Computer-Based Visualization

Data Mining and Visualization

Université de Rennes 1, M2 SIF
Ferran Argelaguet
Alexandre Termier
Peggy Celier

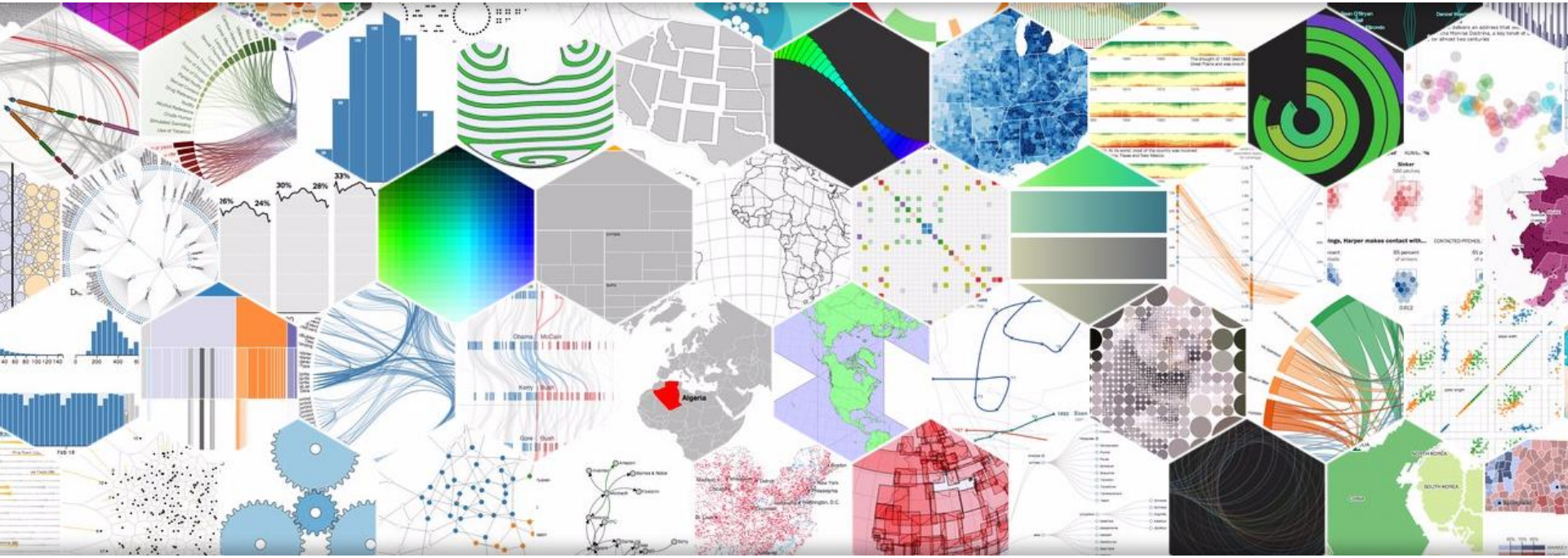
Web site of the course (Remember)

- http://people.irisa.fr/Alexandre.Termier/dmv/

- Web site contains:
  - General information
  - Up-to-date schedule (will be the reference)
  - Links to documents

Definition

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively



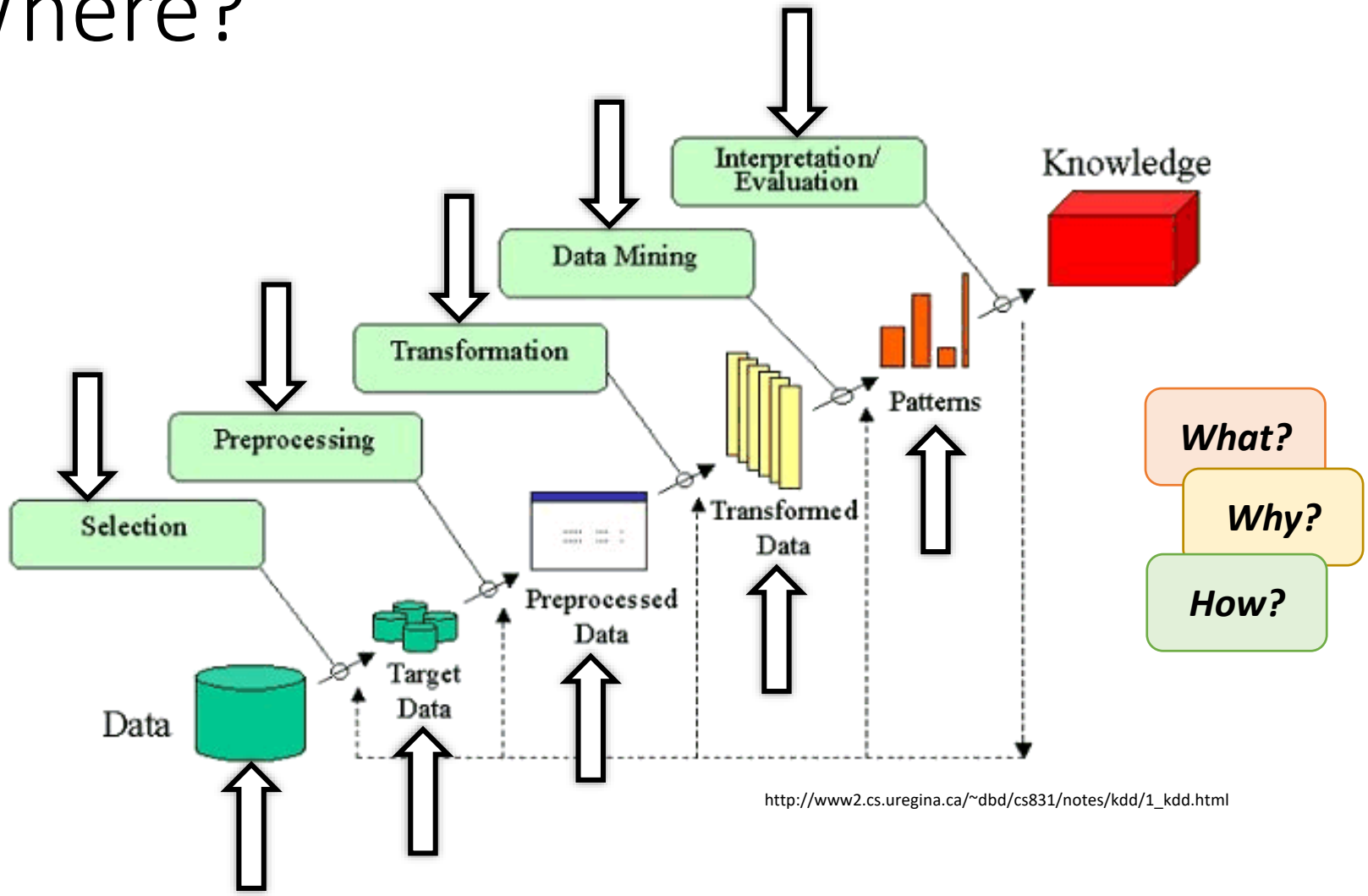
http://3djs.org

Context

Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods

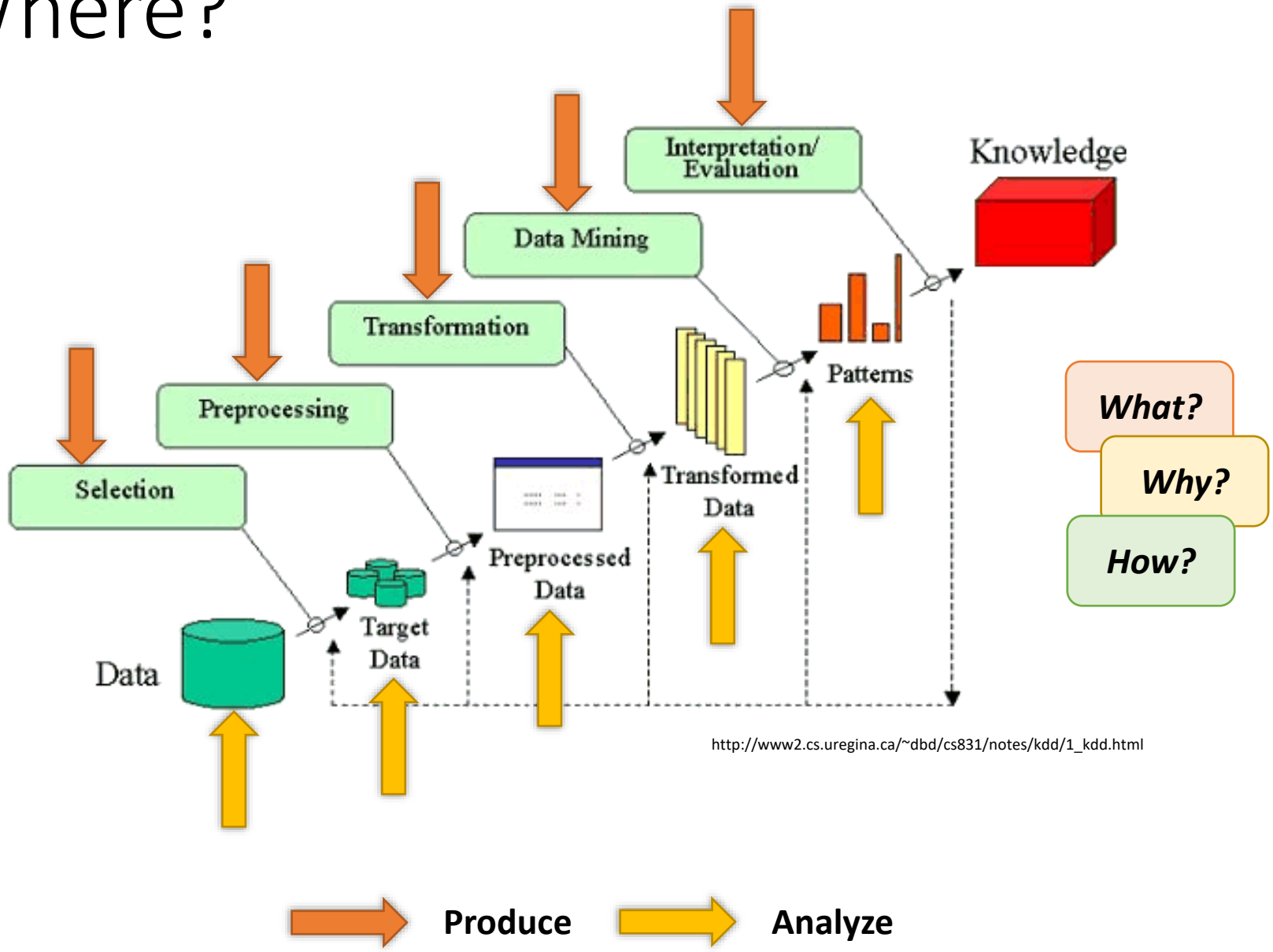
- Need to present results to users -> data visualization
  - Huge lack of communication between data mining / data vis community
  - This course: a small step to improve this communication

Where?



http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html

Where?


Interpretation/
Evaluation
Knowledge
Data Mining
Transformation
Patterns
Preprocessing
Selection
Transformed
Data
Preprocessed
Data
Target
Data
Data
What?
Why?
How?
http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html
Produce
Analyze

Simple Example

1.1.1:MY,NP;MY,MB
1.1.2:MY,ME;ME,MB
1.1.3:MY
1.1.4:MY,ME;MY,CL;MY,GE;MY,MC;MY,MB
1.1.5:MY,MB,ME
1.1.6:ME,MY
1.1.7:MY,CV;MY,MB,ME
1.1.8:SN,MY
1.1.9:MB
1.1.10:MY,GG
1.1.11:MY
1.1.12:MY
1.1.13:MY
1.1.14:MY,SN
1.2.1:JL,JV;JV,MT;MR,JV
…

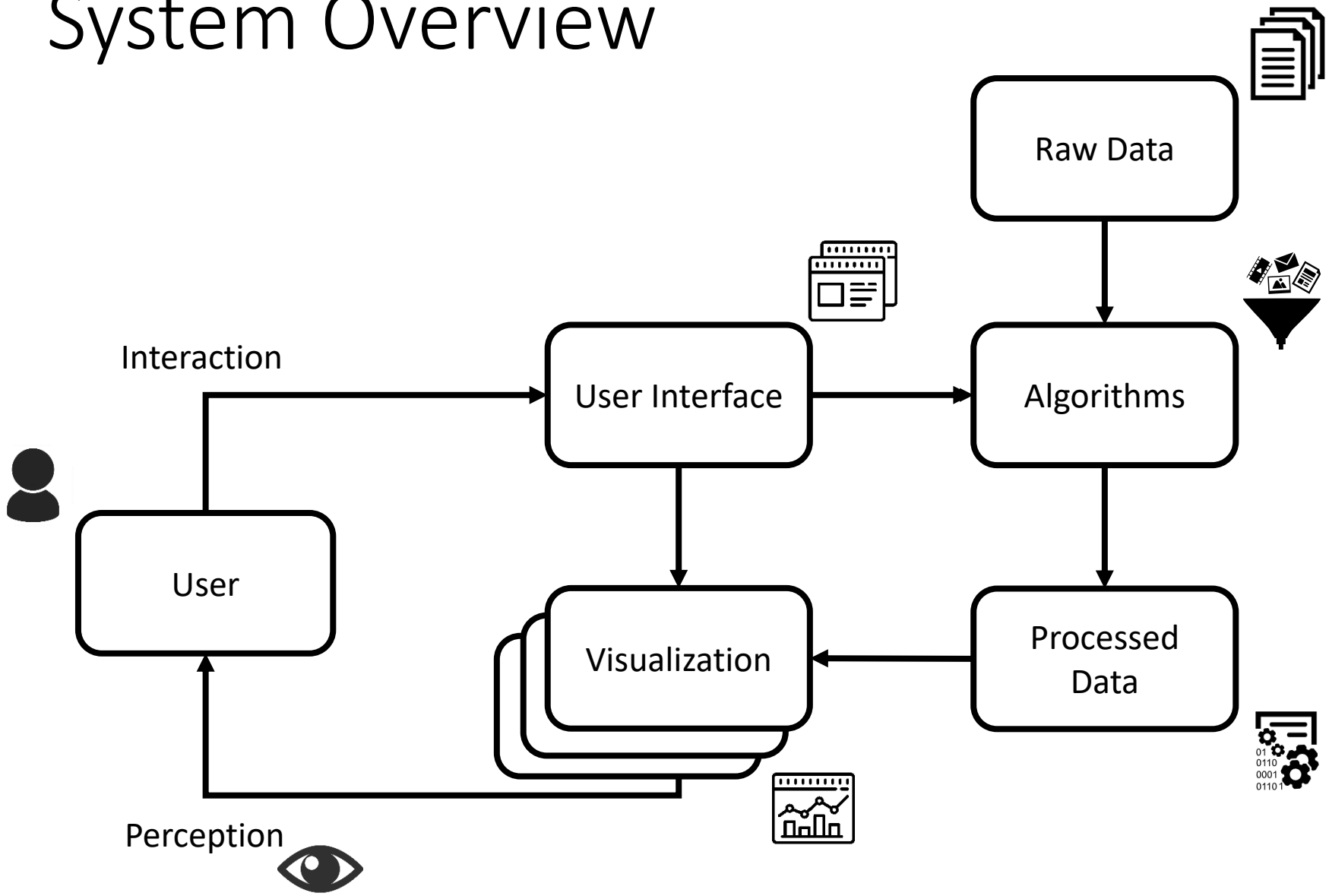
https://bost.ocks.org/mike/miserables/

Context

- Scientific visualization is mostly concerned with:
  - 2, 3, 4 dimensional, spatial or spatio-temporal data
  - Discretized data

- Information visualization focuses on:
  - High-dimensional, abstract data
  - Discrete data
  - Financial, statistical, etc
  - Visualization of large trees, networks, graphs
  - Data mining: finding patterns, clusters, voids, outliers

System Overview



Content

- Introduction
- Data abstraction
- Task abstraction
- Visualization Design
- Validation

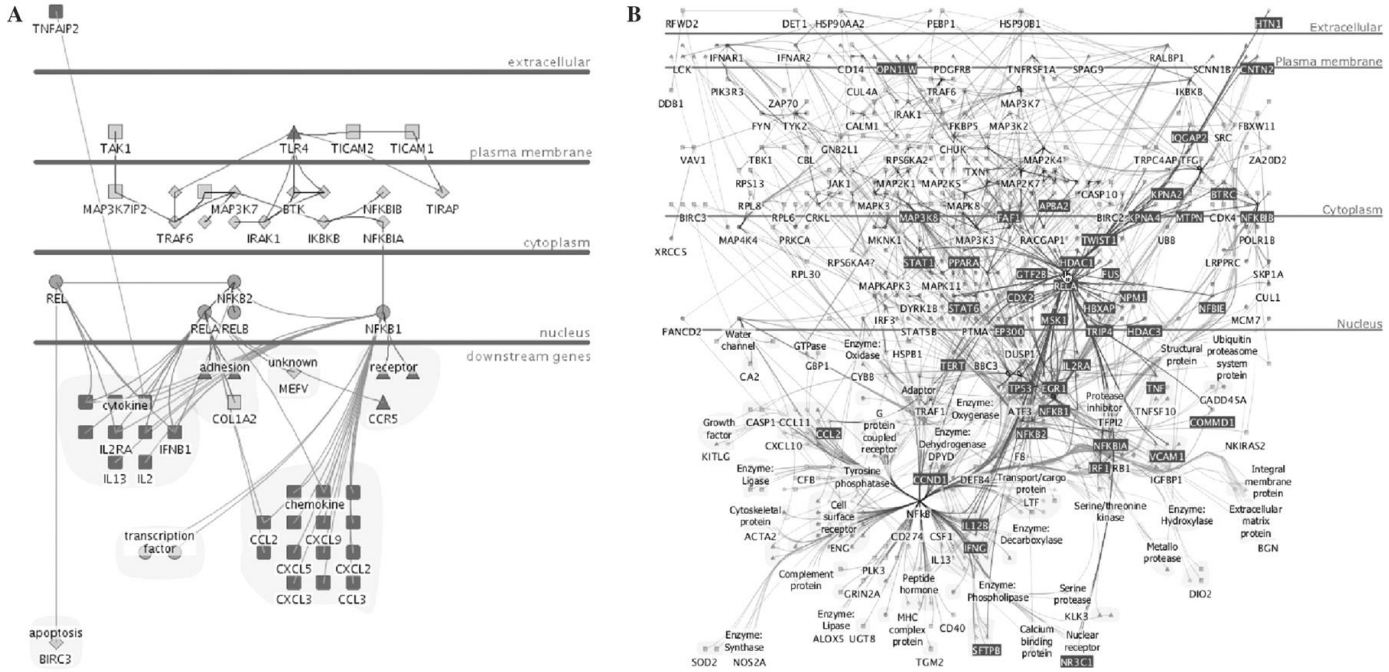
Introduction

Why data vis (in general)?

Overview

Bring the computer in the loop

- Enable the exploration of large datasets
  - Temporal dimension



Cerebral: a Cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation.
Aaron Barsky Jennifer L. Gardy Robert E. W. Hancock Tamara Munzner. Bioinformatics, Volume 23, Issue 8, 2007, 1040–1042,

Bring the computer in the loop

- Computational capacity
  - Scalability: How to deal with increasingly data sizes?

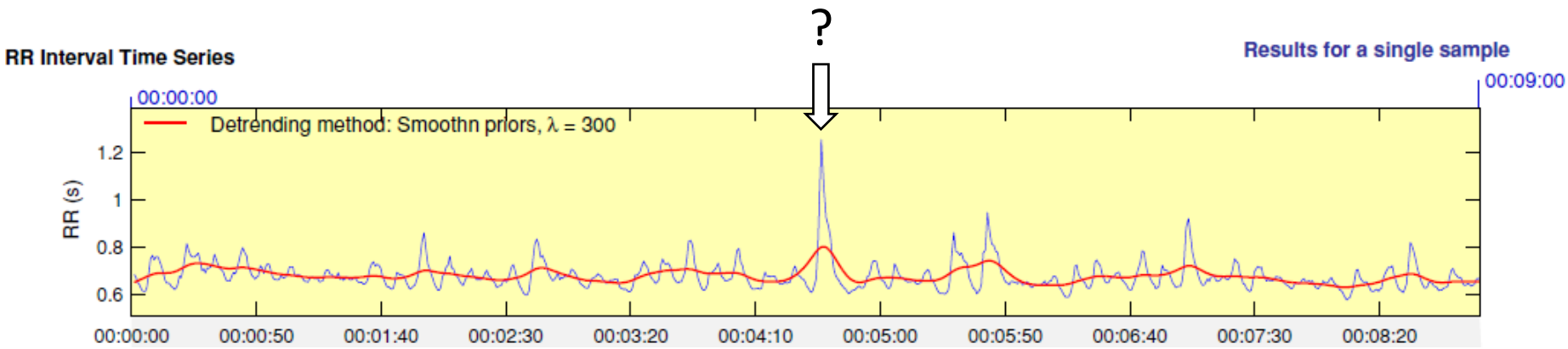
- Display capacity
  - Information density: limit of the displayed information

Bring the human in the loop

Digiscope - Univ. Paris Saclay
http://www.digiscope.fr

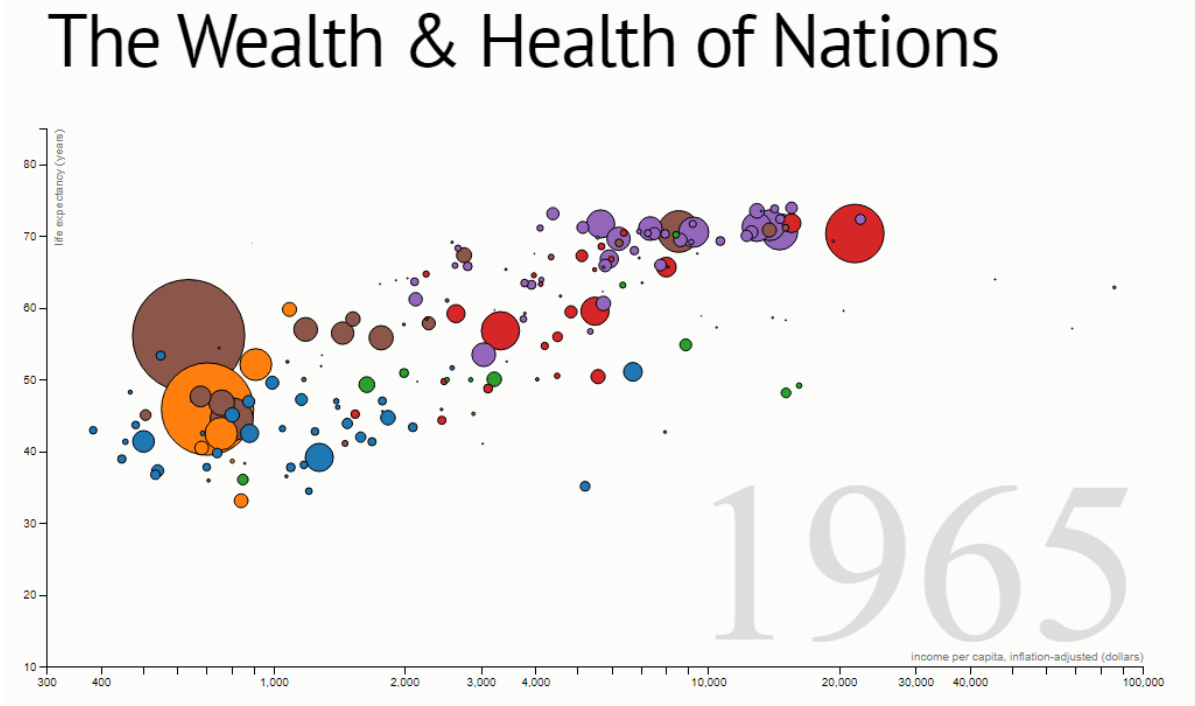
Bring the human in the loop

- The question is not known in advance
  - The analysis process is driven by the user
  - Take advantage of the ability of humans to find patterns
- Augment the human decision process
  - Improve decision making



Provide external representations

- Offload internal cognition and memory usage to the perceptual system
  - Data organized by spatial location
  - Support perceptual inference
  - Select just the relevant information

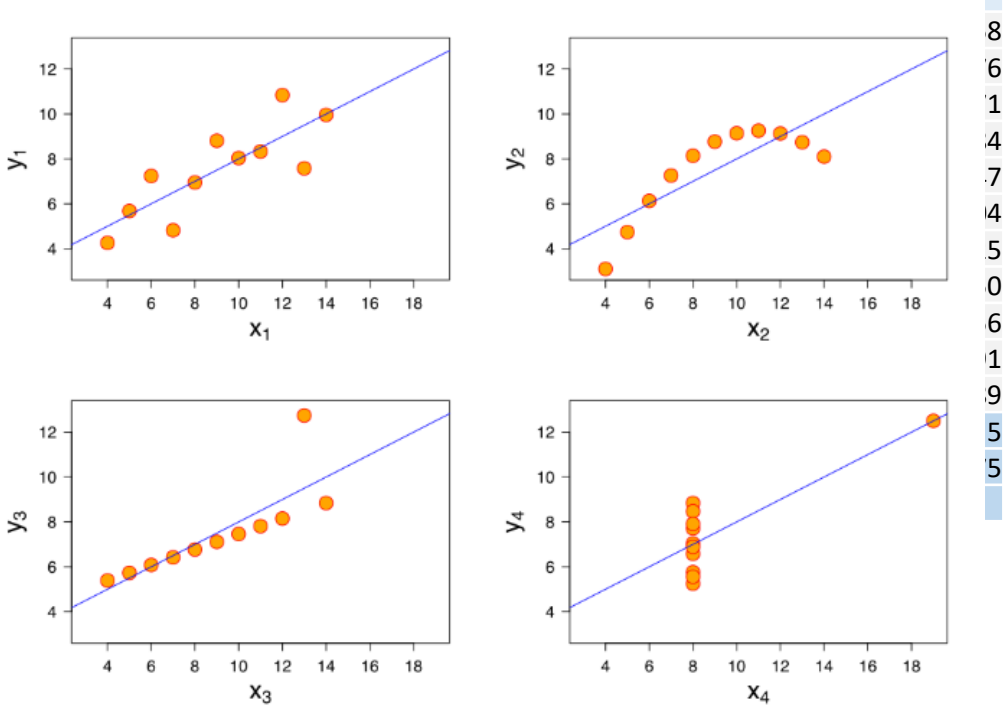


https://bost.ocks.org/mike/nations/

Show the data in detail

- The dataset structure in detail is better just a summary
  - Explore the data to find patterns

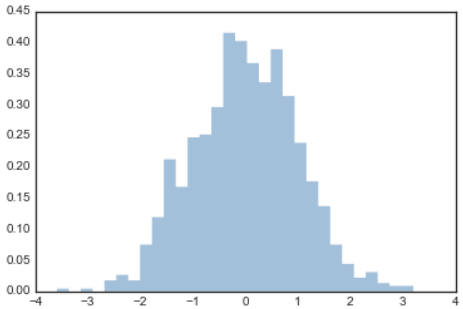
	I		II		III		IV	
	x	y	x	y	x	y	x	y
	10,00							8
	8,00							6
	13,00							1
	9,00							4
	11,00							7
	14,00							4
	6,00							5
	4,00							0
	12,00	1						6
	7,00							1
	5,00							9
Mean	9,0							5
Variance	10,0							5
Linear Regression	y=3 + 0.5x							

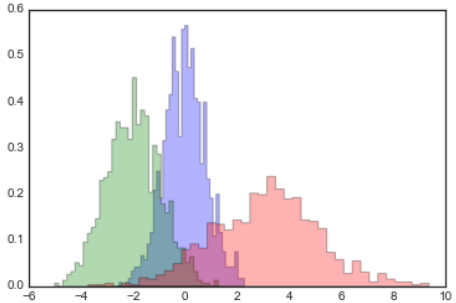


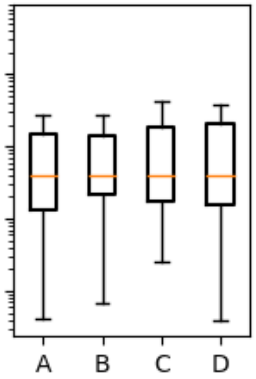
[Anscombe's Quartet]

Show the data in detail

- The dataset structure in detail is better just a summary
  - Explore the data to find patterns



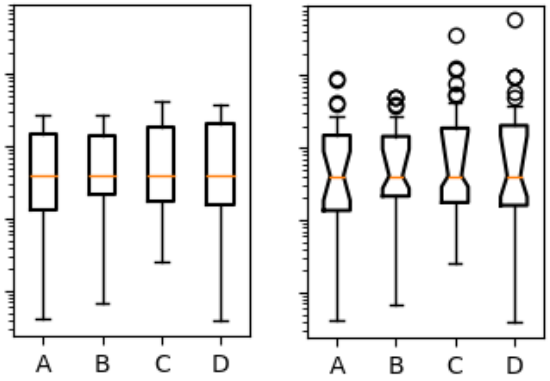


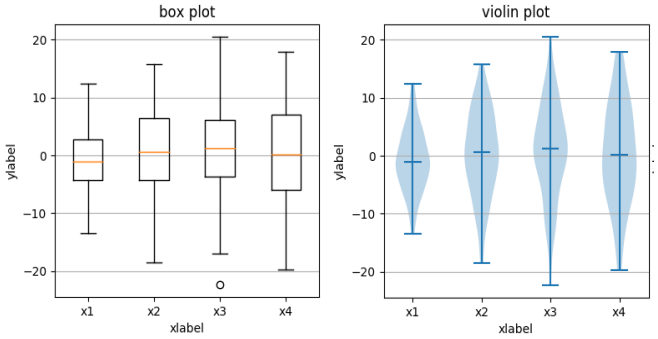


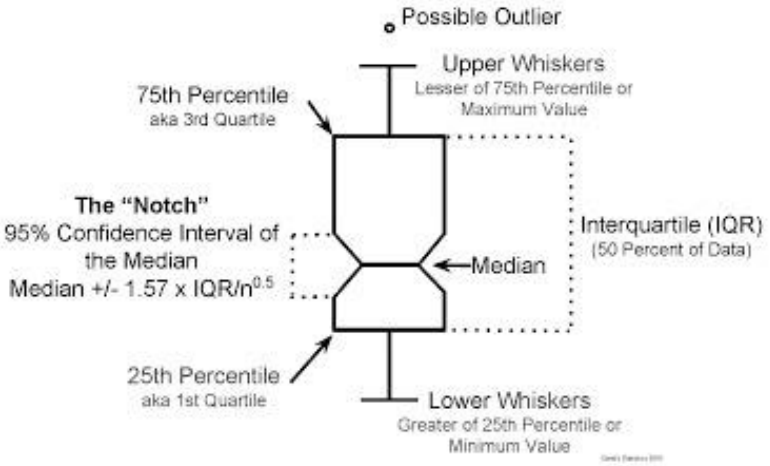
https://matplotlib.org/gallery.html

Show the data in detail

- The dataset structure in detail is better just a summary
  - Explore the data to find patterns





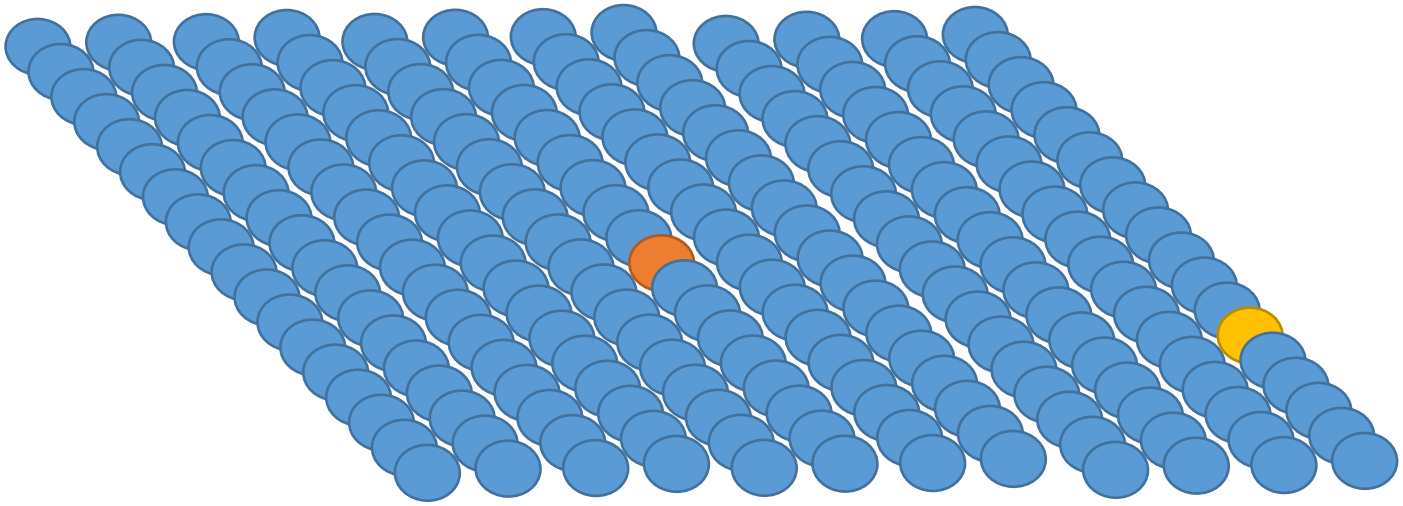


Chambers, John M., William S. Cleveland, Beat Kleiner, and Paul A. Tukey. "**Comparing Data Distributions**." In Graphical Methods for Data Analysis, 62. Belmont, California: Wadsworth International Group;, 1983. ISBN 0-87150-413-8 International ISBN 0-534-98052-X

https://matplotlib.org/gallery.html

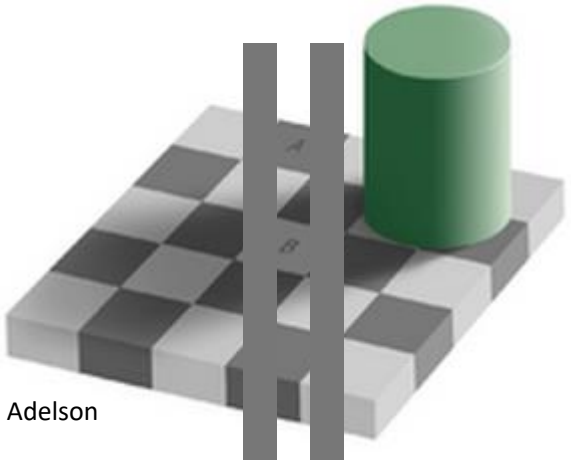
Depend on vision…

- The Visual system is a high-bandwidth channel of information
  - A significant amount of visual information is done in parallel at the preconscious level

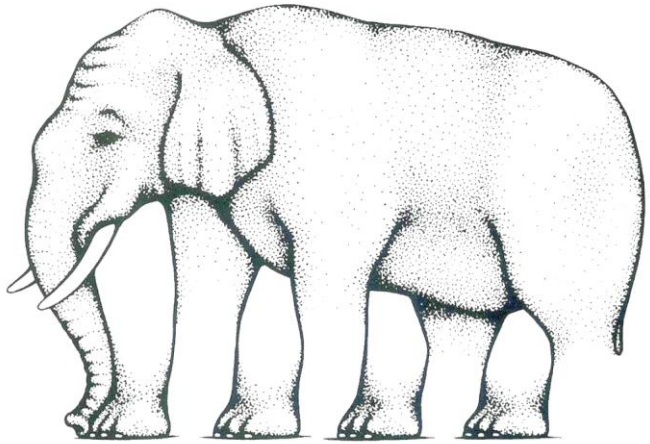


… but be careful

- Need to know its strengths and limitations



Edward H. Adelson



… but be careful

- Need to know its strengths and limitations
- Human perceptual and cognitive capacity limits?

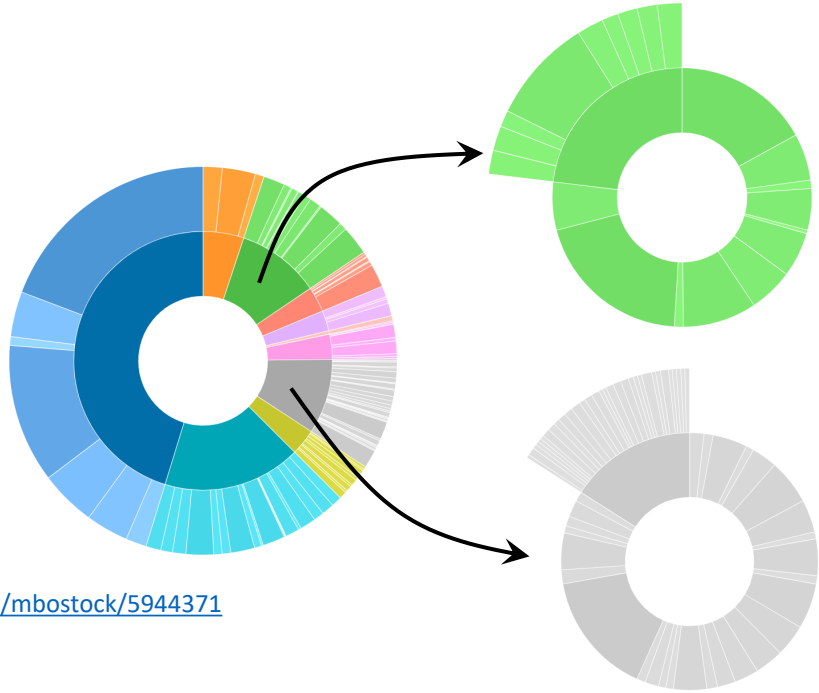
https://www.youtube.com/watch?v=uO8wpm9HSB0

… but be careful

- Need to know its strengths and limitations
- Human perceptual and cognitive capacity limits?
  - Change blindness: the phenomenon where even very large changes are not noticed if we are attending to something else in our view

Interactivity

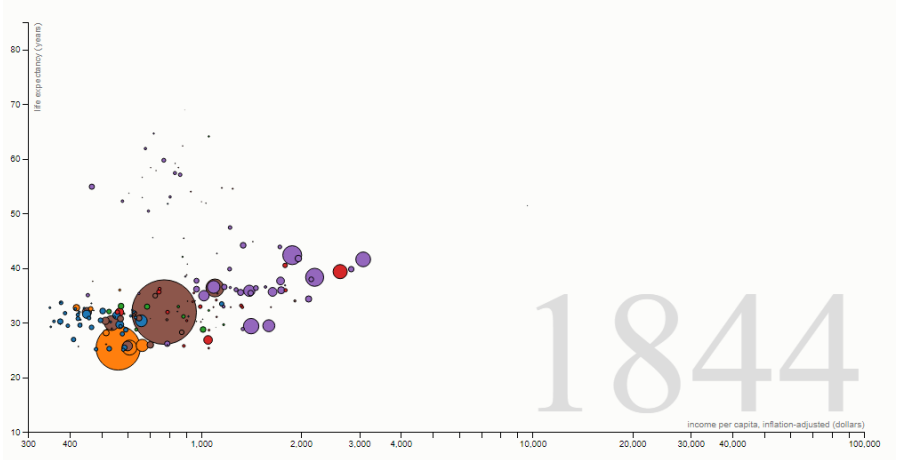
- For big datasets, the user nor the display can enable the visualization of the entire dataset at once

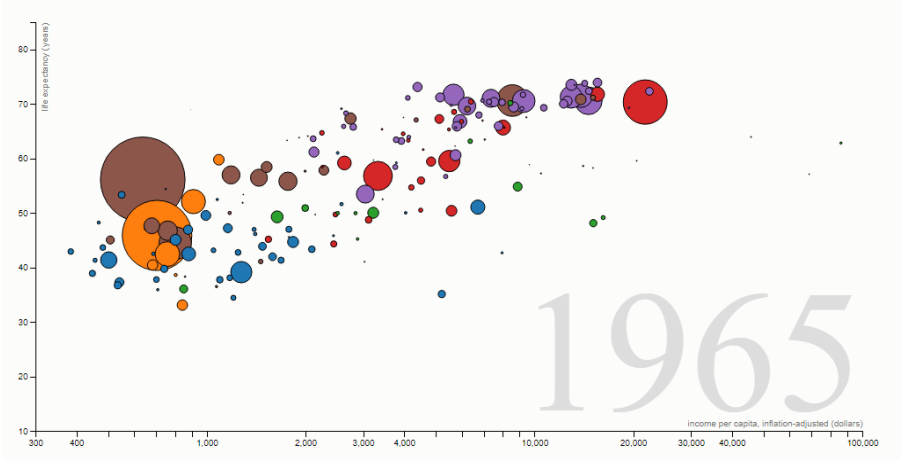


https://bl.ocks.org/mbostock/5944371

Interactivity

- A static view only allows for a single view of a limited number of dimensions

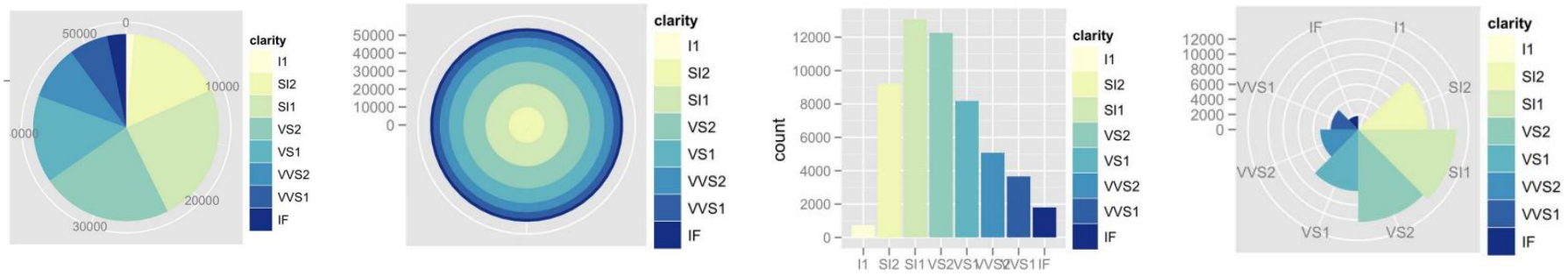




https://bost.ocks.org/mike/nations/

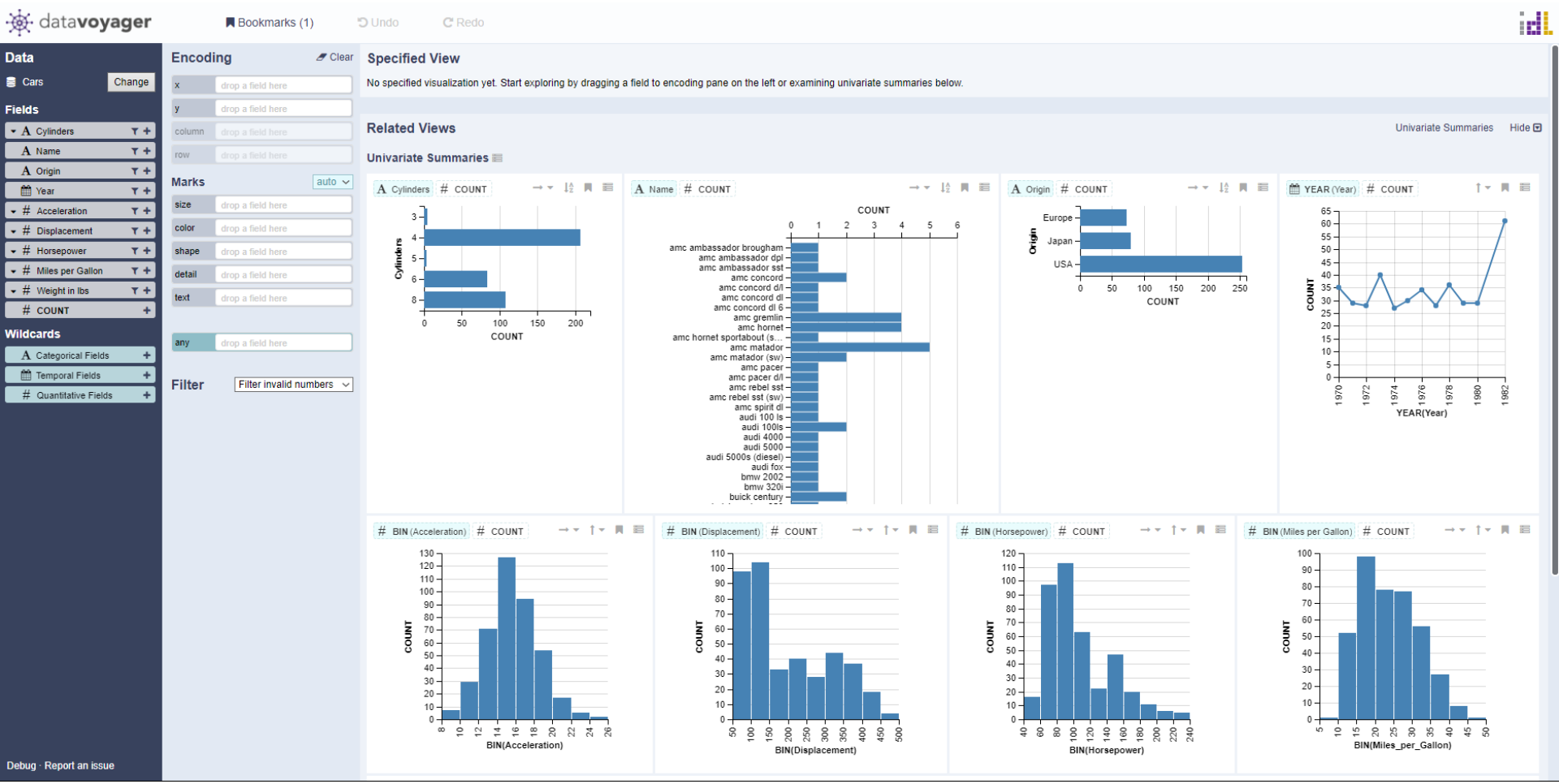
Interactivity

- A static view only allows for a single view of a limited number of dimensions
- The same dataset can be sorted using a number of different criteria



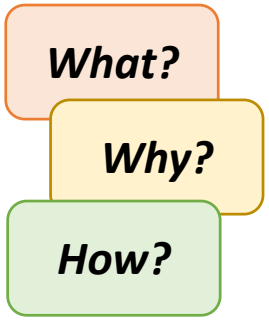
Hadley Wickham **A Layered Grammar of Graphics**. Journal of Computational and Graphical Statistics Vol. 19 , Iss. 1,2010

Interactivity

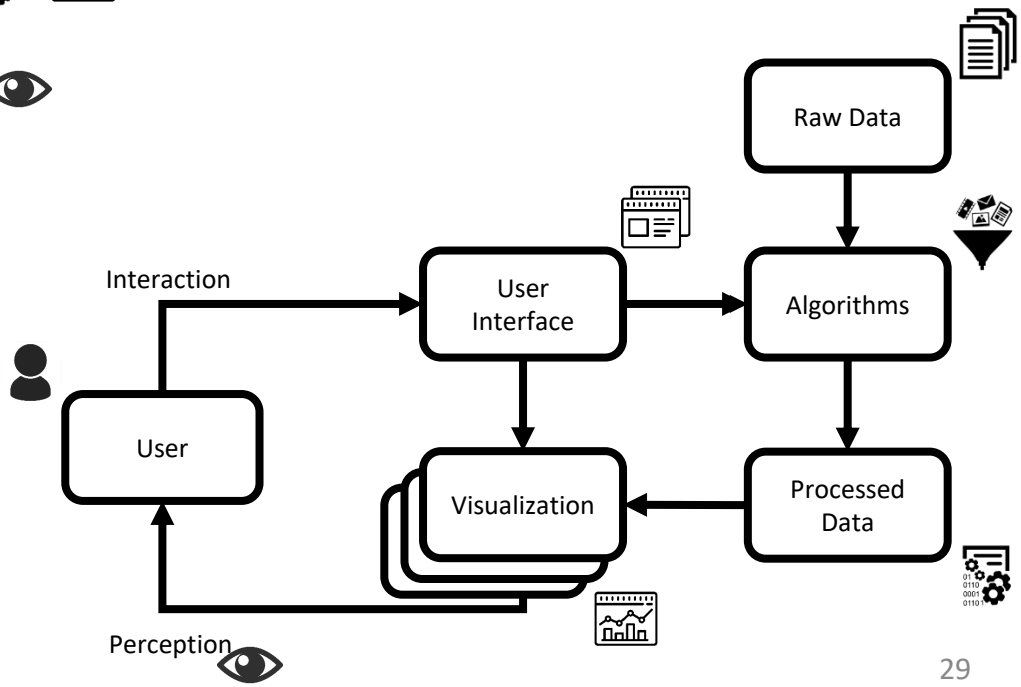


https://vega.github.io/voyager2/

Wrap Up

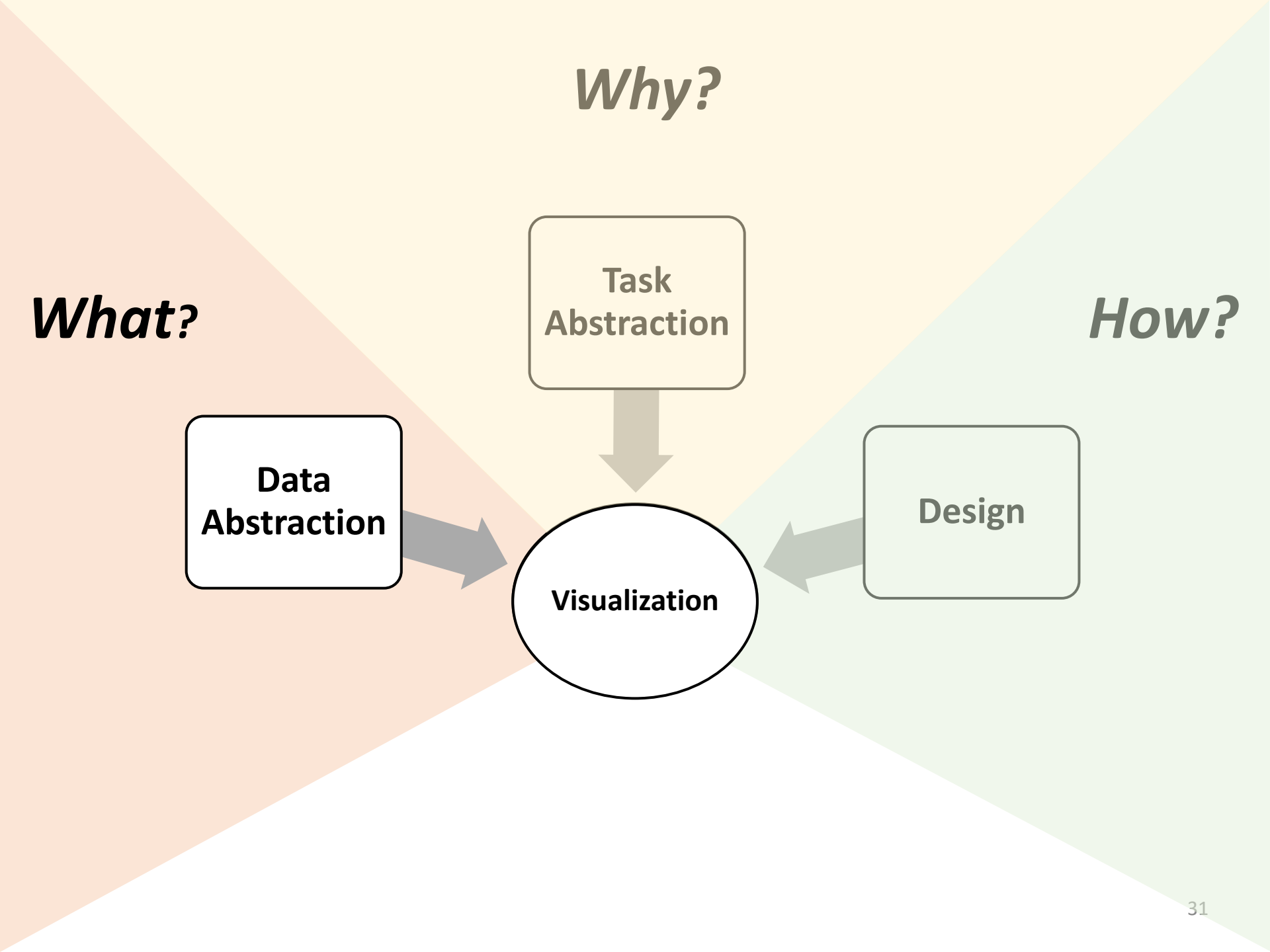


- Bring the computer in the loop
- Bring the human in the loop
- Provide external representations
- Show data in detail
- Depend on vision
- Interactivity



Visualization Design

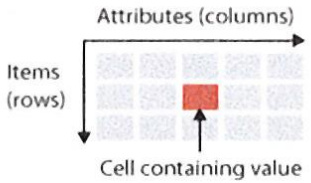
Methodology to design visualizations


Why?
What?
How?
Task Abstraction
Data Abstraction
Design
Visualization

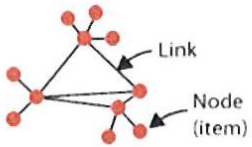
What?

Data and Dataset Types

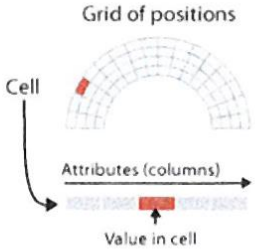
→ Tables



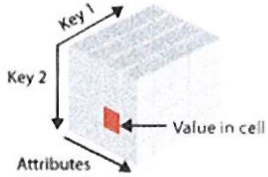
→ Networks



→ Fields (Continuous)



→ *Multidimensional Table*



→ *Trees*

→ Items
→ Attributes
→ Links
→ Positions
→ Grids

→ Geometry (Spatial)

→ **Dataset Availability**

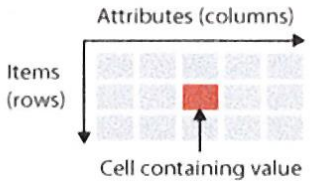
→ Static

→ Dynamic

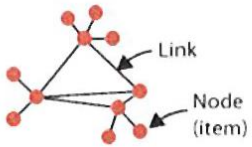
Data and Dataset Types

What?

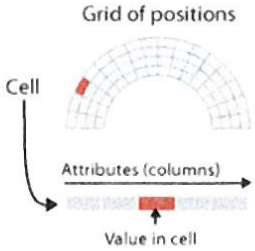
→ Tables



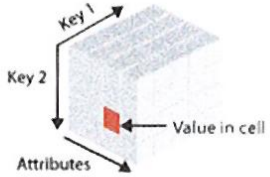
→ Networks



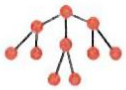
→ Fields (Continuous)



→ *Multidimensional Table*



→ *Trees*



→ Items
→ **Attributes**
→ Links
→ Positions
→ Grids

→ Geometry (Spatial)

→ **Dataset Availability**

→ Static

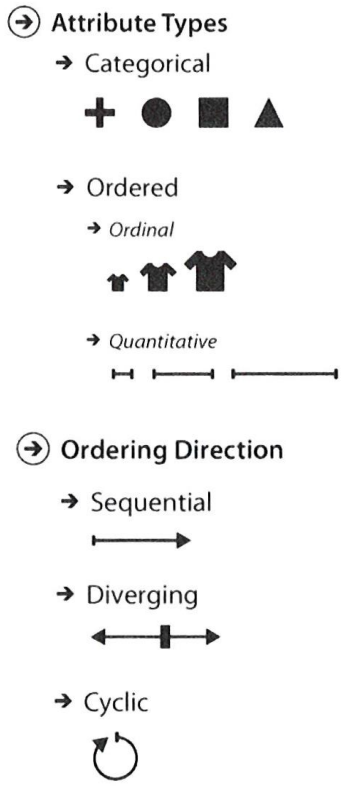
→ Dynamic

Attributes

What?

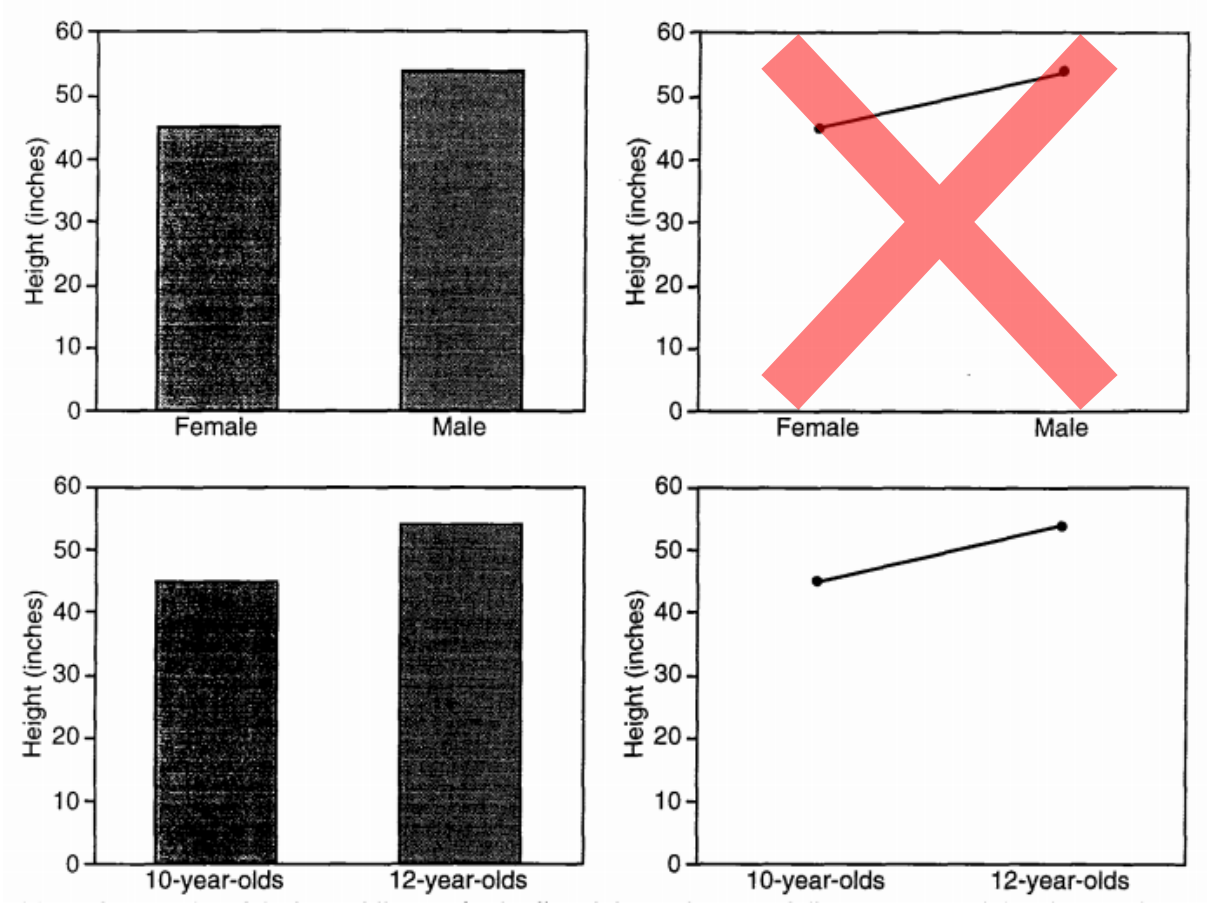
An attribute is some specific property that can be measured, observed or logged

- Categorical
  - Non-numerical, limited number of values, usually fixed
  - Non-ordered, unless explicitly stated
- Ordinal
  - Non-numerical, limited number of values, usually fixed
  - Ordered
- Numerical
  - Quantitative, Infinite number of values
  - Ordered



Bar Charts vs Line Charts

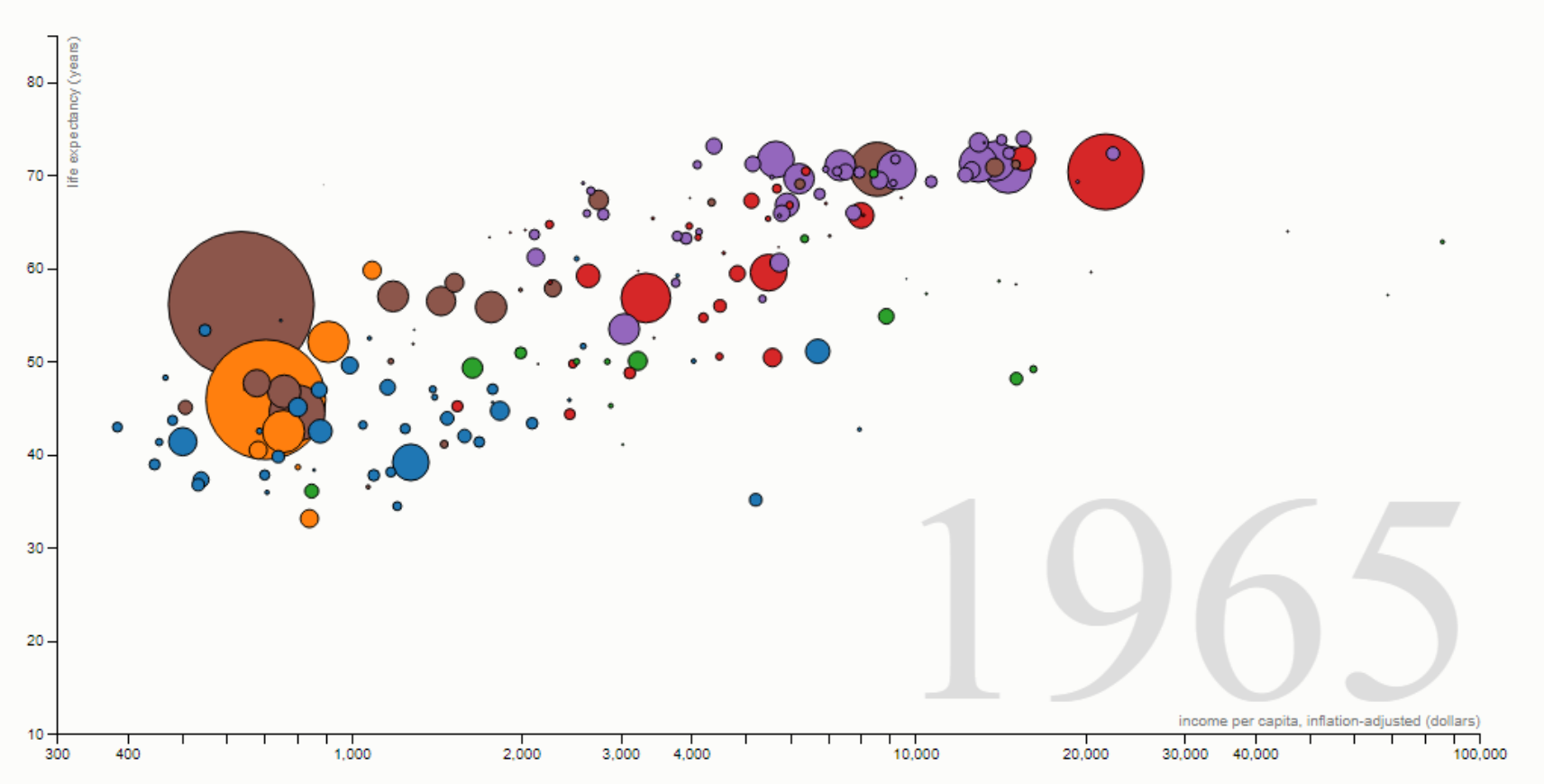
What?



Zacks, Jeff, and Barbara Tversky. **"Bars and lines: A study of graphic communication."** *Memory and Cognition* 27 (1999): 1073-1079.

Scaterplot

What?

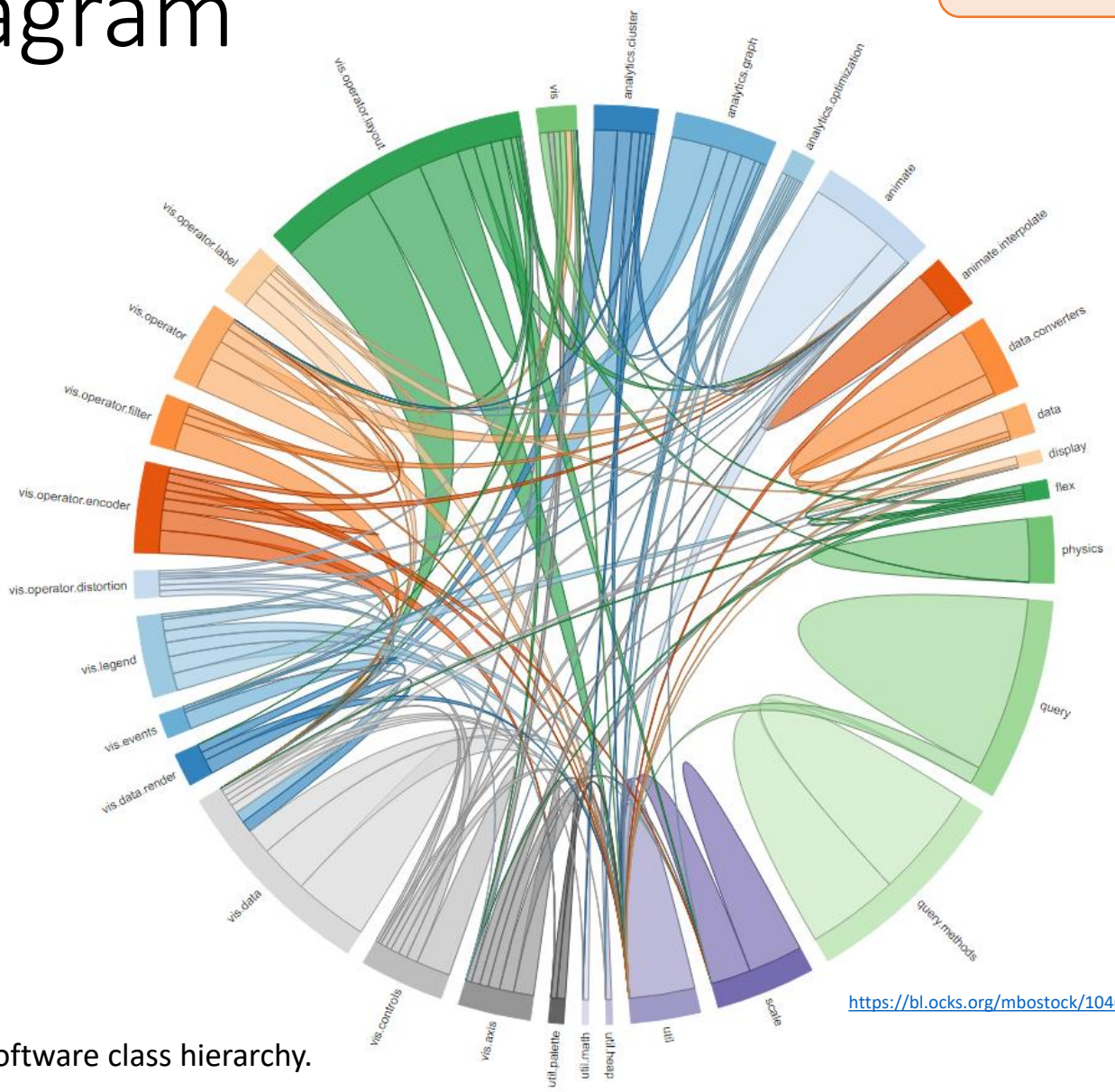


https://bost.ocks.org/mike/nations/

Multi dimensional table: N countries * T time
Attributes: 3 Qualitative

Chord Diagram

What?



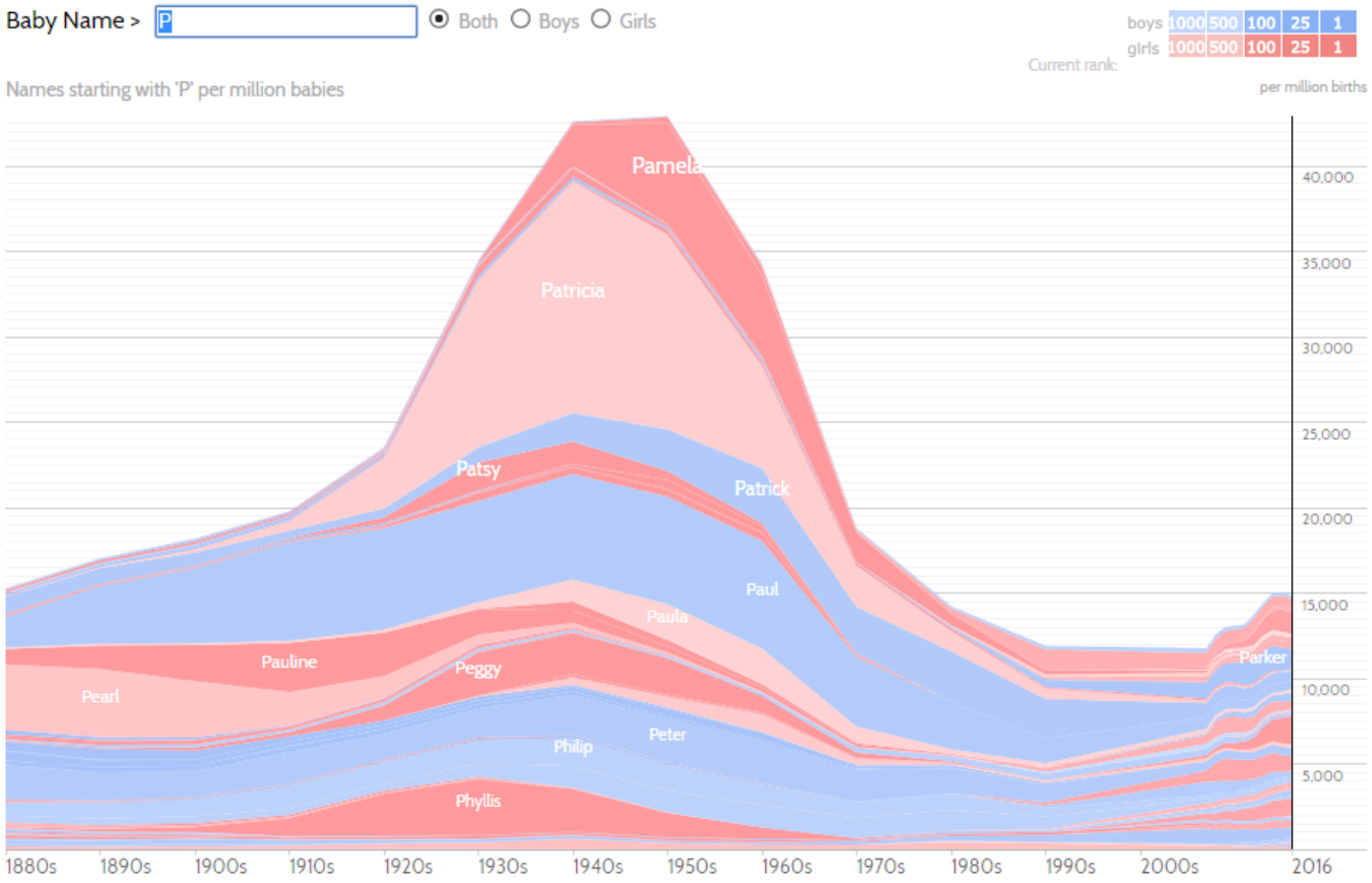
Network

Dependences between classes in a software class hierarchy.

https://bl.ocks.org/mbostock/1046712

Time Varying Data

What?



http://www.babynamewizard.com/voyager

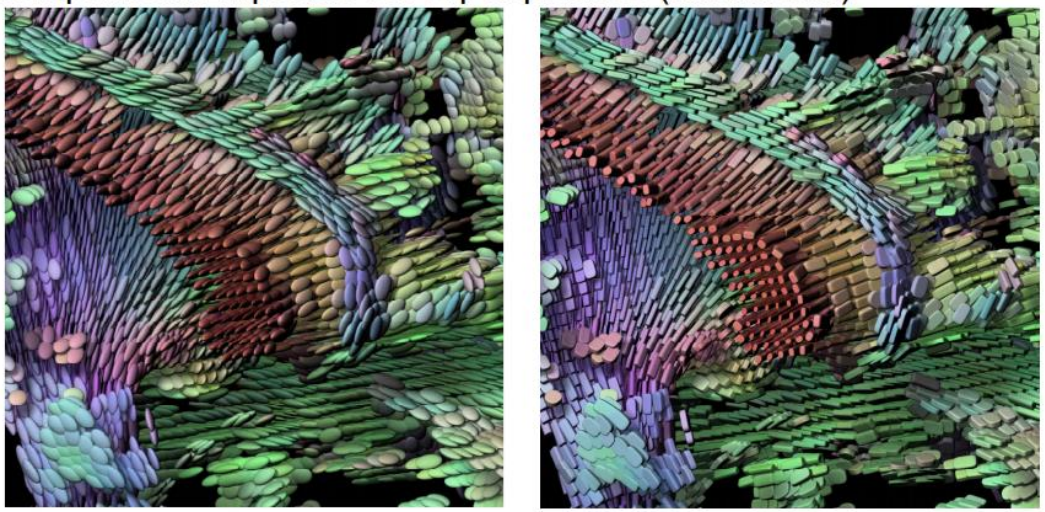
Table: T time

Attributes: 1 Categorical, 1 Quantitative

Tensor Visualization

What?

Comparison: Ellipsoids vs. superquadrics (Kindlmann)



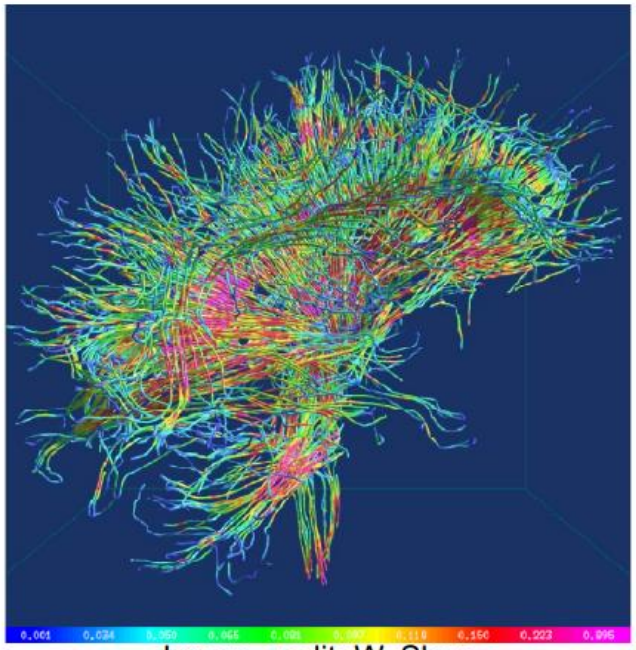
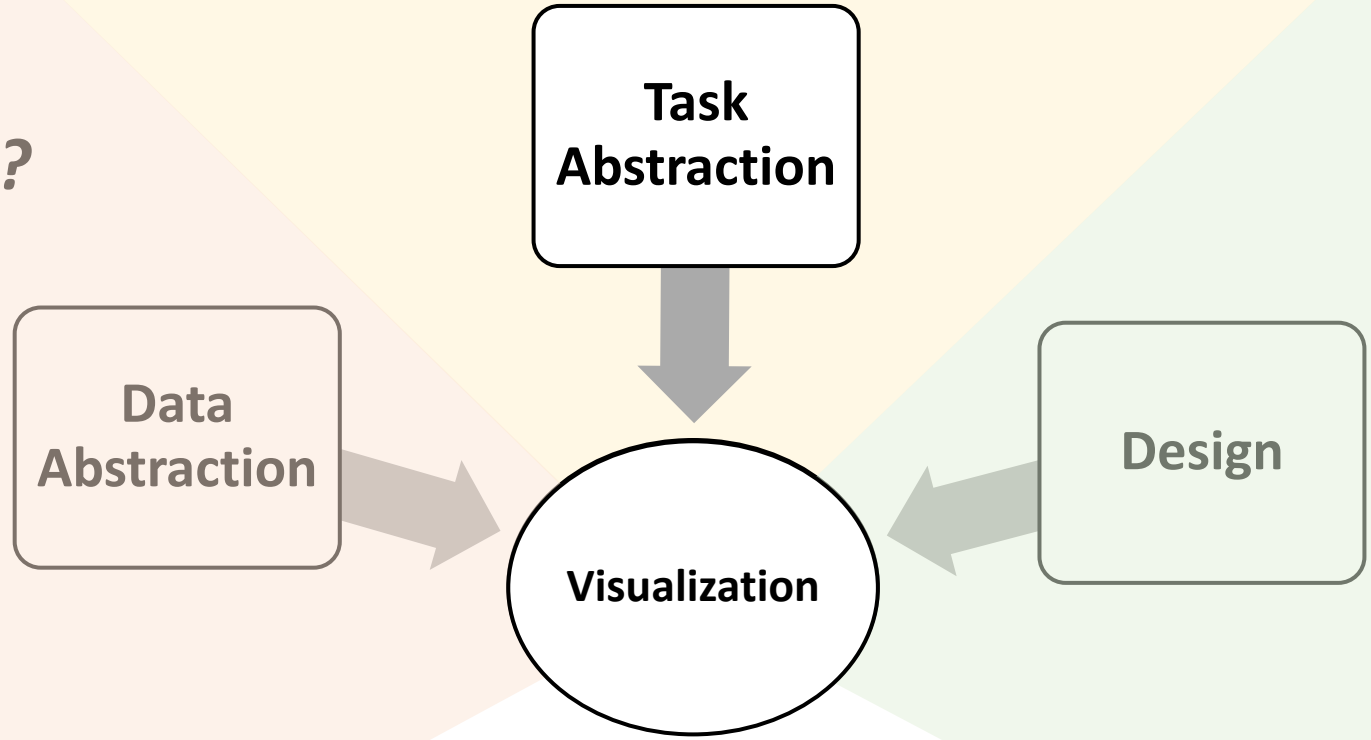


Image credit: W. Shen

Why?

What?

How?



Task Abstraction

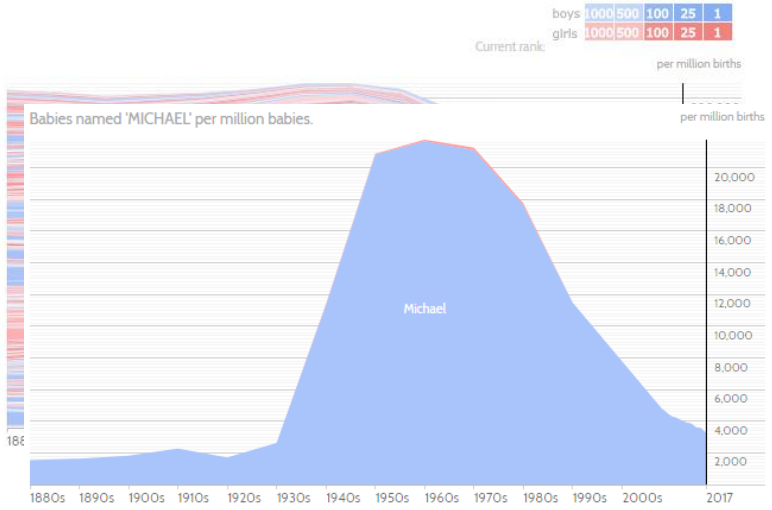
Why?

- Reframe tasks from domain-specific form into abstract form
  - Ease the design process using references
- The focus on tasks enables the improvement of them
  - Making more effective, user-friendly…
- Tasks can be classified in:
  - Actions – Verb - High-level task
  - Targets – Noun - Precise goal

Action Classification

Why?

- Analyze
  - Consume

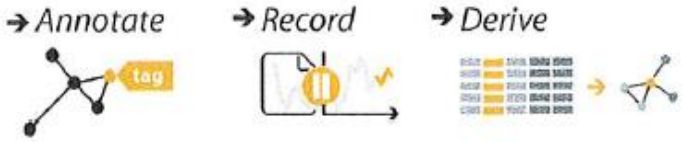


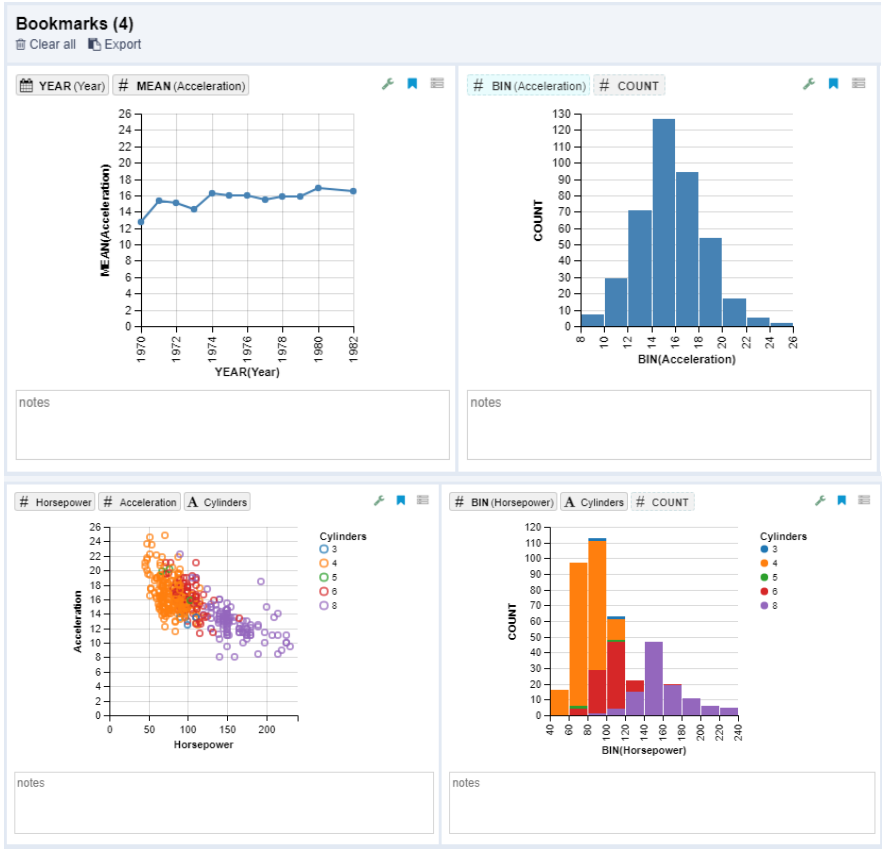
Action Classification

Why?

- Analyze
  - Consume

  - Produce



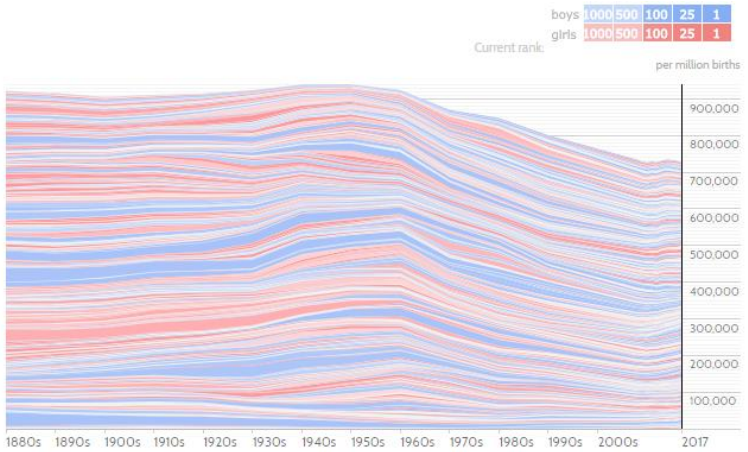


Action Classification

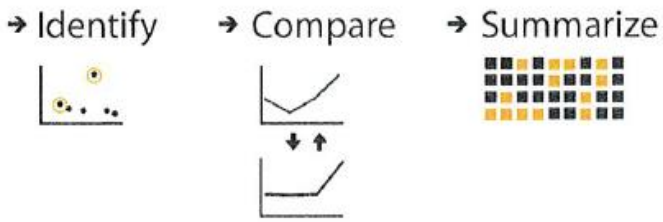
Why?

- Search

	Target known	Target unknown
Location known	Lookup	Browse
Location unknown	Locate	Explore



- Query

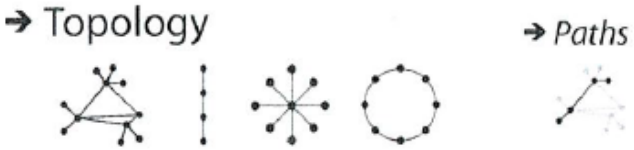


Target Classification

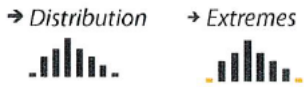
Why?

- All Data

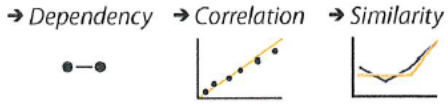
- Network Data



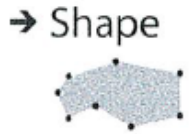
- Attributes
  - One



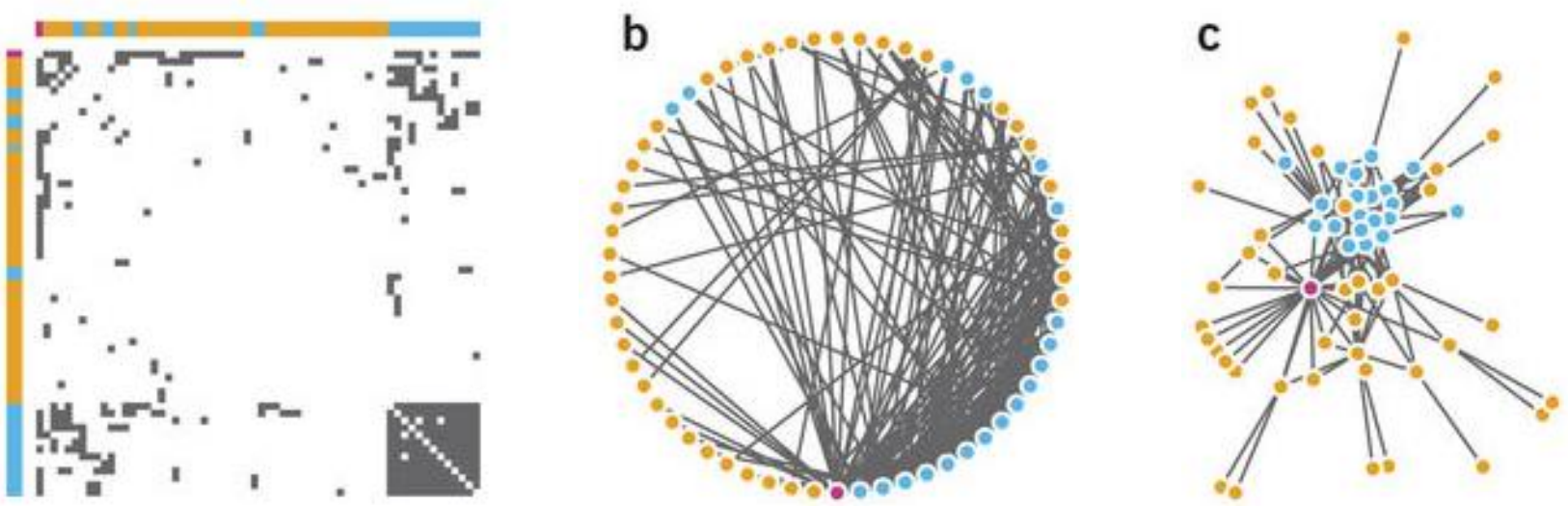
  - Many



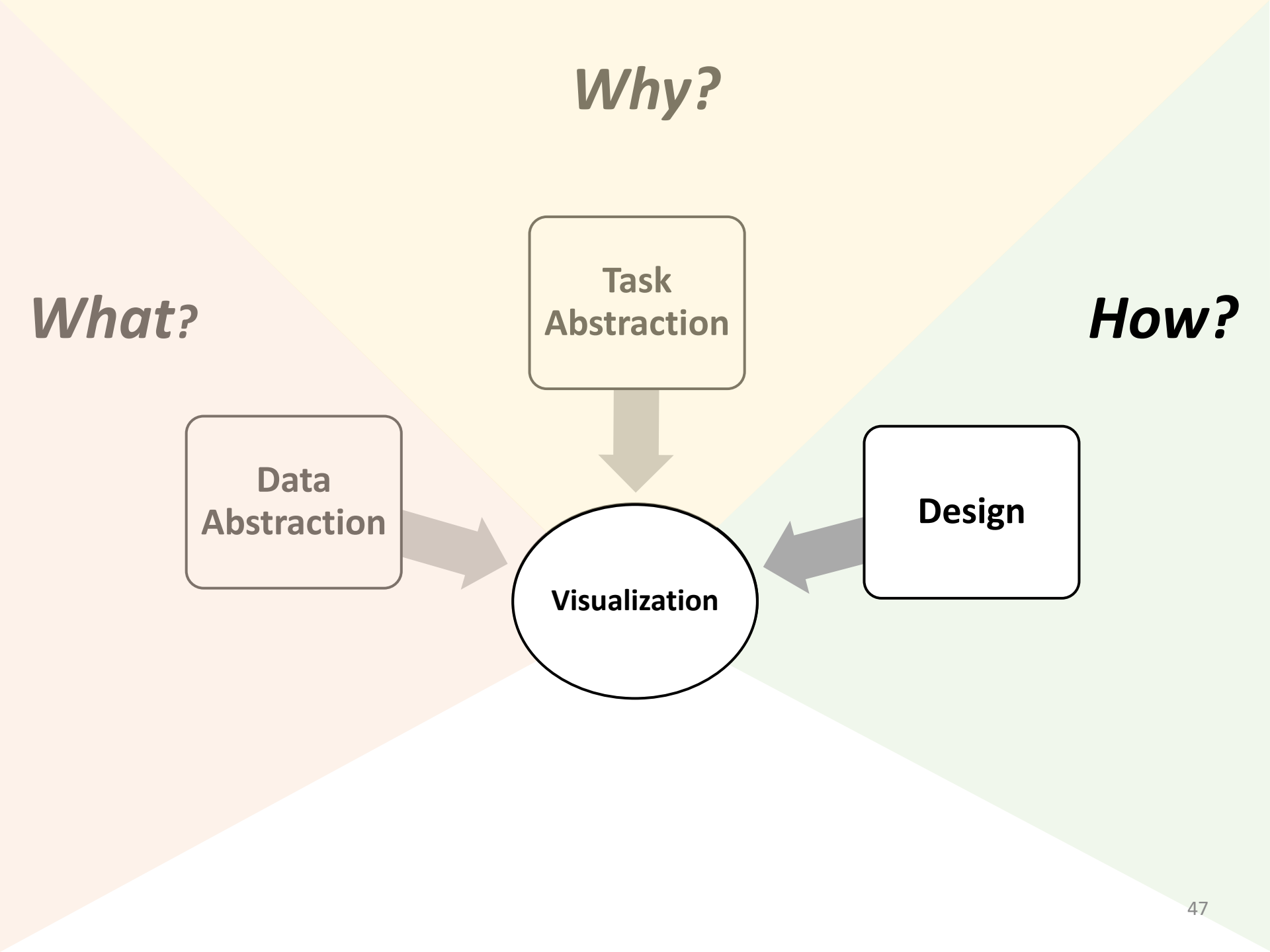
- Spatial Data



Network Visualization



Gehlenborg, Nils, and Bang Wong. **Points of view: Networks** *Nature methods* 9.2 (2012): 115-115.


Why?
What?
How?
Task Abstraction
Data Abstraction
Design
Visualization

How to…

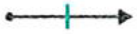
How?

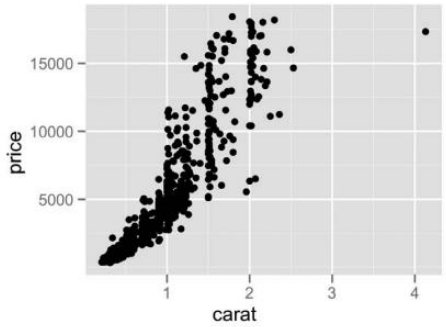
- … **arrange** the data in the view spatially?

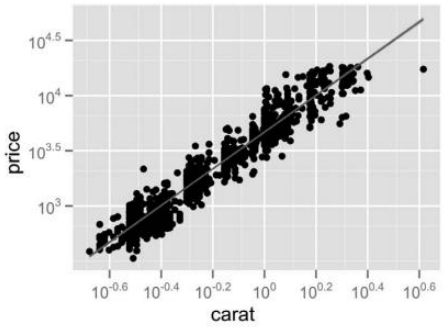
Arrange / Encode

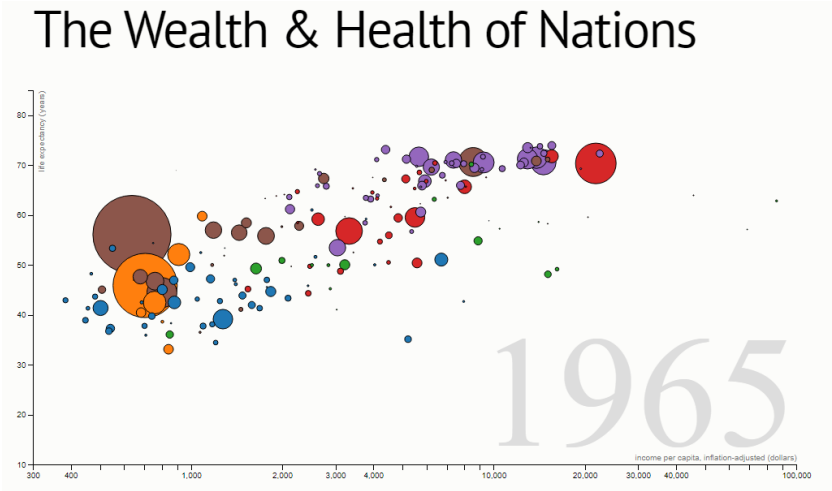
How?

➔ Express





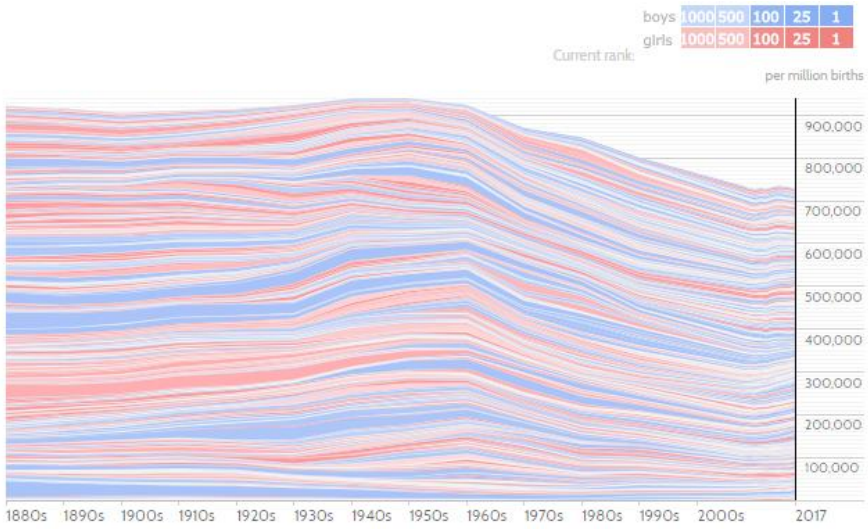




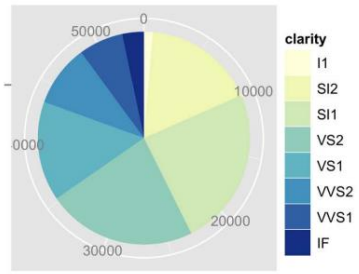
Arrange / Encode

How?

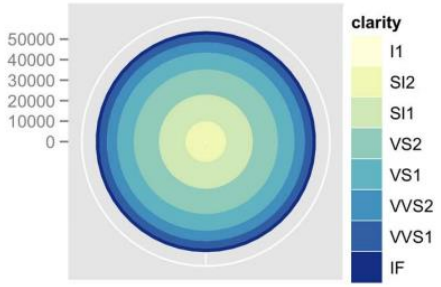
➔ Separate



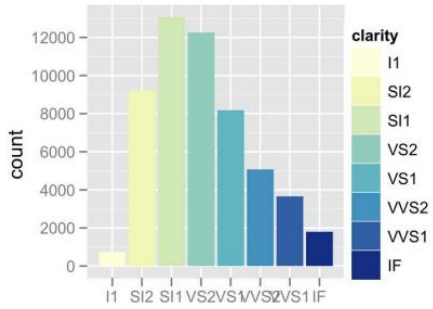
Streamgraphs



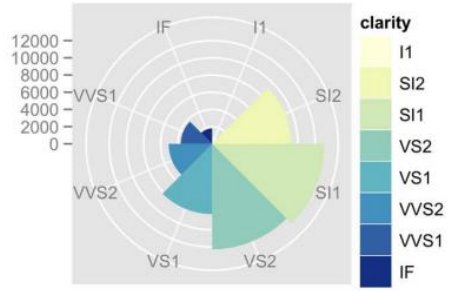
Pie Chart



Bullseye chart



Bar Plot



Coxcomb plot

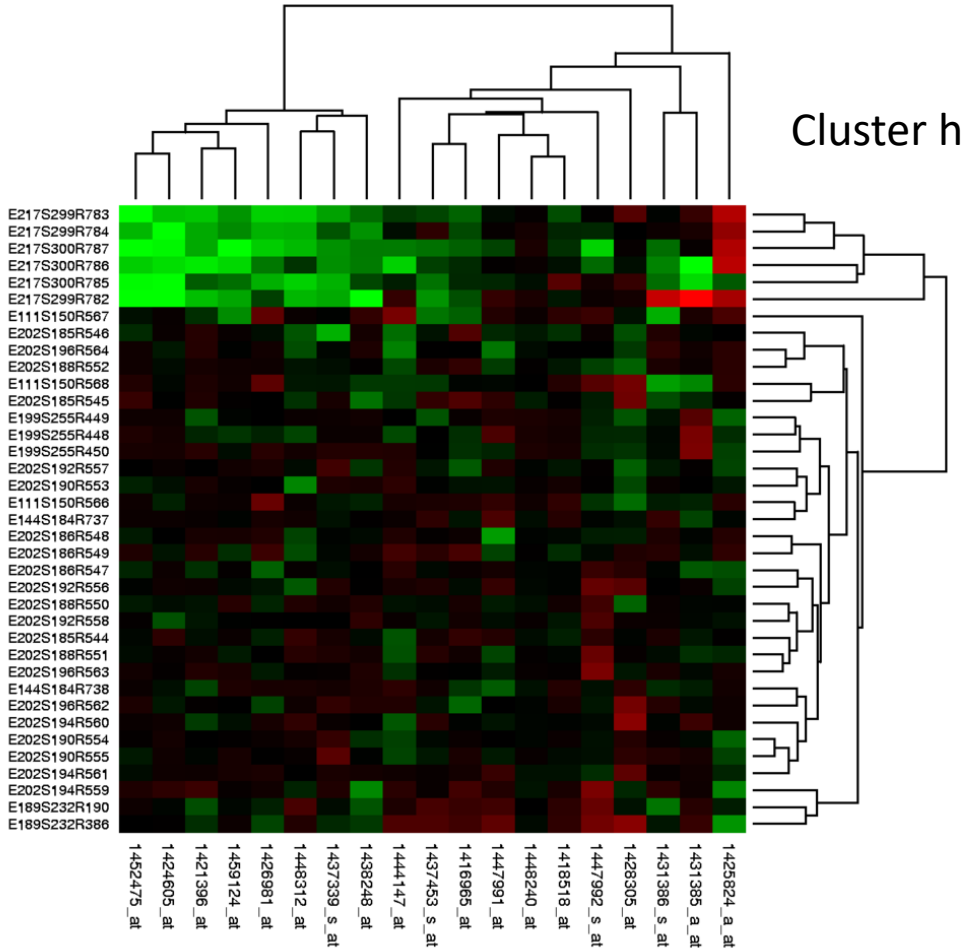
Hadley Wickham **A Layered Grammar of Graphics**. Journal of Computational and Graphical Statistics Vol. 19 , Iss. 1,2010

Arrange / Encode

How?

- Order

- Align



Cluster heatmap

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

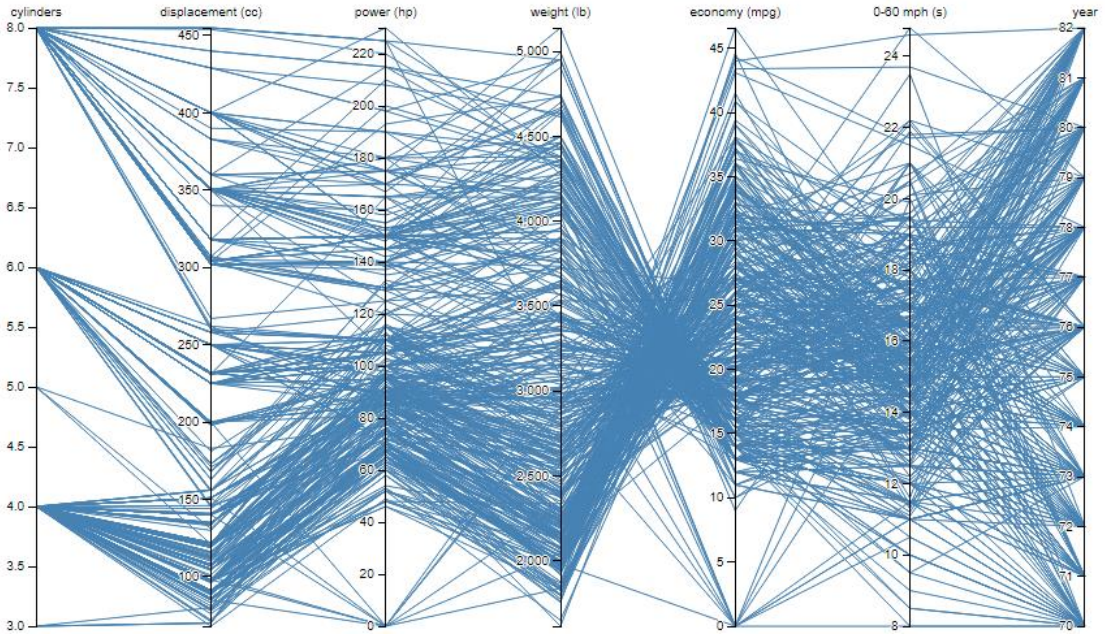
Arrange / Encode

How?

- Order

- Align

Parallel Coordinates



https://bl.ocks.org/jasondavies/1341281

How?

How to…

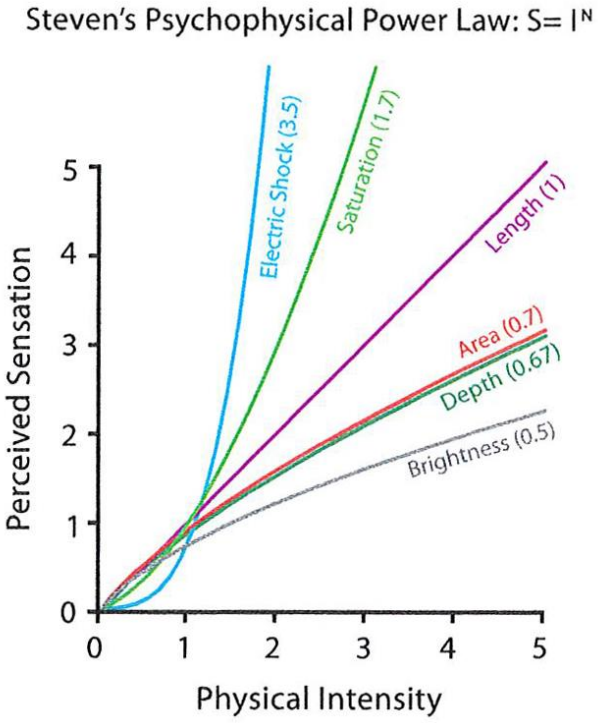
- … **arrange** the data in the view spatially?
- … **map** the data with all non-spatial channels?

Display

How?

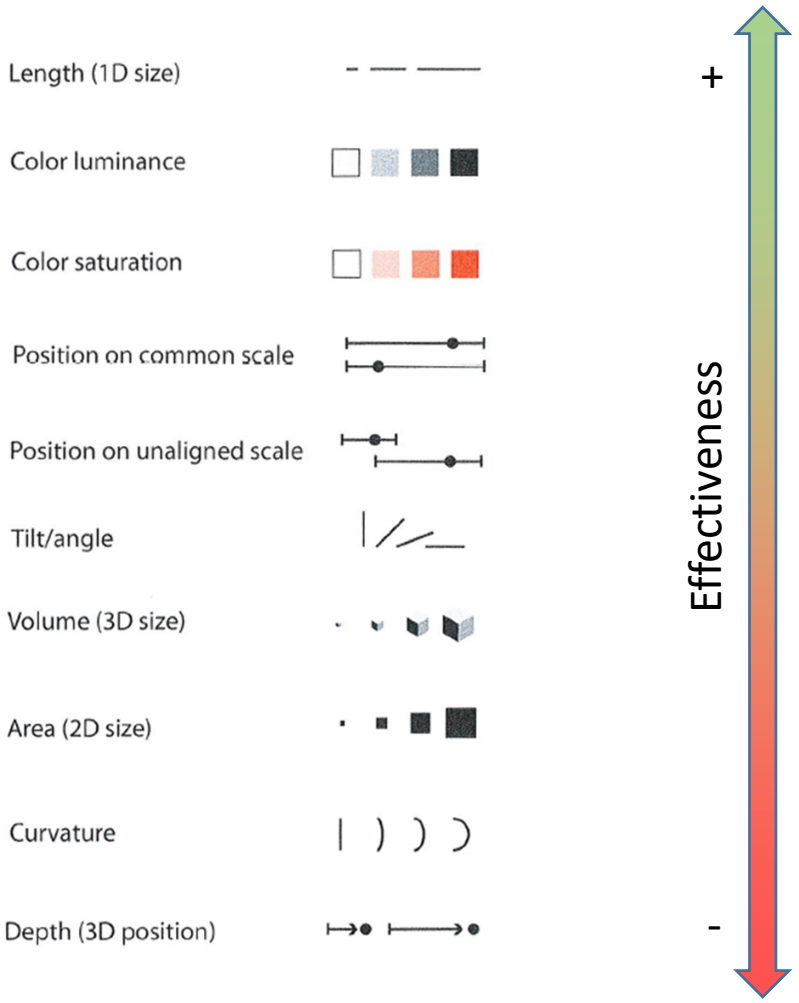
- What is the best way to visually encode the data ?
  - How the human visual system discriminate visual stimuli?

- Two major dimensions
  - Color
  - Shape



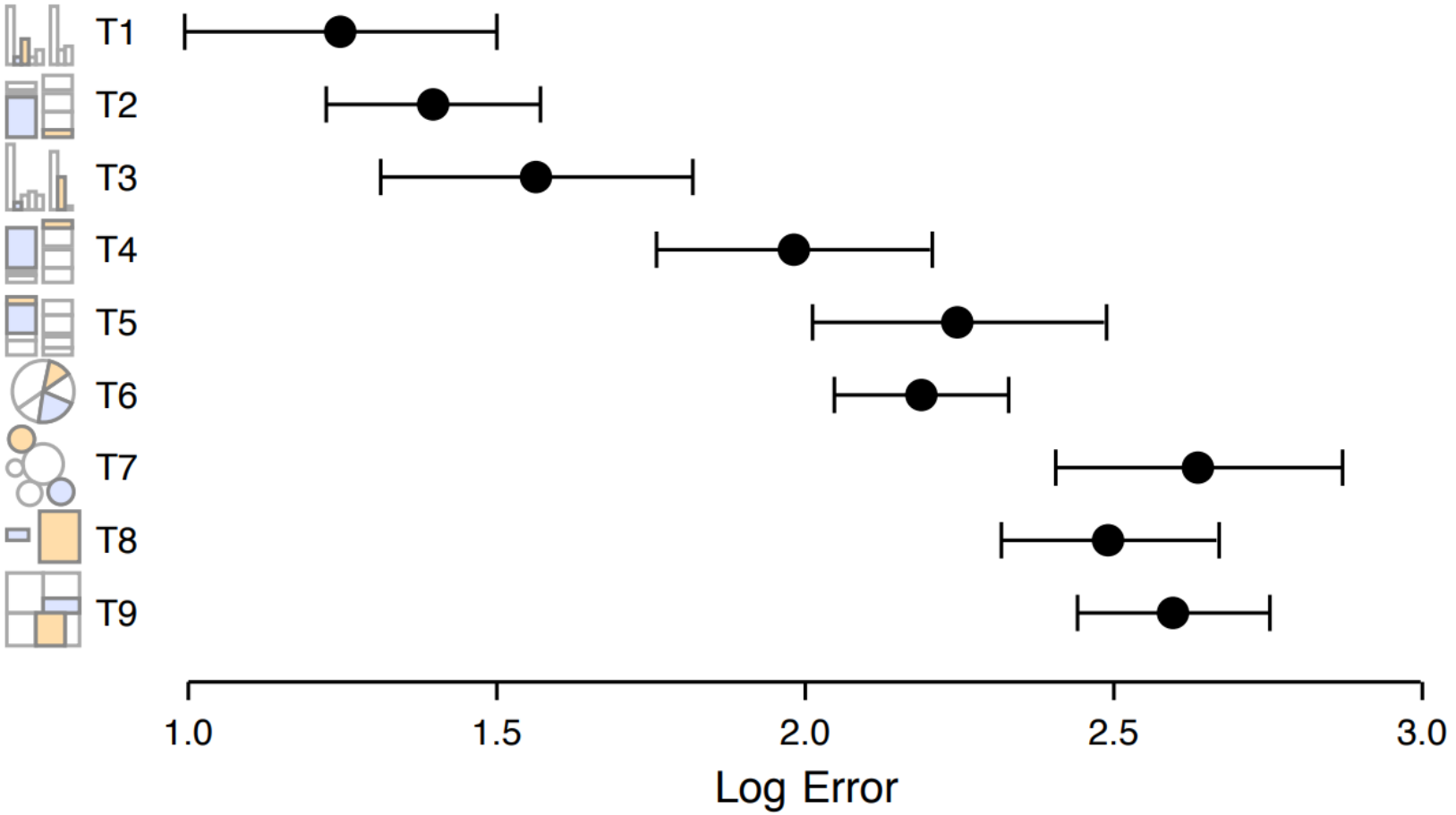
Visual Encoding

How?



Visual discrimination

How?



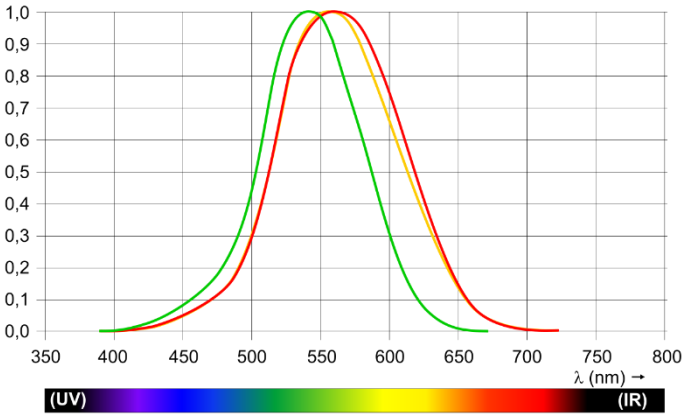
Crowdsourcing graphical perception: using mechanical turk to assess visualization design. Jeffrey Heer and Michael Bostock. 2010. ACM SIGCHI Conference on Human Factors in Computing Systems, p. 203-212

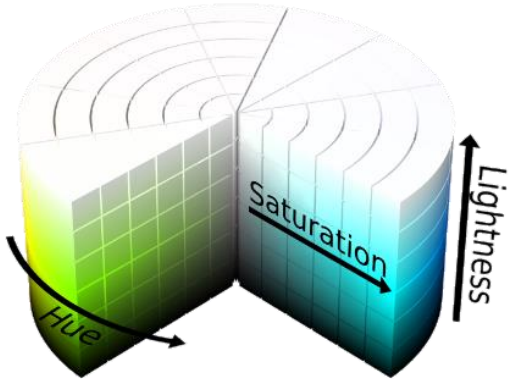
Color Encoding

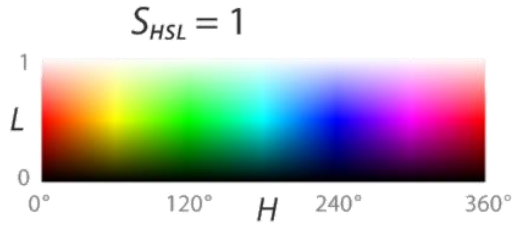
How?

- Color Components

- Color sensitivity



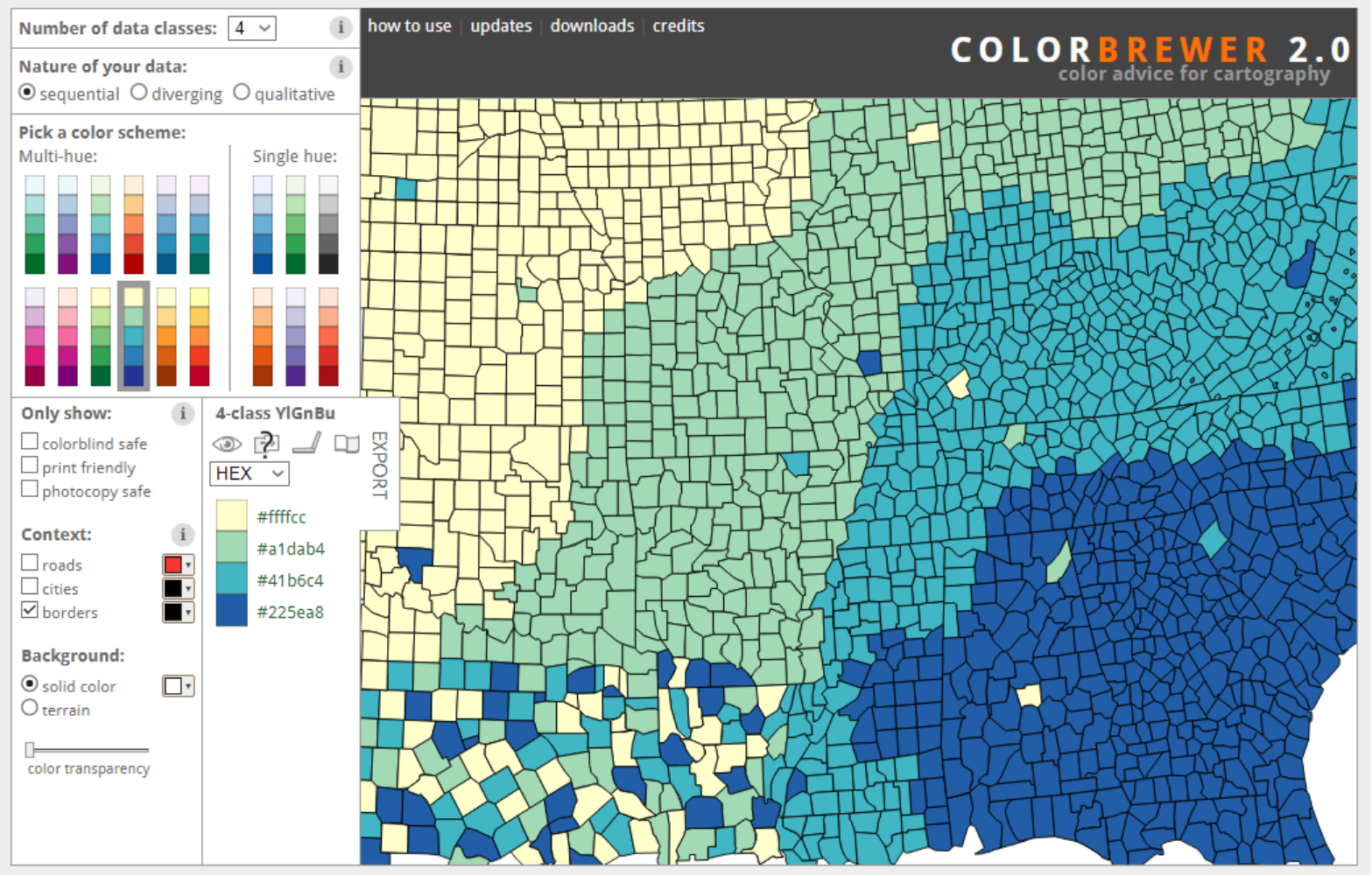




https://commons.wikimedia.org/wiki/File:LuminosityCurve2.svg
https://commons.wikimedia.org/wiki/File:Hsl-hsv_models_b.svg

Color Encoding

How?



http://colorbrewer2.org/

Color Use Guidelines for Data Representation, Brewer, C. A. 1999. Proceedings of the Section on Statistical Graphics, American Statistical Association, Alexandria VA. pp. 55-60.

How?

How to…

- … **arrange** the data in the view spatially?
- … **map** the data with all non-spatial channels?
- … **manipulate** the data in the view?

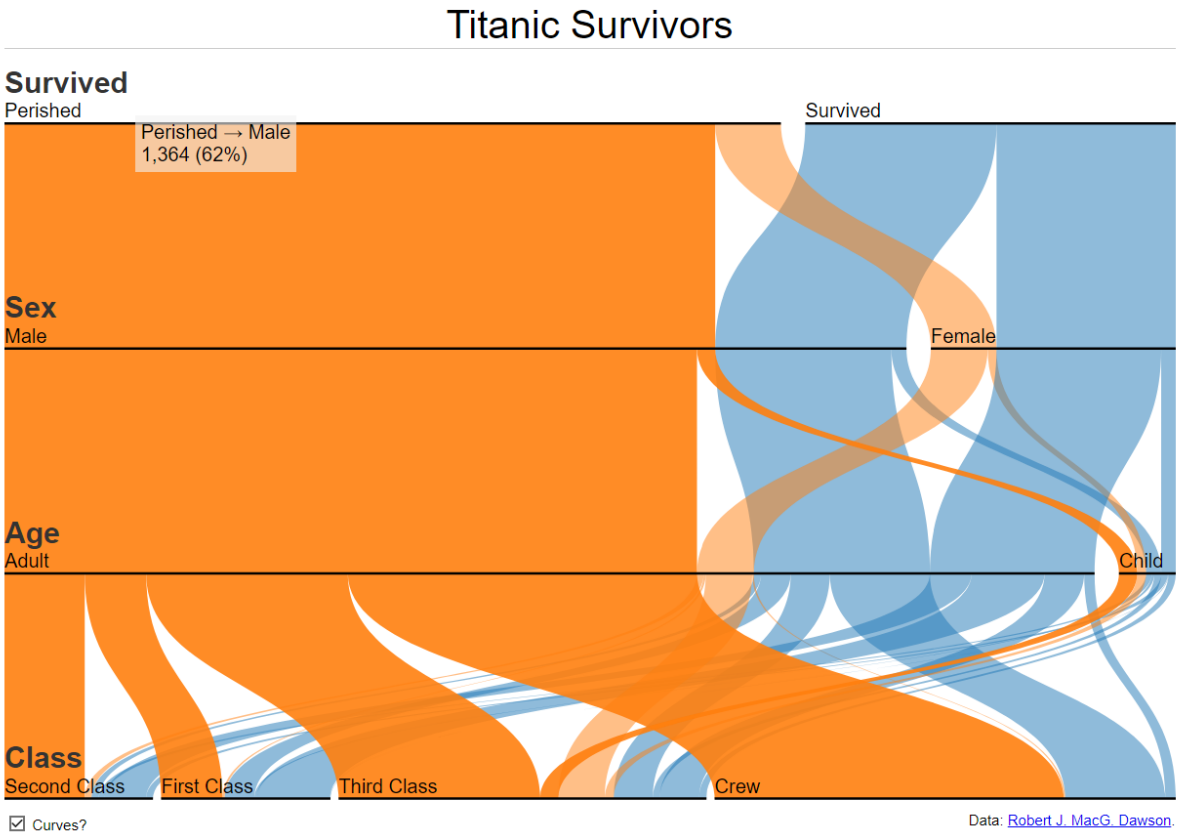
Manipulate

How?

Change

Select

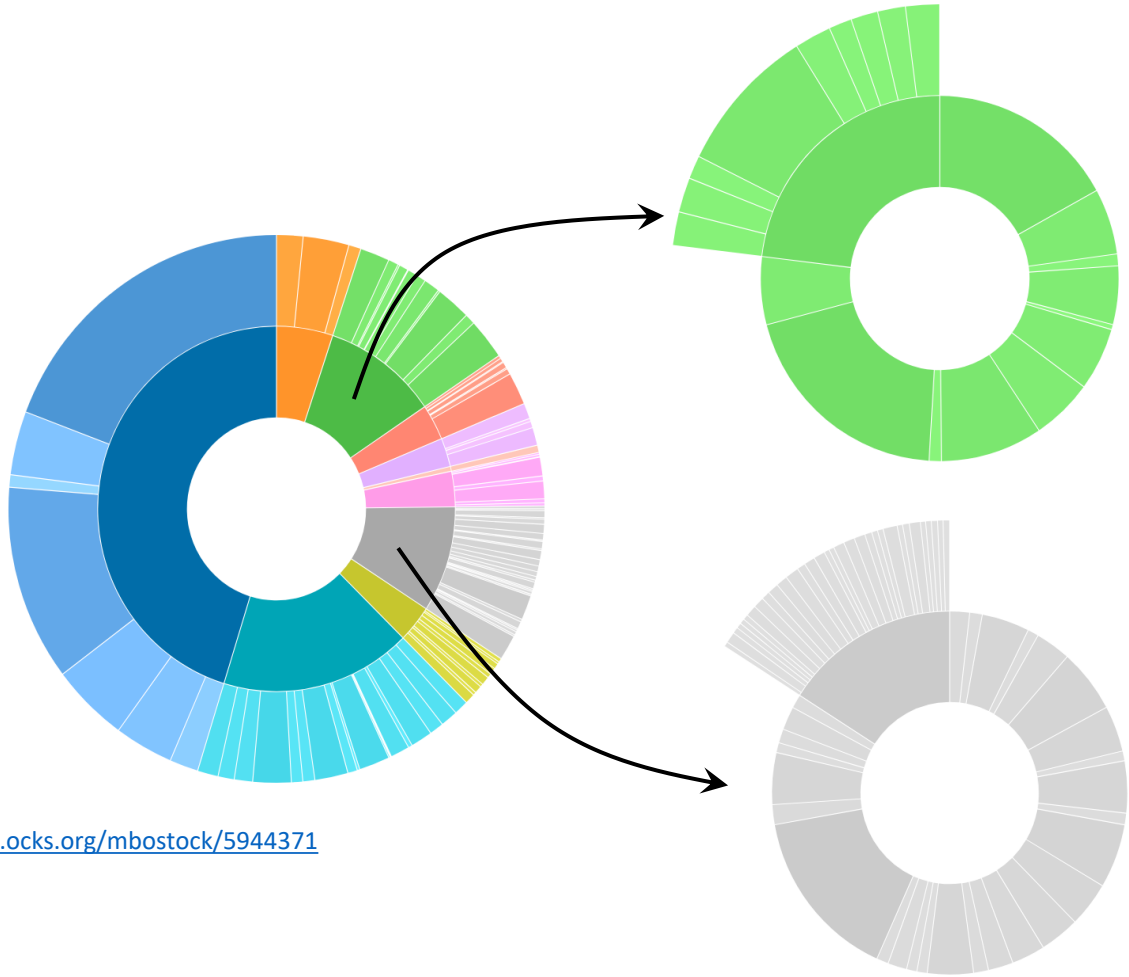
Navigate



https://www.jasondavies.com/parallel-sets

Manipulate

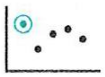
How?



https://bl.ocks.org/mbostock/5944371

Change

Select



Navigate

How?

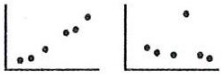
How to…

- … **arrange** the data in the view spatially?
- … **map** the data with all non-spatial channels?
- … **manipulate** the data in the view?
- … **facet** data between views?

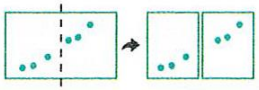
Facet (multi-view)

How?

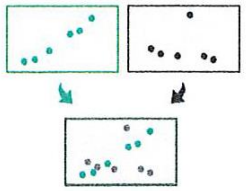
- Juxtapose

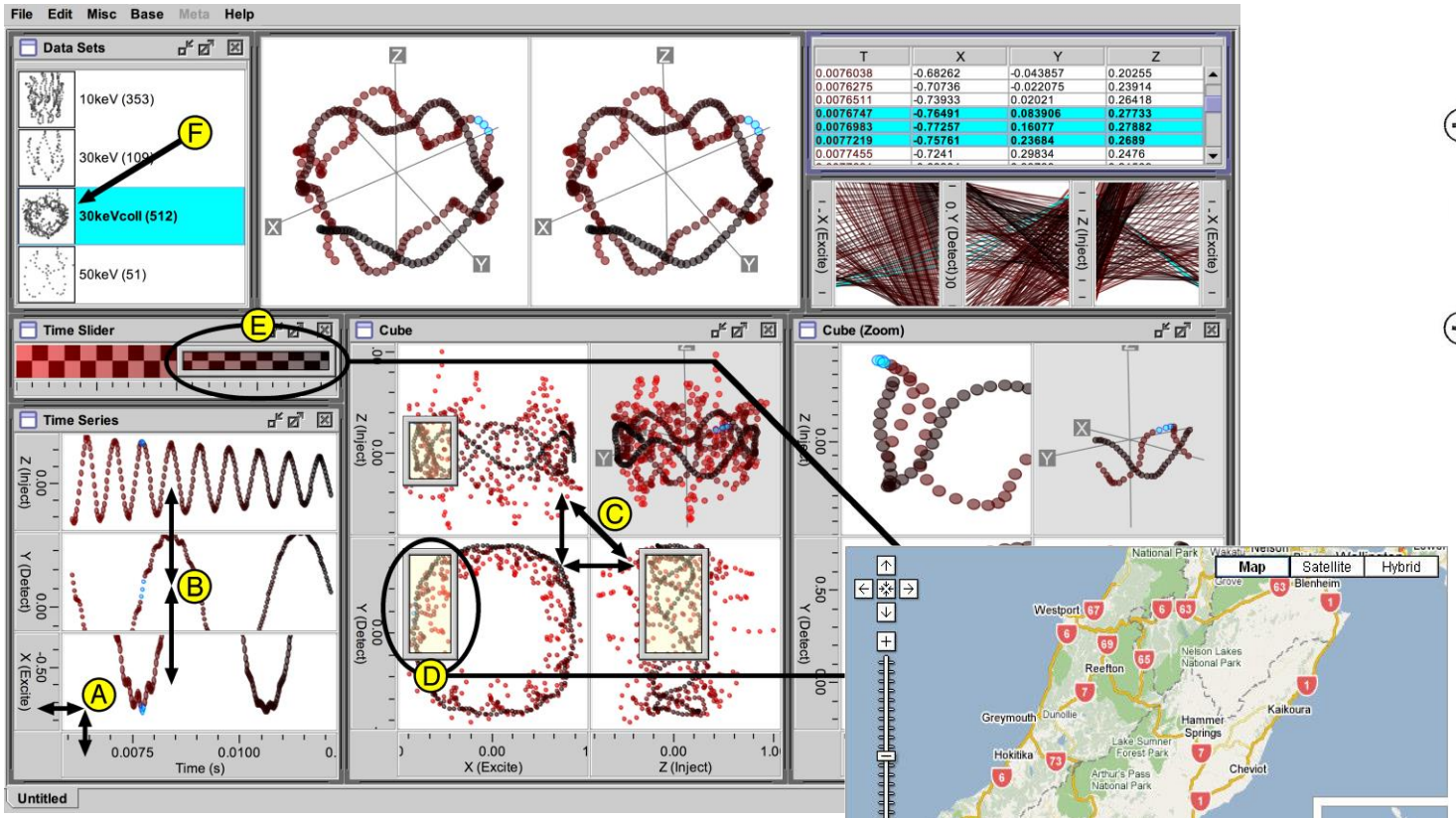


- Partition



- Superimpose



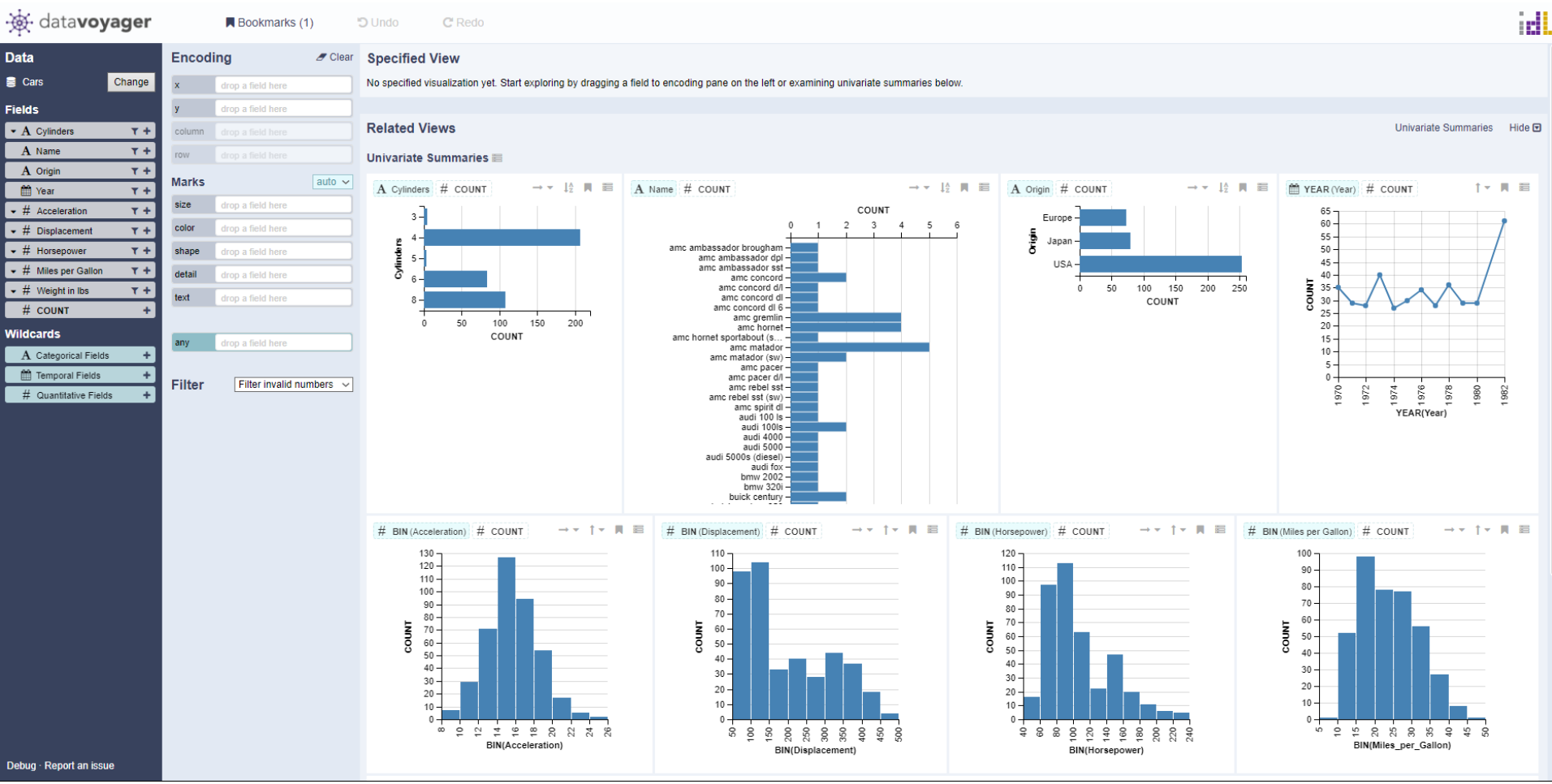


Overview + Detail

Weaver. **Building Highly-Coordinated Visualizations In Improvise**.
Proc. InfoVis 2004, p. 159-166

Facet (multi-view)

How?



https://vega.github.io/voyager2/

How?

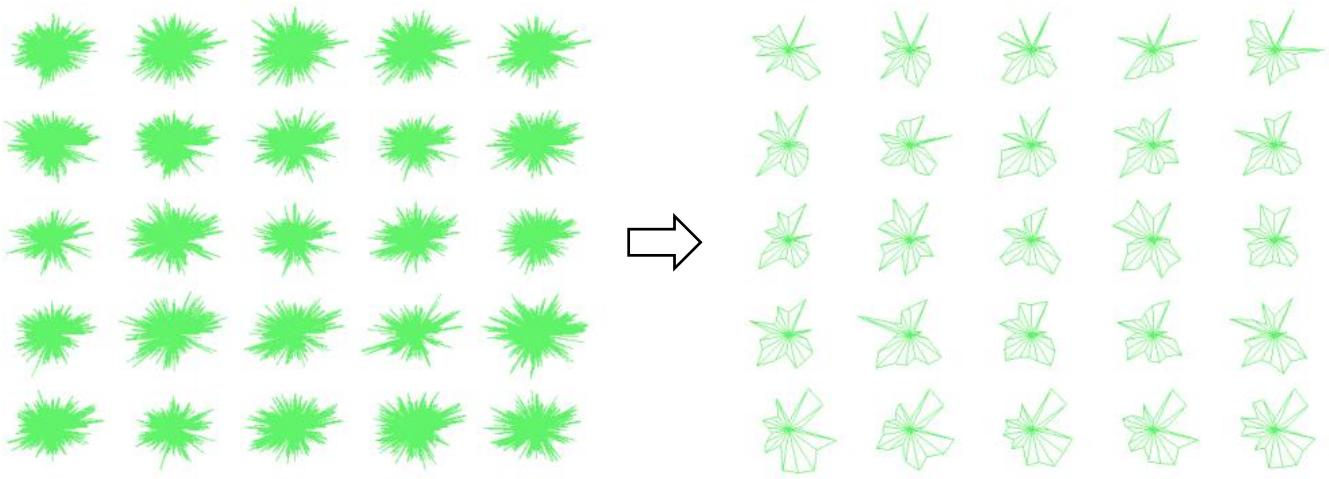
How to…

- … **arrange** the data in the view spatially?
- … **map** the data with all non-spatial channels?
- … **manipulate** the data in the view?
- … **facet** data between views?
- … **reduce** the data in the view?

Reduce

How?

- Filter
- Aggregate
- Embed

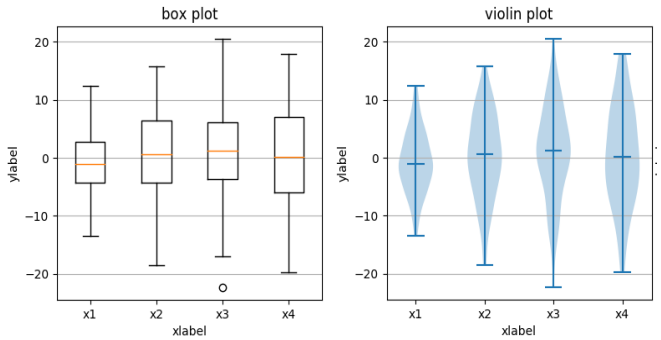


Yang, Jing, et al. "**Interactive hierarchical dimension ordering, spacing and filtering for exploration of high dimensional datasets**." *IEEE Symposium on Information Visualization, 2003*

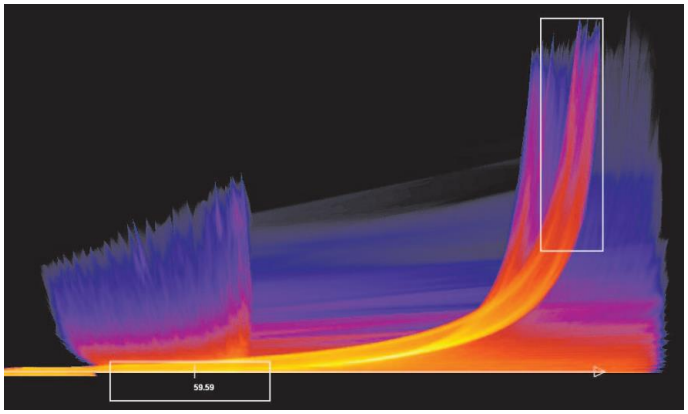
Reduce

How?

- Filter
- Aggregate
- Embed



https://matplotlib.org/gallery.html



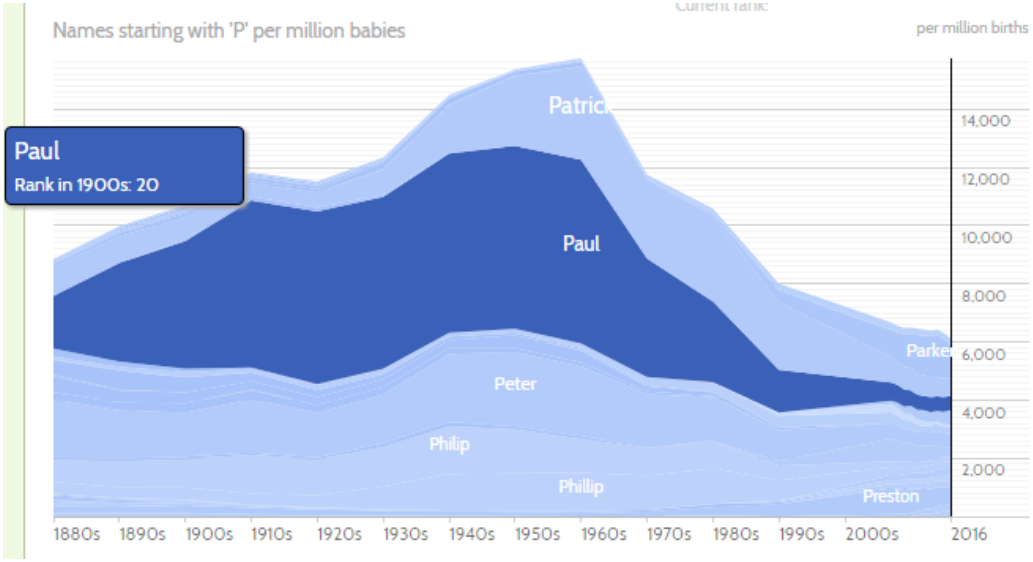
S. Bachthaler and D. Weiskopf, "**Continuous Scatterplots**," in IEEE Transactions on Visualization and Computer Graphics, vol. 14, no. 6, pp. 1428-1435, 2008.

S. Ingram, T. Munzner and M. Olano, "**Glimmer: Multilevel MDS on the GPU**," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 2, pp. 249-261, 2009.

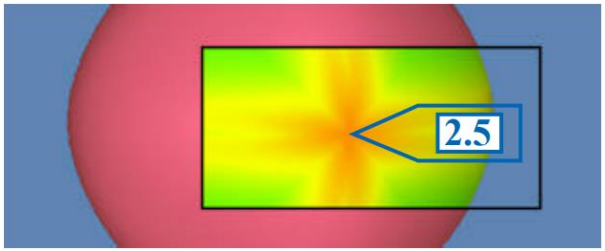
Reduce

How?

- Filter
- Aggregate
- Embed

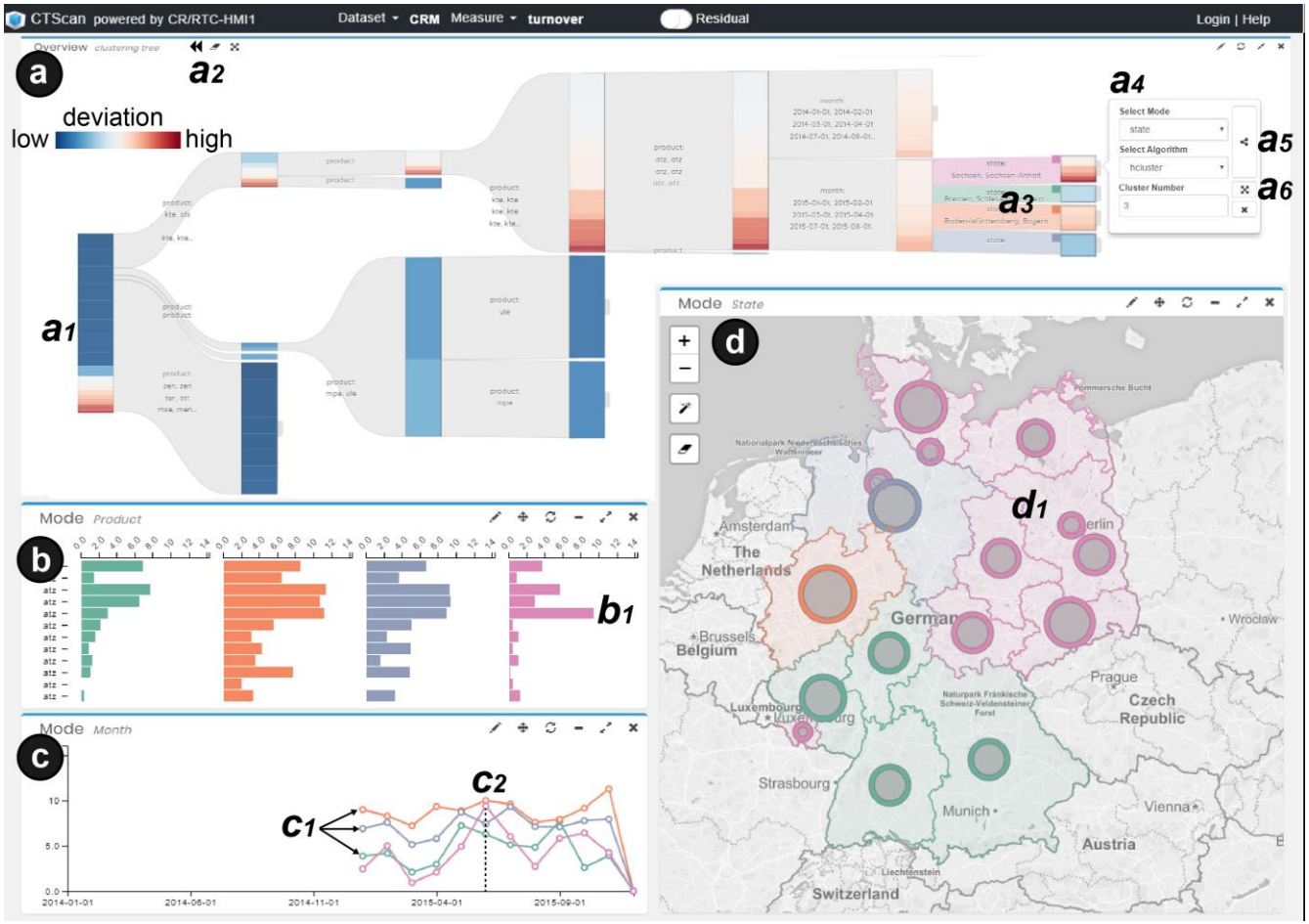


http://www.babynamewizard.com/voyager



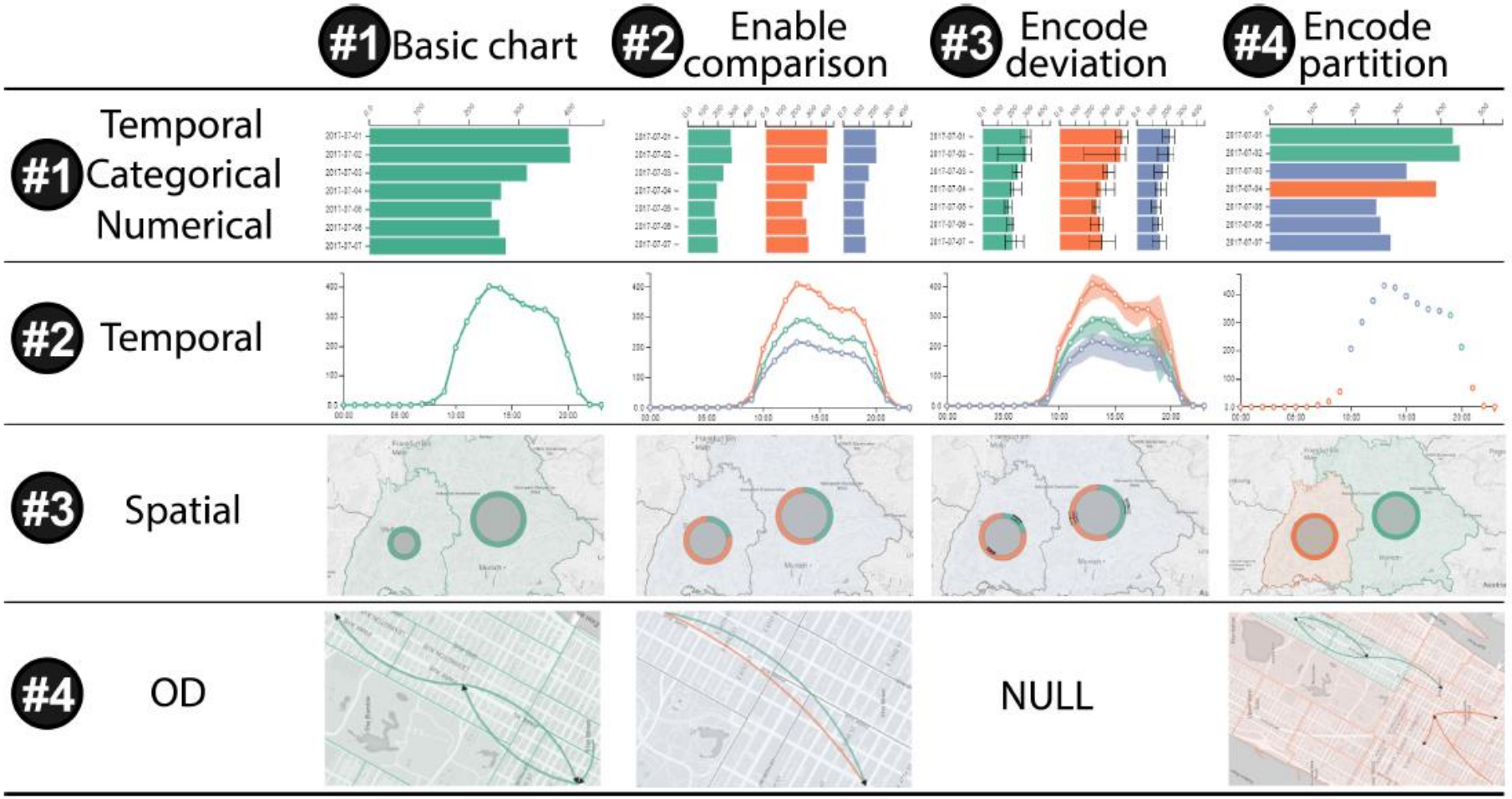
Eric A. Bier, Maureen C. Stone, Ken Pier, William Buxton, and Tony D. DeRose. 1993. **Toolglass and magic lenses: the see-through interface**. ACM SIGGRAPH '93

Example: Progressive Partition and Multidimensional Pattern Extraction for Large-Scale Spatio-Temporal Data Analysis

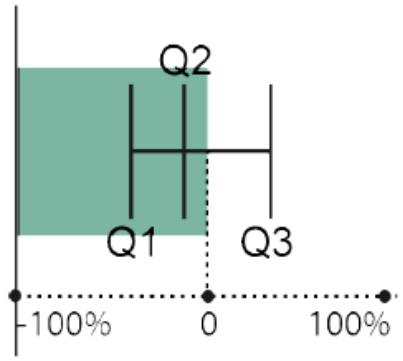


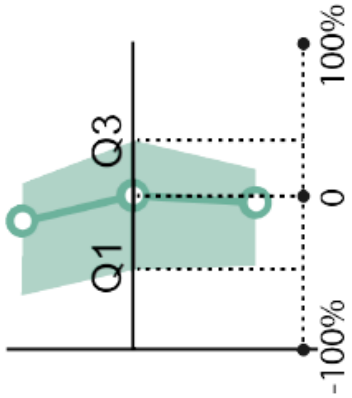
Liu, Dongyu, Panpan Xu, and Liu Ren. "**TPFlow: Progressive partition and multidimensional pattern extraction for large-scale spatio-temporal data analysis**." IEEE transactions on visualization and computer graphics 25.1 (2018): 1-11.

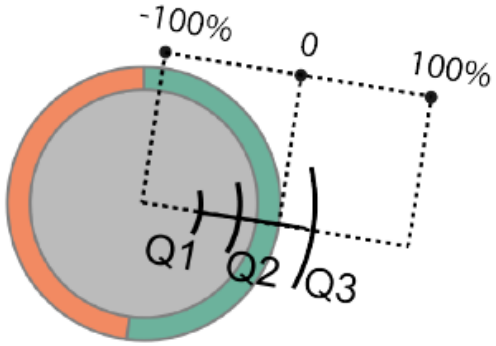
Pattern Comparison

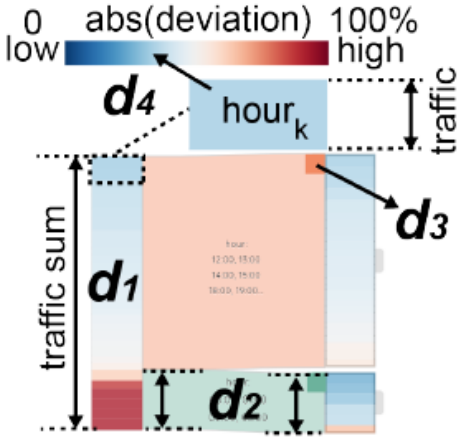


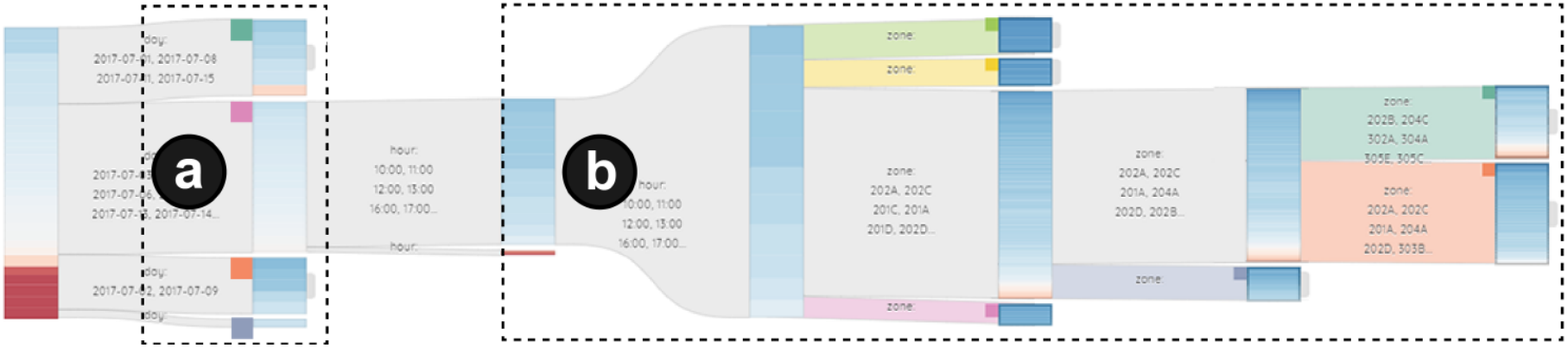
Data Encoding











TPFlow: Progressive Partition and Multidimensional Pattern Extraction for Large-Scale Spatio-Temporal Data Analysis

IEEE VAST 2018 - Conference on Visual Analytics Science and Technology
(TVCG Track*)

Dongyu Liu
Hong Kong University of
Science and Technology, Hong Kong

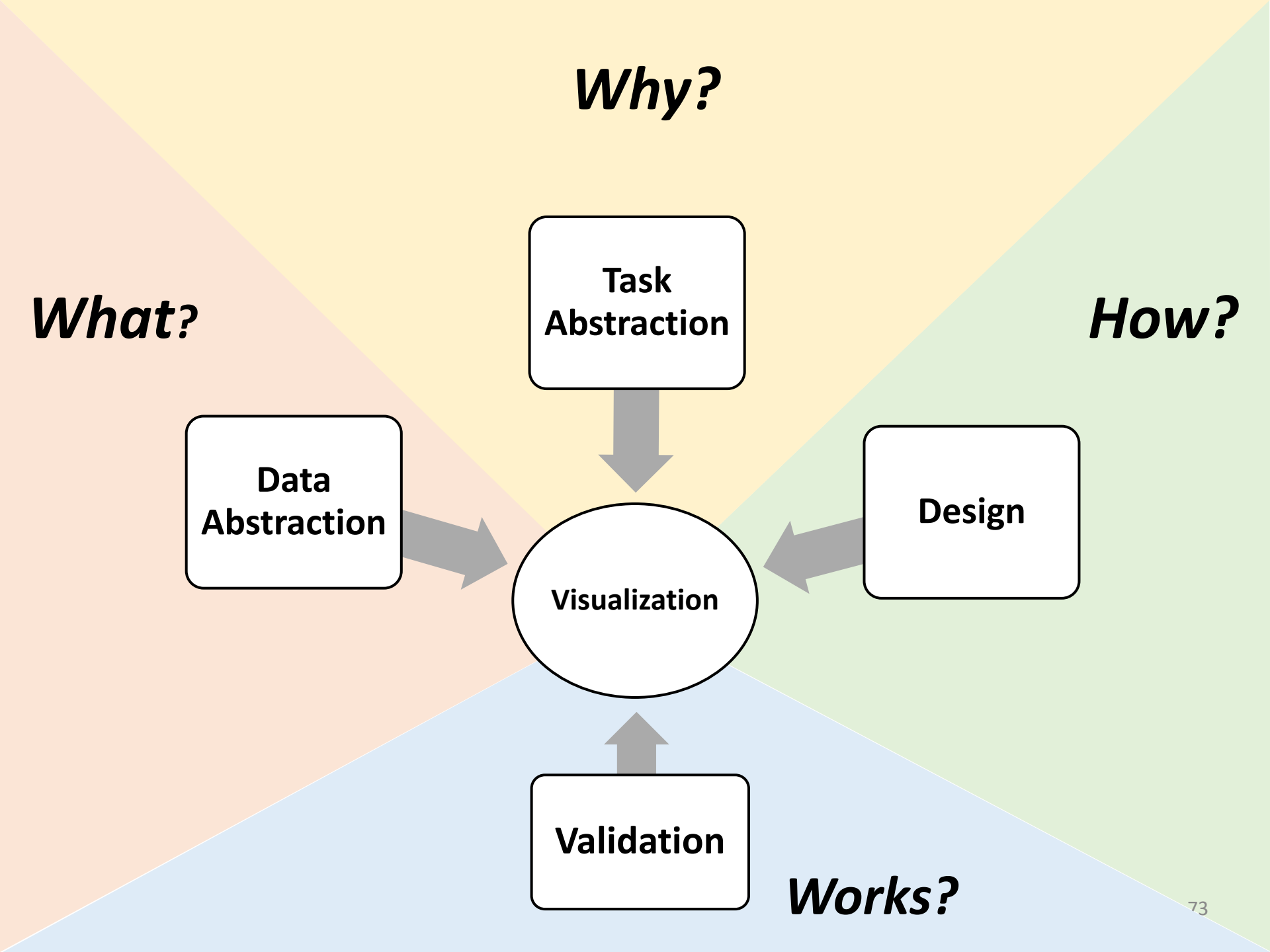
Panpan Xu **Liu Ren**
Bosch Research North America
Sunnyvale, CA

(With Audio)

**Paper selected to IEEE Transations on Visualization and Computer Graphics (TVCG)*

香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

BOSCH
Invented for life


Why?
What?
How?
Task Abstraction
Data Abstraction
Design
Visualization
Validation
Works?

Validation

Works?

- How do you know if it works?
- How do you measure it?
  - How do you decide the benchmark data?
- How do you argue that one design is better or worse than another?
- What does it better mean?
  - Do user get something done faster?

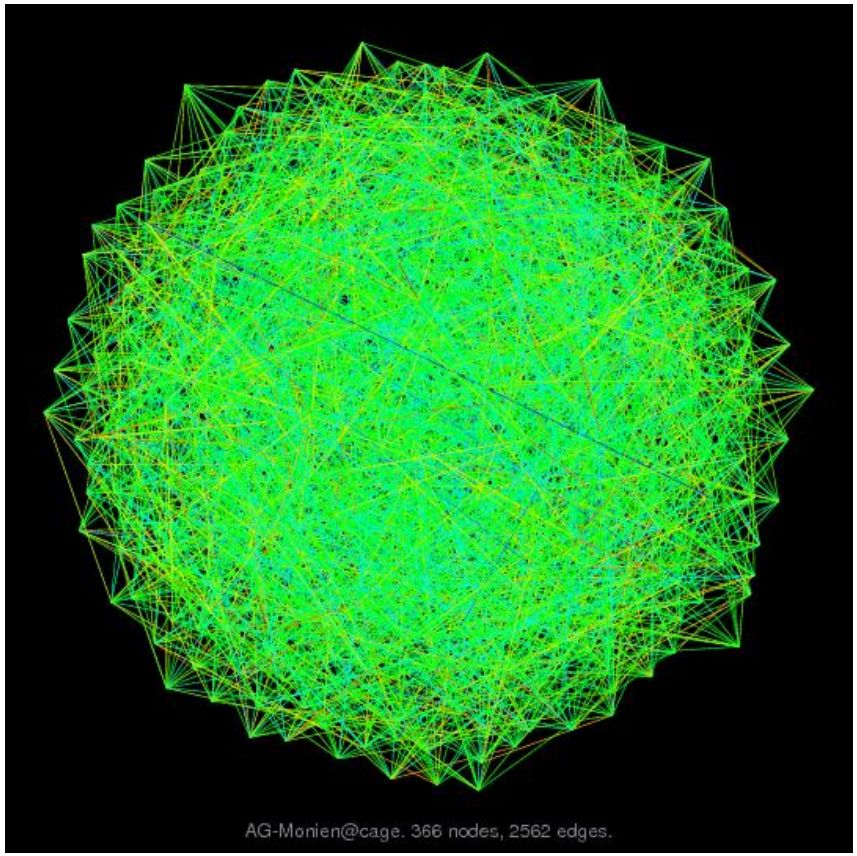
Works?

Why it can go wrong?

- I do not understand
  - Poor match with the properties of the human perceptual and cognitive system
- Lack of context
  - Comprehensible by a human in some other setting
- Scalability problems

Why it can go wrong?

Works?



http://yifanhu.net/GALLERY/GRAPHS

Works?

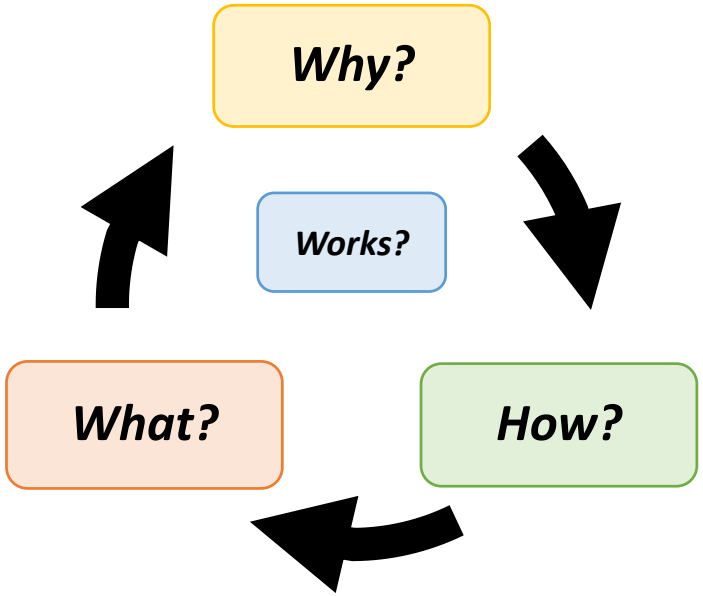
Good Practices

- Do not optimize, satisfy
  - **The perfect visualization does not exist!!**
- Need to know a wide variety of vis techniques
  - Ease the design process using references
- Justification and Alternatives
  - 3D vs 2D?
- Eyes Beat Memory
  - Comparing views better than switch views
- Interactivity and Responsiveness

Works?

Life Cycle

- Follow the **life cycle** of a visualization tool
  - Profile of the user (expert, novice,…)

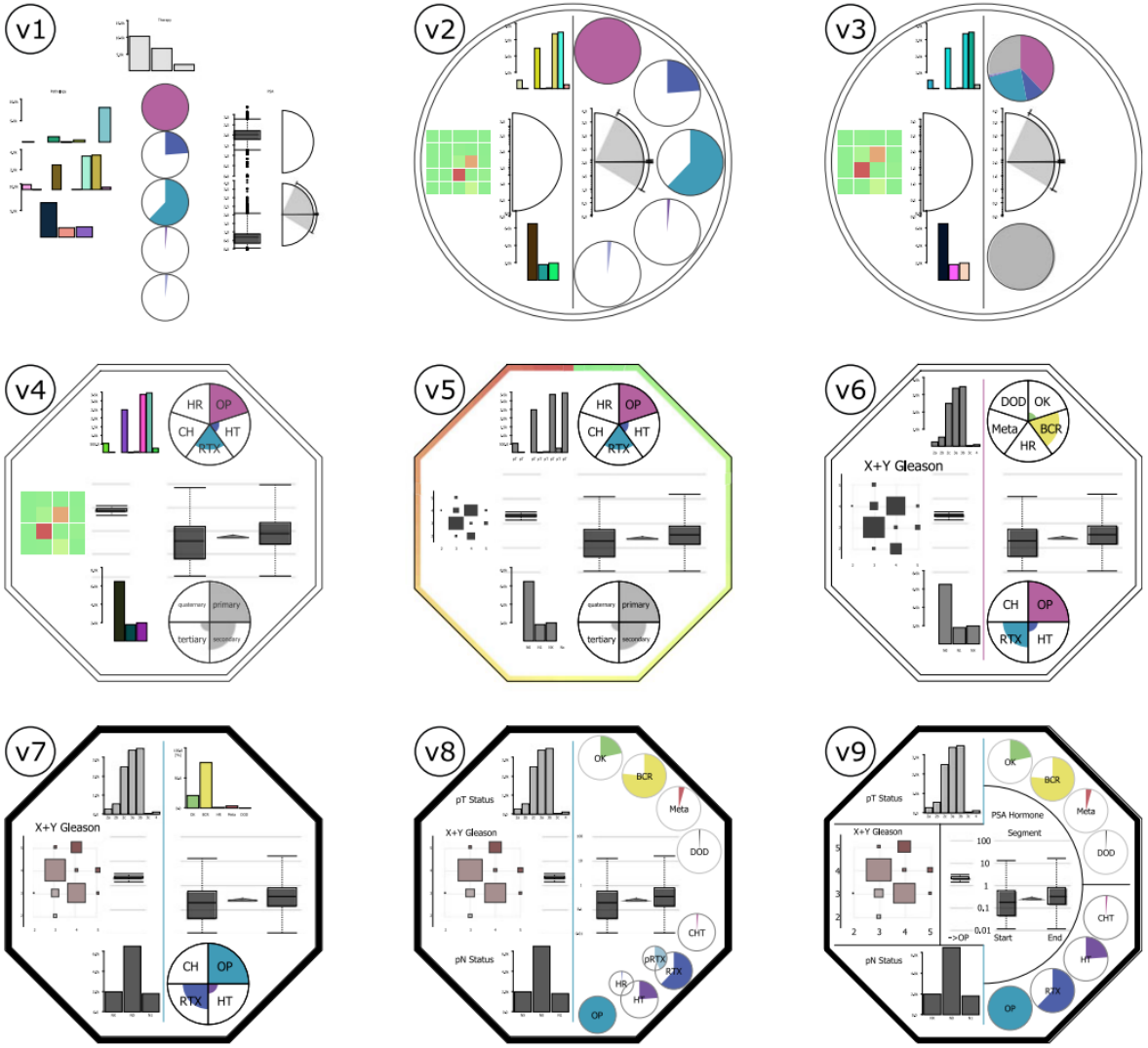


Prototype!

Trust the experts!

Evaluate!

Example: Using Dashboard Networks to Visualize Multiple Patient Histories



Bernard, Jürgen, et al. "**Using dashboard networks to visualize multiple patient histories: a design study on post-operative prostate cancer**." *IEEE transactions on visualization and computer graphics* 25.3 (2018): 1615-1628.

Overview of 15 Most Relevant Attributes Describing the Condition of a Patient

Attribute	Description	Att. Type	Value Domain	Visual Enc.
PSA Pre-OP	The value of the PSA hormone is a most important biological indicator for prostate cancer. Last measurement of the PSA hormone before OP (surgery).	Numerical	$[0;200.0[\frac{ng}{ml}$ $median: 8.7$	
PSA at segment start	Value of the PSA hormone at start of the visualized temporal segment.	Numerical	$[0;4,472.0[\frac{ng}{ml}$ $median: 0.21$	
PSA at segment end	Value of the PSA hormone at the end of the visualized temporal segment. Used to assess the progression of the PSA value within the segment interval.	Numerical	$[0;4,472.0[\frac{ng}{ml}$ $median: 0.26$	
PSA trend in segment	Trend of the PSA hormone value within the duration of a temporal segment. Depicts the delta between the PSA values in a relative way.	Numerical	$]-70.3;4381.68[\frac{ng}{ml}$ $median: 0.0$	
pT-status	Histopathological status of the tumor, obtained by pathology report (pT - pathological tumor). Used for prognoses about the progress of the disease.	Ordinal	$pt2a, pt2b, pt2c, pt3a$ $pt2b, pt3c, pt4$	
pN-status	Histopathological characteristics of the tumor, obtained by pathology report. Used for prognoses about the progress of the disease.	Ordinal	$NX, N0, N1$	
Gleason Score	Histopathological characteristics of the tumor, obtained by pathology report. Used for prognoses about the progress of the disease.	Ordinal	$2+3, 3+2, 3+3, 3+4, 3+5,$ $4+3, 4+4, 4+5, 5+3, 5+4$	
OK after OP	Outcome variable (biological condition) reflecting a good condition after OP. Is active until relapse or metastases.	Boolean	$[false, true]$	
BCR	Outcome variable (biological condition) describing the biochemical recurrence (relapse) of the tumor. Used for prognoses about the progress of the disease.	Boolean	$[false, true]$	
Metastases	Outcome variable (biological condition) describing whether a patient got metastases. Metastases are the most severe biological indicator in the dataset.	Boolean	$[false, true]$	
DOD	Biological end point when a patient died of disease. Often used for cause-effect analyses.	Boolean	$[false, true]$	
OP	Treatment attribute reflecting whether a patient had surgery. In the dataset all patents received OP.	Boolean	$[false, true]$	
RTX	Treatment attribute reflecting whether a patient got radiation therapy. Local treatment, only affecting the prostate area.	Boolean	$[false, true]$	
HT	Treatment attribute reflecting whether a patient had hormone therapy. Systemic treatment, affecting the whole human body.	Boolean	$[false, true]$	
CHT	Treatment attribute reflecting whether a patient got chemotherapy. Often applied in severe situations, e.g., when metastases are detected.	Boolean	$[false, true]$	

Immersive Analytics

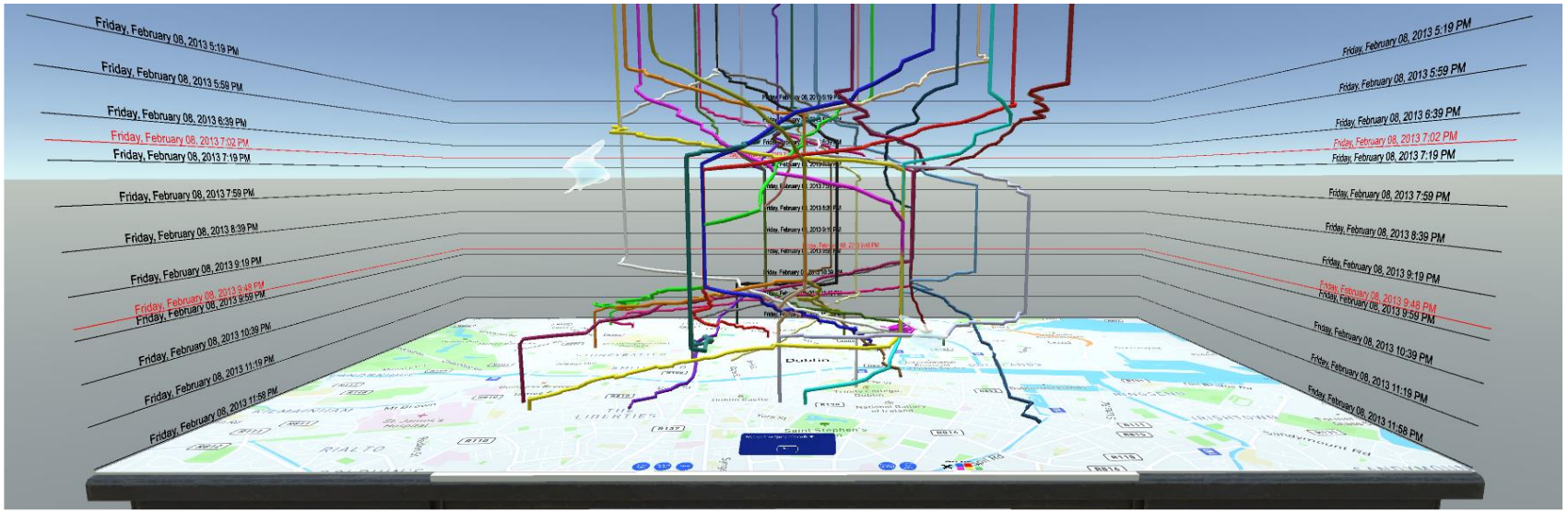
Focus

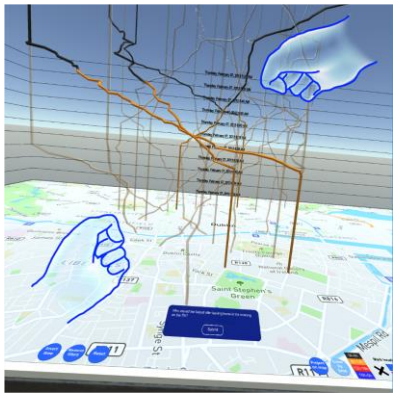
Immersive Analytics

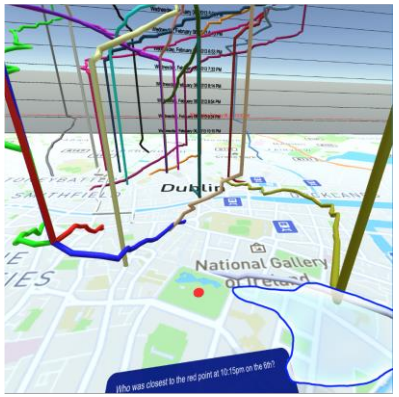
"Science of analytical reasoning facilitated by immersive human-computer interfaces"

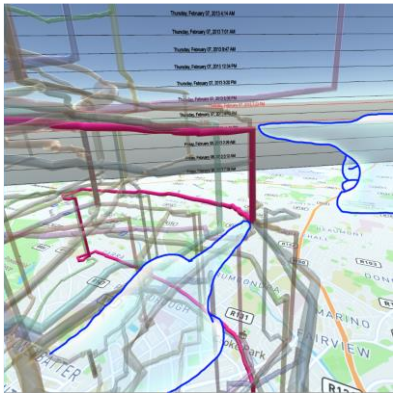
Skarbez R, Polys NF, Ogle JT, North C and Bowman DA (2019) Immersive Analytics: Theory and Research Agenda. Front. Robot. AI 6:82. doi: 10.3389/frobt.2019.00082
Chandler, T., Cordeil, M., Czauderna, T., Dwyer, T., Glowacki, J., Goncu, C., et al. (2015). "Immersive analytics," in 2015 Big Data Visual Analytics (BDVA), (Hobart, TAS: IEEE), 1–8.

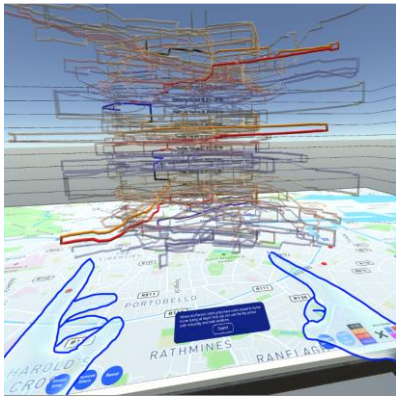
Space-Time Cube Geovisualization











J. A. W. Filho, W. Stuerzlinger and L. Nedel, "Evaluating an Immersive Space-Time Cube Geovisualization for Intuitive Trajectory Data Exploration," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 1, pp. 514-524, Jan. 2020, doi: 10.1109/TVCG.2019.2934415.

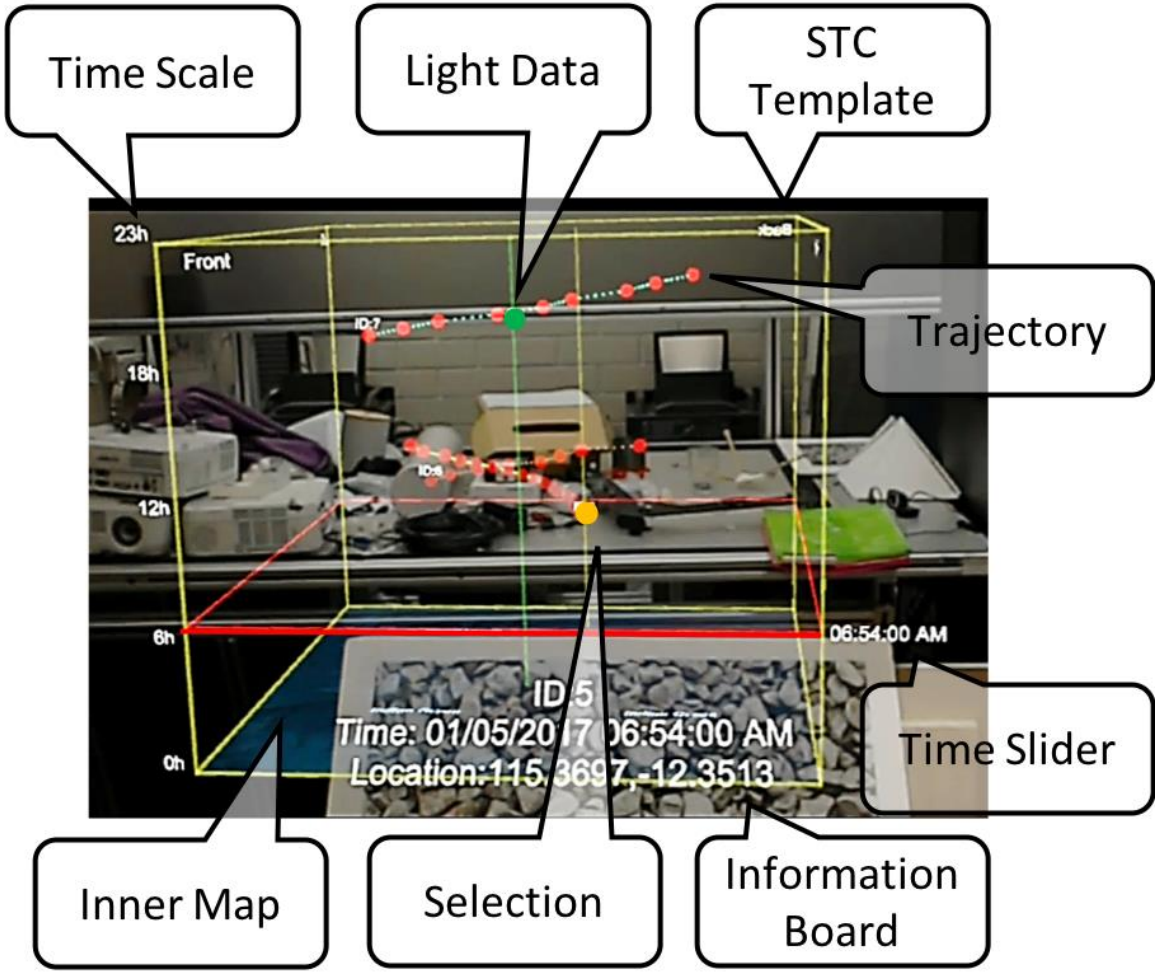
INFOVIS PAPER

Evaluating an Immersive Space-Time Cube Geovisualization for Intuitive Trajectory Data Exploration

Jorge Wagner, Wolfgang Stuerzlinger, Luciana Nedel

VIS 2019
20–25 October 2019
Vancouver, Canada
ieeevis.org

GeoGate

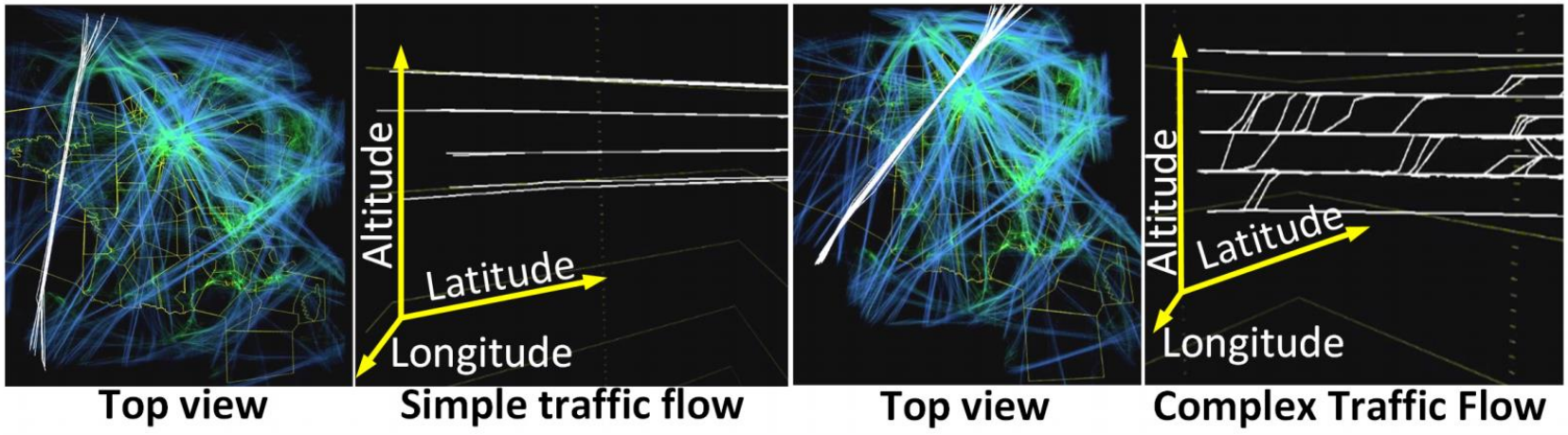


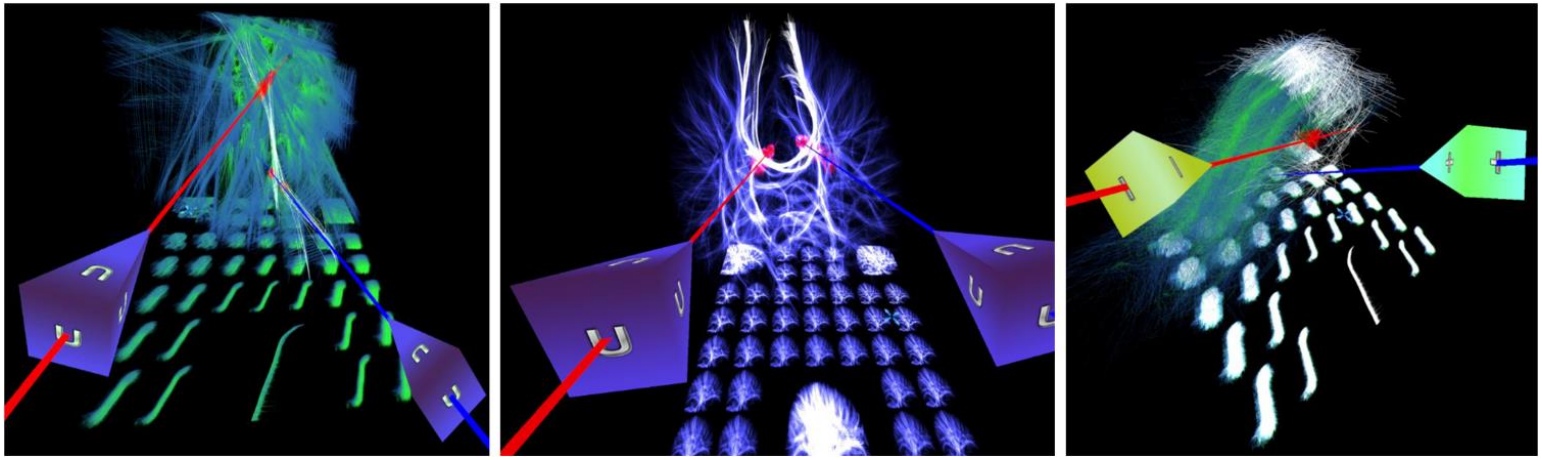
S. Y. Ssin, J. A. Walsh, R. T. Smith, A. Cunningham and B. H. Thomas, "GeoGate: Correlating Geo-Temporal Datasets Using an Augmented Reality Space-Time Cube and Tangible Interactions," *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, Osaka, Japan, 2019, pp. 210-219, doi: 10.1109/VR.2019.8797812.

Bento Box

Johnson S, Orban D, Runesha HB, Meng L, Juhnke B, Erdman A, Samsel F and Keefe DF (2019) Bento Box: An Interactive and Zoomable Small Multiples Technique for Visualizing 4D Simulation Ensembles in Virtual Reality. *Front. Robot. AI* 6:61. doi: 10.3389/frobt.2019.00061

FiberClay





C. Hurter, N. H. Riche, S. M. Drucker, M. Cordeil, R. Alligier and R. Vuillemot, "FiberClay: Sculpting Three Dimensional Trajectories to Reveal Structural Insights," in IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 1, pp. 704-714, Jan. 2019, doi: 10.1109/TVCG.2018.2865191.

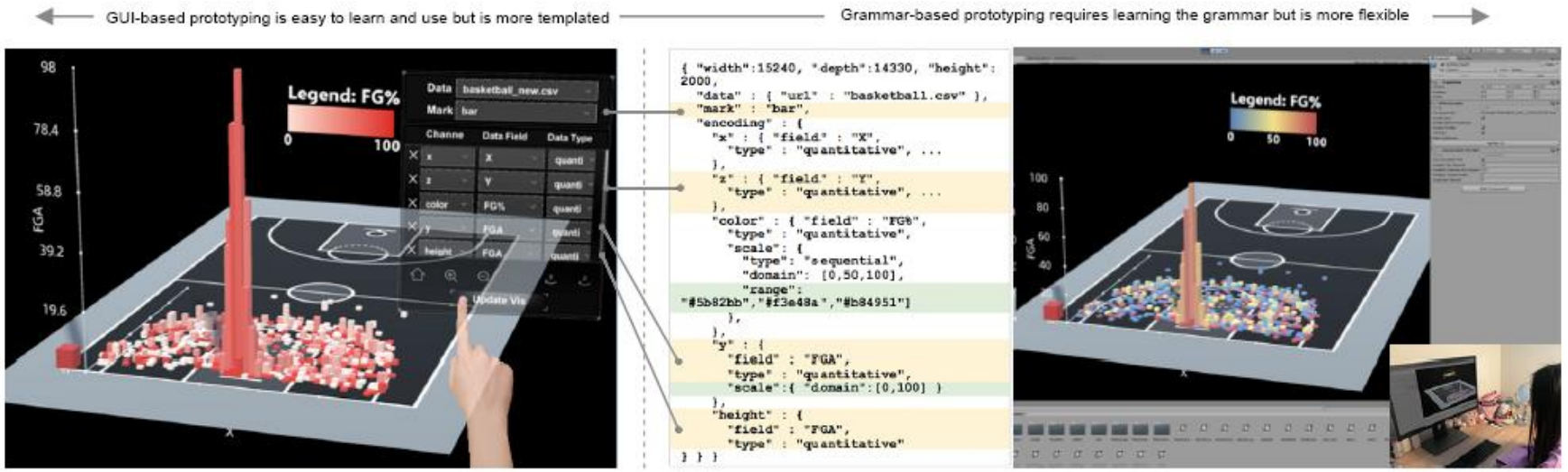
INFOVIS PAPER

FiberClay: Sculpting Three Dimensional Trajectories to Reveal Structural Insights

Christophe Hurter, Nathalie Henry Riche, Steven Drucker, Maxime Cordeil, Richard Alligier, Romain Vuillemot

VIS2018 21-26 October 2018 Berlin, Germany ieeevis.org

Toolkits for Immersive Analitics

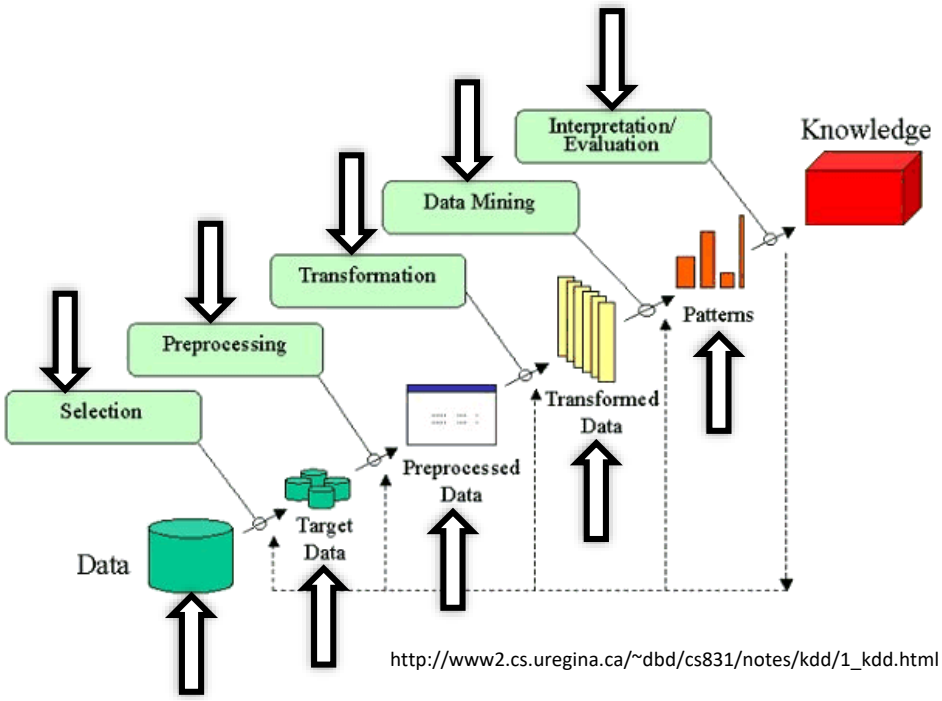


R. Sicat *et al*., "DXR: A Toolkit for Building Immersive Data Visualizations," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 715-725, Jan. 2019, doi: 10.1109/TVCG.2018.2865152.

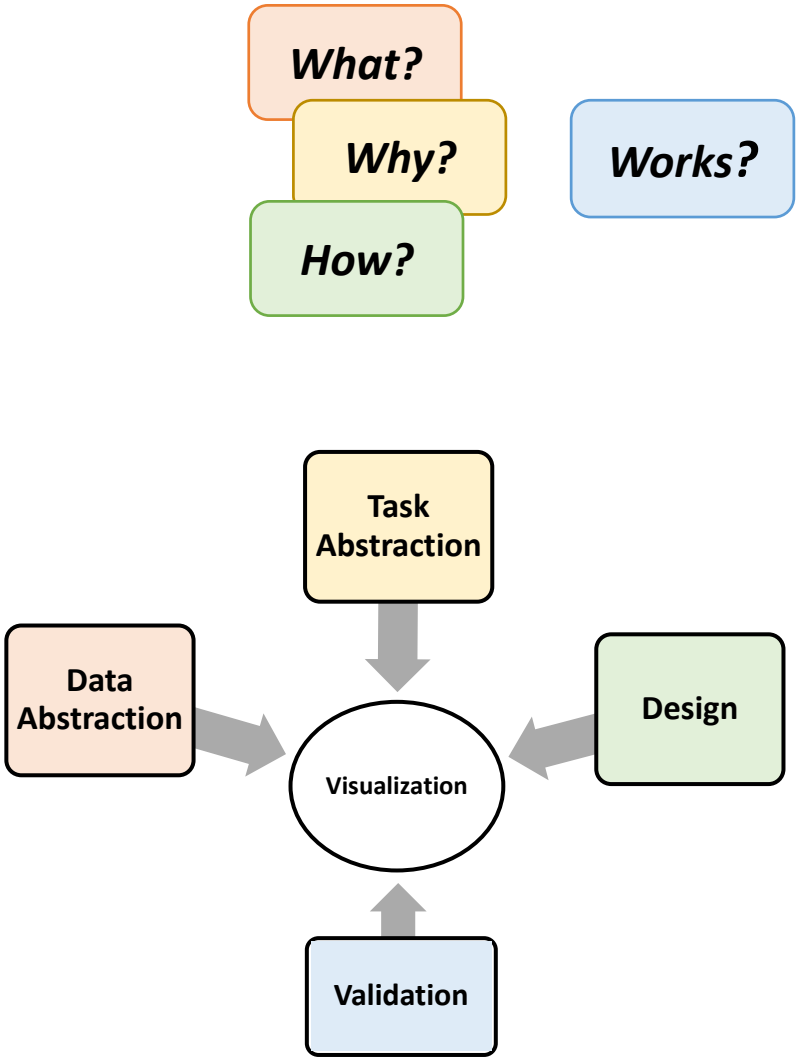
Research Challenges in Immersive Analytics

- Combining Human and Computer Intelligence
- The Utility of Immersion
- Designing Immersive Analytics Systems
- Facilitating Collaboration through Immersion
- Changing the Process of Analysis with Immersion

Wrap Up



http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html



Further reading

- Remy Dautriche. Multi-scale Interaction Techniques for the Interactive Visualization of Execution Traces. Chapter 3.3, Pattern Visualization, 2016
- T. Munzner. **Visualization**. Chapter 27, p 675-707, of Fundamentals of Graphics, Third Edition, by Peter Shirley et al. AK Peters, 2009.
  - http://www.cs.ubc.ca/labs/imager/tr/2009/VisChapter/akp-vischapter.pdf
- T. Munzner. "**A Nested Model for Visualization Design and Validation**," in IEEE Transactions on Visualization and Computer Graphics, vol. 15, no. 6, pp. 921-928, Nov.-Dec. 2009.
  - http://ieeexplore.ieee.org/abstract/document/5290695/
- J. Heer, M. Bostock, and V. Ogievetsky. "**A tour through the visualization zoo**." Queue 8.5 (2010): 20.
  - http://cacm.acm.org/magazines/2010/6/92482-a-tour-through-the-visualization-zoo/pdf
- I. Herman, G. Melançon, and M. S. Marshall. "**Graph visualization and navigation in information visualization: A survey**." IEEE Transactions on visualization and computer graphics 6.1 (2000): 24-43.
  - http://ieeexplore.ieee.org/abstract/document/841119/
- F. Beck et al. **A taxonomy and survey of dynamic graph visualization**. Computer Graphics Forum. Vol. 36. No. 1. 2017.
  - http://onlinelibrary.wiley.com/doi/10.1111/cgf.12791/full

Further reading

- **Visualization Analysis and Design**
- **Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations**
- **The Functional Art: An introduction to information graphics and visualization**