Visualizing Data:
Perception, Interaction, Computation

Dr. Michael Sedlmair
Visualization & Data Analysis Group
universität wien

LMU Munich
Diplom in Computer Science, 2001-2007

BMW Group Research & Technology
PhD in Computer Science, 2007-2010

Univ. of British Columbia, Canada
PostDoc, 2010-2012

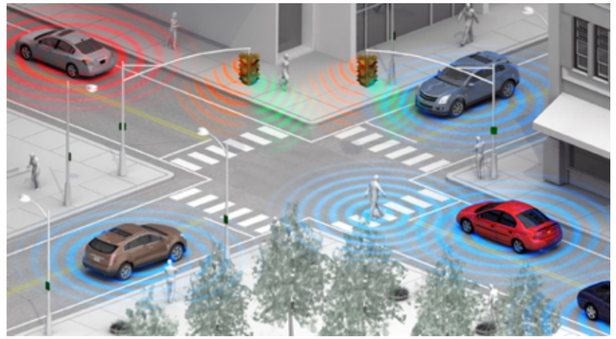
Univ. of Vienna, Austria
Assistant Professor, 2013-now

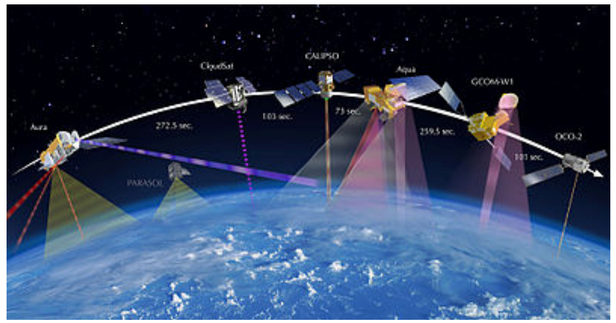
Human-Computer Interaction (HCI)

Information Visualization

Data Science (algorithmic analysis)

Data





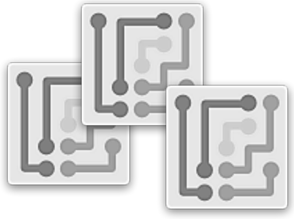
https://en.wikipedia.org/wiki/Earth_observation_satellite

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

Big Data/ Data Science

Data

Algorithms



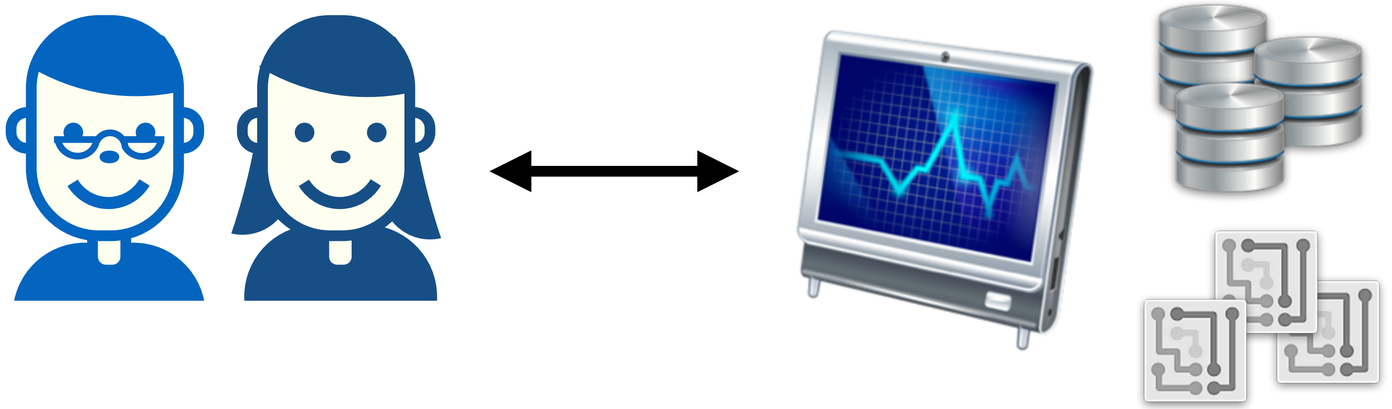
Visualization

Scientists

Engineers

Journalists

Policy Makers



Understand patterns

Gain insights

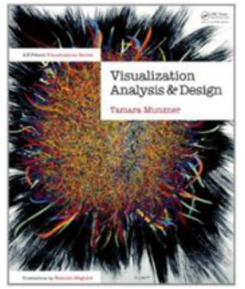
Make decisions

Communicate

Definition

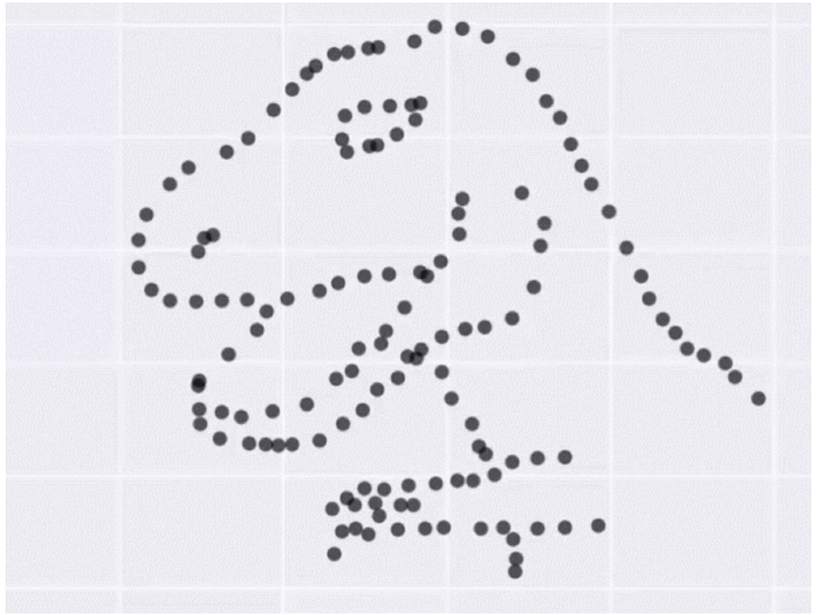
Computer-based visualization systems provide visual representations of datasets intended to help people carry out some task more effectively.

Munzner, 2014



Definition

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



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 CHI, 2017.

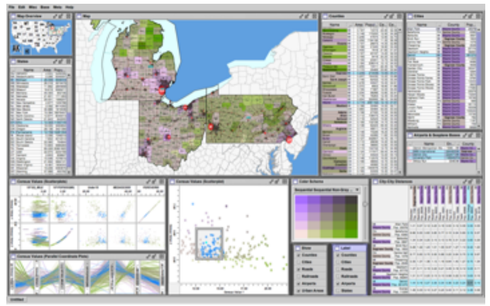
Definition

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

Tableau

SAS

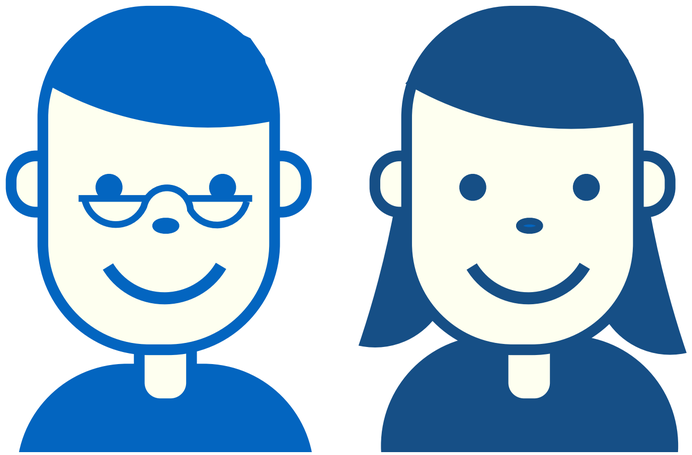
IBM Watson Analytics



Improvise

Definition

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



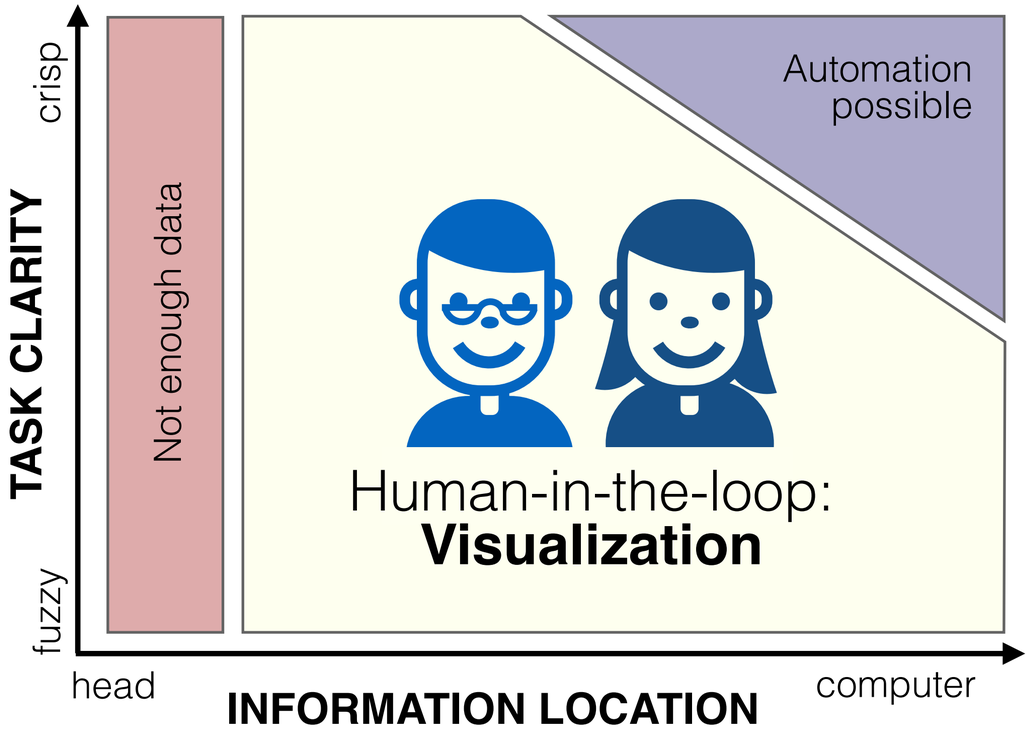
… not necessary if task can be fully automated

Definition

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

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

Why human-in-the-loop?



Examples for “automation possible”:

- handwritten text recognition
- autonomous driving
- …

Examples for ill-defined problems:

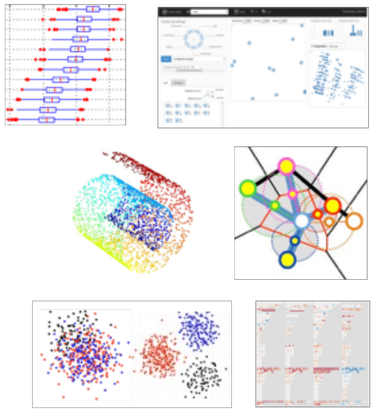
- exploratory analysis of scientific problems
- (collaborative) decision-making problems
- model building & validation
- …

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

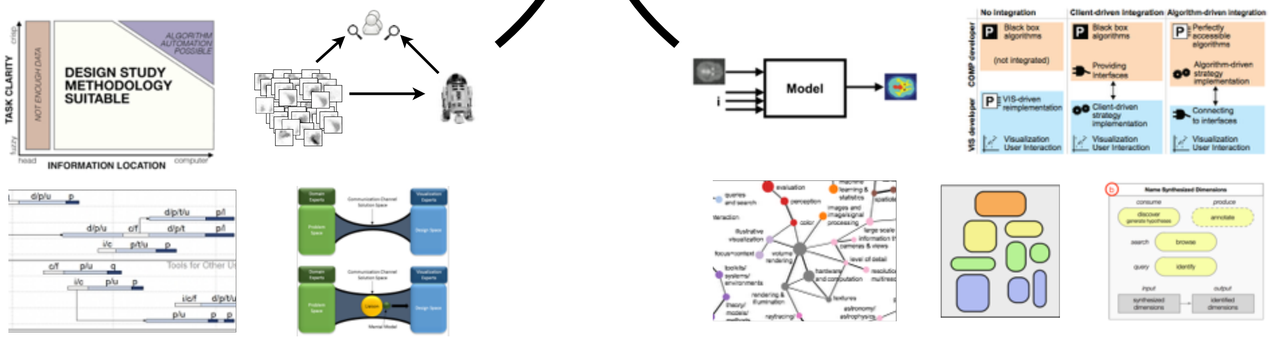
problem-driven

MY WORK

technique-driven



methods & theory



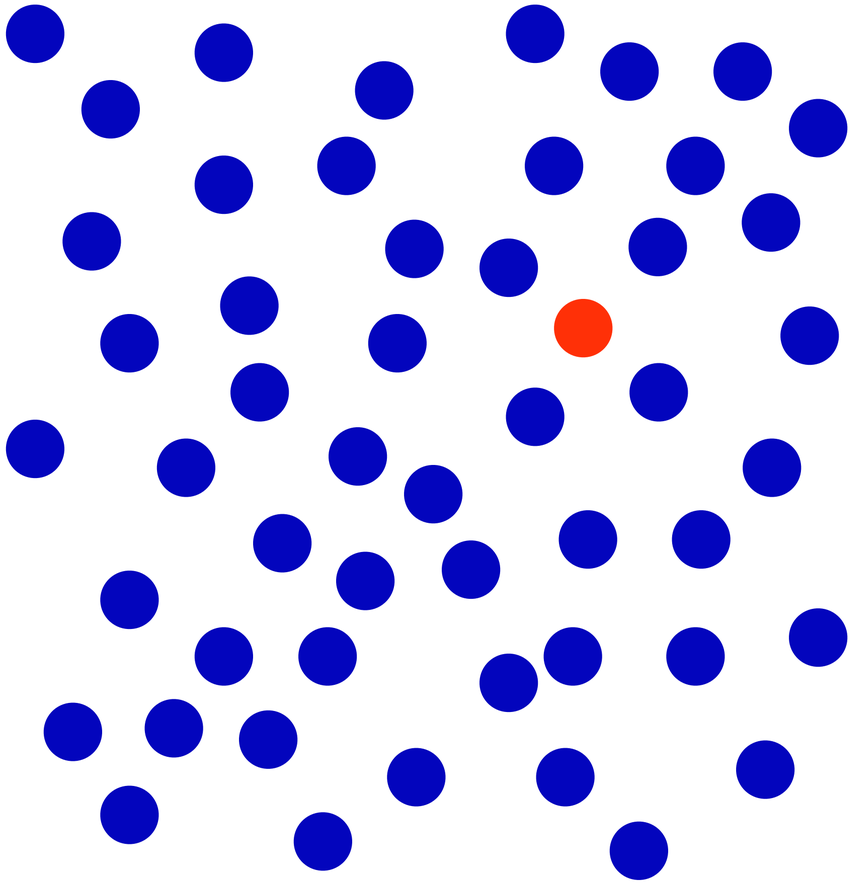
Outline: How to visualize data?

- Leverage perception
- Make it interactive
- Combine computation and visualization

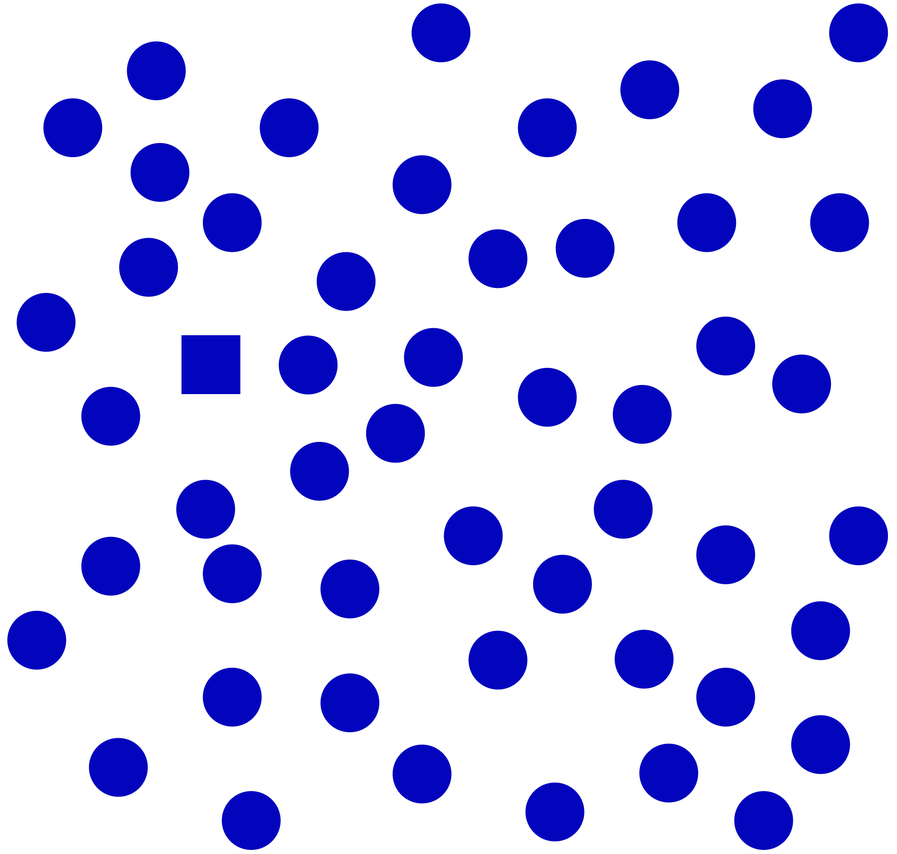
Leverage perception

...

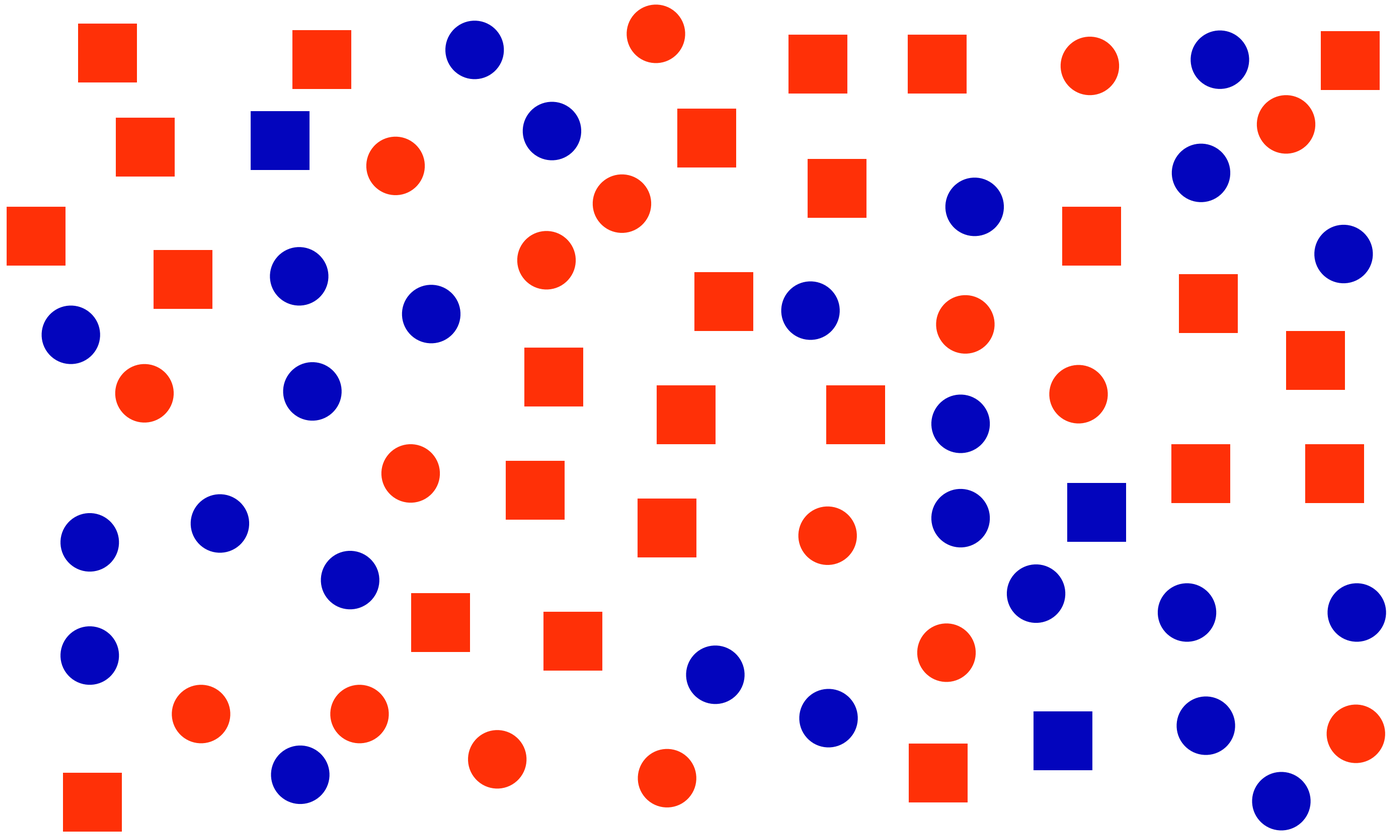
Human perception: Popout



Human perception: Popout

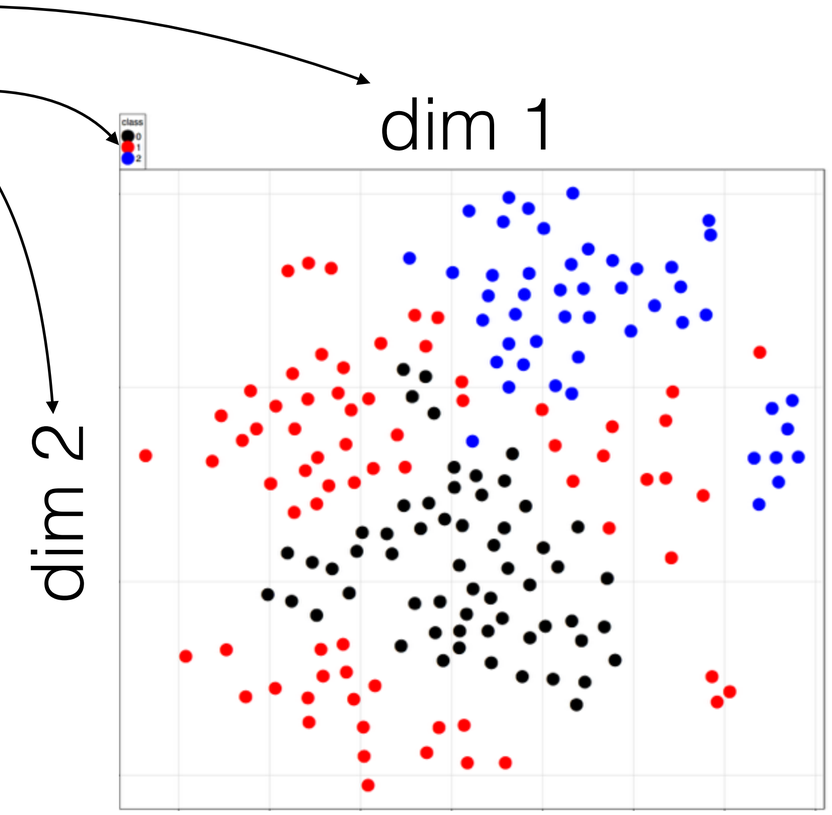


Popout: Combining channels?



2D Scatterplot

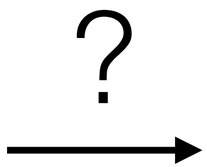
dim 1	dim 2	class
1.2	3.3	0
5.4	7.5	0
3.6	6.4	0
1.4	7.2	0
4.8	5.4	0
7.3	1.3	0
5.6	8.3	1
7.6	2.3	1
5.4	1.1	1
7.2	1.2	1
5.4	5.4	1
1.3	3.6	1
4.8	1.4	2
7.3	4.8	2
5.6	7.3	2
1.2	5.6	2
5.4	7.6	2
3.6	5.4	2
…		



data *visualization*

3D Scatterplot?

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
…			

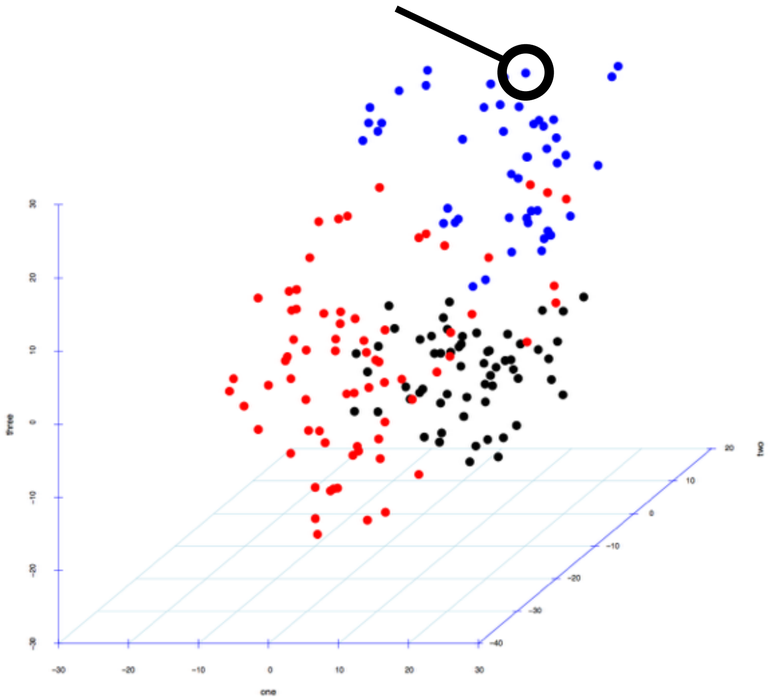


3D Scatterplot?

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
…			

?

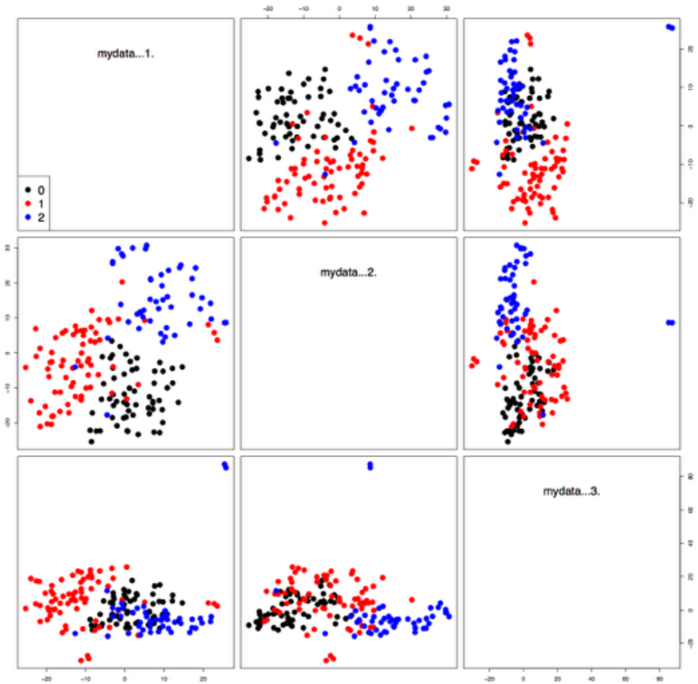
what's the value of this point?



No 3D for abstract data!

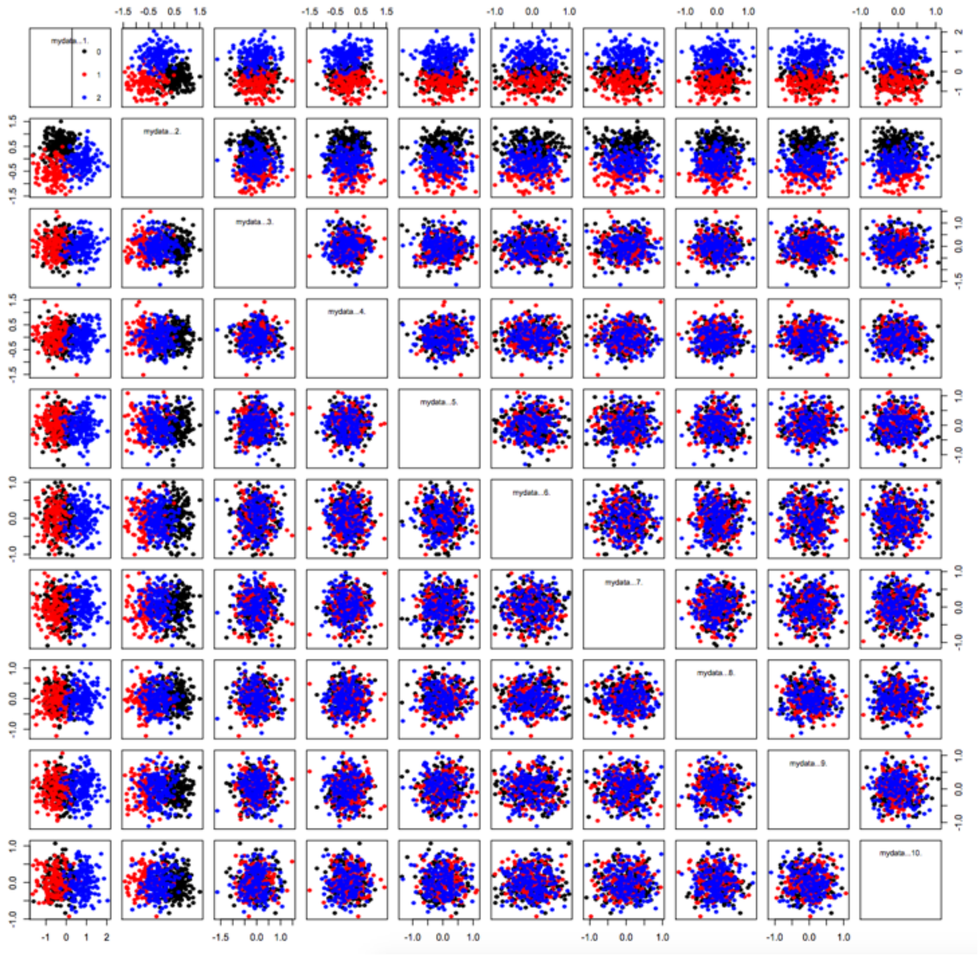
dim 1	dim 2	dim 3	class
1.2	3.3	5.4	0
5.4	7.5	1.3	0
3.6	6.4	4.8	0
1.4	7.2	7.3	0
4.8	5.4	5.6	0
7.3	1.3	1.2	0
5.6	8.3	5.4	1
7.6	2.3	3.6	1
5.4	1.1	3.3	1
7.2	1.2	7.5	1
5.4	5.4	6.4	1
1.3	3.6	7.2	1
4.8	1.4	5.4	2
7.3	4.8	1.3	2
5.6	7.3	8.3	2
1.2	5.6	3.3	2
5.4	7.6	7.5	2
3.6	5.4	1.2	2
…			

Scatterplot Matrix



SedImair, Munzner, Tory: Empirical Guidance on Scatterplot and Dimension Reduction Technique Choices (InfoVis 2013).

Scatterplot Matrix with more dimensions?

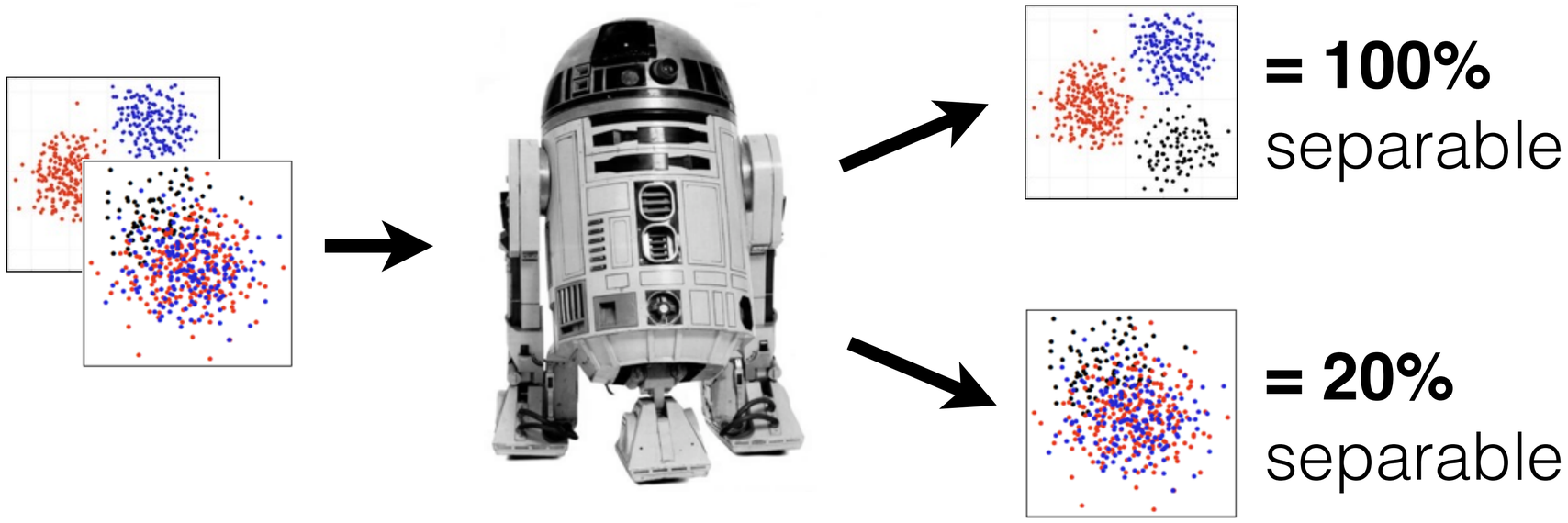


10D —> 45 Views

200D —> ~20k Views

Modeling human perception

e.g. class separation measures

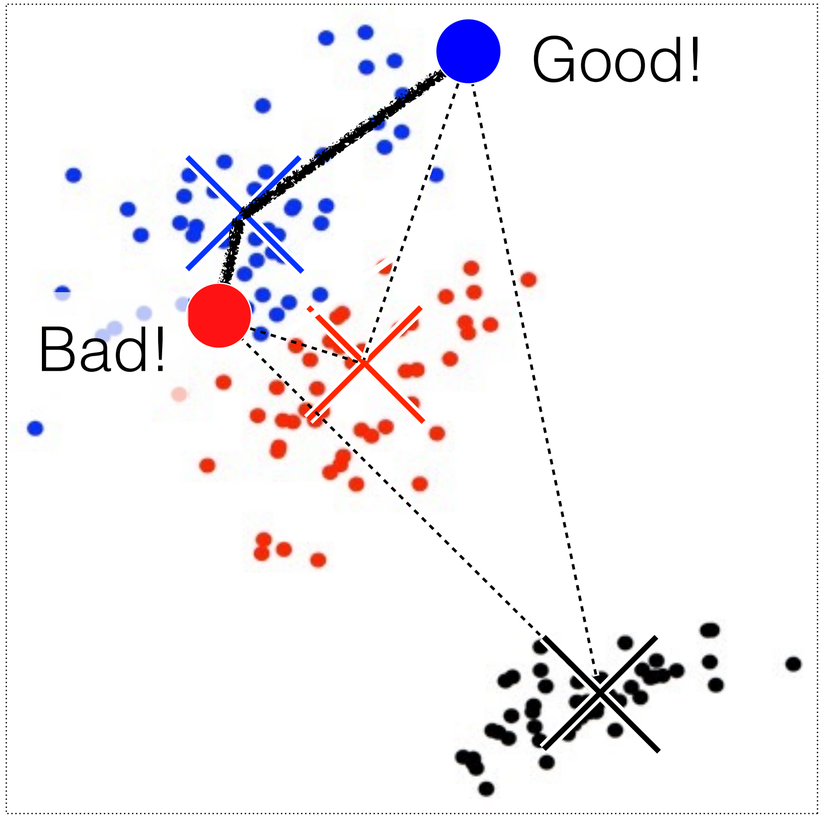


Aupetit, **Sedlmair**: SepMe: 2002 New Visual Separation Measures (PacificVis 2016).

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

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

Example: Centroid measure



Centroid measure
= **93%** separable

Sips et al.: Selecting good views of high-dimensional data using class consistency (EuroVis 2009).

How good is it?

- Evaluation of 35 measures
- How to evaluate?
  - HCI-style user studies?
  - Using a Computer Vision Framework!

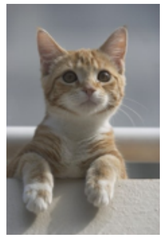
Sedlmair, Aupetit: Data-driven Evaluation of Visual Quality Measures (EuroVis 2015).

A “Computer Vision” framework

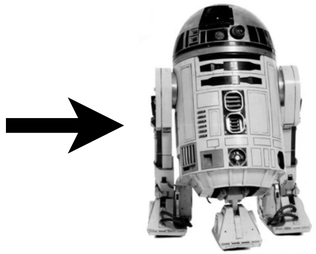
FreeImages.com

or

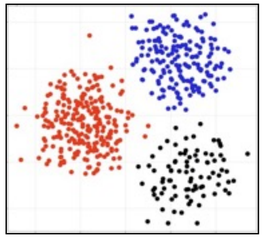
FreeImages.com



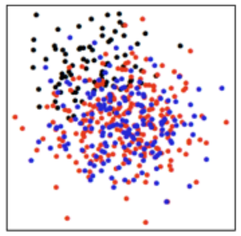
FreeImages.com

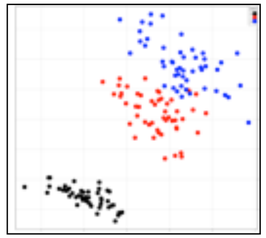


cat!



or

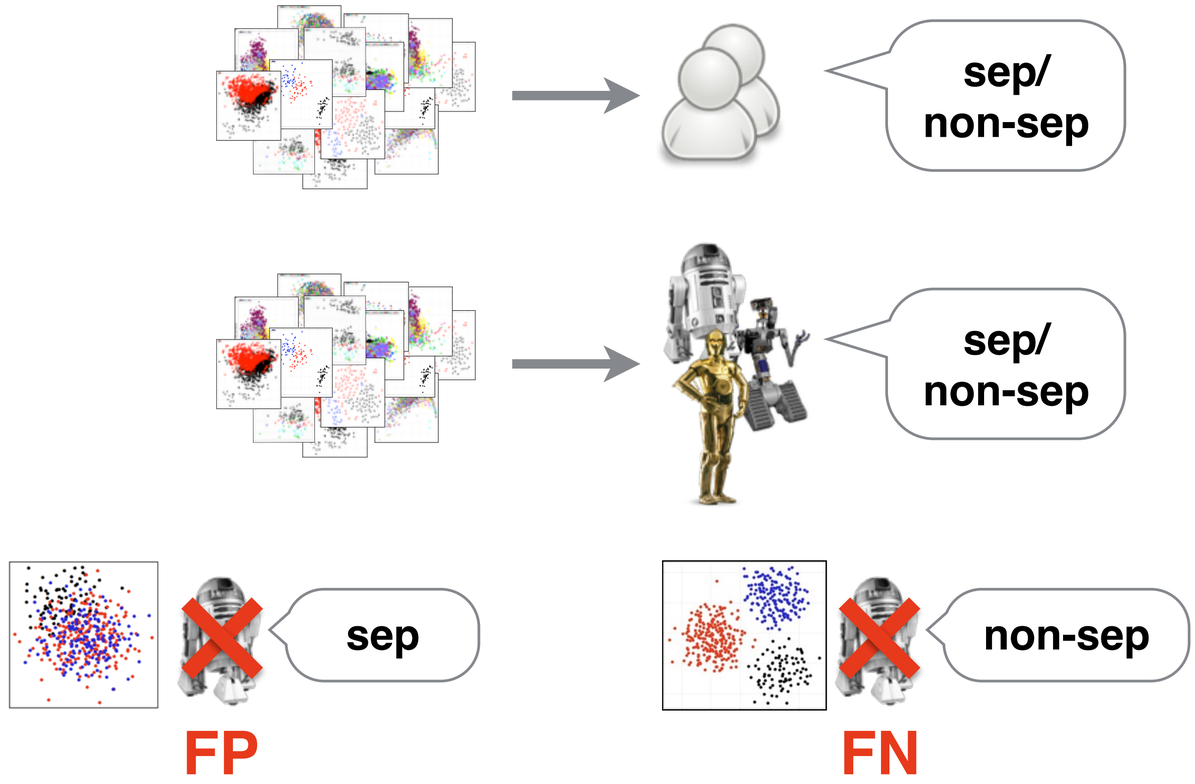




separable!

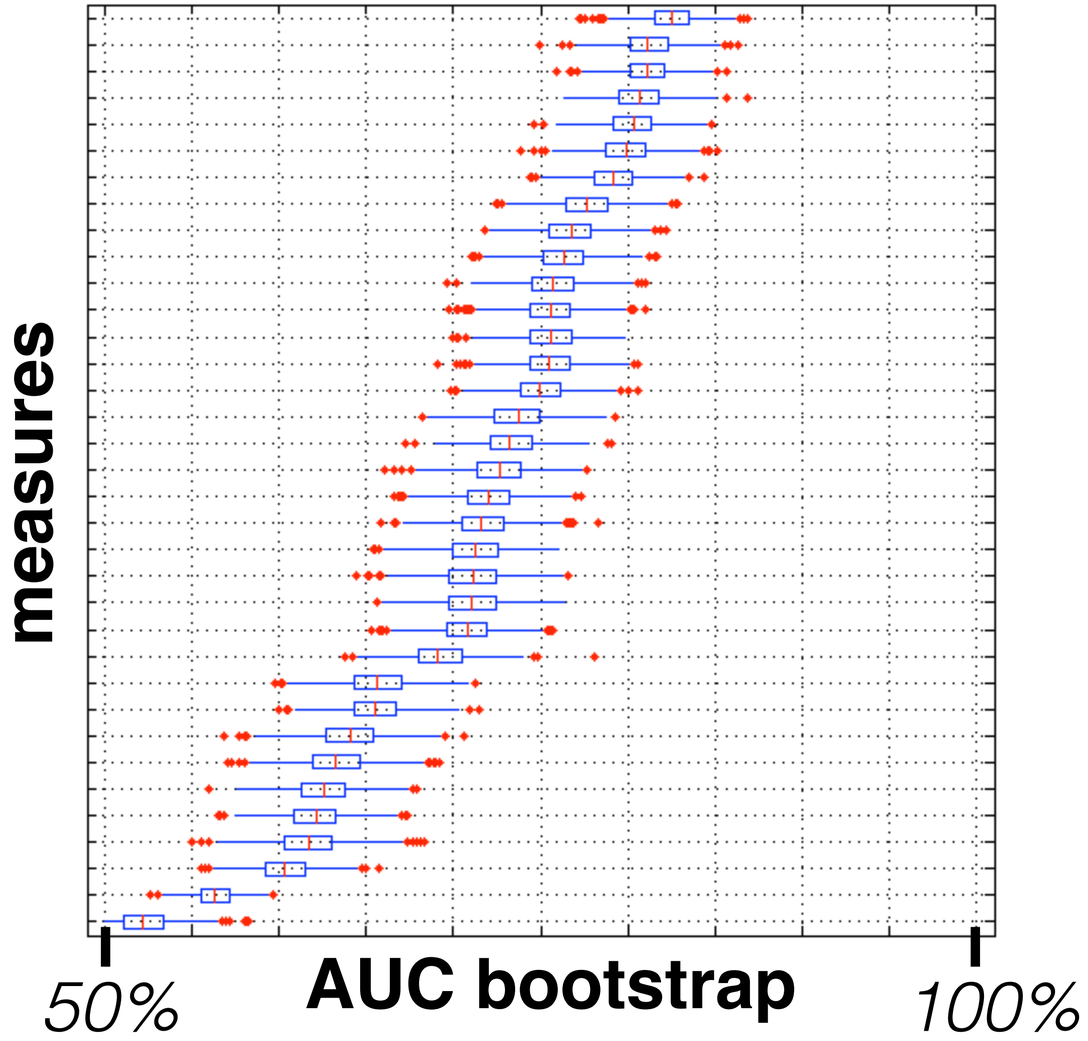
Evaluate with standard Machine Learning pipeline

1. gather large collection of “ground truth” data
2. predict with different measures
3. evaluate quality: FP/FN
4. ROC/AUC
5. bootstrapping to generalize

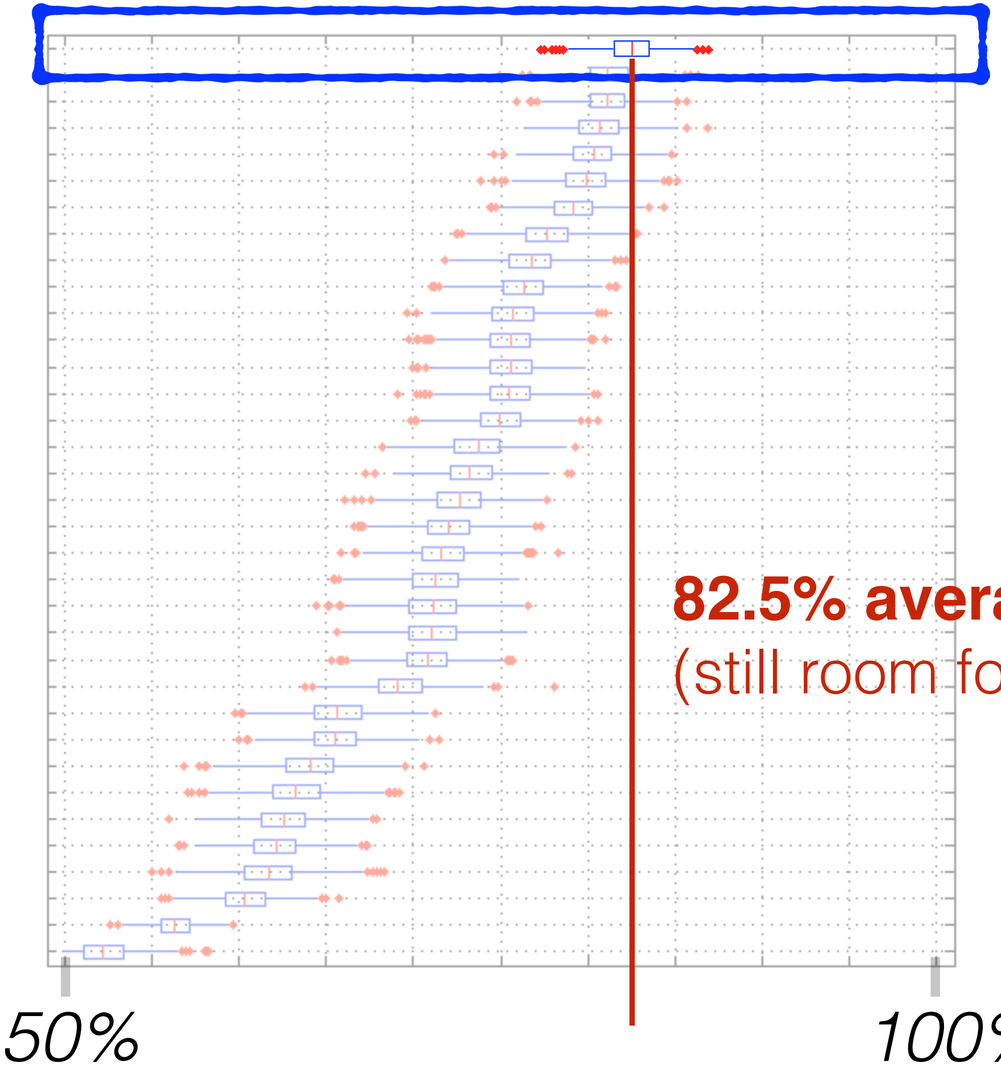


ROC/AUC:
50% = random guess
100% = perfect classification

Results



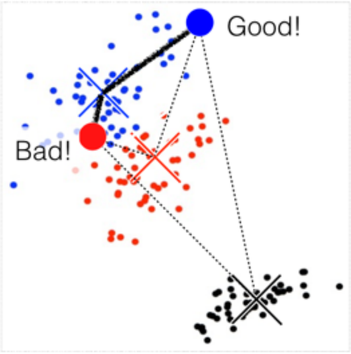
The Winner



Centroid Measure

DSC - Distance Consistency

Sips et al.: Selecting good views of high-dimensional data using class consistency (EuroVis 2009).

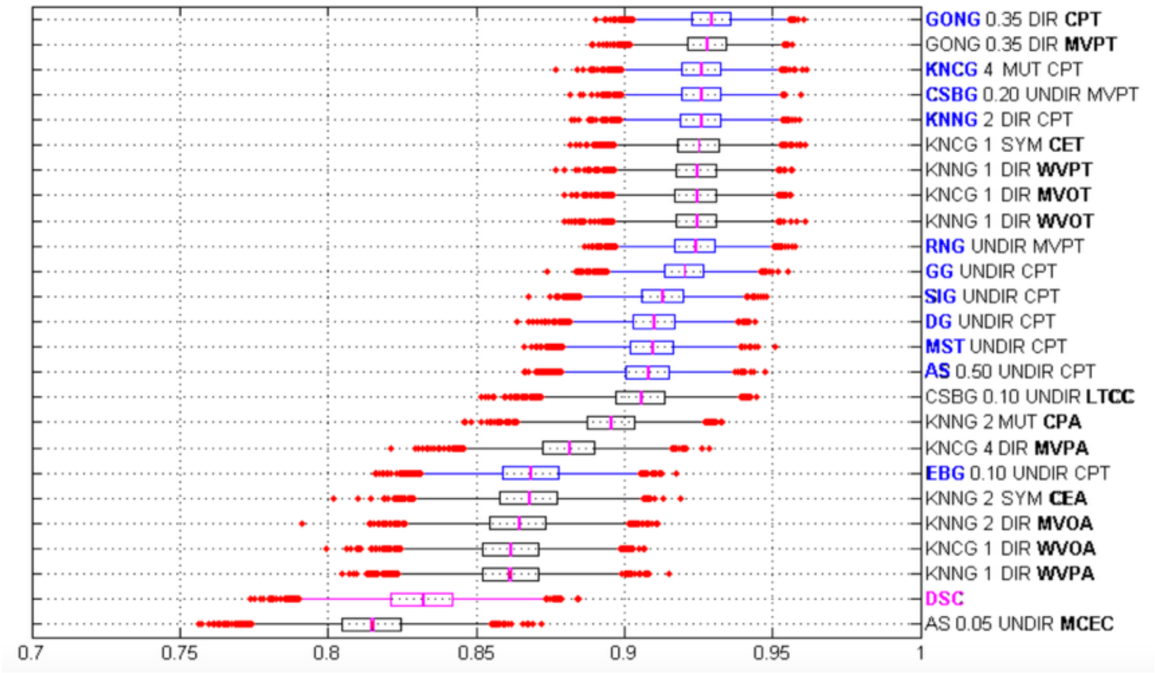


82.5% average AUC

(still room for ~20% improvement)

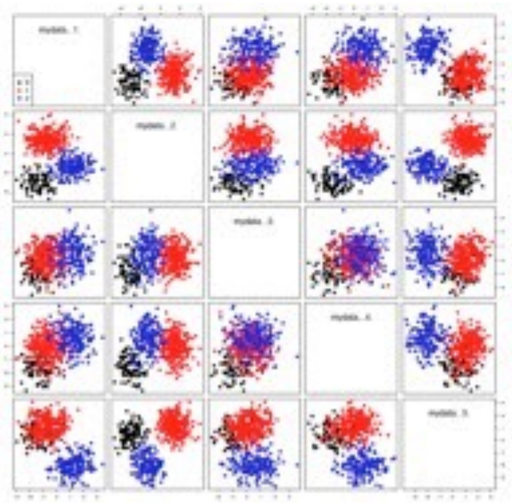
Follow-up work: PacificVis 2016

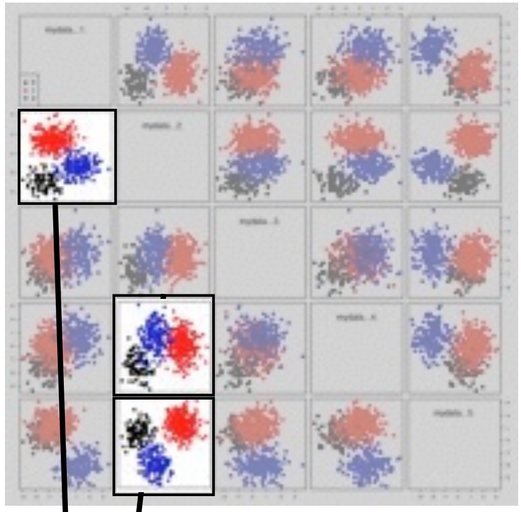
- 2002 New Visual Separation Measures
- 58% of our 2002 new measures outperform best state-of-the-art DSC
- Our best measure
  - GONG 0.35 DIR CPT
  - Prediction accuracy 92.9% (11.7% better than DSC)



Aupetit, **Sedlmair**: SepMe: 2002 New Visual Separation Measures (PacificVis 2016).

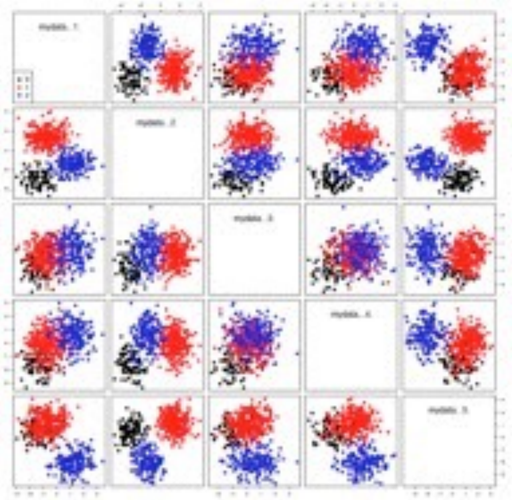
Measures help humans finding the good views

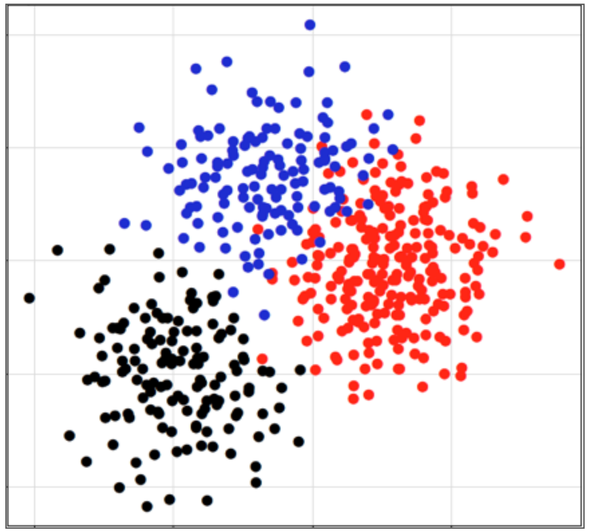




Good!

Perception-driven Dimensionality Reduction



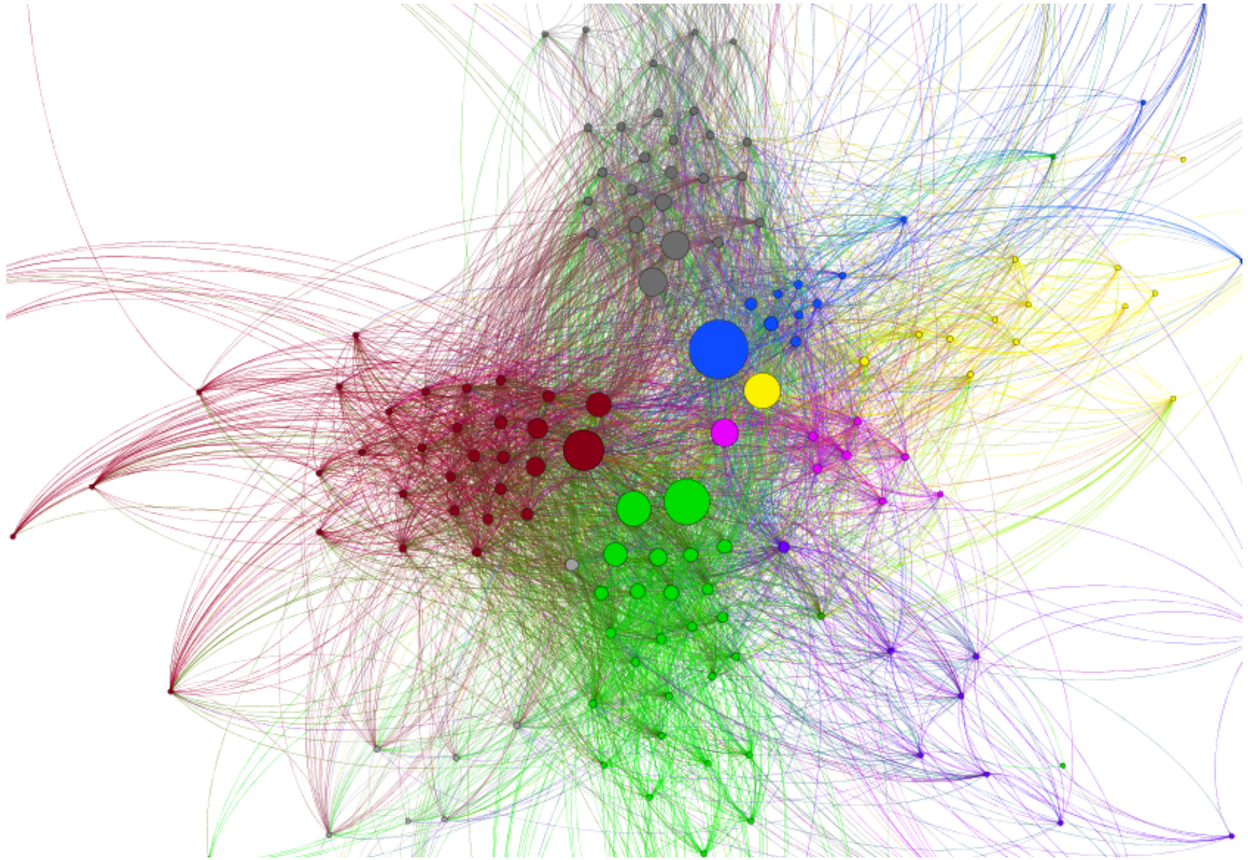


- Find the linear projection that perceptually separates classes as well as possible
- Better than LDA!

Yunhai Wang, Kang Feng, Xiaowei Chu, Jian Zhang, Chi-Wing Fu, **Sedlmair**, Xiaohui Yu, Baoquan Chen.
A Perception-Driven Approach to Supervised Dimensionality Reduction for Visualization. (TVCG 2017)

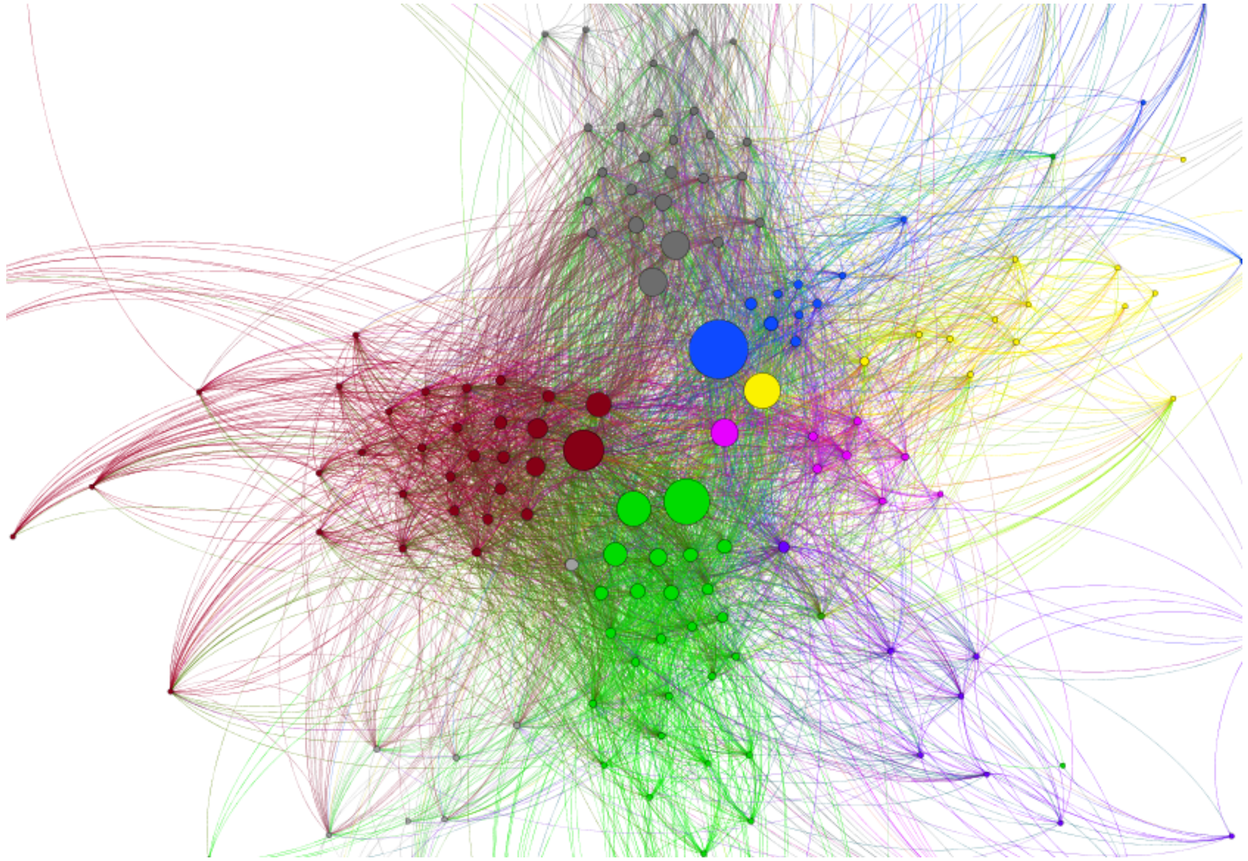
Make it interactive …

Showing all the data?



https://twitter.com/axelmaireder/media

Showing all the data?



https://twitter.com/axelmaireder/media

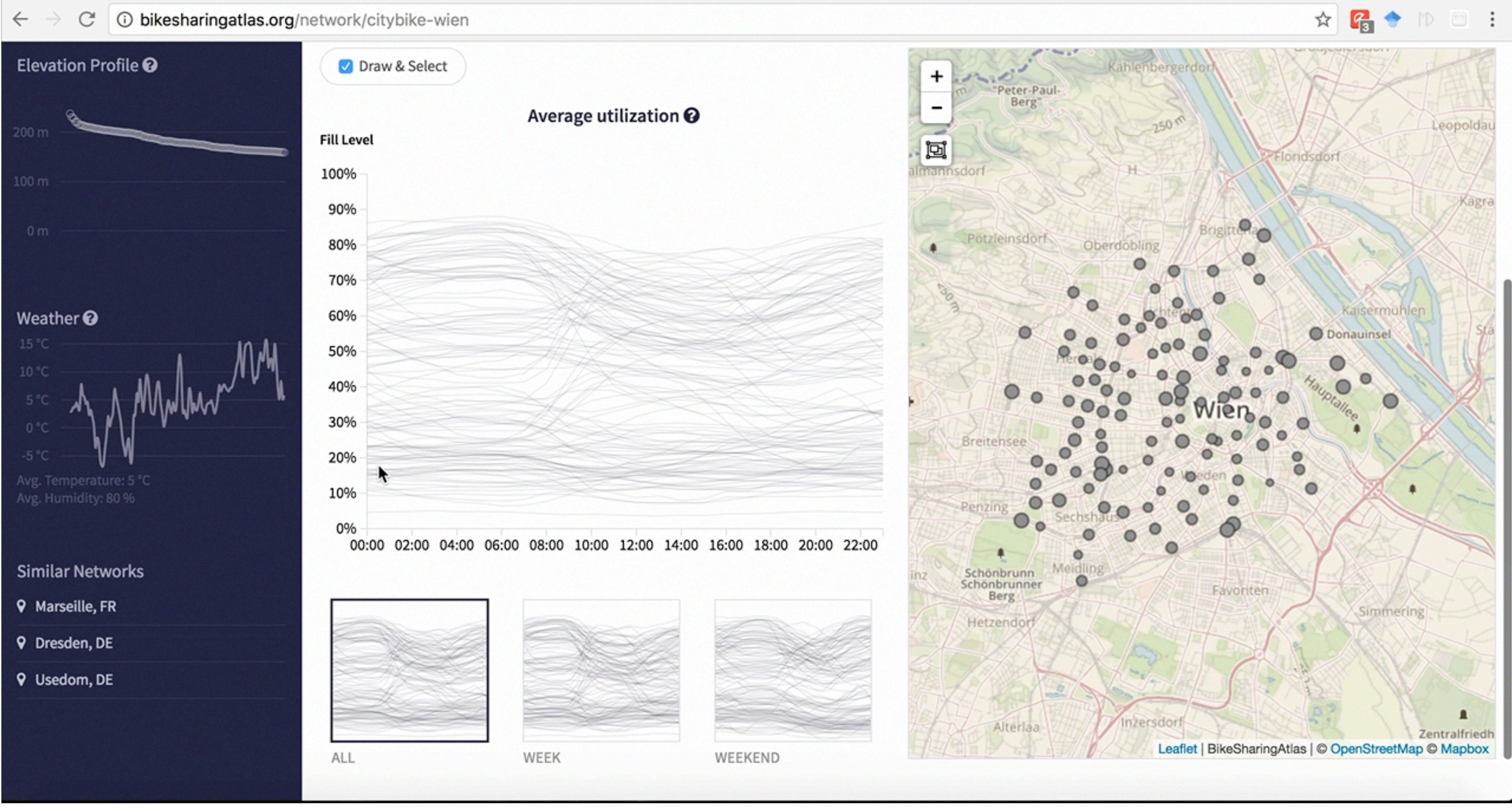
Hairball

Make it interactive

Oppermann, Möller, **Sedlmair**.
To appear in International Journal of Transportation.

example:
bikesharingatlas.org

- 468 networks
- 21.500 stations
- Fill levels every 15 min, for 1.5 years (i.e. >1 billion fill level values)


bikesharingatlas.org/network/citybike-wien
Elevation Profile
200 m
100 m
0 m
Weather
15 °C
10 °C
5 °C
0 °C
-5 °C
Avg. Temperature: 5 °C
Avg. Humidity: 80 %
Similar Networks
Marseille, FR
Dresden, DE
Usedom, DE
Draw & Select
Average utilization
Fill Level
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
00:00 02:00 04:00 06:00 08:00 10:00 12:00 14:00 16:00 18:00 20:00 22:00
ALL
WEEK
WEEKEND
Wien
Donauinsel
Leaflet | BikeSharingAtlas | © OpenStreetMap © Mapbox

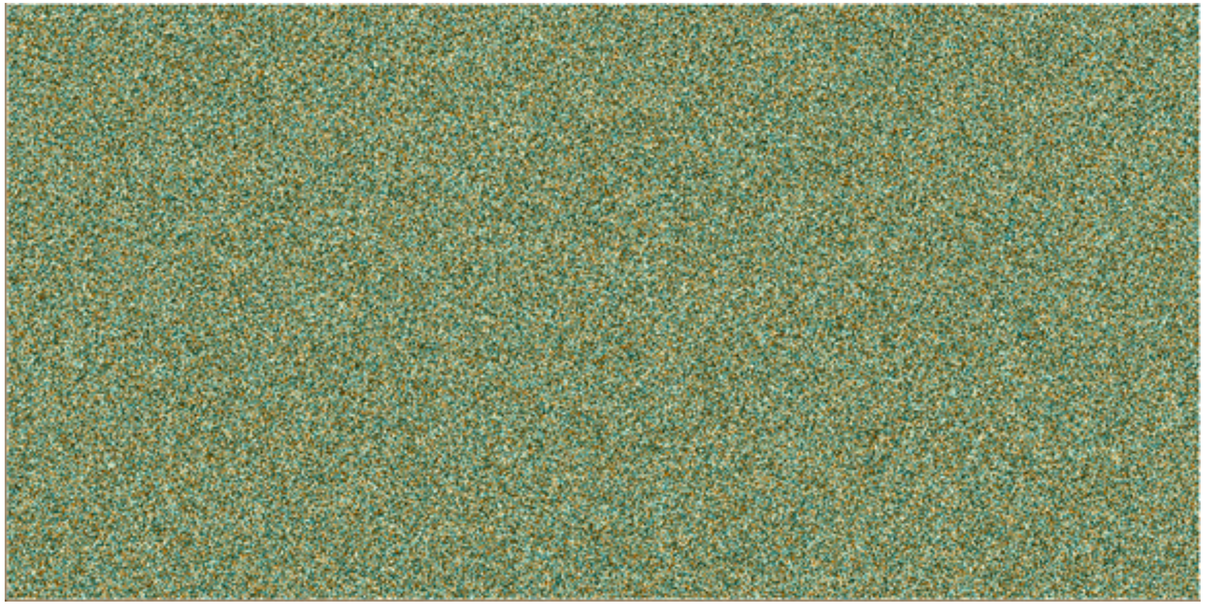
bikesharingatlas.org

Oppermann, Möller, **Sedlmair**.
To appear in International Journal of Transportation.

- Multiple views - linking & brushing
- Novel patterns and insights
- Different user groups
  (network operators, city planners, politicians, journalists, general public, etc.)

Combine
Computation & Visualization

Even too much data for interaction



https://www.geek.com/geek-cetera/pi-visualized-to-4-million-decimal-places-in-one-image-1486849/

Combine visual and computational

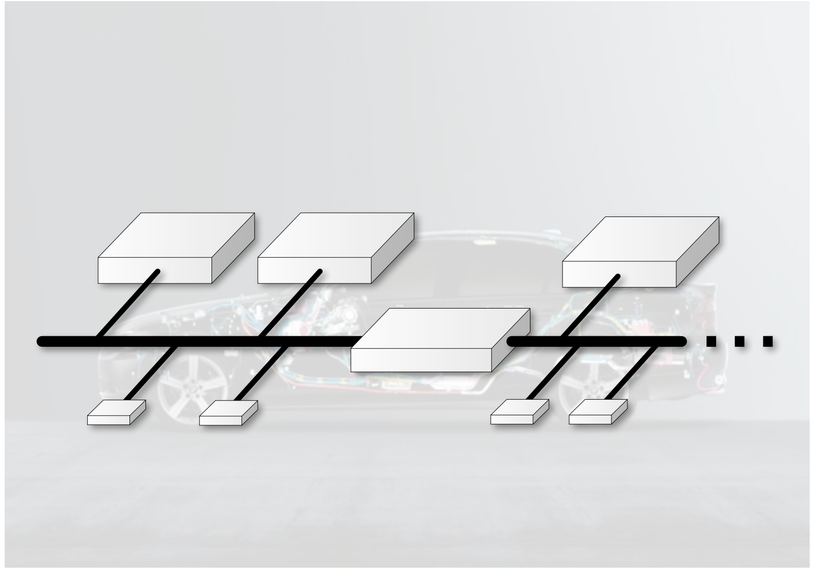
example: Cardiogram @BMW

Sedlmair, Isenberg, Baur, Mauerer, Pigorsch, Butz: Cardiogram: Visual Analytics for Automotive Engineers (CHI 2011).





Combine visual and computational

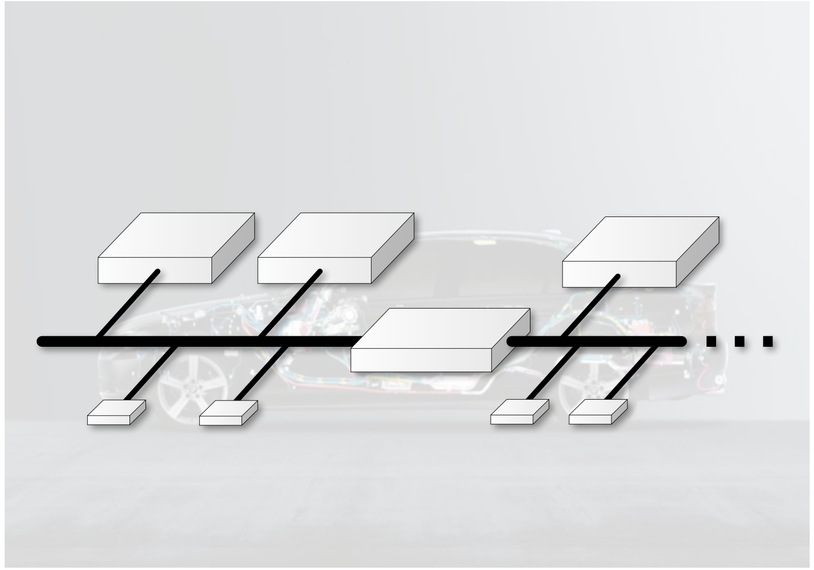


Sedlmair, Isenberg, Baur, Mauerer, Pigorsch, Butz: Cardiogram: Visual Analytics for Automotive Engineers (CHI 2011).

example: Cardiogram @BMW

- 70 interconnected ECUs

Combine visual and computational

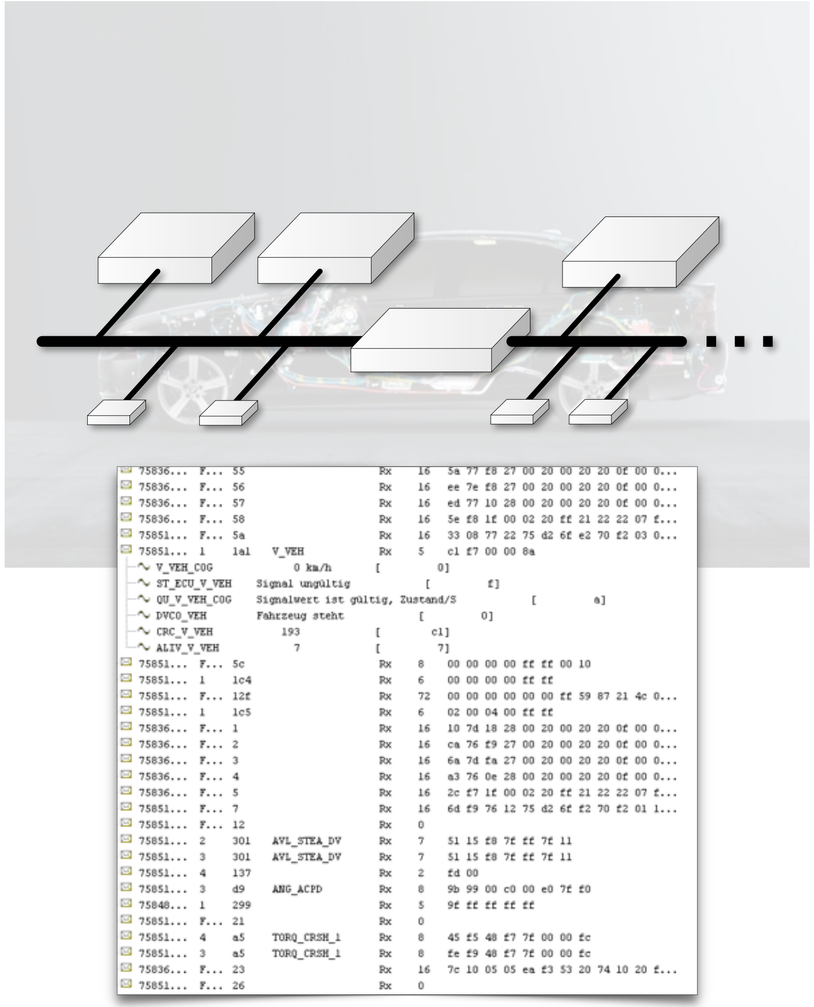


Sedlmair, Isenberg, Baur, Mauerer, Pigorsch, Butz: Cardiogram: Visual Analytics for Automotive Engineers (CHI 2011).

example: Cardiogram @BMW

- 70 interconnected ECUs
- **task:** finding errors
- **process:** analyse test drives
- **data:** recorded traces (15k messages/sec)

Combine visual and computational



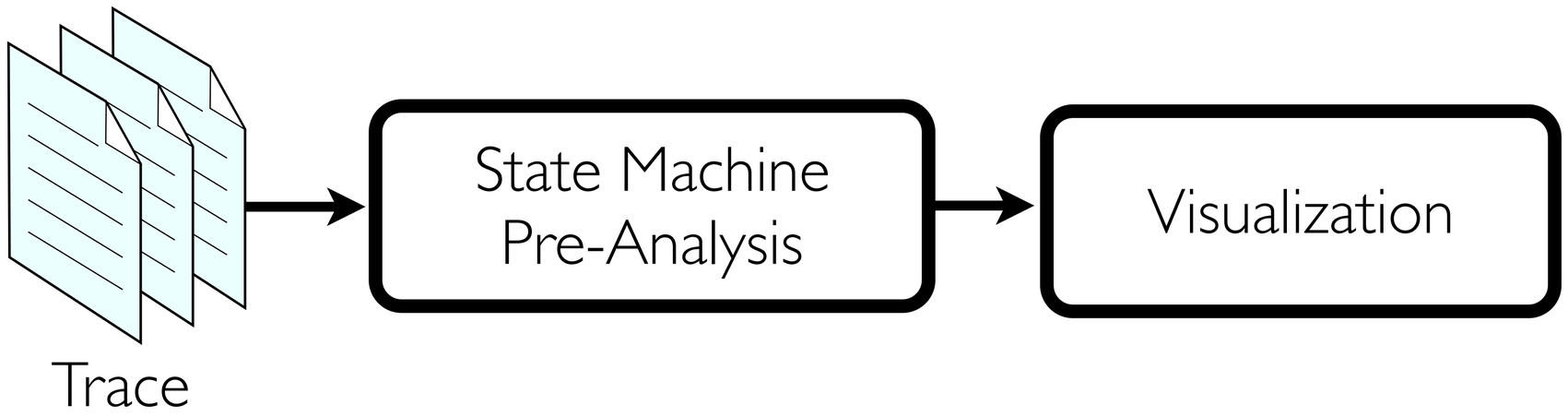
example: Cardiogram @BMW

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

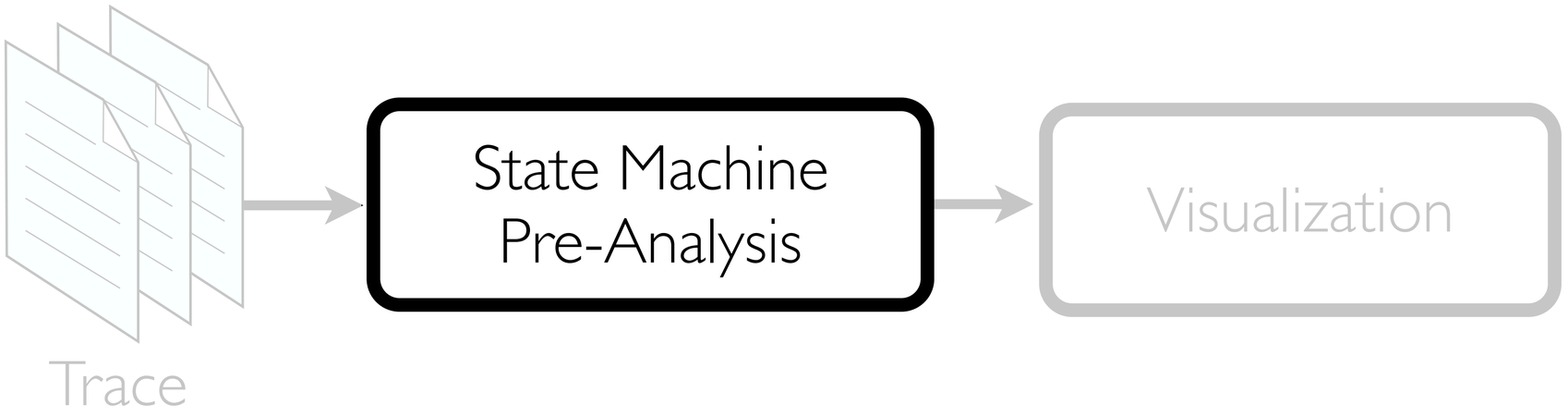
Main challenges

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

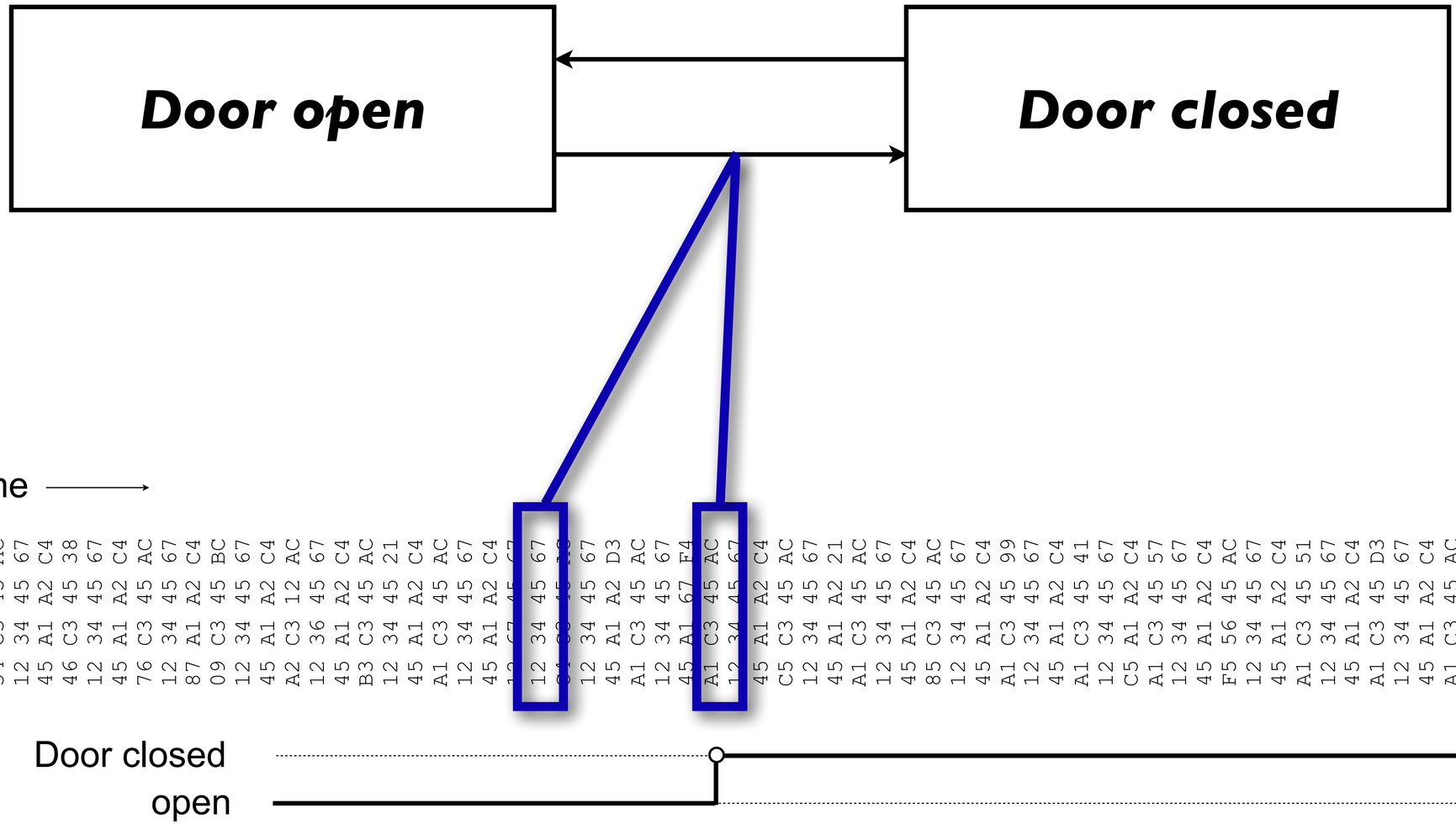
Our solution: **Cardiogram**



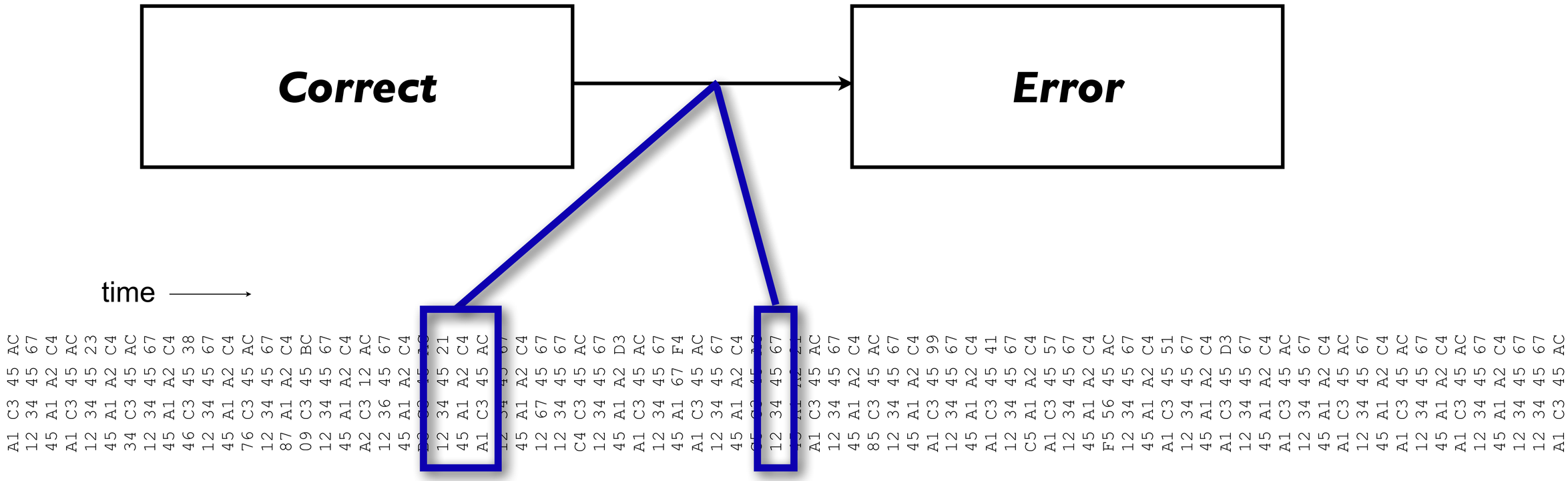
State machine pre-analysis



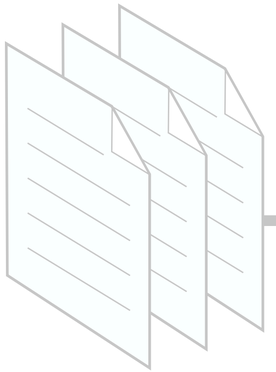
State machines: **Behavior**



State machines: **Error detection**



Create/select state machines

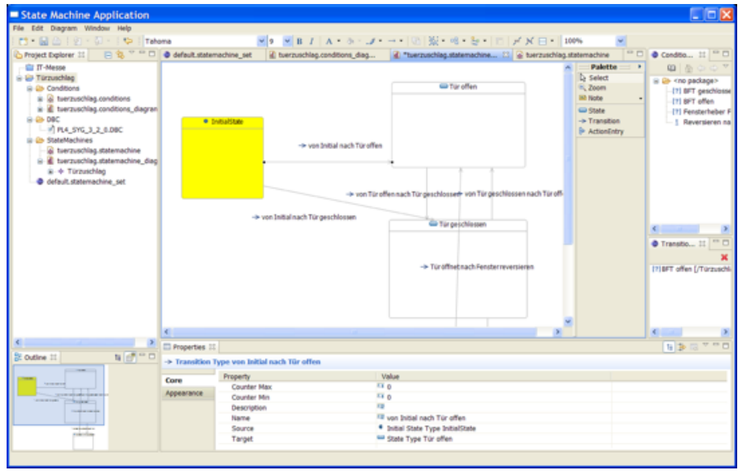


Trace

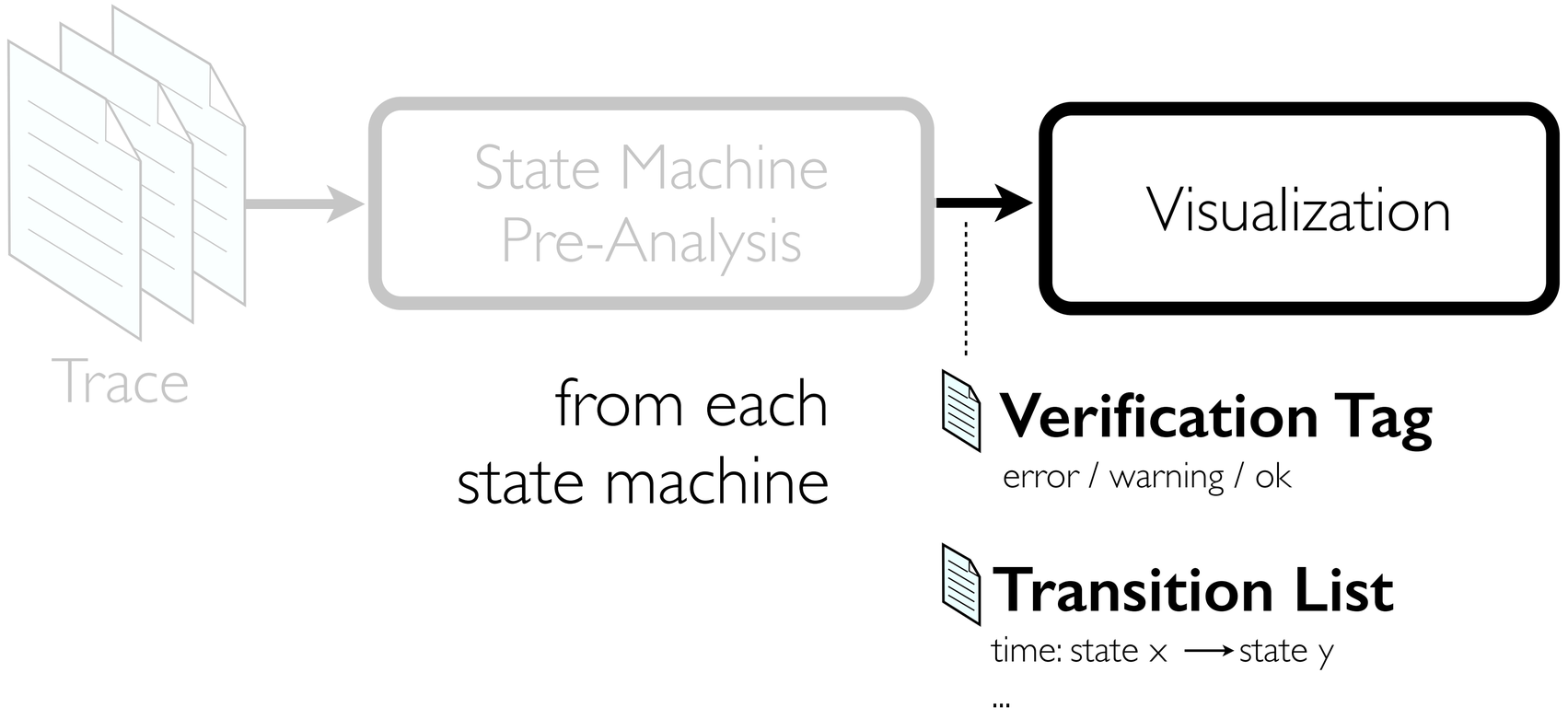
State Machine
Pre-Analysis

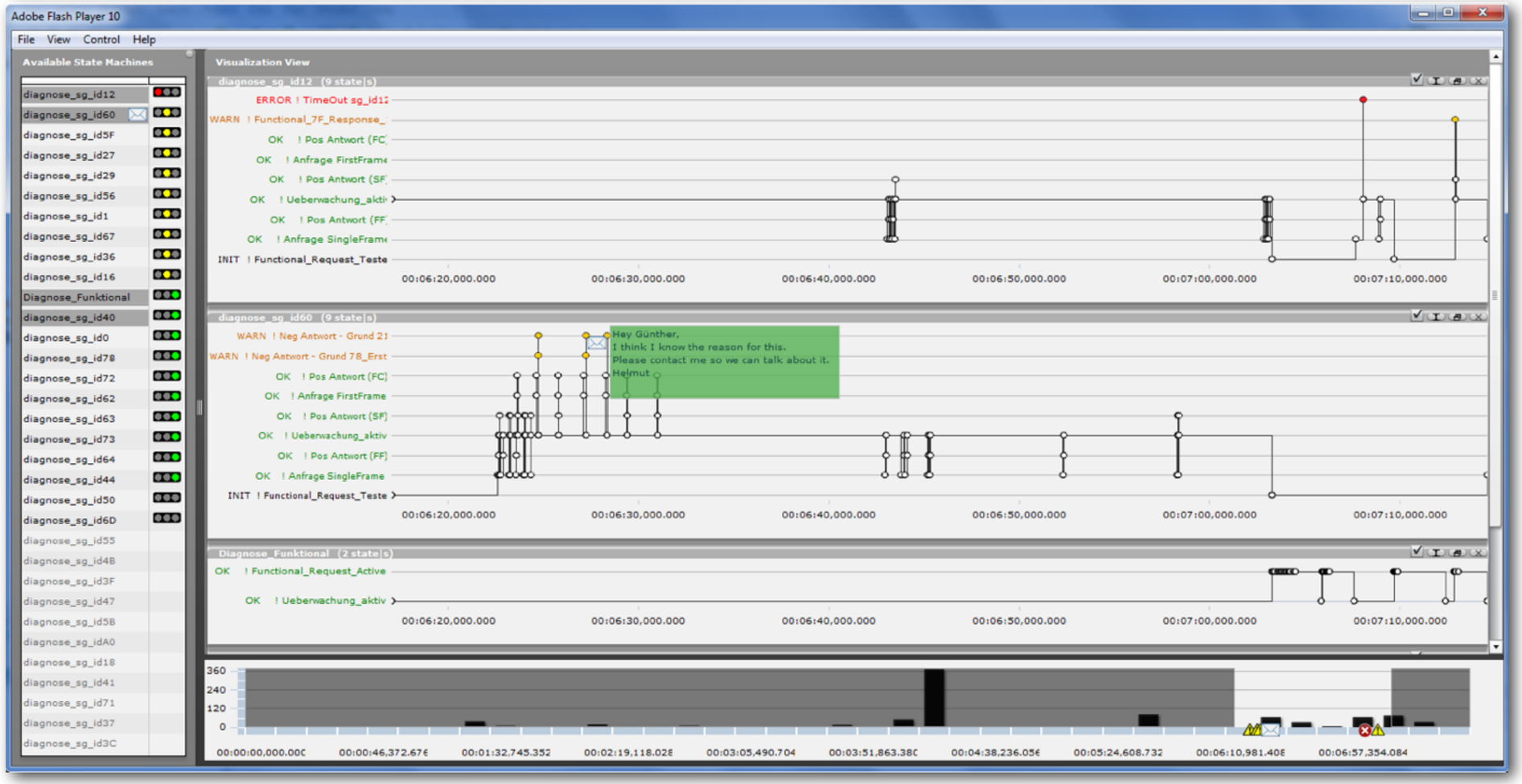
Visualization

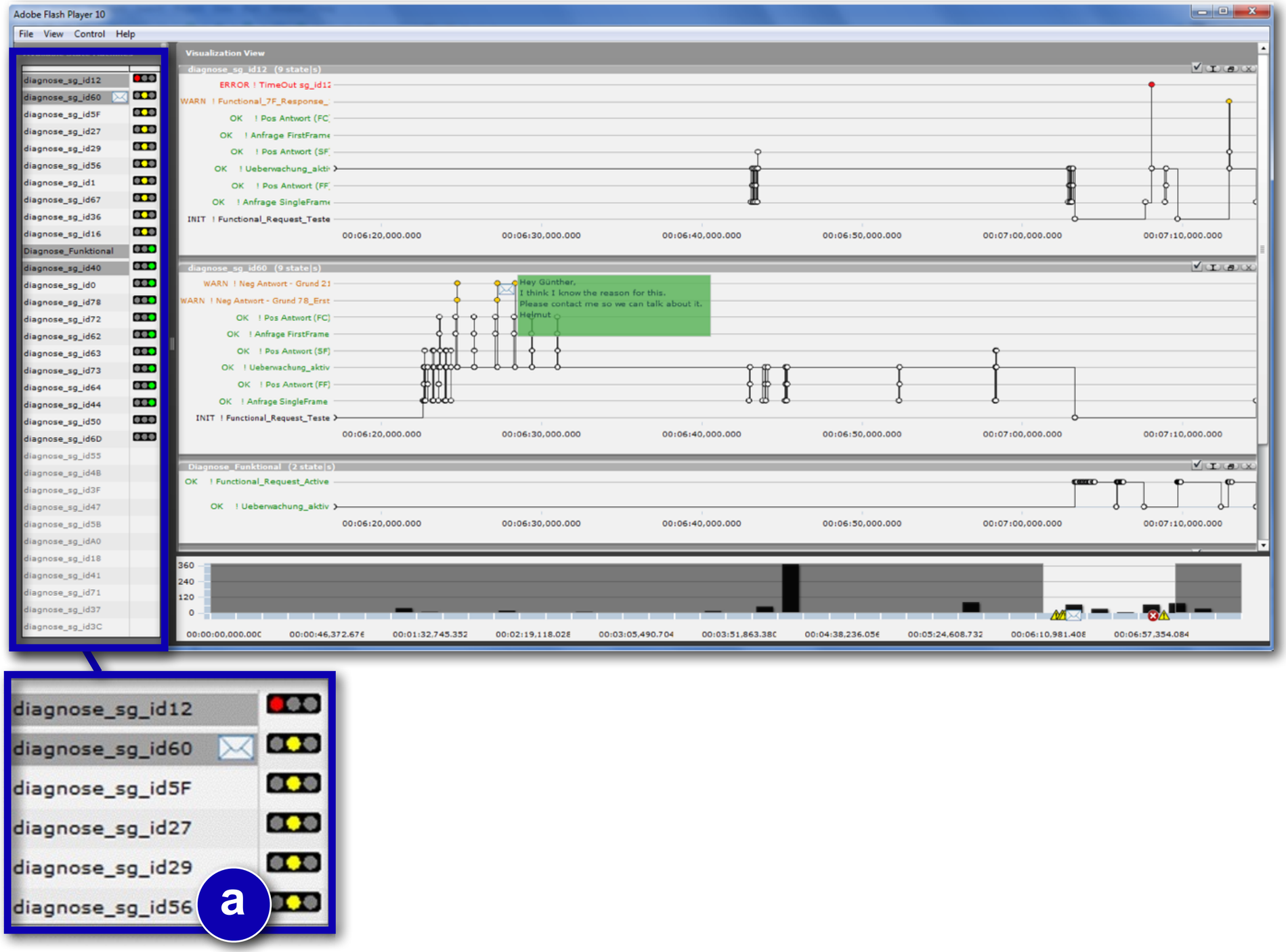
Multiple State Machines
- State Machine 1
- State Machine 2
- State Machine 3
... (dozens)

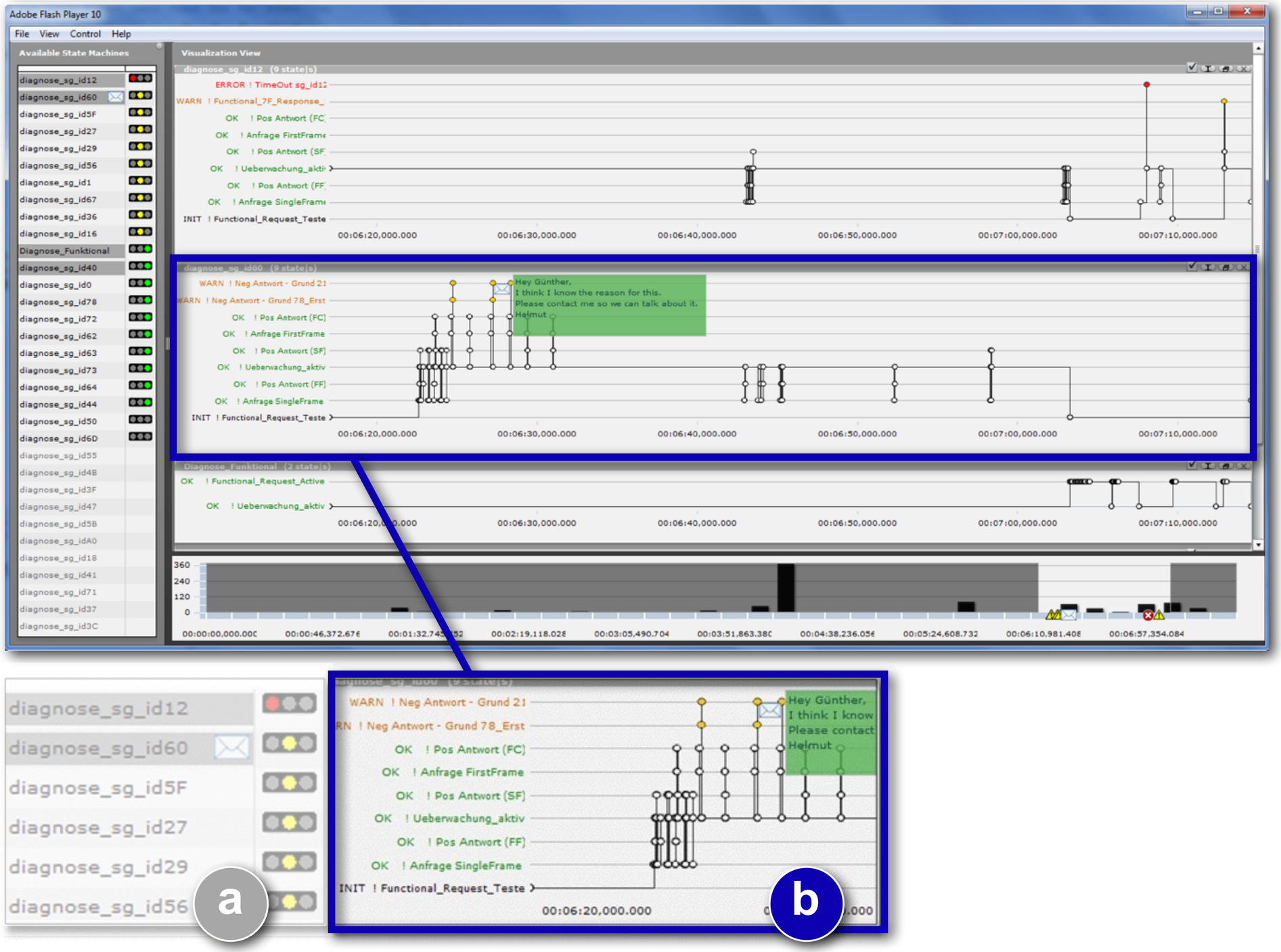


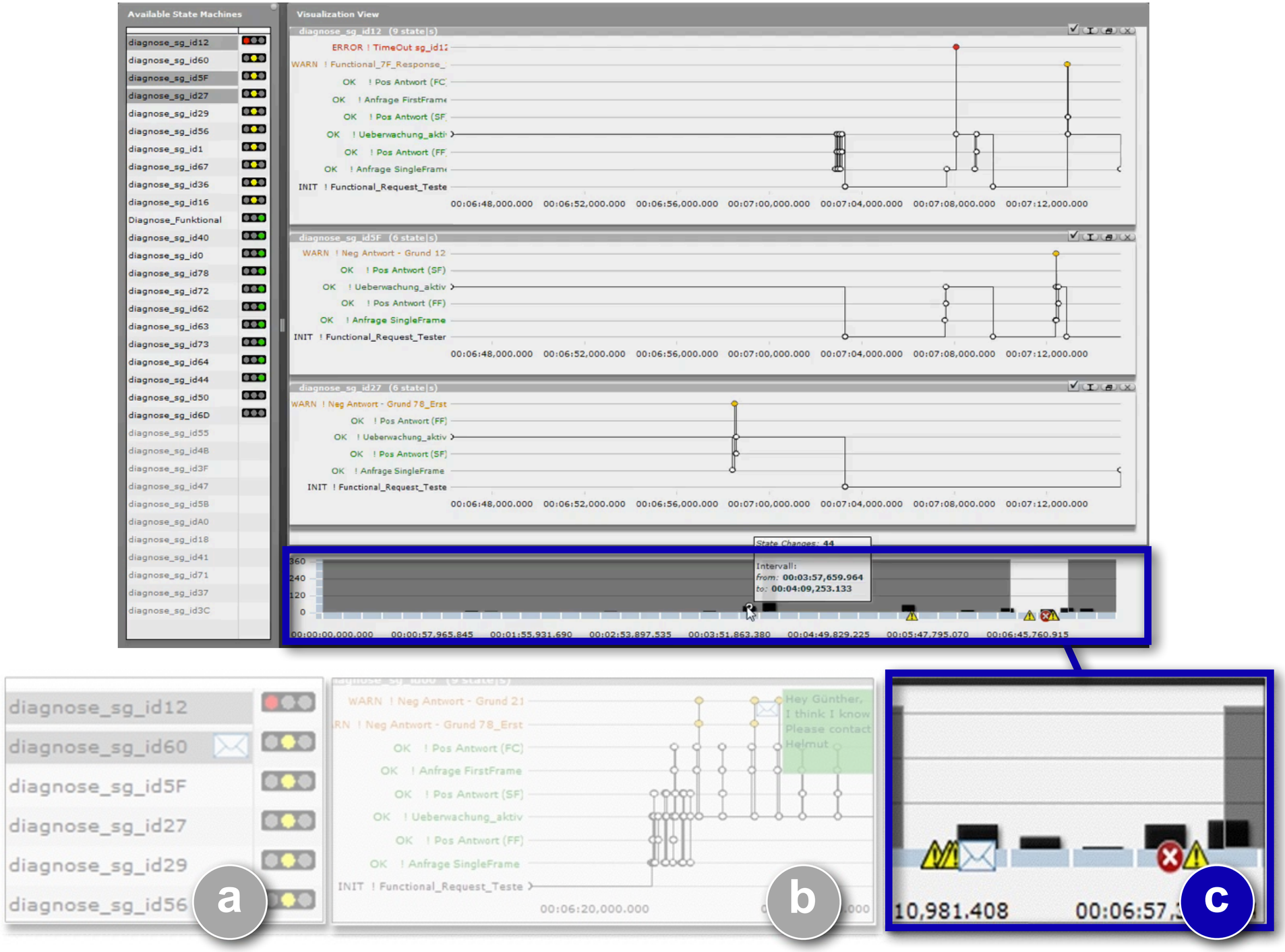
Data abstraction/reduction




Adobe Flash Player 10
File View Control Help
Available State Machines
diagnose_sg_id12
diagnose_sg_id60
diagnose_sg_id5F
diagnose_sg_id27
diagnose_sg_id29
diagnose_sg_id56
diagnose_sg_id1
diagnose_sg_id67
diagnose_sg_id36
diagnose_sg_id16
Diagnose_Funktional
diagnose_sg_id40
diagnose_sg_id0
diagnose_sg_id78
diagnose_sg_id72
diagnose_sg_id62
diagnose_sg_id63
diagnose_sg_id73
diagnose_sg_id64
diagnose_sg_id44
diagnose_sg_id50
diagnose_sg_id6D
diagnose_sg_id55
diagnose_sg_id4B
diagnose_sg_id3F
diagnose_sg_id47
diagnose_sg_id5B
diagnose_sg_idA0
diagnose_sg_id18
diagnose_sg_id41
diagnose_sg_id71
diagnose_sg_id37
diagnose_sg_id3C
Visualization View
diagnose_sg_id12 (9 state|s)
ERROR ! TimeOut sg_id12
WARN ! Functional_7F_Response_
OK ! Pos Antwort (FC
OK ! Anfrage FirstFrame
OK ! Pos Antwort (SF
OK ! Ueberwachung_aktiv
OK ! Pos Antwort (FF
OK ! Anfrage SingleFrame
INIT ! Functional_Request_Teste
00:06:20,000.000
00:06:30,000.000
00:06:40,000.000
00:06:50,000.000
00:07:00,000.000
00:07:10,000.000
diagnose_sg_id60 (9 state|s)
WARN ! Neg Antwort - Grund 21
WARN ! Neg Antwort - Grund 78_Erst
OK ! Pos Antwort (FC)
OK ! Anfrage FirstFrame
OK ! Pos Antwort (SF)
OK ! Ueberwachung_aktiv
OK ! Pos Antwort (FF)
OK ! Anfrage SingleFrame
INIT ! Functional_Request_Teste
Hey Günther,
I think I know the reason for this.
Please contact me so we can talk about it.
Helmut
00:06:20,000.000
00:06:30,000.000
00:06:40,000.000
00:06:50,000.000
00:07:00,000.000
00:07:10,000.000
Diagnose_Funktional (2 state|s)
OK ! Functional_Request_Active
OK ! Ueberwachung_aktiv
00:06:20,000.000
00:06:30,000.000
00:06:40,000.000
00:06:50,000.000
00:07:00,000.000
00:07:10,000.000
360
240
120
0
00:00:00,000.000
00:00:46,372.676
00:01:32,745.352
00:02:19,118.028
00:03:05,490.704
00:03:51,863.380
00:04:38,236.056
00:05:24,608.732
00:06:10,981.408
00:06:57,354.084


Adobe Flash Player 10
File View Control Help
Visualization View
diagnose_sg_id12 (9 state|s)
ERROR ! TimeOut sg_id12
WARN ! Functional_7F_Response_
OK ! Pos Antwort (FC
OK ! Anfrage FirstFrame
OK ! Pos Antwort (SF
OK ! Ueberwachung_aktiv
OK ! Pos Antwort (FF
OK ! Anfrage SingleFrame
INIT ! Functional_Request_Teste
00:06:20,000.000
00:06:30,000.000
00:06:40,000.000
00:06:50,000.000
00:07:00,000.000
00:07:10,000.000
diagnose_sg_id60 (9 state|s)
WARN ! Neg Antwort - Grund 21
WARN ! Neg Antwort - Grund 78_Erst
OK ! Pos Antwort (FC)
OK ! Anfrage FirstFrame
OK ! Pos Antwort (SF)
OK ! Ueberwachung_aktiv
OK ! Pos Antwort (FF)
OK ! Anfrage SingleFrame
INIT ! Functional_Request_Teste
Hey Günther,
I think I know the reason for this.
Please contact me so we can talk about it.
Helmut
Diagnose_Funktional (2 state|s)
OK ! Functional_Request_Active
OK ! Ueberwachung_aktiv
360
240
120
0
00:00:00,000.000
00:00:46,372.676
00:01:32,745.352
00:02:19,118.028
00:03:05,490.704
00:03:51,863.380
00:04:38,236.056
00:05:24,608.732
00:06:10,981.408
00:06:57,354.084
diagnose_sg_id12
diagnose_sg_id60
diagnose_sg_id5F
diagnose_sg_id27
diagnose_sg_id29
diagnose_sg_id56
a


Adobe Flash Player 10
File View Control Help
Available State Machines
diagnose_sg_id12
diagnose_sg_id60
diagnose_sg_id5F
diagnose_sg_id27
diagnose_sg_id29
diagnose_sg_id56
diagnose_sg_id1
diagnose_sg_id67
diagnose_sg_id36
diagnose_sg_id16
Diagnose_Funktional
diagnose_sg_id40
diagnose_sg_id0
diagnose_sg_id78
diagnose_sg_id72
diagnose_sg_id62
diagnose_sg_id63
diagnose_sg_id73
diagnose_sg_id64
diagnose_sg_id44
diagnose_sg_id50
diagnose_sg_id6D
diagnose_sg_id55
diagnose_sg_id4B
diagnose_sg_id3F
diagnose_sg_id47
diagnose_sg_id5B
diagnose_sg_idA0
diagnose_sg_id18
diagnose_sg_id41
diagnose_sg_id71
diagnose_sg_id37
diagnose_sg_id3C
Visualization View
diagnose_sg_id12 (9 state|s)
ERROR ! TimeOut sg_id12
WARN ! Functional_7F_Response_
OK ! Pos Antwort (FC
OK ! Anfrage FirstFrame
OK ! Pos Antwort (SF
OK ! Ueberwachung_aktiv
OK ! Pos Antwort (FF
OK ! Anfrage SingleFrame
INIT ! Functional_Request_Teste
00:06:20,000.000
00:06:30,000.000
00:06:40,000.000
00:06:50,000.000
00:07:00,000.000
00:07:10,000.000
diagnose_sg_id60 (9 state|s)
WARN ! Neg Antwort - Grund 21
WARN ! Neg Antwort - Grund 78_Erst
OK ! Pos Antwort (FC)
OK ! Anfrage FirstFrame
OK ! Pos Antwort (SF)
OK ! Ueberwachung_aktiv
OK ! Pos Antwort (FF)
OK ! Anfrage SingleFrame
INIT ! Functional_Request_Teste
Hey Günther,
I think I know the reason for this.
Please contact me so we can talk about it.
Helmut
Diagnose_Funktional (2 state|s)
OK ! Functional_Request_Active
OK ! Ueberwachung_aktiv
360
240
120
0
00:00:00,000.000
00:00:46,372.676
00:01:32.745.352
00:02:19,118.028
00:03:05,490.704
00:03:51,863.380
00:04:38,236.056
00:05:24,608.732
00:06:10,981.408
00:06:57,354.084
a
b

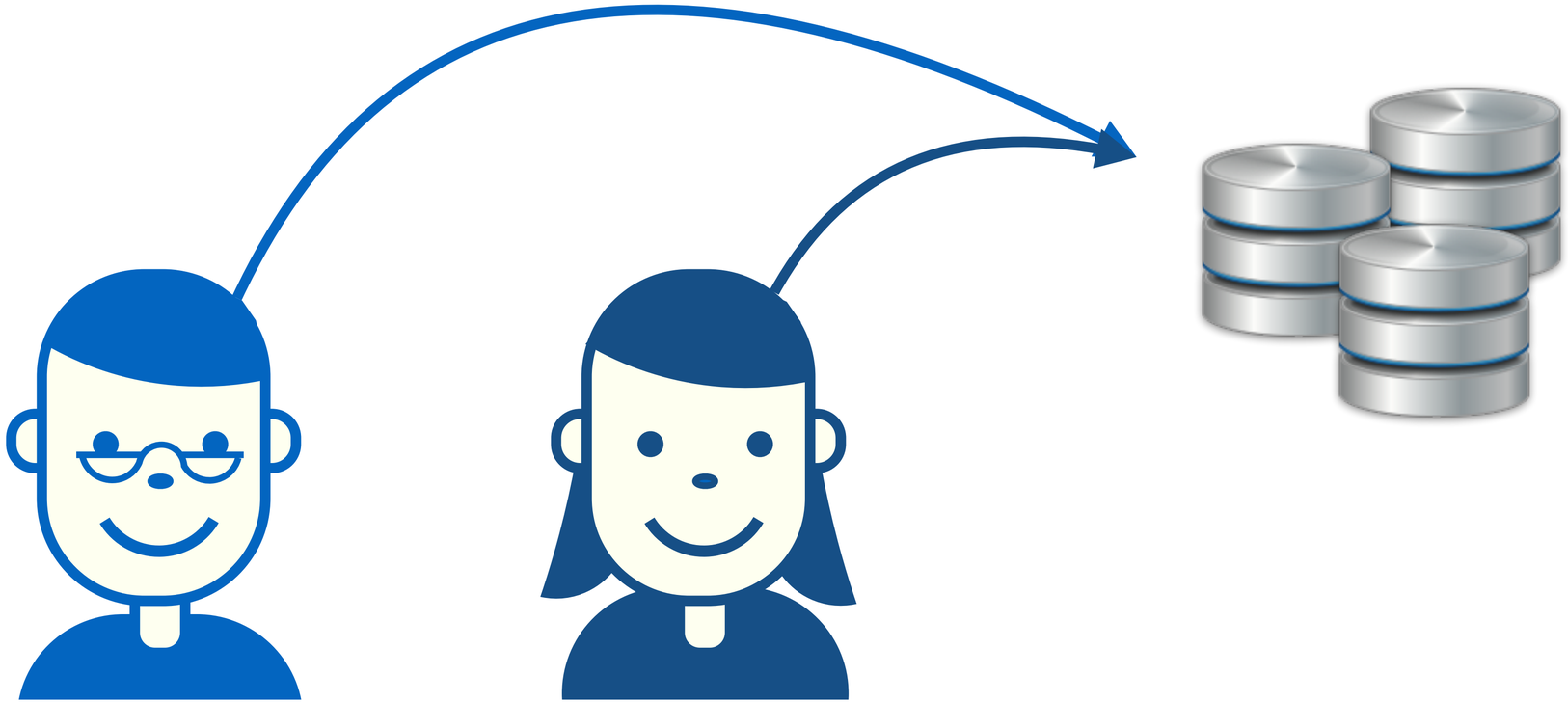

Available State Machines
diagnose_sg_id12
diagnose_sg_id60
diagnose_sg_id5F
diagnose_sg_id27
diagnose_sg_id29
diagnose_sg_id56
diagnose_sg_id1
diagnose_sg_id67
diagnose_sg_id36
diagnose_sg_id16
Diagnose_Funktional
diagnose_sg_id40
diagnose_sg_id0
diagnose_sg_id78
diagnose_sg_id72
diagnose_sg_id62
diagnose_sg_id63
diagnose_sg_id73
diagnose_sg_id64
diagnose_sg_id44
diagnose_sg_id50
diagnose_sg_id6D
diagnose_sg_id55
diagnose_sg_id4B
diagnose_sg_id3F
diagnose_sg_id47
diagnose_sg_id5B
diagnose_sg_idA0
diagnose_sg_id18
diagnose_sg_id41
diagnose_sg_id71
diagnose_sg_id37
diagnose_sg_id3C
Visualization View
diagnose_sg_id12 (9 state(s))
ERROR ! TimeOut sg_id12
WARN ! Functional_7F_Response_
OK ! Pos Antwort (FC
OK ! Anfrage FirstFrame
OK ! Pos Antwort (SF
OK ! Ueberwachung_akti
OK ! Pos Antwort (FF
OK ! Anfrage SingleFram
INIT ! Functional_Request_Teste
00:06:48,000.000
00:06:52,000.000
00:06:56,000.000
00:07:00,000.000
00:07:04,000.000
00:07:08,000.000
00:07:12,000.000
diagnose_sg_id5F (6 state(s))
WARN ! Neg Antwort - Grund 12
OK ! Pos Antwort (SF)
OK ! Ueberwachung_aktiv
OK ! Pos Antwort (FF)
OK ! Anfrage SingleFrame
INIT ! Functional_Request_Tester
00:06:48,000.000
00:06:52,000.000
00:06:56,000.000
00:07:00,000.000
00:07:04,000.000
00:07:08,000.000
00:07:12,000.000
diagnose_sg_id27 (6 state(s))
WARN ! Neg Antwort - Grund 78_Erst
OK ! Pos Antwort (FF)
OK ! Ueberwachung_aktiv
OK ! Pos Antwort (SF)
OK ! Anfrage SingleFrame
INIT ! Functional_Request_Teste
00:06:48,000.000
00:06:52,000.000
00:06:56,000.000
00:07:00,000.000
00:07:04,000.000
00:07:08,000.000
00:07:12,000.000
State Changes: 44
Intervall:
from: 00:03:57,659.964
to: 00:04:09,253.133
360
240
120
0
00:00:00,000.000
00:00:57,965.845
00:01:55,931.690
00:02:53,897.535
00:03:51,863.380
00:04:49,829.225
00:05:47,795.070
00:06:45,760.915
diagnose_sg_id12
diagnose_sg_id60
diagnose_sg_id5F
diagnose_sg_id27
diagnose_sg_id29
diagnose_sg_id56
a
WARN ! Neg Antwort - Grund 21
RN ! Neg Antwort - Grund 78_Erst
OK ! Pos Antwort (FC)
OK ! Anfrage FirstFrame
OK ! Pos Antwort (SF)
OK ! Ueberwachung_aktiv
OK ! Pos Antwort (FF)
OK ! Anfrage SingleFrame
INIT ! Functional_Request_Teste
00:06:20,000.000
b
Hey Günther,
I think I know
Please contact
Helmut
10,981.408
00:06:57,
c

Longitudinal field study

- 1 year, 15 engineers

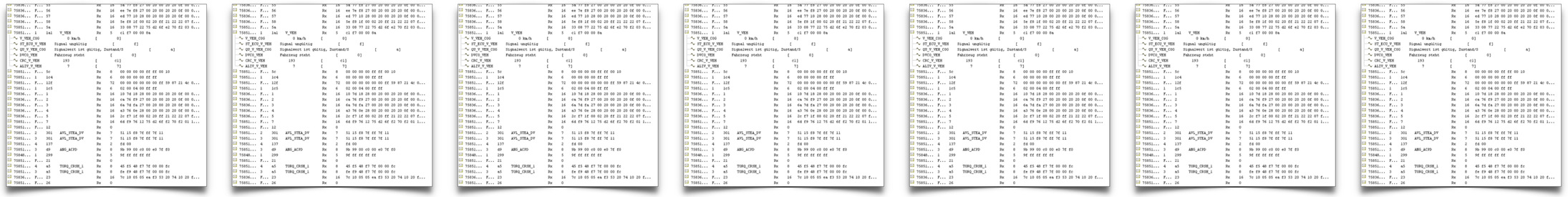
Results

- externalization and sharing of expert knowledge



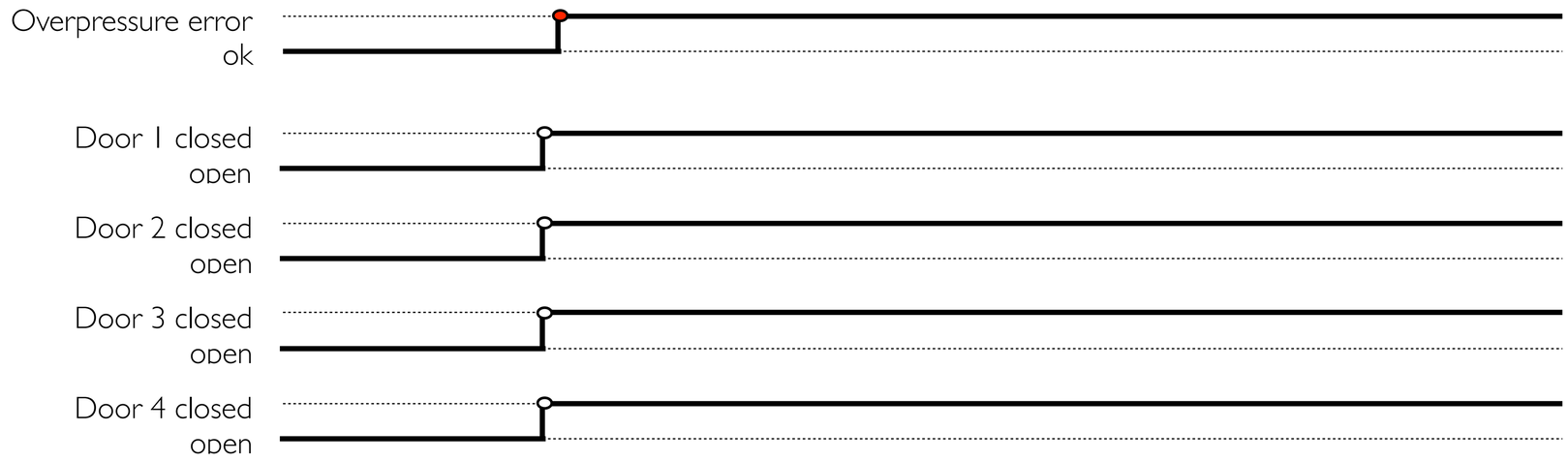
Results

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



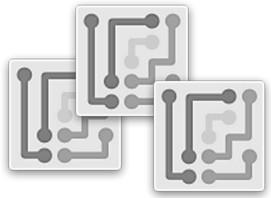
Results

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



http://www.automotiveaddicts.com/5276/bmw-x1-officially-released-images-first-drive-videos/bmw-x1-top-doors-open

Example of:
human-centered data science

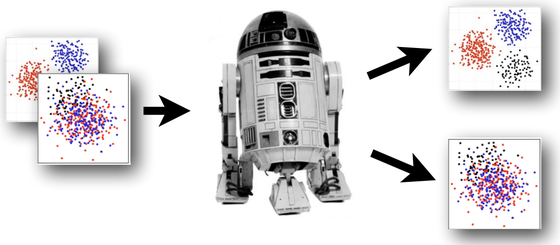


deployed
and
adopted

Summary

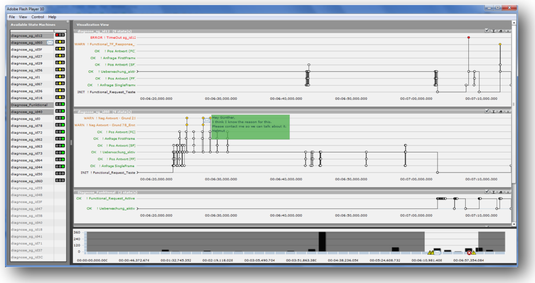
Summary: How to visualize data?

- Leverage perception

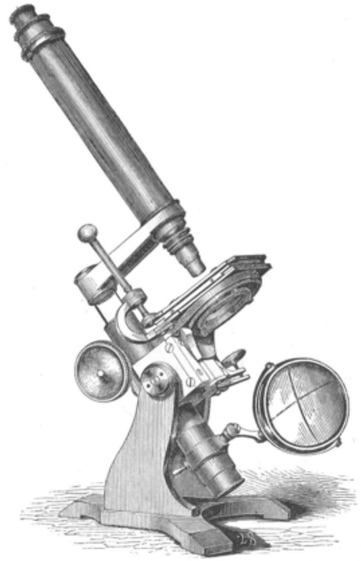


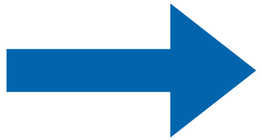
- Make it interactive

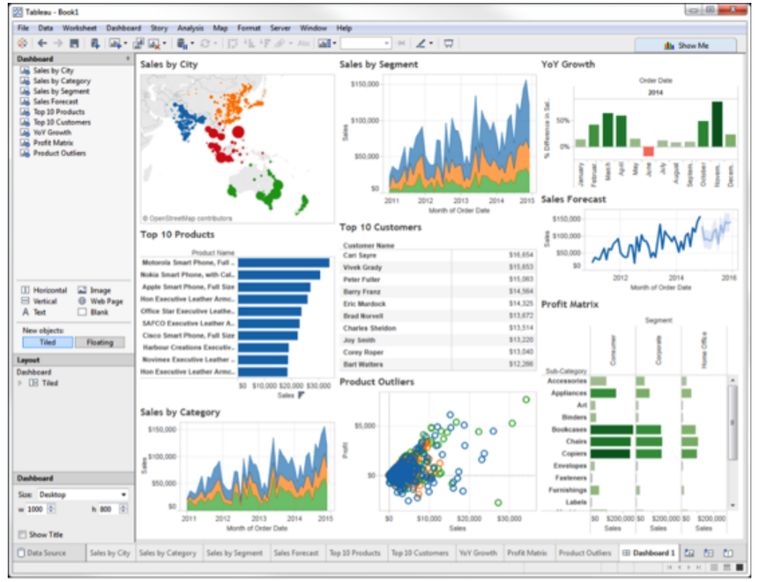
- Combine computation and visualization



Visualization: The Human Lens to Data







Tableau

Thank you!

Dr. Michael Sedlmair
michael.sedlmair@univie.ac.at

slides: http://tinyurl.com/siat2017

universität wien