

Visualization: The Human Lens to Big Data

Dr. Michael Sedlmair

Visualization and Data Analysis group (VDA)



Big Data / Data Science



database

MapReduce

Hadoop



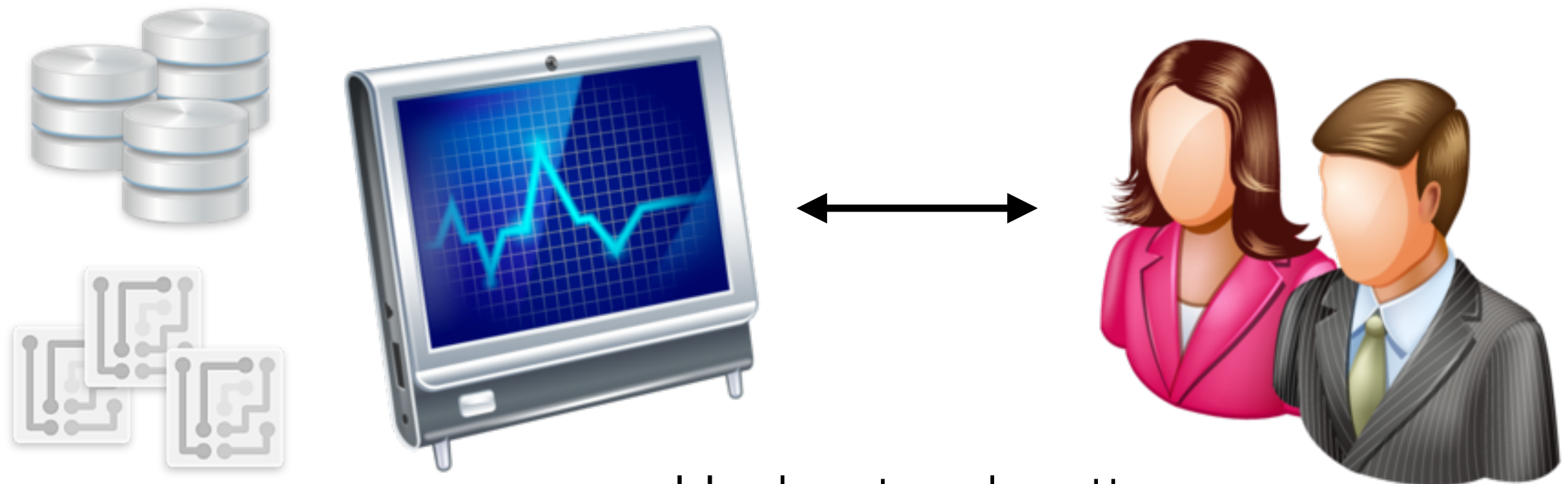
algorithms

Data Mining

Machine Learning

Visualization

Look at / Interact



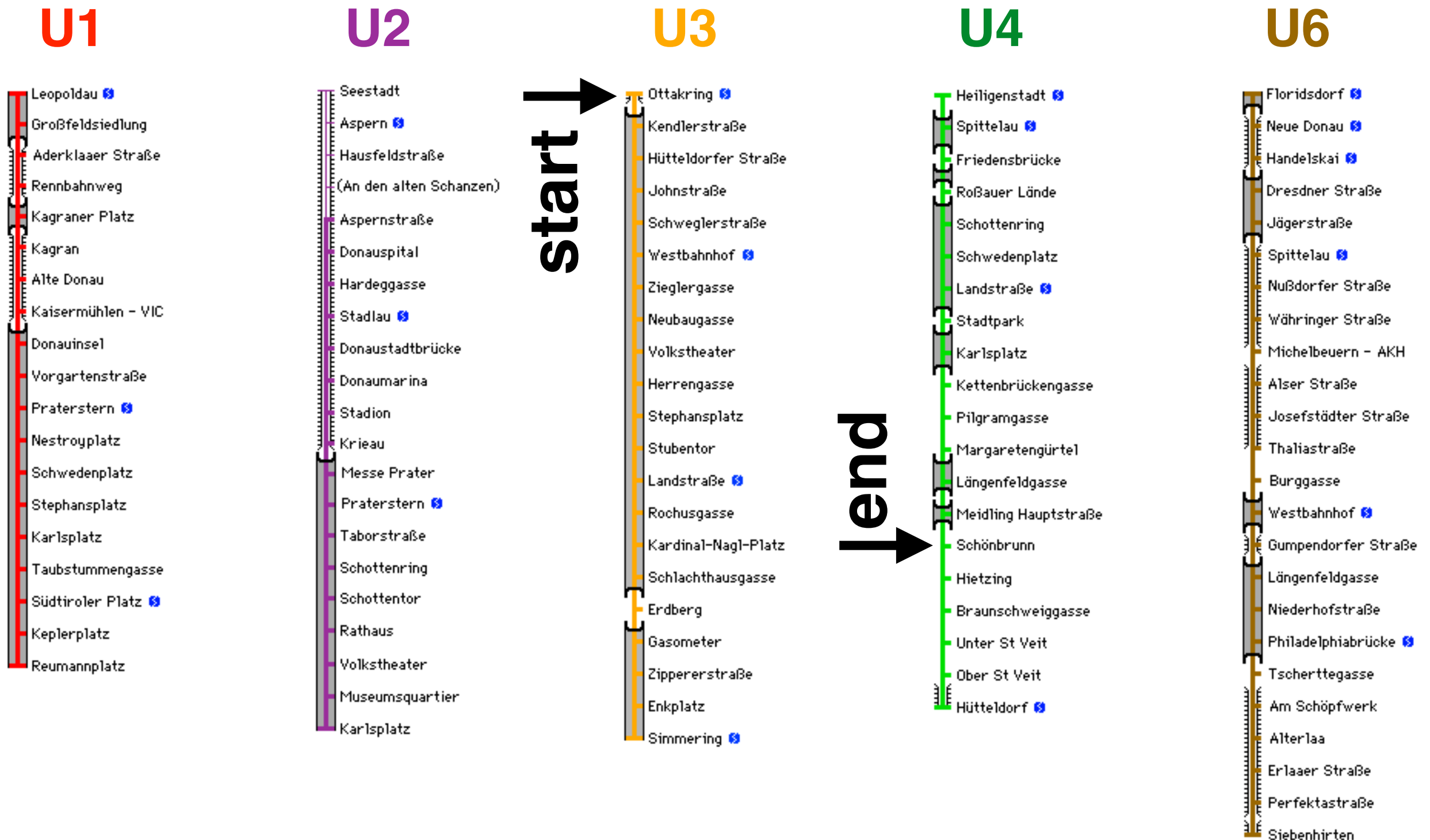
Understand patterns

Gain insights

Make decisions

Communicate

Visualization — Why?

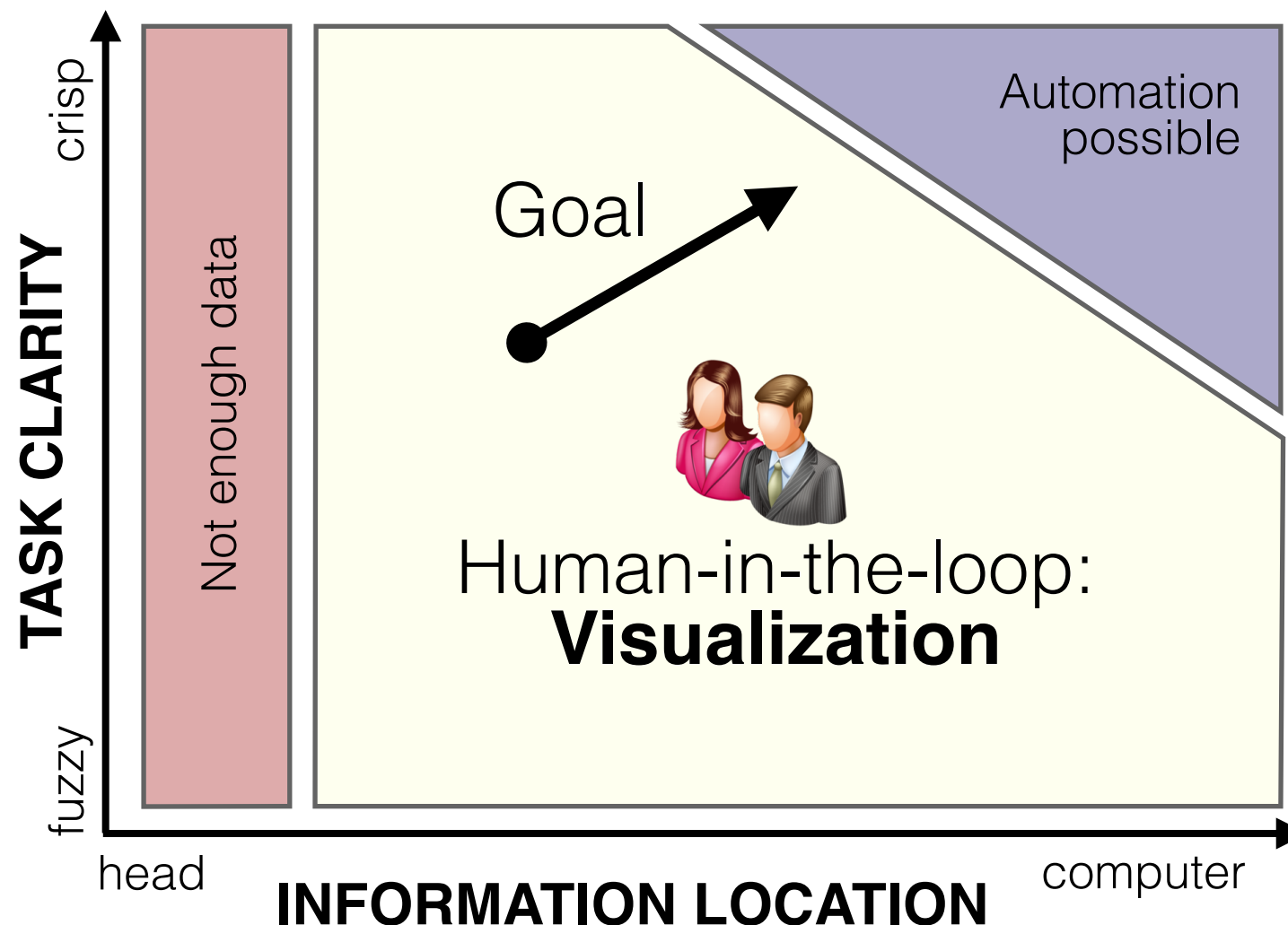


Visualization — Why?

Perception beats cognition



Wait, why humans in the loop? We have great algorithms!



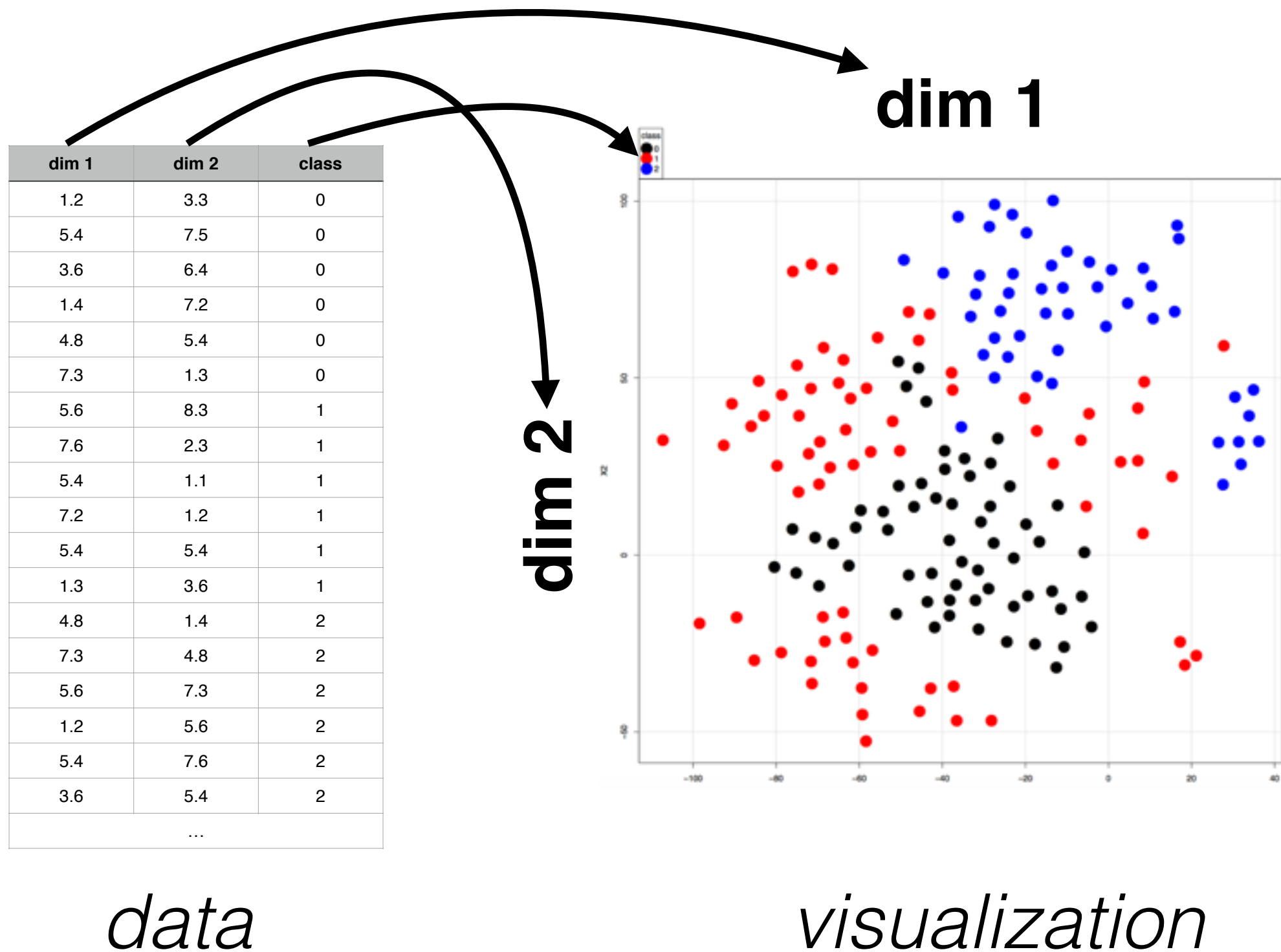
Michael Sedlmair, Miriah Meyer, Tamara Munzner
Design Study Methodology: Reflections from the Trenches and the Stacks (IEEE InfoVis 2012).

Visualization: The human lens to big data

Outline

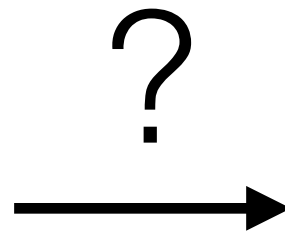
- Visual Encoding: Principle & Pitfalls
- Visualization of Big Data
- Case Study: Cardiogram

Visual Encoding



Visual Encoding: 3D?

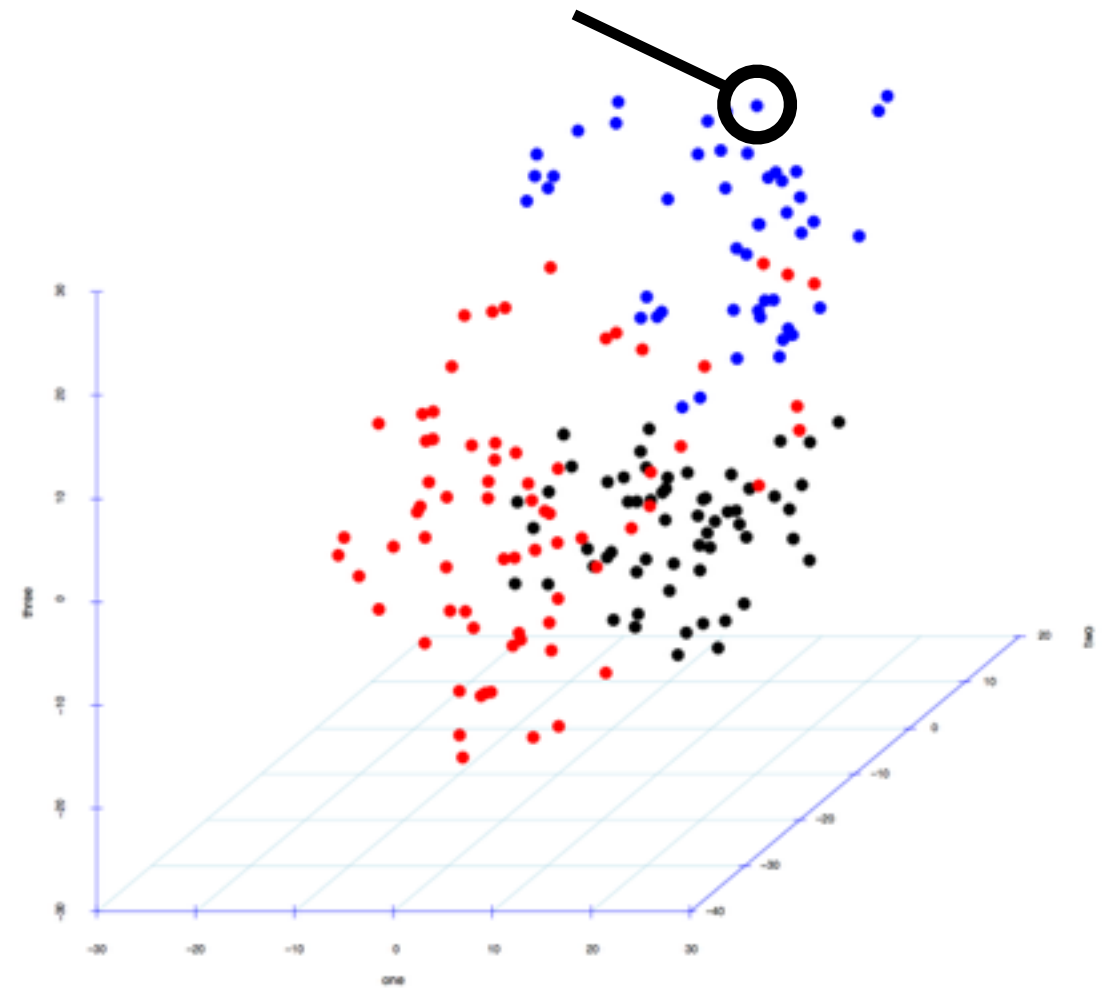
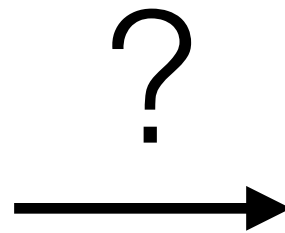
dim 1	dim 2	dim 3	class
1.2	3.3	5.4	0
5.4	7.5	1.3	0
3.6	6.4	4.8	0
1.4	7.2	7.3	0
4.8	5.4	5.6	0
7.3	1.3	1.2	0
5.6	8.3	5.4	1
7.6	2.3	3.6	1
5.4	1.1	3.3	1
7.2	1.2	7.5	1
5.4	5.4	6.4	1
1.3	3.6	7.2	1
4.8	1.4	5.4	2
7.3	4.8	1.3	2
5.6	7.3	8.3	2
1.2	5.6	3.3	2
5.4	7.6	7.5	2
3.6	5.4	1.2	2
...			



Visual Encoding: 3D?

what's the value of this point?

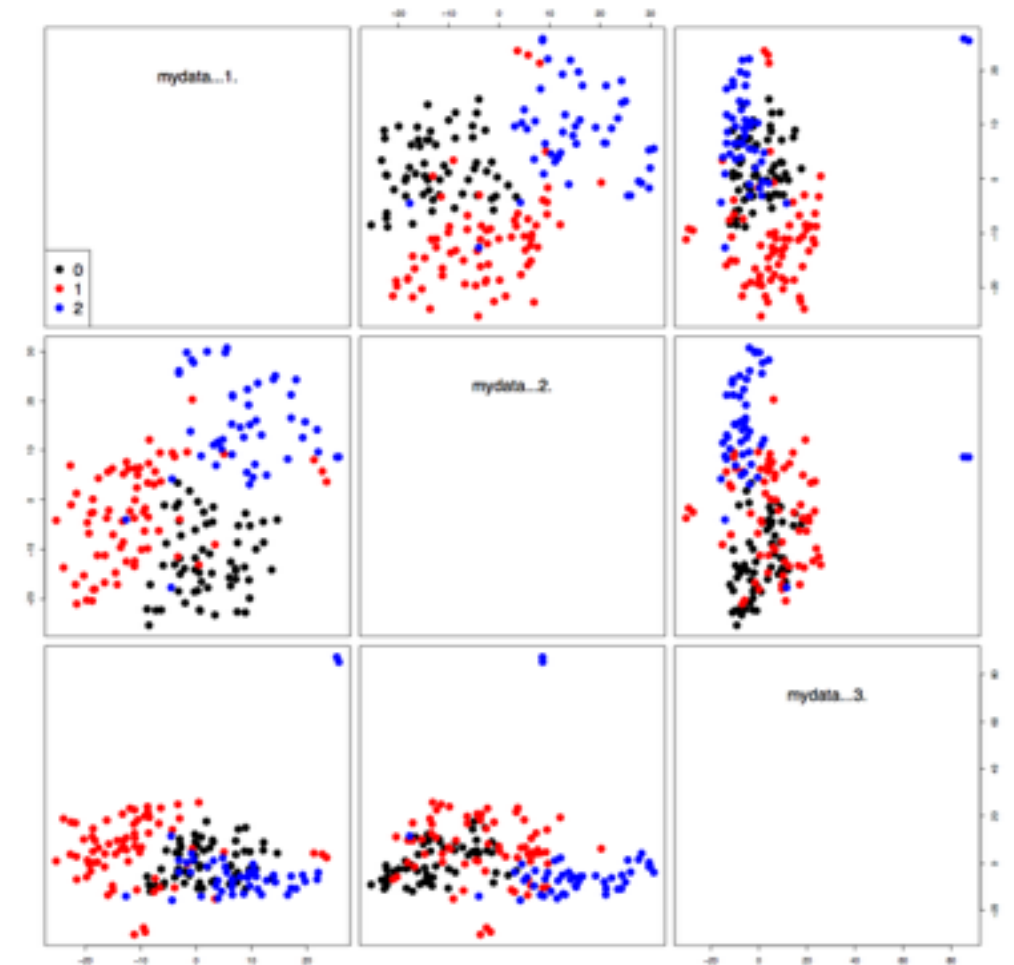
dim 1	dim 2	dim 3	class
1.2	3.3	5.4	0
5.4	7.5	1.3	0
3.6	6.4	4.8	0
1.4	7.2	7.3	0
4.8	5.4	5.6	0
7.3	1.3	1.2	0
5.6	8.3	5.4	1
7.6	2.3	3.6	1
5.4	1.1	3.3	1
7.2	1.2	7.5	1
5.4	5.4	6.4	1
1.3	3.6	7.2	1
4.8	1.4	5.4	2
7.3	4.8	1.3	2
5.6	7.3	8.3	2
1.2	5.6	3.3	2
5.4	7.6	7.5	2
3.6	5.4	1.2	2
...			



Visual Encoding: 3D?

Scatterplot Matrix

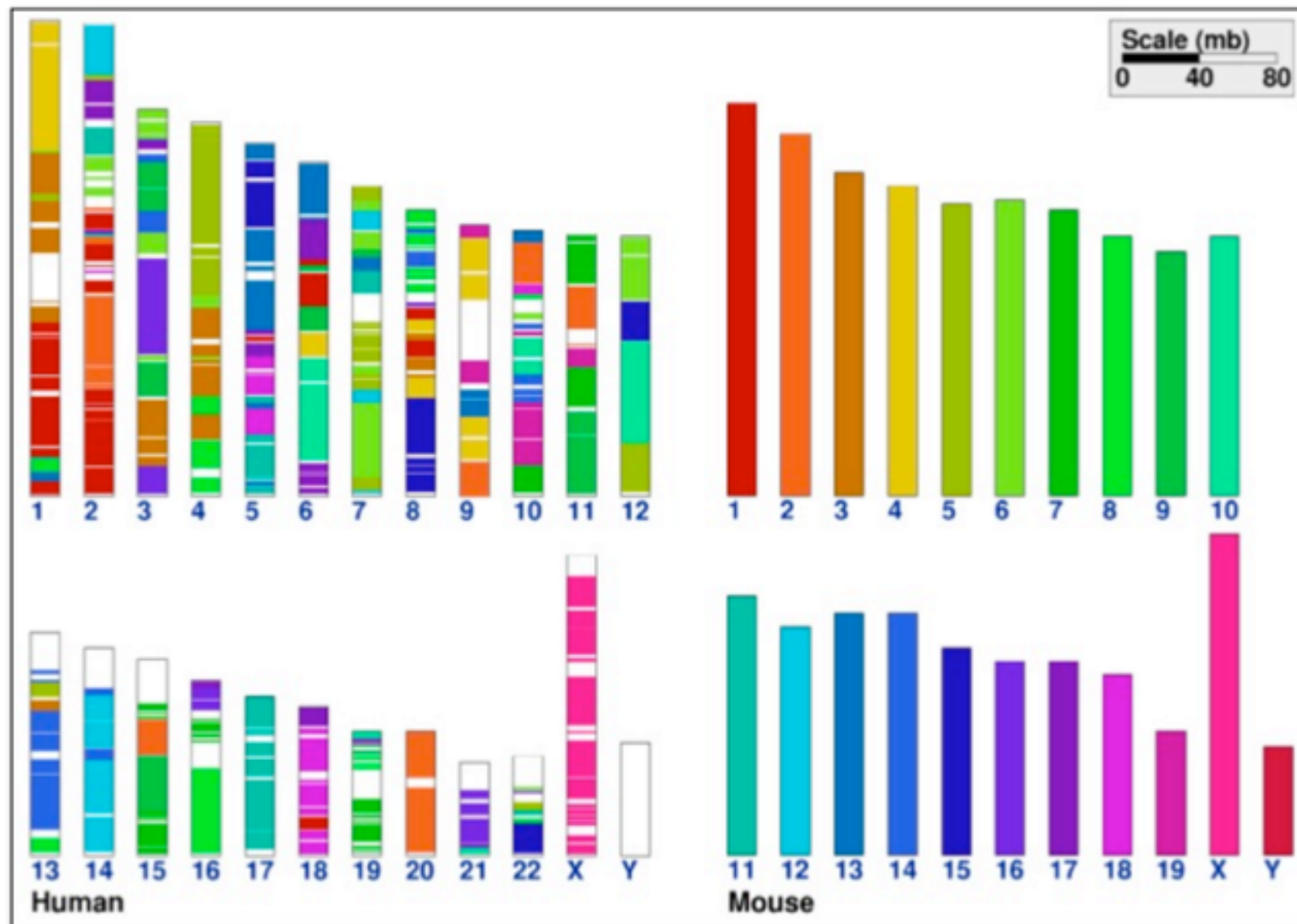
dim 1	dim 2	dim 3	class
1.2	3.3	5.4	0
5.4	7.5	1.3	0
3.6	6.4	4.8	0
1.4	7.2	7.3	0
4.8	5.4	5.6	0
7.3	1.3	1.2	0
5.6	8.3	5.4	1
7.6	2.3	3.6	1
5.4	1.1	3.3	1
7.2	1.2	7.5	1
5.4	5.4	6.4	1
1.3	3.6	7.2	1
4.8	1.4	5.4	2
7.3	4.8	1.3	2
5.6	7.3	8.3	2
1.2	5.6	3.3	2
5.4	7.6	7.5	2
3.6	5.4	1.2	2
...			



No 3D for abstract data!

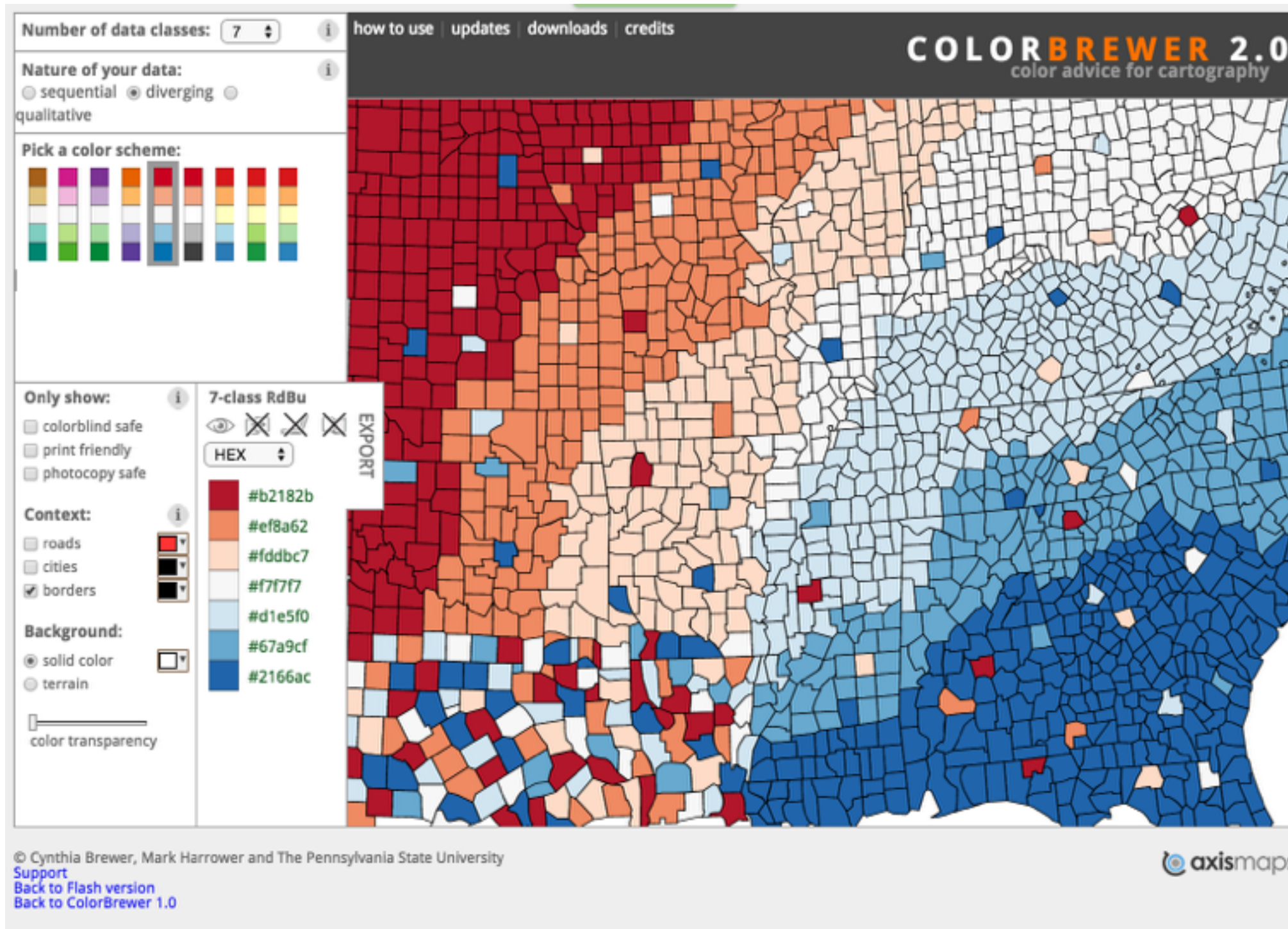
Michael Sedlmair, Tamara Munzner, Melanie Tory
Empirical Guidance on Scatterplot and Dimension Reduction Technique Choices (InfoVis 2013).

Visual Encoding: Color?



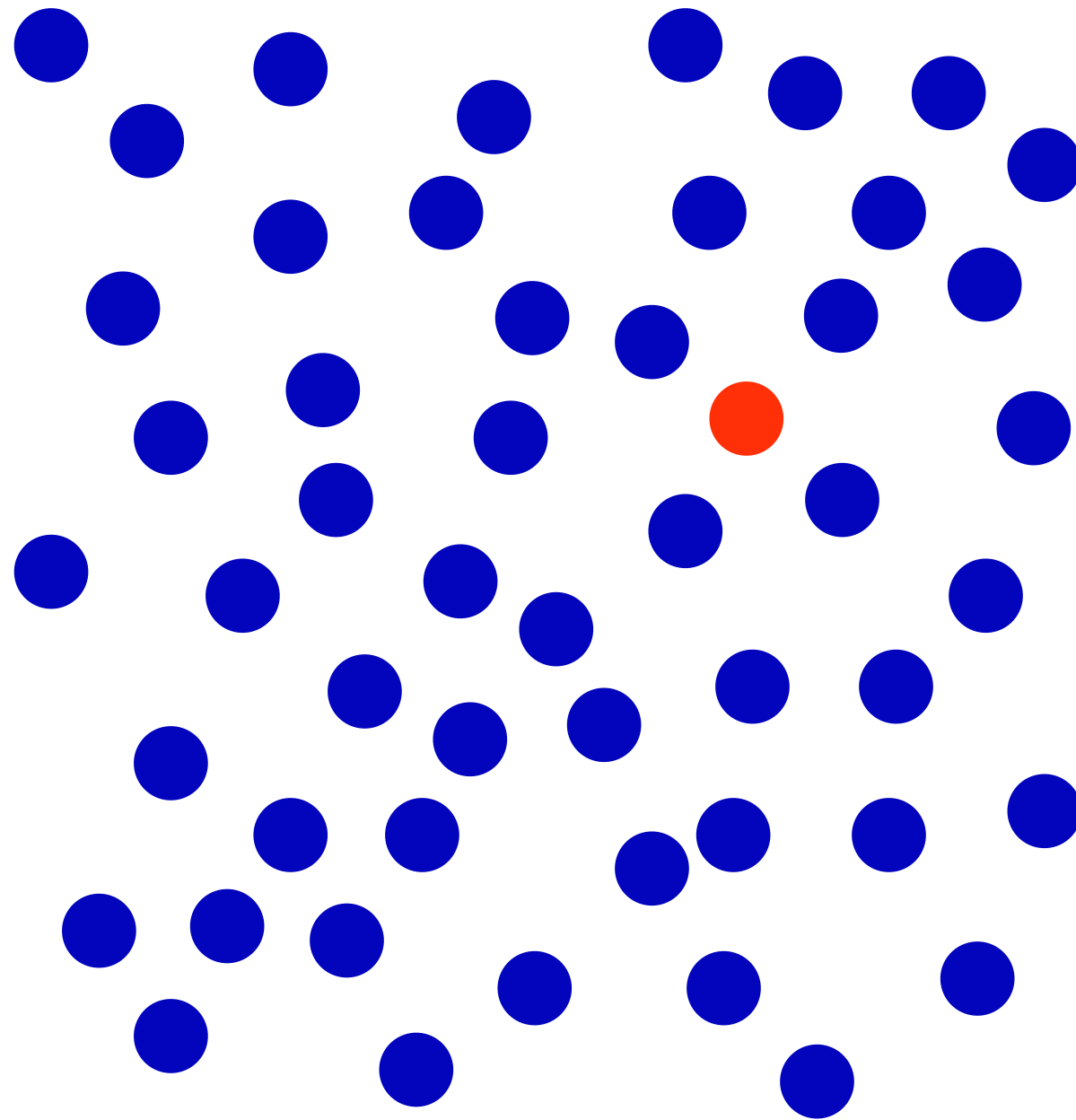
Cinteny: flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. Sinha and Meller. Bioinformatics 2007

ColorBrewer



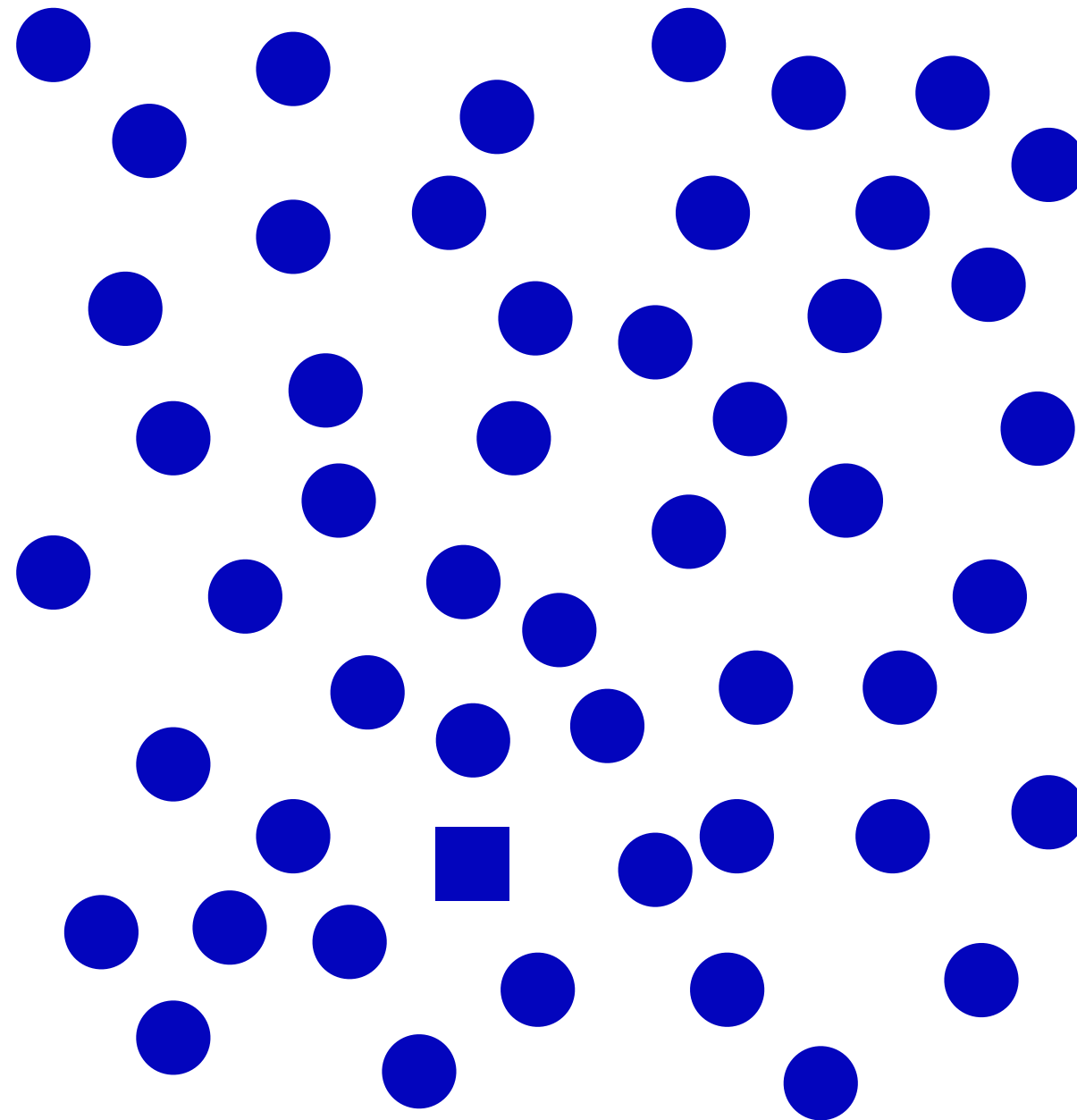
Human Perception

Popout



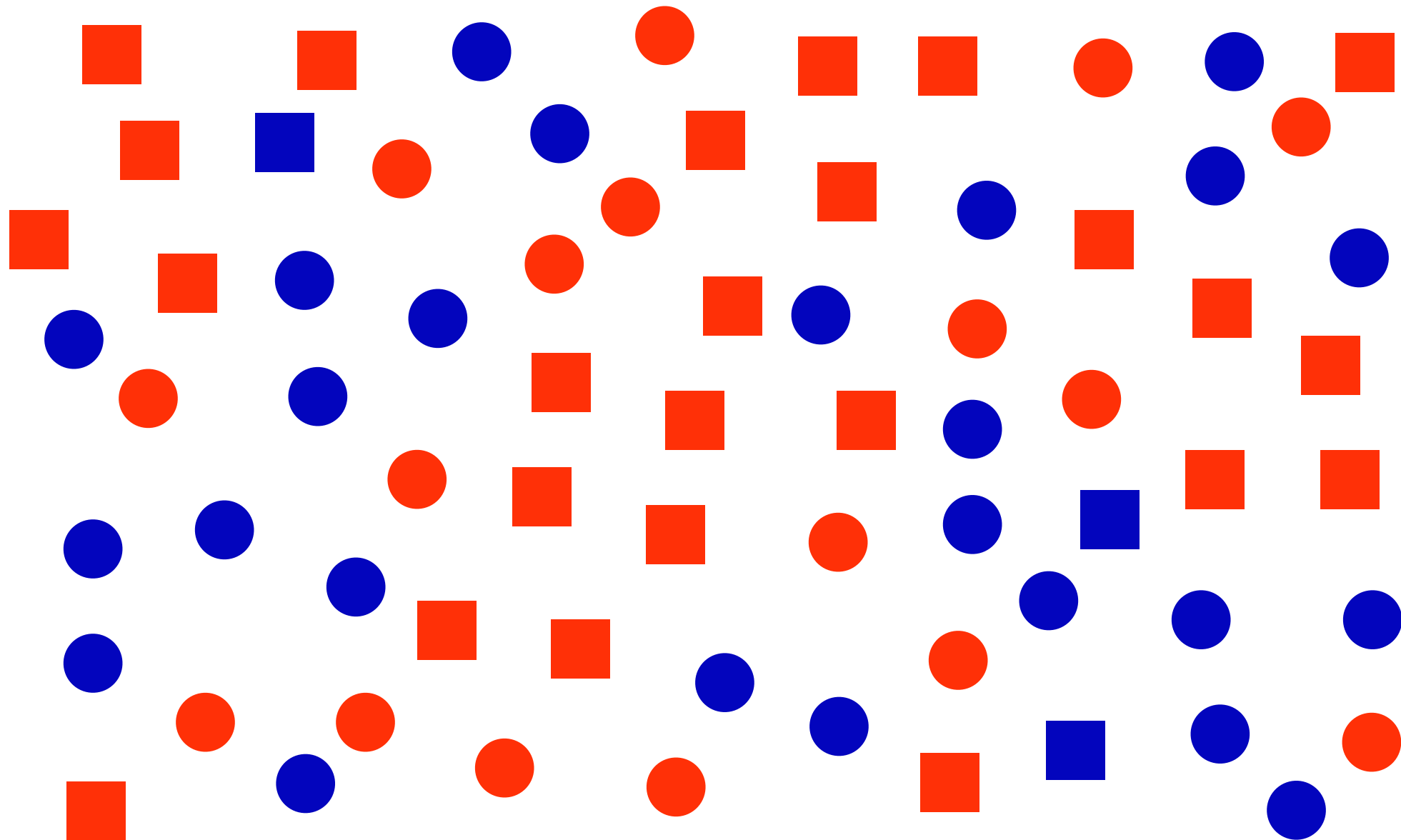
Human Perception

Popout



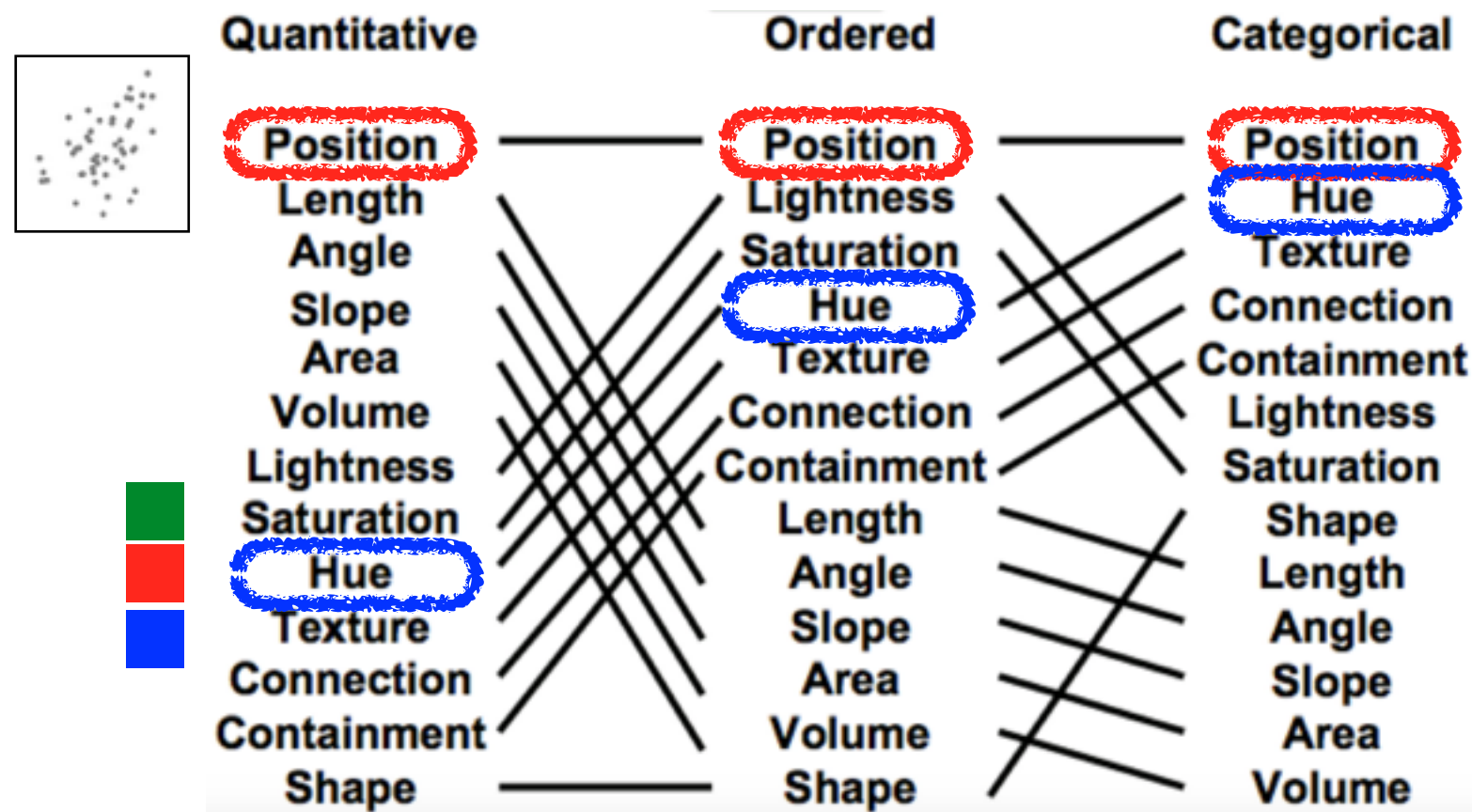
Human Perception

Popout: combining channels?



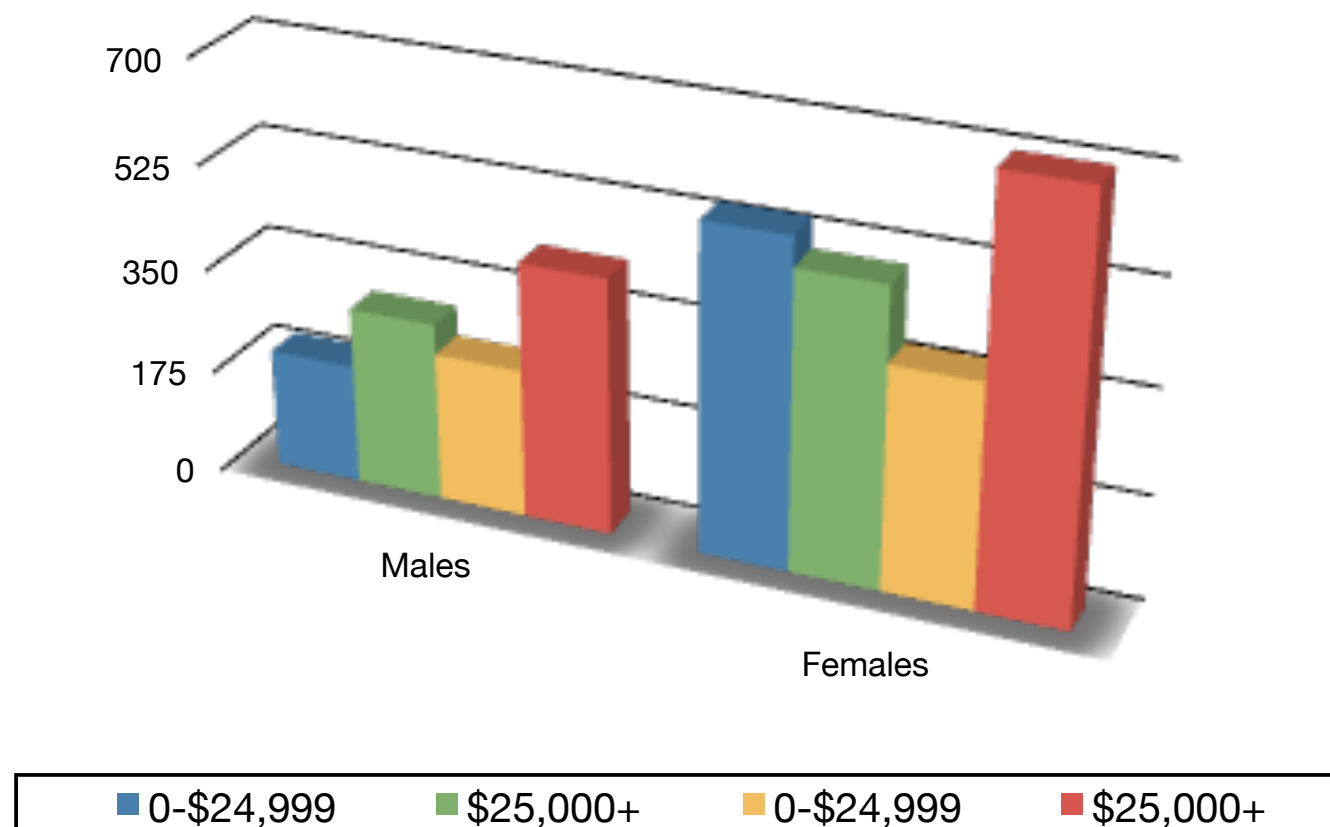
Human Perception

Visual Channels (Jock Mackinlay)



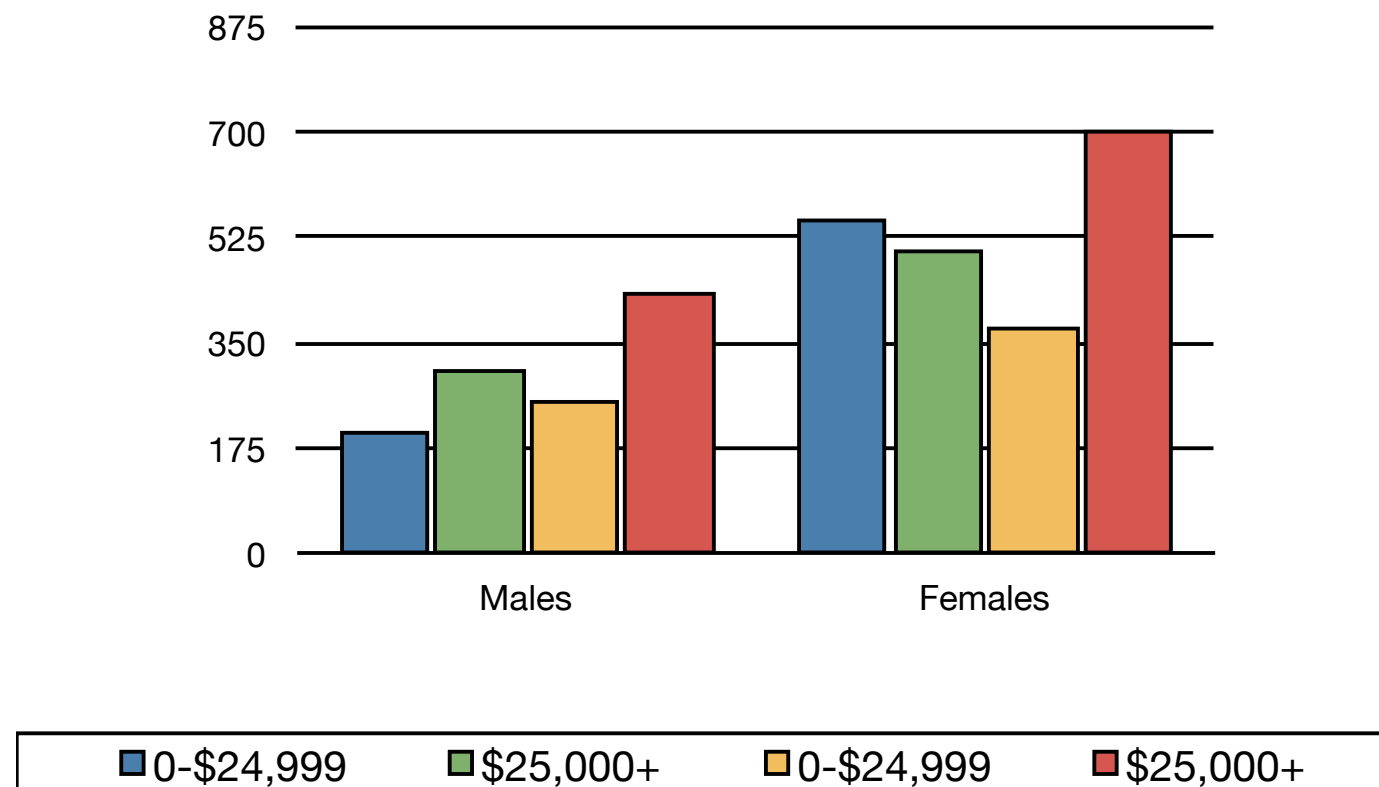
Design

Data-to-Ink Ratio (Edward Tufte)



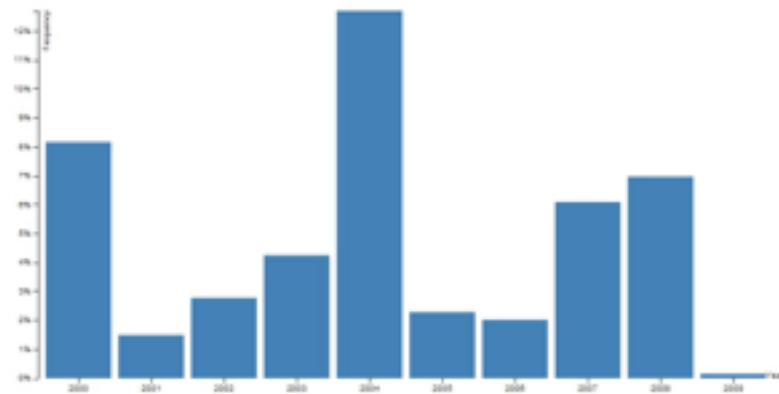
Design

Data-to-Ink Ratio (Edward Tufte)

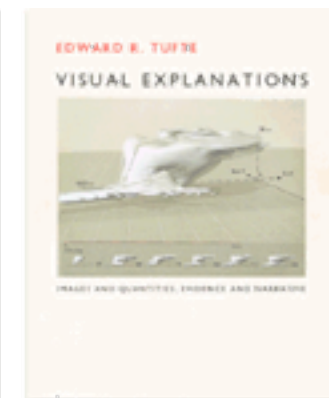
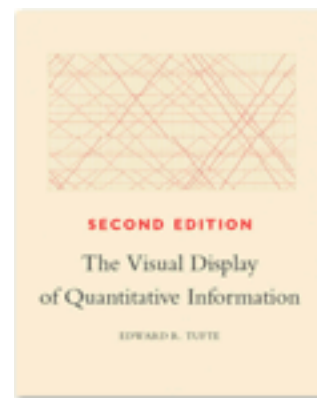
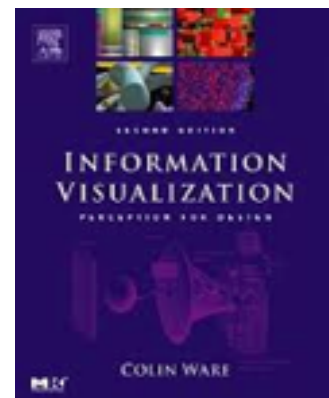
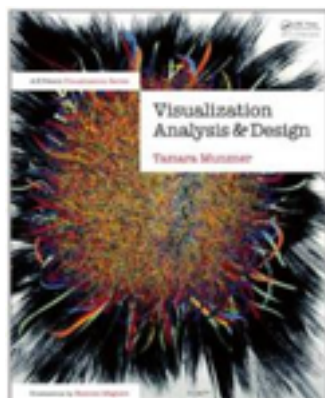


What we can do now ...

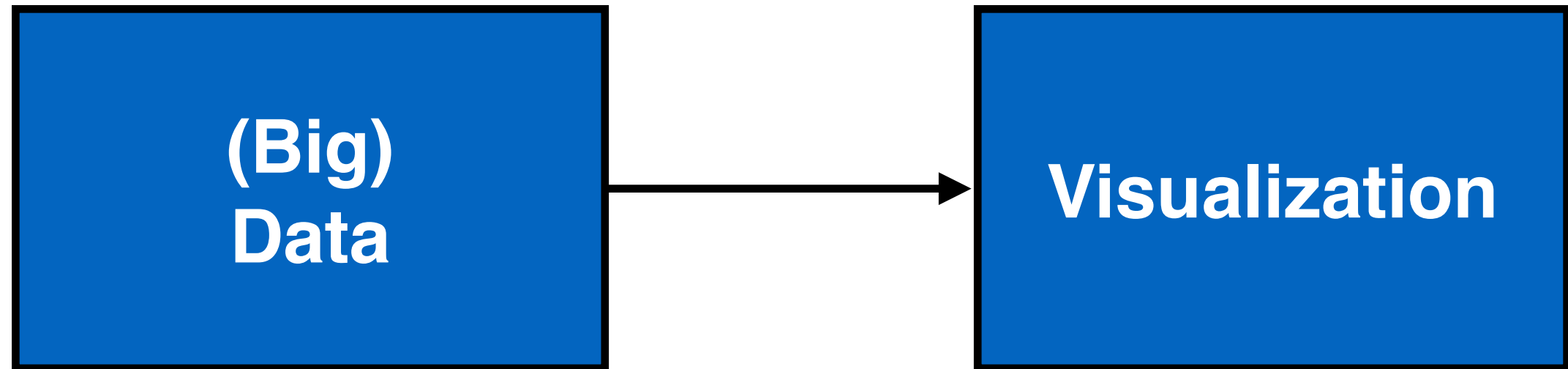
Static representations to present information



<http://andrewgelman.com/wp-content/uploads/2009/07/vouchermapsBAYES2000.png>



BUT:
Big Data?



volume

velocity

variety

veracity

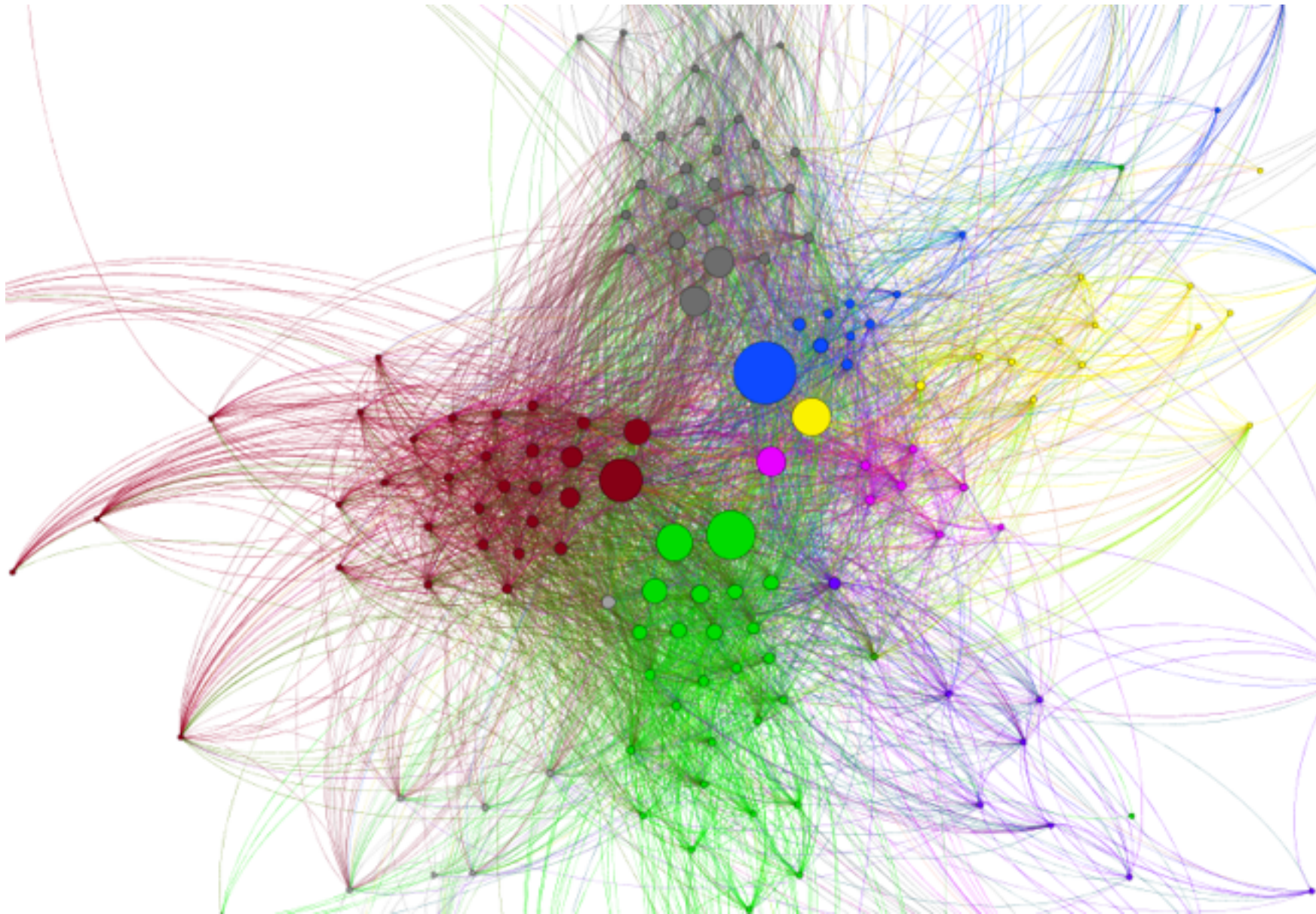
...

perception

design

...

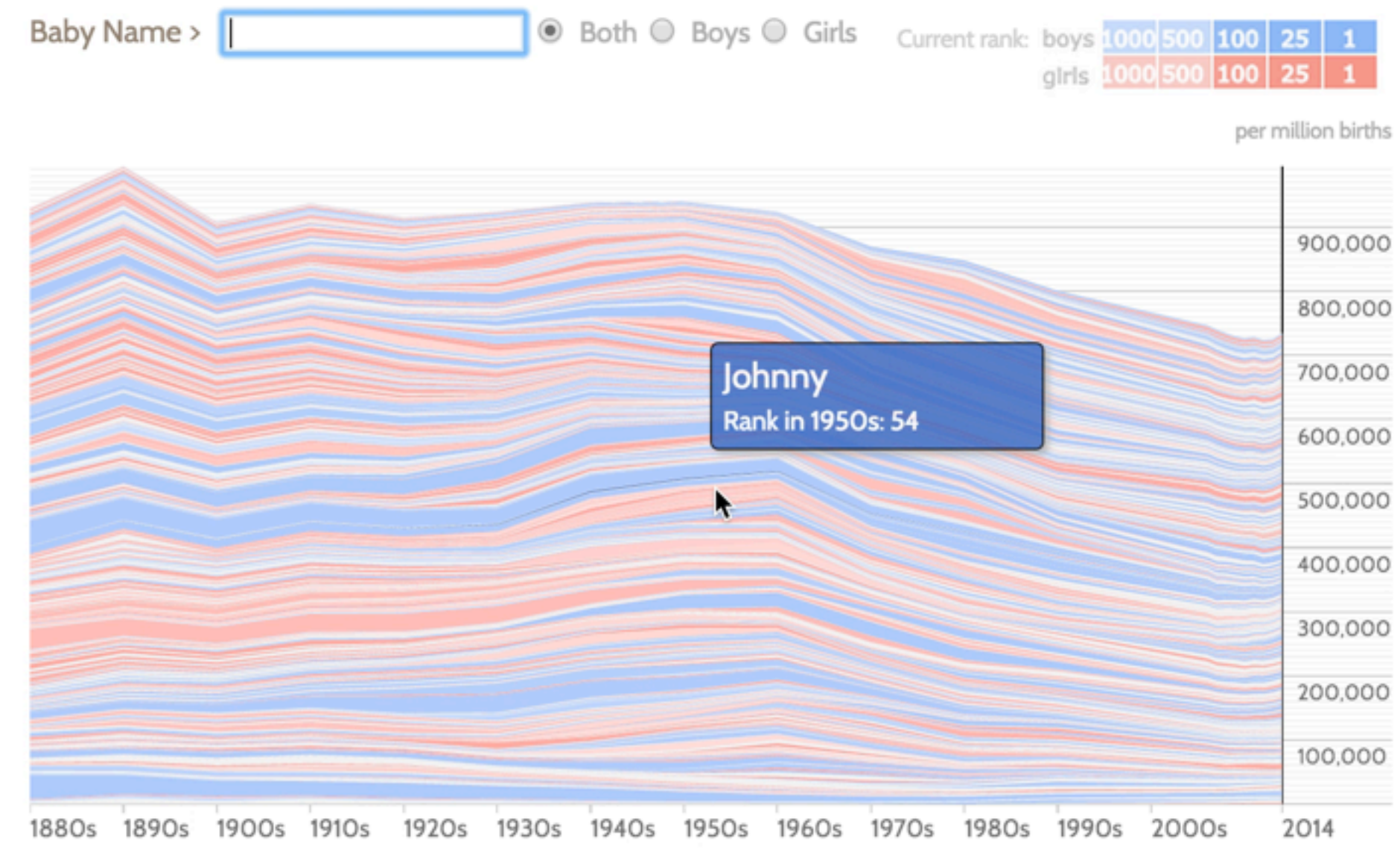
Showing all the data?



<https://twitter.com/axelmaireder/media>

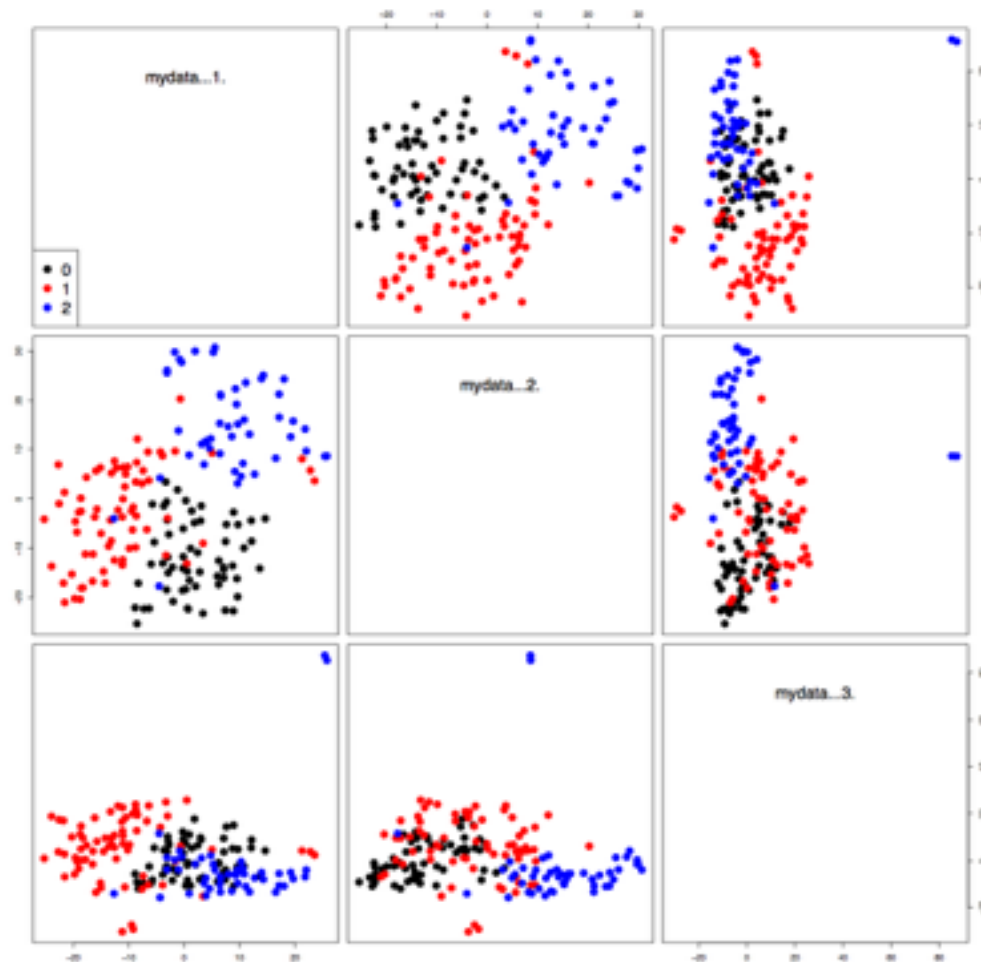
Interaction

example: baby name voyager



Guidance

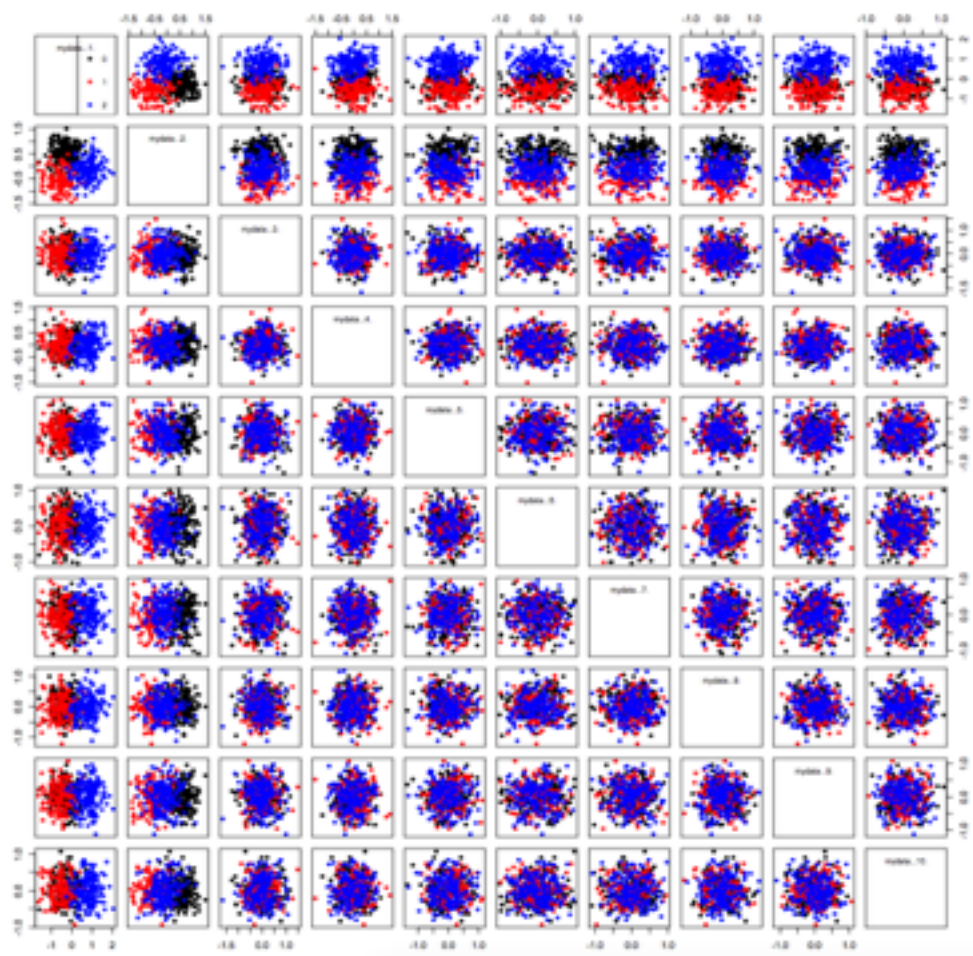
example: finding interesting
projections in high-dim data?



3D Scatterplot Matrix
—> 3 Views

Guidance

example: finding interesting
projections in high-dim data?

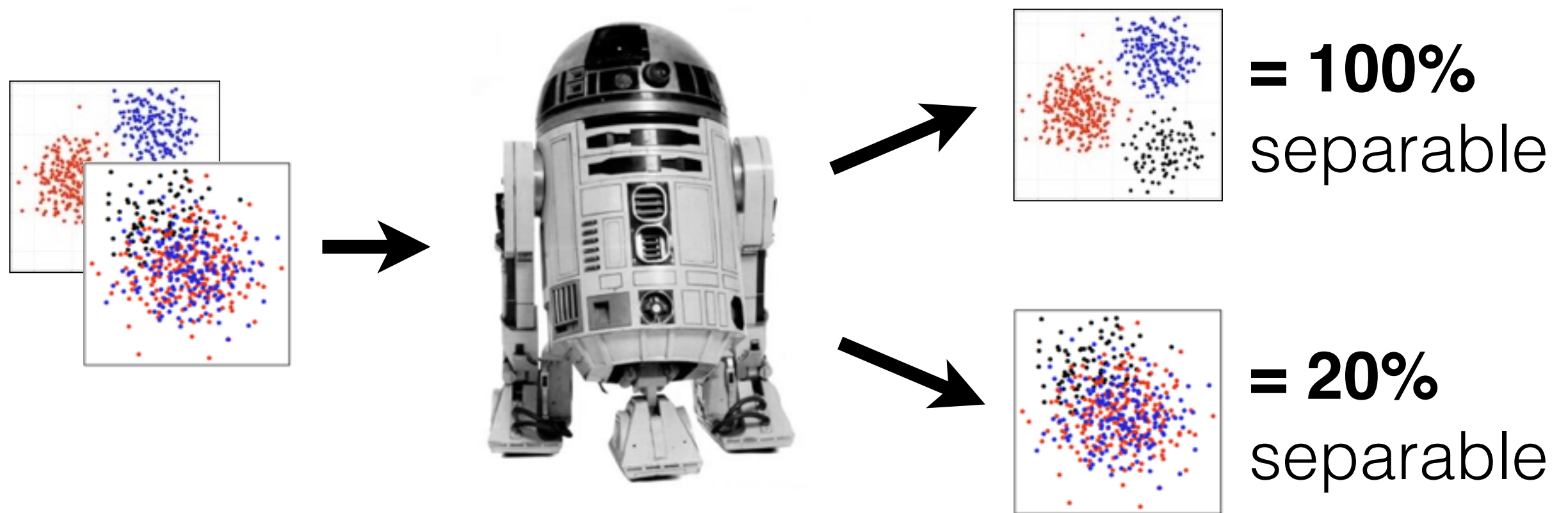


10D $\longrightarrow$ 45 Views

200D $\longrightarrow$ ~20k Views

Guidance

Visual quality measure,
e.g. class separation measures

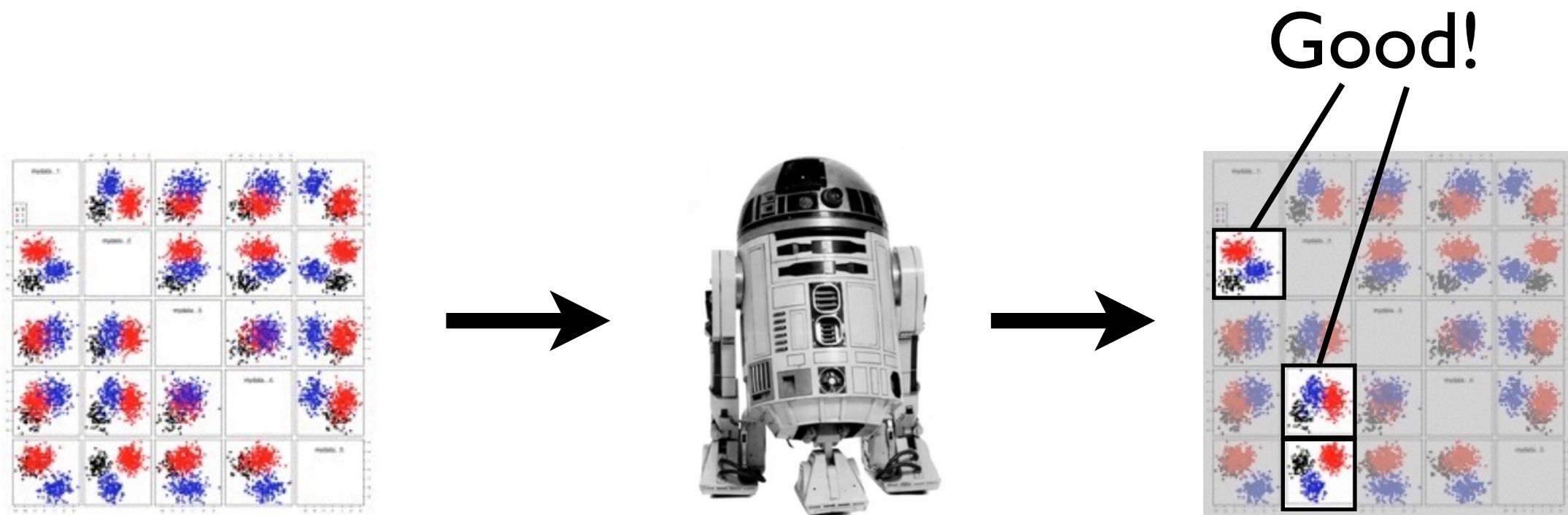


Michael Sedlmair, Michaël Aupetit: Data-driven Evaluation of Visual Quality Measures (EuroVis 2015).

Michael Sedlmair, Andrada Tatu, Tamara Munzner, Melanie Tory: A Taxonomy of Visual Cluster Separation Factors (EuroVis 2012).

Guidance

Visual quality measure,
e.g. class separation measures

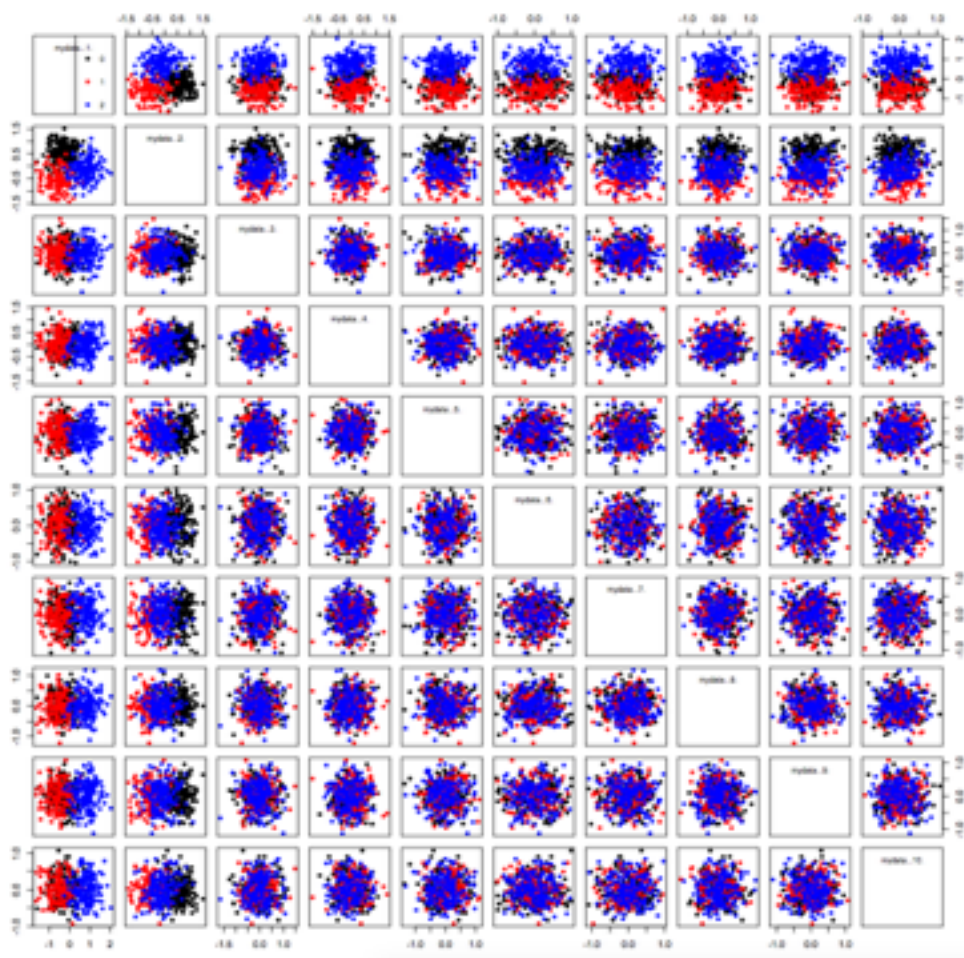


Michael Sedlmair, Michaël Aupetit: Data-driven Evaluation of Visual Quality Measures (EuroVis 2015).

Michael Sedlmair, Andrada Tatu, Tamara Munzner, Melanie Tory: A Taxonomy of Visual Cluster Separation Factors (EuroVis 2012).

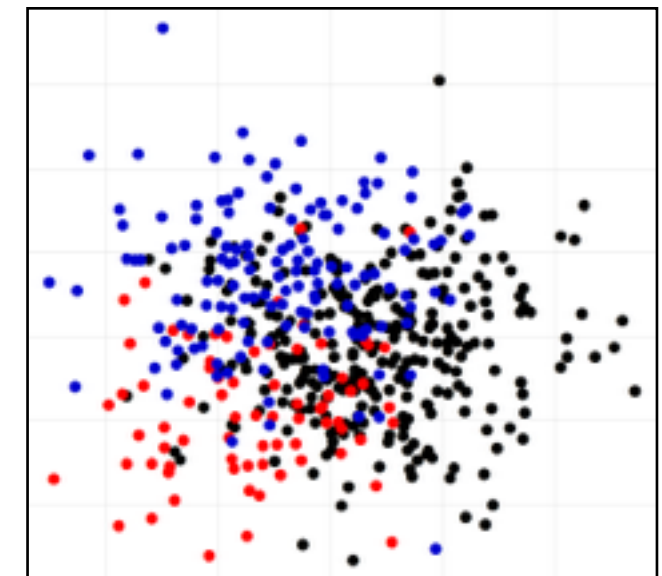
Abstraction/Aggregation

E.g. dimension reduction



PCA
MDS
IsoMap
tSNE

...



Matthew Brehmer, Michael Sedlmair, Stephen Ingram, Tamara Munzner
Visualizing Dimensionally-Reduced Data: Interviews with Analysts and a Characterization of Task Sequences (BELIV 2014).

What we can do now...

- **Combined Visual & Algorithmic Analysis**
 - *called:* Visual Analytics, Visual Data Science, ... or just Visualization
- Build powerful systems for visual analysis to
 - explore the data
 - gain insights
 - support decision making
 - share and communicate

Case Study: Cardiogram

Visual Analytics for Automotive Engineers

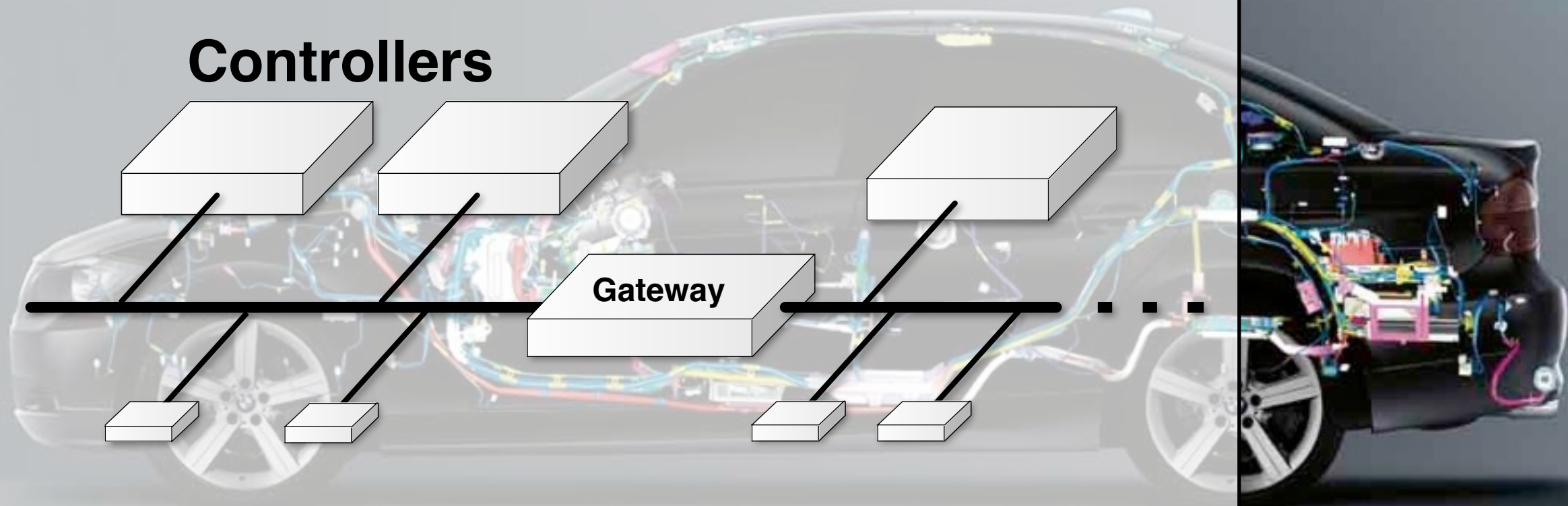


M. Sedlmair, P. Isenberg, D. Baur, M. Mauerer, C. Pigorsch, A. Butz
In Proc. ACM Conference on Human Factors in Computing Systems, p. 1727–1736, 2011, (CHI 2011).

more and more electronics...



**... enabled by in-car
communication networks**



~70 Controllers / Car

Problem Characterization



5836... F... 55	Rx	16	5a 77 f8 27 00 20 00 20 0f 00 0...
5836... F... 56	Rx	16	ee 7e f8 27 00 20 00 20 0f 00 0...
5836... F... 57	Rx	16	ed 77 10 28 00 20 00 20 0f 00 0...
5836... F... 58	Rx	16	5e f8 1f 00 02 20 ff 21 22 22 07 f...
5851... F... 5a	Rx	16	33 08 77 22 75 d2 6f e2 70 e2 03 0...
5851... 1 1a1 V_VEH	Rx	5	c1 f7 00 00 8a
~ V_VEH_COG		[0]	
~ ST_ECU_V_VEH	Signal ungültig	[f]	
~ QU_V_VEH_COG	Signalwert ist gültig, Zustand/S	[a]	
~ DWCO_VEH	Fahrzeug steht	[0]	
~ CRC_V_VEH	193	[c1]	
~ ALIV_V_VEH	7	[7]	
5851... F... 5c	Rx	8	00 00 00 00 ff ff 00 10
5851... 1 1c4	Rx	6	00 00 00 00 ff ff
5851... F... 12f	Rx	72	00 00 00 00 00 00 ff 59 07 21 4c 0...
5851... 1 1c5	Rx	6	02 00 04 00 ff ff
5836... F... 1	Rx	16	10 76 18 28 00 20 00 20 0f 00 0...
5836... F... 2	Rx	16	ca 76 f9 27 00 20 00 20 0f 00 0...
5836... F... 3	Rx	16	6a 76 fa 27 00 20 00 20 0f 00 0...
5836... F... 4	Rx	16	a3 76 0e 28 00 20 00 20 0f 00 0...
5836... F... 5	Rx	16	2c f7 1f 00 02 20 ff 21 22 22 07 f...
5851... F... 7	Rx	16	6d f9 76 12 75 d2 6f e2 70 e2 01 1...
5851... F... 12	Rx	0	
5851... 2 301 AVL_STEA_DV	Rx	7	51 15 f8 7f ff 7f 11
5851... 3 301 AVL_STEA_DV	Rx	7	51 15 f8 7f ff 7f 11
5851... 4 137	Rx	2	f8 00
5851... 3 49 ANG_ACPD	Rx	8	9b 99 00 c0 00 e0 7f f0
5840... 1 299	Rx	5	9f ff ff ff ff
5851... F... 21	Rx	0	
5851... 4 a5 TORQ_CRSH_1	Rx	8	45 f5 48 f7 7f 00 00 fc
5851... 3 a5 TORQ_CRSH_1	Rx	8	fe f9 48 f7 7f 00 00 fc
5836... F... 23	Rx	16	7c 10 05 05 ea f3 53 20 74 10 20 f...
5851... F... 26	Rx	0	

- **data:** recorded traces (15k messages/sec)
- **task:** finding errors
- **process:** test drives
- **current practices:** mainly textual lists

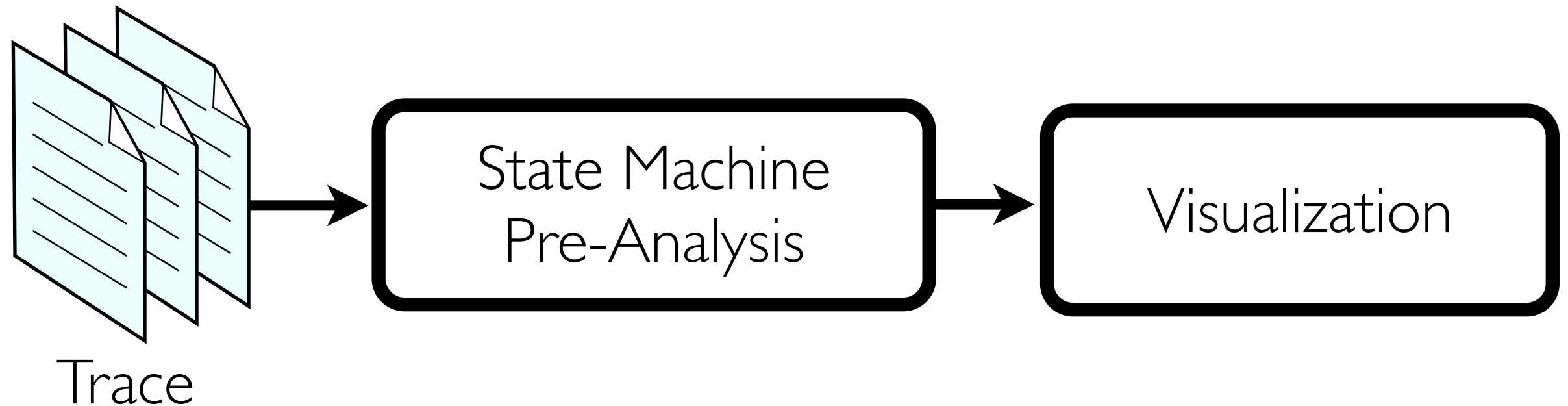
Problem Characterization

- **Main Challenges**

- handling masses of test traces (large and many)
- understanding correlation between trace and car behavior
- ...

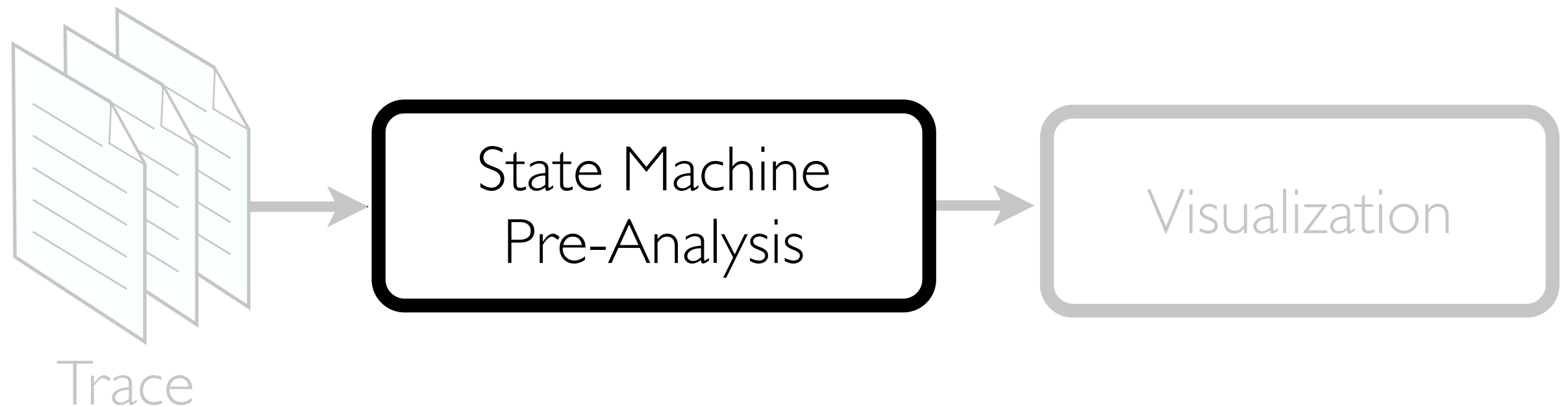
Cardiogram

overview

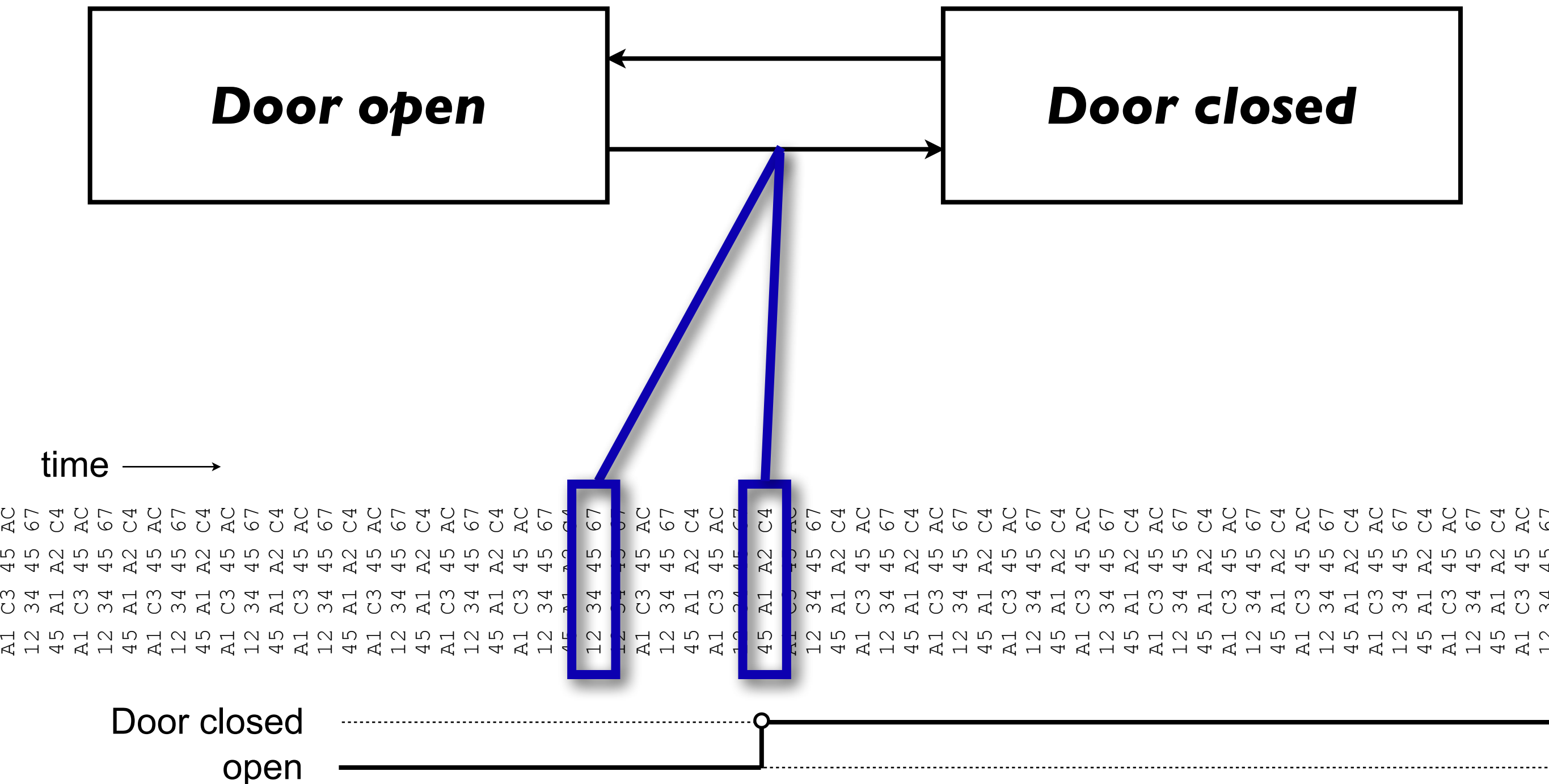


Cardiogram

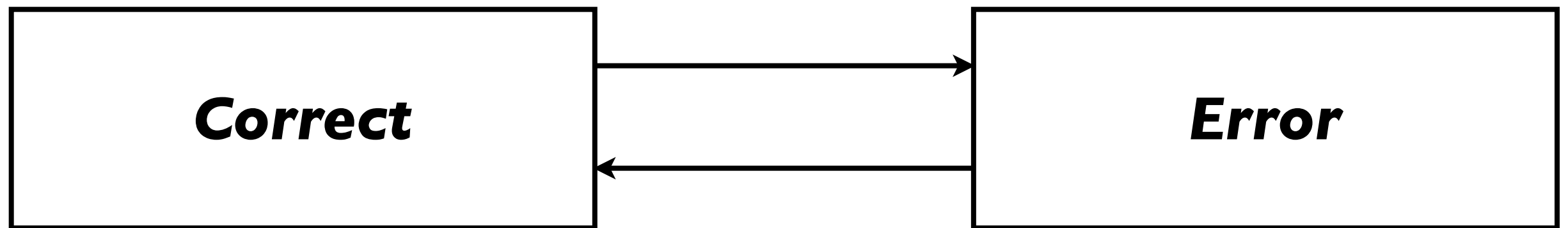
state machine pre-analysis



State Machines: Behavior

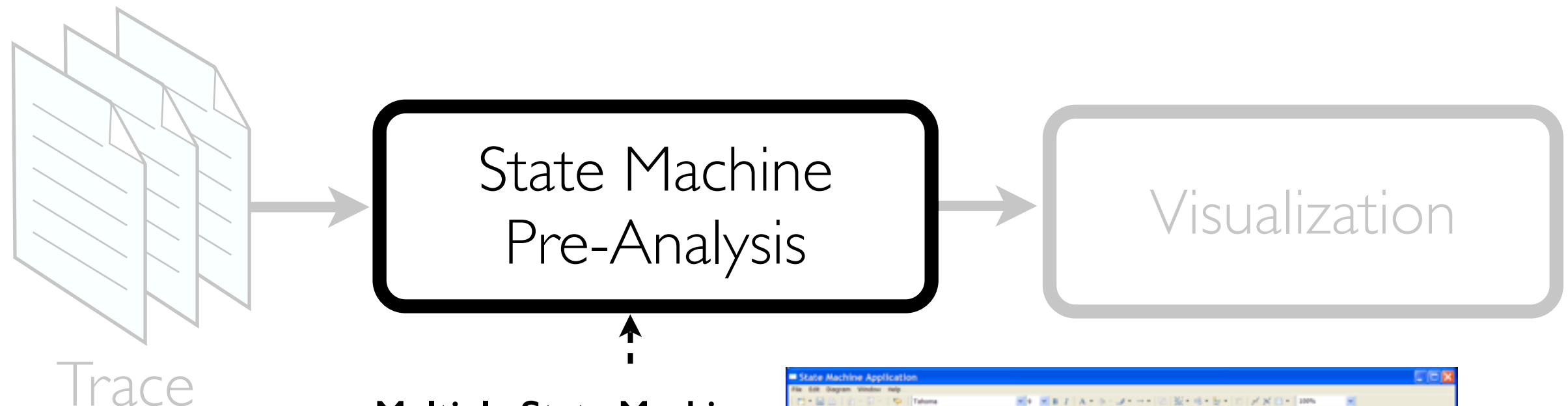


State Machines: Error Detection



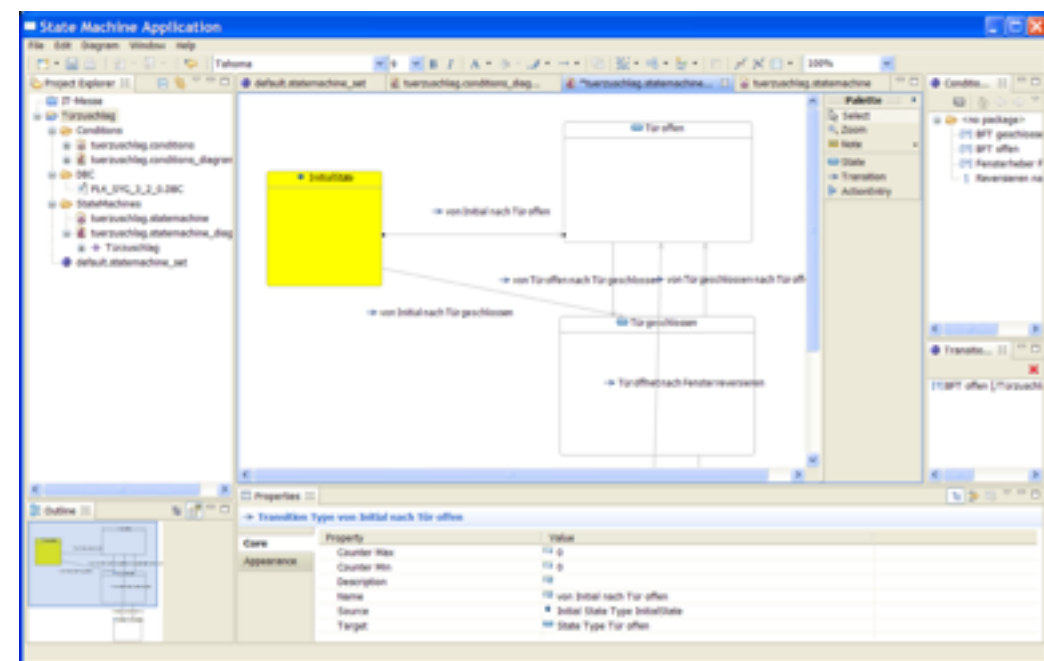
Cardiogram

create/select state machines



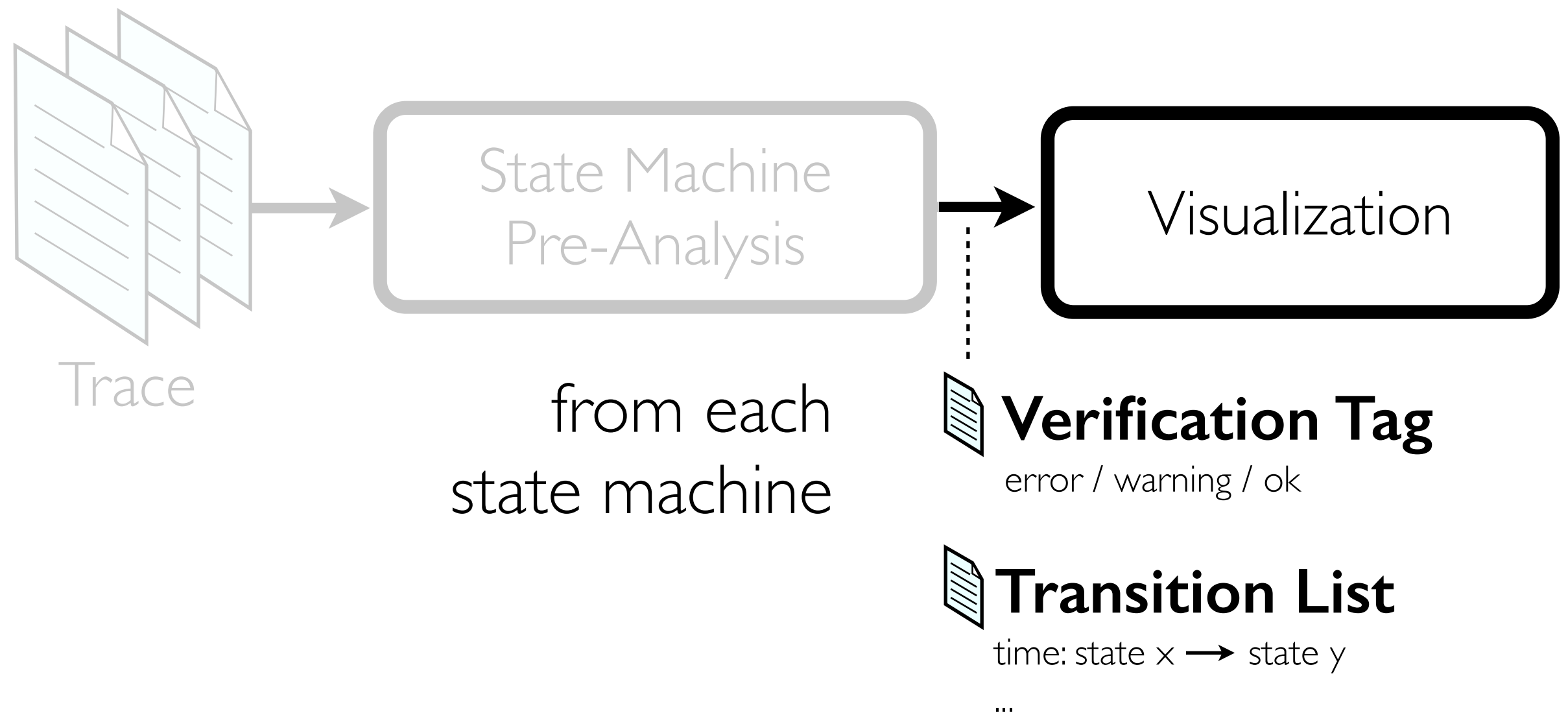
Multiple State Machines

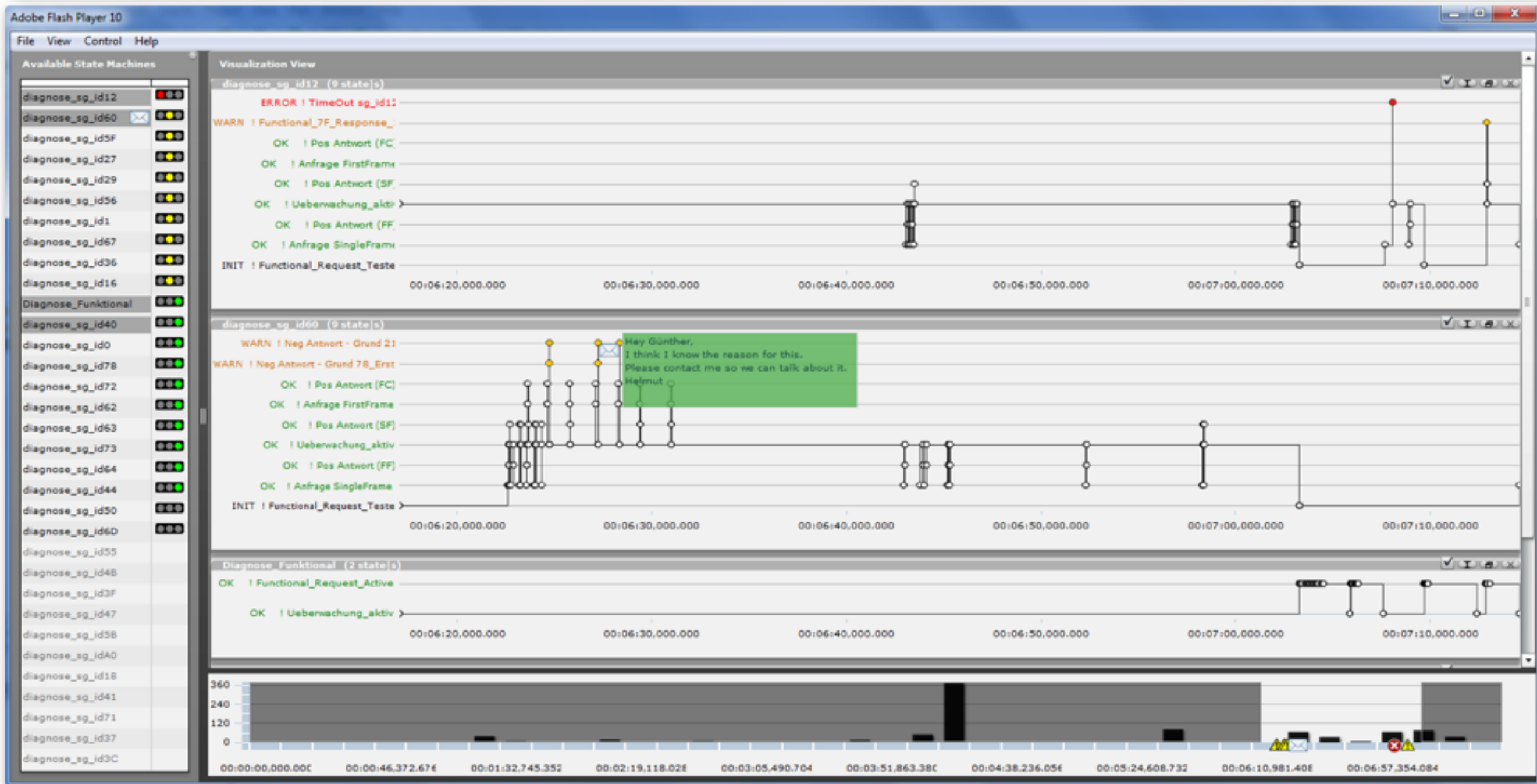
- State Machine 1
- State Machine 2
- State Machine 3
- ... (dozens)

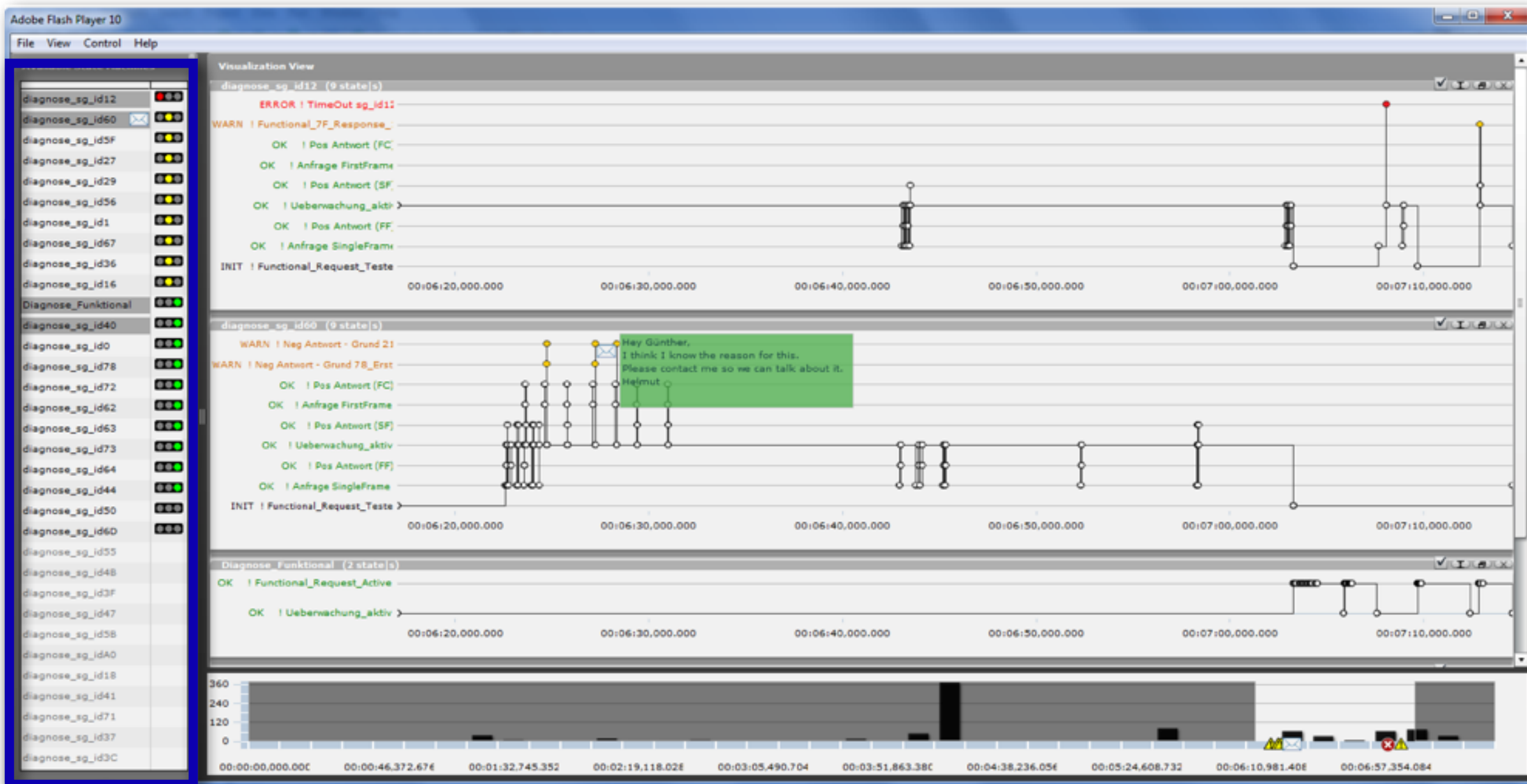


Cardiogram

data abstraction/reduction







diagnose_sg_id12

diagnose_sg_id60

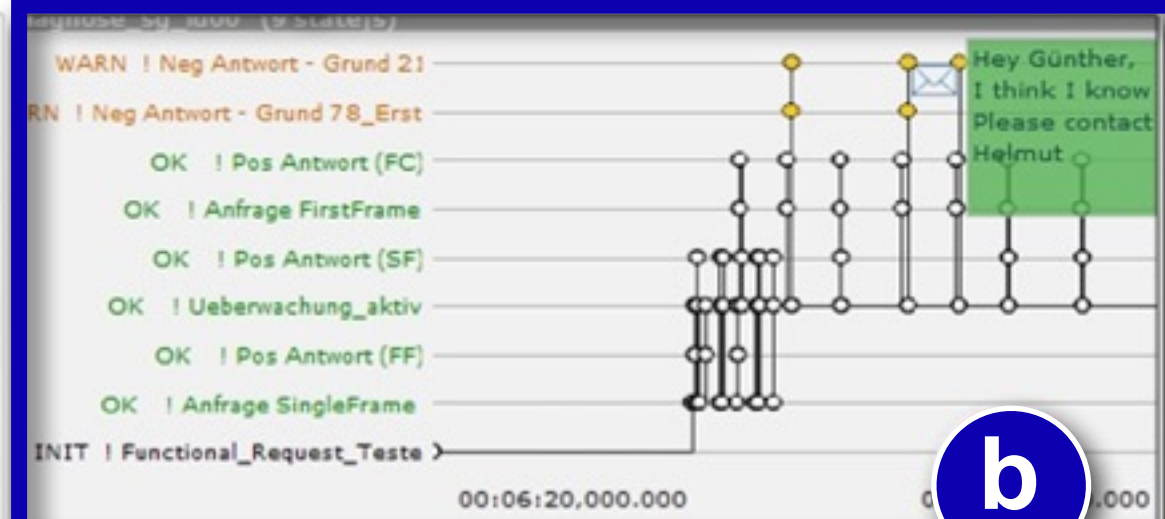
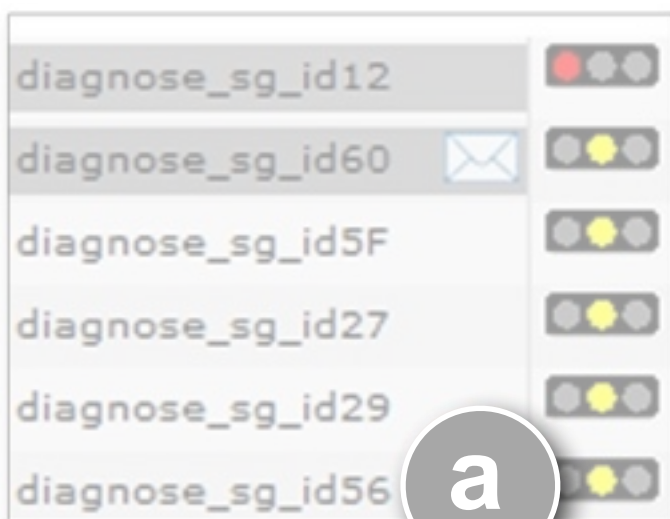
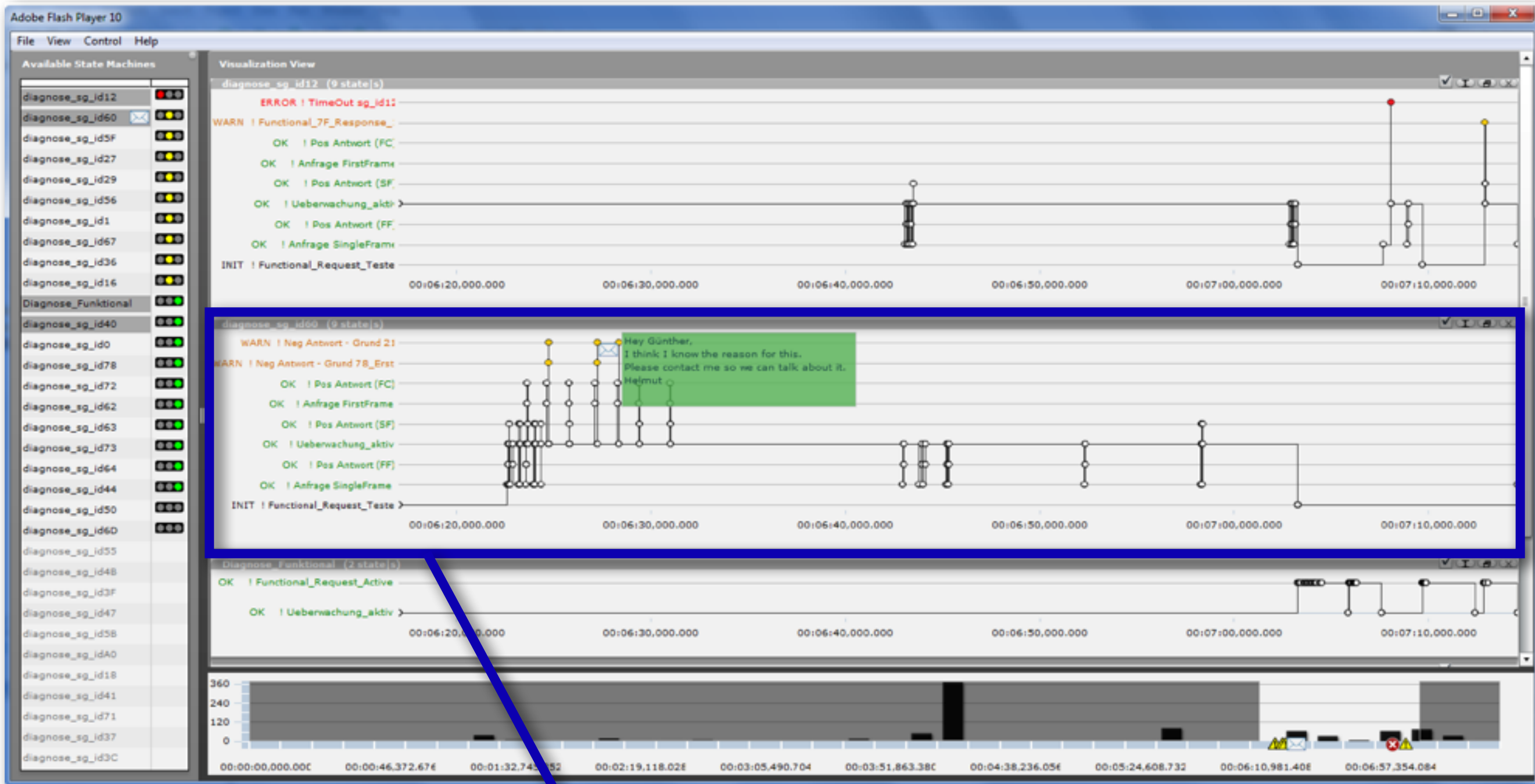
diagnose_sg_id5F

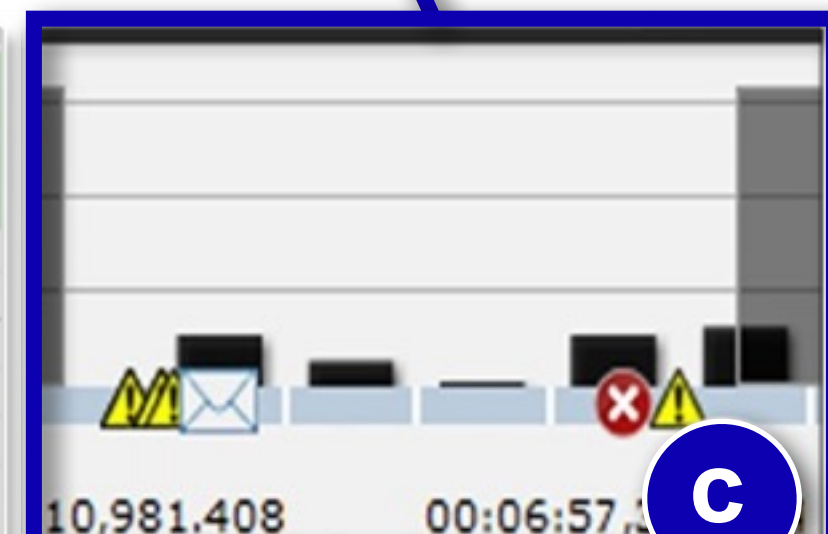
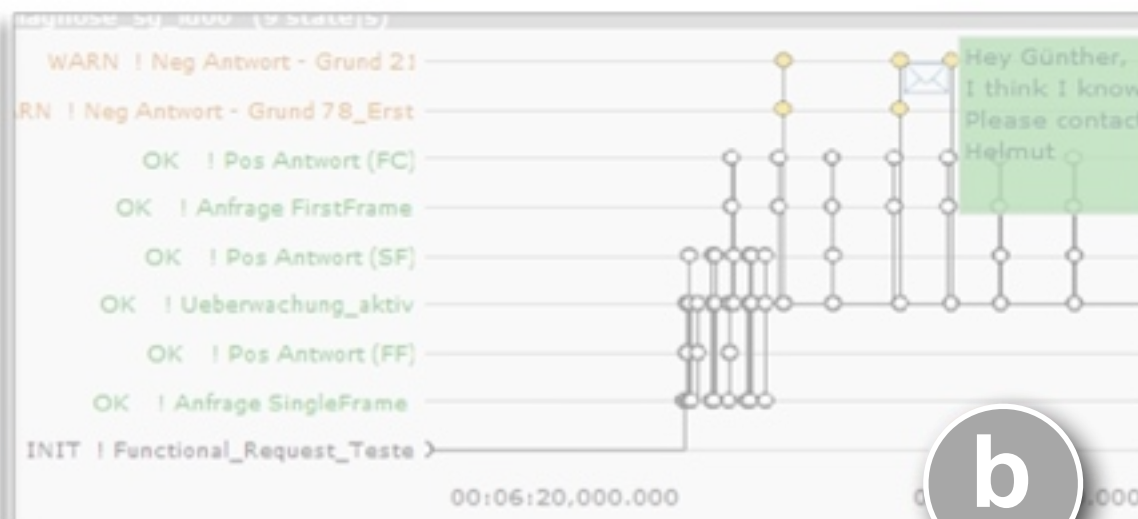
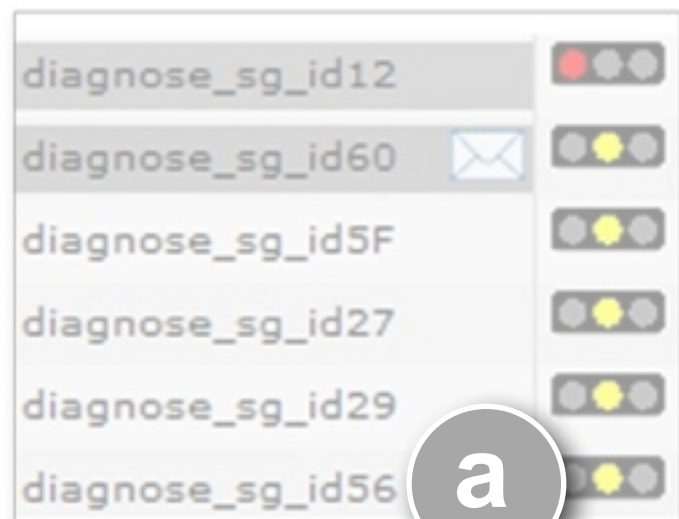
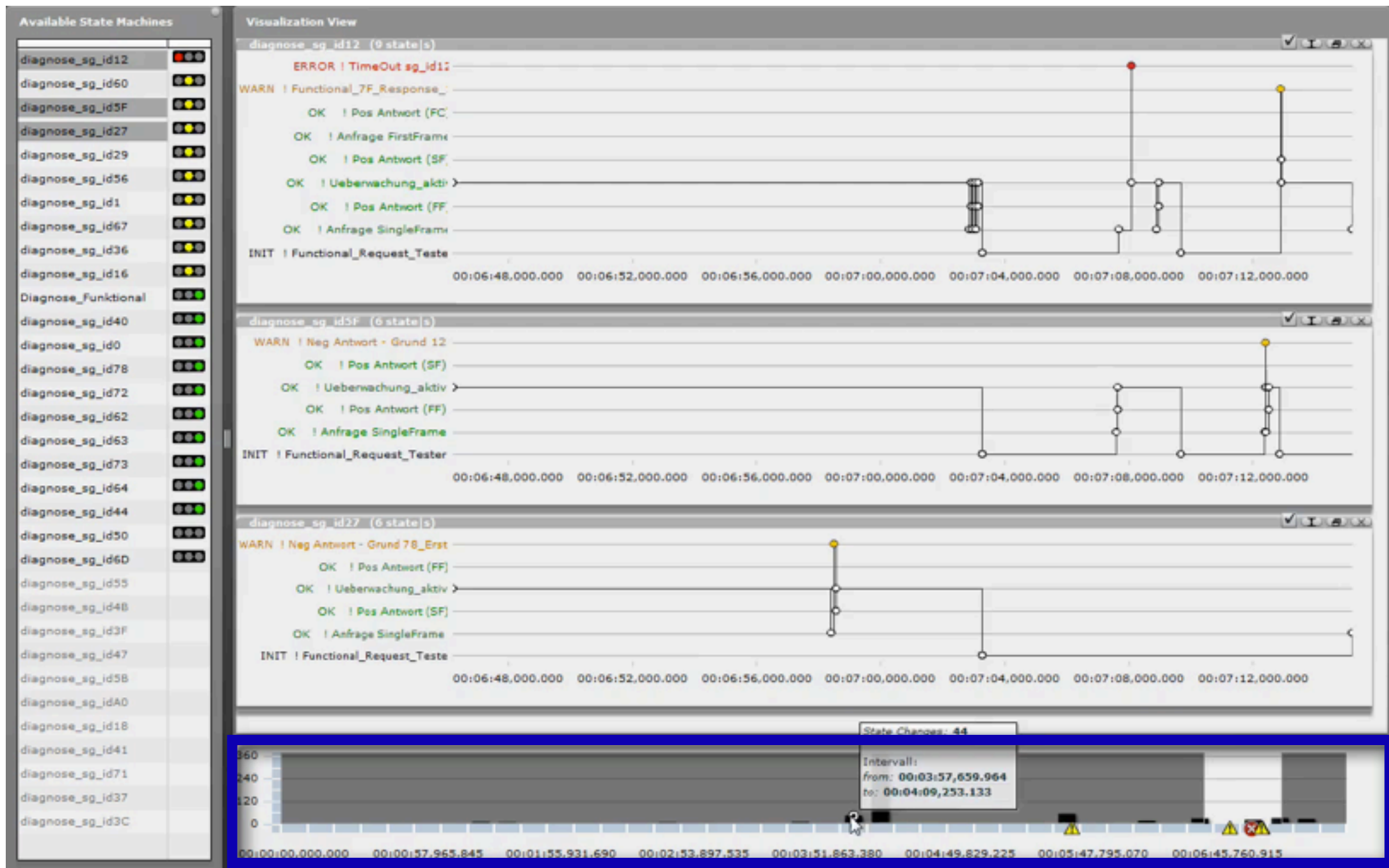
diagnose_sg_id27

diagnose_sg_id29

diagnose_sg_id56

a





Longitudinal field study

1 year, 15 engineers

Longitudinal field study results

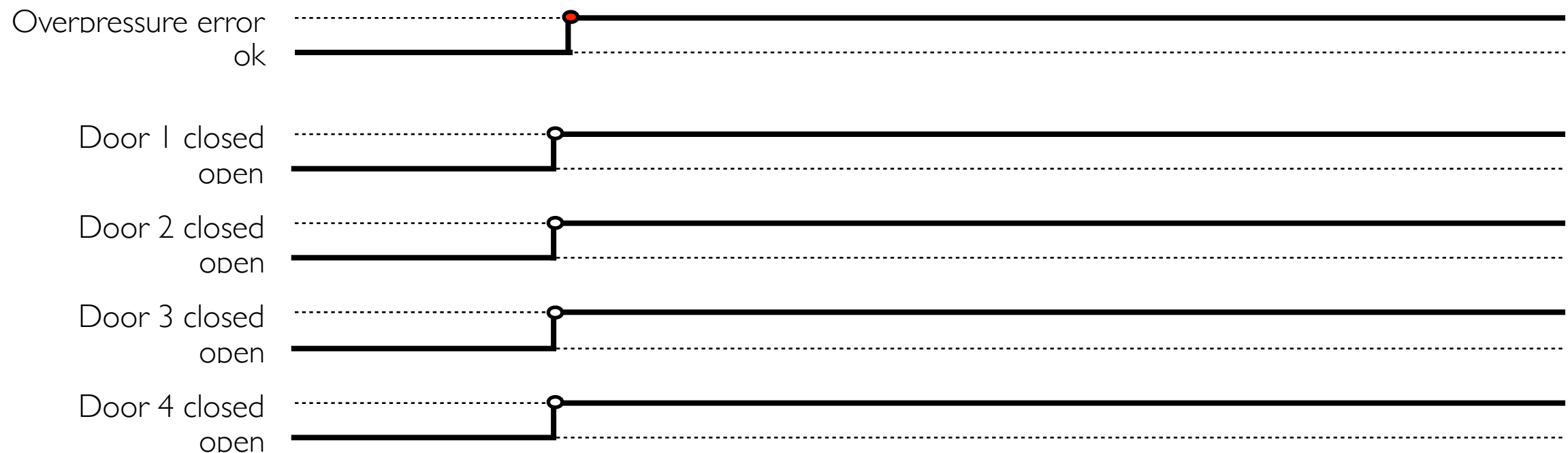
- externalization and sharing of expert knowledge

Longitudinal field study results

- externalization and sharing of expert knowledge
- complete coverage of traces vs. sparse sampling

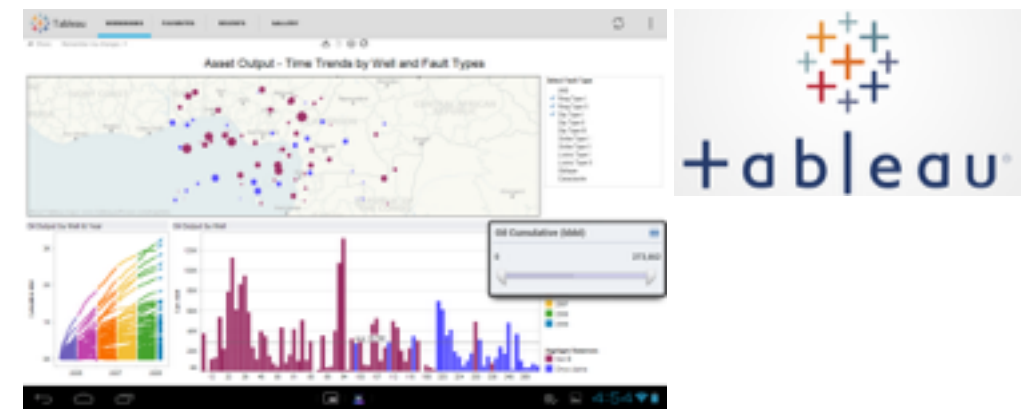
Longitudinal field study results

- externalization and sharing of expert knowledge
- complete coverage of traces vs. sparse sampling
- understand behavioral correlations



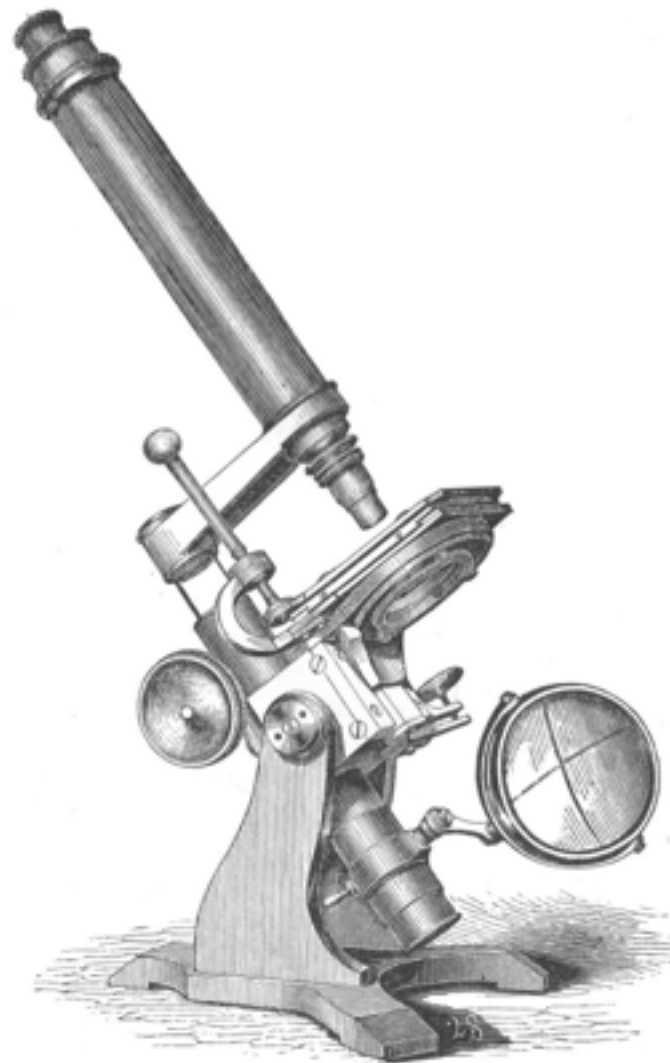
Could we have done it with a general purpose tool?

- Likely not, not yet



- Complex problem space (big data)
- Complex solution space (visual encoding, interaction, abstraction)

Visualization: The Human Lens to Big Data



<http://www.microscope-antiques.com/grunow.html>



<http://galileo.rice.edu/sci/instruments/telescope.html>

Summary

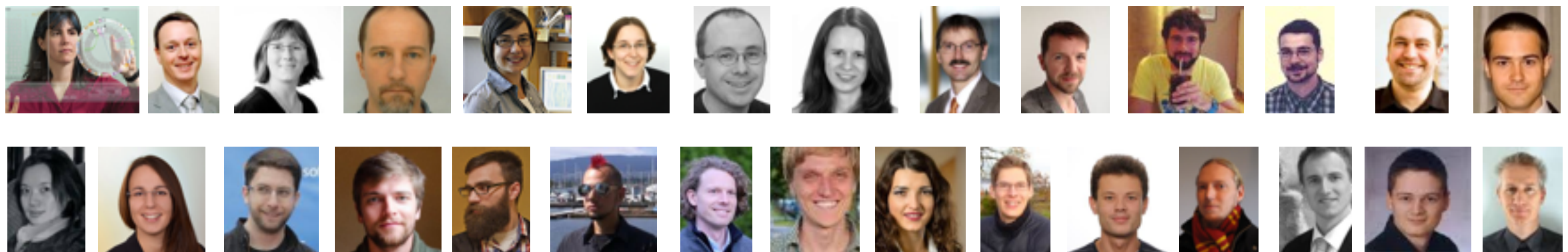
- Visualization for Human-in-the-Loop tasks
- Visualization is not just about pretty images
 - perception, design
 - interaction, guidance, abstraction
- Case study: Cardiogram

Thanks

VDA Group @ University of Vienna



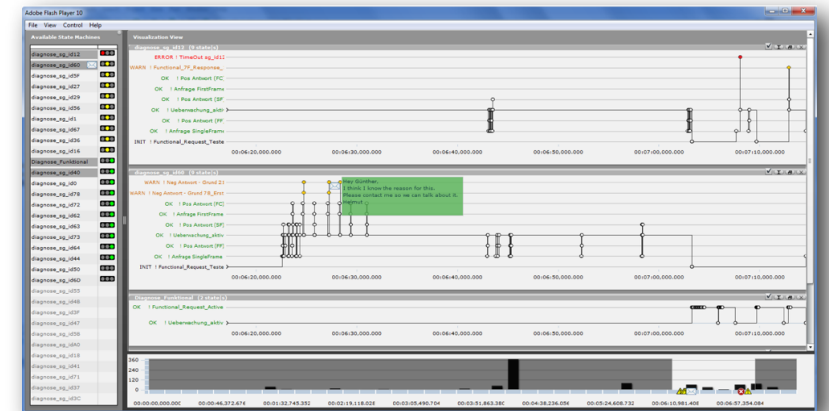
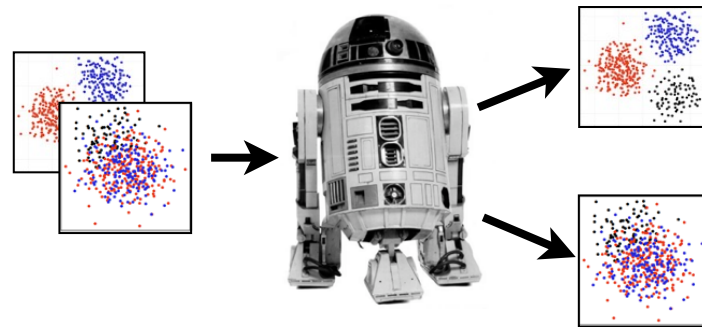
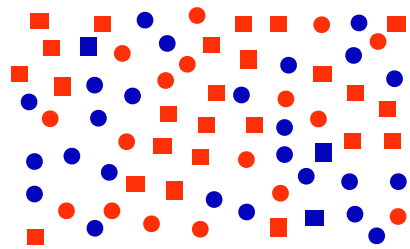
Many, many other collaborators



...

Thanks to you!

Visualization: The Human Lens to Big Data



Dr. Michael Sedlmair