About **Visualization in Bergen**.no and **Interactive Visual Analysis**

Helwig Hauser
University of Bergen

In the Following

1. Briefly about visualization in Bergen, Norway
2. Interactive Visual Analysis (IVA)
3. High-dimensional Data IVA

ii.UiB.no/vis

HH: prof. in visualization (vis)
@ Dept. of Informatics (ii)
@ Univ. of Bergen (UiB)
in Bergen, Norway (.no)

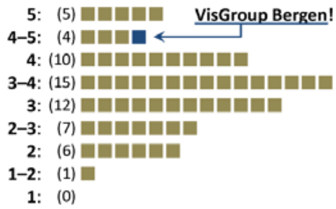
[images: Wikipedia]

UiB VisGroup

- 2007: group of 3: 
- 2009: larger projects start
- 2011: EuroVis in Bergen

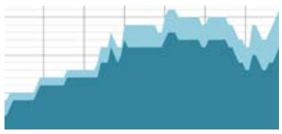
- 2013: new prof.: 

SemSeg · ILLUSTRASOUND · GEOILLUSTRATOR · PHYSIOILLUSTRATION



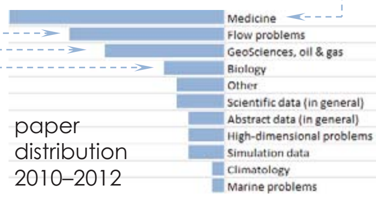
[ranking from NFR's 10-year evaluation in 2011/2012]

ii.UiB.no/vis Research



➢**Application-oriented basic research** in visualization:

1. **Researched visualization methodology** (how to visualize)
   - ➢ **Interactive Visual Analysis, *n*D data** (H. Hauser *et al.*)
   - ➢ **Visual Knowledge Discovery, 3D data** (St. Bruckner *et al.*)
   - ➢ **Illustrative Visualization** (I. Viola *et al.*)
2. **Applications at which this research is oriented** (for whom)
   - ILLUSTRASOUND ➢ **Medical Visualization** (partner in MedViz Bergen, *etc.*)
   - GEOILLUSTRATOR ➢ **GeoSciences / Oil & Gas** (*e.g.*, financed by Statoil's Akademiaavtale)
   - PHYSIOILLUSTRATION ➢ **Biology / Bioinformatics** (with CBU@ii *et al.*)
   - SemSeg ➢ **Fluid Dynamics** (in collab. with FFI.no, for ex.)
   - ➢ **Engineering** (visual analysis of simulation data)



paper distribution 2010–2012

ii.UiB.no/vis Team

Two profs. (HH, StBr) and PostDocs, PhD studs., *et al.*

2007 2008 2009 2010 2011 2012 2013 2014

HH ---

➢ **Stefan Bruckner**

-- **Ivan Viola** -- - - - - - - - - - -

➢ **Július Parulek** -------------------------------------

-- Daniel Patel ---------------

-- Jean-Paul Balabanian --------

➢ Ove Daae Lampe -----------------------

➢ Johannes Kehrer -----------------

➢ Paolo Angelelli ---------------------------- -----------------------

➢ Åsmund Birkeland ---------------------------

➢ Armin Pobitzer -------------------

➢ Veronika Šoltézová ------------- - - - - -

➢ Endre Lidal -------------------------

➢ Çağatay Turkay ------------------------

➢ Andrea Brambilla ------------------------------( )

➢ Mattia Natali ---------------------------------( )

➢➢ Ivan, Andreas --

ii.UiB.no/vis PhDs (10 so far)


Daniel Patel (Oct. 2009):
Expressive Vis. & Rapid Interpr. of Seismic Volumes


Jean-Paul Balabanian (Jan. 2010):
Multi-Aspect Vis.: from Linked to Integrated Views


Johannes Kehrer (May 2011):
IVA of Multi-faceted Scientific Data


Ove Daae Lampe (Nov. 2011):
IVA of Process Data


Armin Pobitzer (June 2012):
IVA of Time-dependent Flows


Paolo Angelelli (June 2012):
Visual Expl. of Human Physiology

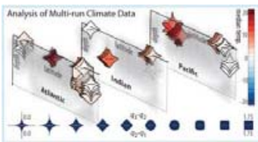

Veronika Šoltészová (Oct. 2012):
Perception-Augmenting Illumination

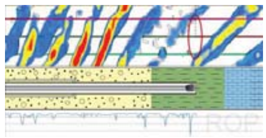
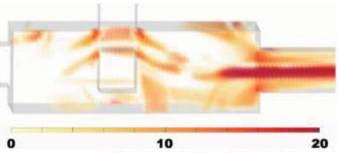

Åsmund Birkeland (May 2013):
Ultrasonic Vessel Vis.: From Extraction to Perception

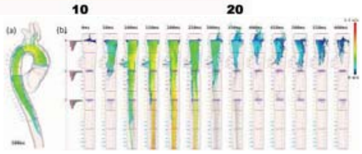

Endre Lidal (May 2013):
Sketch-based Storytelling for Cognitive Problem Solving

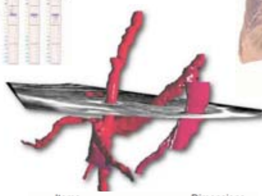
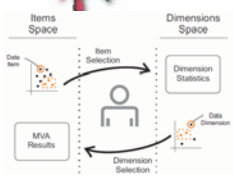

Çağatay Turkay (Nov. 2013):
Interactive Visual Analysis of High-dimensional Data

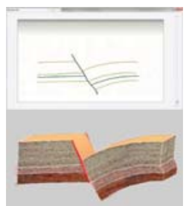






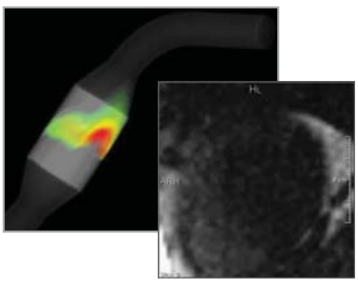


Interactive Visual Analysis (IVA)

- Given data –
  too much and/or *too complex* to be shown at once:
- IVA is an **interactive visualization approach** to facilitate
  - the **exploration** and/or the **analysis** of data (not necessarily the presentation of data), including
    - **hypothesis generation & evaluation, sense making**,
    - **knowledge crystallization**, *etc.*
  - according to the **user's interest/task**, *for ex.*, by interactive feature extraction,
  - navigating between **overview** and **details**, *e.g.*, to enable interactive information drill-down [Shneiderman]
- through an **iterative & interactive visual dialog**

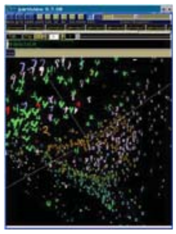
Interactive Visual Analysis ↔ Visual Analytics

- **IVA** ("interactive visual analysis") **since 2000**
- **Tightly related** to **visual analytics**, of course, *e.g.*, *integrating computational & interactive data analysis*
- A **particular methodology** with specific components (*CMV*, *linking & brushing*, *F+C vis.*, *etc.*)
- General enough to work in **many application fields**, but not primarily the VA fields (national security, *etc.*), in particular "**scientific data**" **fields**…

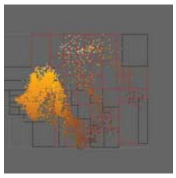


Integrating Interaction & Computation

- **Goal**: to combine the *best of two worlds* [Keim et al.]:
  - data **exploration**/analysis by the **user**, based on **interactive visualization**
  - and **data analysis** by the **computer**, based on **statistics**, **machine learning**, *etc.*
- State of the art / **levels of integration**:
  - **mostly no integration**, still
  - some **vis. of results** of computations
  - also: making **comp. semi-interactive** (here called "**inner integration**")
  - **rare: tight integration**
- **Outer integration** (here!): bundling **interaction & computation in a loop**



[Maniyar & Nabney, 2006]



[Williams & Munzner, 2004]

Target Data Model: "Scientific Data"

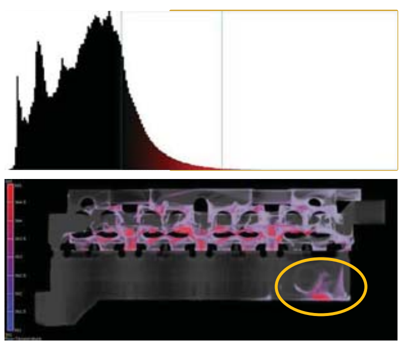
- **Characterized** by a combination of
  - **independent variables**, like **space** and/or **time** (cf. **domain**)
  - and **dependent variables**, like **pressure**, **temp.**, etc. (cf. **range**)
- So we can think of this type of data as **given as d(x)** with **x** ↔ **domain** and **d** ↔ **range** – examples:
  - **CT data** $d(\mathbf{x})$ with $\mathbf{x} \in R^3$ and $d \in R$
  - **unstead 2D flow** $\mathbf{v}(\mathbf{x},t)$ with $\mathbf{x} \in R^2$, $t \in R$, and $\mathbf{v} \in R^2$
  - **num. sim. result** $\mathbf{d}(\mathbf{x},t)$ with $\mathbf{x} \in R^3$, $t \in R$, and $\mathbf{d} \in R^n$
  - **system sim.** $\mathbf{q}(\mathbf{p})$ with $\mathbf{p} \in R^n$ and $\mathbf{q} \in R^m$
- **Common property**:
  - **d** is (at least to a certain degree) **continuous** wrt. **x**

Interactive Visual Analysis of Scientific Data

- **Interactive visual analysis** (as exemplified in this tutorial) **works really well with scientific data**, *e.g.*,
  - results from **numerical simulation** (spatiotemporal)
  - imaging / **measurements** (in particular multivariate)
  - sampled **models**
- When used to study scientific data, **IVA employs**
  - methods from **scientific visualization** (vol. rend., …)
  - methods from **statistical graphics** (scatterplots, …), **information visualization** (parallel coords., *etc.*)
  - **computational tools** (statistics, machine learning, …)
- Applications include
  - **engineering**, **medicine**, **meteorology/climatology**, **biology**, *etc.*

The Iterative Process of IVA

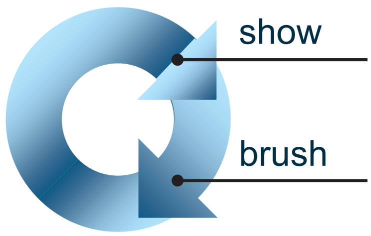
- Loop / bundling of *two complementary parts*:
  - **visualization** – show to the user!
    Something new, or something due to interaction.
  - **interaction** – tell the computer!
    What is interesting? What to show next?
- Basic example (**show – brush** – show – …), cooling jacket context:
  1. show a histogram of temperatures
  2. brush high temperatures (>90°[±2°])
  3. show focus+context vis. in 3D
  4. locate relevant feature(s)



- **KISS-principle IVA**:
  - linking & brushing, focus+context visualization, …

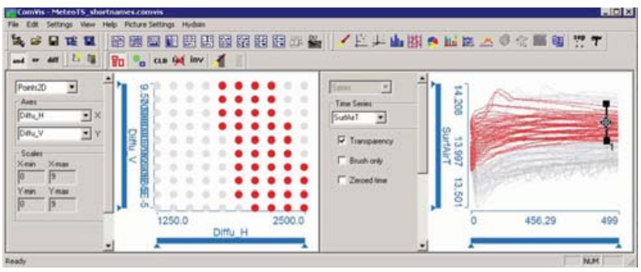
Show & Brush (basic IVA)

- **Tightest IVA loop**
  - **show data** (explicitly represented information)
  - **one brush** (on one view, can work on >1 dims.)



A typical (start into an) **IVA session** of this kind:

- bring up multiple views
  - at least one for **x**, *t*
  - at least one for d_i
- I see (something)!
- brush this "something"
- linked F+C visualization
- first insight!



Show & Brush (basic IVA)

- **Tightest IVA loop**
  - **show data** (explicitly represented information)
  - **one brush** (on one view, can work on >1 dims.)
- **Requires**:
  - multiple views (≥2)
  - interactive brushing capabilities on views (brushes should be editable)
  - focus+context visualization
  - linking between views

… leads to … degree of interest *… requires …* *… is realized via …*

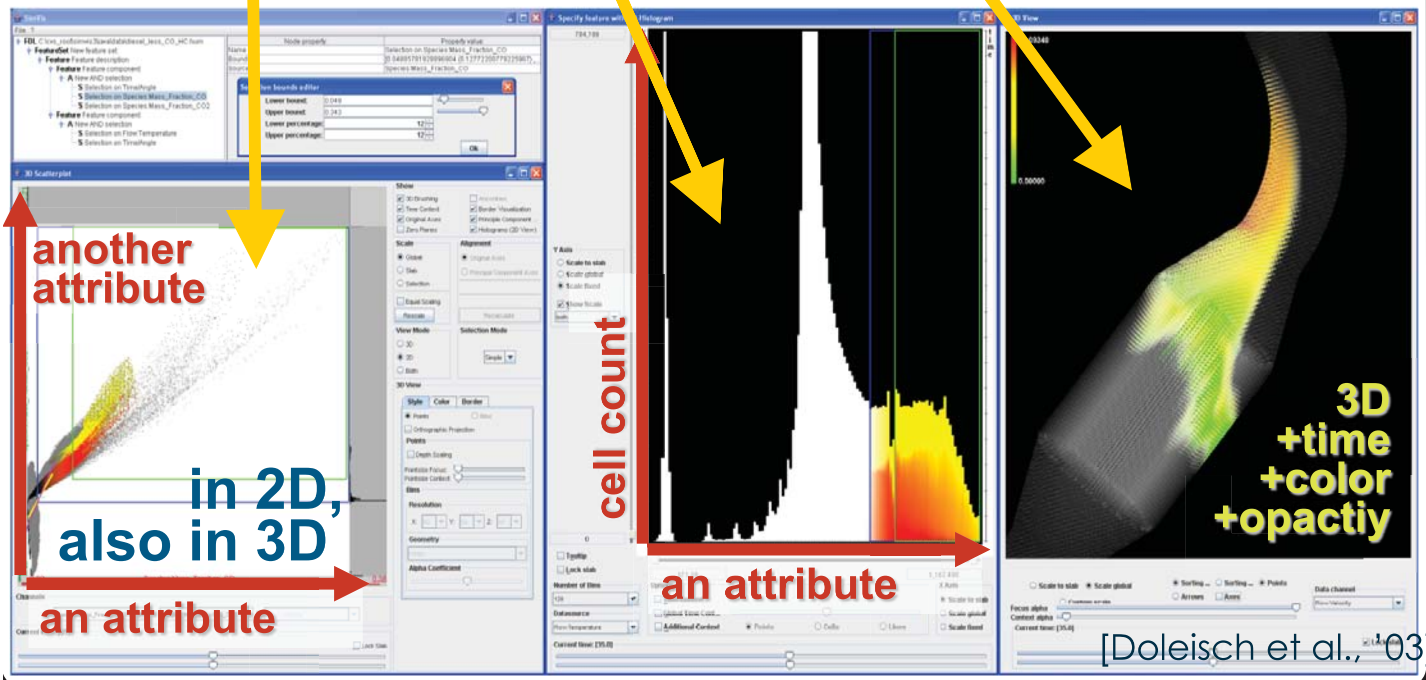
- Allows for **different IVA patterns** (wrt. domain & range)

A typical (start into an) **IVA session** of this kind:

- bring up multiple views
  - at least one for **x**, *t*
  - at least one for d_i
- I see (something)!
- brush this "something"
- linked F+C visualization
- first insight!

IVA: Multiple Views

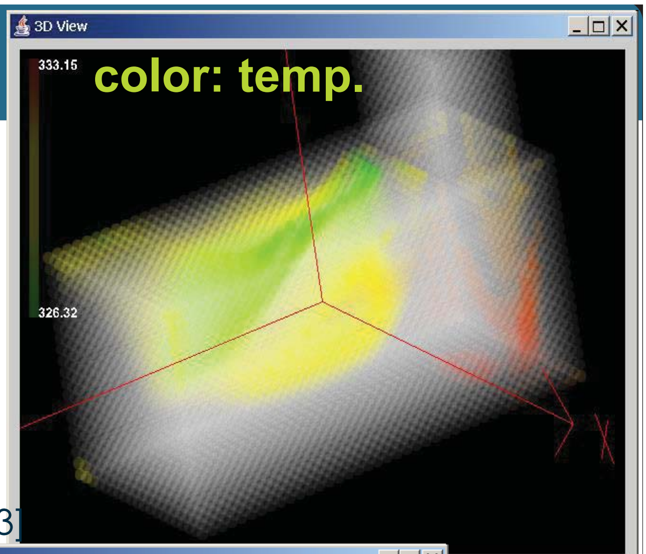
- One dataset, but multiple views
- Scatterplots, histogram, 3D(4D) view, etc.



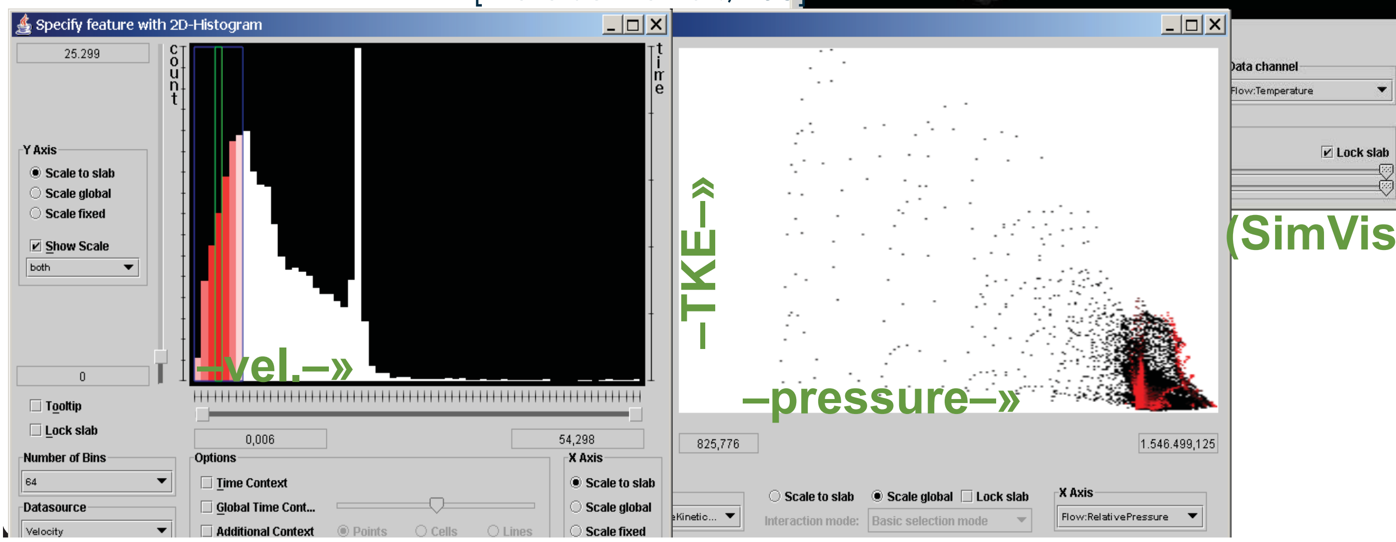
[Doleisch et al., '03]

Interactive Brushing

- Move/alter/extend brush interactively
- Interactively explore/ analyze multiple variates



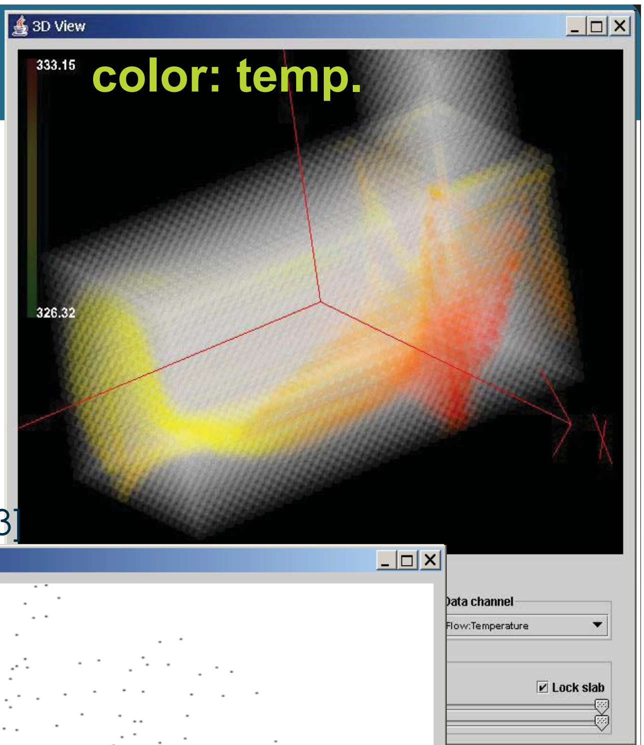
[Doleisch et al., '03]



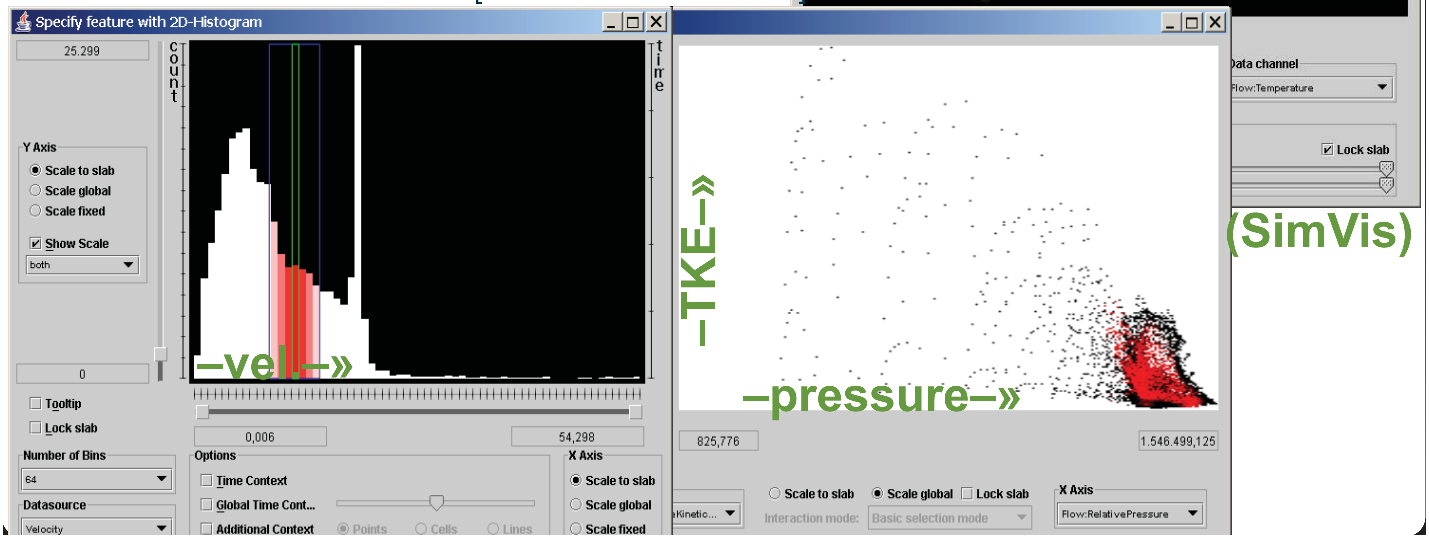
(SimVis)

Interactive Brushing

- Move/alter/extend brush interactively
- Interactively explore/ analyze multiple variates

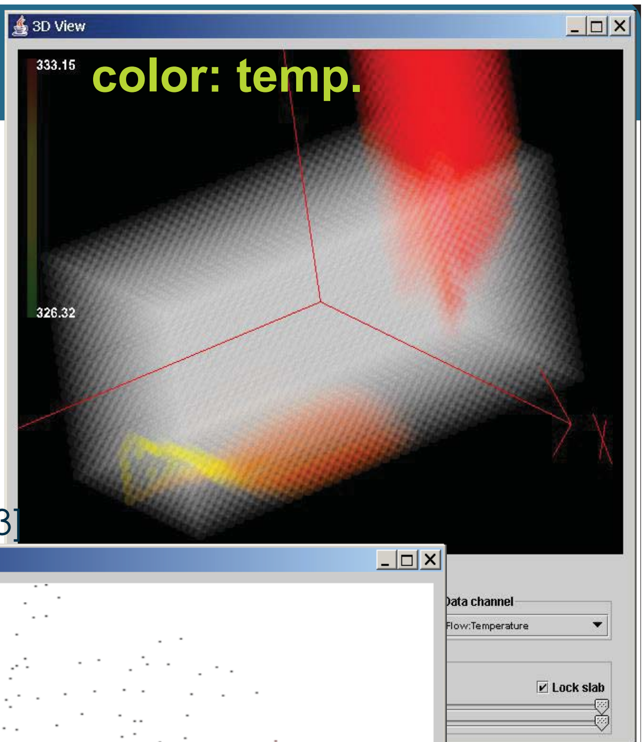


[Doleisch et al., '03]

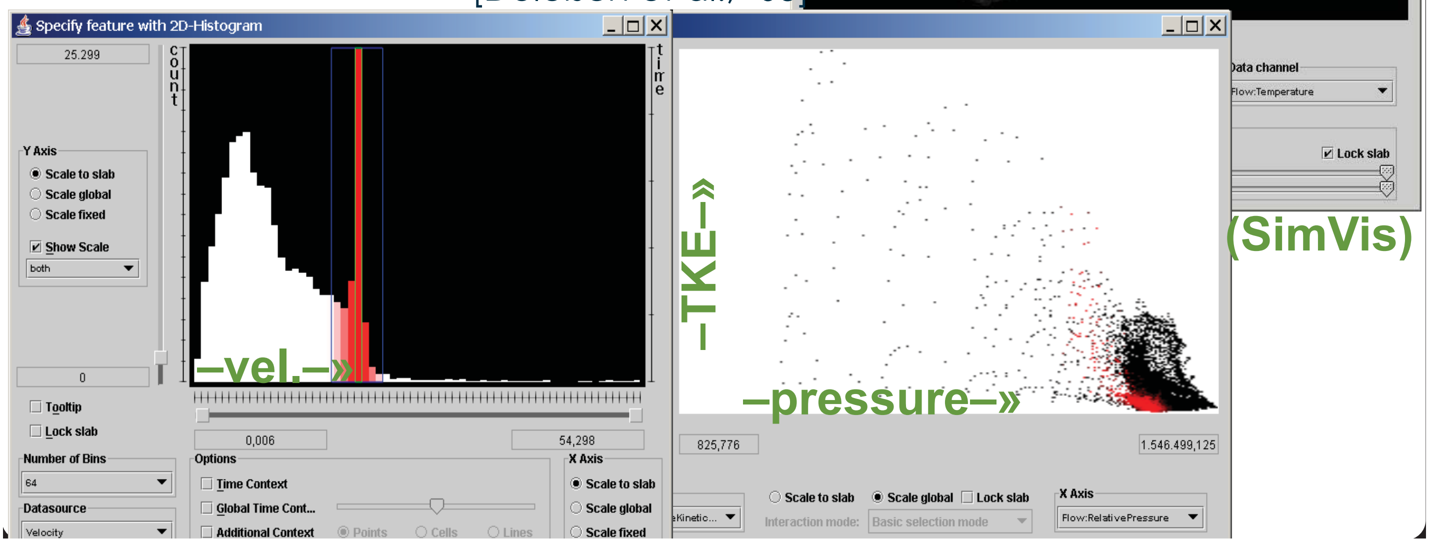


Interactive Brushing

- Move/alter/extend brush interactively
- Interactively explore/ analyze multiple variates



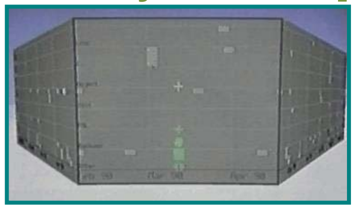
[Doleisch et al., '03]



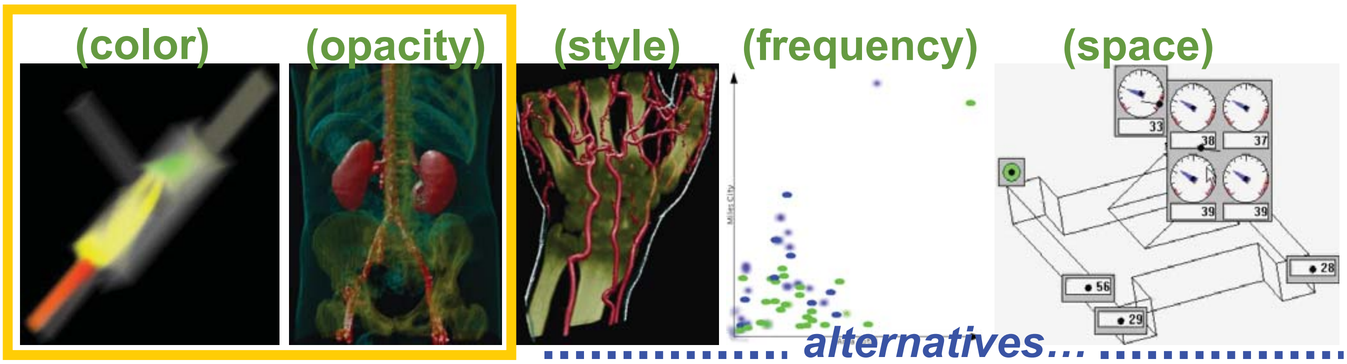
IVA: Focus+Context Visualization

- Traditionally space distortion
  - more space for data of interest
  - rest as context for orientation
- Generalized F+C visualization
  - emphasize data in focus (color,opacity, …)
  - differentiated use of visualization resources

[Mackinlay et al. 1991]



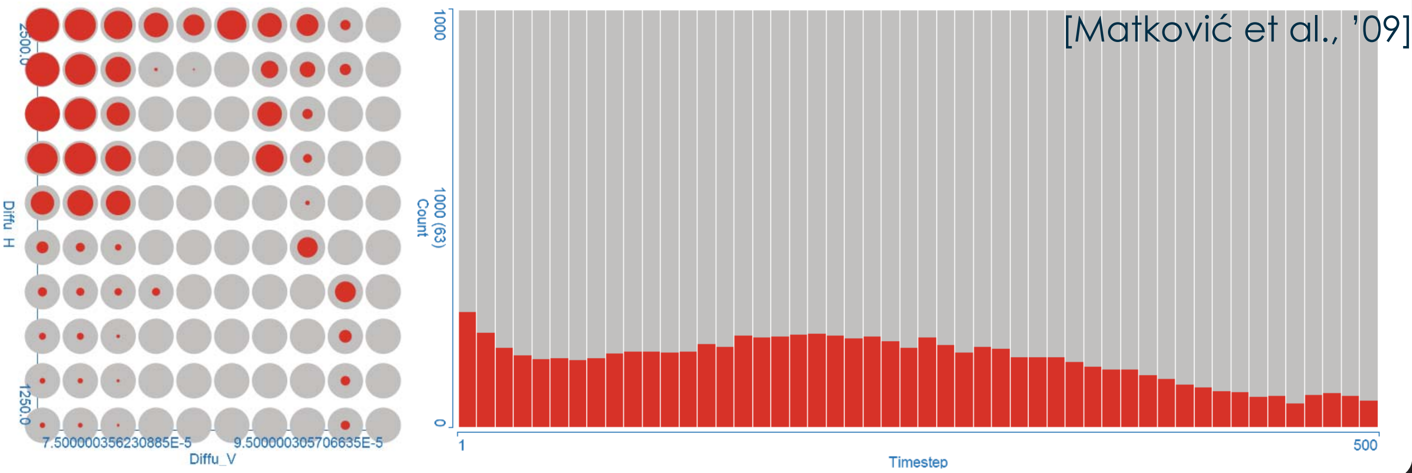
[Hauser… 2001, 2003]



F+C Visualization in IVA Views

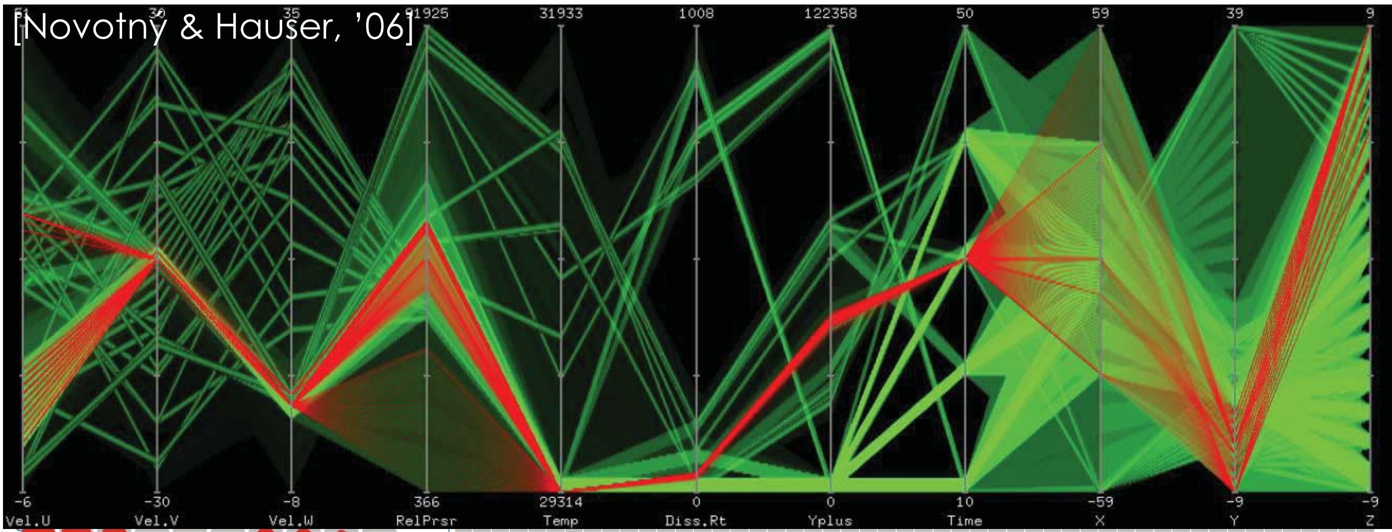
- Colored vs. gray-scale visualization
- Opaque vs. semi-transparent visualization

In a scatterplot (left) or histogram (right): brushed data in red

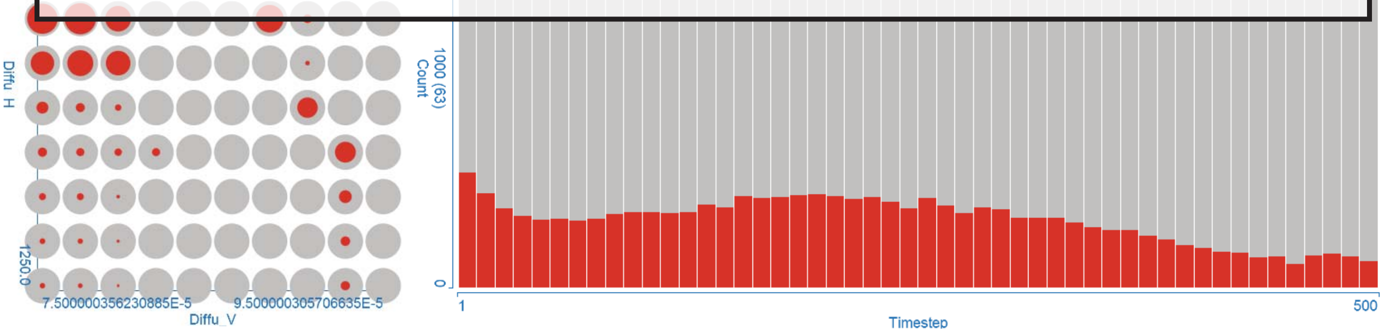


F+C Visualization in IVA Views

[Novotný & Hauser, '06]

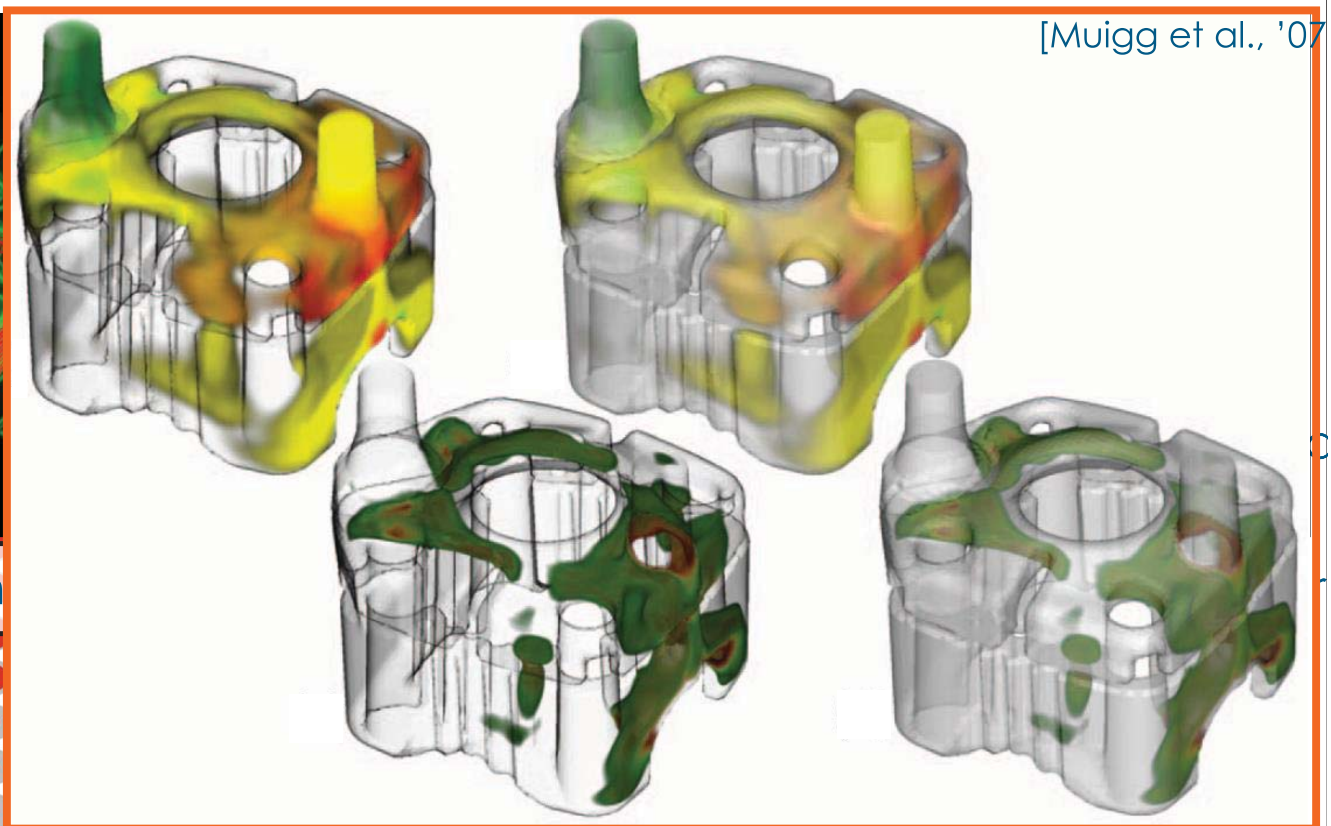


In parallel coordinates (above): brushed data in red & over



F+C Visualization in IVA Views

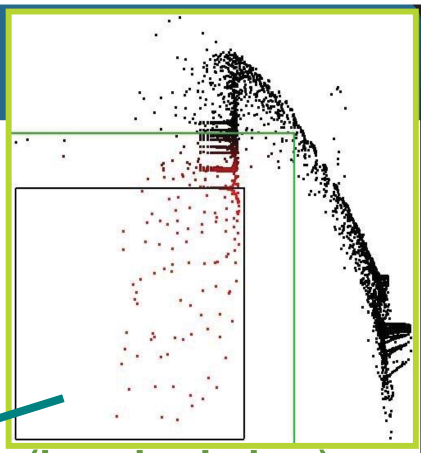
[Muigg et al., '07



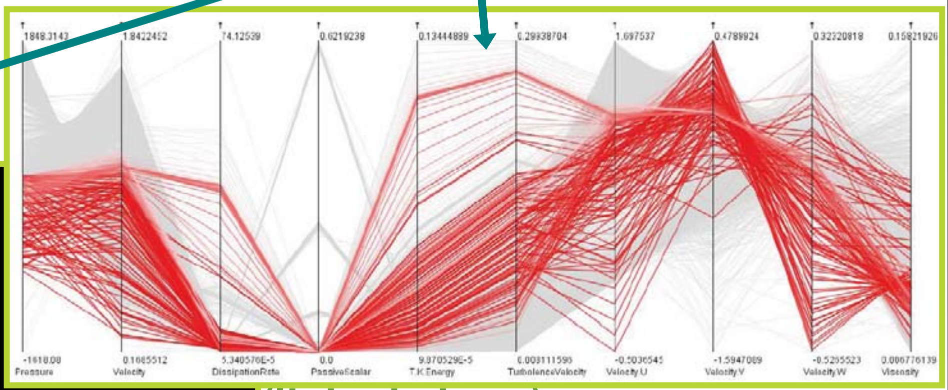
In 3D (above): less transp. & colored, in illustrative context

IVA: Linked Views

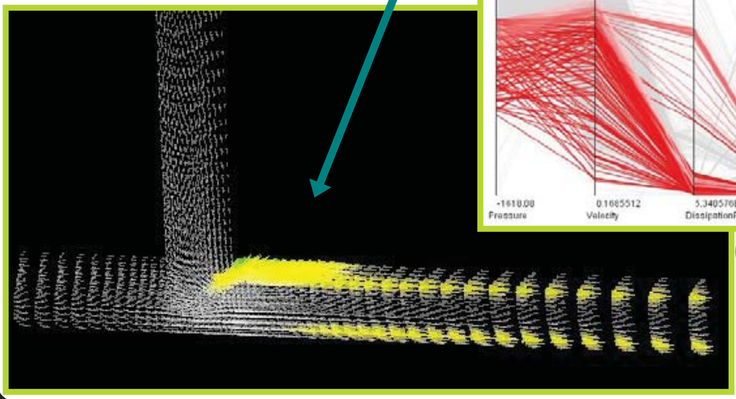
- Brushing: mark data subset as especially intersting
- Linking: enhance brushed data in linked views consistently (F+C)



(brushed view)



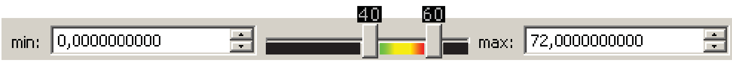
(linked views)



[Doleisch & Hauser, '02

IVA: Degree of Interest (DOI)

- *doi*(.): data items tr_i (table rows) → degree of interest $doi(tr_i) \in [0,1]$
  - $doi(tr_i) = 0 \Rightarrow tr_i$ not interesting ($tr_i \in$ context)
  - $doi(tr_i) = 1 \Rightarrow tr_i$ 100% interesting ($tr_i \in$ focus)
- Specification
  - explicit, e.g., through direct selection
  - implicit, e.g., through a range slider



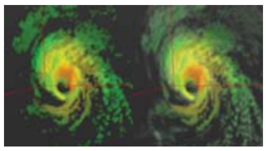
x	*y*	*d1*	*d2*	*doi*
0	0	17,20	-0,22	*0,00*
1	0	12,10	0,10	*0,00*
2	0	7,70	0,45	*0,00*
3	0	2,10	0,90	*0,00*
0	1	24,10	0,02	*0,00*
1	1	21,90	0,36	*0,00*
2	1	15,50	0,87	*0,74*
3	1	11,10	1,20	*1,00*
0	2	27,20	0,12	*0,00*
1	2	24,10	0,66	*0,18*
2	2	17,30	1,35	*1,00*
3	2	12,10	2,20	*0,60*
0	3	35,50	0,67	*0,00*
1	3	30,90	1,30	*0,00*
2	3	24,50	2,10	*0,10*
3	3	20,80	2,90	*0,00*

- Fractional DOI values: $0 \leq doi(tr_i) \leq 1$
  - several levels (0, low, med., …)
  - a continuous measure of interest
  - a probabilistic definition of interest

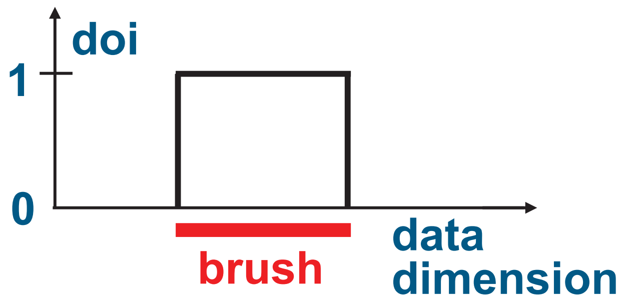
(cont'd on next slid

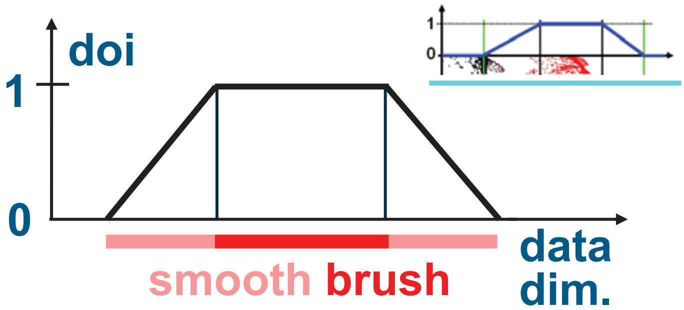
IVA: Smooth Brushing → Fractional DOI

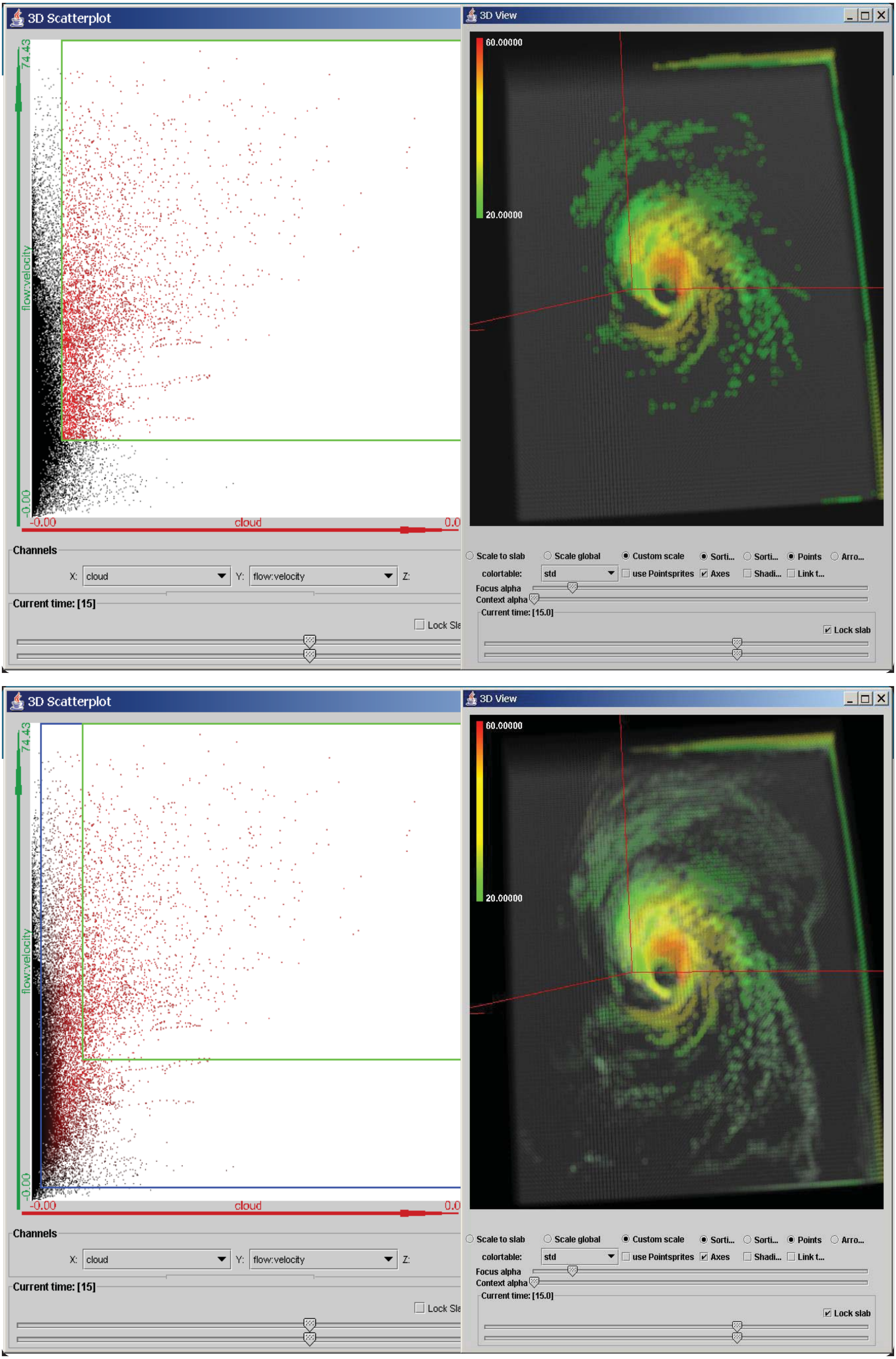
- **Fractional DOI values** esp. useful wrt. **scientific data**: (quasi-)continuous nature of data ↔ smooth borders
- Goes well with gradual focus+context vis. techniques (coloring, semitransparency)



- Specification: **smooth brushing** [Doleisch & Hauser, 2002]
  - "inner" range: all 100% interesting (DOI values of 1)
  - between "inner" & "outer" range: fractional DOI values
  - outside "outer" range: not interesting (DOI values of 0)

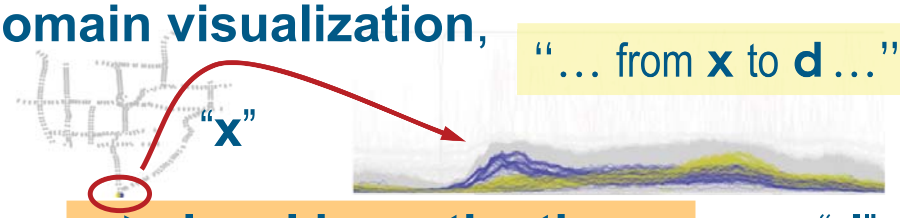





3D Scatterplot
74.43
flow:velocity
-0.00
-0.00
cloud
0.0
Channels
X: cloud
Y: flow:velocity
Z:
Current time: [15]
Lock Sla
3D View
60.00000
20.00000
Scale to slab
Scale global
Custom scale
Sorti...
Sorti...
Points
Arro...
colortable:
std
use Pointsprites
Axes
Shadi...
Link t...
Focus alpha
Context alpha
Current time: [15.0]
Lock slab
3D Scatterplot
74.43
flow:velocity
-0.00
-0.00
cloud
0.0
Channels
X: cloud
Y: flow:velocity
Z:
Current time: [15]
Lock Sla
3D View
60.00000
20.00000
Scale to slab
Scale global
Custom scale
Sorti...
Sorti...
Points
Arro...
colortable:
std
use Pointsprites
Axes
Shadi...
Link t...
Focus alpha
Context alpha
Current time: [15.0]
Lock slab

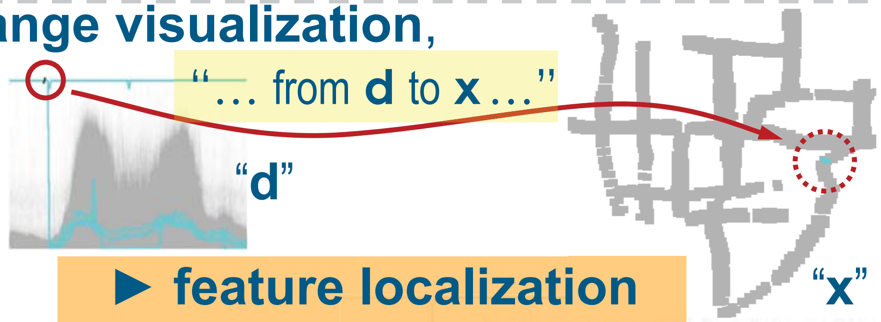
Three Patterns of SciData IVA

- Preliminary: domain **x** & range **d** visualized (≥2 views)
  - **brushing on domain visualization,** *e.g.*, brushing special locations in the map view



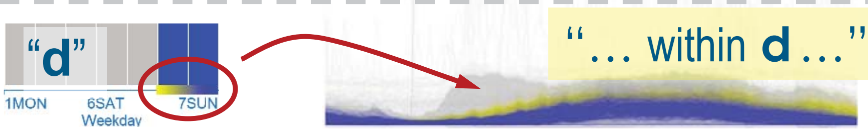
❶ ▶ **local investigation**

  - **brushing on range visualization,** *e.g.*, brushing outlier curves in a function graph view



❷ ▶ **feature localization**

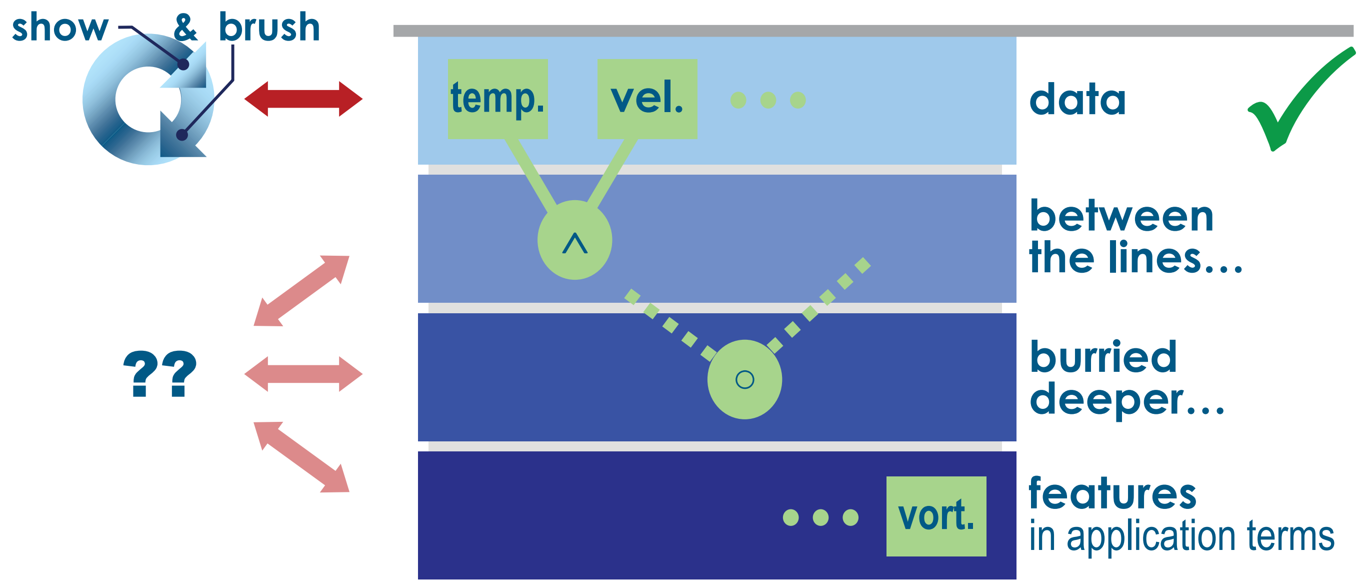
  relating multiple range variates



❸ ▶ **multi-variate analysis**

IVA – Levels of Complexity (1/4)

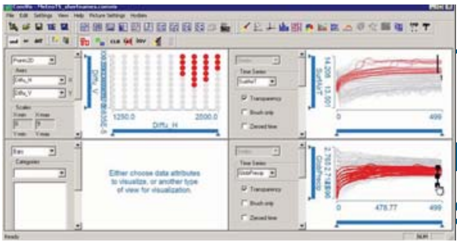
- A ***lot*** can be done with basic IVA, already! [pareto rule]
- We can consider a **layered information space**: from **explicitly** represented information (the **data**) to **implicitly** contained information, **features**, …



IVA – Levels of Complexity (2/4)

- A ***lot*** can be done with basic IVA, already! ✓ [pareto rule]
- **For more advanced** exploration/analysis **tasks**, we extend it (in seveal steps):
  - IVA, level 2: **logical combinations of brushes**, *e.g.*, utilizing the *feature definition language* [Doleisch et al., 2003]
  - IVA, l. 3: **attribute derivation**; **advanced brushing**, with interactive formula editor; *e.g.*, similarity brushing
  - IVA, l4: **application-specific feature extraction**, *e.g.*, based on vortex extraction methods for flow analysis
- Level 2: like **advanced verbal feature description**
  - ex.: "**hot** flow, also **slow**, near **boundary**" (cooling j.)
  - brushes comb. with **logical operators** (AND, OR, SUB)
  - in a **tree**, or **iteratively** ((((b_0 op$_1$ b_1) op$_2$ b_2) op$_3$ b_3) …)

IVA – Levels of Complexity (2/4)



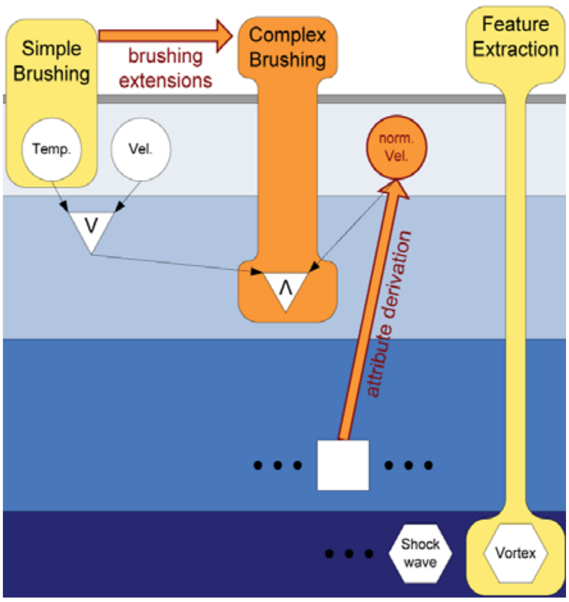
- A lot can be done with basic IVA, already! ✓ [pareto rule]
- **For more advanced** exploration/analysis **tasks**, we extend it (in seveal steps):
  - IVA, level 2: **logical combinations of brushes**, *e.g.*, utilizing the *feature definition language* [Doleisch et al., 2003]
  - IVA, l. 3: **attribute derivation**; **advanced brushing**, with interactive formula editor; *e.g.*, similarity brushing
  - IVA, l4: **application-specific feature extraction**, *e.g.*, based on vortex extraction methods for flow analysis
- Level 2: like **advanced verbal feature description**
  - ex.: "**hot** flow, also **slow**, near **boundary**" (cooling j.)
  - brushes comb. with **logical operators** (AND, OR, SUB)
  - in a **tree**, or **iteratively** ((((b_0 op$_1$ b_1) op$_2$ b_2) op$_3$ b_3) …)

IVA – Levels of Complexity (3/4)

- A *lot* can be done with basic IVA, already! [pareto rule]
- **For more advanced** exploration/analysis **tasks**, we extend it (in seveal steps):
  - IVA, level 2: **logical combinations of brushes**, *e.g.*, utilizing the *feature definition language* [Doleisch et al. 2003]
  - IVA, l. 3: **attribute derivation**; **advanced brushing**, with interactive formula editor; *e.g.*, similarity brushing
  - IVA, l4: **application-specific feature extraction**, *e.g.*, based on vortex extraction methods for flow analysis
- Level 3: using **general info extraction** mechanisms, two (partially complementary) approaches:
  1. **derive additional attribute**(s), then show & brush
  2. use an **advanced brush** to select “hidden” relations

IVA – Levels of Complexity (3/4)

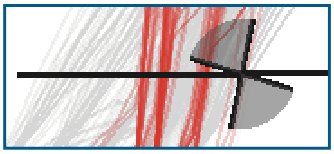
- A *lot* can be done with basic I
- **For more advanced** explorat we extend it (in seveal steps):
  - IVA, level 2: **logical combin** utilizing the *feature definitior*
  - IVA, l. 3: **attribute derivatio** with interactive formula editc
  - IVA, l4: **application-specifi** based on vortex extraction n



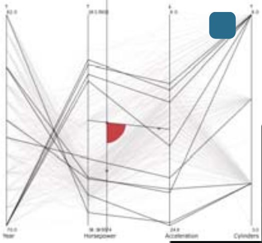
- Level 3: using **general info extraction** mechanisms, two (partially complementary) approaches:
  1. **derive additional attribute**(s), then show & brush
  2. use an **advanced brush** to select “hidden” relations

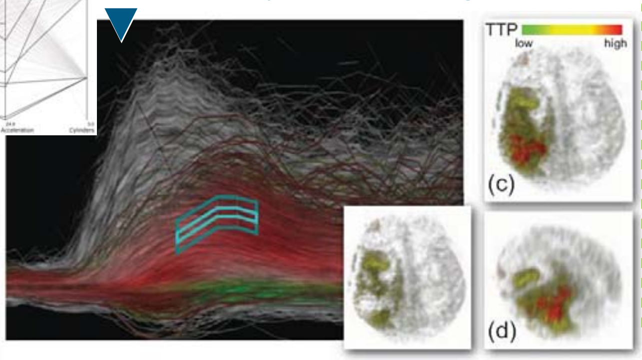
IVA (level 3): Advanced Brushing

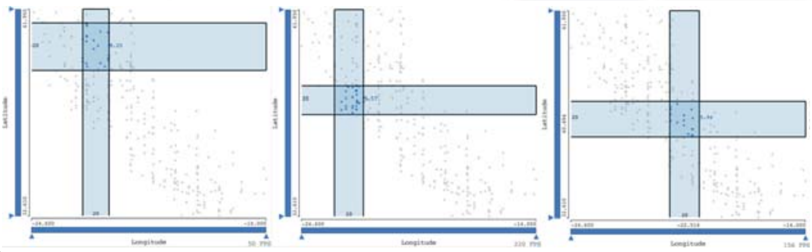
- **Std. brush**: brush 1:1 what you see
  Adv. brush: executes additional function ("intelligent"?)

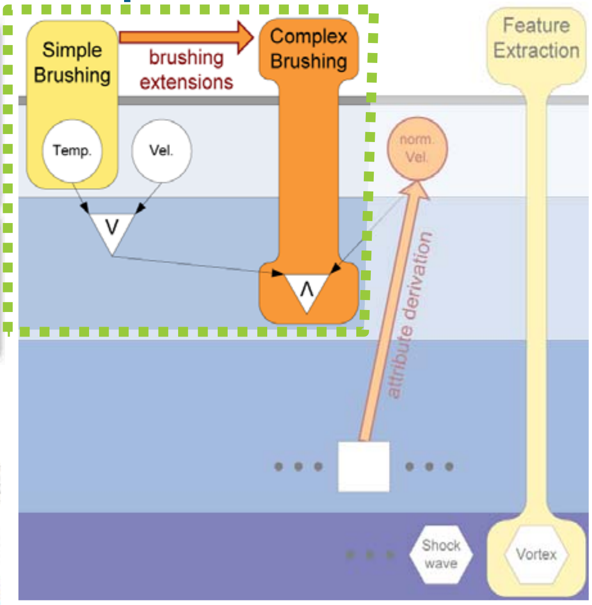


- **Examples**:
  - angular brushing [Hauser et al., 2002]
    similarity brushing [Muigg et al., 2008]



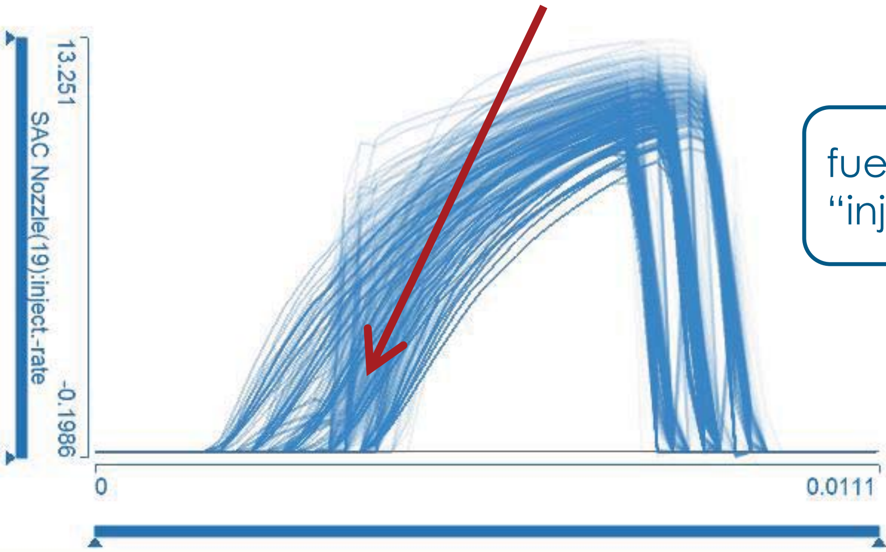






3rd level IVA, adv. brushing example

- Considering a visualization of a family of function graphs:
  - select the steeply rising graphs



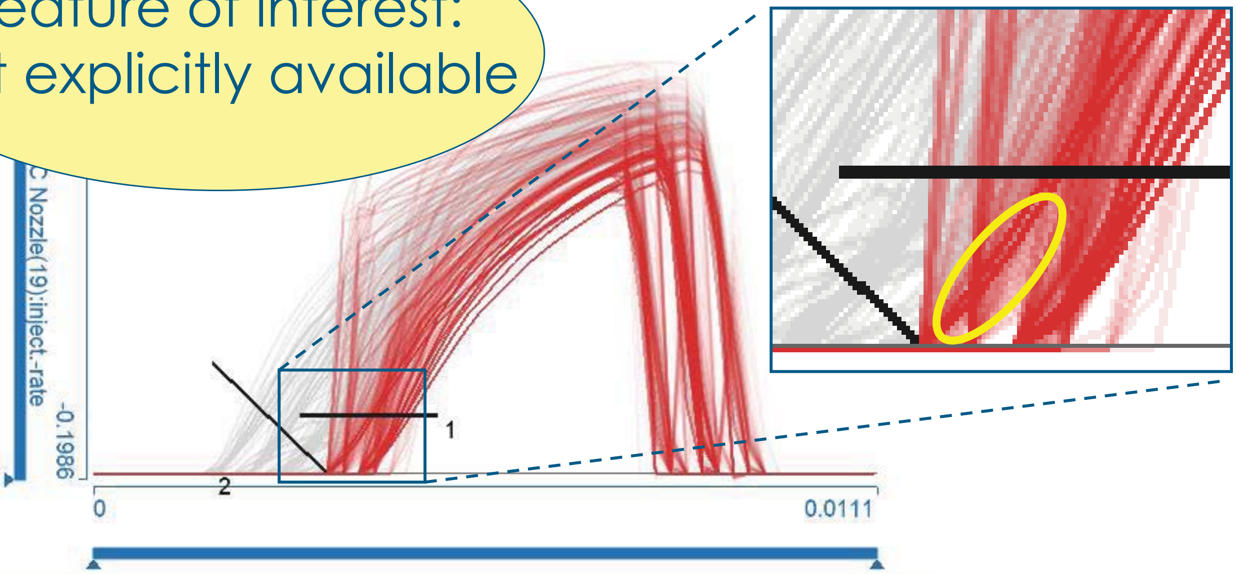
fuel injection simulation
"injection rate"

example prepared by Konyha, Zolt

3rd level IVA, adv. brushing example

- A simple line brush is not enough
- Combining line brushes does not work, either

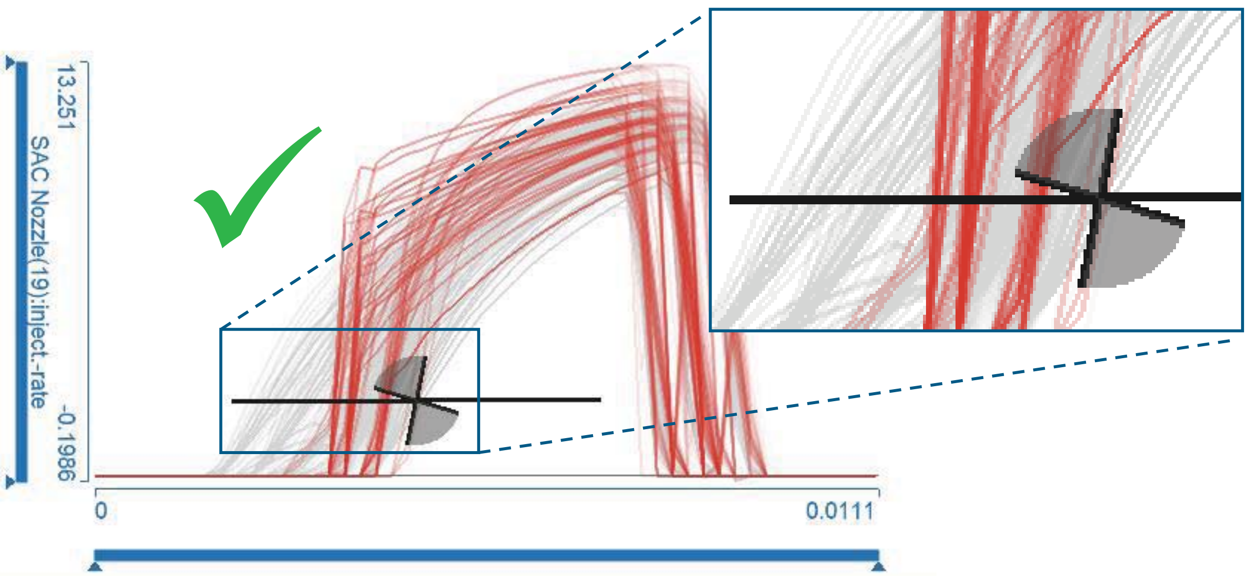
feature of interest:
not explicitly available



example prepared by Konyha, Zolt

3rd level IVA, adv. brushing example

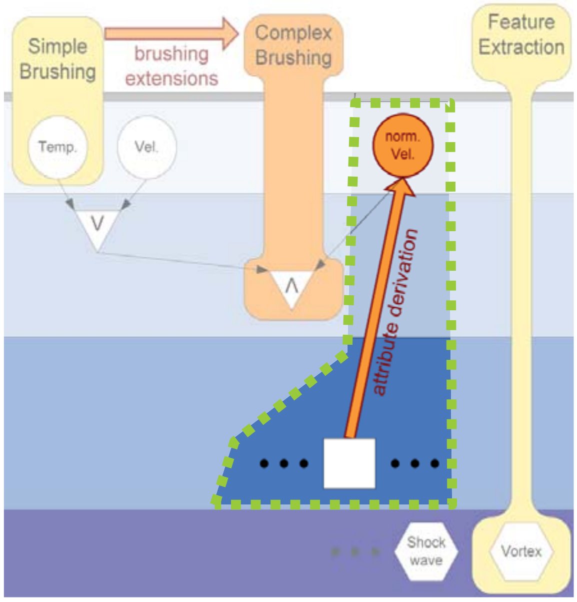
- The *angular line brush* (a specialized brush) selects the intended function graphs
  - that it intersects, and
  - the angle is in a given threshold



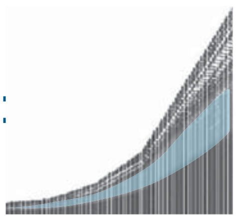
example prepared by Konyha, Zolt

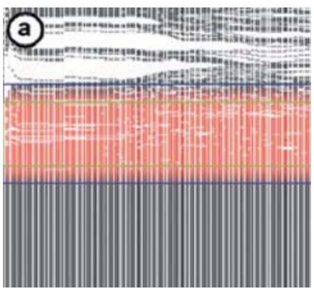
IVA (level 3): Attribute Derivation

- **Principle** (in the context of iterative IVA):
  - **see** some data feature Φ of interest in a visualization
  - **identify** a **mechanism** T **to describe** Φ
  - **execute** (interactively!) an **attribute derivation** step to represent Φ explicitly (as new, synthetic attribute[s] d_φ)
  - **brush** d_φ to get Φ
- **Tools** T to describe Φ from:
  - **numerical mathematics**
  - **statistics**, **data mining**
  - *etc.*
  - ➢ **scientific computing**
- **IVA** w/ **T** ↔ **visual computing**

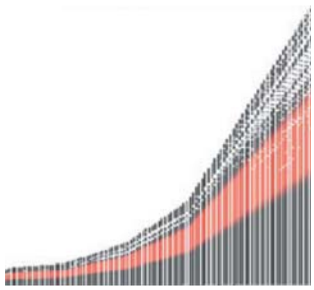


Attribute Derivation ↔ User Task / example

- The tools T, available in an IVA system, must reflect/match the **analytical steps of the user**:
- **Example**:
  - **first vis.**:  ↔ user wishes to select the "band" in the middle
  - **so?** an advanced brush? a lasso maybe?
  - **ah!** → let's normalize *y* and then brush (a)



  - **leading to the wished selection**:



What user wishes to reflect?

- Many **generic wishes** – users interest in:
  - something **relative** (instead of some absolute values), example: show me the *top-15%*
  - **change** (instead of current values), ex.: show me *regions with increasing temperature*
  - some **non-local property**, ex.: show me regions with *high average temperature*
  - **statistical properties**, ex.: show me *outliers*
  - **ratios/differences**, ex.: show me population per area, difference from trend
  - *etc.*
- **Common characteristic** here:
  - **questions/tools generic**, not application-dependent!

How to reflect these user wishes?

- Many **generic wishes** – users interest in:
  - something **relative** (instead of some absolute values), example: show me the *top-15%* ⇒ ***use**, e.g., **normalization***
  - **change** (instead of current values), ex.: show me *regions with increasing temperature* ⇒ ***derivative estimation***
  - some **non-local property**, ex.: show me regions with *high average temperature* ⇒ ***numerical integration***
  - **statistical properties**, ex.: show me *outliers* ⇒ ***descriptive statistics***
  - **ratios/differences**, ex.: show me population per area, difference from trend ⇒ ***calculus***
  - *etc.* ⇒ ***data mining*** (fast enough?)
- **Common characteristic** here:
  - **questions/tools generic**, not application-dependent!

Some useful tools for 3rd-level IVA

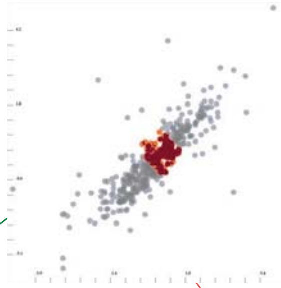
- From **analysis**, **calculus**, **num. math**:
  - **linear filtering** (convolve the data with some linear filter on demand, e.g., to smooth, for derivative estimation, *etc.*)
  - **calculus** (use an interactive formula editor for computing simple relations between data attributes; +, −, ·, ⁄, *etc.*)
  - **gradient estimation**, **numerical integration** (*e.g.*, wrt. space and/or time) ⇒ example
  - **fitting/resampling** via **interpolation/approximation**
- From **statistics**, **data mining**: ⇒ example
  - **descriptive statistics** (compute the statistical moments, also robust, measures of outlyingness, detrending, *etc.*) ⇒ example
  - **embedding** (project into a lower-dim. space, *e.g.*, with PCA for a subset of the attribs., *etc.*) ⇒ example
- **Important**: **executed on demand**, after prev. vis.

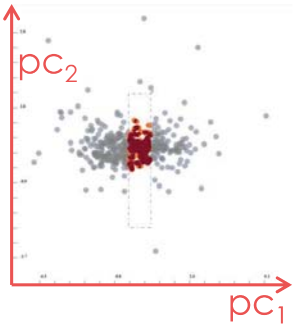
3rd-level IVA – Sample Iterations (1/2)

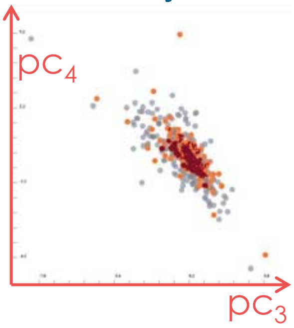
- The **Iterative Process of 3rd-level IVA**:
  - Example 1:
    - you look at some *temp. distribution over some region*
    - you are *interested in raising temperatures*, ***but not temperature fluctuations***
    - you use a **temporal derivate estimator**, *for ex.*, central differences $t_{change} = (t_{future} – t_{past})$ / len(future−past)
    - you plot t_{change}, *e.g.*, in a **histogram** and **brush** whatever change you are interested in
    - maybe you see some frequency amplification due to derivation, ***so you go back*** and
    - **use an appropriate smoothing filter** to *remove high frequencies from the temp. data*, leading to a derived new $\tau = t_{smooth}$ data attribute
    - selecting from a **histogram** of τ_{change} (computed like above) is then less sensitive to temperature fluctuations

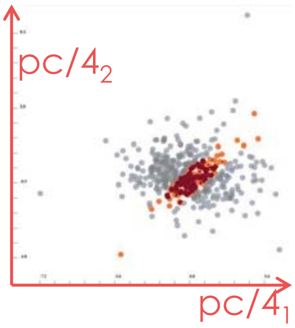
3rd-level IVA – Sample Iterations (2/2)

- The **Iterative Process of 3rd-level IVA**:
  - Example exploiting PCA:
    - you bring up a scatterplot of d_1 vs. d_2: (from an ECG dataset [Frank, Asuncion; 2010])
    - obviously, d_1 and d_2 are correlated, our interest: the **data center** *wrt. the main trend*
    - we ask for a (local) **PCA** of d_1 and d_2:
    - then we **brush the data center**
    - we get the wished selection
    - from here further steps are possible…, incl. study of other PCA-results, *etc.*









[IEEE Vis, 2008]

Brushing of Attribute Clouds for the Visualization of Multivariate Data

Heike Jänicke, Michael Böttinger, and Gerik Scheuermann, *Member, IEEE*

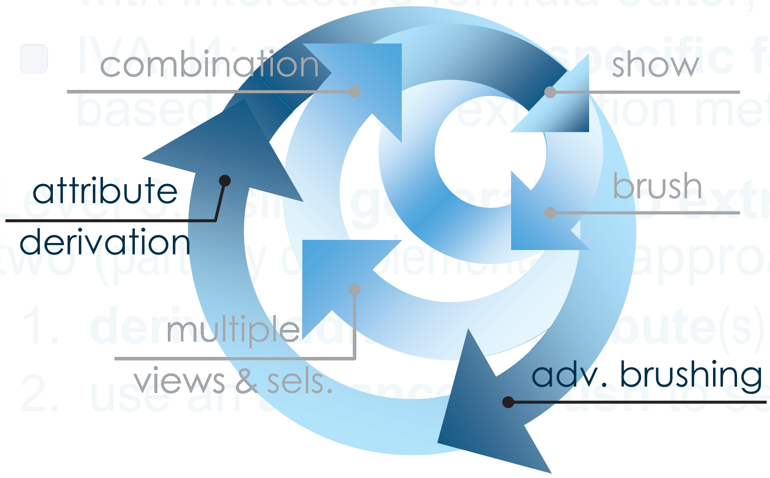
2D embedding: the attribute cloud

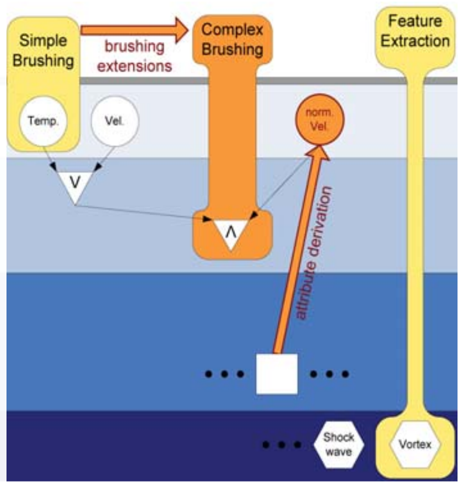
brushed cloud:

corresponding feature(s):

IVA – Levels of Complexity (4/4)

- A *lot* can be done with basic IVA, already! ✓ [pare rule
- **For more advanced** exploration/analysis **tasks**, we extend it (in seveal steps):
  - IVA, level 2: **logical combinations of brushes**, e.g., utilizing the *feature definition language* [Doleisch et al. 200] ✓
  - IVA, l. 3: **attribute derivation**; **advanced brushing**,



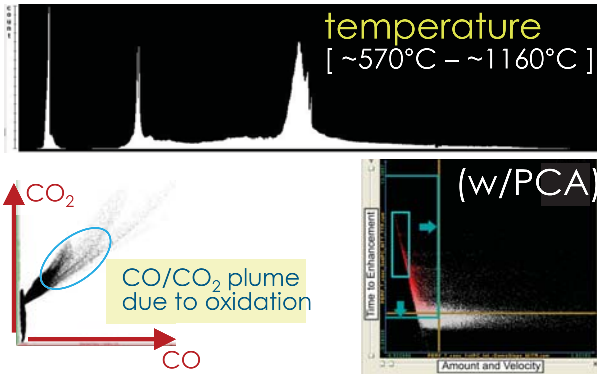


IVA – Levels of Complexity (4/4)

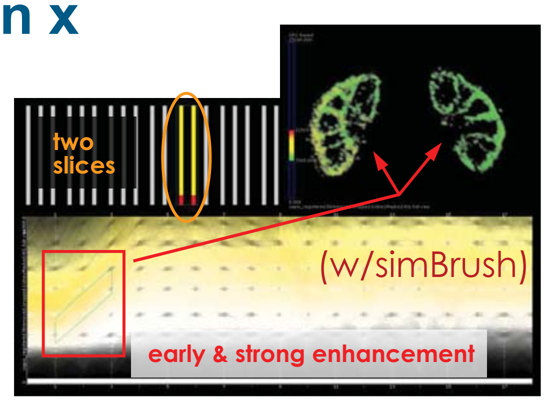
- A *lot* can be done with basic IVA, already! ✓ [pare rule
- **For more advanced** exploration/analysis **tasks**, we extend it (in seveal steps):
  - IVA, level 2: **logical combinations of brushes**, e.g., utilizing the *feature definition language* [Doleisch et al. 200] ✓
  - IVA, l. 3: **attribute derivation**; **advanced brushing**, with interactive formula editor; *e.g.*, similarity ✓
  - IVA, l4: **application-specific feature extraction**, e.g., based on vortex extraction methods for flow analysis
- Level 4: **application-specific procedures**
  - tailored solutions (for a specific problem)
  - "deep" information drill-down
  - *etc.*

Interactive Visual Analysis – delivery

- Understanding data **wrt. range d**
  - which distribution has data attribute $\mathbf{d}_i$?
  - how do $\mathbf{d}_i$ and $\mathbf{d}_j$ relate to each other? (**multivariate analysis**)
  - which $\mathbf{d}_k$ discriminate data features?



- Understanding data **wrt. domain x**
  - **where** are relevant features? (**feature localization**)
  - **which** values at specific **x**? (**local analysis**)
  - how are they **related to parameters**?



The Iterative Process of IVA…

…is a **very useful methodology** for **data exploration & analysis**

…is **very general** and can be (has already been) applied to **many different application fields** (in this talk the focus was on scientific data)

…**meets scientific computing** as a complementary methodology (with the **important difference** that in IVA the **user** with his/her **perception/cognition** is **in the loop** at **different frequencies**, also many fps)

…is **not yet fully implemented** (we've done something, e.g., in the context of **SimVis**, **ComVis**, *etc.*) – from here: different possible paths, incl. InteractiveVisualMatlab, IVR, *etc.*)

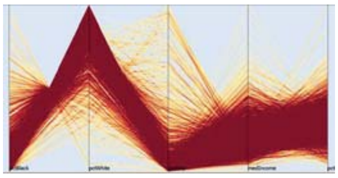
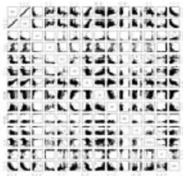
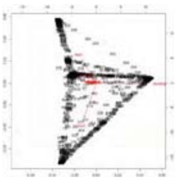
The **Dual Analysis Framework** for **High-dimensional Data IVA**

Çağatay Turkay, Helwig Hauser
University of Bergen

Target: High-dimensional Data

High- vs. multi-dimensional data

– **multi-dimensional**: >3D, 4D, 6D, 12D, ..., 24D(?)

– **high-dimensional**: ..., 40D, 80D, **240D**, **1200D**, ...
  - std. tools for multi-dim. data vis. don't work
  - lots of statistics, etc., do not work properly, esp. when #dims. > #items

Where?

– **Biology data** (e.g., from genomics/proteomics), **astronomy data** (e.g., spectral imaging data), **survey data** (many questions), ...

Understanding *n*D for (really) large *n*

Curse of dimensionality is a problem, when *n* large

– *n*D distances become meaningless
– with that distances-based project methods
– statistics of wide tables don't work

Hypothesis:

– there is valuable information in the «**space of dimensions**»
– the «space of dimensions» is **structured**, **heterogeneous**
– it's worthwhile to understand this «space of dimensions» in order to do a **better informed IVA of the data items**

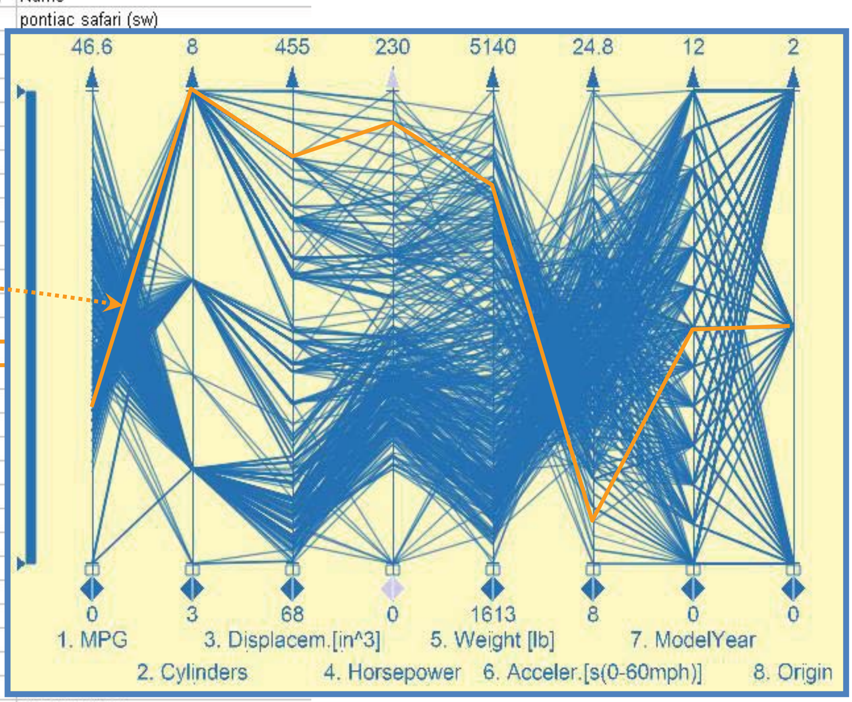
But how to understand this «space of dimensions»?

Can we visualize the dimensions of a dataset?

Traditional Vis = Items Visualization

Almost all of visualization is about
visualizing the (multi-dimensional) **data items**

MPG	Cyls.	Displace-ment [in³]	Horse-power	Weight [lb]	Seconds ´til 60mph	Model Year	Origin	Name
13	8	400	175	5140	12	71	US	pontiac safari (sw)
11	8	400	150	4997	14	73	US	
12	8	383	180	4955	11,5	71	US	
12	8	429	198	4952	11,5	73	US	
12	8	455	225	4951	11	73	US	
12	8	400	167	4906	12,5	73	US	
13	8	400	170	4746	12	71	US	
13	8	440	215	4735	11	73	US	
9	8	304	193	4732	18,5	70	US	
13	8	350	150	4699	14,5	74	US	
16	8	400	170	4668	11,5	75	US	
14	8	351	148	4657	13,5	75	US	
13	8	360	170	4654	13	73	US	
14	8	302	140	4638	16	74	US	
11	8	429	208	4633	11	72	US	
10	8	360	215	4615	14	70	US	
13	8	350	155	4502	13,5	72	US	
12	8	350	180	4499	12,5	73	US	
16	8	318	150	4498	14,5	75	US	
14	8	400	175	4464	11,5	71	US	
13	8	400	150	4464	12	73	US	
14	8	318	150	4457	13,5	74	US	
12	8	350	160	4456	13,5	72	US	
15	8	350	145	4440	14	75	US	
14	8	455	225	4425	10	70	US	
13	8	400	190	4422	12,5	72	US	
14	8	400	175	4385	12	72	US	
11	8	318	210	4382	13,5	70	US	
16,5	8	350	180	4380	12,1	76	US	

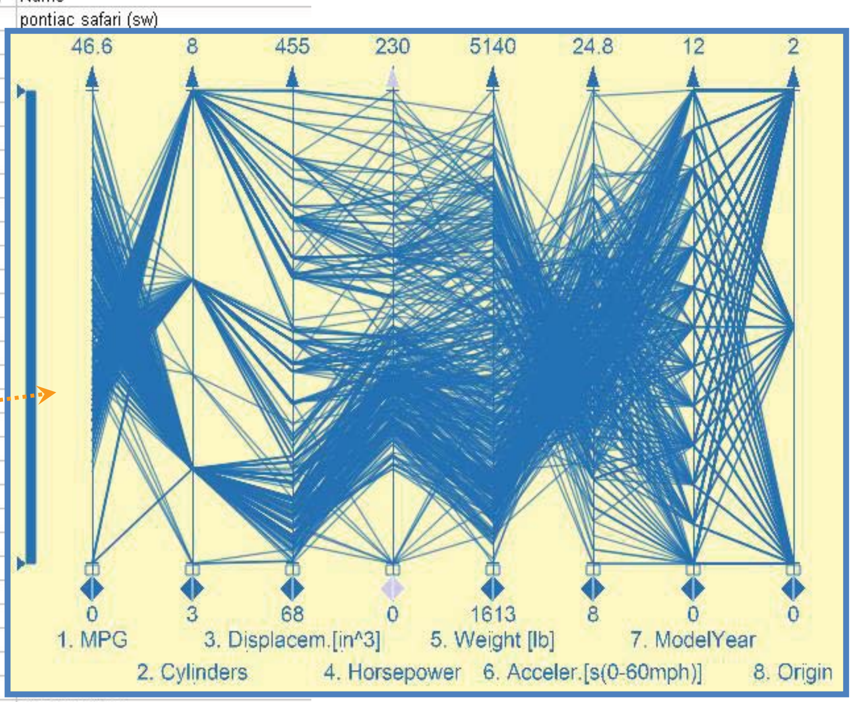


A new perspective: Dims. Visualization

Alternatively, and esp., when we have so many dims.,
we could visualize the **data dimensions** themselves!

MPG	Cyls.	Displace-ment [in³]	Horse-power	Weight [lb]	Seconds ´til 60mph	Model Year	Origin	Name
13	8	400	175	5140	12	71	US	pontiac safari (sw)
11	8	400	150	4997	14	73	US	
12	8	383	180	4955	11,5	71	US	
12	8	429	198	4952	11,5	73	US	
12	8	455	225	4951	11	73	US	
12	8	400	167	4906	12,5	73	US	
13	8	400	170	4746	12	71	US	
13	8	440	215	4735	11	73	US	
9	8	304	193	4732	18,5	70	US	
13	8	350	150	4699	14,5	74	US	
16	8	400	170	4668	11,5	75	US	
14	8	351	148	4657	13,5	75	US	
13	8	360	170	4654	13	73	US	
14	8	302	140	4638	16	74	US	
11	8	429	208	4633	11	72	US	
10	8	360	215	4615	14	70	US	
13	8	350	155	4502	13,5	72	US	
12	8	350	180	4499	12,5	73	US	
16	8	318	150	4498	14,5	75	US	
14	8	400	175	4464	11,5	71	US	
13	8	400	150	4464	12	73	US	
14	8	318	150	4457	13,5	74	US	
12	8	350	160	4456	13,5	72	US	
15	8	350	145	4440	14	75	US	
14	8	455	225	4425	10	70	US	
13	8	400	190	4422	12,5	72	US	
14	8	400	175	4385	12	72	US	
11	8	318	210	4382	13,5	70	US	
16,5	8	350	180	4380	12,1	76	US	

?



Naïve Approach

Transposing the data table should do it, right? :–)

Not really…

Naïve Approach

Transposing the data table should do it, right? :–)

Not really…

	MPG	Cylinders	Displacem.[in³]	Horsepower	Weight [lb]	Acceler.[s(0-60	ModelYear	Origin	Name
Car001	18	8	307	130	3504	12	70	US	chevrolet chevelle malibu
Car002	15	8	350	165	3693	11,5	70	US	buick skylark 320
Car003	18	8	318	150	3436	11	70	US	plymouth satellite
Car004	16	8	304	150	3433	12	70	US	amc rebel sst
Car005	17	8	302	140	3449	10,5	70	US	ford torino
Car006	15	8	429	198	4341	10	70	US	ford galaxie 500
Car007	14	8	454	220	4354	9	70	US	chevrolet impala
Car008	14	8	440	215	4312	8,5	70	US	plymouth fury iii
Car009	14	[illegible]	[illegible]	[illegible]	4425	10	70	US	pontiac catalina
Car010	15	8	390	190	3850	8,5	70	US	amc ambassador dpl
Car011	NA	4	[illegible]	[illegible]	[illegible]	[illegible]	70	Europe	citroen ds-21 pallas
Car012	NA	[illegible]	350	165	4142	11,5	70	US	chevrolet chevelle concours
Car013	NA	[illegible]	351	153	4034	11	70	US	ford torino (sw)
Car014	NA	8	383	175	4166	10,5	70	US	plymouth satellite (sw)
Car015	NA	8	360	175	3850	11	70	US	amc rebel sst (sw)
Car016	15	8	383	170	3563	10	70	US	dodge challenger se
Car017	14	8	340	160	3609	8	70	US	plymouth 'cuda 340
Car018	NA	8	302	140	3353	8	70	US	ford mustang boss 302
Car019	15	8	400	150	3761	9,5	70	US	chevrolet monte carlo
Car020	14	8	455	225	3086	10	70	US	buick estate wagon (sw)
Car021	24	4	113	95	2372	15	70	Japan	toyota corona mark ii

just for illustration:

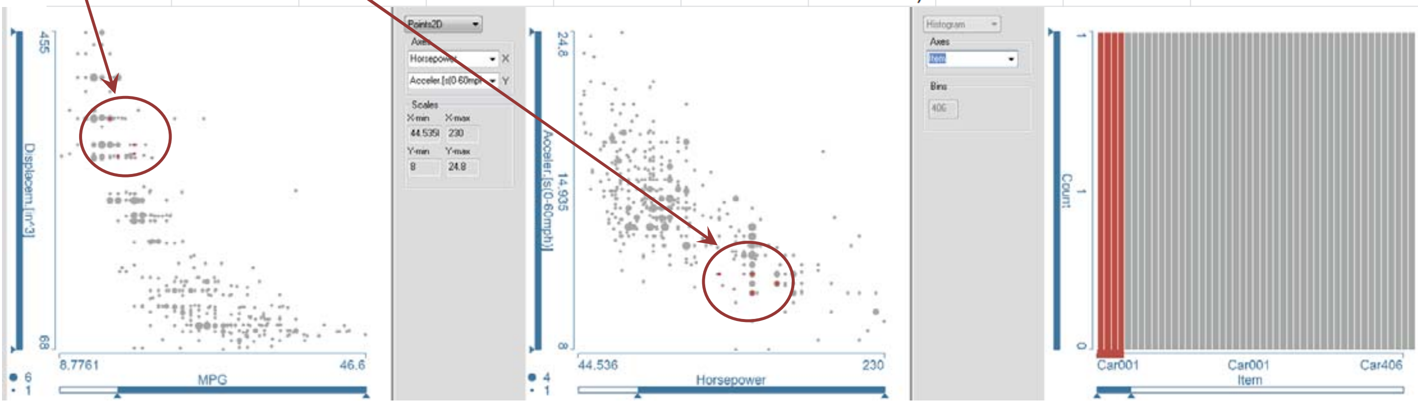
406 cars, 7 numeric dimensions

(not a high-dimensional dataset!)

Naïve Approach

Visualizing rows (cars) from this table is standard InfoVis:

	MPG	Cylinders	Displacem.[in'	Horsepower	Weight [lb]	Acceler.[s(0-6(	ModelYear	Origin	Name
Car001	18	8	307	130	3504	12	70	US	chevrolet chevelle malibu
Car002	15	8	350	165	3693	11,5	70	US	buick skylark 320
Car003	18	8	318	150	3436	11	70	US	plymouth satellite
Car004	16	8	304	150	3433	12	70	US	amc rebel sst
Car005	17	8	302	140	3449	10,5	70	US	ford torino
Car006	15	8	429	198	4341	10	70	US	ford galaxie 500
Car007	14	8	454	220	4354	9	70	US	chevrolet impala
Car008	14	8	440	215	4312	8,5	70	US	plymouth fury iii
Car009	14	8	455	225	4425	10	70	US	pontiac catalina
Car010	15	8	390	190	3850	8,5	70	US	amc ambassador dpl
Car011	NA	4	133	115	3090	17,5	70	Europe	citroen ds-21 pallas
Car012	NA	8	350	165	4142	11,5	70	US	chevrolet chevelle concour



Naïve Approach

Data transposition makes the dimensions to rows:

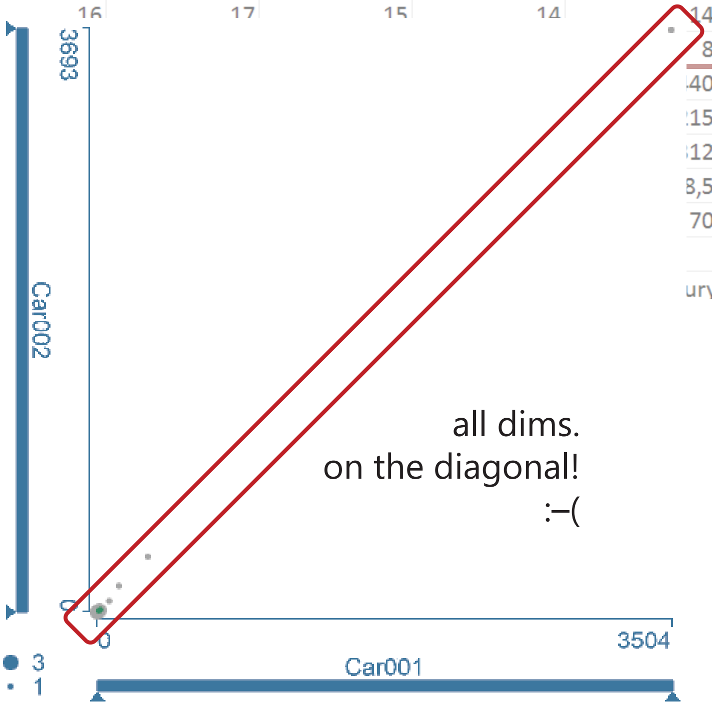
	Car001	Car002	Car003	Car004	Car005	Car006	Car007	Car008
MPG	18	15	18	16	17	15	14	14
Cylinders	8	8	8	8	8	8	8	8
Displacem.[in'	307	350	318	304	302	429	454	440
Horsepower	130	165	150	150	140	198	220	215
Weight [lb]	3504	3693	3436	3433	3449	4341	4354	4312
Acceler.[s(0-6(	12	11,5	11	12	10,5	10	9	8,5
ModelYear	70	70	70	70	70	70	70	70
Origin	US	US	US	US	US	US	US	US
Name	chevrolet chev	buick skylark 3	plymouth sate	amc rebel sst	ford torino	ford galaxie 5(	chevrolet imp;	plymouth fury

So what about visualizing this table?

Naïve Approach

Data transposition makes the dimensions to rows:

	Car001	Car002	Car003	Car004	Car005	Car006	Car007	Car008
MPG	18	15	18	16	17	15	14	14
Cylinders	8	8	8					8
Displacem.[in'	307	350	318					40
Horsepower	130	165	150					15
Weight [lb]	3504	3693	3436					12
Acceler.[s(0-60	12	11,5	11					3,5
ModelYear	70	70	70					70
Origin	US	US	US	US				
Name	chevrolet chev	buick skylark 3	plymouth sate	amc				ury

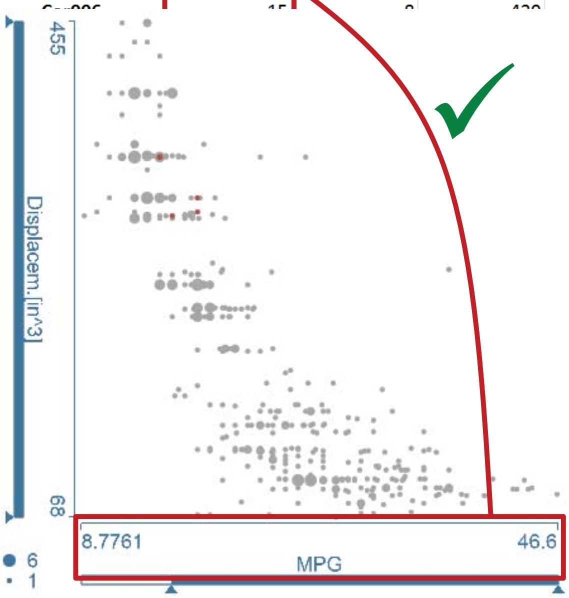


So why doesn't this work?

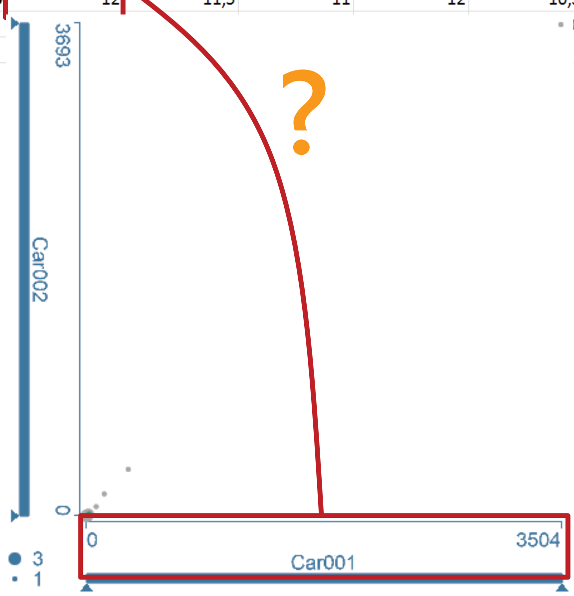
Naïve Approach

No comparable values in the columns after transposition!

	MPG	Cylinders	Displacem.[in'
Car001	18	8	307
Car002	15	8	350
Car003	18	8	318
Car004	16	8	304
Car005	17	8	302



	Car001	Car002	Car003	Car004	Car005
MPG	18	15	18	16	17
Cylinders	8	8	8	8	8
Displacem.[in	307	350	318	304	302
Horsepower	130	165	150	150	140
Weight [lb]	3504	3693	3436	3433	3449
Acceler.[s(0-6	12	11,5	11	12	10,5
ModelYear					0
Origin					
Name					



Dual Analysis Framework

What do we do, when we visualize data items?

- per data item $\mathbf{p}_i$,
  we map properties/attributes of $\mathbf{p}_i$ to vis.-cues, f.i., x and y →
- we see, how the $\mathbf{p}_i$ relate to each other wrt. to their props.!

Translating this to visualizing dimensions:

- per data dimension $\mathbf{d}_j$,
  we map properties/attributes of $\mathbf{d}_j$ to vis.-cues, f.i., x and y →
- we see, how the $\mathbf{d}_j$ relate to each other wrt. to their props.!

Expressive properties of dimensions $\mathbf{d}_j$ (selection):

- descriptive statistics, like mean and std.-der.
- measures of outlyingness

Dual Analysis Framework

So: constructing the properties table for dims. $\mathbf{d}_j$:

- normalization first, then feature extraction

	n(MPG)	n(Cylinders)	n(Displacem.[i	n(Horsepower	n(Weight [lb])	n(Acceler.[s(0-	n(ModelYear)	Origin	Name
#Outl	40	0	31	42	44	50	35		
#Outl_high	33	0	31	40	42	25	0		
#Outl_low	7	0	0	2	2	25	35		
high_outl_th	0,698	1,013	0,735	0,638	0,748	0,700	1,008		
low_outl_th	0,076	-0,013	-0,078	0,007	0,029	0,200	0,072		
IQR	0,309	0,800	0,510	0,298	0,396	0,210	0,500		
q_3	0,532	1,000	0,605	0,457	0,570	0,548	0,750		
Med.	0,372	0,200	0,214	0,266	0,343	0,446	0,500		
q_1	0,223	0,200	0,095	0,159	0,174	0,338	0,250		
Kurt.	-0,511	-1,411	-0,811	0,541	-0,821	0,373	-1,200		
Skew.	0,457	0,506	0,694	1,034	0,506	0,230	0,021		
Std.-Dev.	0,208	0,342	0,271	0,210	0,240	0,167	0,312		
Avg.	0,39	0,50	0,33	0,32	0,39	0,45	0,54		
Max.	1,00	1,00	1,00	1,00	1,00	1,00	1,00		
Min.	0,00	0,00	0,00	0,00	0,00	0,00	0,00		
Car001	0,24	1,00	0,62	0,46	0,54	0,24	0,00	US	chevrolet chevelle malibu
Car002	0,16	1,00	0,73	0,65	0,59	0,21	0,00	US	buick skylark 320
Car003	0,24	1,00	0,65	0,57	0,52	0,18	0,00	US	plymouth satellite
Car004	0,19	1,00	0,61	0,57	0,52	0,24	0,00	US	amc rebel sst
Car005	0,21	1,00	0,60	0,51	0,52	0,15	0,00	US	ford torino
Car006	0,16	1,00	0,93	0,83	0,77	0,12	0,00	US	ford galaxie 500
Car007	0,13	1,00	1,00	0,95	0,78	0,06	0,00	US	chevrolet impala
Car008	0,13	1,00	0,96	0,92	0,77	0,03	0,00	US	plymouth fury iii
Car009	0,13	1,00	1,00	0,97	0,80	0,12	0,00	US	pontiac catalina

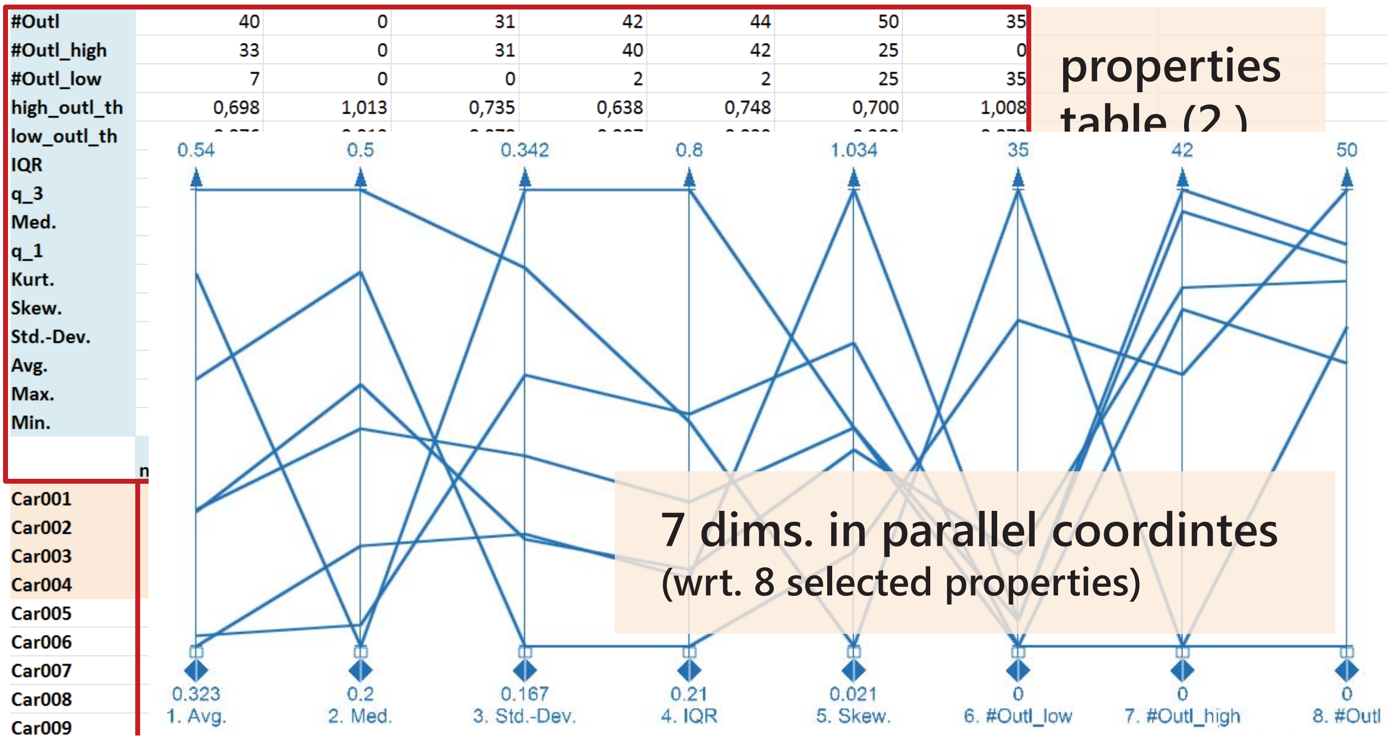
properties table (2.)

normalized data (1.)

Dual Analysis Framework

Now we can visualize the dims. $\mathbf{d}_j$!

– mapping coherent and expressive properties to vis.-cues!



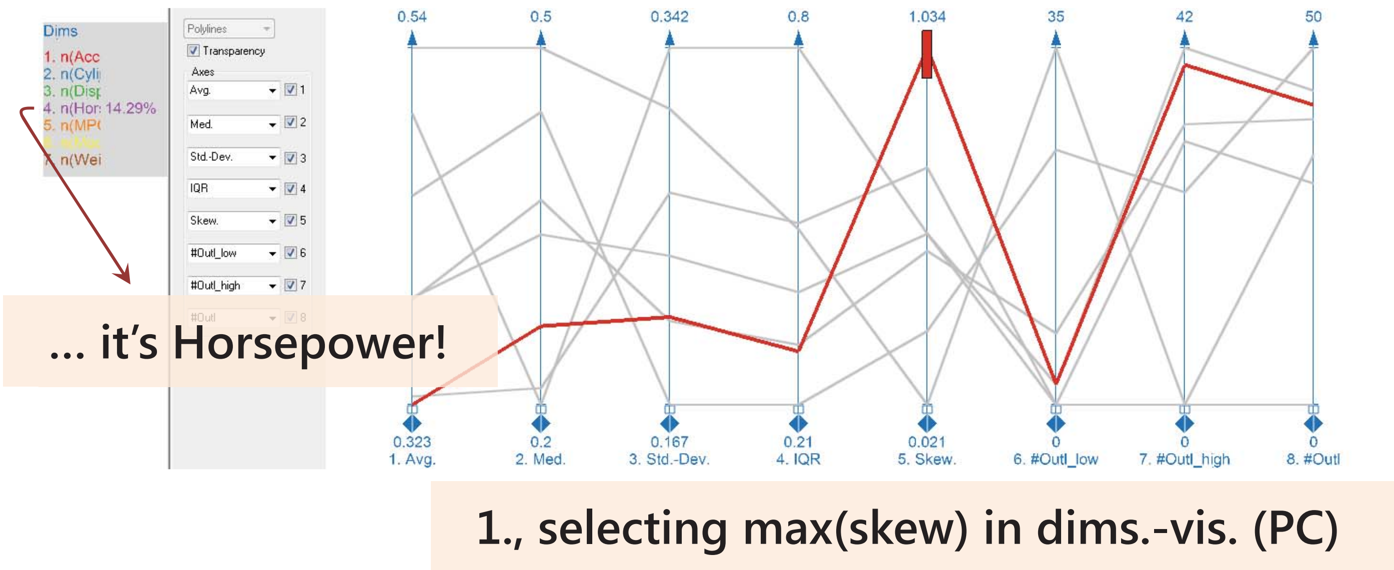
Dual Analysis Framework

Now the dual analysis can start!

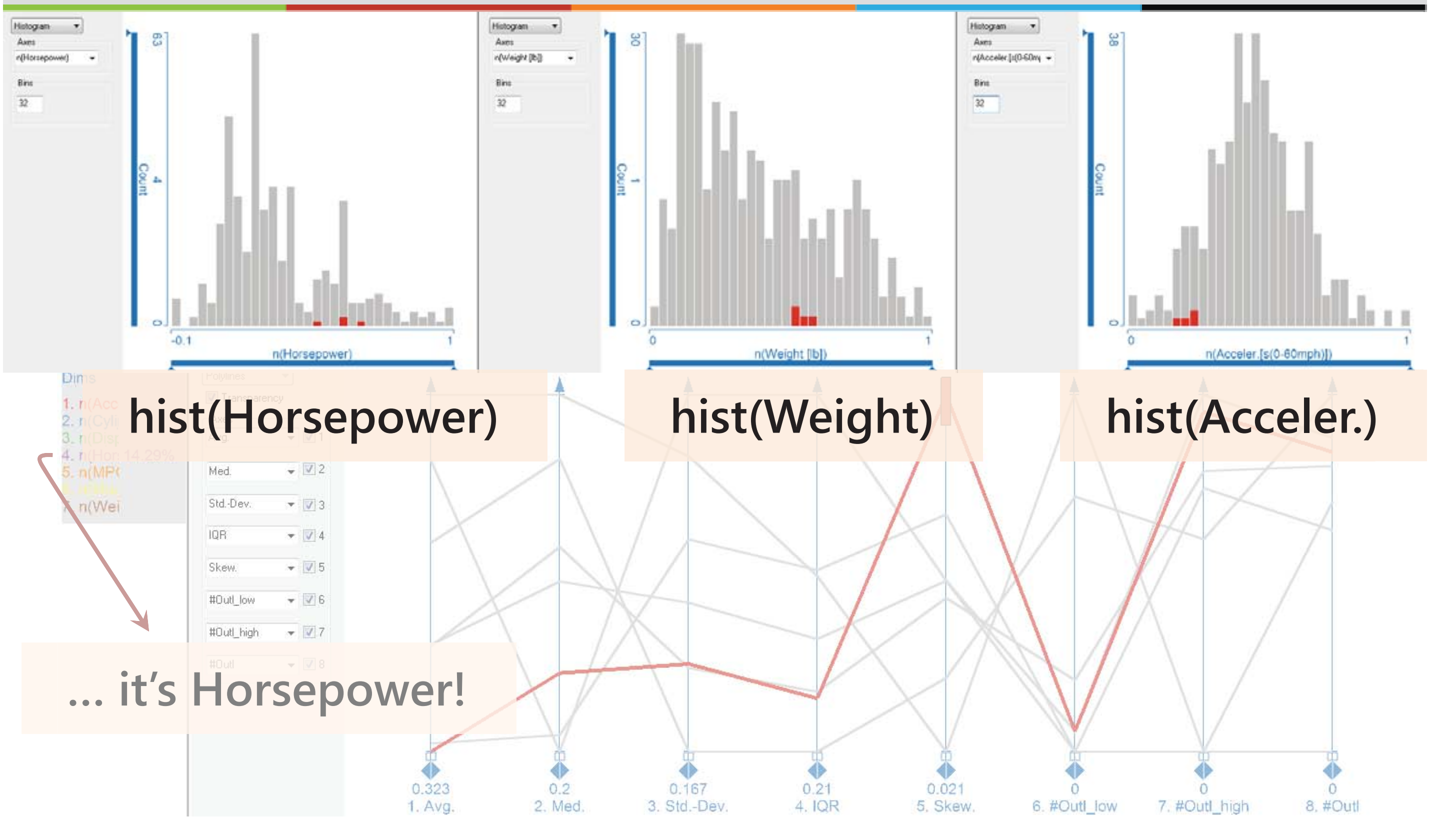
– look up informative properties in the dims.-vis.

– do related items-visualization, accordingly

Example 1: exploring the most skewed dimension



Dual Analysis Framework



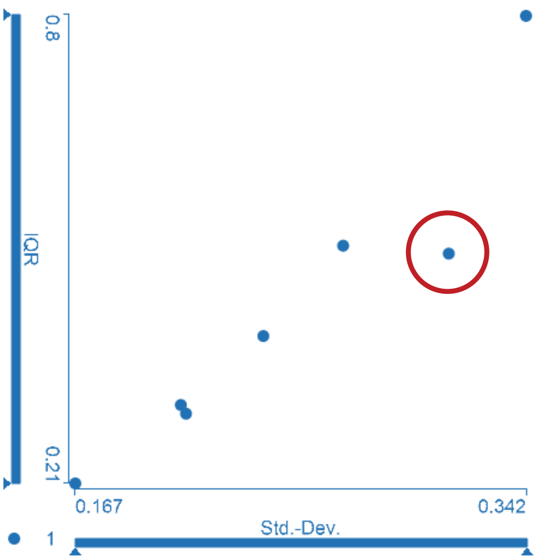
1., selecting max(skew) in dims.-vis. (PC)

Dual Analysis Framework

Now the dual analysis can start!

- look up informative properties in the dims.-vis.
- do related items-visualization, accordingly

Example 2: comparing Gaussian & ranking-based stats.



1., seeing that both measures for the spread (std.-dev. vs. IQR) agree for all but one dim.

2., selecting it (not shown) reveals: it's ModelYear, a discrete data dimension…

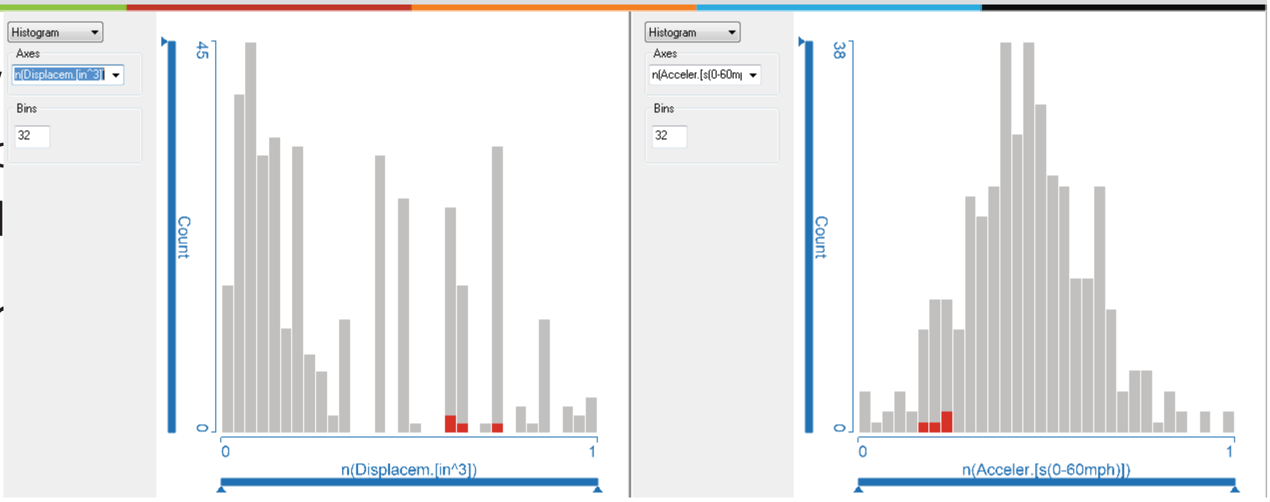
Dual Analysis Framework

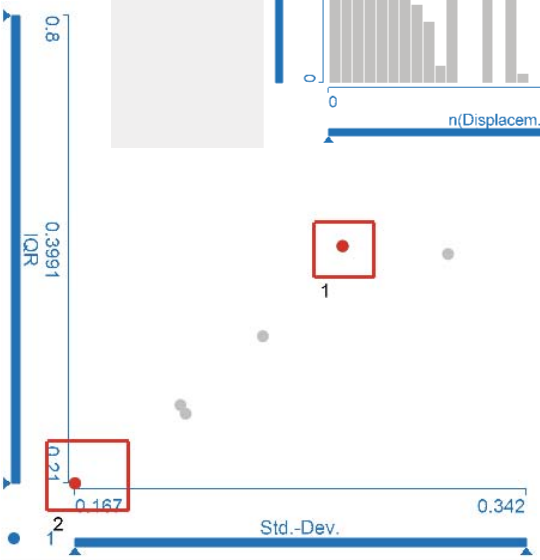
Now

– lo

– d

Exam





next:
comparing a low-spread (Accel.) vs. a high-spread dim. (Displacem.)

... the top-right dim. – max. spread – is Cylinders – basically categorical...

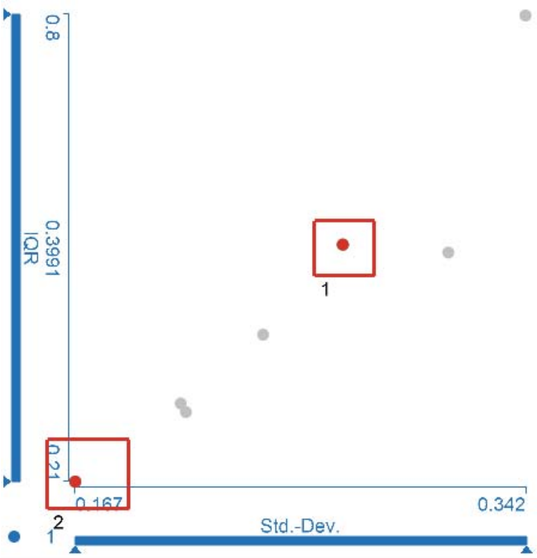
Dual Analysis Framework

Now the dual analysis can start!

– look up informative properties in the dims.-vis.

– do related items-visualization, accordingly

Example 2: comparing Gaussian & ranking-based stats.



next:
comparing a low-spread (Accel.) vs. a high-spread dim. (Displacem.)

... the top-right dim. – max. spread – is Cylinders – basically categorical...

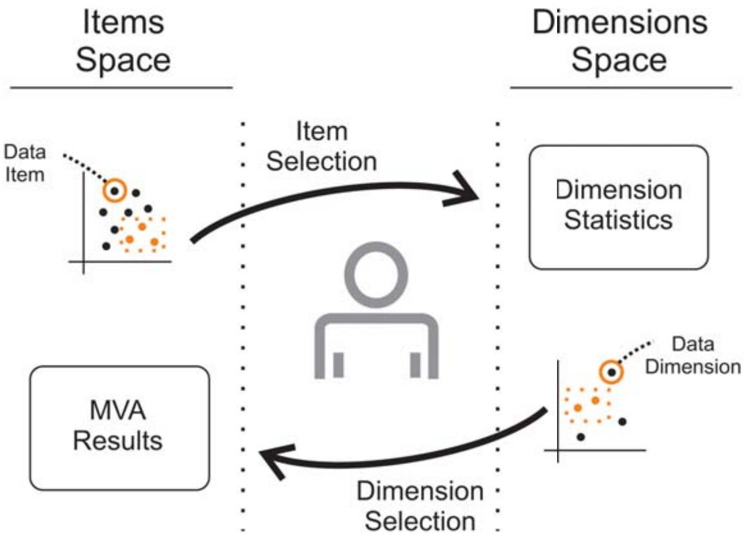
Dual Analysis Framework

Quickly, we explore the dims. according to their props.

- hundreds or thousands of dims. → no problem! :–)
- dozens of properties → std. InfoVis is fine!

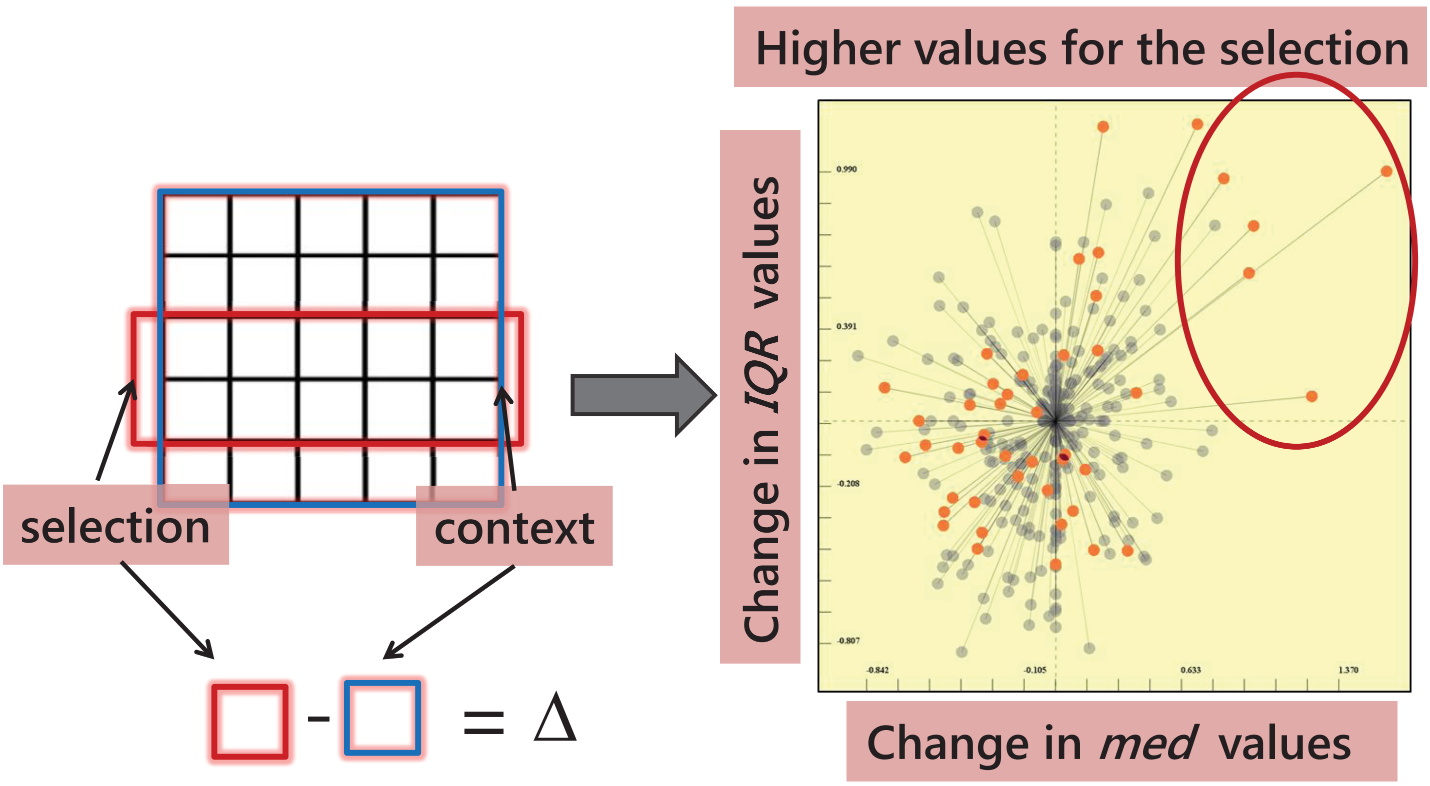
The Dual Analysis emerges through iteration:

- one key tool:
  the difference view



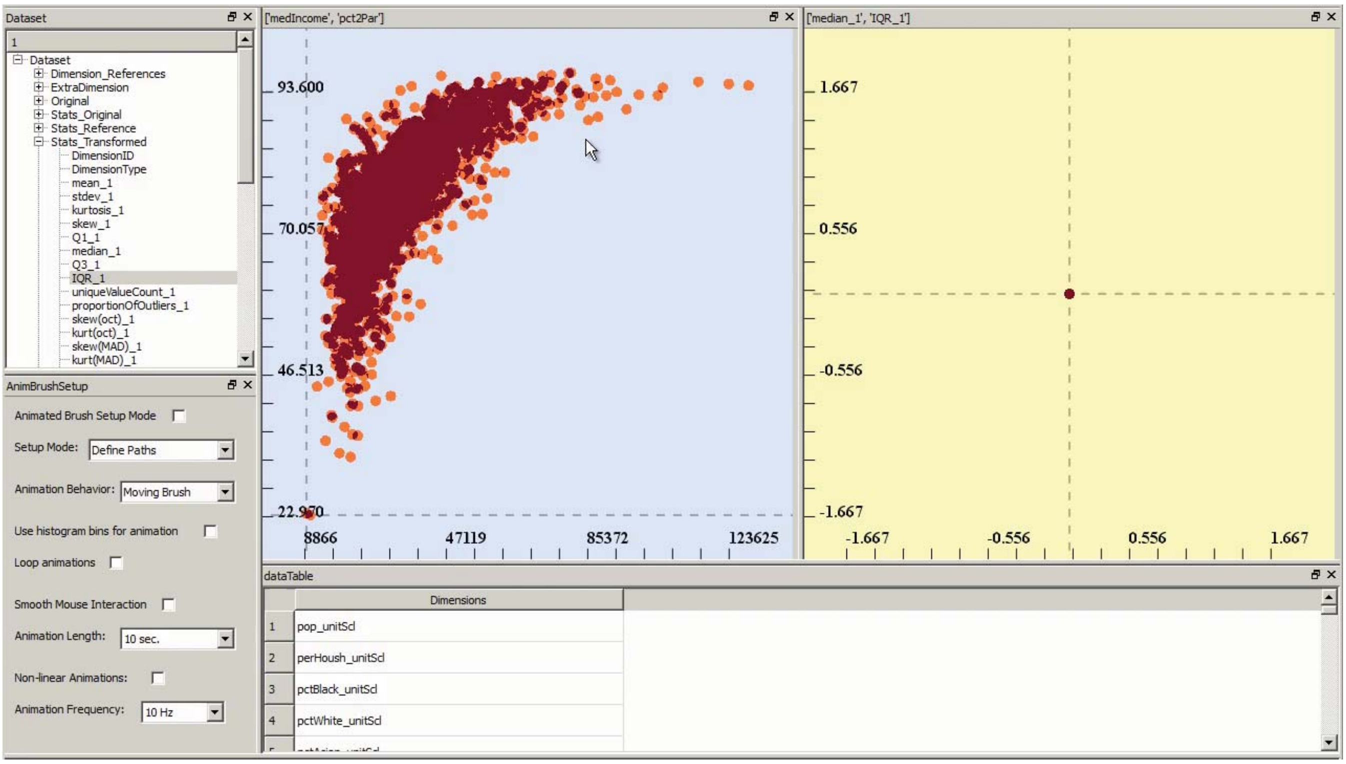
Dual Analysis Framework

Working with the difference view



Dual Analysis Framework

Difference View in action:



Dual Analysis Framework

More in the thesis / papers of Çağatay Turkay *et al.*

C. Turkay, P. Filzmoser, and H. Hauser. **Brushing Dimensions–A Dual Visual Analysis Model for High-Dimensional Data.** *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2591–2599, 2011.

C. Turkay, A. Lundervold, A.J. Lundervold, and H. Hauser. **Representative Factor Generation for the Interactive Visual Analysis of High-Dimensional Data.** *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2621–2630, 2012.

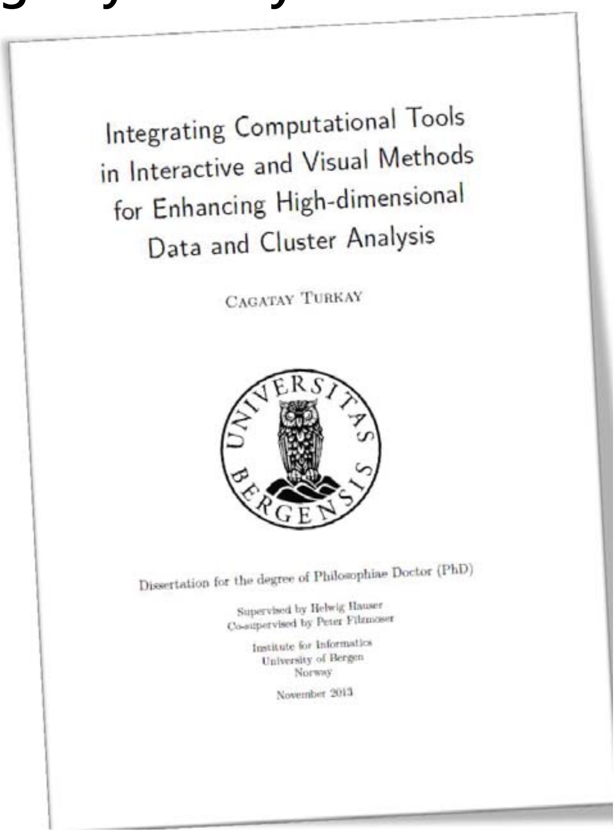
C. Turkay, P. Angelelli, P. Filzmoser, and H. Hauser. **Outlier Dimensions – Outlier Aware Analysis of High-dimensional Data.** In submission to: *IEEE Transactions on Visualization and Computer Graphics*, 2013.

C. Turkay and H. Hauser. **Optimizing Processes in Visual Analytics to Meet the Three Human Time Constants.** In submission to: *Computers and Graphics*, 2013.

C. Turkay, A. Lundervold, A.J. Lundervold, and H. Hauser. **Hypothesis Generation by Interactive Visual Exploration of Heterogeneous Medical Data.** *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data. Lecture Notes in Computer Science*, Volume 7947:1–12, 2013.

C. Turkay, A. Lex, M. Streit, H.P. Pfister, and H. Hauser. **Characterizing Cancer Subtypes using the Dual Analysis Approach in Caleydo.** In submission to: *IEEE Computer Graphics and Applications*, 2013.

C. Turkay, J. Parulek, N. Reuter, and H. Hauser. **Interactive Visual Analysis of Temporal Cluster Structures.** *Computer Graphics Forum*, 30(3):711–720, 2011.



Acknowledgements

You!

Luis Gustavo!

We'd like to hire
- **1 PostDoc**

Collaborators:
H. Doleisch, R. Fuchs/Bürger, J. Kehrer, Ç. Turkay, Z. Konyha, Kr. Matković, P. Filzmoser, *et al.*

UNIVERSITY OF BERGEN

Upcoming Job Opportunities in Visualization at the University of Bergen

The visualization research group at the University of Bergen's Department of Informatics, Norway (UiB), is preparing the announcement of two open positions. We invite all potentially interested applicants to contact us directly for further details.

- **1 PhD Fellowship (4 years):**
  We are looking for a highly motivated individual with interest in the visualization of medical data. Suitable applicants should have a background in visualization, human-computer interaction, and/or computer graphics and very good programming skills. Since eligible candidates are required to have completed their Master's degree prior to the upcoming job announcement, we ask all interested candidates to contact us as soon as possible.
  Contact: Prof. Stefan Bruckner (stefan.bruckner@uib.no)
- **1 Postdoctoral Fellowship (4 years):**
  We are looking for a promising scholar to fill a postdoctoral fellowship in visualization. This strategic position is not associated with a specific research project, but meant to strengthen the core visualization research activities in Bergen – considerable flexibility in terms of filling this appointment is envisioned. Applicants should have a completed PhD degree in visualization or a closely related subject, a strong publication record, a strong interest in high-quality research and also substantial interest in teaching of visualization and related topics.
  Contact: Prof. Helwig Hauser (helwig.hauser@uib.no)

The University of Bergen

The University of Bergen, Norway (UiB), has approximately 14,500 enrolled students and 3,200 faculty and staff. Six faculties cover most of the traditional university disciplines. Within the faculties there are 40 different specialized departments, multi-disciplinary research centers, and institutes. The University is engaged in the European Union's Framework programs for research and technological development and has been designated as a European Research Infrastructure and a Research Training Site in several scientific fields. Since 1997 more than 500 European researchers (professors, senior researchers, post docs and PhD candidates) have visited Bergen on EU grants, making UiB one of the most international universities, setting out to attract both established and junior scientists to contribute to research teams and work in multidisciplinary research groups. **Further information:** http://www.uib.no/en

The Visualization Group

The Department of Informatics at UiB is a dynamic international environment. The visualization research group was established in 2007 and focuses on new solutions for the effective and efficient visualization of large and complex data from measurements (e.g., from medical imaging or from seismic/sonar sensors), computational simulation (e.g., from computational fluid dynamics), or from analytic modeling (e.g., in the form of differential equations) for the purpose of data exploration, analysis, and presentation. Despite being a still relatively young group, the recent ten-year evaluation of all 62 informatics research groups in Norway conducted by an international evaluation panel on behalf of Norway's Research Council, found that the group's "[...] accomplishments made so far are excellent." and concluded with an overall assessment of "very good to excellent". **Further information:** http://www.ii.uib.no/vis

The City of Bergen

Bergen is Norway's second largest city with over 260,000 inhabitants and is located in the county of Hordaland on the south-western coast of Norway. Known as the "gateway to the fjords", Bergen is surrounded by spectacular scenery. As one of the offices of the Hanseatic League, Bergen was for several hundred years the center of prosperous trade between Norway and the rest of Europe. Today, Bergen is a lively and vibrant city with an international flair. Every semester the city with its university and colleges are proud and enthusiastic to welcome students from all over the world. Bergen is one of the most popular student cities in Scandinavia with 30,000 students in many different educational institutions. Its international atmosphere, together with the fact that Norway has been voted #1 in standard of living by the UNDP's Human Development Index for several years, make Bergen a very attractive place for life and study. **Further information:** http://www.visitbergen.com/en