INTRODUCTION TO DATA VISUALIZATION

Hans-Jörg Schulz

ABOUT YOUR LECTURER

Hans-Jörg Schulz

eMail: hjschulz@cs.au.dk

Assoc. Professor @ CS Department, HCC Section

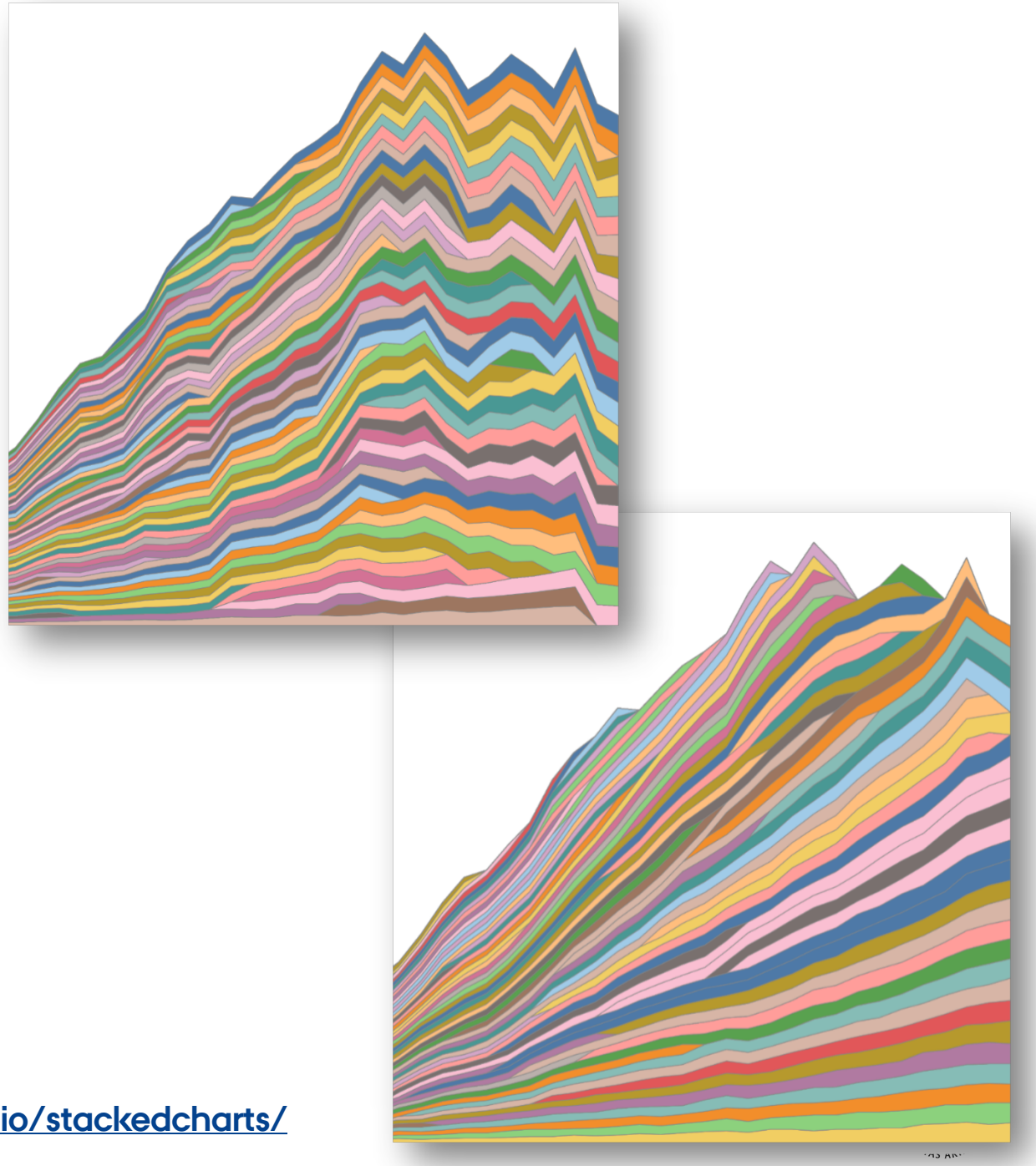
46 yrs., married, 6yo. daughter + 2wo. son

>20 years of experience in visualization gathered at

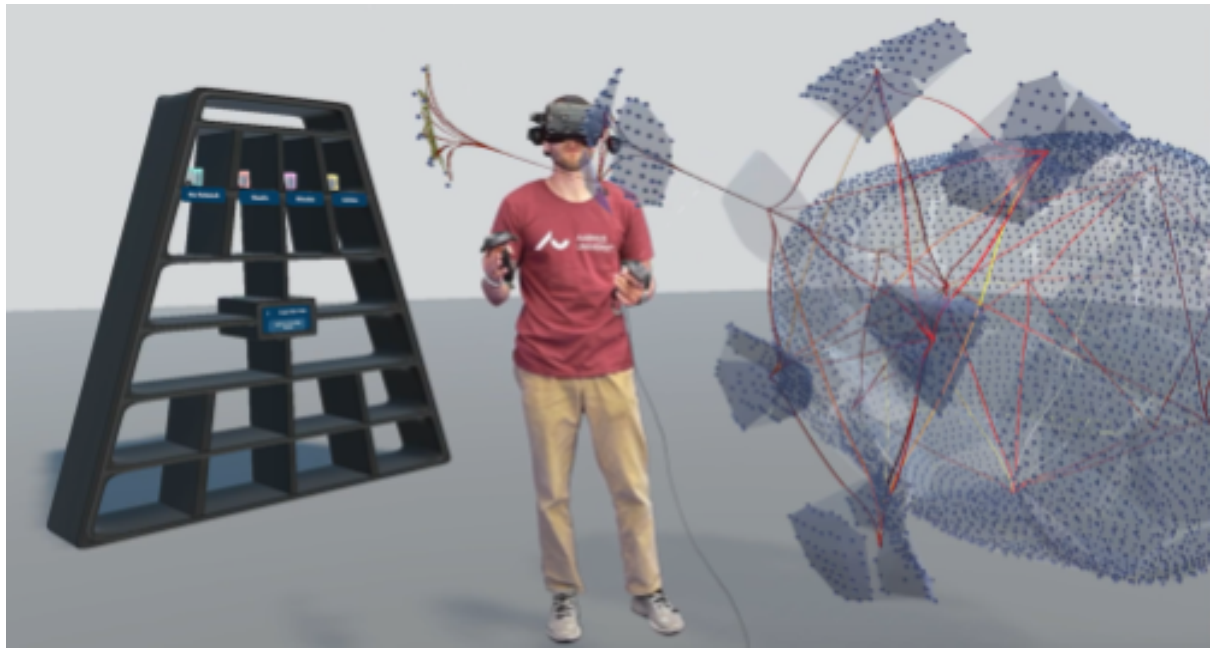
- University of Rostock, Germany
- TU Graz, Austria
- University of Calgary, Canada
- Fraunhofer IGD, Germany

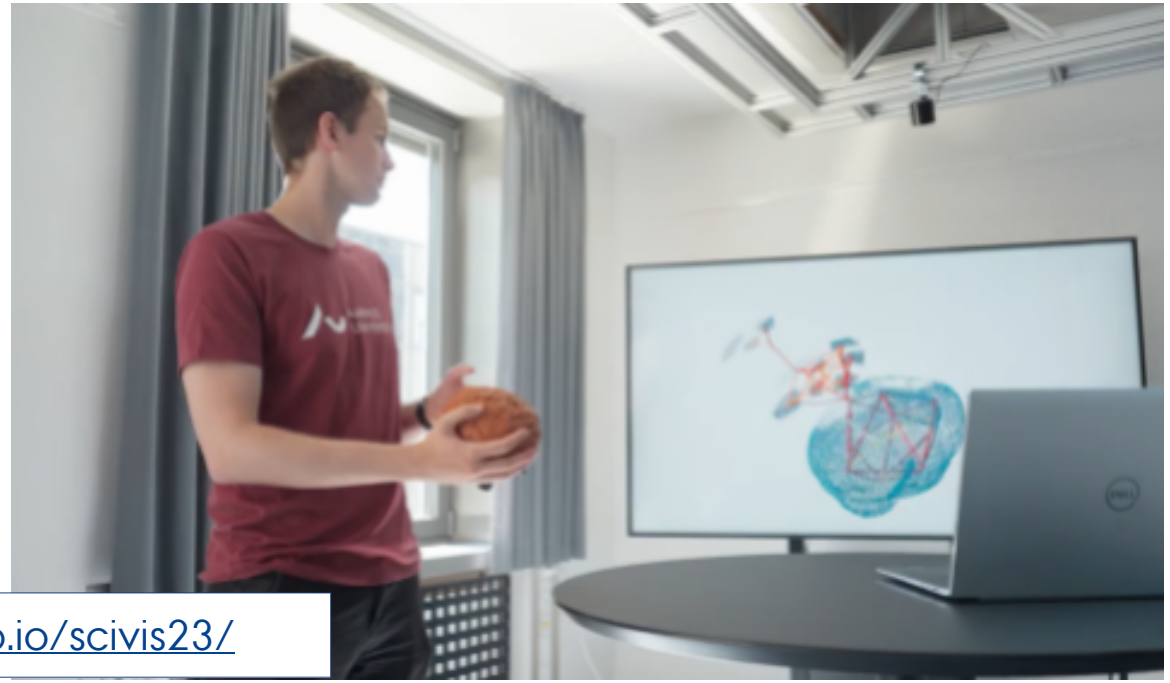
Algorithm 2 FindBestPosition

```
1:  procedure FINDBESTPOSITION(order, f_i)
2:      /* Preprocessing Stage */
3:      gBelow ← 0
4:      gAbove ← f_i
5:      costBelow, costAbove, costLayer ← []
6:      for pos = 0 to length(order) − 1 do
7:          costBelow.add(cost_layer(order[pos], gBelow))
8:          costAbove.add(cost_layer(order[pos], gAbove))
9:          costLayer.add(cost_layer(f_i, gBelow))
10:         gBelow ← gBelow + order[pos]
11:         gAbove ← gAbove + order[pos]
12:     end for
13:     costLayer.add(cost_layer(f_i, gBelow))
14:
15:     /* Testing Stage */
16:     currentCost ← costLayer[0] + Σ_{l=0}^{j−2} costAbove[l]
17:     bestIndex ← 0, bestCost ← currentCost
18:     for pos = 1 to length(order) − 1 do
19:         currentCost += costBelow[pos − 1]
20:         currentCost -= costAbove[pos − 1]
21:         currentCost += costLayer[pos]
22:         currentCost -= costLayer[pos − 1]
23:         if currentCost < bestCost then
24:             bestIndex ← pos, bestCost ← currentCost
25:         end if
26:     end for
27:     return bestIndex
28: end procedure
```

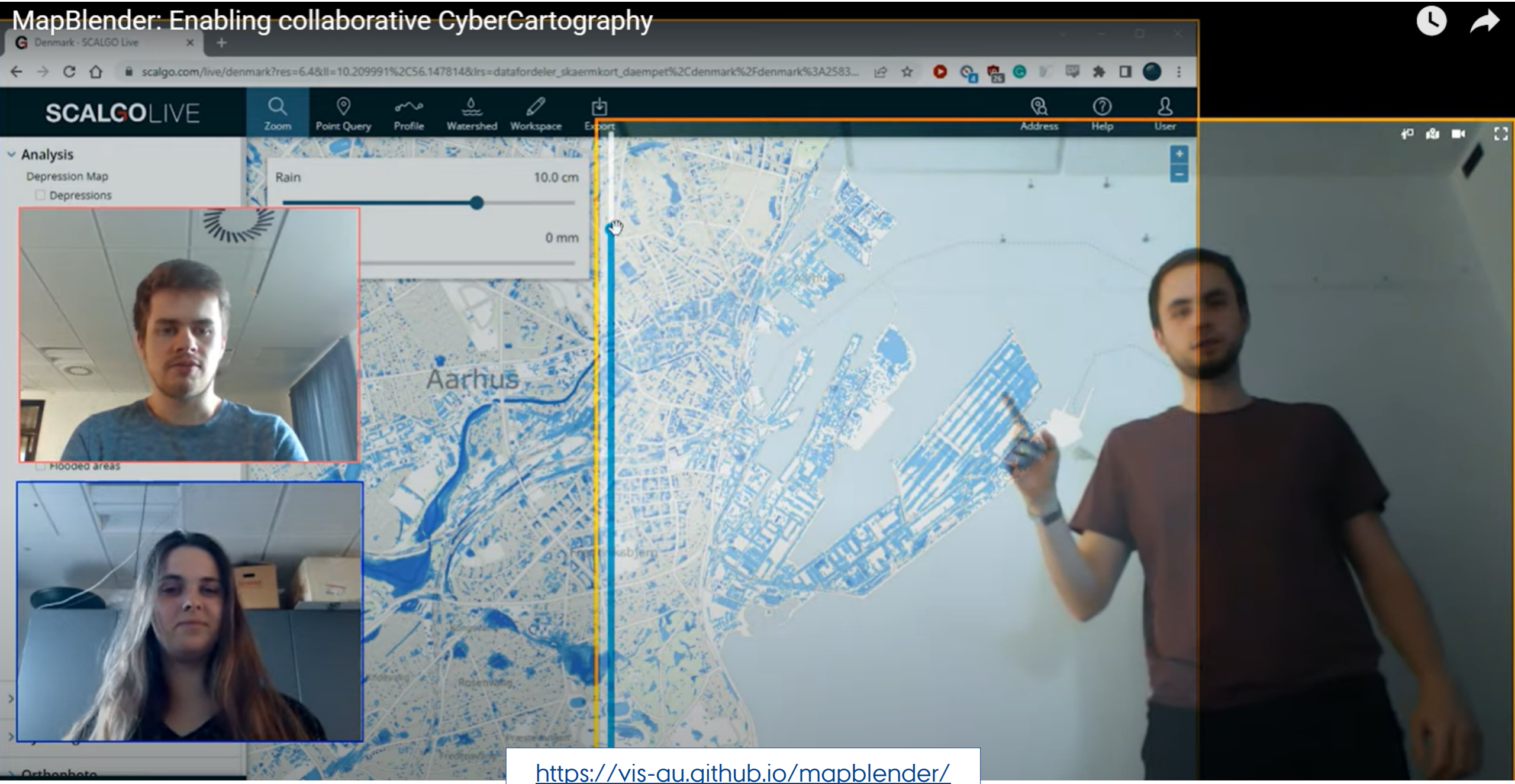


https://vis-au.github.io/stackedcharts/





https://vis-au.github.io/scivis23/


MapBlender: Enabling collaborative CyberCartography
Denmark - SCALGO Live
scalgo.com/live/denmark?res=6.4&ll=10.209991%2C56.147814&lrs=datafordeler_skaermkort_daempet%2Cdenmark%2Fdenmark%3A2583...
SCALGOLIVE
Zoom
Point Query
Profile
Watershed
Workspace
Export
Address
Help
User
Analysis
Depression Map
Depressions
Rain
10.0 cm
0 mm
Flooded areas
Aarhus
https://vis-au.github.io/mapblender/

VAOS

LOAD DATA

Input

Selected file
@M1_f1hz.xlsx
@M2_f1hz.xlsx
@M3_f1hz.xlsx

Stress Filtering

Highest Harmonic to cons
3

EXPLORE HARMONI

Points pr. Quarter Cycle i
300

APPLY

EXPORT GRAPHS

EXPORT DATA

DASHBOARD LISSAJOUS CURVES PIPKIN DIAGRAMS LAYOVER PLOTS RATIOS SIMILARITIES

Pipkin across Samples: Stress vs. Strain

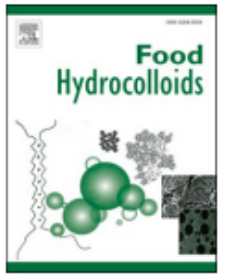
Max Stress Stiffening Thickening Hide Numbers

Centerpoint Distances: Stress vs. Strain

Contents lists available at ScienceDirect

Food Hydrocolloids

journal homepage: www.elsevier.com/locate/foodhyd



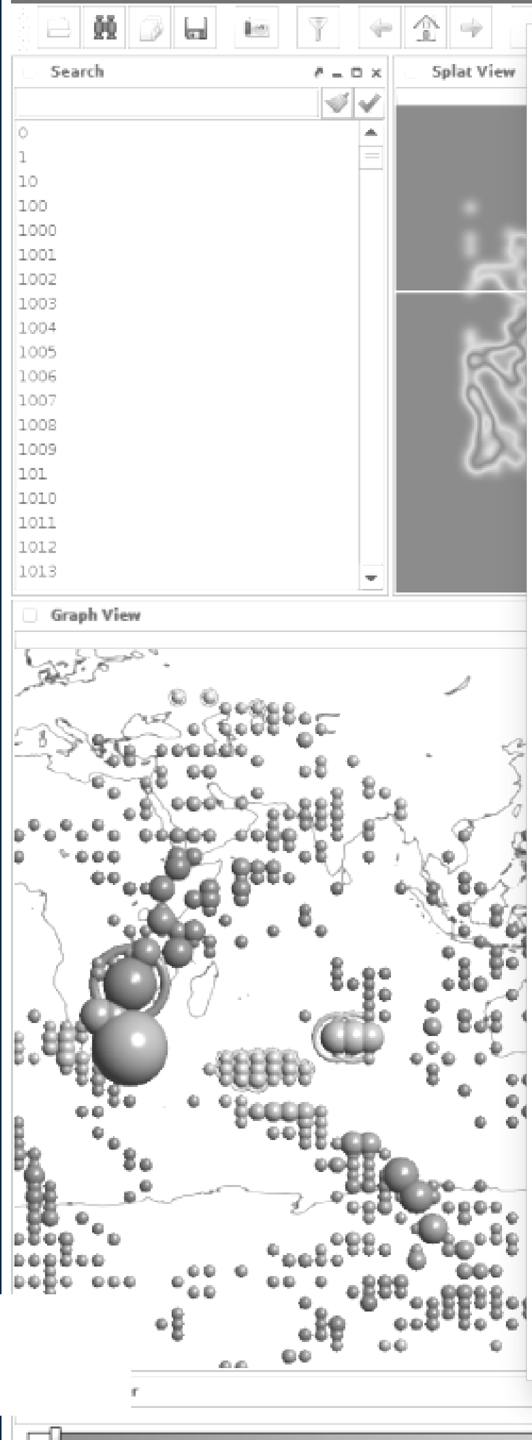
Check for updates

Small and large deformation rheology on pizza cheese as an example of application to study anisotropic properties of food soft materials

Julie Frost Dahl [a,*], Sandra Beyer Gregersen [a], Ulf Andersen [b], Hans-Jörg Schulz [c], Milena Corredig [a,**]

[a] Department of Food Science, CiFood, Center for Innovative Food, Aarhus University, 8200, Aarhus N, Denmark
[b] Arla Foods Amba, 8200, Aarhus N, Denmark
[c] Department of Computer Science, Aarhus University, 8200, Aarhus N, Denmark

https://vis-au.github.io/vaos/



Nonlin. Processes Geophys., 22, 545–570, 2015
www.nonlin-processes-geophys.net/22/545/2015/
doi:10.5194/npg-22-545-2015

Nonlinear Processes in Geophysics
Open Access

Review: visual analytics of climate networks

T. Nocke[1], S. Buschmann[2], J. F. Donges[1,3], N. Marwan[1], H.-J. Schulz[4], and C. Tominski[5]

[1]Potsdam Institute for Climate Impact Research, Telegrafenberg A31, 14473 Potsdam, Germany
[2]Hasso Plattner Institute, Prof.-Dr.-Helmert-Str. 2–3, 14482 Potsdam-Babelsberg, Germany
[3]Stockholm Resilience Centre, Stockholm University, Kräftriket 2B, 114 19 Stockholm, Sweden
[4]Fraunhofer Institute for Computer Graphics Research, Joachim-Jungius-Str. 11, 18059 Rostock, Germany
[5]Institute for Computer Science, University of Rostock, Albert-Einstein-Str. 22, 18059 Rostock, Germany

Correspondence to: T. Nocke (nocke@pik-potsdam.de)

Received: 1 April 2015 – Published in Nonlin. Processes Geophys. Discuss.: 30 April 2015
Revised: 12 August 2015 – Accepted: 19 August 2015 – Published: 23 September 2015

Abstract. Network analysis has become an important approach in studying complex spatiotemporal behaviour within geophysical observation and simulation data. This new field produces increasing numbers of large geo-referenced networks to be analysed. Particular focus lies currently on the network analysis of the complex statistical interrelationship structure within climatological fields. The standard procedure for such network analyses is the extraction of network measures in combination with static standard visualisation methods. Existing interactive visualisation methods and tools for geo-referenced network exploration are often either not amples from global, regional, and multi-layered climate networks.

1 Introduction

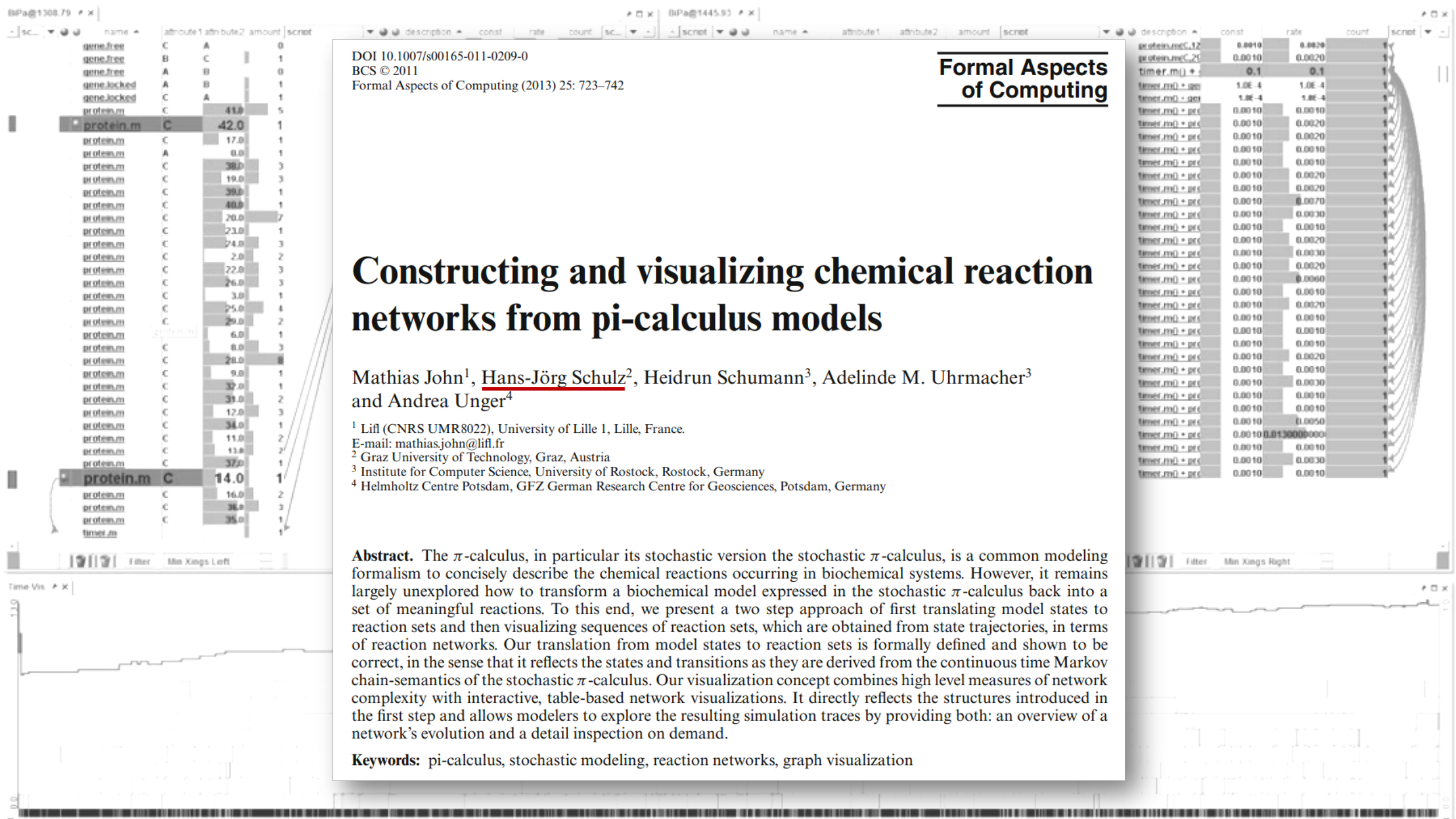
Data visualisation created within scientific contexts aims at the provision of meaningful visual representations that support the exchange of working results and provide scientists with appropriate tools to reveal relations and hidden patterns

Towards a Contextualized Visual Analysis of Heterogeneous Manufacturing Data

Mario Aehnelt[1], Hans-Jörg Schulz[2], and Bodo Urban[1]

[1] Fraunhofer IGD Rostock, Germany
{mario.aehnelt,bodo.urban}@igd-r.fraunhofer.de
[2] University of Rostock, Germany
hjschulz@informatik.uni-rostock.de

Abstract. Visual analysis spanning multiple data sources usually requires the integration of multiple specialized applications to handle their heterogeneity. This is also true in manufacturing, where data about orders, personnel, workloads, maintenance, etc. must be analyzed together to make well-founded management decisions. Yet, the orchestration of multiple data sources and applications poses challenges to the software infrastructure and to the analyst. We present a three-tiered approach to cope with these challenges. In a first step, we establish a domain-dependent workflow as the mental model of the analyst. Based on the novel concept of contextualization, we then align the different applications with this model for their meaningful integration. In a third step, we incorporate the data according to its use in the aligned applications by means of a service-based architecture. By starting the integration on the user level, we are able to pragmatically target and streamline the required integration to a degree that is technically achievable and interactively manageable. We exemplify our approach with the Plant@Hand system for integrating manufacturing data and applications.



DOI 10.1007/s00165-011-0209-0
BCS © 2011
Formal Aspects of Computing (2013) 25: 723–742

Formal Aspects of Computing

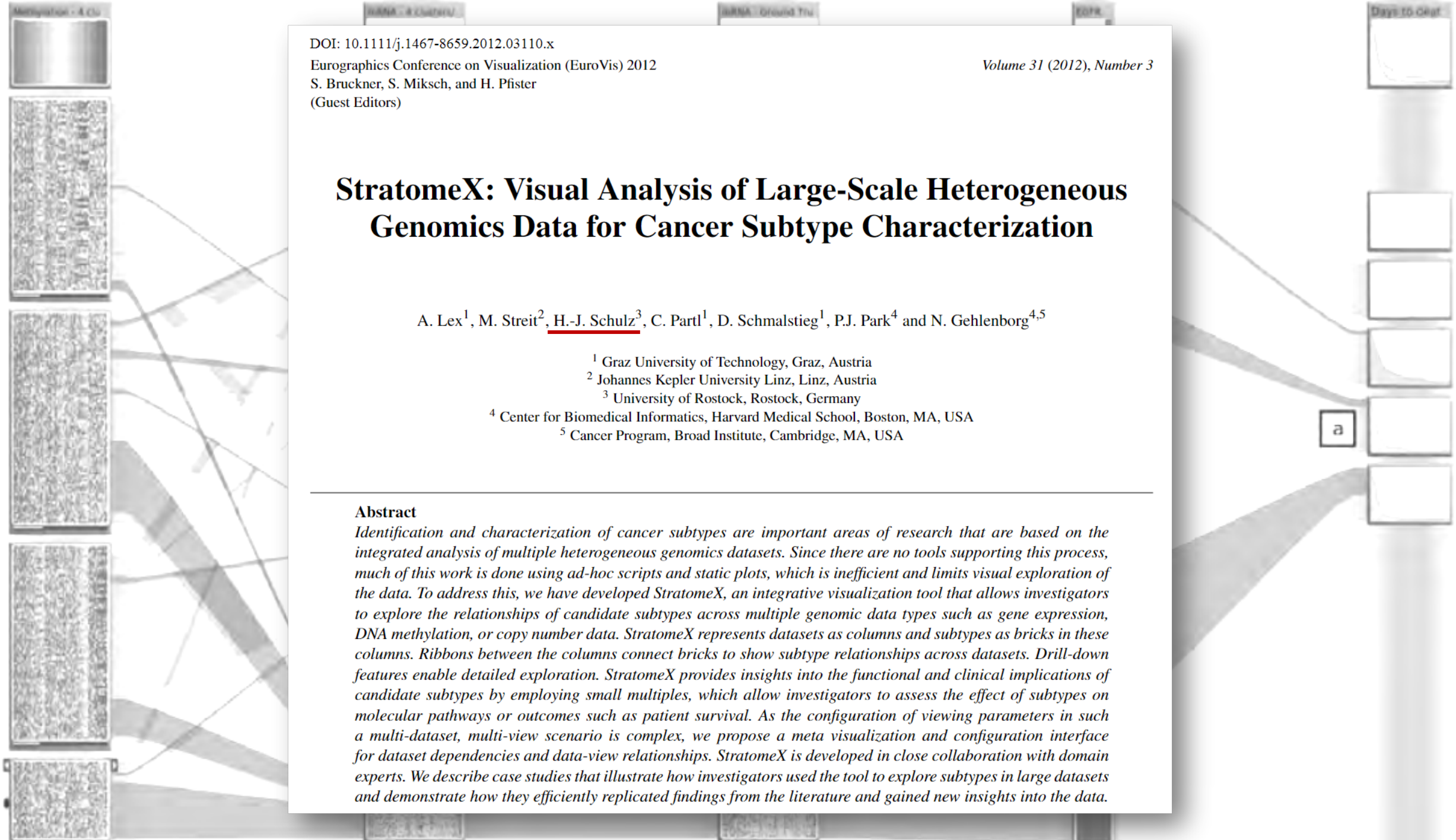
Constructing and visualizing chemical reaction networks from pi-calculus models

Mathias John[1], Hans-Jörg Schulz[2], Heidrun Schumann[3], Adelinde M. Uhrmacher[3] and Andrea Unger[4]

[1] Lifl (CNRS UMR8022), University of Lille 1, Lille, France.
E-mail: mathias.john@lifl.fr
[2] Graz University of Technology, Graz, Austria
[3] Institute for Computer Science, University of Rostock, Rostock, Germany
[4] Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Potsdam, Germany

Abstract. The π-calculus, in particular its stochastic version the stochastic π-calculus, is a common modeling formalism to concisely describe the chemical reactions occurring in biochemical systems. However, it remains largely unexplored how to transform a biochemical model expressed in the stochastic π-calculus back into a set of meaningful reactions. To this end, we present a two step approach of first translating model states to reaction sets and then visualizing sequences of reaction sets, which are obtained from state trajectories, in terms of reaction networks. Our translation from model states to reaction sets is formally defined and shown to be correct, in the sense that it reflects the states and transitions as they are derived from the continuous time Markov chain-semantics of the stochastic π-calculus. Our visualization concept combines high level measures of network complexity with interactive, table-based network visualizations. It directly reflects the structures introduced in the first step and allows modelers to explore the resulting simulation traces by providing both: an overview of a network's evolution and a detail inspection on demand.

Keywords: pi-calculus, stochastic modeling, reaction networks, graph visualization



DOI: 10.1111/j.1467-8659.2012.03110.x
Eurographics Conference on Visualization (EuroVis) 2012
S. Bruckner, S. Miksch, and H. Pfister
(Guest Editors)

Volume 31 (2012), Number 3

StratomeX: Visual Analysis of Large-Scale Heterogeneous Genomics Data for Cancer Subtype Characterization

A. Lex[1], M. Streit[2], H.-J. Schulz[3], C. Partl[1], D. Schmalstieg[1], P.J. Park[4] and N. Gehlenborg[4,5]

[1] Graz University of Technology, Graz, Austria
[2] Johannes Kepler University Linz, Linz, Austria
[3] University of Rostock, Rostock, Germany
[4] Center for Biomedical Informatics, Harvard Medical School, Boston, MA, USA
[5] Cancer Program, Broad Institute, Cambridge, MA, USA

Abstract

Identification and characterization of cancer subtypes are important areas of research that are based on the integrated analysis of multiple heterogeneous genomics datasets. Since there are no tools supporting this process, much of this work is done using ad-hoc scripts and static plots, which is inefficient and limits visual exploration of the data. To address this, we have developed StratomeX, an integrative visualization tool that allows investigators to explore the relationships of candidate subtypes across multiple genomic data types such as gene expression, DNA methylation, or copy number data. StratomeX represents datasets as columns and subtypes as bricks in these columns. Ribbons between the columns connect bricks to show subtype relationships across datasets. Drill-down features enable detailed exploration. StratomeX provides insights into the functional and clinical implications of candidate subtypes by employing small multiples, which allow investigators to assess the effect of subtypes on molecular pathways or outcomes such as patient survival. As the configuration of viewing parameters in such a multi-dataset, multi-view scenario is complex, we propose a meta visualization and configuration interface for dataset dependencies and data-view relationships. StratomeX is developed in close collaboration with domain experts. We describe case studies that illustrate how investigators used the tool to explore subtypes in large datasets and demonstrate how they efficiently replicated findings from the literature and gained new insights into the data.

How to cite this site?

Check out other surveys!

treevis.net - A Visual Bibliography of Tree Visualization 2.0 by Hans-Jörg Schulz

v.04-SEP-2023

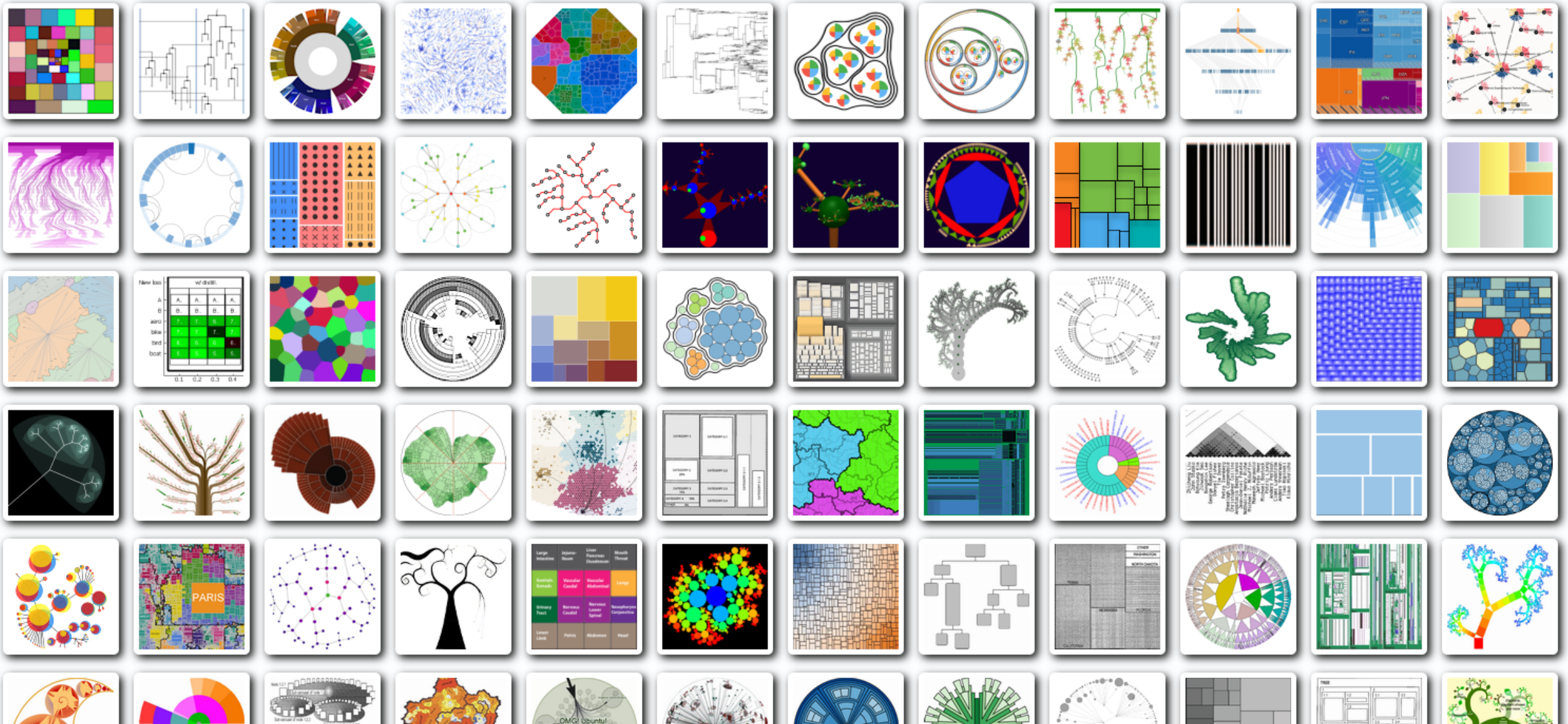
Dimensionality: All

Representation: All

Alignment: All

Fulltext Search

Techniques Shown: 339



Arvind Srinivasan

AARHUS UNIVERSITY

Who are you?

- 🇩🇰 **2024 – present**: **Ph.D. Student** in **Department of Computer Science** at the **Center for Anytime Anywhere Analytics** in the Ubiquitous Computing Group, supervised by Dr. Niklas Elmqvist. Current research interests include Attention Aware Visualizations and Collaborative Immersive Analytics
- 🇺🇸**2023**: Received a **Masters in Human Computer Interaction** at the **University of Maryland**.
- Besides **JavaScript**, **Python** and **Go**, I'm best with **D3**, good with **Vega-Lite**, familiar with **Tableau**.
- With me, you can be assured that: At best, I solve your problem, at worst, you're inspired towards it.

How do I reach you?

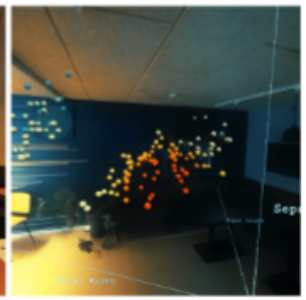
Email: arvind@cs.au.dk
Office: Room 126, Building 5346
Åbogade 34, 8200 Aarhus N

When are you available?

Mondays and **Wednesdays** after class in my office / opposite to Hans'.

If you have any purely technical questions, feel free to reach out!

(a) AAV2D condition.



(b) AAV3D condition.

Peichen Liu

- **Basic Info:**
  - First year PhD Student in Human-centered Computing, CS Department
  - Email: pliu@cs.au.dk
- **Research Interest:**
  - Visual Analytic
  - Health Data Processing
- **Office Time:**
  - Every Wednesday 14:00-16:00 in Hopper-118, Building 5432
- **Technology Stack:**
  - C++(Qt), Python
- **Education Background:**
  - MSc in Computer Science and Engineering @DTU, DK
- **During the Course:**
  - I will support you in exercises and assignments, and I look forward to learning and exploring ideas together with you.

This is YOUR course!

YOU AND YOUR COURSE MATES

Bioinformatics	2
Business Economy	7
Cognitive Science	1
Computer Engineering	16
Computer Science (MSc)	41
Data Science (BSc)	2
Data Science (MSc)	16
Electrical Engineering	2
IT Product Development (BSc)	1
IT Product Development (MSc)	1
Mathematics	1
(blank)	8

Total = 98

(as of 20-AUG-2025)

Brightspace = 102

(as of 24-AUG-2025)

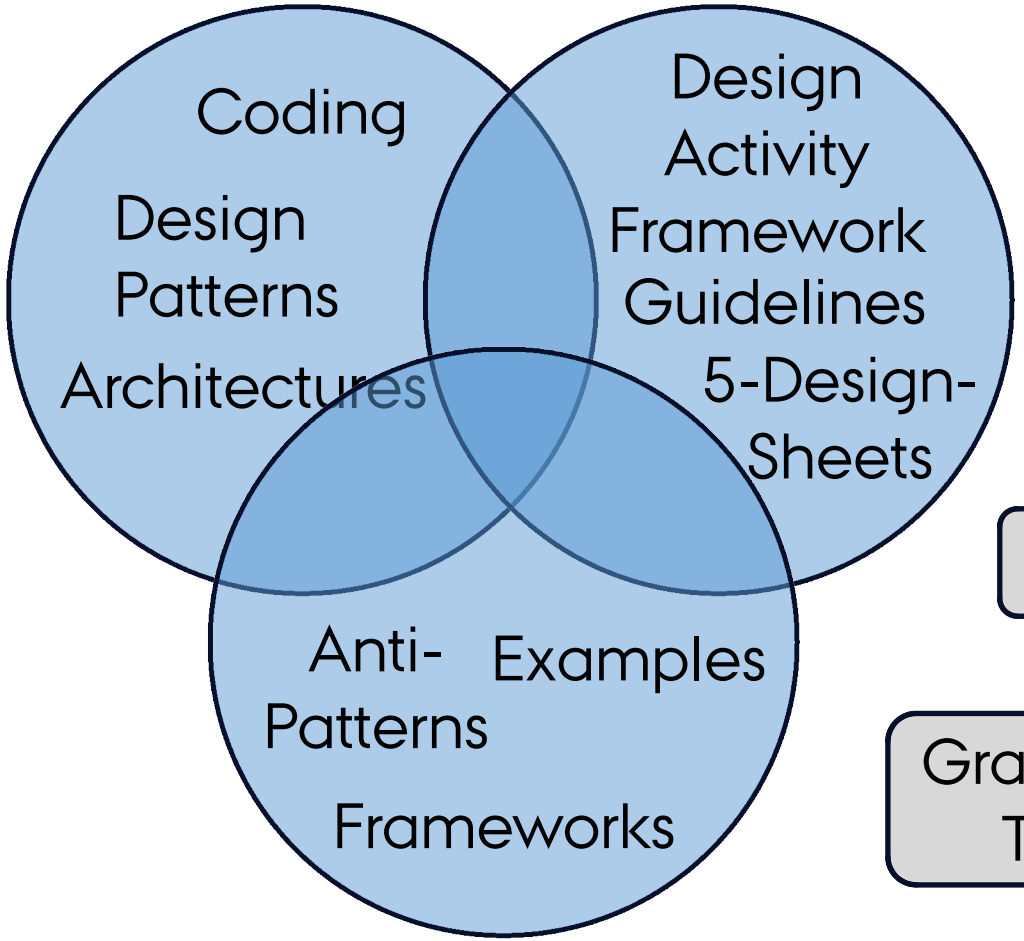
VISUALIZATION IS FACETED

Statistics & Machine Learning

Databases

Computer Graphics

Implementing Visualizations



Designing Visualizations

Web Development

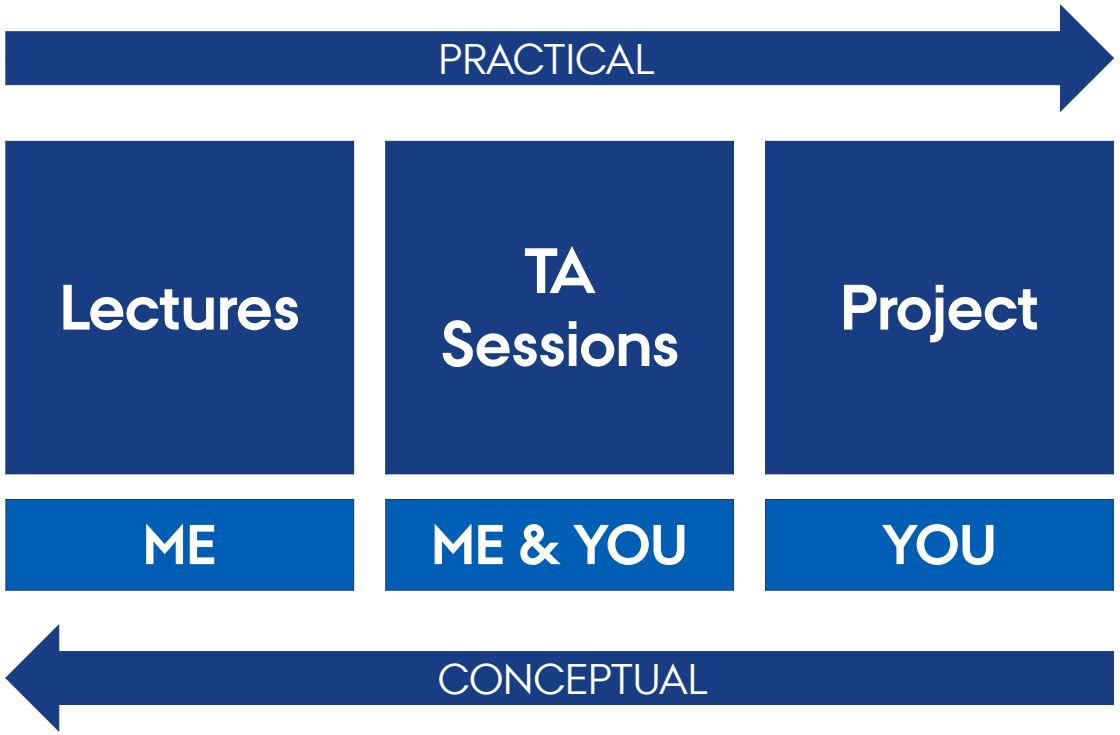
Cognitive Psychology

Interaction Design

Graphic Design & Typography

Critiquing & Evaluating Visualizations

STRUCTURE OF THIS LECTURE



Lectures:

- Visualization Concepts
- Visualization Techniques
- Visualization Algorithms

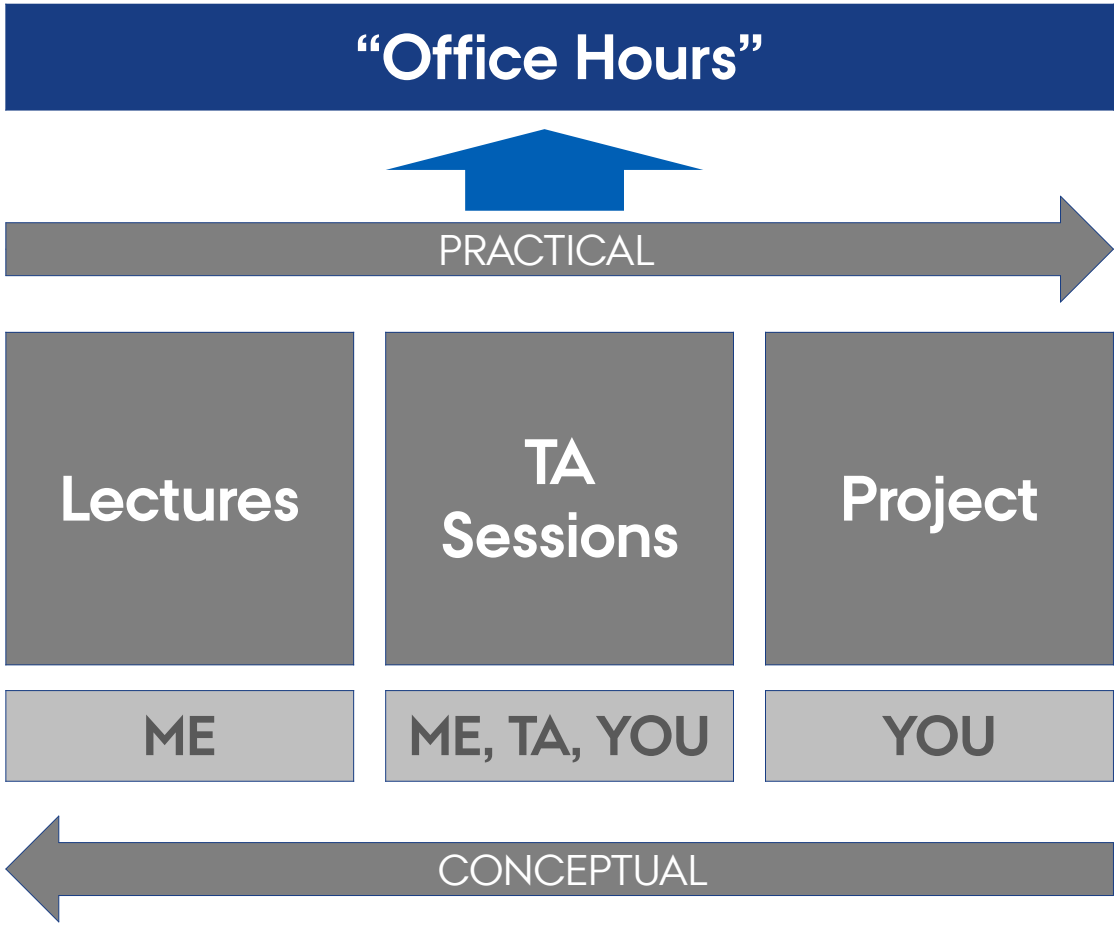
TA Sessions:

- Practical Exercises
- Visualization Critique
- Quizzes

Project:

- Visualization Design
- Implementation
- Actual Use & Users

STRUCTURE OF THIS LECTURE



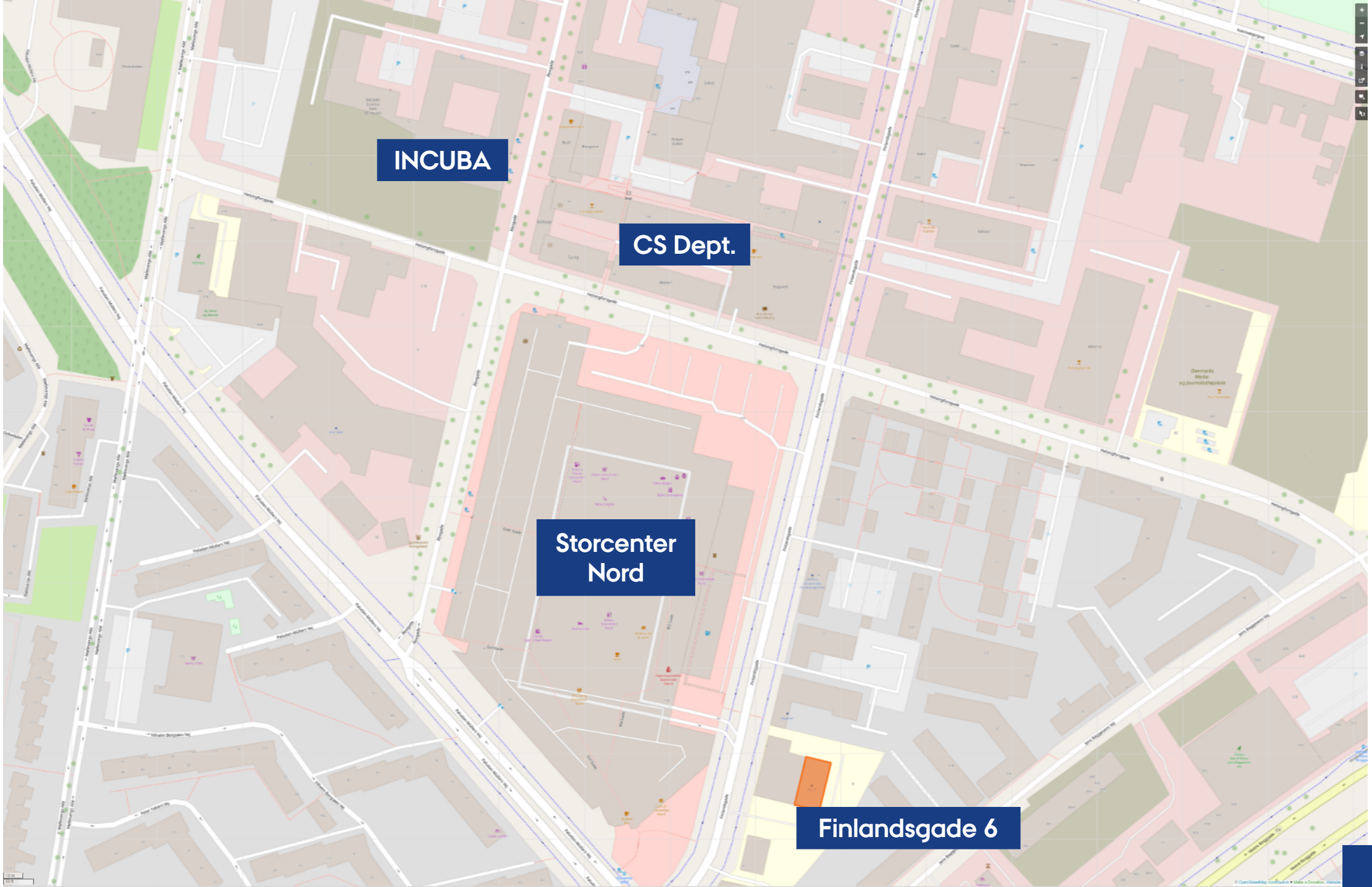
"Office Hours"

- Questions about lecture content
- Feedback (giving & getting)
- Bring-your-own-vis discussions
- Topics not covered in the lecture
- ...

TIMES & DATES

1st Half of the Semester: Visualization Building Blocks

Week	Lecture Mondays 9-11	3rd Hour Mondays 11-12	TA Session Wednesdays 12-14
35	Introduction to DataVis	The DataVis Project	Visualization Design I
36	Visual Encoding	Advanced Color Coding	Visualization Critique I
37	Basic Charts	Vis Critique Exercise 1+2	Visualization Critique II
38	Interaction for DataVis	Vis Critique Exercise 3+4	Color & Chart Exercises
39	Visualization Design	Vis Critique Exercise 5+6	5-Design-Sheets Exercise
40	Data Preprocessing	Black Hat Vis Preparations	Black Hat Vis Presentations
41	Round-up of 1st Half	Quiz	Exam Simulation


INCUBA
CS Dept.
Storcenter Nord
Finlandsgade 6
iNano

TIMES & DATES

1st Half of the Semester: Visualization Building Blocks

Week	Lecture Mondays 9-11	3rd Hour Mondays 11-12	TA Session Wednesdays 12-14
35	Introduction to DataVis	The DataVis Project	
36	Perception	Analysis Reporting	Visualization Critique I
37	Basic Charts	Vis Critique Exercise 1+2	Visualization Critique II
38	Idiom for DataVis	Vis Critique Exercise 3+4	Color & Chart Exercises
39	Visualization Design	Vis Critique Exercise 5+6	5-Design-Sheets Exercise
40	Data Preprocessing	Black Hat Vis Preparations	Black Hat Vis Presentations
41	Round-up of 1st Half	Quiz	Exam Simulation

Recordings: Panopto -> linked on Brightspace

Live Stream: on Zoom

This is thought as a fall-back and not as a replacement for the physical lectures / TA sessions!

- Assignments
- Discussions
- Meetings - Zoom
- Quizzes
- Portfolio
- Videos - Panopto

Minimize setup

This course is **open** fo ts

Course setup

Activity Feed

Create a post...

Latest Posts

Me • Message • 24 August 2025 18:47

Yes, despite rumors to the contrary, there will be a lecture and TA session in the coming week (Monday 25-AUG-2025 and Wednesday 27-AUG-2025). See you tomorrow morning, Hans-Jörg

Calendar

Sunday, 24 August 2025

Upcoming events

There are no events to display.

Create an event

TIMES & DATES

Office Hours:

Hans-Jörg: Mondays 12-14, Åbogade 34, Hopper Bldg., **5346-117**

Arvind: Mondays 12-14, Åbogade 34, Hopper Bldg., **5346-126**

Peichen: Wednesdays 14-16, Åbogade 34, Hopper Bldg., **5346-118**

DAVI ORAL EXAM

Qualification: written report on course project + video presentation

Exam structure: 20 minutes, internal co-examiner

1. oral presentation of course project + Q&A
2. answering questions about the lecture material
3. critiquing a visualization

PRO TIPS

- Read the material before the lecture
- Lecture > Life Stream > Recorded Video > Sped-up Video

ELSEVIER

Computers & Education

Volume 55, Issue 1, August 2010, Pages 313-319

Online lecture accessibility and its influence on performance in skills-based courses

Ada Le [a], Steve Joordens [a], Sophie Chrysostomou [b], Raymond Grinnell [b]

https://doi.org/10.1016/j.compedu.2010.01.017

Abstract

At the University of Toronto at Scarborough, we provide enhanced flexibility to our students using a blended-learning approach (i.e., the webOption) whereby students can attend lectures live, watch them online at their convenience, or both. The current research examines the use of pause and seeks features afforded by the webOption interface and how these features are related to students' learning approaches and their performance in calculus courses. These courses emphasize the teaching of mathematical proofs; cognitive skills that are enhanced with practice (Schneider & Shiffrin, 1977). Access to online lectures allows students to re-experience the professor as they teach these skills. Given this, it was predicted that use of the webOption might be especially potent in these learning contexts. The results we report here do not confirm that prediction. Students do use and appreciate the features of the webOption, however, those students who augmented their class attendance with online viewing, and those who used the lecture-control features the most, were actually the students who performed most poorly. We interpreted the results to be due to different learning strategies and the manner in which these strategies interact with course content. Our results suggest that using the pause feature is related to a surface strategy of learning, which is in turn related to poorer performance in the course.

ELSEVIER

Computers & Education

Volume 61, February 2013, Pages 185-192

Evaluating the use and impact of lecture recording in undergraduates: Evidence for distinct approaches by different groups of students

Wendy Leadbeater [a], Tom Shuttleworth [b], John Couperthwaite [c], Karl P. Nightingale [b]

https://doi.org/10.1016/j.compedu.2012.09.011

Abstract

Lecture recordings are increasingly used to supplement lecture attendance within higher education, but their impact on student learning remains unclear. Here we describe a study to evaluate student use of lecture recordings and quantify their impact on academic performance. Questionnaire responses and online monitoring of student's access to recordings indicate that ~75% students use this material, the majority in a targeted manner. In contrast, a small subset of students (~5%) are highly dependent on recordings downloading every lecture, and viewing the material for long periods, such that this represents a large proportion of their independent study. This *'high user'* group is atypical, as it contains a high proportion of dyslexic and Non-English Speaking Background students. Despite high usage, lecture recordings do not have a significant impact on academic performance, either across the cohort or with students that use the recordings. Overall, this approach appears to be beneficial, but may reduce lecture attendance and encourage surface learning approaches in a minority of students.

EMAIL FROM A FORMER STUDENT

“I had no clue why the second part went so poorly until I realized that the multivariate visualization was covered right after the Autumn break while I was still away on an extended holiday; I watched the lecture online but as we know and is backed up by research - and as I now have firsthand experience with -, that is much less efficient for memorizing things. I of course went through the course material while preparing but that lecture just didn't stick as much as the others.”

PRO TIPS

- Read the material before the lecture
- Lecture > Life Stream > Recorded Video > Sped-up Video
- Pick a new seat every time

The Variation Effect: How seating plans might be undermining learning

> *Observe always that everything is the result of a change, and get used to thinking that there is nothing Nature loves so well as to change existing forms and to make new ones like them.*
>
> *Marcus Aurelius*

It is a truth universally acknowledged that a teacher in possession of a large roomful of children must be in want of a carefully crafted seating plan. Secondary schools in particular have normalised the idea that children should sit in the seat where they will be least distracted and best able to learn. There are many excellent reasons to use seating plans: they're a great way to learn students' names, keep order and establish routines. These are all sound and it is not my intention to unpick them.

PRO TIPS

- Read the material before the lecture
- Lecture > Life Stream > Recorded Video > Sped-up Video
- Pick a new seat every time
- Take notes

ELSEVIER

Computers & Education

Volume 85, July 2015, Pages 14-22

To provide or not to provide course PowerPoint slides? The impact of instructor-provided slides upon student attendance and performance

Debra L. Worthington [a], David G. Levasseur [b]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.compedu.2015.02.002

Highlights

- We examined student progress across two semesters comparing three slide conditions: No vs. partial vs. full slides.
- We measured actual student attendance: Instructor-provided slides had no impact on actual class attendance.
- We examined student use of partial vs. full slides in class note-taking and studying for exams.
- Instructor-provided slides adversely impacted student course performance on exam items.
- There were no differences by academic group (high/med/low GPA) on use of slides.

ELSEVIER

Computers & Education

Volume 173, November 2021, 104283

Impact of the provision of PowerPoint slides on learning

Samuel P. León [a], Inmaculada García-Martínez [b]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.compedu.2021.104283

Highlights

- First within-subject study design to analyse this topic.
- Providing slides to students impacts negatively on their academic performance.
- The slide availability adversely impacts on the students' free attendance to class.
- Students' academic engagement and study strategies modulated student performance.
- The slides are a supplement created by teachers for teaching.

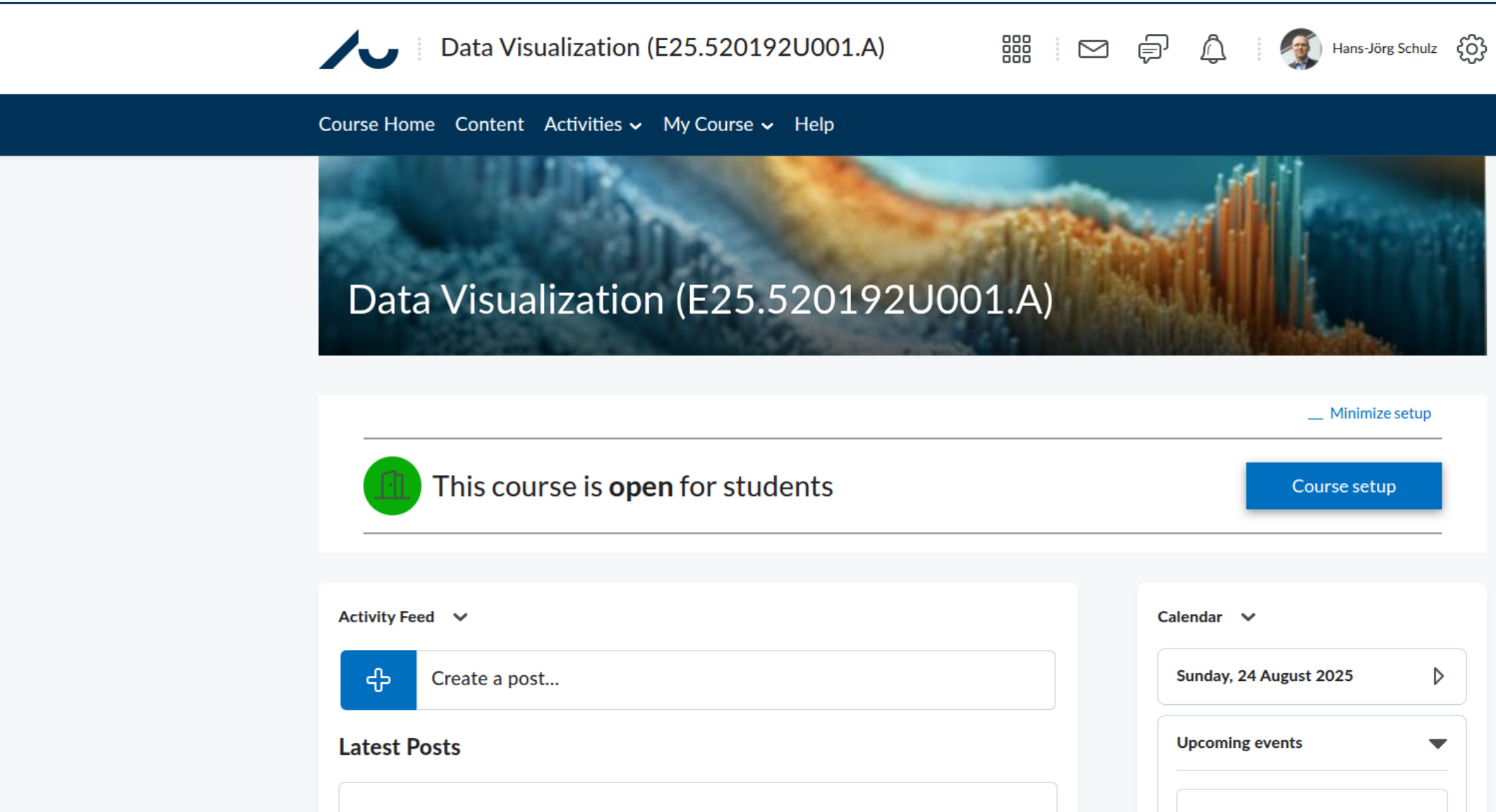
Abstract

PowerPoint is a basic tool for university teaching. Teachers use it extensively for presenting material. At times, it is used as a guide for organizing lessons, at other

PRO TIPS

- Read the material before the lecture
- Lecture > Life Stream > Recorded Video > Sped-up Video
- Pick a new seat every time
- Take notes
- Make use of the office hours
- Use the General Discussion Board on Brightspace

A TOUR THROUGH THE BRIGHTSPACE SITE



DATA VISUALIZATION

AARHUS
UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

HANS-JÖRG SCHULZ
ASSOCIATE PROFESSOR

WHAT IS DATAVIS?

"Visual representations of data enable us to **communicate a large amount of information** to our viewers." [Marty 2009]

All taken from Chen et al. "What Is Visualization Really For?", Springer 2014

VISUALIZATION FOR COMMUNICATION

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Mean(x) = 9.0

Mean(y) = 7.5

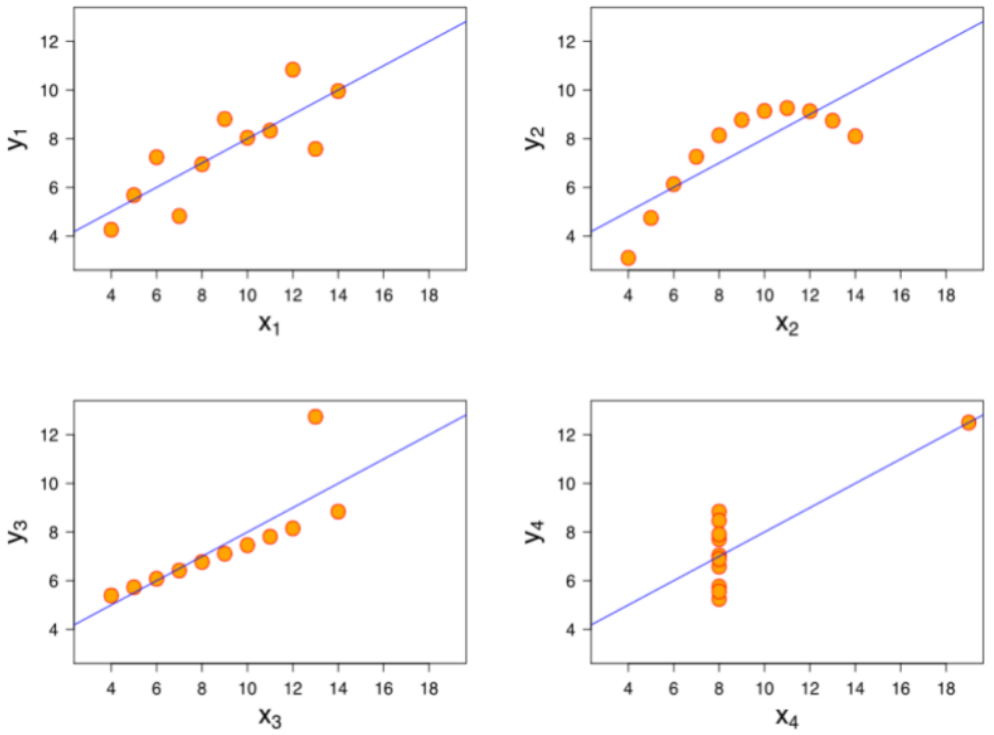
Variance(x) = 11.0

Variance(y) = 4.1

Correlation(x,y) = 0.816

Regression line:

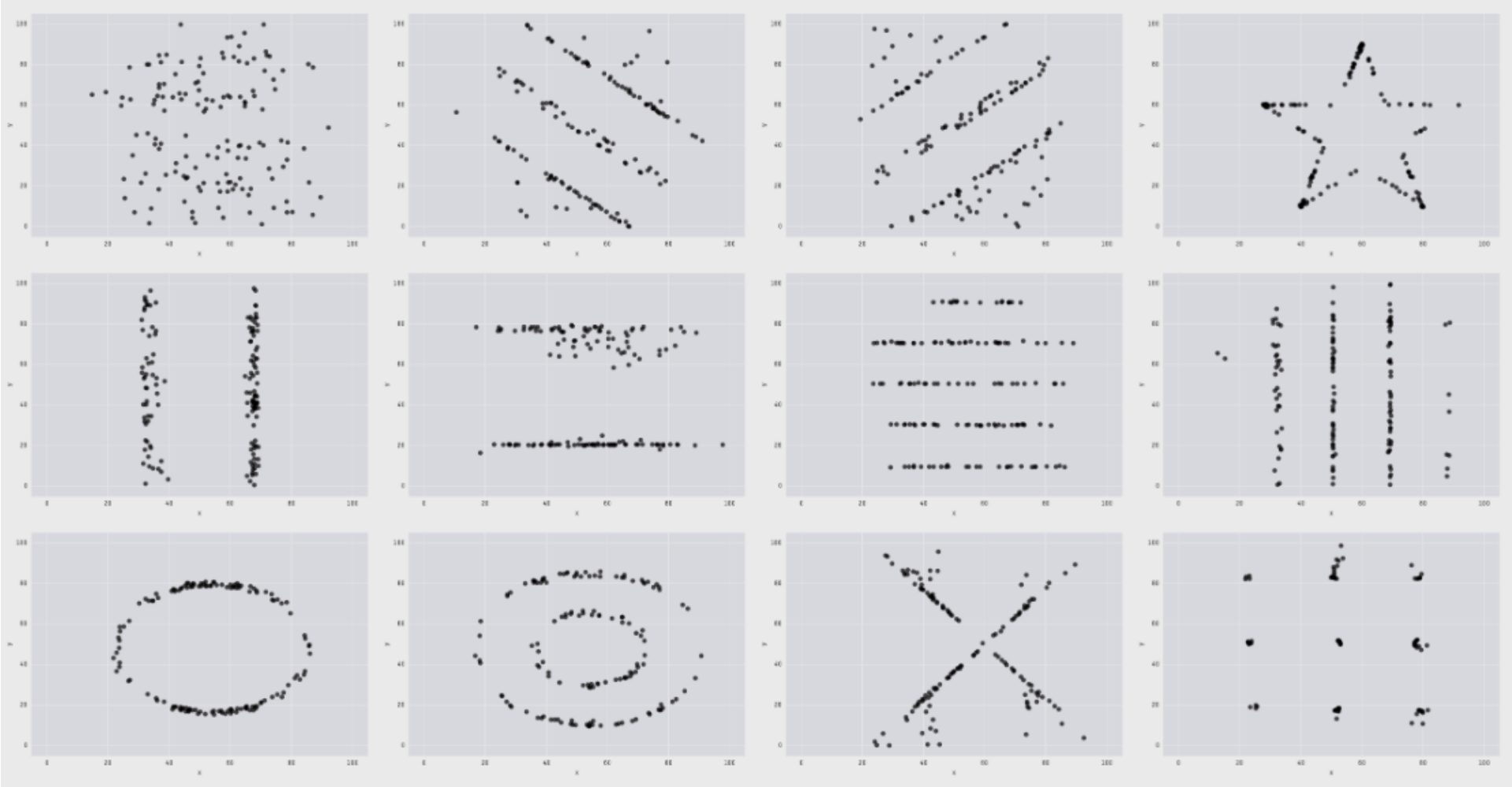
$y = 3.0 + 0.5x$



Source: [Anscombe 1973]

Image Source: Wikipedia

THE DATASAUR



Mean(x) = 54.26

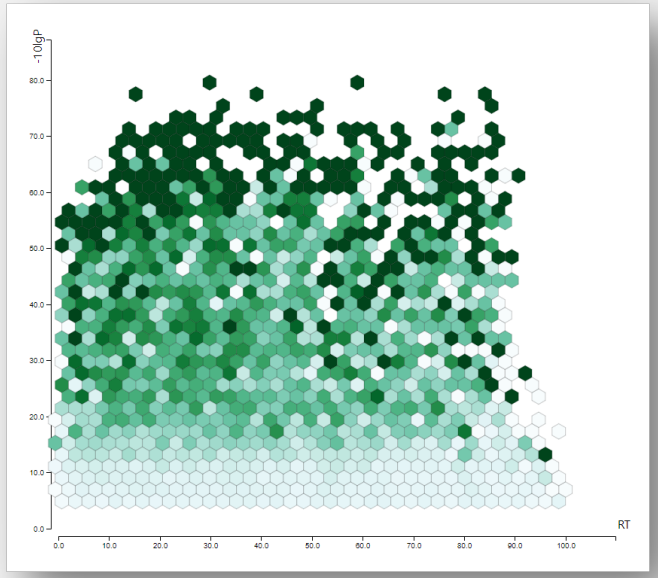
Mean(y) = 47.83

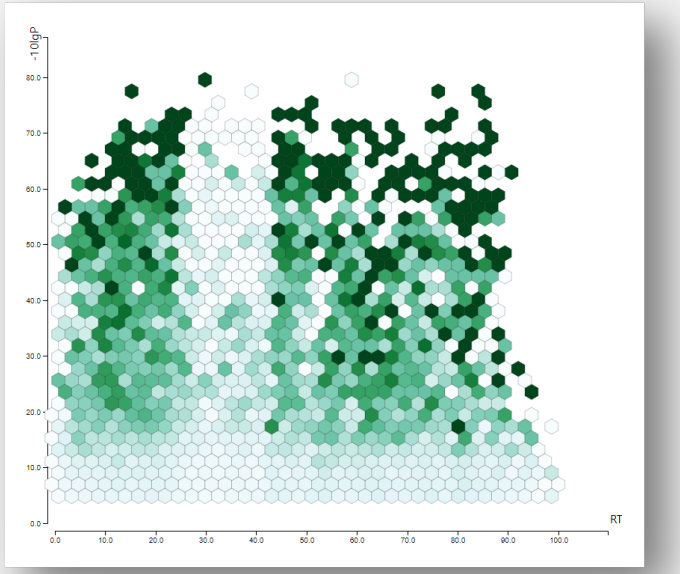
Std. Dev(x) = 16.76

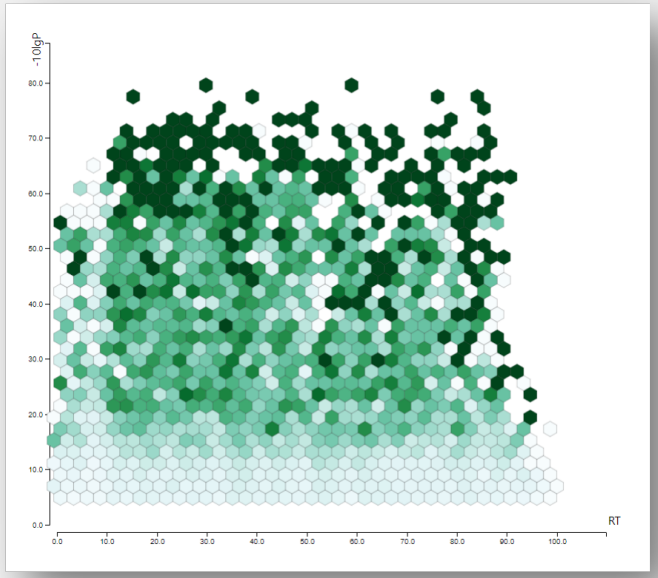
Std. Dev(y) = 26.93

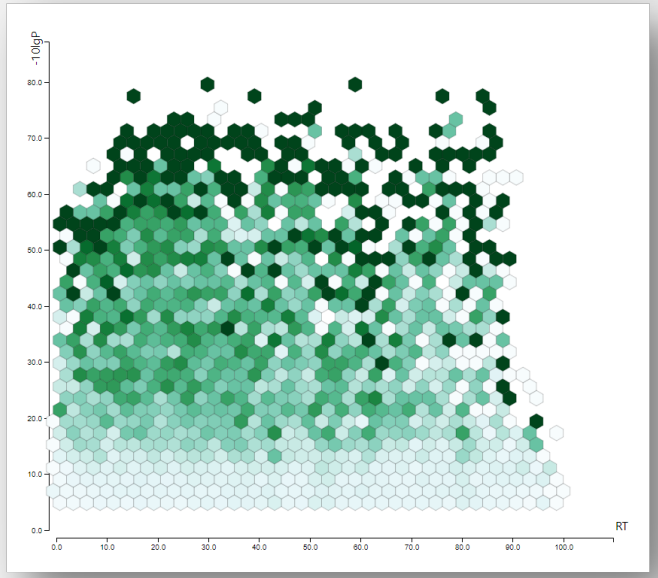
Correlation(x,y) = -0.06

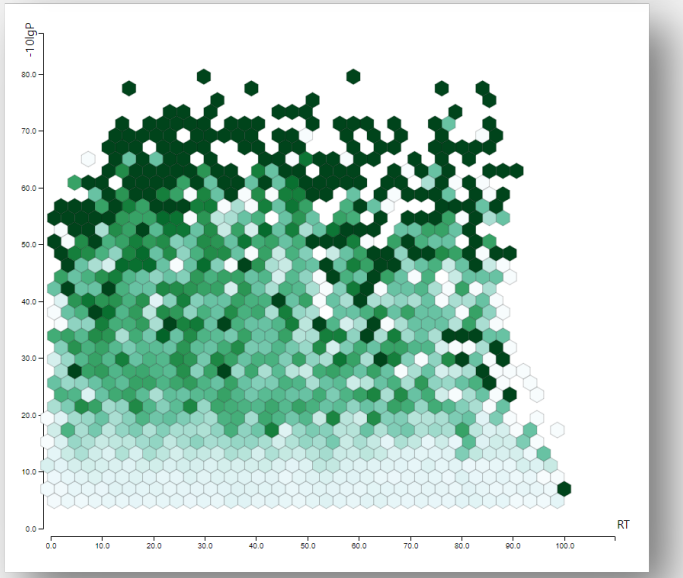
Source: [Cairo 2016]

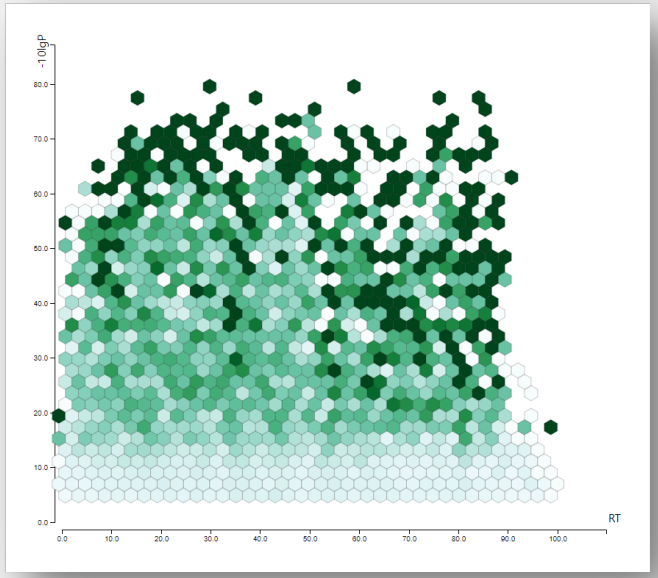












Data: Søren Drud-Heydary Nielsen | Visualization: Harith Rathish

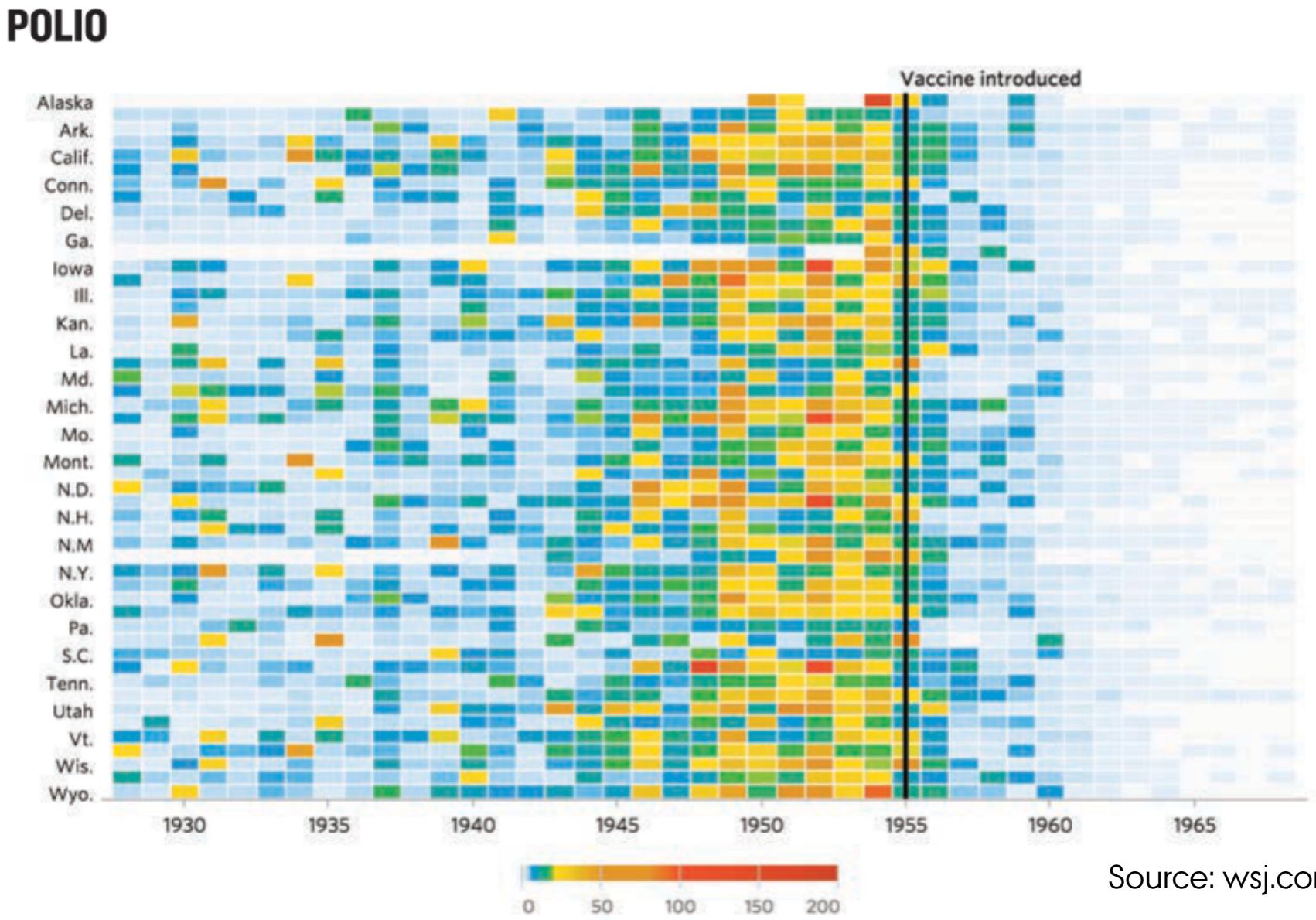
WHAT IS DATAVIS?

"Visual representations of data enable us to **communicate a large amount of information** to our viewers." [Marty 2009]

"The purpose of information visualization is to **amplify cognitive performance**, not just to create interesting pictures. Information visualizations should do for the mind what automobiles do for the feet." [Card 2008]

All taken from Chen et al. "What Is Visualization Really For?", Springer 2014

VISUALIZATION TO AMPLIFY COGNITION



Source: wsj.com

WHAT IS DATAVIS?

"Visual representations of data enable us to **communicate a large amount of information** to our viewers." [Marty 2009]

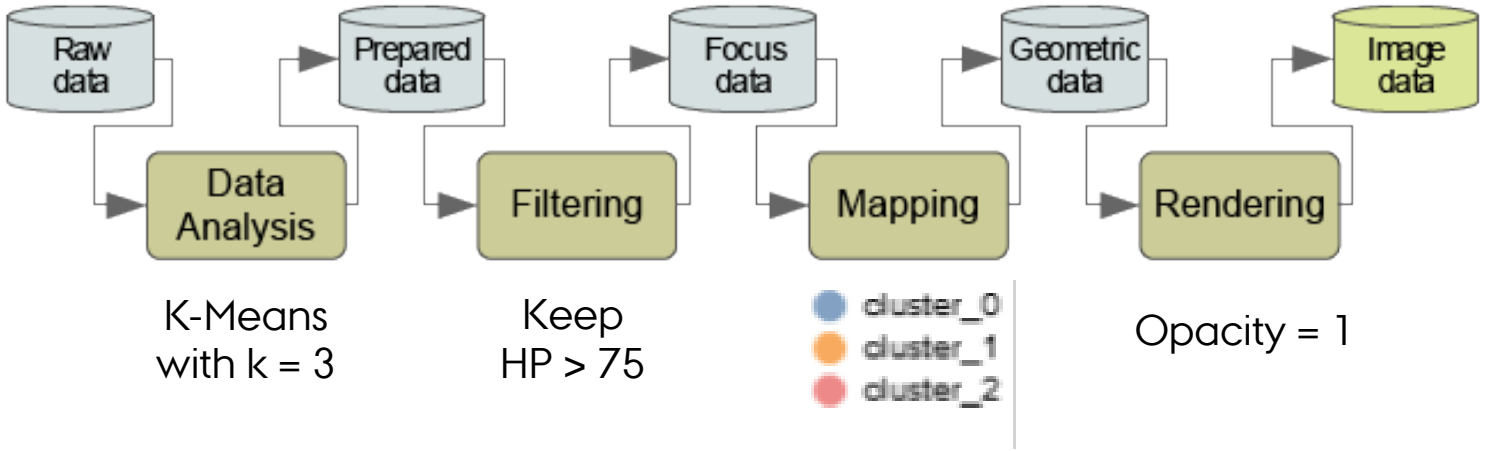
"The purpose of information visualization is to **amplify cognitive performance**, not just to create interesting pictures. Information visualizations should do for the mind what automobiles do for the feet." [Card 2008]

"Visualization is a method of computing. It **transforms the symbolic into the geometric**, … Visualization offers a method for seeing the unseen. It enriches the process of scientific discovery and fosters profound and unexpected insights." [McCormick et al. 1987]

All taken from Chen et al. "What Is Visualization Really For?", Springer 2014

THE VISUALIZATION PIPELINE

Image source: InfoVis Wiki



WHAT IS DATAVIS?

"Visual representations of data enable us to **communicate a large amount of information** to our viewers." [Marty 2009]

"The purpose of information visualization is to **amplify cognitive performance**, not just to create interesting pictures. Information visualizations should do for the mind what automobiles do for the feet." [Card 2008]

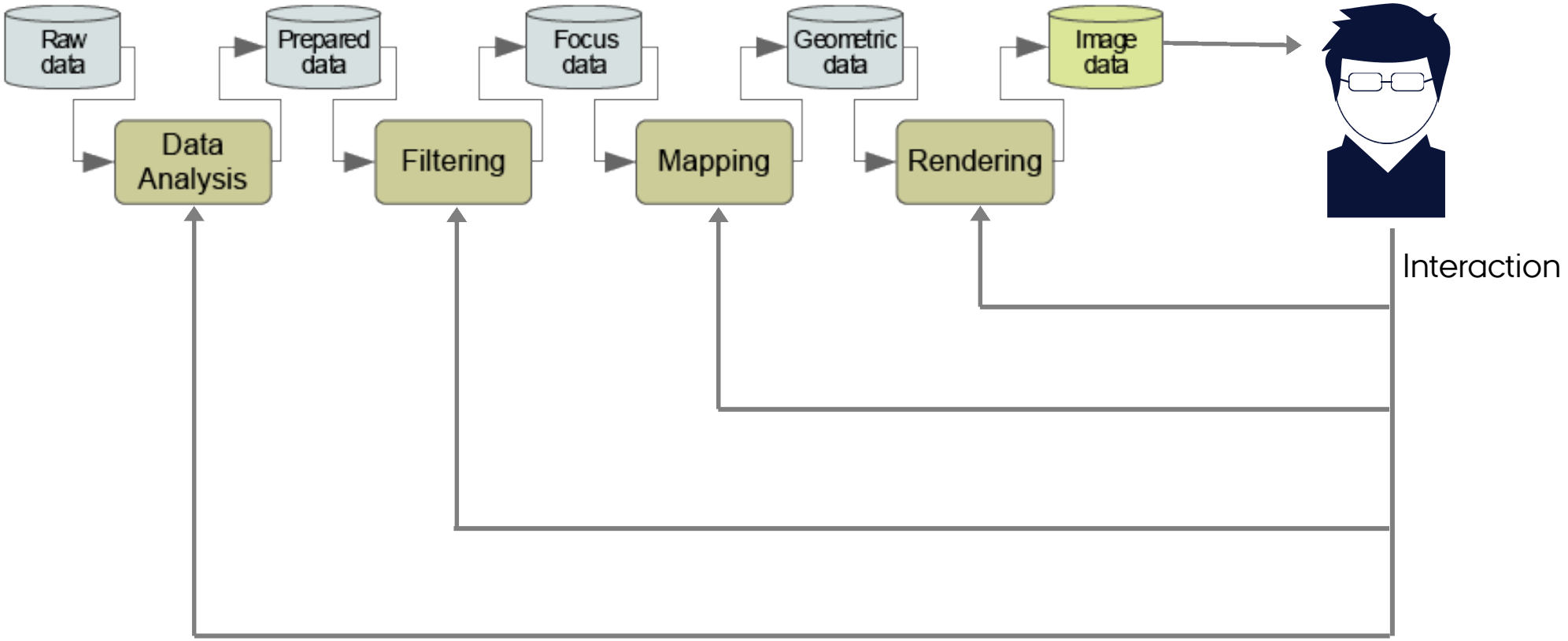
"Visualization is a method of computing. It **transforms the symbolic into the geometric**, … Visualization offers a method for seeing the unseen. It enriches the process of scientific discovery and fosters profound and unexpected insights." [McCormick et al. 1987]

"Visualization is essentially a **mapping process from computer representations to perceptual representations**, choosing encoding techniques to maximize human understanding and communication." [Owen 1999]

All taken from Chen et al. "What Is Visualization Really For?", Springer 2014

THE VISUALIZATION PIPELINE

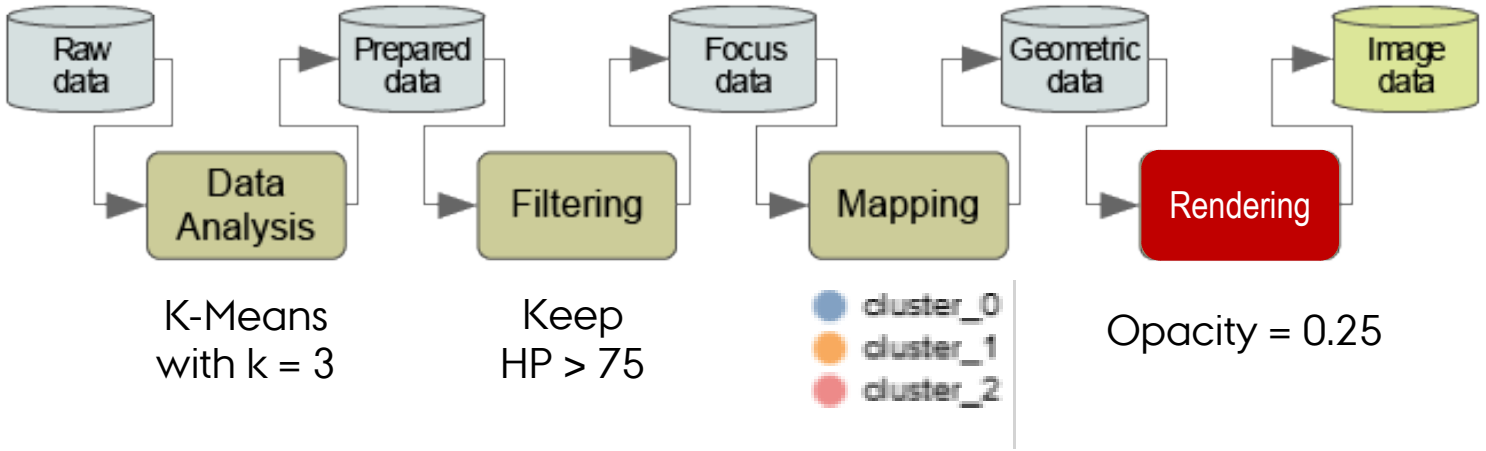
Image source: InfoVis Wiki

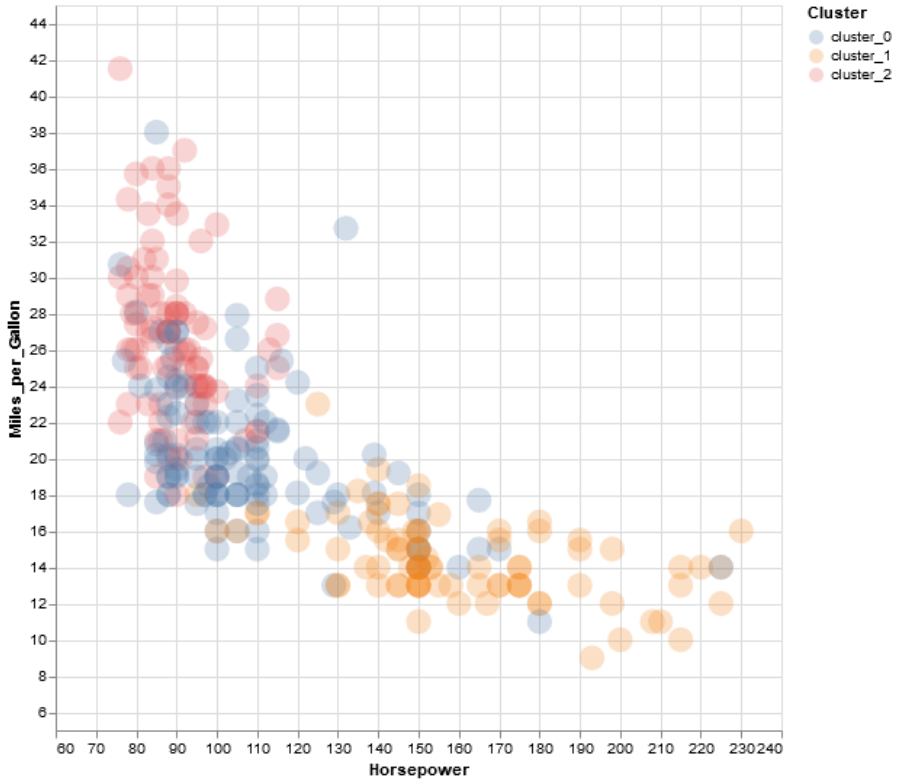


[Haber, McNabb 1990] and [Card, Mackinley 1999]

THE VISUALIZATION PIPELINE

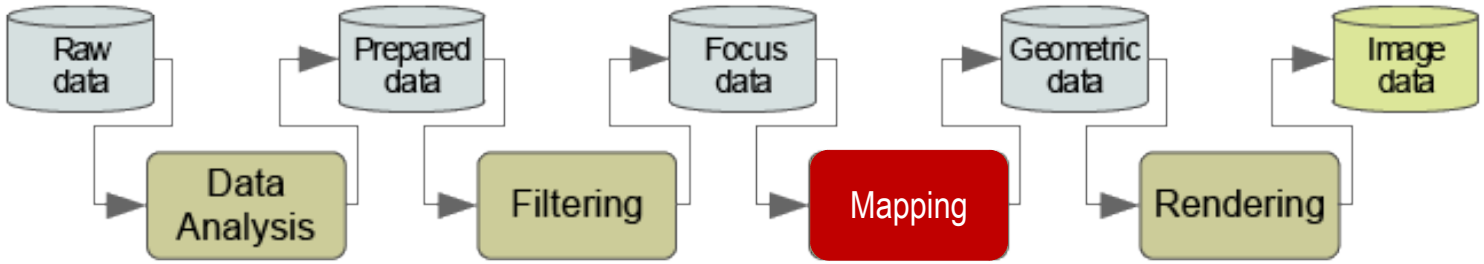
Image source: InfoVis Wiki





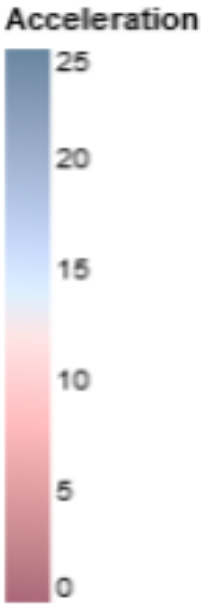
THE VISUALIZATION PIPELINE

Image source: InfoVis Wiki

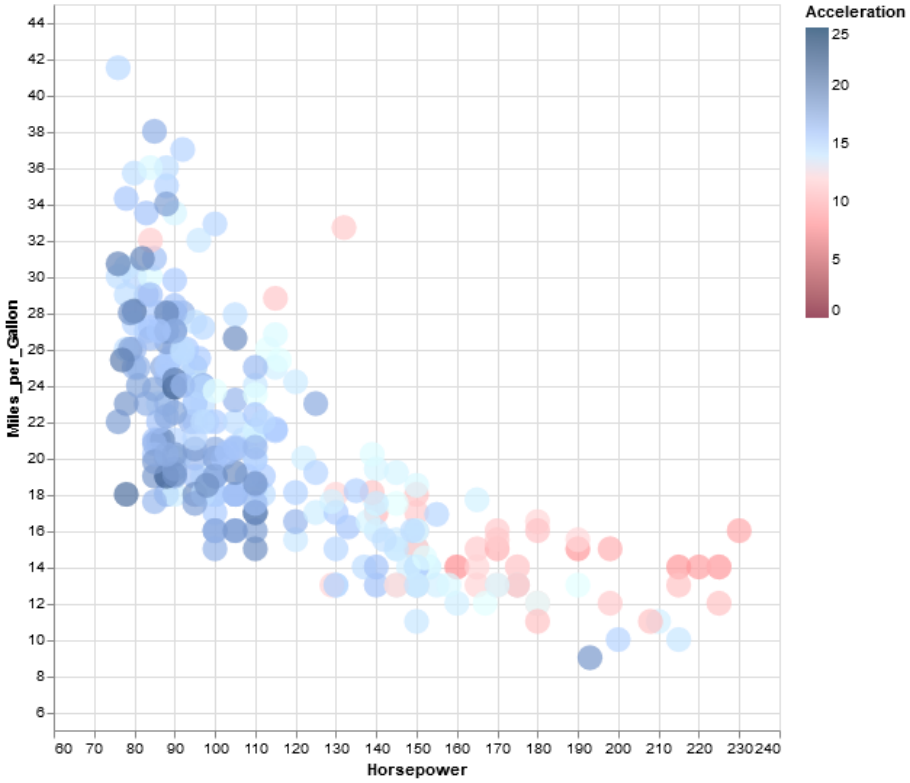


K-Means with k = 3

Keep HP > 75

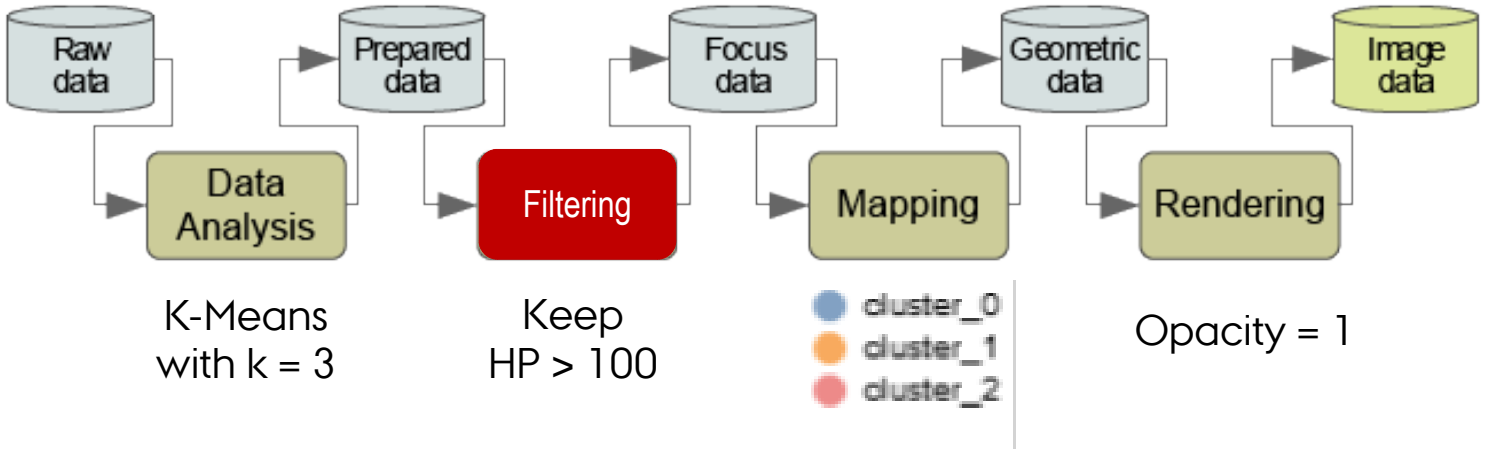


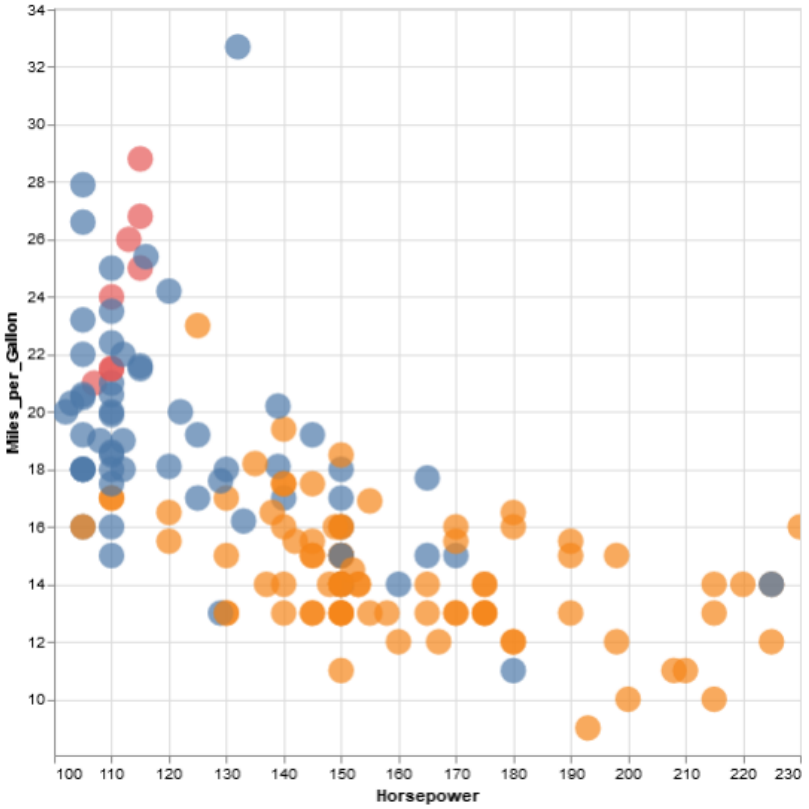
Opacity = 1



THE VISUALIZATION PIPELINE

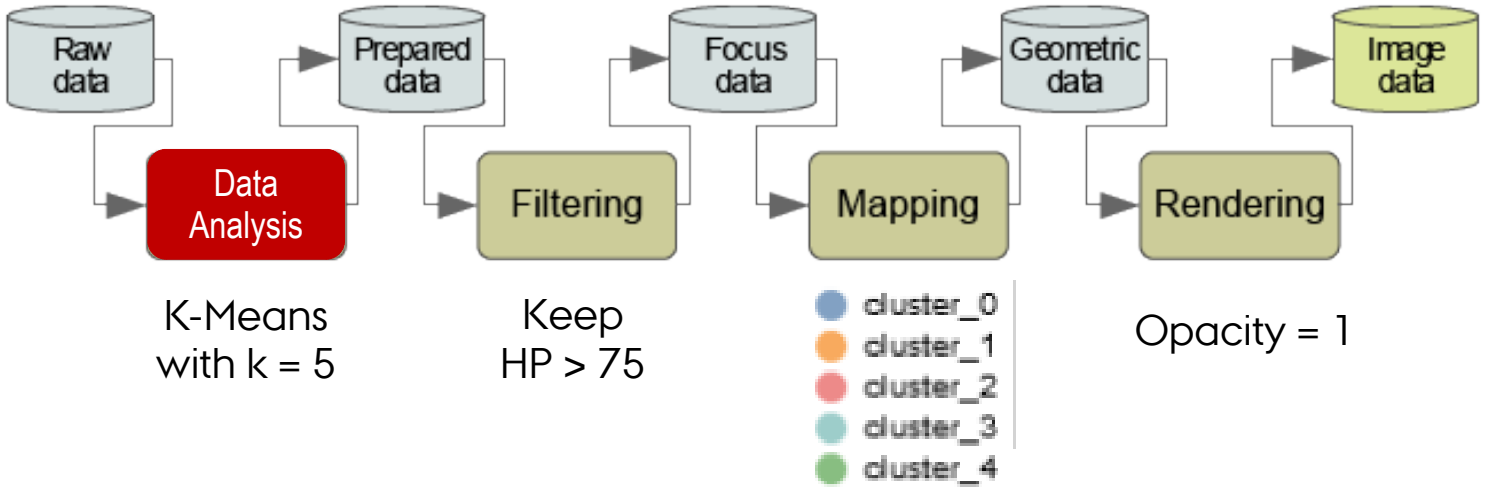
Image source: InfoVis Wiki

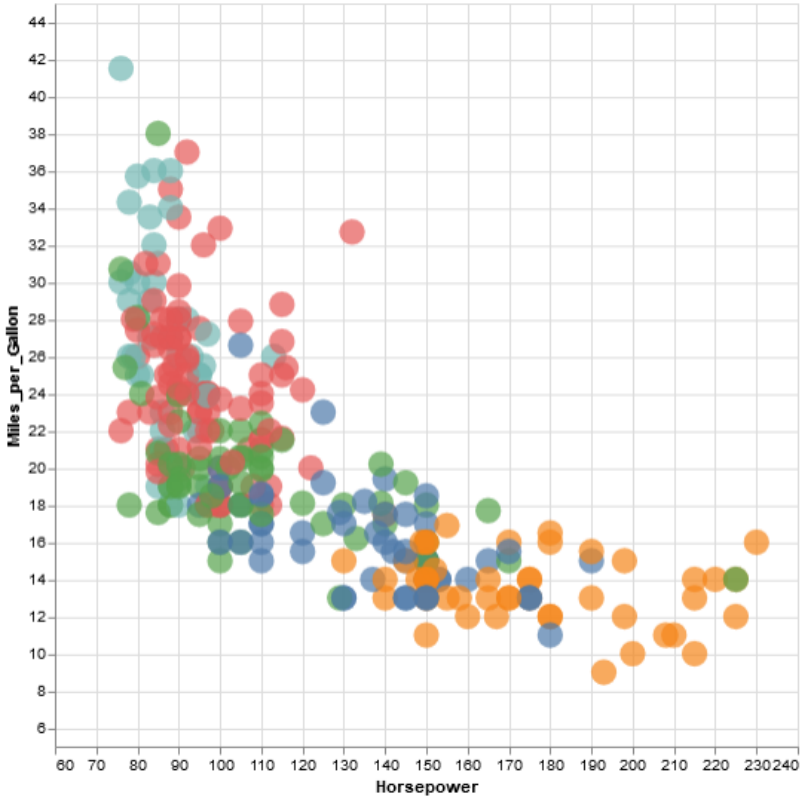




THE VISUALIZATION PIPELINE

Image source: InfoVis Wiki





WHAT IS DATAVIS?

"Visual representations of data enable us to **communicate a large amount of information** to our viewers." [Marty 2009]

"The purpose of information visuali **cognitive performance**, not just to create interesting pictures. Info izations should do for the mind what automobiles do for the feet." [C]

"Visualization is a method of computing. It **transform e symbolic into the geometric**, … Visualization offers a method for seeing the un iches the process of scientific discovery and fosters profound and unex [McCormick et al. 1987]

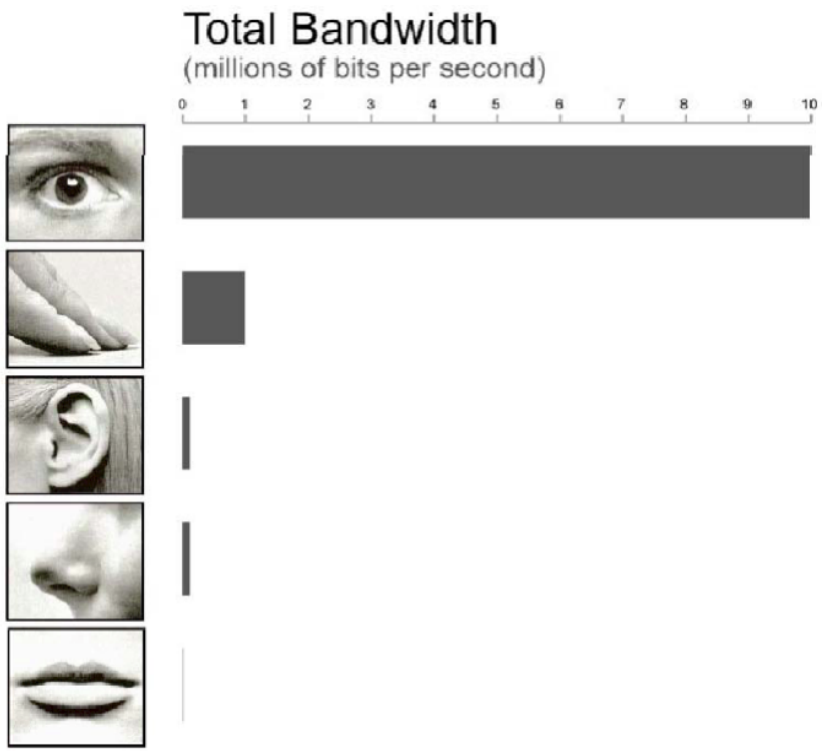
"Visualization is essentially a **ma ess from computer representations to perceptual representations**, choo encoding techniques to maximize human understanding and communication." [Owen 1999]

PRODUCT

PROCESS

All taken from Chen et al. "What Is Visualization Really For?", Springer 2014

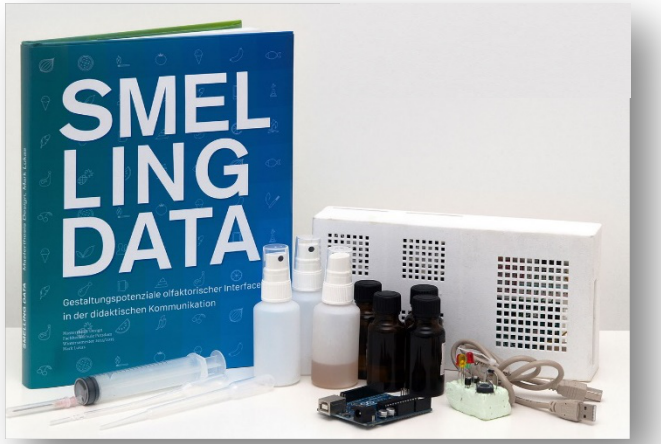
WHY DATAVIS?



Source: Tor Nørretranders “The User Illusion” (1998)

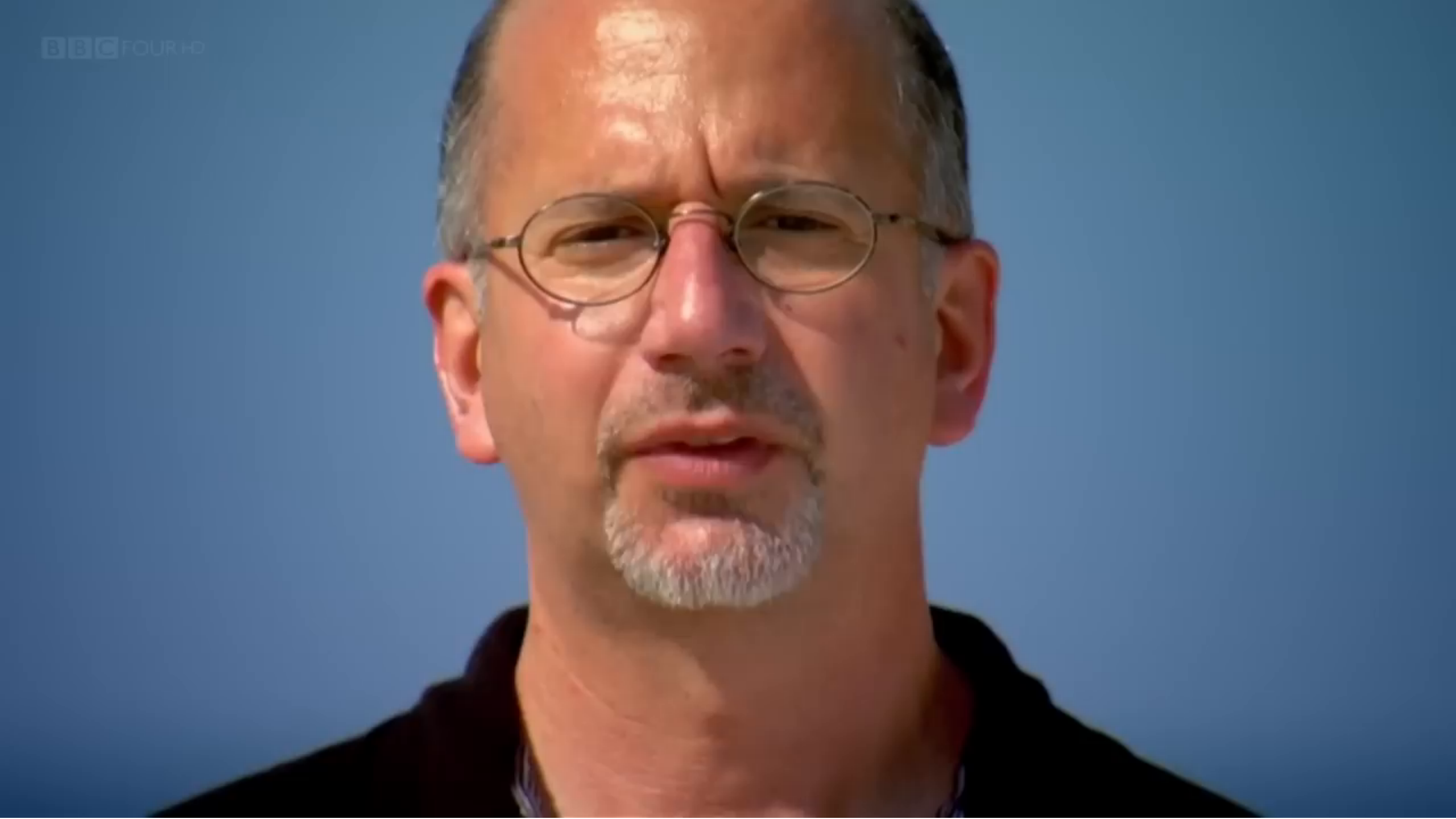
Haptic Displays – [Leithinger et al. 2011]

Data Sonification



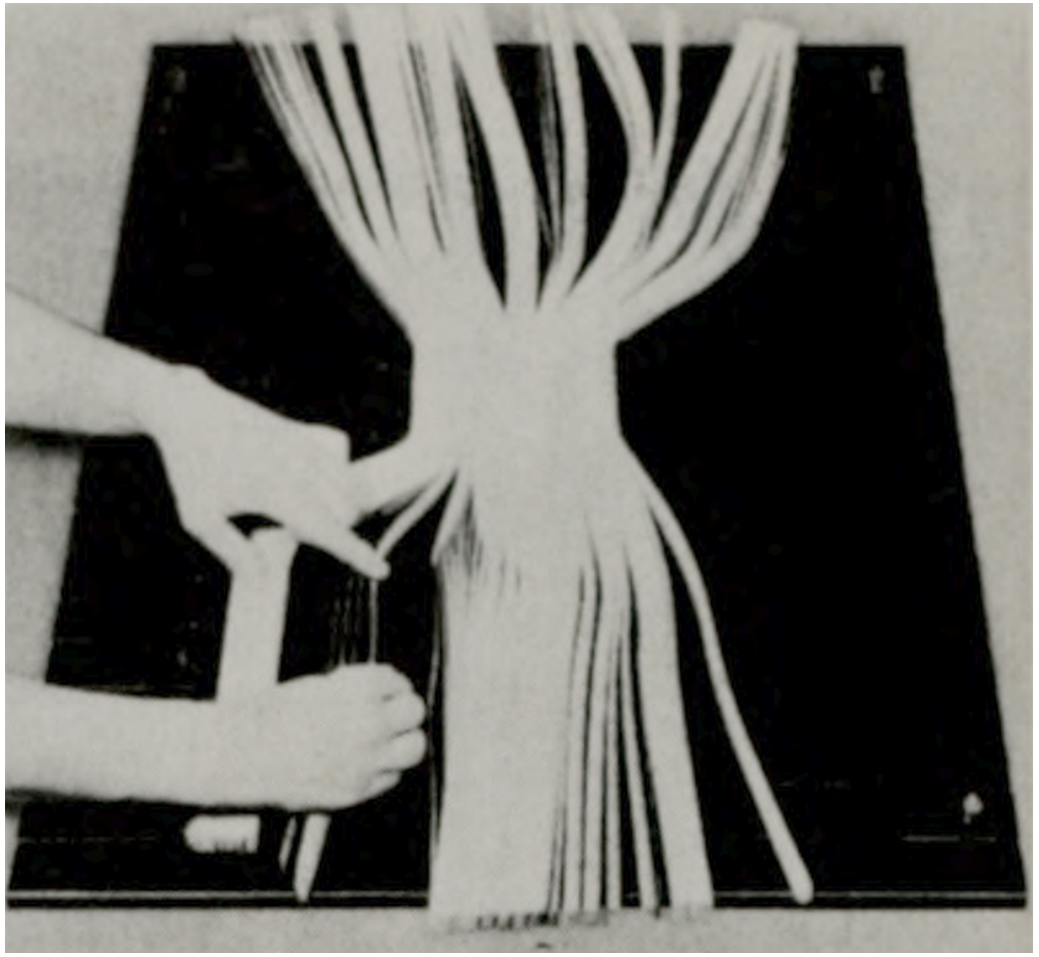
Olfactoric Interfaces – [Lukas 2015]

Edible Diagrams/Data Dishes – [Himmelspach 2016]


BBC FOUR HD

WHY DATAVIS? – PHYSICAL VISUALIZATIONS

Jacques Bertin's Reorderable Matrices (1968)



IBM Cosmograph (1933)

DATA VISUALIZATION IN CONTEXT

AARHUS
UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

HANS-JÖRG SCHULZ
ASSOCIATE PROFESSOR

5 Essential Tips for Creative Storytelling Through Data Visualization

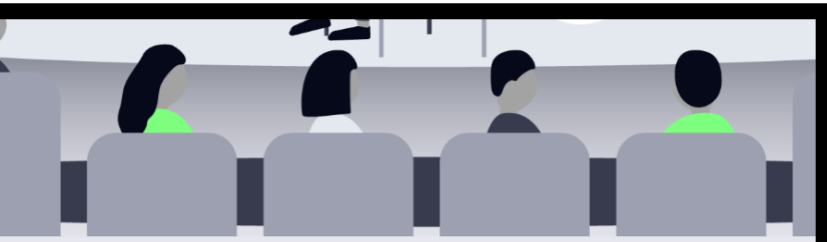
While some companies are collecting so much data that they don't know what they're even collecting,

CATEGORY: DATA VISUALIZATION

Persuasive Storytelling with Data Visualization

MAY 29, 2018 - Nick Mannon

31 31 SHARES



Data Storytelling: Best Practices and Visualization Tips To Create Remarkable Reports

by John Ostrowski on Growth Mindset & Strategy, Tools & Resources, August 14, 2019

Data storytelling is changing the way we communicate data insights – for the better. By shaping our facts using the right techniques we improve understanding and actions. In this Data Storytelling Guide, we give our 2019 updated point of view

Forbes

Nov 1, 2018, 11:40am EDT

Data Visualization: How To Tell A Story With Data

Nicole Martin Former Contributor

AI & Big Data

I write about digital marketing, data and privacy concerns.

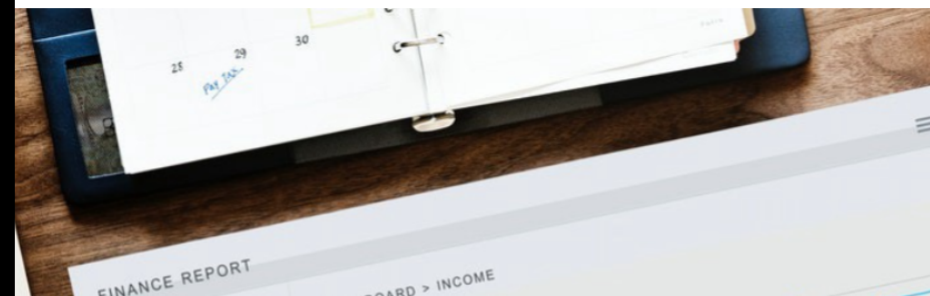
SCIENCE MACHINE LEARNING PROGRAMMING VISUALIZATION VIDEO ★ ABOUT | CONTRIBUTE

Storytelling with Data: A Data Visualization Guide for Business Professionals

For anyone who needs to communicate something to someone using data

Admond Lee Follow

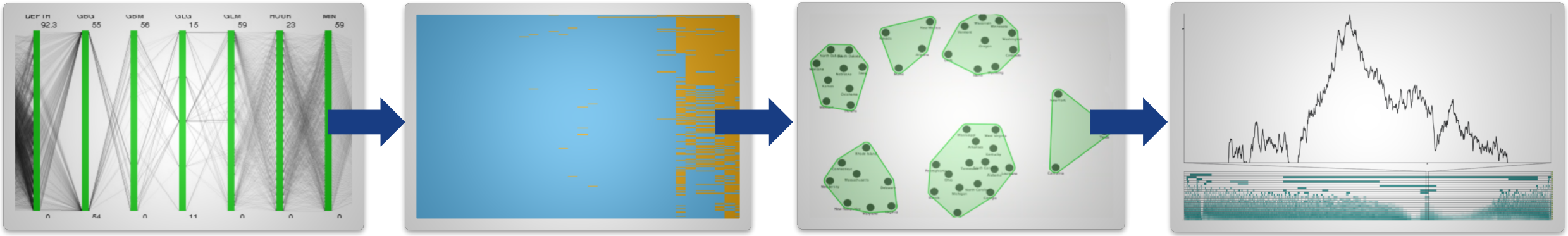
Oct 9, 2018 · 6 min read



DATAVIS IS SO MUCH MORE!

For many, this is the only point of visualization!

For us, it is the last step in a visualization supported data analysis pipeline:



Only once we identified and cross-checked a finding in the data, THEN we have a story to tell!

THE VISUALIZATION DESIGN SPACE

[BERINATO 2016]

Is the information *conceptual* or *data-driven?*

CONCEPTUAL

FOCUS: Ideas

GOALS: Simplify, teach ("Here's how our organization is structured.")

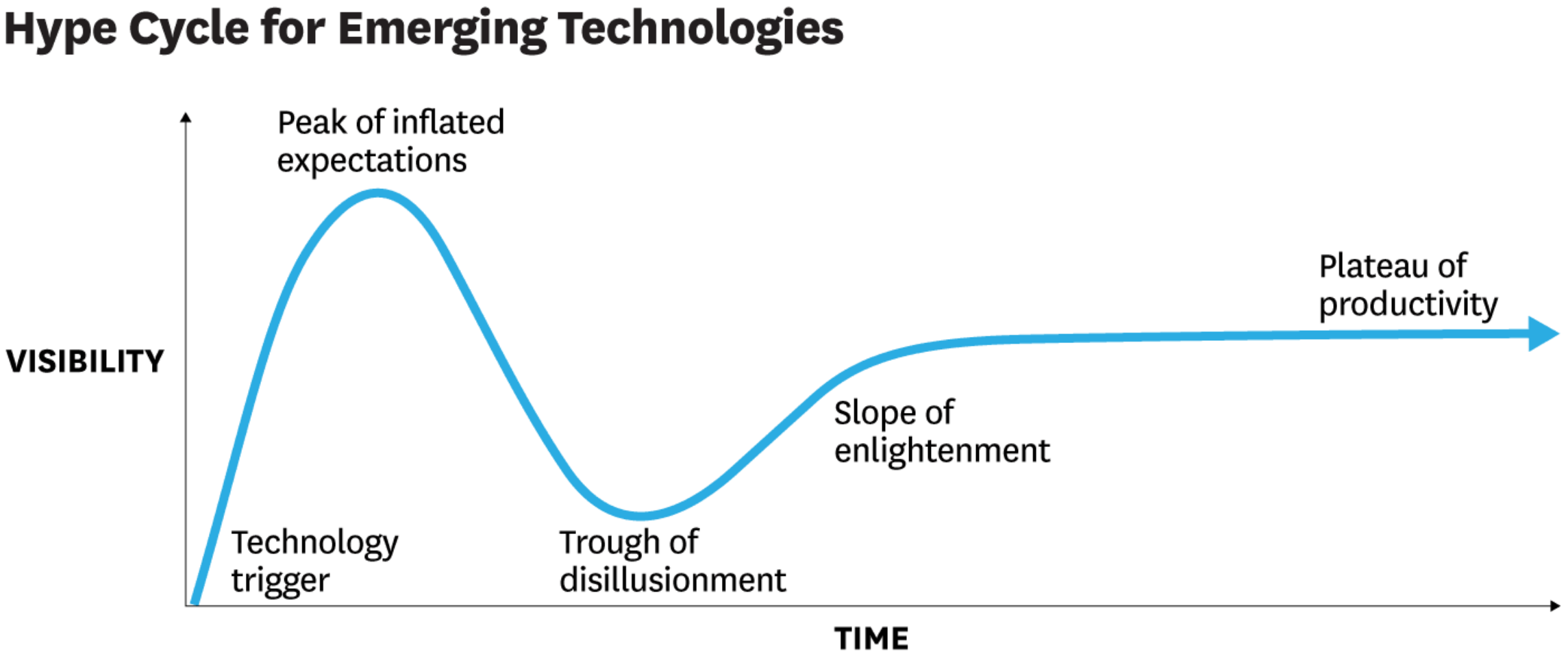
DATA-DRIVEN

FOCUS: Statistics

GOALS: Inform, enlighten ("Here are our revenues for the past two years.")

DATA-DRIVEN INFORMATION

CONCEPTUAL INFORMATION



THE VISUALIZATION DESIGN SPACE

[BERINATO 2016]

Is the information *conceptual* or *data-driven?*

CONCEPTUAL

FOCUS: Ideas

GOALS: Simplify, teach ("Here's how our organization is structured.")

DATA-DRIVEN

FOCUS: Statistics

GOALS: Inform, enlighten ("Here are our revenues for the past two years.")

Am I *declaring* something or *exploring* something?

DECLARATIVE

FOCUS: Documenting, designing

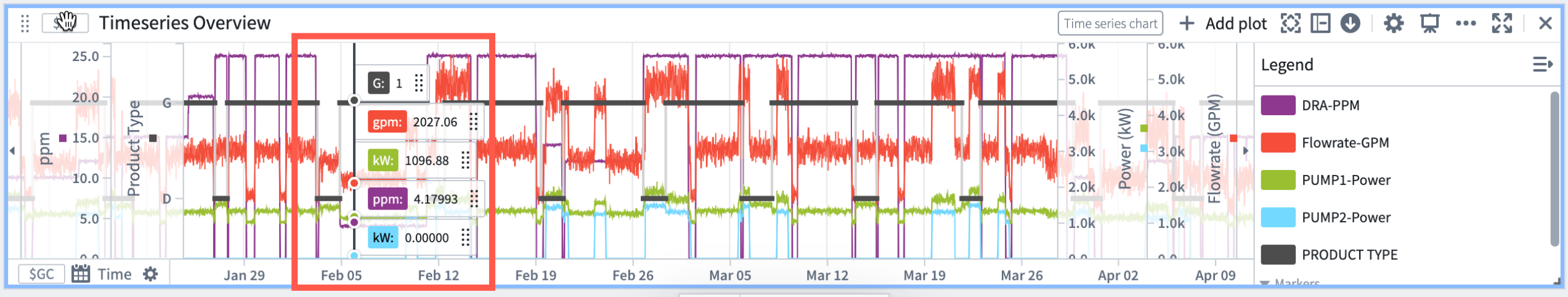
GOALS: Affirm ("Here is our budget by department.")

EXPLORATORY

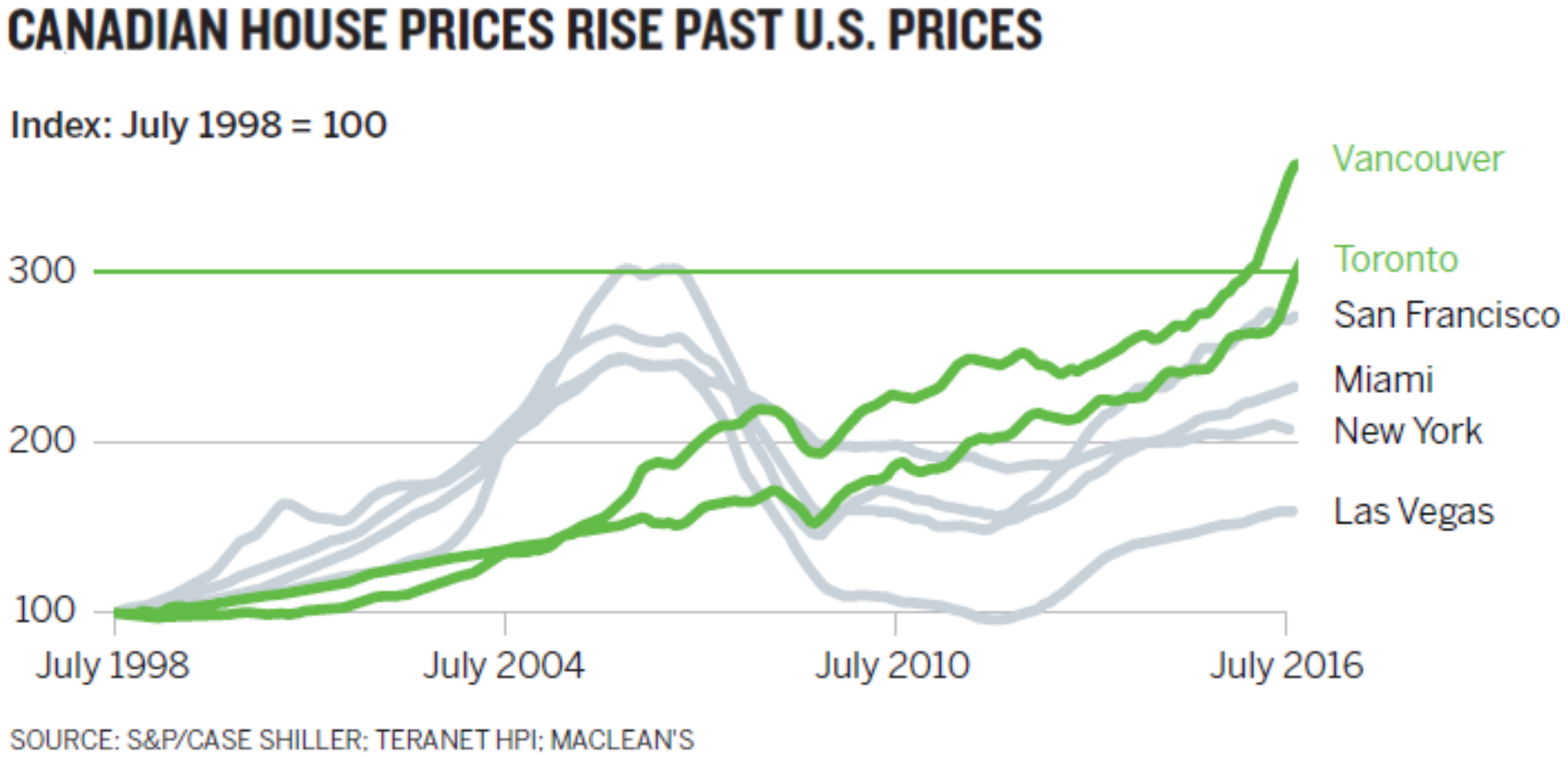
FOCUS: Prototyping, iterating, interacting, automating

GOALS: Confirm ("Let's see if marketing investments contributed to rising profits.") and discover ("What would we see if we visualized customer purchases by gender, location, and purchase amount in real time?")

EXPLORING SOMETHING

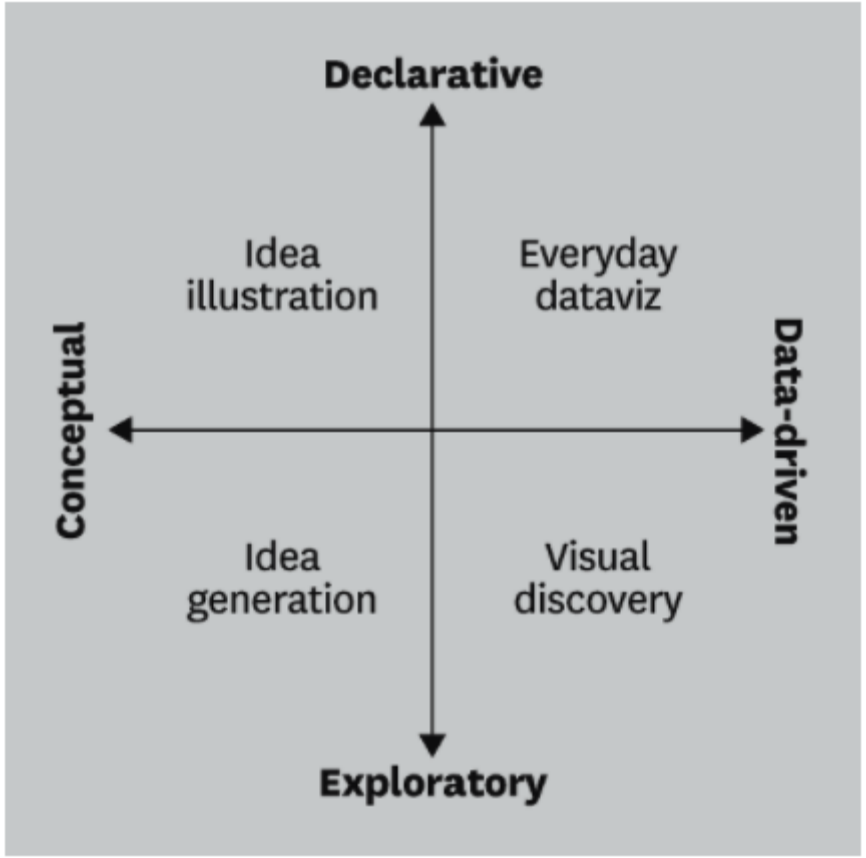


DECLARING SOMETHING

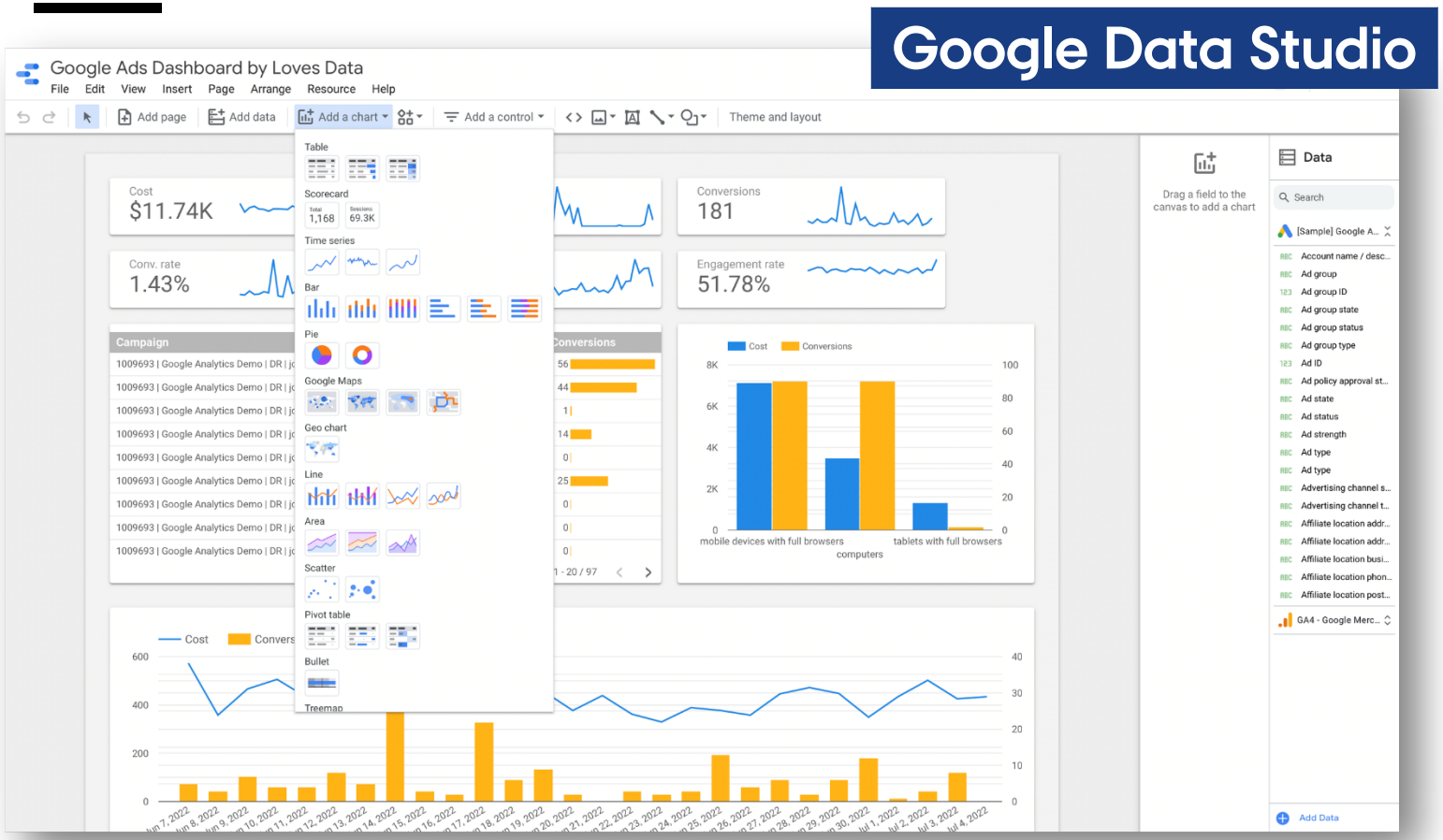


THE VISUALIZATION DESIGN SPACE

[BERINATO 2016]



DATA-DRIVEN & EXPLORATORY: VISUAL DISCOVERY



VISUAL DISCOVERY

INFO TYPE: Big data, complex, dynamic

TYPICAL SETTING: Working sessions, testing, analysis

PRIMARY SKILLS: Business intelligence, programming, paired analysis

GOALS: Trend spotting, sense making, deep analysis

DATA-DRIVEN & DECLARATIVE: EVERYDAY DATAVIS

EVERYDAY DATAVIZ

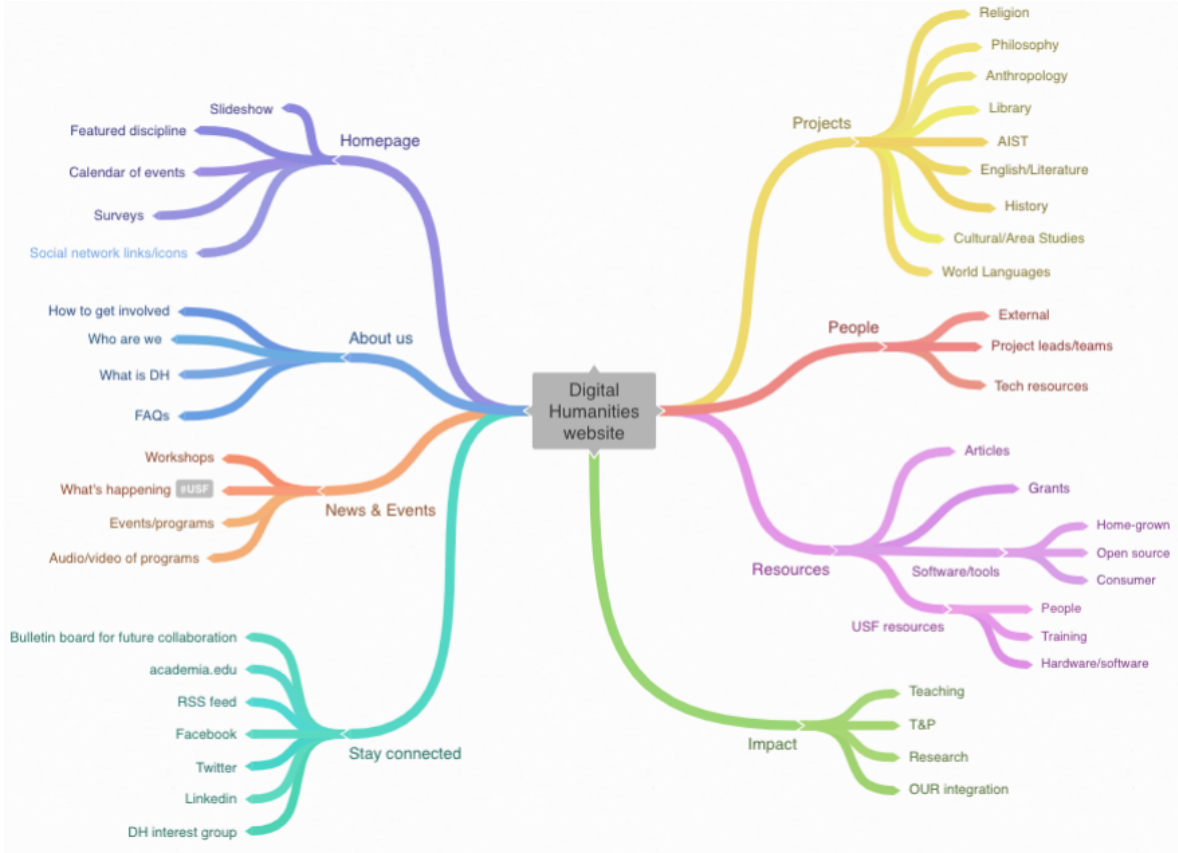
INFO TYPE: Simple, low volume

TYPICAL SETTING: Formal, presentations

PRIMARY SKILLS: Design, storytelling

GOALS: Affirming, setting context

CONCEPTUAL & EXPLORATORY: IDEA GENERATION



IDEA GENERATION

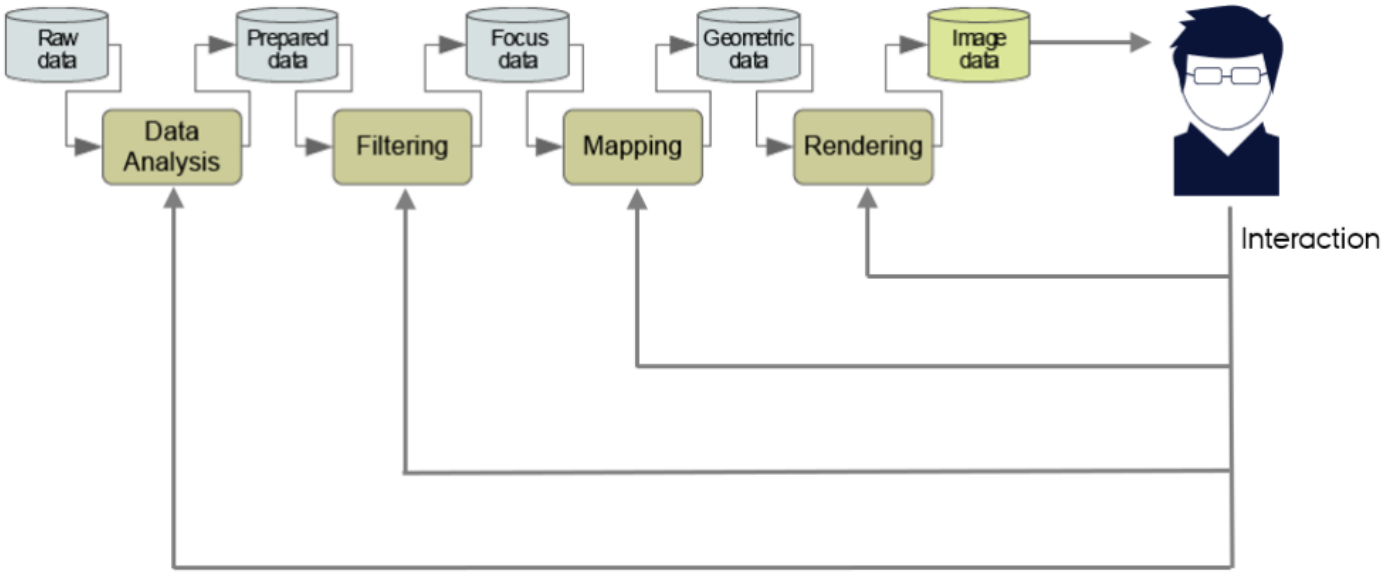
INFO TYPE: Complex, undefined

TYPICAL SETTING: Working session, brainstorming

PRIMARY SKILLS: Team-building, facilitation

GOALS: Problem solving, discovery, innovation

CONCEPTUAL & DECLARATIVE: IDEA ILLUSTRATION



IDEA ILLUSTRATION

INFO TYPE: Process, framework

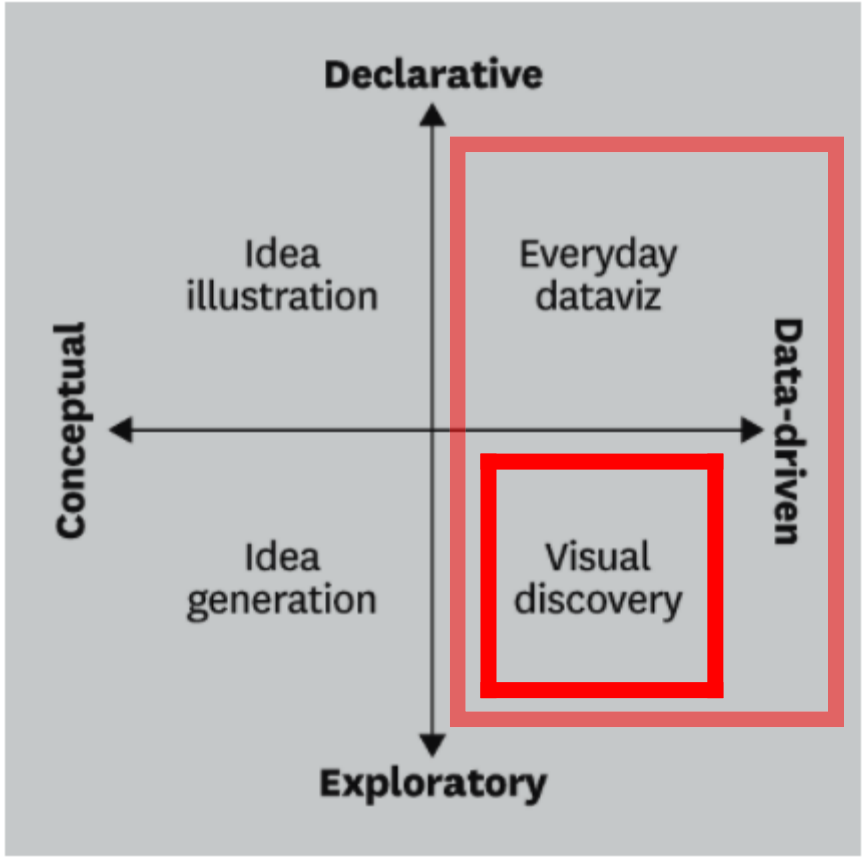
TYPICAL SETTING: Presentations, teaching

PRIMARY SKILLS: Design, editing

GOALS: Learning, simplifying, explaining

THE VISUALIZATION DESIGN SPACE

[BERINATO 2016]



BUT WHY LEARN ABOUT DATA VISUALIZATION?

VISUALIZATION OF DATA DISTRIBUTIONS

Example: given the birth year of every person in a cohort you are to visualize the age distribution

-> How would you go about it? **And why?**

-> Which visualization / chart type would you use? **And why?**

-> How exactly would you draw this chart? **And why?**

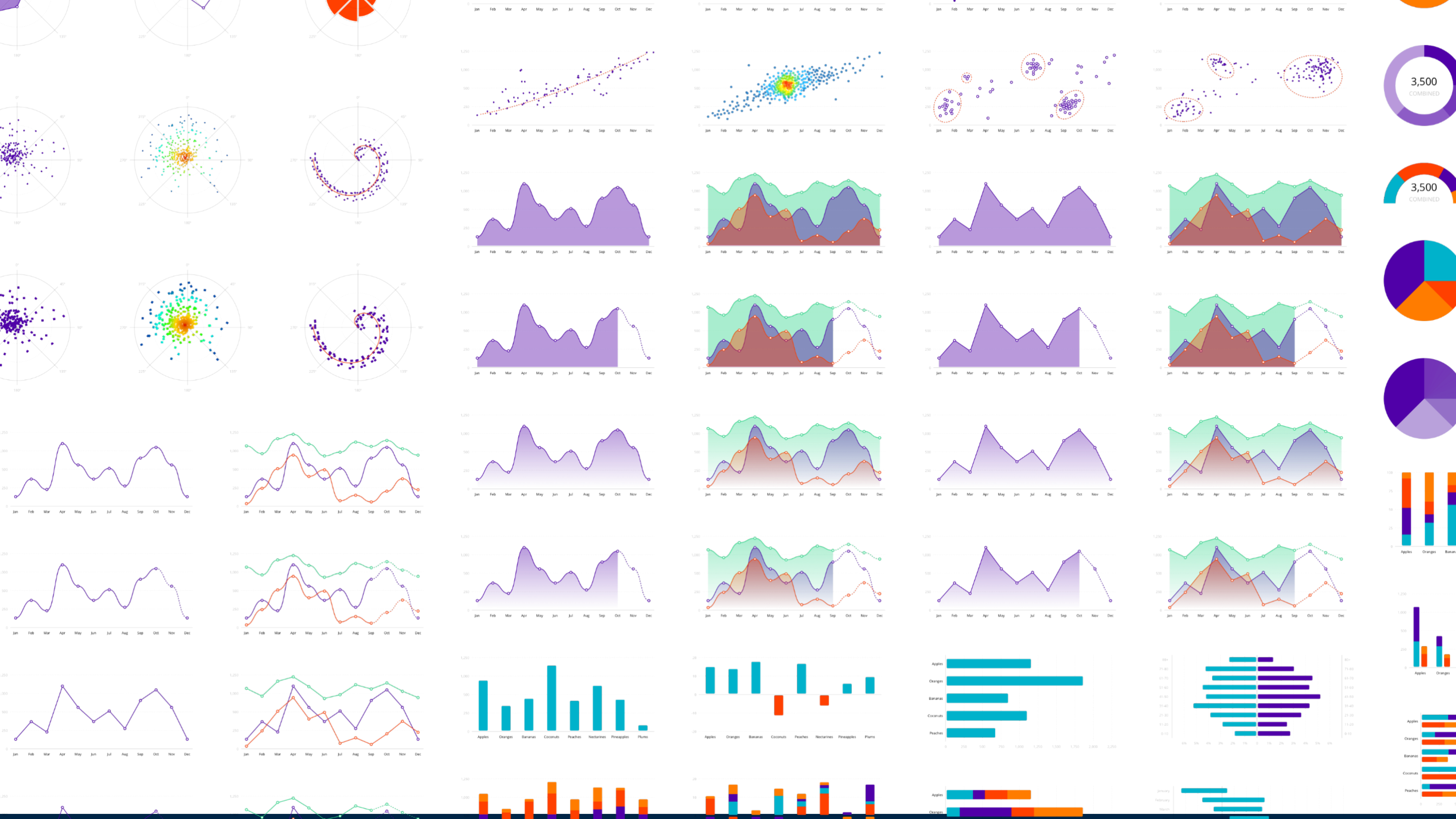
HOW WOULD YOU GO ABOUT IT?

- Data-driven process

  (dense vs. sparse, 10 vs. 1.000 vs. 1.000.000 data points,...)
- Task-driven process

  (find median + outliers, compare against other age distribution...)
- Design-driven process

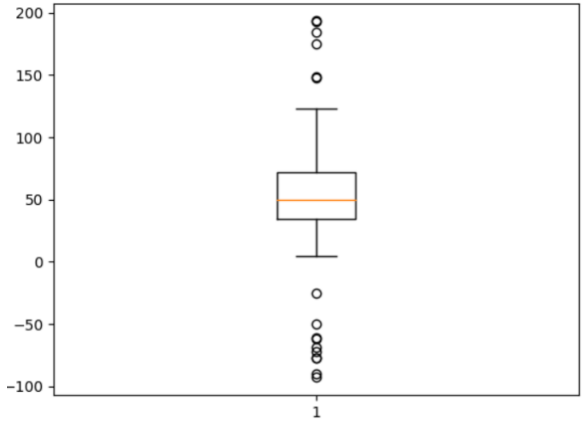
  (sketching, rapid prototyping, output-oriented process,...)


3,500
COMBINED
3,500
COMBINED

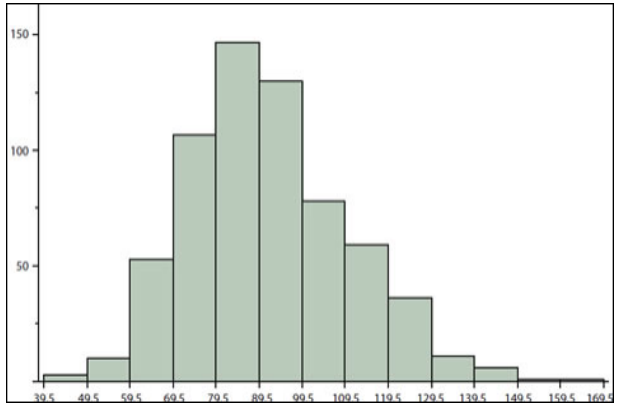
HOW WOULD YOU GO ABOUT IT?

- Data-driven process
  (dense vs. sparse, 10 vs. 1.000 vs. 1.000.000 data points,...)
- Task-driven process
  (find median + outliers, compare against other age distribution...)
- Design-driven process
  (sketching, rapid prototyping, output-oriented process,...)
- User-driven process
  (your mom vs. domain expert, different domains,...)

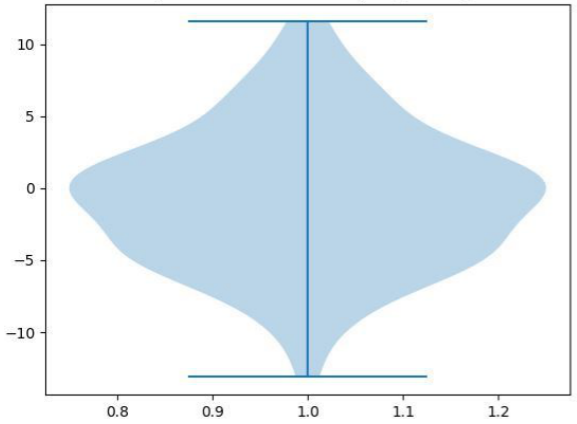
WHICH CHART TYPE TO USE?



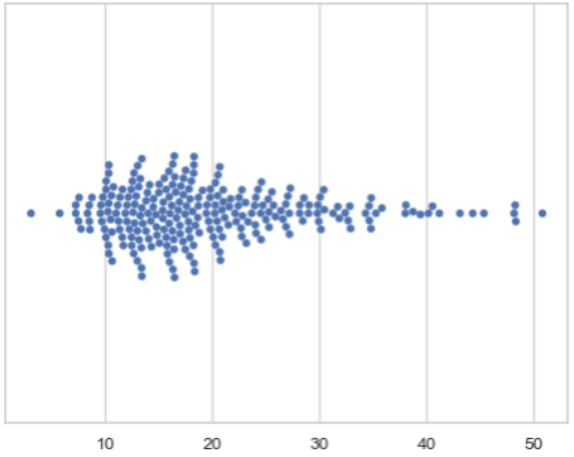
Boxplot /Box-Whisker-Plot



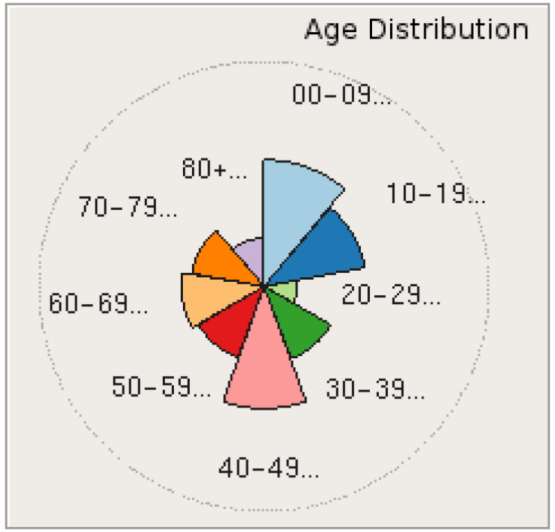
Histogram



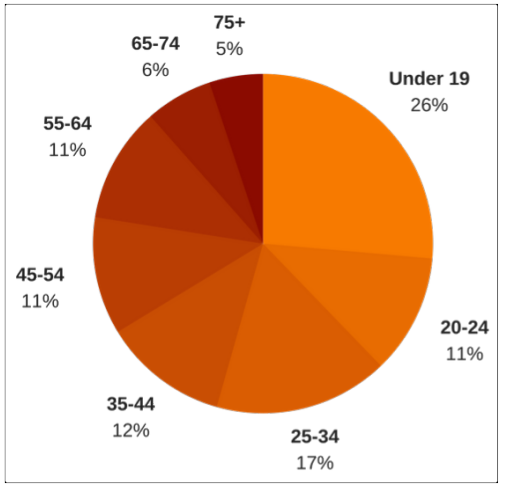
Violin Plot



Beeswarm Plot

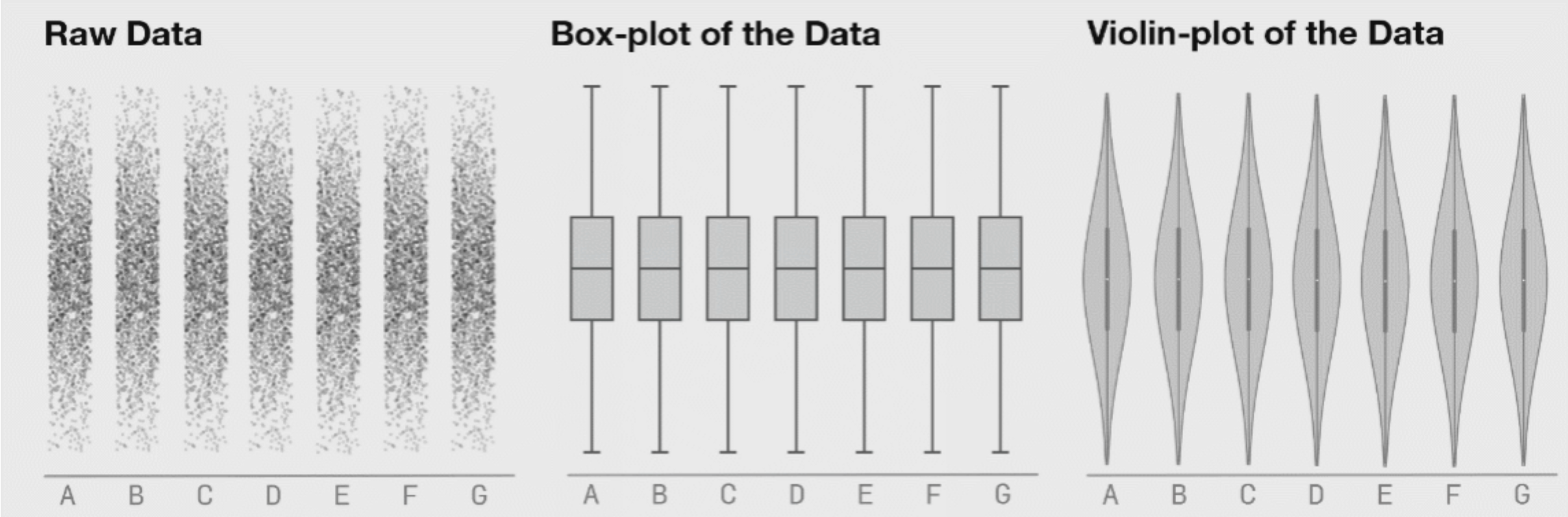


Windrose Chart / Coxcomb



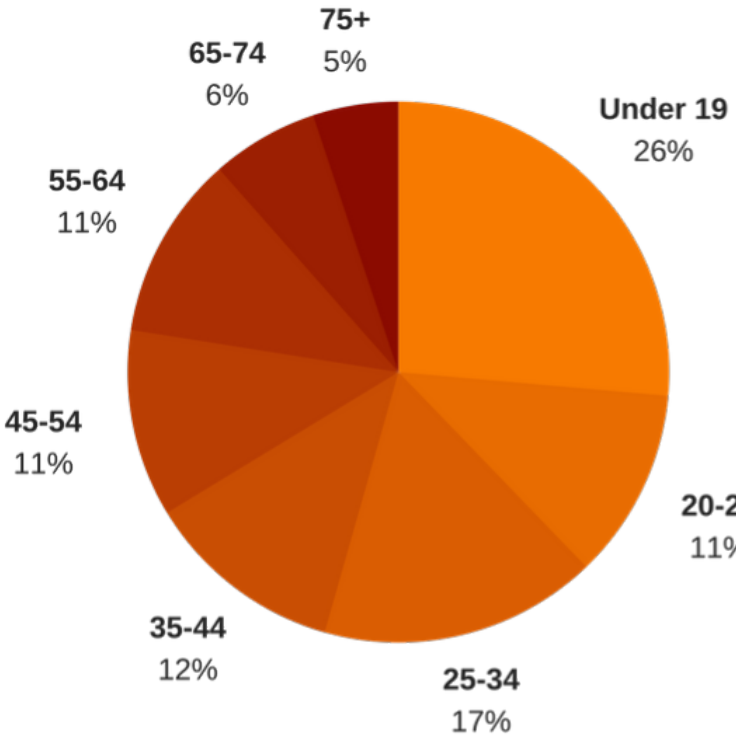
Piechart

BOX PLOT VS. VIOLIN PLOT

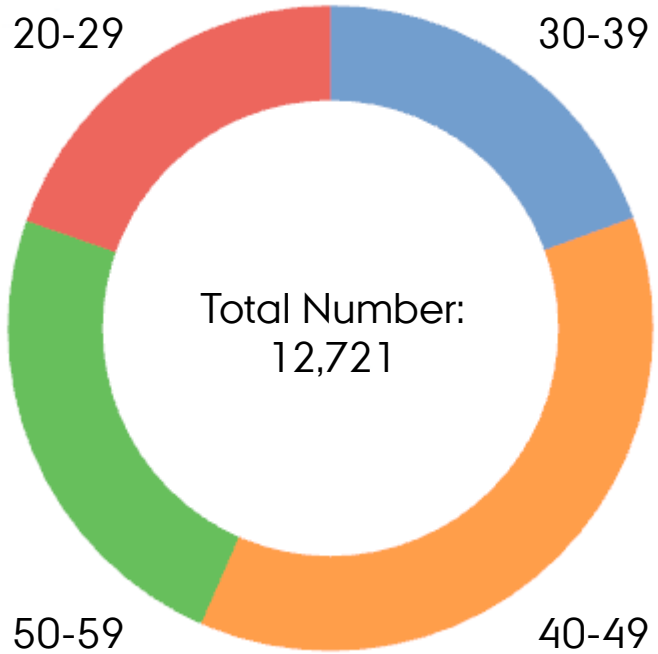


Source: Autodesk Research

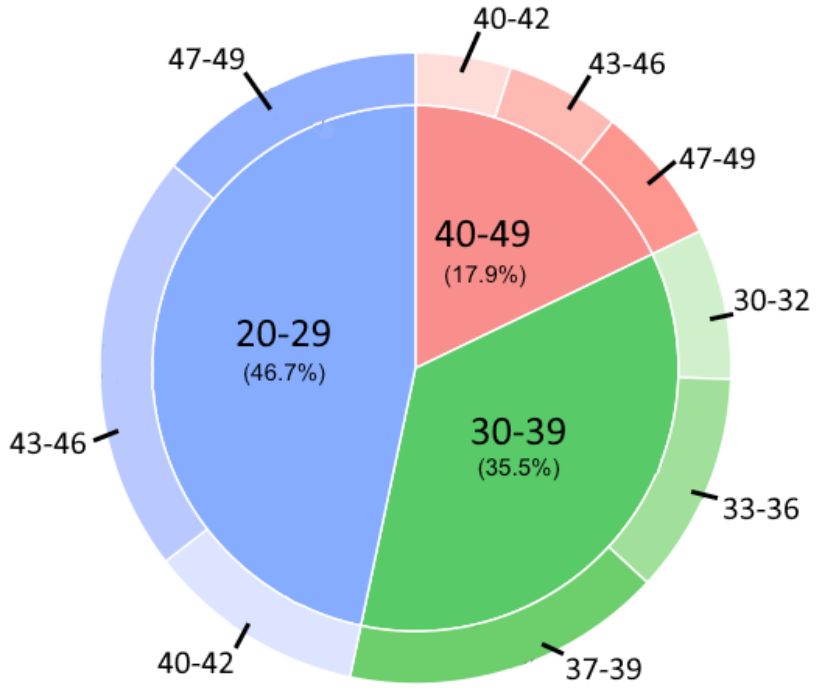
WHICH CHART "VARIANT" TO USE?



Pie Chart

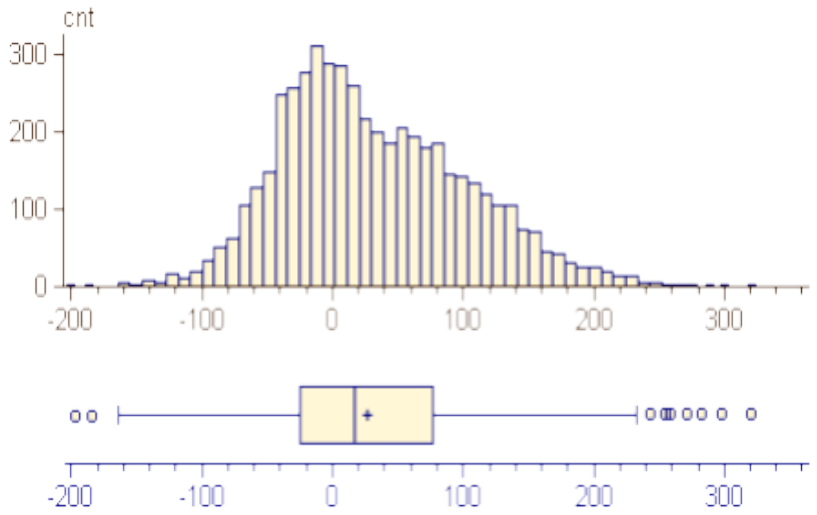


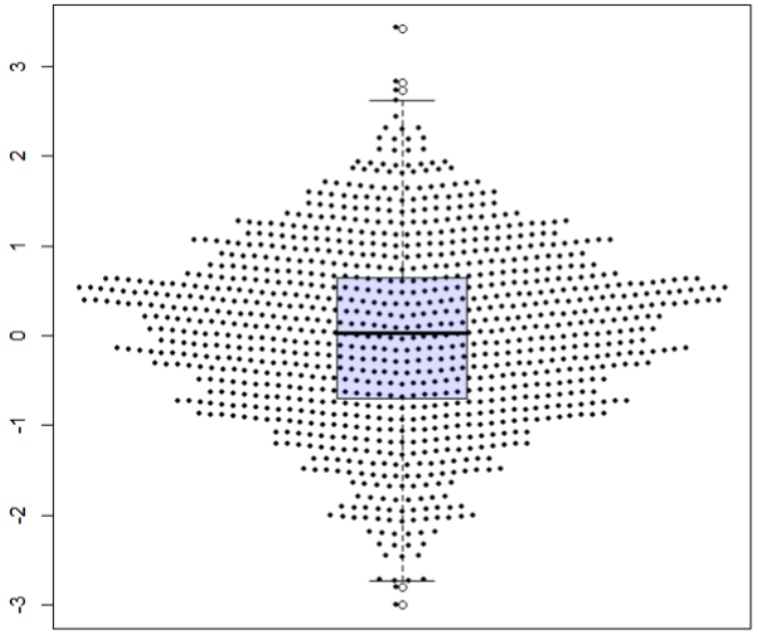
Doughnut Chart

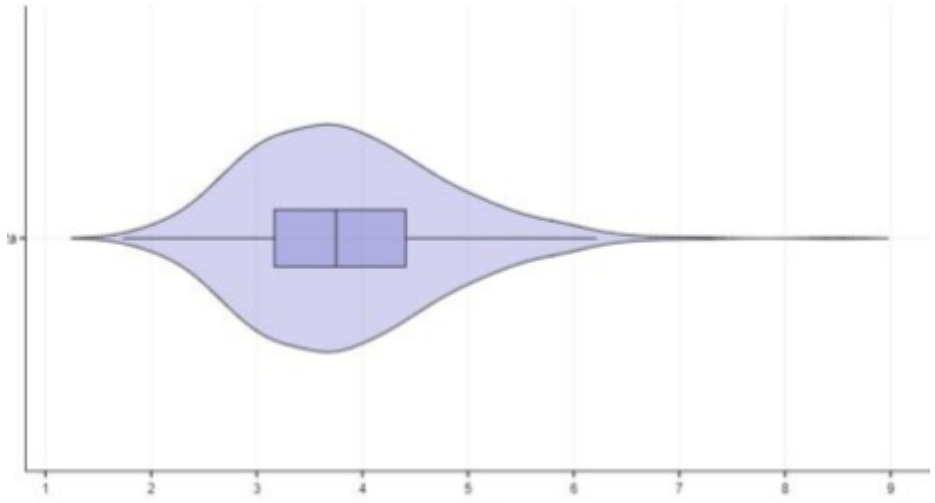


Hierarchical Pie Chart / Sunburst

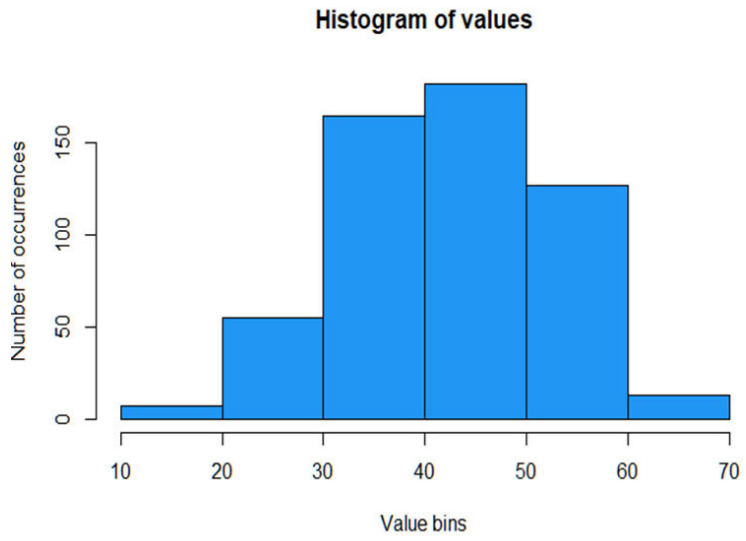
DO YOU WANT TO COMBINE CHARTS?

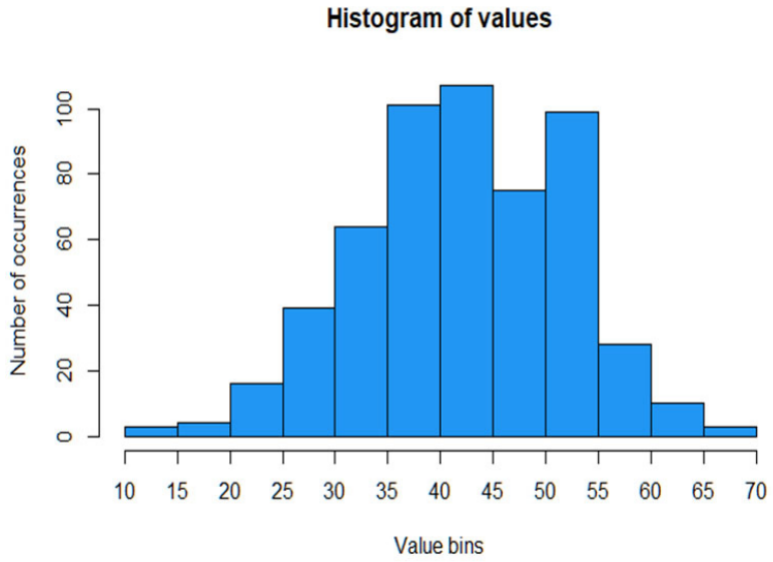


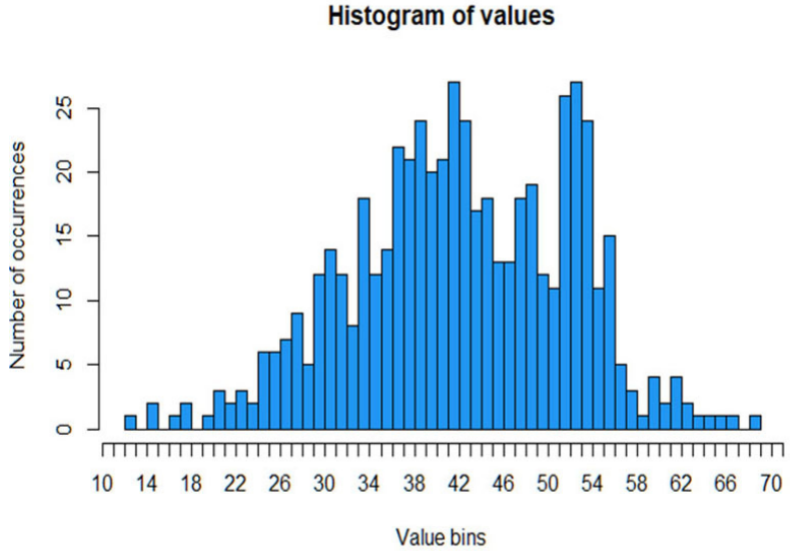




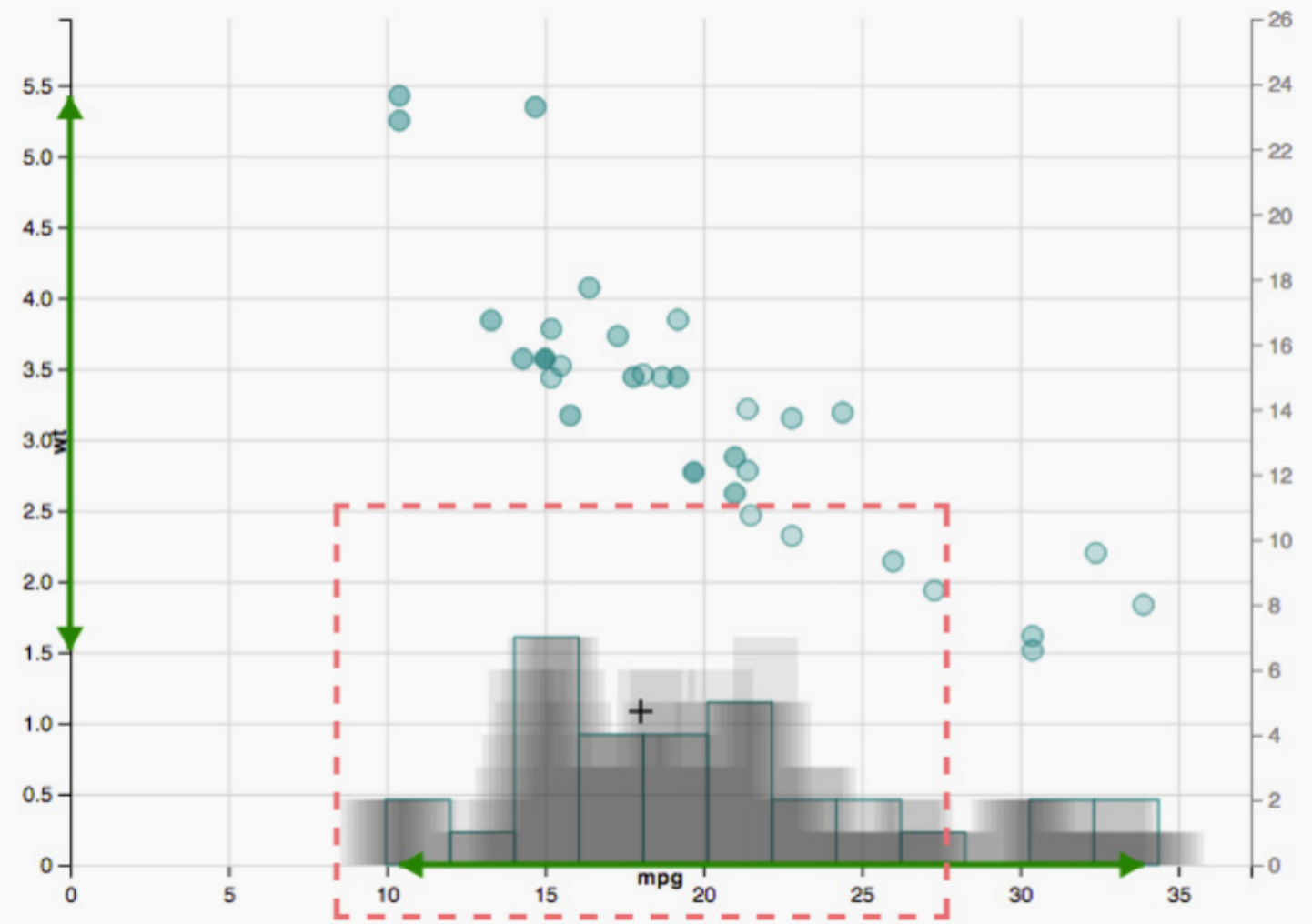
HOW TO DRAW THE CHART?





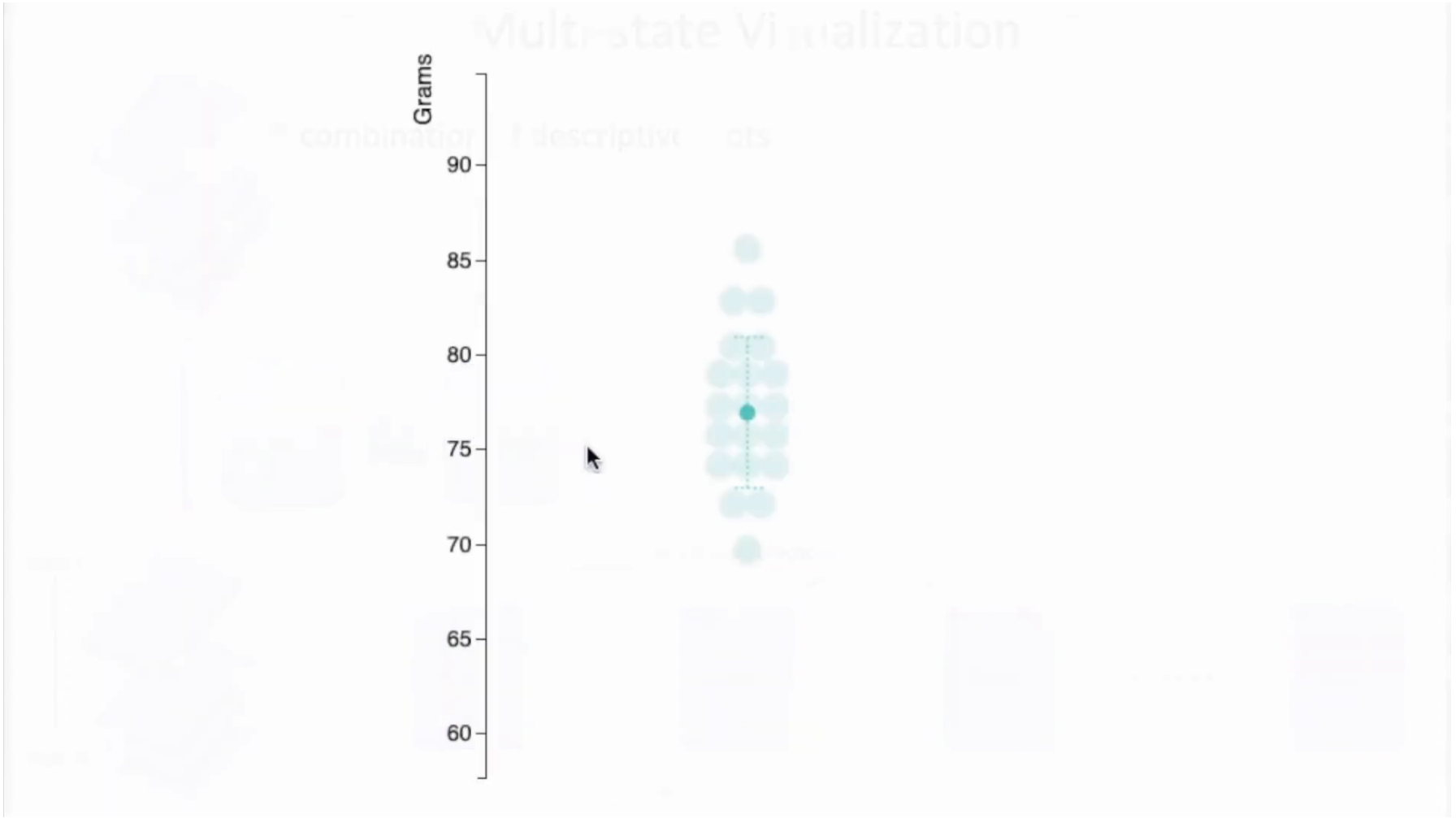


HOW TO DRAW THE CHART?



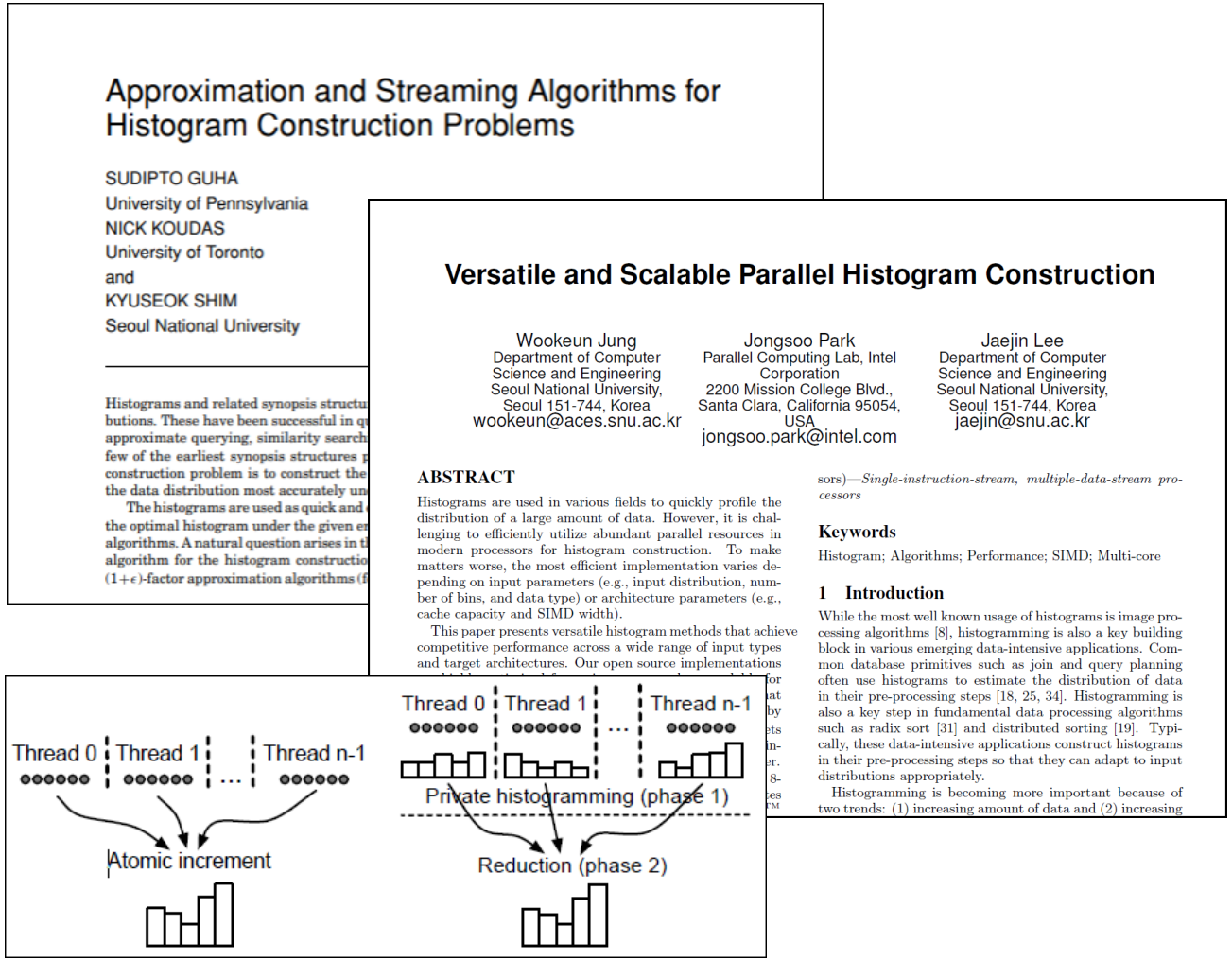
Source: Lunzer, McNamara 2014

HOW TO INTERACT WITH THE CHART?



HOW TO ENGINEER THE CHART?

	Method formula bins	Method formula width
Square-root	$\sqrt{n}$	$\frac{\max(values) - \min(values)}{\sqrt{n}}$
Sturges 1926	$\text{ceil}(\log_2 n) + 1$	$\frac{\max(values) - \min(values)}{\text{ceil}(\log_2 n) + 1}$
Rice 1944	$2 * \sqrt[3]{n}$	$\frac{\max(values) - \min(values)}{2 * \sqrt[3]{n}}$
Scott 1979	$\frac{\max(values) - \min(values)}{3.5 * \frac{stdev(values)}{\sqrt[3]{n}}}$	$3.5 * \frac{stdev(values)}{\sqrt[3]{n}}$
Freedman-Diaconis 1981	$\frac{\max(values) - \min(values)}{2 * \frac{IQR(values)}{\sqrt[3]{n}}}$	$2 * \frac{IQR(values)}{\sqrt[3]{n}}$



COURSE OVERVIEW

HANS-JÖRG SCHULZ
ASSOCIATE PROFESSOR

1. INTRODUCTION

2. VISUAL ENCODING

3. BASIC CHART TYPES

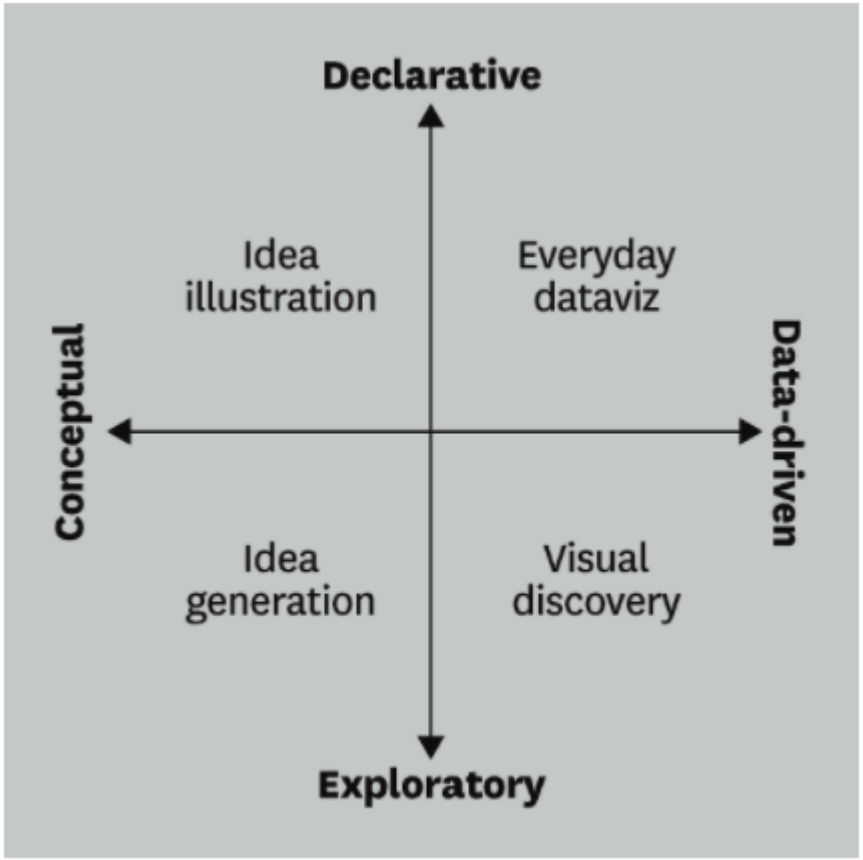
Basics

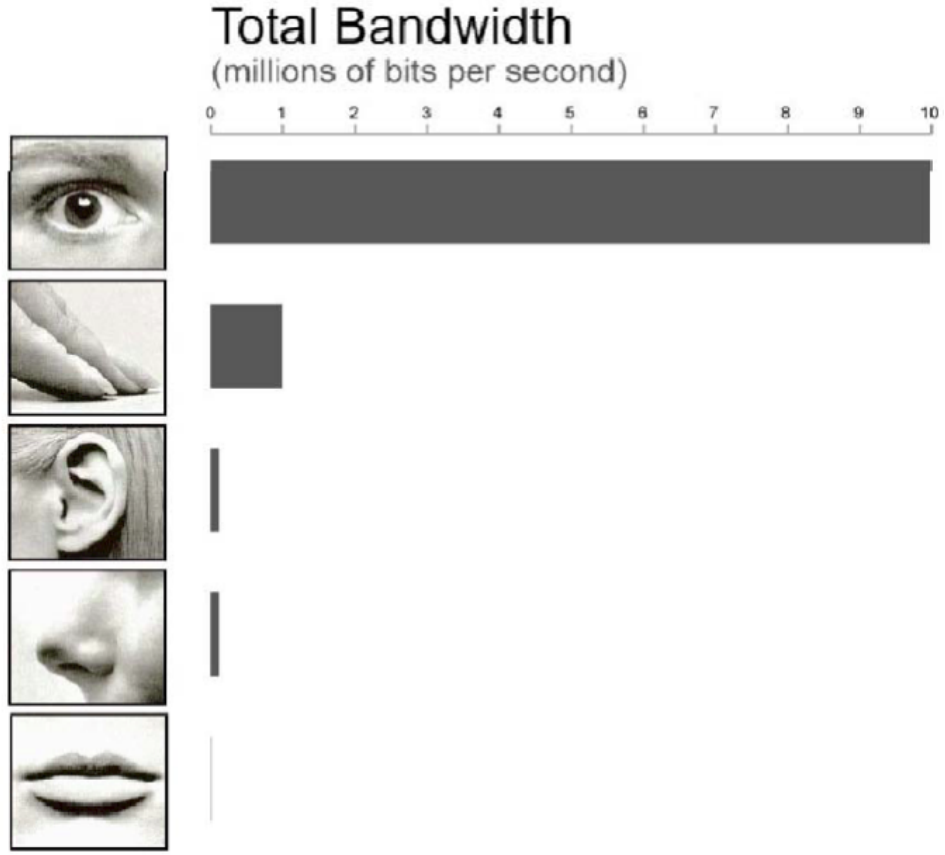
Visualization
Building Blocks
& Processes

Visualization
Techniques

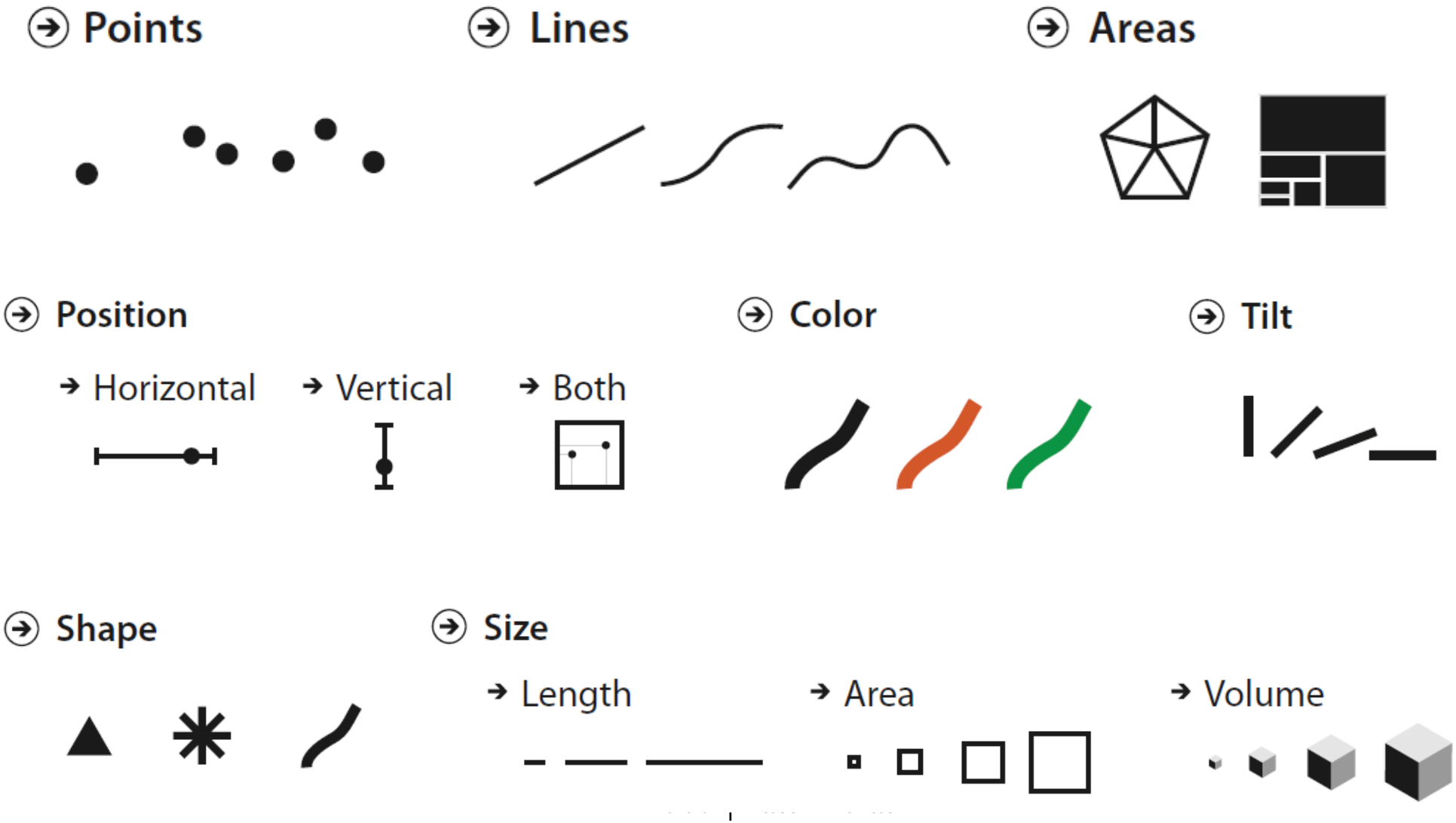
Cross-cutting
Topics

1. INTRODUCTION TO VISUALIZATION





2. VISUAL MAPPING & ENCODING



3. BASIC CHART TYPES

Deviation

Emphasise variations (+/-) from a fixed reference point. Typically the reference point is zero but it can also be a target or a long-term average. Can also be used to show sentiment (positive/neutral/negative).

Example FT uses
Trade surplus/deficit; climate change

Diverging bar

A simple standard bar chart that can handle both negative and positive magnitude values.

Diverging stacked bar

Perfect for presenting survey results which involve sentiment (eg disagree/neutral/agree).

Spine

Splits a single value into two contrasting components (eg male/female).

Surplus/deficit filled line

The shaded area of these charts allows a balance to be shown – either against a baseline or between two series.

Correlation

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume the relationships you show them to be causal (i.e. one causes the other).

Example FT uses
Inflation and unemployment, income and life expectancy

Scatterplot

The standard way to show the relationship between two continuous variables, each of which has its own axis.

Column + line timeline

A good way of showing the relationship between an amount (columns) and a rate (line).

Connected scatterplot

Usually used to show how the relationship between 2 variables has changed over time.

Bubble

Like a scatterplot, but adds additional detail by sizing the circles according to a third variable.

XY heatmap

A good way of showing the patterns between 2 categories of data, less effective at showing fine differences in amounts.

Ranking

Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest.

Example FT uses
Wealth, deprivation, league tables, constituency election results

Ordered bar

Standard bar charts display the ranks of values much more easily when sorted into order.

Ordered column

See above.

Ordered proportional symbol

Use when there are big variations between values and/or seeing fine differences between data is not so important.

Dot strip plot

Dots placed in order on a strip are a space-efficient method of laying out ranks across multiple categories.

Slope

Perfect for showing how ranks have changed over time or vary between categories.

Lollipop

Lollipops draw more attention to the data value than standard bar/column and can also show rank and value effectively.

Bump

Effective for showing changing rankings across multiple dates. For large datasets, consider grouping lines using colour.

Distribution

Show values in a dataset and how often they occur. The shape (or 'skew') of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data.

Example FT uses
Income distribution, population (age/sex) distribution, revealing inequality

Histogram

The standard way to show a statistical distribution - keep the gaps between columns small to highlight the 'shape' of the data.

Dot plot

A simple way of showing the change or range (min/max) of data across multiple categories.

Dot strip plot

Good for showing individual values in a distribution, can be a problem when too many dots have the same value.

Barcode plot

Like dot strip plots, good for displaying all the data in a table, they work best when highlighting individual values.

Boxplot

Summarise multiple distributions by showing the median (centre) and range of the data

Violin plot

Similar to a boxplot but more effective with complex distributions (data that cannot be summarised with simple average).

Population pyramid

A standard way for showing the age and sex breakdown of a population distribution; effectively, back to back histograms.

Cumulative curve

A good way of showing how unequal a distribution is: y axis is always cumulative frequency, x axis is always a measure.

Frequency polygons

For displaying multiple distributions of data. Like a regular line

Change over Time

Give emphasis to changing trends. These can be short (intra-day) movements or extended series traversing decades or centuries. Choosing the correct time period is important to provide suitable context for the reader.

Example FT uses
Share price movements, economic time series, sectoral changes in a market

Line

The standard way to show a changing time series. If data are irregular, consider markers to represent data points.

Column

Columns work well for showing change over time - but usually best with only one series of data at a time.

Column + line timeline

A good way of showing the relationship over time between an amount (columns) and a rate (line).

Slope

Good for showing changing data as long as the data can be simplified into 2 or 3 points without missing a key part of story.

Area chart

Use with care – these are good at showing changes to total, but seeing change in components can be very difficult.

Candlestick

Usually focused on day-to-day activity, these charts show opening/closing and high/low points of each day.

Fan chart (projection)

Use to show the uncertainty in future projections – usually this grows the further forward to projection.

Connected scatterplot

A good way of showing changing data for two variables whenever there is a relatively clear pattern of progression.

Calendar heatmap

A great way of showing temporal patterns (daily, weekly, monthly)

Magnitude

Show size comparisons. These can be relative (just being able to see larger/bigger) or absolute (need to see fine differences). Usually these show a 'counted' number (for example, barrels, dollars or people) rather than a calculated rate or per cent.

Example FT uses
Commodity production, market capitalisation, volumes in general

Column

The standard way to compare the size of things. Must always start at 0 on the axis.

Bar

See above. Good when the data are not time series and labels have long category names.

Paired column

As per standard column but allows for multiple series. Can become tricky to read with more than 2 series.

Paired bar

See above.

Marimekko

A good way of showing the size and proportion of data at the same time – as long as the data are not too complicated.

Proportional symbol

Use when there are big variations between values and/or seeing fine differences between data is not so important.

Isotype (pictogram)

Excellent solution in some instances – use only with whole numbers (do not slice off an arm to represent a decimal).

Lollipop

Lollipop charts draw more attention to the data value than standard bar/column – does not have to start at zero (but preferable).

Radar

A space-efficient way of showing value of multiple variables – but

Part-to-whole

Show how a single entity can be broken down into its component elements. If the reader's interest is solely in the size of the components, consider a magnitude-type chart instead.

Example FT uses
Fiscal budgets, company structures, national election results

Stacked column/bar

A simple way of showing part-to-whole relationships but can be difficult to read with more than a few components.

Marimekko

A good way of showing the size and proportion of data at the same time – as long as the data are not too complicated.

Pie

A common way of showing part-to-whole data – but be aware that it's difficult to accurately compare the size of the segments.

Donut

Similar to a pie chart – but the centre can be a good way of making space to include more information about the data (eg total).

Treemap

Use for hierarchical part-to-whole relationships; can be difficult to read when there are many small segments.

Voronoi

A way of turning points into areas – any point within each area is closer to the central point than any other centroid.

Arc

A hemicycle, often used for visualising parliamentary composition by number of seats.

Gridplot

Good for showing % information, they work best when used on whole numbers and work well in small multiple layout form.

Venn

Generally only used for schematic representation.

Spatial

Aside from locator maps only used when precise locations or geographical patterns in data are more important to the reader than anything else.

Example FT uses
Population density, natural resource locations, natural disaster risk/impact, catchment areas, variation in election results

Basic choropleth (rate/ratio)

The standard approach for plotting data on a map – should always be rates rather than totals and use a sensible base geography

Proportional symbol (count/magnitude)

Use for totals rather than rates – be wary that small differences in data will be hard to see.

Flow map

For showing unambiguous movement across a map.

Contour map

For showing areas of equal value on a map. Can use deviation colour schemes for showing +/- values

Equalised cartogram

Converting each unit on a map to a regular and equally-sized shape – good for representing voting regions with equal value.

Scaled cartogram (value)

Stretching and shrinking a map so that each area is sized according to a particular value.

Dot density

Used to show the location of individual events/locations – make sure to annotate any patterns the reader should see.

Heat map

Grid-based data values mapped with an intensity colour scale. As choropleth map – but not snapped to an admin/political unit

Flow

Show the reader volumes or intensity of movement between two or more states or conditions. These might be logical sequences or geographical locations.

Example FT uses
Movement of funds, trade, migrants, lawsuits, information; relationship graphs.

Sankey

Shows changes in flows from one condition to at least one other; good for tracing the eventual outcome of a complex process.

Waterfall

Designed to show the sequencing of data through a flow process, typically budgets. Can include +/- components.

Chord

A complex but powerful diagram which can illustrate 2-way flows (and net winner) in a matrix.

Network

Used for showing the strength and inter-connectedness of relationships of varying types.

1. INTRODUCTION

2. VISUAL ENCODING

3. BASIC CHART TYPES

4. INTERACTION

5. VISUALIZATION DESIGN

6. DATA PREPROCESSING

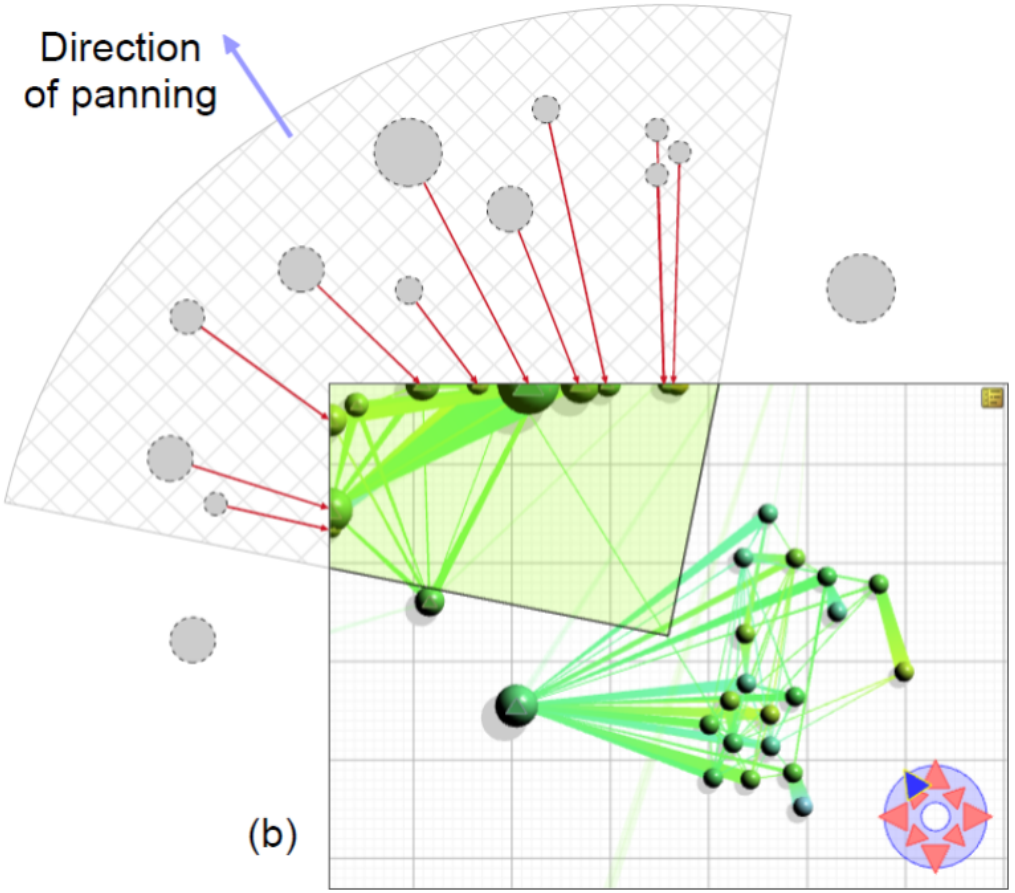
Basics

Visualization Building Blocks & Processes

Visualization Techniques

Cross-cutting Topics

4. INTERACTION WITH VISUALIZATION



5. VISUALIZATION DESIGN

Design Activity Framework

***u**nderstand* | ***i**deate* | ***m**ake* | ***d**eploy*

Exploration Actions

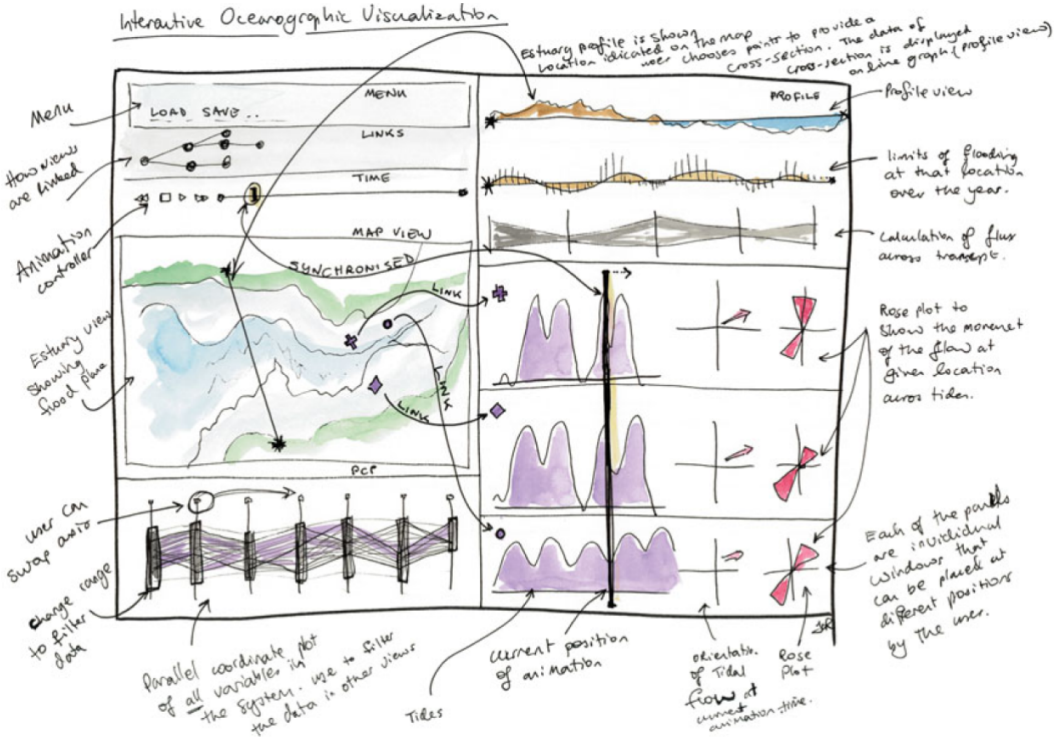
- **Data Exploration Actions**
  - Filter
  - Inspect
  - Query
  - Restore
- **Visual Exploration Actions**
  - Brush
  - Change-Metaphor
  - Change-Range
    - Zoom
    - Pan
  - Merge
  - Sort
  - Split

Insight Actions

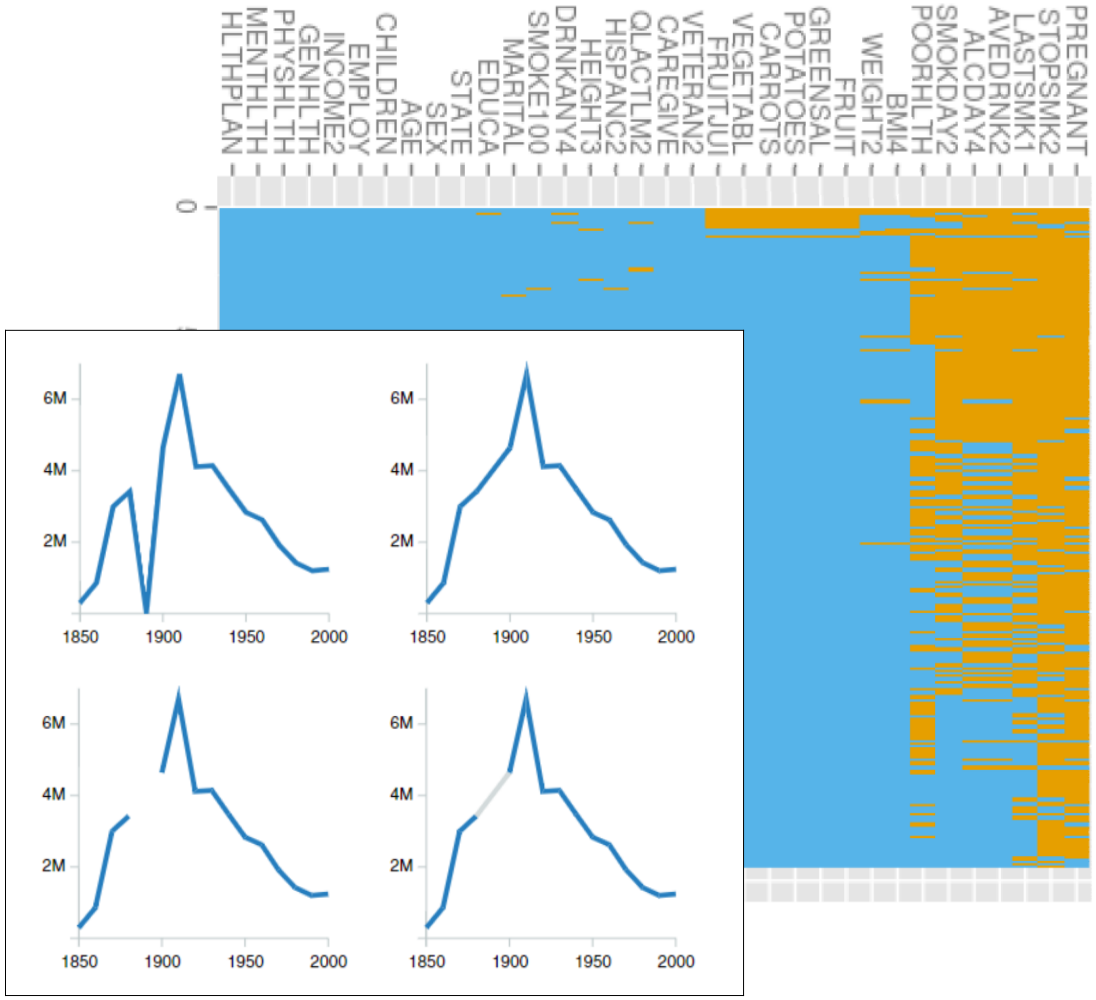
- **Visual Insight Actions**
  - Annotate
  - Bookmark
- **Knowledge Insight Actions**
  - Create
  - Modify
  - Remove

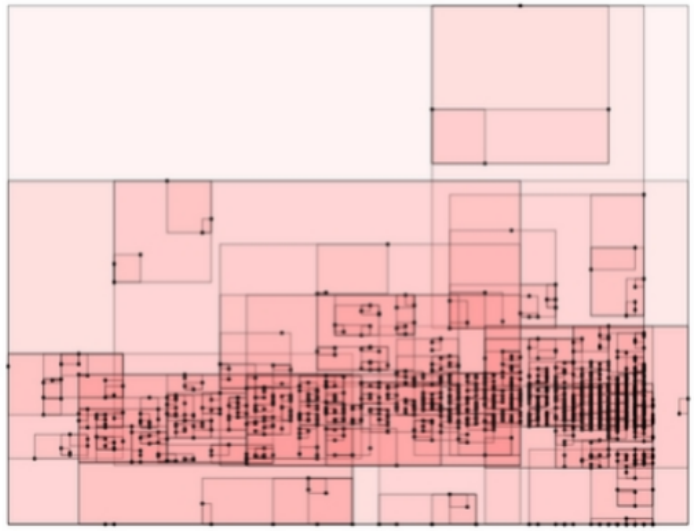
Meta Actions

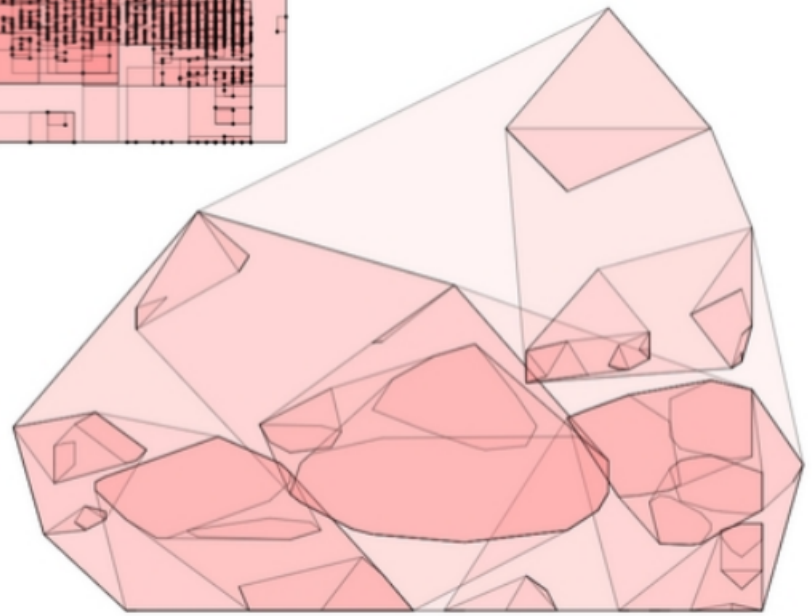
- Delete
- Edit
- Redo
- Revisit
- Undo



6. DATA PREPROCESSING







1. INTRODUCTION

2. VISUAL ENCODING

3. BASIC CHART TYPES

4. INTERACTION

5. VISUALIZATION DESIGN

6. DATA PREPROCESSING

7. RECAP 1st Half

Basics

Visualization Building Blocks & Processes

Visualization Techniques

Cross-cutting Topics


FALL BREAK

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

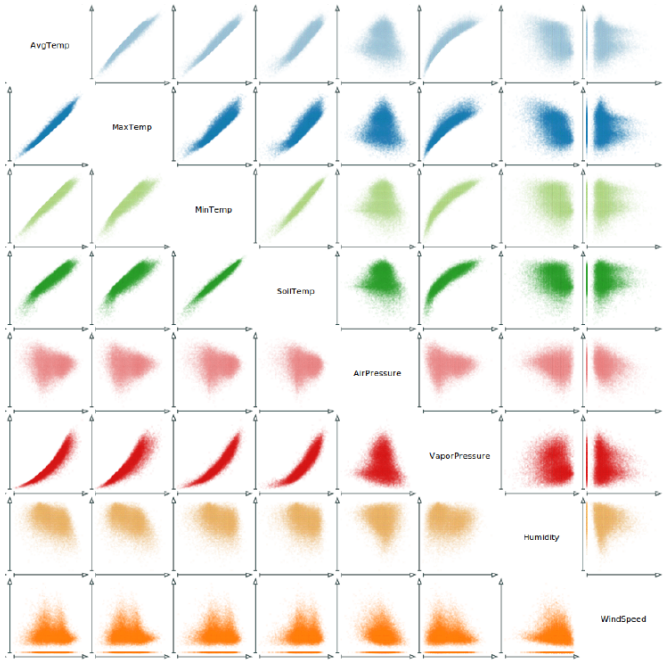
10. GEOSPATIAL DATAVIS

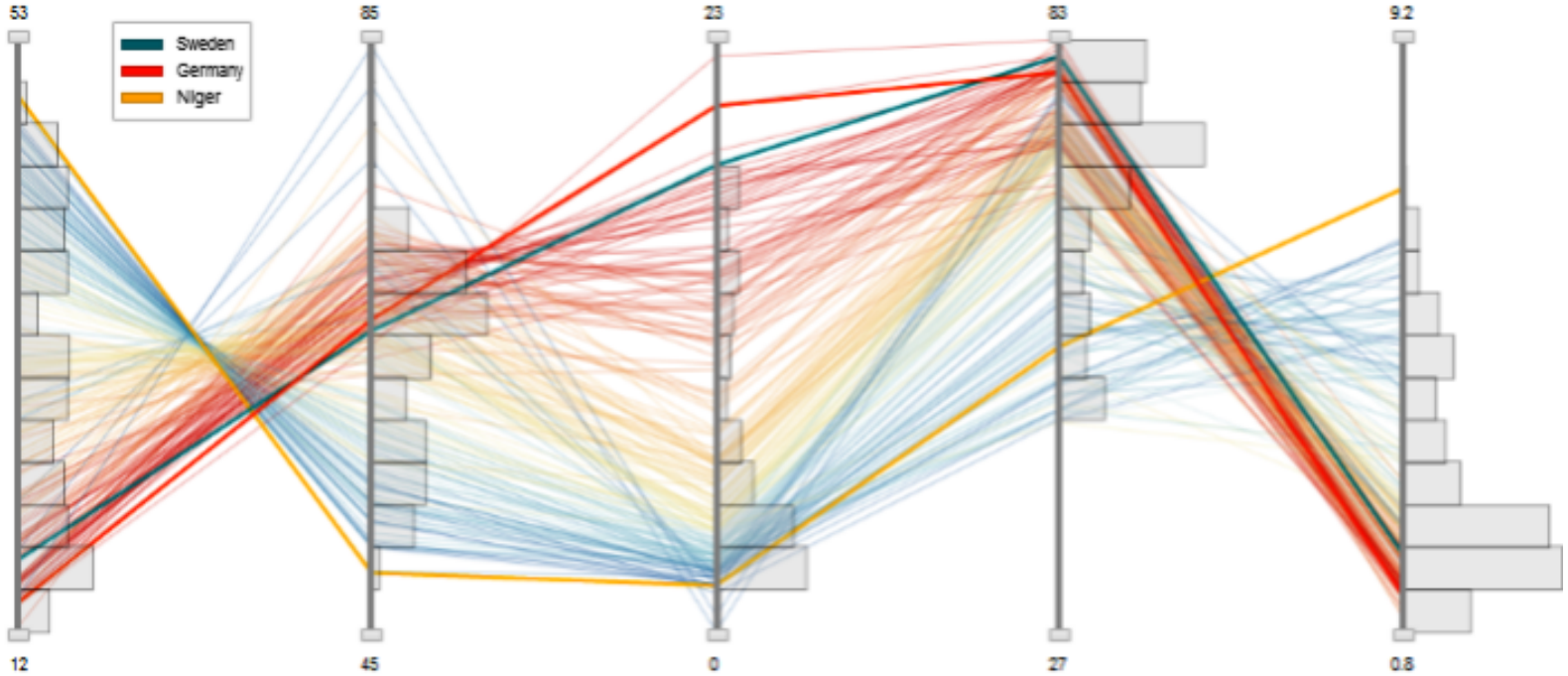
11. GRAPH DATAVIS

Visualization Techniques

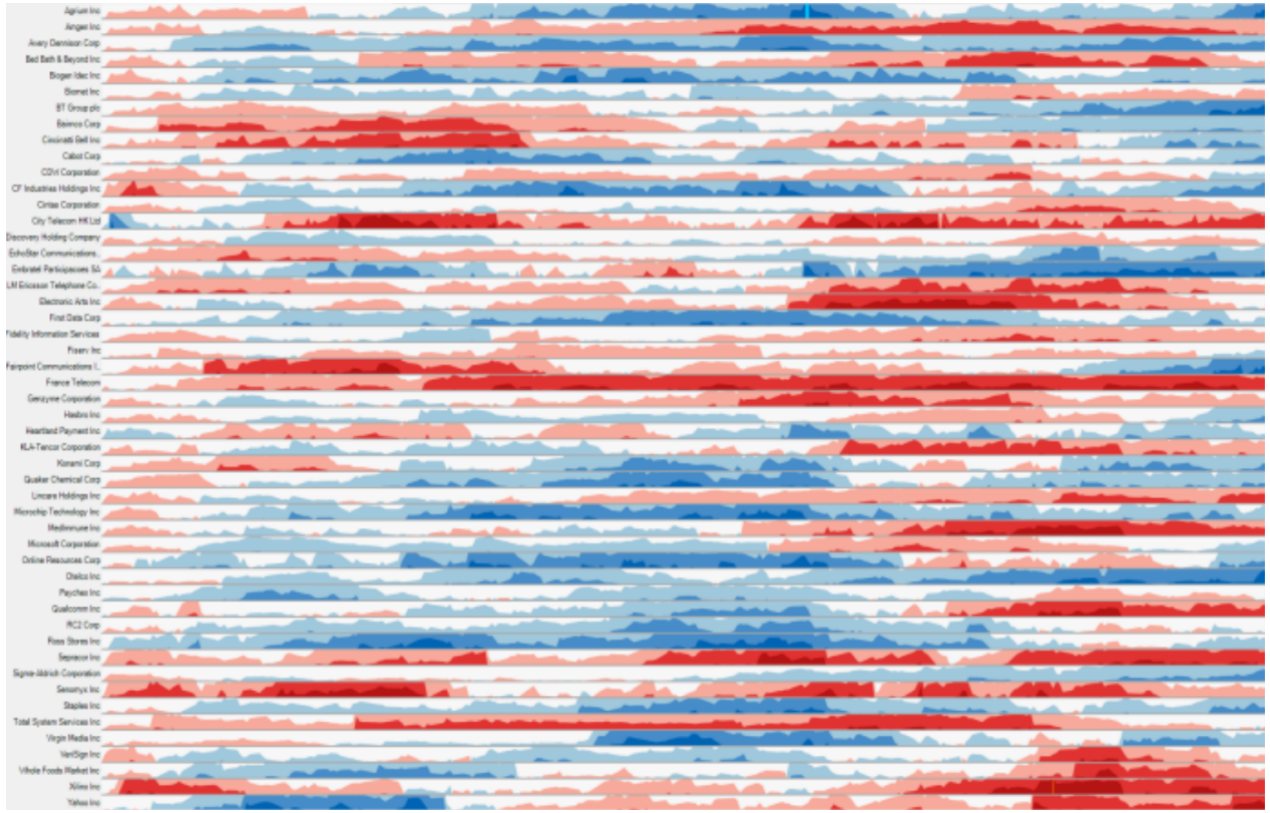
Visualization Applications

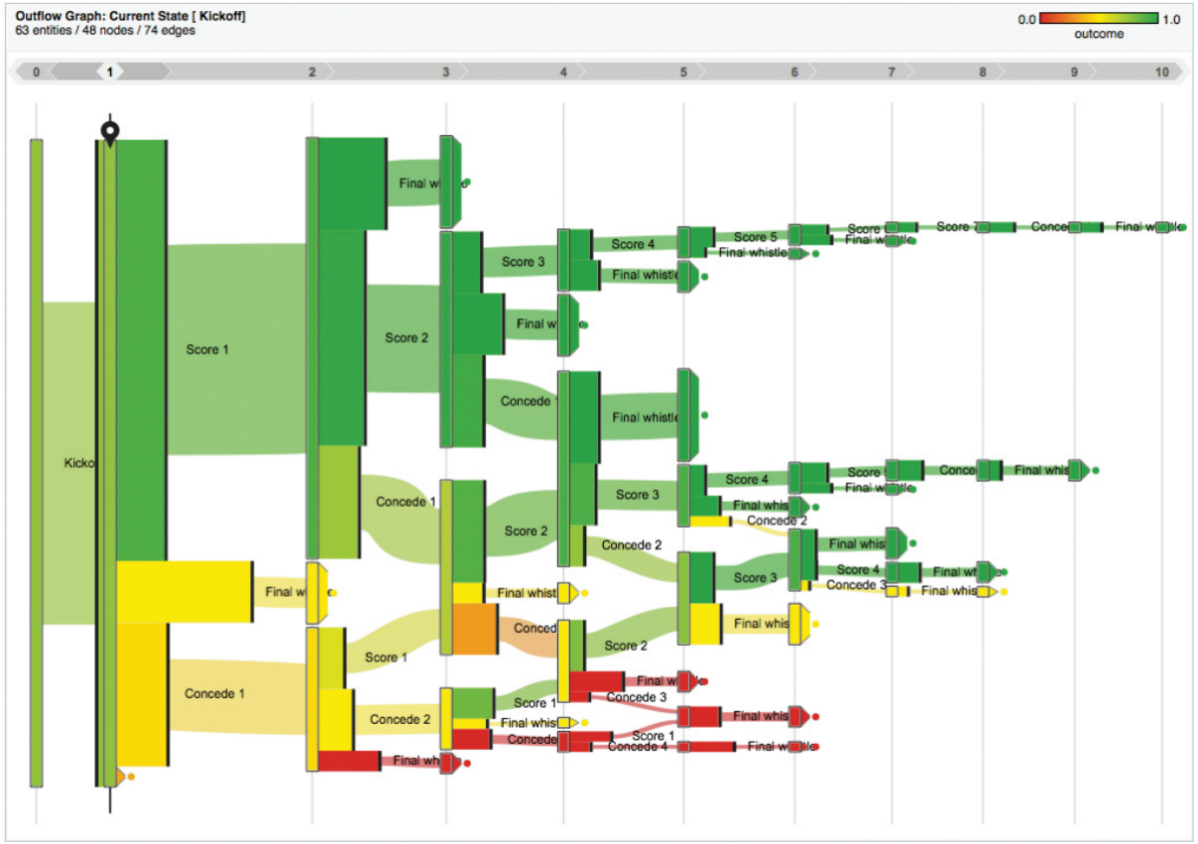
8. MULTIVARIATE VISUALIZATION



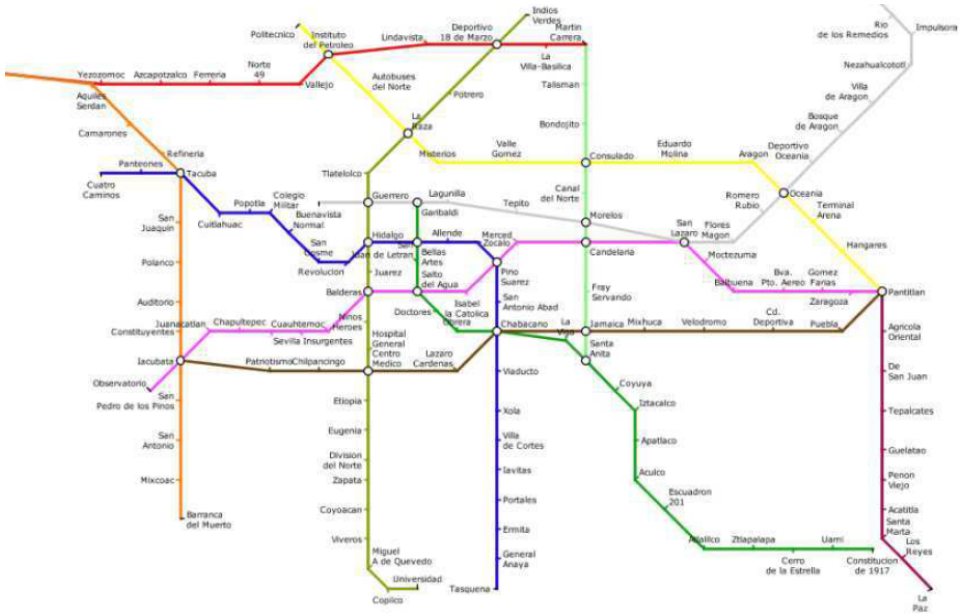


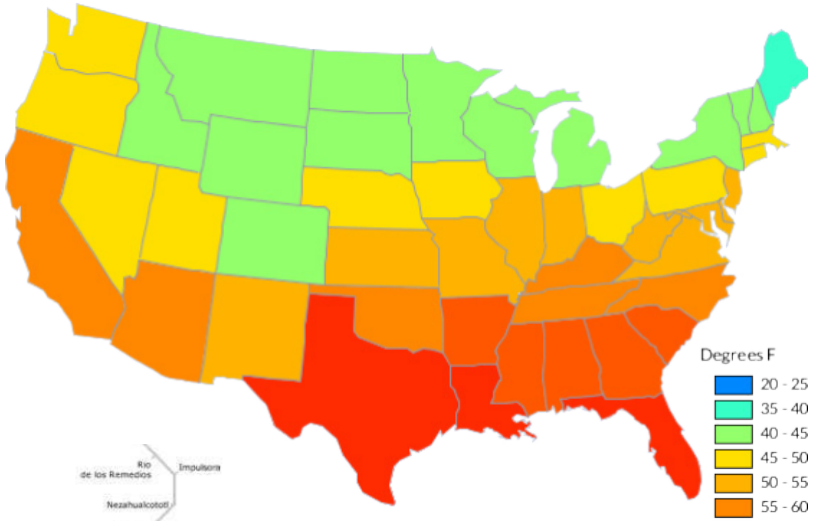
9. VISUALIZATION OF TEMPORAL DATA

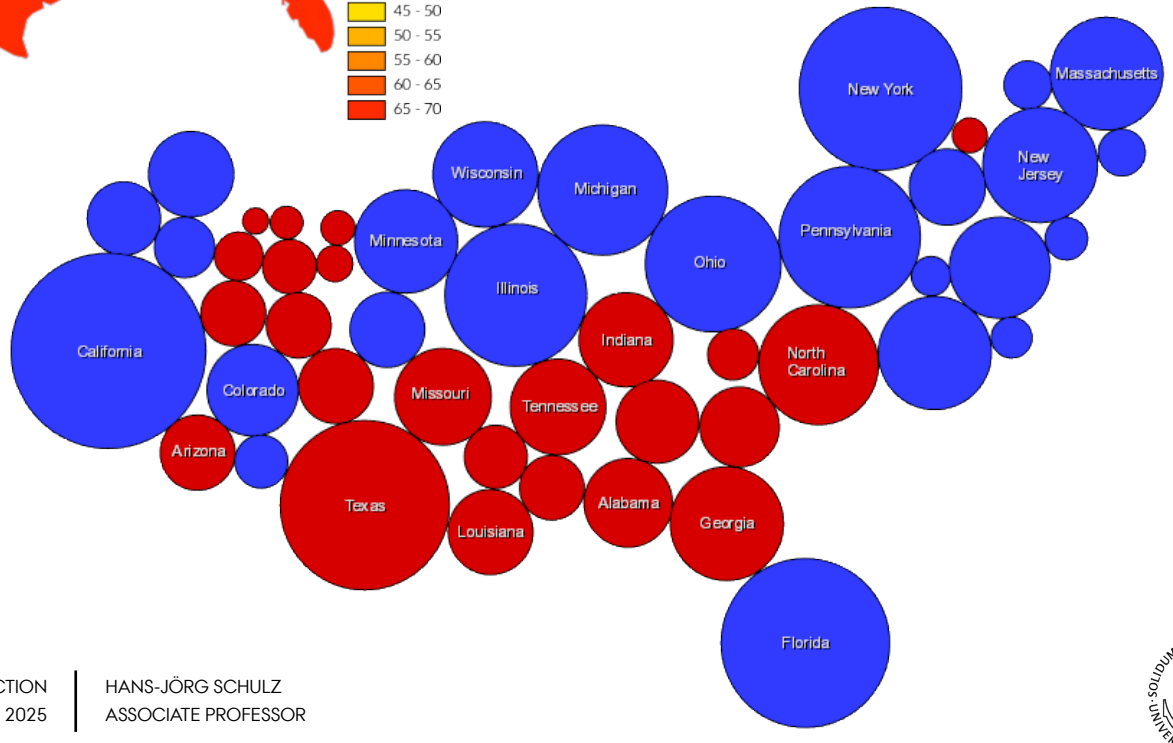




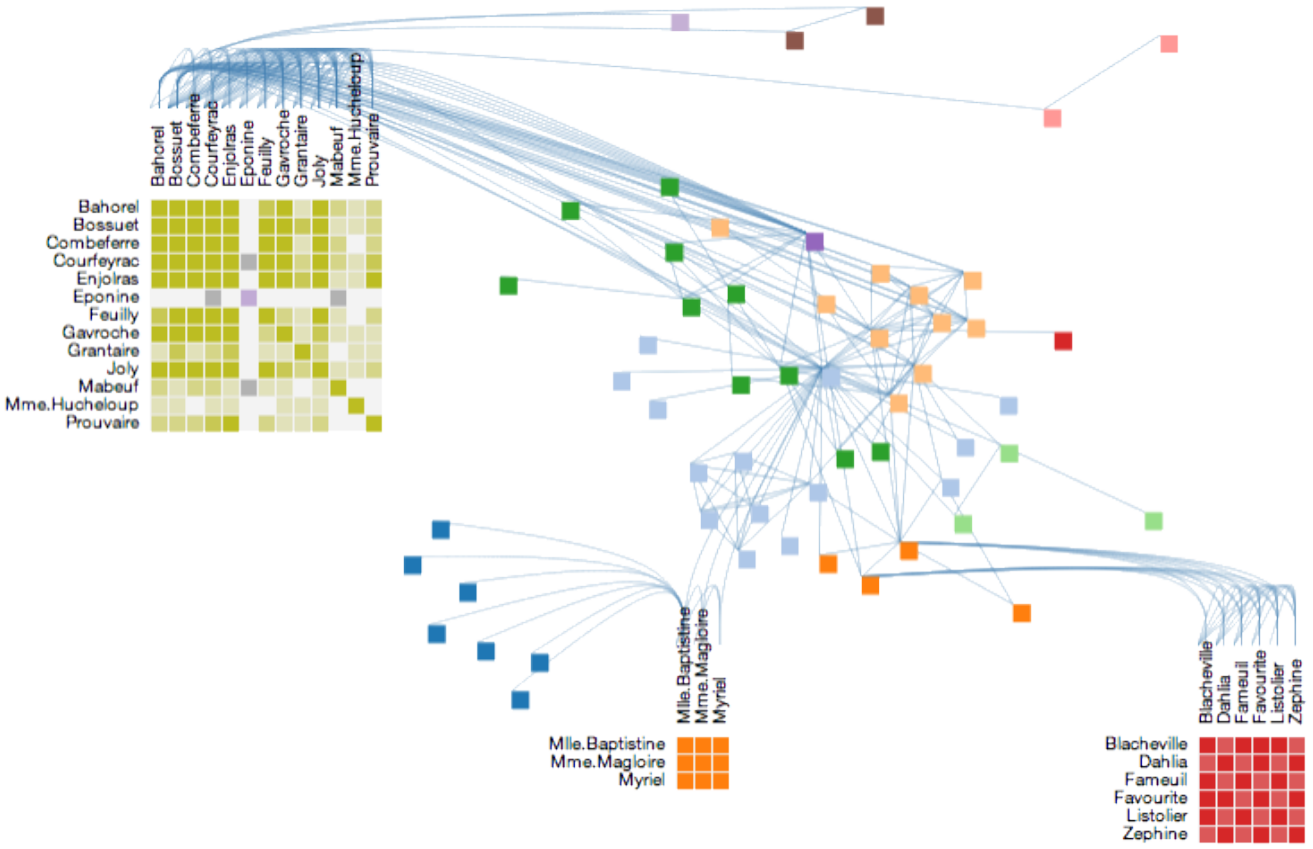
10. VISUALIZATION OF GEOSPATIAL DATA

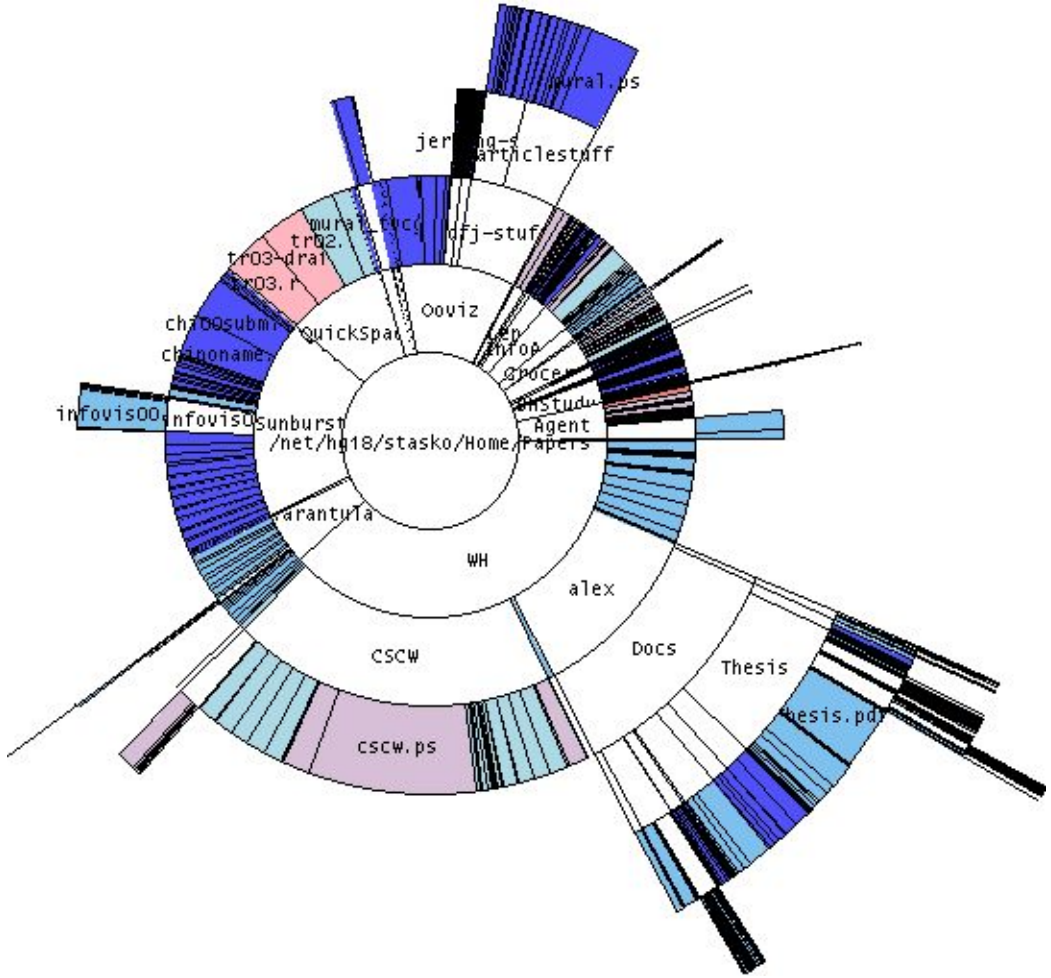






11. VISUALIZATION OF GRAPH DATA





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

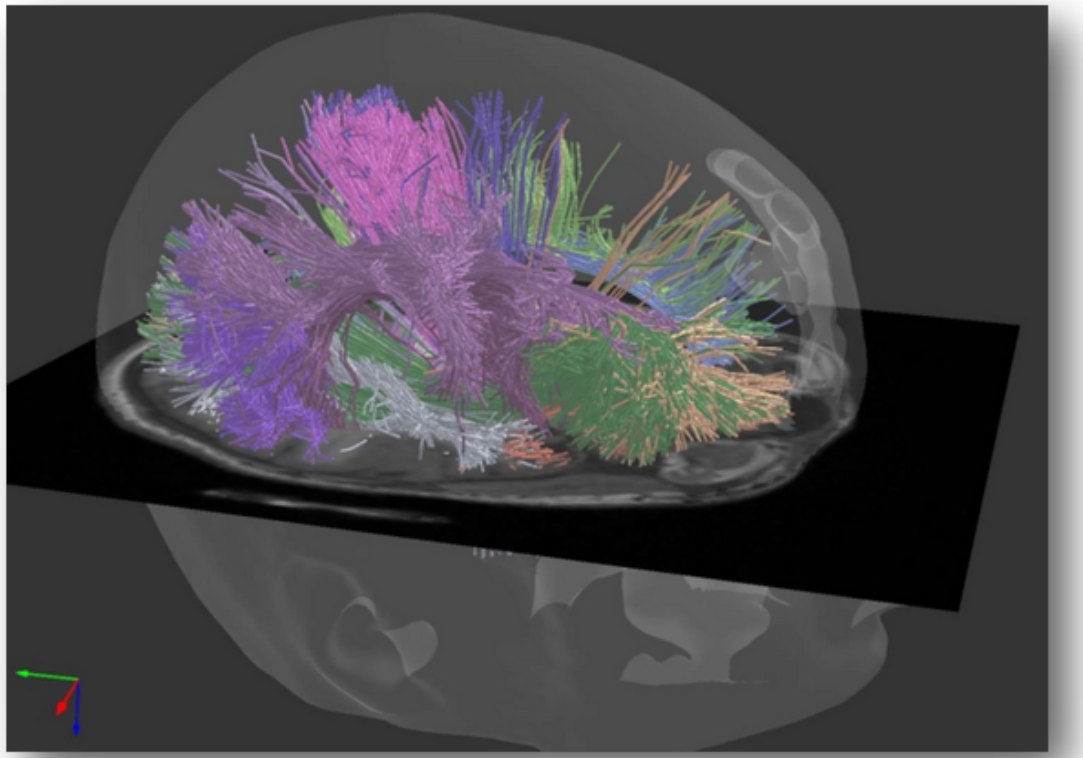
Visualization Techniques

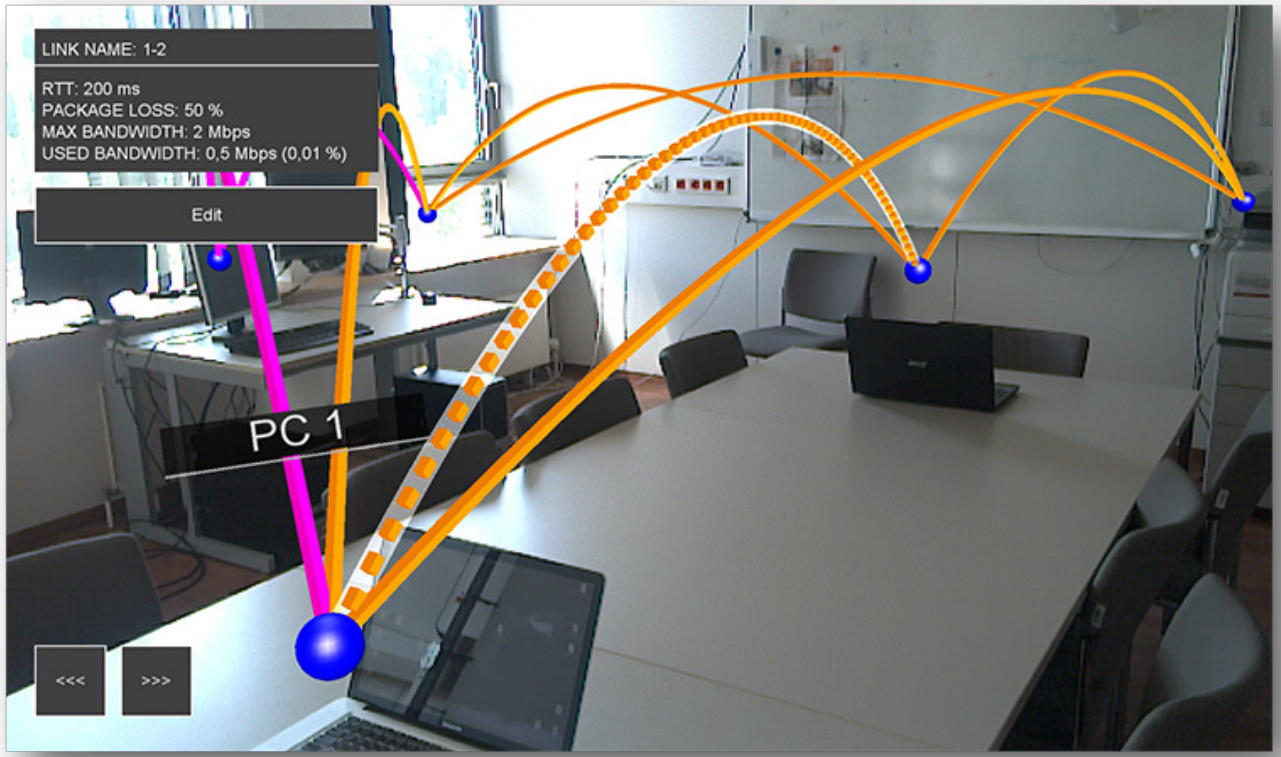
12. 3D DATAVIS

13. VISUAL ANALYTICS

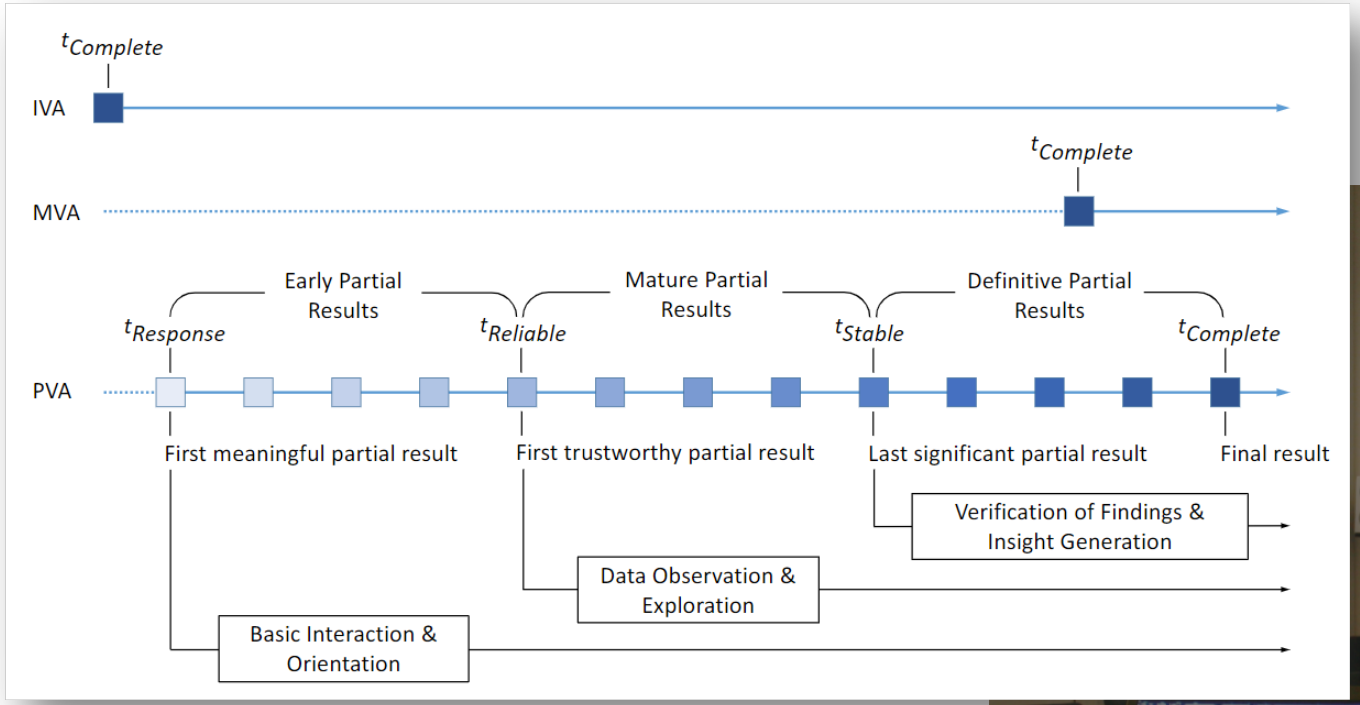
Visualization Applications

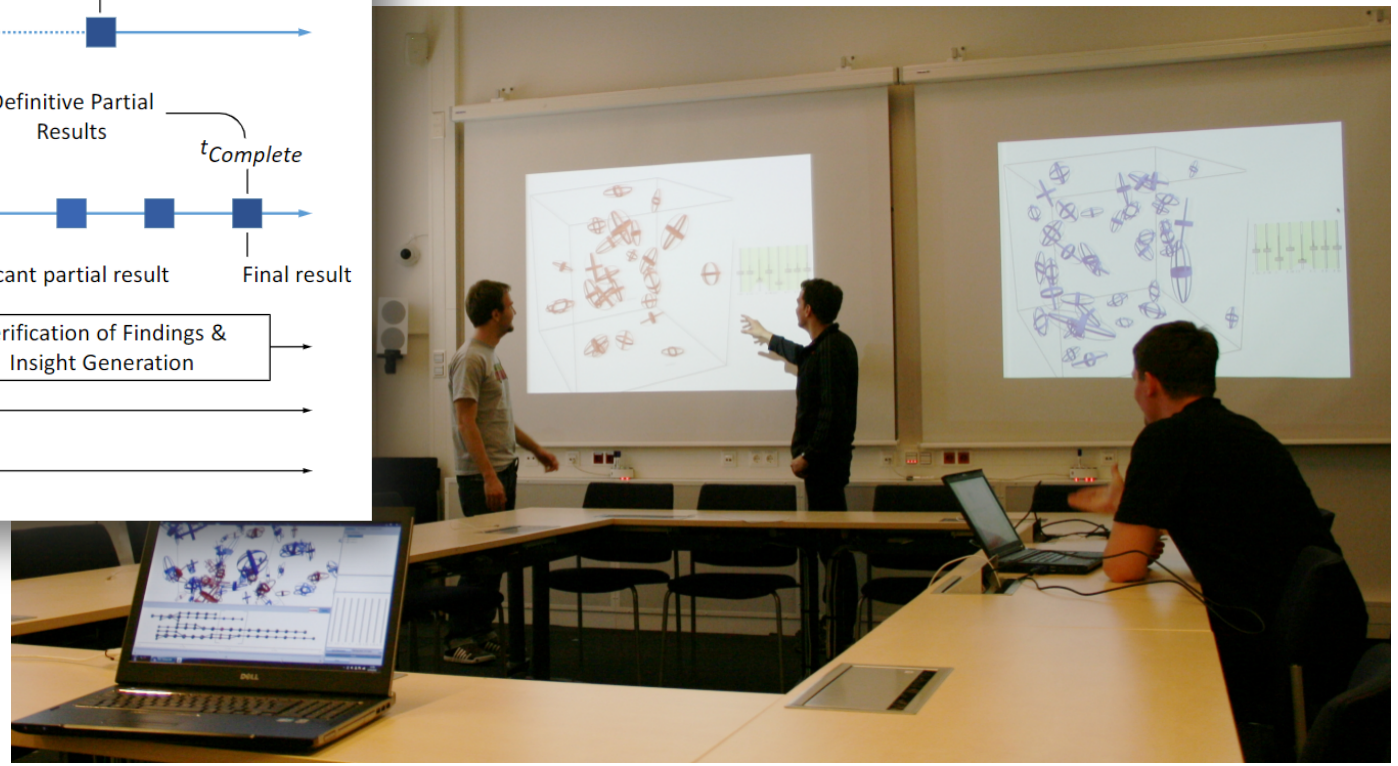
12. 3D DATA VIS





13. VISUAL ANALYTICS





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

READING FOR TODAY

CHAPTER 1

Introduction

CONTENTS

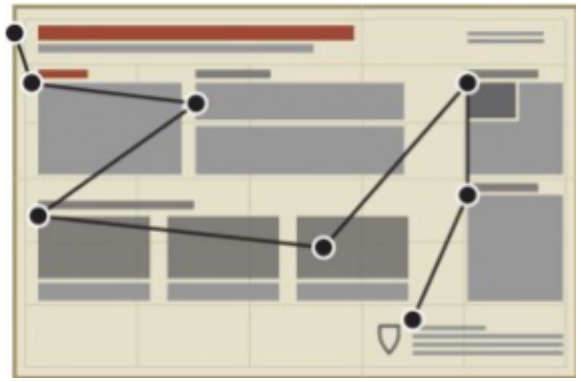
READING FOR NEXT WEEK

CHAPTER 3

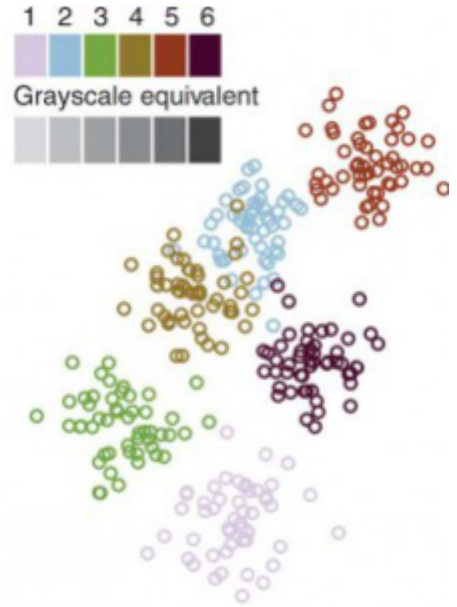
Visualization Methods and Techniques

CONTENTS

NATURE METHODS: “POINTS OF VIEW”

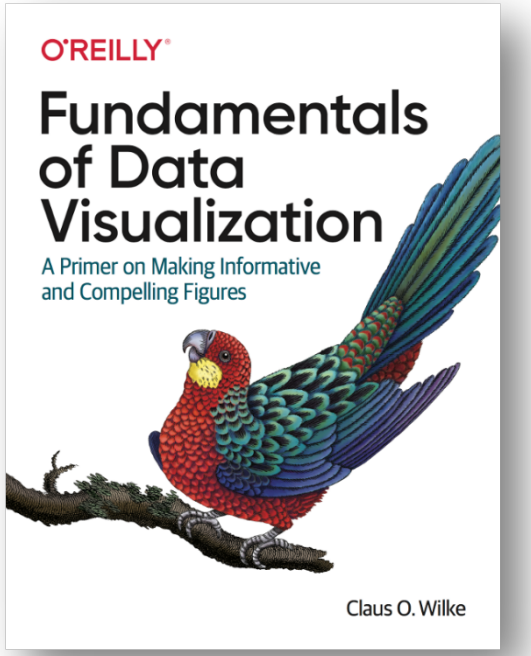


- Visual Design & Style
- Layout
- Negative Space
- Typography
- Labels & Callouts
- Color Coding
- Axes, Ticks, Grids
- Arrows
- Plot Types

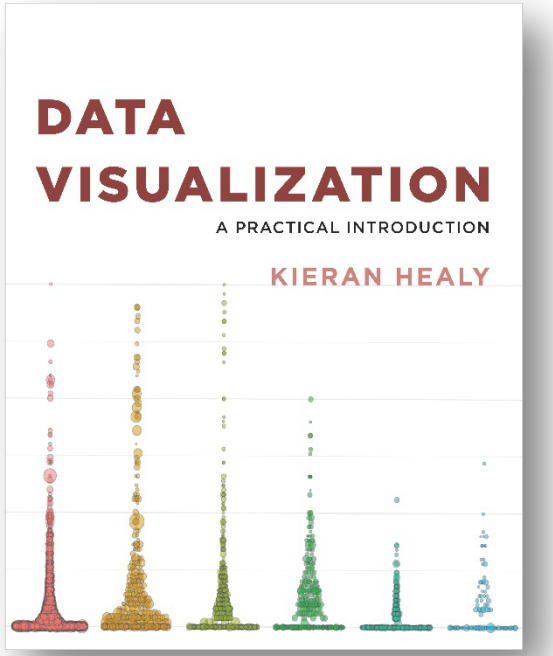


http://blogs.nature.com/methagora/2013/07/data-visualization-points-of-view.html

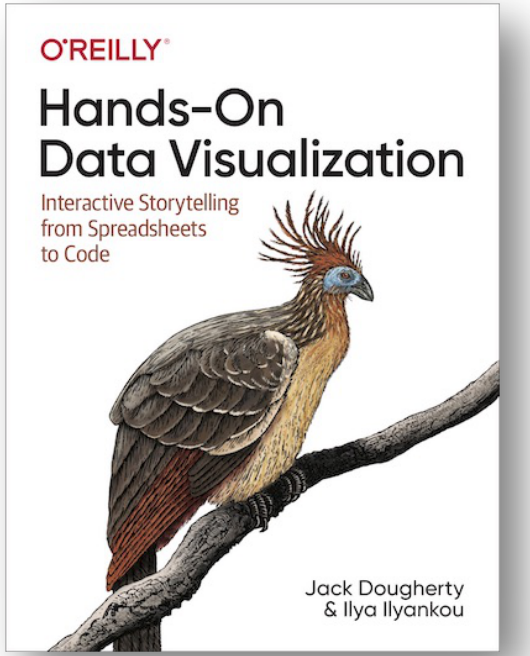
FREE BOOKS...



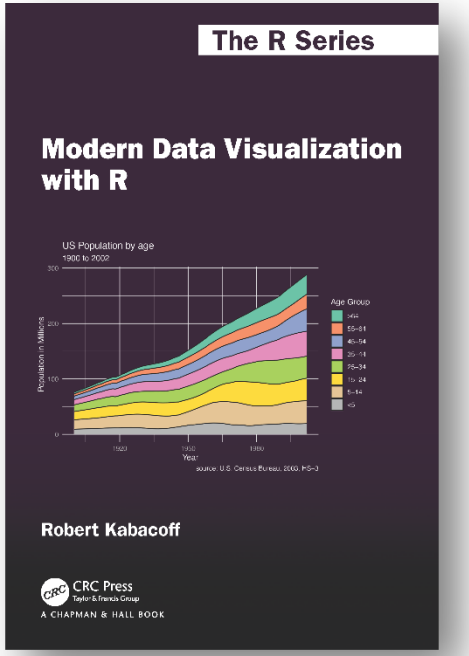
https://serialmentor.com/dataviz/



https://socviz.co



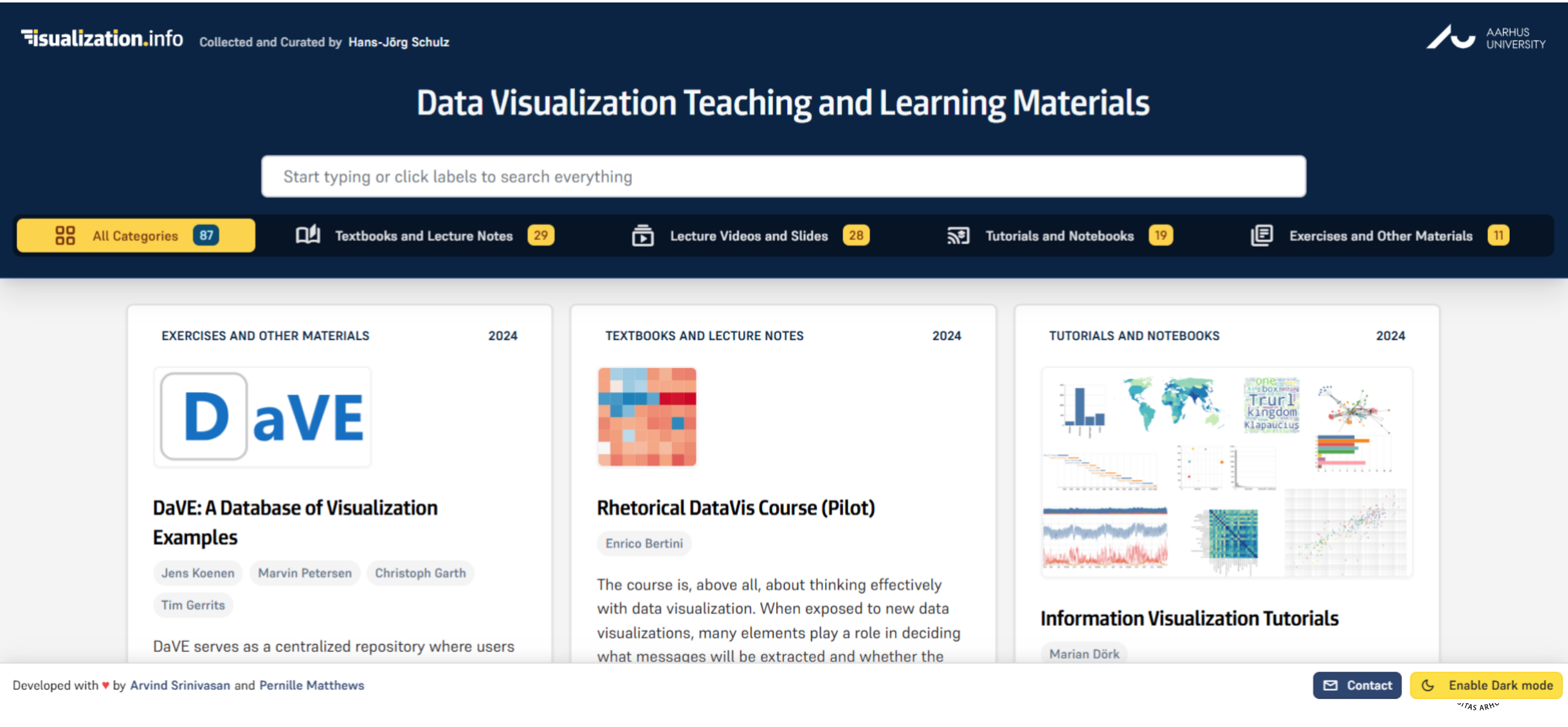
https://handsondataviz.org



https://rkabacoff.github.io/datavis/

FREE LEARNING RESOURCES

https://visualization.info



LIST OF LITERATURE SOURCES

- Anscombe 1973: https://www.jstor.org/stable/2682899
- Berinato 2016: https://hbr.org/2016/06/visualizations-that-really-work
- Cairo 2016: https://cran.r-project.org/web/packages/datasauRus/vignettes/Datasaurus.html
- Card 2008: chapter 26 in https://doi.org/10.1201/9781410615862
- Card, Mackinley 1999: Fig.1.23 in https://hci.ucsd.edu/220/CMSChap1_Using_Vision_to_Think.pdf
- Chen et al. 2014: https://doi.org/10.1007/978-3-319-07121-3_5
- Haber, McNabb 1990: https://www.academia.edu/download/40903302/HaberMcNabb1990.pdf
- Himmelspach 2016: https://data-cuisine.net/data-dishes/onionland
- Leithinger et al. 2011: https://doi.org/10.1145/2047196.2047268
- Lukas 2015: https://visualprogramming.net/blog/2015/smelling-data--translating-data-into-scents/
- Lunzner, MacNamara 2014: https://tinlizzie.org/VPRIPapers/m2014002_robust.pdf
- Marty 2009: ISBN=0321510100
- McCormick et al. 1987: http://www.sci.utah.edu/vrc2005/McCormick-1987-VSC.pdf
- McGurk Effect: https://doi.org/10.3389/fpsyg.2014.00725
- Owen 1999: https://education.siggraph.org/static/HyperVis/hypervis.htm

VISUALIZATION PROJECT

AARHUS
UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

HANS-JÖRG SCHULZ
ASSOCIATE PROFESSOR

WHY DO A PRACTICAL PROJECT?

- To understand how the concepts, algorithms, and frameworks from the lecture relate to practical questions and issues of visualization
- To test-drive your visualization skills in an environment of frequent constructive input and guidance from the TAs
- To improve your visual literacy and critical thinking skills
- To showcase what you've learned in this lecture beyond the mere grade (e.g., as part of a project portfolio for job applications)
- To find out whether data visualization is actually your thing and which aspects of it you like best (graphic design, software engineering, data analytics, visual storytelling / data journalism,...)
- **To prove to me that you have engaged with the lecture contents!**

HOW DOES THE DAVI PROJECT WORK?

Group project: 3-4 students per group

3-persons: 9.9
2-persons: 9.4
1-person: 7.8

Why a group project?

- More works hours / project => better outcomes:
  - more interesting/complicated problems
  - more mature results
- More diversity / project => better outcomes:
  - larger skillset/expertise
  - more eyes/opinions

Higher Grades

PRECONDITIONS & EXCEPTIONS

Precondition: You must align on your group's ambition!

Exceptions under which you may also perform a solo project:

- You are a PhD student and want to use the project's outcome for your thesis
- You want to integrate the project with a side job you have
- You want to combine the project with another course project
- You want to visualize confidential data
- You cannot work in a team for obvious reasons (e.g., remote participation)

HOW DOES THE DAVI PROJECT WORK?

You pick a dataset and build an interactive visualization for it.

Must-haves:

- Make sure you are an expert on that data!
- Make sure the data is big enough to warrant visualization!
- Make sure to pick data of reasonable quality!

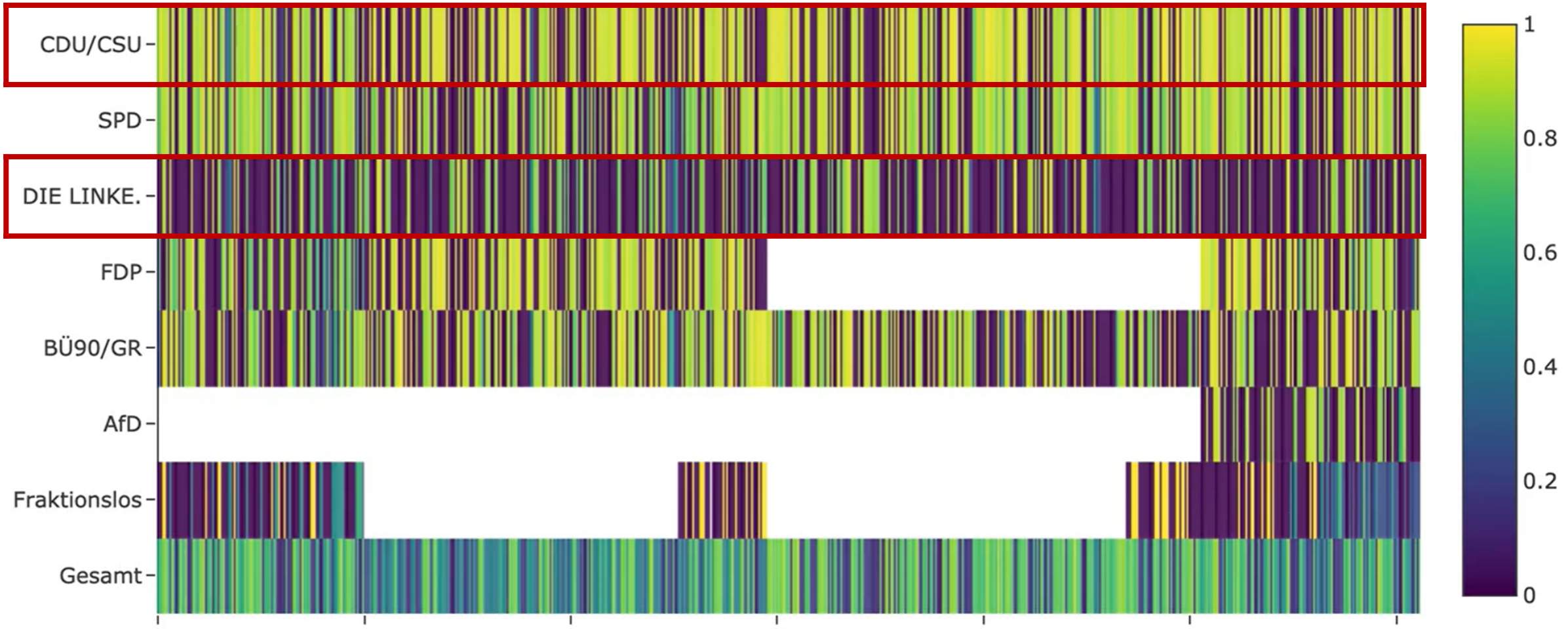
Nice-to-haves:

- Align the project with other work you have to do.
- Choose the project so that there is a high possibility for follow-up work.
- Use the project to jump-start your career.

FOR YOU TO DECIDE

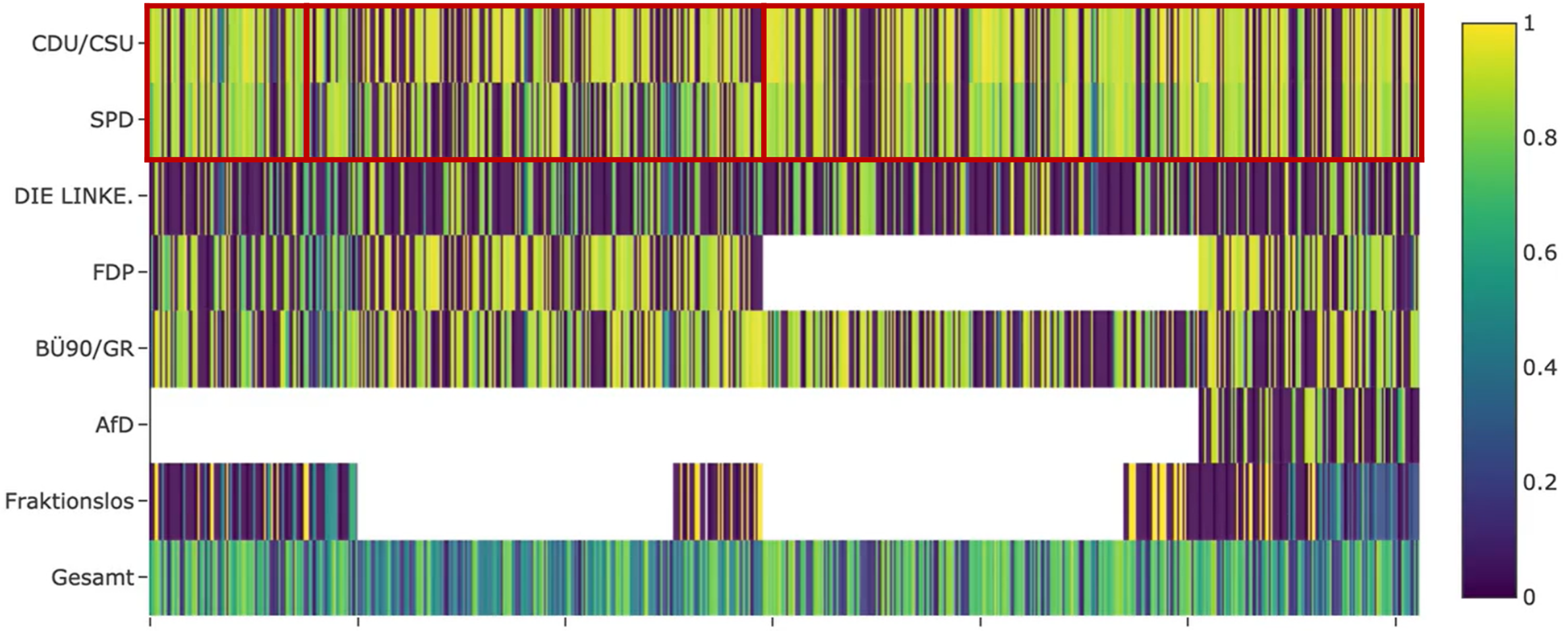
- Your project team (try to complement each other)
- Dataset (large and complex, but reasonable quality)
- What you want to do with it (goal, question)
- Tooling (must fit your problem)
- Your users and mode of feedback (the more the merrier)
- If and how to make the result public
  (Github, Live Demo on a webserver, Publication,...)

GERMAN PARLIAMENT VOTING BEHAVIOR



Courtesy Nils Simon & Niklas Müller 2019

GERMAN PARLIAMENT VOTING BEHAVIOR



Courtesy Nils Simon & Niklas Müller 2019

Dash

127.0.0.1:8050/#anchor

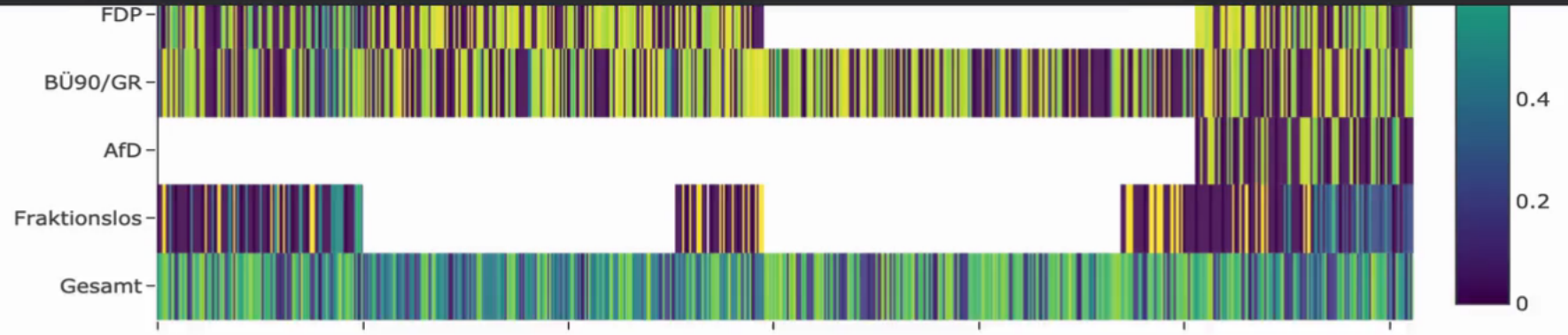
Select...

Anträge durchsuchen:

ZOOM IN CLEAR VIEW

Information

Diese Visualisierung entstand im Rahmen des Kurses Data Visualization an der Aarhus Universität unter Betreuung von Prof. Dr. Hans-Jörg Schulz im Winter 2019. Erstellt wurde diese von Nils Simon und Niklas Müller. Dieses Tool visualisiert die Abstimmungsergebnisse der Parteien zu den Anträgen im Zeitraum 2007-2019.Individuelle Exploration der Daten wird durch Interaktivität ermöglicht. Mittels der DRS im vollständigen Antragstext können die original Dokumente unter pdok.bundestag.de



Antrag vollständig

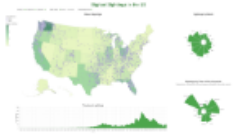
Antrag auswählen um vollständige Information und ein **Balkendiagramm links** anzuzeigen.

Course Projects

2022

Explorable Cluster Workload Visualization *Malthe Thingholm, Johan Kjær Nielsen, Ivan Luchev*

Companies such as Google manage vast quantities of data over their various services. One such service is BORG, their computing cluster service. The dataset contains information about machines, deploys, and metrics, but it's so big that getting any useful information or even understanding what you're looking at is incredibly difficult and time-consuming. This visualization approaches the management and analysis of this data through data exploration. The final result is an interactive, expressive visualization that shows the user where the most CPU usage hours lie and allows the user to navigate to every corner of the data. On a larger scale, it allows the user to see relations in the data that would otherwise be obscured.

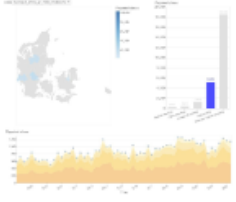


Data Visualization for Exploring Spatial-Temporal Distributions of Bigfoot Sightings in the US *Stinna Weisberg Danger, Johannes Ellemose*

We constructed a visualization of Bigfoot sightings in the US using data from the Bigfoot Field Research Organization. The visualization consists of 4 views including a choropleth map, two radial charts, and an area chart all views the distribution of reported Bigfoot behavior over space and time. The visualization's interactivity consists of cross-filtering between these views, allowing the user to explore the data and gain insights on sasquatches' behavioral patterns.

Electricity Visualization for German Energy Data *Michael Jörg Allgaier, Anthon Løbner, Denis Leu*

The Electricity Visualization for Germany contains data on electricity generation, consumption, import/export and emissions per kilowatt hour for the German electricity sector from 2015 to 2022. The data is combined and graphically purified into a comprehensive visualization utilizing several charts. Spanning from time series and column charts over a heatmap to a fully self-composable scatterplot, we enable the user to explore the German electricity sector from various perspectives. With the visualization, the user is optimally equipped to discover patterns, observe outliers and to develop an in-depth insight into the German electricity sector.



Visualizing Denmark's Crime Statistics *Joachim Brendborg, Asger Ullersted Rasmussen*

This project was about visualizing the crime rates in different areas of Denmark. It shows over 5 different categories, how the crime rate has developed over time and across municipalities.

EVALUATION CRITERIA

- MSc level course => scholarly, academic project
- 10 ECTS => ~250-280 hrs. workload / student
  => ~150 hrs. for the course project
- Visualization project => focus on analytic data visualization
  => correctness of visualization
- Visualization design decisions are argued for
- Resulting visualization must be “interactive”
- Hand-in: Written report (max. 6 pages) + Video demo

Due: 05-DEC-2025 Late submissions until 12-DEC-2025

PROJECT IDEAS & POTENTIAL DATASETS

Dataset Hubs:

- https://github.com/awesomedata/awesome-public-datasets
- https://data.fivethirtyeight.com
- https://github.com/BuzzFeedNews
- https://cloud.google.com/bigquery/public-data/
- https://datasetsearch.research.google.com
- https://www.datavejviser.dk or https://www.opendata.dk

Possible Datasets to explore:

- Formula 1 Data: https://openf1.org
- Linux Kernel Releases: https://kernelnewbies.org/LinuxVersions (starting point)
- Wikipedia edits


SOLIDUM PETIT IN PROFUNDIS
UNIVERSITAS ARHUSIENSIS