

CS49000-VIZ - Fall 2020

Introduction to Data Visualization

Introduction

Lecture 1

August 21, 2020

Instructor

Xavier Tricoche xmt@purdue.edu

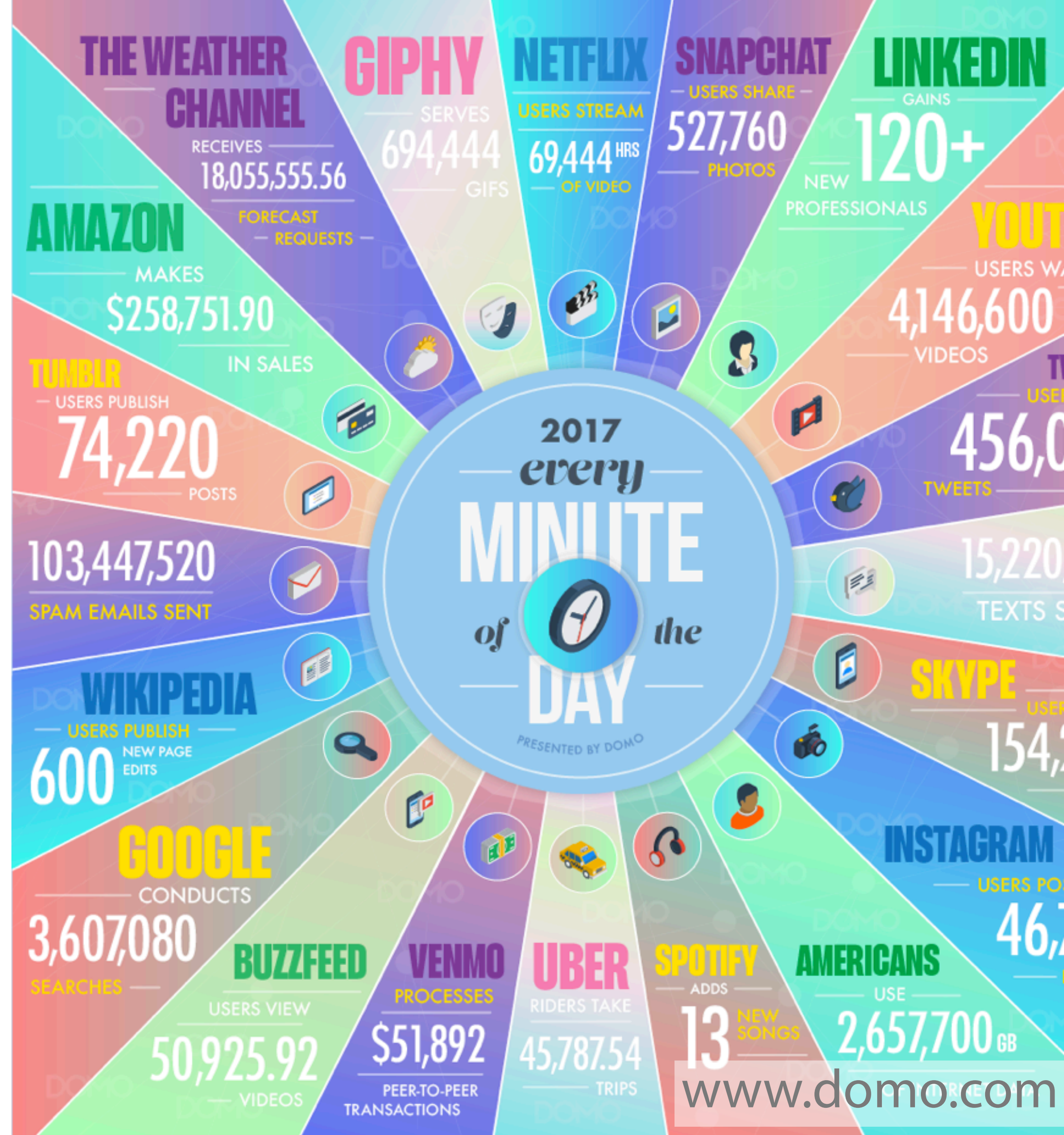
- Assoc. Prof. of Computer Science
- Research interests: Visualization, Data Analysis, Scientific Computing
- Office: LWSN 3154P

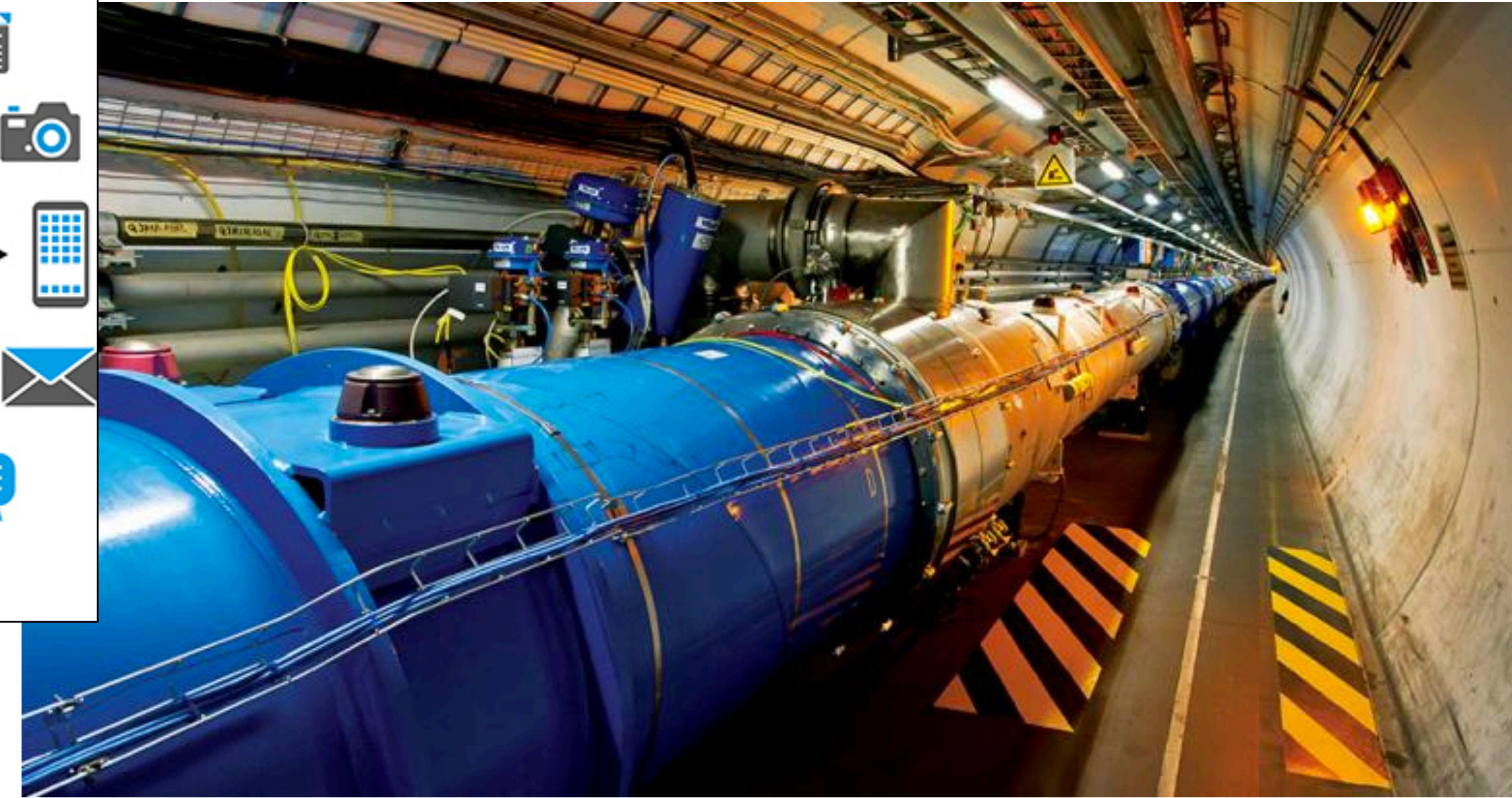
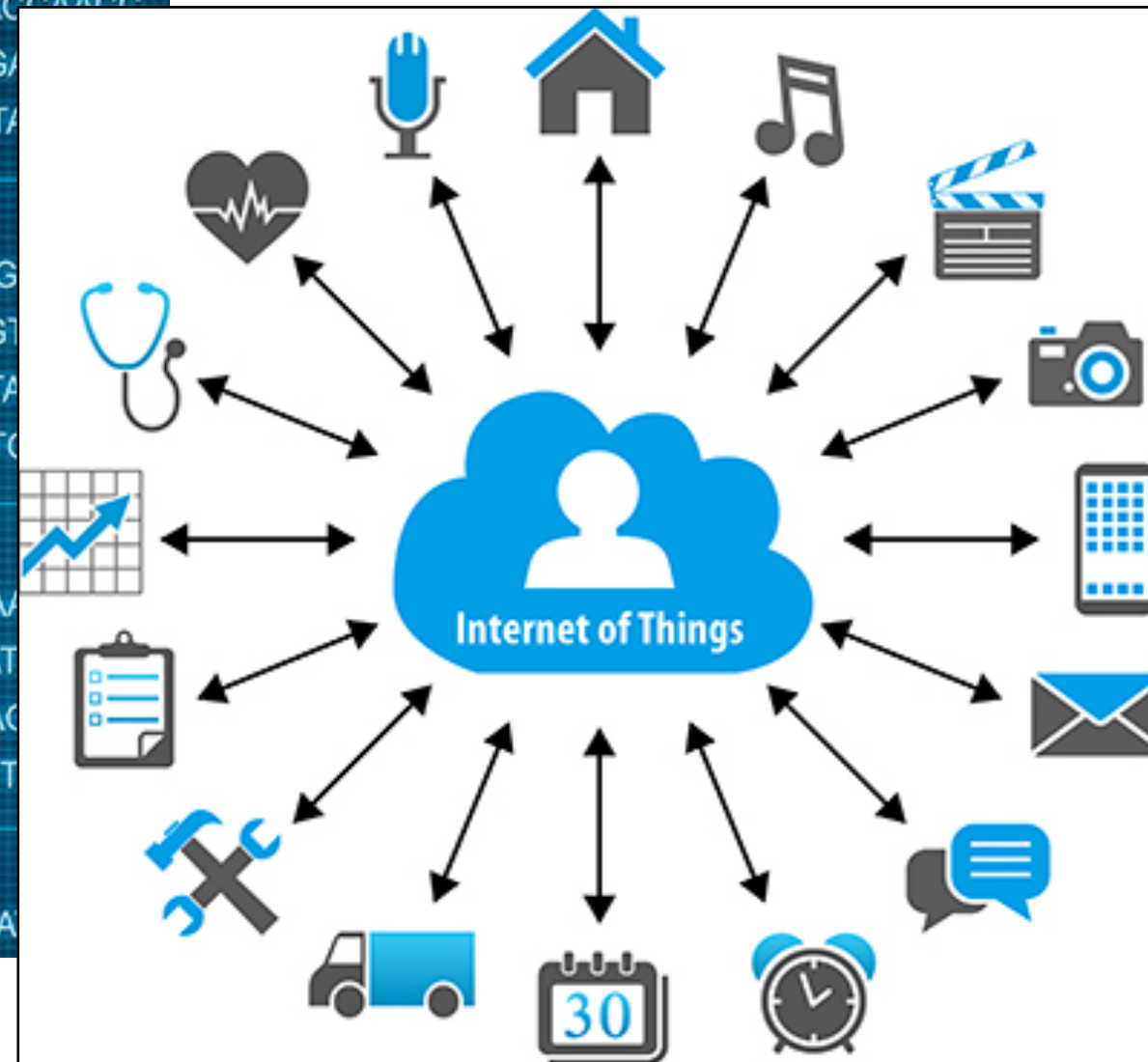
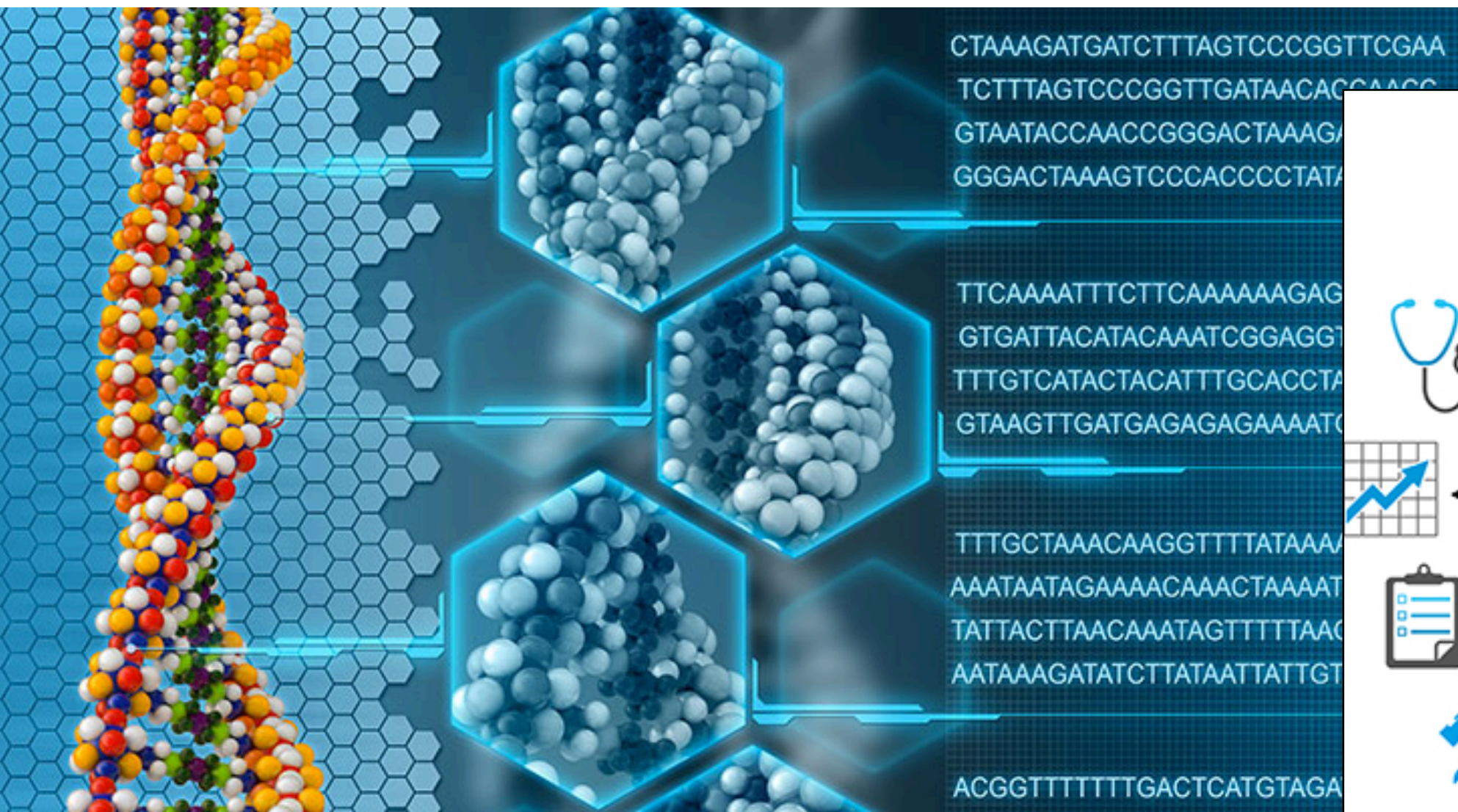
Today

- Introduction
- Syllabus
- Class overview
- Visualization libraries (1/2)
- Homework

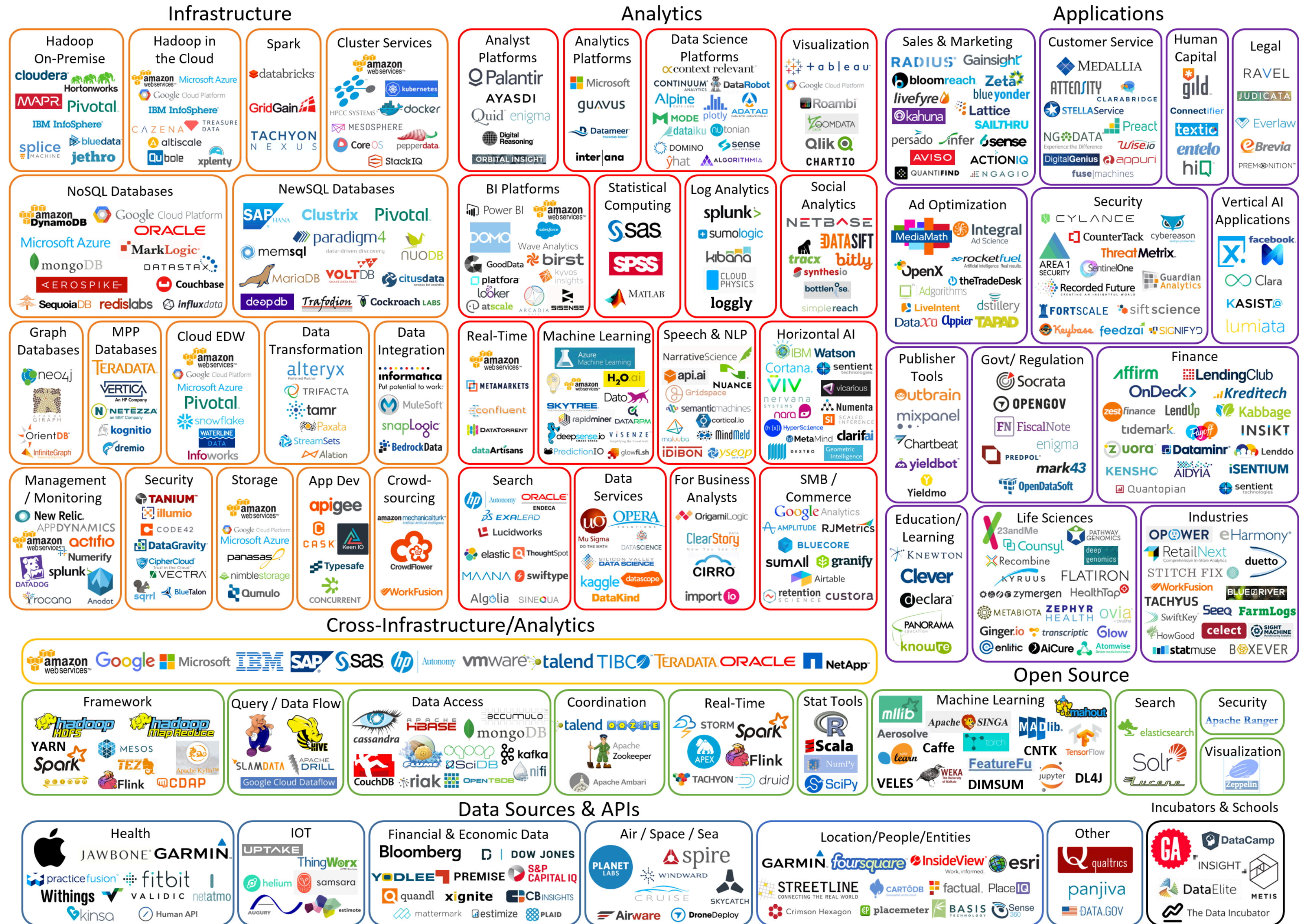
Big Data

- 2.5×10^{18} bytes/min
- 90% of world data created in last 2 years





Big Data Landscape 2016



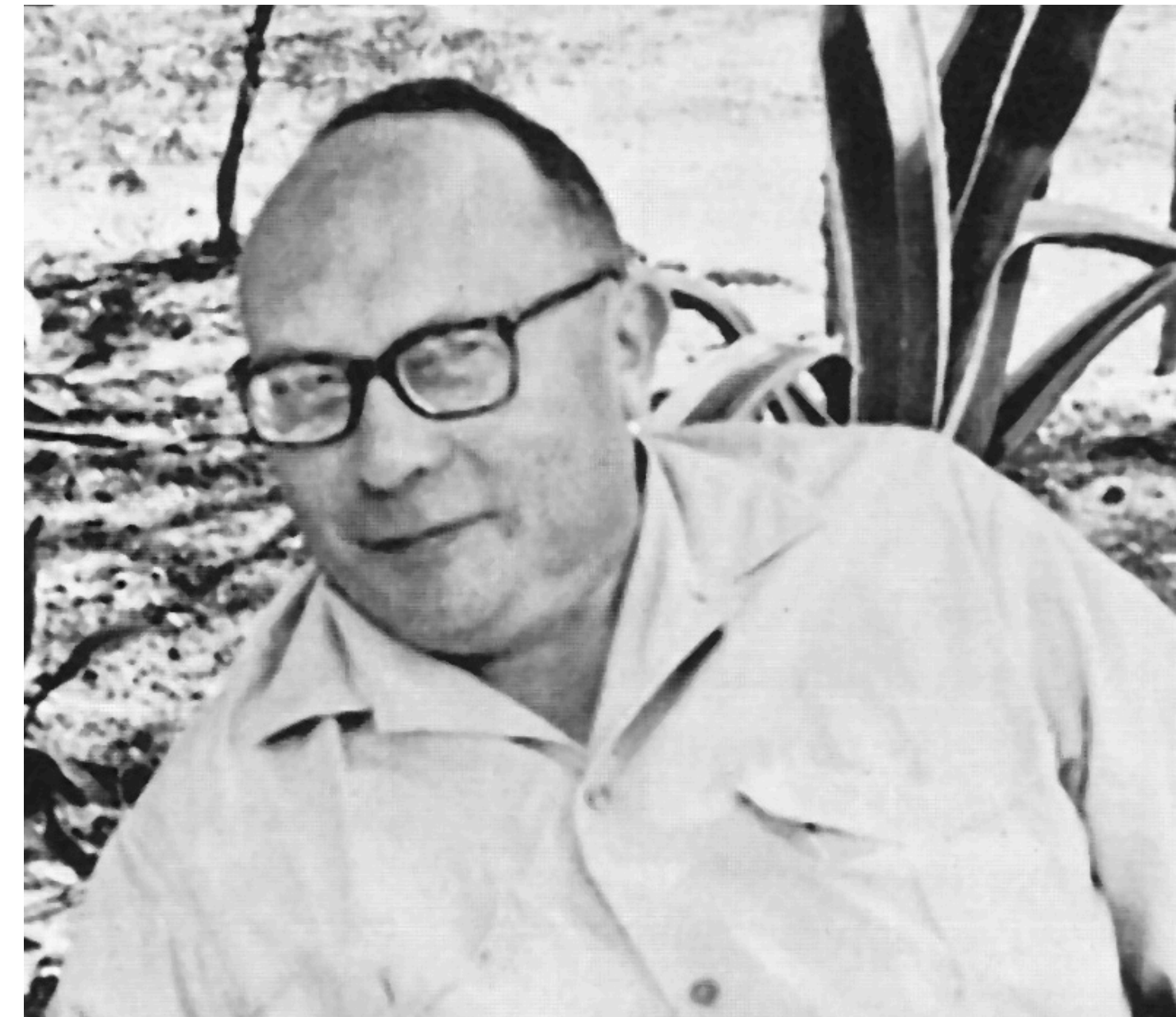
© Matt Turck (@mattturk), Jim Hao (@jimrhao), & FirstMark Capital (@firstmarkcap)

FIRSTMARK

Classical Definitions

“... the artificial memory that best supports our natural means of perception.”

[Bertin 1967]



Classical Definitions

“Transformation of the symbolic into the geometric”

[McCormick et al. 1987]

Executive Summary

Panel Report on Visualization in Scientific Computing

Visualization in Scientific Computing (ViSC) is emerging as a major computer-based field, with a body of problems, a commonality of tools and terminology, boundaries, and a cohort of trained personnel. As a tool for applying computers to science, it offers a way to see the unseen. As a technology, Visualization in Scientific Computing promises radical improvements in the human/computer interface and may make human-in-the-loop problems approachable.

Visualization in Scientific Computing can bring enormous leverage to bear on scientific productivity and the potential for major scientific breakthroughs, at a level of influence comparable to that of supercomputers themselves. It can bring advanced methods into technologically intensive industries and promote the effectiveness of the American scientific and engineering communities. Major advances in Visualization in Scientific Computing and effective national diffusion of its technologies will drive techniques for understanding how models evolve computationally, for tightening design cycles, integrating hardware and software tools, and standardizing user interfaces.

Visualization in Scientific Computing will also provide techniques for exploring an important class of computational science problems, relying on cognitive pattern recognition or human-in-the-loop decision making. New methods may include guiding simulations interactively and charting their parameter space graphically in real time. Significantly more complexity can be comprehended through Visualization in Scientific Computing techniques than through classical ones.

The university/industrial research and development cycle is found to be inadequate for Visualization in Scientific Computing. The programs and facilities are not in place for researchers to identify and address problems far enough in advance, even though the emerging

discipline of Visualization in Scientific Computing is found to be critically important to a portion of the country's domestic and export trade threatened by foreign competition. At the present rate of growth, the capabilities of networks, displays, and storage systems will not be adequate for the demands Visualization in Scientific Computing will place on them.

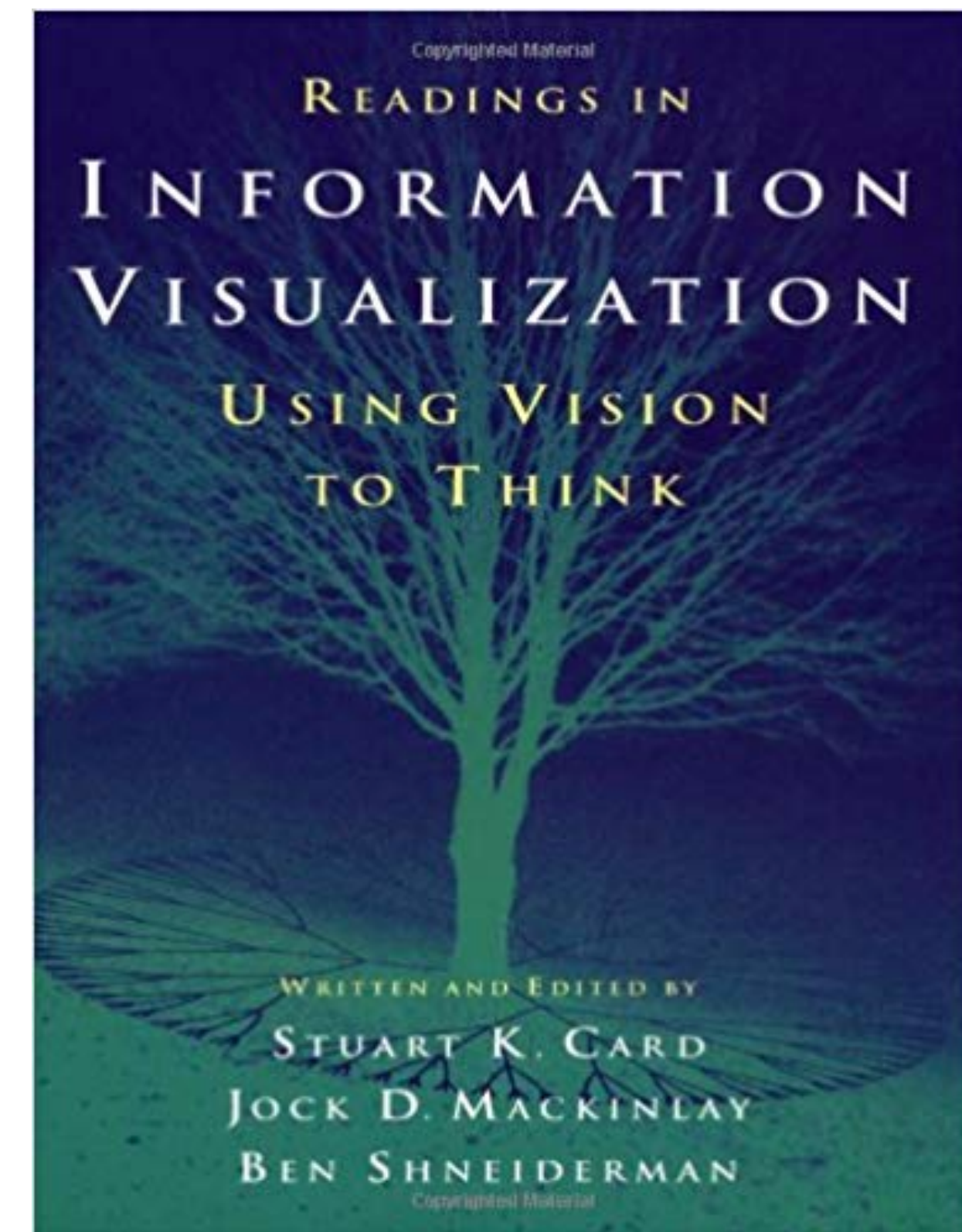
The gigabit bandwidth of the eye/visual cortex system permits much faster perception of geometric and spatial relationships than any other mode, making the power of supercomputers more accessible. Users from industry, universities, medicine and government are largely unable to comprehend or influence the "fire hoses" of data, produced by contemporary sources such as supercomputers and satellites, because of inadequate Visualization in Scientific Computing tools. The current allocation of resources at the national supercomputer centers is considered unbalanced against visualization, in competition with demands for more memory and disks, faster machines, faster networks, and so forth, although all need to be improved.

The Panel recommends a new initiative in Visualization in Scientific Computing, to get visualization tools into "the hands and minds" of scientists. Scientists and engineers would team up with visualization researchers in order to solve graphics, image processing, human/computer interface, or representational problems grounded in the needs and methods of an explicit discipline. The expectation is that visualization tools solving hard, driving problems in one computational science would be portable to problems in another. Proposals would be peer reviewed, and awarded for both facilities and projects at national supercomputer centers and elsewhere. Other agencies of government are encouraged to recognize the value of Visualization in Scientific Computing in their missions and support its development accordingly.

Classical Definitions

“The use of computer-generated, interactive, visual representations of data to amplify cognition.”

[Card, Mackinlay, & Shneiderman 1999]



More Definitions

“Creation of images that convey salient information about underlying data and processes”

[Hansen & Johnson, 2005]

“Communication of information using graphical representations”

[Ward et al., 2010]

More Definitions

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

[Munzner, 2014]

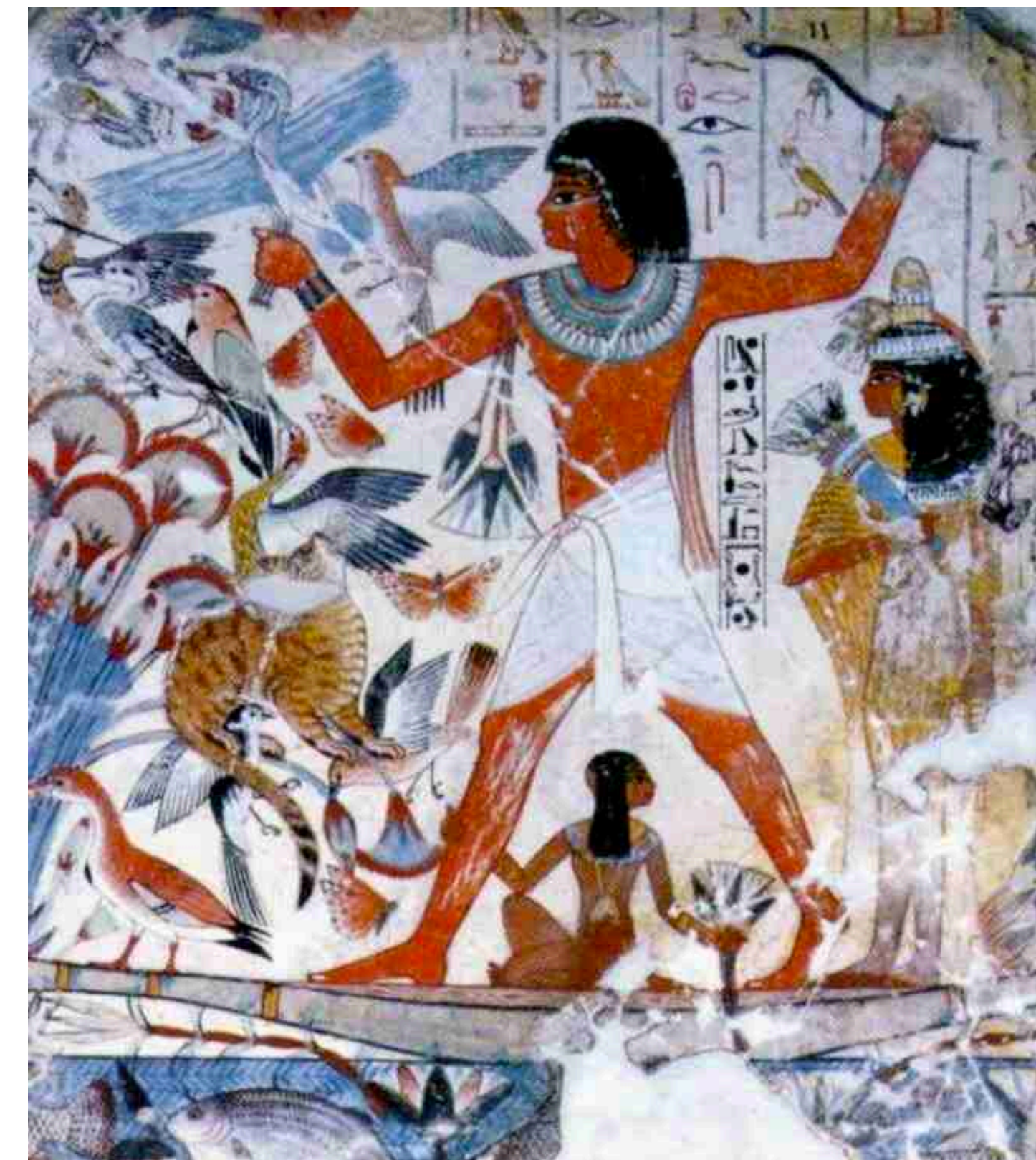


Visualization Objectives

- Record information
- Analyze data to support reasoning
- Confirm hypotheses
- Communicate ideas to others

Why Visualize?

To record information

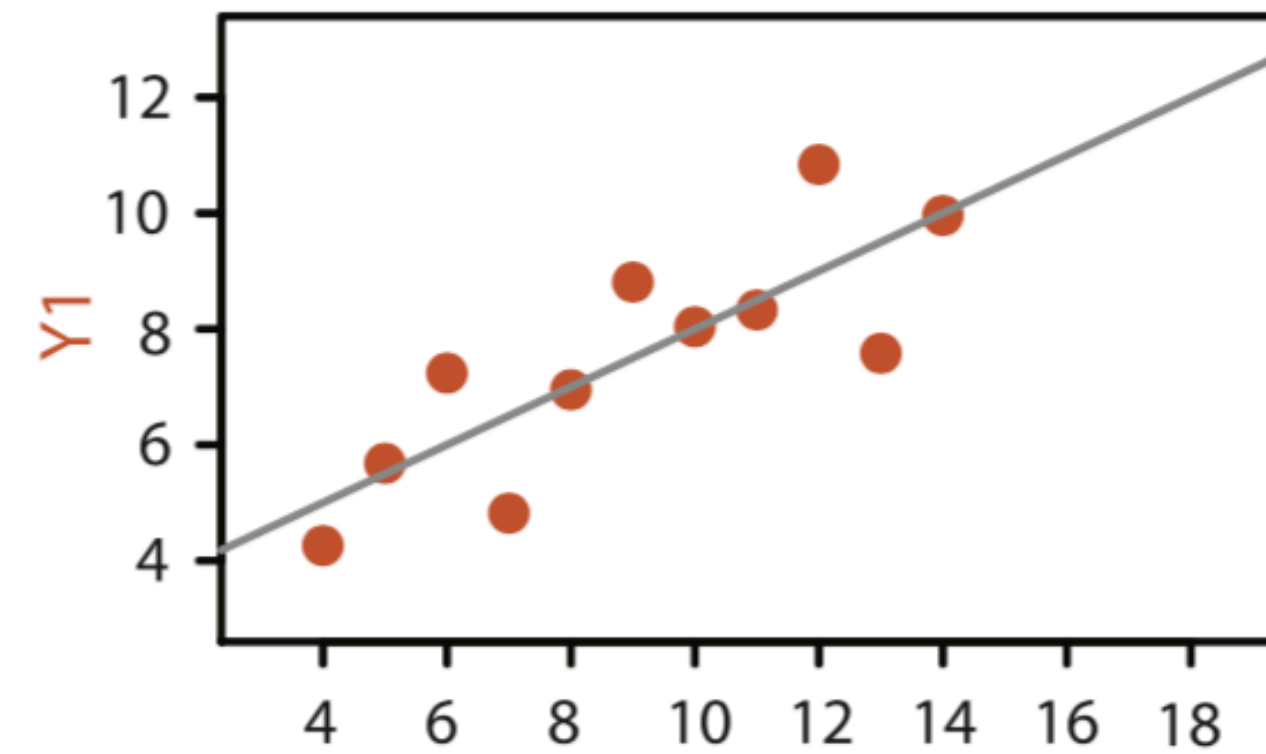


Why Visualize?

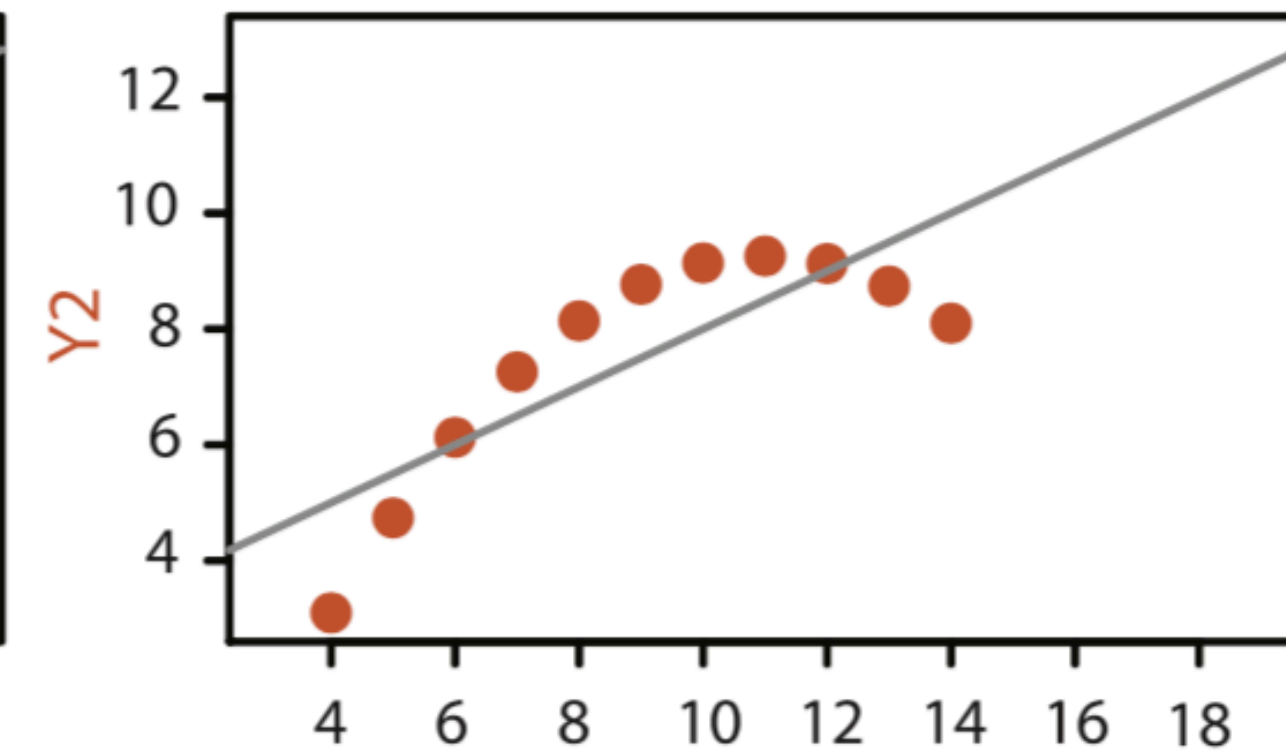
To show data in detail

Anscombe's Quartet: Raw Data

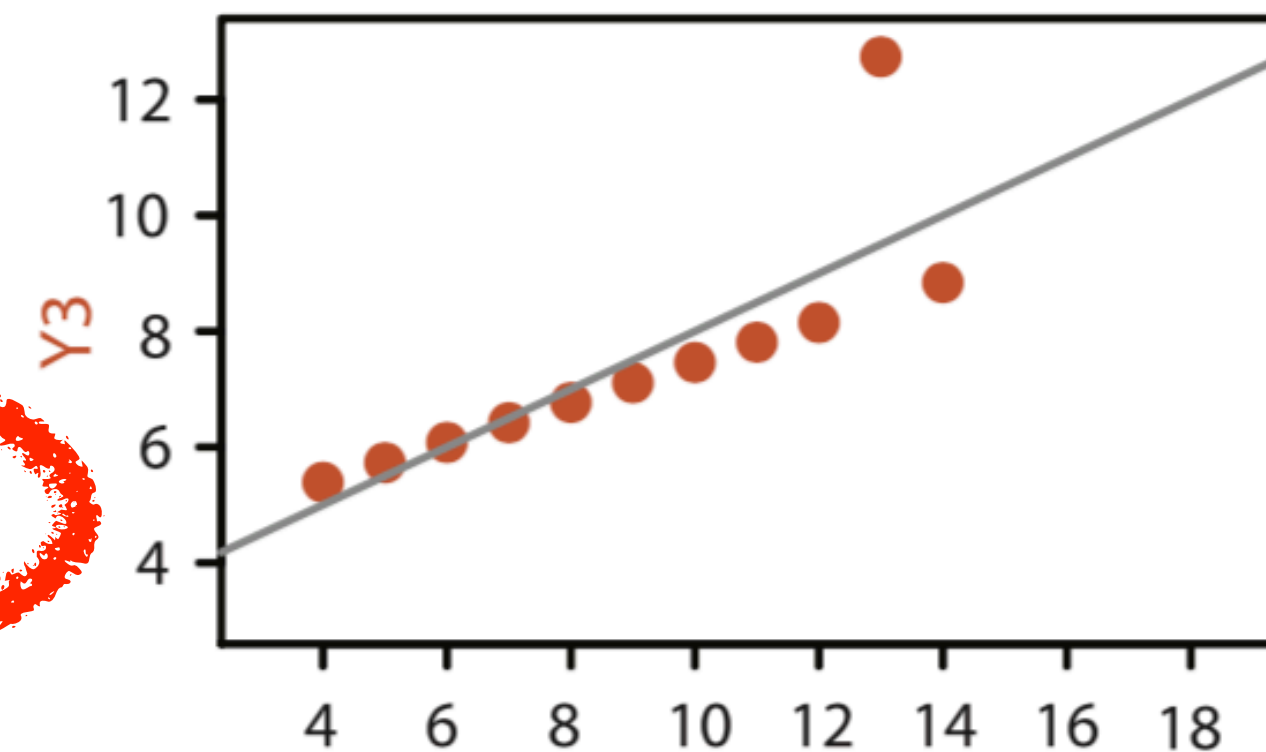
	1		2		3		4	
	X	Y	X	Y	X	Y	X	Y
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	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	5.55	5.0	4.74	5.0	5.73	8.0	6.89
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816	



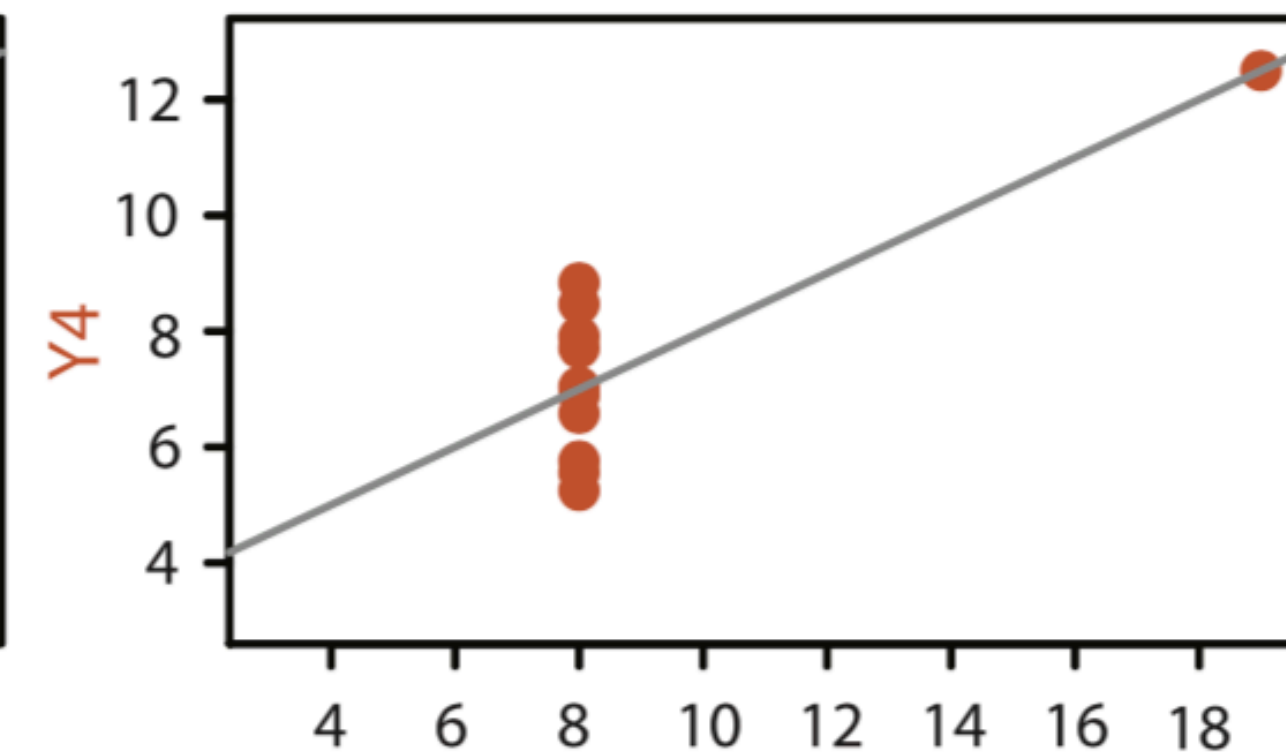
X1



X2



X3



X4
Anscombe, 1973



Why Visualize?

To point out interesting things

MTHIVLWYADCEQGHKILKMTWYN
ARDCAIREQGHVLMFPSTWYARN
GFPSVCEILQGKMFPSNDRCEQDIFP
SGHLMFHKMVPSTWYACEQTWRN

Why Visualize?

To point out interesting things

MTHI**V**LWYADCEQGHKILKMTWYN
ARDCAIREQGH**L**KMFPSTWYARN
GFPS**V**CEILQGKMFPNSNDRCEQDIFP
SGHLMFH**KM**VPSTWYACEQTWRN

Why Visualize?

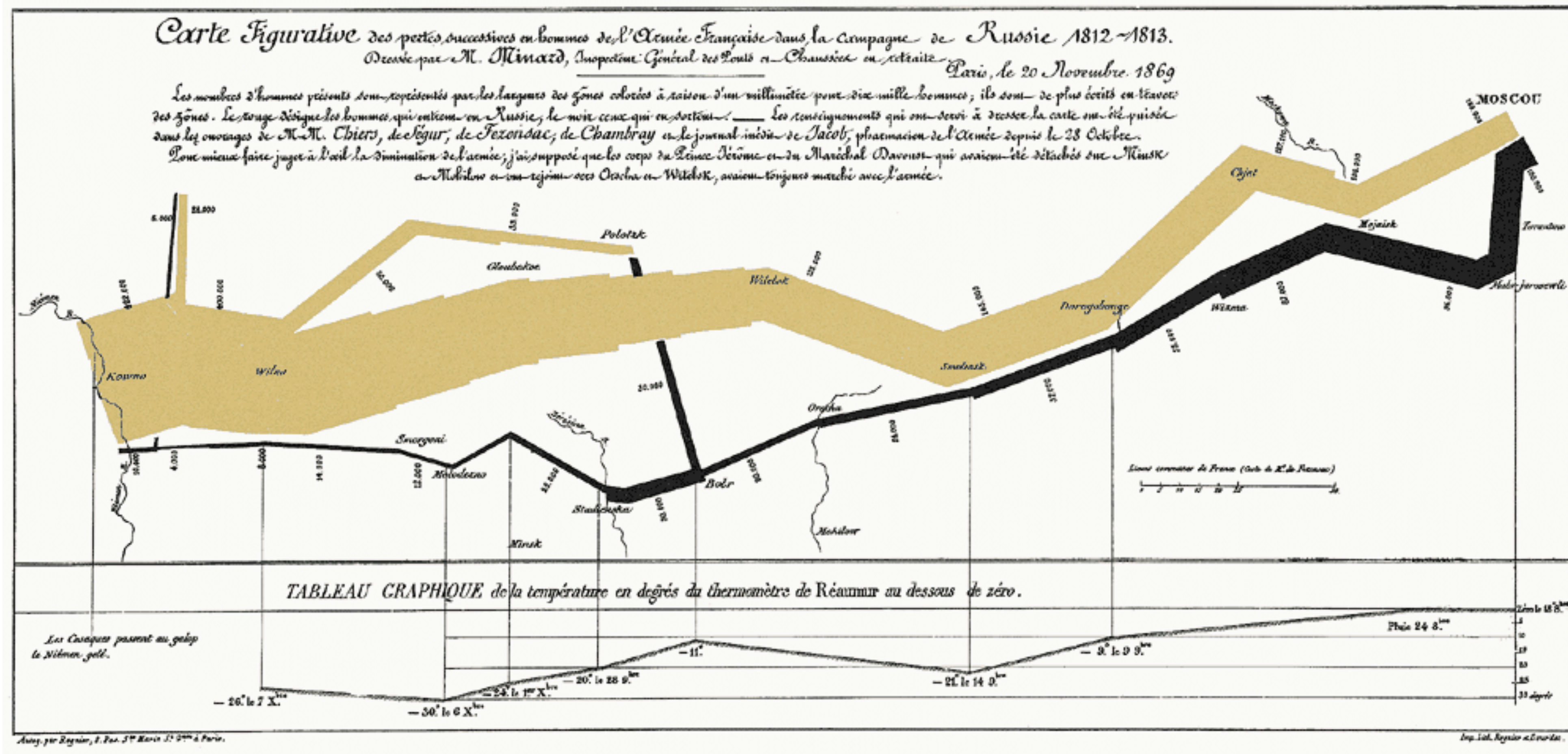
To communicate information



1812

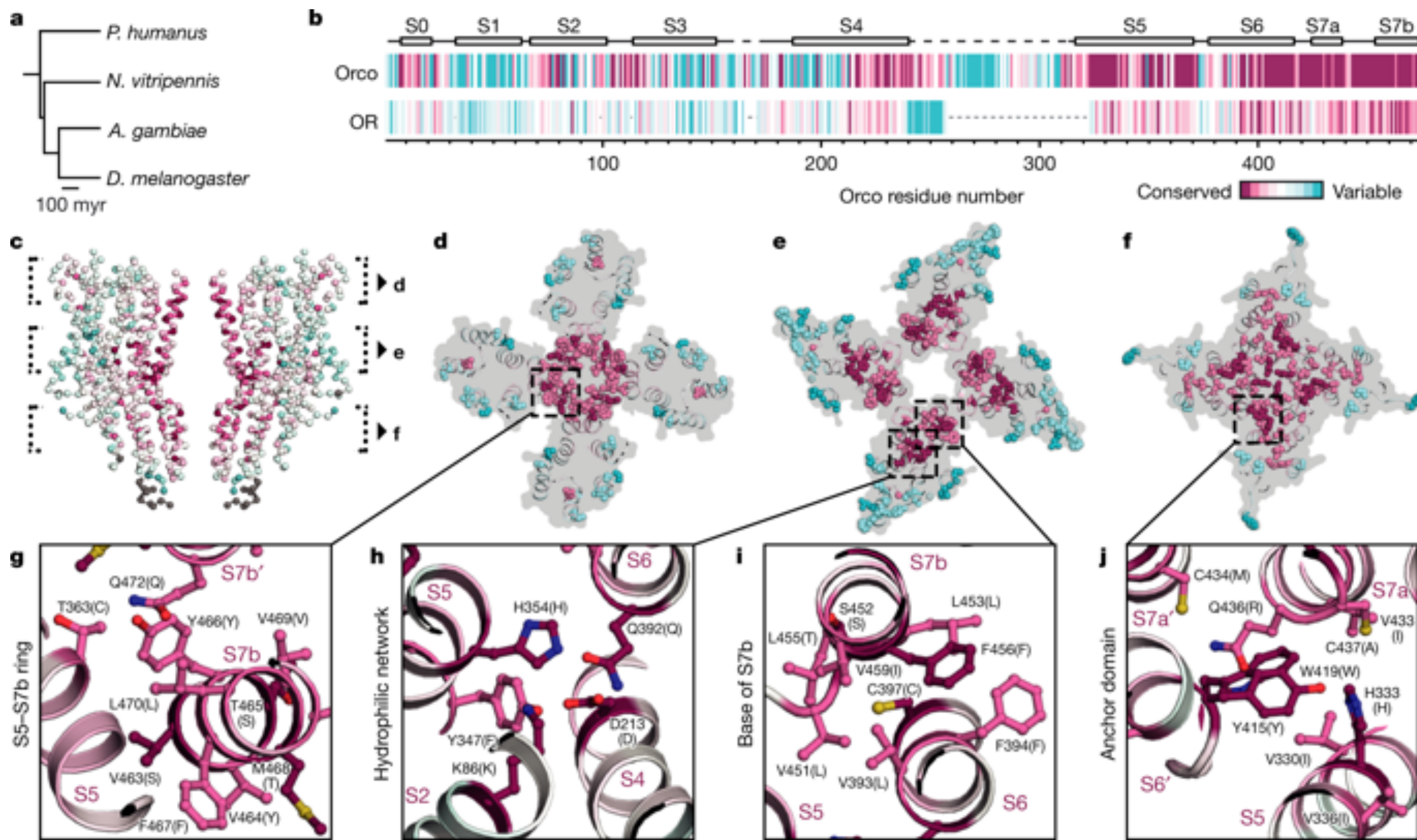
WHEN NAPOLEON
VENTURED EAST

HOW THE 1812 PATRIOTIC WAR TURNED NAPOLEON'S GRAND ARMY
INTO A HANDFUL OF SURVIVORS

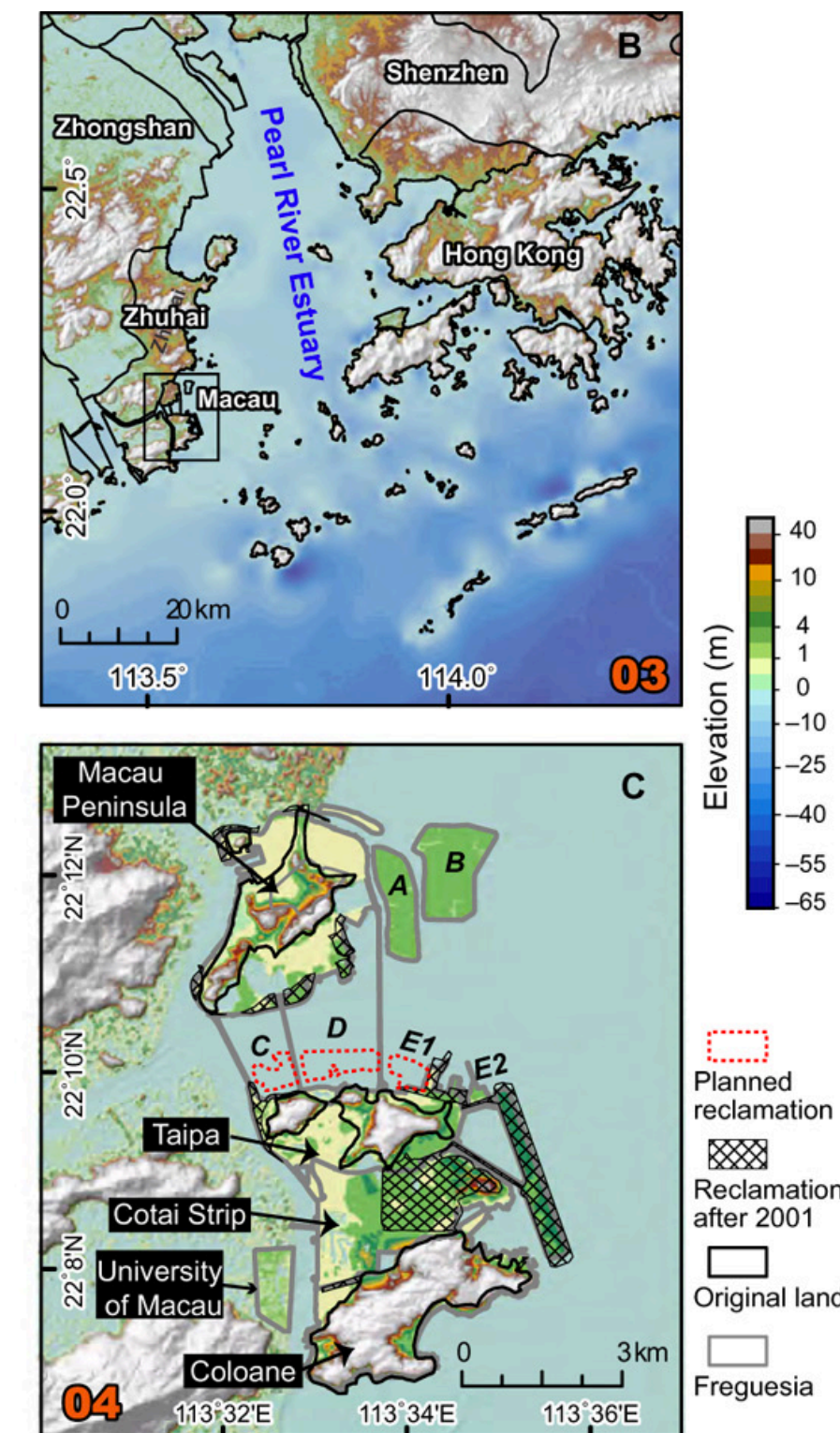


Why Visualize?

To communicate information



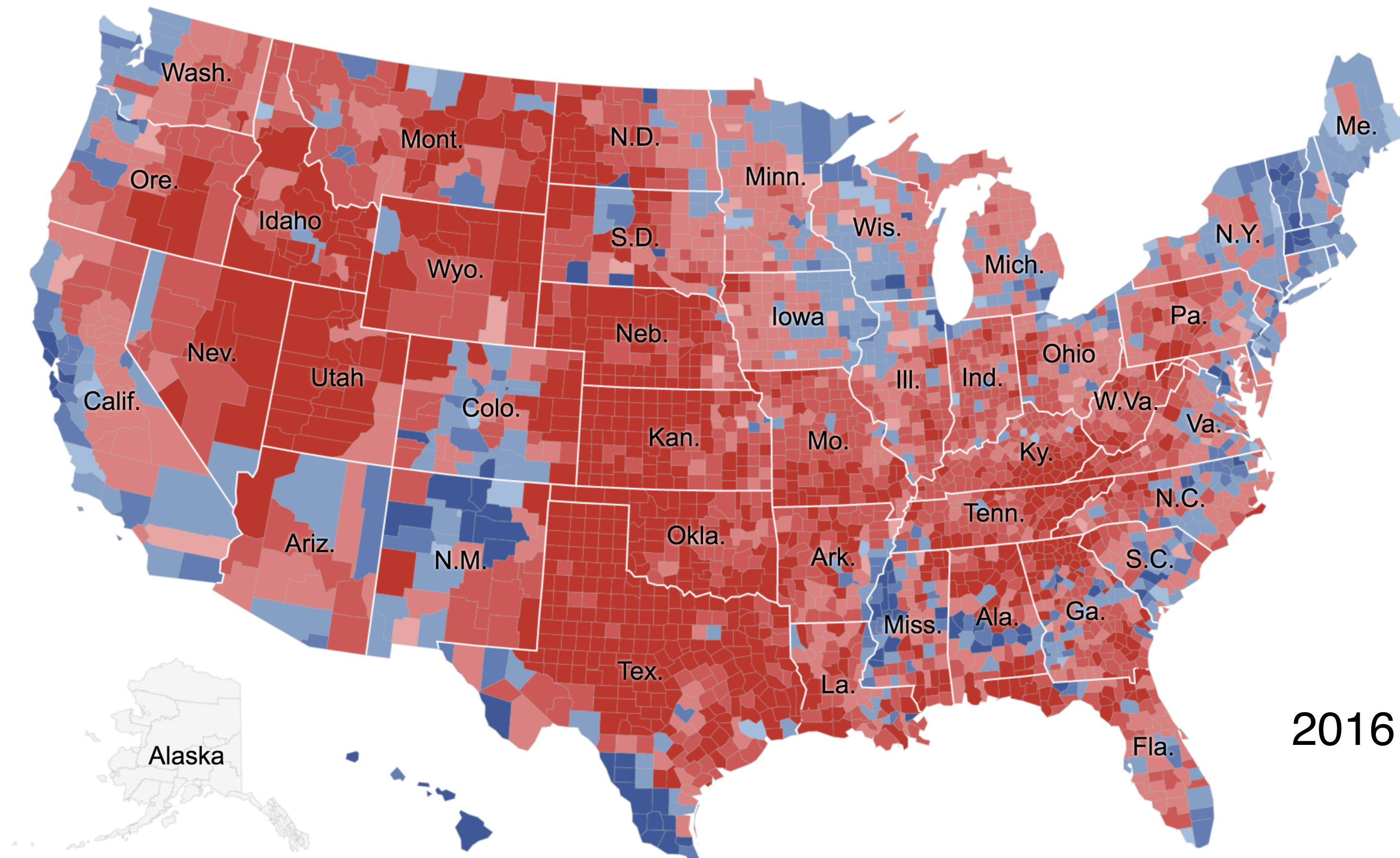
nature.com



sciencemag.com

Why Visualize?

- To analyze data



2016 US Presidential Election
NYTimes 2018

Search an address, Zip code or city

An Extremely Detailed Map of the 2016 Election

By **MATTHEW BLOCH, LARRY BUCHANAN, JOSH KATZ** and **KEVIN QUEALY** JULY 25, 2018

Share Tweet

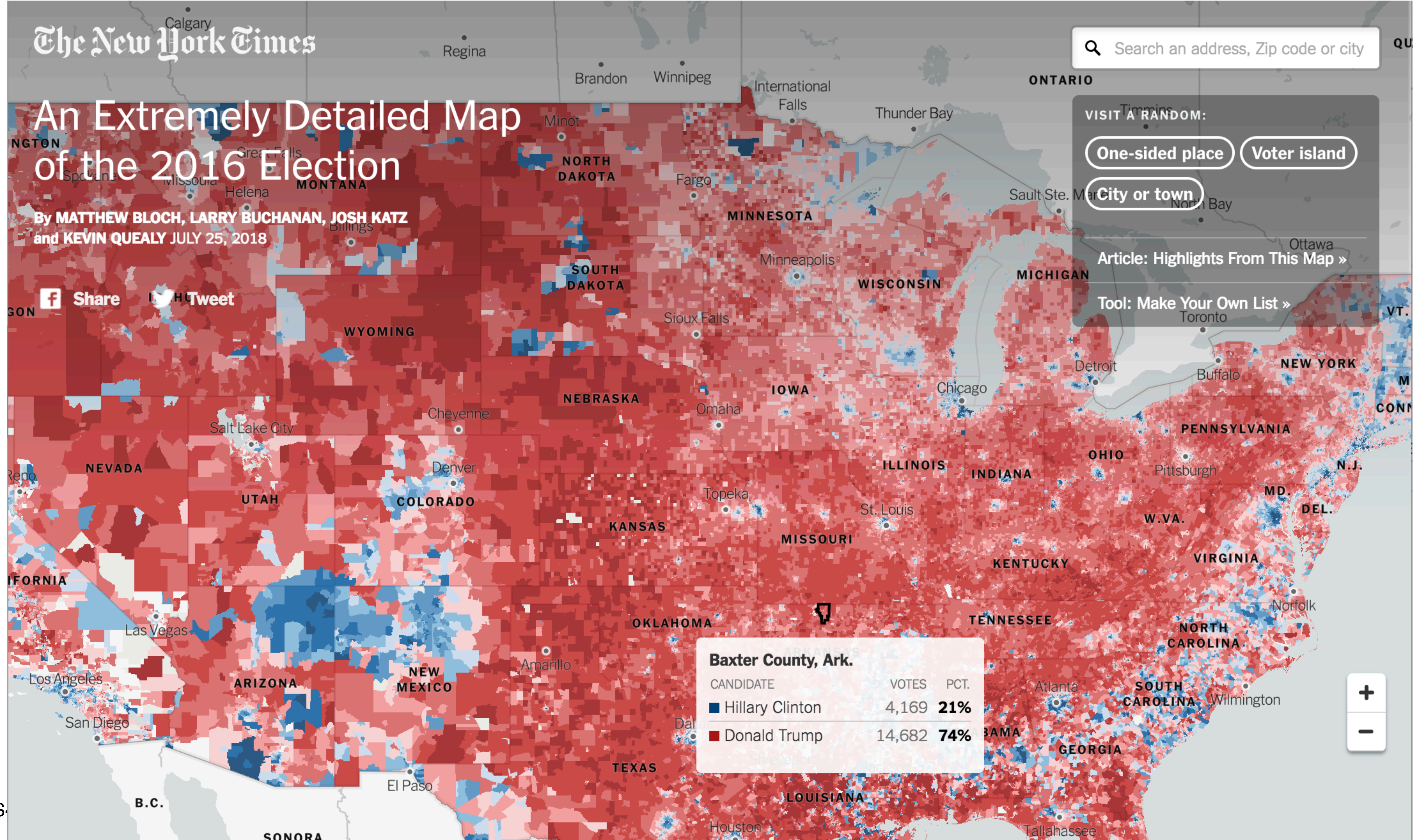
VISIT A RANDOM:

[One-sided place](#) [Voter island](#)

[City or town](#)

[Article: Highlights From This Map »](#)

[Tool: Make Your Own List »](#)



Baxter County, Ark.

CANDIDATE	VOTES	PCT.
■ Hillary Clinton	4,169	21%
■ Donald Trump	14,682	74%



An Extremely Detailed Map of the 2016 Election

By MATTHEW BLOCH, LARRY BUCHANAN, JOSH KATZ and KEVIN QUEALY JULY 25, 2018

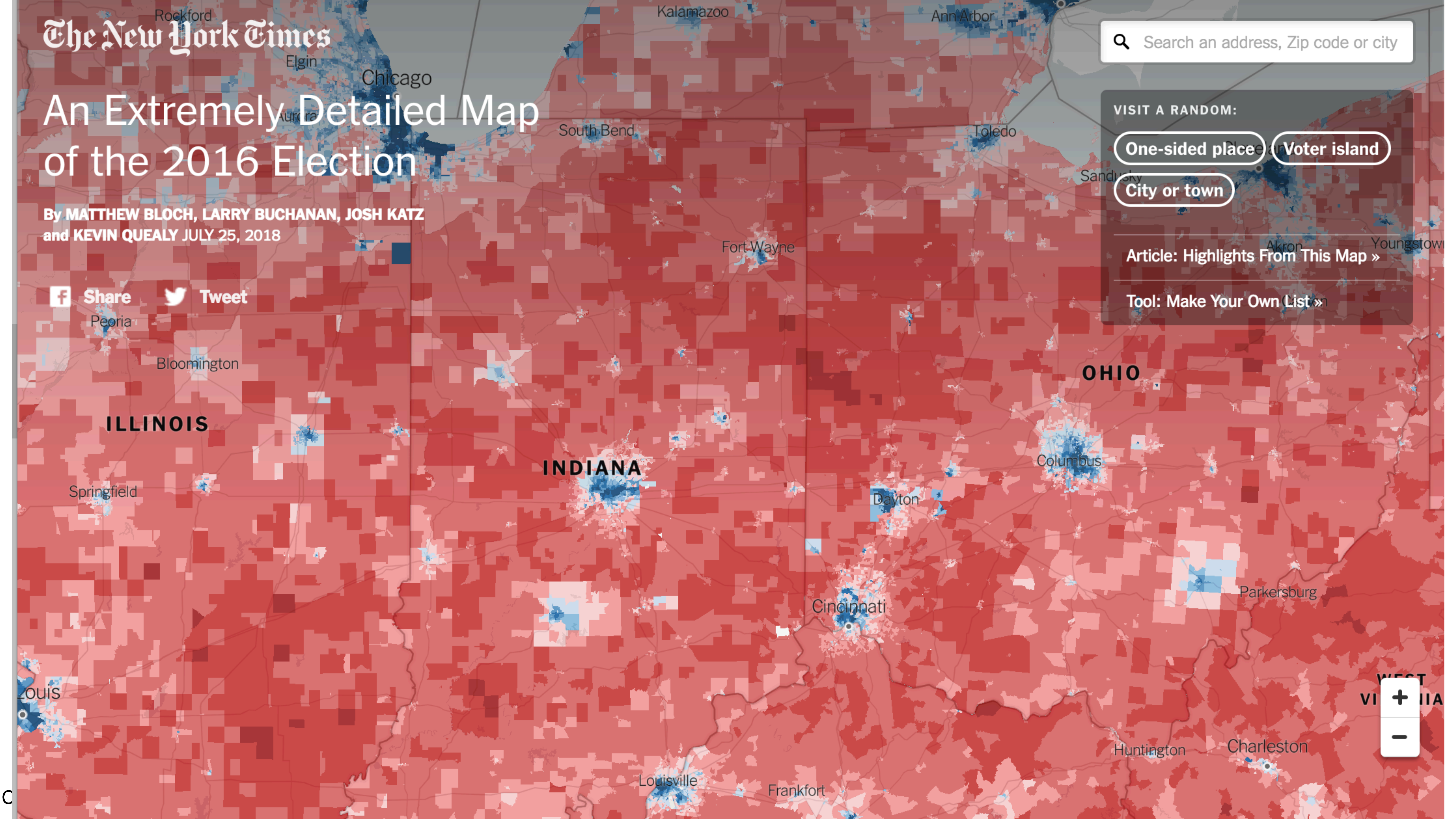
Share Tweet

VISIT A RANDOM:

- One-sided place
- Voter island
- City or town

Article: Highlights From This Map »

Tool: Make Your Own List »



An Extremely Detailed Map of the 2016 Election

By **MATTHEW BLOCH, LARRY BUCHANAN, JOSH KATZ**
and **KEVIN QUEALY** JULY 25, 2018

 Share  Tweet

 Search an address, Zip code or city

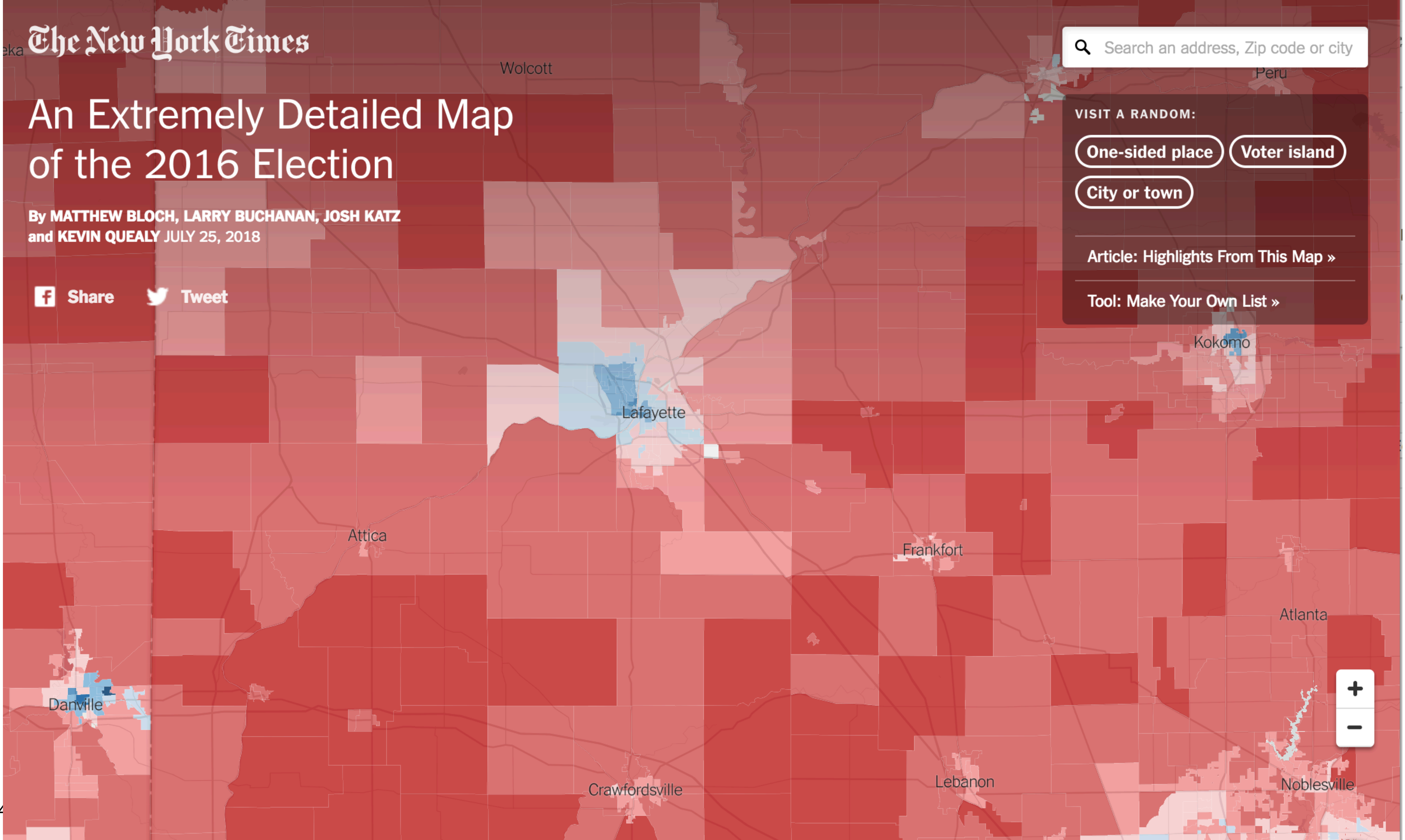
VISIT A RANDOM:

[One-sided place](#) [Voter island](#)

[City or town](#)

[Article: Highlights From This Map »](#)

[Tool: Make Your Own List »](#)



An Extremely Detailed Map of the 2016 Election

By **MATTHEW BLOCH, LARRY BUCHANAN, JOSH KATZ** and **KEVIN QUEALY** JULY 25, 2018

Share Tweet

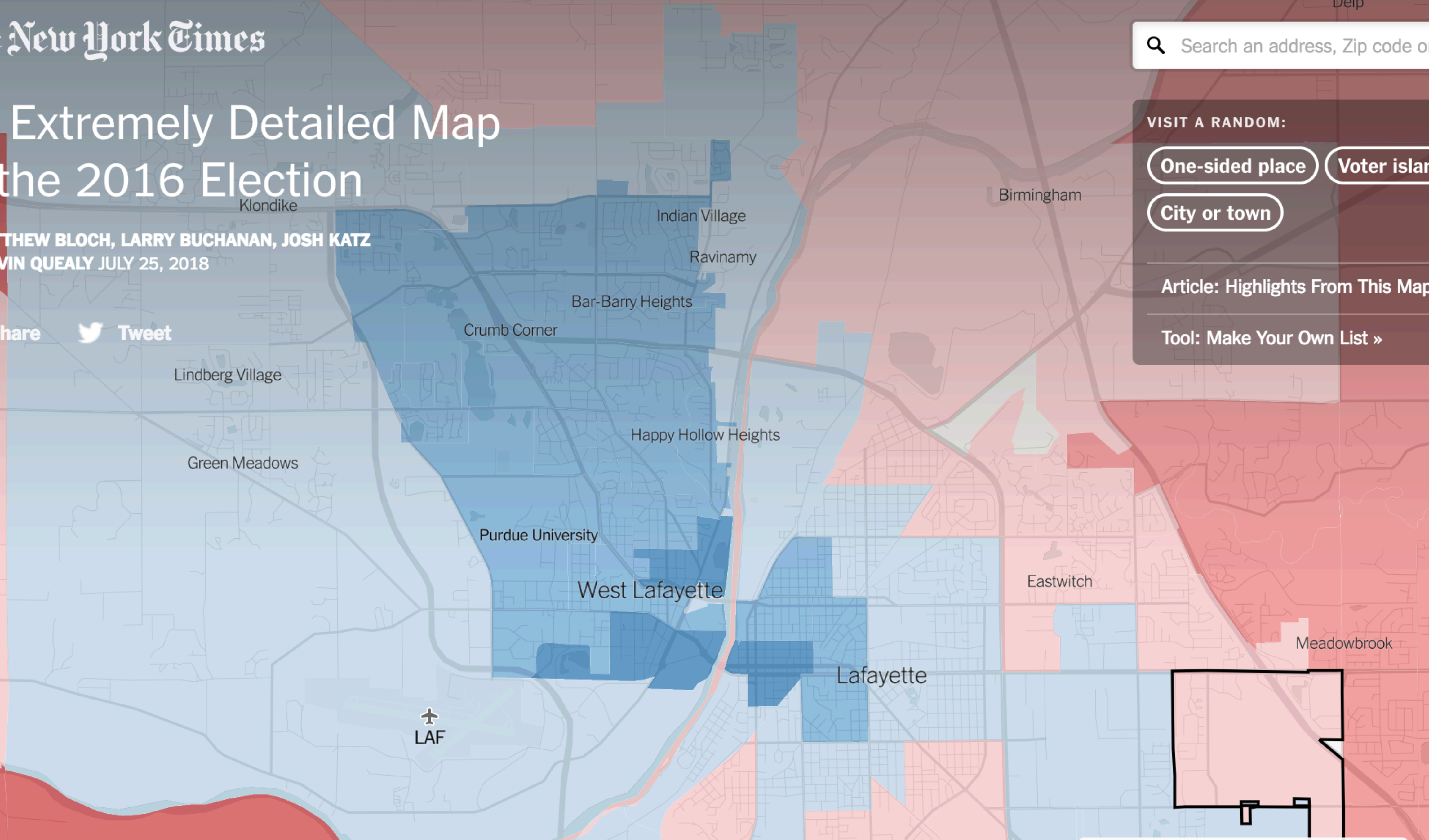
VISIT A RANDOM:

[One-sided place](#) [Voter island](#)

[City or town](#)

[Article: Highlights From This Map »](#)

[Tool: Make Your Own List »](#)

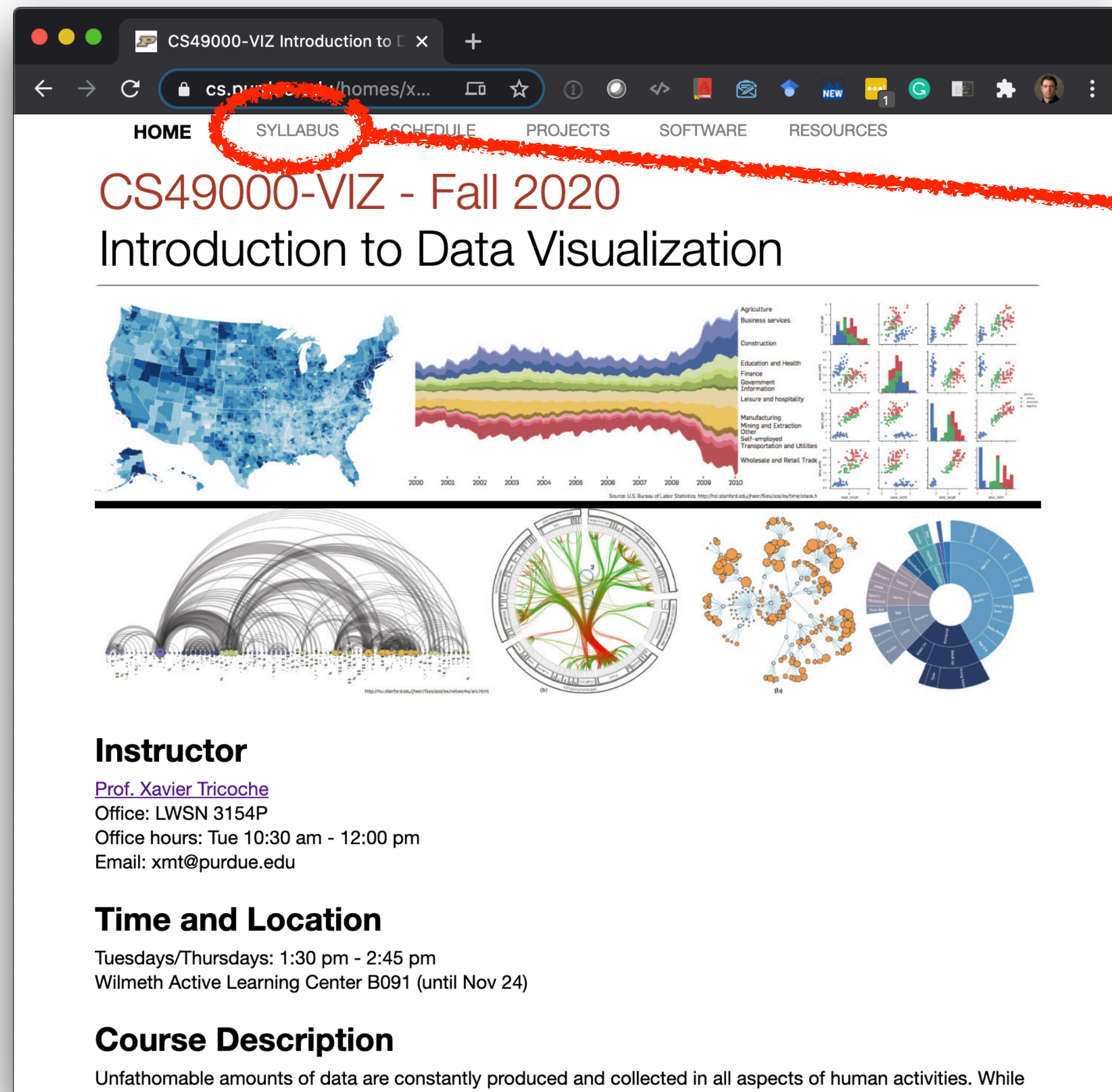


CANDIDATE	VOTES	PCT.
Clinton	101	50%
Trump	101	50%
Biden	101	50%



Syllabus

Web Page & Syllabus

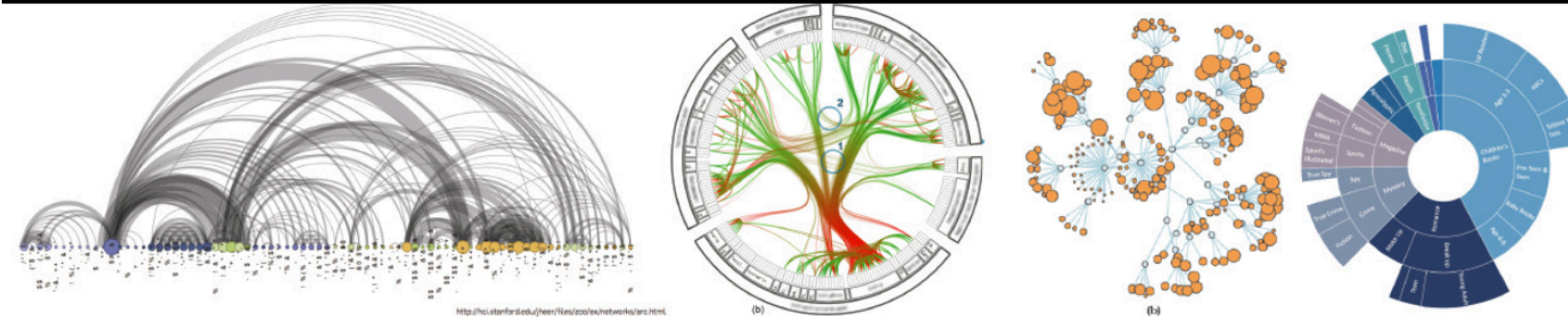
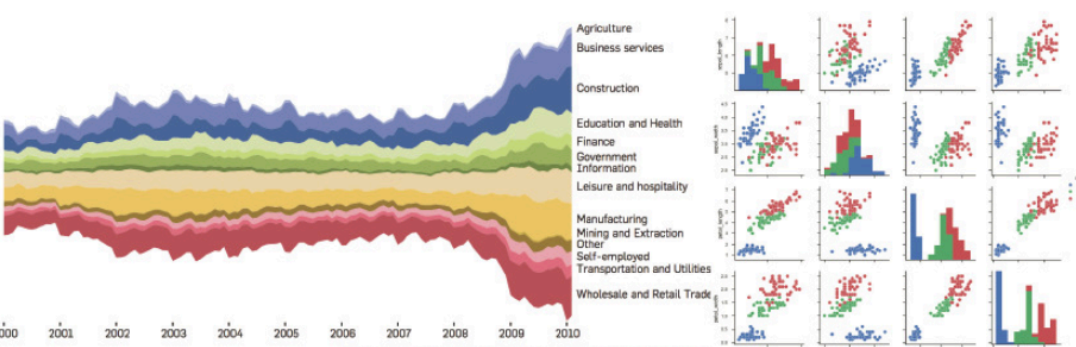
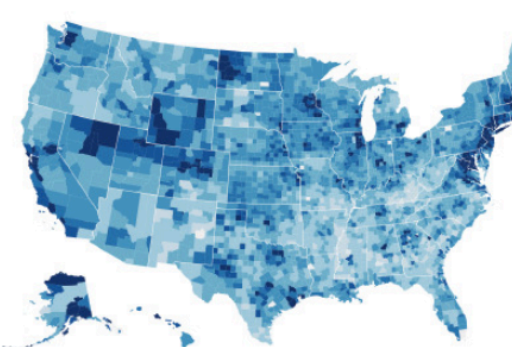


The screenshot shows the course homepage for CS49000-VIZ - Fall 2020. The navigation menu includes HOME, SYLLABUS, SCHEDULE, PROJECTS, SOFTWARE, and RESOURCES. The SYLLABUS link is circled in red, with a red arrow pointing to the syllabus page. The main content area features a title, a map of the United States, a horizontal bar chart, and several network diagrams.

HOME SYLLABUS SCHEDULE PROJECTS SOFTWARE RESOURCES

CS49000-VIZ - Fall 2020

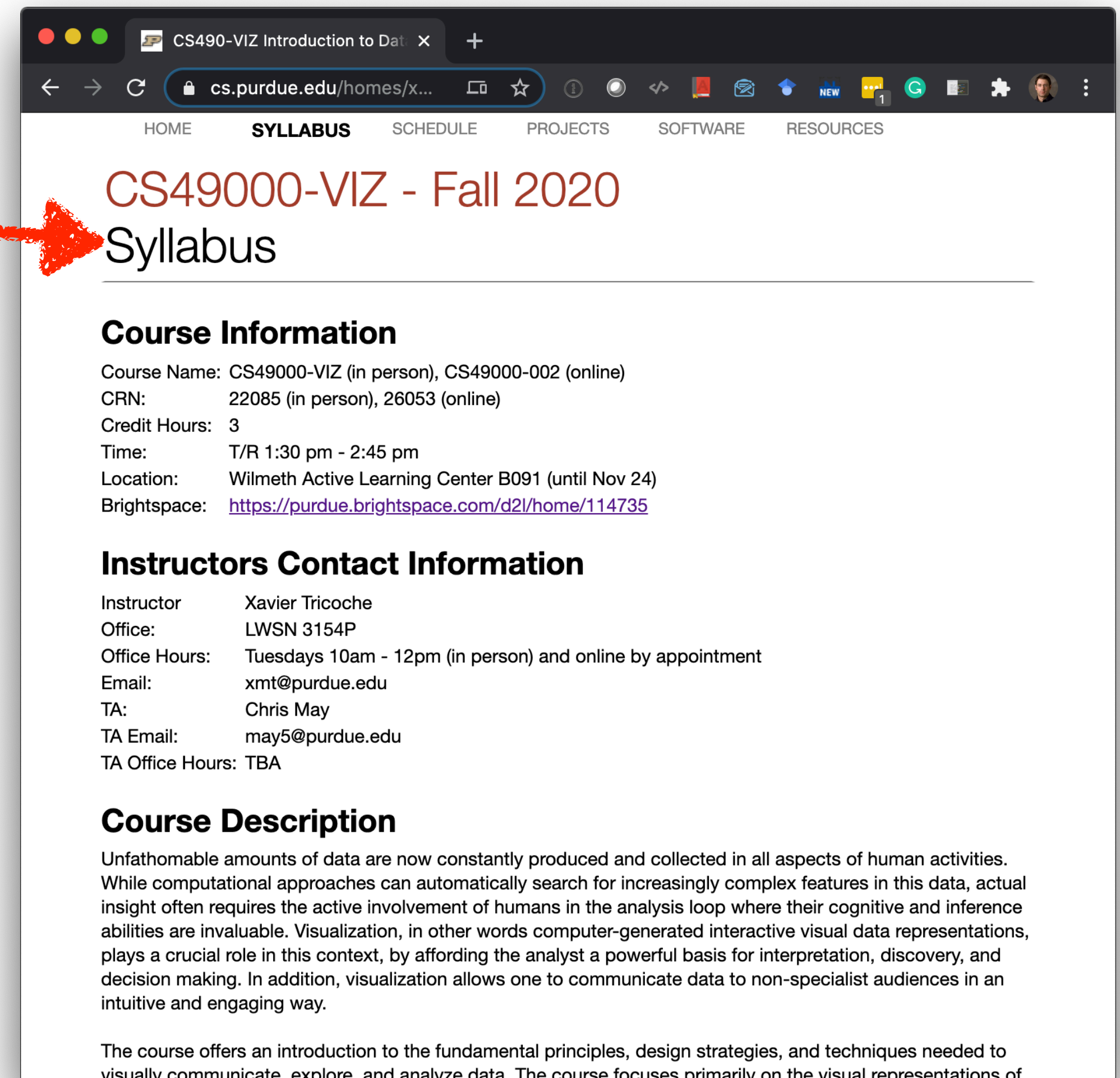
Introduction to Data Visualization



Instructor
[Prof. Xavier Tricoche](#)
Office: LWSN 3154P
Office hours: Tue 10:30 am - 12:00 pm
Email: xmt@purdue.edu

Time and Location
Tuesdays/Thursdays: 1:30 pm - 2:45 pm
Wilmeth Active Learning Center B091 (until Nov 24)

Course Description
Unfathomable amounts of data are constantly produced and collected in all aspects of human activities. While



The screenshot shows the course syllabus page for CS49000-VIZ - Fall 2020. The navigation menu is the same as the homepage. The main content area includes a title, a 'Course Information' section with details like course name, CRN, and credit hours, an 'Instructors Contact Information' section with instructor and TA details, and a 'Course Description' section.

HOME **SYLLABUS** SCHEDULE PROJECTS SOFTWARE RESOURCES

CS49000-VIZ - Fall 2020

Syllabus

Course Information

Course Name: CS49000-VIZ (in person), CS49000-002 (online)
CRN: 22085 (in person), 26053 (online)
Credit Hours: 3
Time: T/R 1:30 pm - 2:45 pm
Location: Wilmeth Active Learning Center B091 (until Nov 24)
Brightspace: <https://purdue.brightspace.com/d2l/home/114735>

Instructors Contact Information

Instructor: Xavier Tricoche
Office: LWSN 3154P
Office Hours: Tuesdays 10am - 12pm (in person) and online by appointment
Email: xmt@purdue.edu
TA: Chris May
TA Email: may5@purdue.edu
TA Office Hours: TBA

Course Description

Unfathomable amounts of data are now constantly produced and collected in all aspects of human activities. While computational approaches can automatically search for increasingly complex features in this data, actual insight often requires the active involvement of humans in the analysis loop where their cognitive and inference abilities are invaluable. Visualization, in other words computer-generated interactive visual data representations, plays a crucial role in this context, by affording the analyst a powerful basis for interpretation, discovery, and decision making. In addition, visualization allows one to communicate data to non-specialist audiences in an intuitive and engaging way.

The course offers an introduction to the fundamental principles, design strategies, and techniques needed to visually communicate, explore, and analyze data. The course focuses primarily on the visual representations of

<https://www.cs.purdue.edu/homes/xmt/classes/CS490-VIZ-Fall2020>

Basic Facts

- Lectures: T/Th 1:30 pm - 2:45 pm
- Office hours: T 10 am - 12 pm
- Performance evaluation:
 - 5 (programming) assignments (total: **60%**)
 - Midterm exam: **20%**
 - Final project: **20%**

Learning Objectives

1. Know **main visualization techniques** and understand their use cases
2. Discern **strengths and weaknesses** of existing visualizations
3. **Design effective visualizations** in a principled and structured fashion
4. **Program interactive visualizations** using open source software

Textbook

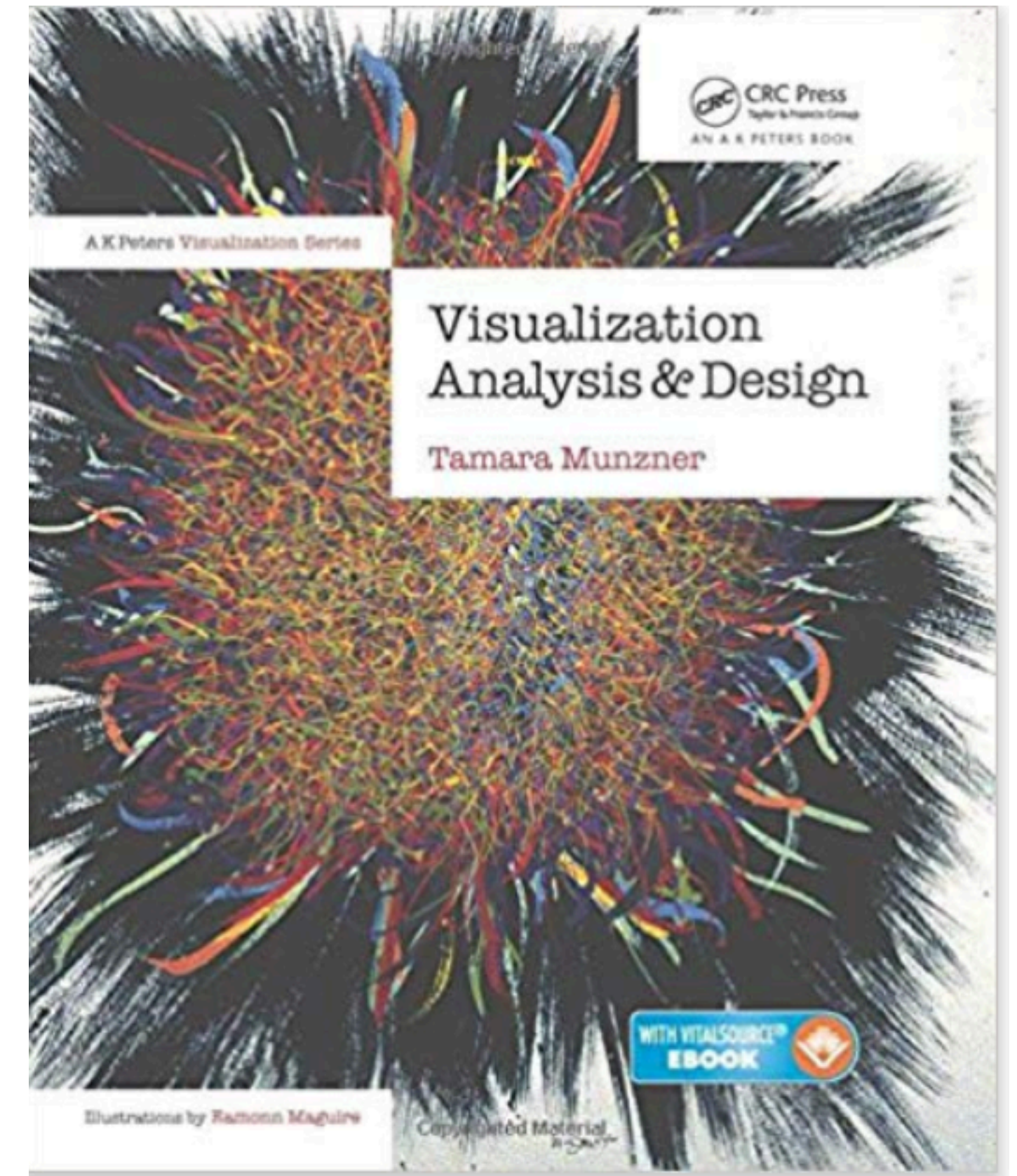
Visualization Analysis and Design

by **Tamara Munzner**

AK Peters Visualization Series

CRC Press, Nov. 2014.

Required



Communication

- Discussions and Q&A's will be on Piazza




<https://piazza.com/purdue/fall2020/cs49000viz/home>

- Sign up!



Software

- Python libraries
 - **matplotlib.pyplot:**
 - standard general-purpose data visualization library
 - fairly low-level plotting control
 - **seaborn:**
 - high-level interface to matplotlib
 - primarily meant for statistical data visualization

Software

- Python Libraries
 -  **bokeh**:
 - high-level visualization library
 - creates javascript code: visualization runs on web browsers

Software

-  (Data Driven Document):
 - javascript library for web visualization
 - non-trivial to use
-  **+ a b | e a u**[®]: widely adopted commercial visualization tool (*no programming*)
 - will be used for Project 5

Lateness Policy

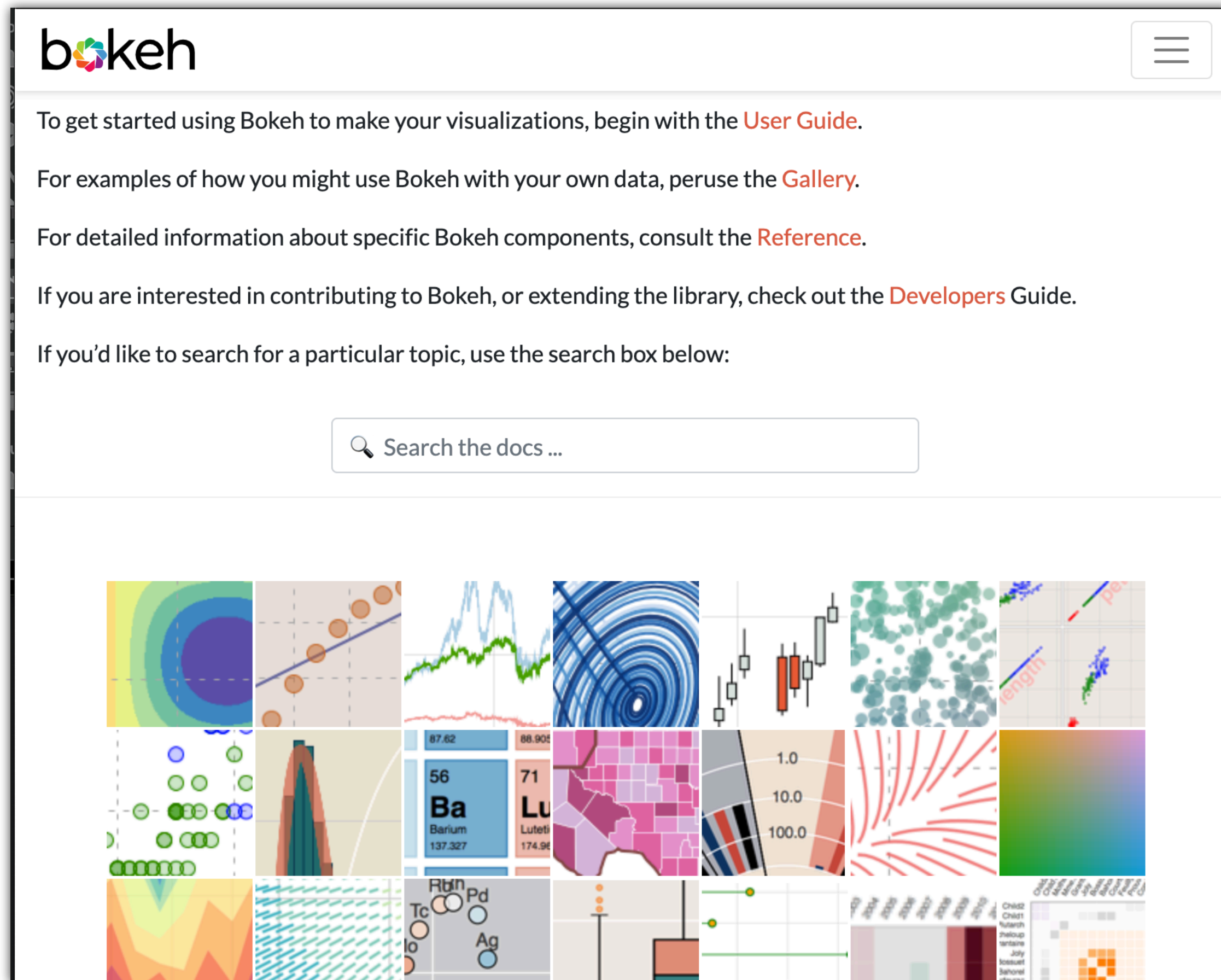
- Assignments due by given deadline
- 5 extension days granted to each student
- 20% penalty for each late day
- No submission accepted past 3rd late day

Final Project

- Individually or in pair of students
- Roadmap:
 1. Select dataset
 2. Identify question(s) to answer or task(s) to perform with visualization
 3. Sketch a visualization solution
 4. Final presentation with live demo during dead week.

Topics Overview

This week: Visualization Libraries



The screenshot shows the Bokeh website's landing page. At the top left is the Bokeh logo. Below it, there are four lines of introductory text with links to the User Guide, Gallery, Reference, and Developers Guide. A search box is provided for finding specific topics. At the bottom, there is a large grid of various data visualization examples, including contour plots, scatter plots, line graphs, and heatmaps.

bokeh

To get started using Bokeh to make your visualizations, begin with the [User Guide](#).

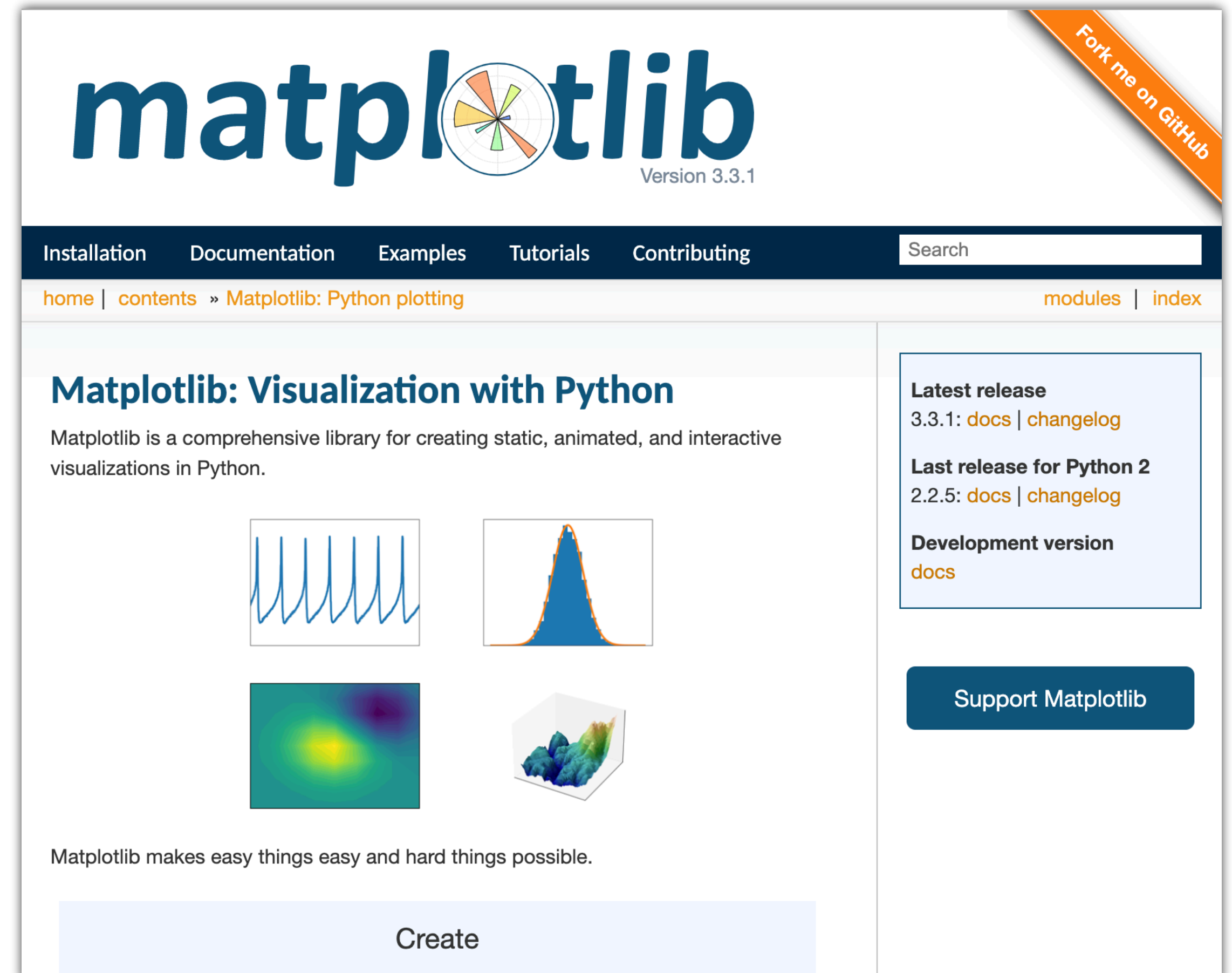
For examples of how you might use Bokeh with your own data, peruse the [Gallery](#).

For detailed information about specific Bokeh components, consult the [Reference](#).

If you are interested in contributing to Bokeh, or extending the library, check out the [Developers Guide](#).

If you'd like to search for a particular topic, use the search box below:

Search the docs ...



The screenshot shows the Matplotlib website's landing page. At the top is the Matplotlib logo with the version number 3.3.1. A navigation bar includes links for Installation, Documentation, Examples, Tutorials, and Contributing, along with a search box. Below the navigation bar, there are links for home, contents, and Matplotlib: Python plotting. The main content area features the title 'Matplotlib: Visualization with Python' and a brief description. There are four small images showing different types of plots: a line plot, a histogram, a heatmap, and a 3D surface plot. A sidebar on the right contains information about the latest release (3.3.1), the last release for Python 2 (2.2.5), and the development version. A 'Support Matplotlib' button is also present. At the bottom, there is a large 'Create' button.

matplotlib Version 3.3.1

Fork me on GitHub

Installation Documentation Examples Tutorials Contributing Search

home | contents » Matplotlib: Python plotting modules | index

Matplotlib: Visualization with Python

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.

Latest release
3.3.1: [docs](#) | [changelog](#)

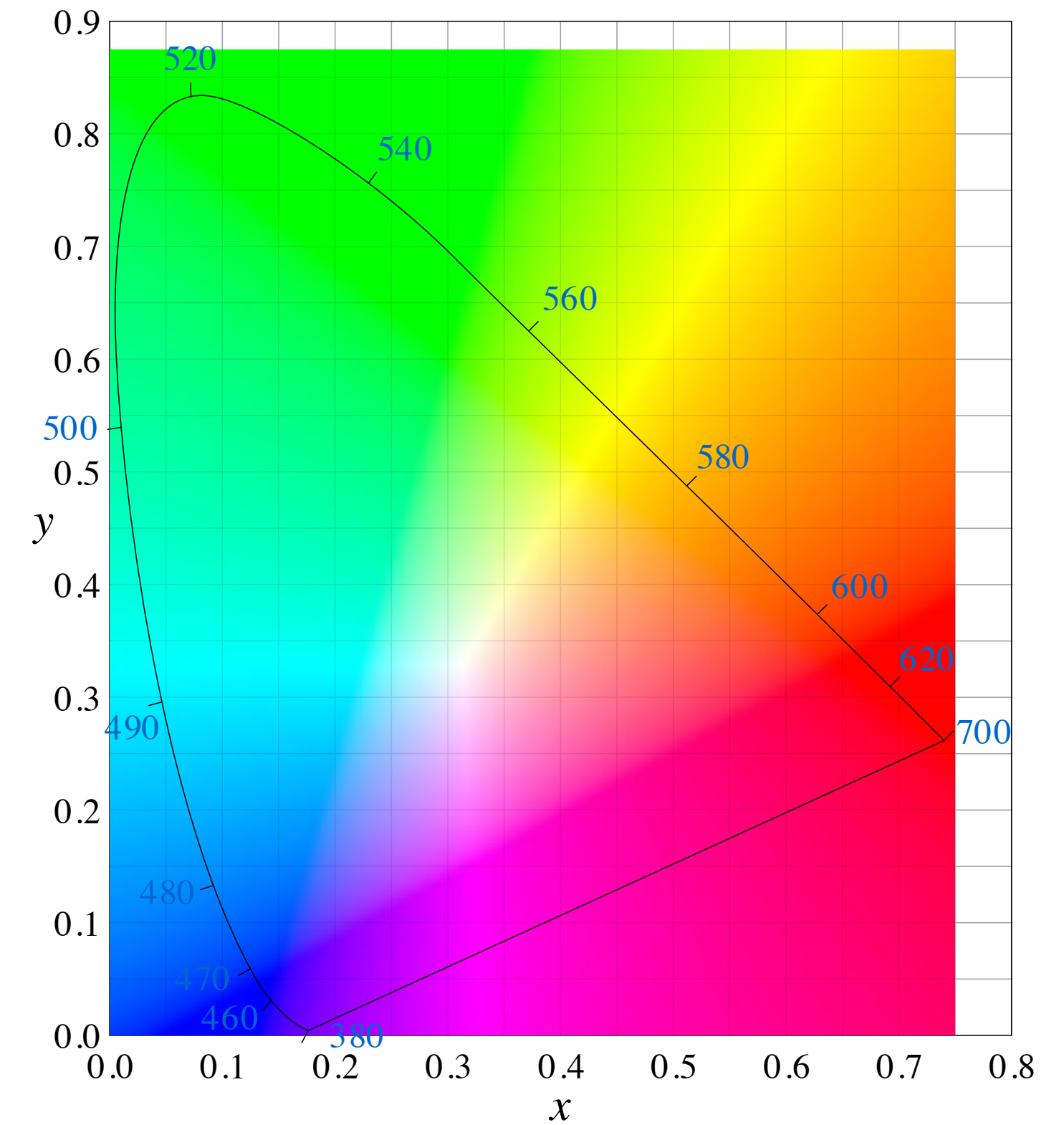
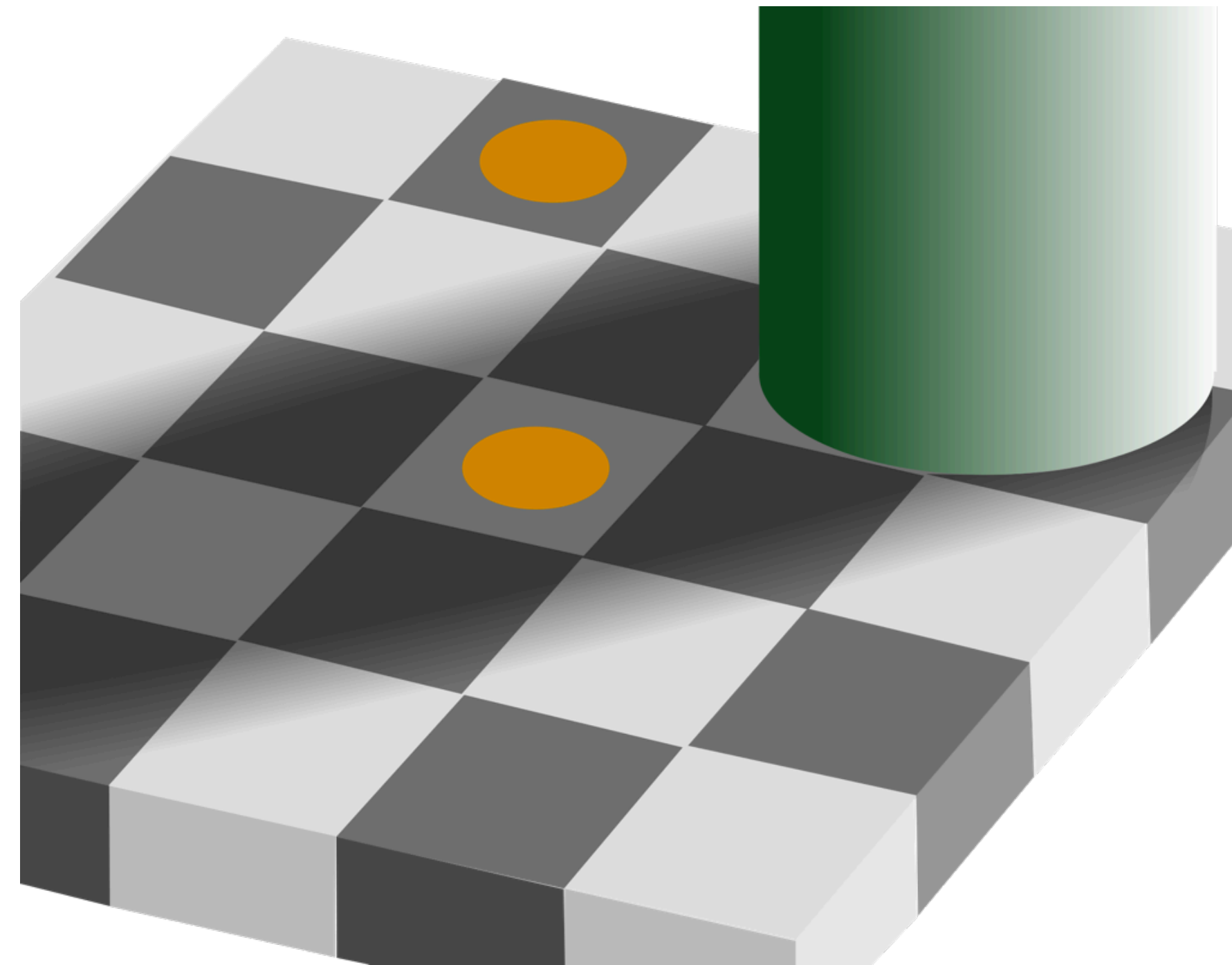
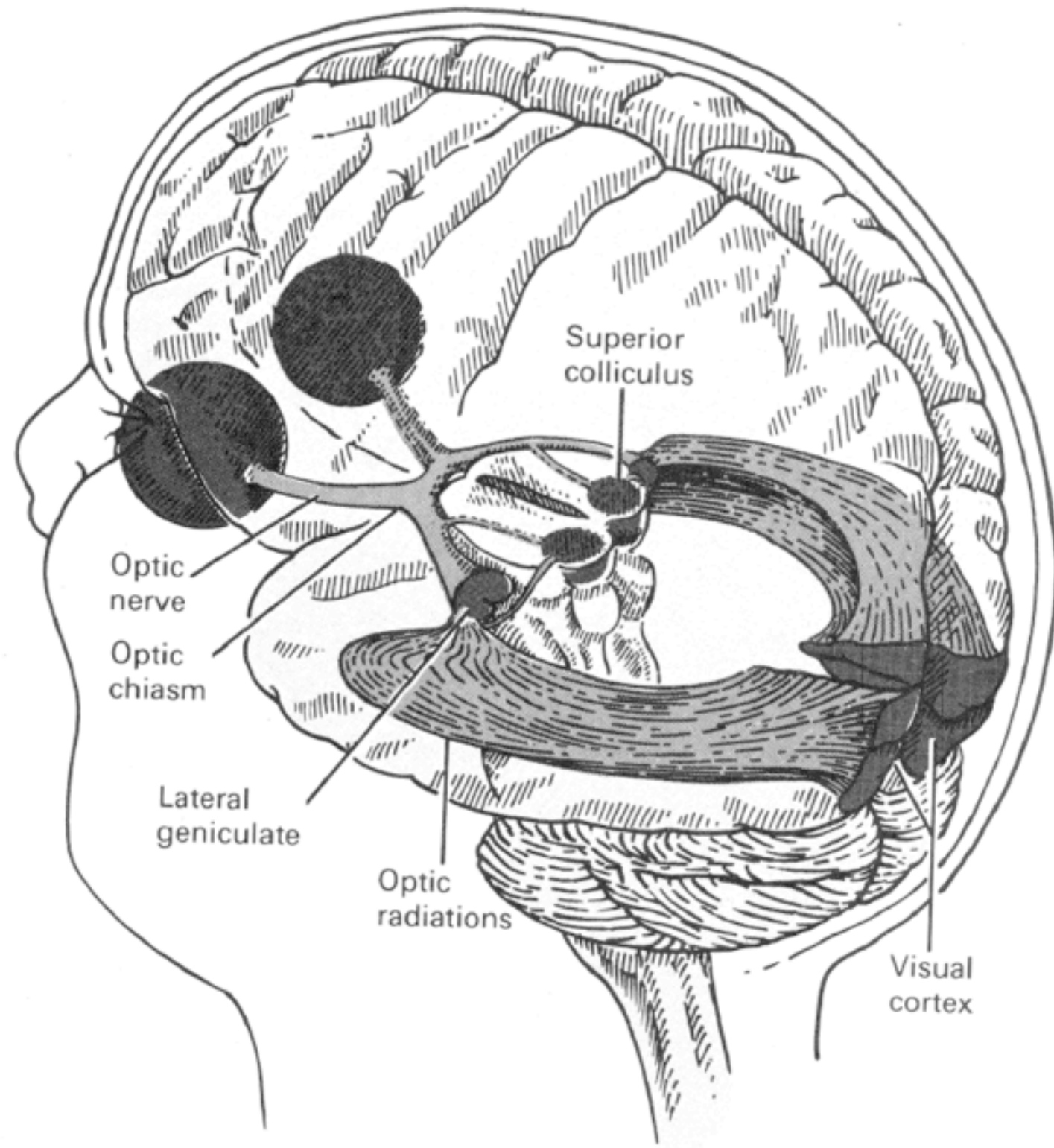
Last release for Python 2
2.2.5: [docs](#) | [changelog](#)

Development version
[docs](#)

Support Matplotlib

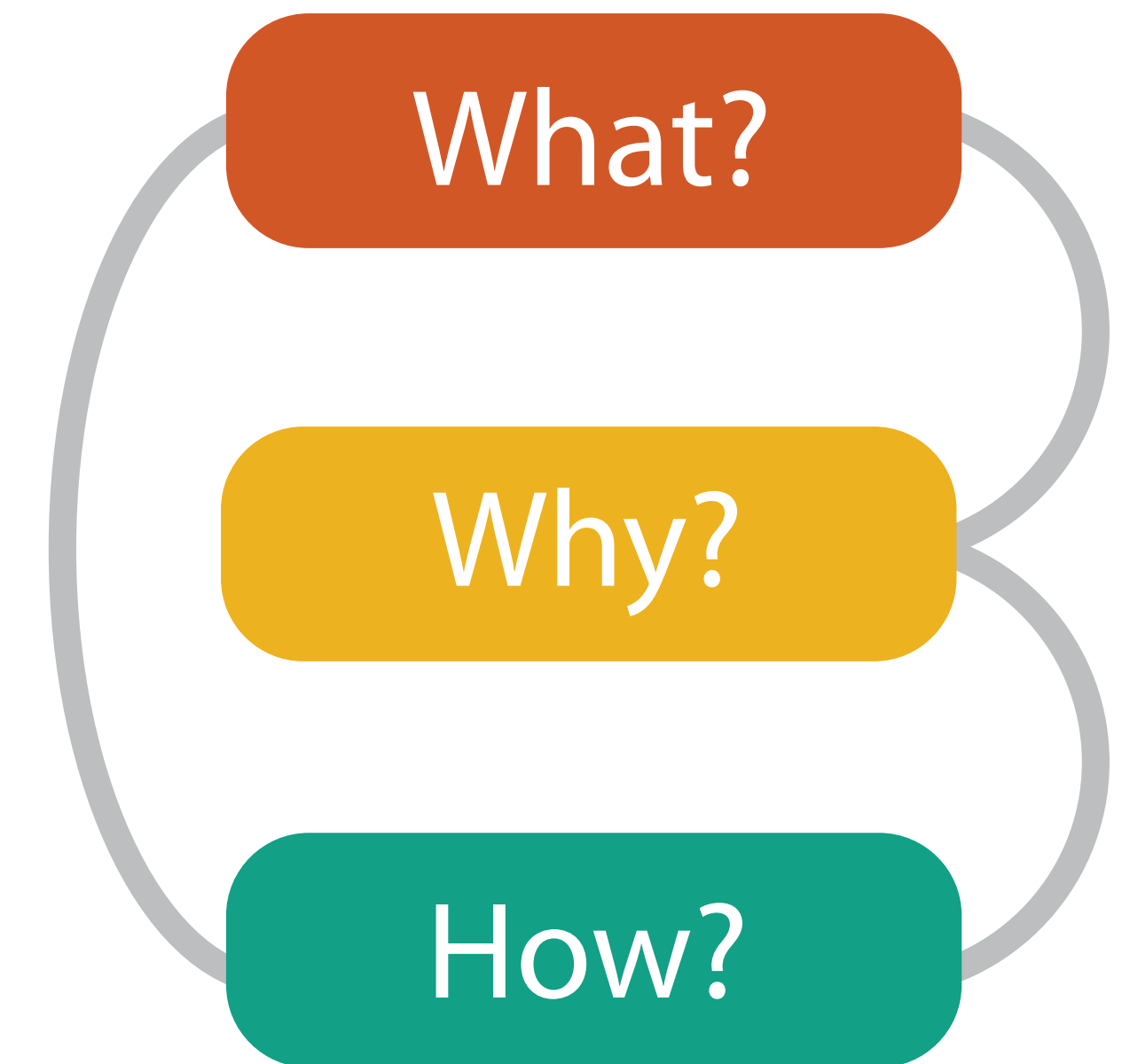
Create

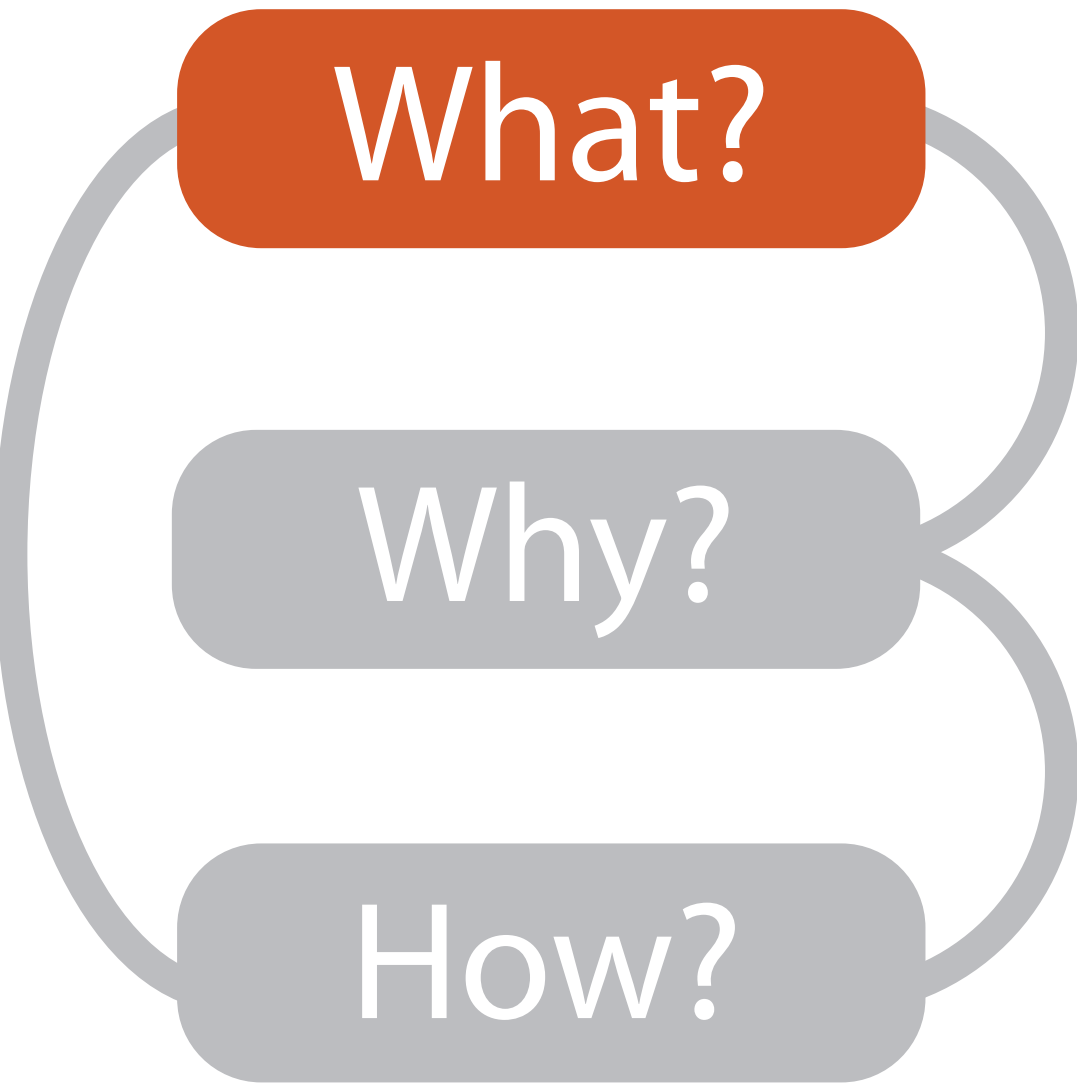
Perception & Color Vision



Methodology

- **what** is shown?
 - **data** abstraction
- **why** is the user looking at it?
 - **task** abstraction
- **how** is it shown?
 - **idiom**: visual encoding and interaction





What?

Datasets

Attributes

- ➔ Data Types
 - ➔ Items
 - ➔ Attributes
 - ➔ Links
 - ➔ Positions
 - ➔ Grids

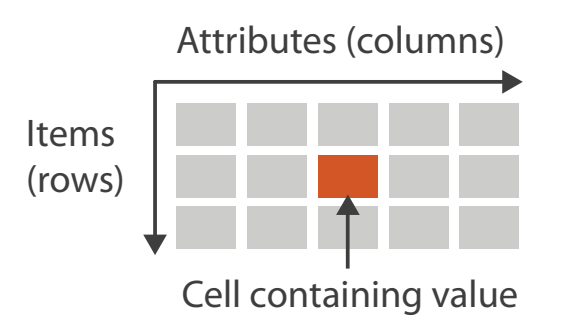
- ➔ Attribute Types
 - ➔ Categorical
 - + ● ■ ▲
 - ➔ Ordered
 - ➔ Ordinal
 - 👕 👕 👕
 - ➔ Quantitative
 - — —

Data and Dataset Types

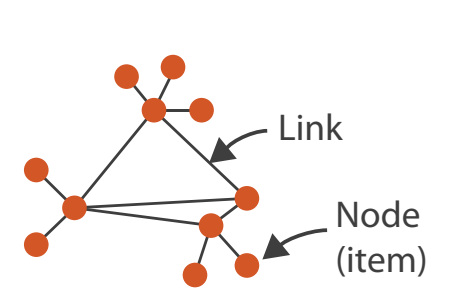
Tables	Networks & Trees	Fields	Geometry	Clusters, sets, lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Dataset Types

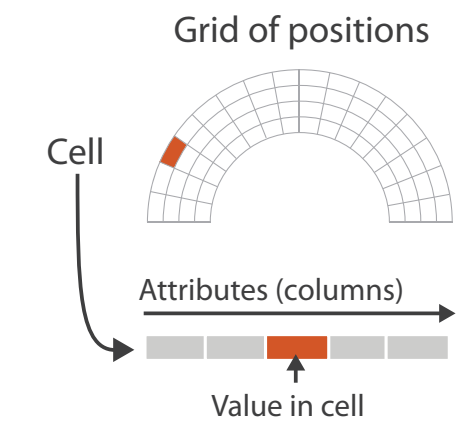
➔ Tables



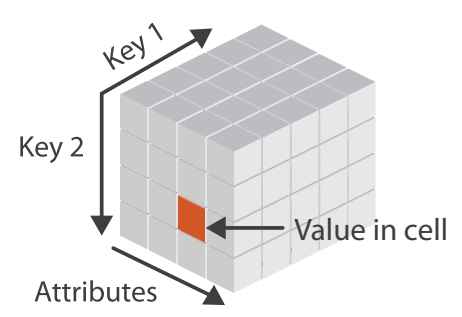
➔ Networks



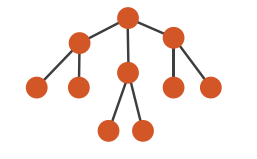
➔ Fields (Continuous)



➔ Multidimensional Table



➔ Trees



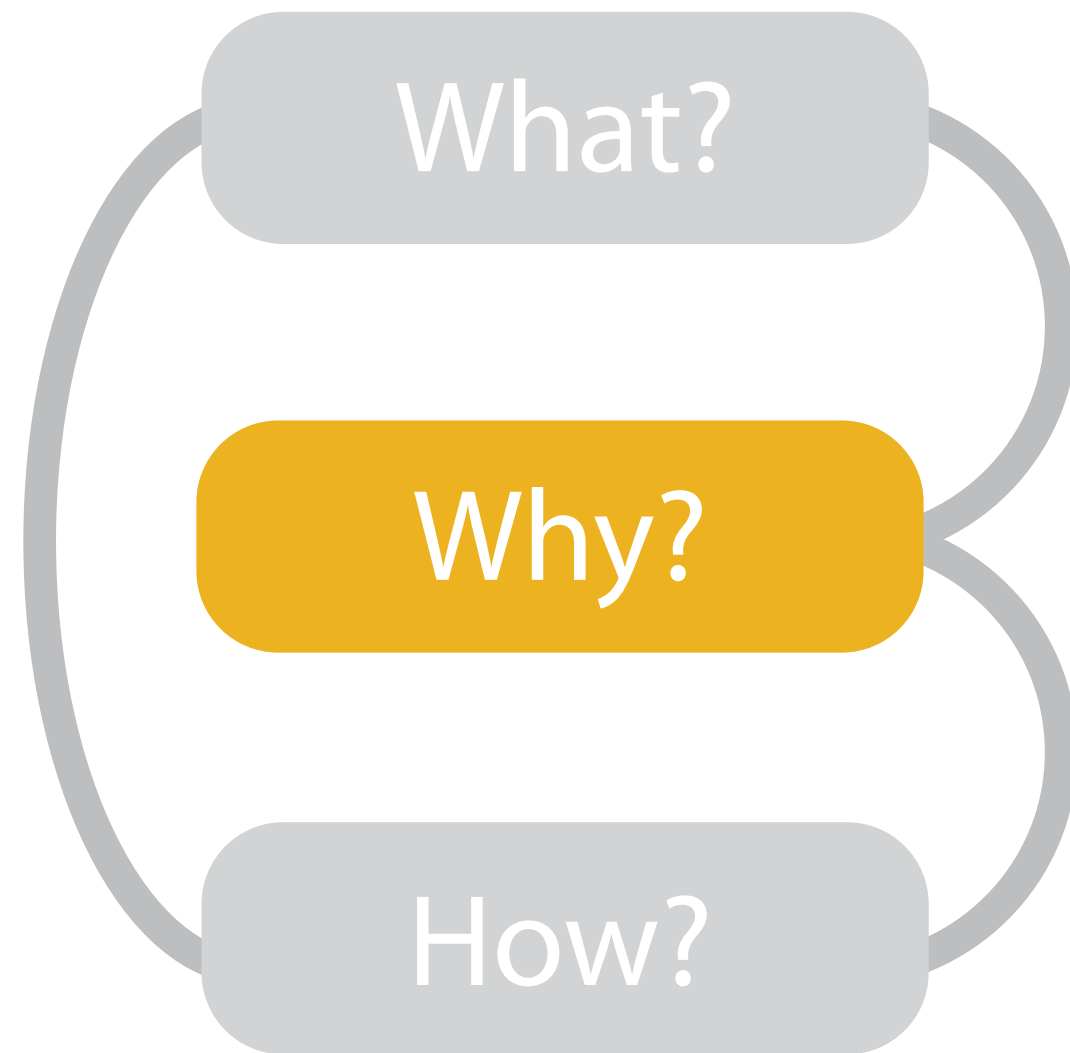
➔ Geometry (Spatial)



Ordering Direction

- ➔ Sequential
 -
- ➔ Diverging
 - ← →
- ➔ Cyclic
 - ↻

Why?



👉 Actions

🎯 Targets

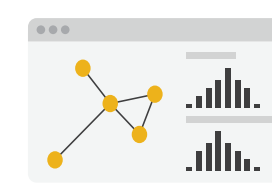
➔ Analyze

➔ Consume

➔ Discover



➔ Present

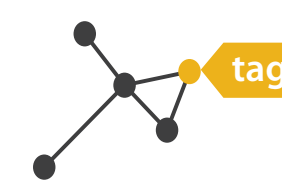


➔ Enjoy

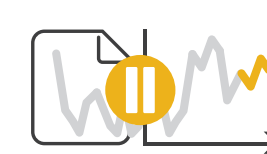


➔ Produce

➔ Annotate



➔ Record



➔ Derive

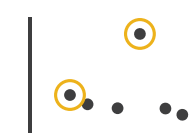


➔ Search

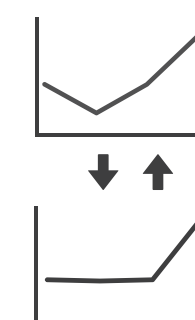
	Target known	Target unknown
Location known	<i>Lookup</i>	<i>Browse</i>
Location unknown	<i>Locate</i>	<i>Explore</i>

➔ Query

➔ Identify



➔ Compare

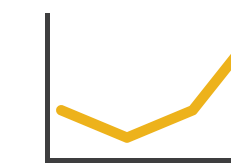


➔ Summarise



➔ All Data

➔ Trends



➔ Outliers



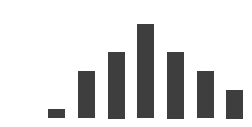
➔ Features



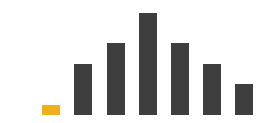
➔ Attributes

➔ One

➔ Distribution



➔ Extremes

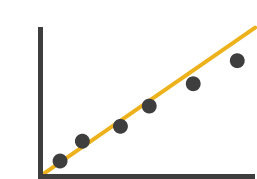


➔ Many

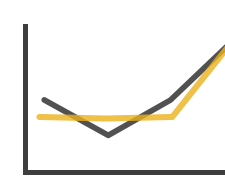
➔ Dependency



➔ Correlation

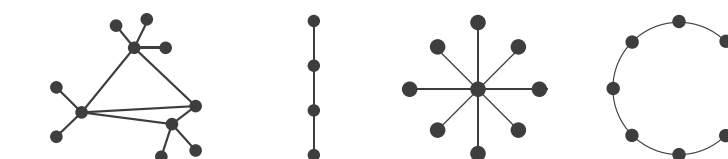


➔ Similarity



➔ Network Data

➔ Topology

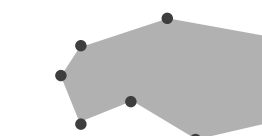


➔ Paths




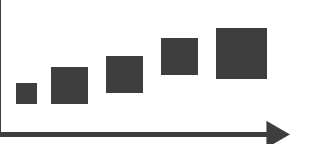



➔ Spatial Data






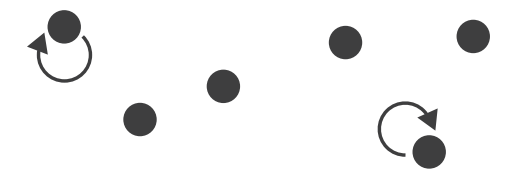
➔ Shape



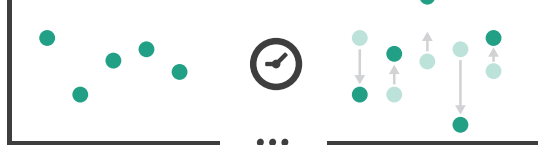
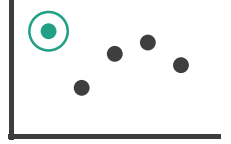

How?

Encode

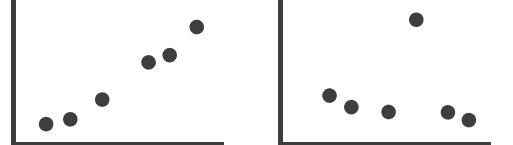

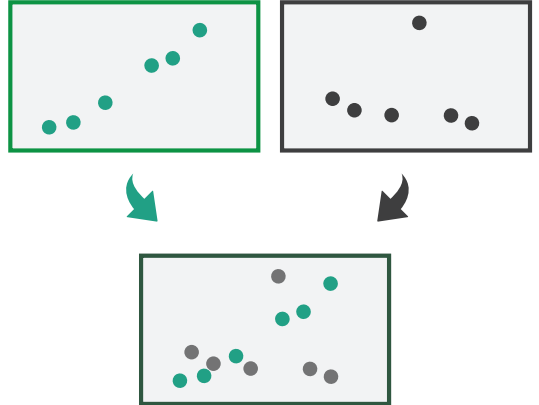
- ➔ **Arrange**
 - ➔ Express 
 - ➔ Order 
 - ➔ Use 
- ➔ **Separate** 
- ➔ **Align** 

- ➔ **Map** from **categorical** and **ordered** attributes
 - ➔ **Color**
 - ➔ *Hue* 
 - ➔ *Saturation* 
 - ➔ *Luminance* 
 - ➔ *Size, Angle, Curvature, ...*
 - 
 - ➔ **Shape**
 - 
 - ➔ **Motion**
 - Direction, Rate, Frequency, ...*
 - 


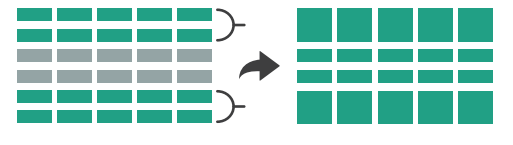

Manipulate

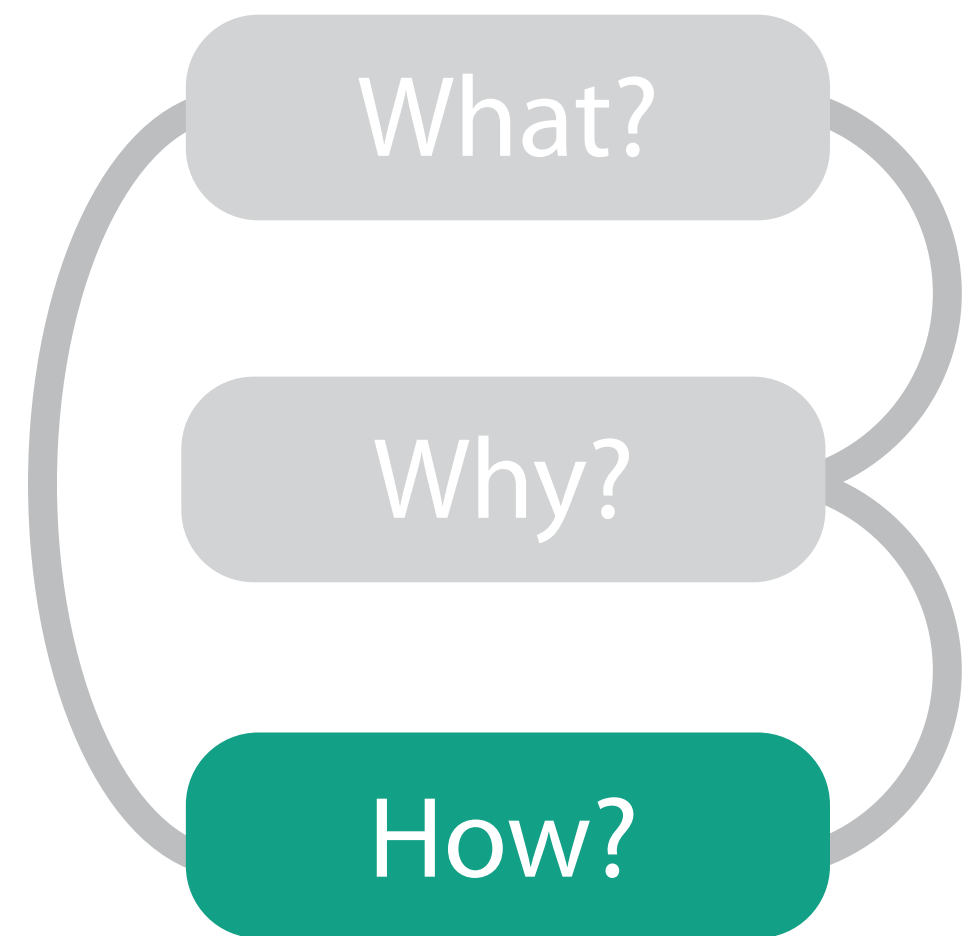
- ➔ **Change** 
- ➔ **Select** 
- ➔ **Navigate** 

Facet

- ➔ **Juxtapose** 
- ➔ **Partition** 
- ➔ **Superimpose** 

Reduce

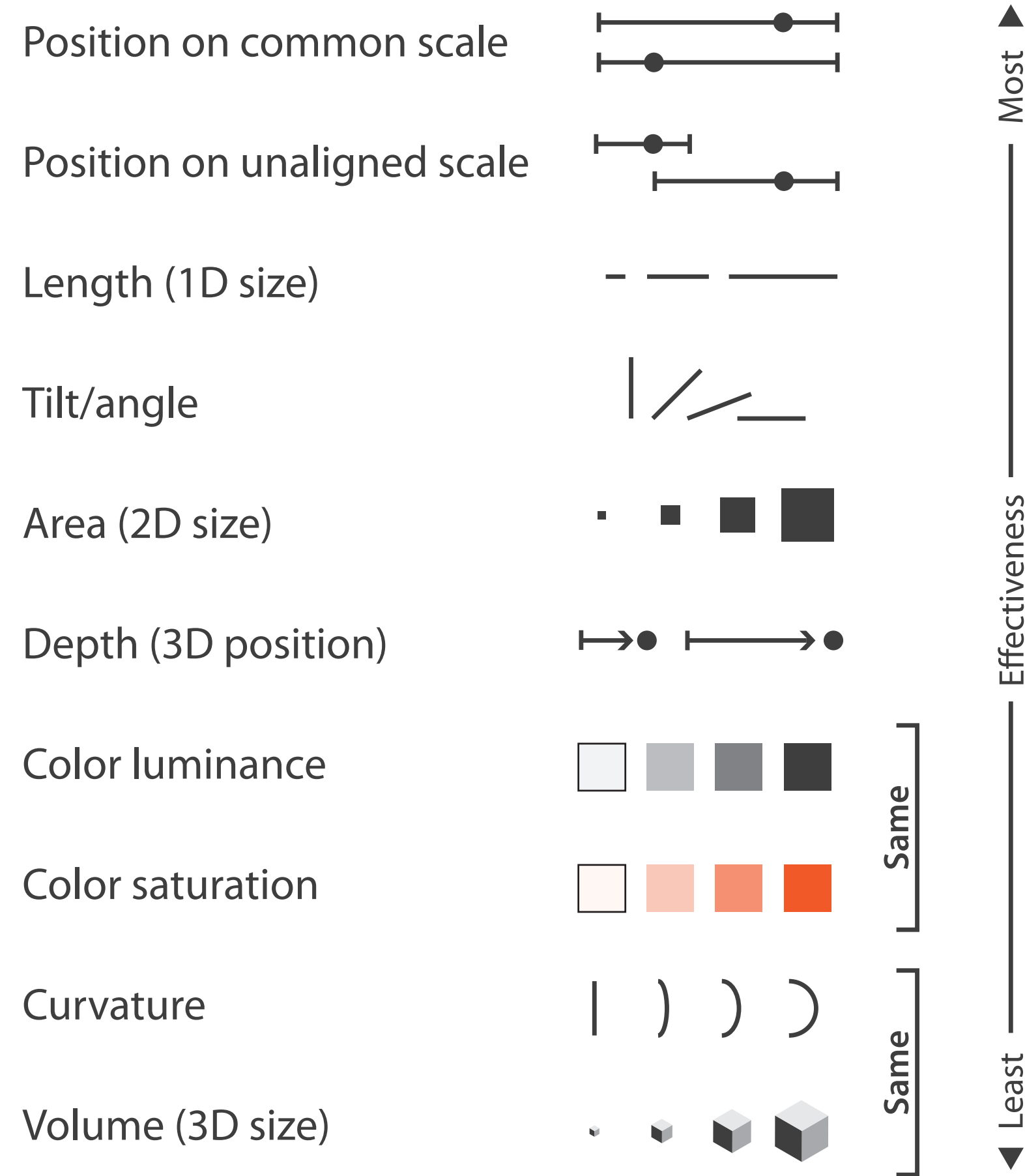
- ➔ **Filter** 
- ➔ **Aggregate** 
- ➔ **Embed** 



Marks and Channels

Channels: Expressiveness Types and Effectiveness Ranks

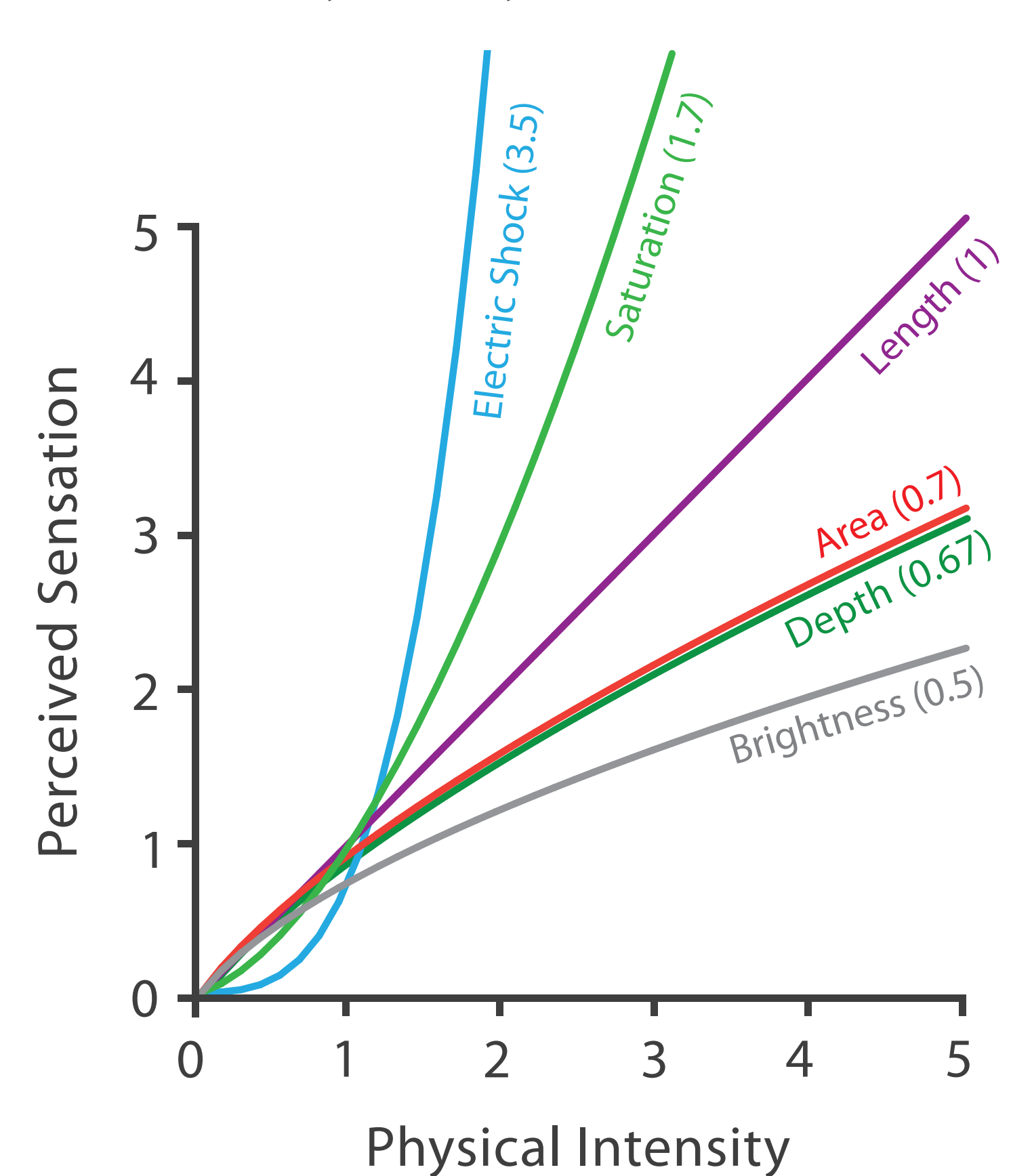
➔ **Magnitude Channels: Ordered Attributes**



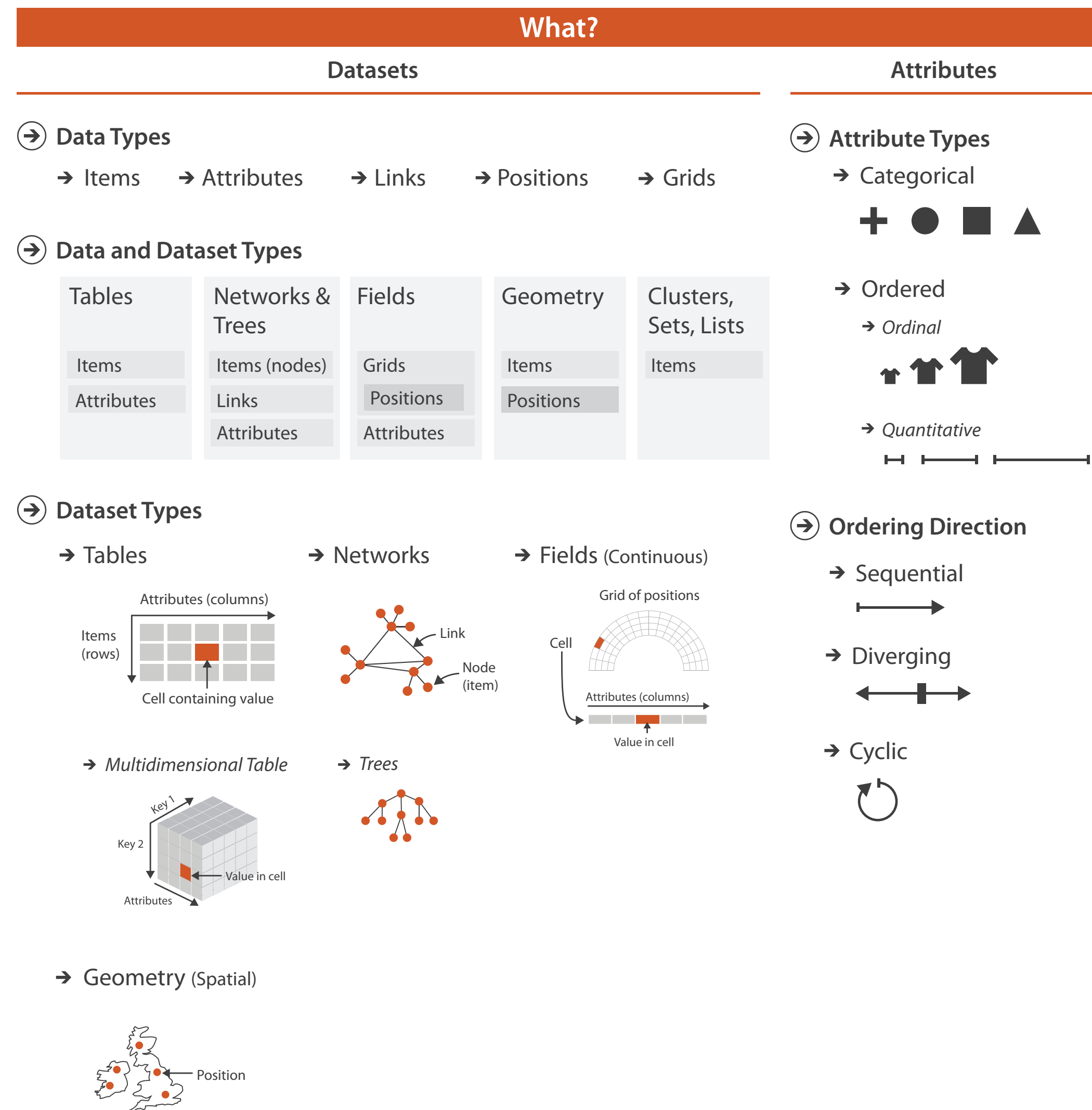
➔ **Identity Channels: Categorical Attributes**



Steven's Psychophysical Power Law: $S = I^N$



What: Data Abstraction



Why: Tasks Abstraction

Why?

Actions

Targets

→ Analyze

→ Consume

→ Discover



→ Present

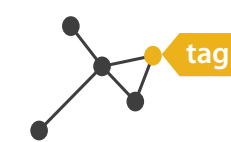


→ Enjoy

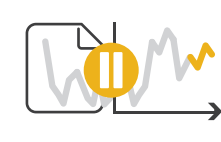


→ Produce

→ Annotate



→ Record



→ Derive

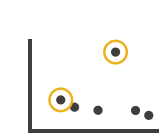


→ Search

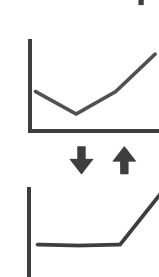
	Target known	Target unknown
Location known	<i>Lookup</i>	<i>Browse</i>
Location unknown	<i>Locate</i>	<i>Explore</i>

→ Query

→ Identify



→ Compare

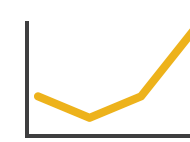


→ Summarize

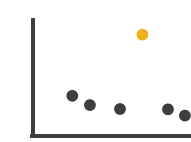


→ All Data

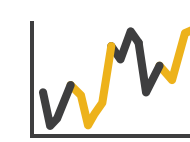
→ Trends



→ Outliers



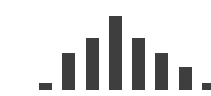
→ Features



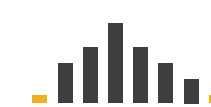
→ Attributes

→ One

→ Distribution

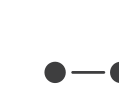


→ Extremes

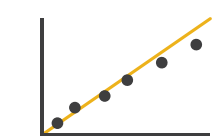


→ Many

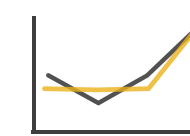
→ Dependency



→ Correlation

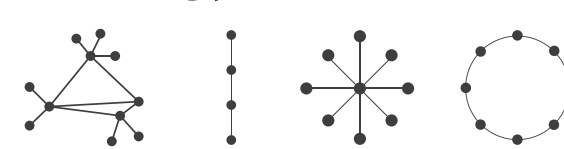


→ Similarity

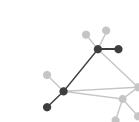


→ Network Data

→ Topology

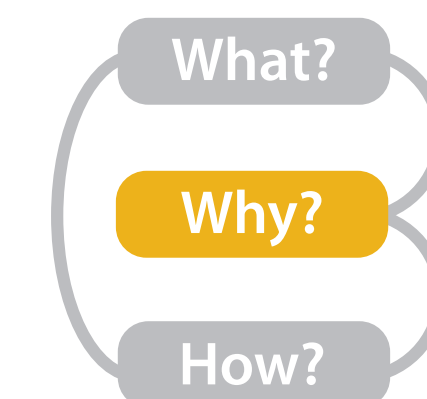


→ Paths



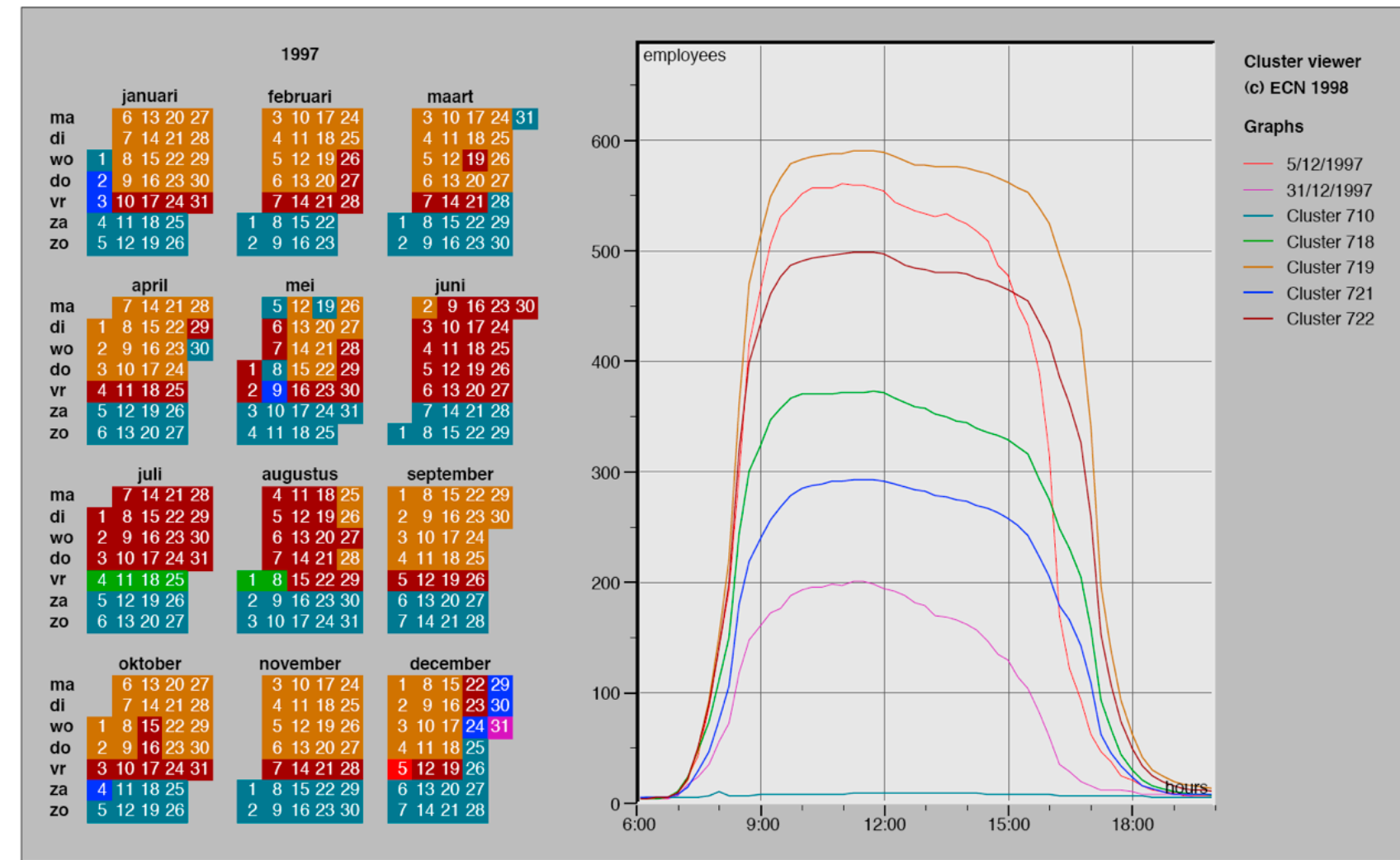
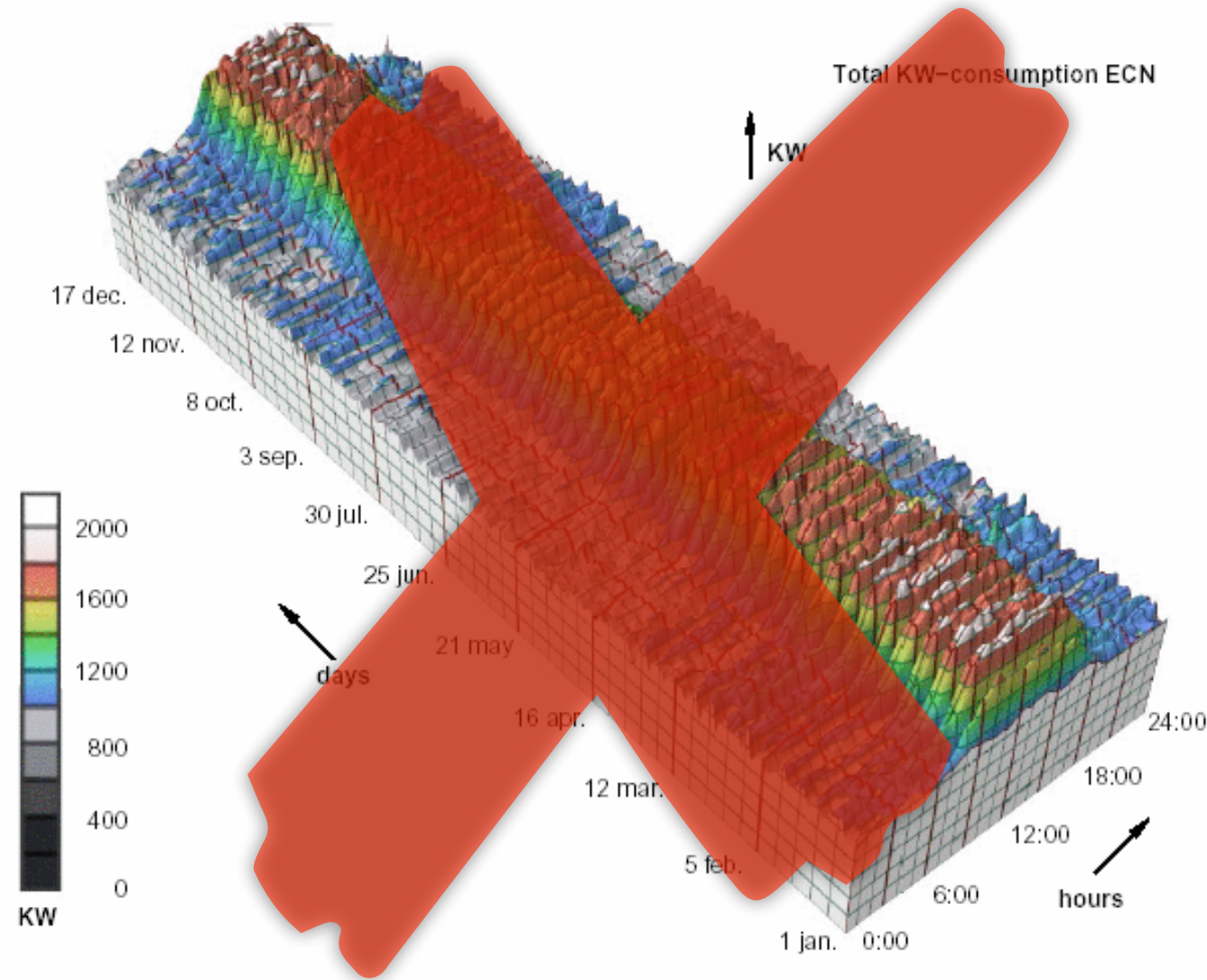
→ Spatial Data

→ Shape



Design Guidelines

T. Munzner VAD 2014



PROXIMITY

When objects placed together, the eye perceives them as a group.



SIMILARITY

When objects look similar to one another, the eye perceives them as a group or pattern.



CONTINUANCE

The eye is compelled to move from one object through another.



CLOSURE

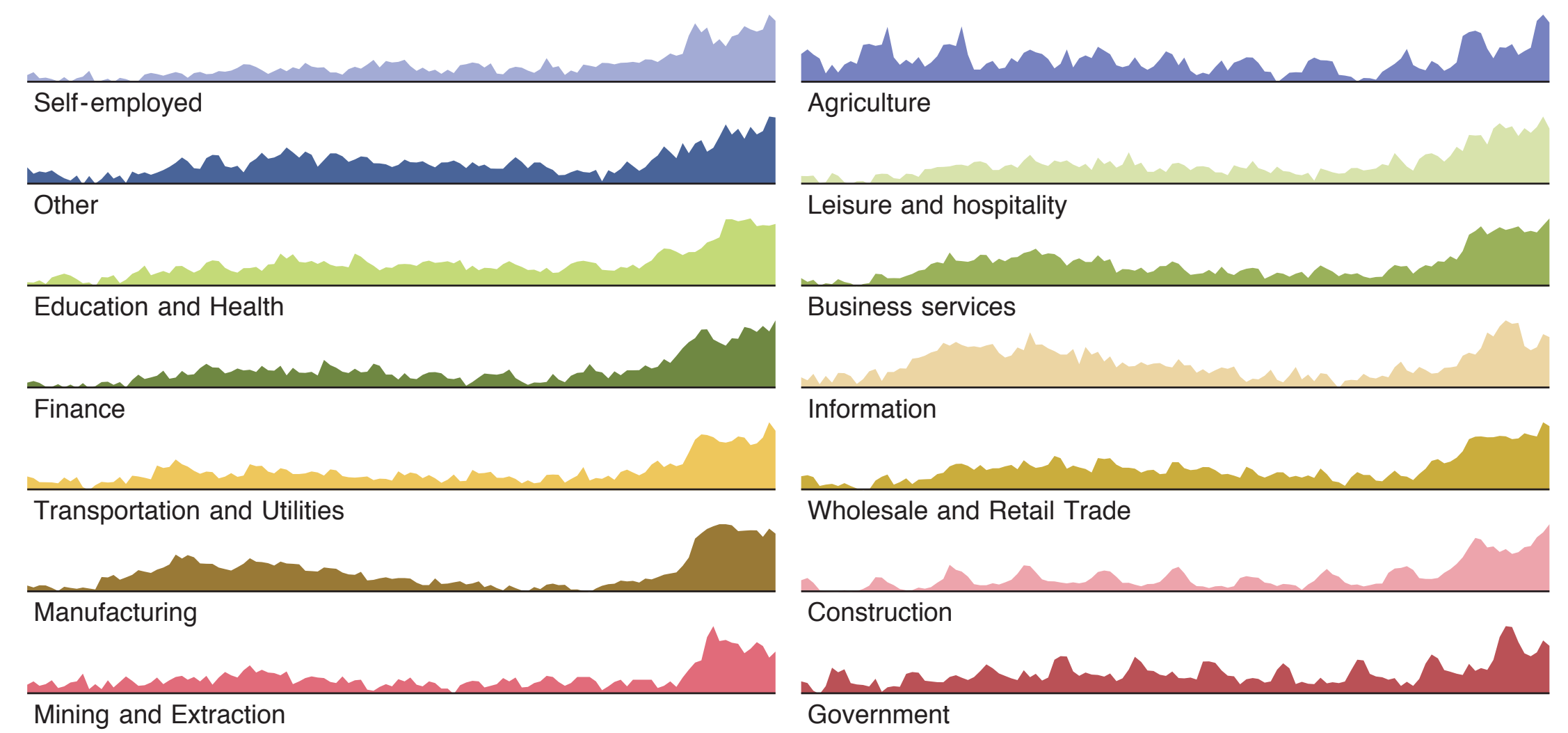
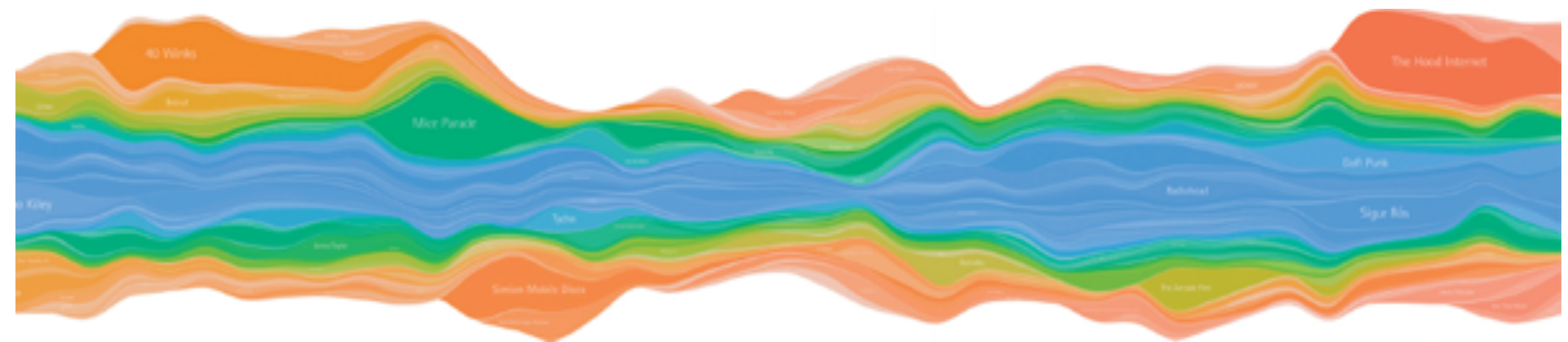
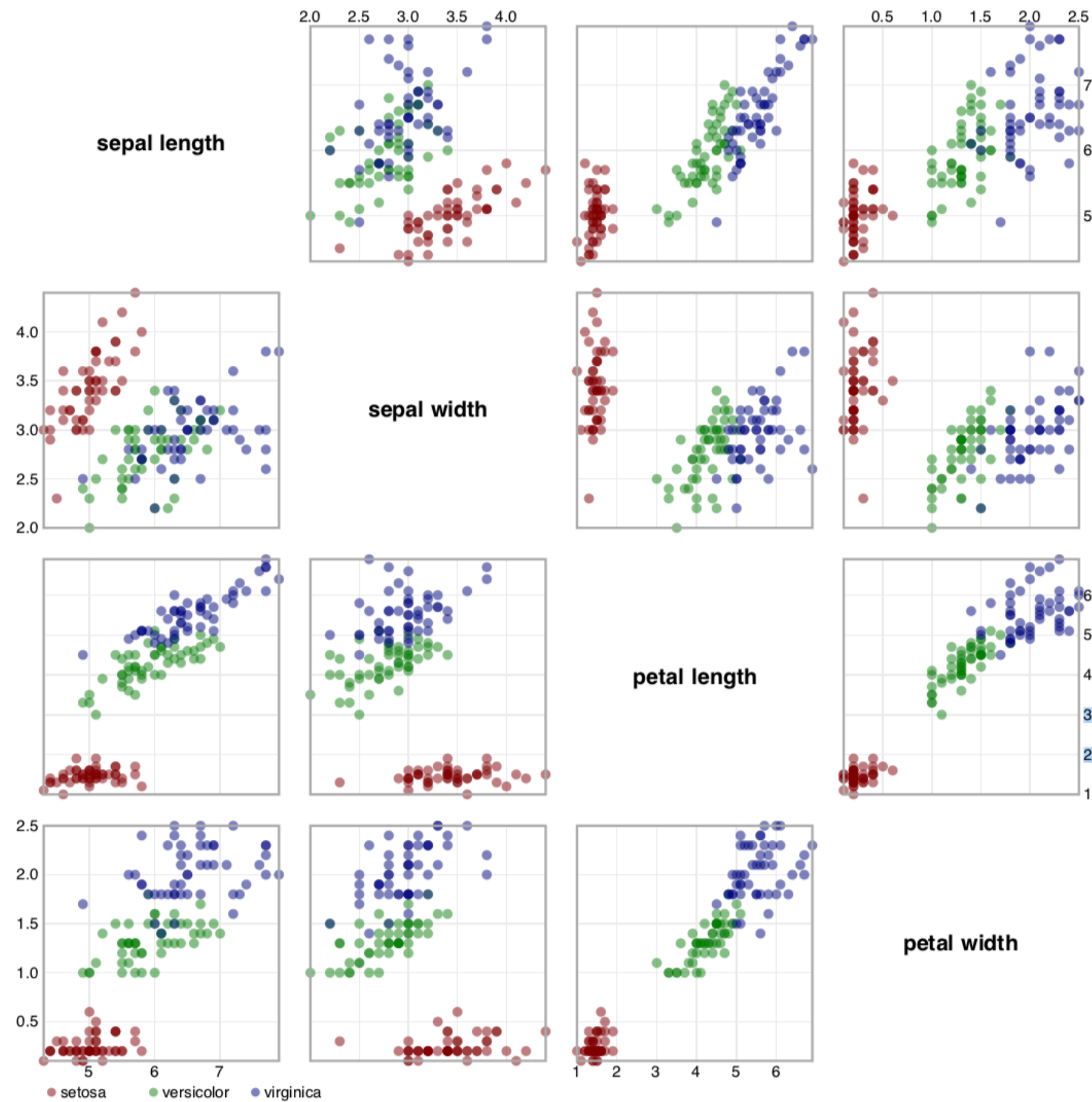
When an object is incomplete or not completely enclosed.



FIGURE & GROUND

When the eye differentiates an object from its surrounding area.

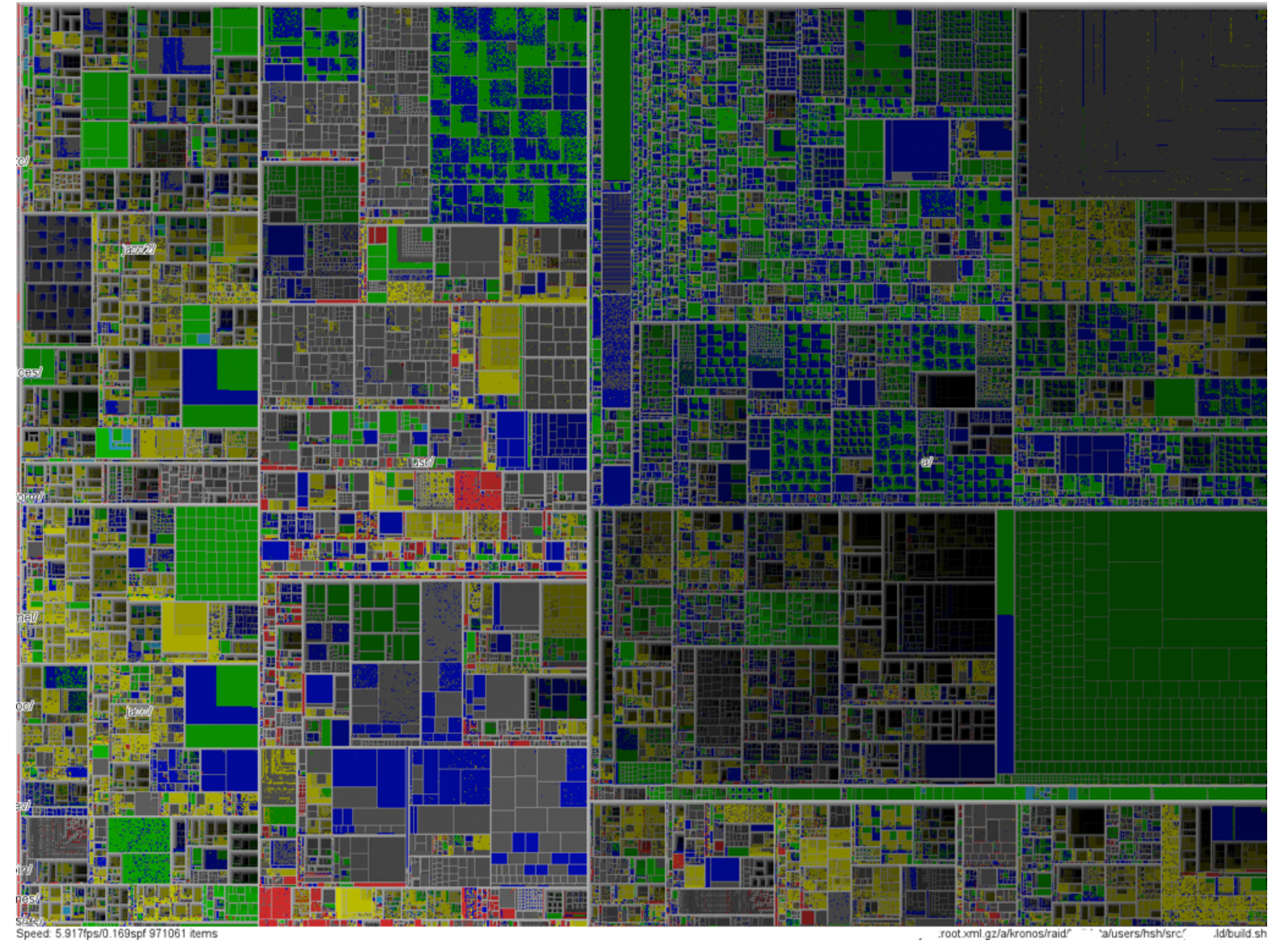
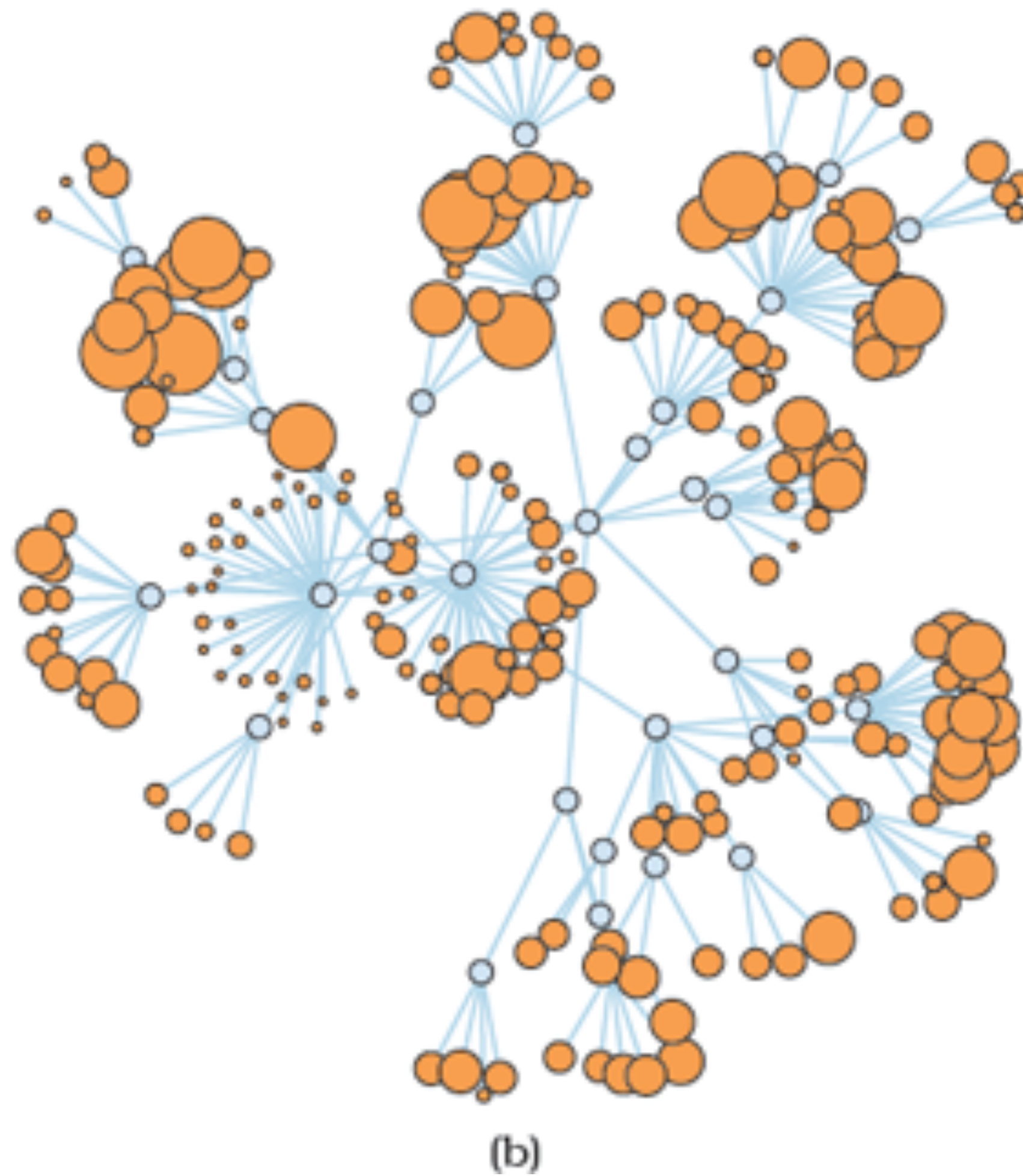
Tables



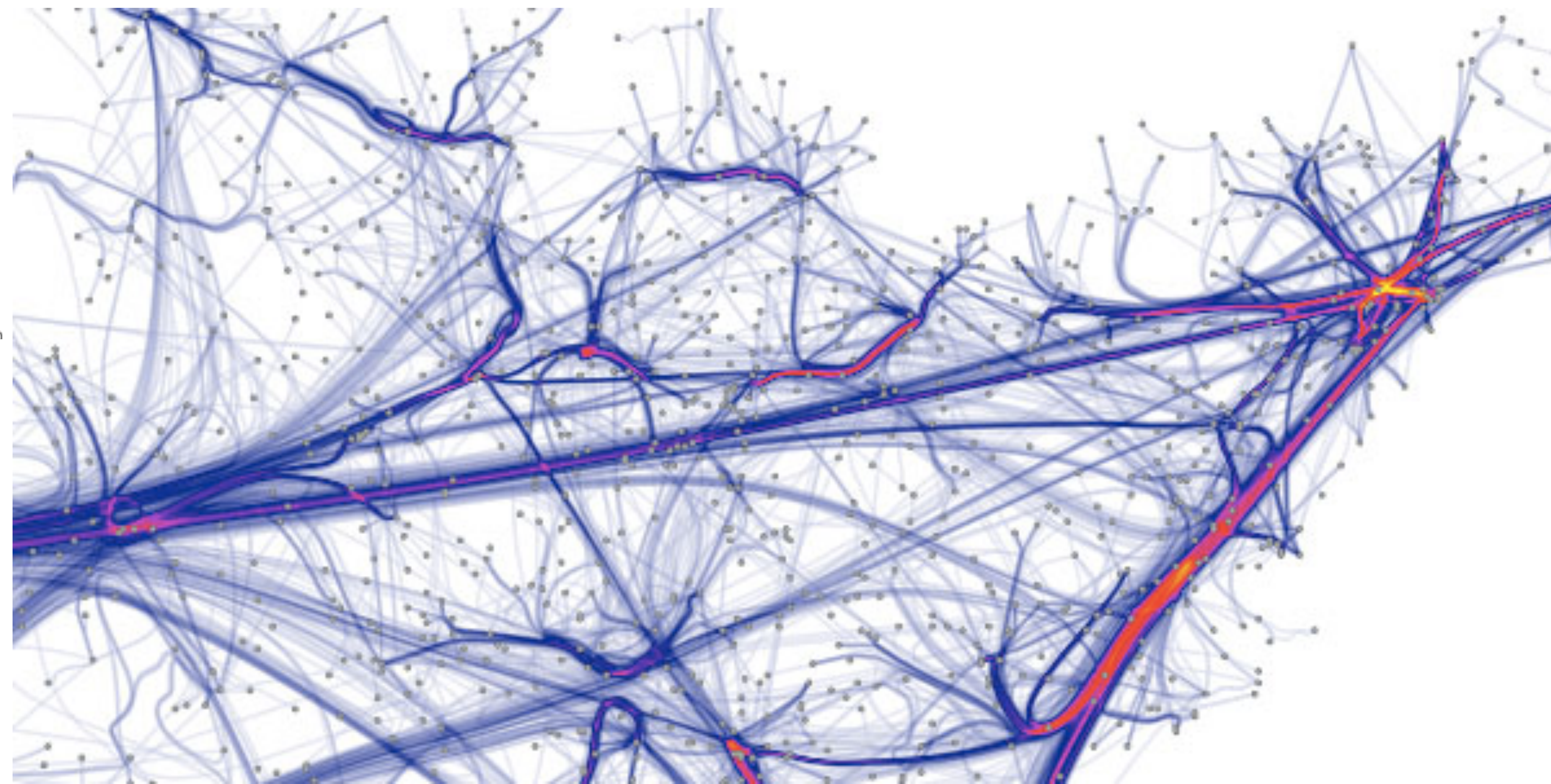
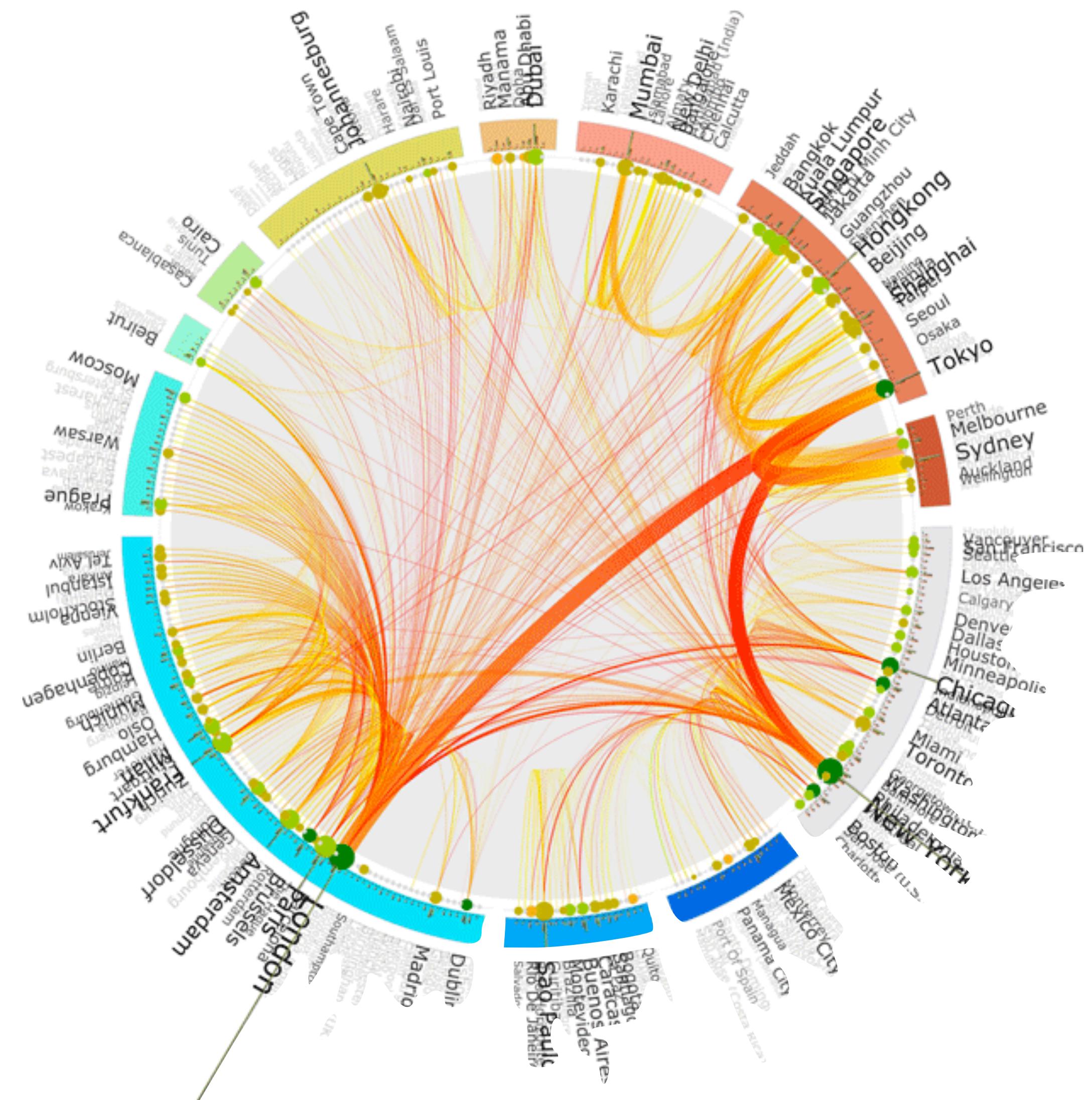
Source: U.S. Bureau of Labor Statistics; <http://hci.stanford.edu/jheer/files/zoo/ex/time/multiples.html>

Trees

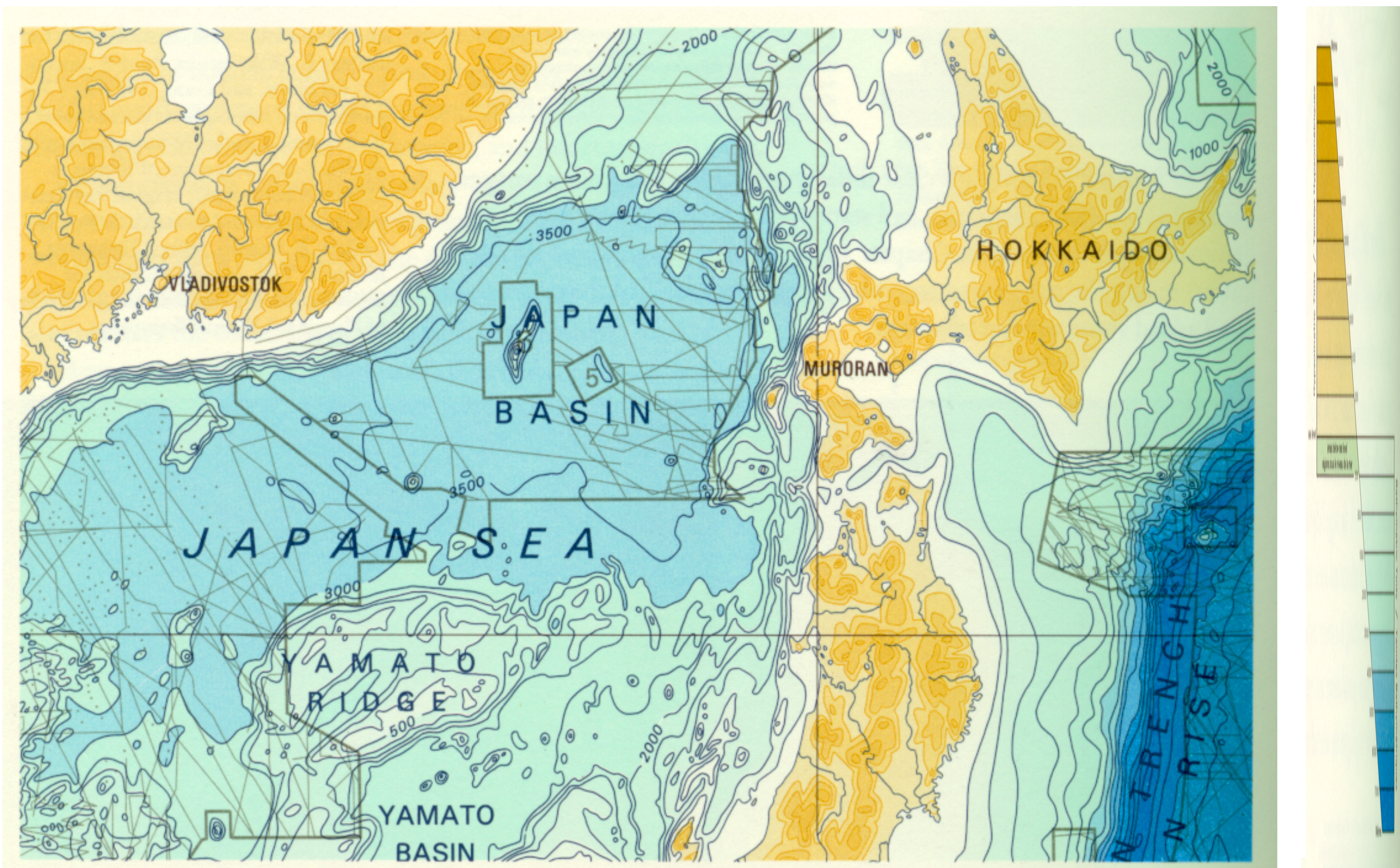
T. Munzner, VAD 2014



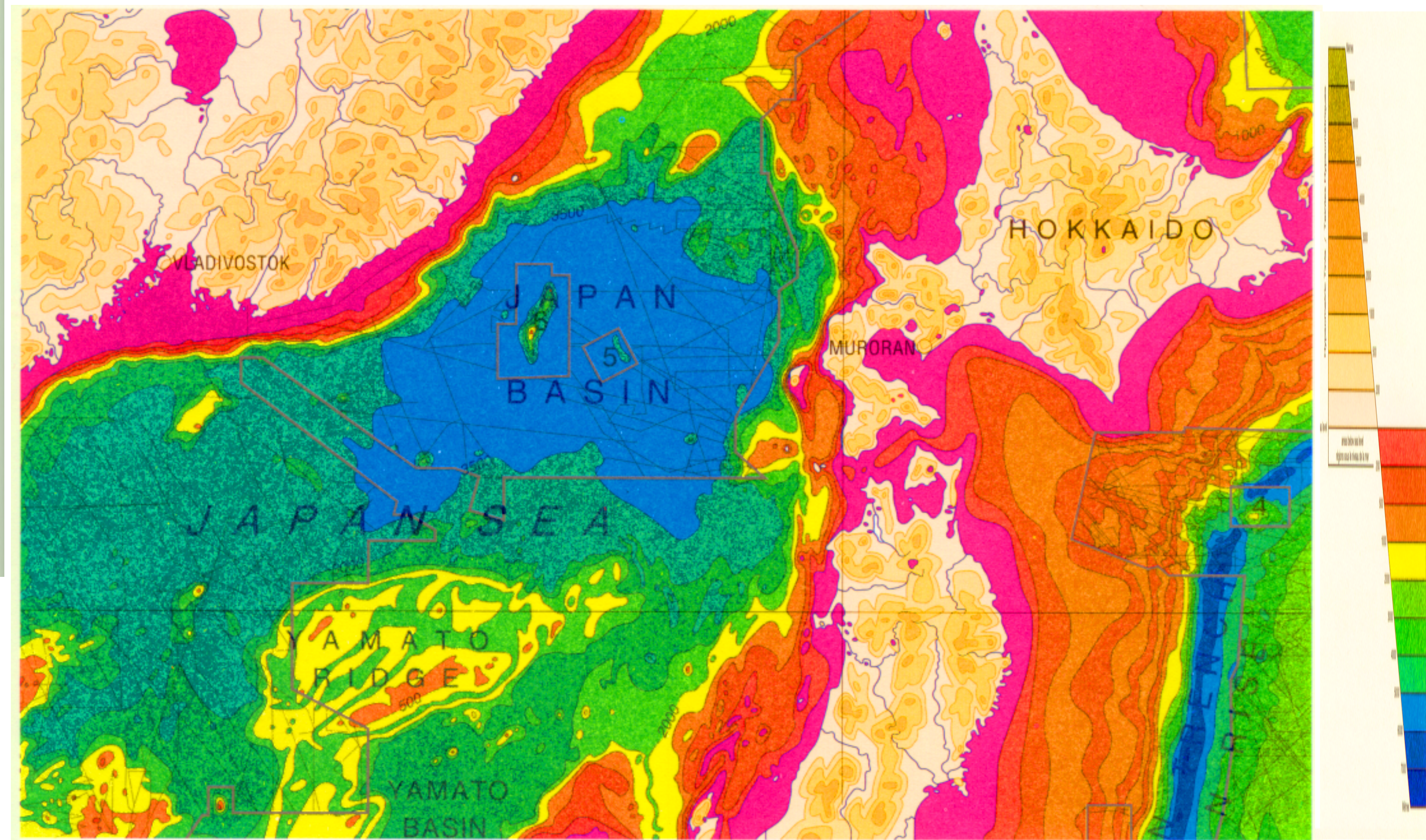
Networks / Graphs



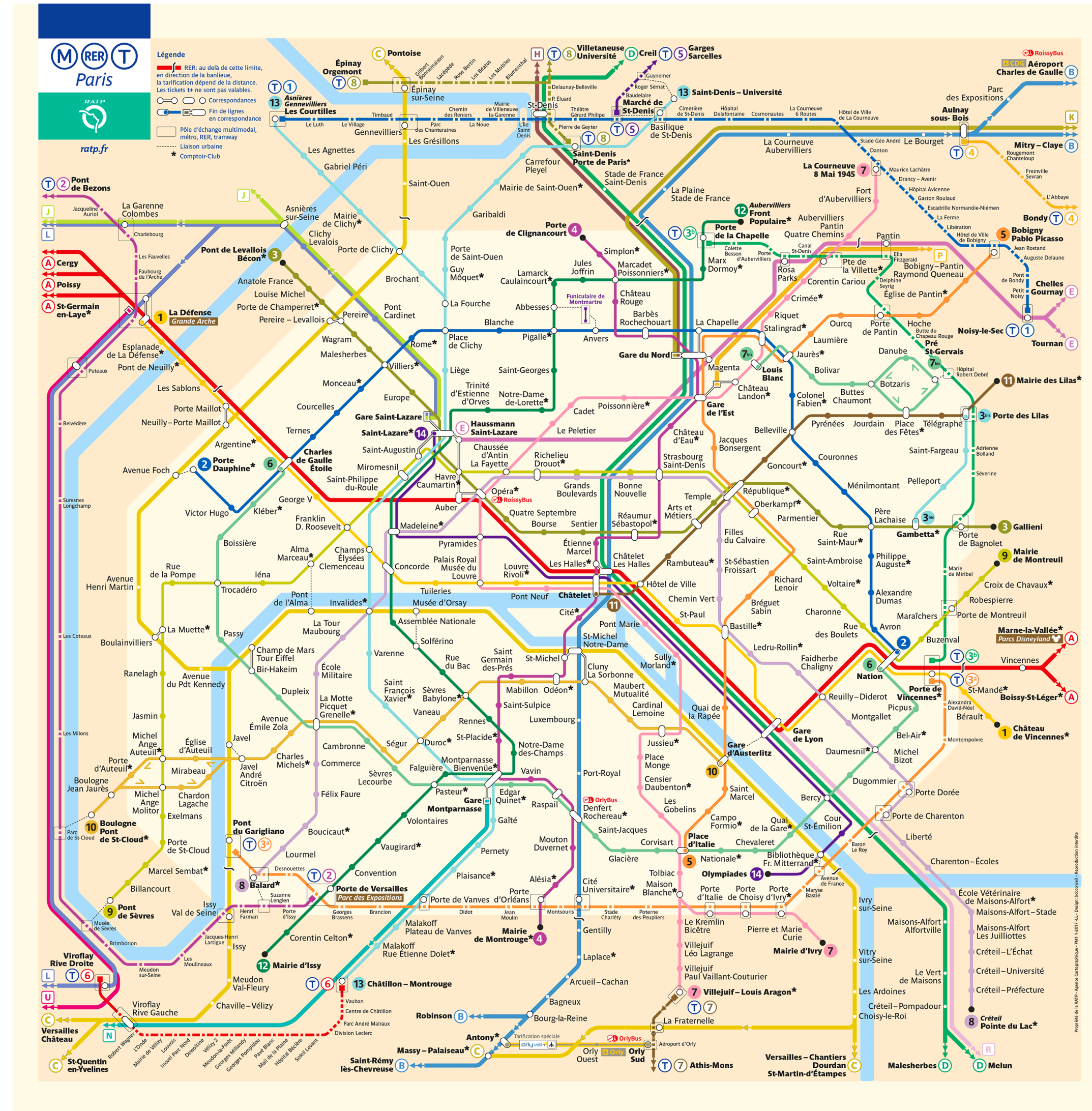
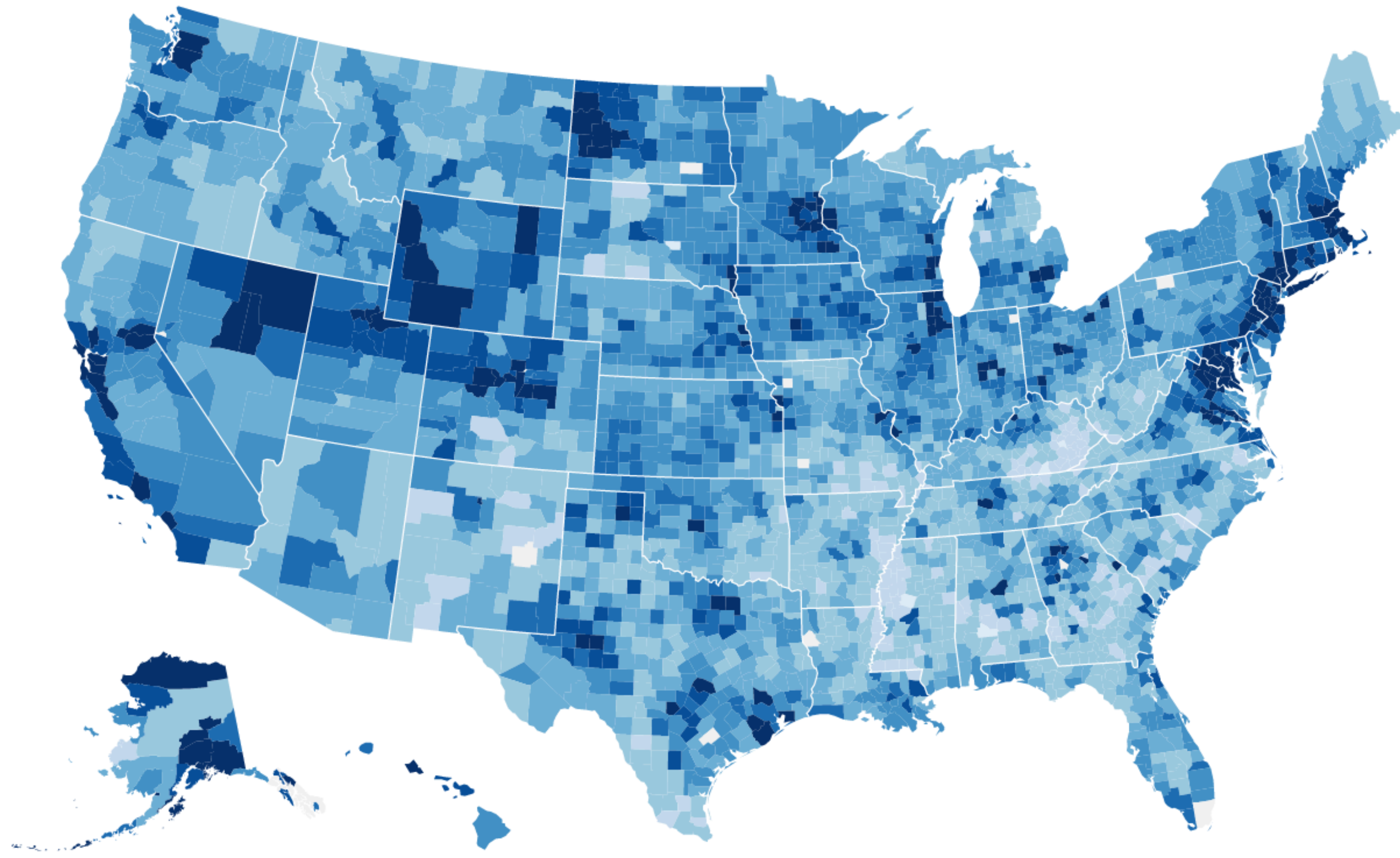
Colors



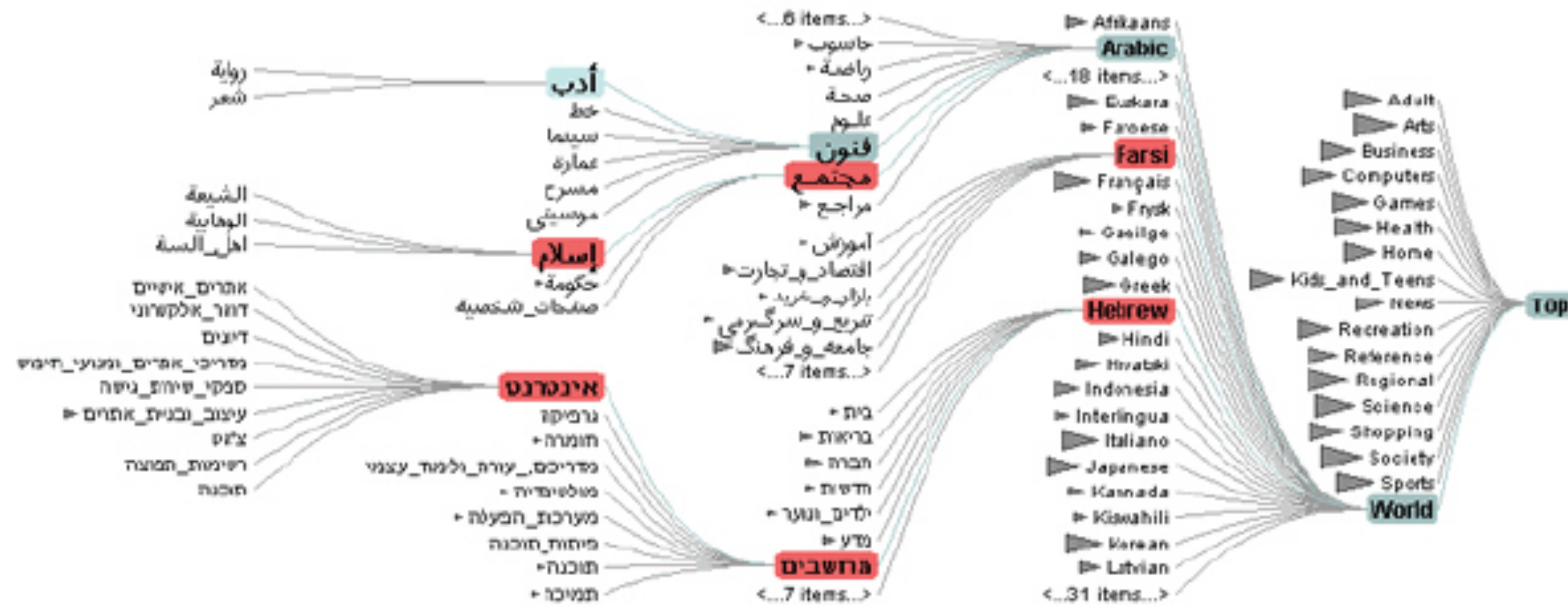
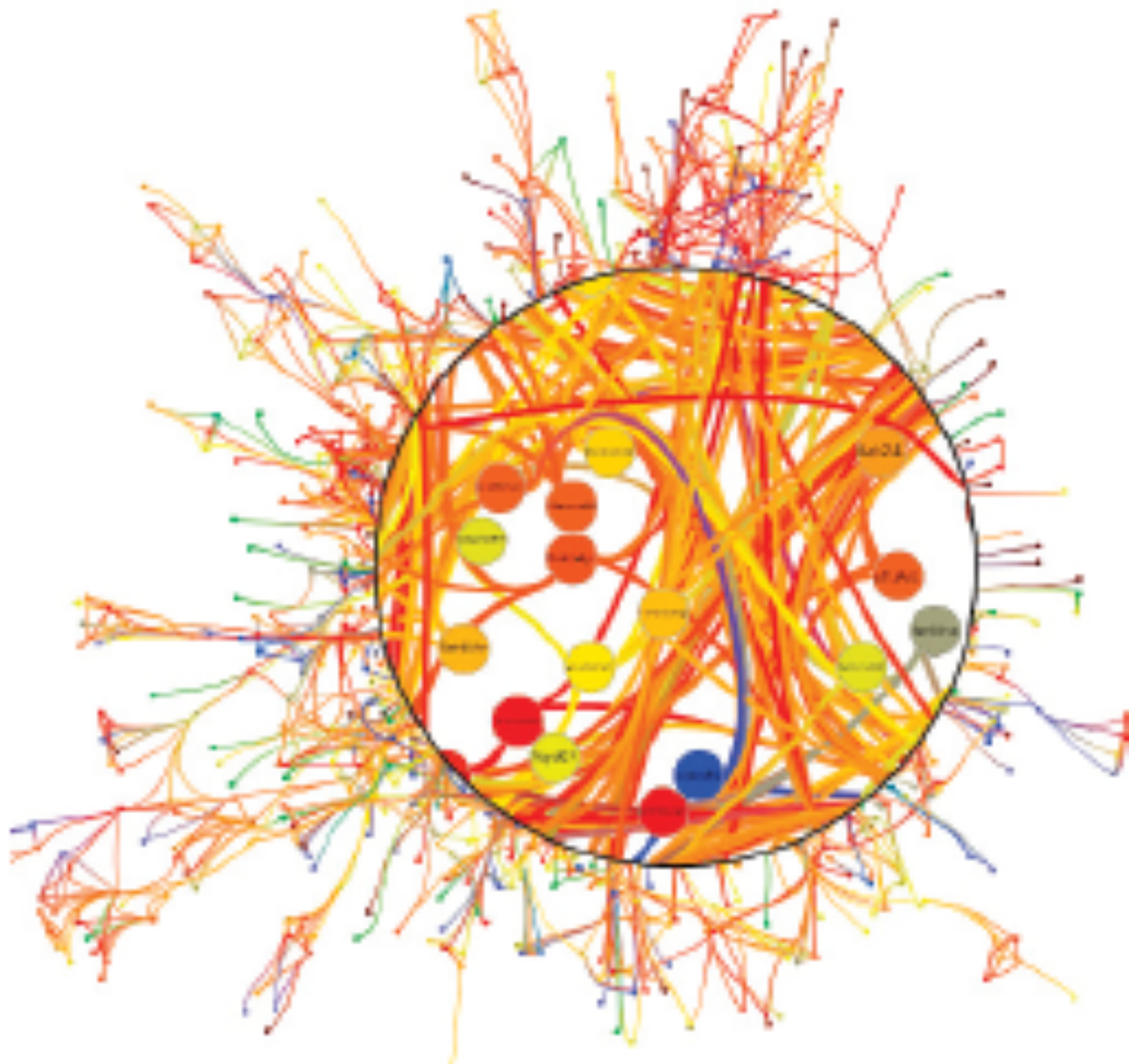
Colin Ware



Maps

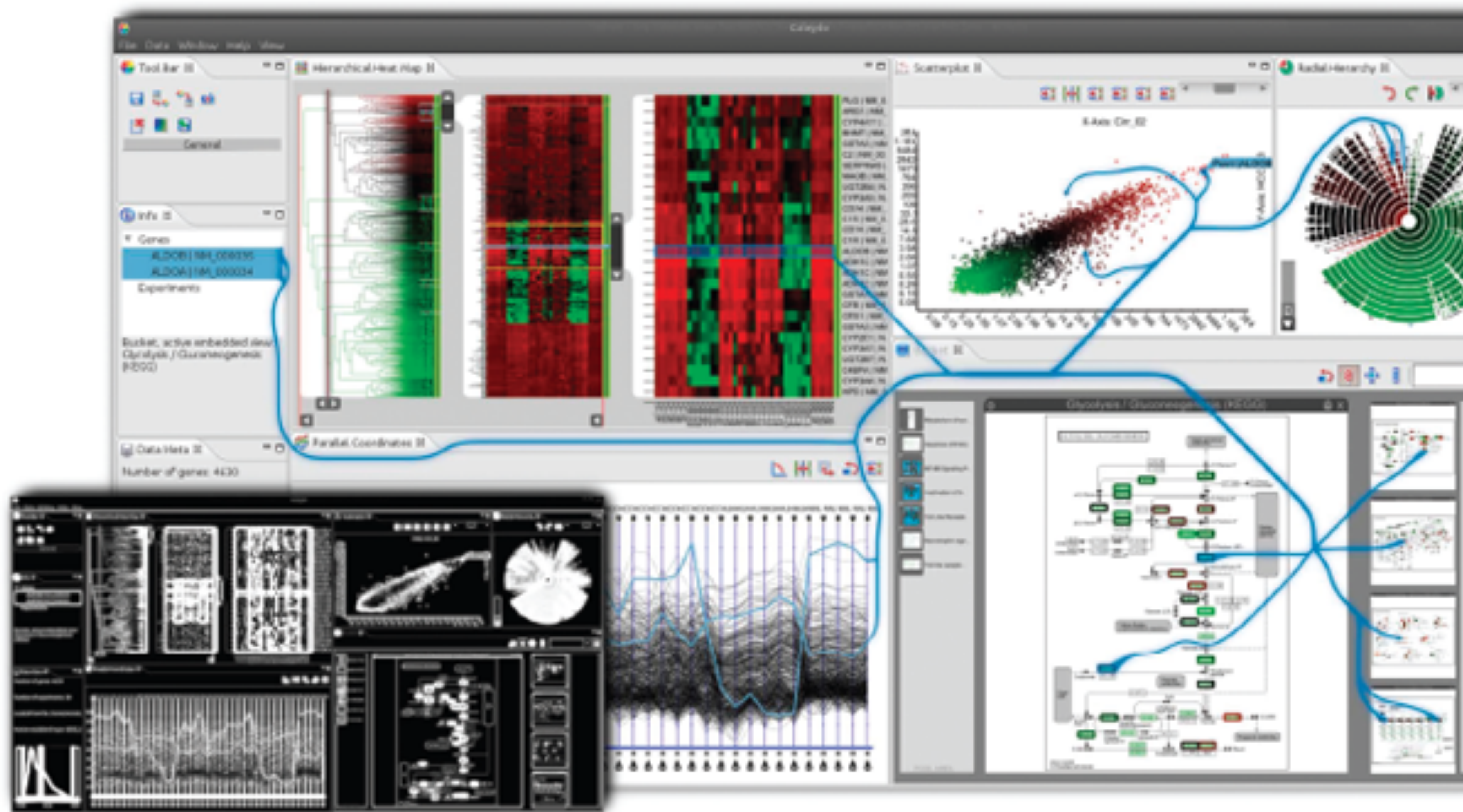


Focus + Context



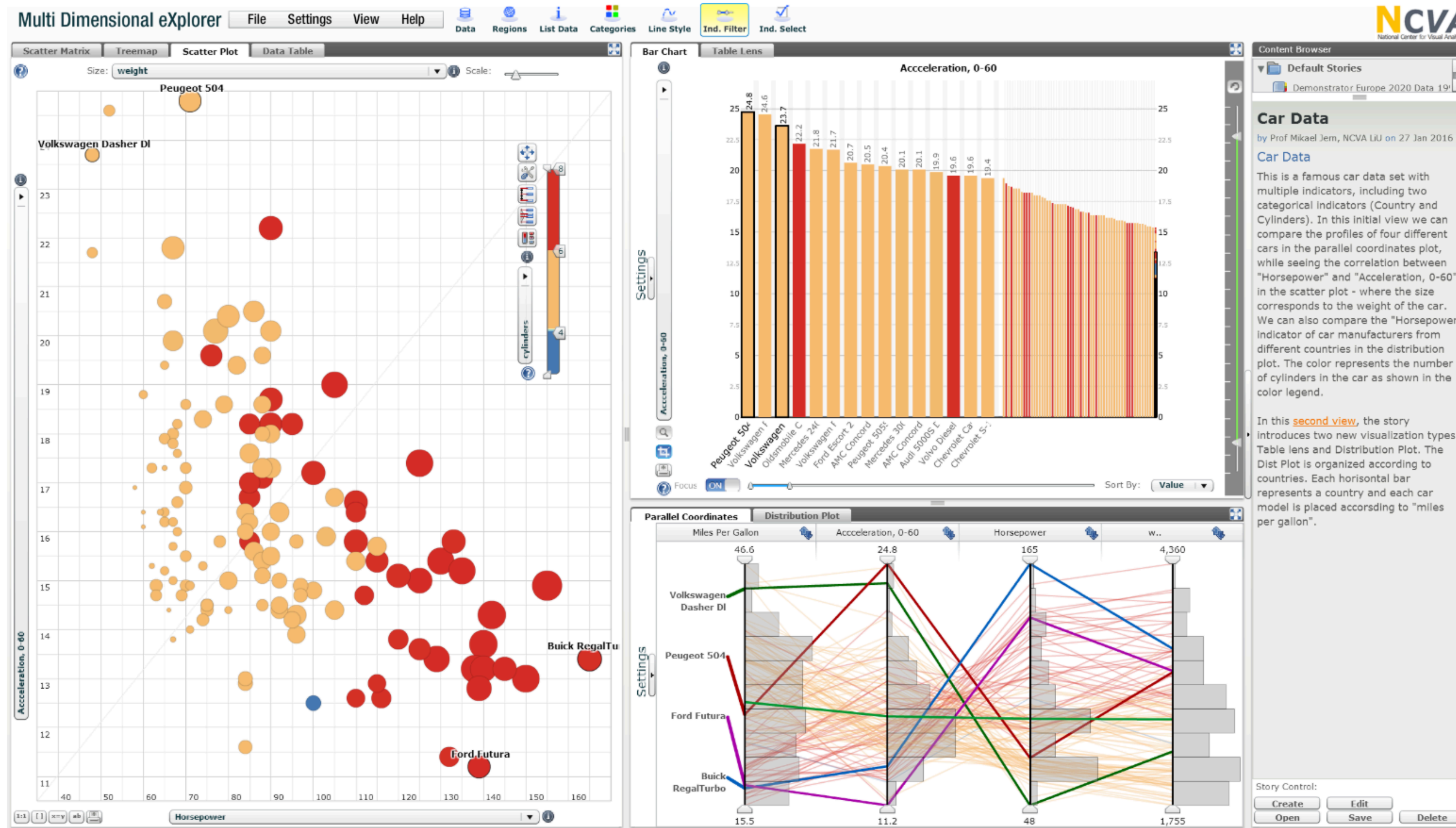
T. Munzner, VAD 2014

Interaction



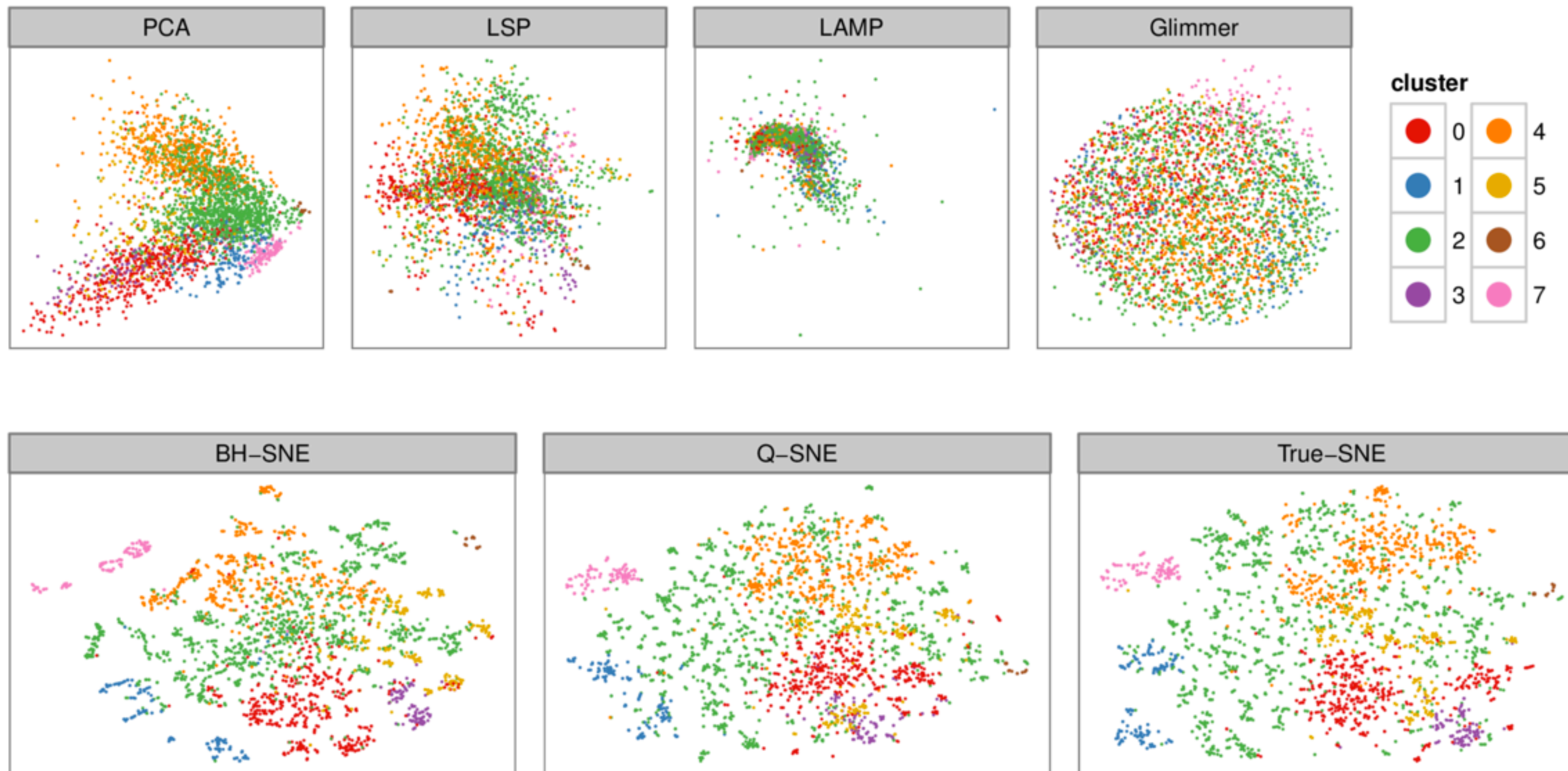
T. Munzner, VAD 2014

Multiple Views



<http://ncva.itn.liu.se/explorer/mdim-explorer>

Reduction

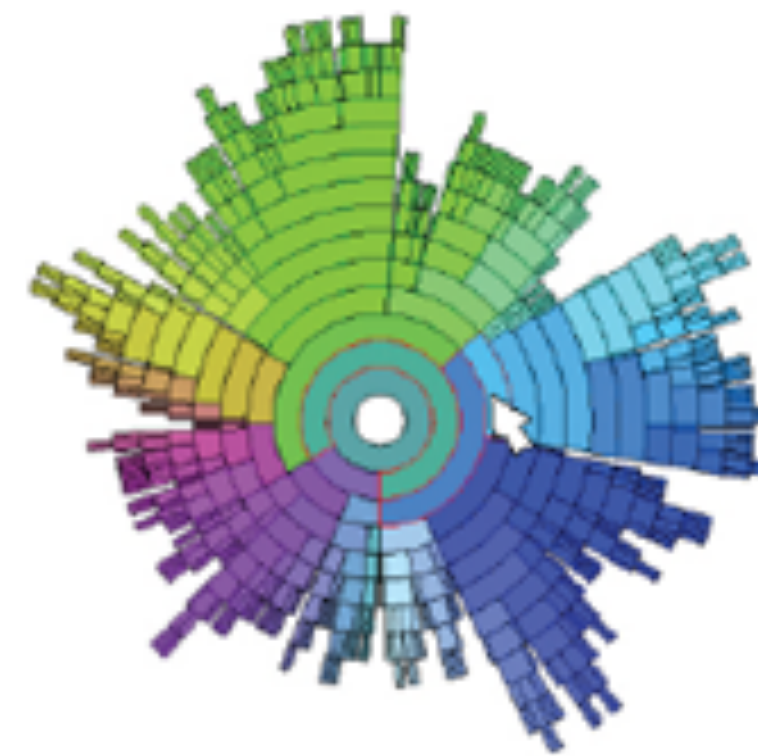
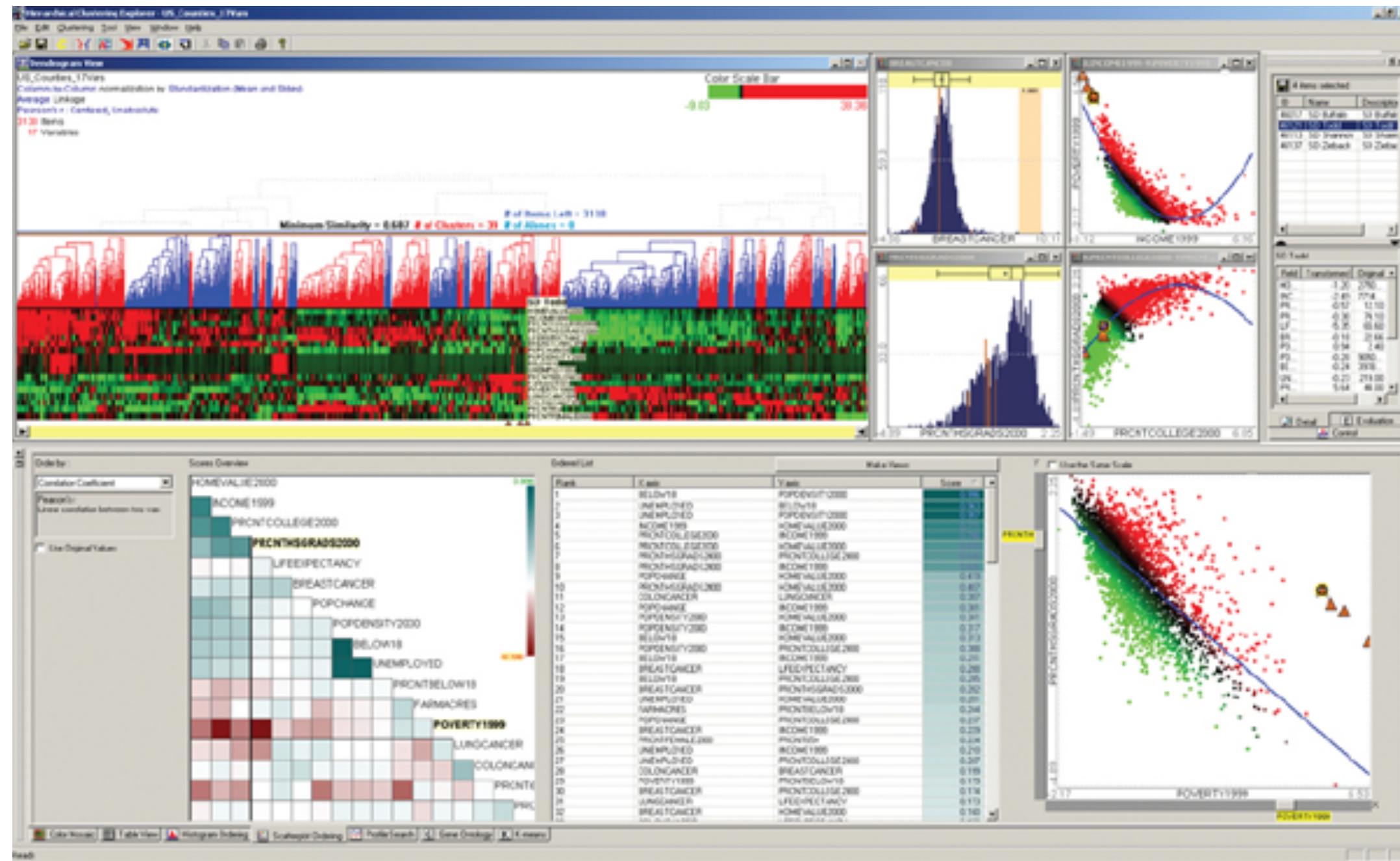


Ingram and Munzner, 2015

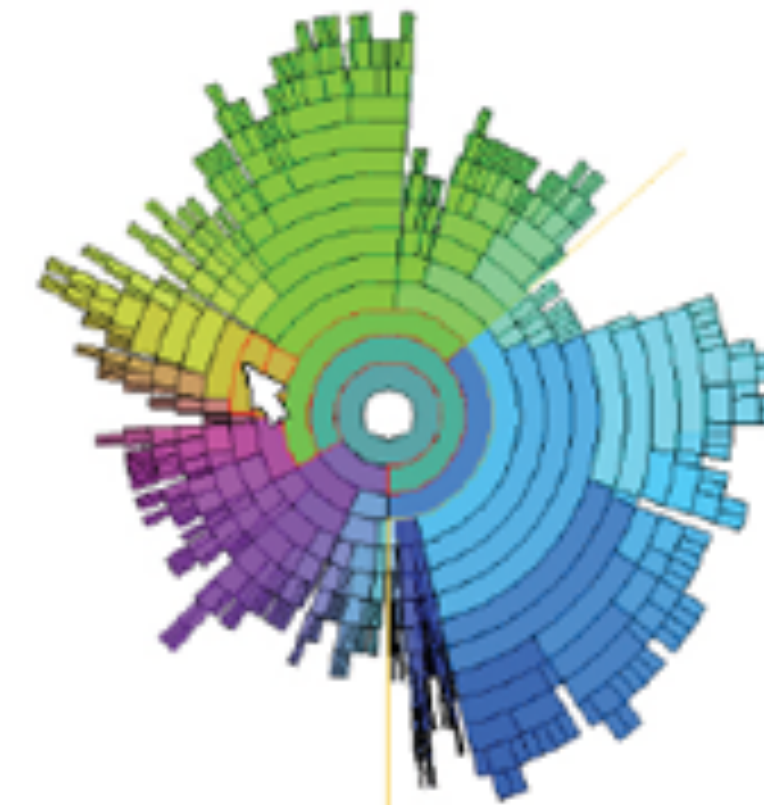
Visualization Tools



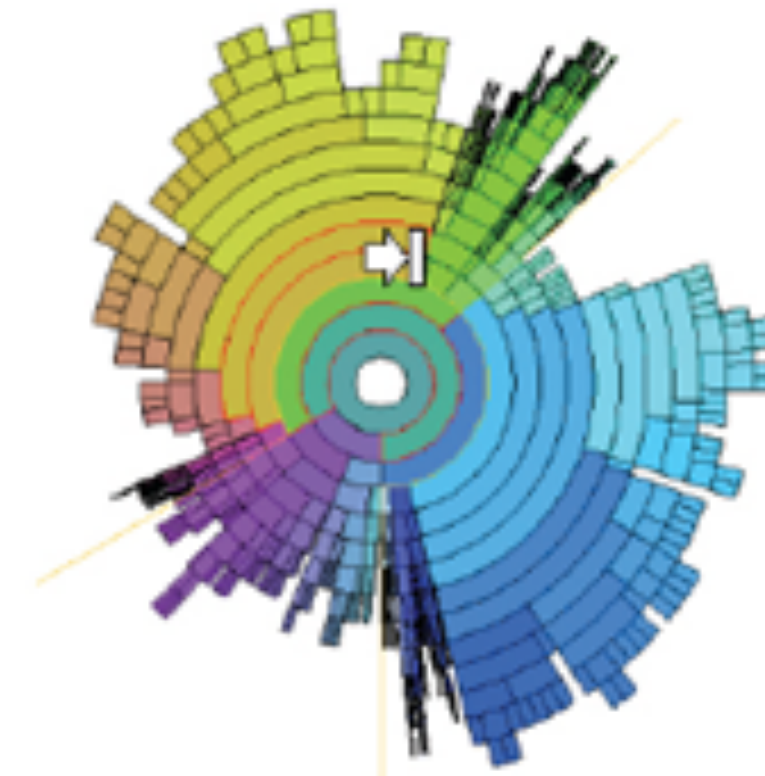
Design Studies



(a)



(b)



(c)

Next Lecture: **Visualization Libs**



- Read VAD Chapter 1
- Install and experiment with your visualization library