

第一讲：信息可视化简介

信息可视化

曹楠（教授）

<https://idvxlab.com>

同济大学

课程大纲

- 什么是可视化?
- 为什么要可视化?
- 可视化的类型与发展趋势
- 怎样对数据进行可视化?
- 如何对可视化进行评估?
- 怎样学习可视化?
- 可视化应用案例

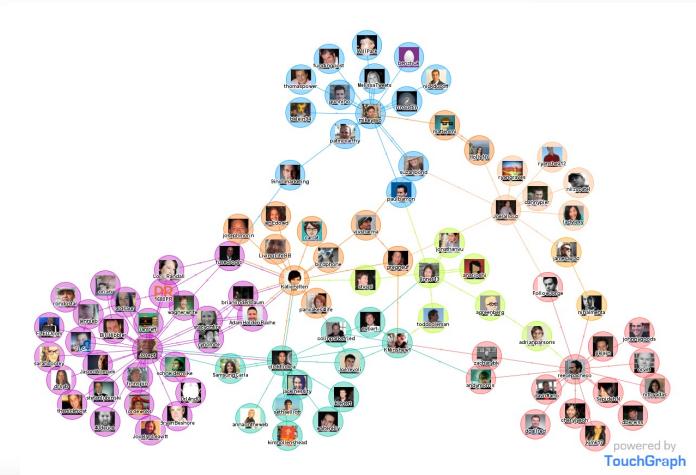
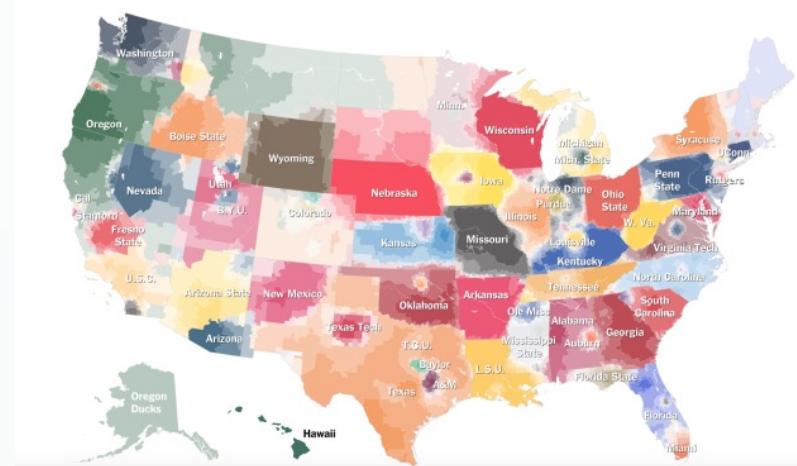
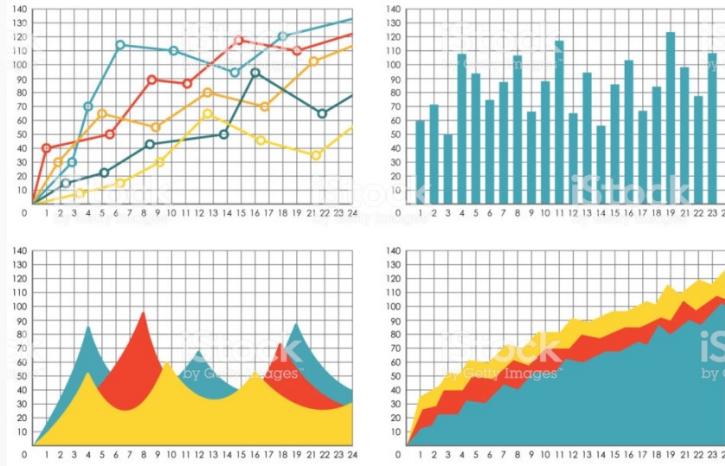
什么是可视化?



“用视觉的方式传达信息被称之为信息传达设计”

设计师是信息的发送者，传达对象是信息的接受者；传达是指人与人之间利用“看”的形式所进行的交流，是通过视觉语言进行信息表达与传播

什么是可视化?



当信息来自数据时，可以用数据可视化技术来呈现信息的内涵

输入数据，自动生成对数据内容的图形化表达

课程大纲

- 什么是可视化?
- 为什么要可视化?
- 可视化的类型与发展趋势
- 怎样对数据进行可视化?
- 如何对可视化进行评估?
- 怎样学习可视化?
- 可视化应用案例

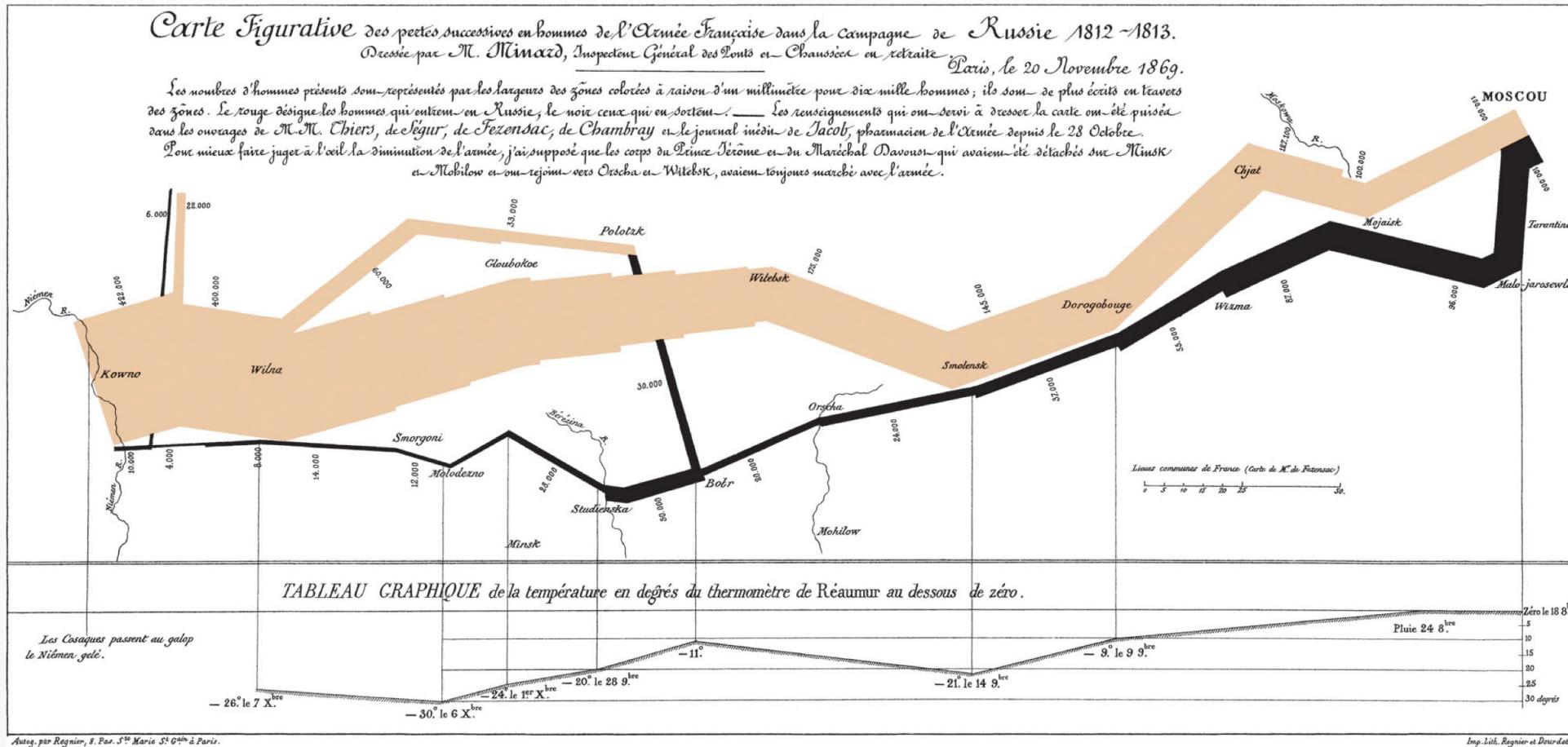
为什么要可视化?

- 直观的展示信息



为什么要可视化?

- 直观的展示信息



为什么要可视化?

123569323487043297346702739
373436793091025942425523

字符串中有几个3?

为什么要可视化?

**123569323487043297346702739
373436793091025942425523**

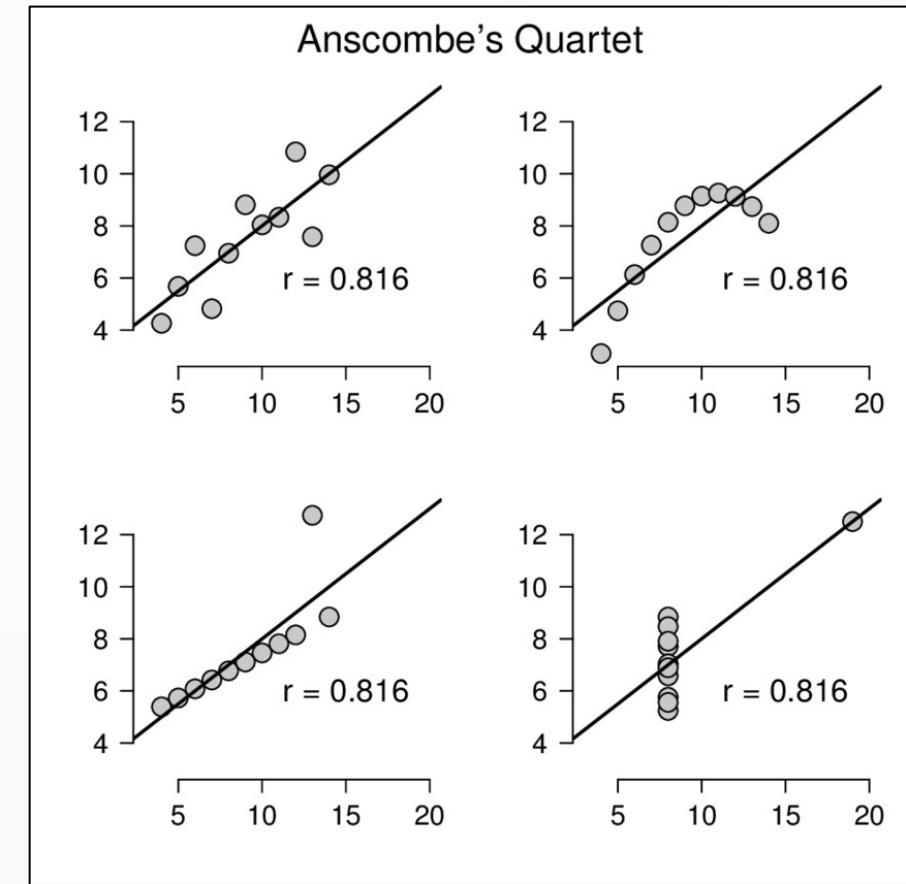
字符串中有几个3?

为什么要可视化?

- 直观的展示信息
- 帮助数据分析与推理

	Set A		Set B		Set C		Set D	
	X	Y	X	Y	X	Y	X	Y
0	10	8.04	10	9.14	10	7.46	8	6.58
1	8	6.95	8	8.14	8	6.77	8	5.76
2	13	7.58	13	8.74	13	12.74	8	7.71
3	9	8.81	9	8.77	9	7.11	8	8.84
4	11	8.33	11	9.26	11	7.81	8	8.47
5	14	9.96	14	8.10	14	8.84	8	7.04
6	6	7.24	6	6.13	6	6.08	8	5.25
7	4	4.26	4	3.10	4	5.39	19	12.50
8	12	10.84	12	9.13	12	8.15	8	5.56
9	7	4.82	7	7.26	7	6.42	8	7.91
10	5	5.68	5	4.74	5	5.73	8	6.89
mean	9.00	7.50	9.00	7.50	9.00	7.50	9.00	7.50
std	3.32	2.03	3.32	2.03	3.32	2.03	3.32	2.03
corr	0.82		0.82		0.82		0.82	
lin. reg.	$y = 3.00 + 0.500x$		$y = 3.00 + 0.500x$		$y = 3.00 + 0.500x$		$y = 3.00 + 0.500x$	

安斯庫姆四重奏 (Anscombe's Quartet)



为什么要可视化?

- 直观的展示信息
- 帮助数据分析与推理

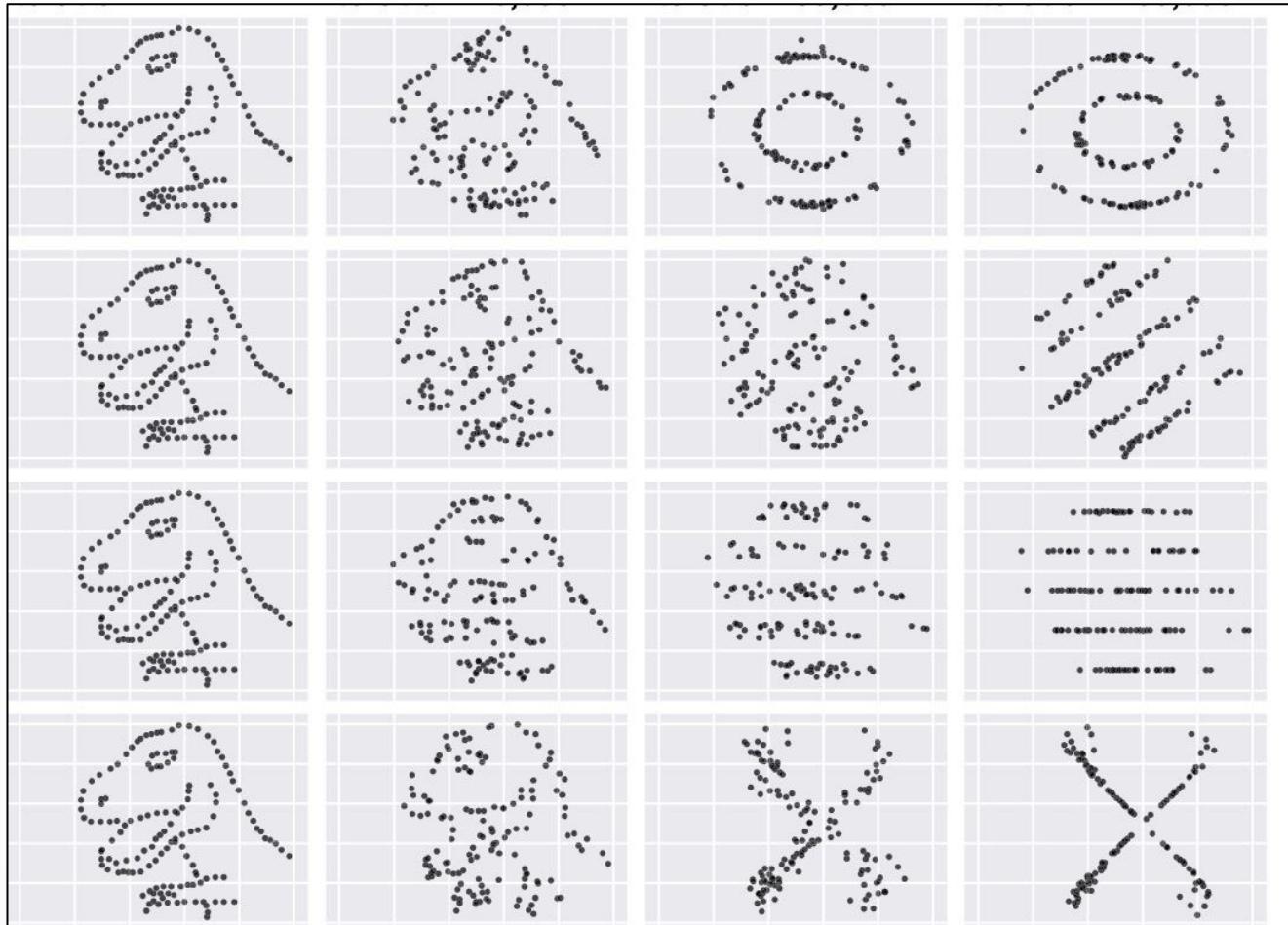
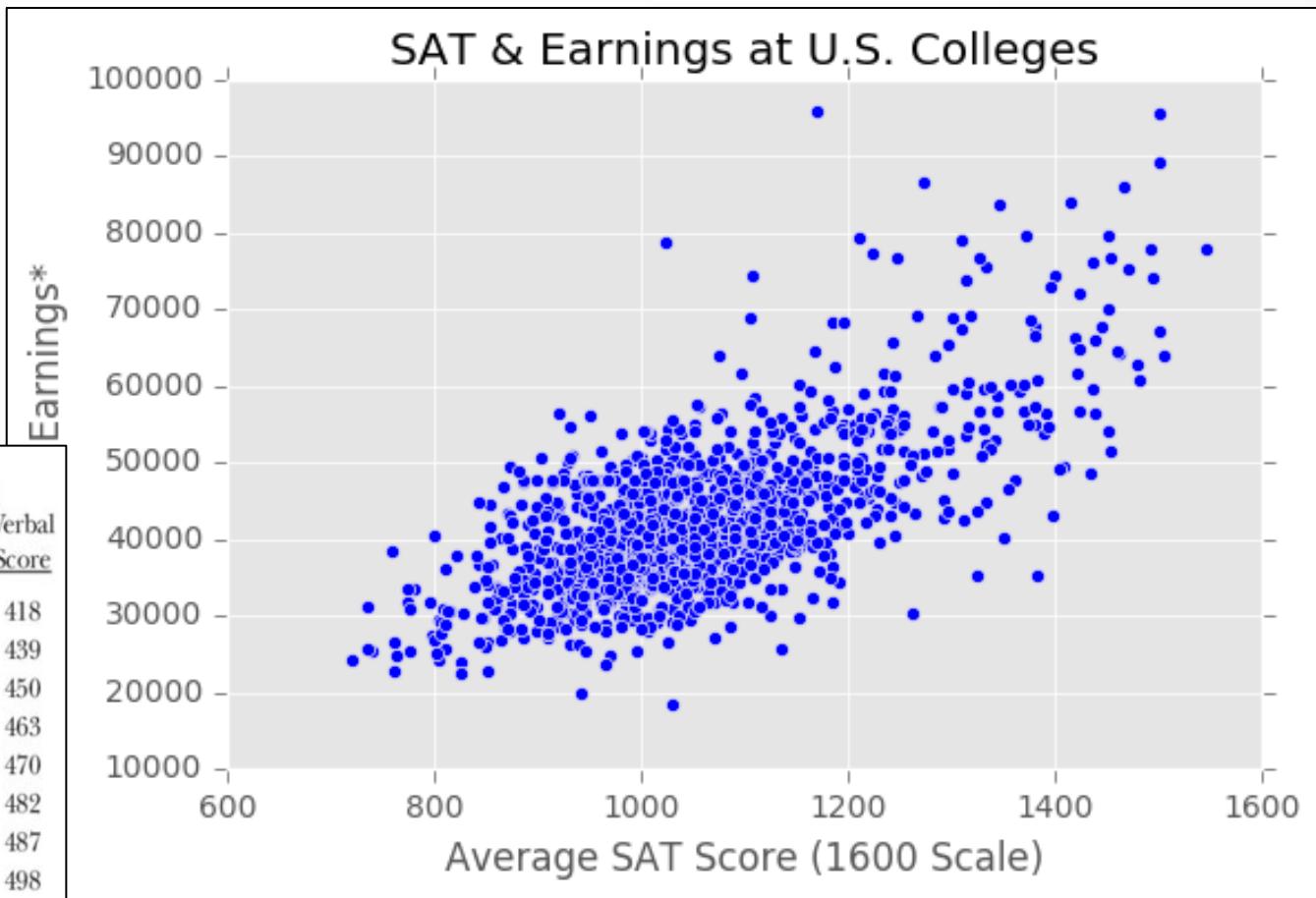


Figure 6. Creating a collection of datasets based on the “dinosaurus” dataset. Each dataset has the same summary statistics to two decimal places: ($\bar{x} = 54.26$, $\bar{y} = 47.83$, $sd_x = 16.76$, $sd_y = 26.93$, Pearson’s $r = -0.06$).

为什么要可视化?

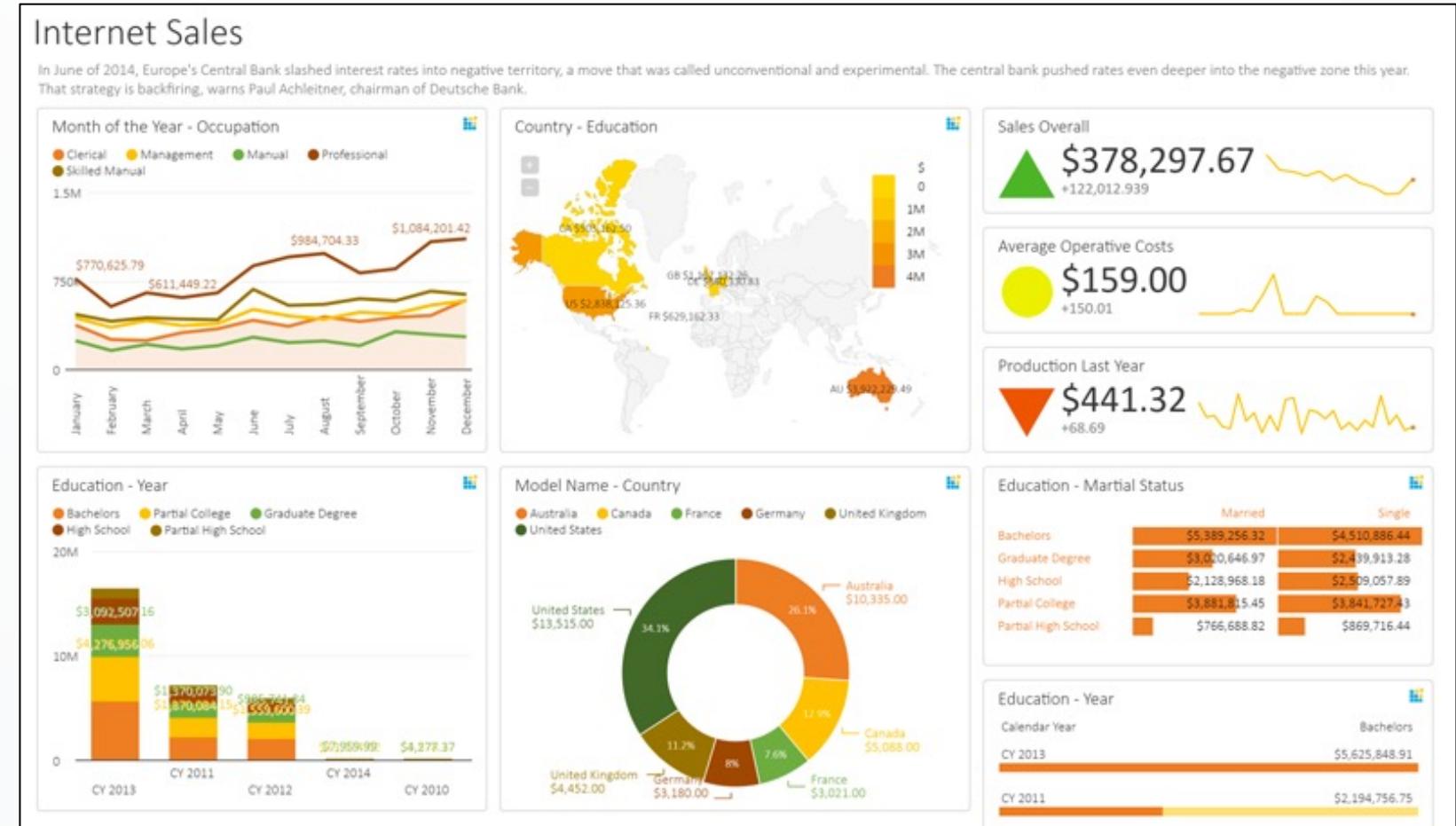
- 直观的展示信息
- 帮助数据分析与推理

<u>Family Income</u>	Black Test-takers		White Test-takers		Total Sample	
	<u>Math Score</u>	<u>Verbal Score</u>	<u>Math Score</u>	<u>Verbal Score</u>	<u>Math Score</u>	<u>Verbal Score</u>
Less than \$10,000	382	381	478	480	418	418
\$10,000 to \$15,000	395	398	478	481	435	439
\$15,000 to \$20,000	400	405	485	488	446	450
\$20,000 to \$25,000	409	413	493	495	460	463
\$25,000 to \$30,000	411	419	495	497	466	470
\$30,000 to \$35,000	419	426	502	504	479	482
\$35,000 to \$40,000	422	430	504	505	484	487
\$40,000 to \$50,000	431	438	510	510	496	498
\$50,000 to \$60,000	441	450	516	514	507	506
\$60,000 to \$70,000	440	450	521	519	512	512
\$70,000 to \$80,000	448	457	528	524	521	518
\$80,000 to \$100,000	461	468	539	534	533	529
More than \$100,000	490	495	568	557	564	555



为什么要可视化?

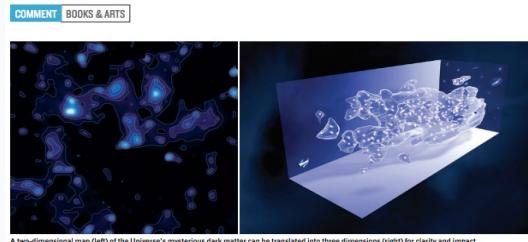
- 直观的展示信息
- 帮助数据分析与推理
- 促进对数据的沟通与交流



信息图、可视化大屏、及可视化报表，就好似对数据照了一张照片，让我们能够通过可视化展示便利的传达、沟通数据

为什么需要可视化?

数据可视化研究如何用图形符号呈现数据信息，是帮助人们快速辨识复杂数据模式的重要手段，是大数据领域不可缺少的重要分支，是人类在探索自然、科学发现过程中不可缺少的重要技术



Picturing science

Katy Börner weighs up a lavish, lab-friendly guide to transforming dry data into insightful images.

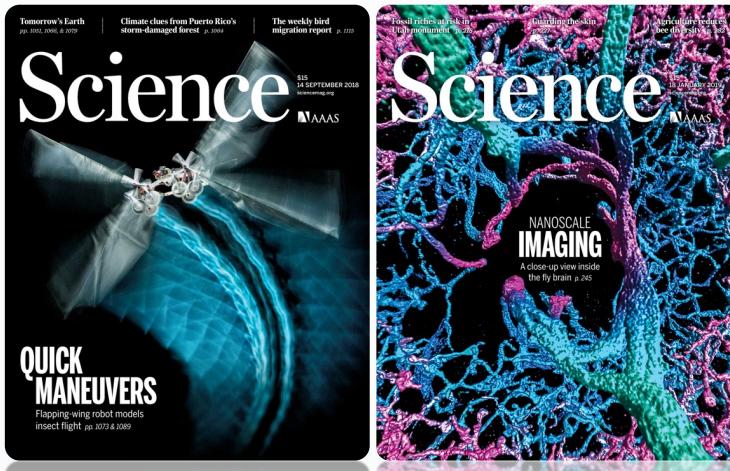
Images of scientific results matter. They support data exploration and communication, and are particularly valuable in the age of big data. They help to transcend disciplinary, cultural and language barriers, and can be used to break down barriers in art museums. Others have gone viral on the Internet, such as Darwin's finches — and the way their beak shape — have changed our view of the world forever.

Given the importance of imagery in science, it is surprising that few scholars are properly trained in graphic design or

重要性：尽管可视化在科学发现过程中至关重要，但令人惊奇的是，很少有学者接受了可视化设计相关的培训

"Given the importance of imagery in science, it is surprising that few scholars are trained in graphic design."

430 | **Nature** | VOL 487 | 26 JULY 2012



科学性：Science连续发文指出：借助可视化手段将人机智能有机结合，形成可视分析环境，可有效提升数据关联分析的效率

高效性："A good user interface is worth a thousand petabytes". The Big Picture for Big Data: Visualization, Science 14 Feb 2014: Vol. 343, Issue 6172, pp. 730

急迫性：

2018年美国商务部工业安全署将“可视化、自动分析算法、上下文感知计算”列入出口管制框架，对我国实施技术封锁

2022年美国国家科学技术委员会(NSTC)发布了新版《关键和新兴技术清单》进一步加强了对上述技术的管控

国家《十四五》规划

可视化被列入数字经济重点产业，并明确指出要“**推动大数据**采集、清洗、存储、挖掘、分析、**可视化算法等技术创新**。”

中国科技创新2030“新一代人工智能”和“大数据”专项均将可视化列为大数据智能急需突破的关键共性技术。

课程大纲

- 什么是可视化?
- 为什么要可视化?
- **可视化的类型与发展趋势**
- 怎样对数据进行可视化?
- 如何对可视化进行评估?
- 怎样学习可视化?
- 可视化应用案例

可视化的类型及发展趋势

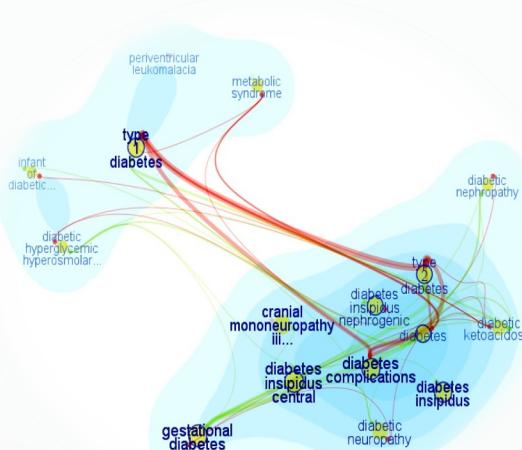
可视化自诞生之初到现在，其发展经历了三个阶段

解决数据看不见的问题



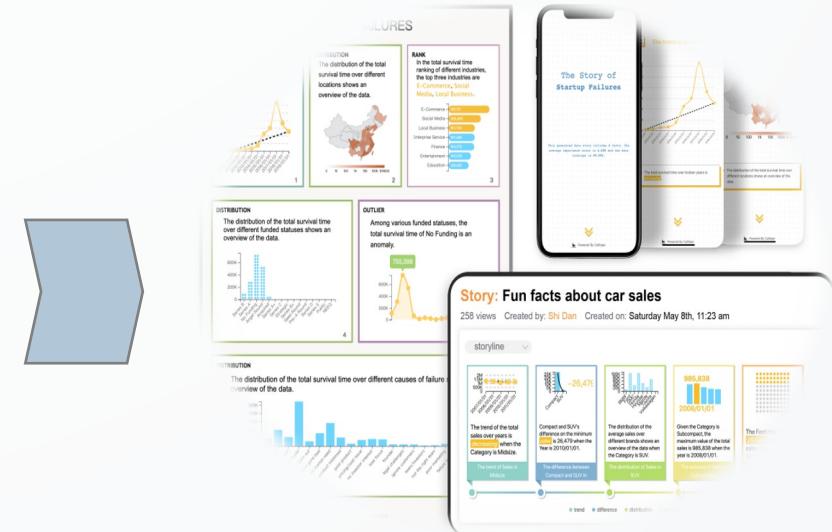
科学研究需求驱动下的
科学可视化
1990

解决数据难理解的问题



大数据时代的
信息可视化与可视分析
2012 → 现在

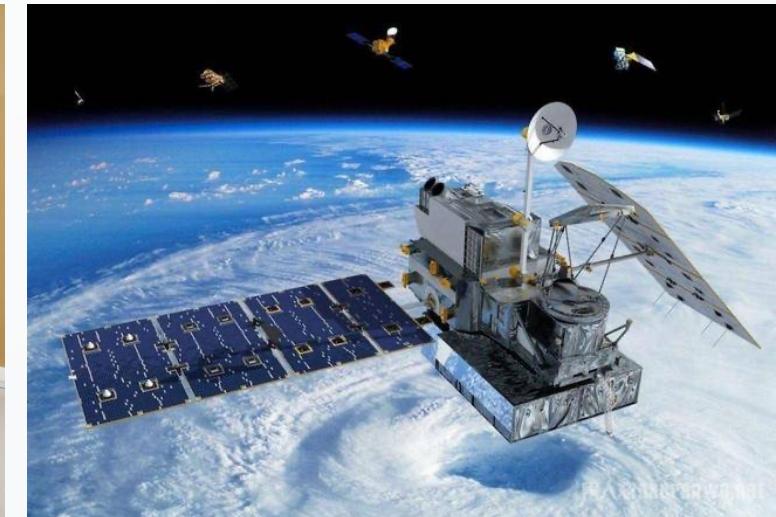
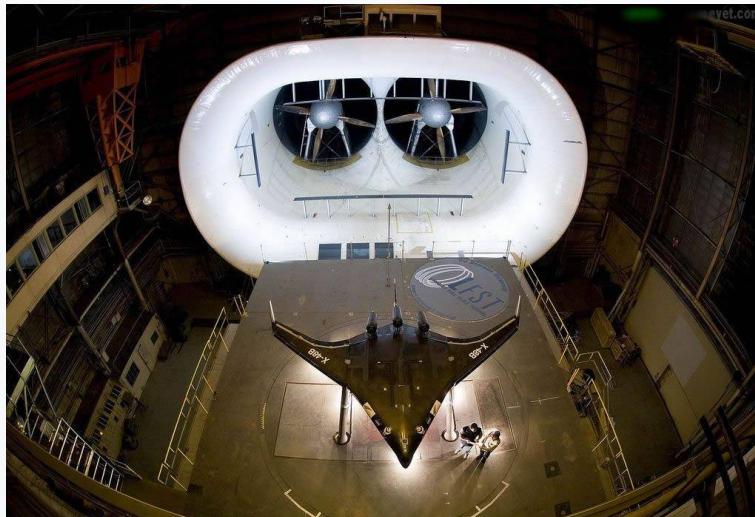
解决在线实时分析与决策的问题

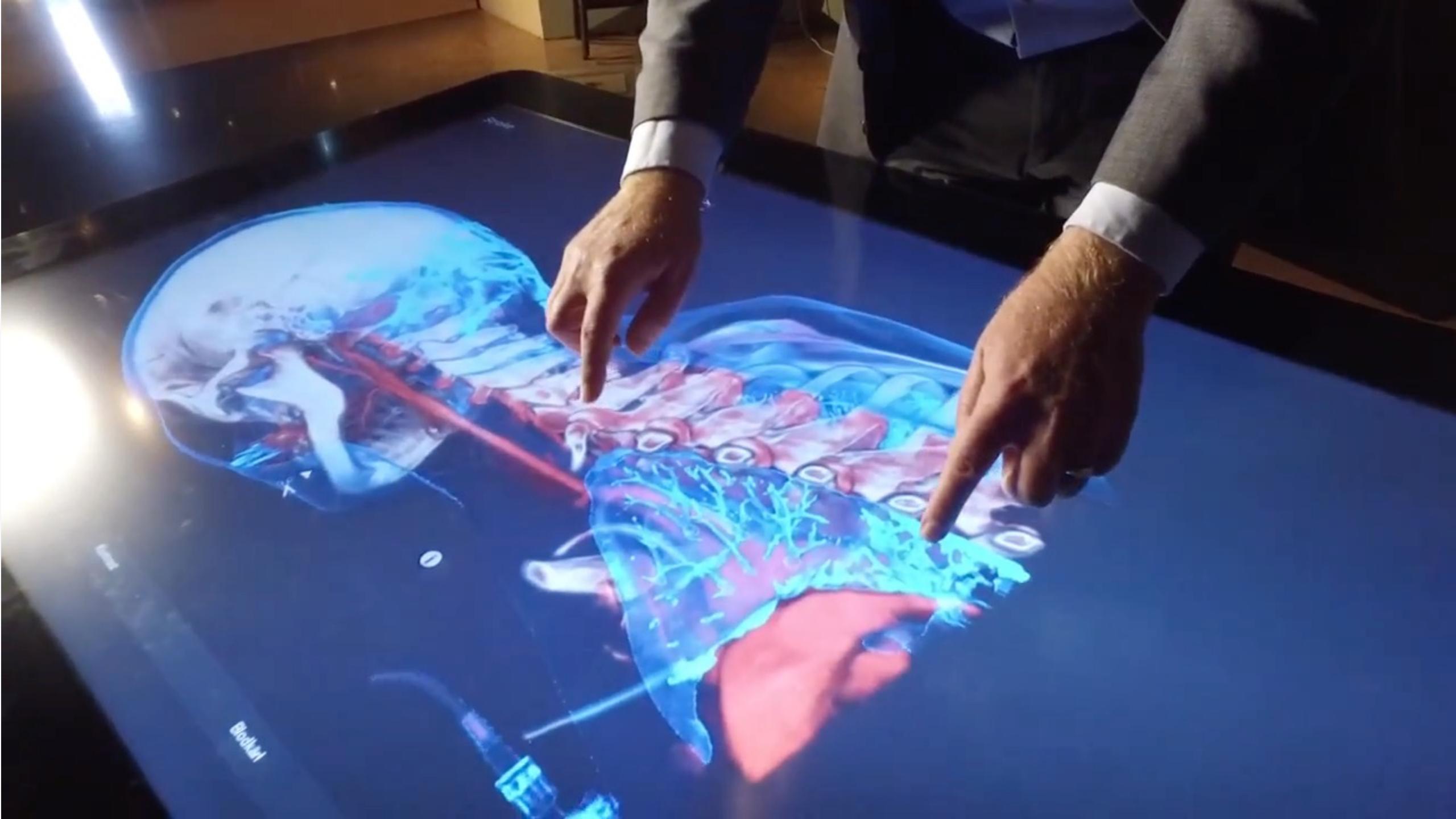


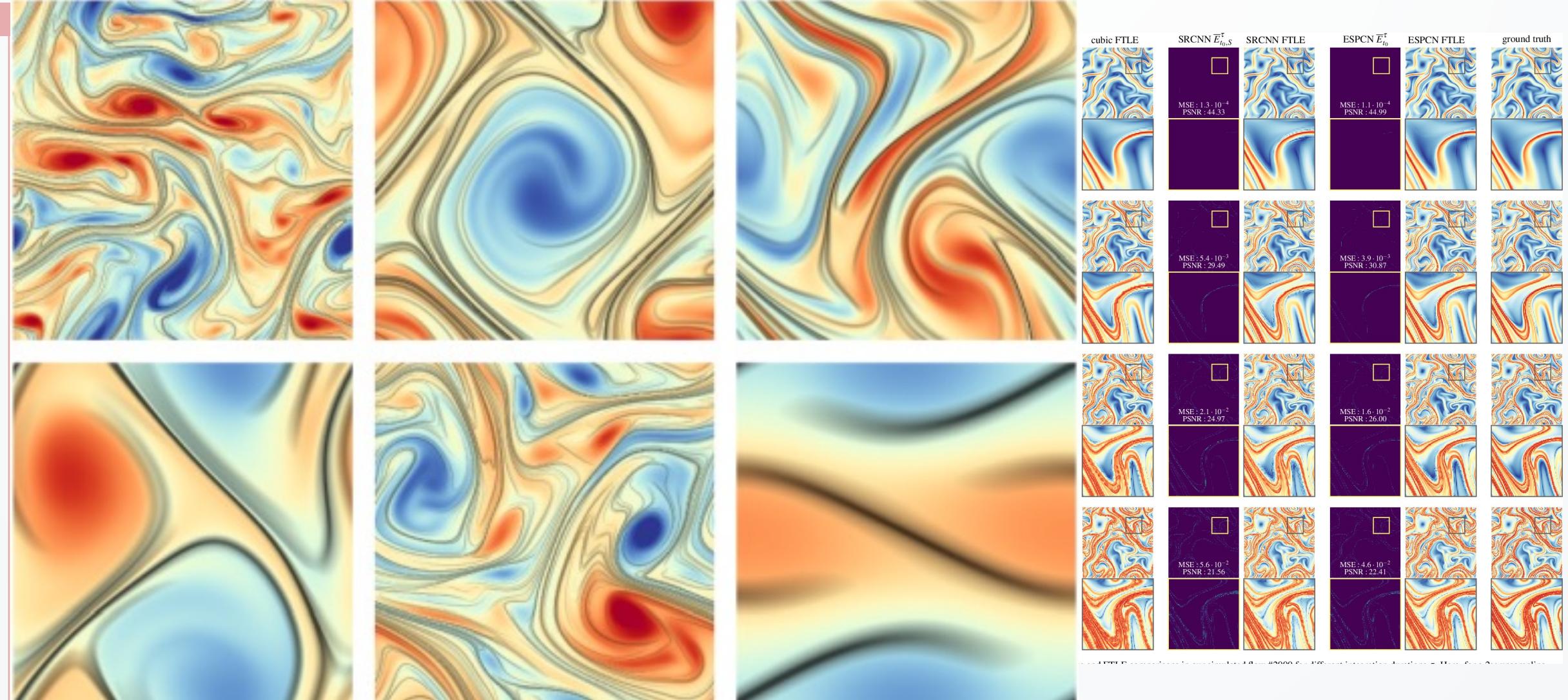
人工智能时代的
先进数据分析与智能可视化
现在 -> 未来

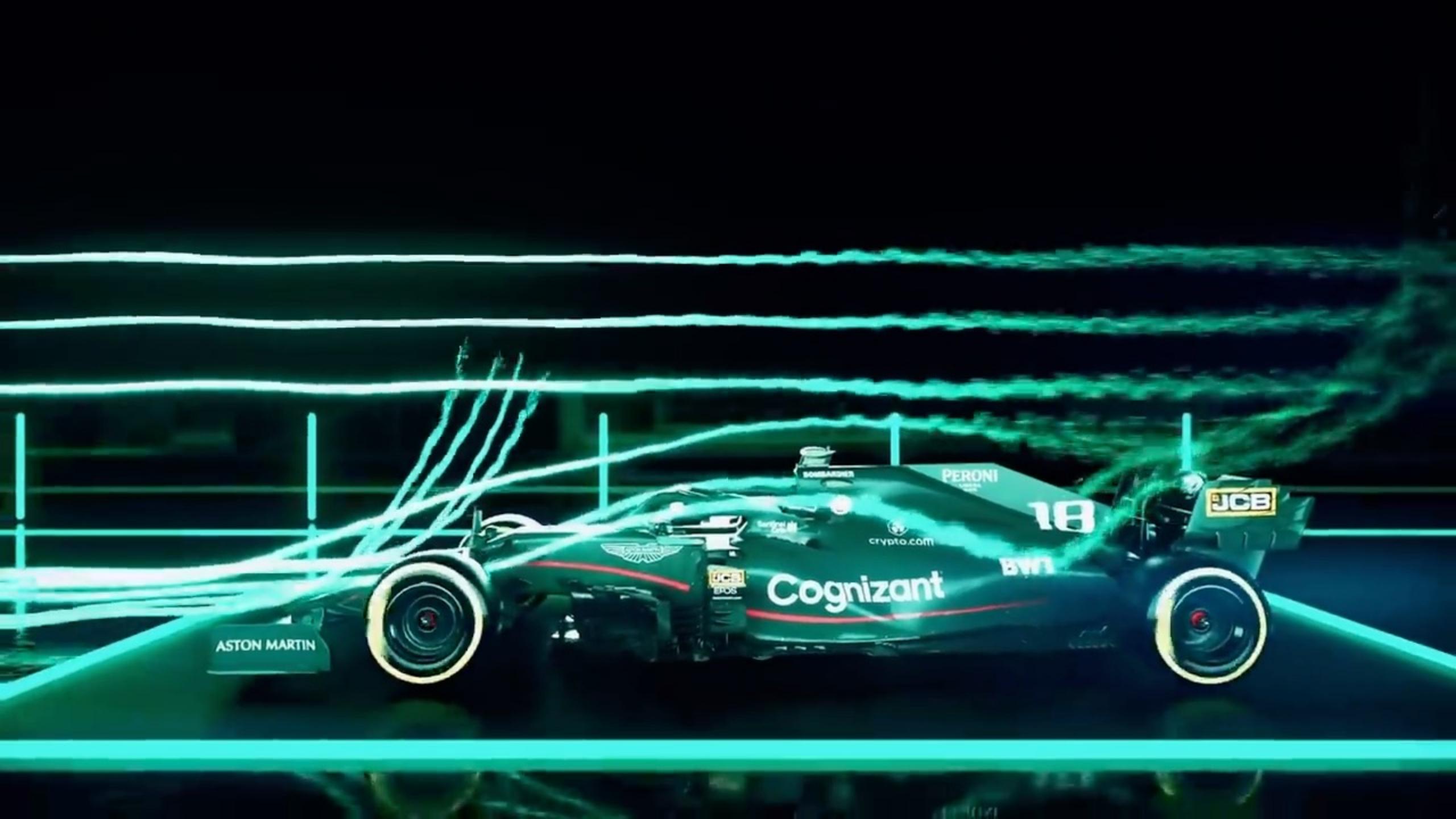
科学可视化

- 通过科学仪器采集到的数据 / 通过科学实验探测到的数据
- 解决科学数据看不见的问题









ASTON MARTIN

Cognizant

18

JCB

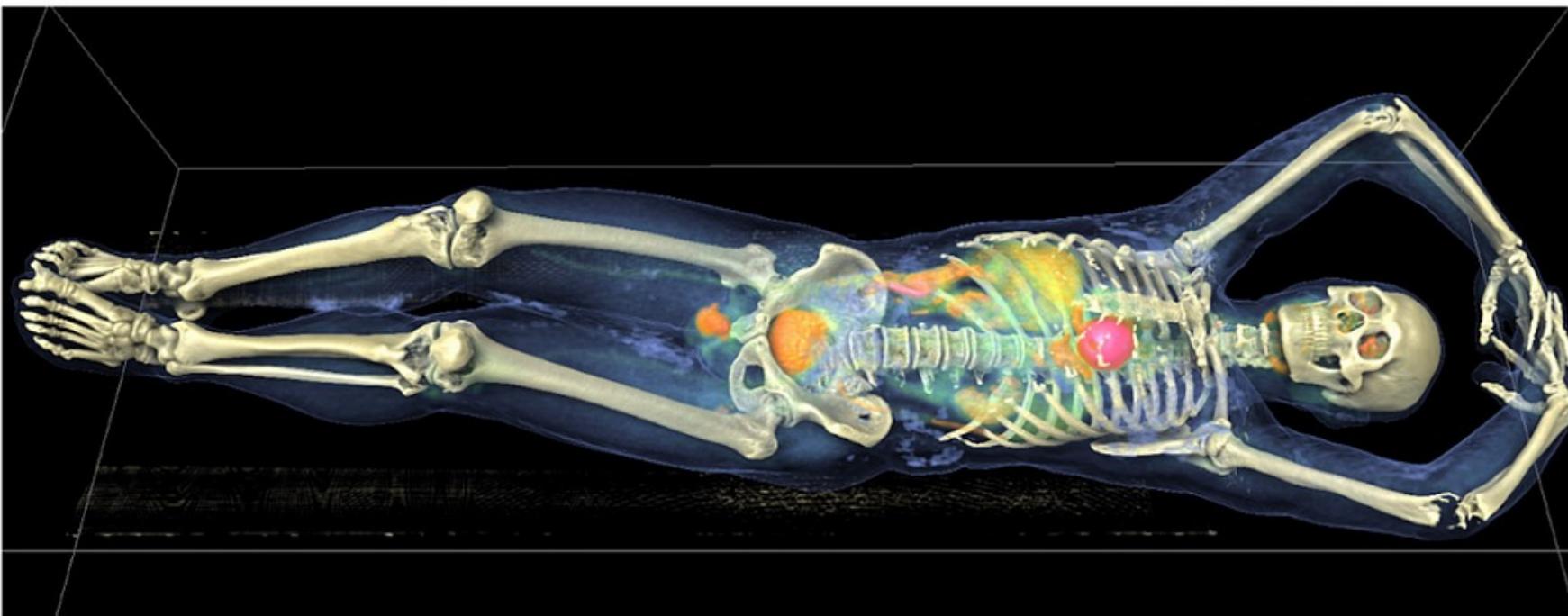
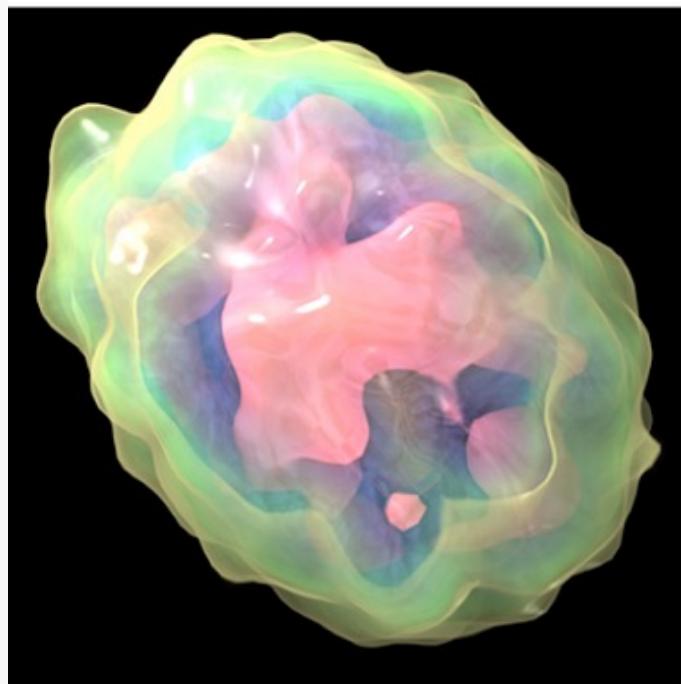
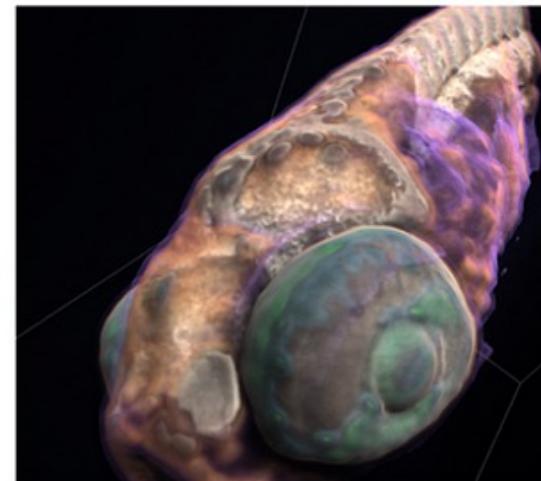
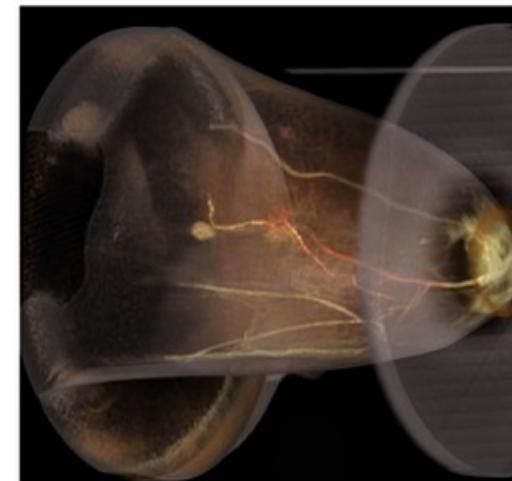
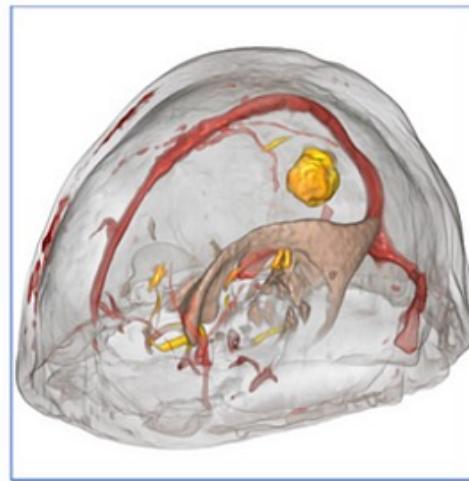
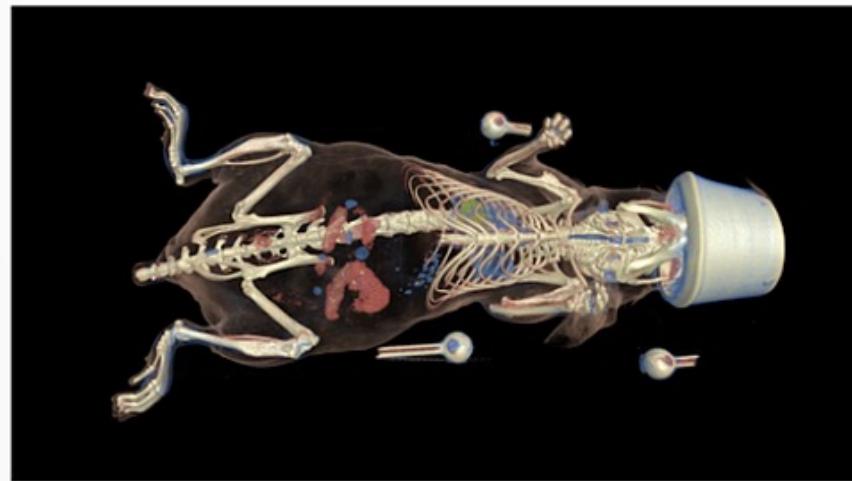
BMW

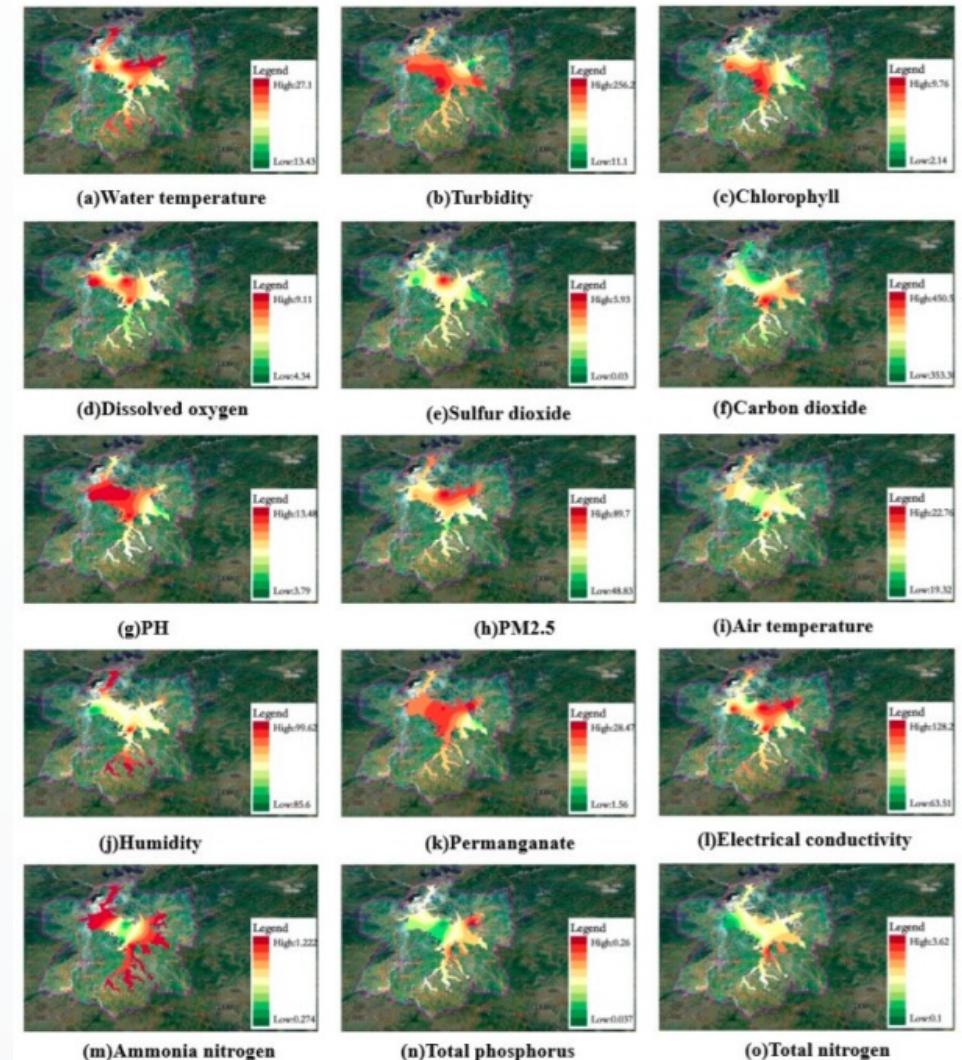
UFC EPOS

PERONI

crypto.com

SCHUMACHER





可视化的类型及发展趋势

可视化自诞生之初到现在，其发展经历了三个阶段

解决数据看不见的问题



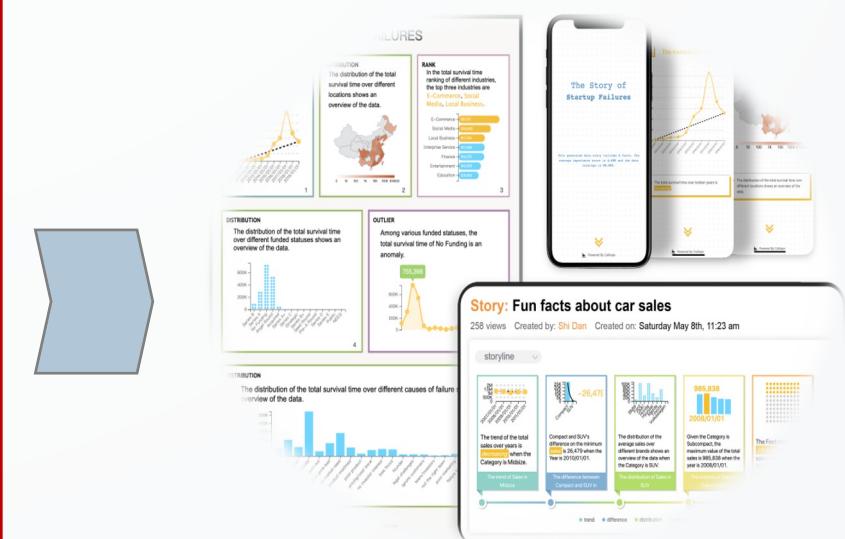
科学研究需求驱动下的
科学可视化
1990

解决抽象数据难理解的问题



大数据时代的
信息可视化与可视分析
2012 → 现在

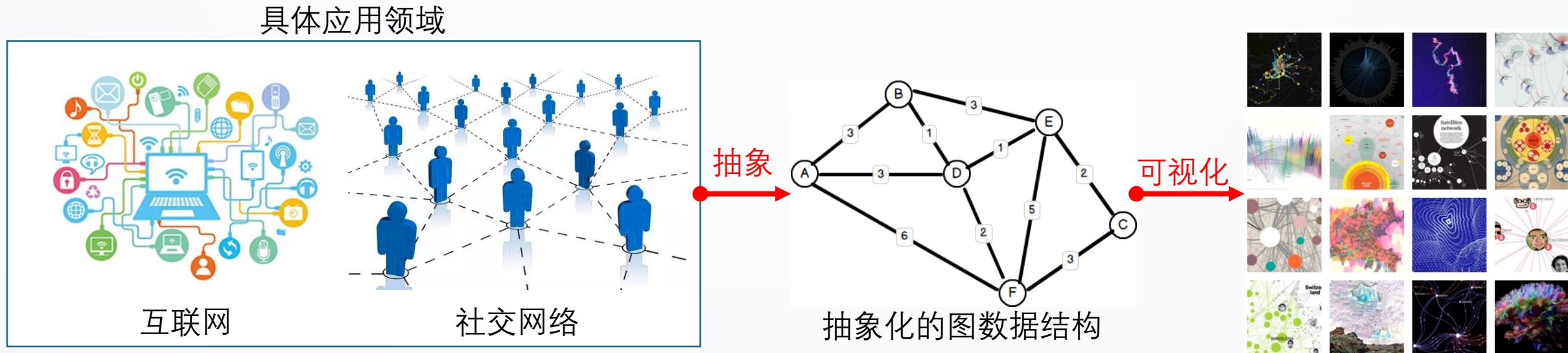
解决在线实时分析与决策的问题



人工智能时代的
先进数据分析与智能可视化
现在 -> 未来

信息可视化

- 信息可视化 - 是对抽象数据 直观的图形化展现
- 抽象数据 - 是对具体应用数据的高度概括与抽象，往往存储在数据结构当中



keithurbahn

brianstelter



SOCIALFLOW



facebook

December 2010



**The Contiguous United States
Visualized by distance to the nearest McDonald's**

Created by Stephen Von Worley

<http://www.weathersealed.com/tags/maps/>

Location data courtesy of AggData

<http://www.aggdata.com/>

Vision Statement

Six Ways to Find Value in Twitter's Noise

Text by Scott Berinato; data and visualization by Jeff Clark

It's easy to dismiss Twitter as jabber, but smart marketers will recognize it as a stream of free consumer data to be mined in near-real time. Online visualization tools can help pinpoint what consumers are reading and sharing, elucidate memes in the chatter, and unearth trends. To show marketers how they can gain insight from Twitter, we captured more than a half million tweets containing the word "iPad" that were broadcast during the product's launch weekend in April. We then mapped key words that appeared in those tweets on the graph below.

HBR Reprint F10062

The iPad Launch by the Numbers

547,898

TOTAL TWEETS USING "IPAD" DURING PRODUCT'S LAUNCH WEEKEND

13%

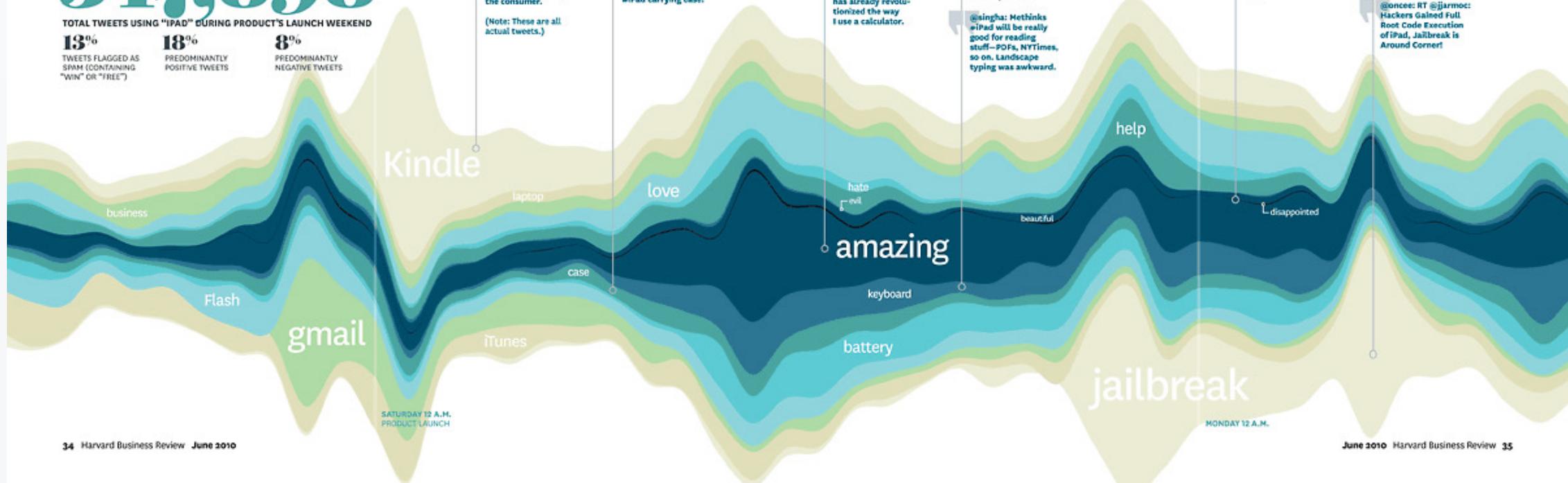
Tweets flagged as spam (containing "win" or "free")

18%

Predominantly positive tweets

8%

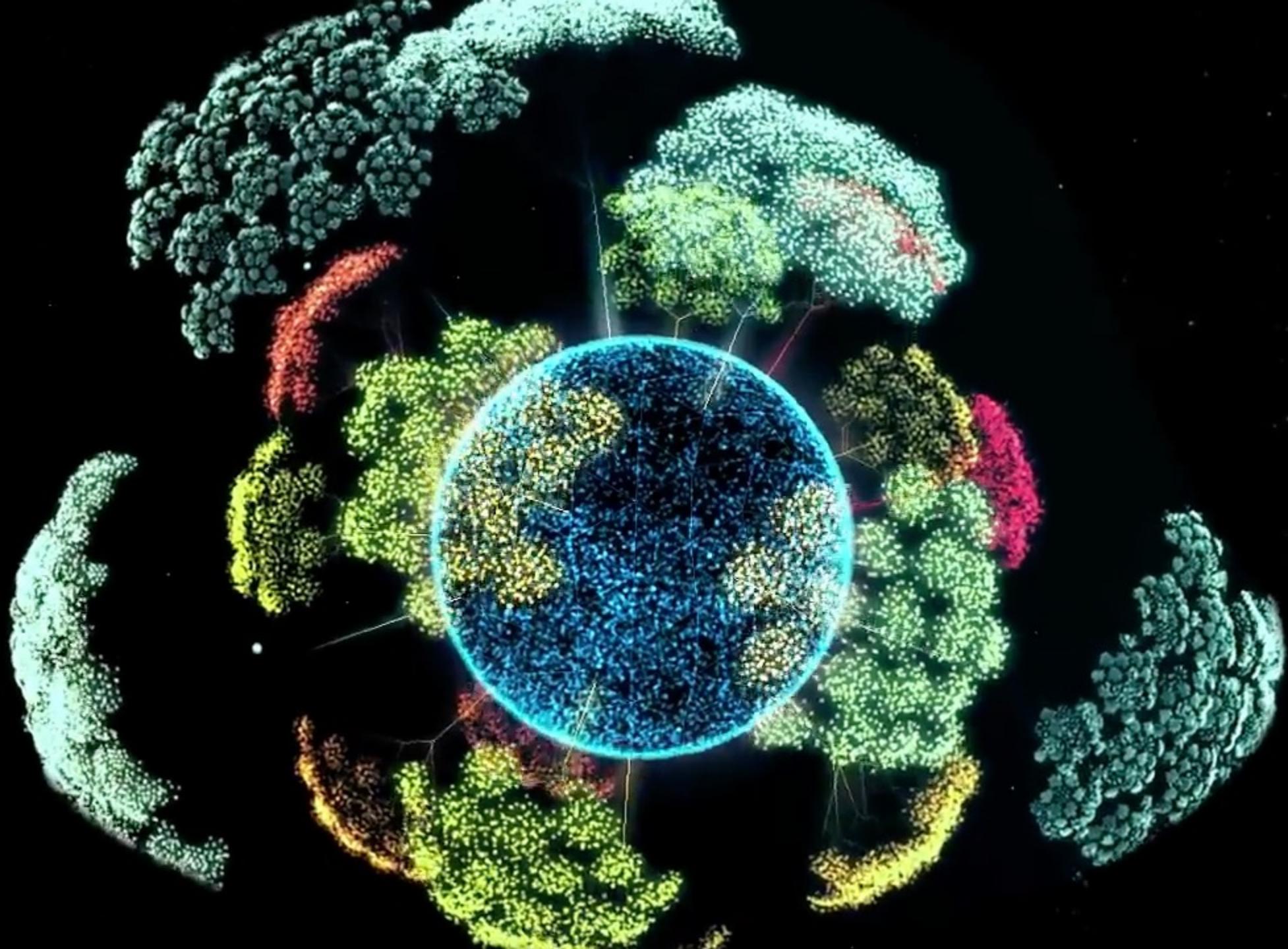
Predominantly negative tweets



Scott Berinato is an editor at HBR. Jeff Clark is a developer and information visualizer based in Toronto.

How to Read This Graph

This stream graph shows tweet volume over time. Each rivulet illustrates the proportion of iPad tweets containing a given word, such as **HELP**. Color is used only to distinguish rivulets.



可视化的类型及发展趋势

可视化自诞生之初到现在，其发展经历了三个阶段

解决数据看不见的问题



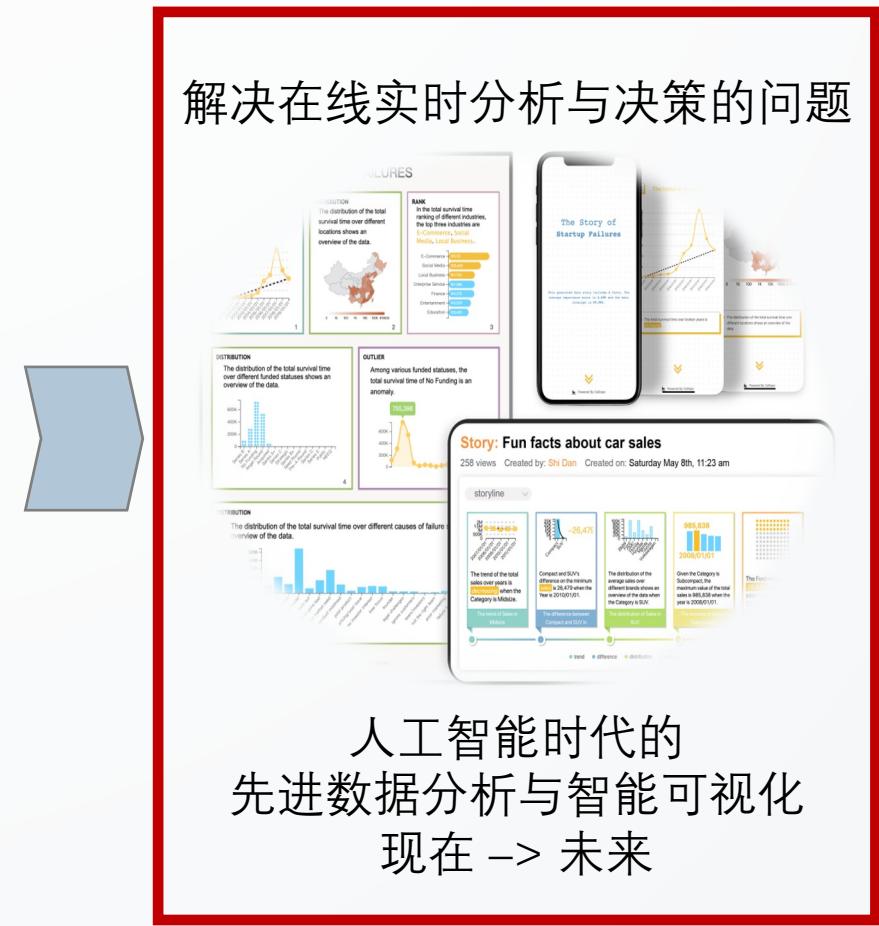
科学研究需求驱动下的
科学可视化
1990

解决抽象数据难理解的问题



大数据时代的
信息可视化与可视分析
2012 → 现在

解决在线实时分析与决策的问题



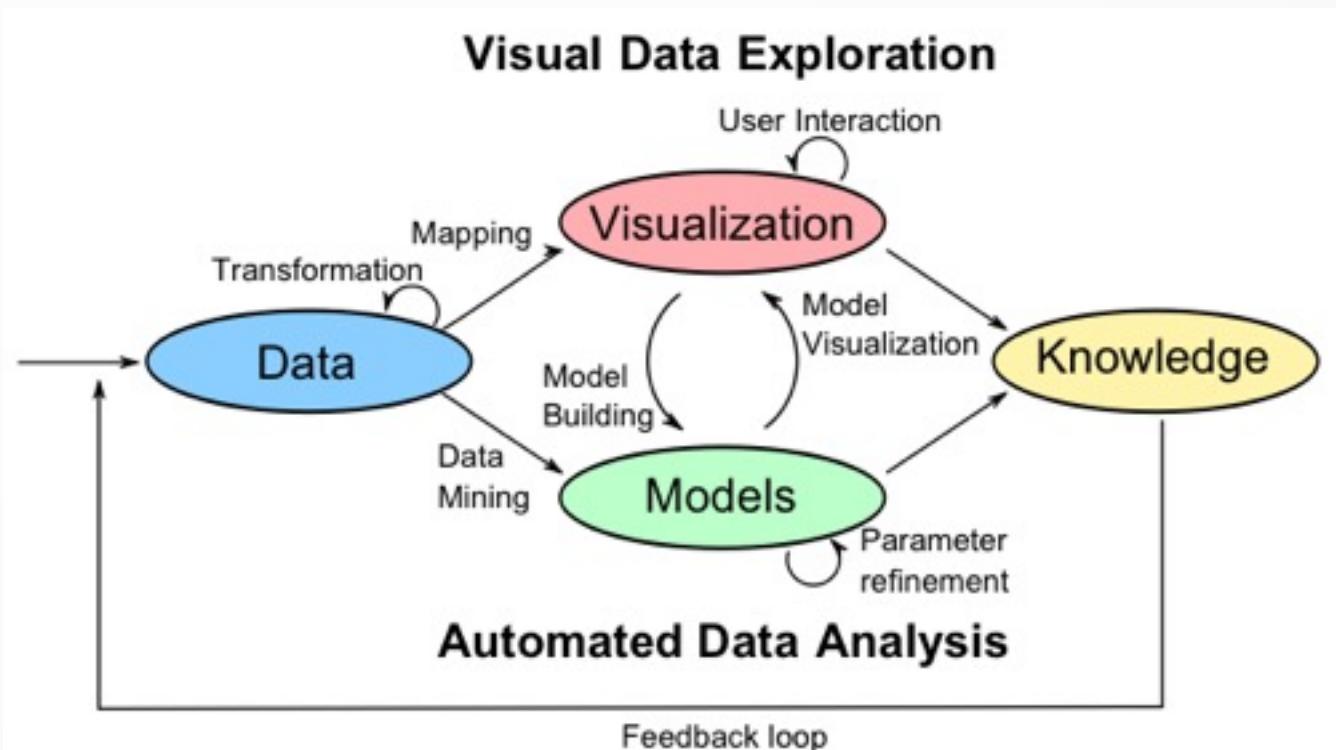
人工智能时代的
先进数据分析与智能可视化
现在 -> 未来

可视分析

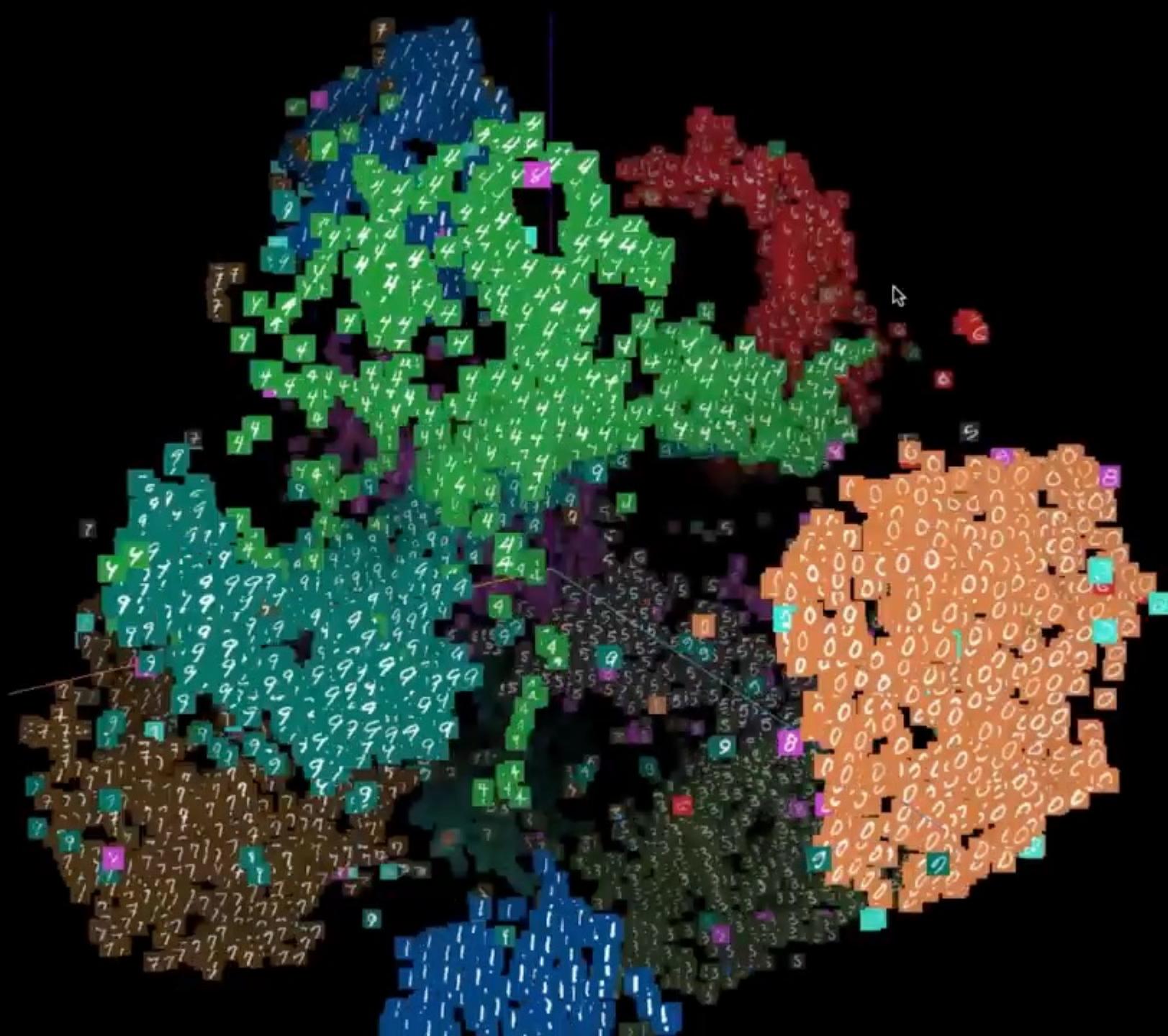
- 可视化 + 数据分析 + 人机交互 (人在回路的数据分析 (Keep Human in the Analysis Loop))
- 可视化能够帮助人们发现数据规律 (What), 而可视分析则帮助人们找到原因 (Why)



对分析结果的直观展现，及交互式反馈，是一个跨领域的方向





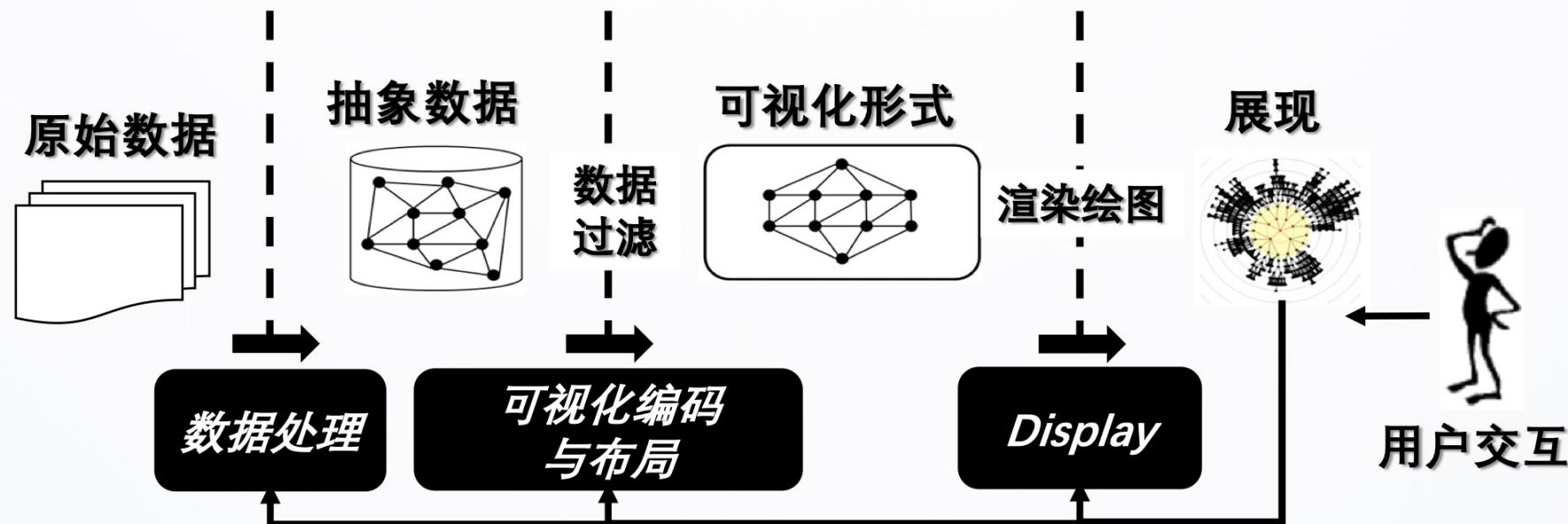


课程大纲

- 什么是可视化?
- 为什么要可视化?
- 可视化的类型与发展趋势
- **怎样对数据进行可视化?**
- 如何对可视化进行评估?
- 怎样学习可视化?
- 可视化应用案例

怎样对数据进行可视化?

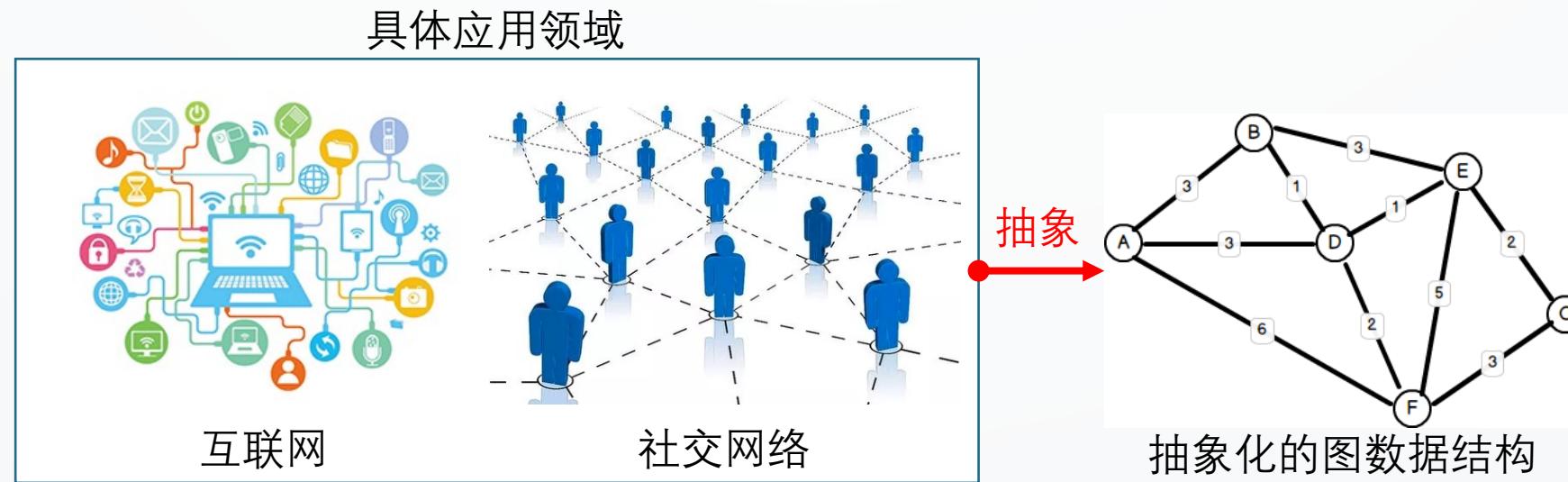
从技术角度而言



信息可视化参考模型

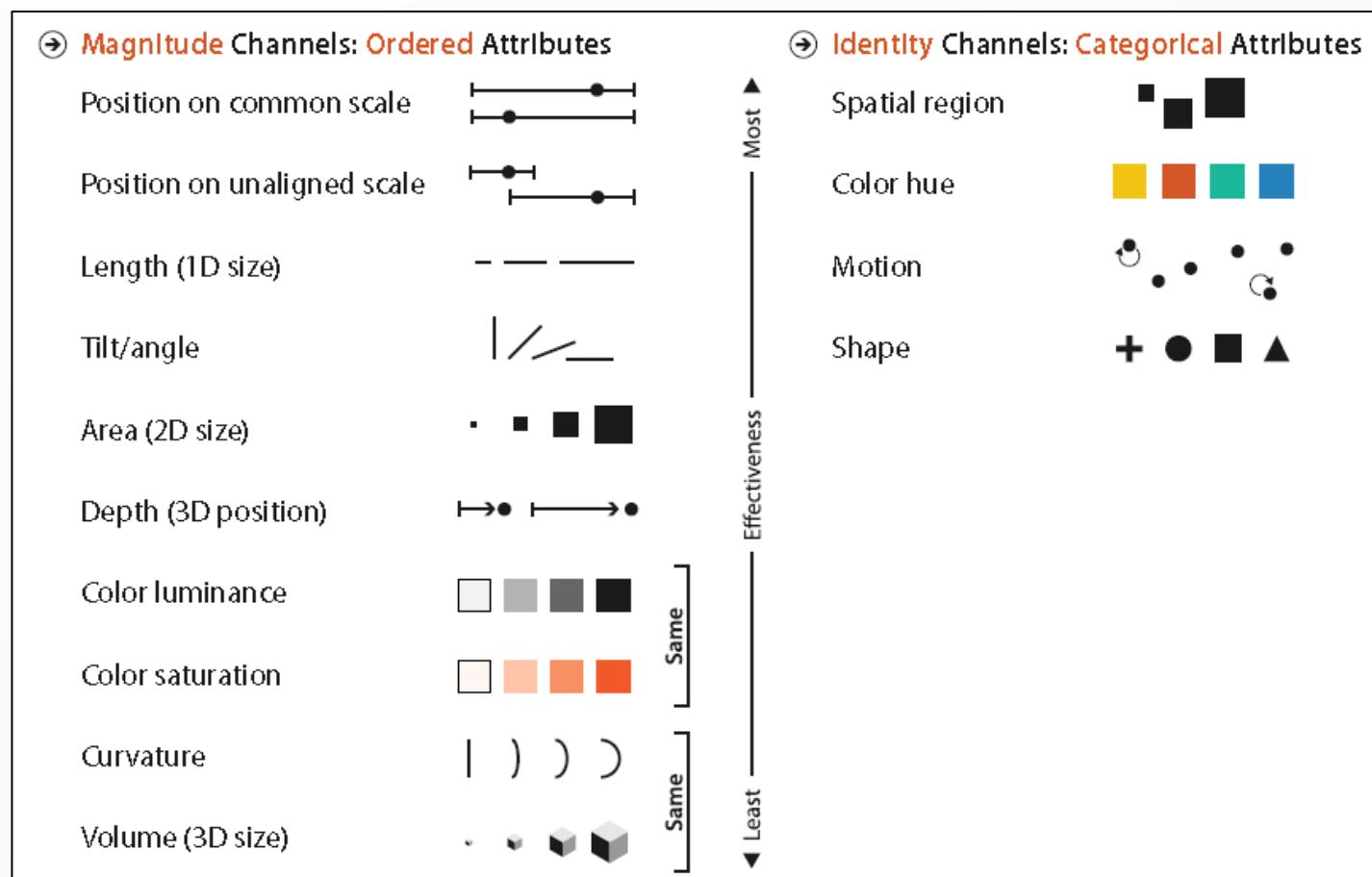
第一步：将数据转换为抽象形式

- **数据清洗**: 借助于各种手段（过滤，采样，压缩，分类，聚类等）对数据进行清理，去除数据中的噪声，提取有用信息
- **数据变换**: 改变数据的表达存储形式，例如 无结构数据的结构化转换
- **数据抽象**: 提取应用数的抽象数据展现形式



第二步：可视化设计

- 数据本身并没有形态的，需要通过设计才能够进行图形化的表达
- 可视化设计的基本准则：**信**（可视化需要准确表达数据）、**达**（可视化需要高效传达数据中的信息）、**雅**（可视化需要美观）
- 可视化编码：如何用图形符号（点、线、面）及属性（颜色、形状、大小等）展示数据元素及属性



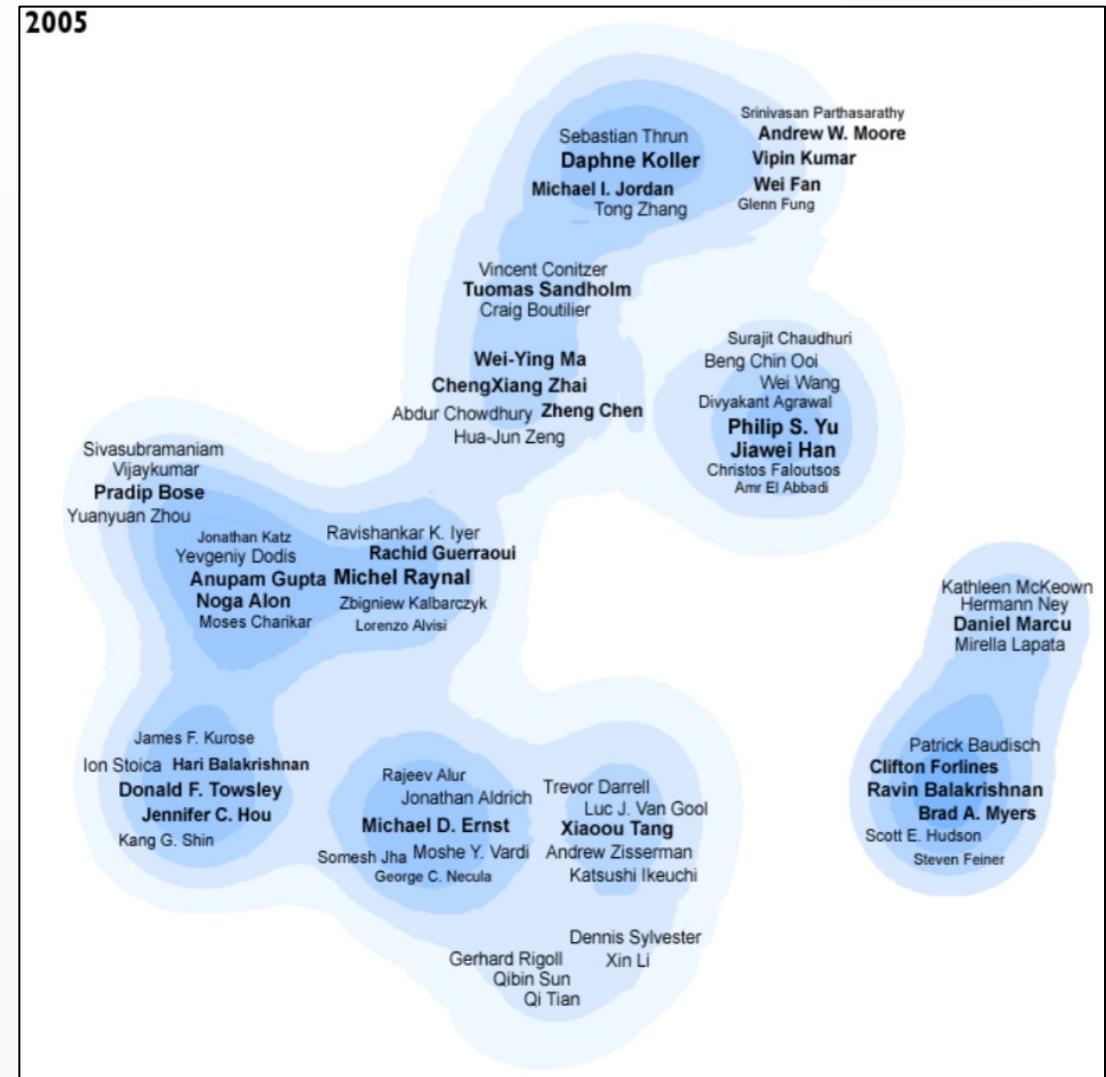
第三步：可视化布局



- 通过算法自动计算可视化元素在屏幕上的位置
- 往往是一个优化过程，需要考虑诸多信息呈现及美观方面的因素
- 例如左图所示的可视化布局需要避免点的遮挡与线的交叉

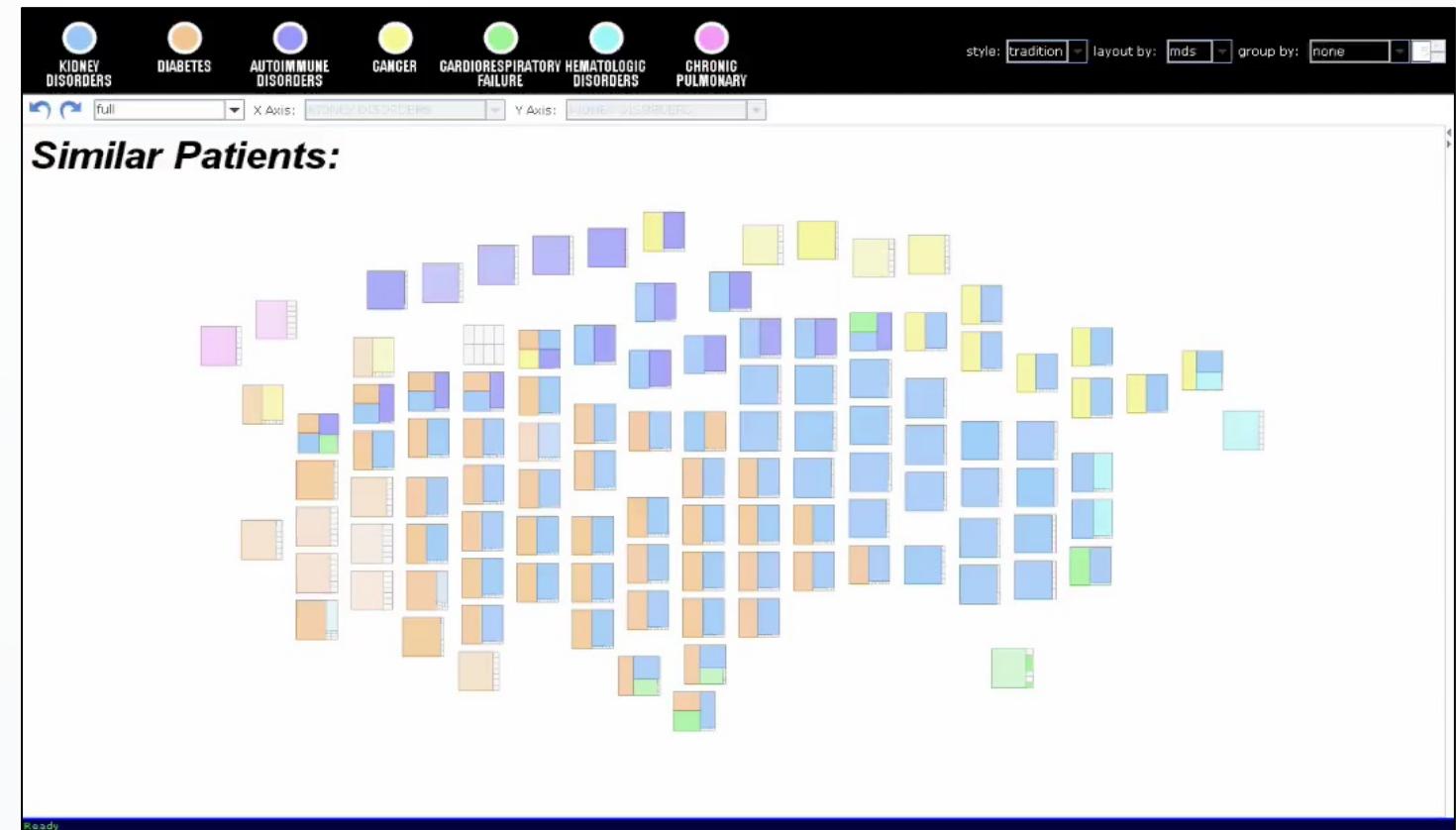
第四步：可视化渲染

- 根据设计及布局，将数据显示在屏幕上之上
- 渲染使用了计算机图形学中的相关技术，可视化往往是计算机前端的工作，渲染往往采用支持前端网页显示的渲染技术，例如，SVG，H5 Canvas，以及 Web GL 等



第五步：可视化交互与动画

- 当数据量较大，无法完整显示在可视化视图当中时，交互能够帮助用户完整的浏览数据
- 在进行可是分析的过程中，当用户对可视化所展示的数据分析结果有异议时，交互能够帮助系统收集必要的用户反馈
- 当数据发生变化时，还需要设计开发相应的动画，用于平滑的展现数据中的变动，以便用户追踪变动的数据，并分析相应的原因



常用的可视化开发工具

D3.js

<https://d3js.org/>



The screenshot shows the official D3.js website. At the top, there is a navigation bar with links to 'Overview', 'Examples', 'Documentation', and 'Source'. Below the navigation is the D3.js logo ('d3') and the text 'Data-Driven Documents'. A prominent orange diagonal banner on the right says 'Fork me on GitHub'. The main content area features a large, dense collage of various data visualizations, including maps, treemaps, network graphs, and statistical charts, demonstrating the library's capabilities.

D3.js is a JavaScript library for manipulating documents based on data. **D3** helps you bring data to life using HTML, SVG, and CSS. D3's emphasis on web standards gives you the full capabilities of modern browsers without tying yourself to a proprietary framework, combining powerful visualization components and a data-driven approach to DOM manipulation.

See [more examples](#).

常用的可视化开发工具

Three.js

<http://threejs.org/>

The image displays a grid of 3D visualizations created using the Three.js library. The grid is organized into four rows and seven columns. Each cell in the grid contains a thumbnail of a different project. The projects include:

- Row 1: VOID, RESONANCE, WE ARE ACTIVATION NODEPLUS, a dark car, a circular graphic with 'OMA'.
- Row 2: A globe with a banner, DENNIS, THE Fallen of World War II, THE BOAT.
- Row 3: A red car, a landscape scene with a river, a checkered floor, a blue flower, GIVE SHAPE.
- Row 4: THE DILLA DIMENSION, a futuristic interior, WAY TO GO, URBAN GALAXY ONLINE, LEIA, a figure holding a baton, BULLSEY'S PLAYGROUND, a flamingo's head.

On the left side of the grid, there is a sidebar with the following links:

- three.js ^{r77}
- documentation
- examples
- download
- github
- stackoverflow
- irc
- editor

Below these links are two small thumbnail images:

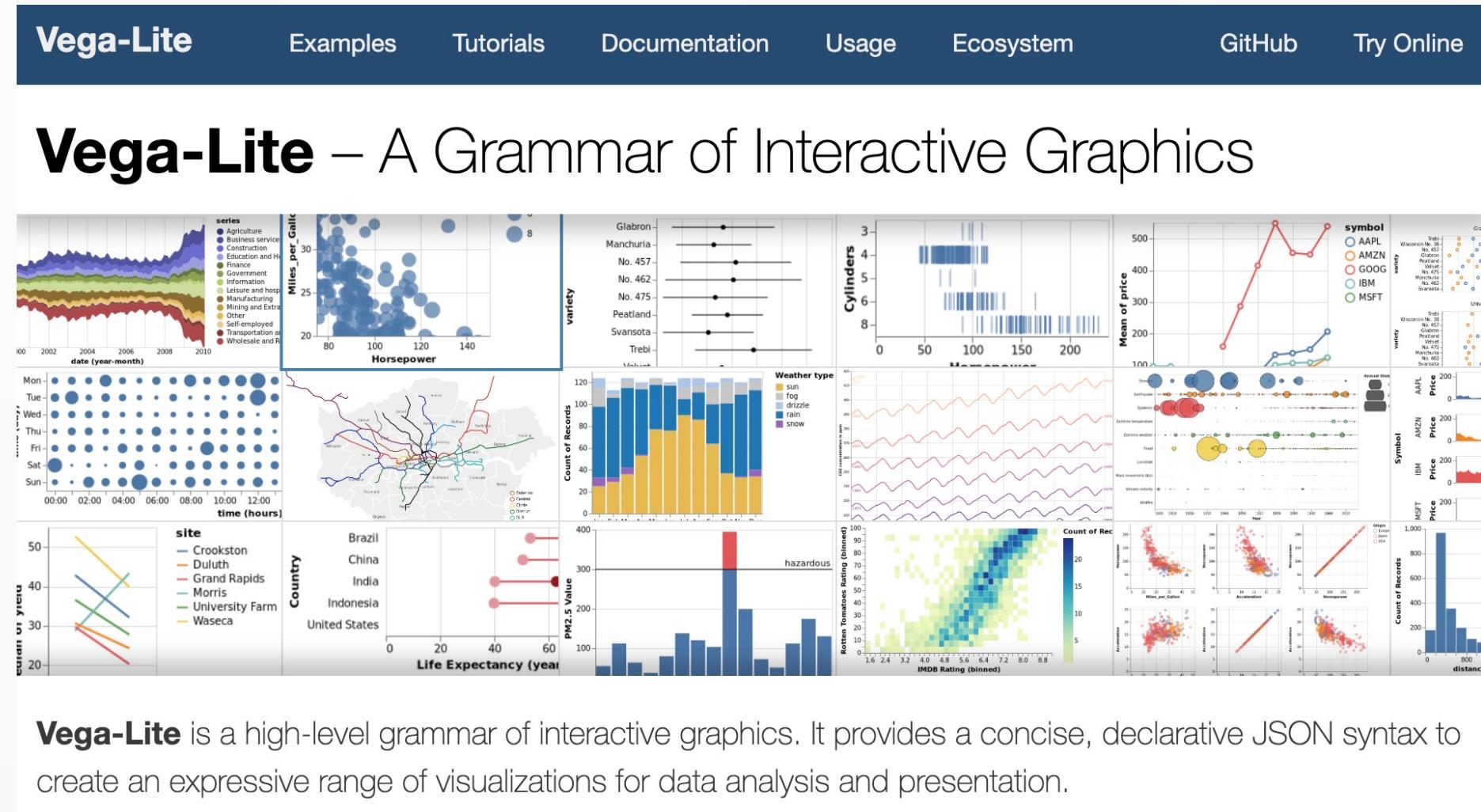
- Intro to WebGL with Three.js
- Game Development with Three.js

At the top right of the grid, there is a link labeled "more projects".

常用的可视化图表库及工具

Vega-Lite

<https://vega.github.io/vega-lite/>

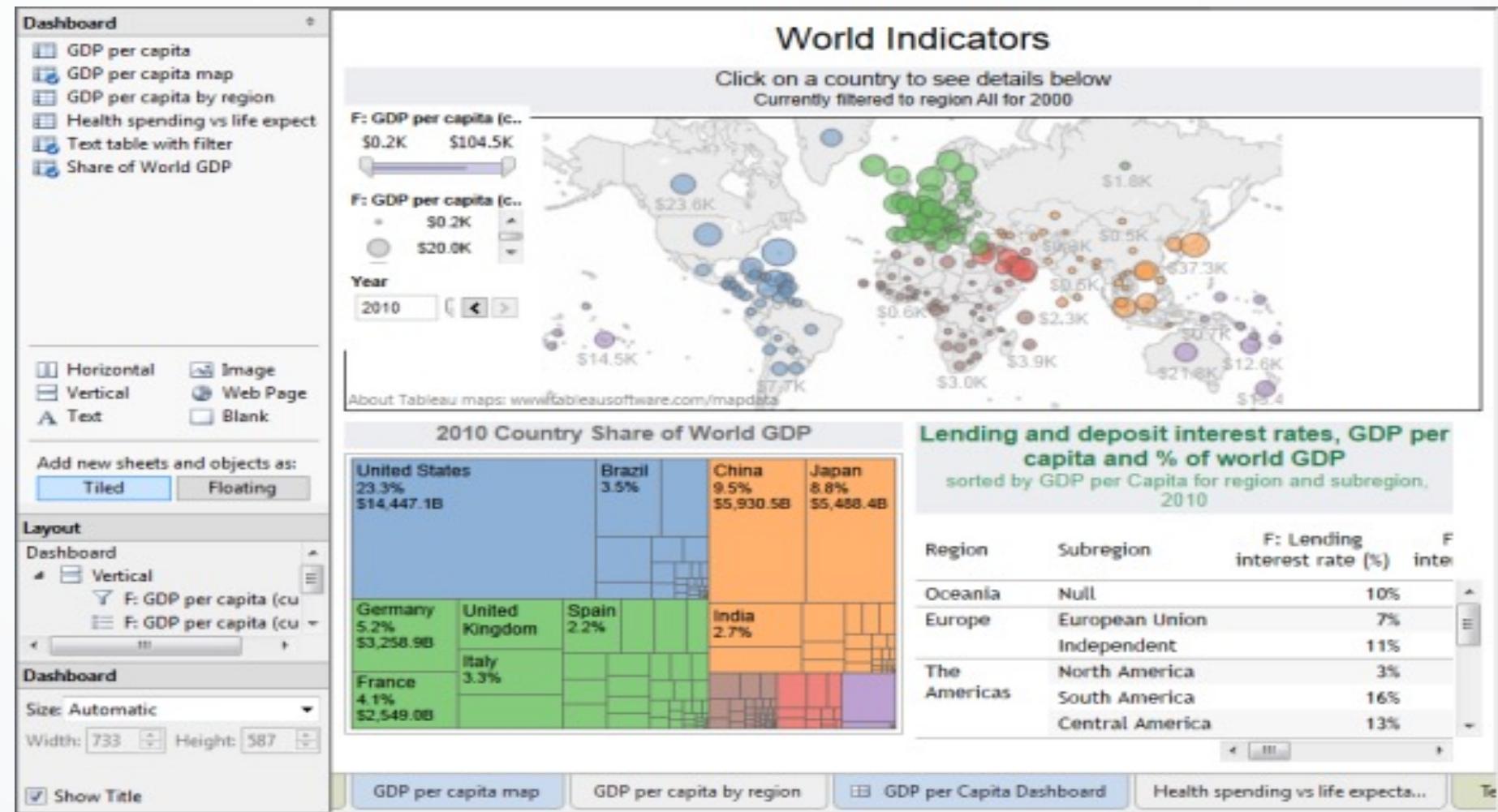


常用的可视化图表库及工具

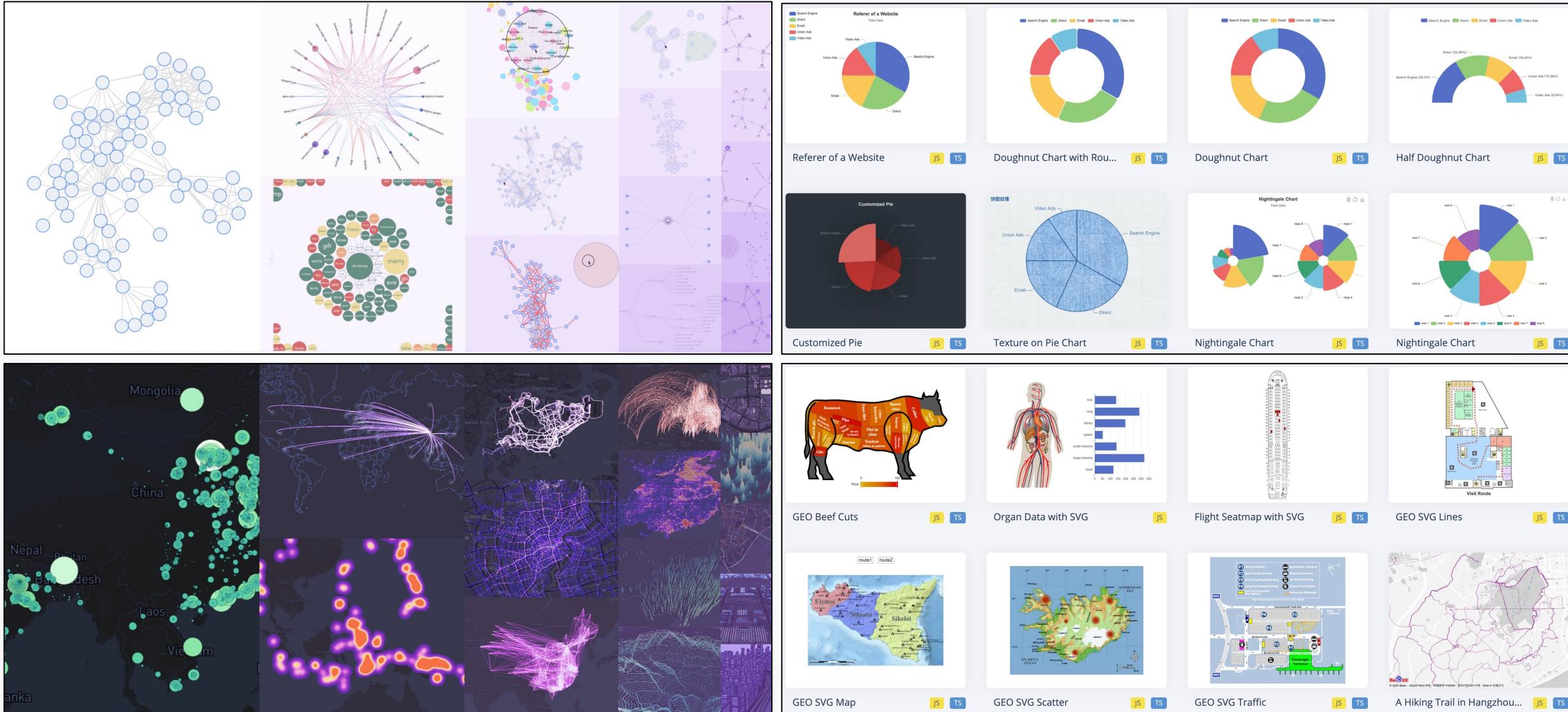
Tableau

<https://www.tableau.com>

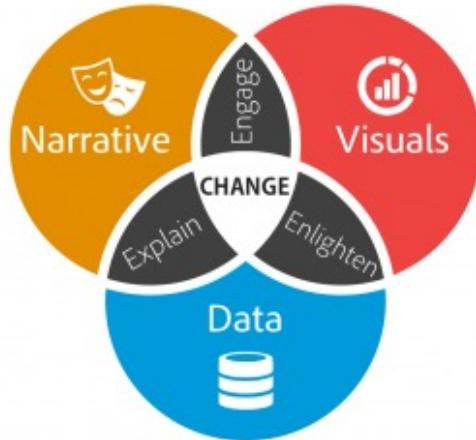
<https://public.tableau.com/s/>



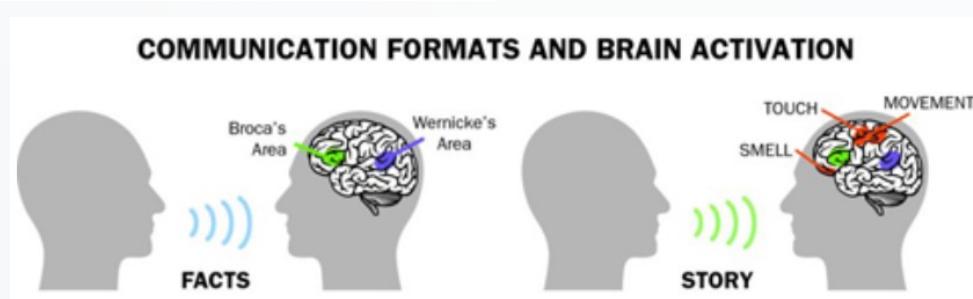
各种图表库



「叙事」与「表达」为可视化注入了灵魂



叙事是讲述数据中所包含信息的一种手段。它通过将关键信息按照逻辑串联，并融合在一个连贯的上下文中进行表达，以可视化形式加以呈现，从而达到沟通的目的。



仅仅呈现数据内容只会激活大脑中与信息处理有关的区域。而叙事性内容则可以点亮大脑的多个感知区域，帮助理解并促进决策。

脑神经学家发现，叙述性（narrative）的表述与呈现会比独立的事实（facts）点亮更多大脑的区域（González et al., 2006; Lacey et al., 2012）。



Krish Sathian,
宾夕法尼亚州立
大学神经学与心理
学教授，神经
科学研究所主任



CÉSAR
ÁVILA, 西班
牙海梅一世
大学实验心
理学教授

认知心理学家发现，叙事会促使大脑释放皮质醇和催产素，这两种激素可以帮助提升注意力、共情力和行动力（Zak, 2012）。并指出，叙事有助于人们理解信息（Stephens et al., 2010），帮助抽象和复杂信息的传达（Sperber et al., 1995）。



Paul J. Zak, 美国
克莱蒙研究大学
神经济研究中心主任，心理
学、经济科学和
管理学教授



Uri Hasson,
普林斯顿大学
心理学与神经
科学教授，普
林斯顿神经科
学研究所成员



Dan Sperber,
法国著名认知
心理科学家，
布达佩斯中欧
大学认知科学
和哲学系教授



难点与挑战

核心挑战：如何 **设计** “行之有效”的 叙事性可视化 方案？

有内涵

让可视化言之有物：如何自动洞察大规模、多变化的复杂数据，
以获得重要且有意义的数据内涵？

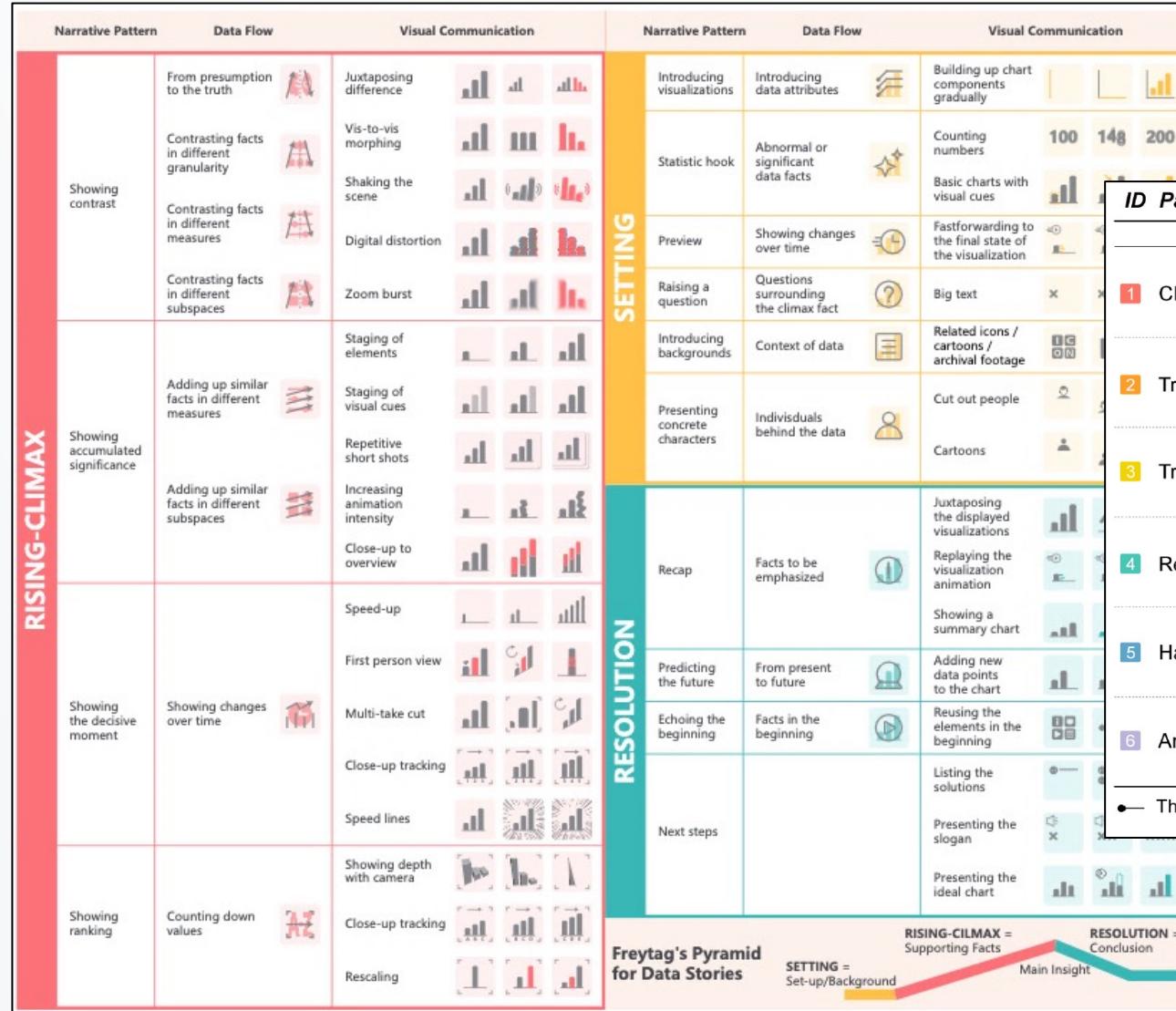
好理解

让可视化易于理解：如何自动设计并生成直观易懂的可视化图
表以呈现重要的数据信息？

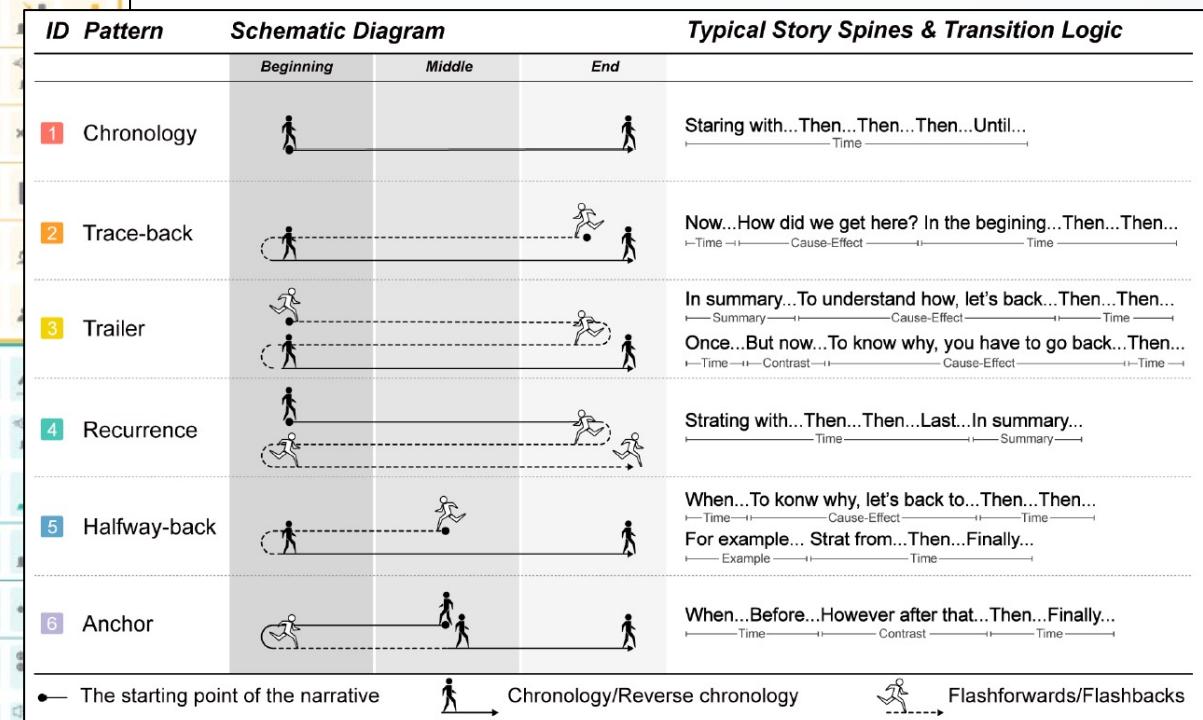
易传达

让可视化便于描述：如何自动生成结构鲜明且逻辑清晰的可视
化叙事方案，以便信息的描述与传达？

寻找可视化叙事范式、构建设计空间体系



Narrative Structures



可视化叙事逻辑结构及设计空间

课程大纲

- 什么是可视化?
- 为什么要可视化?
- 可视化的类型与发展趋势
- 怎样对数据进行可视化?
- **如何对可视化进行评估?**
- 怎样学习可视化?
- 可视化应用案例

如何对可视化进行评估

设计准则比对 (Design Principles)

非正式用户调研 (Informal Study)

受控用户调研 (Controlled User Study)

实际应用及案例分析 (Case Study)

理论分析 (Theoretical Analysis)

基准化分析法 (Benchmark)

如何对可视化进行评估

- 设计准则比对 (Design Principles)
- 非正式用户调研 (Informal Study)
- 受控用户调研 (Controlled User Study)
- 实际应用及案例分析 (Case Study)
- 理论分析 (Theoretical Analysis)
- 基准化分析法 (Benchmark)

检验可视化的设计
是否符合一般性的设计准则

如何对可视化进行评估

设计准则比对 (Design Principles)

非正式用户调研 (Informal Study)

受控用户调研 (Controlled User Study)

实际应用及案例分析 (Case Study)

理论分析 (Theoretical Analysis)

基准化分析法 (Benchmark)

展示可视化结果给用户
并收集他们的反馈

如何对可视化进行评估

设计准则比对 (Design Principles)

非正式用户调研 (Informal Study)

受控用户调研 (Controlled User Study)

实际应用及案例分析 (Case Study)

理论分析 (Theoretical Analysis)

基准化分析法 (Benchmark)

设计比对试验，计算用户在完成相同任务时耗费的时间及任务的正确率是否有显著差异

如何对可视化进行评估

设计准则比对 (Design Principles)

非正式用户调研 (Informal Study)

受控用户调研 (Controlled User Study)

实际应用及案例分析 (Case Study)

理论分析 (Theoretical Analysis)

基准化分析法 (Benchmark)

将可视化应用于实际案例，通过展示所得到的数据模式，及有趣发现证明可视化设计的有用性

如何对可视化进行评估

设计准则比对 (Design Principles)

非正式用户调研 (Informal Study)

受控用户调研 (Controlled User Study)

实际应用及案例分析 (Case Study)

理论分析 (Theoretical Analysis)

基准化分析法 (Benchmark)

分析可视化布局算法复杂度、及
可视化系统的综合性能等

如何对可视化进行评估

设计准则比对 (Design Principles)

非正式用户调研 (Informal Study)

受控用户调研 (Controlled User Study)

实际应用及案例分析 (Case Study)

理论分析 (Theoretical Analysis)

基准化分析法 (Benchmark)

与现有最佳解决
方案进行直观对比

课程大纲

- 什么是可视化?
- 为什么要可视化?
- 可视化的类型与发展趋势
- 怎样对数据进行可视化?
- 如何对可视化进行评估?
- **怎样学习可视化?**
- 可视化应用案例

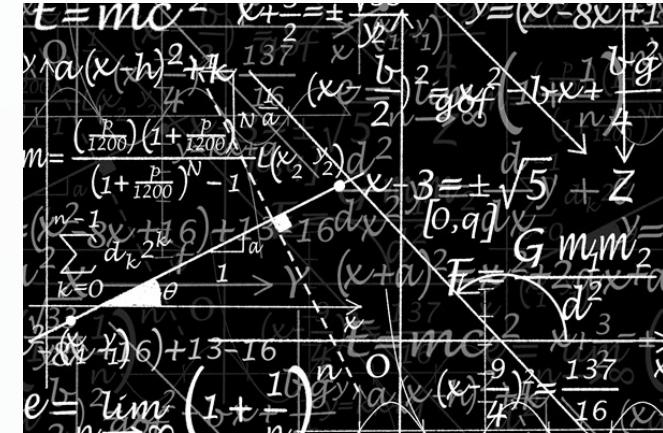
需要学习什么？



设计 (Web, UI & UX, Visual Design)



编程

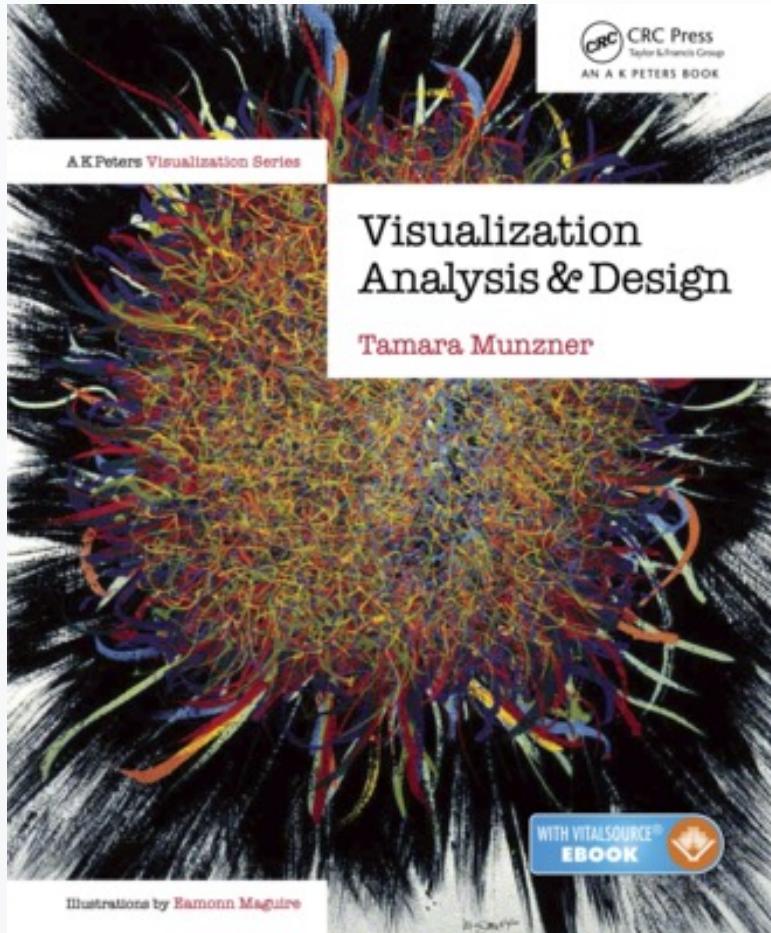


数学与算法

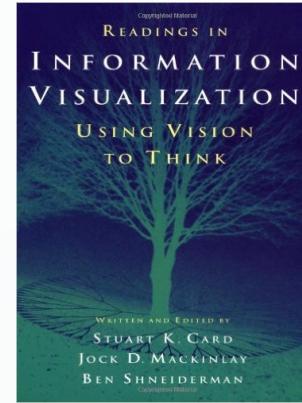
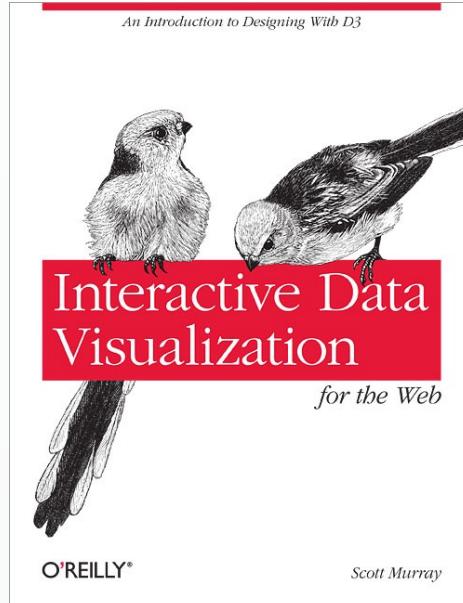


技术 (机器学习, 数据挖掘)

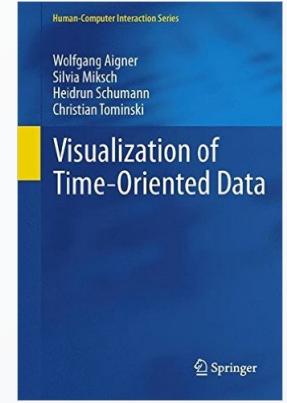
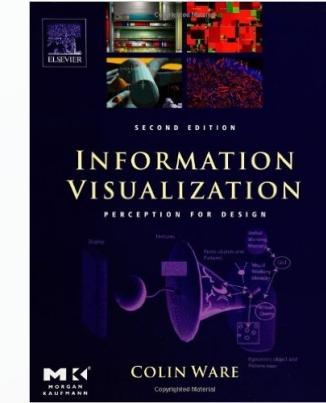
有哪些参考书？



可视化编程



一些其他的参考书



<http://www.cs.ubc.ca/~tmm/vadbook/>

如何获得前沿知识？

顶级学术期刊

年度学术会议



IEEE InfoVis / VAST / SciVis Conferences

IEEE TRANSACTIONS ON
VISUALIZATION AND COMPUTER GRAPHICS

A CATALYST FOR ADVANCED VISUALIZATION

From the September 2013 issue
Scheduling in Heterogeneous Computing Environments for Proximity Queries

By Duksu Kim, Jinkyu Lee, Junghwan Lee, Insik Shin, John Kim, Sung-Eui Yoon

We present a novel, linear programming (LP)-based scheduling algorithm that exploits heterogeneous multicore architectures such as CPUs and GPUs to accelerate a wide variety of proximity queries. To represent complicated performance relationships between heterogeneous architectures and different computations of proximity queries, we propose a simple, yet accurate model that measures the expected running time of these computations. Based on this model, we formulate an optimization problem that minimizes the largest time spent on computing resources, and propose a novel, iterative LP-based scheduling algorithm. Since our method is general, we are able to apply our method into various proximity queries used in five different applications that have different characteristics. Our method achieves an order of magnitude performance improvement by using four different GPUs and two hexa-core CPUs over using a hexa-core CPU only.

IEEE Transaction on Visualization and Computer Graphics

课程大纲

- 什么是可视化?
- 为什么要可视化?
- 可视化的类型与发展趋势
- 怎样对数据进行可视化?
- 如何对可视化进行评估?
- 怎样学习可视化?
- **可视化应用案例**

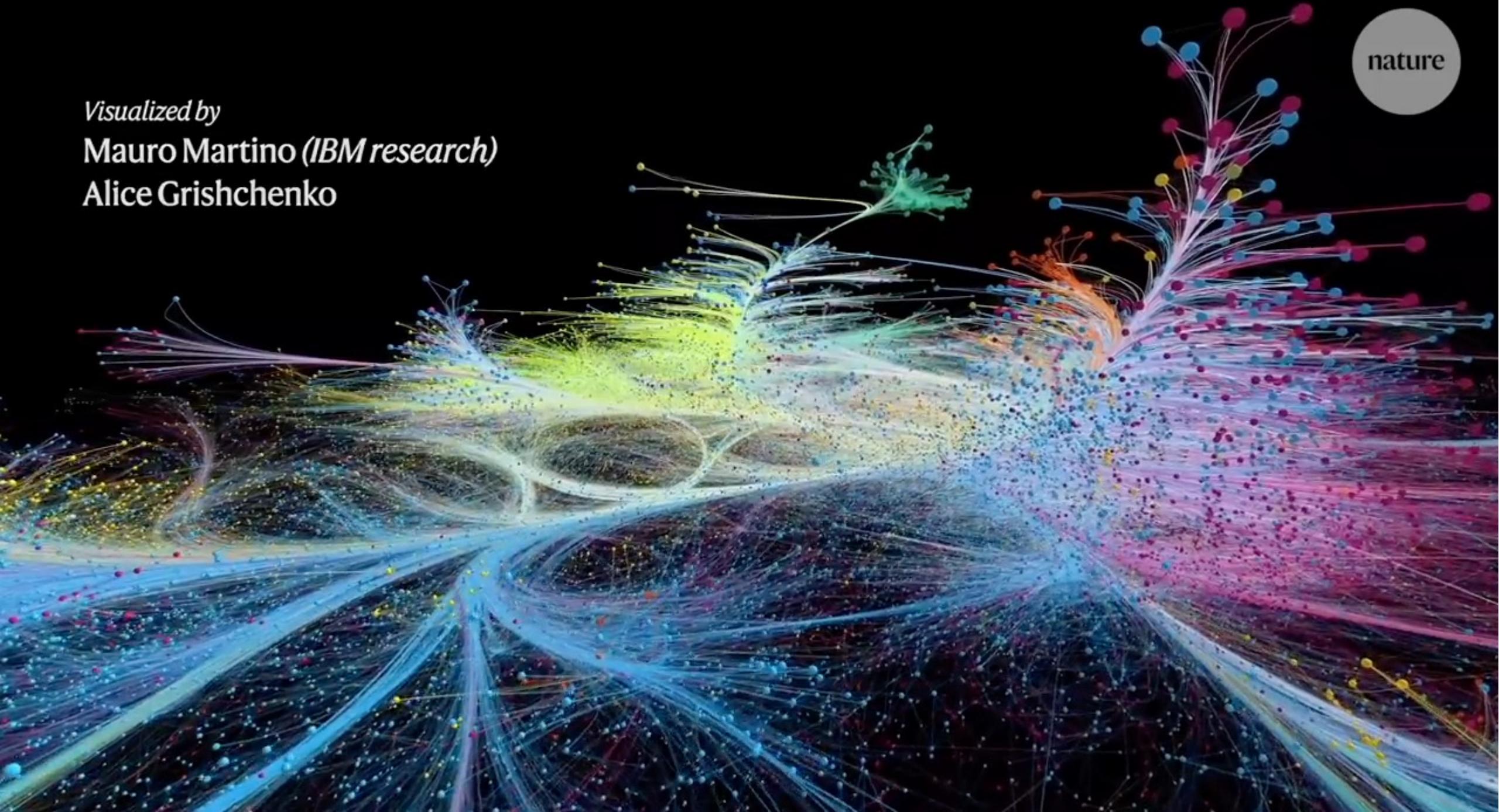


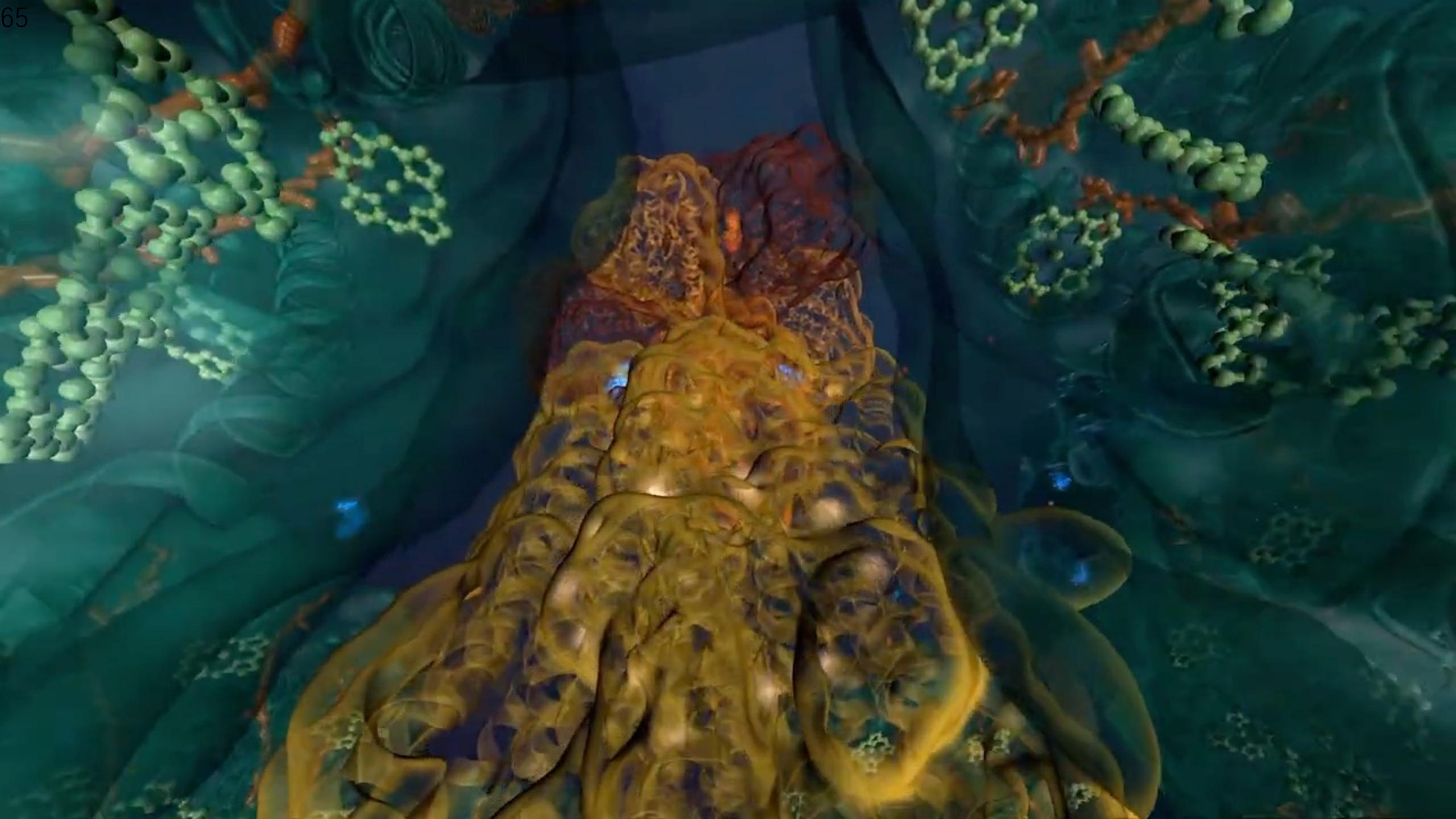
idv

同济智能大数据可视化实验室
TONGJI INTELLIGENT BIG DATA VISUALIZATION LAB

案例1：科学与艺术

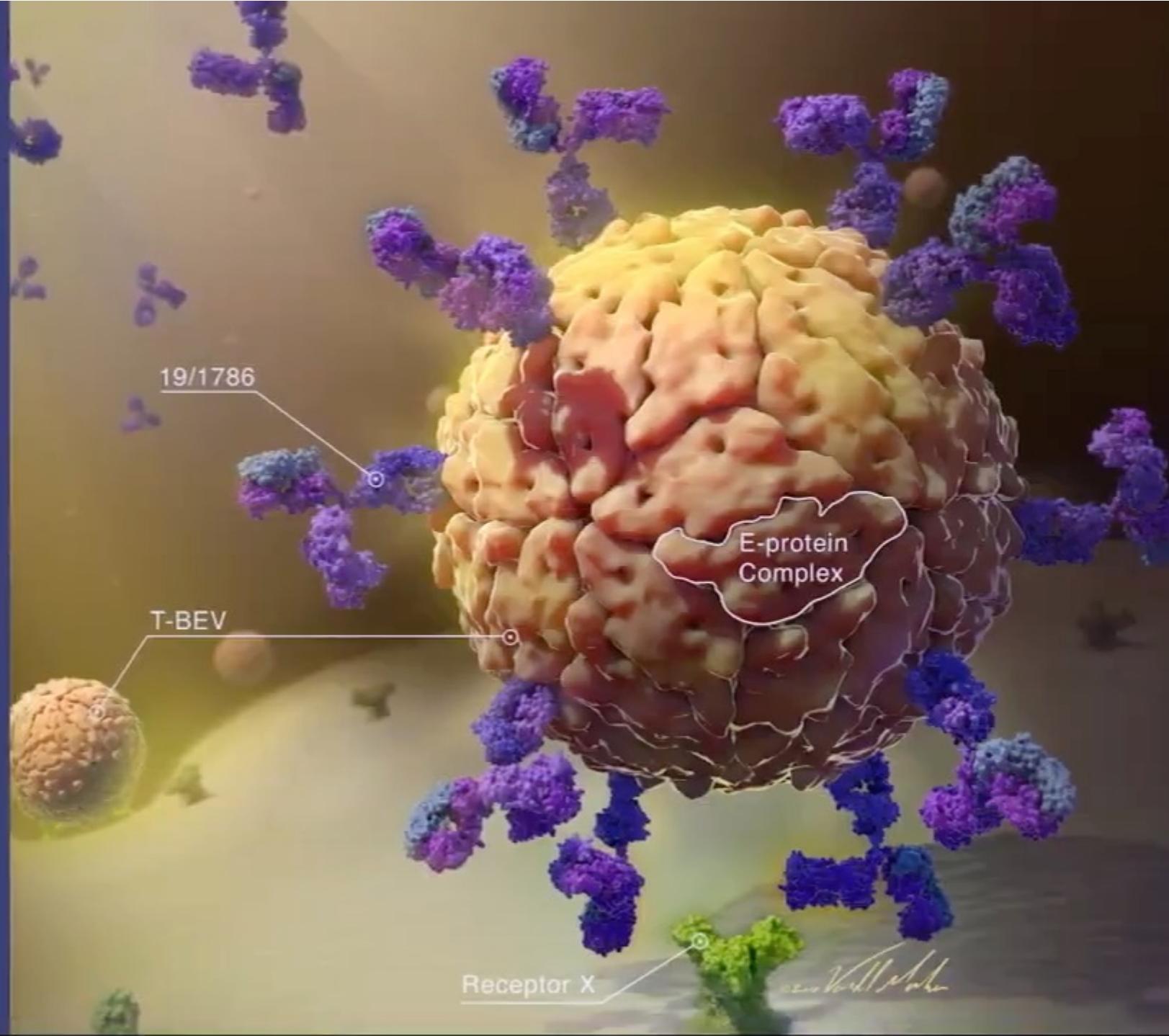
Visualized by
Mauro Martino (*IBM research*)
Alice Grishchenko







MODIFIED TEETH OF THE *(Anser anser domesticus)*



Anser anser domesticus, an edentulous waterbird, lost its ability to produce teeth about **116 million years ago**. Tooth-loss was caused by inactivation of the enamel and dentin-producing genes. These genes are still present in its genome but remain inactive. Instead of teeth, *Anser anser domesticus* has bony processes called **tomia**, which do not contain any enamel or dentin. Along with tooth-loss, the bill (rostrum) was also modified to more of a horn-like structure, lacking any tooth-like structures at its anterior region. These modifications evolved together, occurring anterior to posterior which is noticeable in today's domestic goose.





idv

同济智能大数据可视化实验室
TONGJI INTELLIGENT BIG DATA VISUALIZATION LAB

案例2：历史与文化

胡天汉月 古道千家

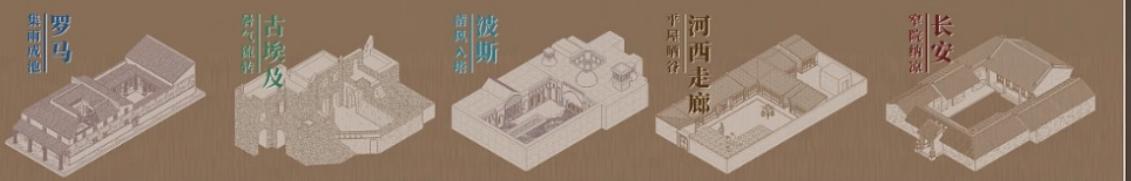
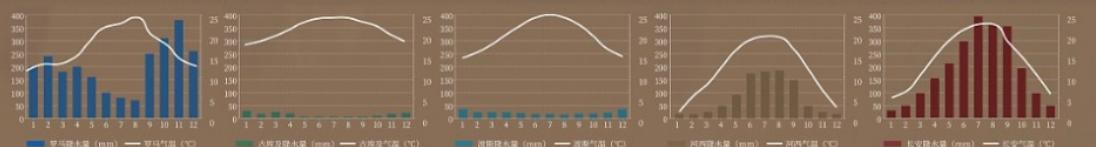
丝绸之路传统民居的生态智慧

Ecological Adaptation of Silk Road Vernacular Dwellings

丝路历史

西汉时，张骞出使西域，以长安为起点，经甘肃、新疆，到中亚、西亚，并联结地中海各国的陆上通道，在通过这条漫漫长途进行贸易交接的货物中，中国的丝绸最具代表性，“丝绸之路”因此得名。丝绸之路不仅是古代亚欧互通有无的商贸大道，还是促进亚欧各国和中国友好往来、沟通东西方文化的友谊之路。

丝绸之路上下跨越历史2000多年，涉及陆路与海路，所以按历史划分为先秦、汉唐、宋元、明清4个时期。按线路分陆上丝绸之路与海上丝绸之路。陆上丝绸之路走向不一，又分为“北方丝绸之路”与“南方丝绸之路”。陆上丝路所经地区的地理环境差异很大，人们又把它细分为“草原森林丝路”、“高山峡谷丝路”和“沙漠绿洲丝路”。



幅。辽阔的亚欧大陆，南北纬度跨度广，南北热差差异明显，体现了从热带、亚热带、暖温带、中温带到寒温带的气候过渡，具有丰富的地理差异。同时，亚欧大陆东西跨度范围广，从东临大海到西北内陆，再到地中海沿岸、大西洋东岸，各个地区的降水差异十分显著，体现了从湿润、半湿润、半干旱到干旱地区的降水的递变。这种复杂的、巨大的气候差异深刻地影响分布其中的传统民居建筑形态。民居建筑是产生是最近生活，最需要高能效建筑造就，是蕴含着中留着的具有历史文化、科学、艺术价值的珍贵实体遗存，同时也包含反映城市文化、民俗和城市风貌特征的宝贵实体遗存。

本信息图将以丝绸之路为引线，串联起沿线城市的民居建筑风貌，用可视化的方式探索民居建筑与气候环境的和谐共生之道，展现古代劳动人民因地制宜而生的建筑智慧，启发当代建筑可持续发展道路的探索与研究。



气候与建筑

本信息图旨在探究丝绸之路路上五个典型气候分区内典型地点的民居建筑特征与当地气候之间的关系，下面展示了我们的研究思路。

首先通过对各气候区较为明显的气候特征进行数据可视化，观察分析各地气候的差异性。基于数据来源特征的原因，我们制定了温度、湿度、日照辐射、降水五类气候特征。

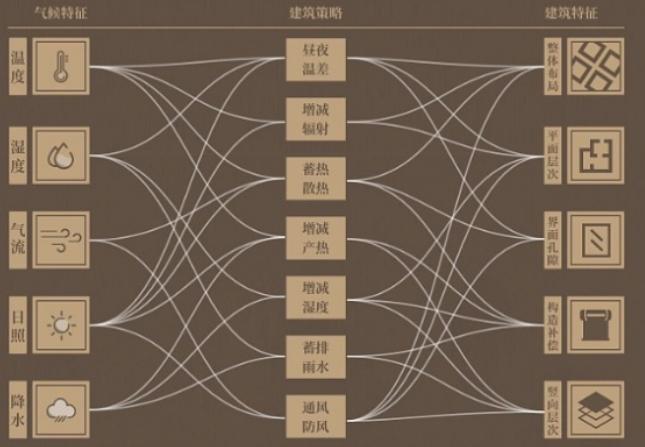
其次，我们研究了建筑气候适应性的各式方法，包括增加或减少热量、储蓄或排放自然降水、控制昼夜温差等建筑策略。

最后，在我们所选择的五处民居中，分别提取整体布局、平面层次、界面孔隙、空间层次和构造补位五个维度进行比较，分析其各自所对应的建筑策略。



在气候数据方面，我们选取的数据按照从大陆-国家-省-市区的层次进行细分，对获取的EPW文件进行数据统计和可视化表达。这些气象信息分别包含：温度与湿度、日照和云层、风速、空气湿度图、自然通风能力等。对于某些地点，以上气候特征并不鲜明，对建筑成形的影响较小，于是针对各地特征又补充了特有的气候数据进行分析，例如：部分埃及的篇章通过增加云量图来辅助说明当地日夜温差巨大的原因。

为了后面能够对各地的气象数据进行横向的对比，以及在Tableau等可视化平台上进行可视化。我们将EPW文件转换为CSV格式进行处理。最终生成图表类型包括日辐射强度曲线图、风玫瑰图、温度热力图、年降水量曲线图等。



长安·窄院纳凉

E 108.93° N 34.27°
合院建筑+大陆性季风气候



胡天汉月
古道千家

丝绸之路传统民居的生态智慧

Ecological Adaptation of Silk Road Vernacular Dwellings

窄院的建筑+减少太阳辐射

长安地区的气候干燥，夏季炎热，冬季寒冷，因此在建筑设计中特别重视遮阳和隔热。居民在房屋设计时会考虑朝向，尽量减少直射阳光，同时增加外窗面积，利用自然通风降温，从而解决夏季高温的问题。

丝路遗迹+丝路起点的民族融合国际大都市

长安作为丝绸之路的起点，是当时世界上最大的国际大都市之一。它不仅是一个商业中心，也是一个文化交流的中心，吸引了来自世界各地的商人、学者和艺术家。

降雨特征

屋面特征

在三个季节中，雨水充沛，雨水充沛，于是长安常见的坡面和檐沟排水系统。

通道特征

长安的风向频率图显示，风从三个方向十分普遍。为了减少风对建筑的影响，居民在走廊中设置屏障以阻挡风，以及使用屏障以减少风对建筑的影响。

日照特征

长安的日照强度图显示，夏季日照时间长，冬季日照时间短。居民在房屋设计时会考虑朝向，以充分利用日照。

开窗特征

受地形影响，长安全年东北风频发，因此在房屋设计时会考虑朝向，以减少风对建筑的影响。

风象特征

受地形影响，长安全年东北风频发，因此在房屋设计时会考虑朝向，以减少风对建筑的影响。

墙体特征

长安墙体材料以砖石为主，墙体厚度较大，起到良好的保温效果。



入选 2022 中国数据内容年度案例
(信息设计类第一名)
中国数据内容大赛

河西走廊·平屋晒谷

E 102.88 N 37.47°

平顶建筑 | 温带大陆性气候 (干旱)

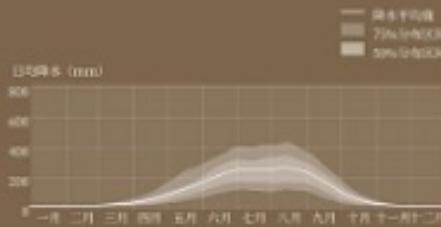
建筑名称	薛家村
建筑地址	武威市古浪县土门镇薛家村
始建年代	1970
建筑特点	三合院上廊式民居



丝绸之路传统民居的生态智慧
Ecological Interpretation of Silk Road Vernacular Dwellings

平屋晒谷 | 利用光照资源

河西走廊地区降水量少且蒸发量大，多年平均降水量不到 200 毫米，是甘肃省降水最少的地区。古籍记载：“因西村少雨，不瓦不甃。用松柏木密搭晒谷巷以屋檐覆以瓦砾。”平屋顶为村民在夏天的晒粮、饲料提供了有利条件。



河西走廊木材总量稀少，木质差、胸径小。当地村民选择用“密肋式”结构建造房屋。用较细的木梁和木椽子搭在墙上，不设柱子。而肋式屋架省材，轻便，耐用，较好地发挥了屋架承重效果。

当地一年内降水量变化大，降水量主要集中在夏季，占年降水量的 47%-64%；冬季降水量少，降水量约为 2-8mm，占年降水量的 2%-8%。

丝路期间 | 丝绸路上不用愁没地方打尖住店



丝绸之路是河西走廊区域经济发展的主轴线，也是对外联系的通道。在开通丝绸之路时期，沿途驿站建设也随之展开。汉代每六十里设置一所驿亭，唐代时随着丝绸之路更是形成了“五里一亭，十里一驿”的格局。河西走廊地区的丝绸之路古道上设立了重重关卡的驿站机构。丝绸之路先后经过河西走廊的绝大部分城镇，使得这些城镇自身在执行军务和驿站功能的同时也得到发展。



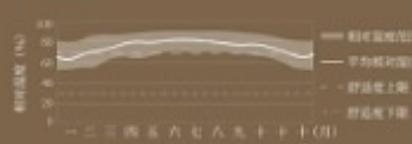
温度特征
河西走廊地区为大陆性干旱气候，虽然年平均温度较低，但夏季最高温可高达 42 摄氏度，冬季严寒，最低温可低达 -20 度。在温差巨大的气候条件下，墙体需要承担保温和隔热双重的功能。

墙体特征
早期生土民居土坯墙下多用砖块夯实作为墙体基础，但易吸水水浸透松软，严重威胁建筑的安全。后来逐渐发展成用石为主的基脚称做方式，在具体地基差时，一般先对地基槽进行夯实。两壁石块或者砖土之间筑成地槽 30-40 厘米以上以防止雨水的侵袭。

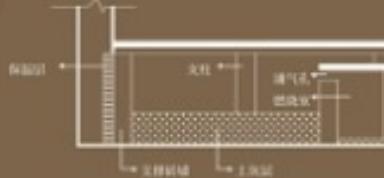
降温特征



温度特征



防寒特征



日照特征



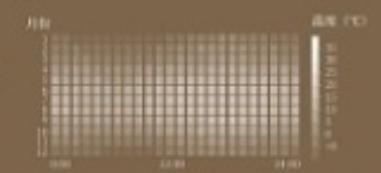
风象特征

河西走廊风向多变。嘉峪关以西的玉门、敦煌等地以东风为主；武威一代以西北风为主。其中民勤沙暴日一年有 50 天以上，安西路大风一年高达 80 天，有“风库”之称。

开窗特征

河西走廊民居为防风沙和日晒，造型上采用内向型平面，建筑四周不开窗，聚居未分离防御居住一体的古代“碉楼”。

温度特征



波斯·清风入塔

E 54° N 31°

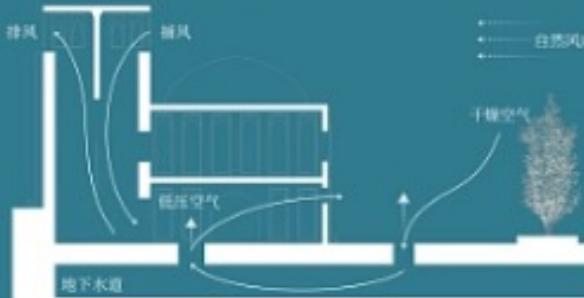
风塔建筑 | 温带大陆性气候 (半干旱)

建筑名称	Karimi House
建筑地点	伊朗伊斯法罕
始建年代	德加王朝时期
建筑特点	风塔式自然通风结构



丝绸之路传统民居的生态智慧
Ecological Interpretation of Silk Road Residential Buildings

清风入塔 | 增加室内舒适度



同时太阳辐射热引起，地表的空气受热膨胀后往上升，低温的冷空气相对流入。上升的空气经过冷却便相对降低，这种空气的流动就是风。伊朗（古波斯）建筑在顶部设置风塔，由于高密度、风塔内部形成风压，具有拔风的效果。地表干燥的热空气被吸进地下木渠，慢慢变成潮湿空气，然后上升至房间，降低室内温度湿度，热交换完成后，通过风塔顶部排出。

丝路趣闻 | 象征吉祥多福的石榴从何而来



石榴是如何从西方传入我国的？历史资料中的记载并不多，但是我们仍然能够通过“安石榴”这一石榴古代的名称推断出植物传播在我国的大致时间和路线。被后人引用最多的记载是西晋张华的《博物志》：“汉张骞出使西域，得涂林安石国种以归。”石榴在中国古代被称为“安石榴”。而波斯一带被称为安息国。古汉书“安息”与“安石”发音相似，“安”则是波斯语“小粒”的音译。所以，“安石榴”的字面意思应该是“安息国小粒的繁花”。

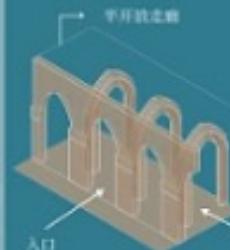


穹顶特征



由于伊朗地区气候干燥炎热，以穹顶作为室内提高区域，下方常设置通风孔以降温增湿，促进室内空气流动。

日照特征



通道特征

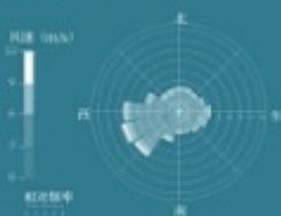
波斯白天太阳辐射强，地面加热迅速，太阳直射的温度可高达65-70℃。因此民居构造需要避开过多太阳直射面积。

中庭特征

通过夜间通风冷却的“麦卡大墙”和中庭利用自然风的流动带走室内热空气。



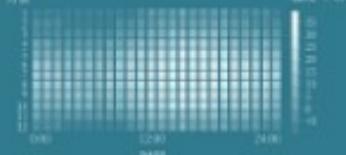
风向特征



温度特征

波斯夏季高温冬季寒冷，全年温差较大。同时由于当地地理环境多砾石，日夜温差十分巨大，最高可达50摄氏度以上。

月份



墙体特征

民居墙体多用粘土压制干燥制成空心的粘土砖为基层，外部贴石灰、沙子和水制作的灰浆进行涂抹覆盖，并以灰泥糊墙作为建筑外表面的装饰。空心砌筑的材料制成厚厚的墙体，保证炎热季节为室内降温，寒冷季节保暖防寒。

古埃及·暑气流转

E 15° N 24°

挑高建筑 | 热带沙漠气候

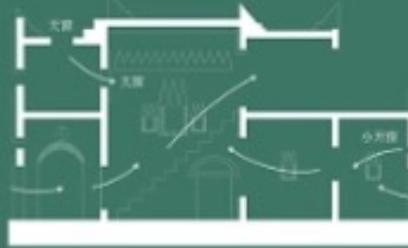
建筑名称	加达梅斯民居
建筑地点	叙利亚
始建年代	公元前2世纪前
建筑特点	墙体涂料与开窗



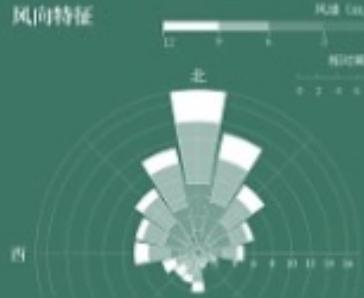
丝绸之路传统民居的生态智慧
Ecological Interpretation of Silk Road Traditional Dwellings

暑气流转 | 利用风力资源

风向特征



埃及建筑师哈希·法蒂赫说，古埃及传统民居建筑依赖自然通风的思想，通过小开窗、大窗等方式将自然界的风引入屋内，达到抵触建筑的建造目的。



埃及的风资源非常充足，其北向风速高，出现频率高。因此即使位于沙漠地区，也有充足的自然风力帮助进行建筑内外的热量交换。

丝路趣闻 | 埃及艳后穿丝绸也得精打细算



埃及人是否埃及艳后克利奥帕特拉的长裙解体有所争议。但那时埃及艳后克利奥帕特拉的女王身上的那件丝质长袍津津乐道。当时为了节省珍贵的丝线，克利奥帕特拉把从中国移植埃及的桑蚕，交由宫廷养蚕，先是将从中国买来的桑树服装一并一齐地解开，再重新织成或脱成点细线的布品。



屋顶特征



建筑立面图（局部）

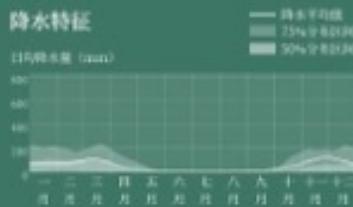
此处的传统民居为露天的平台，平台四面有围墙围绕，有道路相连，是女性的专用活动空间，她们可以在上面自由交谈、走动和聊天。

采光特征



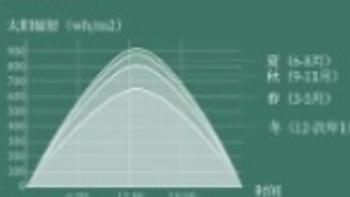
利用屋顶格栅小孔引入日光，并让其折射到室内的每个角落，室内将被抹上白色的石灰粉或熟石膏，使得整个房间更加明亮。

降水特征



埃及属于热带沙漠气候，全年降水少，在六月至九月之间每月平均降水量甚至低于100mm。这样干旱的气候允许埃及民居拥有平坦开阔的屋顶。

日照特征



埃及的日照强度在夏秋两季十分强烈，为了避免靠近赤道的太阳直射，民居建筑开窗在阳面较小。

室内特征



埃及云量雨量全年较低，昼夜温差较大，因此开窗小而细。

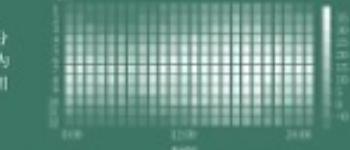
云层特征



温度特征

埃及的夏天十分干燥炎热，正午时分的温度甚至能高达35摄氏度之上。为了避暑降温，民居的墙体也往往采用特殊的材料。

温度特征



墙体特征

传统建筑完全用泥土建造，部分墙体涂有白色石灰粉，用以反射阳光。这样的结构可以让居民在夏天保持凉爽，在寒冷的冬季保持温暖。

罗马·集雨成池

E 12° N 41°

双庭建筑 | 地中海气候

建筑名称	Domas Romana
建筑地点	罗马
始建于年代	公元前500年
建筑特点	内庭

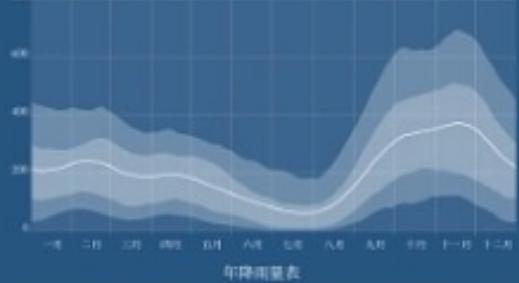


丝绸之路传统民居的生态智慧
Ecological Intuition of Silk Road Vernacular Dwellings

集雨成池 | 积蓄水资源

降雨量表 (mm)

— 年平均降雨 — 月均分佈 — 季节分佈



罗马地处地中海沿岸，是典型的地中海气候，冬季潮湿多雨，夏季高温少雨。虽然全年雨水量充足，但分布十分不均匀。古罗马早期城市里的水管工程并不成熟，还托排水主要依赖收集雨水。前厅的屋檐沟槽处，将雨水引入下方0.3米深的蓄水池中。

丝路趣闻 | 为赶时是一掷千金买丝绸



由于运输路途的漫长与艰险，加之其间商人多次转手而遭受的暴利盘剥，罗马人进口中国丝织品必须付出大量黄金。在汉朝，一匹高级丝绸料子的10尺要价12两黄金，盈盈达到百倍。丝绸大量入境造成罗马帝国每年黄金外流达10万盎司，罗马帝国不得不颁布法令，禁止穿丝绸，但这依旧没有阻止丝绸在罗马的风靡。

All roads lead to Rome All roads lead



降水特征



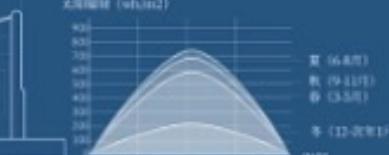
湿度特征

相对湿度指空气中水汽压和水汽饱和百分比。该比例越高环境湿度越大。罗马一年中四分之三的时间相对湿度比高达70%以上。通过人体舒适指数上积，由此诞生了房屋内部防潮结构的雨水。

通风特征



日照特征



罗马的日照在春秋秋三季较为温和，在冬季由于云层变厚雨水偏多，日照强度偏低。民房内保留露天阳台以保证冬季的采光量，同时设置地暖在夏天用于降温。

开窗特征

罗马的风向分布较为平均，不存在特定方向较为突出的风向。而且全年的平均风力在中小风力之间，于是罗马庭院的开窗设计上较少考虑结构对风力的防御。

风向特征



墙体特征

罗马是典型的地中海气候，全年温和，温差小。古罗马混凝土用天然火山灰砌成，火山灰砌成，小化拌匀之后再凝固。在多层建筑中，底层用粗颗粒石作骨料，二层用灰华石，上层则用火山喷发物产生的琨斑块多孔石。因此建筑越往下越结实。越往上越轻盈。此类墙体对保持室内恒温起很大作用。

温度特征





idv

同济智能大数据可视化实验室
TONGJI INTELLIGENT BIG DATA VISUALIZATION LAB

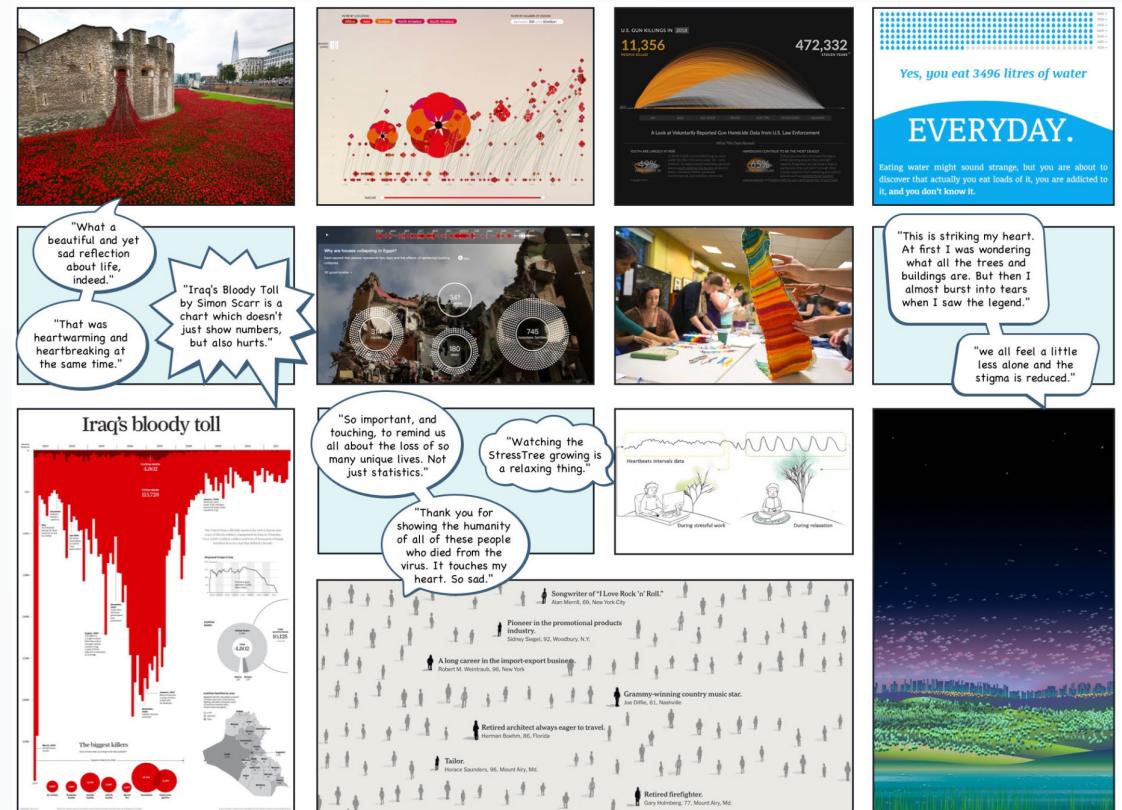
案例3：情感与表达

情感的可视化表达

情感可视化可以被定义为，那些与情感相关、由情感而生、或影响观众情感的可视化表达形式

对象

情感作为被可视化的对象



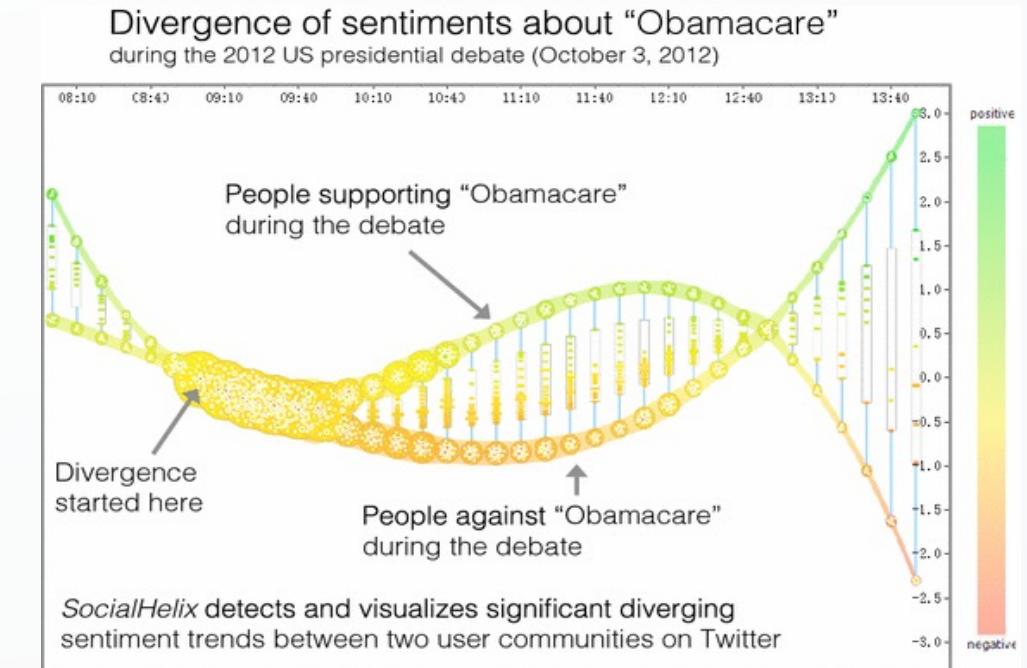
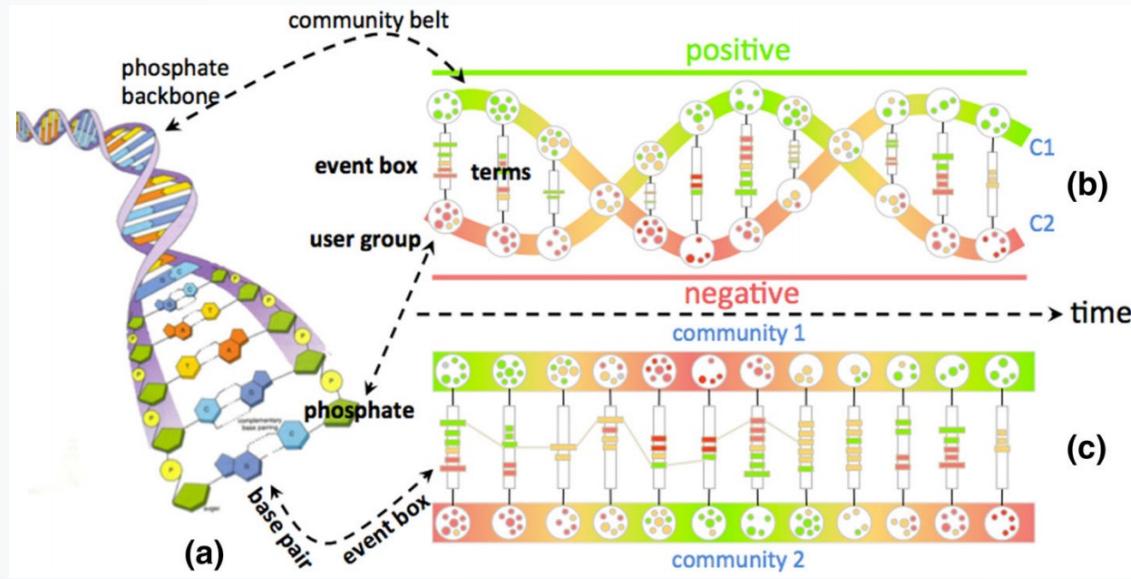
条件

情感作为阅读可视化的条件

结果

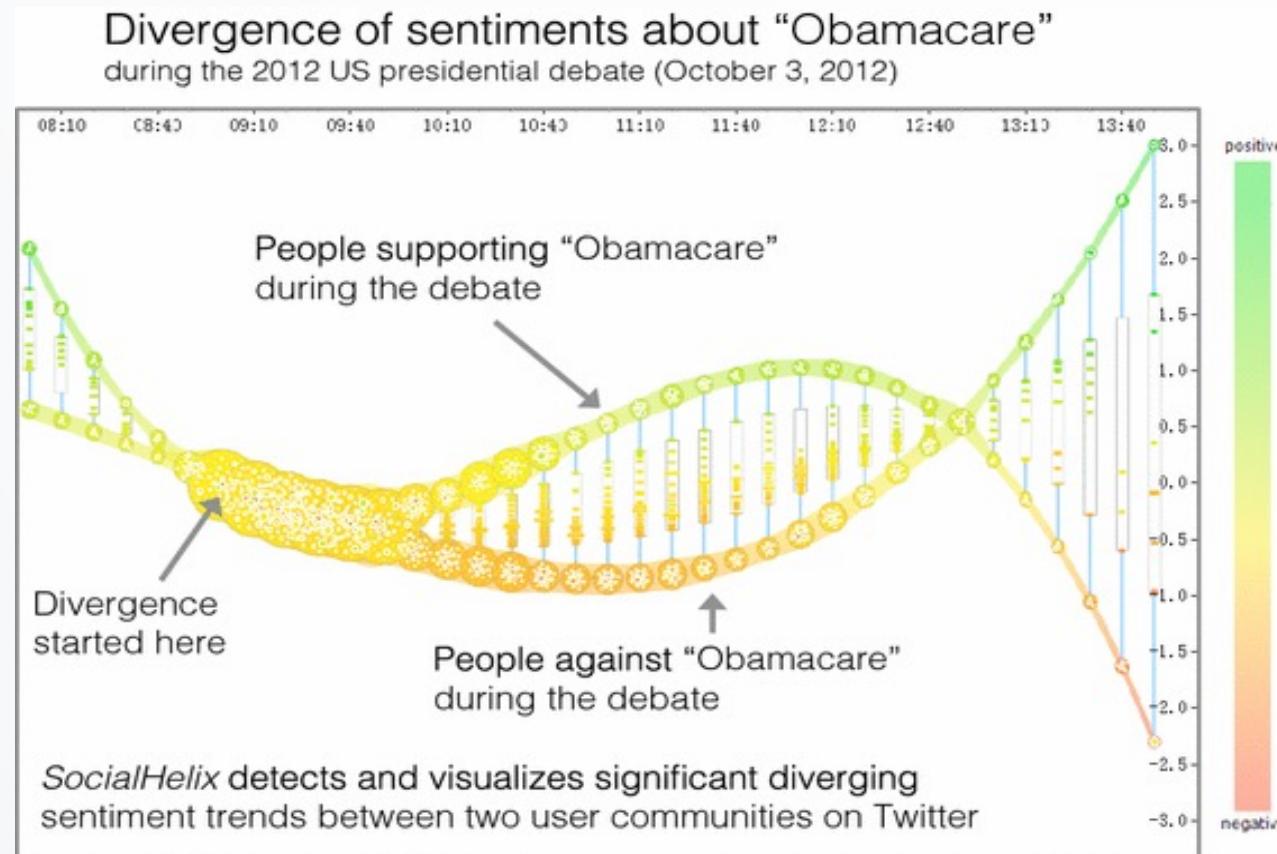
情感作为可视化设计的结果

分析对立情感的波动变化模式



“SocialHelix: Visual Analysis of **Sentiment Divergence** in Social Media”,
 Journal of Visualization, May 2015, Volume 18, Issue 2, pp 221-235

分析对立情感的波动变化模式



例如，上图中的可视化是Twitter中人们围绕“奥巴马医疗保险”改革而产生的意见分歧。图中的横轴代表时间，每一条带子代表持有相同观点的一组Twitter用户，带子的颜色与位置代表了针对该事件的情感（红色/底部：负面情感，绿色/上部：正面情感），两根带子中间的连接线代表了人们争论的具体内容中的关键词。

- 针对特定事件的异常舆情状况检测与分析
- 基于可视化的动态情感数据分析及实时直观展现
- 基于智能交互的异常情感波动模式探索与比对

Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.
 Dressée par M. Minard, Inspecteur Général des Ponts et Chaussees en retraite. Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie; le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Segur, de Fezensac, de Chambray et le journal médical de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk et Mobilow et qui rejoignirent Witebsk, avaient toujours marché avec l'armée.

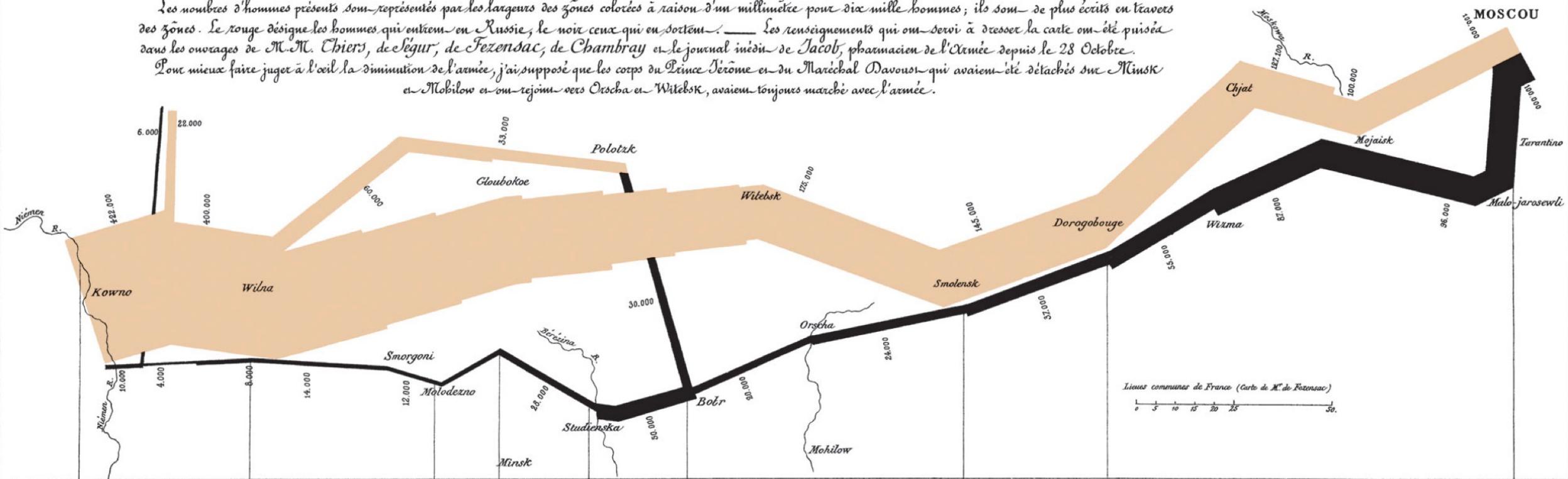


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les Cosaques passent au galop
le Niemen gelé.

— 26° le 7 X.^{bre}
— 30° le 6 X.^{bre}
— 24° le 1^{er} X.^{bre}
— 20° le 28 9.^{bre}
— 11°

— 21° le 14 9.^{bre}

— 5° le 9 9.^{bre}

Pluie 24 8.^{bre}

Zéro le 18 8.^{bre}
5
10
15
20
25
30 degrés

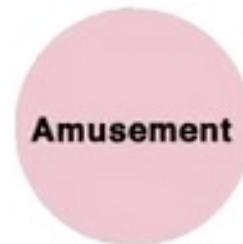
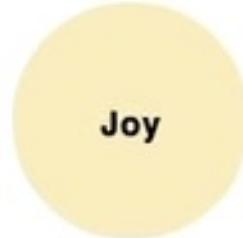
Autog. par Regnier, 8. Pas. S^ete Marie St^e Gait^e à Paris.

Imp. Lith. Regnier et Dourdet.

"In Minard's Napoleon chart, the shocking loss of life that he visualized with thick black bars receding from Moscow tell a foreboding story about foreign military adventures as well as serve as a grim memorial to his contemporaries who served on this calamitous campaign." - Kostelnick

带有情感的信息叙事表达

Kineticharts



Kineticharts are composed of five types of affect and three types of chart (**bar charts, line charts, and pie charts**).

Xingyu Lan, Yang Shi, Yanqiu Wu, Xiaohan Jiao, Nan Cao

Kineticharts: Augmenting Affective Expressiveness of Charts in Data Stories with Animation Design
IEEE Transactions on Visualization and Computer Graphics (IEEE Vis2021)

带有情感的信息叙事表达



What if
the world only has 100 people?

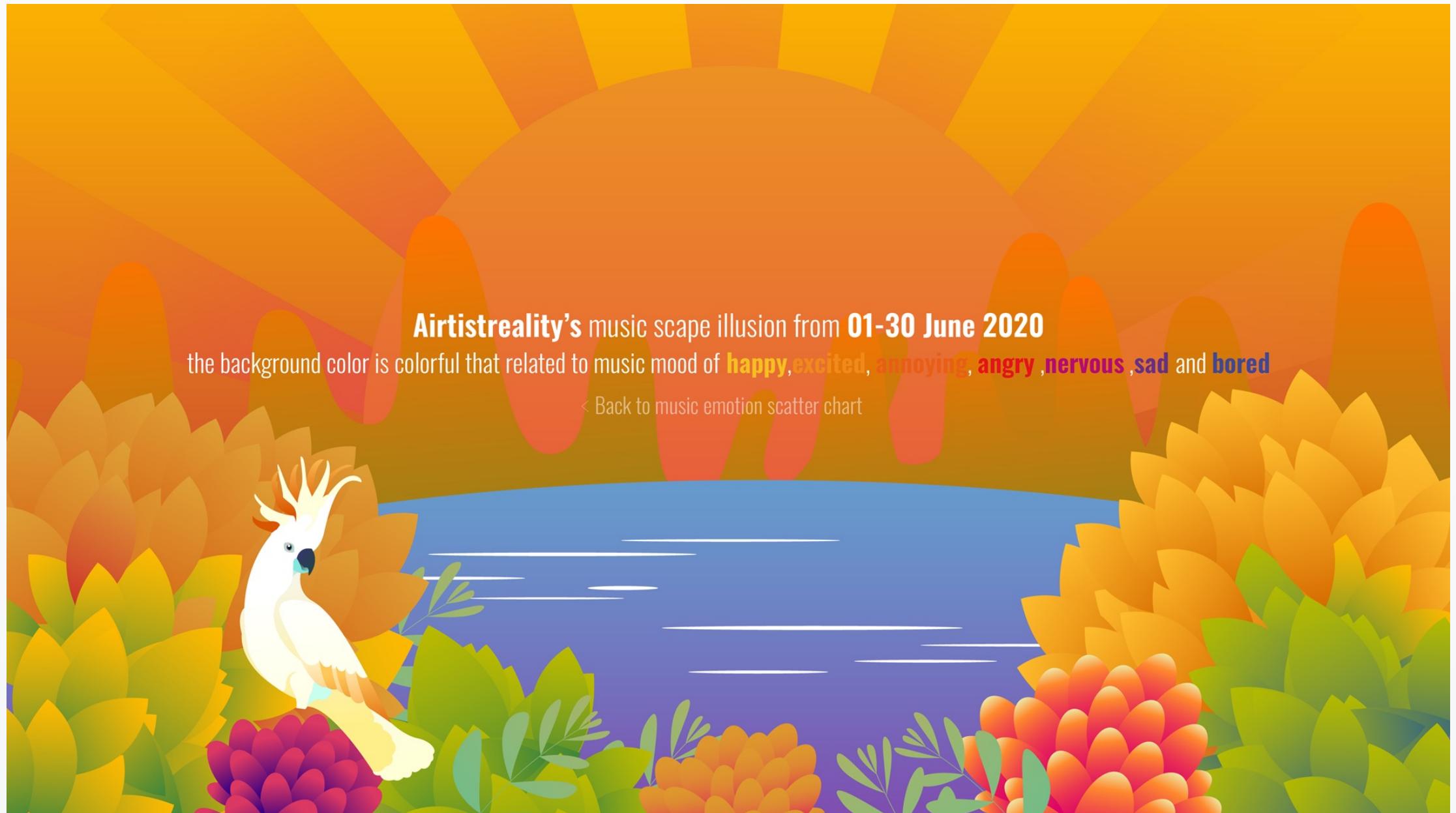
Xingyu Lan, Yang Shi, Yanqiu Wu, Xiaohan Jiao, Nan Cao

Kineticcharts: Augmenting Affective Expressiveness of Charts in Data Stories with Animation Design
IEEE Transactions on Visualization and Computer Graphics (IEEE Vis2021)



MUSIC SCAPE









idv

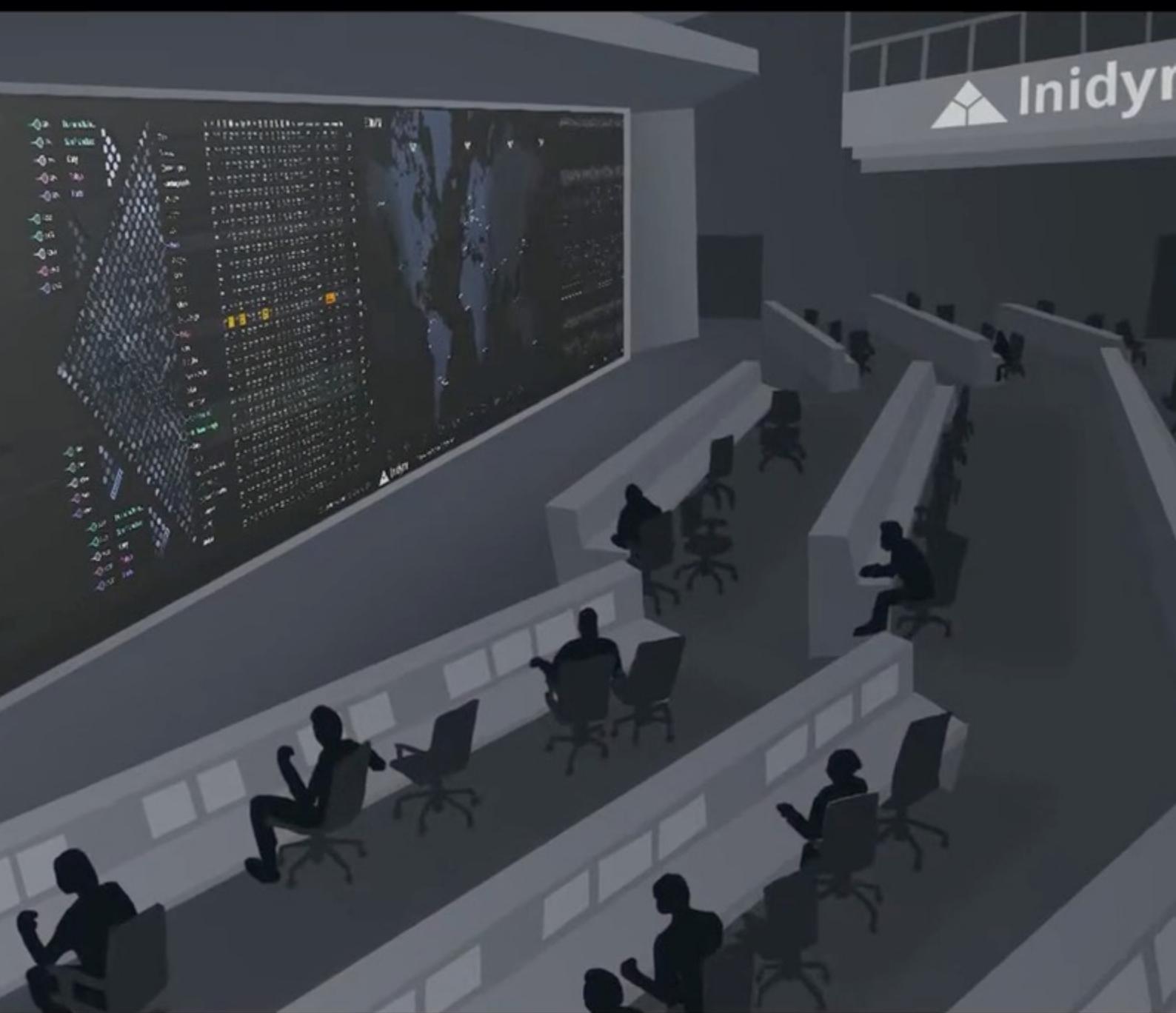
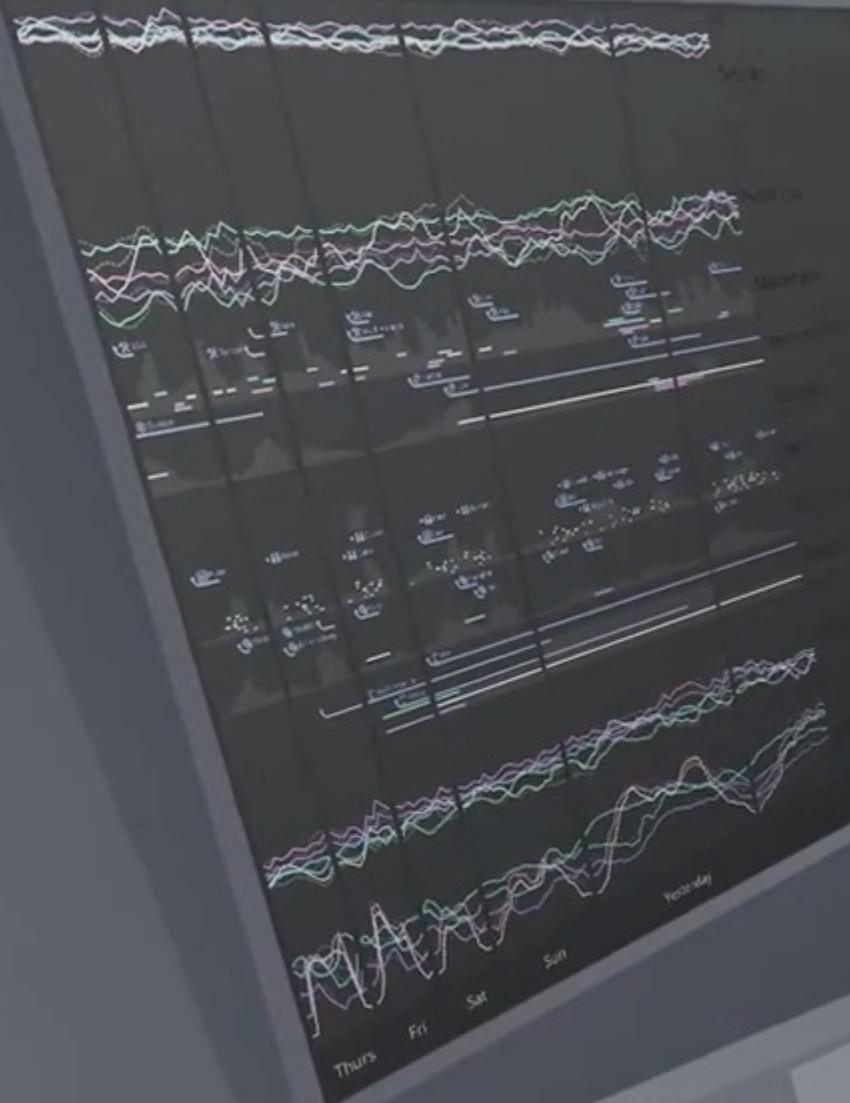
同济智能大数据可视化实验室
TONGJI INTELLIGENT BIG DATA VISUALIZATION LAB

案例4：探索与发现





Inidyn



The Growth of Social Media



Do you trust your friends on social media ?

“On the Internet, Nobody Knows You’re a Dog”: A Twitter Case Study of Anonymity in Social Networks

Sai Teja Peddinti*
psaiteja@nyu.edu

Keith W. Ross*†
keithwross@nyu.edu

Justin Cappos*
jcappos@nyu.edu

*Dept. of Computer Science and Engineering, NYU
Brooklyn, New York, USA

†NYU Shanghai
Shanghai, China

(ACM Conference on Online Social Networks, 2014)

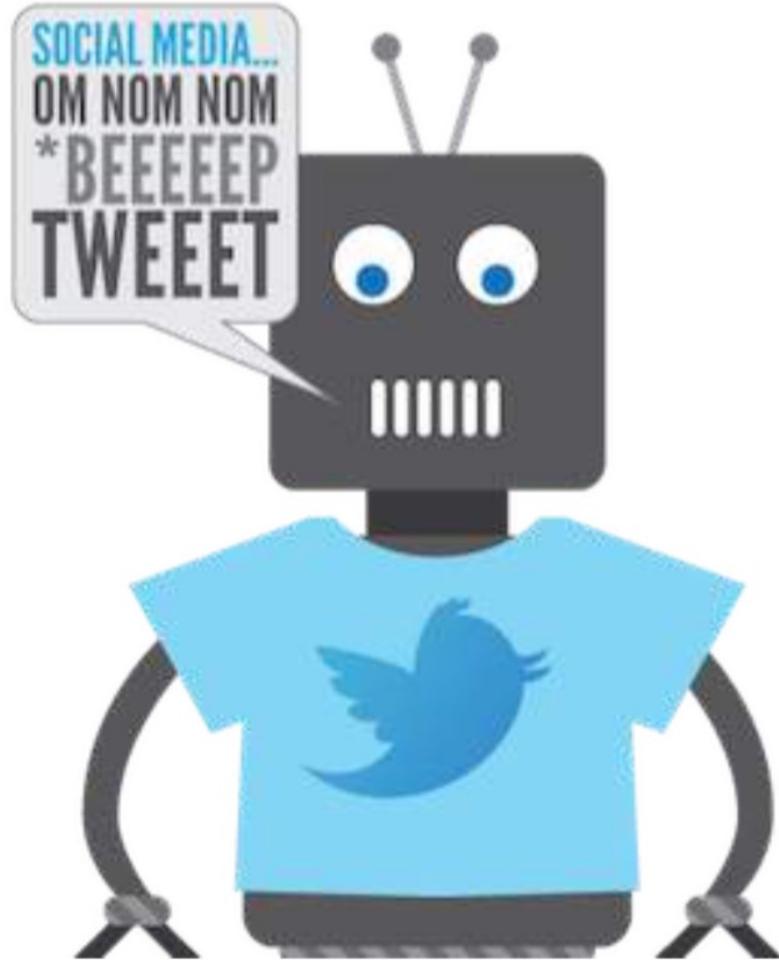


"On the Internet, nobody knows you're a dog."

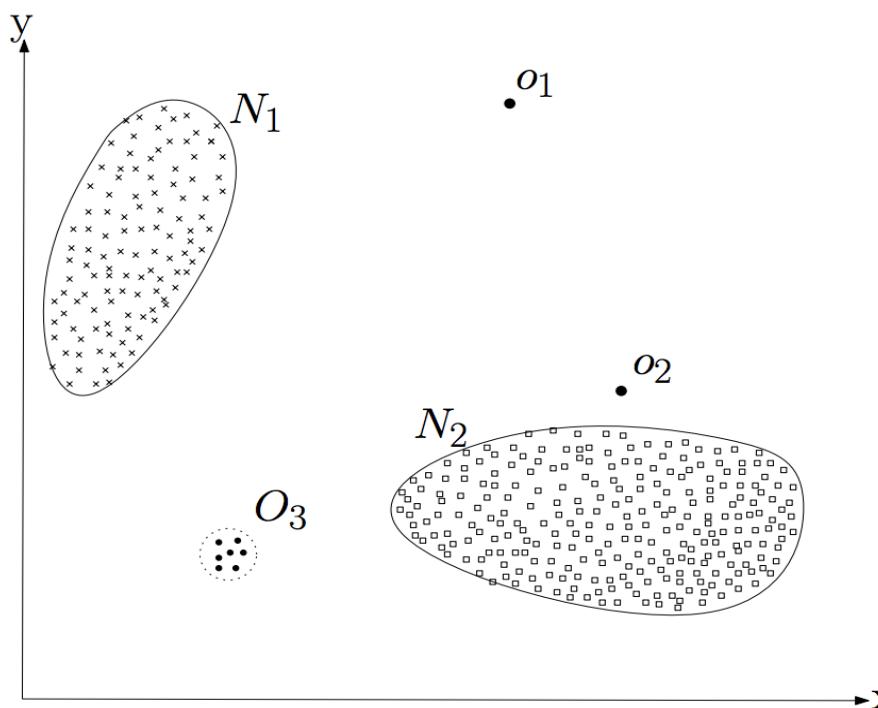
An adage since 1993

*Anonymous users are potential
threats to the society*

Ultimate goal: Catch users with anomalous behaviors



Anomaly Detection



Anomaly detection (or outlier detection) is the identification of items, events or observations which do not conform to an expected pattern or other items in a dataset

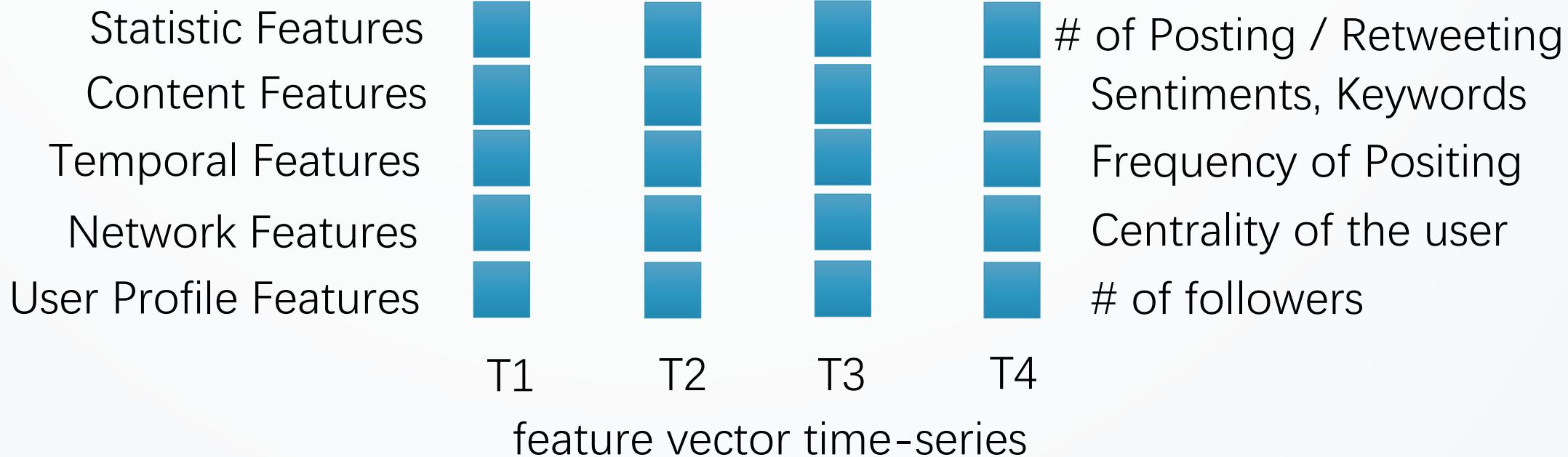
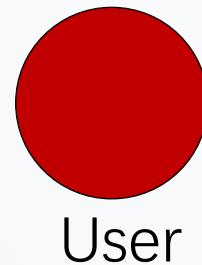
Key Challenges

- It is difficult to define what is normal or abnormal
- Unavailable of ground truth or labelled data making results validation difficult

User Behaviors

- Posting (Tweeting)
 - Create a message and post it to others
- Responding (Replying / Retweeting)
 - Spread the messages posted by others

Capturing User Behaviors via Features



Anomaly Detection

TLOF: Temporal Local Outlier Factor

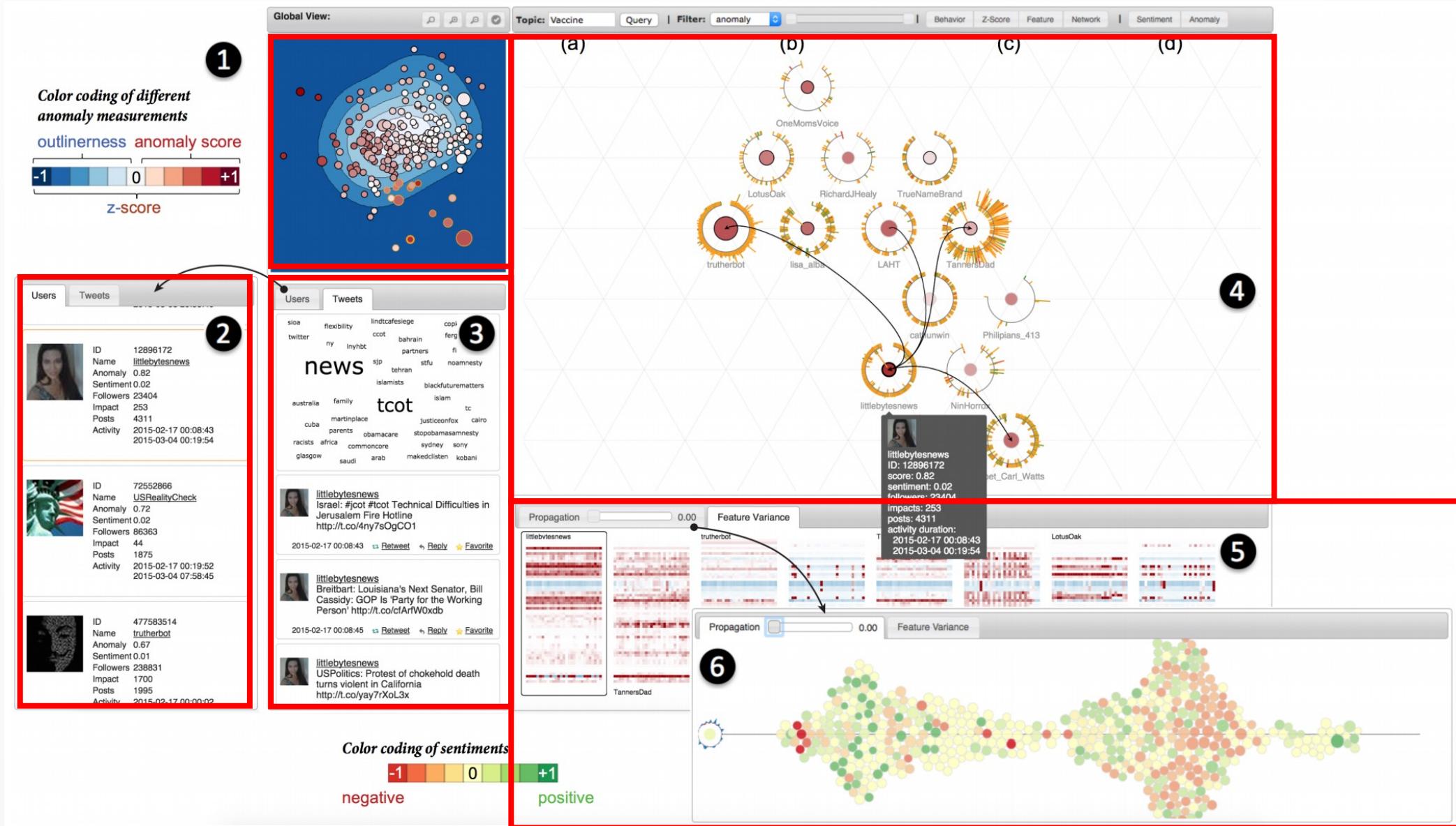
$$s(X) = \alpha \cdot Z_1(X) + (1 - \alpha) \cdot Z_2(X)$$

$$Z_1(X) = LOF(x_T) - \sum_{t=T-W}^{t=T-1} LOF(x_t)/W$$

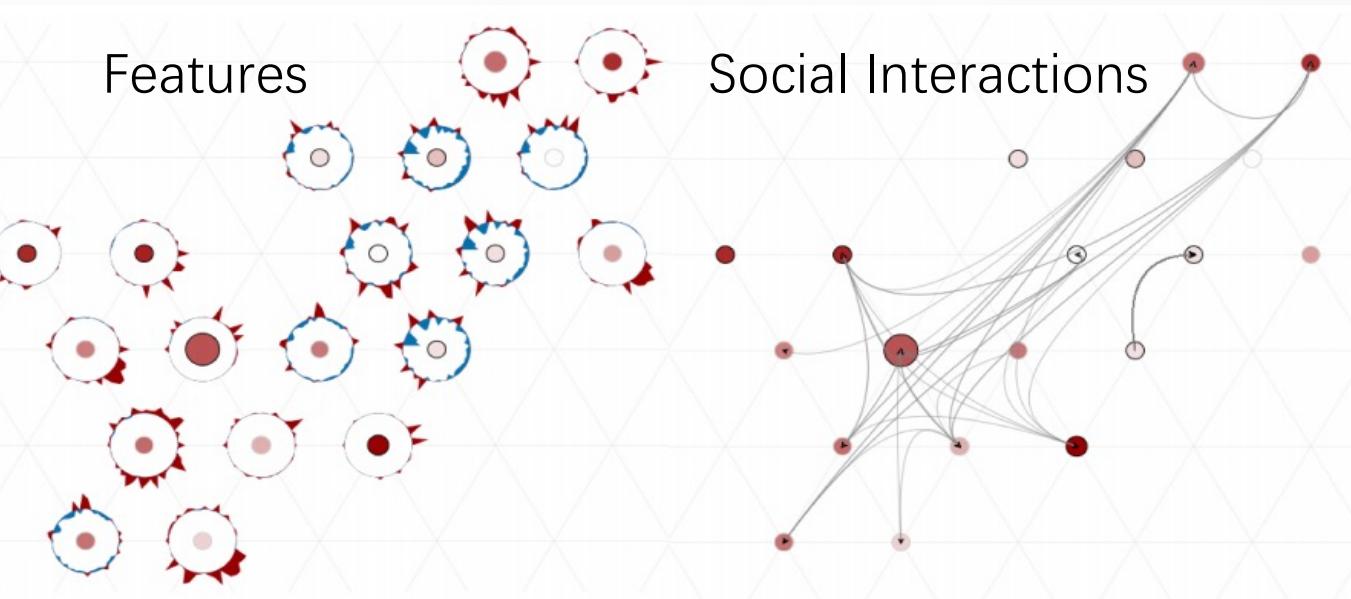
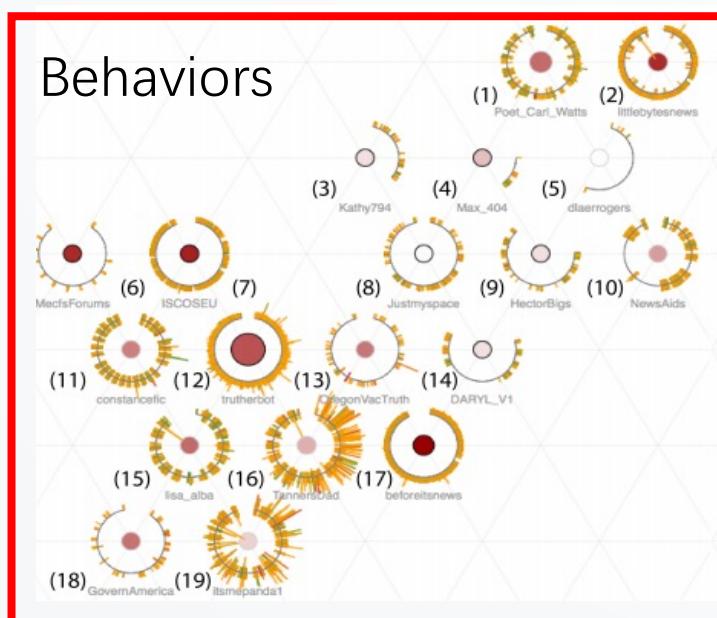
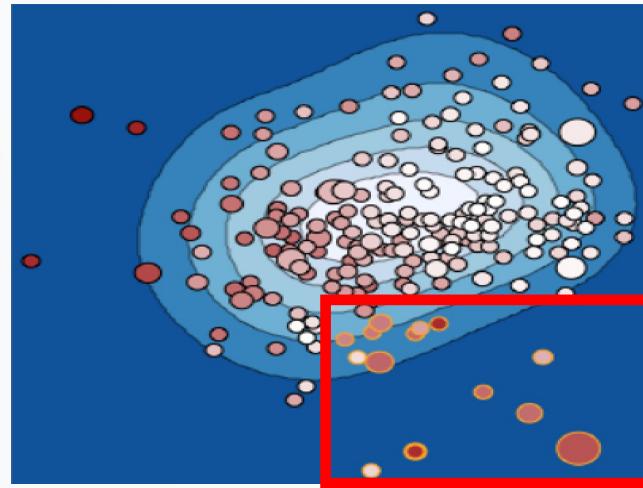
$$Z_2(X) = 1 - P_N(LOF(x_T), \mu, \sigma)$$

The TLOF gives an anomaly measurement for every users by identifying the features that are significantly different from other users in the test data and the past history of his own

Breunig, Markus M., et al. "LOF: identifying density-based local outliers." ACM sigMOD record. Vol. 29. No. 2. ACM, 2000.

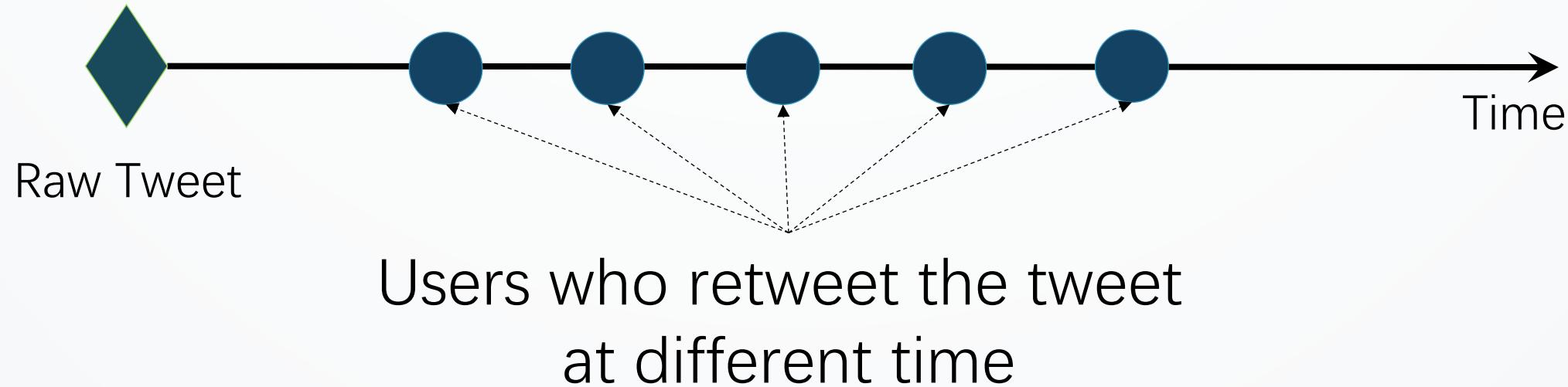


The Inspection View

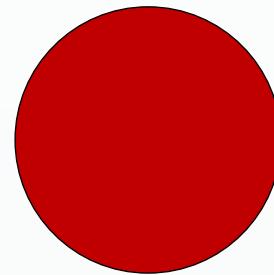


Showing Posting and Retweeting Activities

Activity Thread

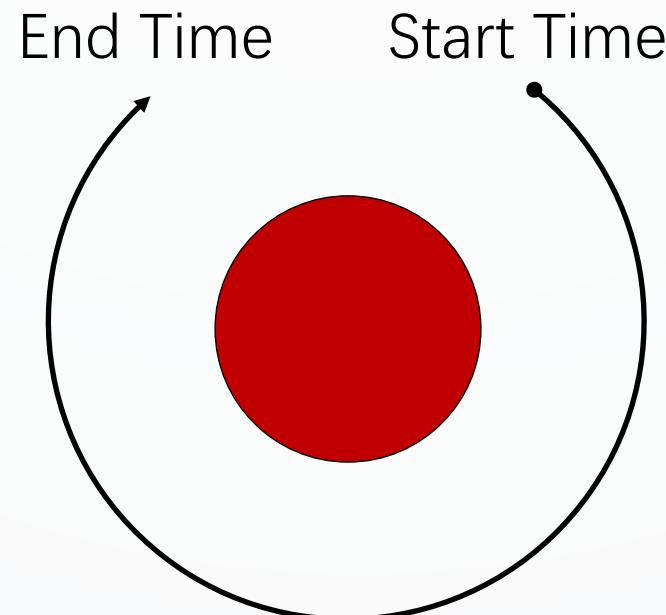


Showing Posting and Retweeting Activities



A user is visualized as circle sized by their importance and colored by their anomaly score

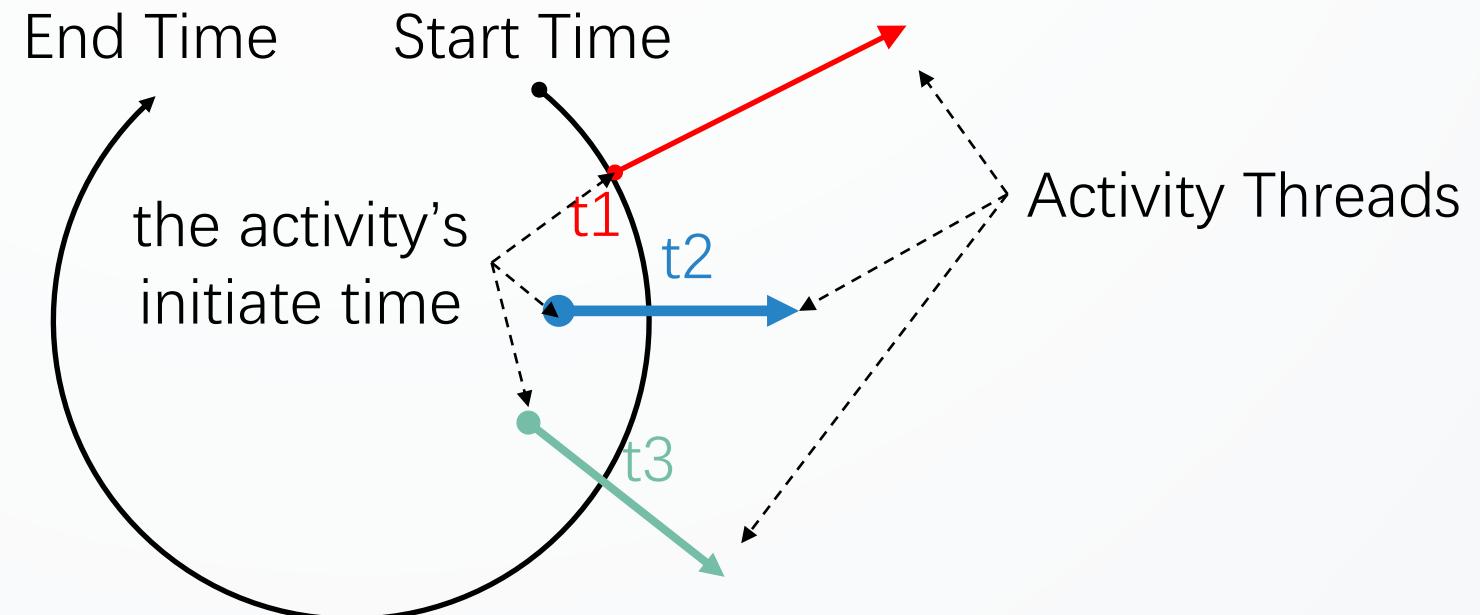
Showing Posting and Retweeting Activities



The centric user' s activity history is recorded in
a circular timeline

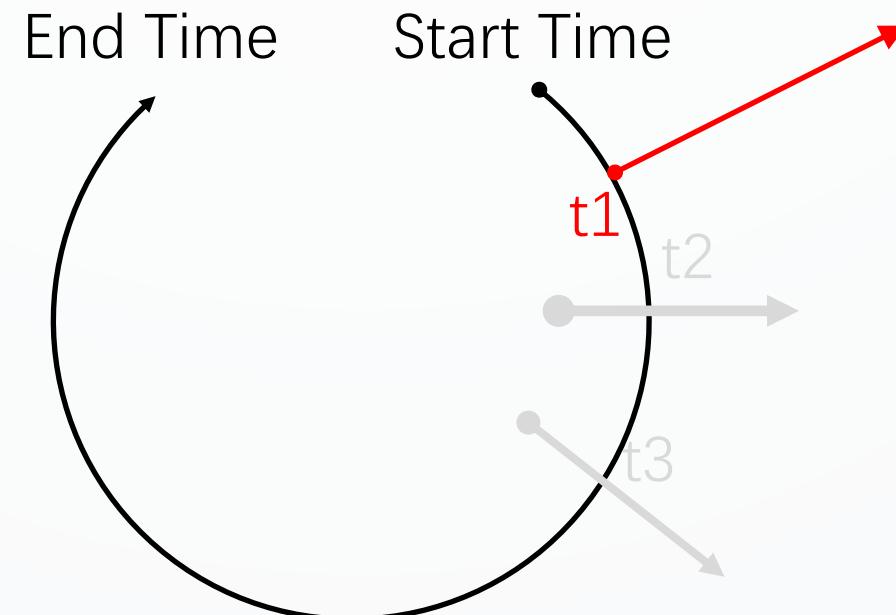
Showing Posting and Retweeting Activities

Color: sentiments
Thickness: # of involving users

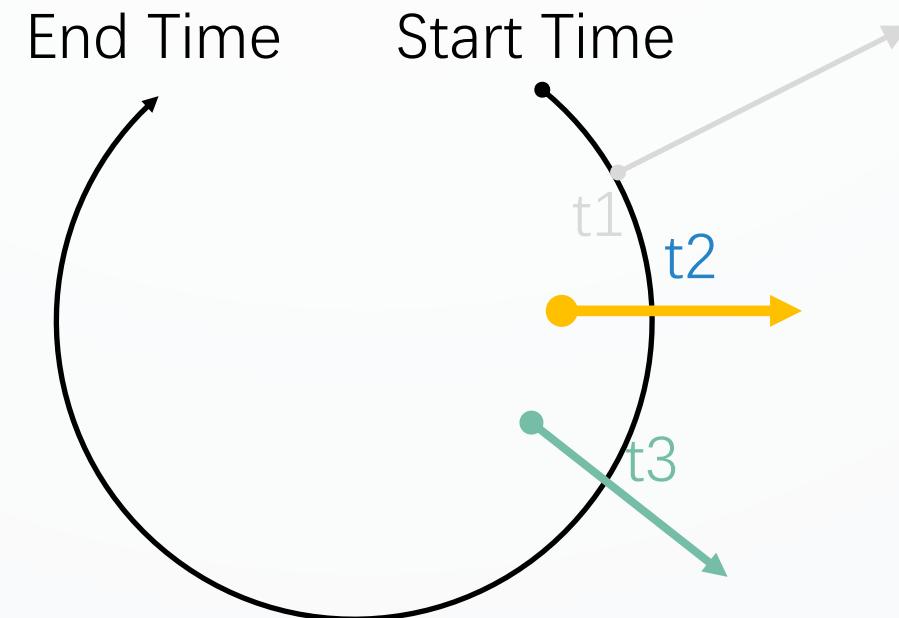


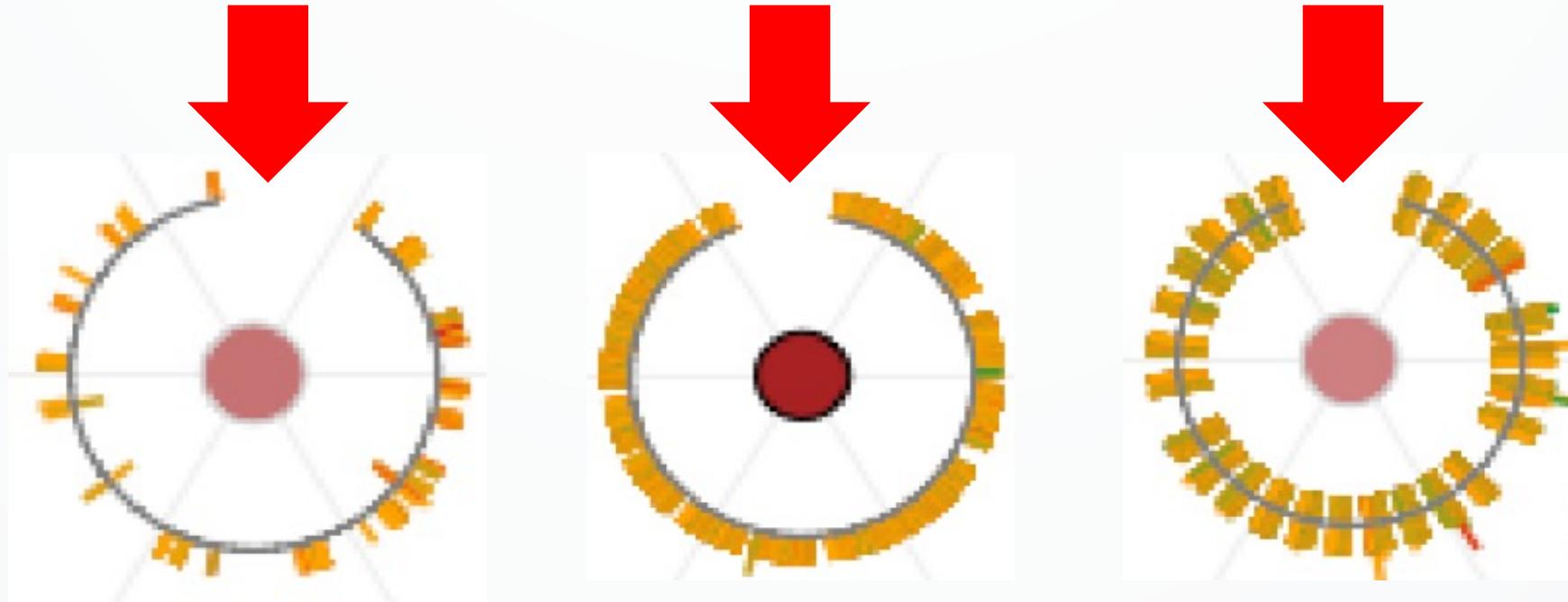
When the user posts or retweets a tweet, we draw the corresponding activity threads perpendicular to the time arc at the point when the activity occurred

Showing Posting and Retweeting Activities



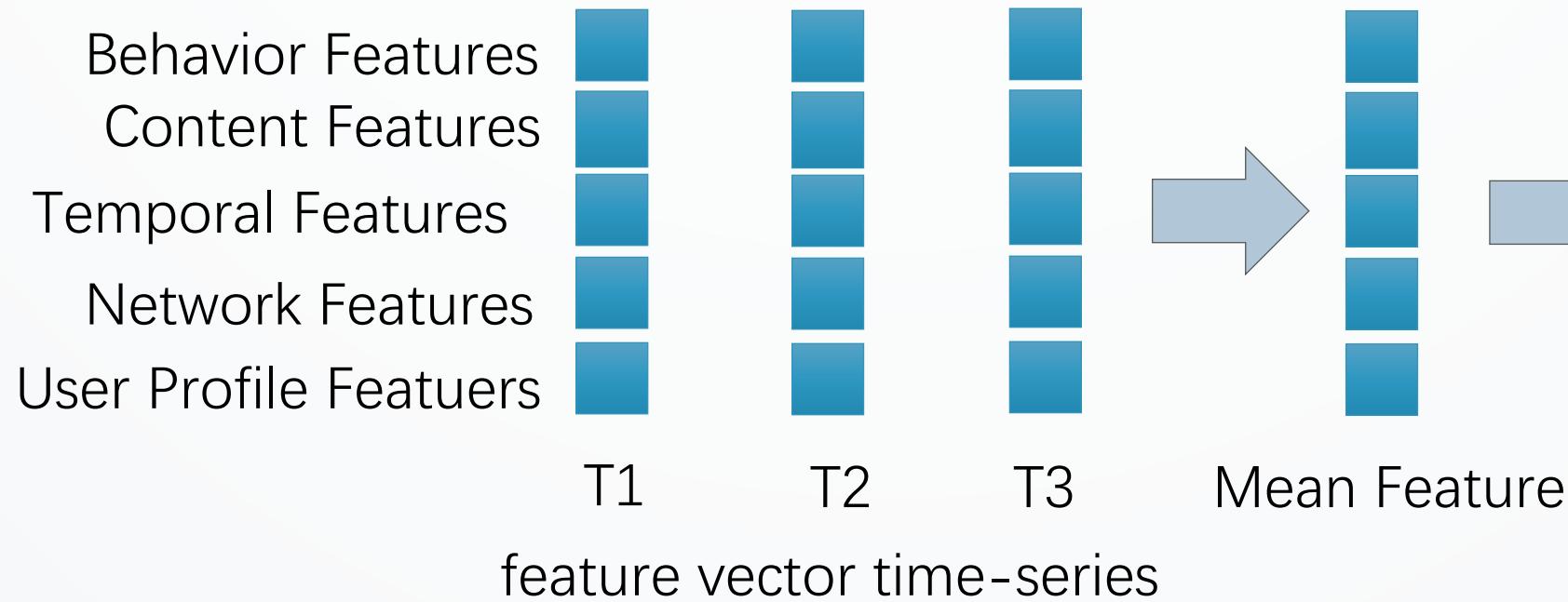
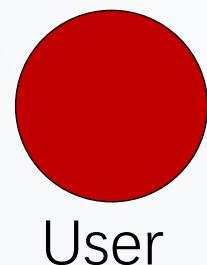
Showing Posting and Retweeting Activities



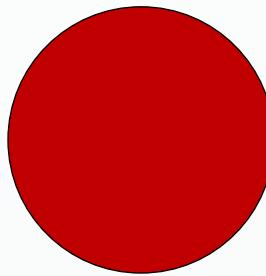


Who is the normal user?

Showing a User' s Features

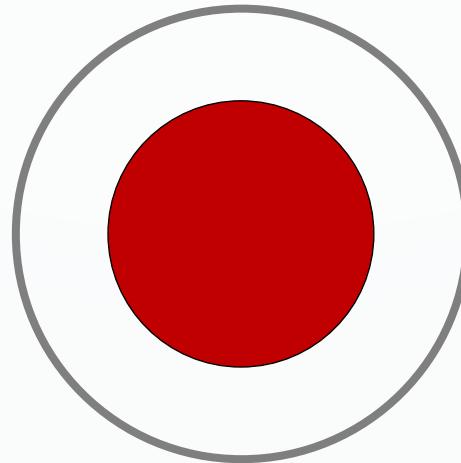


Showing a User' s Features



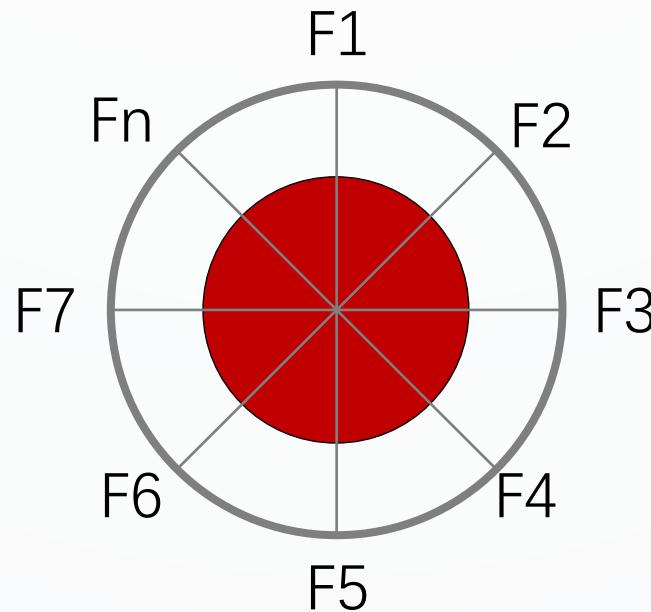
A user is visualized as circle sized by their importance and colored by their anomaly score

Showing a User' s Features



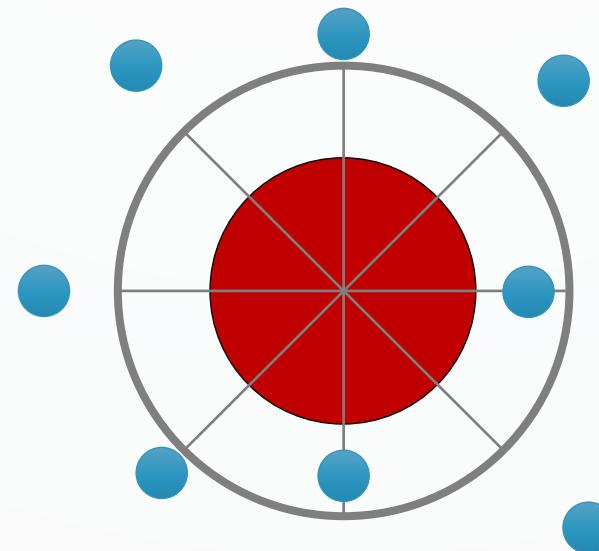
Using a baseline circle to indicate the mean feature values over all the users

Showing a User' s Features



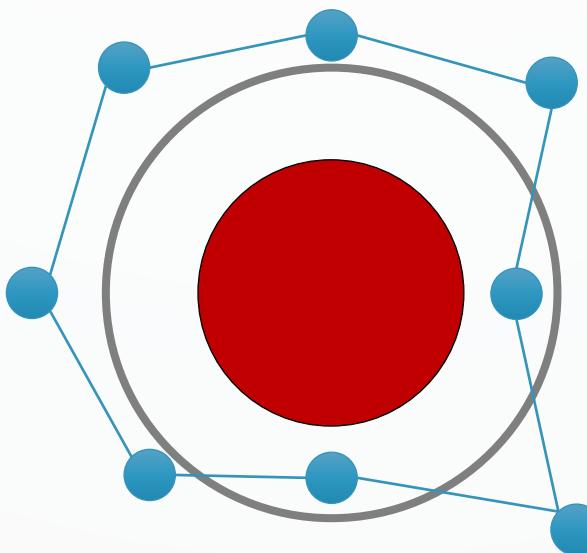
Feature axes are radially arranged

Showing a User' s Features

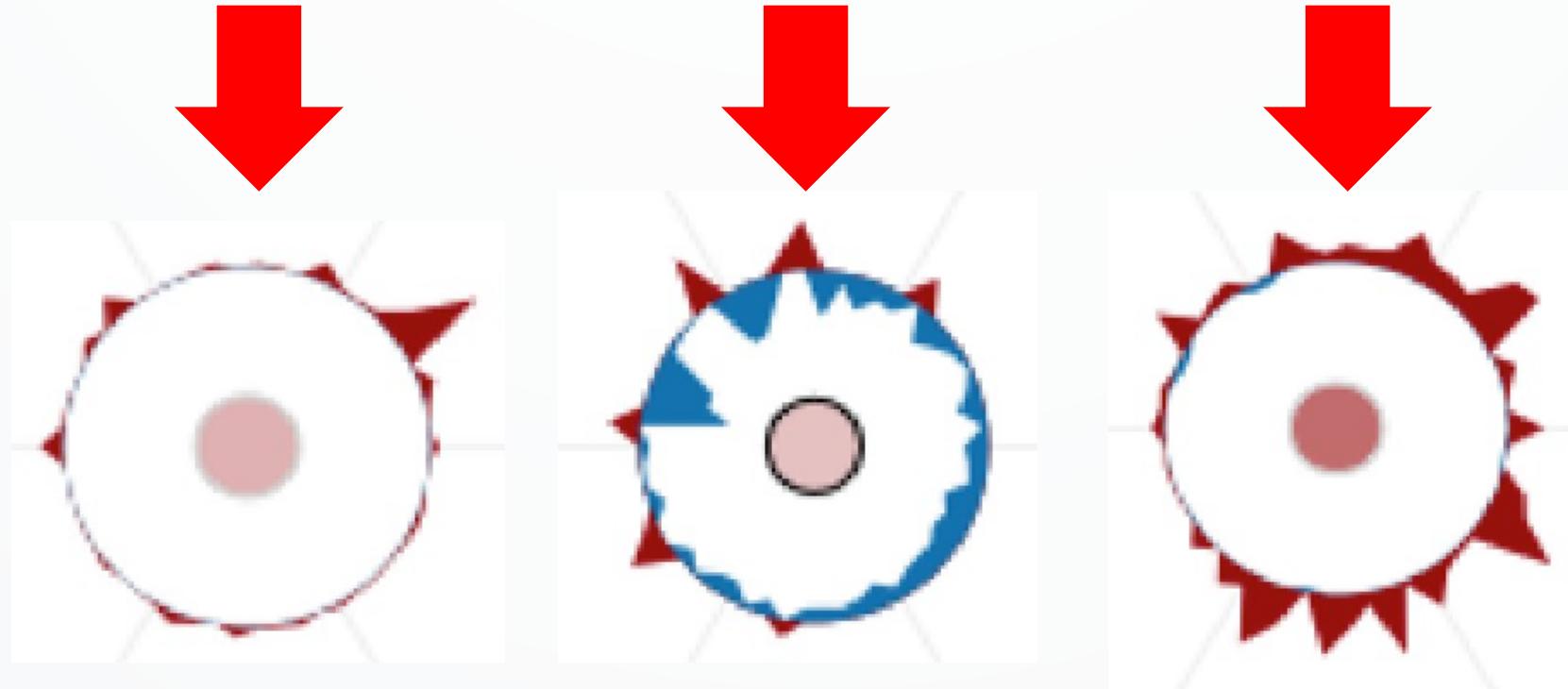


Plot the user' s feature values along the feature axes surrounding the baseline

Showing a User' s Features

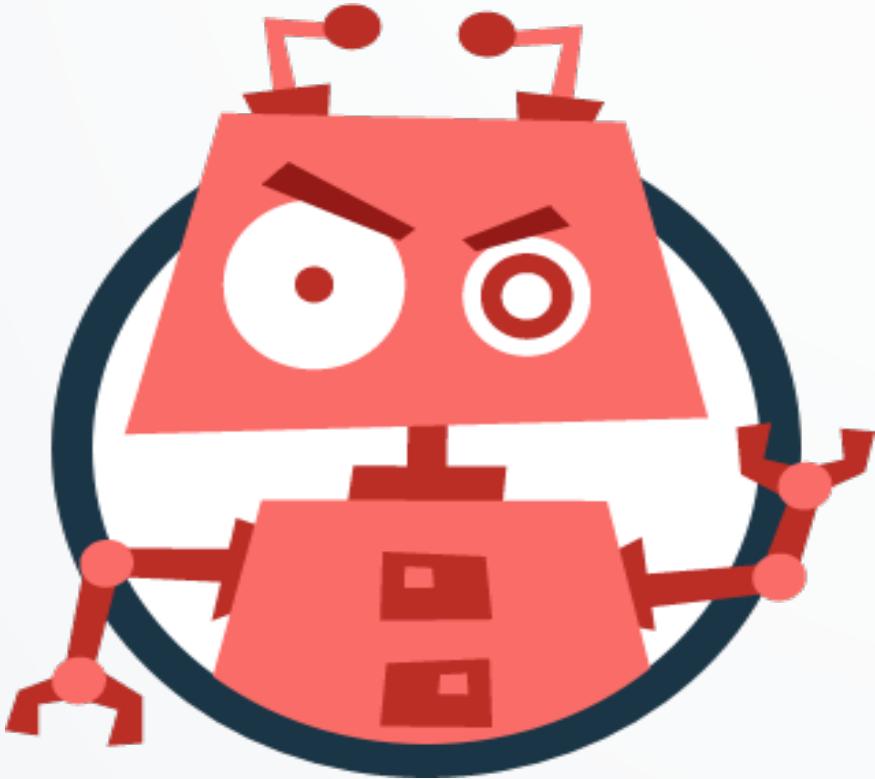


Connecting the data points to produce the feature glyph



Who is the normal user?

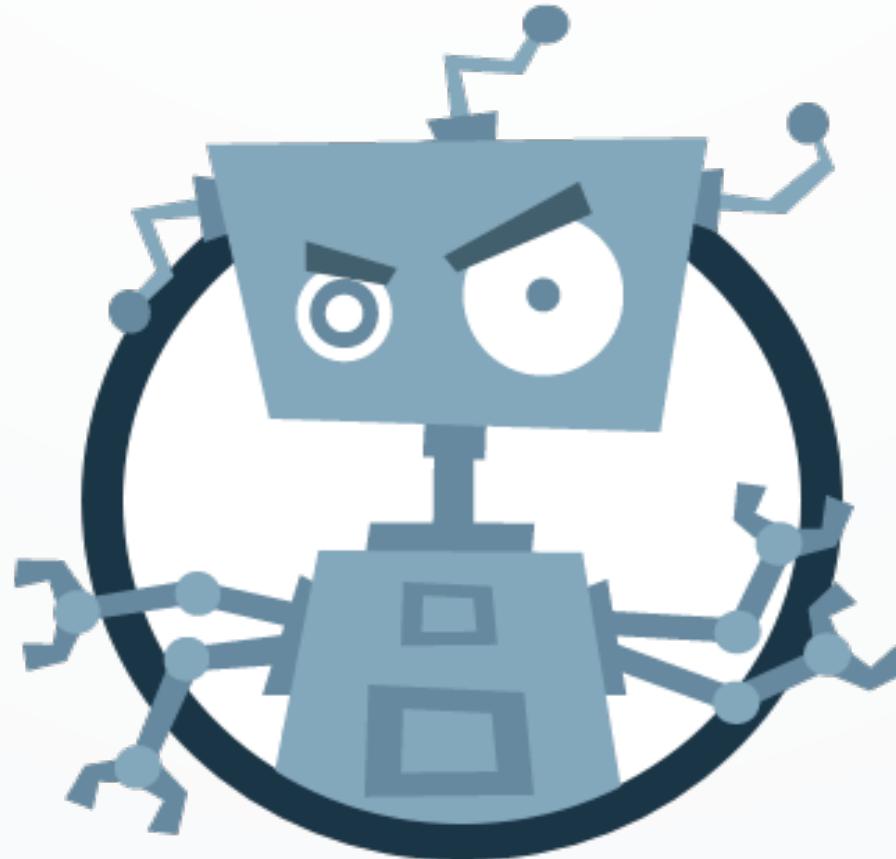
Evaluation: Bot Detection Challenge



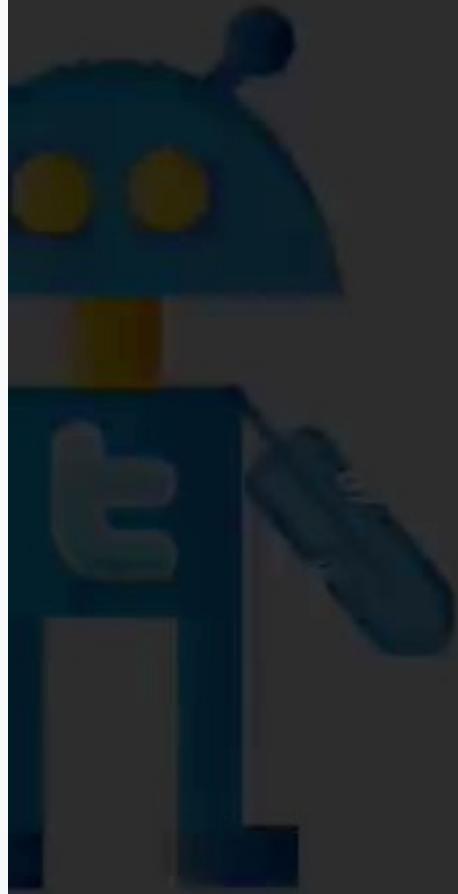
Bot Influence Challenge

- The goal of the influence challenge was to
 - design social bots in Twitter to promote the advantages of vaccination
 - influence a target network of users who are supporters of anti-vaccine
- Lasted for a month during Dec 2014
- 8000 target users, 4 million tweets

Evaluation



Bot Detection Challenge

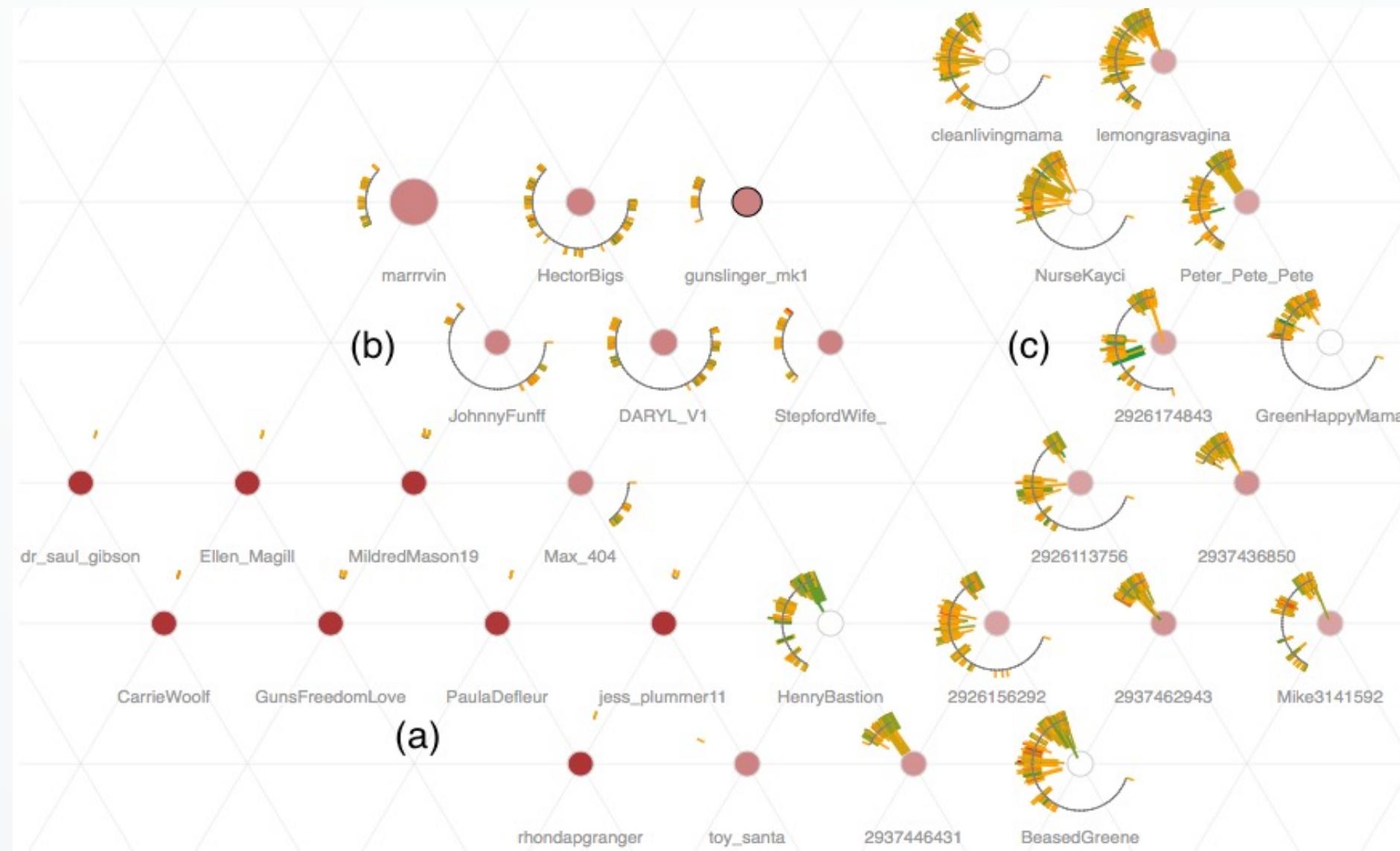


Goal:

Social Bot Detection Challenge

Detect fully automated social bot accounts on Twitter to promote the advantages of vaccination and influence a target network of users who tweet or retweet messages of anti-vaccine nature.

Final Results



课程总结

- 本节课
 - 什么是可视化?
 - 为什么要可视化?
 - 可视化的类型与发展趋势
 - 怎样对数据进行可视化?
 - 如何对可视化进行评估?
 - 怎样学习可视化?
 - 可视化应用案例
- 下节课
 - 数据与数据分析基础