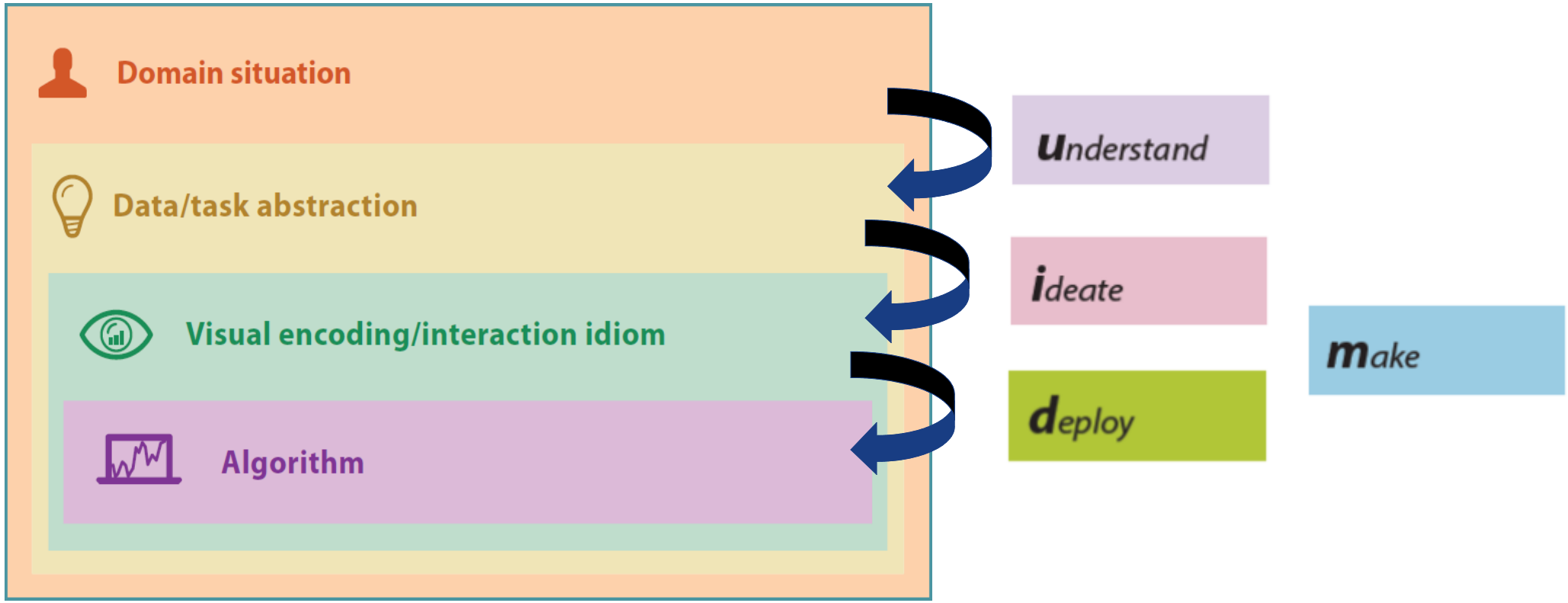
FIELD STUDY (AKA USE CASE)

For evaluating a novel visualization if there is no known prior for the same data / task characterization by showing the visualization's utility through

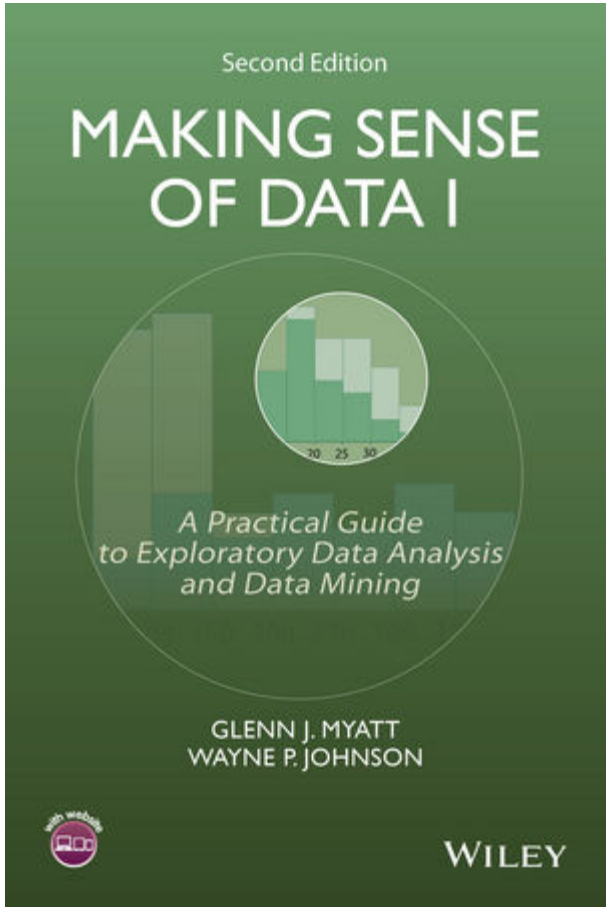
- finding / revealing a pattern in the data that could not have been found with another visualization
- showing the possibility and ease of integrating it in a pre-existing workflow that was not possible before

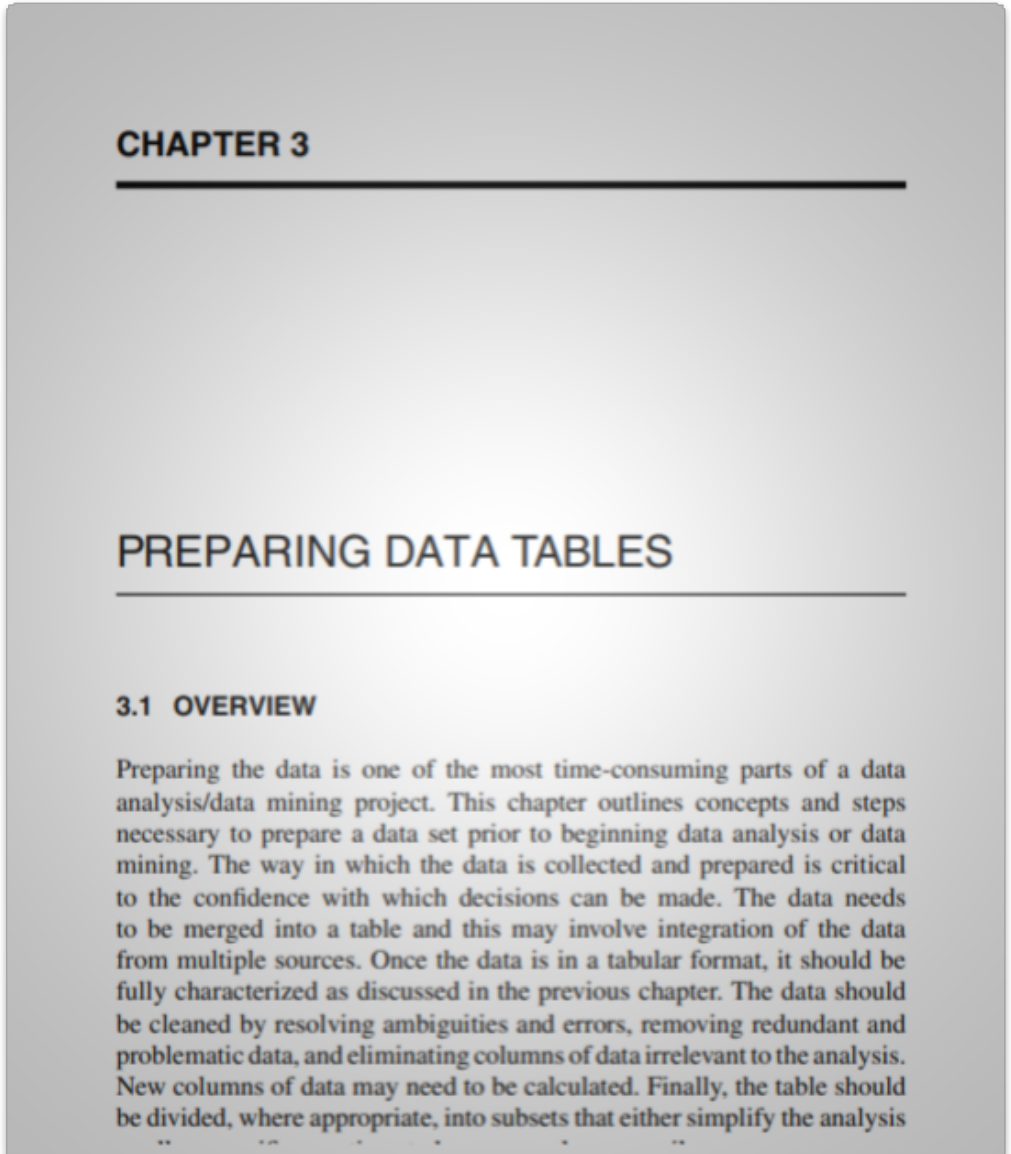
Note that the result is qualitative and often anecdotal evidence of utility!

MUNZNER'S MODEL VS. DESIGN ACTIVITY FRAMEWORK



WEEKLY READING





CHAPTER 3

PREPARING DATA TABLES

3.1 OVERVIEW

Preparing the data is one of the most time-consuming parts of a data analysis/data mining project. This chapter outlines concepts and steps necessary to prepare a data set prior to beginning data analysis or data mining. The way in which the data is collected and prepared is critical to the confidence with which decisions can be made. The data needs to be merged into a table and this may involve integration of the data from multiple sources. Once the data is in a tabular format, it should be fully characterized as discussed in the previous chapter. The data should be cleaned by resolving ambiguities and errors, removing redundant and problematic data, and eliminating columns of data irrelevant to the analysis. New columns of data may need to be calculated. Finally, the table should be divided, where appropriate, into subsets that either simplify the analysis

LIST OF LITERATURE SOURCES

- Min Cheng et al. 2009: https://doi.org/10.1109/MCG.2009.6
- Cui et al. 2019: https://doi.org/10.48550/arXiv.1906.10017
- Schulz et al. 2017: https://doi.org/10.1177/1473871616667767
- MacEachren 1992: https://doi.org/10.14714/CP13.1000
- Donald Shepard 1968: https://doi.org/10.1145/800186.810616
- Stevens 1946: https://doi.org/10.1126/science.103.2684.677
- Nielles-Vallespin et al. 2013: https://doi.org/10.1002/mrm.24488
- Gotz+Zhou 2008: https://doi.org/10.1109/VAST.2008.4677365
- Schulz et al. 2013: https://doi.org/10.1109/TVCG.2013.120
- Rind et al. 2016: https://doi.org/10.1177/1473871615621602
- Andrienko 2006: https://doi.org/10.1007/3-540-31190-4_3
- Gleicher 2018: https://doi.org/10.1109/TVCG.2017.2744199
- McKenna et al. 2014: https://doi.org/10.1109/TVCG.2014.2346331
  Website with worksheets: https://mckennapsean.com/projects/design-activity-framework/
- The 5 Design Sheet Methodology: https://doi.org/10.1007/978-3-319-55627-7
  Website with worksheets: http://fds.design
- T.Munzner 2009: https://doi.org/10.1109/TVCG.2009.111


SOLIDUM PETIT IN PROFUNDIS
UNIVERSITAS ARHUSIENSIS