

Matthew Brehmer



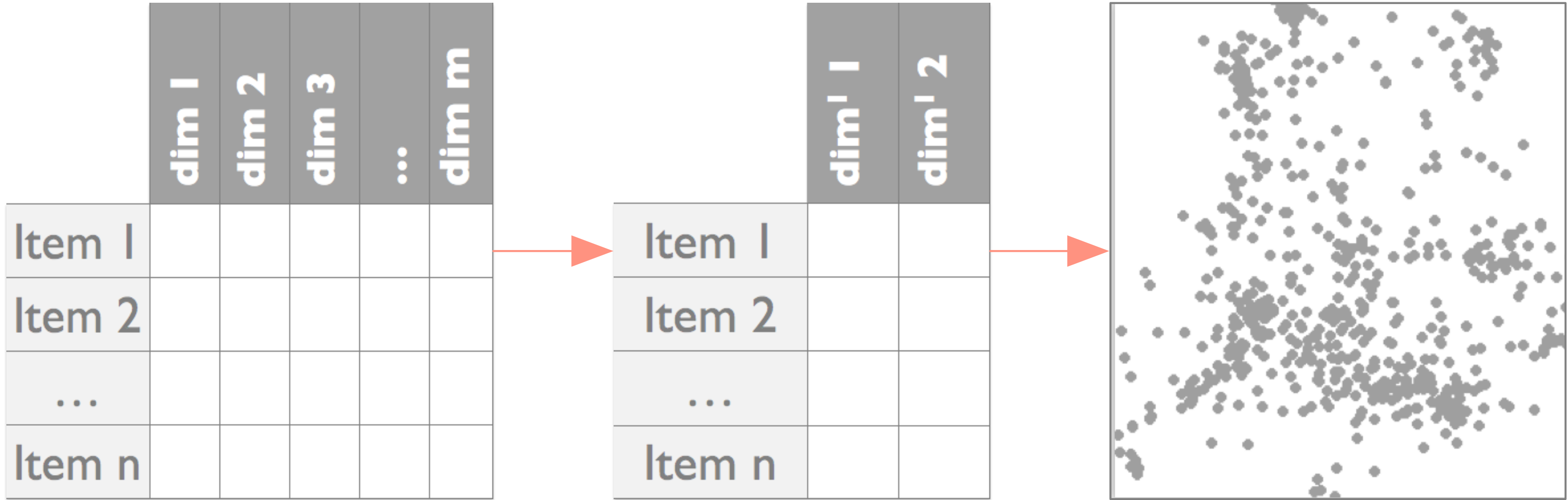
Michael Sedlmair



Stephen Ingram



Tamara Munzner



# Visualizing Dimensionally-Reduced Data: Interviews with Analysts and a Characterization of Task Sequences



Matthew Brehmer



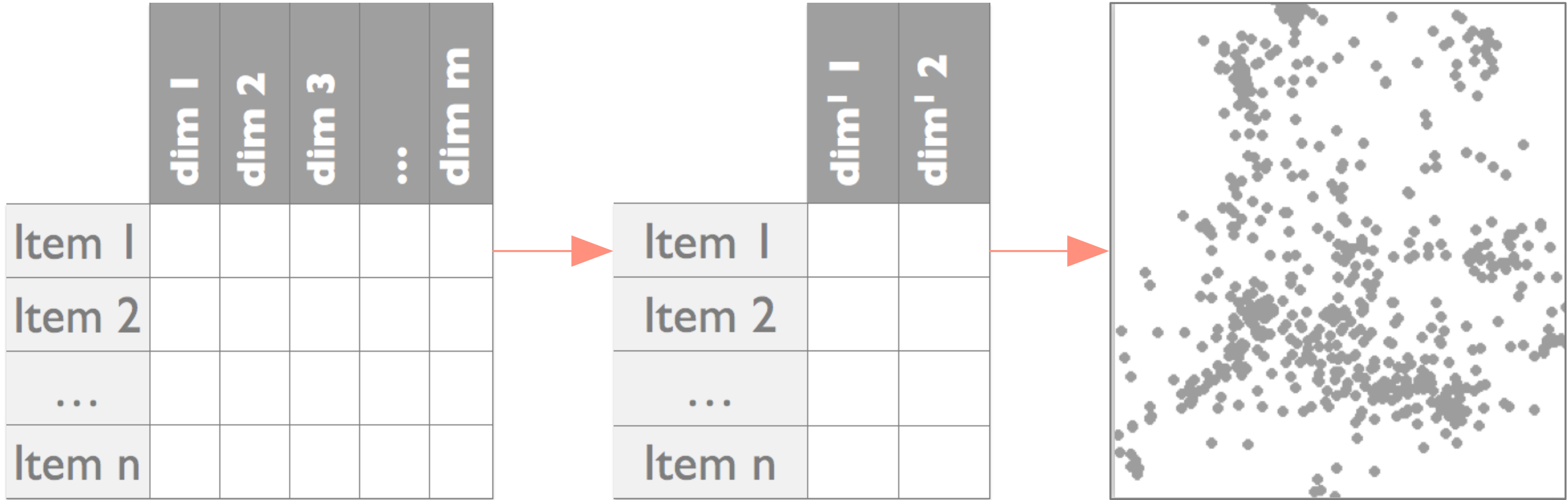
Michael Sedlmair



Stephen Ingram



Tamara Munzner



**Visualizing Dimensionally-Reduced Data:**  
 Interviews with Analysts and a  
 Characterization of Task Sequences



**A need for abstract task characterization...**

**...yet specific to **data type****

# AN ABUNDANCE OF VIS TASK CHARACTERIZATION

- Amar & Stasko (2004)  
Amar, Eagan, & Stasko (2005)  
Andrienko & Andrienko (2006)  
Brehmer & Munzner (2013)  
Buja et al. (1996)  
Card, Mackinlay, Shneiderman (1999)  
Casner (1991)  
Chi & Riedl (1998)  
Chuah & Roth (1996)  
Dix & Ellis (1998)  
Gotz & Zhou (2008)  
Heer & Shneiderman (2012)  
Keim (2002)  
Klein, Moon, & Hoffman (2006)  
Liu & Stasko (2010)  
Mullins & Treu (1993)  
Pike, Stasko, et al. (2009)  
Pirolli & Card (2005)  
Schulz et al. (2013)  
Spence (2007)  
Springmeyer et al. (1992)  
RE Roth (2013)  
Roth & Mattis (1990)  
Shneiderman (1996)  
Tweedie (1997)  
Valiati et al. (2006)  
Ward & Yang (2004)  
Wehrend & Lewis (1990)  
Yi, Stasko, et al. (2007)  
Zhou & Feiner (1998)

# DATA-TYPE SPECIFIC TASK CHARACTERIZATION

## Vis Tasks for **1D, 2D, 3D, Multi-Dim, Temporal, Tree, & Network Data**

Shneiderman. (1996) IEEE Symp. Visual Languages

## Vis Tasks for **Tabular Data**

Henry & Fekete. (2006) ACM **BELIV** Workshop

## Vis Tasks for **Graph Data**

Lee et al. (2006) ACM **BELIV** Workshop

## Vis Tasks for **Time-Oriented Data**

Lammarsch et al. (2012) EuroVA Workshop

# DATA-TYPE SPECIFIC TASK CHARACTERIZATION

## Vis Tasks for **1D, 2D, 3D, Multi-Dim, Temporal, Tree, & Network Data**

Shneiderman. (1996) IEEE Symp. Visual Languages

## Vis Tasks for **Tabular Data**

Henry & Fekete. (2006) ACM **BELIV** Workshop

## Vis Tasks for **Graph Data**

Lee et al. (2006) ACM **BELIV** Workshop

## Vis Tasks for **Time-Oriented Data**

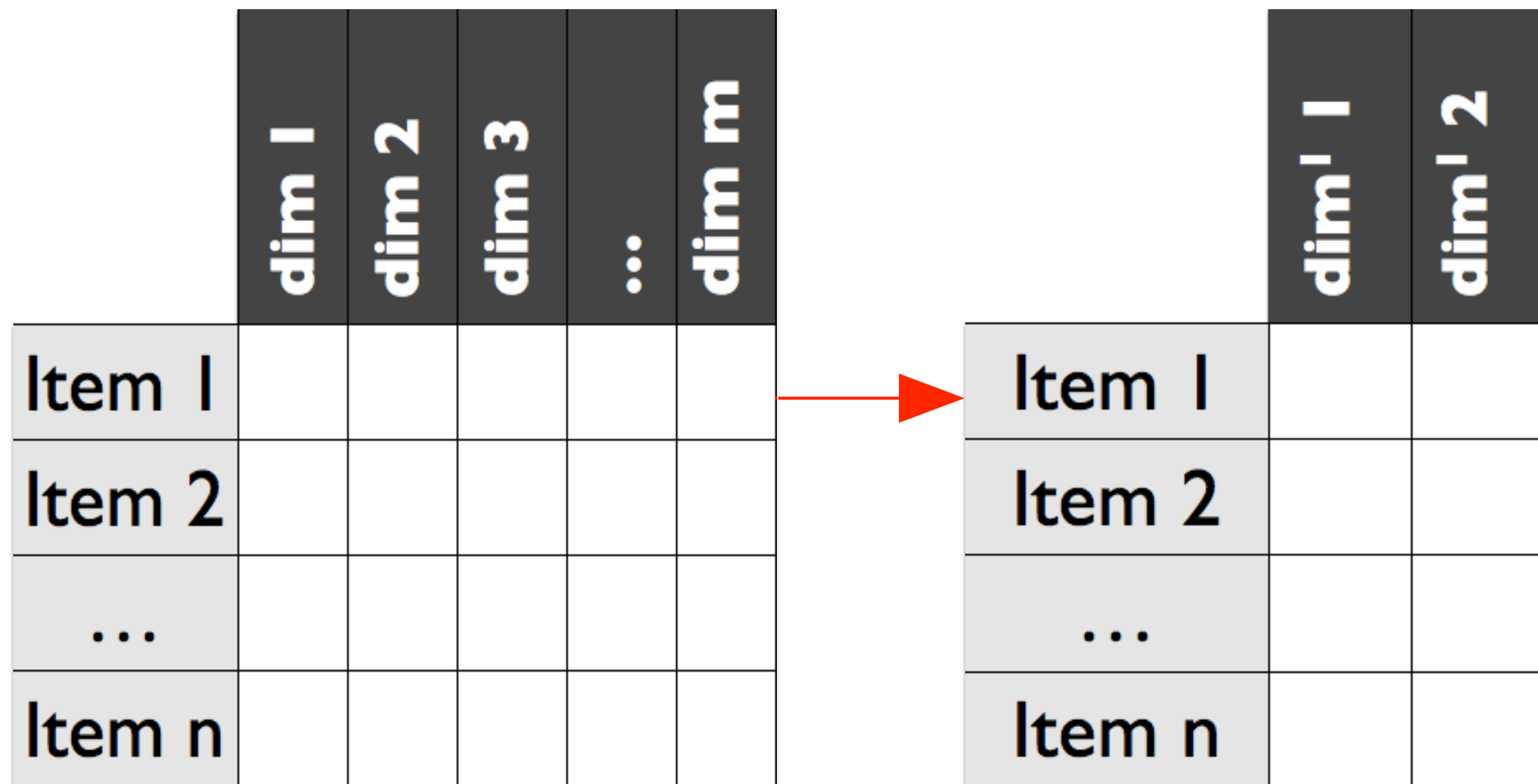
Lammarsch et al. (2012) EuroVA Workshop

...what about **DR data?**

# DIMENSIONALITY REDUCTION (E.G. PCA, MDS) & VIS

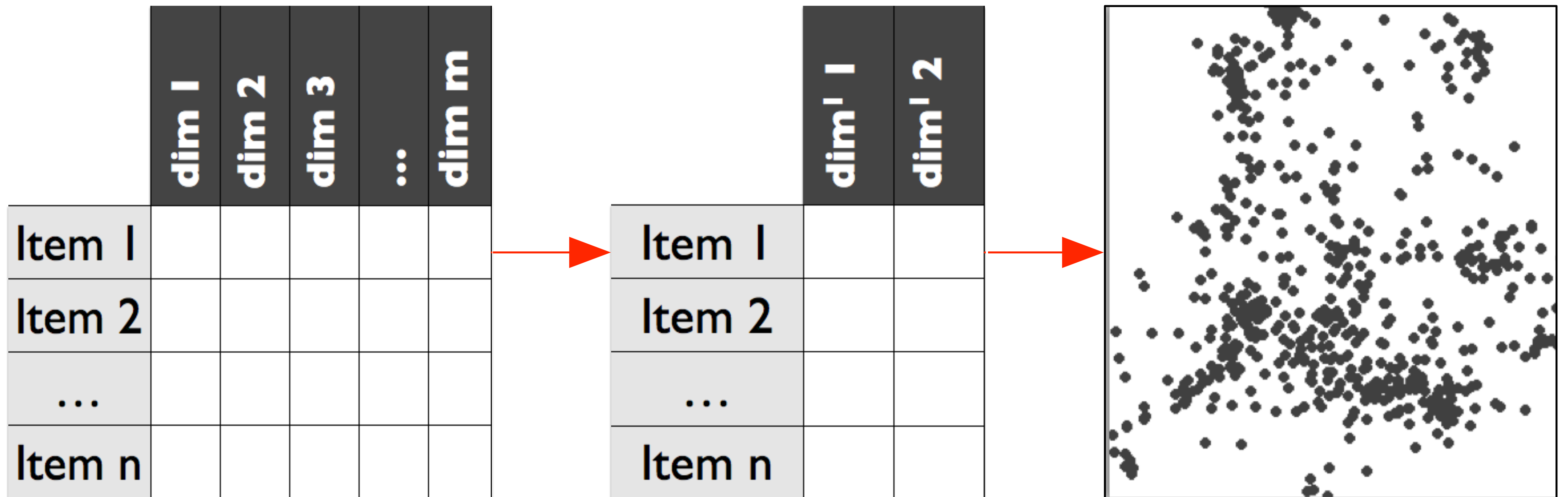
	dim 1	dim 2	dim 3	...	dim m
Item 1					
Item 2					
...					
Item n					

# DIMENSIONALITY REDUCTION (E.G. PCA, MDS) & VIS



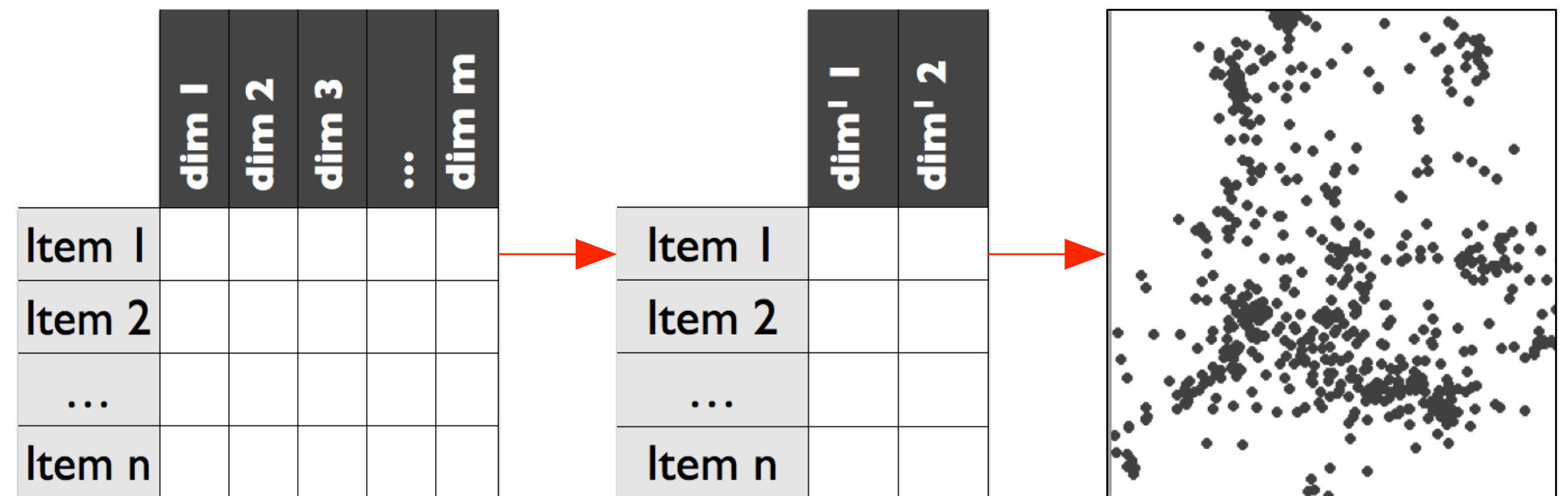


# DIMENSIONALITY REDUCTION (E.G. PCA, MDS) & VIS

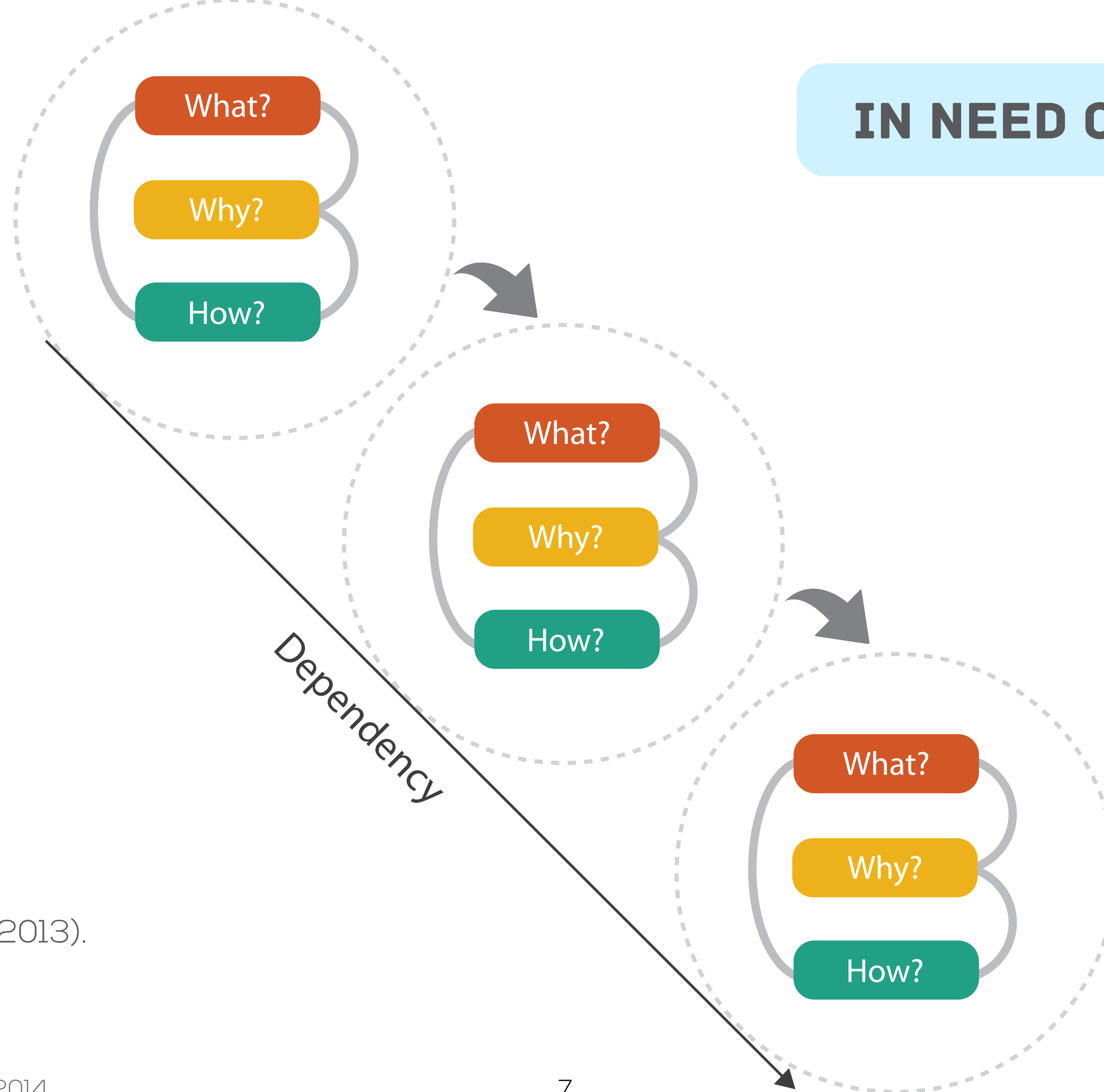


# 10 ANALYST INTERVIEWEES, 6 DOMAINS

Human computer interaction (x3)  
Bioinformatics (x3)  
Policy analysis  
Computational chemistry  
Social network analysis  
Investigative journalism



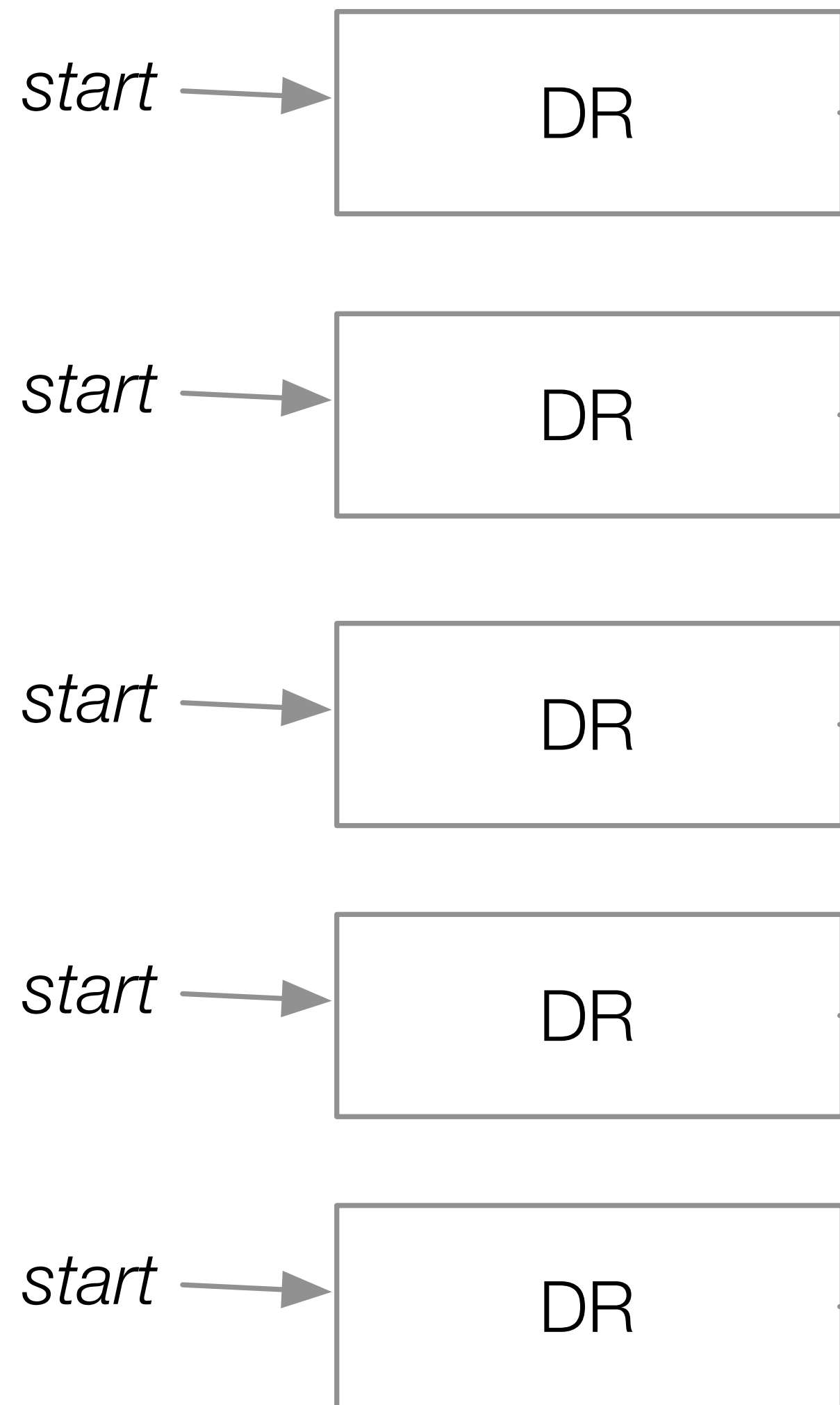
# IN NEED OF A FRAMEWORK



Brehmer & Munzner.  
IEEE TVCG / Proc. InfoVis (2013).

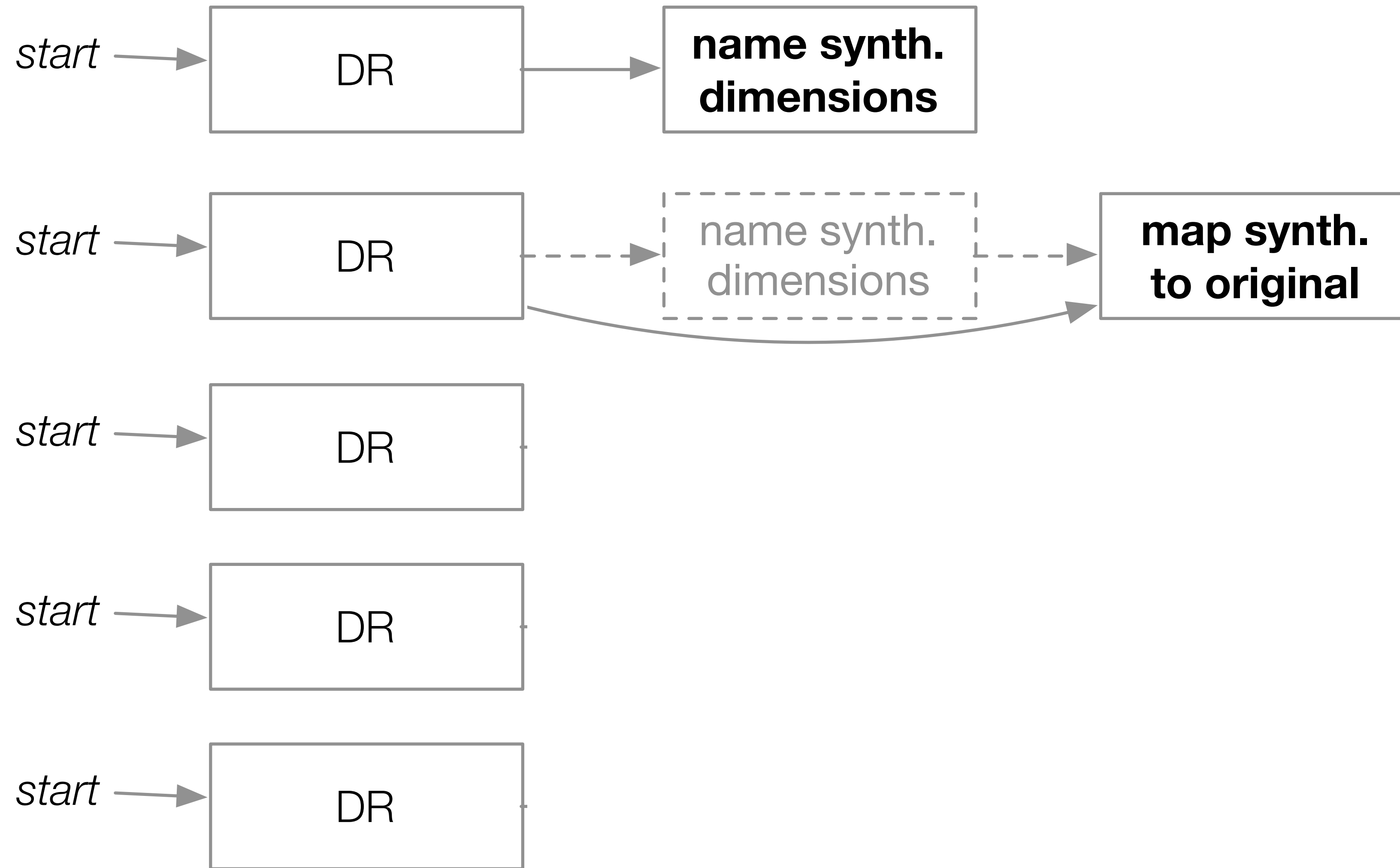
Munzner (2014)

# CONTRIBUTION: VIS TASK SEQUENCES FOR DR DATA



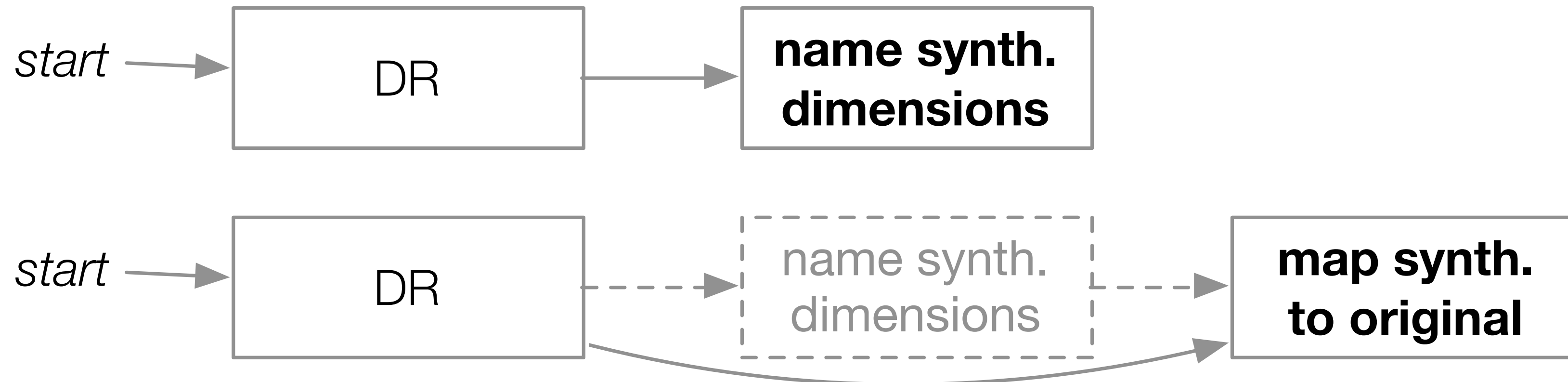
# CONTRIBUTION: VIS TASK SEQUENCES FOR DR DATA

2 dimension-oriented sequences

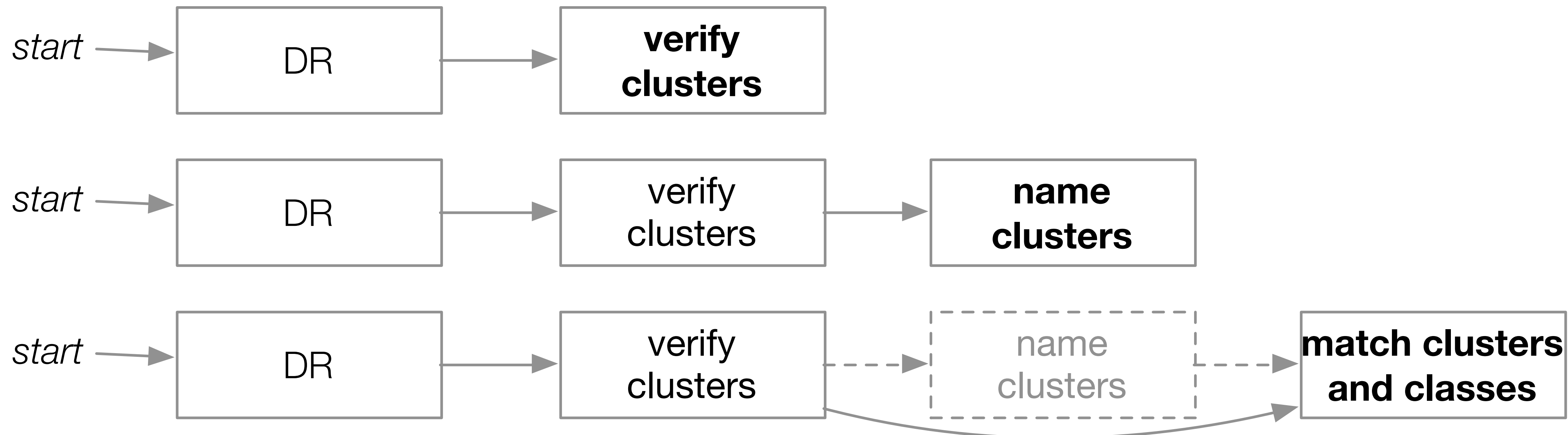


# CONTRIBUTION: VIS TASK SEQUENCES FOR DR DATA

2 dimension-oriented sequences



3 cluster-oriented sequences



# IMPLICATIONS FOR VIS EVALUATION

understanding work practices  
evaluating visual data analysis & reasoning  
evaluating communication through vis  
evaluating collaborative data analysis  
evaluating user performance  
evaluating user experience  
evaluating vis algorithms

“Seven Scenarios”: Lam et al. IEEE TVCG 2012.

## Empirical Studies in Information Visualization: Seven Scenarios

Heidi Lam, Enrico Bertini, Petra Isenberg, Catherine Plaisant, and Sheelagh Carpendale

**Abstract**—We take a new, scenario-based look at evaluation in information visualization. Our seven scenarios, evaluating visual data analysis and reasoning, evaluating user performance, evaluating user experience, evaluating environments and work practices, evaluating communication through visualization, evaluating visualization algorithms, and evaluating collaborative data analysis were derived through an extensive literature review of over 800 visualization publications. These scenarios distinguish different study goals and types of research questions and are illustrated through example studies. Through this broad survey and the distillation of these scenarios, we make two contributions. One, we encapsulate the current practices in the information visualization research community and, two, we provide a different approach to reaching decisions about what might be the most effective evaluation of a given information visualization. Scenarios can be used to choose appropriate research questions and goals and the provided examples can be consulted for guidance on how to design one's own study.

**Index Terms**—Information visualization, evaluation.

### 1 INTRODUCTION

EVALUATION in information visualization is complex since, for a thorough understanding of a tool, it not only involves assessing the visualizations themselves, but also the complex processes that a tool is meant to support. Examples of such processes are exploratory data analysis and reasoning, communication through visualization, or collaborative data analysis. Researchers and practitioners in the field have long identified many of the challenges faced when planning, conducting, and executing an evaluation of a visualization tool or system [10], [41], [54], [63]. It can be daunting for evaluators to identify the right evaluation questions to ask, to choose the right variables to evaluate, to pick the right tasks, users, or data sets to test, and to pick appropriate evaluation methods. Literature guidelines exist that can help with these problems but they are almost exclusively focused on methods—“structured as an enumeration of methods with focus on *how* to carry them out, without prescriptive advice for *when* to choose between them.” ([54, p.1], author's own emphasis).

This paper takes a different approach: instead of focusing on evaluation methods, we provide an in-depth

discussion of evaluation scenarios, categorized into those for understanding data analysis processes and those which evaluate visualizations themselves.

The scenarios for understanding data analysis are

- Understanding environments and work practices (UWP),
- evaluating visual data analysis and reasoning (VDAR),
- evaluating communication through visualization (CTV), and
- evaluating collaborative data analysis (CDA).

The scenarios for understanding visualizations are

- Evaluating user performance (UP),
- evaluating user experience (UE), and
- evaluating visualization algorithms (VA).

Our goal is to provide an overview of different types of evaluation scenarios and to help practitioners in setting the right evaluation goals, picking the right questions to ask, and to consider a variety of methodological alternatives to evaluation for the chosen goals and questions. Our scenarios were derived from a systematic analysis of 850 papers (361 with evaluation) from the information visualization research literature (Section 5). For each evaluation scenario, we list the most common evaluation goals and outputs, evaluation questions, and common approaches in Section 6. We illustrate each scenario with representative published evaluation examples from the information visualization community. In cases where there are gaps in our community's evaluation approaches, we suggest examples from other fields. We strive to provide a wide coverage of the methodology space in our scenarios to offer a diverse set of evaluation options. Yet, the “Methods and Examples” lists in this paper are not meant to be comprehensive as our focus is on choosing among evaluation scenarios. Instead, we direct the interested reader

- H. Lam is with Google, Inc, Mountain View, CA. E-mail: heidi.lam@gmail.com.
- E. Bertini is with the Department of Computer and Information Science, University of Konstanz, Box 78, Konstanz 78457, Germany. E-mail: enrico.bertini@uni-konstanz.de.
- P. Isenberg is with INRIA, Université Paris-Sud, Team Aviz, Bat 650, Saclay, Orsay Cedex 91405, France. E-mail: petra.isenberg@inria.fr.
- C. Plaisant is with the University of Maryland, 2117C Hornbake South Wing, College Park, MD 20742. E-mail: plaisant@cs.umd.edu.
- S. Carpendale is with the Department of Computer Science, University of Calgary, 2500 University Dr. NW, Calgary, AB T2N 1N4, Canada. E-mail: sheelagh@ucalgary.ca.

Manuscript received 8 Sept. 2010; revised 6 Nov. 2011; accepted 9 Nov. 2011; published online 30 Nov. 2011.

Recommended for acceptance by C. North.

For information on obtaining reprints of this article, please send e-mail to: [tcvg@computer.org](mailto:tcvg@computer.org), and reference IEEECS Log Number TVCG-2010-09-0224. Digital Object Identifier no. 10.1109/TVCG.2011.279.

1077-2626/12/\$31.00 © 2012 IEEE

Published by the IEEE Computer Society

# CONCLUSION

**Domain-Agnostic** and **Data-Type-Specific**  
Task Characterization

**Tasks in Sequence**, not in Isolation

Tasks Characterization for **BELIV**

[cs.ubc.ca/labs/imager/tr/2014/DRVisTasks/](http://cs.ubc.ca/labs/imager/tr/2014/DRVisTasks/)

**thanks:** UBC InfoVis group. UBC Multimodal User Experience group



Matthew Brehmer



Michael Sedlmair



Stephen Ingram



Tamara Munzner





# DR Vis Tasks: Supplemental

# IMPLICATIONS FOR VIS EVALUATION

Lam et al. IEEE TVCG 2012.



# IMPLICATIONS FOR VIS EVALUATION

Lam et al. IEEE TVCG 2012.

**understanding  
work practices**

pre-design  
**requirements  
analysis**, especially  
in problem-driven  
**design studies**; use  
task sequences as  
code set

# IMPLICATIONS FOR VIS EVALUATION

Lam et al. IEEE TVCG 2012.

## understanding work practices

pre-design  
**requirements analysis**, especially in problem-driven  
**design studies**; use task sequences as code set

## evaluating user performance

task sequences can inform **experimental design** and participant instructions when evaluating techniques combining DR + Vis

# IMPLICATIONS FOR VIS EVALUATION

Lam et al. IEEE TVCG 2012.

understanding work practices	evaluating user performance	evaluating user experience
pre-design <b>requirements analysis</b> , especially in problem-driven <b>design studies</b> ; use task sequences as code set	task sequences can inform <b>experimental design</b> and participant instructions when evaluating techniques combining DR + Vis	inform participant instructions in a think-aloud evaluation <b>focus</b> questionnaire <b>questions...</b>



# IMPLICATIONS FOR VIS EVALUATION

Lam et al. IEEE TVCG 2012.

understanding work practices	evaluating user performance	evaluating user experience	evaluating visual data analysis & reasoning
pre-design <b>requirements analysis</b> , especially in problem-driven <b>design studies</b> ; use task sequences as code set	task sequences can inform <b>experimental design</b> and participant instructions when evaluating techniques combining DR + Vis	inform participant instructions in a think-aloud evaluation <b>focus questionnaire questions...</b>	analyze the use of deployed DR + Vis tools <i>in the wild</i> ; use task sequences as code ... <b>focus diary / interview questions</b>



## 10 ANALYST INTERVIEWEES, 6 DOMAINS

Human computer interaction (x3)  
Bioinformatics (x3)  
Policy analysis  
Computational chemistry  
Social network analysis  
Investigative journalism

# 10 ANALYST INTERVIEWEES, 6 DOMAINS

Human computer interaction (x3)  
Bioinformatics (x3)  
Policy analysis  
Computational chemistry  
Social network analysis  
Investigative journalism

+ 4 known  
use cases  
from  
DR + Vis  
technique  
papers

## Morphable model of quadrupeds skeletons for animating 3D animals

Lionel Reveret, Laurent Favreau, Christine Depraz, Marie-Paule Cani

GRAVIR, INRIA

## Visualization Methodology for Multidimensional Scaling

ANDREAS BUJA<sup>1</sup> and DEBORAH F. SWAYNE<sup>2</sup>

March 30, 2004

We discuss the application of interactive visualization techniques to multidimensional scaling (MDS). MDS in its conventional batch implementations is prone to uncertainties with regard to the choice of the number of dimensions. These uncertainties are addressed by a system, called the "XGobi", used here for

These uncertainties are addressed by a system, called the "XGobi", used here for

### 1 Introduction

We describe methodology for interactive visualization techniques to multidimensional scaling (MDS). MDS in its conventional batch implementations is prone to uncertainties with regard to the choice of the number of dimensions. These uncertainties are addressed by a system, called the "XGobi", used here for

<sup>1</sup>Andreas Buja is at AT&T, Park, NJ 07932-0971.  
<sup>2</sup>Deborah F. Swayne is at AT&T, Florham Park, NJ 07932-0971.

### A Data-Driven Reflectance Model

Wojciech Matusik<sup>\*</sup> Hanspeter Pfister<sup>†</sup> Matt Brand<sup>‡</sup> Leonard McMillan<sup>§</sup>

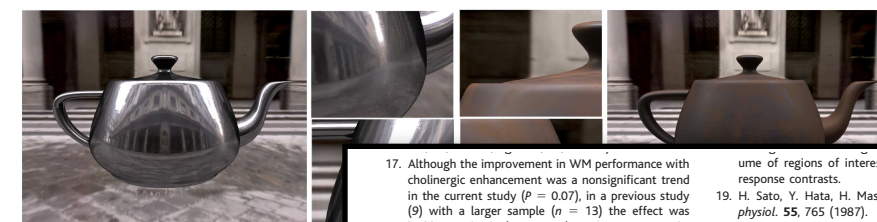


Figure 1: Renditions of materials generated using the data-driven reflectance model. Closeup pictures in the center. We used a spatially varying reflectance model.

### Abstract

We present a generative model for isotropic bidirectional reflectance distribution functions (BRDFs) based on reflectance data. Instead of using analytical reflectance models, we represent each BRDF as a dense set of measurements. We use these measurements to create new BRDFs. We treat each BRDF as a single high-dimensional vector taken from a space of possible BRDFs. We apply both linear (subspace) and nonlinear dimensionality reduction tools in an effort to find a low-dimensional representation that characterizes our data. We let users define perceptually meaningful parameters to navigate in the reduced-dimension BRDF space. This allows us to interpolate and extrapolate in the space of possible BRDFs. We apply both linear (subspace) and nonlinear dimensionality reduction tools in an effort to find a low-dimensional representation that characterizes our data. We let users define perceptually meaningful parameters to navigate in the reduced-dimension BRDF space. This allows us to interpolate and extrapolate in the space of possible BRDFs.

**Keywords:** Light Reflection Models, Photometric Model, Reflectance, BRDF, Image-based Modeling

### 1 Introduction

A fundamental problem of computer graphics rendering is how light is reflected from surfaces. A class of functions that describe the reflectance of a surface is the bidirectional reflectance distribution function (BRDF). The BRDF is a function of the incident and outgoing directions of light and the surface normal. The BRDF is a function of the incident and outgoing directions of light and the surface normal. The BRDF is a function of the incident and outgoing directions of light and the surface normal.

<sup>\*</sup>MIT, Cambridge, MA.  
Email: wjcm@graphics.cs.mit.edu  
<sup>†</sup>MERL, Cambridge, MA.  
Email: [pfister,brand]@merl.com  
<sup>‡</sup>UNC, Chapel Hill, NC.  
Email: mcmillan@cs.unc.edu

The goal of this work is to design a morphable model of quadruped skeletons for animating 3D animals. This model enables the automatic generation of skeletons from a few simple measurements performed on real animals. Our approach for constructing the morphable skeletons designed by an expert animator. This raises the question of how to represent the skeletons and the use of a quaternion representation for rotations and the use of a quaternion representation for rotations and the use of a quaternion representation for rotations.

Our approach for constructing the morphable skeletons designed by an expert animator. This raises the question of how to represent the skeletons and the use of a quaternion representation for rotations and the use of a quaternion representation for rotations.

Categories and Subject Descriptors: I.3.7 [Computer Graphics]: Animation

Models of face 3D shapes and accurate laser scans of faces. We use these models to generate intuitive parameters for animating faces. We use these models to generate intuitive parameters for animating faces.

We investigate the generation of skeletons for animating 3D animals. We investigate the generation of skeletons for animating 3D animals.

## A Global Geometric Framework for Nonlinear Dimensionality Reduction

Joshua B. Tenenbaum,<sup>1\*</sup> Vin de Silva,<sup>2</sup> John C. Langford<sup>3</sup>

Scientists working with large volumes of high-dimensional data, such as global climate patterns, stellar spectra, or human gene distributions, regularly confront the problem of dimensionality reduction: finding meaningful low-dimensional structures hidden in their high-dimensional observations. The human brain confronts the same problem in everyday perception, extracting from its high-dimensional sensory inputs—30,000 auditory nerve fibers or 10<sup>6</sup> optic nerve fibers—a manageable small number of perceptually relevant features. Here we describe an approach to solving dimensionality reduction problems that uses easily measured local metric information to learn the underlying global geometry of a data set. Unlike classical techniques such as principal component analysis (PCA) and multidimensional scaling (MDS), our approach is capable of discovering the nonlinear degrees of freedom that underlie complex natural observations, such as human handwriting or images of a face under different viewing conditions. In contrast to previous algorithms for nonlinear dimensionality reduction, ours efficiently computes a globally optimal solution, and, for an important class of data manifolds, is guaranteed to converge asymptotically to the true structure.

A canonical problem in dimensionality reduction from the domain of visual perception is illustrated in Fig. 1A. The input consists of many images of a person's face observed under different pose and lighting conditions, in no particular order. These images can be thought of as points in a high-dimensional vector space, with each input dimension corresponding to the brightness of one pixel in the image or the firing rate of one retinal ganglion cell. Although the input dimensionality may be quite high (e.g., 4096 for these 64 pixel by 64 pixel images), the perceptually meaningful structure of these images has many fewer independent degrees of freedom. Within the 4096-dimensional input space, all of the images lie on an intrinsically three-dimensional manifold, or constraint surface, that can be parameterized by two pose variables plus an azimuthal lighting angle. Our goal is to discover, given only the unordered high-dimensional inputs, low-dimensional representations such as Fig. 1A with coordinates that capture the intrinsic degrees of freedom of a data set. This problem is of central importance not only in studies of vision (1–5), but also in speech (6, 7), motor control (8, 9), and a range of other physical and biological systems (10, 11).

<sup>1</sup>Department of Psychology and <sup>2</sup>Department of Mathematics, Stanford University, Stanford, CA 94305, USA. <sup>3</sup>Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15217, USA.  
<sup>\*</sup>To whom correspondence should be addressed. E-mail: jtenenb@leland.stanford.edu

The classical techniques for dimensionality reduction, PCA and MDS, are simple to implement, efficiently computable, and guaranteed to discover the true structure of data lying on or near a linear subspace of the high-dimensional input space (1,3). PCA finds a low-dimensional embedding of the data points that best preserves their variance as measured in the high-dimensional input space. Classical MDS finds an embedding that preserves the interpoint distances, equivalent to PCA when those distances are Euclidean. However, many data sets contain essential nonlinear structures that are invisible to PCA and MDS (4, 5, 11, 14). For example, both methods fail to detect the true degrees of freedom of the face data set (Fig. 1A), or even its intrinsic three-dimensionality (Fig. 2A).

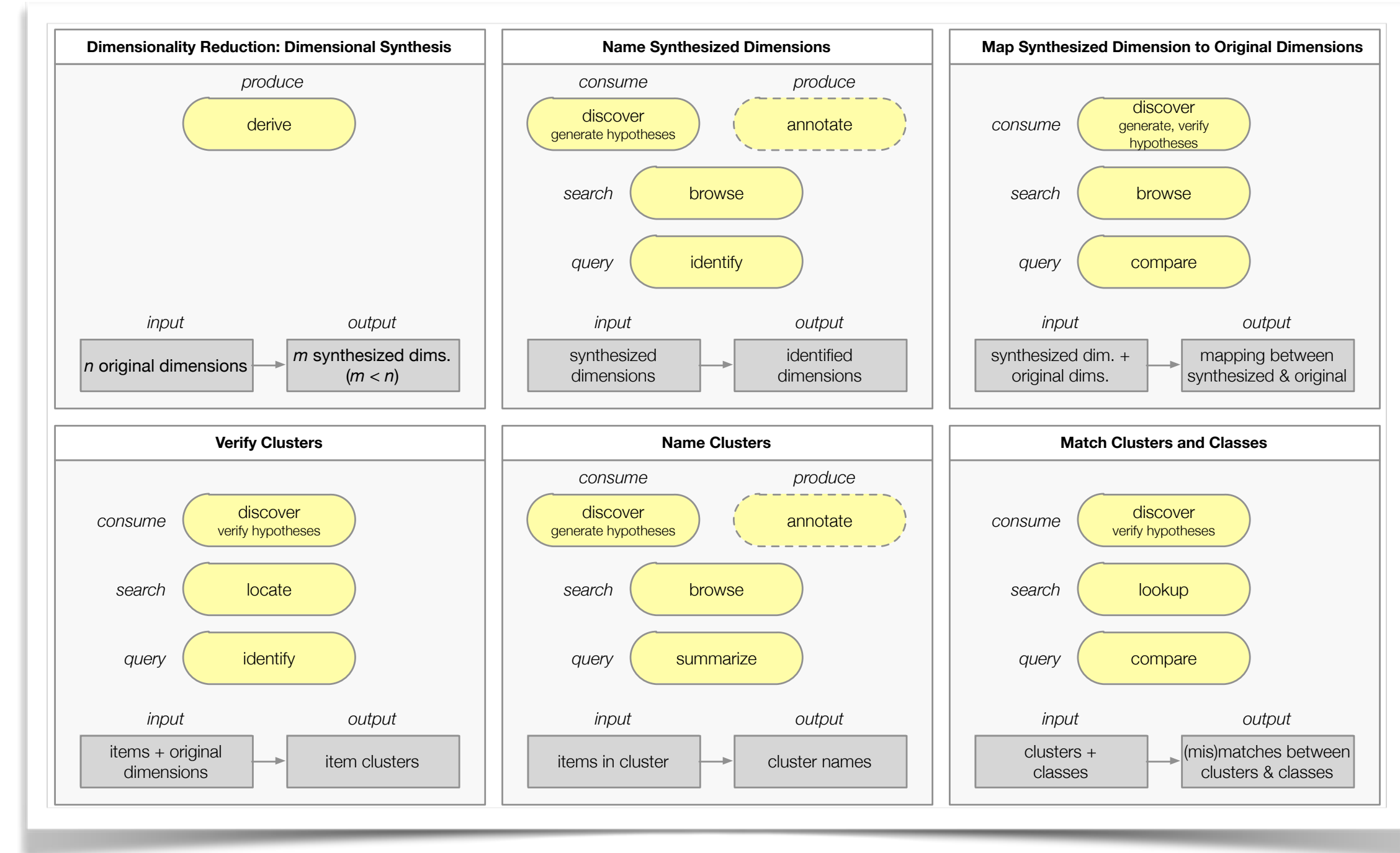
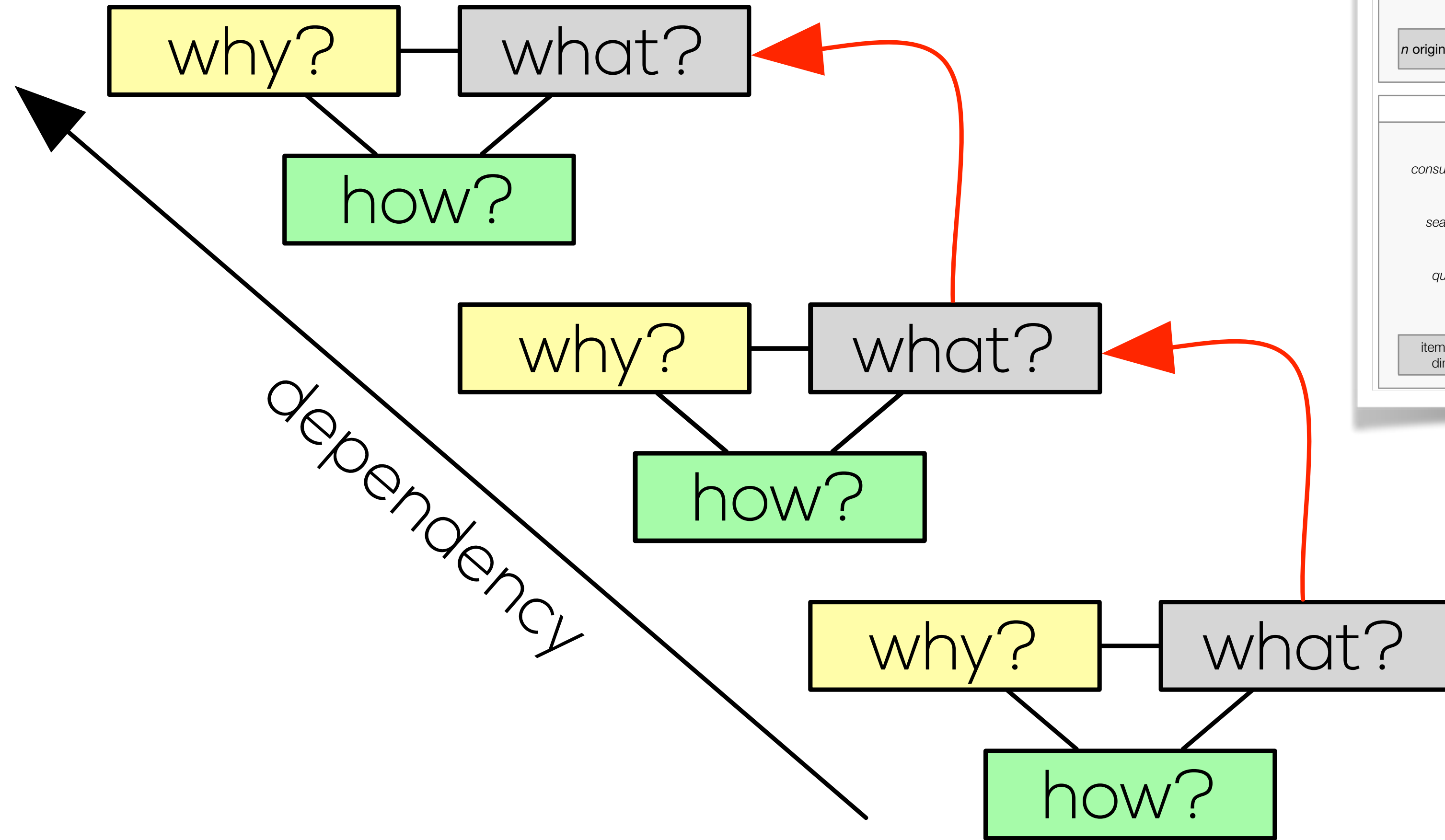
Here we describe an approach that combines the major algorithmic features of PCA and MDS—computational efficiency, global optimality, and asymptotic convergence guarantees—with the flexibility to learn a broad class of nonlinear manifolds. Figure 3A illustrates the challenge of nonlinearity with data lying on a two-dimensional “Swiss roll”: points far apart on the underlying manifold, as measured by their geodesic, or shortest path, distances, may appear deceptively close in the high-dimensional input space, as measured by their straight-line Euclidean distance. Only the geodesic distances reflect the true low-dimensional geometry of the manifold, but PCA and MDS effectively see just the Euclidean structure; thus, they fail to detect the intrinsic two-dimensionality (Fig. 2B).

Our approach builds on classical MDS but seeks to preserve the intrinsic geometry of the data, as captured in the geodesic manifold distances between all pairs of data points. The crux is estimating the geodesic distance between faraway points, given only input-space distances. For neighboring points, input-space distances provide a good approximation.





# A COMMON LEXICON FOR ANALYSIS



**domain-agnostic** yet  
**data-type-specific** task  
characterization

