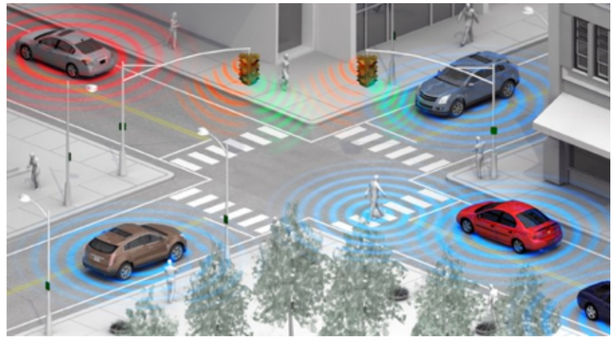
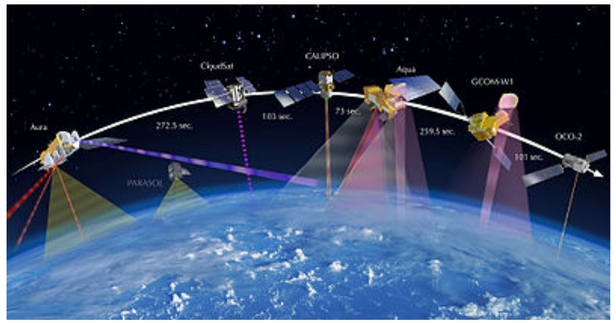
Visualization Design Studies

Michael Sedlmair
Visualization & Data Analysis Group
universität wien

Data





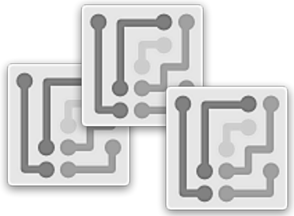
https://en.wikipedia.org/wiki/Earth_observation_satellite

https://www.extremetech.com/extreme/226739-how-smart-cities-will-work

Data Science

Data

Algorithms



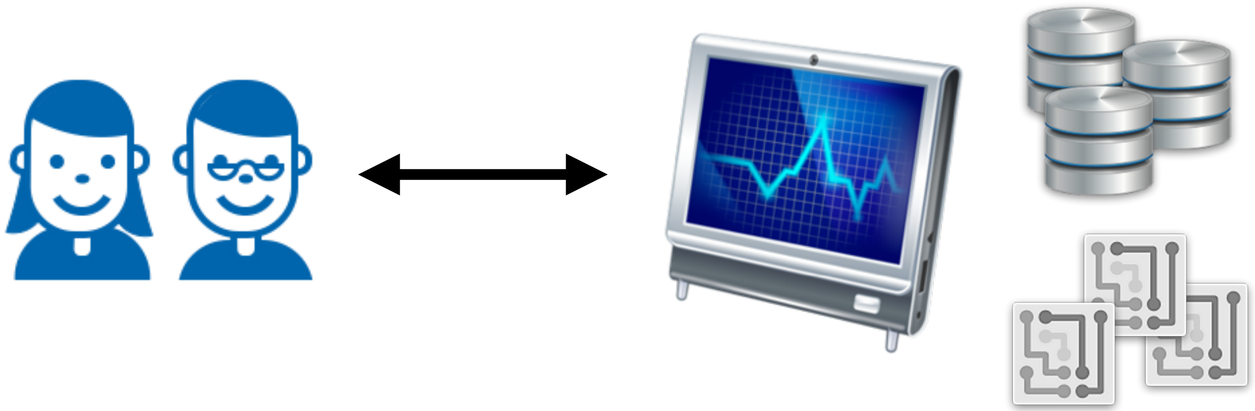
Human-centered Data Science

Scientists

Engineers

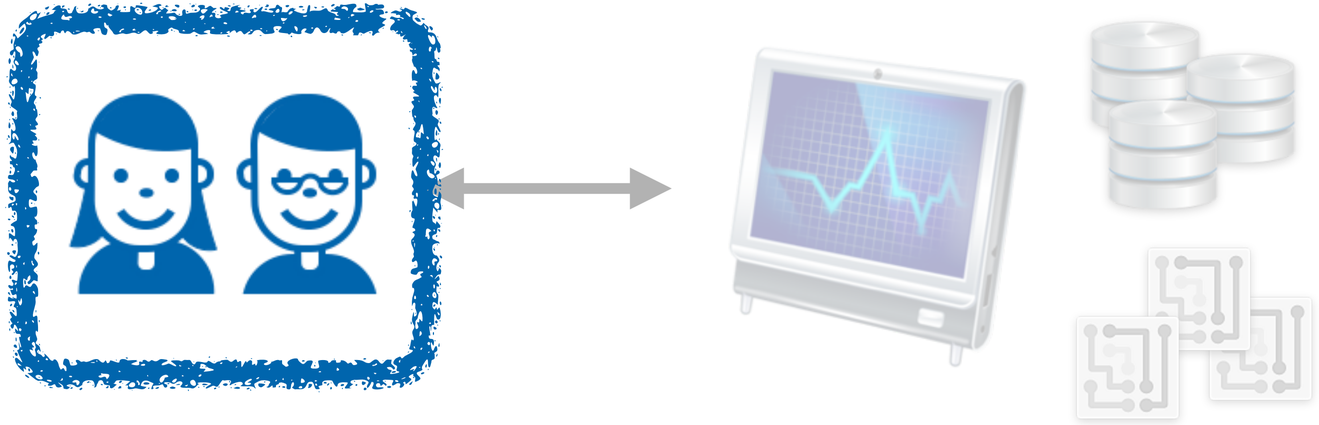
Journalists

Policy Makers



My goals

1. Human as a first class citizen in Data Science research

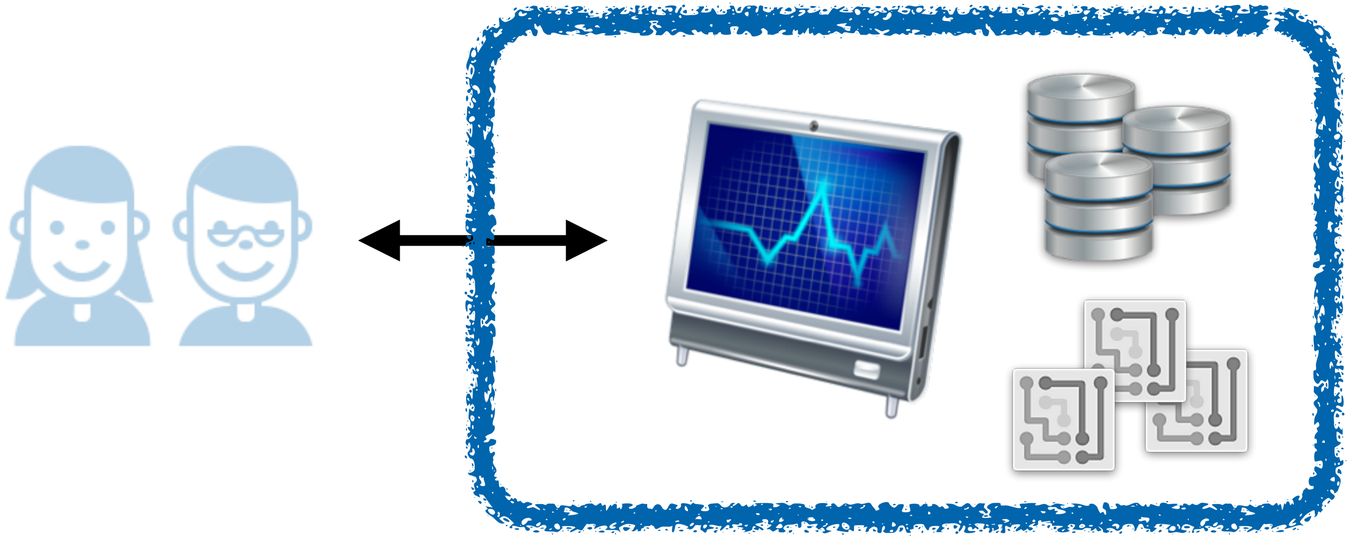


My goals

2. Human-centered tools & techniques

➡ interactive visualization

➡ algorithmic analysis



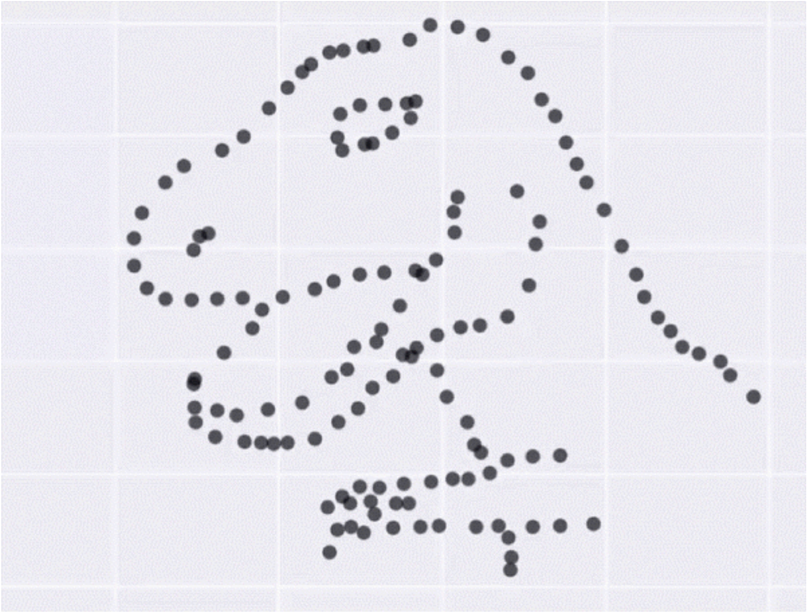
Help to:

understand patterns, gain insights, make decisions, …

Why interactive visualization?

- **humans** can see patterns that algorithm cannot

Different graphs …



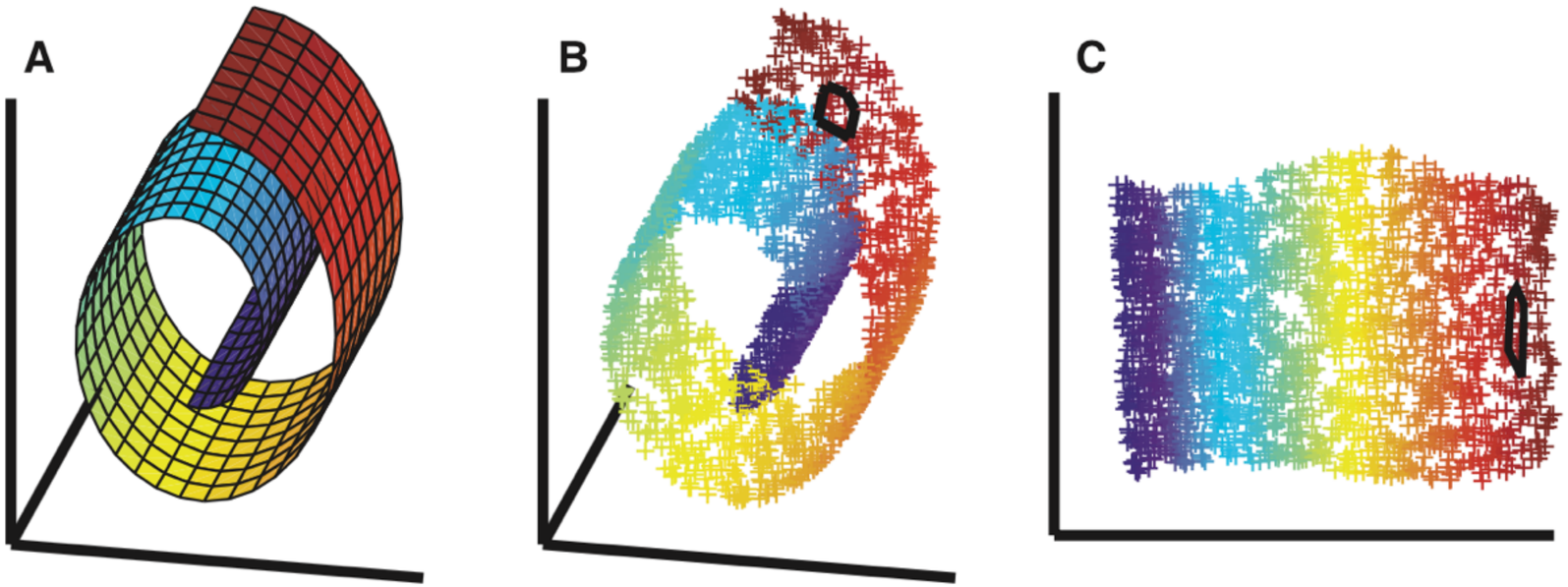
Same stats

x mean	**54**
y mean	**48**
x SD	**17**
y SD	**27**
x/y correlation	**-0.1**

Matejka, Fitzmaurice.
Same Stats, Different Graphs: Generating Datasets with Varied
ACM SIGCHI Conference on Human Factors in Computing Systems, 2017, to appear.

Why algorithmic analysis?

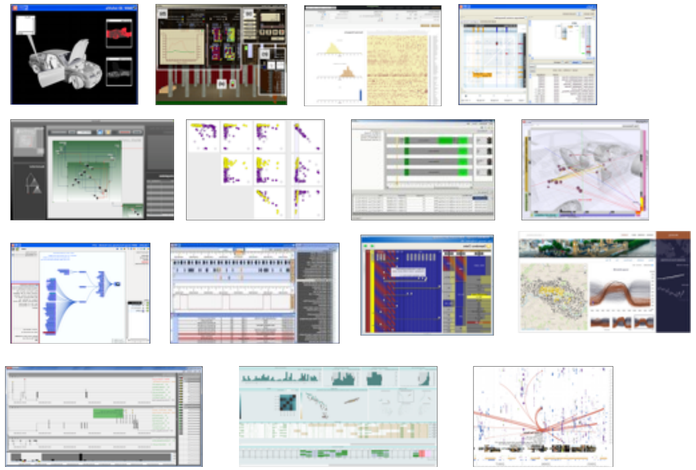
- **algorithms** can reveal patterns that humans cannot see

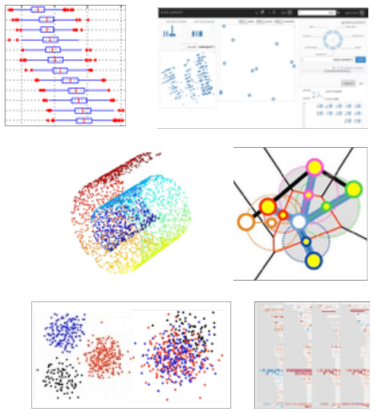


Dimension reduction
e.g., unfolding swiss roll

Roweis, Saul.
Nonlinear dimensionality reduction by locally linear embedding.
Science, p. 2323-2326, 2000.

My Work



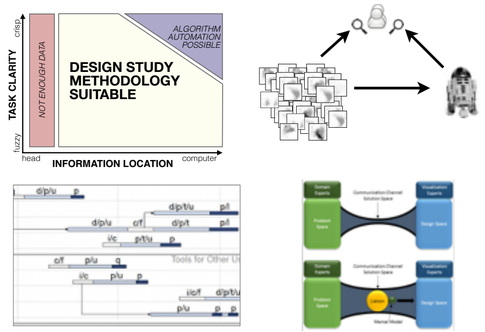


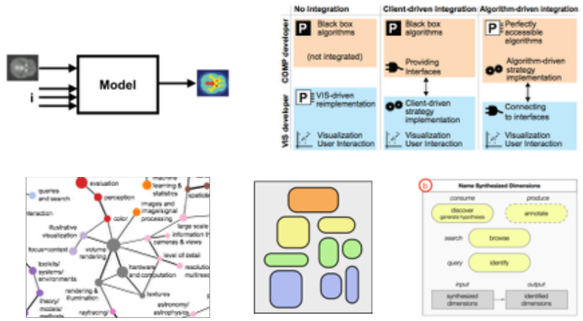
problem-driven

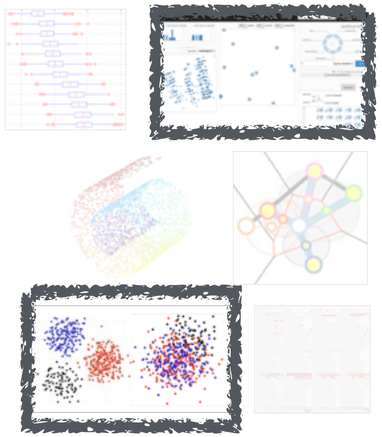
technique-driven

research methods

theory

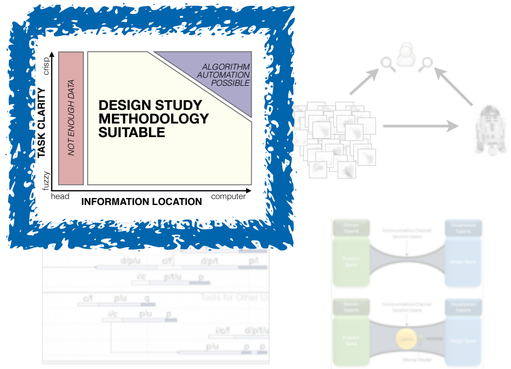


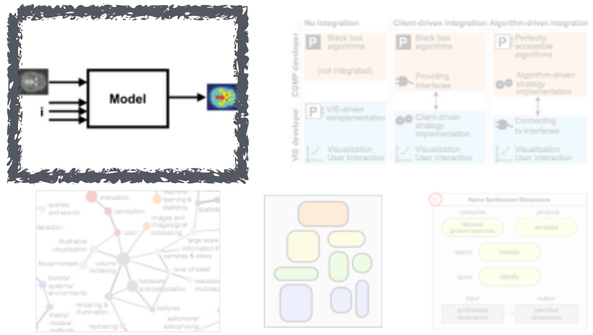




problem-driven

Today's main focus:
Visualization Design Studies





problem-driven

Today's main focus: **Visualization Design Studies**

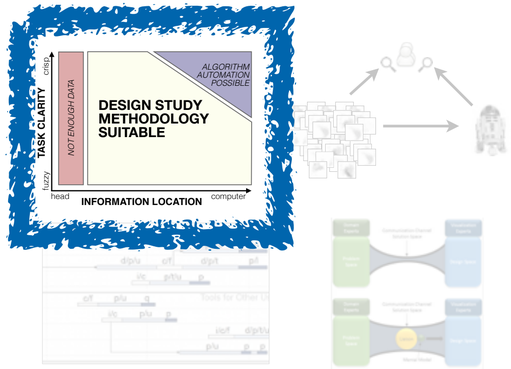
"A design study is a project in which visualization researchers analyze a specific

real-world problem *faced by domain experts,*

design *a visualization system that supports solving this problem,*

validate *the design, and*

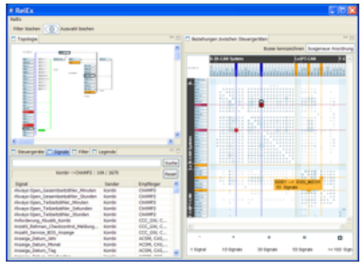
reflect *about lessons learned in order to refine visualization design guidelines."*



Sedlmair, Meyer, Munzner.
Design Study Methodology: Reflections from the Trenches and the Stacks
IEEE TVCG (Proc. InfoVis), 18(12): 2431–2440, 2012.

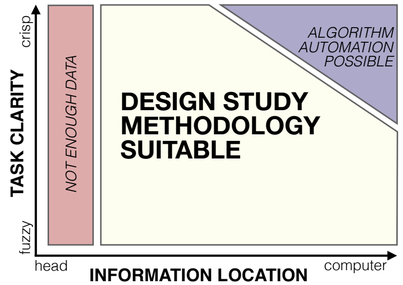
Outline

- RelEx
  - example design study with automotive engineers



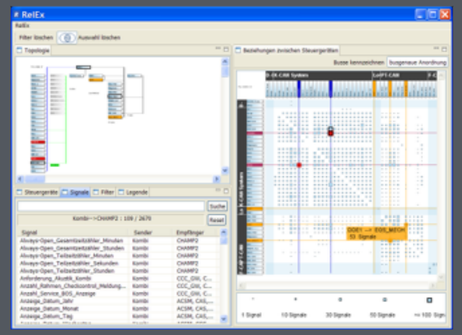
[IEEE TVCG, 2012]

- Design Study Methodology
  - meta-paper: how to do design studies



[IEEE TVCG, 2012]

- The way forward
  - why design studies matter



RelEx:
Example of a Design Study

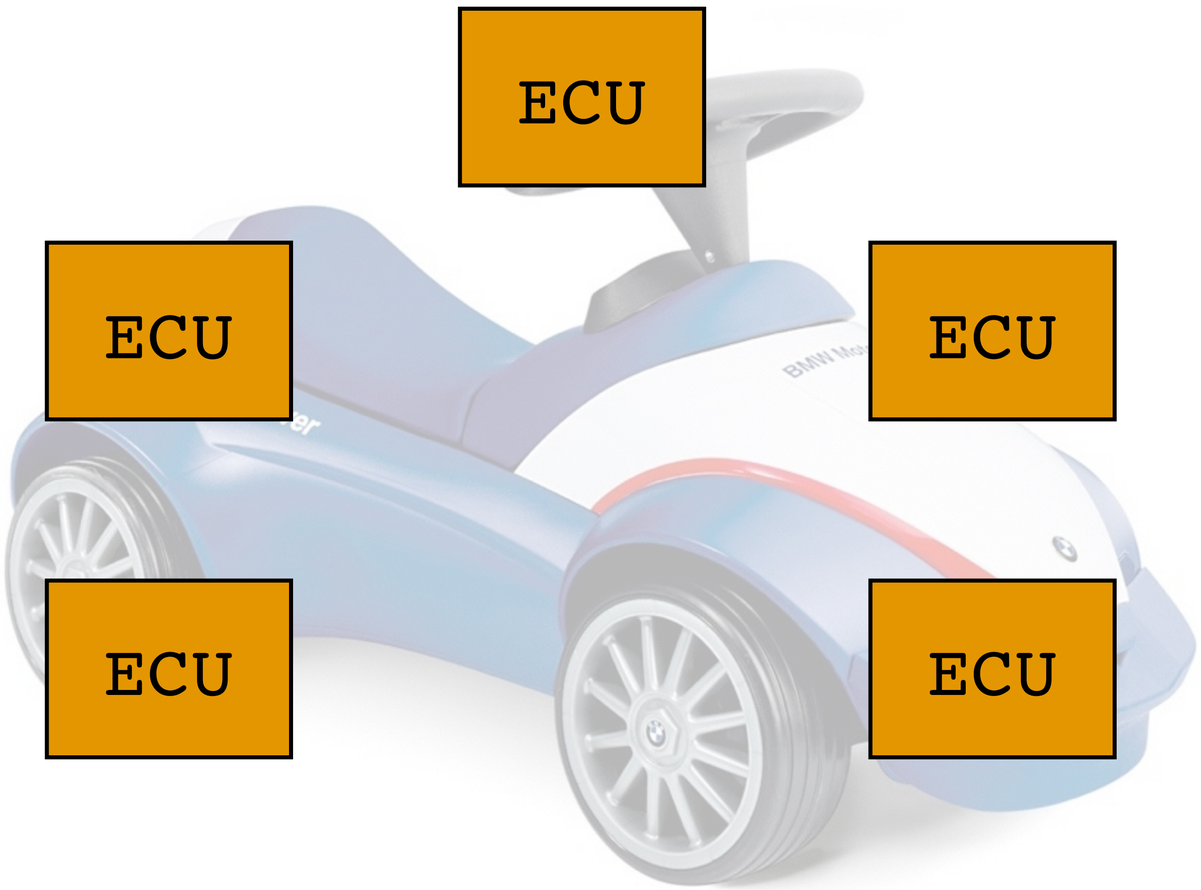
RelEx: Visualization for Actively Changing Overlay Network Specifications

M. Sedlmair, A. Frank, T. Munzner, A. Butz

IEEE TVCG 18(12): 2729-2738, 2012 (Proc. InfoVis).

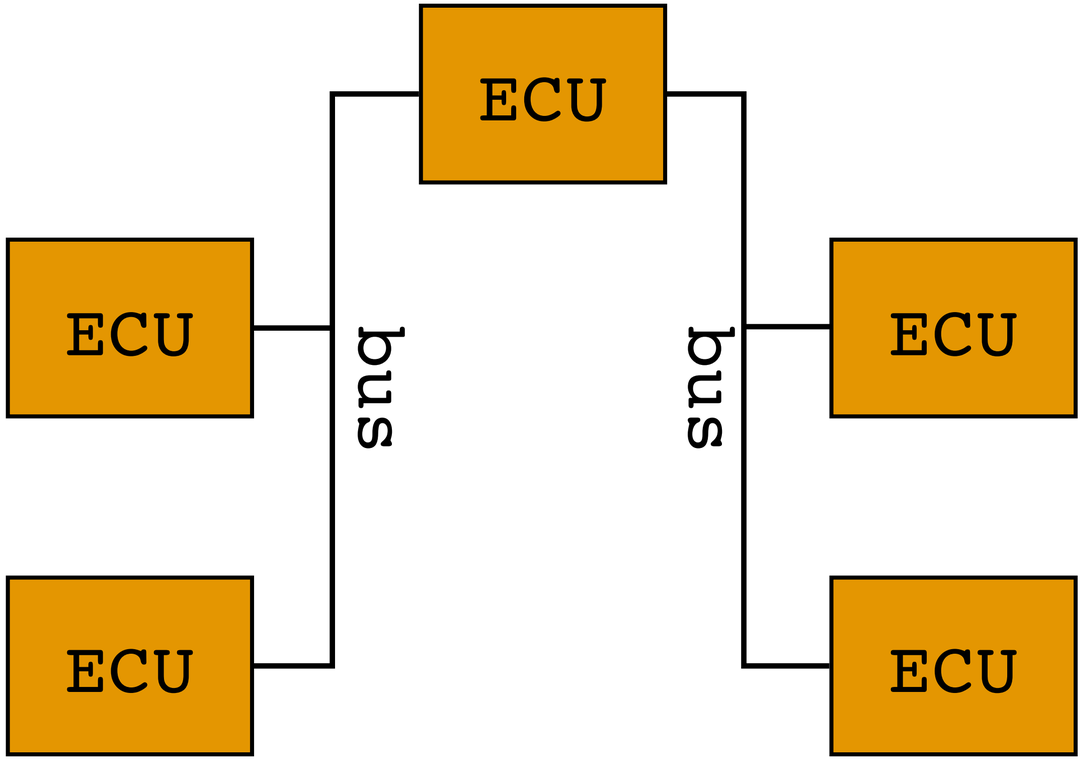
Domain: In-car Electronics

Physical network **(Hardware)**



~100 ECU (nodes)

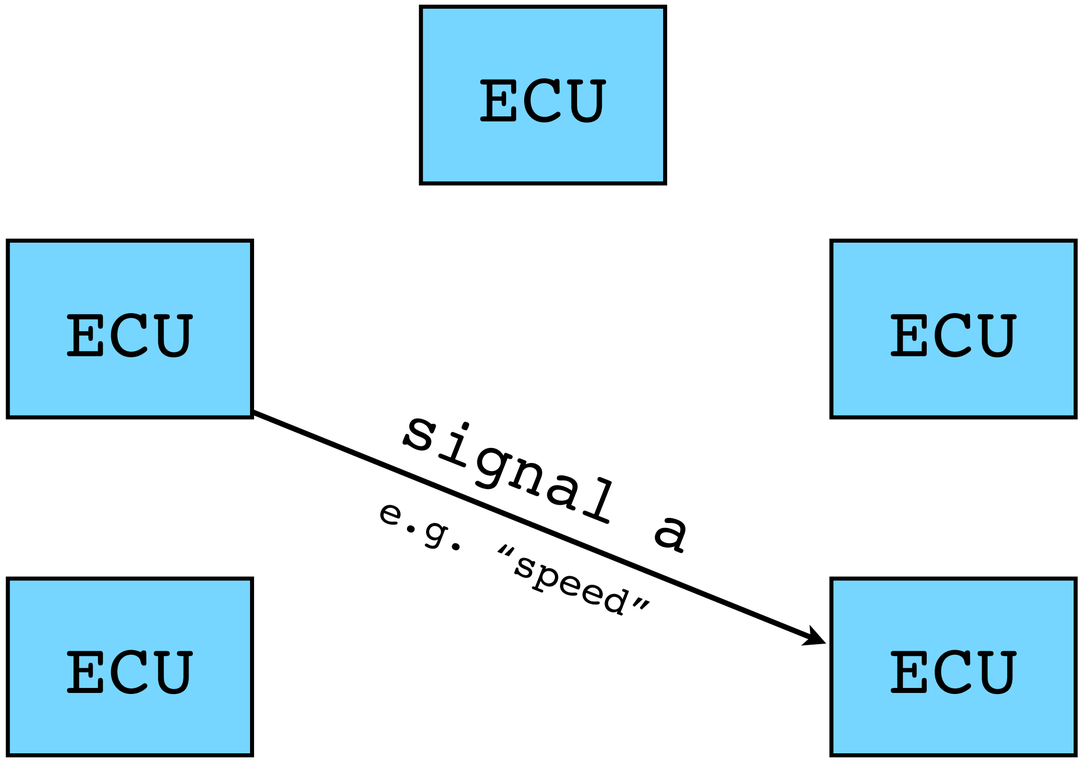
Physical network **(Hardware)**



~100 ECU (nodes)

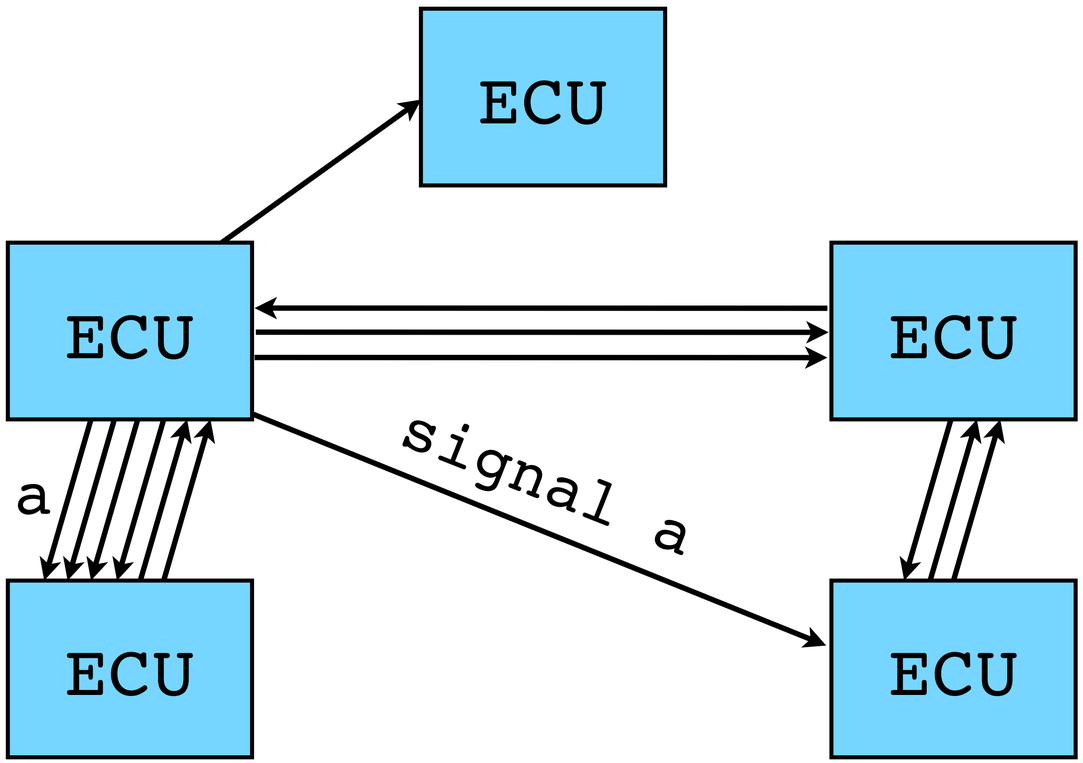
10-15 Bus systems (edges)

Logical Network **(Software)**



~100 ECU (nodes)

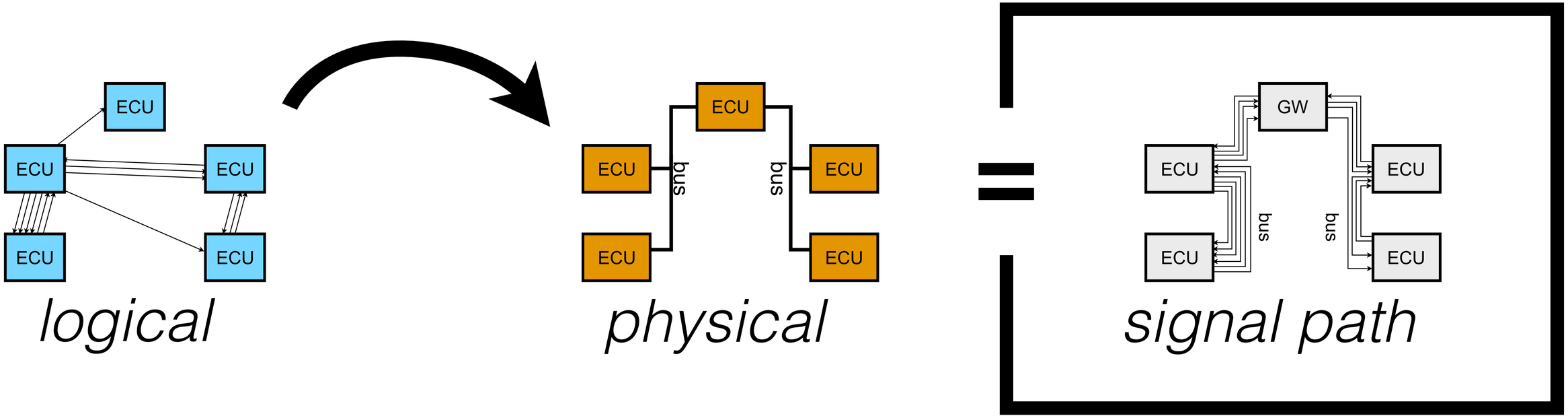
Logical Network **(Software)**



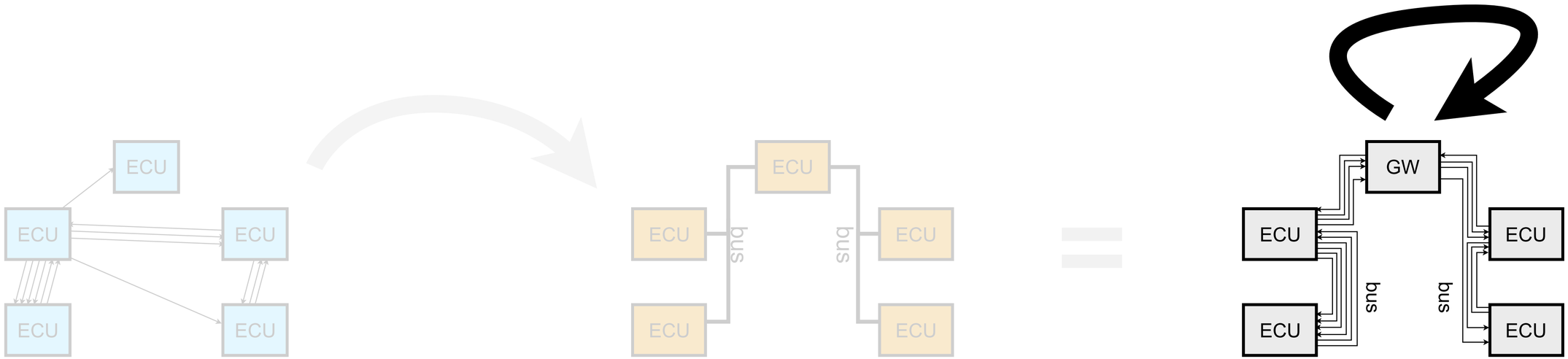
~100 ECU (nodes)

~10k signals (edges)

Tasks: Mapping



Tasks: Multi-objective optimization

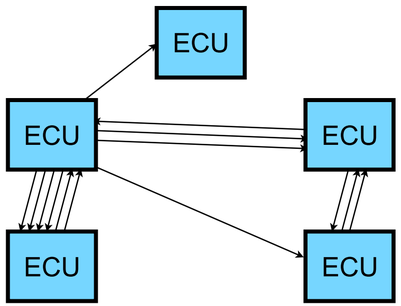


Ill-defined constraints

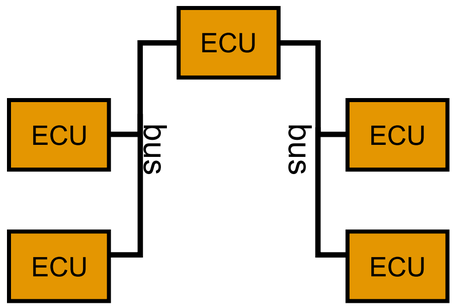
bandwidth ... delay/real time ...
path length ... load balance ...
reliability ... money ...

-- engineer, BMW --

Tasks: External Change Requests

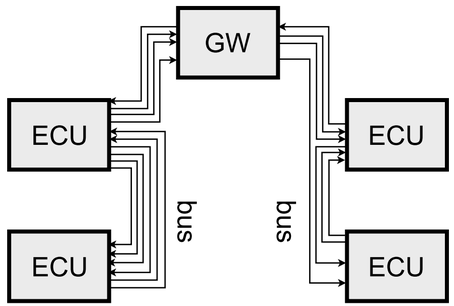


logical



physical

=



signal path

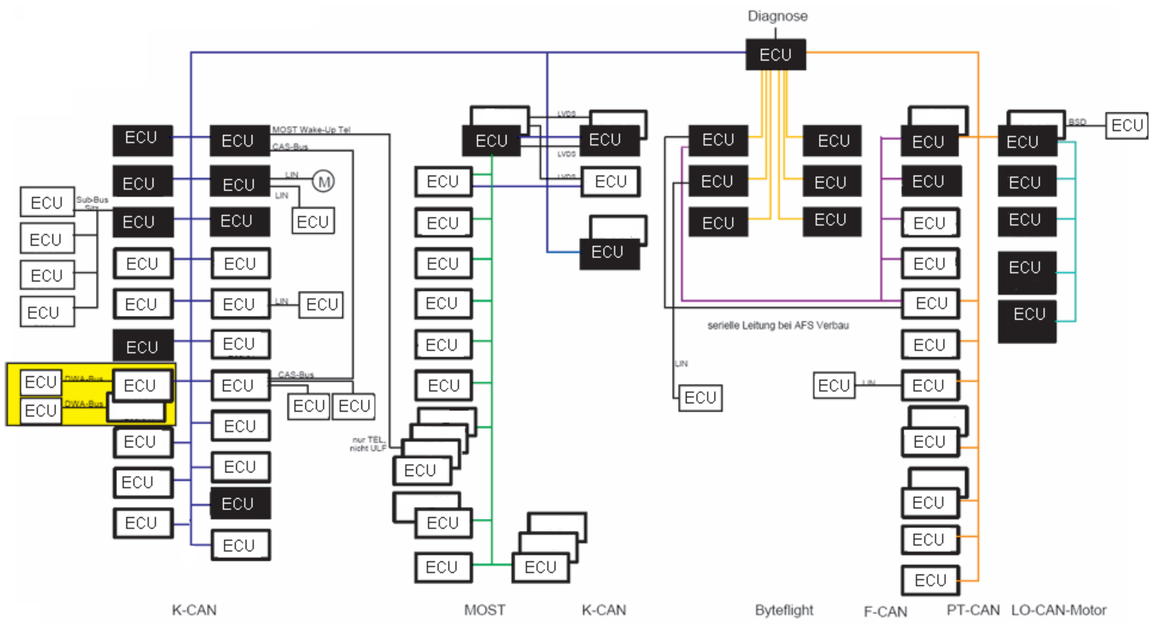
Change request

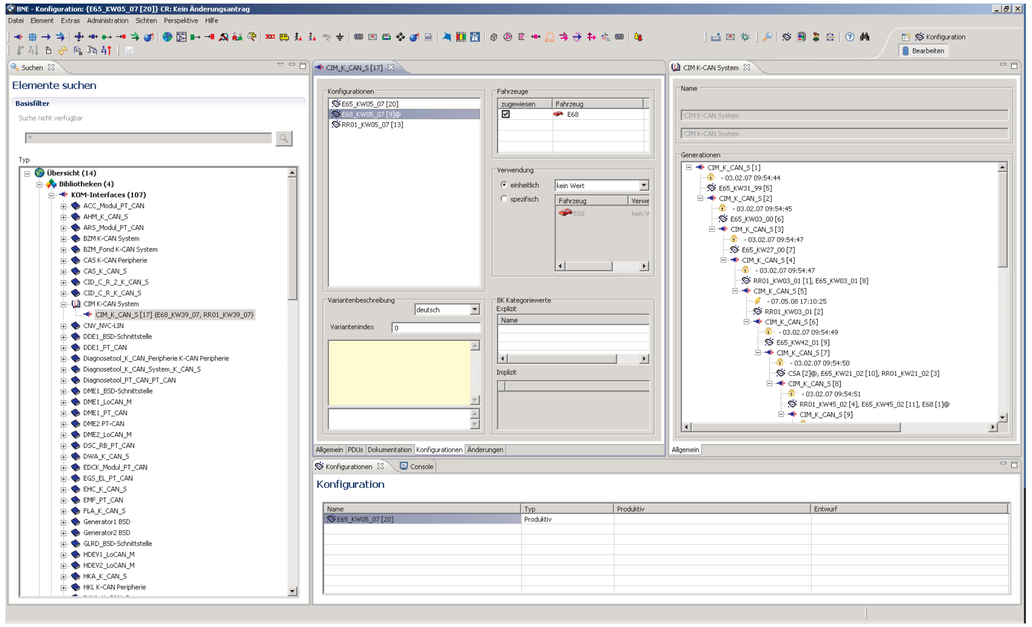
Change request

Change

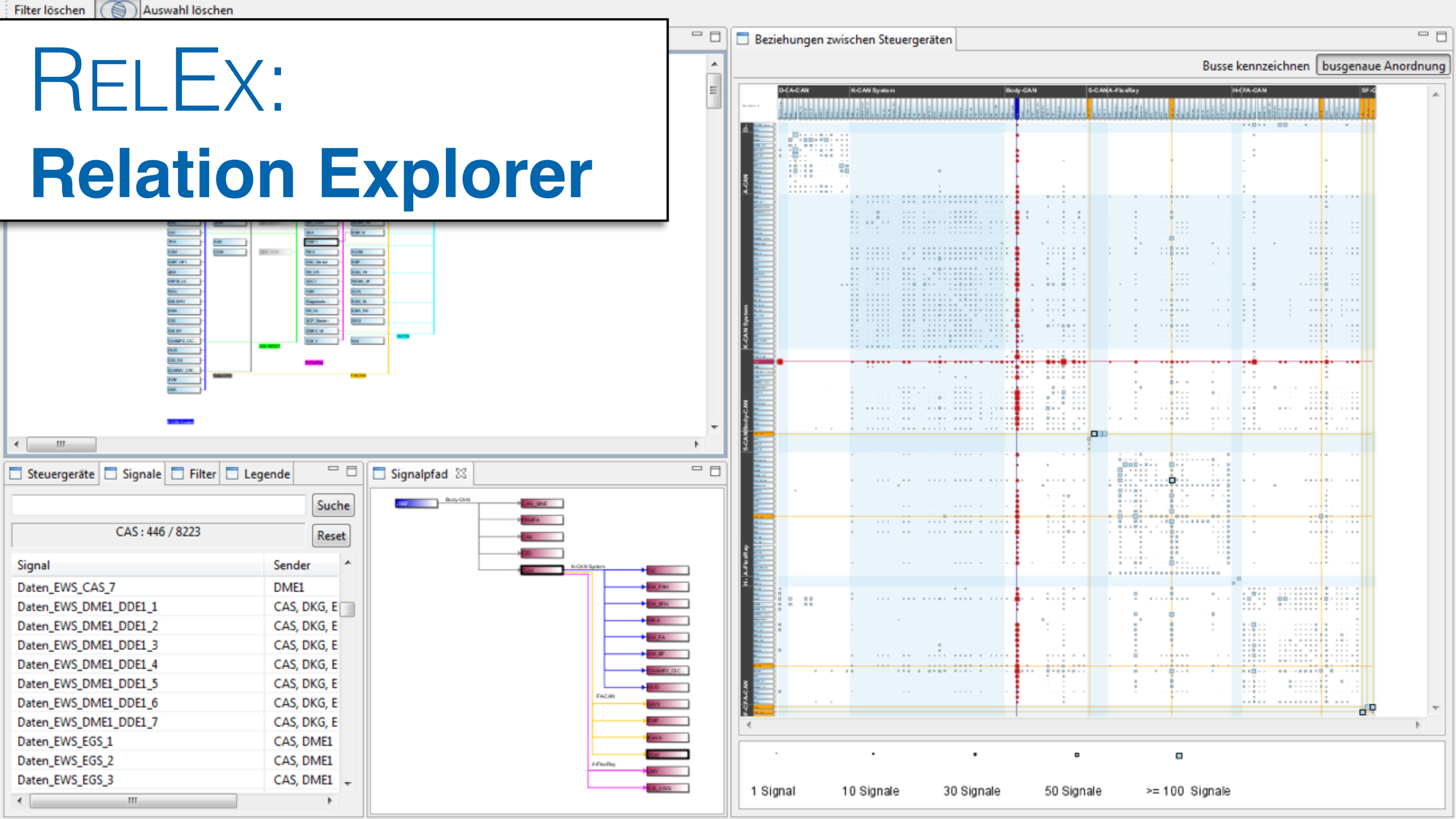
(trivial requests might lead to complex changes)

Not well supported by current practices:



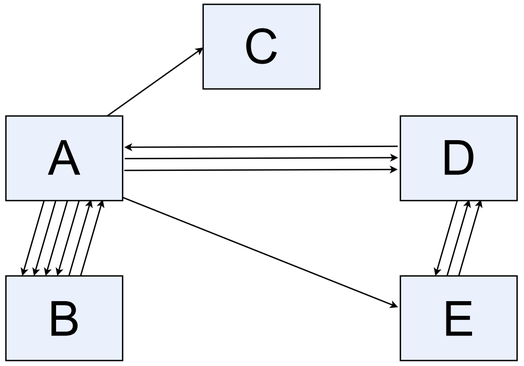


RELEX:
Relation Explorer



RELEX:
Logical Overview

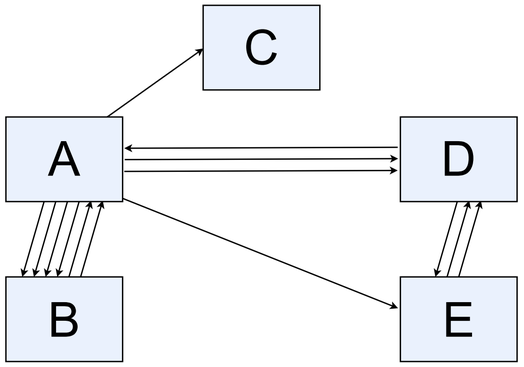
LOGICAL NETWORK



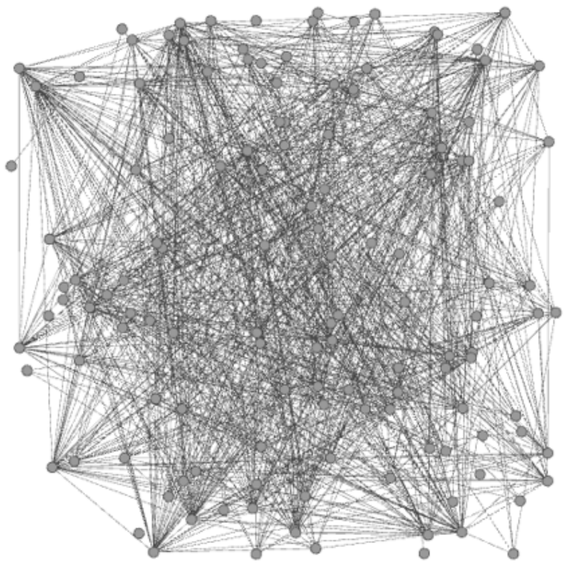
- multigraph
- 100 nodes/10k edges

RELEX:
Logical Overview

LOGICAL NETWORK



- multigraph
- 100 nodes/10k edges

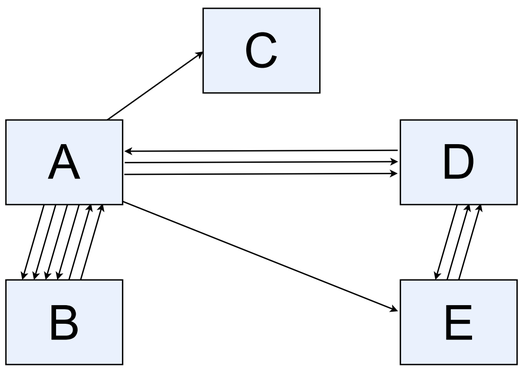


Hairball

RELEX: Logical Overview

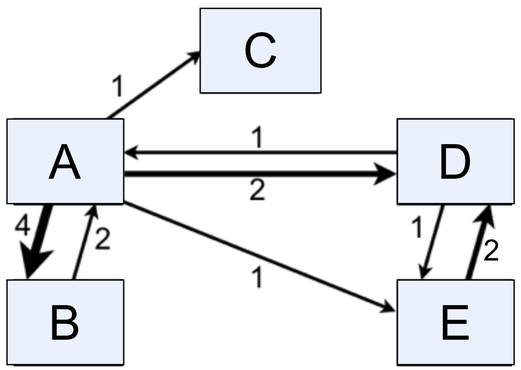
Guideline [Ghoniem 2004]
Matrix for dense graphs

LOGICAL NETWORK



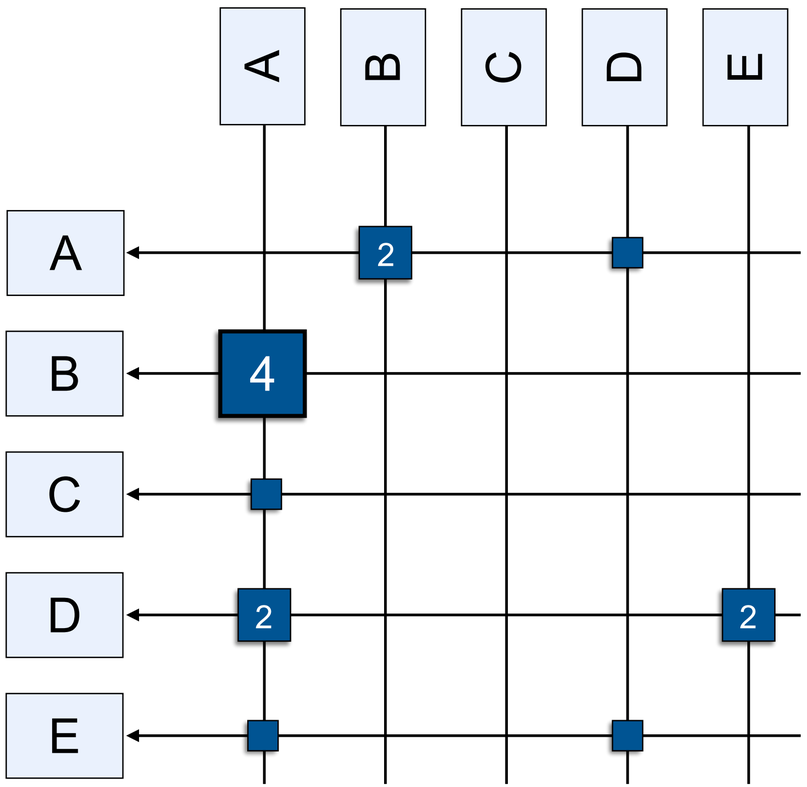
- multigraph
- 100 nodes /10k edges

SIGNAL COUNT NETWORK

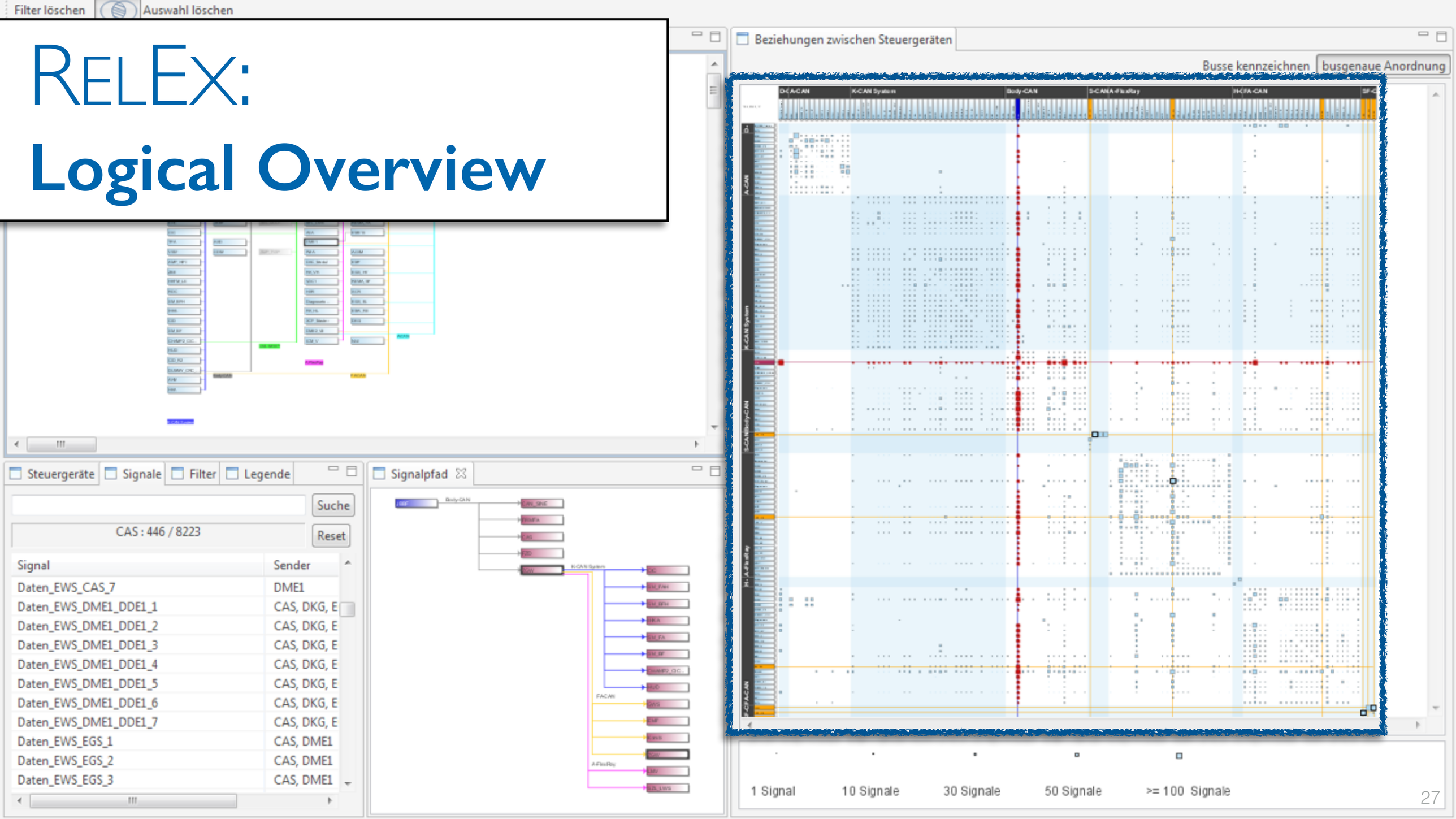


- directed graph
- 1k weighted edges

VISUAL ENCODING: SIZE-CODED MATRIX



RELEX:
Logical Overview



RelEx

RelEx

Filter löschen | Auswahl löschen

Topologie

Steuergeräte | Signale | Filter | Legende

Suche

XBU : 158 / 8222

Reset

Signal	Sender	Empfänger
ABHN-Signal	KWR	TMM-ECU, XBU
AONQ-Signal	DNJ	KWR, XBU
APWC-Signal	KWR	XBU
ARNL-Signal	KWR	AGW, CSD, ELP, MSE-ECU, MWD,...
AWKZ-Signal	DNJ	ELP, FAV, KWR, MWQ, PVV, QXE, ...
AXYQ-Signal	KWR	BQM, CSD, IBC, JHH, MSE-ECU, X...

Beziehungen zwischen Steuergeräten

Busse kennzeichnen | busgenaue Anordnung

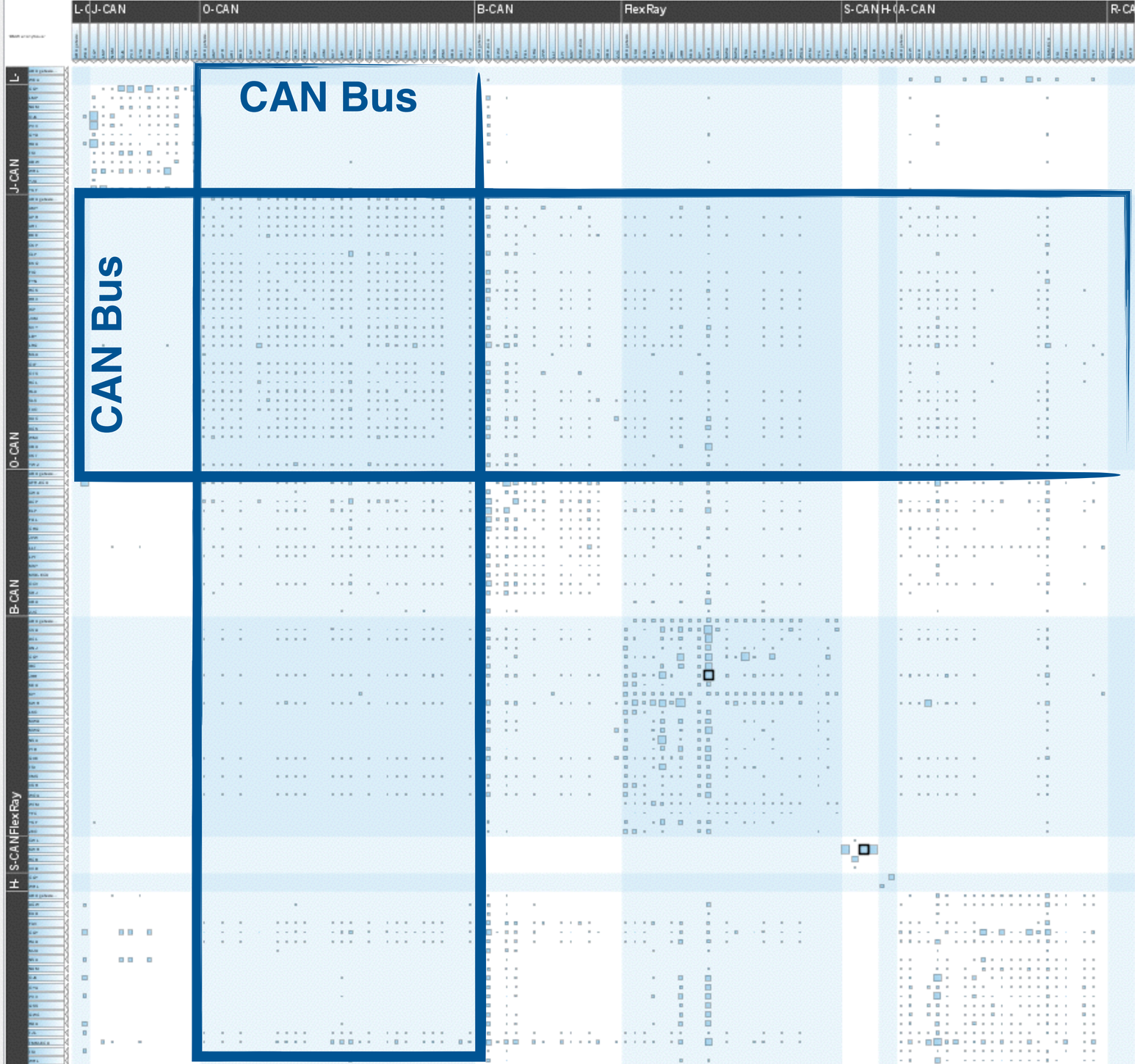
FlexRay

1 Signal | 10 Signale | 30 Signale | 50 Signale | >= 100 Signale

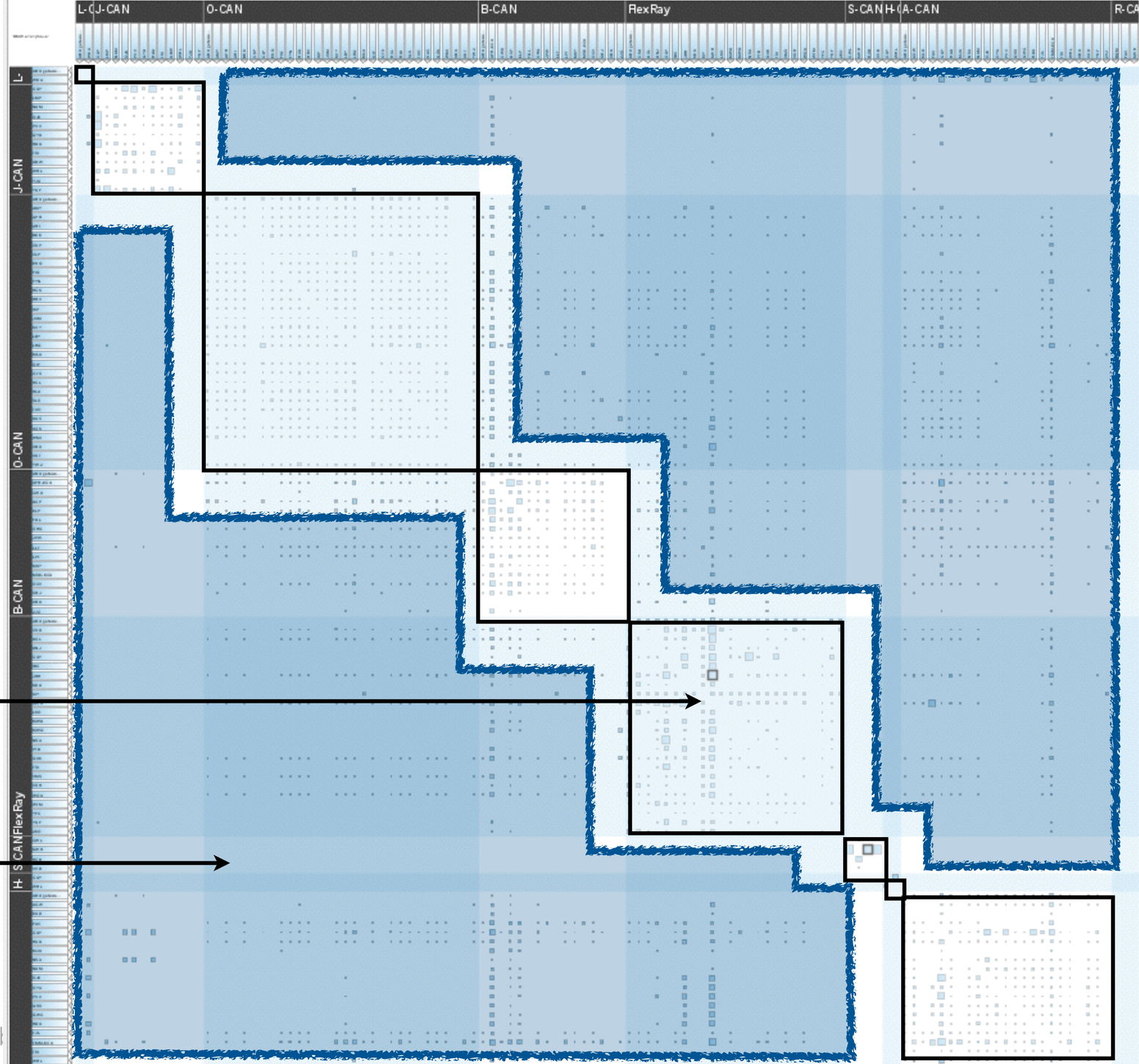
Evaluation of RelEx

- field & lab studies with final release
  - overall 17 engineers
  - 5 weeks

Novel insights:
Bus communication patterns



Novel insights:
Bus communication patterns

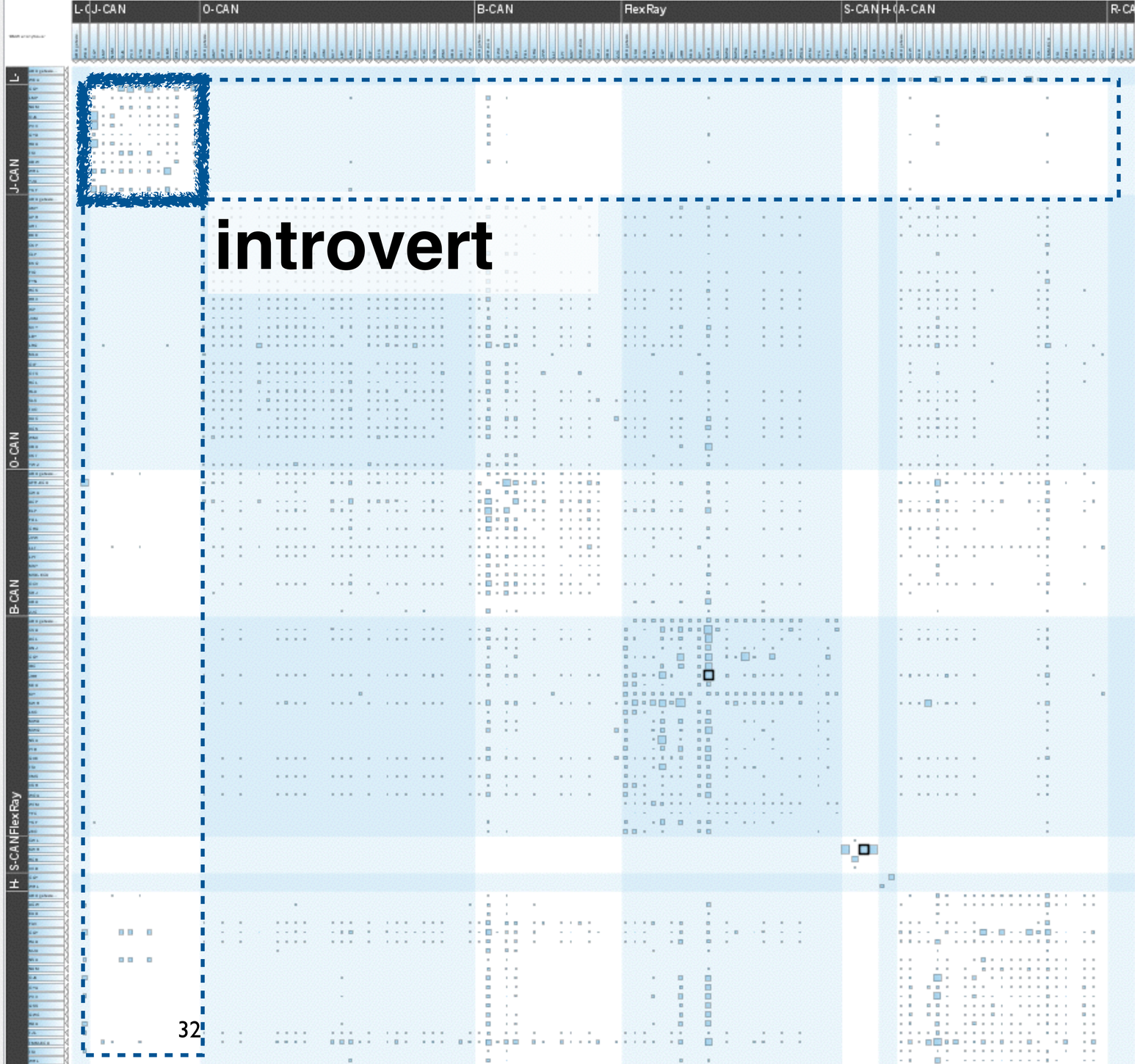


Within-bus

Between-bus

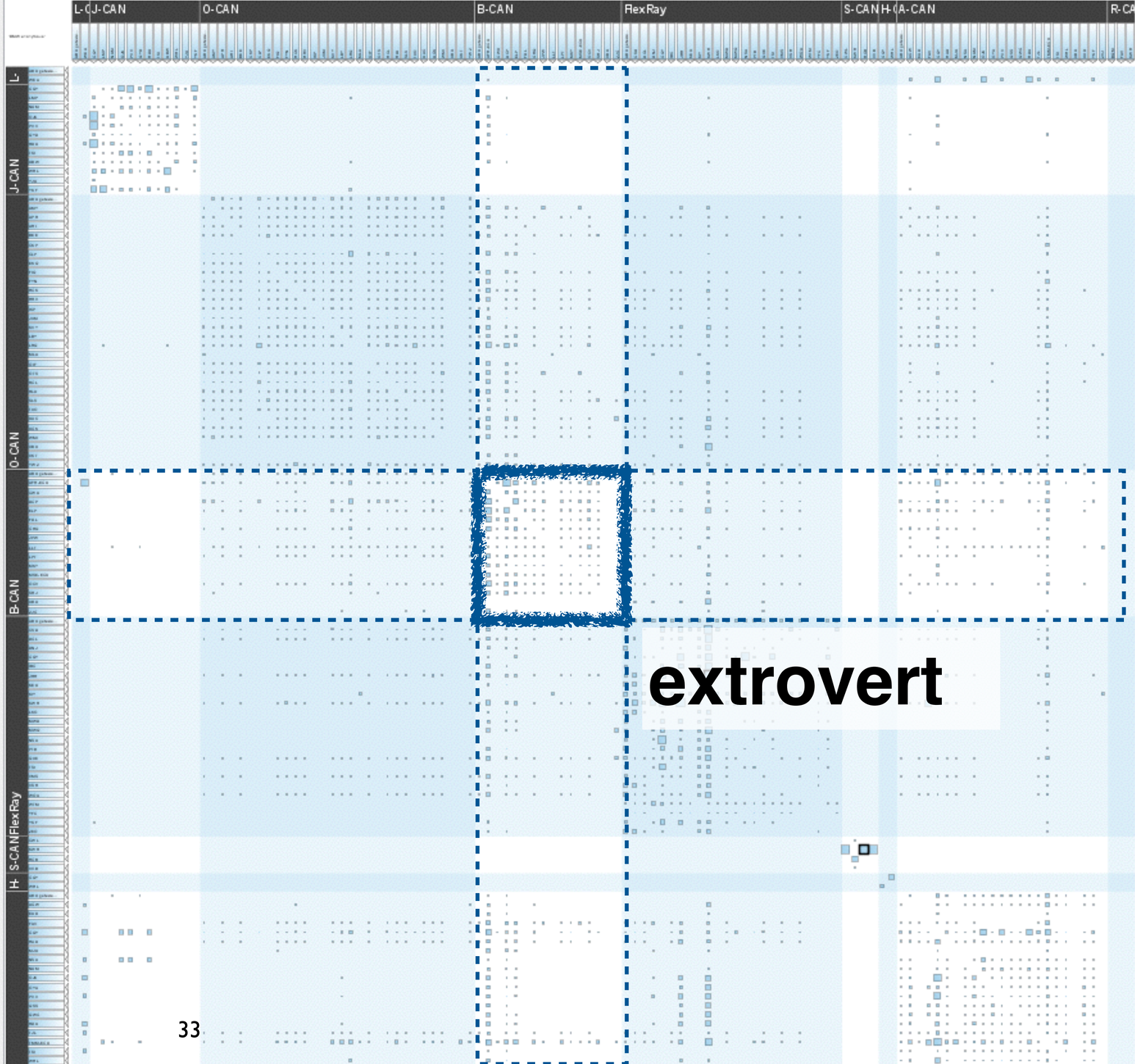
Novel insights:
Bus communication patterns

introvert
vs.
extrovert



Novel insights:
Bus communication patterns

introvert
vs.
extrovert



Speedup

"RelEx gives me a more compact, way faster access to the information I need."

BMW engineer
(translated from German).

Adoption

- 3-month post study
- 15+ engineers

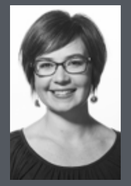
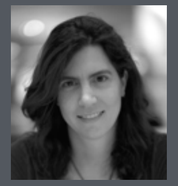
Design Study Methodology

Design Study Methodology: Reflection from the Trenches and the Stacks

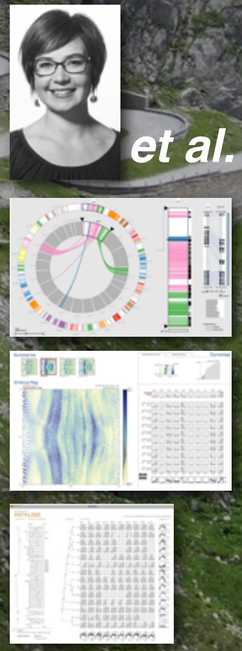
M. Sedlmair, M. Meyer, T. Munzner

IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis).

Best paper honorable mention.

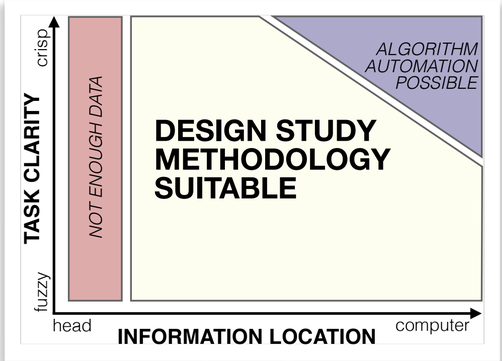



Design studies:
long and winding road with many pitfalls!

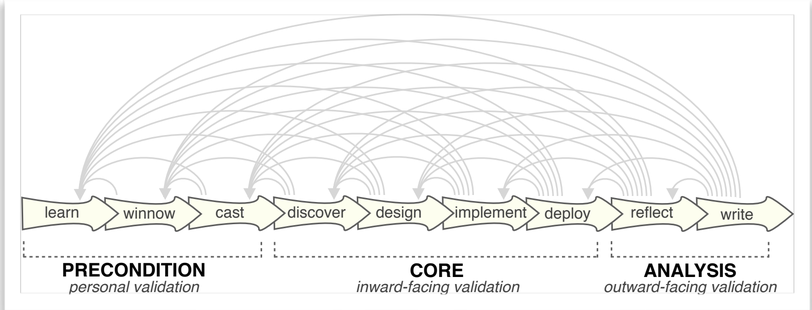


How to do design studies?

- Definitions



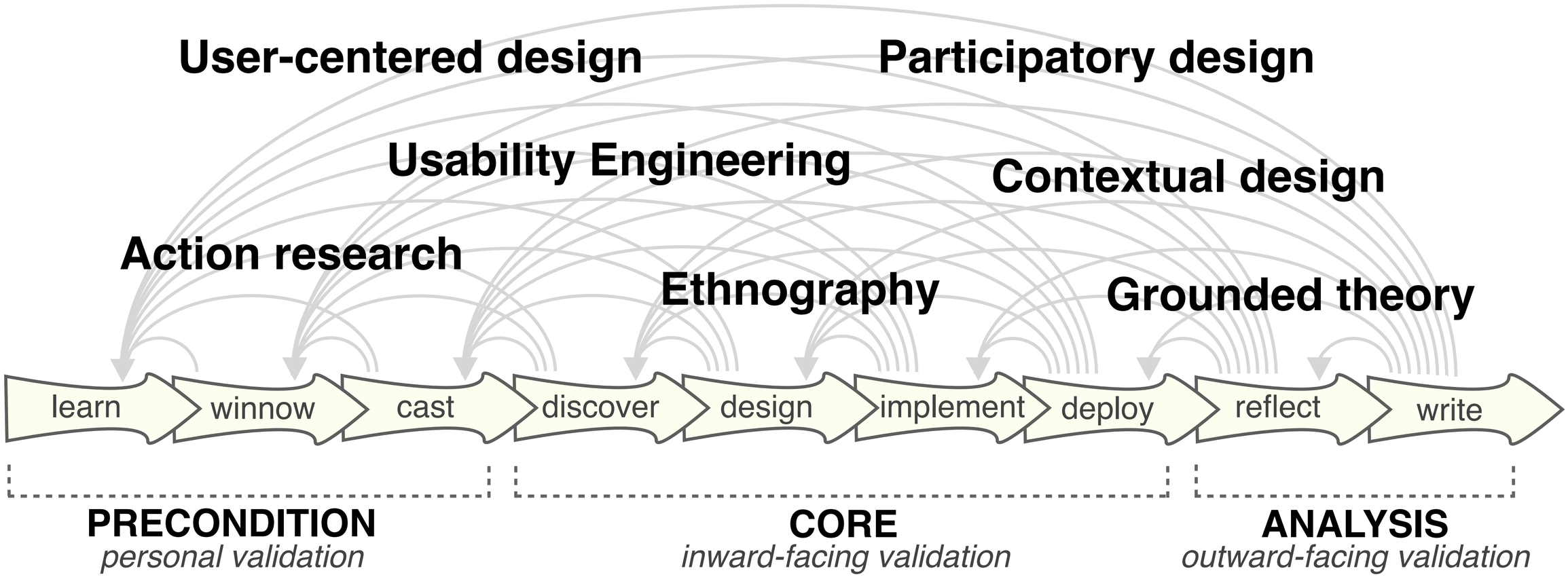
- 9-stage framework



- 32 pitfalls

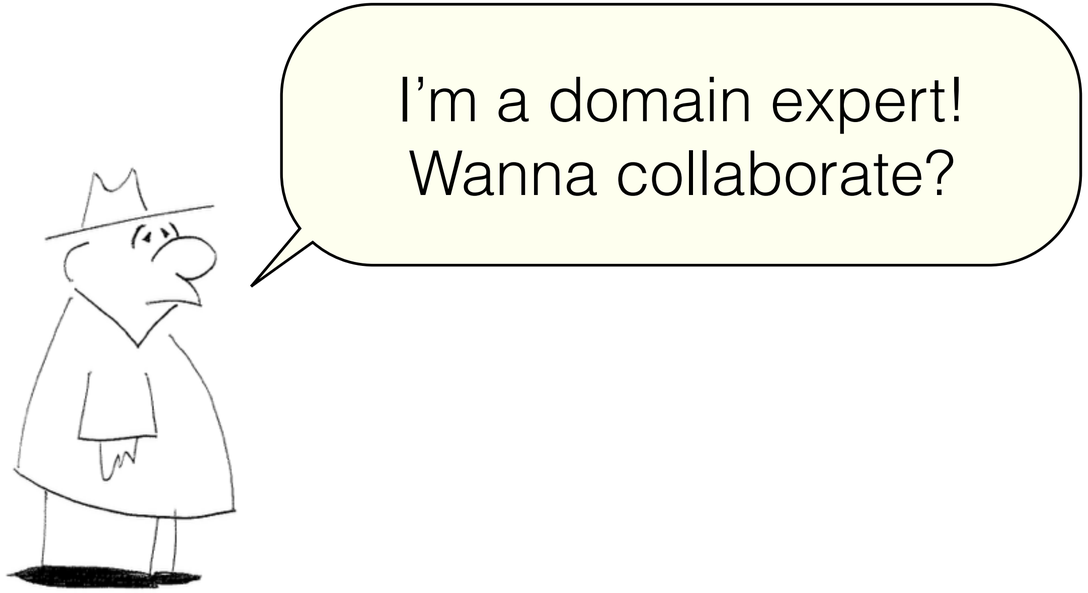
PF-1	premature advance: jumping forward over stages	general
PF-2	premature start: insufficient knowledge of vis literature	learn
PF-3	premature commitment: collaboration with wrong people	winnow
PF-4	no real data available (yet)	winnow
PF-5	insufficient time available from potential collaborators	winnow
PF-6	no need for visualization: problem can be automated	winnow
PF-7	researcher expertise does not match domain problem	winnow
PF-8	no need for research: engineering vs. research project	winnow
PF-9	no need for change: existing tools are good enough	winnow

9-stage framework

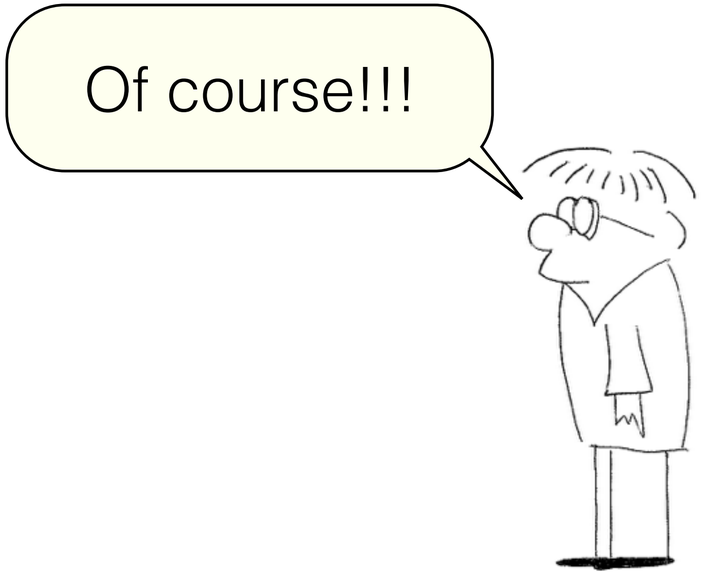


32 pitfalls

Example: premature commitment



Collaborator



Scientist

Important considerations!

Collaborator

Have **data**?
Have **need**?
Does the **problem** fit?
...

Scientist

32 pitfalls

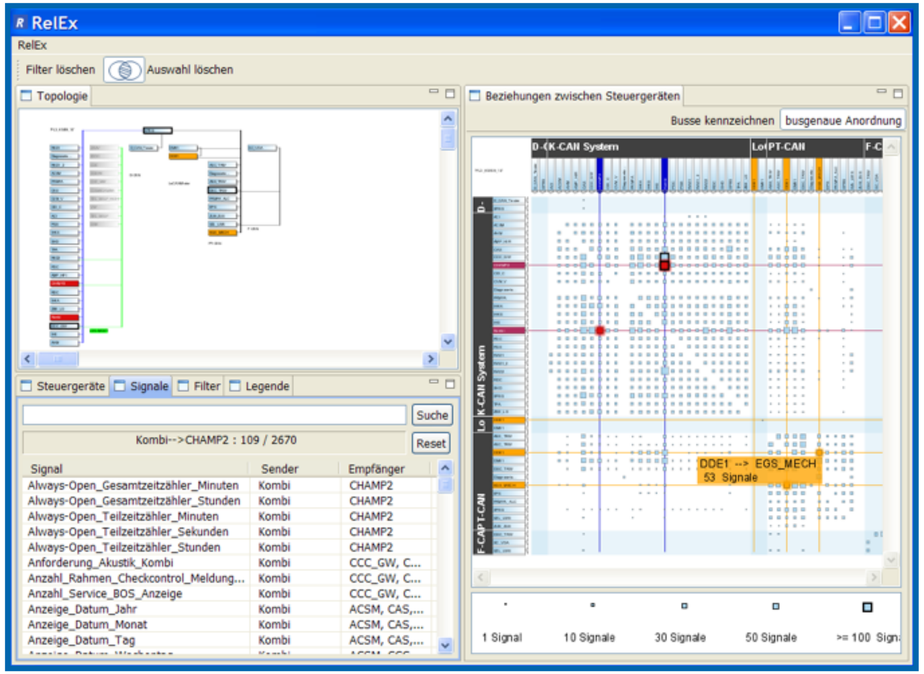
Example: no reflection

Reflection is where research emerges from engineering

Transferability: relate to other design studies

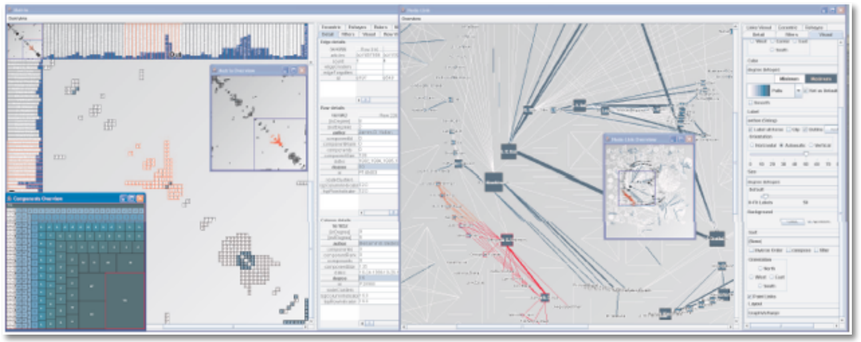
Sedlmair.
Design study contributions come in different guises: Seven guiding scenarios.
Beyond Time and Errors: Novel Evaluation Methods for Visualization (BELIV). ACM Press, 2016.

Example:
Reflection in RelEx

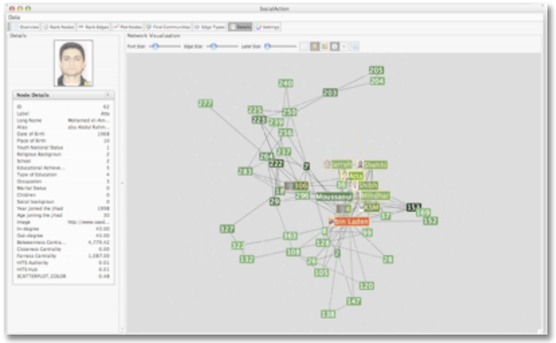


Previous work:
Focus on social network analysis

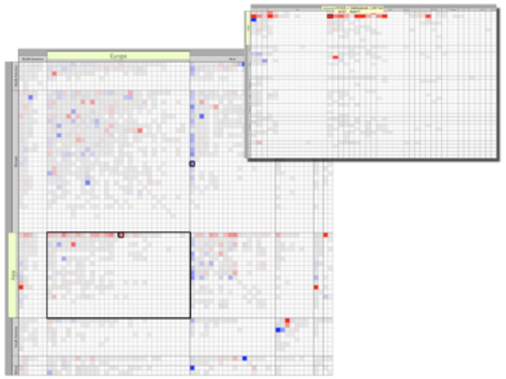
- radically different task and data abstractions



MatrixExplorer



SocialAction

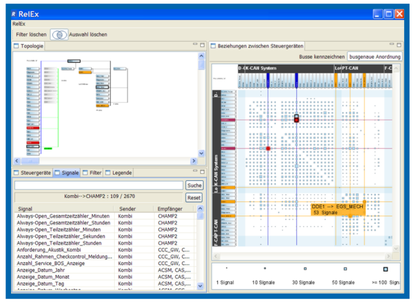


Honeycomb

vizster

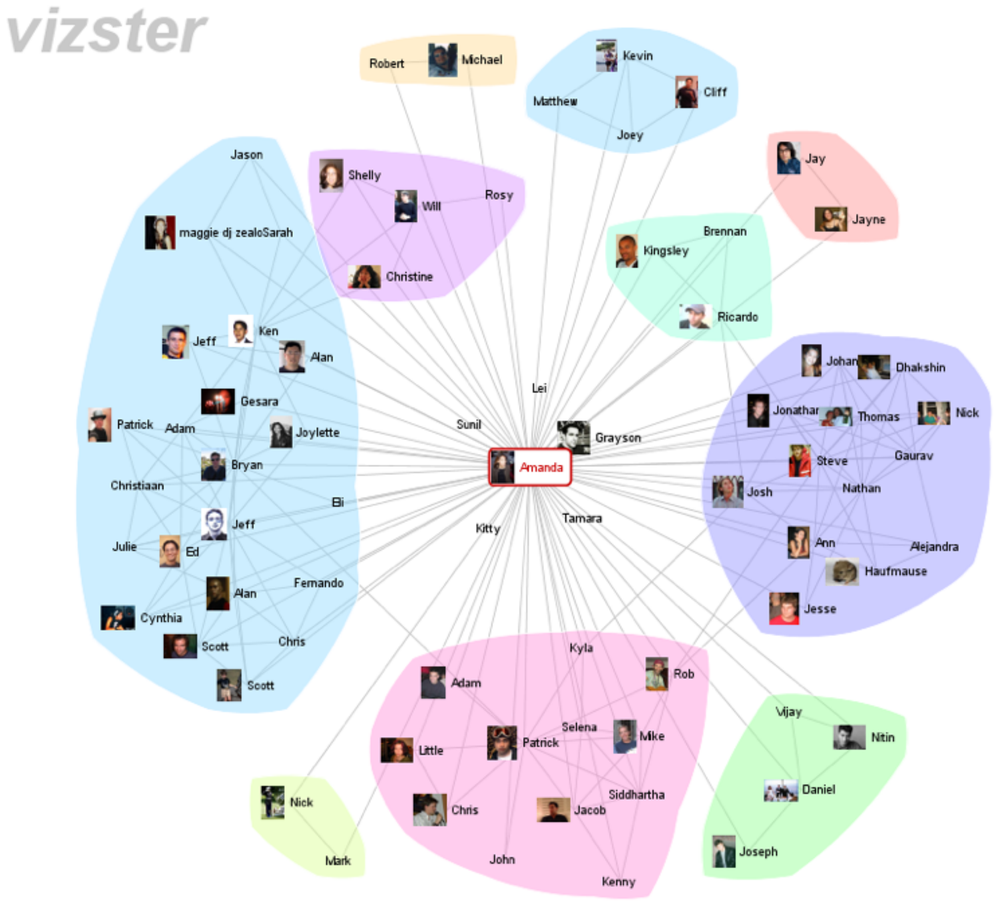
Image: http://www.orbitalrpm.com/wp-content/uploads/2007/10/sna-graphic-7.jpg

Previous work:
Social network analysis



Task abstraction:

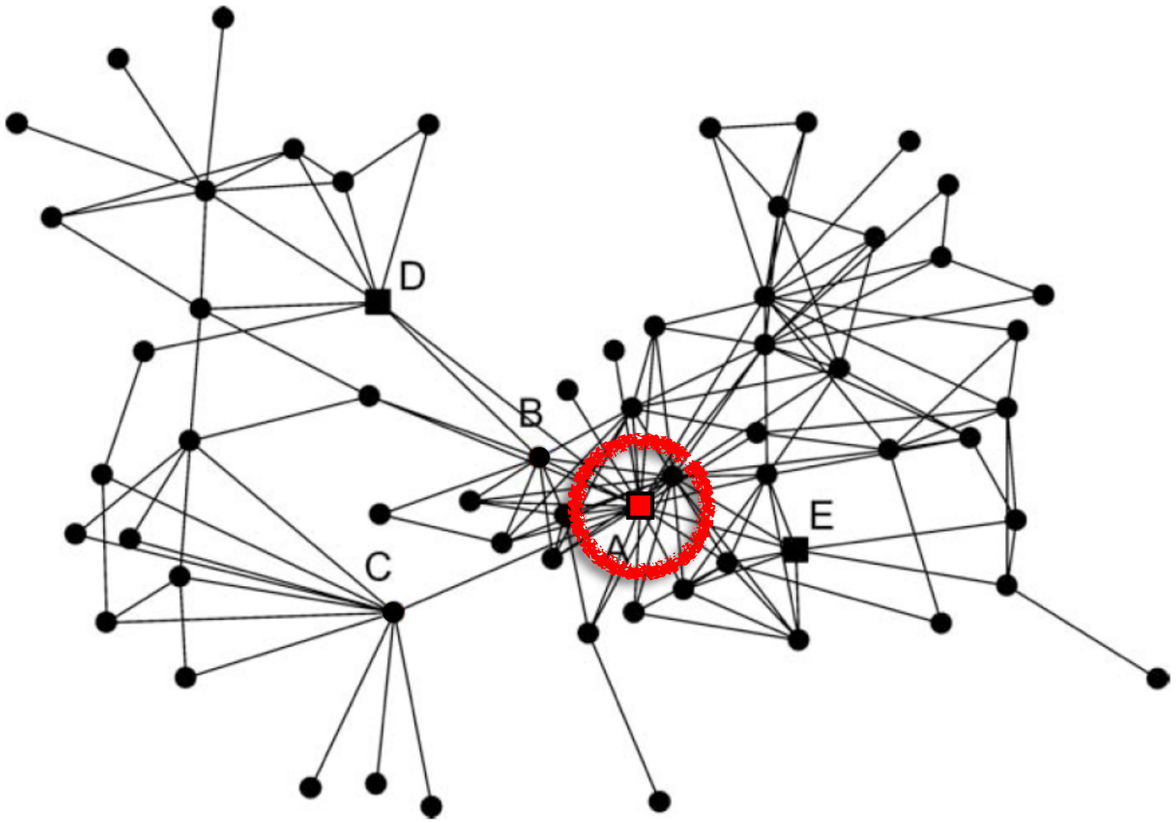
- find clusters



Previous work:
Social network analysis

Task abstraction:

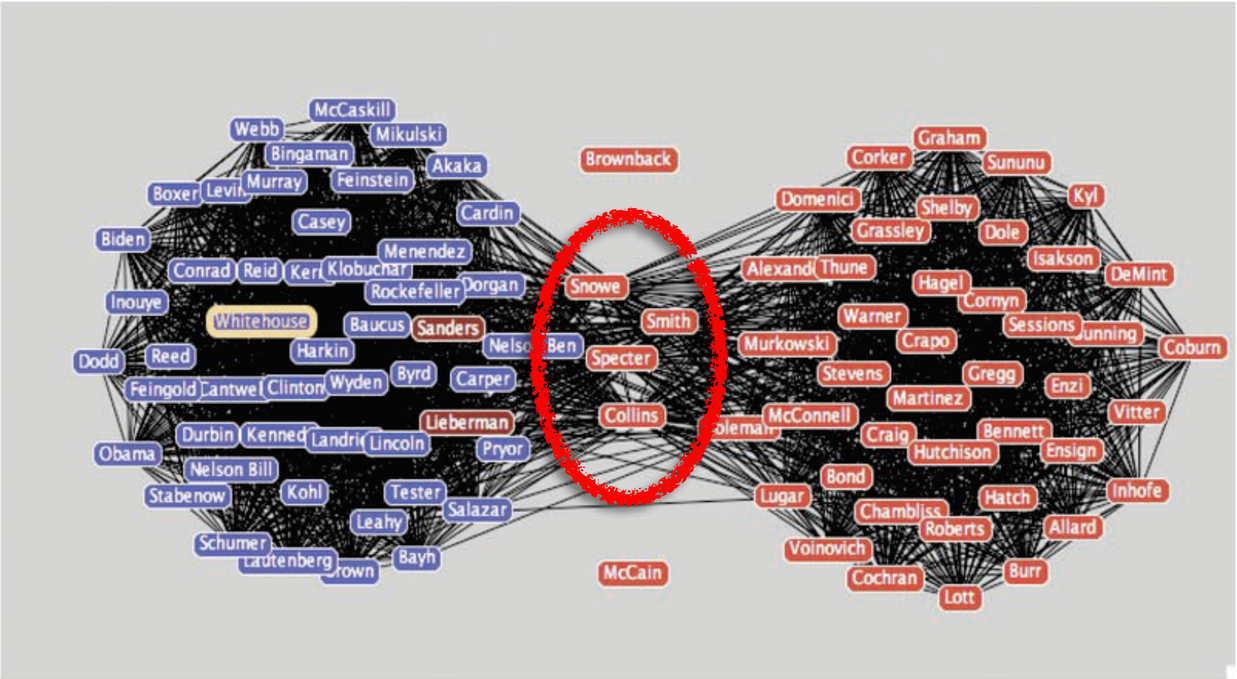
- find clusters
- find high-degree nodes



Previous work:
Social network analysis

Task abstraction:

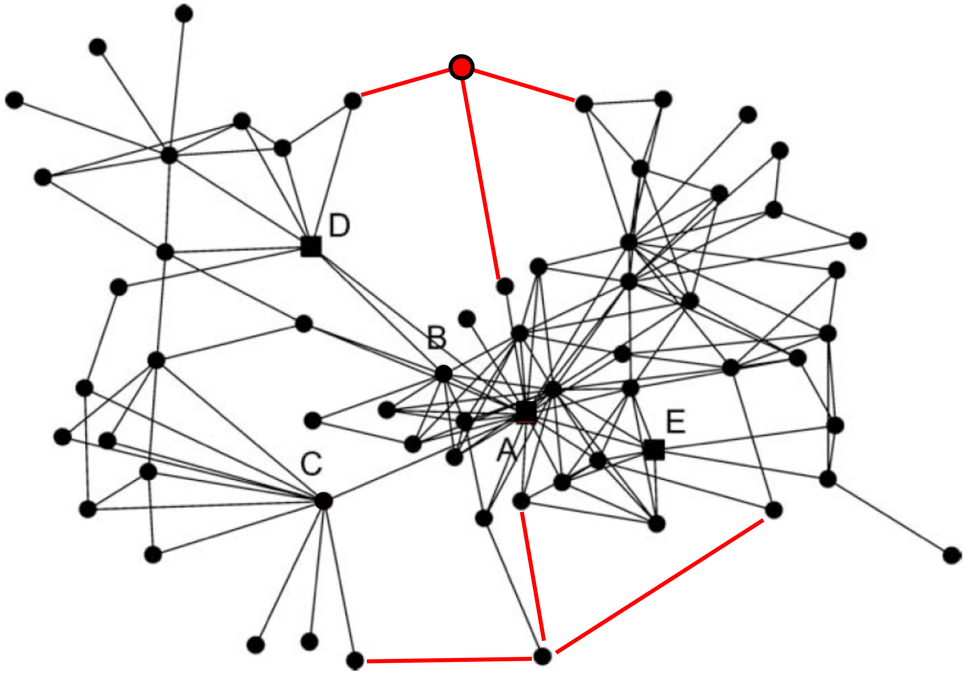
- find clusters
- find high-degree nodes
- find bridging nodes



Previous work:
Social network analysis

Task abstraction:

- find clusters
- find high-degree nodes
- find bridging nodes
- understand temporal dynamics
  - passively notice changes



Abstraction Innovation

Social network analysis

Tasks

- find clusters, high-degree/bridge nodes
- passive changes

Data

- single network

- node scalability
  - sparse edges

Overlay network optimization

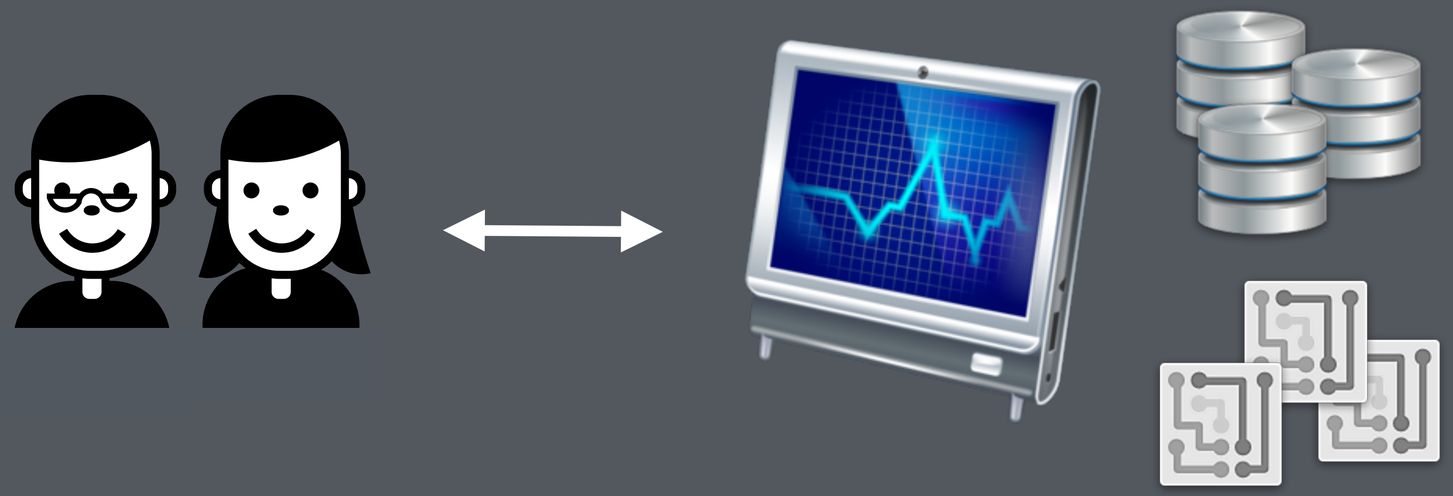
Tasks

- traffic optimization
- active changes

Data

- three related networks
  - physical, logical, overlay
- path scalability
  - dense edges, few nodes

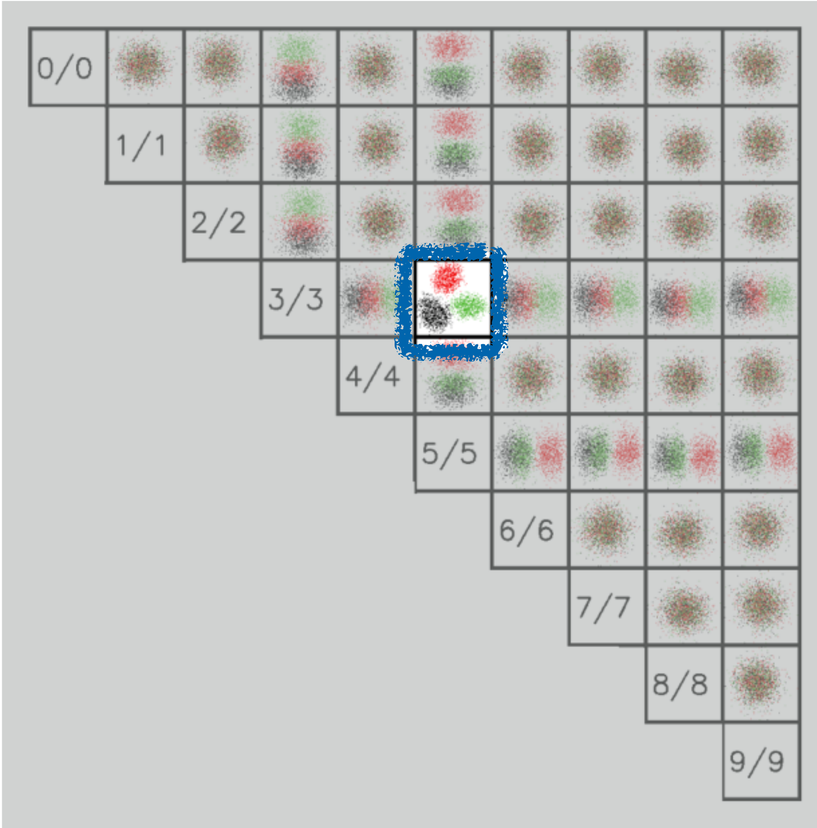
The way forward:
Why design studies matter?



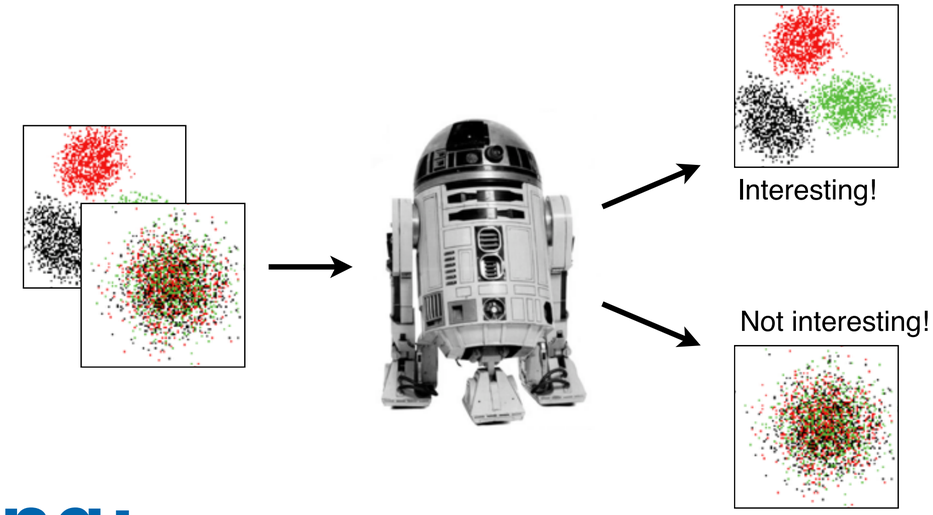
Why design studies matter?

1.
Inform & ground technique-driven work

Example:
Visual separation measures



Albuquerque et al. Synthetic Generation of High-Dimensional Datasets, TVCG 2011

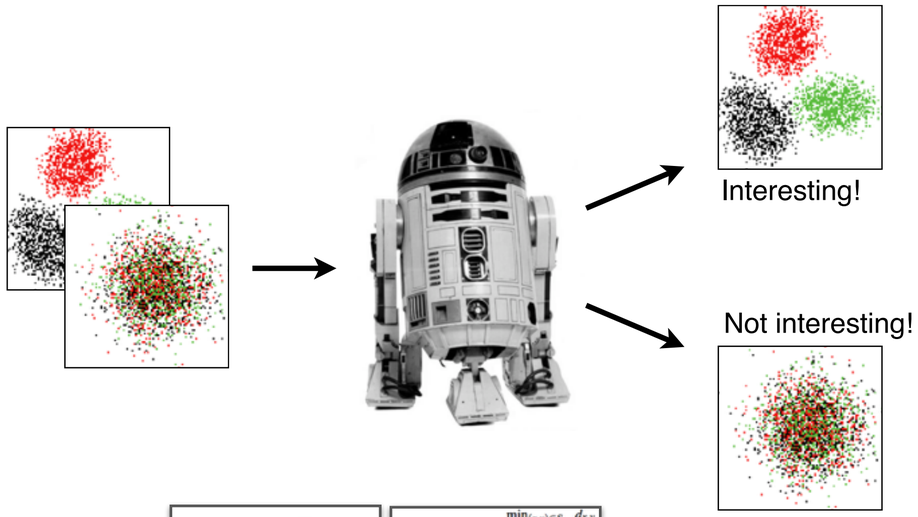


Finding:
State-of-the-art measures do not perform well under real conditions

Sedlmair, Tatu, Munzner, Tory
A Taxonomy of Visual Cluster Separation Factors
Computer Graphics Forum (Proc. EuroVis), 31(3): 1335-1344, 2012.

Example:
Visual separation measures

modeling human perception, grounded in real data



$$AWTN(\mathcal{C}_s) = \frac{\sum_{u \in \{t,o\}} \mathcal{N}_u D_{\mathcal{C}_u}}{\sum_{u \in \{t,o\}} \mathcal{N}_u}$$

$$DUNN(\mathcal{C}_s) = \frac{\min_{(x,y) \in S_{inc}} d_{x,y}}{\max_{(x,y) \in S_{sim}} d_{x,y}}$$

$$DC(\mathcal{C}_s) = 1 - \frac{\sum_{z \in \mathcal{T}} |eNN(z)| H_{eNN}(z)}{\sum_{z \in \mathcal{T}} |eNN(z)|}$$

$$CS(\mathcal{C}_s) = \frac{1}{N_s} \sum_{u \in \{t,o\}} \sum_{x \in \mathcal{C}_u} \frac{|ExMST(x) \cap \mathcal{C}_u|}{|ExMST(x)|}$$

$$CDM(\mathcal{C}_s) = \sum_{z \in \mathcal{T}} |\rho_{\mathcal{C}_t}(z,K) - \rho_{\mathcal{C}_o}(z,K)|$$

$$DC(\mathcal{C}_s) = 1 - \frac{\sum_{z \in \mathcal{T}} |eNN(z)| H_{eNN}(z)}{\sum_{z \in \mathcal{T}} |eNN(z)|}$$

$$GAM(\mathcal{C}_s) = \frac{d^+ - d^-}{d^+ + d^-}$$

$$CAL(\mathcal{C}_s) = \frac{\sum_{u \in \{t,o\}} N_u d^2_{\mathcal{C}_u, \mathcal{C}_s}}{\sum_{u \in \{t,o\}} \sum_{x \in \mathcal{C}_u} d^2_{\mathcal{C}_u, x}}$$

etc. etc.

Sedlmair, Munzner, Tory
Empirical Guidance on Scatterplot and Dimension Reduction Technique Choices.
IEEE TVCG (Proc. InfoVis), 19(12): 2634-2643, 2013.

Sedlmair, Aupetit
Data-driven Evaluation of Visual Quality Measures.
Computer Graphics Forum (Proc. EuroVis), 34(3): 201-210, 2015.

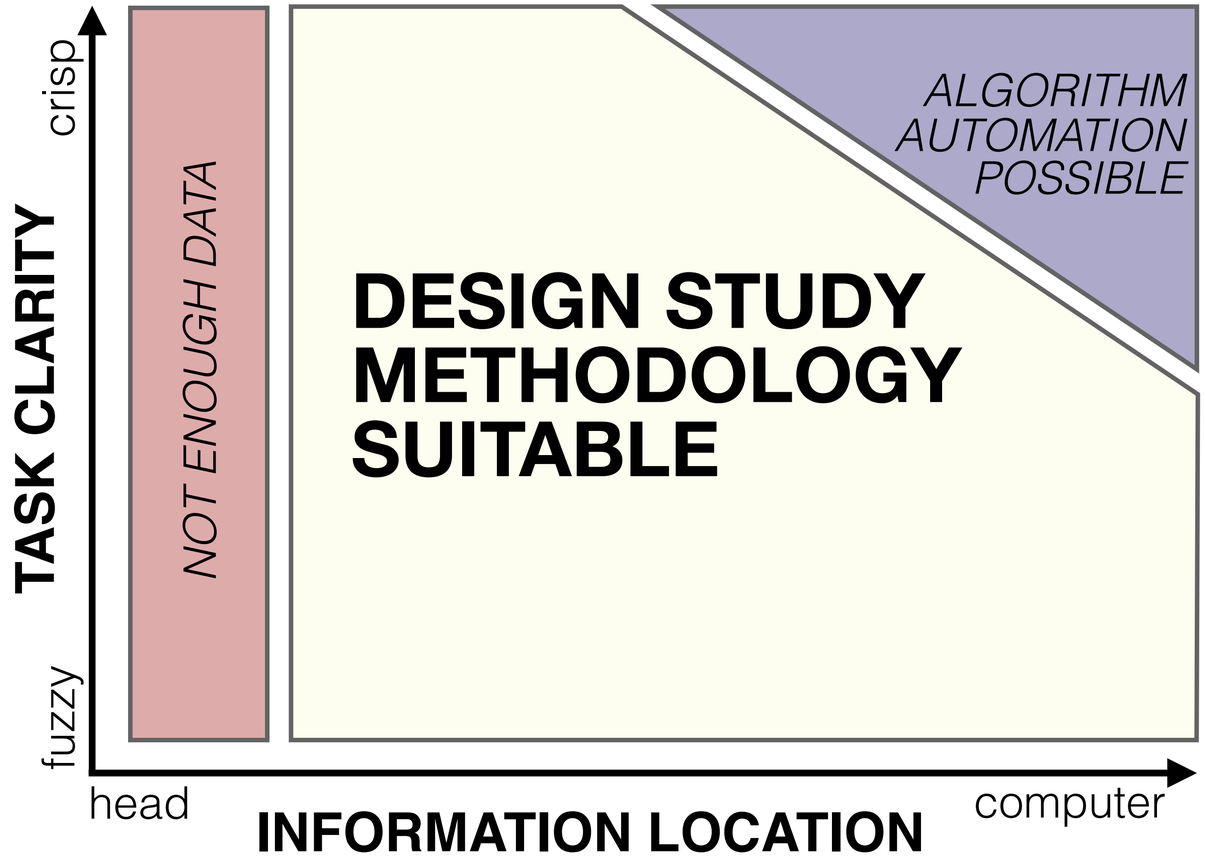
Aupetit, **Sedlmair**
SepMe: 2002 New Visual Separation Measures
IEEE Pacific Visualization Symposium, 2016 (in submission).

Wang, Feng, Chu, Zhang, Fu, **Sedlmair**, Yu, Chen
A Perception-Driven Approach to Supervised Dimensionality Reduction for Visualization.
IEEE TVCG, 2017, to appear.

Why design studies matter?

2.
Better understand ill-defined analysis problems

Design study contributions



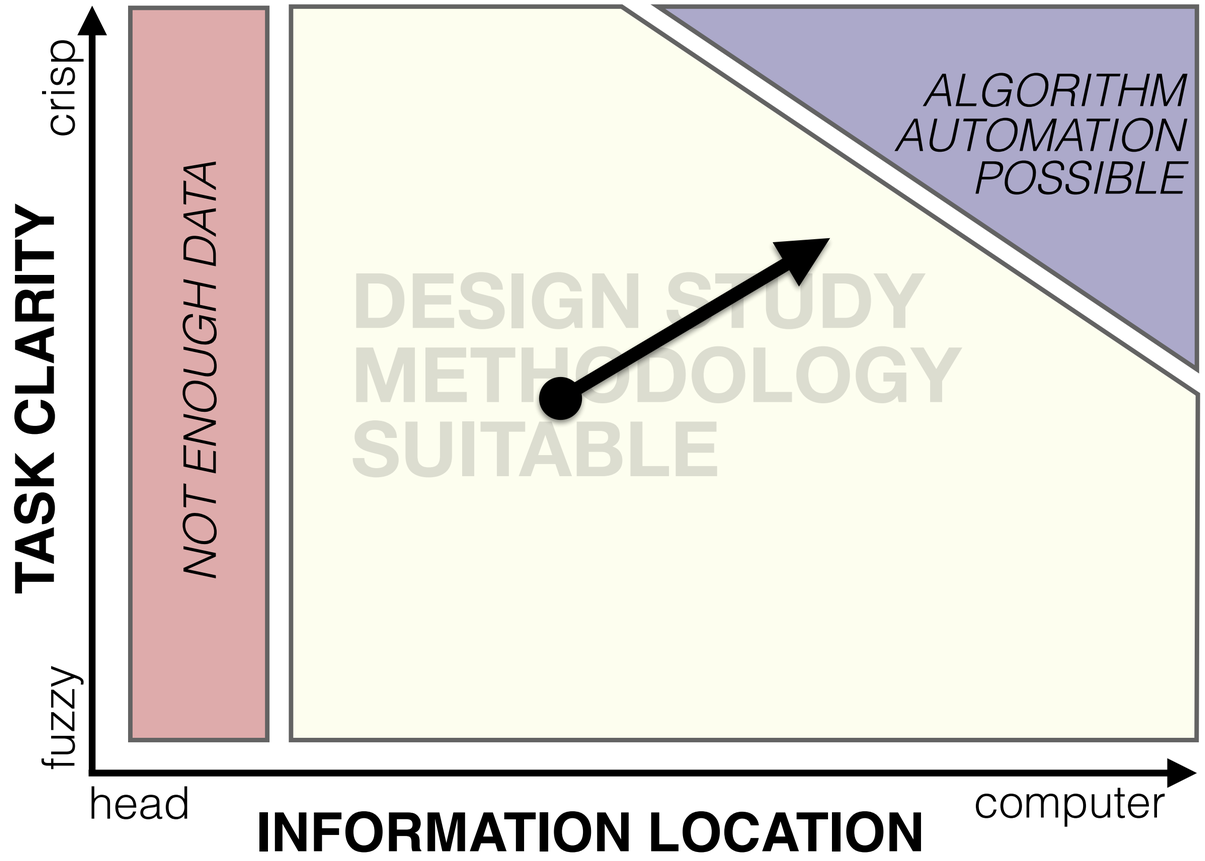
ill-defined tasks:

- multi-objective mapping (RelEx)
- exploratory analysis of scientific problems
- decision-making problems
- model building & validation

not all information externalized

Sedlmair, Meyer, Munzner.
Design Study Methodology: Reflections from the Trenches and the Stacks
IEEE TVCG (Proc. InfoVis), 18(12): 2431–2440, 2012.

Design study contributions



Contributions:

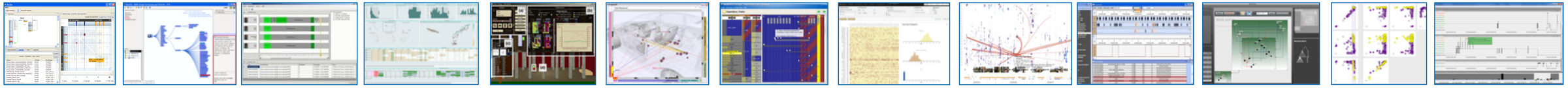
- visual analysis tool
- reflection
- **problem characterization**

Sedlmair, Meyer, Munzner.
Design Study Methodology: Reflections from the Trenches and the Stacks
IEEE TVCG (Proc. InfoVis), 18(12): 2431–2440, 2012.

Why design studies matter?

3.
Refine methods

No one size fits all method
Work on different problems!



Sedlmair et al., A Dual-View Visualization of In-Car Communication Processes, *IV 2008.*

Ruhland et al., LibViz: Data Visualisation of the old Library, *VSMM 2008.*

Sedlmair et al., User-centered Development of a Visual Exploration System for In-Car Communication, Smart Graphics, *2009.*

Sedlmair et al., MostVis: An Interactive Visualization Supporting Automotive Engineers in MOST Catalog Exploration, *IV 2009.*

Baur et al., The Streams of Our Lives: Visualizing Listening Histories in Context, *IEEE TVCG (Proc. InfoVis) 2010.*

Sedlmair et al., Cardiogram: Visual Analytics for Automotive Engineers, ACM CHI *2011.*

Sedlmair et al., RelEx: Visualization for Actively Changing Overlay Network Specifications, *IEEE TVCG (Proc. InfoVis) 2012.*

Bergner et al., Paraglide: Interactive Parameter Space Partitioning for Computer Simulations, *IEEE TVGC 2013.*

Hund et al., Visual Analytics for Concept Exploration in Subspaces of Patient Groups, *to appear, BRIN 2016.*

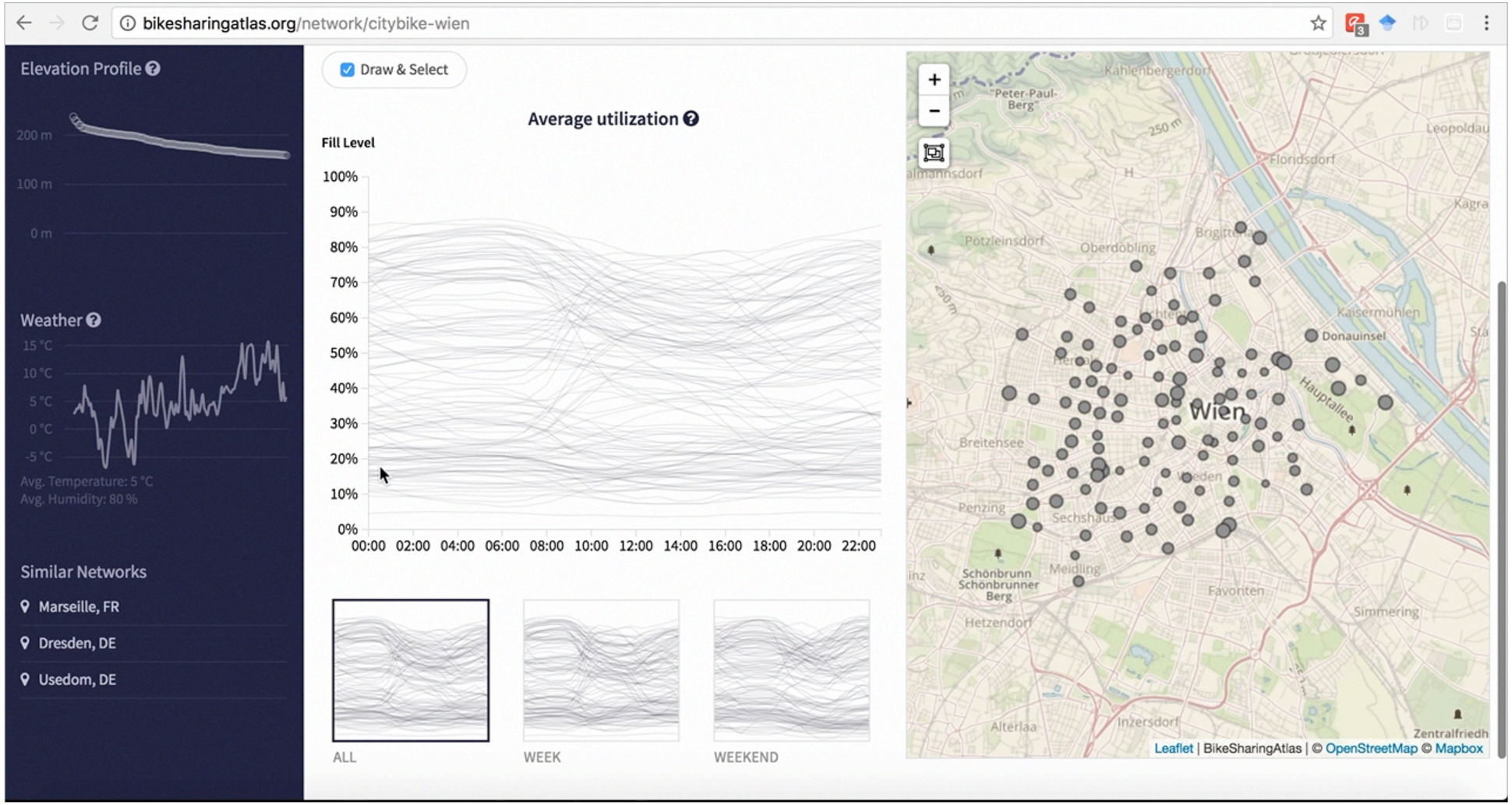
Another Example: Mobility!

bikesharingatlas.org

Oppermann, Möller, **Sedlmair**.
ongoing work.

Another Example: Mobility!

bikesharingatlas.org



Another Example: Mobility!

bikesharingatlas.org

Oppermann, Möller, **Sedlmair**.
ongoing work.

Methodologically:

- started with data
- different user groups
  (network operators, city planners, politicians, journalists, general public, etc.)
- problem generalization
  (car data, etc.)

Why design studies matter?

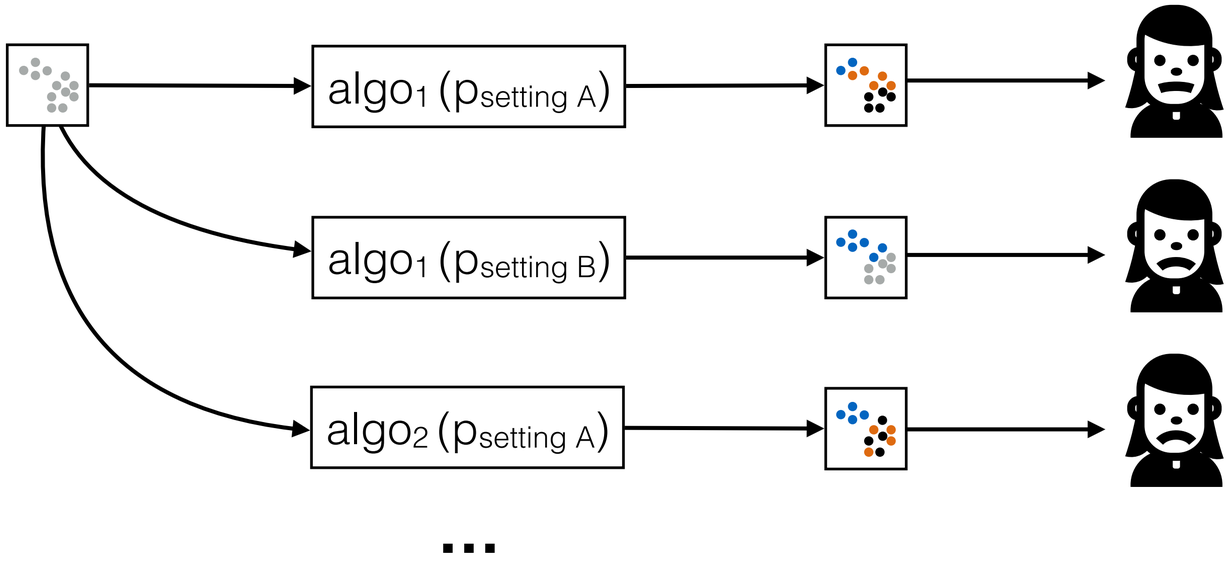
4.
Build visualization theory

Goal

Derive theory from multiple real-world design studies

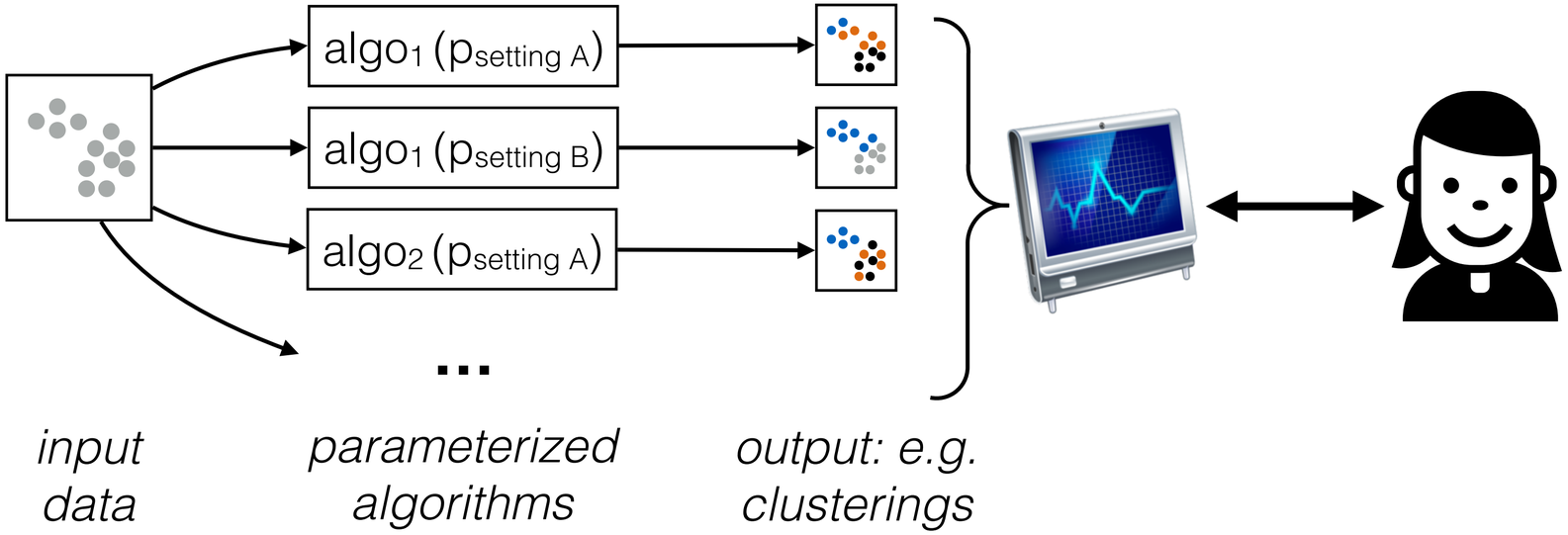
Example:
Visual parameter space analysis

- traditional approach: trial and error (e.g., clustering)

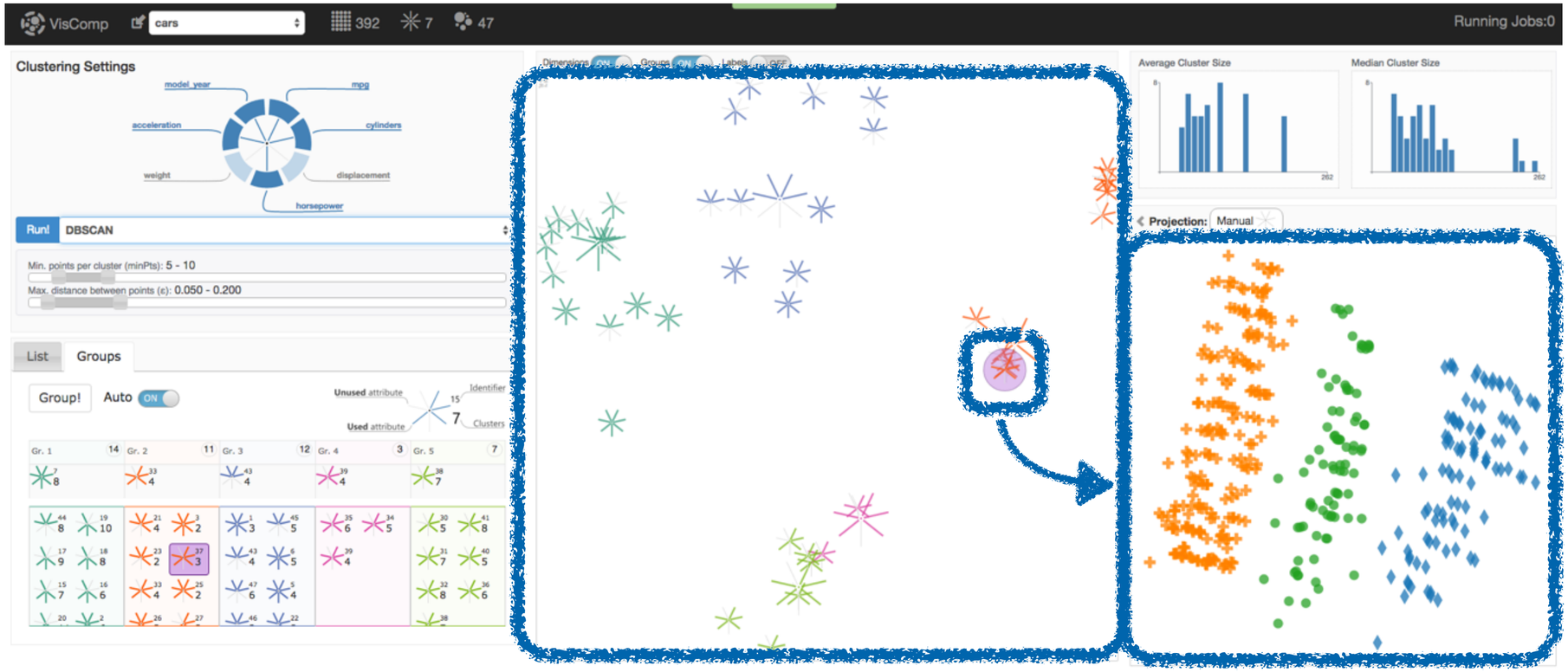


Example:
Visual parameter space analysis

- traditional approach: trial and error (e.g., clustering)
- instead: visual parameter space analysis



Example:
Visual parameter space analysis



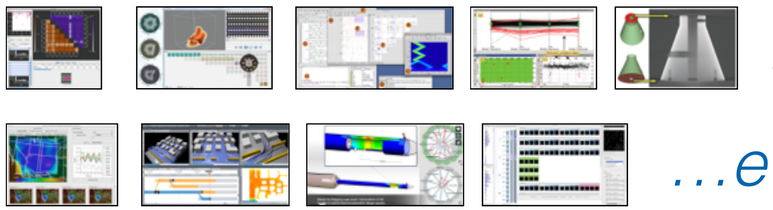
Herzner, Stusak, **Sedlmair.**
ClustComp: Visual Comparison of Clusterings.
Master thesis, 2014.

Overview of clusterings (distance = similarity)

Show cluster results (upon selecting one)

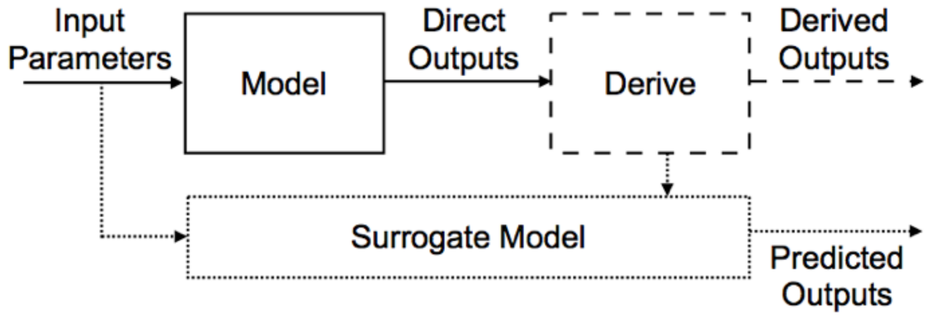
Example:
Visual parameter space analysis

Theoretical framework
(from meta analysis of 23 design study papers)



…etc. (us)

…etc. (others)

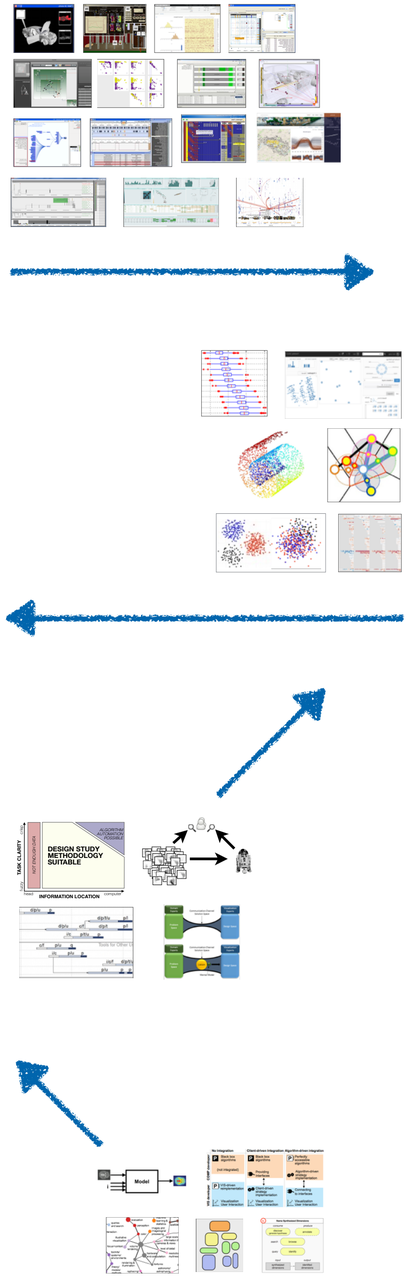


Sedlmair, Heinzl, Bruckner, Piringer, Möller.
Visual Parameter Space Analysis: A Conceptual Framework.
IEEE TVCG (Proc. InfoVis), 20(12): 2161-2170, 2014.

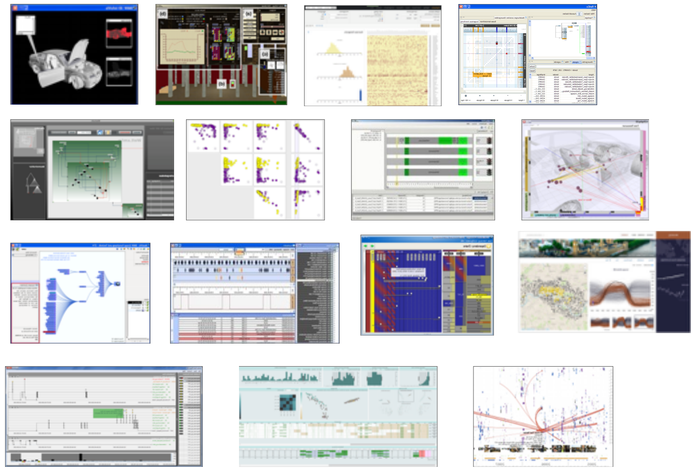
The way forward:
Design studies as a core building block of visualization research

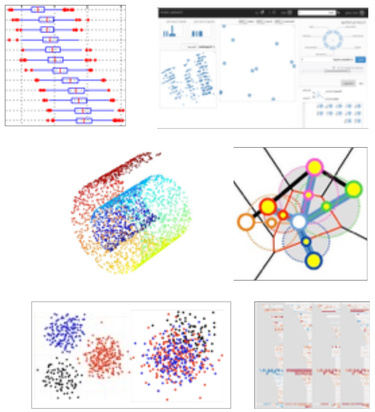
The way forward

- Need for design studies in different domains
- Come from different angles (problem- **and** technique-driven)
- Need to further improve methodology
- Build up much needed theoretical underpinnings



My Work



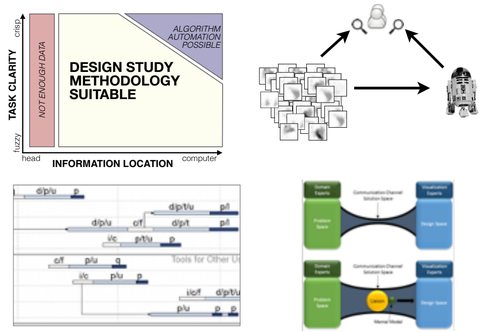


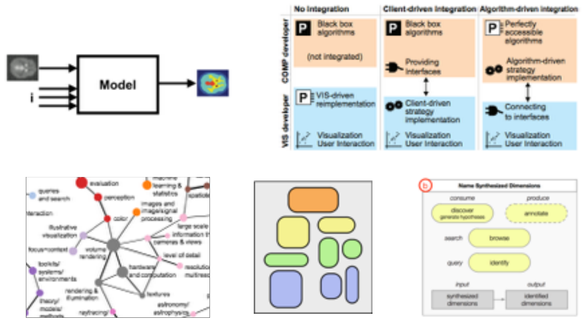
problem-driven

technique-driven

research methods

theory





Vision: “Visualization Cookbook”

abstract data & task

↕

data analysis technique

Vision: “Visualization Cookbook”

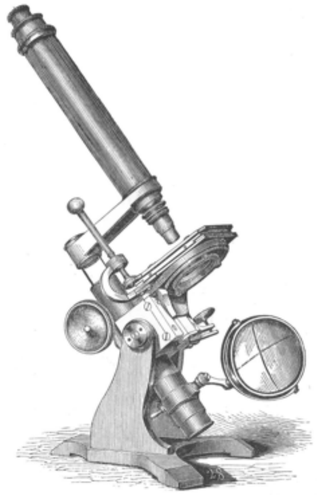
vast amounts of design studies necessary.

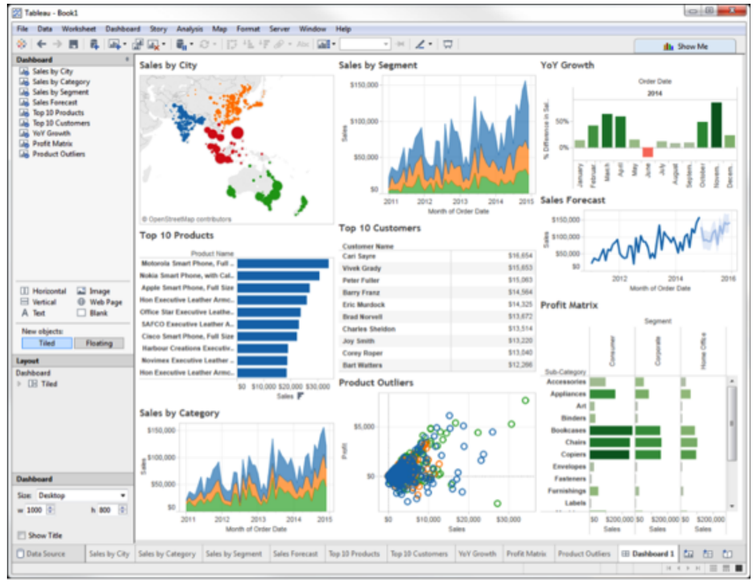
natural science vs. design research

Sir Francis Bacon 1561-1626

Vision: “Visualization Cookbook”

new and more powerful analysis tools for humans.

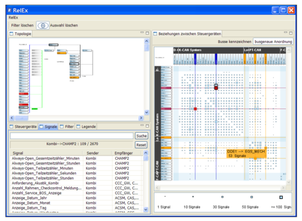




Tableau

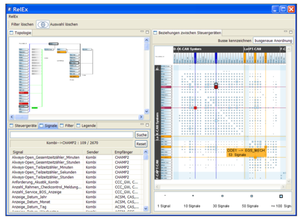
Summary

Visualization Design Studies

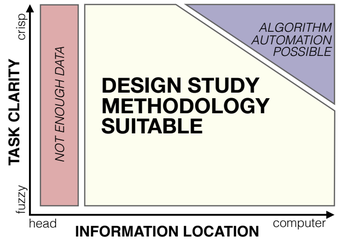


- RelEx Design Study
  - impact: adopted by BMW engineers

Visualization Design Studies



- RelEx Design Study
  - impact: adopted by BMW engineers



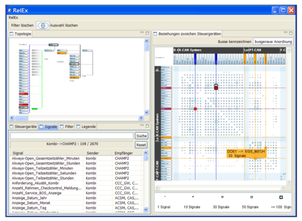
- Design Study Methodology
  - impact: used and well-cited

```
hi folks,

I just read the design study paper by sedlmair (thanks manu for sending
it to me) and it was quite revealing for me. I am so glad that I read it
at the beginning of my phd. If you have any papers that you consider as
a must read for every phd-student, then pls let know.

haichao
```

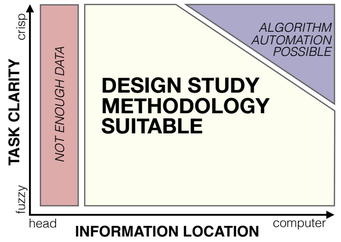
Haichao Miao,
PhD student, TU Vienna
(forwarded by Manu Waldner)

bold and color added

Visualization Design Studies



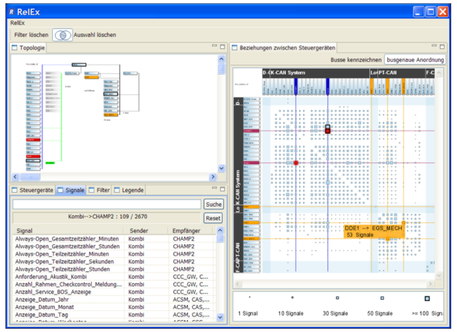
- RelEx Design Study
  - impact: adopted by BMW engineers

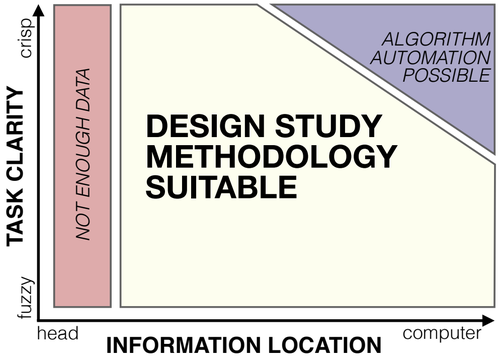


- Design Study Methodology
  - impact: used and well-cited

- The way forward
  - why design studies matter

Thank you!





Michael Sedlmair

Visualization & Data Analysis Group

universität wien

Appendix

Role of a Data Scientist

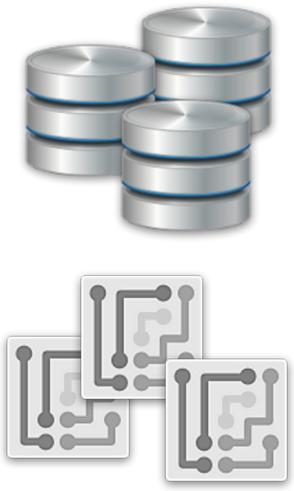
Current Approach

Scientists

Engineers

Journalists

Decision Makers

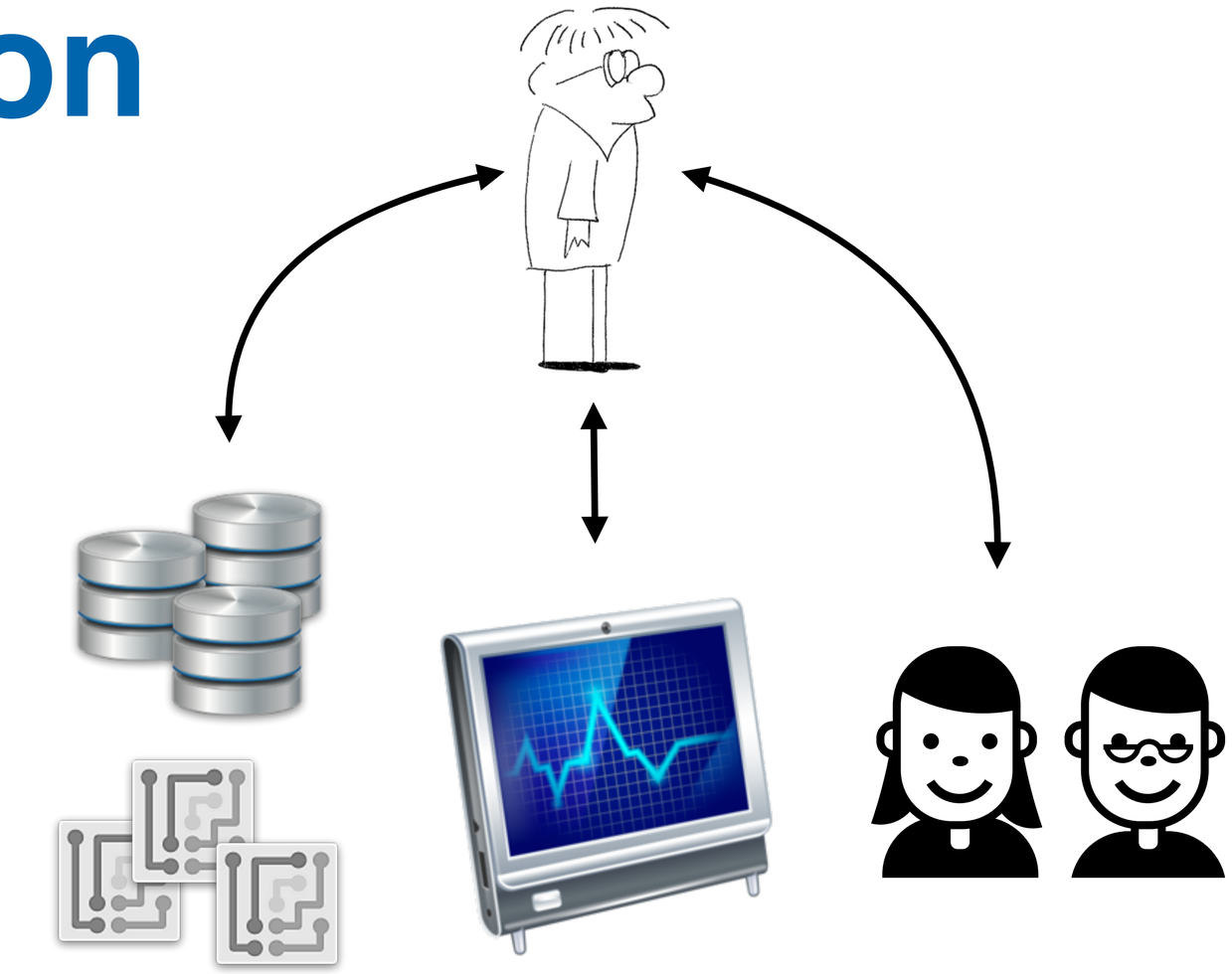


Data &
Algorithms

Data/Vis
Scientist

Domain
User

My vision



Data & Algorithms

Data/Vis **Tool**

Domain User

My vision:
Human-centered data science

Data &
Algorithms

Data/Vis
Tool

Domain
User

Design Study Methodology

Definition: Design Studies

"A design study is a project in which visualization researchers analyze a specific

real-world problem *faced by domain experts,*

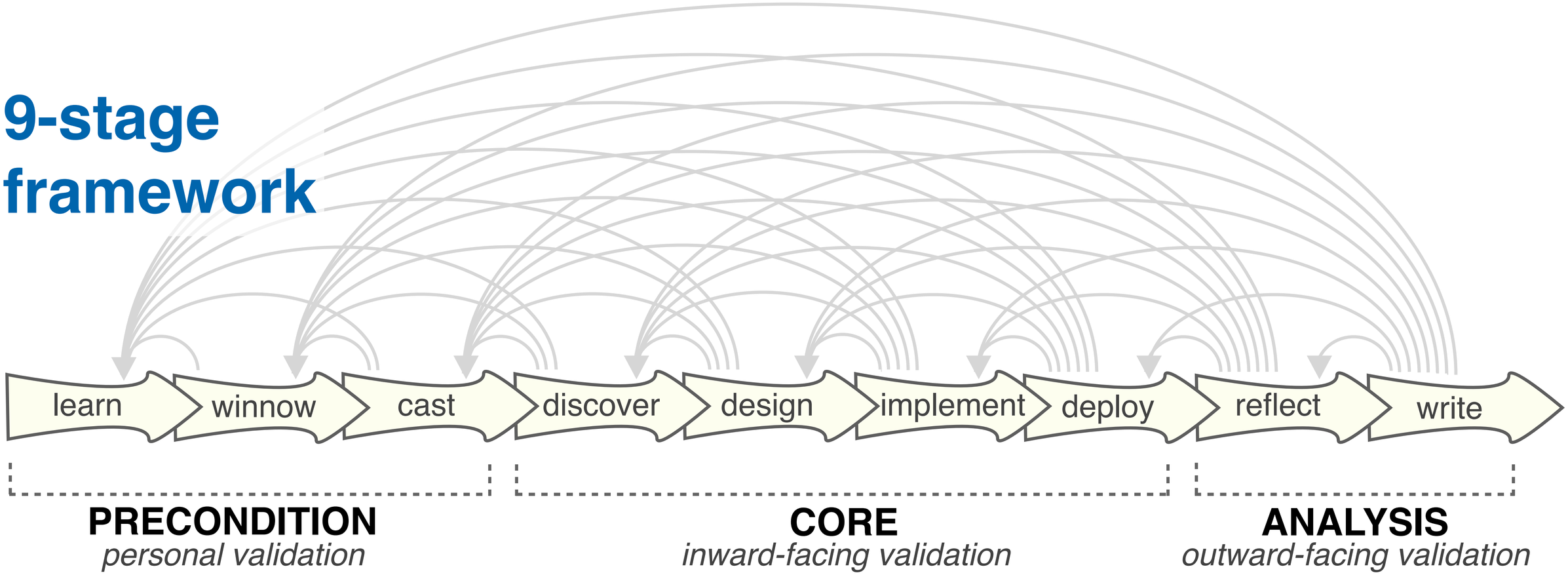
design *a visualization system that supports solving this problem,*

validate *the design, and*

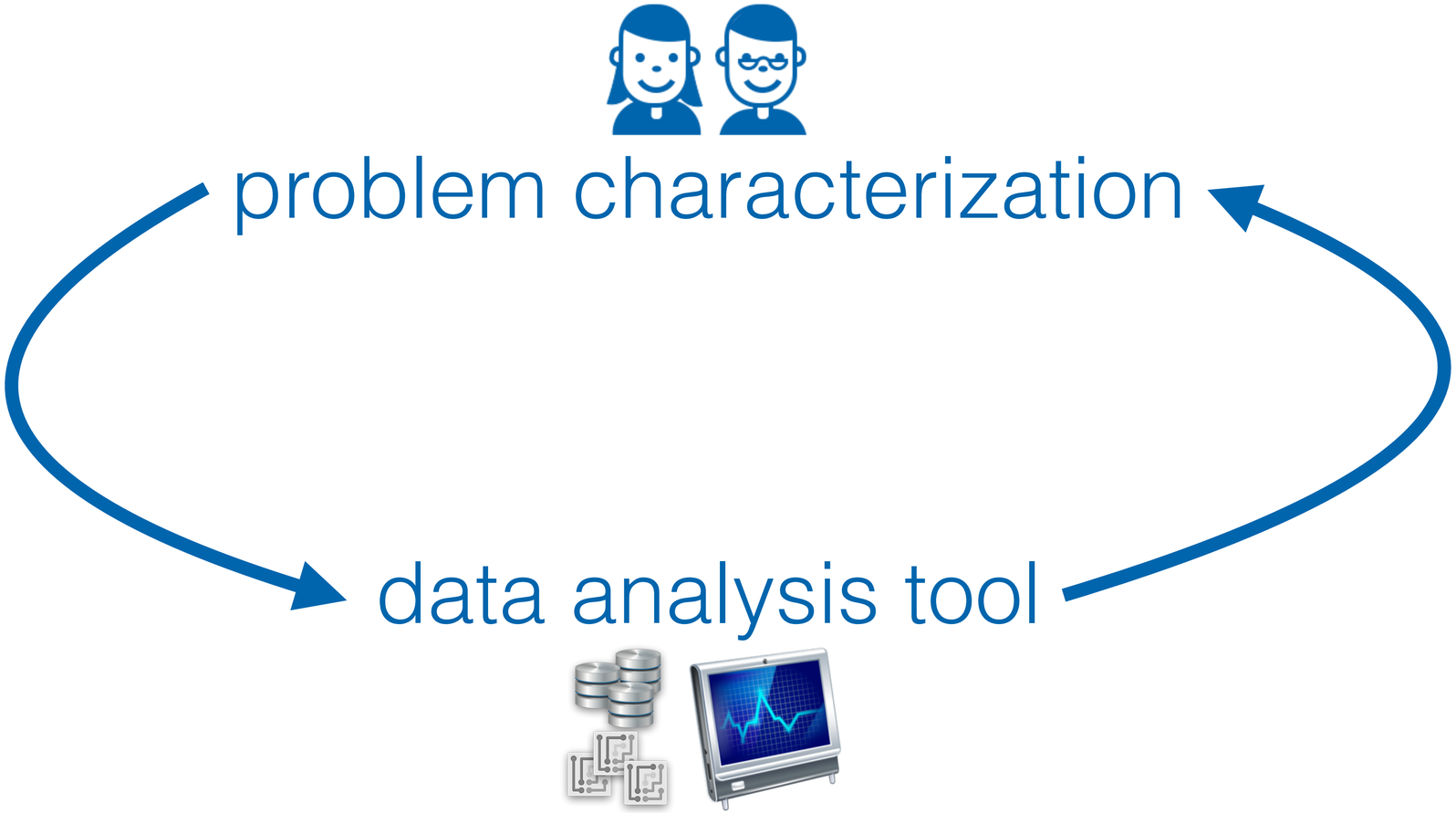
reflect *about lessons learned in order to refine visualization design guidelines."*

Why another methodology?

9-stage framework



Why not just a social scientist?



Example pitfall: premature publishing

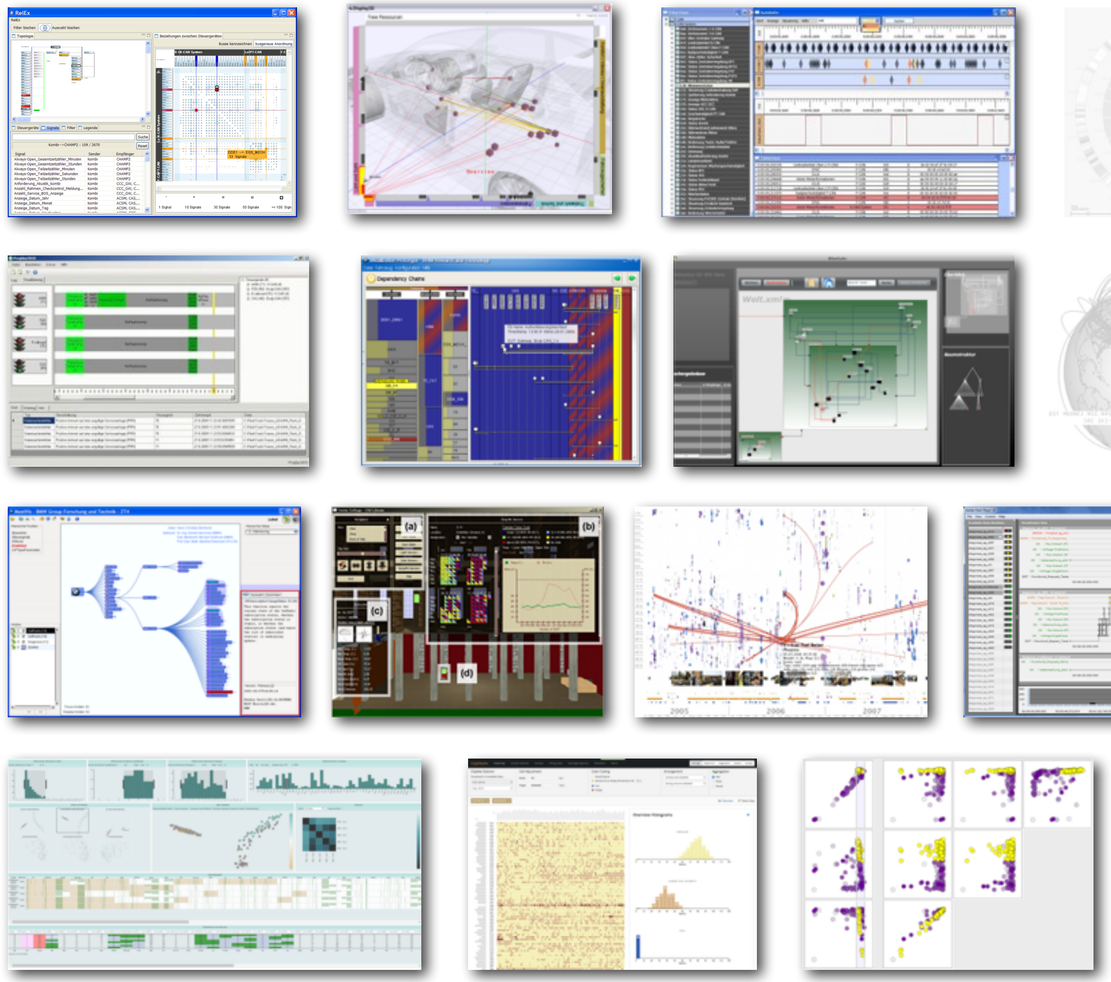
Must be first!

technique-driven

Am I ready?

problem-driven

My Design Studies



- Different domains
  - automotive engineering
  - simulation models
  - biomedical analysis
  - bike sharing
  - music listening histories

Definition of Visualization

Defintion: Visualization (Vis)

“Computer-based visualization systems provide visual representations of datasets intended to help people carry out some task more effectively.”

Tamara Munzner.
Visualization Analysis and Design.
A K Peters Visualization Series, CRC Press, 2014.

Interactive visual analysis

"Computer-based visualization systems provide visual representations of datasets intended to help people carry out some task more effectively."

Tableau

SAS

Improvise

IBM Watson Analytics

not pretty pictures

Why a human in the loop?

"Computer-based visualization systems provide visual representations of datasets intended to help people carry out some task more effectively."

not needed if fully automatic solution exists (and is trusted)

Ill-defined / ill-structured problems:

- no single optimal solution
- no clear objective measures

Herbert A. Simon.
The structure of ill-structured problems.
Artificial Intelligence, 4, 181-204. 1973.

Examples:

- exploratory analysis of scientific problems
- (collaborative) decision-making problems
- model building & validation

Effective support of ill-defined tasks!

"Computer-based visualization systems provide visual representations of datasets intended to help people carry out some task more effectively."

Metrics of success:

- Novel — entirely new insights
- Faster — speed up common workflows

Why visual representations?

"Computer-based visualization systems provide visual representations of datasets intended to help people carry out some task more effectively."

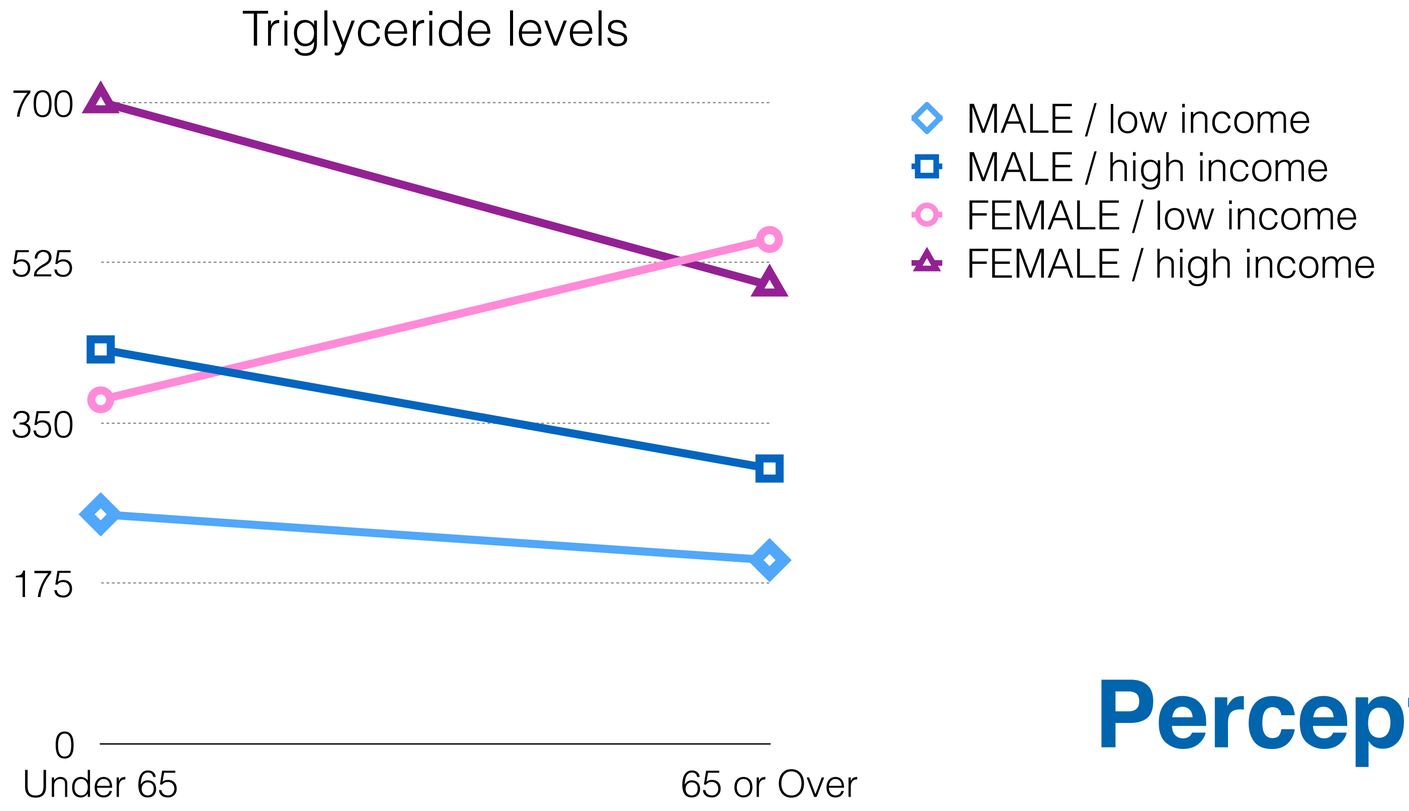
Triglyceride levels

	Males		Females	
Income Group	**Under 65**	**65 or Over**	**Under 65**	**65 or Over**
low income	250	200	375	550
high income	430	300	700	500

Q: Which gender or income level group shows different effects of age on triglyceride levels?

Why visual representations?

"Computer-based visualization systems provide visual representations of datasets intended to help people carry out some task more effectively."



Perception beats Cognition!