VISUAL ANALYTICS

Hans-Jörg Schulz

AARHUS
UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

1. INTRODUCTION

2. VISUAL ENCODING

3. BASIC CHART TYPES

Basics

4. INTERACTION

5. VISUALIZATION DESIGN

6. DATA PREPROCESSING

Visualization Building Blocks & Processes

7. RECAP 1st Half

8. MULTIVARIATE DATAVIS

9. TEMPORAL DATAVIS

10. GEOSPATIAL DATAVIS

11. GRAPH DATAVIS

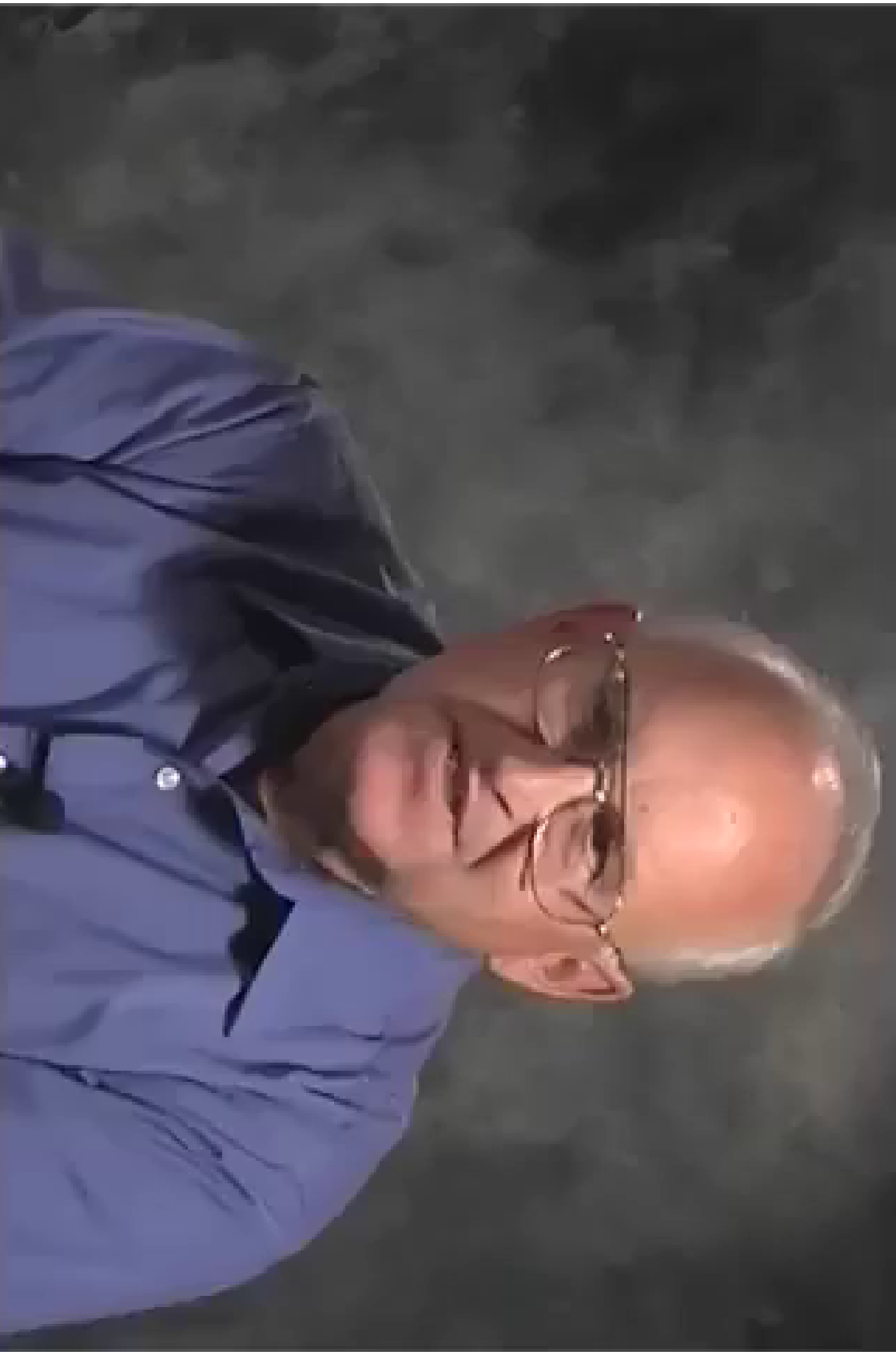
Visualization Techniques

12. 3D DATAVIS

13. VISUAL ANALYTICS

Visualization Applications

14. RECAP 2nd Half

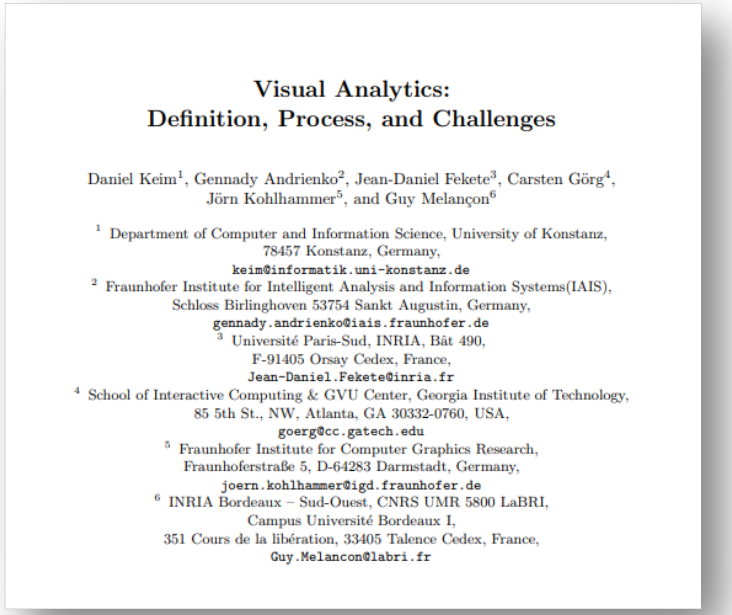


video source: https://youtu.be/5uGRGqCFryg

VISUAL ANALYTICS

"Visual analytics combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex datasets."

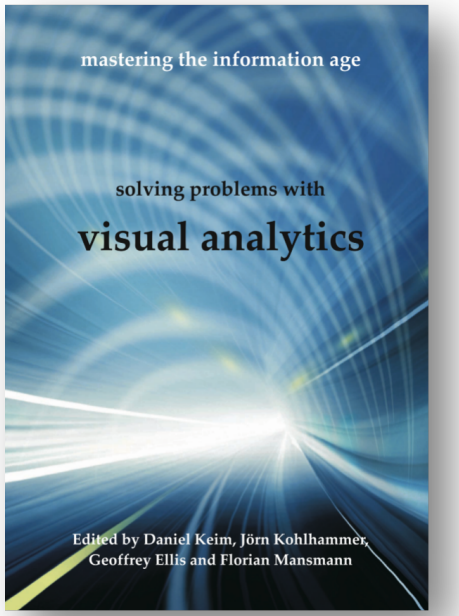
Keim et al. 2008



Visual Analytics:
Definition, Process, and Challenges

Daniel Keim[1], Gennady Andrienko[2], Jean-Daniel Fekete[3], Carsten Görg[4], Jörn Kohlhammer[5], and Guy Melançon[6]

[1] Department of Computer and Information Science, University of Konstanz, 78457 Konstanz, Germany, keim@informatik.uni-konstanz.de
[2] Fraunhofer Institute for Intelligent Analysis and Information Systems(IAIS), Schloss Birlinghoven 53754 Sankt Augustin, Germany, gennady.andrienko@iais.fraunhofer.de
[3] Université Paris-Sud, INRIA, Bât 490, F-91405 Orsay Cedex, France, Jean-Daniel.Fekete@inria.fr
[4] School of Interactive Computing & GVU Center, Georgia Institute of Technology, 85 5th St., NW, Atlanta, GA 30332-0760, USA, goerg@cc.gatech.edu
[5] Fraunhofer Institute for Computer Graphics Research, Fraunhoferstraße 5, D-64283 Darmstadt, Germany, joern.kohlhammer@igd.fraunhofer.de
[6] INRIA Bordeaux – Sud-Ouest, CNRS UMR 5800 LaBRI, Campus Université Bordeaux I, 351 Cours de la libération, 33405 Talence Cedex, France, Guy.Melancon@labri.fr



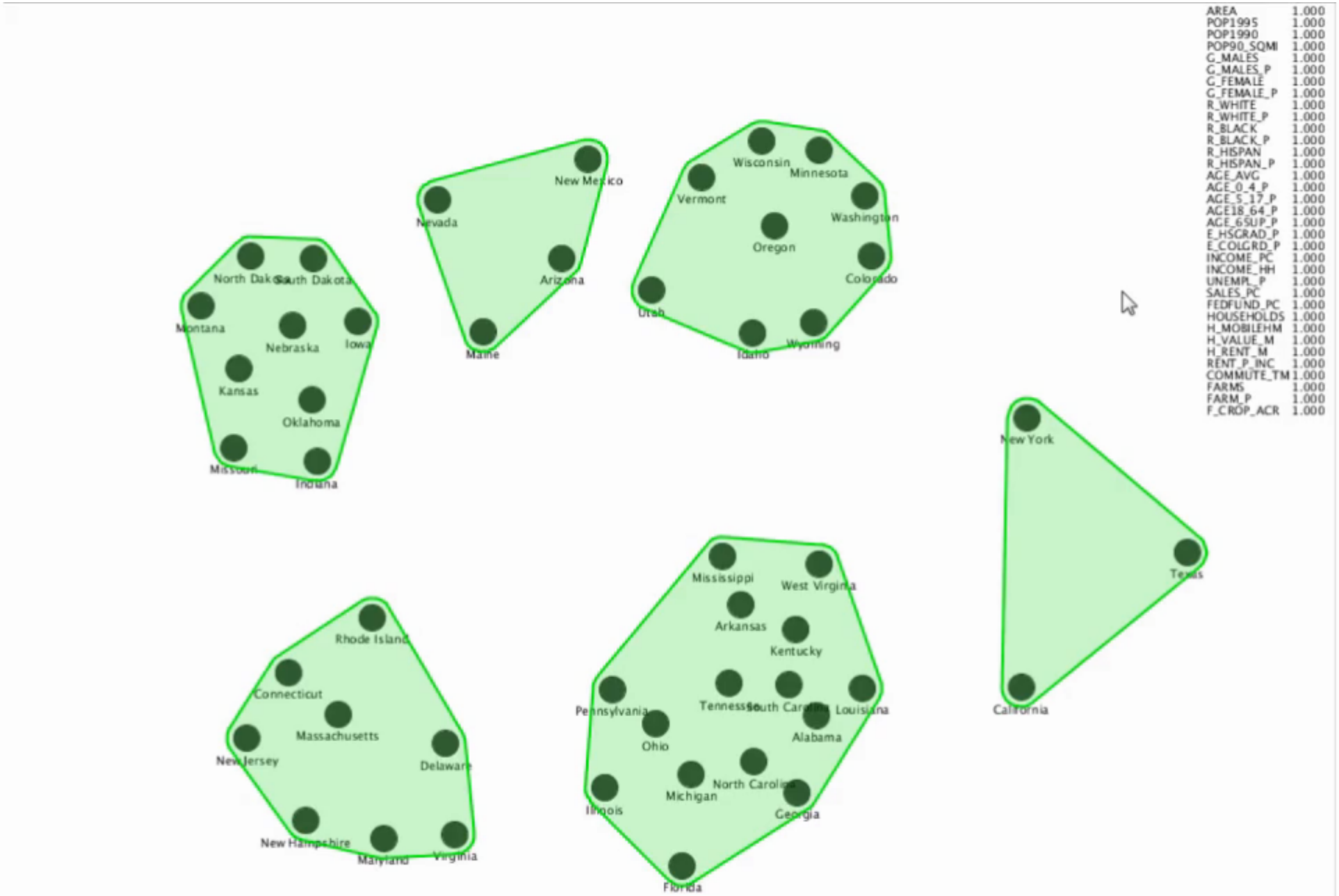
[https://doi.org/10.1007/978-3-540-70956-5_7]

[https://doi.org/10.2312/14803]

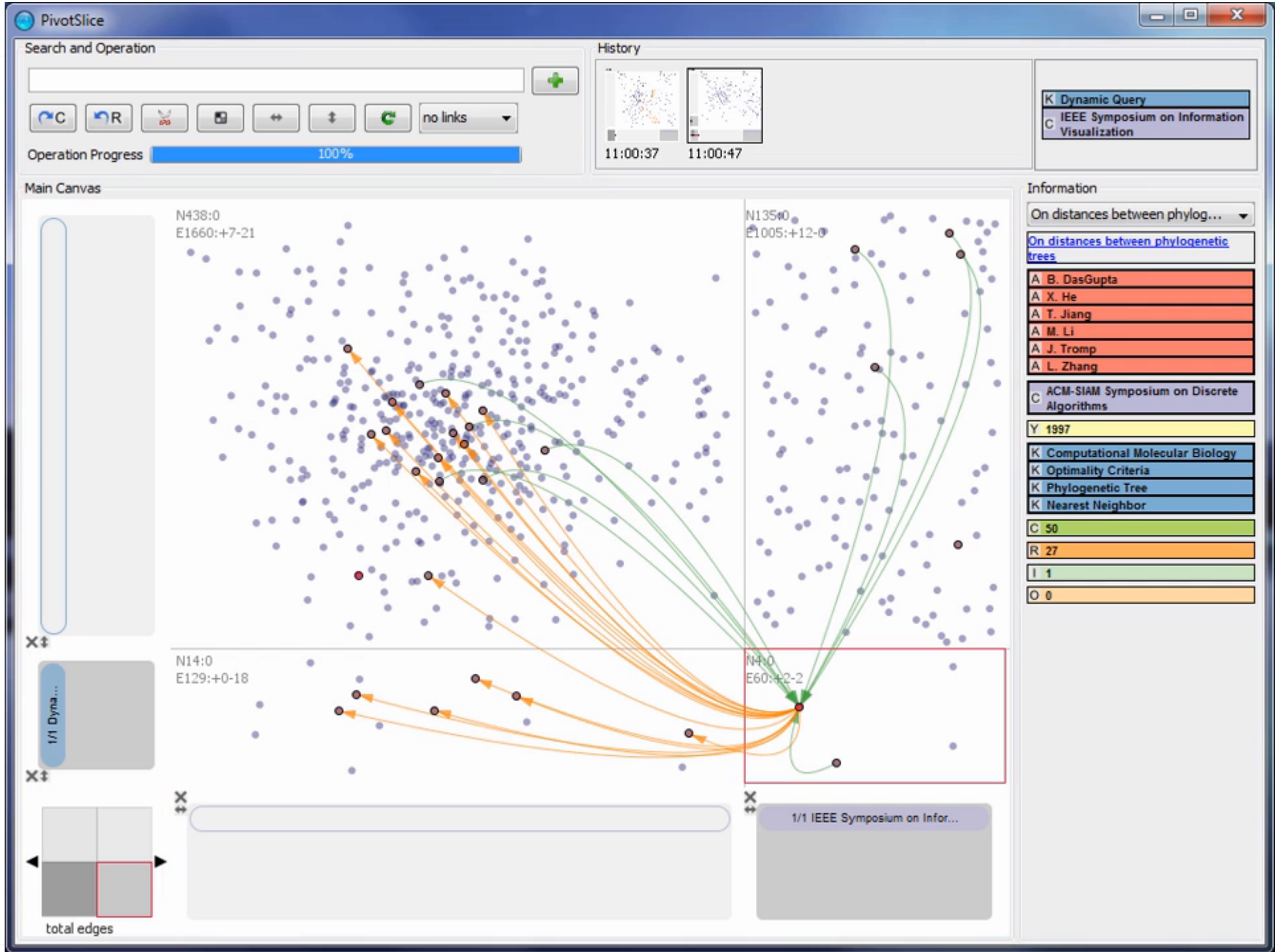
WHY TALK ABOUT VISUAL ANALYTICS?

VA comes with a series of challenges in its own right:

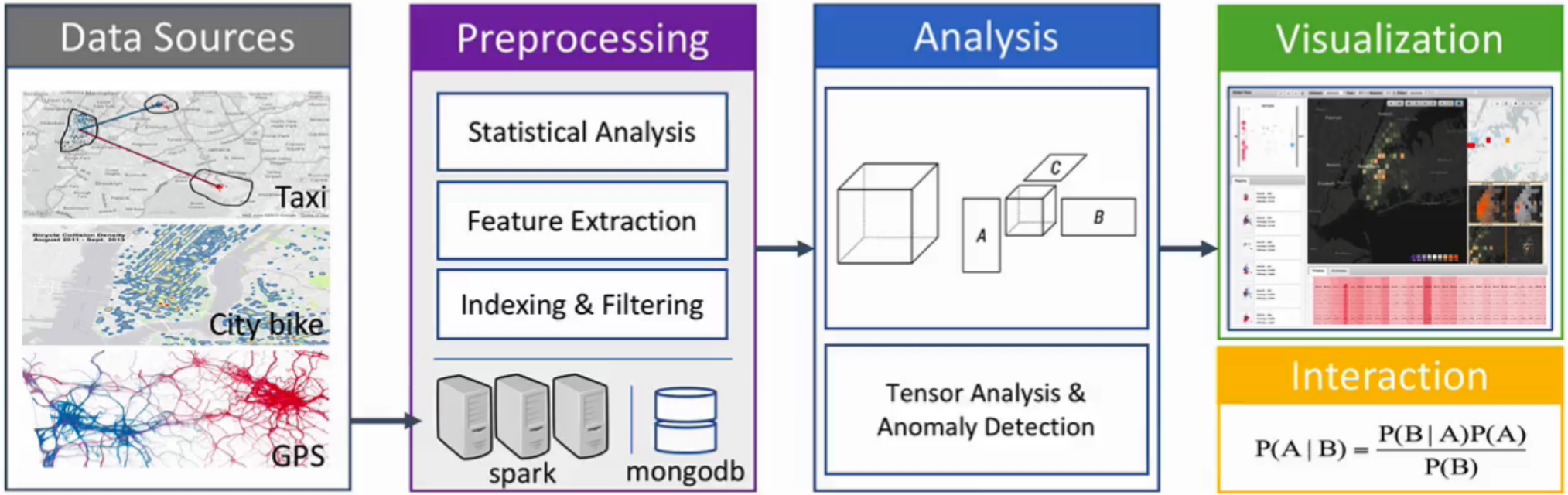
- Computational analysis is often a long-running, time consuming process, while visual data analysis is fast-paced and interactive.
- Visualization must not only be expressive of the underlying data, but the underlying computational processes
- The aim of VA is
  - ill-defined: “generate insight in the data”
  - subjective: “confirm the expected and discover the unexpected”
  - elusive / iterative: the moment you reach an analysis goal, it shifts…
  - constrained: results must be ready to base timely decisions on them



https://doi.org/10.1109/VDS48975.2019.8973381



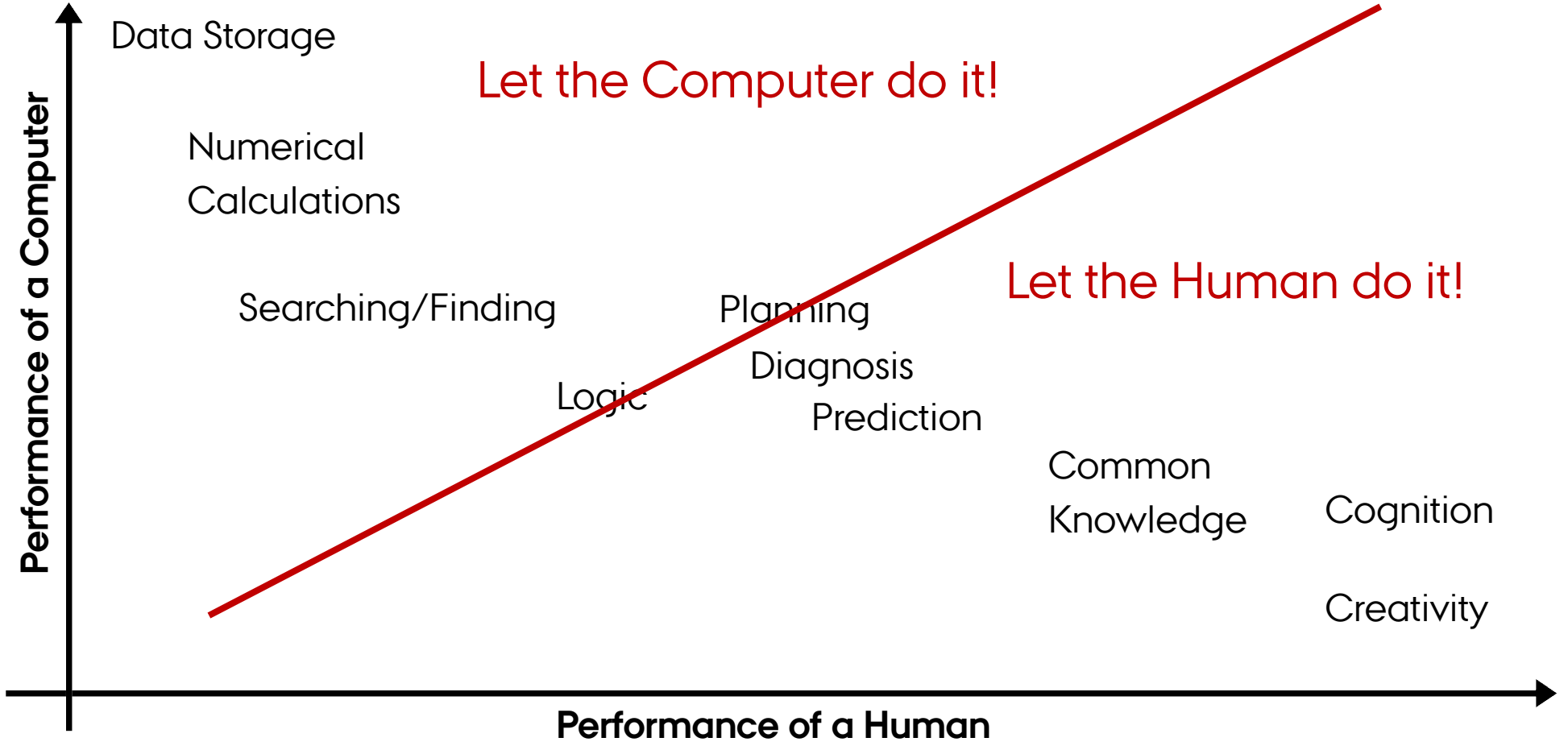
https://doi.org/10.1109/TVCG.2013.167



https://doi.org/10.1109/TVCG.2017.2744419

THEORETICAL BACKGROUND

KEIM'S ABILITY MATRIX



adapted from Daniel Keim, Uni. Konstanz
[https://bib.dbvis.de/uploadedFiles/233.pdf]

YET ANOTHER VA DEFINITION...

"Visual analytics combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex datasets."

Keim et al. 2008

[https://doi.org/10.1007/978-3-540-70956-5_7]

"Visual analytics combines the strengths of both worlds: On the one hand they take advantage of intelligent algorithms and vast computational power of modern computers and on the other hand they integrate human background knowledge and intuition to find a good solution."

Keim et al. 2009

[https://doi.org/10.1145/1809400.1809403]

INFORMATION FORAGING

- We are “Informavores”. (see George Miller: “Informavores” in F. Machlup & U. Mansfield (Eds.), The Study of Information: Interdisciplinary Messages (pp. 111-113) Wiley 1983)

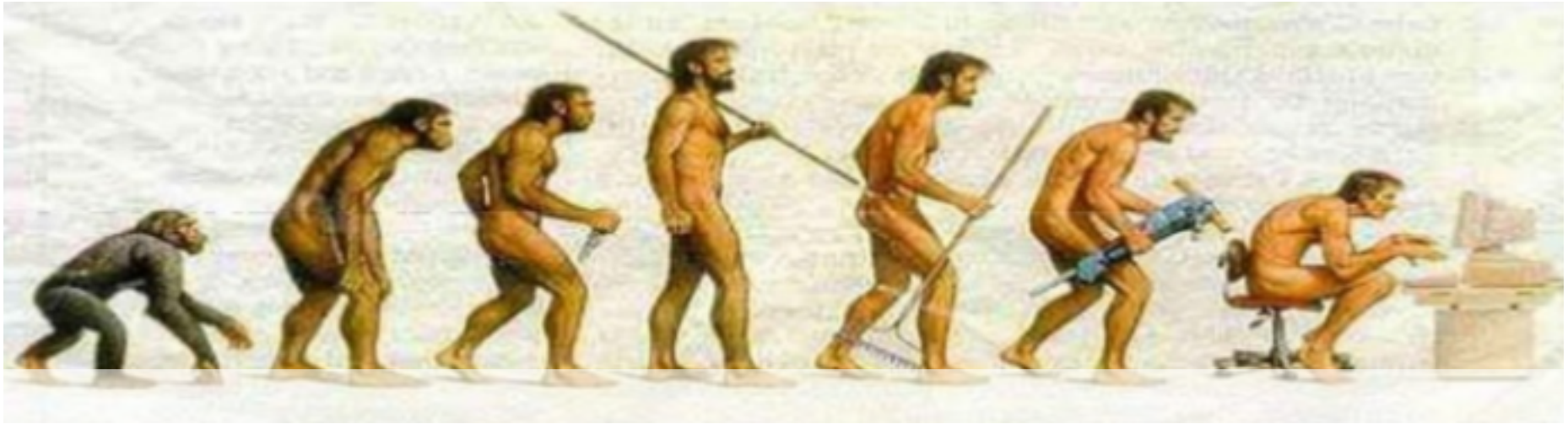


image source: Catherine Tiphanie, http://www.catherinetiphanie.co.uk/ergonomics.html

- We “hunt” for information in similar ways as predators hunt for prey. (see Pirolli+Card, 1999) https://doi.org/10.1037/0033-295X.106.4.643

INFORMATION FORAGING

ANIMAL FORAGING		INFORMATION FORAGING	
Food	**Goal**	Information	
A site containing one or more potential sources of food	**Patch**	A website (or other source of information)	
Search for food	**Forage**	Search for information	
The animal's assessment of how likely it is that a given patch will provide food	**Scent**	How promising a potential source of information appears to the user	
The totality of food types that an animal may consider in order to satisfy hunger	**Diet**	The totality of the information sources that a user may consider in order to satisfy an information need	

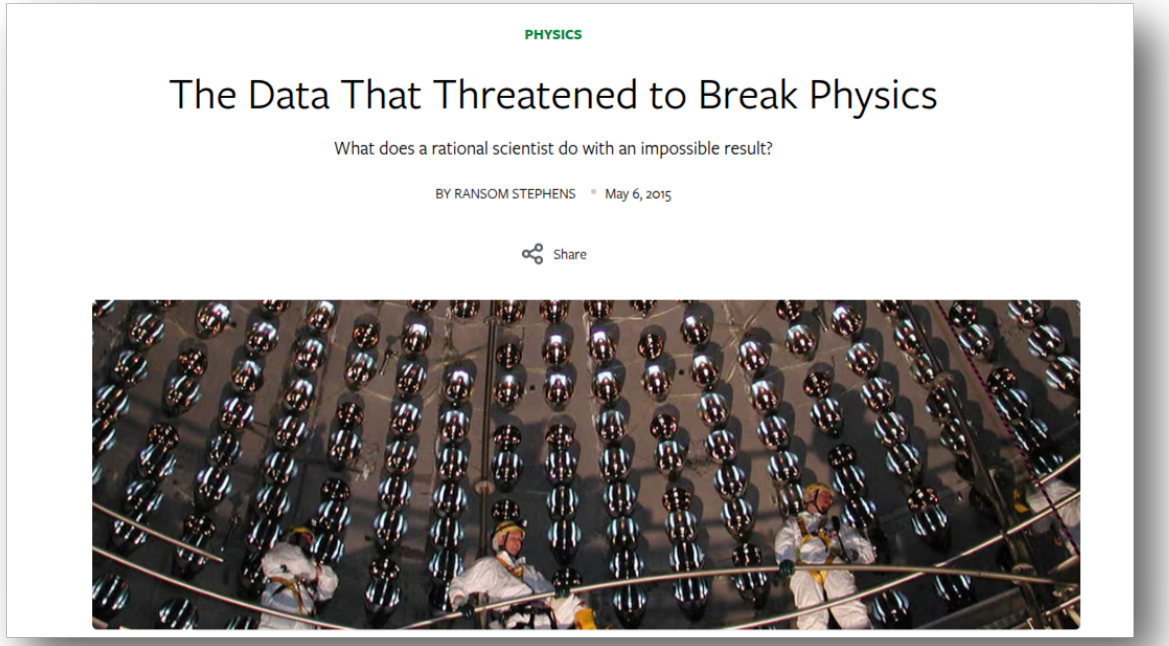
Image source: https://www.nngroup.com/articles/information-foraging/

SENSEMAKING

https://doi.org/10.1109/MIS.2006.75
https://doi.org/10.1109/MIS.2006.100

Sensemaking is an active two-way process of fitting data into a frame (mental model) and fitting a frame around the data. Neither data nor frame comes first; data evoke frames and frames select and connect data. When there is no adequate fit, the data may be reconsidered or an existing frame may be revised.

[adapted from Klein et al. "Making sense of sensemaking", 2003]



https://nautil.us/the-data-that-threatened-to-break-physics-235415/

SENSEMAKING

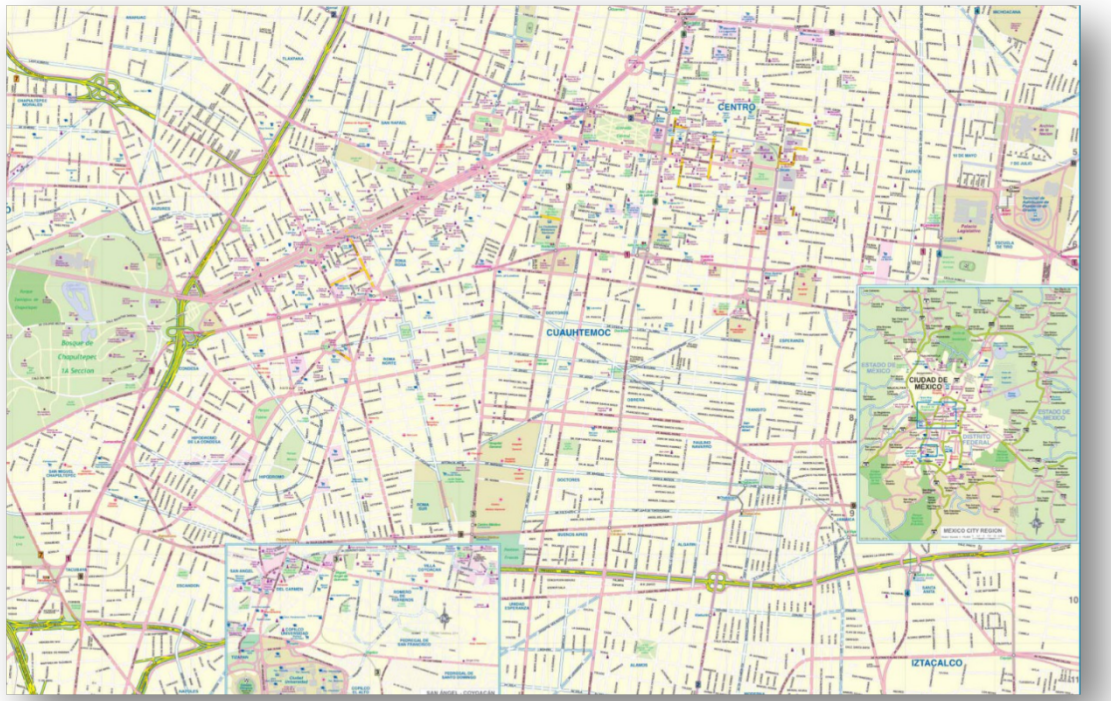
There are 5 cognitive triggers that change the way we think and that lead to insights:

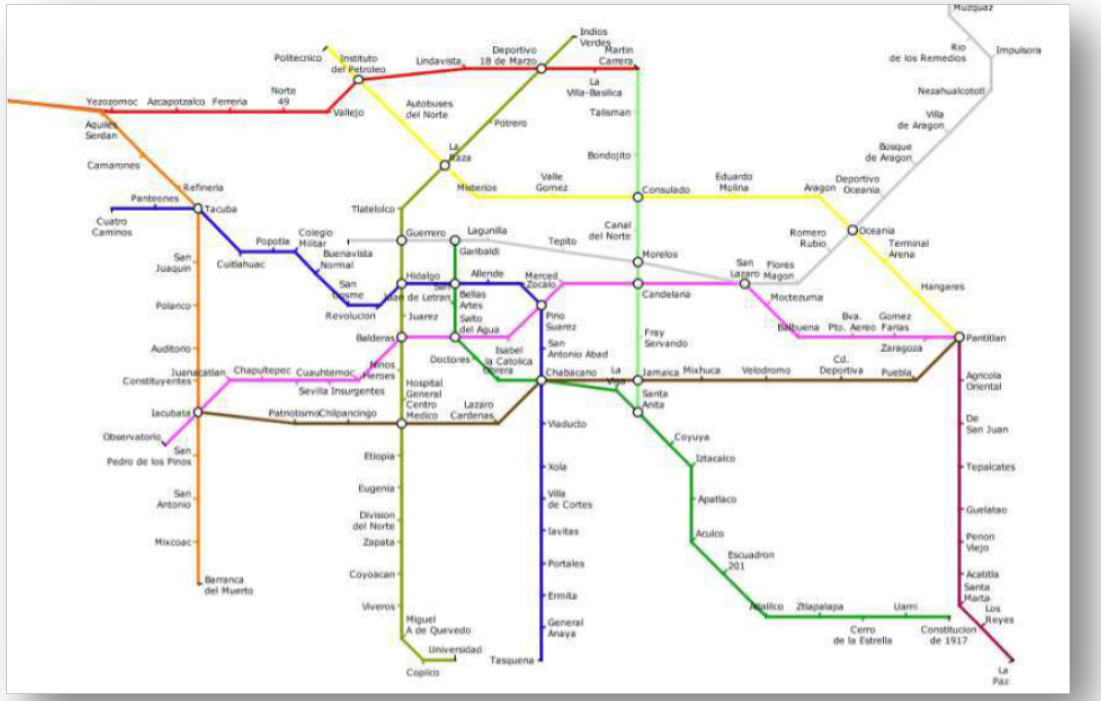
- **making connections:** *a new piece of information combines with other information we already have*
- **finding coincidences:** *spotting a relation that we don't quite understand (chance or pattern?)*
- **spotting contradictions:** *puts our mental or computational model in question, creates doubt*
- **emerging curiosities:** *a single event / data item provoking further investigation, creates wonder*
- being in a state of creative desperation: *extreme need for a breakthrough*
  -> try unorthodox approaches

[G. Klein: Seeing What Others Don't: The Remarkable Ways We Gain Insights, 2013]

GUIDING PRINCIPLES (NORMAN, TVERSKY)

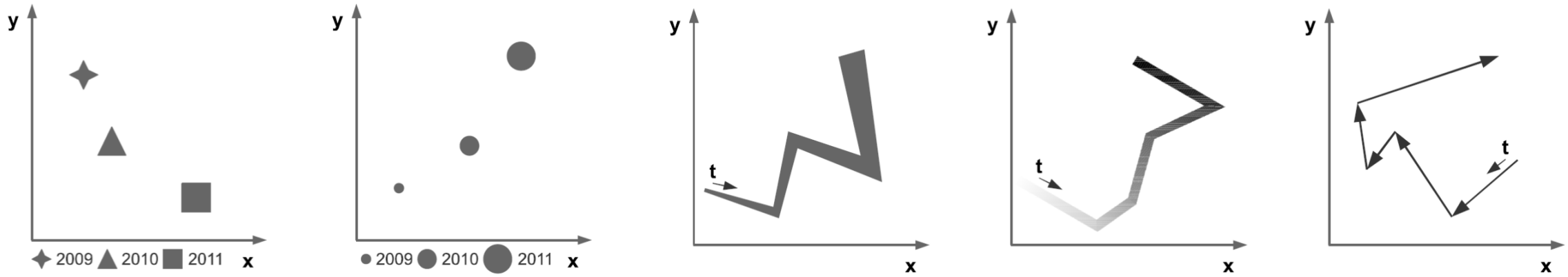
- **Appropriateness Principle:** Visual analytics should provide neither more nor less information than that needed for solving the problem





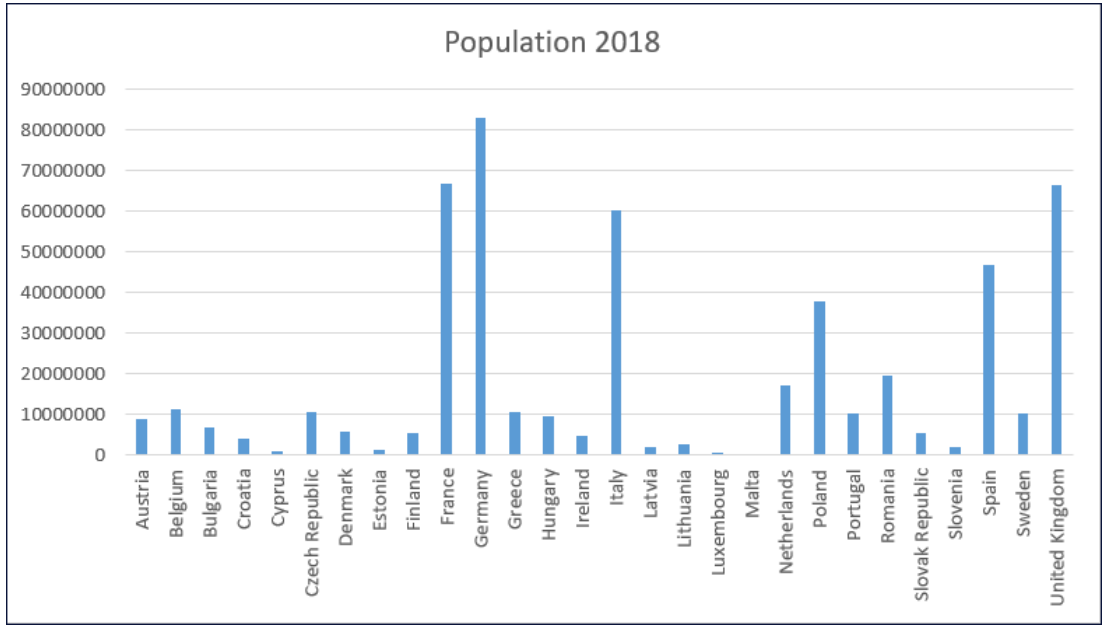
GUIDING PRINCIPLES (NORMAN, TVERSKY)

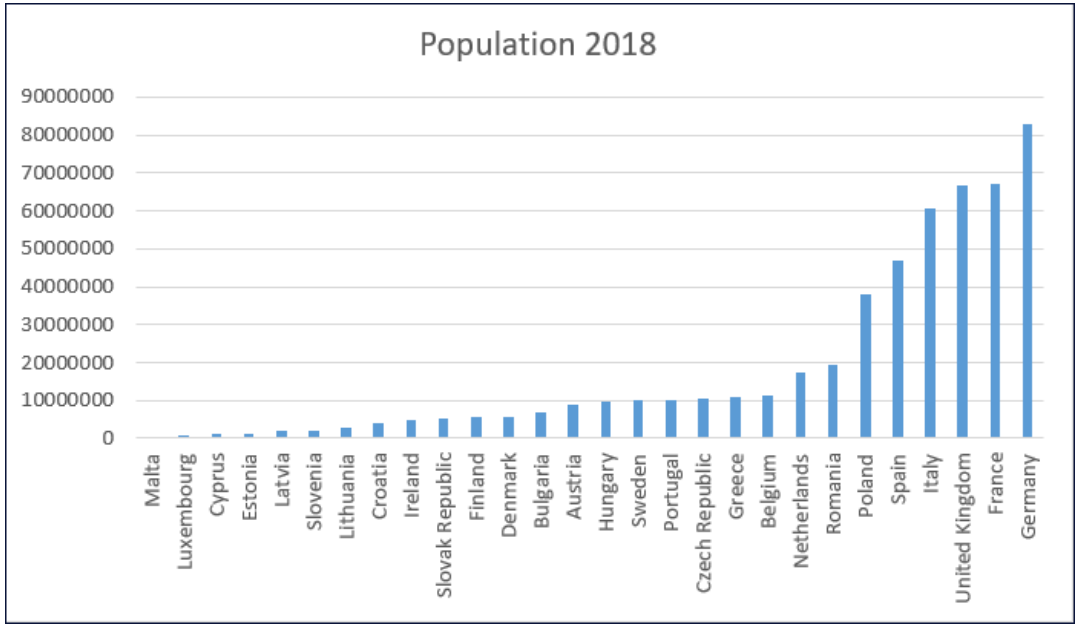
- **Naturalness Principle:** Visual cognition is most effective when the representation most closely matches the information being represented.



GUIDING PRINCIPLES (NORMAN, TVERSKY)

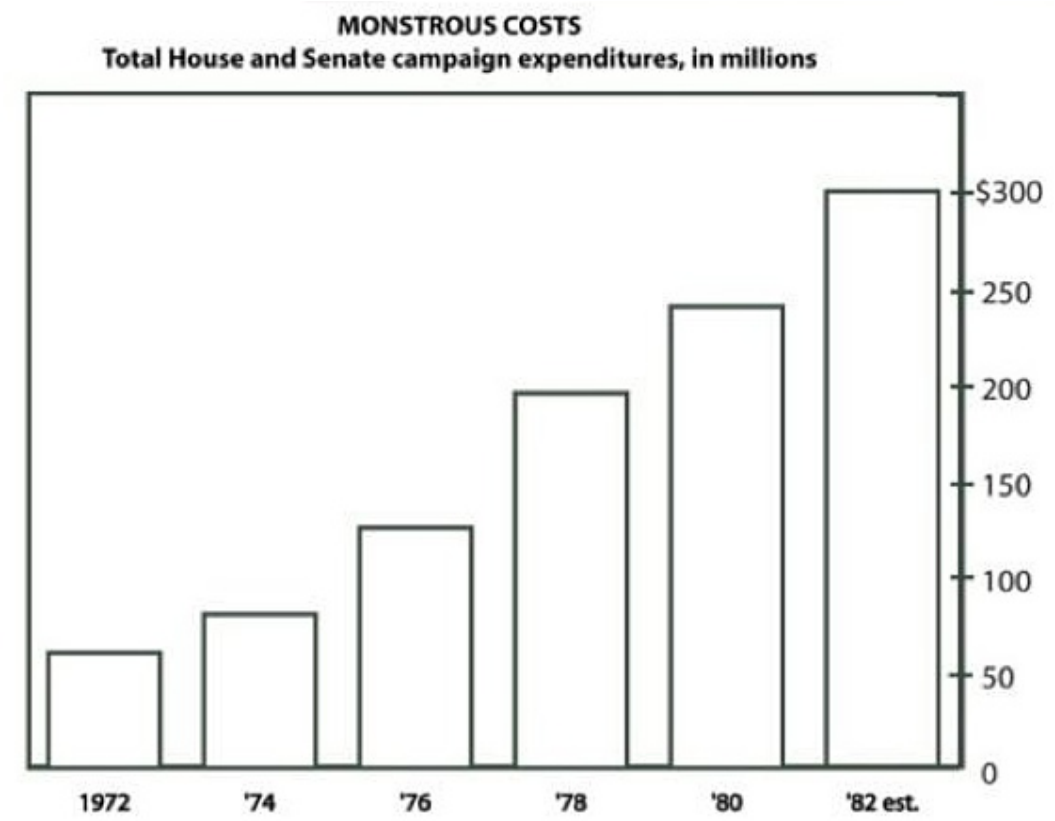
- **Matching Principle:** Visual analytics must match the task to be performed.





GUIDING PRINCIPLES (NORMAN, TVERSKY)

- **Apprehension Principle:** The content of the representation should be accurately and easily perceived.



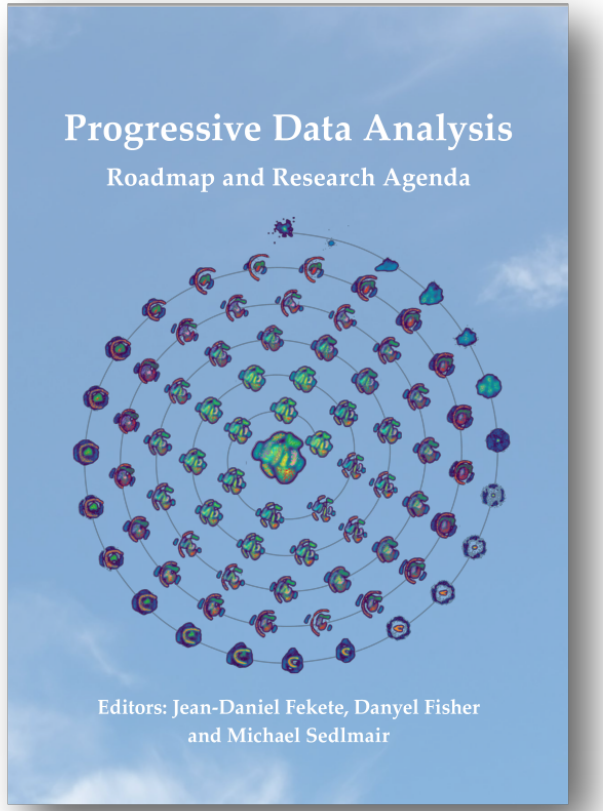
PROGRESSIVE VISUAL ANALYTICS

AARHUS
UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

HANS-JÖRG SCHULZ
ASSOCIATE PROFESSOR

AGENDA

1. PVA Fundamentals – The What and Why of PVA
2. Using PVA
    - PVA Roles
3. Realizing PVA
    - Interaction
    - Visualization
    - Software Architecture
4. Outlook on Transient Visual Analytics



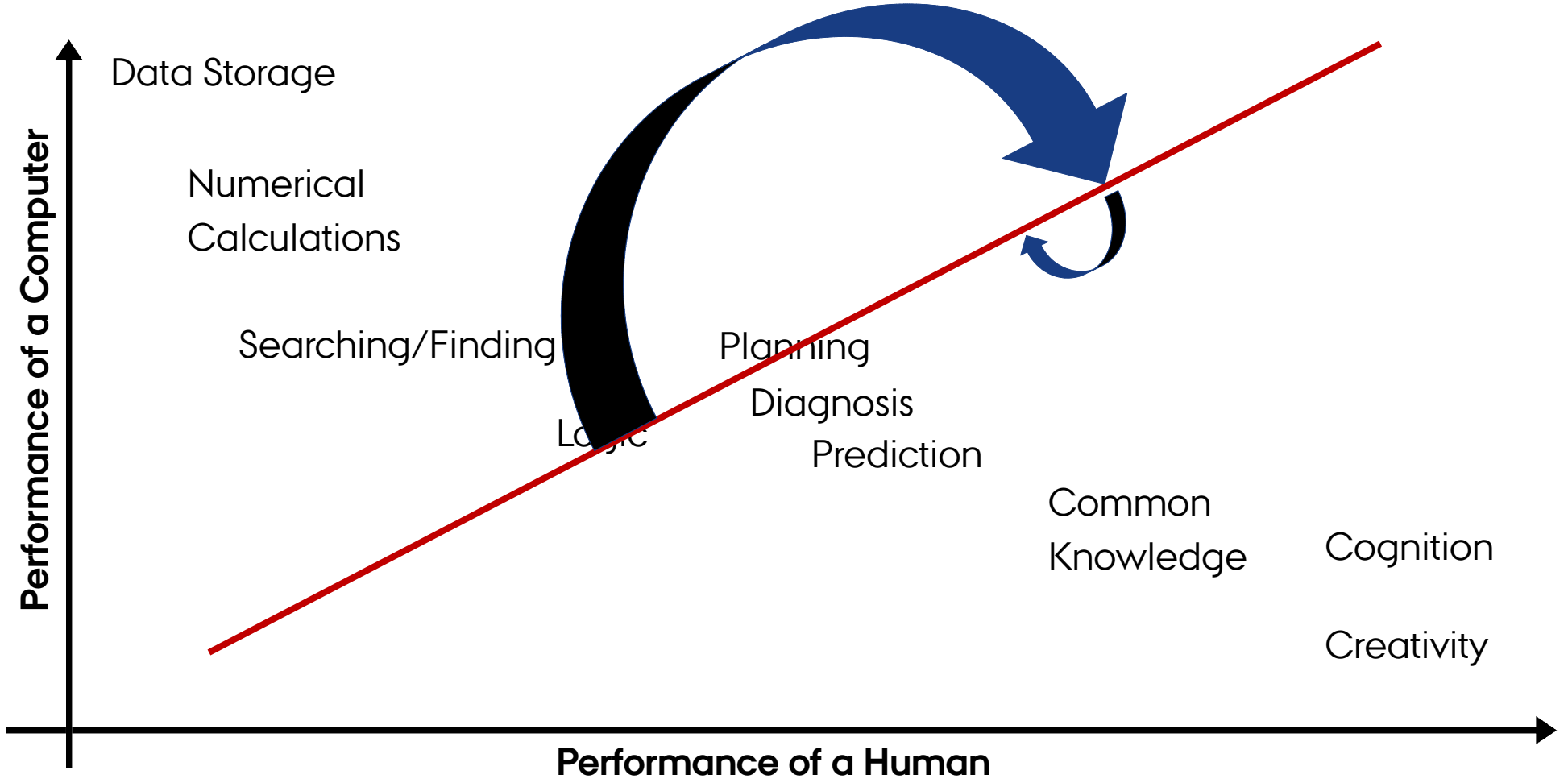
[https://doi.org/10.2312/pda.20242707]

PVA FUNDAMENTALS

AARHUS
UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

HANS-JÖRG SCHULZ
ASSOCIATE PROFESSOR

THE CHALLENGE INHERENT IN VA



adapted from Daniel Keim, Uni. Konstanz
[https://bib.dbvis.de/uploadedFiles/233.pdf]

HUMAN TIME CONSTRAINTS

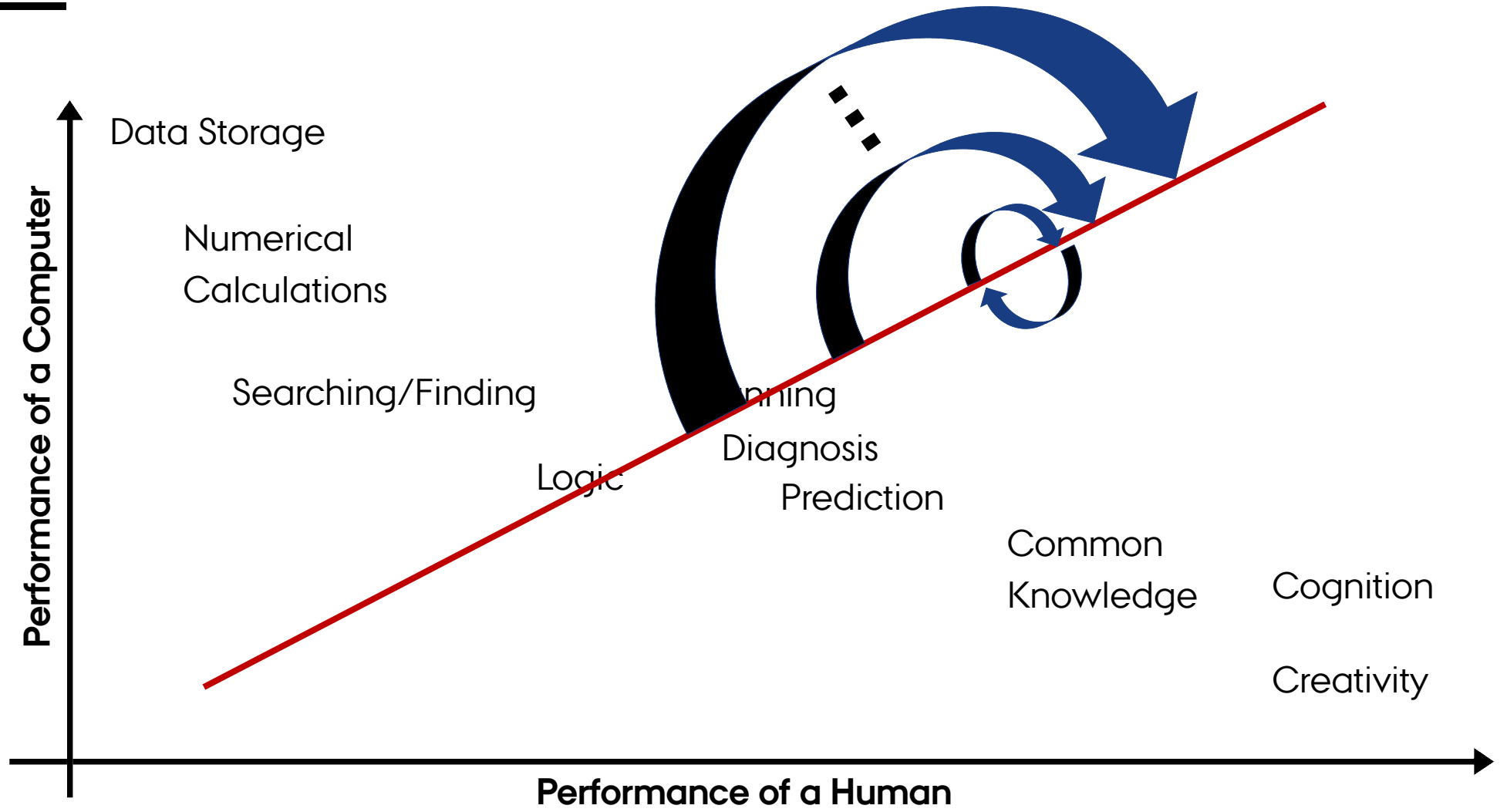
- **Task completion: ~10 sec (command line interaction)**
  Initiating a computational task, such as a query or complex filter operation
- **Immediate response: ~1 sec (indirect interaction)**
  Tuning computational parameters in a GUI, such as clustering parameters
- **Perceptual update: ~100 ms (direct interaction)**
  Computations initiated through direct interaction, such as a smart lens

[https://doi.org/10.1177/1473871611413180]
[https://doi.org/10.1145/2514.2517]

Violating these constraints -> Disrupting the user's analytic flow
-> Reduced discovery rates & data coverage

[https://doi.org/10.1109/TVCG.2016.2607714]

PROGRESSIVE VISUAL ANALYTICS (PVA)

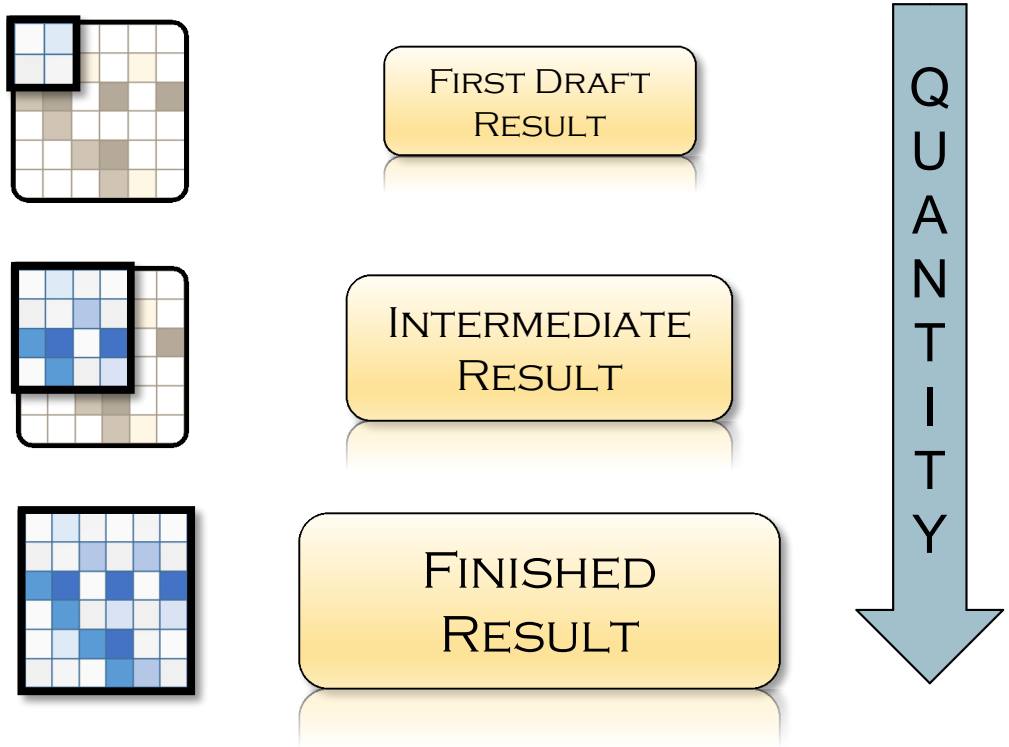


adapted from Daniel Keim, Uni. Konstanz
[https://bib.dbvis.de/uploadedFiles/233.pdf]

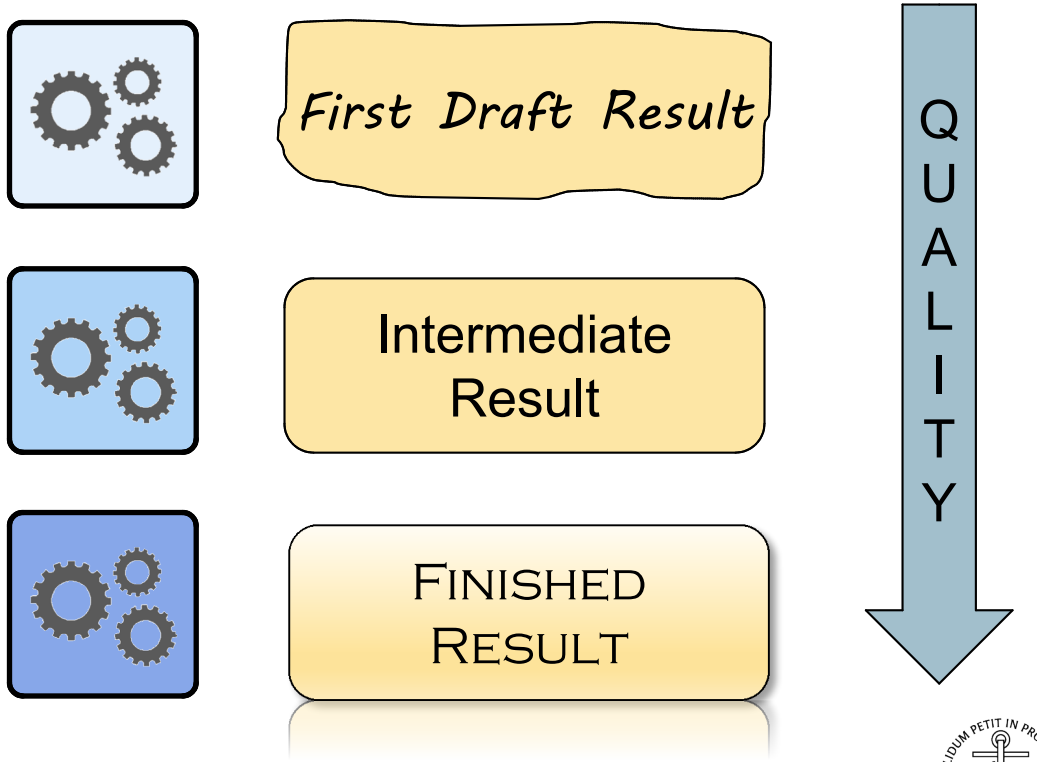
TWO PRINCIPAL TYPES OF PVA

Progressive Visual Analytics := refinement of processing result to be shown

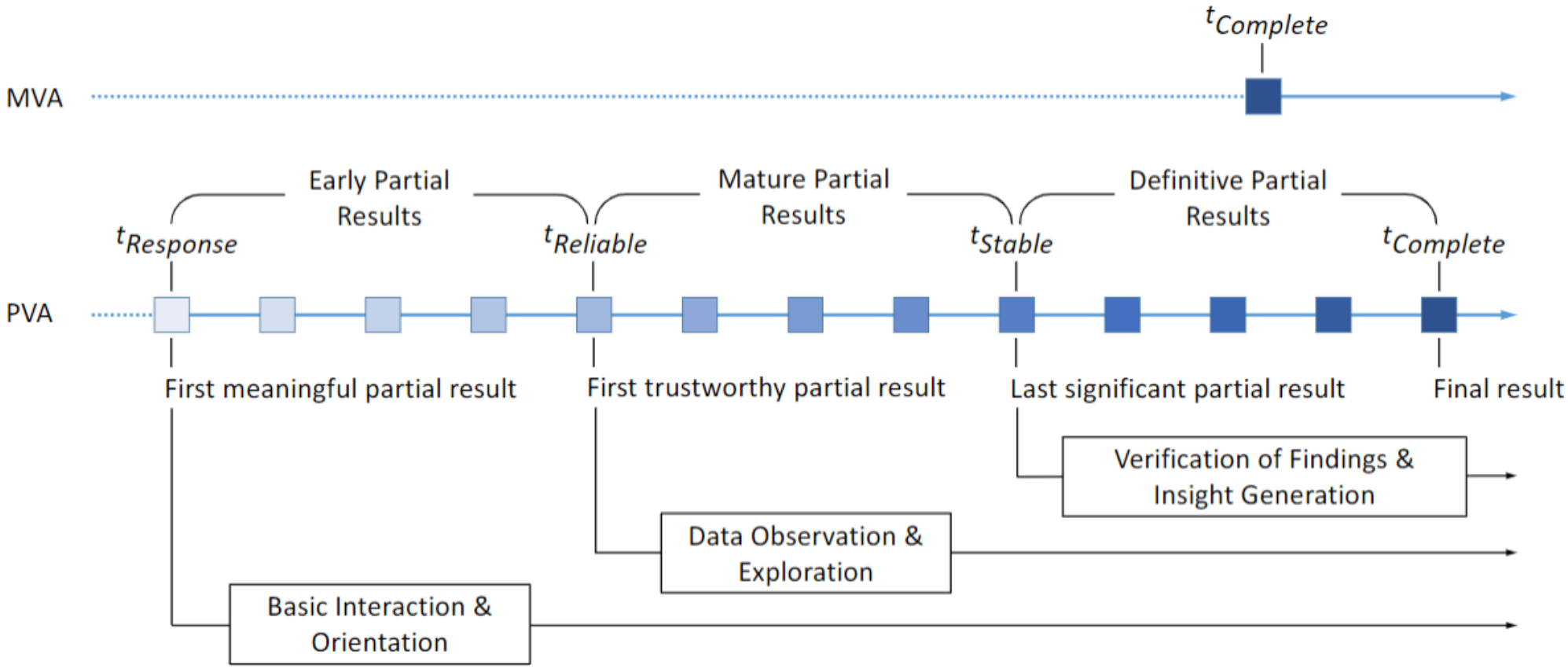
INCREMENTAL PVA / DATA CHUNKING



ITERATIVE PVA / PROCESS CHUNKING



THE PVA PROCESS



[https://doi.org/10.3390/informatics5030031]

A DEFINITION FOR PVA

Progressive Visual Analytics (PVA) is a type of Visual Analytics where the computational analysis produces intermediate results as approximations of the final result, so as to fulfill the human time constraints imposed by the visual-interactive analysis and thus maintain the user's analytic flow.

PVA can be (either) an incremental process that subdivides the data, processes it chunk-by-chunk, and thus yields intermediate results of increasing quantity / completeness – or it can be an iterative process that subdivides the computation, runs it step-by-step, and thus yields intermediate results of increasing quality / correctness.

VARIOUS USES OF THE PROGRESSION

Overcoming bottlenecks:

- algorithmic complexity
- big data
- slow network connection
- combinatorial explosion of search space

Resource-conscious VA:

- find good-enough solutions (reduce time to decision, power consumption)
- adjust to available screen space (reduce network load, cognitive load)

USING PVA

USER ROLES & TASKS

	Roles	R1: The Observer	R2: The Searcher	R3: The Explorer
Tasks		**T1:** Ascertain suitable quantity of processed data and stability of the computation results **T2:** View large information spaces progressively **T3:** Understand an algorithm and its inner workings	**T4:** Analyze an approximate or partial result of a costly query **T5:** Refine search space based on the intermediate results provided by the progression **T6:** Compare different executions of the computation	**T7:** Gain an overview of an unfamiliar information space **T8:** Identify possibilities for furthering the computation by integrating the user's tacit knowledge or preferences **T9:** Investigate alternative scenarios
Focus Space	**F1: Data Space** *(Application Expert)*	**U1: Observe Data Progression** • Visual sedimentation[1] • Data loading per properties[2]	**U2: Query Data Progressively** • Incremental queries[5,6] • Successive querying[7]	**U3: Explore Data Progressively** • Exploring data subsets[10] • Exploring patterns[11]
	F2: Algorithm Space *(Data Scientist)*	**U4: Observe Algorithm Progression** • Viewing algorithm animation[3] • Viewing inner workings[4]	**U5: Adjust Algorithm Runs** • Choosing parameters[8] • Debugging algorithm[9]	**U6: Explore Algorithm at Runtime** • Integrating user knowledge[12] • Investigating ongoing What-If scenarios[13]

[extended from https://doi.org/10.2312/evs.20191164]

R1: THE OBSERVER

Interest: the final result of the analysis process and how it came about

Benefits from PVA:

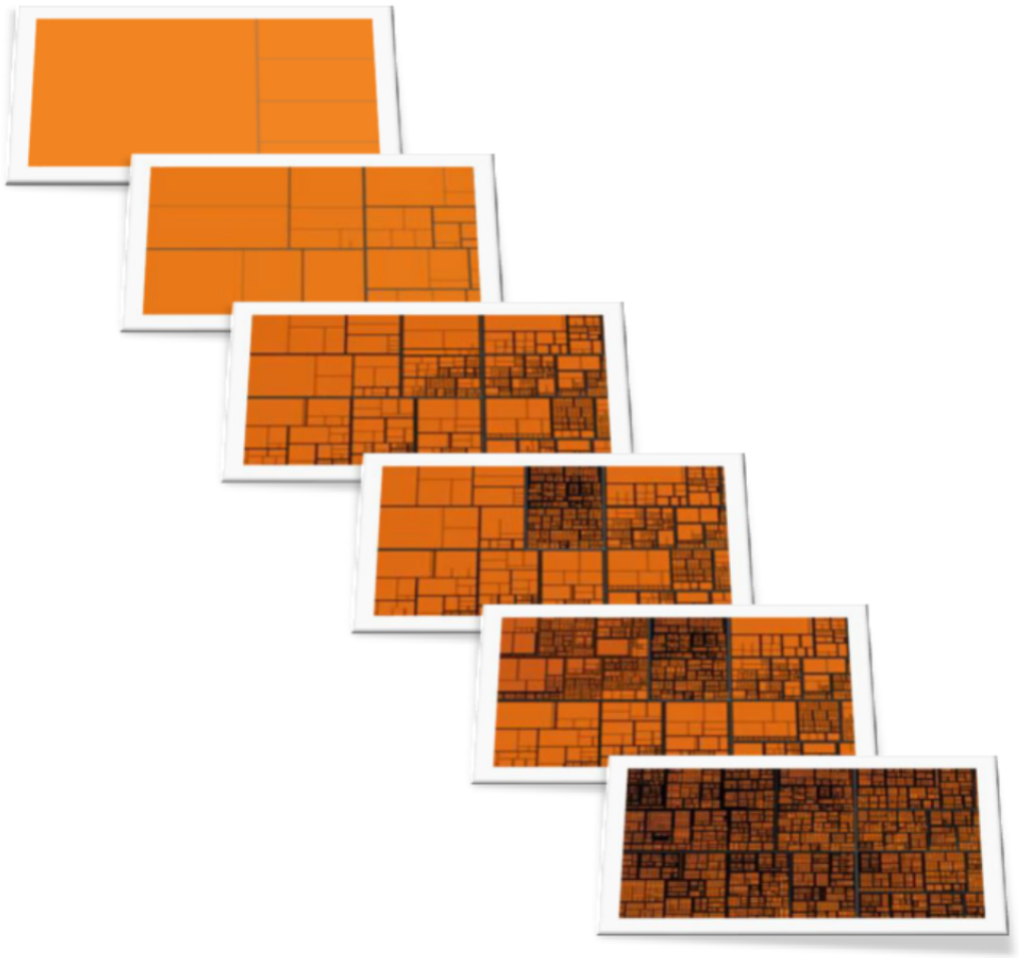
- stay informed about the state of the computation (aliveness)
- estimate how long it will take until it is done (progress)
- watch the algorithm working and see the final result unfold (provenance)

Expectation: PVA as a steadily improving output that arrives at a predictable end

Analogies: loading process, fancy progress bar

Requirements: Make the progression as stable as possible! + Pause/Play

R1: THE OBSERVER – EXAMPLE



[https://doi.org/10.1007/978-3-642-10520-3_7]

- Deterministic (“loading process”)
- Watching evolution over time
  -> to better understand how final result comes about
  -> to have an “aliveness” indicator
- Ensures that internal nodes are visible for a short while at least, before they get overdrawn by their children
  -> ability to pause / unpause the progression to inspect such intermediate states that will no longer be visible afterwards

R2: THE SEARCHER

Interest: a trustworthy intermediate result that is just good enough to solve a particular problem within a large information space (data, param, model)

Benefits from PVA:

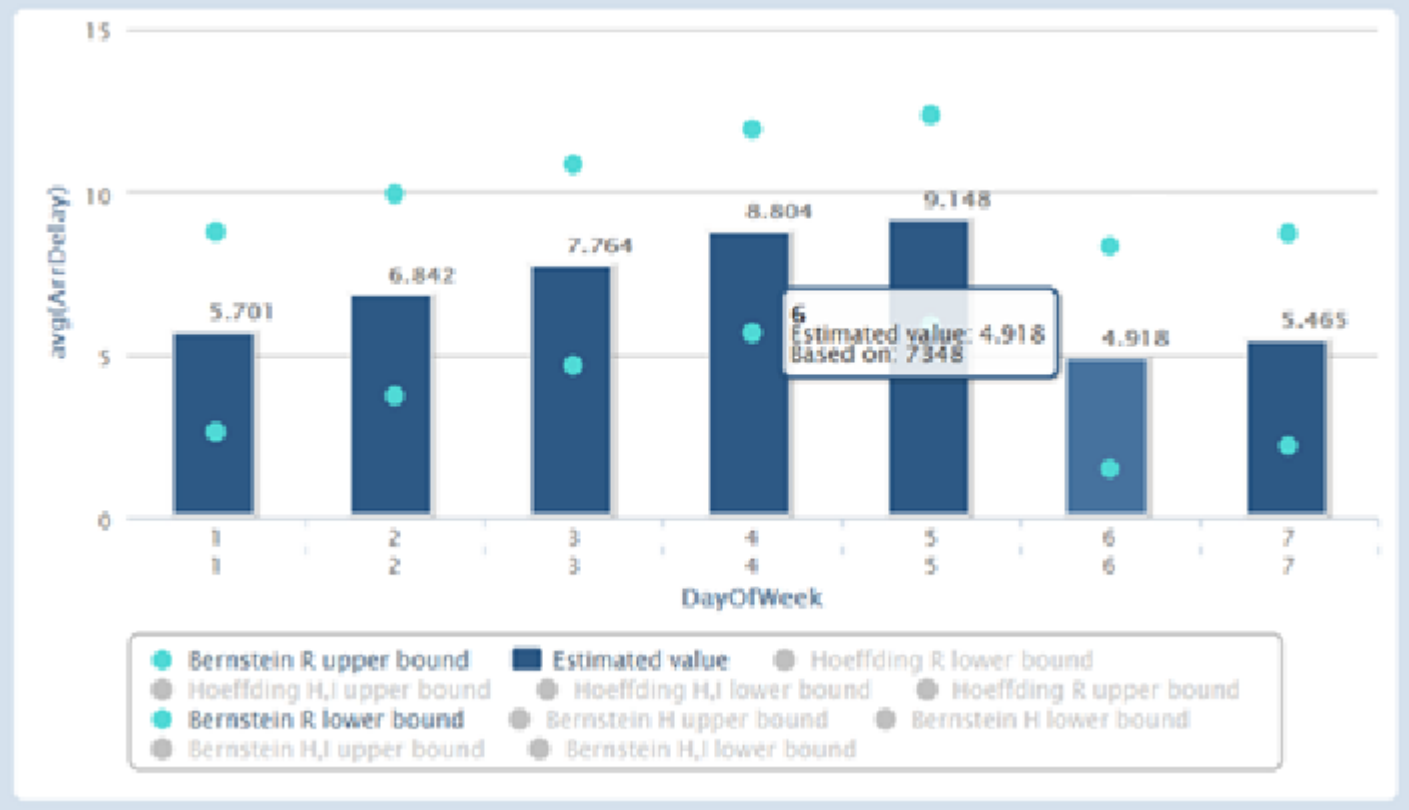
- fully interactive intermediate results allowing their visual exploration
- cancel the running process, saving the system a costly exhaustive search (early termination)

Expectation: PVA as step-wise process that produces constantly updated results which can be used in place of a final result for subsequent analytic operations

Analogies: internet search engine, product search

Requirements: Make the progression as useful as possible as early as possible!

R2: THE SEARCHER – EXAMPLE



[https://doi.org/10.1145/2207676.2208294]

- On which day to fly?
  -> show flight delays
  (FAA database, 17.5 million entries)
- after 0.32% completion
  -> weekday / weekend pattern emerges
- Either: good enough result to choose a weekend day
- Or: useful to prune the search space to include only Sat, Sun, Mon -> “incremental querying”

R3: THE EXPLORER

Interest: a comprehensive understanding of the data and process itself as observed by the sequence of intermediate results (sensitivity, perturbation)

Benefits from PVA:

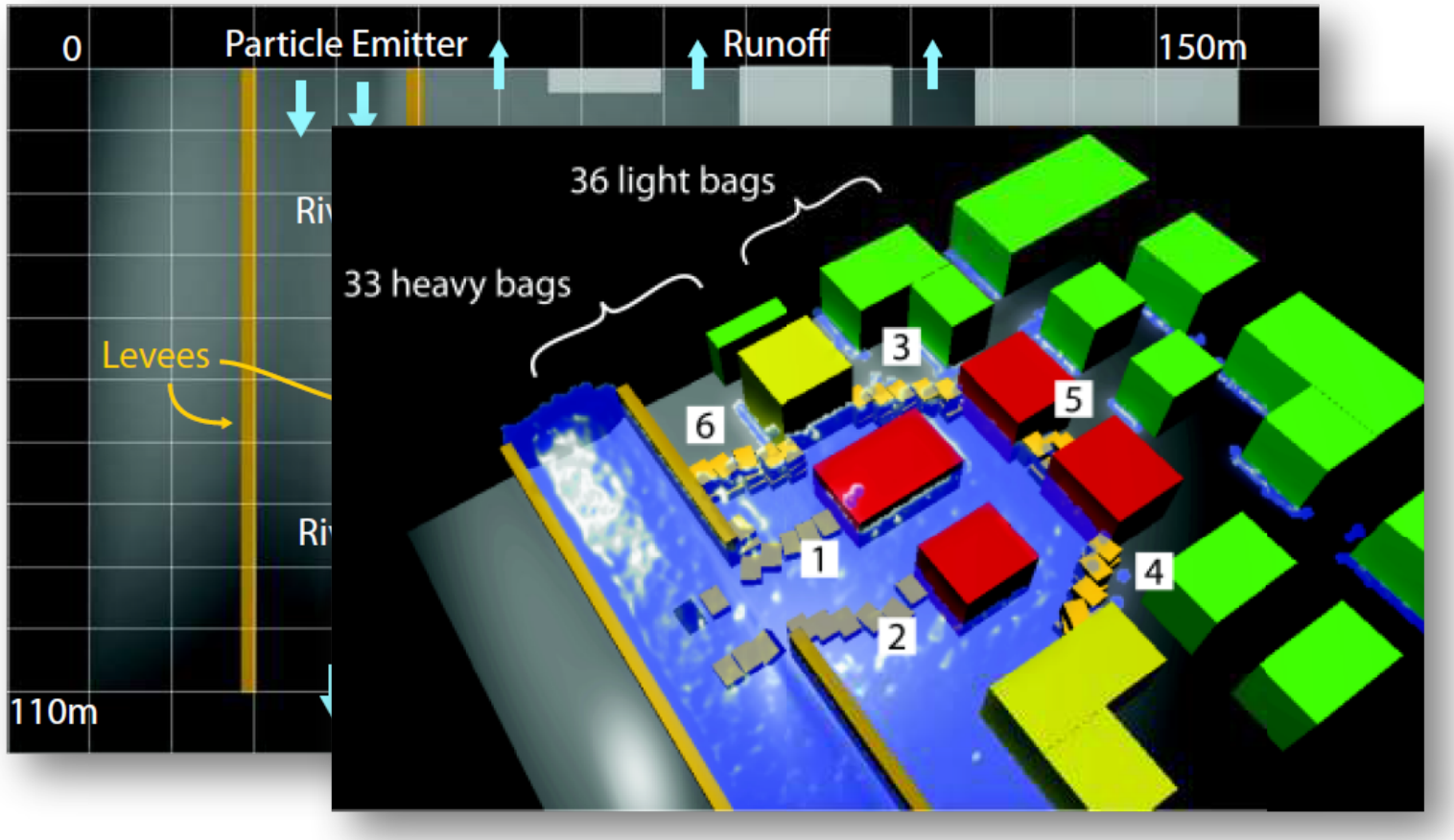
- orient themselves and decide in which direction(s) to steer the “tour” through the data or parameter space, swiftly adjusting process settings at runtime
- branching-off alternatives & comparing their trajectories (multiverse analysis)

Expectation: PVA as malleable process that can be redirected, re-parametrized, re-prioritized, refocused, and interactively perturbed

Analogies: physics simulation, What-if analyses

Requirements: Make the running progression as flexible/interactive as possible!

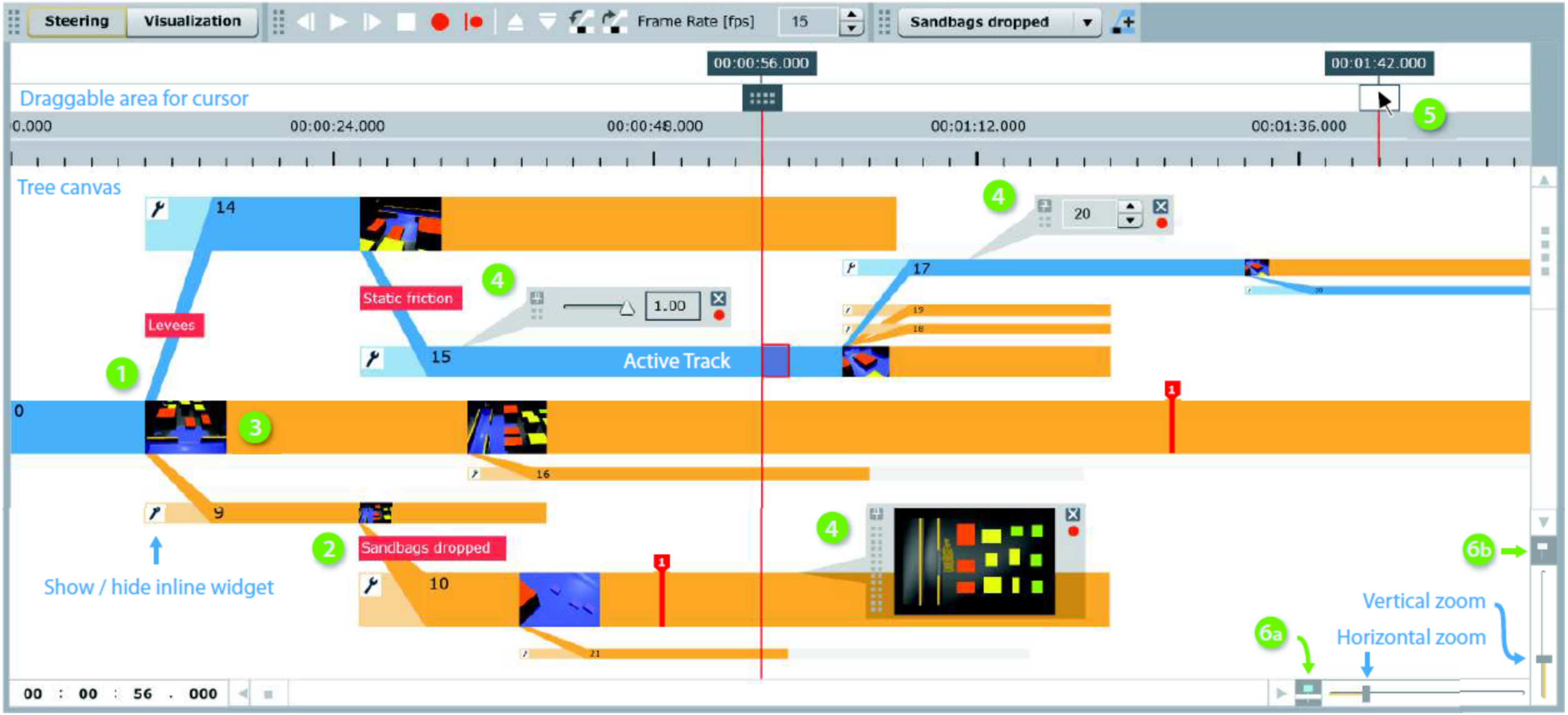
R3: THE EXPLORER – EXAMPLE



- Scenario-based analyses
- Adjust at runtime

  - parameters (water speed/volume)
  - configuration (placement of sandbags)
- Branching for What-If analyses

[https://doi.org/10.1109/TVCG.2010.223]



[https://doi.org/10.1109/TVCG.2010.223]

REALIZING PVA - INTERACTION

AARHUS
UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

HANS-JÖRG SCHULZ
ASSOCIATE PROFESSOR

INTERACTION – PAUSE/UNPAUSE

- Pause/Play – as in a video player
- Playhead to navigate to previous time point

[https://doi.org/10.1111/cgf.13205]

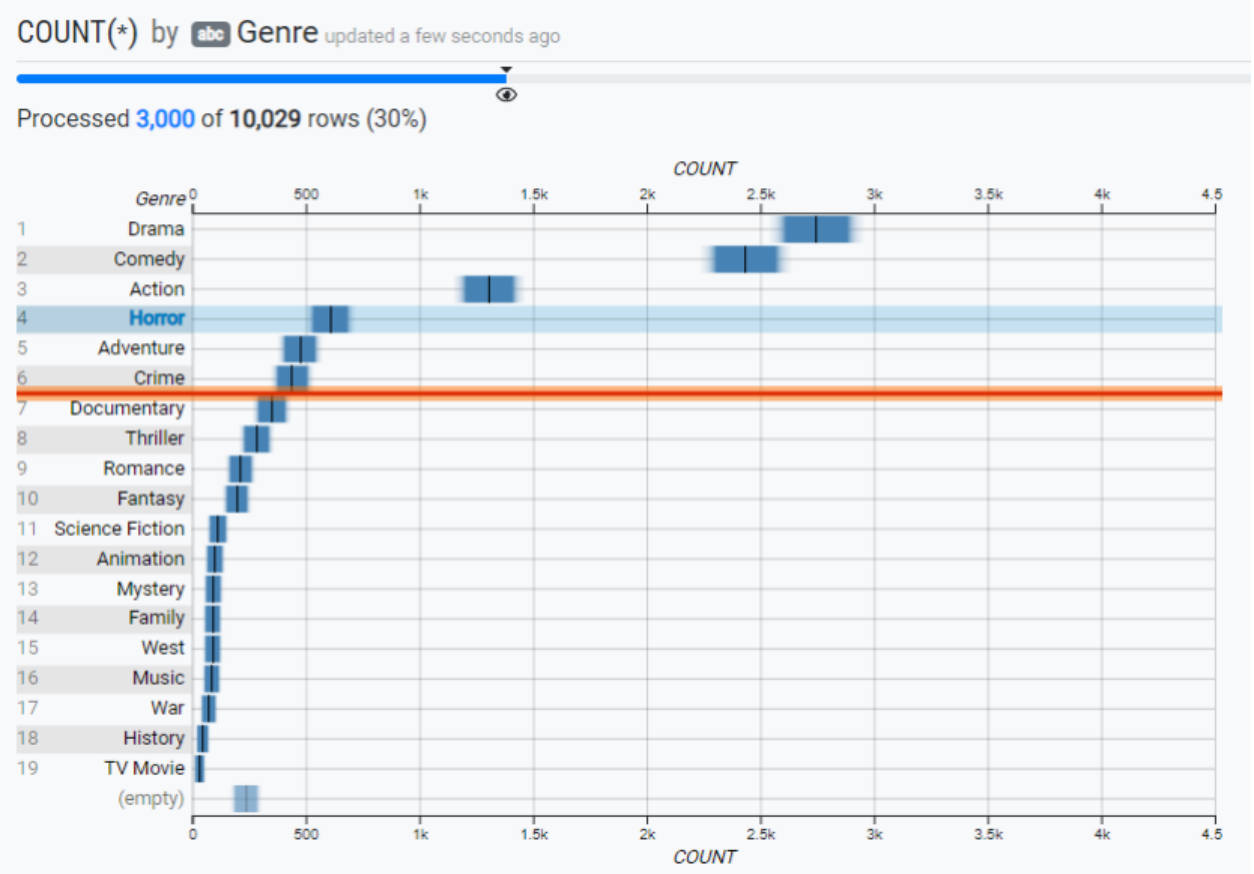
- Pause – freezes the visualization, but progression continues in the background
- Play – then just displays the results that have been generated in the meantime
- Update – moves playhead to most recent time point but stays paused
- Alert – when difference between shown results and new results becomes too large

[https://doi.org/10.1109/TVCG.2014.2346574]

INTERACTION – EARLY TERMINATION

Leave behind "Progressive Guards" (https://proreveal.github.io/ProReveal/)

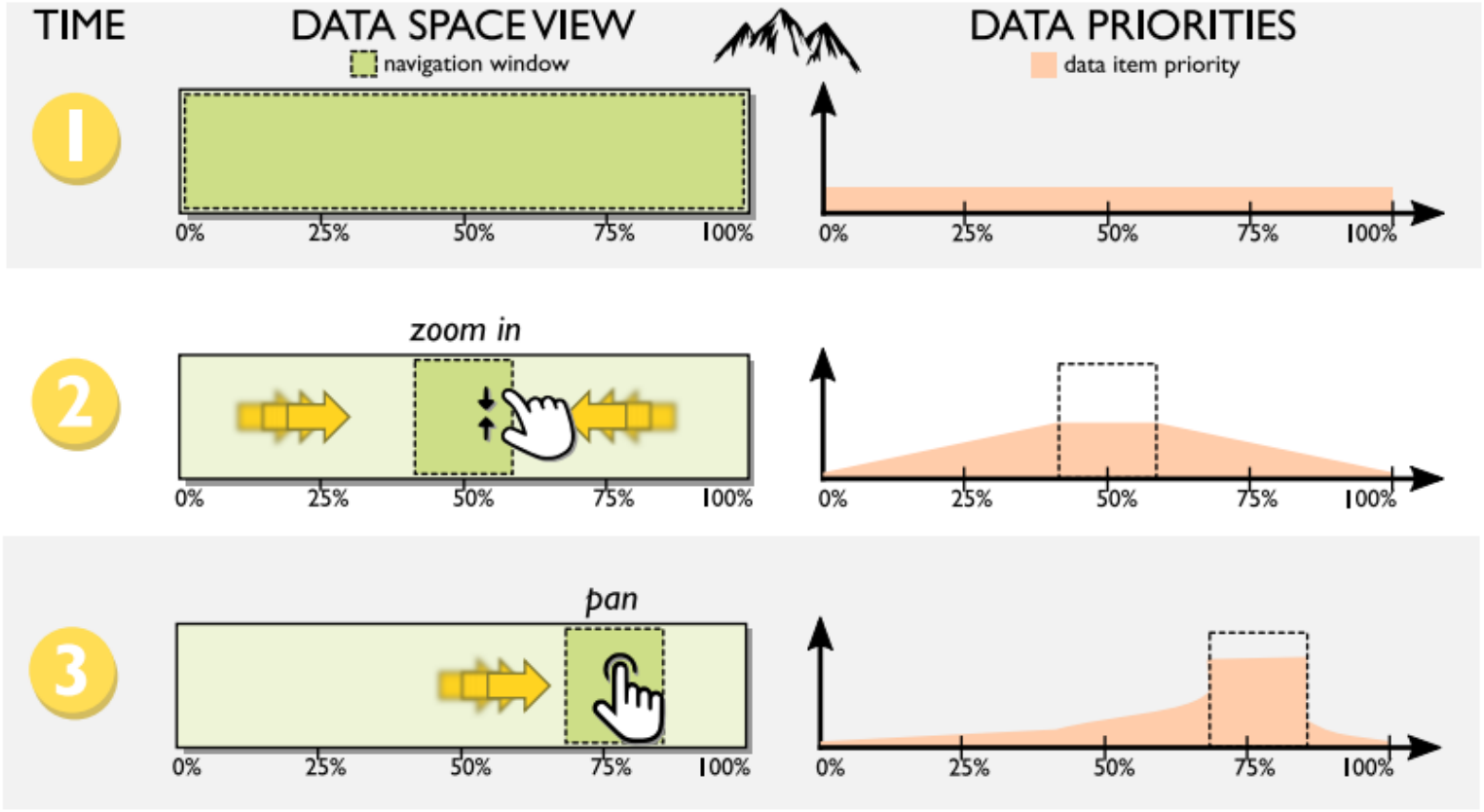
Name	Example ($\langle variable \rangle$ $\langle operator \rangle$ $\langle constant \rangle$)
Value	$Price(Apple) \leq \$2$
Rank	$Rank(Price(Apple)) \leq 10$
Range	$Price(Apple) \in [\$1, \$3]$
Comparative	$Price(Melon) \geq Price(Apple)$
Power Law	$Prices\ of\ Fruit \sim a \times Rank^{-k}$
Normal	$Prices\ of\ Fruit \sim \mathcal{N}(\mu, \sigma^2)$
Linear	$Prices\ of\ Fruit \propto Sizes\ of\ Fruit$



[https://doi.org/10.1109/tvcg.2019.2962404]

INTERACTION – STEERING/PRIORITIZE

SHERPA – coupling view space navigation & data prioritization



[https://doi.org/10.1109/VDS48975.2019.8973384]

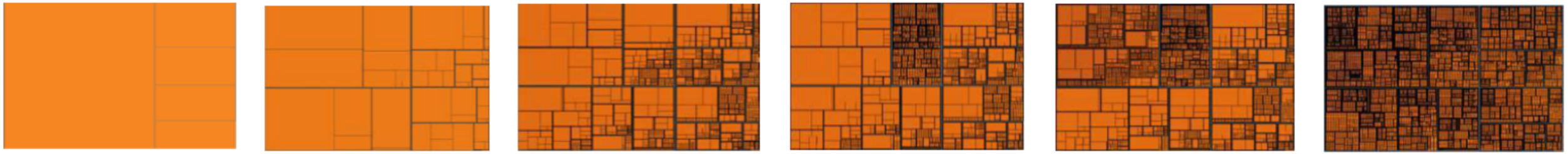
REALIZING PVA - VISUALIZATION

AARHUS
UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

HANS-JÖRG SCHULZ
ASSOCIATE PROFESSOR

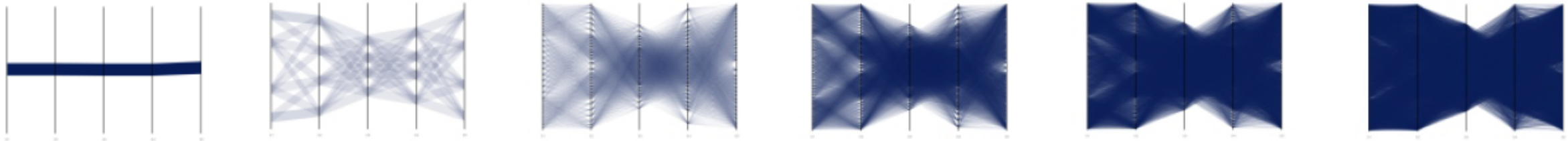
VISUALIZATION – 1 DATA ITEM = 1 MARK

Progressive Treemaps



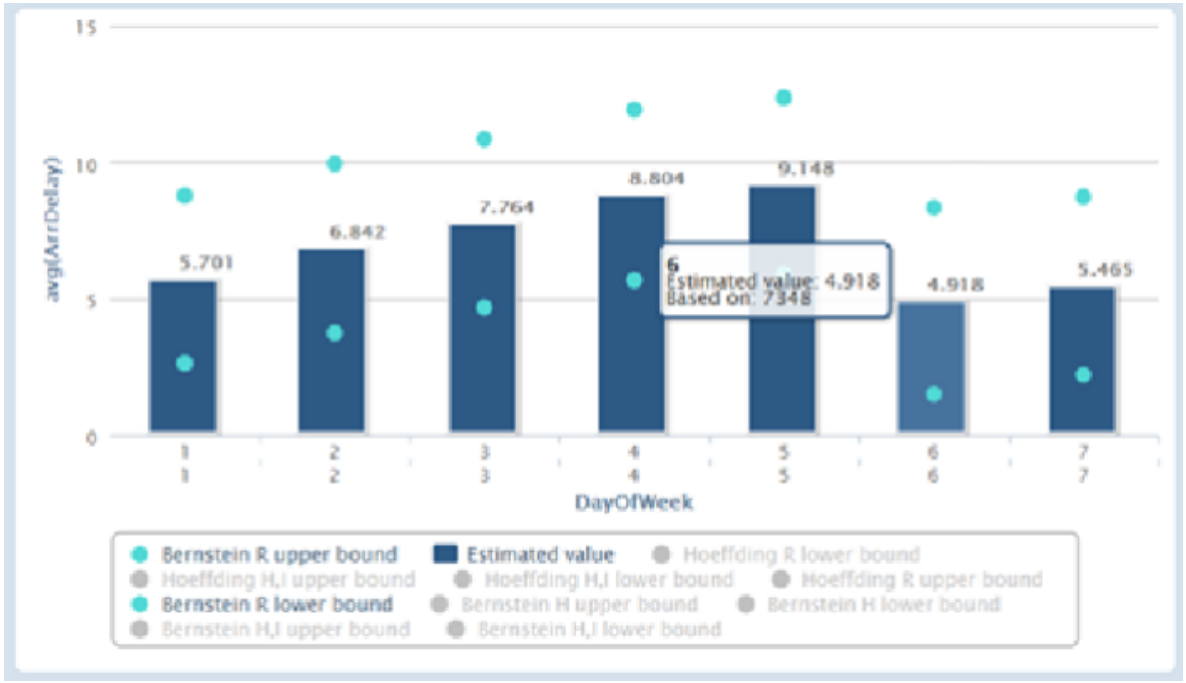
[https://doi.org/10.1007/978-3-642-10520-3_7]

Progressive Parallel Coordinates

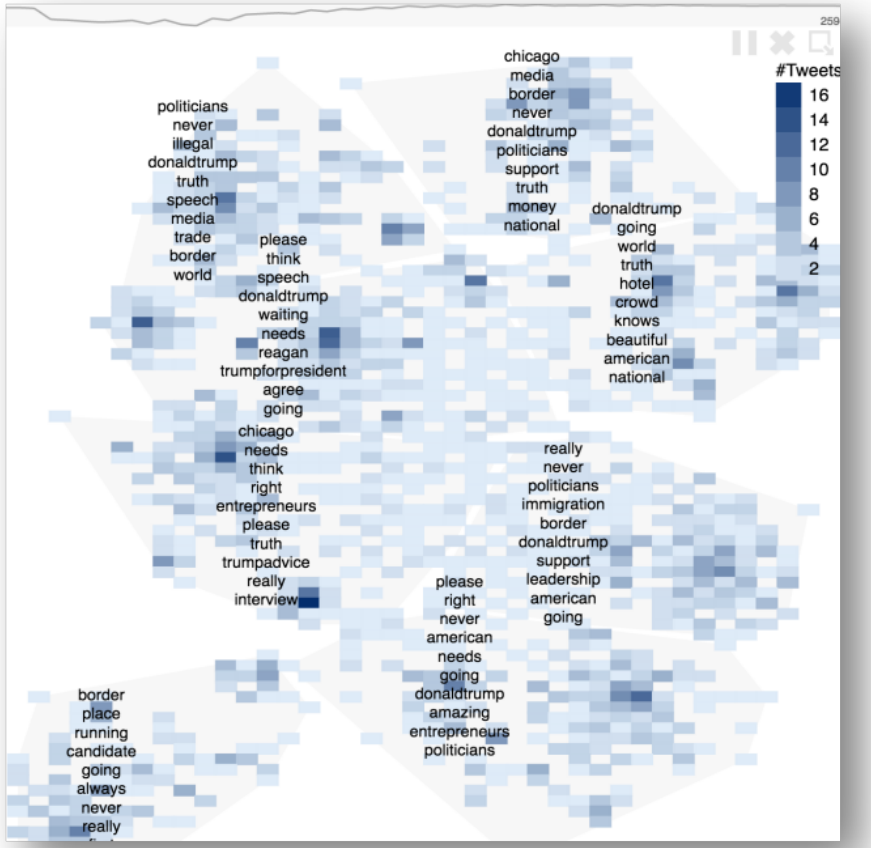


[https://doi.org/10.1109/PacificVis.2012.6183570]

VISUALIZATION – AGGREGATE DATA ITEMS



[https://doi.org/10.1145/2207676.2208294]

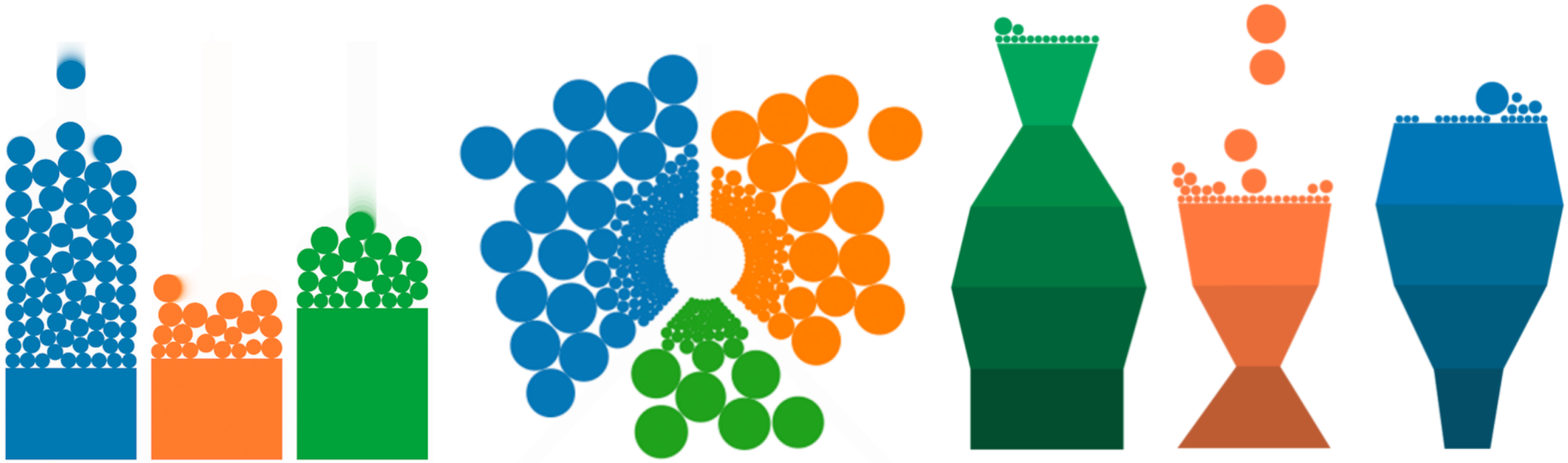


[https://doi.org/10.1111/cgf.13205]

VISUALIZATION – A COMBINATION

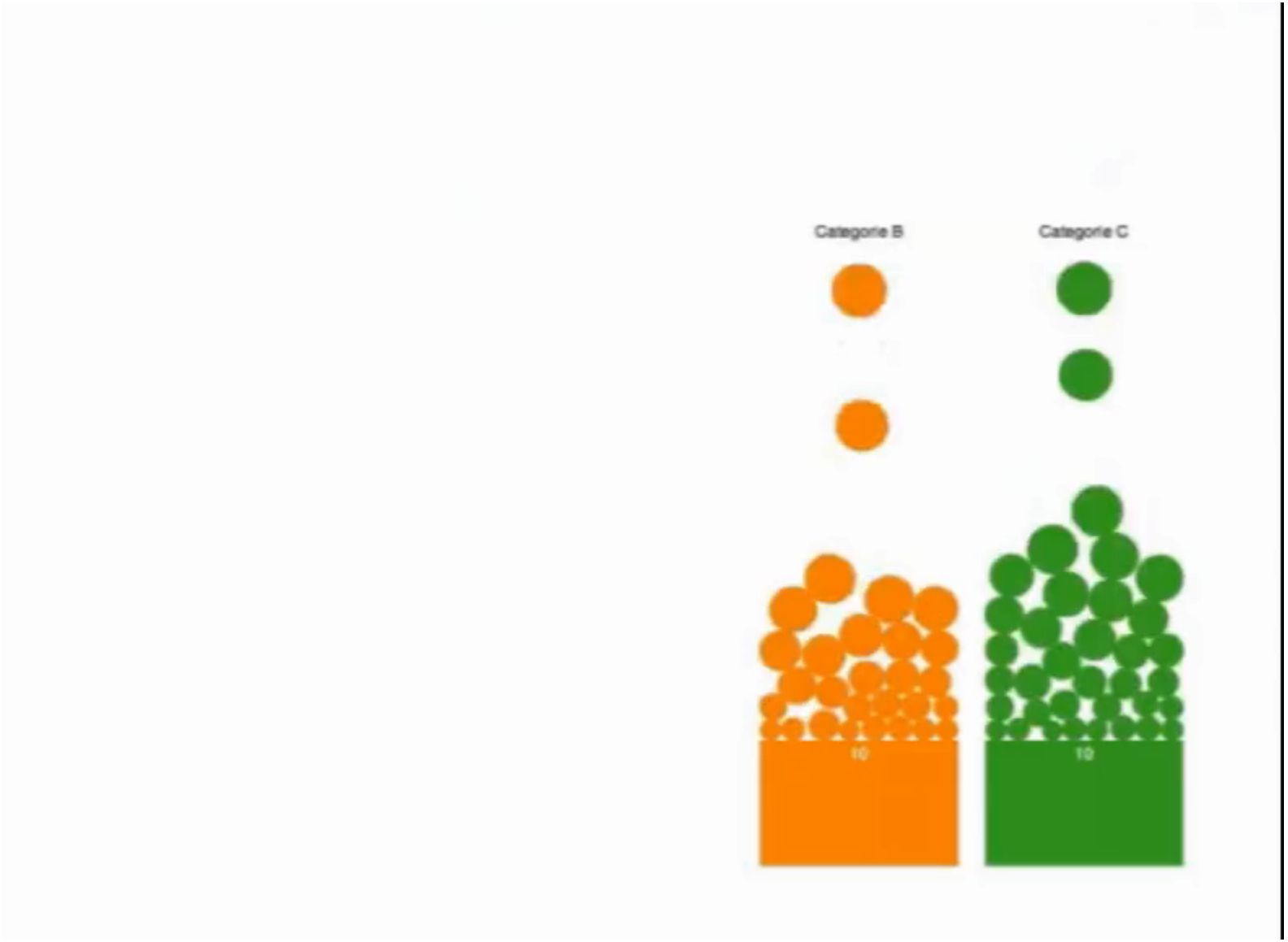
Visual Sedimentation

https://peltiertech.com/anybody-but-romney-snakeskin-chart/



[https://doi.org/10.1109/TVCG.2013.227]

VISUAL SEDIMENTATION



REALIZING PVA – SOFTWARE ARCHITECTURE

AARHUS
UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

VISUAL ANALYTICS
DATAVIS FALL 2025

HANS-JÖRG SCHULZ
ASSOCIATE PROFESSOR

HOW TO BUILD VIS/PVA SYSTEMS?

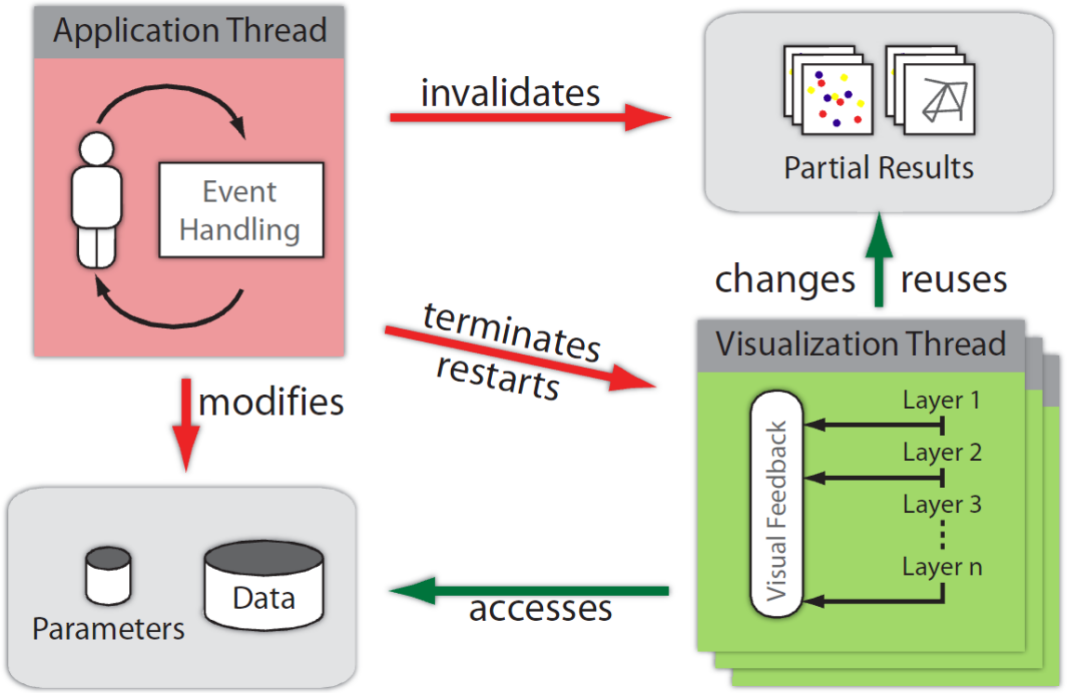
Multi-threading Architecture

-Separate Application thread and Visualization thread(s)

-Each View gets its own Visualization thread

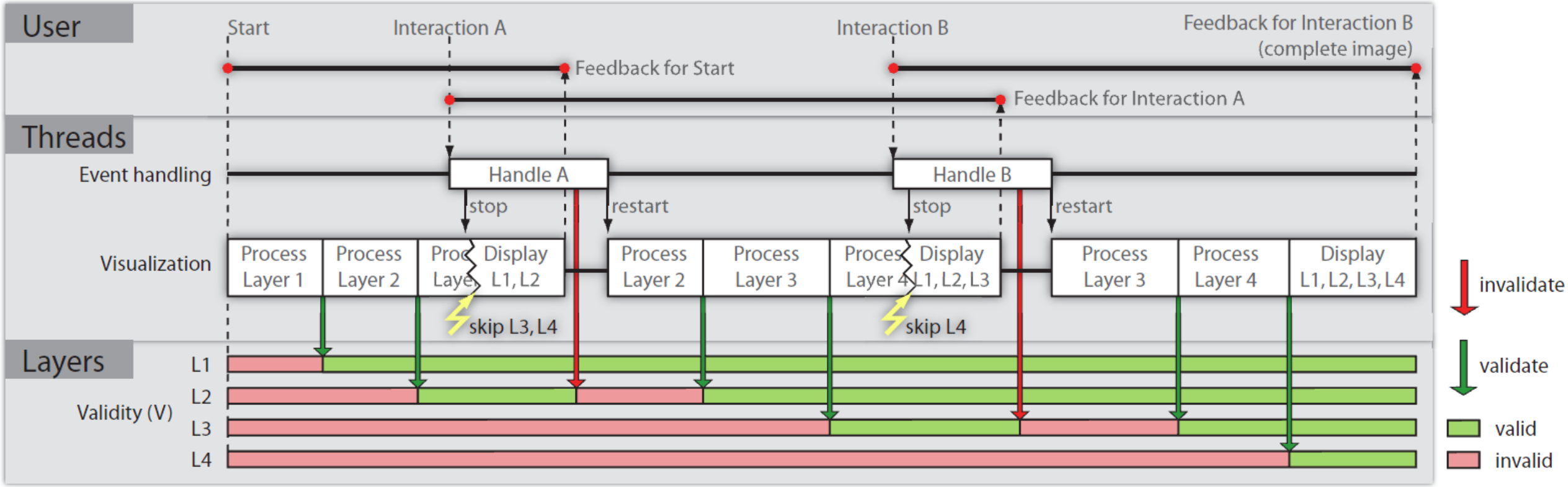
-Visualization thread progressively refines the view until finished or cancelled

-Progression/refinement defined using layers



HOW TO BUILD VIS/PVA SYSTEMS?

Visualizations must be layered to allow reuse of partially drawn elements:



HOW TO BUILD VIS SYSTEMS?

Common layering mechanisms:

- **Semantic layers:**
  background (map, grid,...), coordinate axes, data items, labels,...
  -> sort by decreasing relevance or increasing effort

- **Incremental layers:**
  sampled data items – e.g., every 100th, every 10th, and finally every item

- **Level-of-Detail layers:**
  -> on data level – clusters, subclusters,...
  -> on image level – first at low resolution, then at high resolution, then with anti-aliasing
  Note: here layers get replaced, not blended onto drawn ones

OUTLOOK ON TRANSIENT VISUAL ANALYTICS

AARHUS
UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

HANS-JÖRG SCHULZ
ASSOCIATE PROFESSOR

LET'S RECAP WHAT PVA DOES...

Over time, PVA adds more data

and more data

and more data

and more data

and more data

and more data

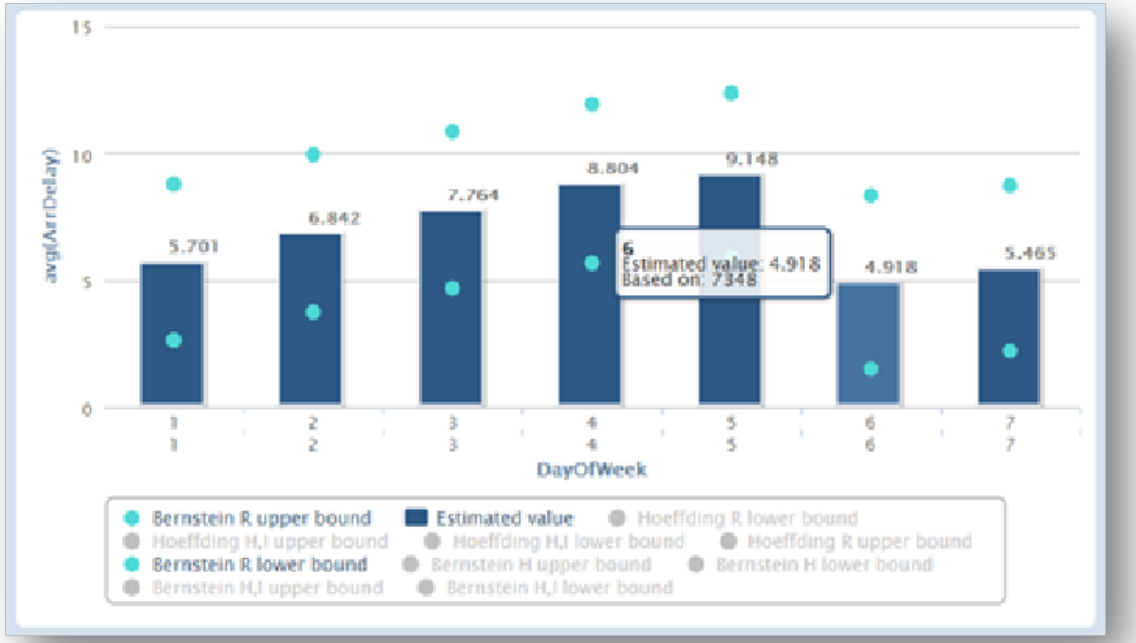
and more data

and more data

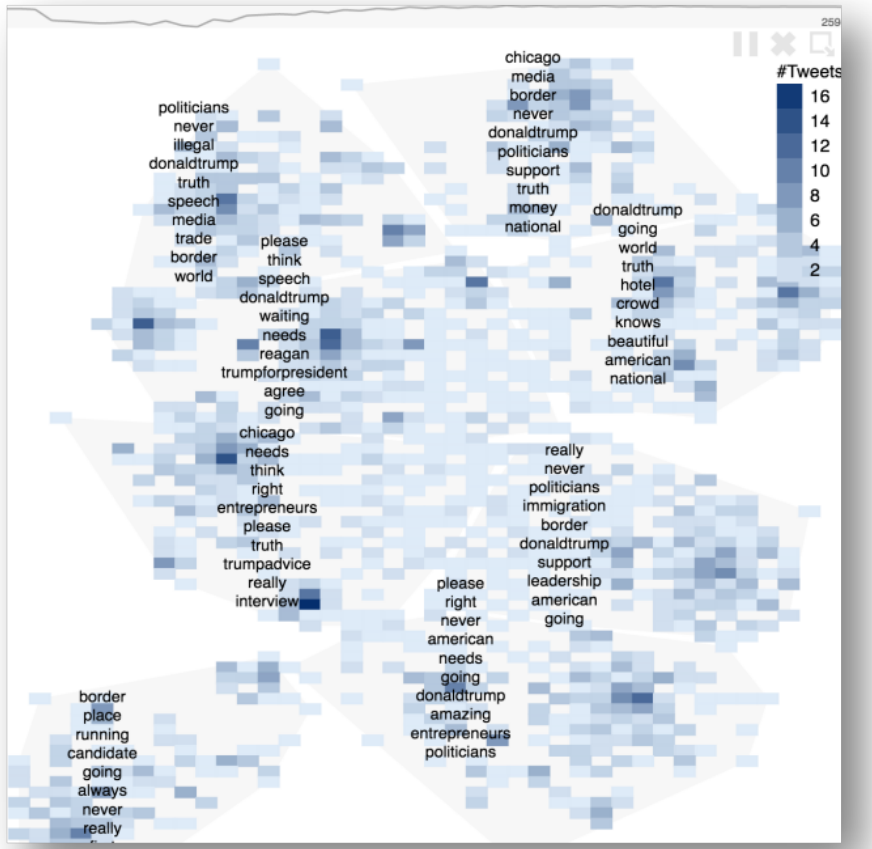
and more data.

Out of Memory / Out of Screen Space

HOW NOT TO RUN OUT OF SPACE?



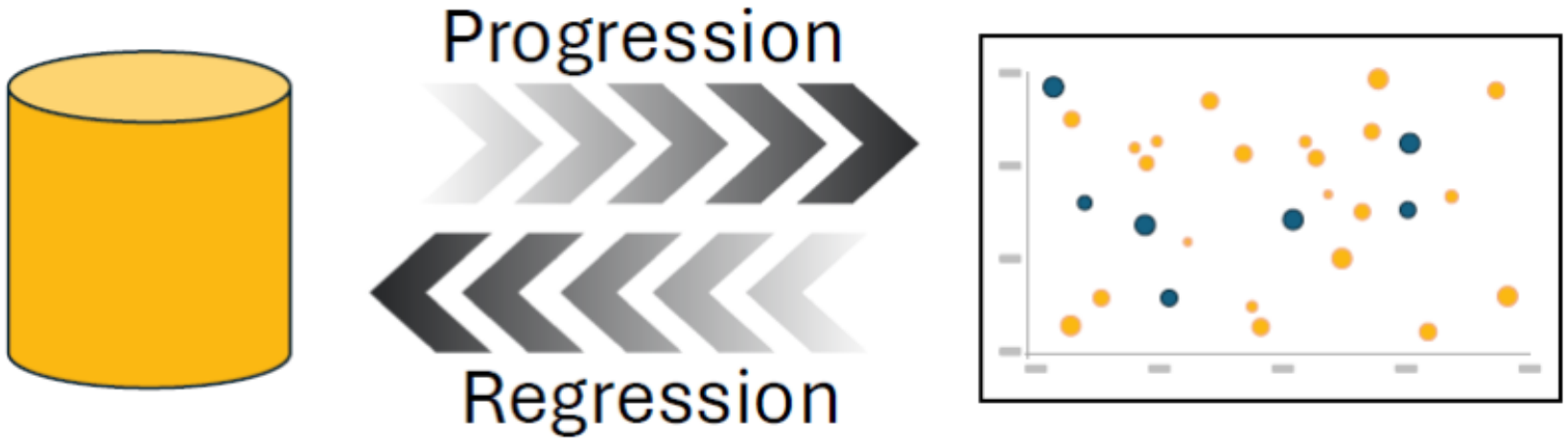
"Trust me, I'm partially right" by D.Fisher et al. 2012



"Steering the Craft"
by S.K.Badam et al. 2017

THE IDEA OF TRANSIENT VISUAL ANALYTICS

[https://doi.org/10.2312/eurova.20241108]

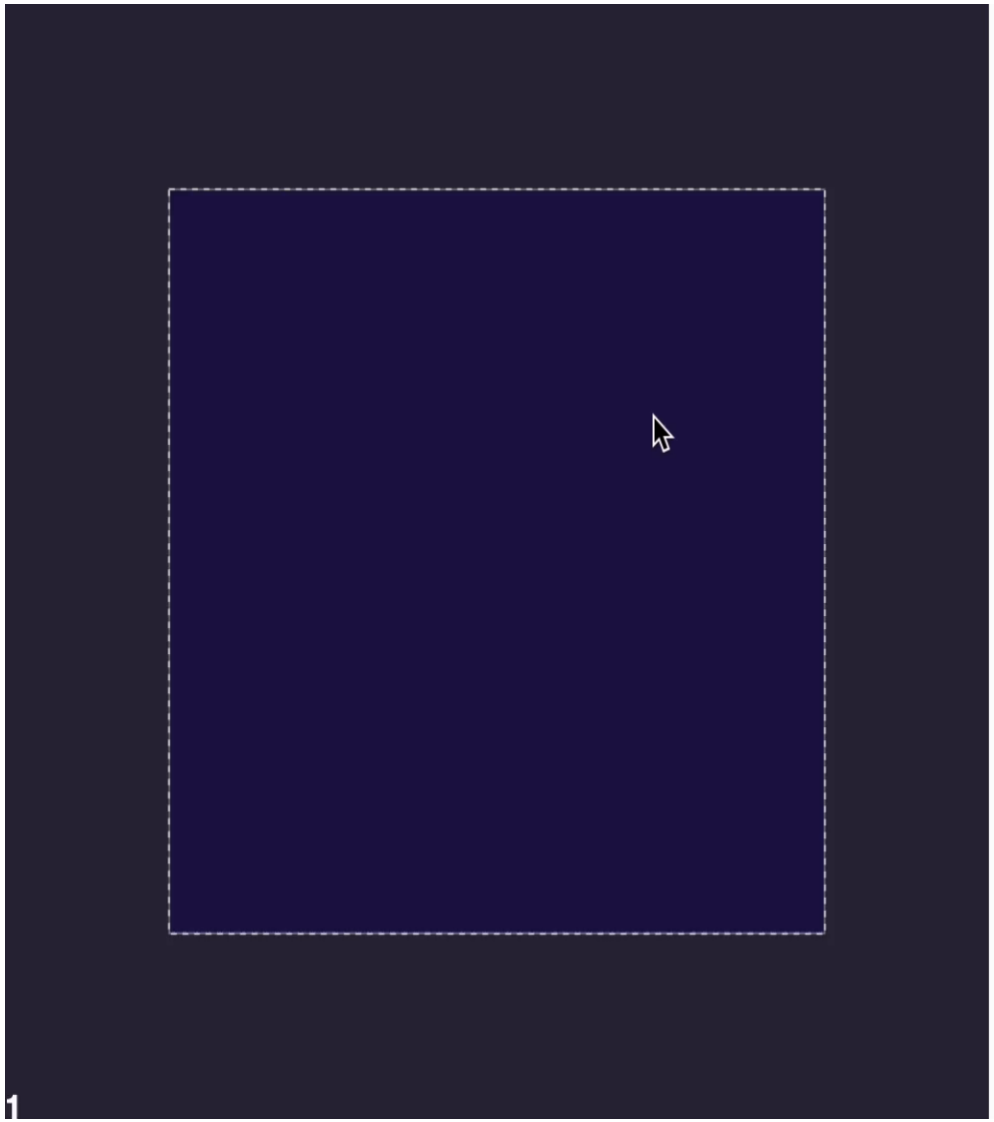


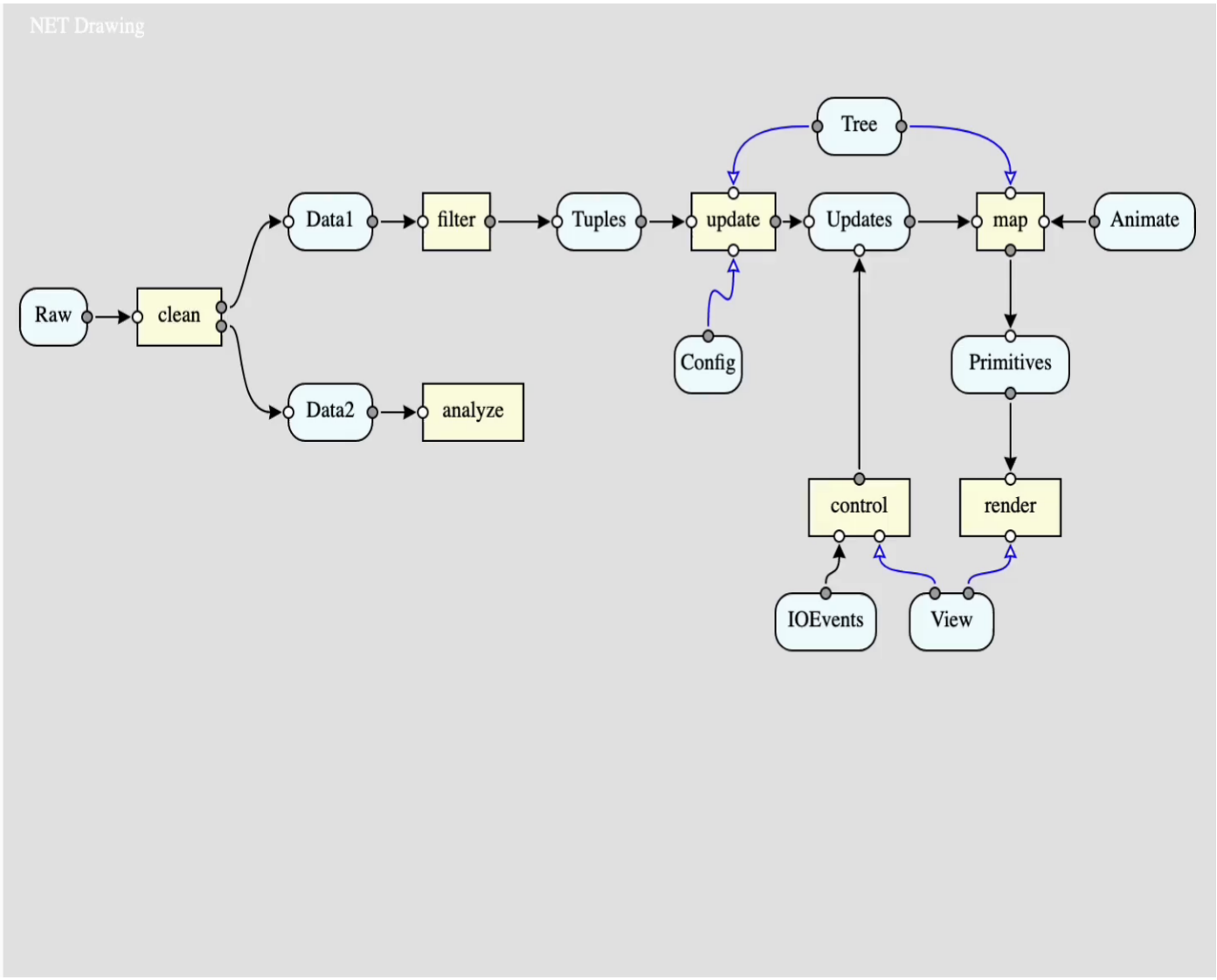
PVA VS. TVA

[https://doi.org/10.2312/eurova.20241108]

Progressive VA	Transient VA
Gradually maturing, converging towards the full view	Stays forever intermediate, never completes but provides the right portion of the data at the right time of the analysis
Tailored to support a single analysis task/interest	Changes and adapts its contents to follow the meandering and unplannable flow of analysis

A FIRST TVA PROTOTYPE






SOLIDUM PETIT IN PROFUNDIS
UNIVERSITAS ARHUSIENSIS