# Compromises and Added Value in Visual Analytics

**Helwig Hauser** (Univ. of Bergen)

## Plan

~~So what is Visual Analytics?~~

*OK, OK, ..., not this question, again, ...* ☺

Instead:

**Five selected characteristics (C1–C5) of Visual Analytics** (VA solutions)

leading to

a discussion of

**– compromises**

**– chances**

C1: **problem size**

C2: **visualization richness**

C3: **interaction pace**

C4: **computational analysis**

C5: **comprehensiveness**

* So let's start with **C1: Problem Size...**

## C1: Large Data

- A recurring statement: there's **sooo much data**...
  ... therefore **visual analytics**!

Daniel Keim, 2007:

Dagstuhl Seminar Talk

**Challenge of the Information Age**

- 100 million FedEx transactions per day
- 150 million VISA credit card transactionen per day
- 300 million long distance calls in ATT's network per day
- 50 billion e-mails worldwide per day
- 600 billion IP packets per day DE-CIX backbone

Visual Analytics – Keim – May, 31 2007

**Why is the topic highly relevant today?**

- Very Large Data Collections are available in Databases and Data Warehouses
- On the Basis of the Data Complex Decisions have to made in a timely fashion
- Pure Visualization Methods (Information Visualisation) do not work for Billions of Data Records
- Full Automatic Knowledge Discovery Approaches only work for well-defined and clearly specifiable problems.
- Especially for adversarial situations: *Fraud, Viruses, SPAM, Attacks, Competition, ...*

Visual Analytics – Keim – May, 31 2007

Thomas Ertl, 2009:

Visual Computing Trends, Wien

**Information Explosion**

Univ. Stuttgart, Visualization and Interactive Systems Institute

- Today: the Peta ($10^{15}$) era
- Computer Simulations
  - Weather — 1 Petabyte per year
  - High-performance computing center — 1 Petaflop per second
  - Game PC graphics card — 1 Teraflop per second
- Sensors
  - Satellite astronomy — 1 Petabyte per year
  - Particle physics (CERN) — 1 Petabyte per year
  - Bio-chemical high-troughput — 1 Petabyte per year
- Digital Events
  - Credit card transactions — 100 Mio per day
  - Long distance calls ATT — 300 Mio per day
  - Internet packets through DE-CIX — 500 Bill per day

**Information Explosion**

Univ. Stuttgart, Visualization and Interactive Systems Institute

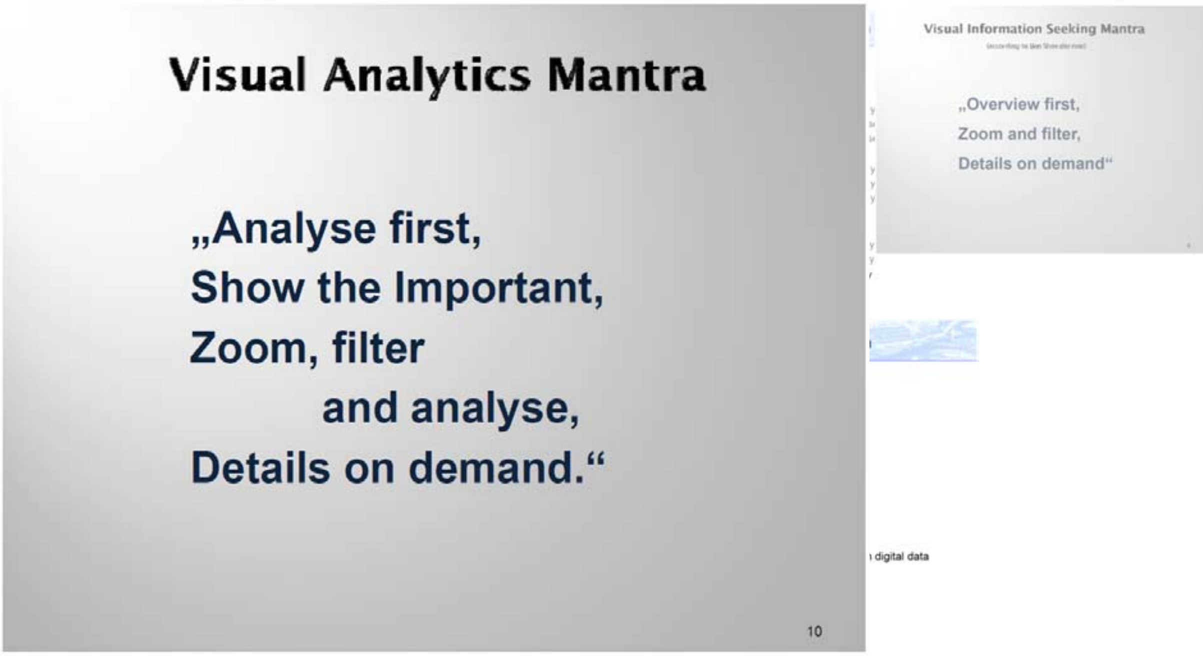
- Tomorrow: the era of Exa($10^{18}$) and Zeta ($10^{21}$)
- Digital Information (created, captured, replicated)
  - 2002 — 22 Exabyte
  - 2006 — 160 Exabyte
  - 2010 — almost 1 Zetabyte increase per year (28 trillion e-mails/year, totaling about 6 EB of data)
- Information "stored" by humans
  - est.1 Petabyte in entire life time
  - 80% through vision (space, form, color, texture, ..)
  - through display with 1-100 Megapixels
- **Visualization** plays a significant role in dealing with digital data
- **Interaction** and **abstraction** are the keys for the visualization of huge data sets

## C1: Large Data

- A recurring statement: there's **sooo much data**...
  ... therefore **visual analytics**!

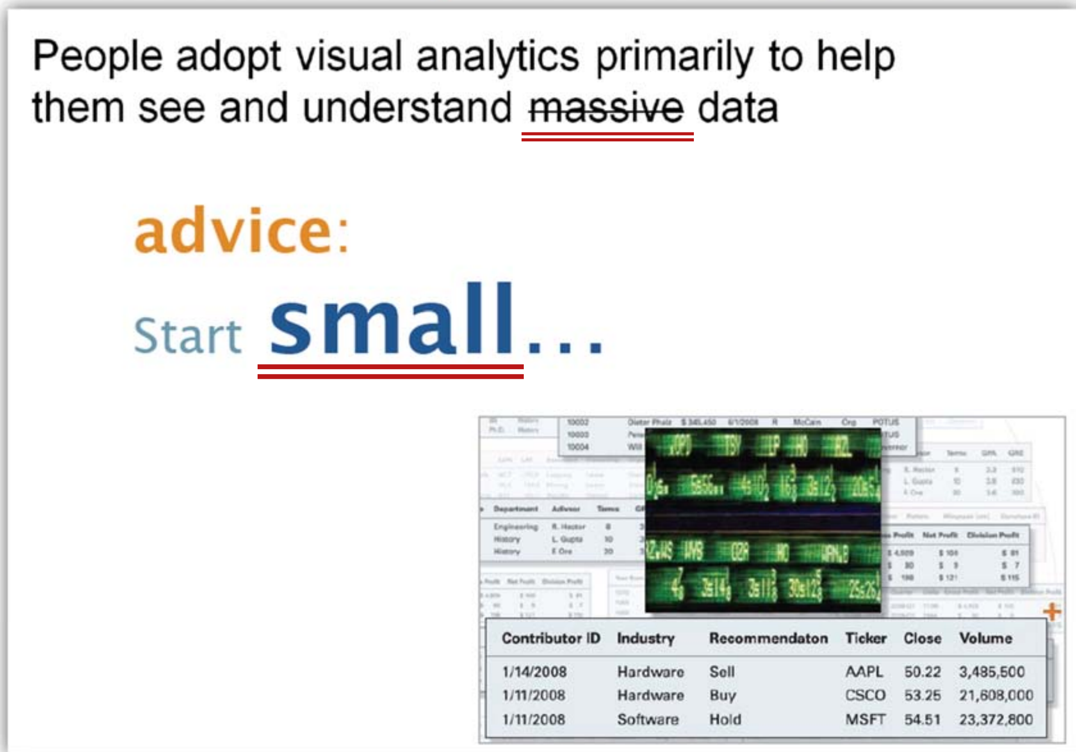
  Daniel Keim, 2009:

Dagstuhl Seminar Talk




## C1: Large Data – a question mark...

- Christian Chabot (CEO of Tableau), 2008:




VAST Keynote

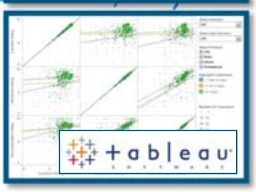
## C1: Problem Size – some categorization

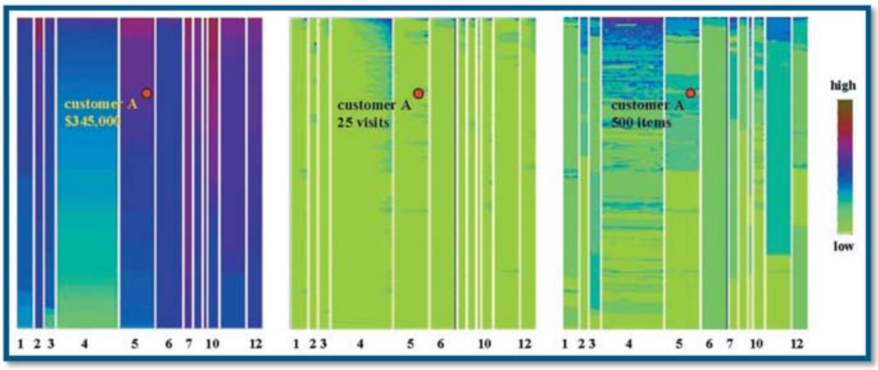
- **Small ▶ moderate ▶ large ▶ very large (huge)**
  - no VA needed for (really) small problems
  - lots of solutions do work for moderately sized cases
    // e.g., an Excel sheet of data (hundreds, thousands, ...)
  - selected solutions address large problems
    // tens to hundreds of thousands, etc.
  - very few focus on (really) huge problems
- Important difference between large and huge cases!

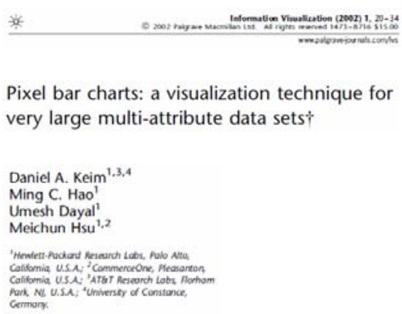

## C1: Problem Size – some categorization

- **Small ▶ moderate ▶ large ▶ very large (huge)**
  - no VA needed for (really) small problems
  - lots of solutions do work for moderately siz
    // e.g., an Excel sheet of data (hundreds, thous
  - selected solutions address large problems
    // tens to hundreds of thousands, etc.
  - very few focus on (really) huge problems







Information Visualization (2002) 1, 20–34
© 2002 Palgrave Macmillan Ltd. All rights reserved 1473-8716 $15.00
www.palgrave-journals.com/ivs

Pixel bar charts: a visualization technique for very large multi-attribute data sets†

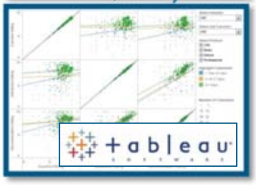
Daniel A. Keim[1,3,4]
Ming C. Hao[1]
Umesh Dayal[1]
Meichun Hsu[1,2]

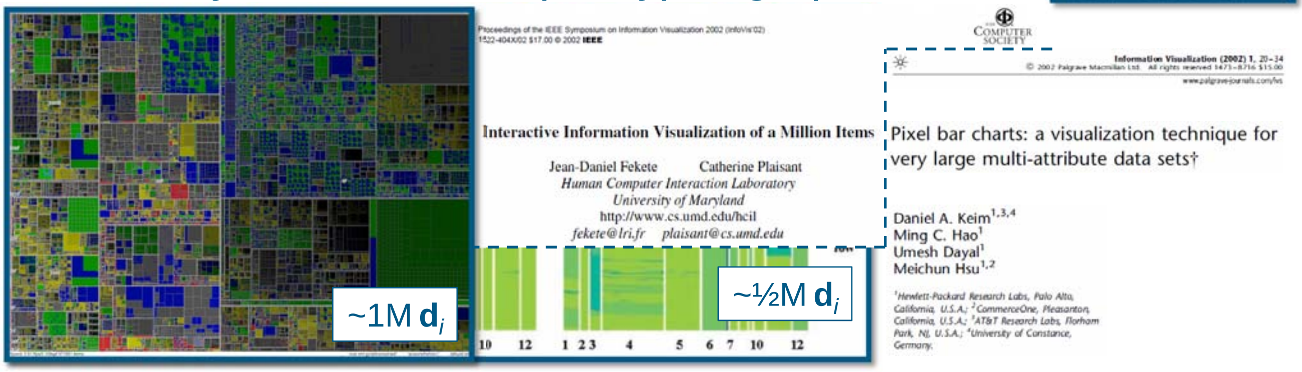
[1]Hewlett-Packard Research Labs, Palo Alto, California, U.S.A.; [2]CommerceOne, Pleasanton, California, U.S.A.; [3]AT&T Research Labs, Florham Park, NJ, U.S.A.; [4]University of Constance, Germany.


- Important difference between large and huge cases!

# C1: Problem Size – some categorization

- **Small ► moderate ► large ► very large (huge)**
  - no VA needed for (really) small problems
  - lots of solutions do work for moderately siz
    // e.g., an Excel sheet of data (hundreds, thous
  - selected solutions address large problems
    // tens to hundreds of thousands, etc.
  - very few focus on (really) huge problems

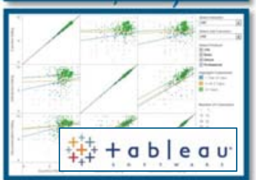




- Important difference between large and huge cases!

# C1: Problem Size – some categorization

- **Small ► moderate ► large ► very large (huge)**
  - no VA needed for (really) small problems
  - lots of solutions do work for moderately siz
    // e.g., an Excel sheet of data (hundreds, thous
  - selected solutions address large problems
    // tens to hundreds of thousands, etc.
  - very few focus on (really) huge problems



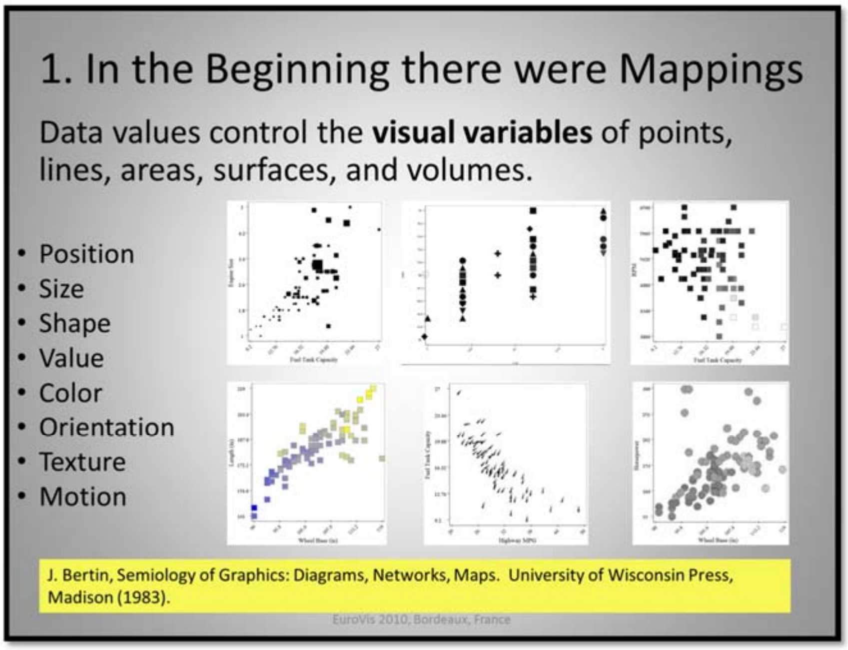
- Important difference between large and huge cases!

* Next **C2: Visualization Richness...**

# C2: Visualization

- The **perceptual** and **cognitive power** of users should not be left unutilized!
- Matt Ward, 2010:

EuroVis Keynote



# C2: Visualization

- The sho
- Mat

EuroVis Keynote

1. I

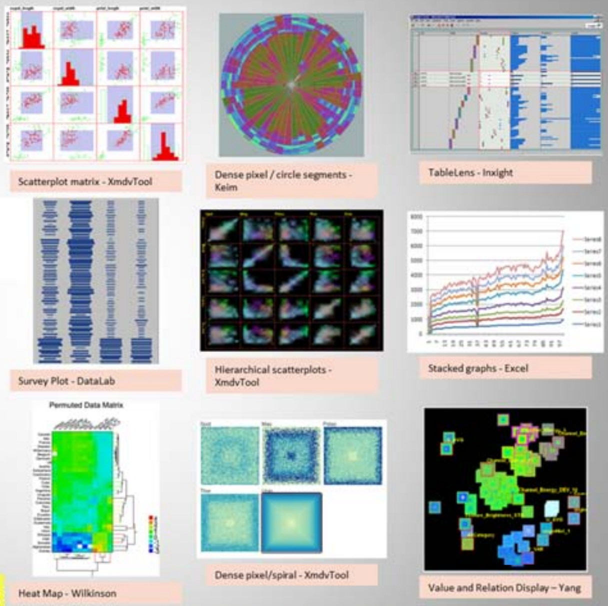
Data lines

- Positi
- Size
- Shape
- Value
- Color
- Orien
- Textu
- Motio

J. Bertin Madiso

## Dealing with Dimensions

- Many categorizations of dimension organization (see below paper for an early one)
- My categories:
  - **Subsetting** (e.g., SPLOMs, dense pixels)
  - **Reorganization** (e.g., parallel coords, glyphs)
  - **Embedding** (dimensional stacking, stacked bar charts, trellis displays)
  - **Reduction** (PCA, MDS, RadViz)




P. Wong and R. D. Bergeron,``30 years
Scientific Visualization: Overviews, Methodologies, and Techniques, edited by Nielson, Hagen, and Mueller (1994). pp. 3-33.

EuroVis 2010, Bordeaux, France

# C2: Visualization

- The sho
- Mat

EuroVis Keynote

1. I

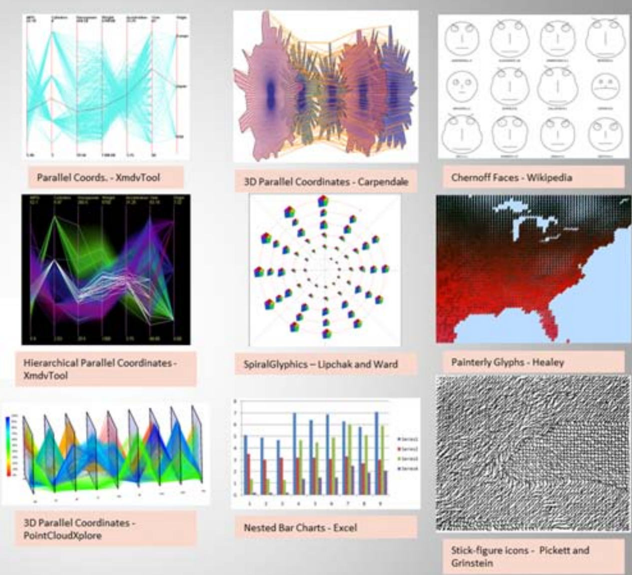
Data lines

- Positi
- Size
- Shape
- Value
- Color
- Orien
- Textu
- Motio

J. Bertin Madiso

## Dealing with Dimensions

- Many categorizations of dimension organization (see below paper for an early one)
- My categories:
  - **Subsetting** (e.g., SPLOMs, dense pixels)
  - **Reorganization** (e.g., parallel coords, glyphs)
  - **Embedding** (dimensional stacking, stacked bar charts, trellis displays)
  - **Reduction** (PCA, MDS, RadViz)




P. Wong and R. D. Bergeron,``30 years of multidimensional multivariate visualization.'' in Scientific Visualization: Overviews, Methodologies, and Techniques, edited by Nielson, Hagen, and Mueller (1994). pp. 3-33.

EuroVis 2010, Bordeaux, France

# C2: Visualization

- The
  sho
- Mat

EuroVis Keynote

1. I

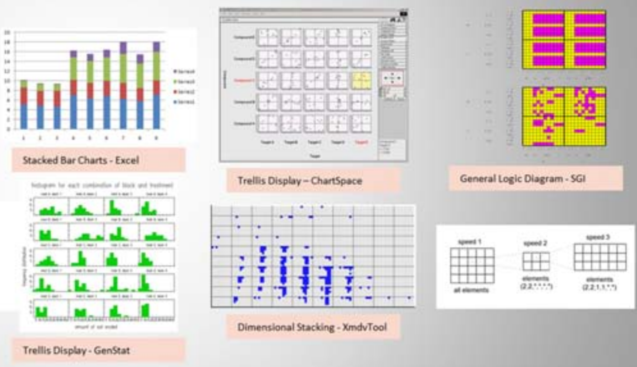
Data
lines

- Positi
- Size
- Shape
- Value
- Color
- Orien
- Textu
- Motio

J. Bertin
Madiso

## Dealing with Dimensions

- Many categorizations of dimension organization
  (see below paper for an early one)
- My categories:
  - **Subsetting**
    (e.g., SPLOMs, dense pixels)
  - **Reorganization**
    (e.g., parallel coords, glyphs)
  - **Embedding**
    (dimensional stacking, stacked bar charts, trellis displays)
  - **Reduction**
    (PCA, MDS, RadViz)




P. Wong and R. D. Bergeron,``30 years of multidimensional multivariate visualization.'' in Scientific Visualization: Overviews, Methodologies, and Techniques, edited by Nielson, Hagen, and Mueller (1994). pp. 3-33.

EuroVis 2010, Bordeaux, France

# C2: Visualization

- The
  sho
- Mat

EuroVis Keynote

1. I

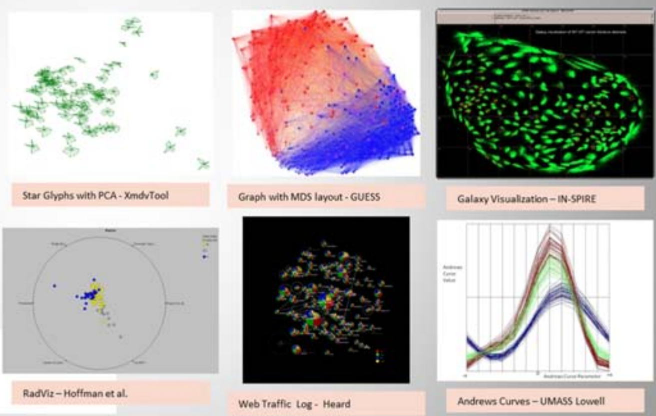
Data
lines

- Positi
- Size
- Shape
- Value
- Color
- Orien
- Textu
- Motio

J. Bertin
Madiso

## Dealing with Dimensions

- Many categorizations of dimension organization
  (see below paper for an early one)
- My categories:
  - **Subsetting**
    (e.g., SPLOMs, dense pixels)
  - **Reorganization**
    (e.g., parallel coords, glyphs)
  - **Embedding**
    (dimensional stacking, stacked bar charts, trellis displays)
  - **Reduction**
    (PCA, MDS, RadViz)



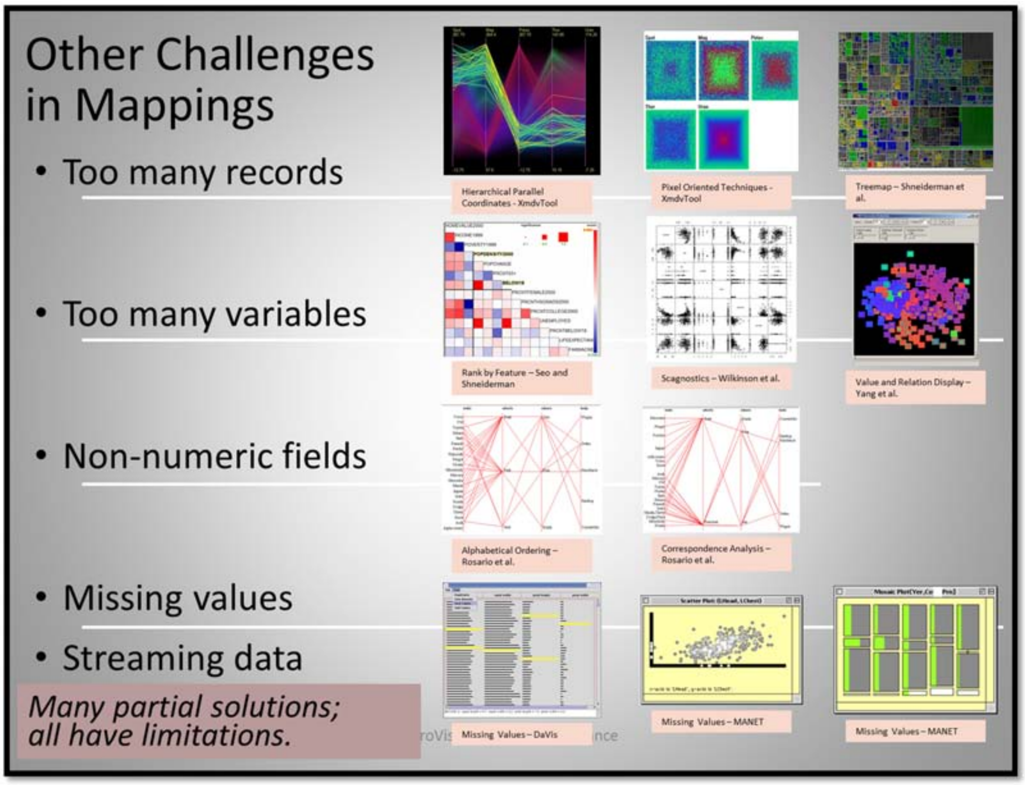

P. Wong and R. D. Bergeron,``30 years of multidimensional multivariate visualization.'' in Scientific Visualization: Overviews, Methodologies, and Techniques, edited by Nielson, Hagen, and Mueller (1994). pp. 3-33.

EuroVis 2010, Bordeaux, France

## C2: Visualization – more...

- Also from Matt Ward's talk:

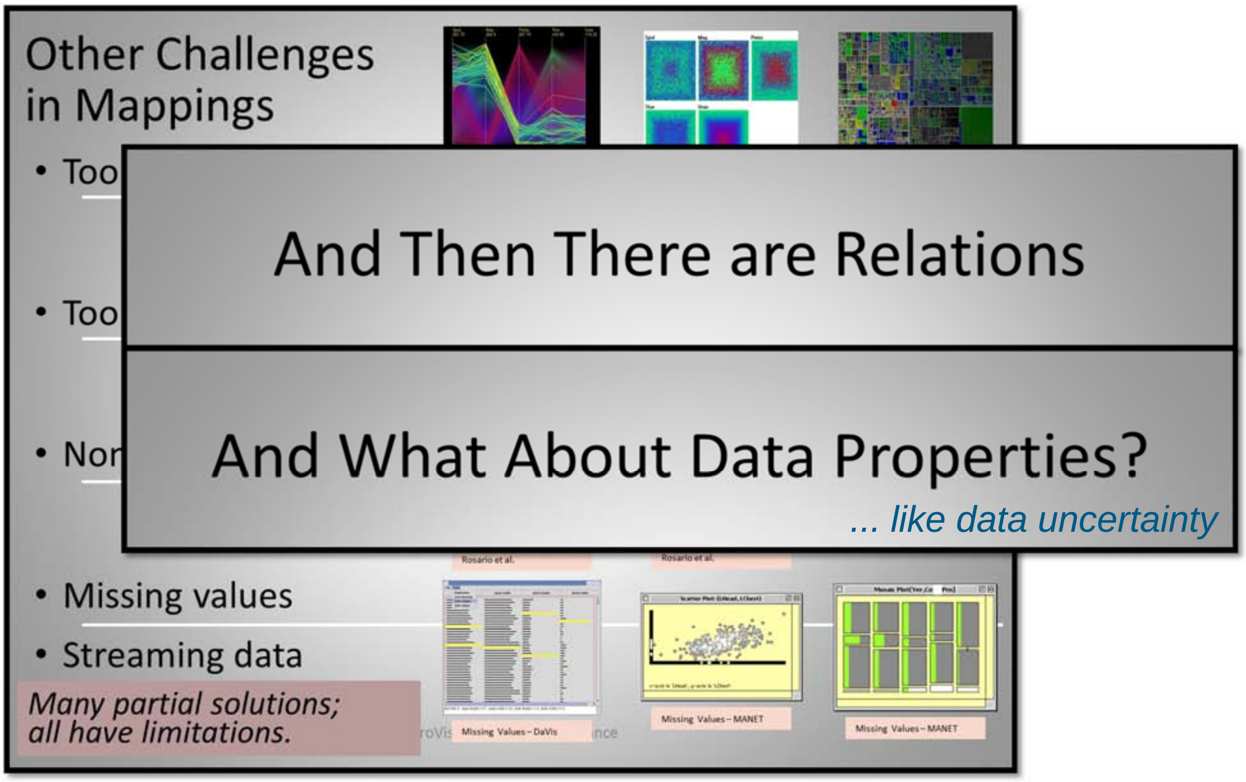
EuroVis 2010 Keynote




## C2: Visualization – more...

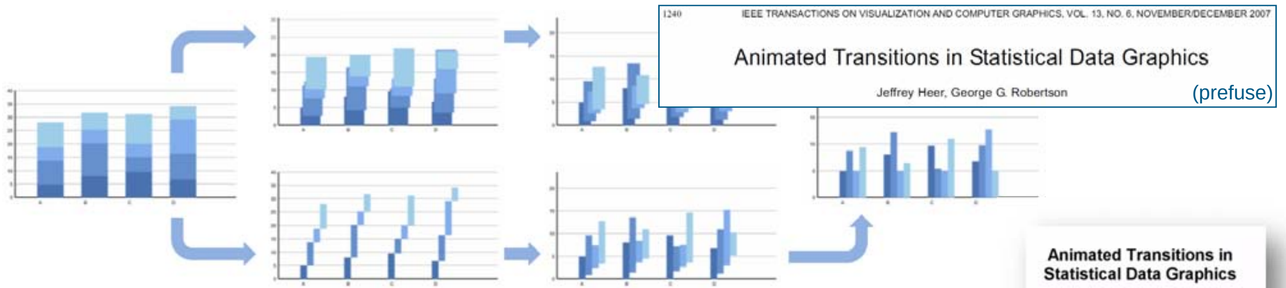
- Also from Matt Ward's talk:

EuroVis 2010 Keynote



# C2: Visualization Richness

- **None / primitive ► advanced ► rich**
  - none or at least some primitive vis (bar charts, etc.) are the minimum – state-of-the-art, in particular outside visualization
  - advanced: state-of-the-art wrt. visualization, in particular ***selected*** advanced visualization
  - rich: an extensive spectrum of available vis. – there is a choice of various advanced vis. techniques

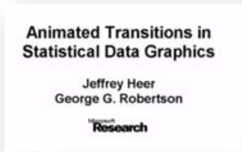


1240 IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 13, NO. 6, NOVEMBER/DECEMBER 2007

Animated Transitions in Statistical Data Graphics

Jeffrey Heer, George G. Robertson

(prefuse)




- Rich: often through library-sort of solutions

# After Mapping Comes Interaction

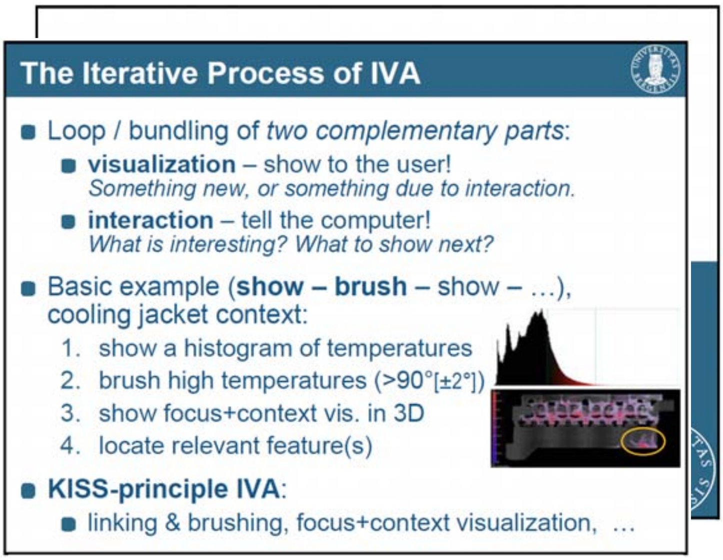
Visualization without interaction
is like a sports car with no engine!

Nice to look at,
but not good for much! ☺



# C3: Interaction

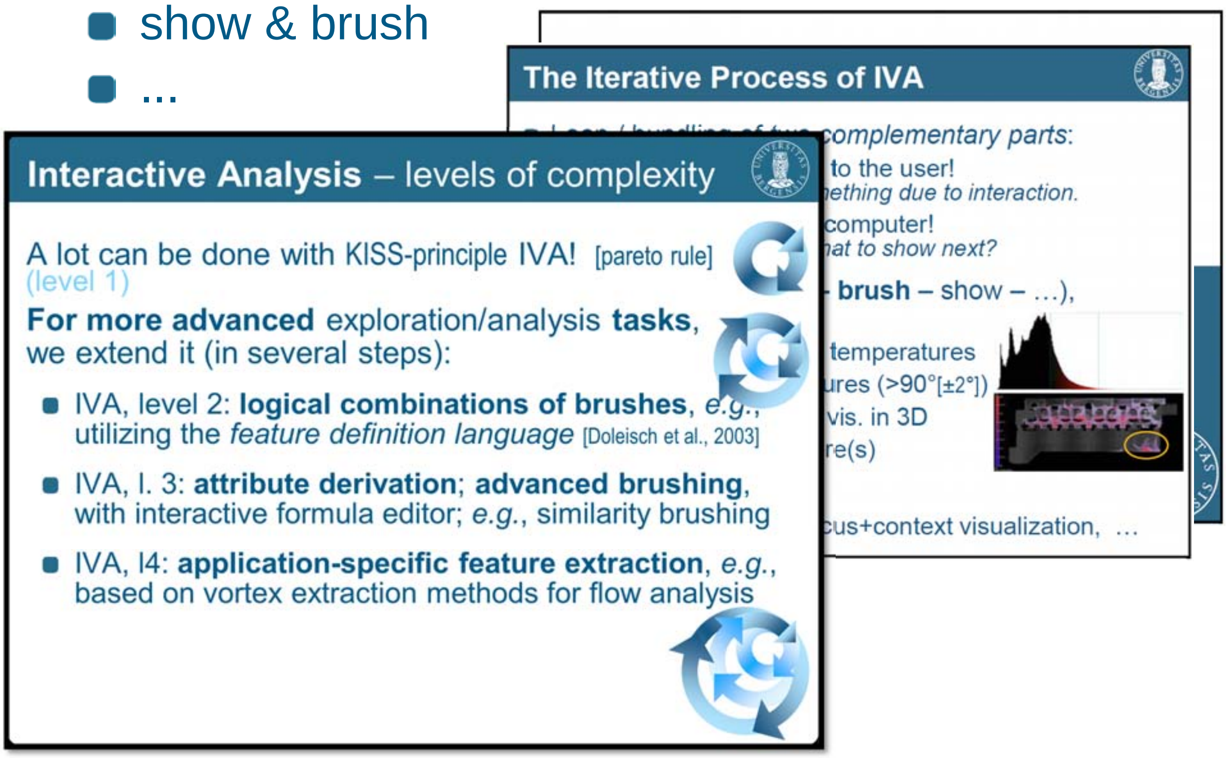
- Making VA a **loop**, an **interactive visual dialog**, like
  - show & brush
  - ...




HH, EuroVA 2012 Keynote

## C3: Interaction

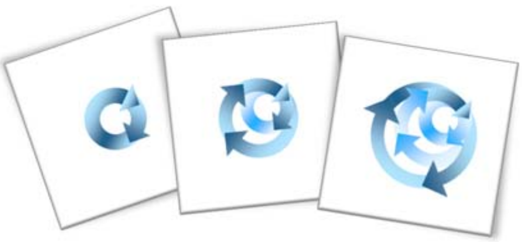
- Making VA a **loop**, an **interactive visual dialog**, like
  - show & brush
  - ...




HH, EuroVA 2012 Keynote

## C3: Interaction

- Making VA a **loop**, an **interactive visual dialog**, like
  - show & brush
  - ...
- A really important question is: **how fast** is **one such loop**?
- Jean-Daniel Fekete, 2012:



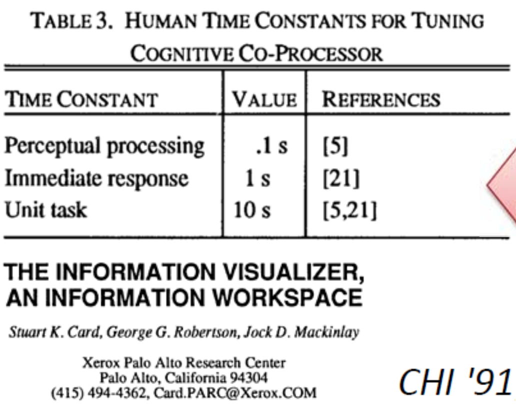


TABLE 3. HUMAN TIME CONSTANTS FOR TUNING COGNITIVE CO-PROCESSOR

| TIME CONSTANT | VALUE | REFERENCES |
|---|---|---|
| Perceptual processing | .1 s | [5] |
| Immediate response | 1 s | [21] |
| Unit task | 10 s | [5,21] |

THE INFORMATION VISUALIZER, AN INFORMATION WORKSPACE

*Stuart K. Card, George G. Robertson, Jock D. Mackinlay*

Xerox Palo Alto Research Center
Palo Alto, California 94304
(415) 494-4362, Card.PARC@Xerox.COM

*CHI '91*

### Response Times

- 0.1 sec - animation, visual continuity, sliders
- 1 sec - system response, conversation break
- 10 sec - cognitive response

Stuart K. Card, George G. Robertson, Jock D. Mackinlay. The information visualizer, an information workspace. *Proc. CHI '91*, 181-186, 1991.

- Beyond 20 sec, users wait and loose attention
  - Forget their goals and plans
  - **Progress bar needed!**

Dagstuhl Seminar Talk

# C3: Interaction

Dagstuhl Seminar Talk

- Jean-Daniel Fekete, 2012:

## Interaction / Cognition

- Main issue: managing short-term memory efficiently
  - 7 items ±2
  - Needed to maintain planning, hypotheses
- Typing/language use several items
  - Scripting / SQL interfere with exploration
- Other issue: avoid distracting, understanding is cognitively demanding!

# C3: Interaction

Dagstuhl Seminar Talk

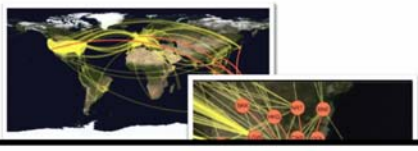
- Jean-Daniel Fekete, 2012:

## Examples of Interac

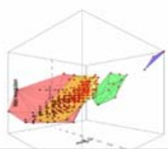
- View-level
  - Topology-Aware Navigation in Larg
- Visual-form level
  - ScatterDice, GraphDice
- Data level
  - Dynamic Queries, Brushing His
- Machine Learning Level

### Topology-Aware Navigation in Large Networks

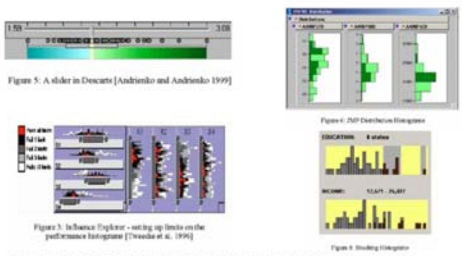
T. Moscovich, F. Chevalier, N. Henry Riche, E. Pietriga, J.-D. Fekete, ACM CHI 09



### Visual Form Interaction: ScatterDice



### Brushing Histograms

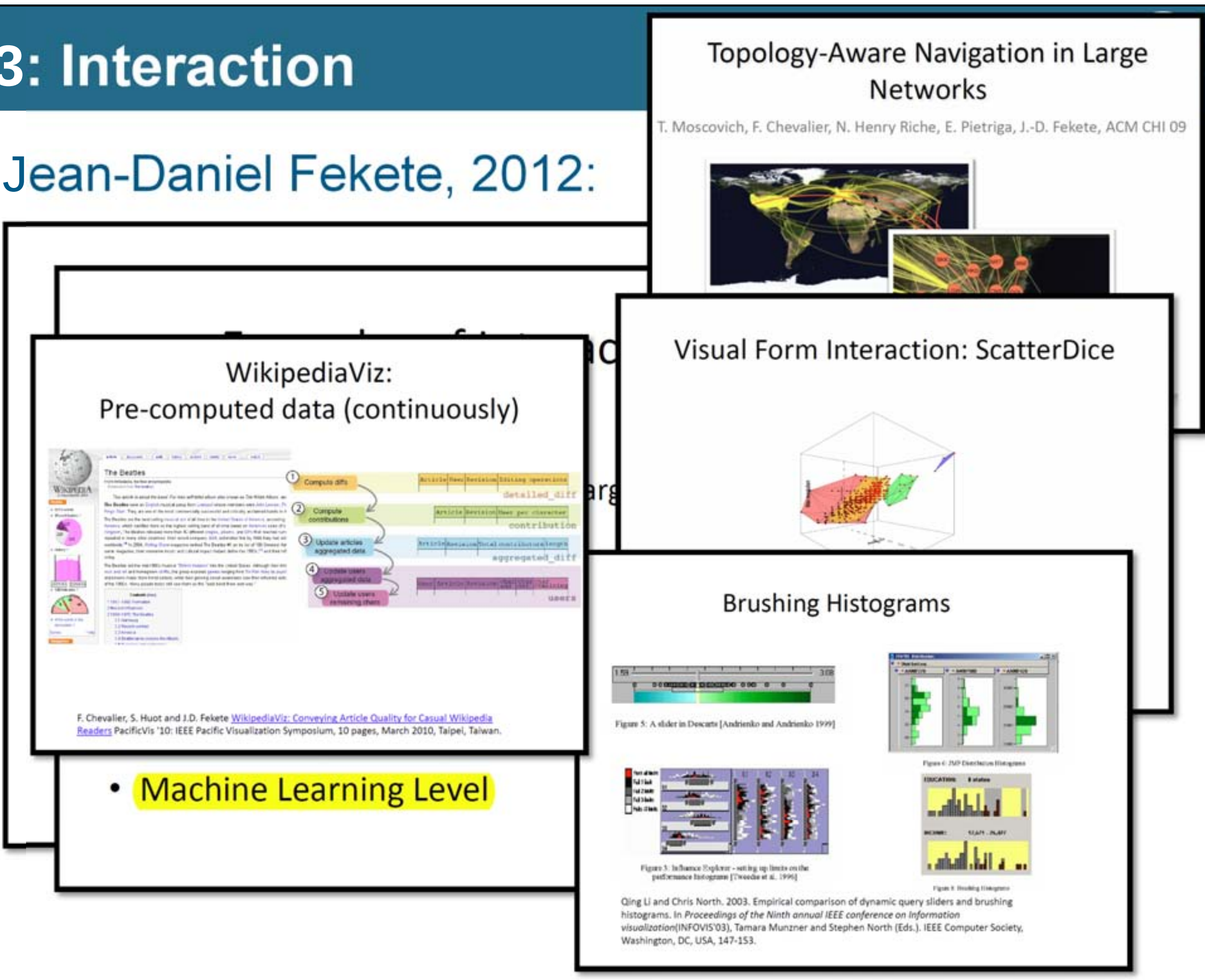


Qing Li and Chris North. 2003. Empirical comparison of dynamic query sliders and brushing histograms. In *Proceedings of the Ninth annual IEEE conference on Information visualization*(INFOVIS'03), Tamara Munzner and Stephen North (Eds.). IEEE Computer Society, Washington, DC, USA, 147-153.

# C3: Interaction

- Jean-Daniel Fekete, 2012:

Dagstuhl Seminar Talk




# C3: Interaction Pace

- **Separate ► unit task ► immediate ► continuous**
  - separate: offline processing
  - unit task [Card et al., '91]: ≈10s – before attention breaks!
  - immediate: ≈1s – maintains an interplay, a conversation
  - continuous: ≈0.1s – smooth in the eye (perception)

***The unit task time constant.*** Finally, we seek to make it possible for the user to complete some elementary task act within 10 sec (say, 5~30 sec) [5,21], about the pacing of a point and click editor. Information agents may require considerable time to complete some complicated request, but the user, in this paradigm, always stays active. He or she can begin the next request as soon as sufficient information has developed from the last or even in parallel with it.

***The immediate response time constant.*** A person can make an unprepared response to some stimulus within about a second [21]. If there is more than a second, then either the listening party makes a backchannel response to indicate that he his listening (e.g., "uh-huh") or the speaking party makes a response (e.g., "uh...") to indicate he is still thinking of the next speech. These serve to keep the parties of the interaction informed that they are still engaged in an interaction. In the Cognitive Co-processor, we attempt to have agents provide status feedback at intervals no longer than this constant. Immediate response animations (e.g., swinging the branches of a 3D tree into view) are designed to take about a second. If the time were much shorter, then the user would lose object constancy and would have to reorient himself. If they were much longer, then the user would get bored waiting for the response.

***The perceptual processing time constant.*** The Cognitive Co-processor is based on a continuously-running scheduler loop and double-buffered graphics. In order to maintain the illusion of animation in the world, the screen must be repainted at least every .1 sec [5]. The Cognitive Co-processor therefore has a *Governor* mechanism that monitors the basic cycle time. When the cycle time becomes too high, cooperating rendering processes reduce the quality of rendering (e.g., leaving off most of the text during motion) so that the cycle speed is increased.

- Really important differences on the user side!

## * Sure, **C4: Computational Analysis...**

## C4: Computational Analysis

- VA is about the **integration** of **interactive visual** analysis means **and computational analysis**

**Humans and Computers**

*"Computers are incredibly fast, accurate, and stupid;*
*humans are incredibly slow, inaccurate, and brilliant;*
*together they are powerful beyond imagination."*

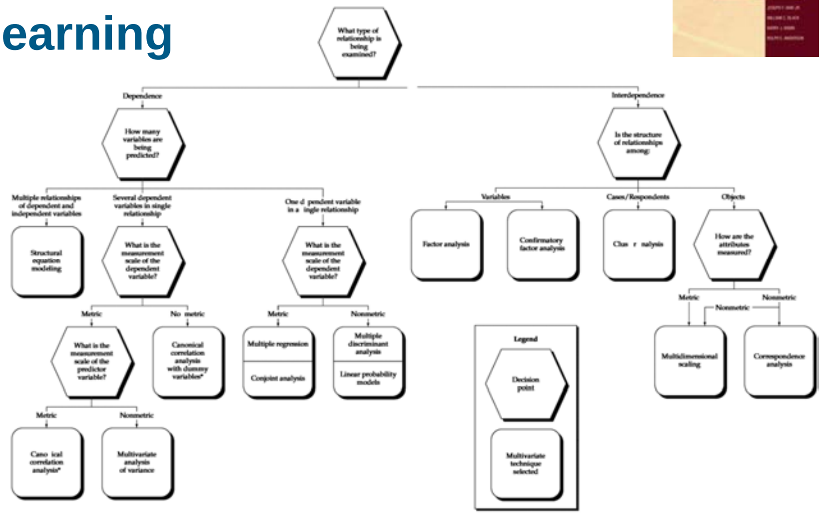
attributed to Albert Einstein

D. Keim, F. Mansmann | Dagstuhl Seminar 12081 | Information Visualization, Visual Data Mining and Machine Learning | 7

D. Keim, Dagstuhl Seminar Talk, 2012

# C4: Computational Analysis

- VA is about the **integration** of **interactive visual** analysis means **and computational analysis**
- So then: computational analysis...
  - from **statistics**
  - from **data mining**
  - from **machine learning**
  - from ...

# C4: Computational Analysis

- **None ► selected ► rich**
  - vis.-and-only-vis. solutions do not offer comp. analysis

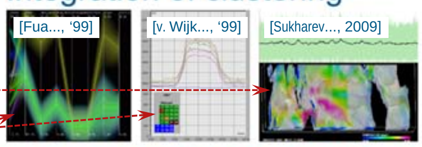
HH, Dagstuhl Seminar, 2012

## Integrated Methods

- Clustering
  - k-means
  - hierarchical clustering me
  - etc.
- Projections (embeddings), e.g.,
  - PCA
  - MDS
  - etc. (SOM)
- Classification, regression
  - decision trees
  - SVM
  - etc. (gen. alg.)
- Etc.

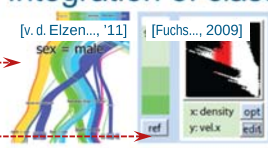
## Some Examples

- Integration of clustering




- Integration of projection/embedding

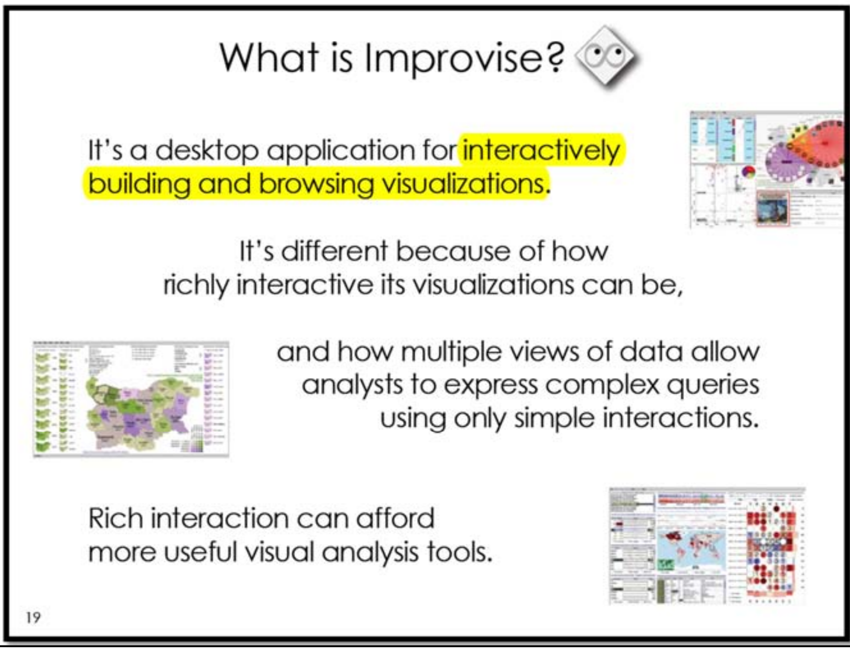

- Integration of classifcation/learning



* Last, here, **C5: Comprehensiveness…**

## C5: Comprehensiveness

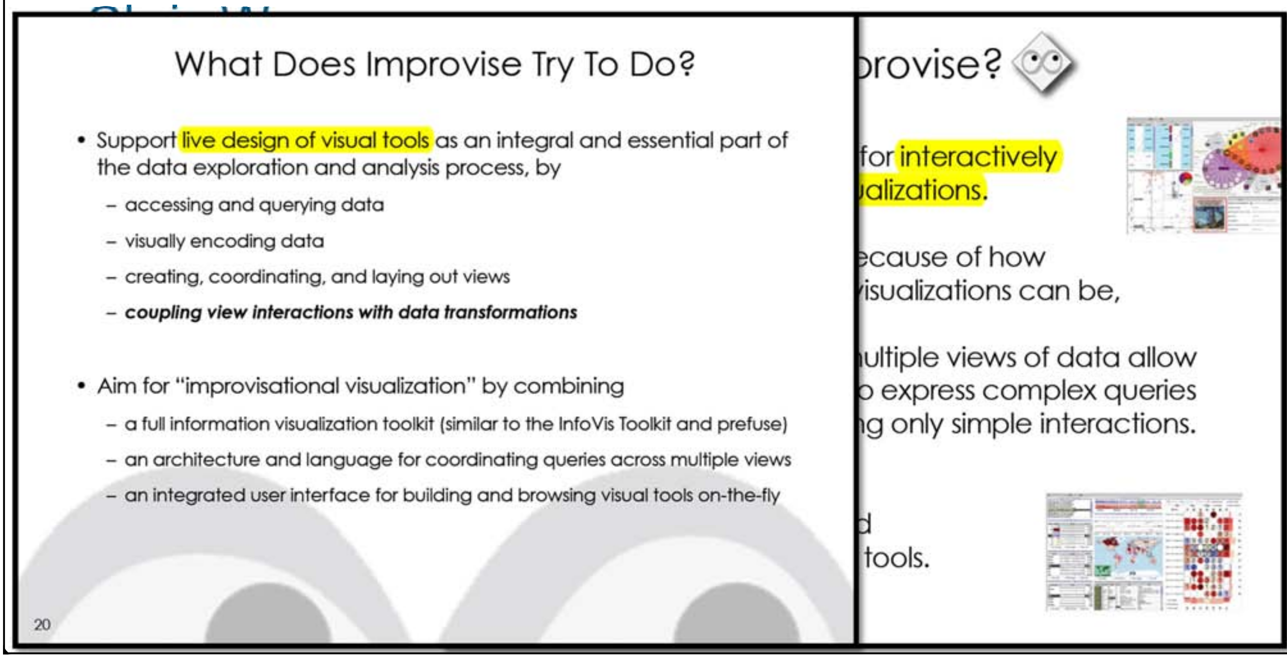
- How **specialized** is the VA solution?
  Does it cover **heterogenous** aspects?
  Is the VA solution **open**?
- Chris Weaver, 2011:




Oklahoma State, CS Colloquium, 2011

# C5: Comprehensiveness

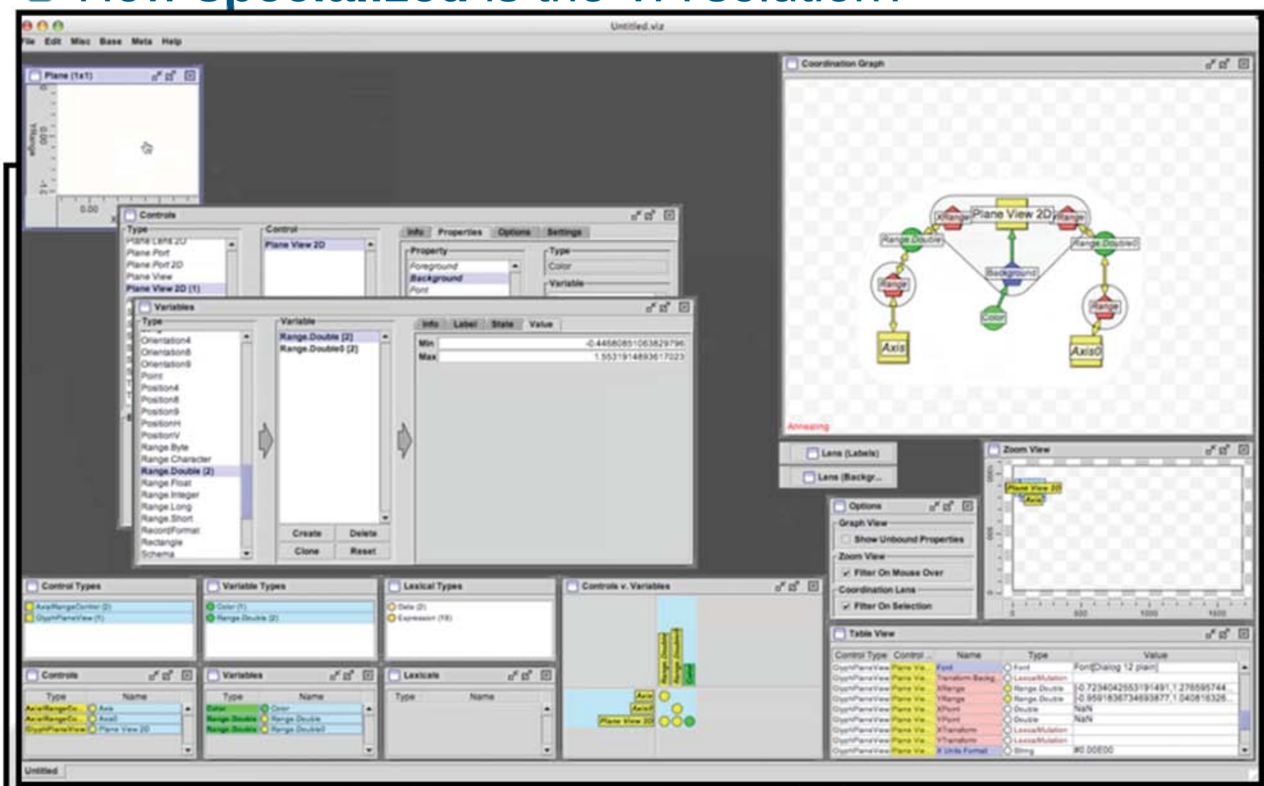
- How **specialized** is the VA solution?
  Does it cover **heterogenous** aspects?
  Is the VA solution **open**?




Oklahoma State, CS Colloquium, 2011

# C5: Comprehensiveness

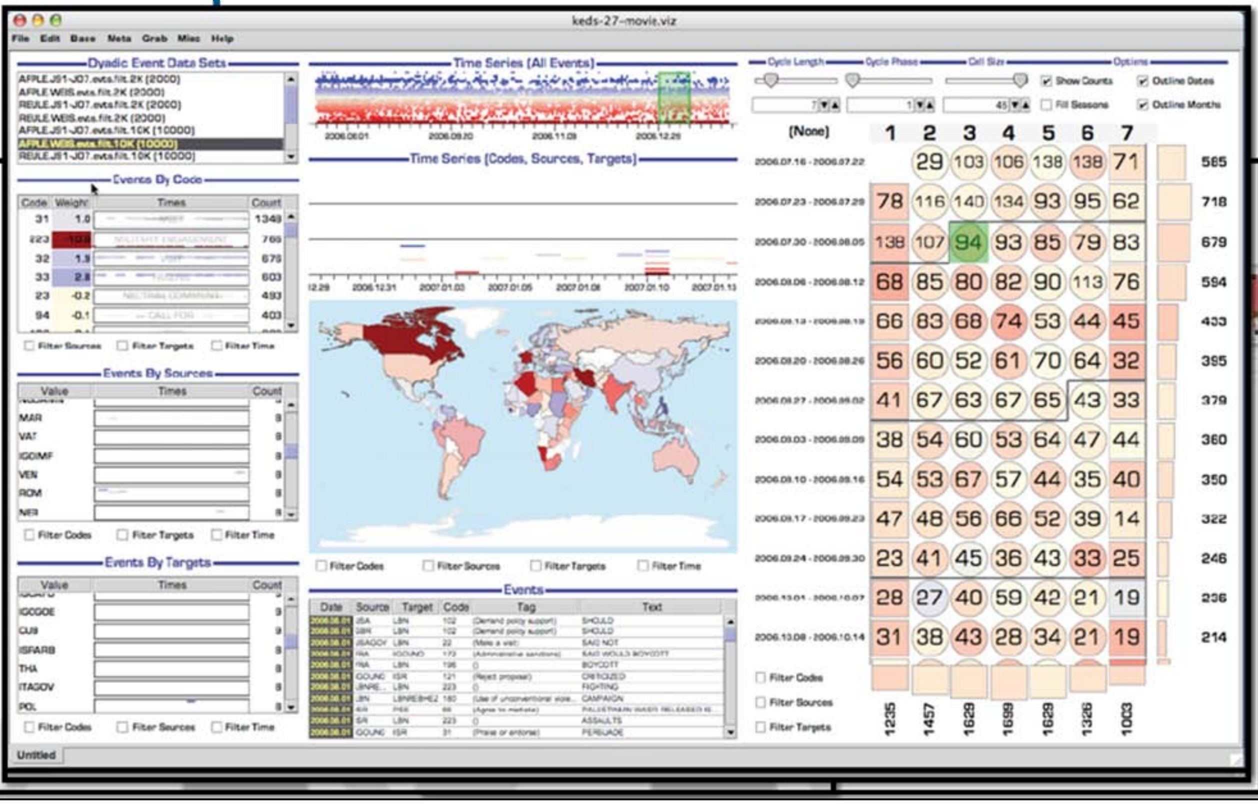
- How **specialized** is the VA solution?




Oklahoma State, CS Colloquium, 2011

# C5: Comprehensiveness

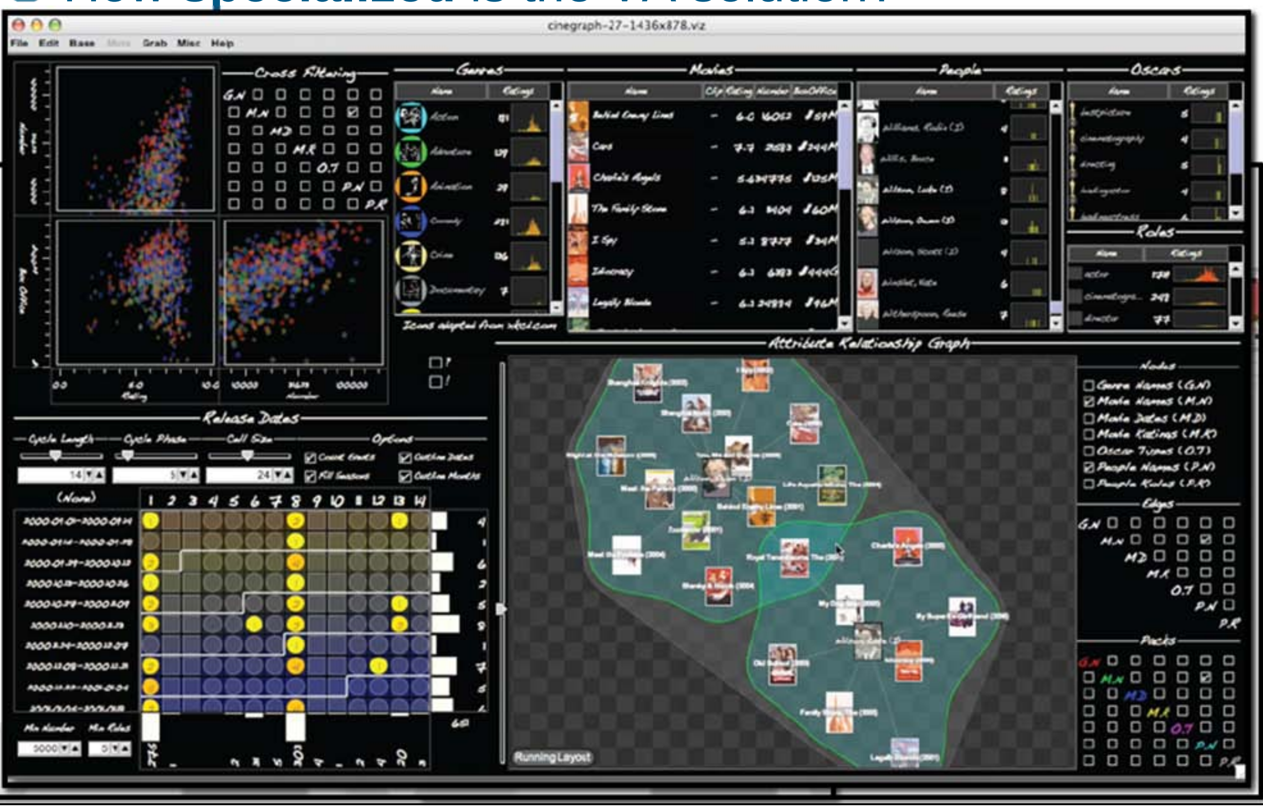
- How **specialized** is the VA solution?




Oklahoma State, CS Colloquium, 2011

# C5: Comprehensiveness

- How **specialized** is the VA solution?




Oklahoma State, CS Colloquium, 2011

## C5: Comprehensiveness

- **Targeted ► semi-flexible ► open**
  - targeted: one specific problem context, tailored / optimized solution
  - semi-flexible: general wrt. a certain type of problem
  - open: broad variety of problems, also broad variety of problem aspects (can treat heterogeneous problems)

## Five Characteristics – summary

C1: **problem size**
small ▶ moderate ▶ large ▶ very large (huge)

C2: **visualization richness**
none / primitive ▶ advanced ▶ rich

C3: **interaction pace**
separate ▶ unit task ▶ immediate ▶ continuous

C4: **computational analysis**
none ▶ selected ▶ rich

C5: **comprehensiveness**
targeted ▶ semi-flexible ▶ open

## Five Characteristics – summary

C1: **problem size**
small ▶ moderate ▶ large ▶ very large (huge)

C2: **visualization richness**
none / primitive ▶ advanced ▶ rich

C3: **interaction pace**
separate ▶ unit task ▶ immediate ▶ continuous

C4: **computational analysis**
none ▶ selected ▶ rich

C5: **comprehensiveness**
targeted ▶ semi-flexible ▶ open

## Five Characteristics – summary

C1: **problem size**
small ▶ moderate ▶ large ▶ very large (huge)
▶ distinguish between large and huge!

C2: **visualization richness**
none / primitive ▶ advanced ▶ rich
▶ different views for different purposes!

C3: **interaction pace**
separate ▶ unit task ▶ immediate ▶ continuous
▶ respect the *human* time constants!

C4: **computational analysis**
none ▶ selected ▶ rich
▶ huge potential for VA!

C5: **comprehensiveness**
targeted ▶ semi-flexible ▶ open
▶ so many VA cases are, in fact, heterogeneous!

approaches to VA: ▶ from visualization, adding analysis
▶ from analysis, adding visualization

## Acknowledgements

- **You!**
- Daniel Keim, Thomas Ertl, Christian Chabot, Matt Ward, Jean-Daniel Fekete, Chris Weaver, …
- Çağatay Turkay
- **Questions?**
Discussion?