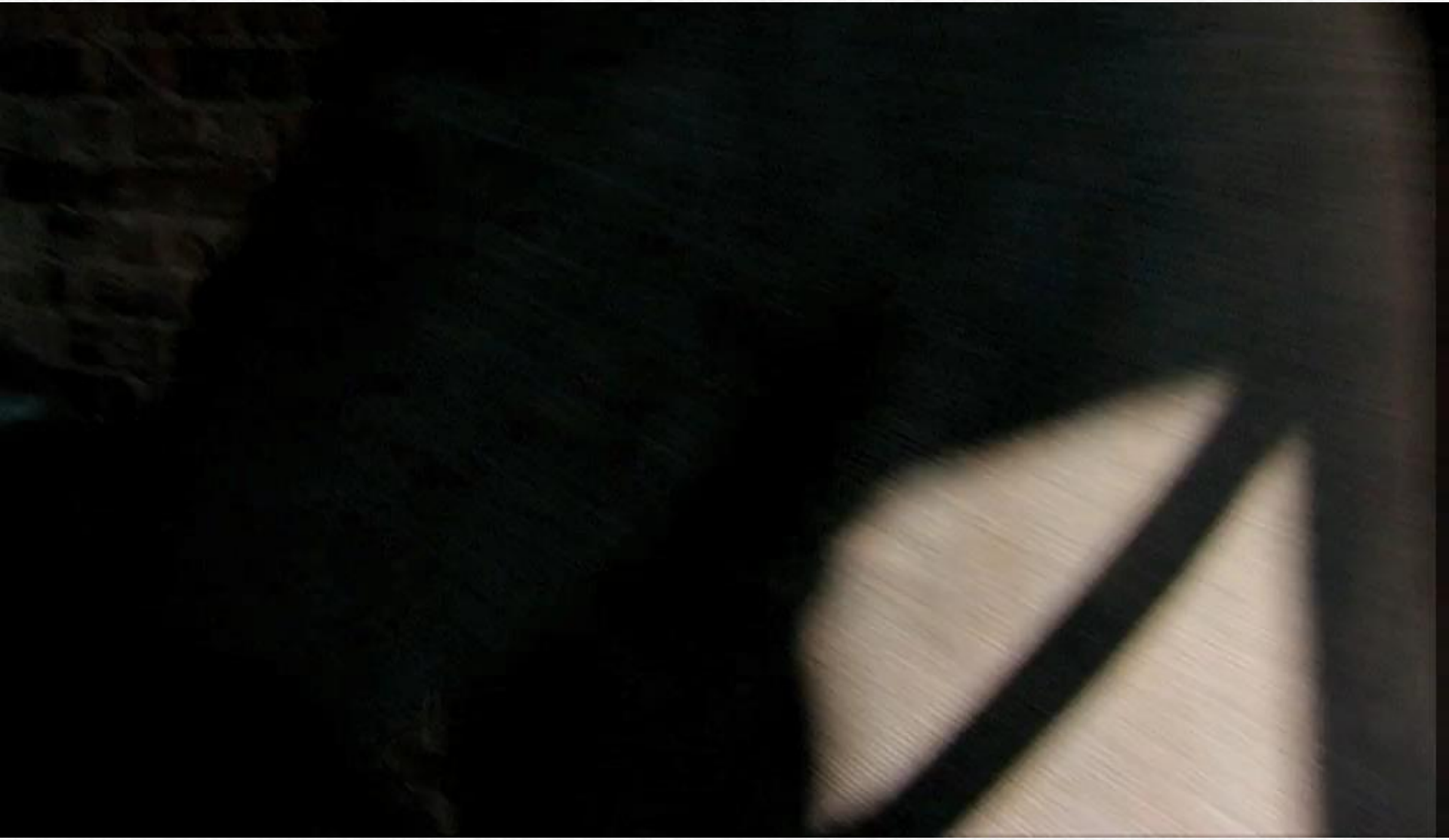
大数据可视化

挑战与趋势

陈为
chenwei@cad.zju.edu.cn
浙江大学CAD&CG国家重点实验室

数据可视化

五维统计数据的生动可视化



数据可视化

- 创建并研究数据的视觉表达 (Visual Representation)
  - 输入：数据 (data)
  - 输出：视觉形式 (visual form)
  - 目标：深入理解 (insight)

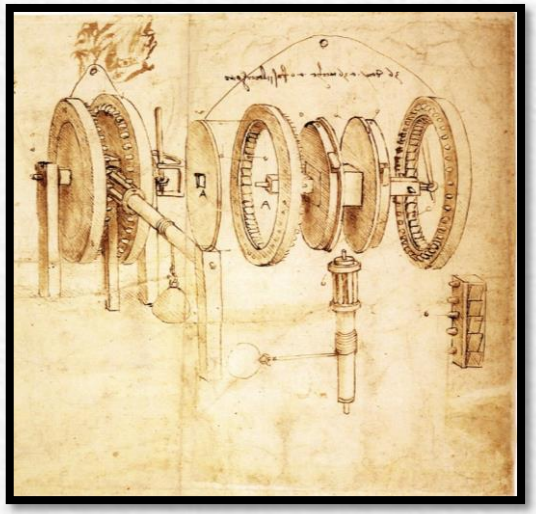
数据

视觉形式

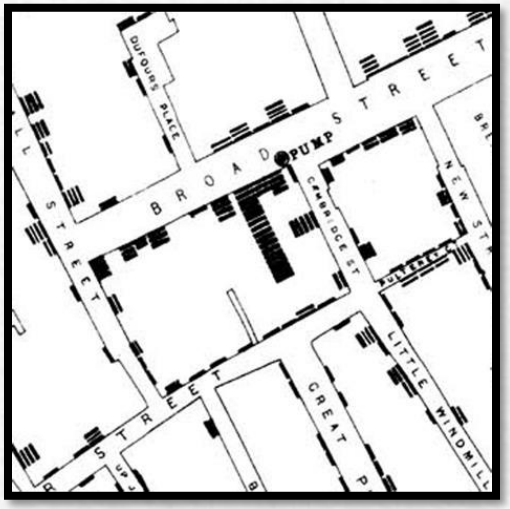
深入理解

数据可视化的主要任务

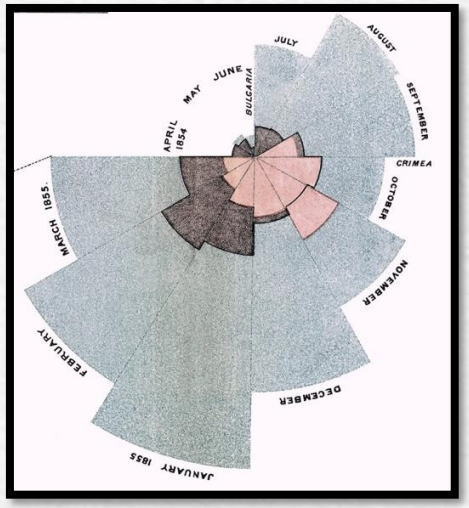
- 表示数据 - Represent
- 分析数据 - Analyze
- 交流数据 - Communicate



三维素描图



霍乱病例的分布

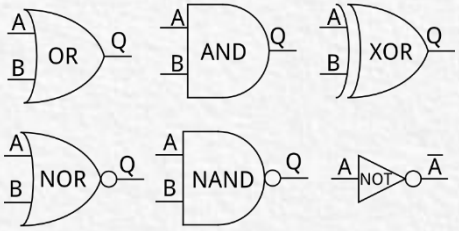


英国东征士兵死亡原因

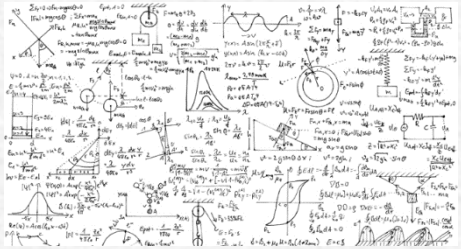
思维系统

表达 (符号) + 操作规则

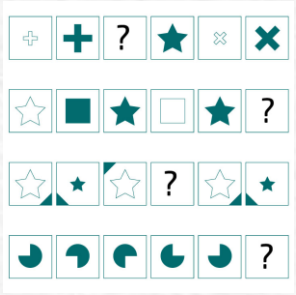
语言



逻辑

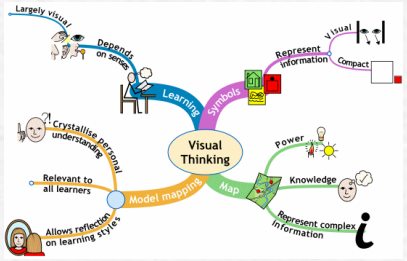


数学



推理与统计

计算



可视化与形象思维（认知）

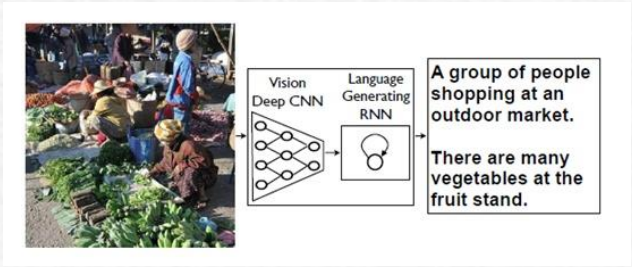
可视化与形象思维（认知）

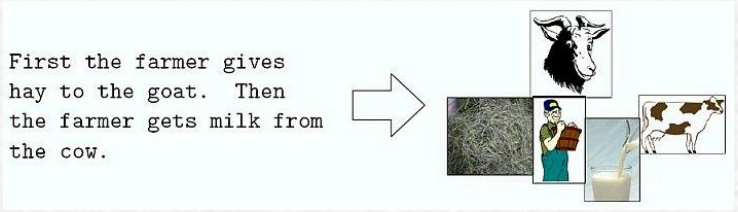
Why is a Diagram (Sometimes
Worth 10,000 Words

Larkin and Simon, Cognitive Science, 1987

一图胜千言

图表表达与句型表达具有信息和计算上的等价性
[Simon 1978]

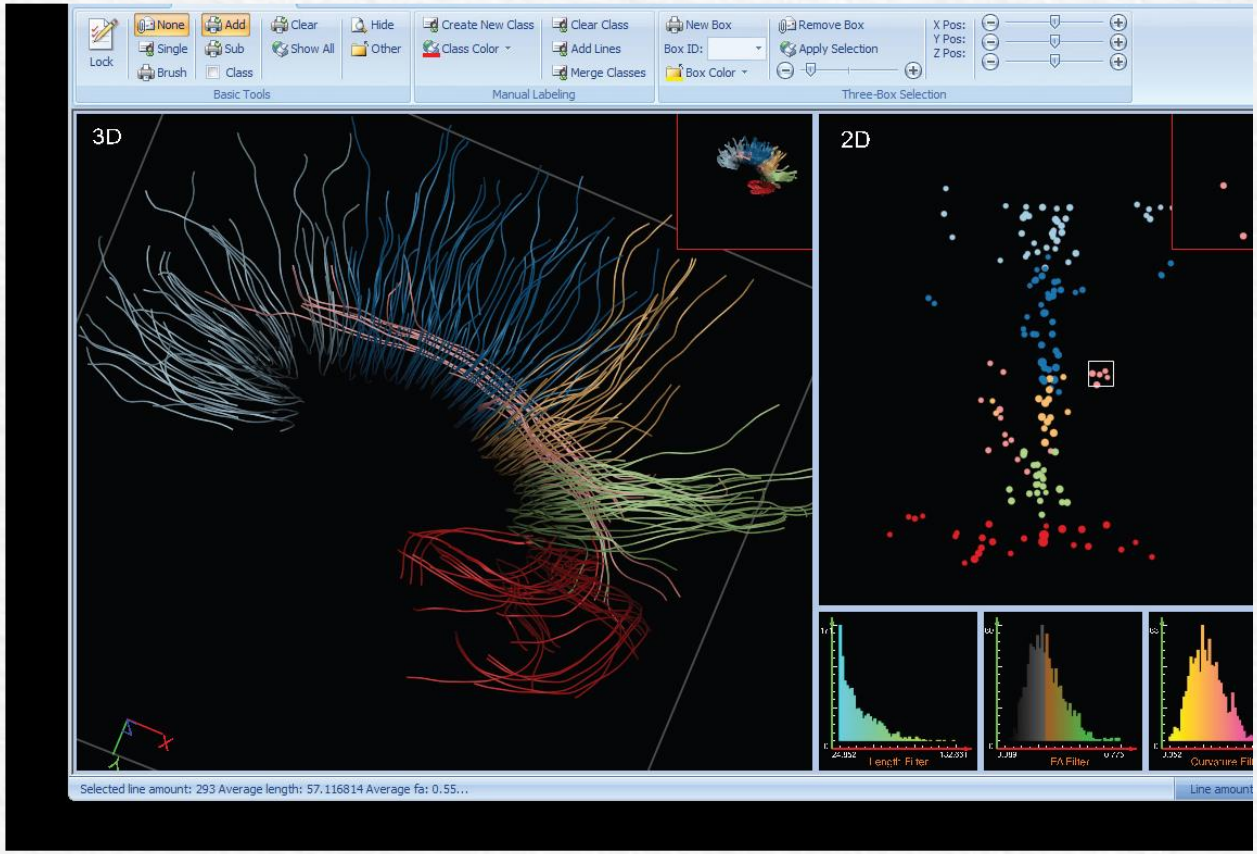




图像到文字的互相转换

一图胜千言

图表具有拓扑和几何的关联，将信息基于位置进行索引，所见处即所得；



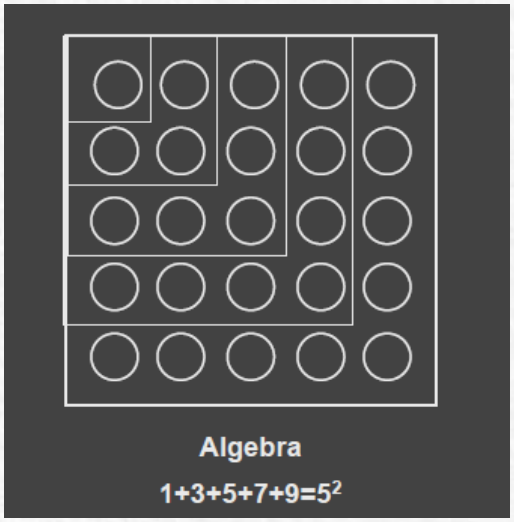
一图胜千言

- 句型表达具有时间或逻辑方面的序列，显式地表达了单个元素。
- 句型表达假设每句话是串行阵列；而图表表达有一个简洁的语义网络，认知时只需要在不同的节点间定位。

重要人物的关系分析系统

一图胜千言

在求解问题时，图表表达可以提供搜索与认知的便利；句型表达在搜索时需要记住更多的信息。



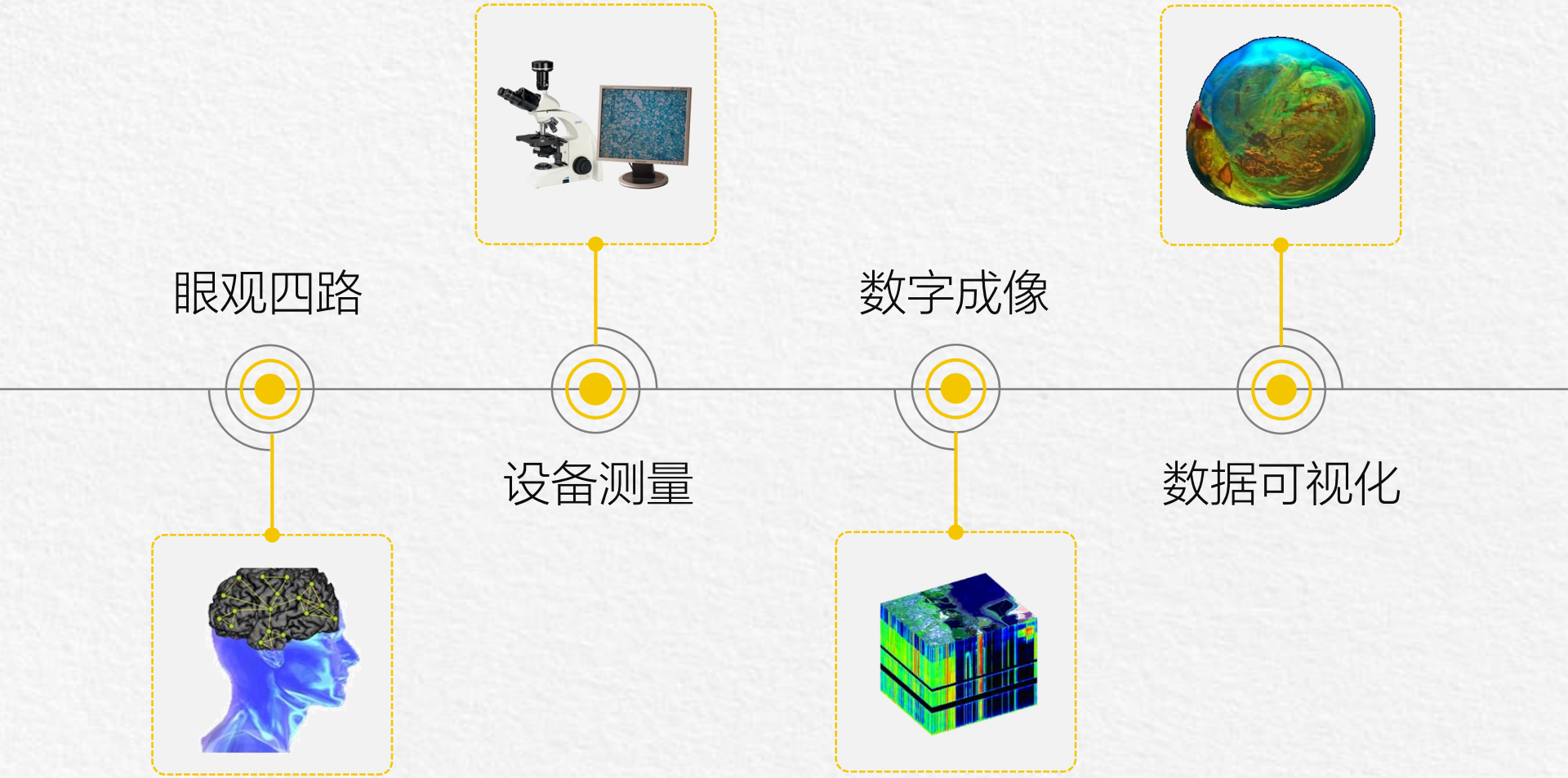
https://seeing-theory.brown.edu/bayesian-inference/index.html

可视化是一种外部认知的方法，即：如何利用人眼的感知能力和人脑之外的资源，提升人脑的认知能力。

Stuart Card，1946-

美国工程院院士
1970年代就职于美国施乐公司，
鼠标和GUI的主要推动者

可视化：大数据时代的成像利器



交互智能可视分析

不可信

不存在可信的自动分析方法

未定义

分析任务没有良好的定义,甚至不知道任务是什么

人更有效

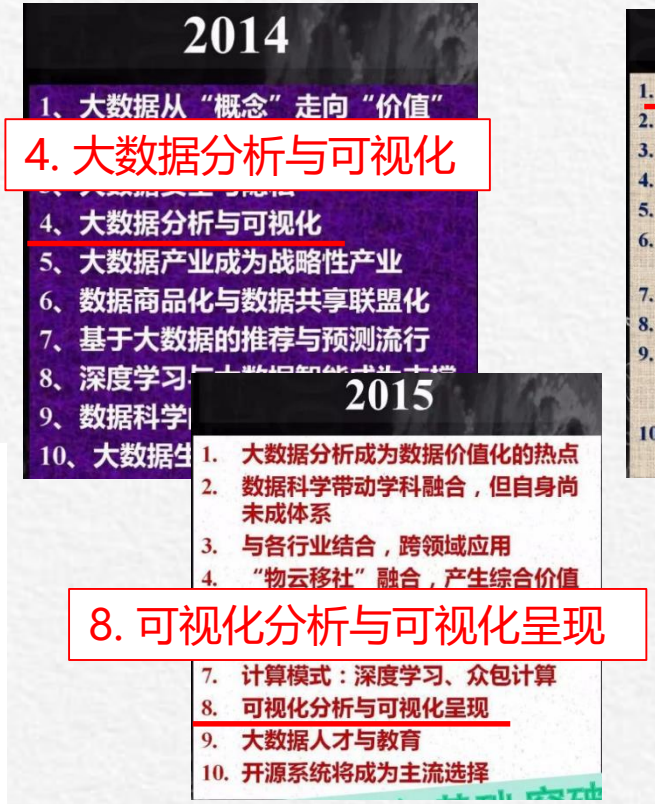
应急、复杂环境和对抗事件等条件下,人的智能更有效和可靠

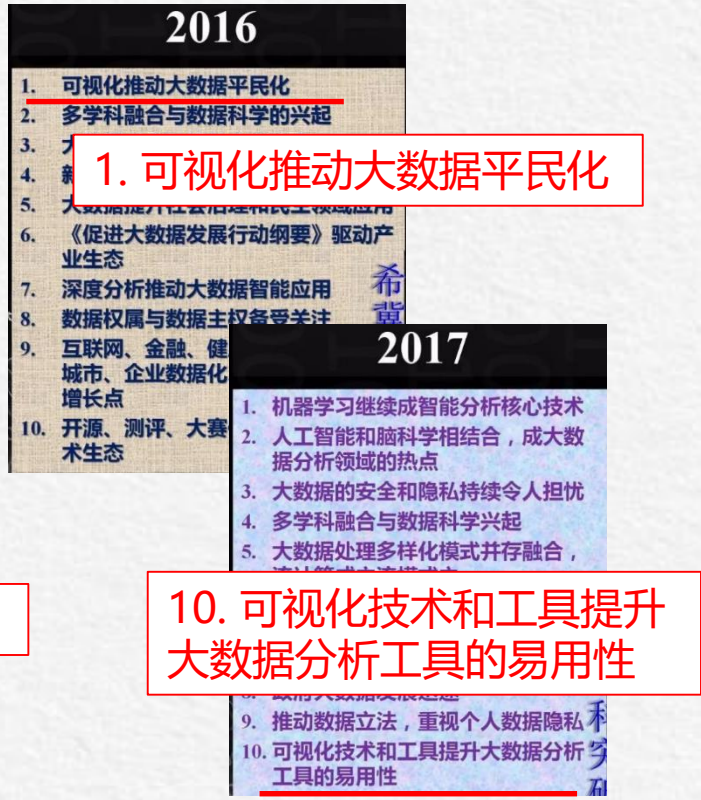
大数据可视化
现状

新一代人工智能规划中大数据智能

可视化与可视分析是人类理解数据的导航仪：运用与人类视认知相一致的图形展示数据内在结构与规律，增强理解和分析效率。

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

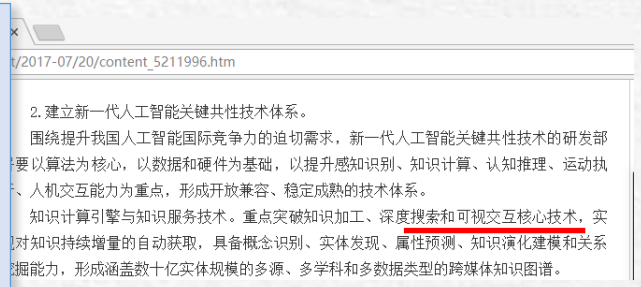




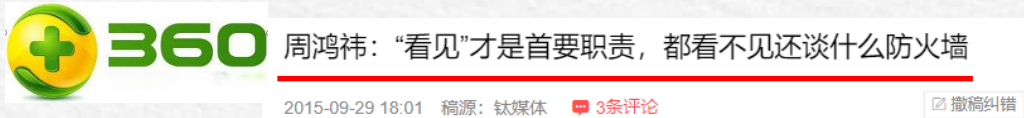
重要性：中国计算机学会每年发布十大大数据发展趋势报告

新一代人工智能规划中大数据智能

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



2008年后，美、欧盟、日均成立国家可视分析研究中心。国内外著名企业成立独立部门，研发新兴可视化与可视分析技术。



急迫性：2018年11月，美国拟提议的最新14类技术出口管制之第六条：数据分析（可视化、自动分析算法、上下文感知计算）

新一代人工智能规划中大数据智能

可视化与可视分析的核心挑战是：如何利用人的感知能力，增强有限的认知能力，以应对理解和分析复杂数据的迫切需求。

新挑战	新理论	新技术
数据复杂 数据尺度海量 时域演化多变 内蕴动态关联	**高效可视表达与呈现方法**	**通用信息可视化框架与软件**
认知低效 显示空间固定 视觉感知有限 任务交互无限	**探索式可视分析理论与方法**	**沉浸式可视分析软件与装备**

可视化与人工智能 1.0

可视化与从数据到知识的链条中的每个环节互助互利

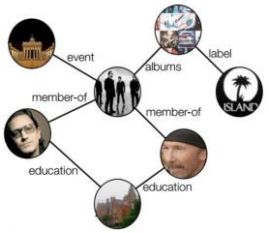
数据 → 数据清洗 → 数据整合 → 数据组织 → 数据分析 → 知识

Trifacta

Dataiku

IBM DSW

flickr

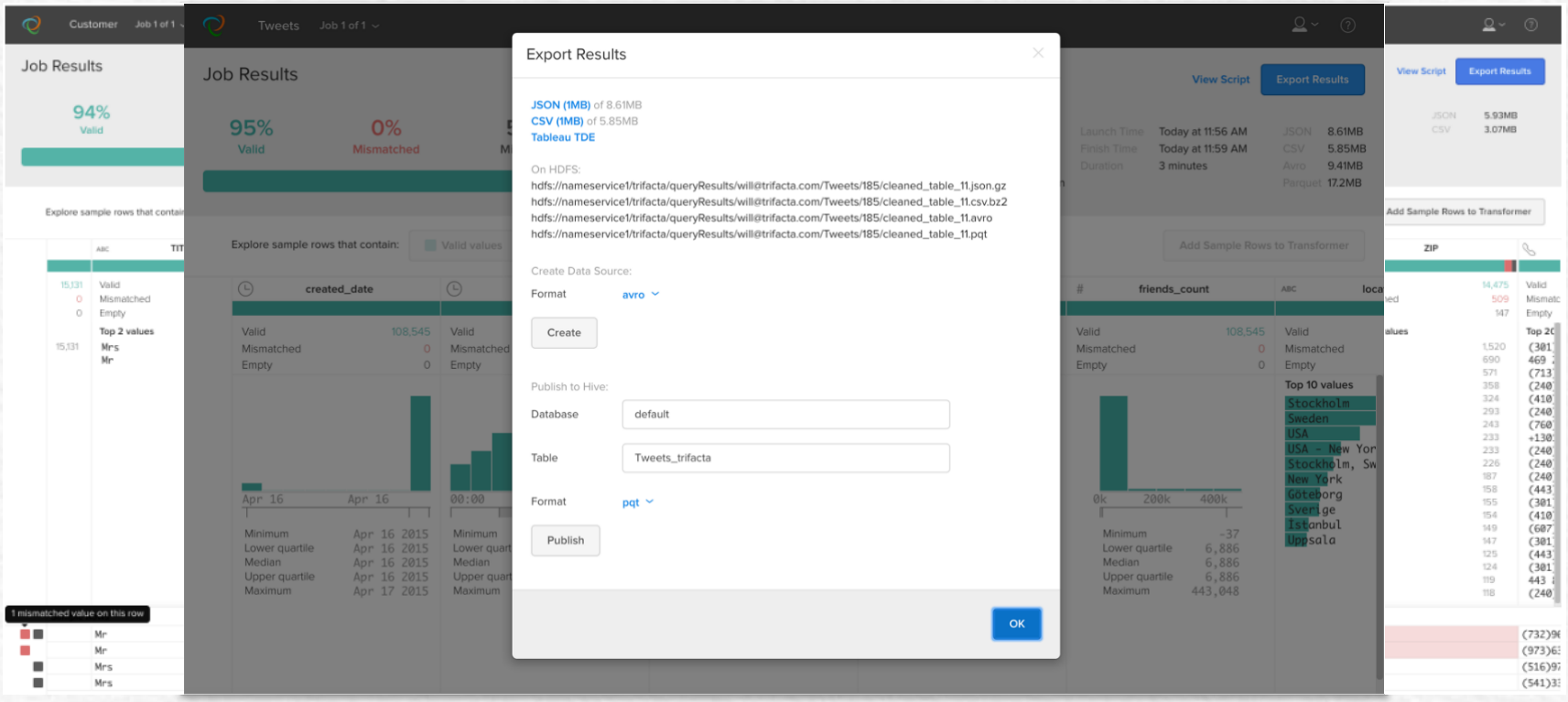


可视化与人工智能 1.0：可视数据清洗

TRIFACTA

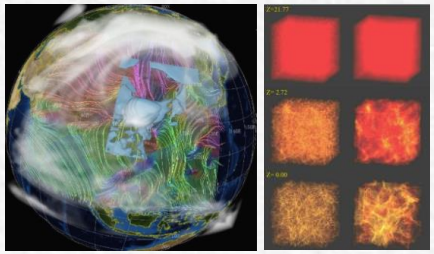
www.trifacta.com

探索　结构化　清洗　丰富　验证　发布



可视化重要应用

大科学



大安全

大工程

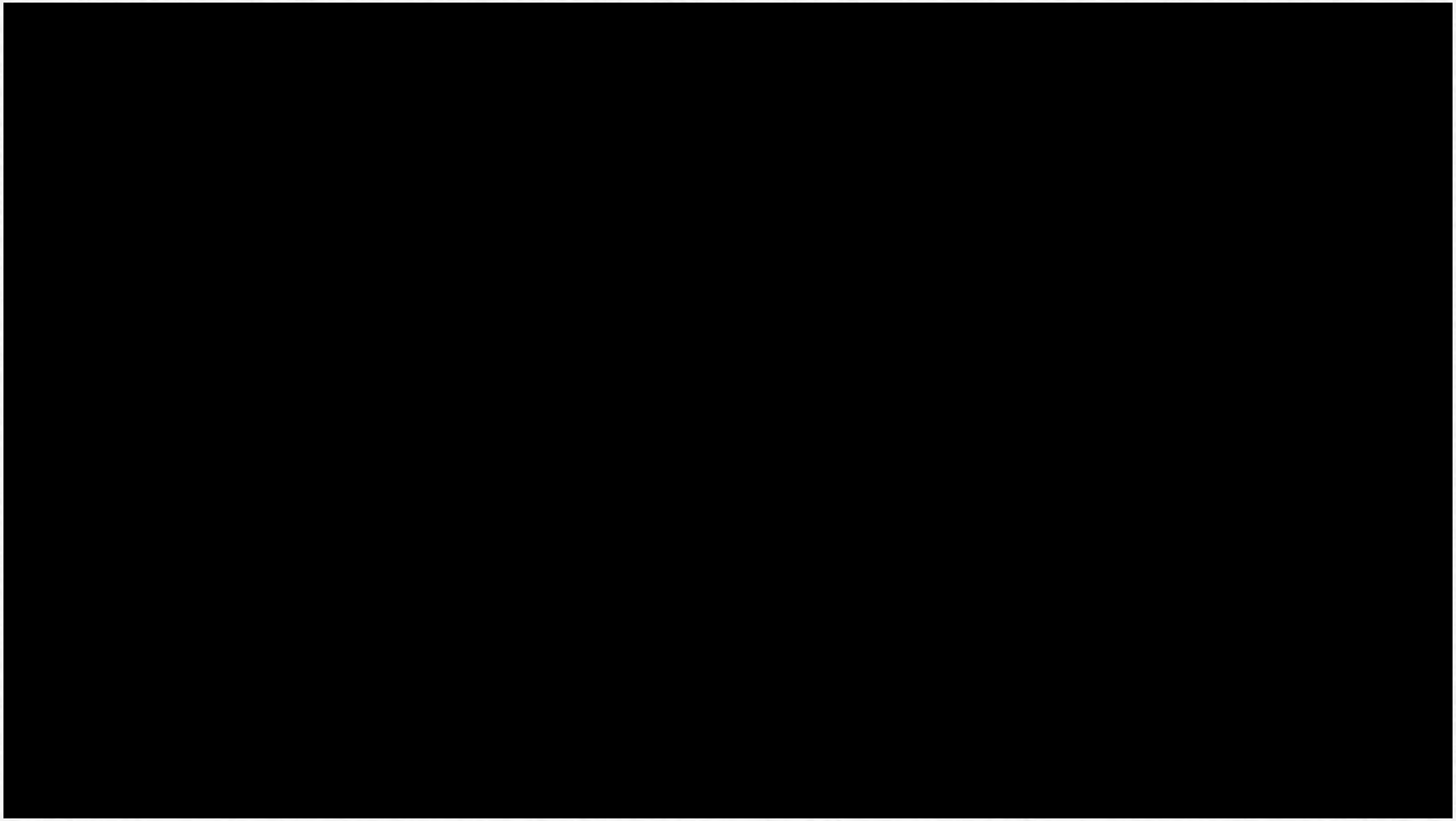
大数据
可视化

物联网与智慧城市

互联网与社交媒体

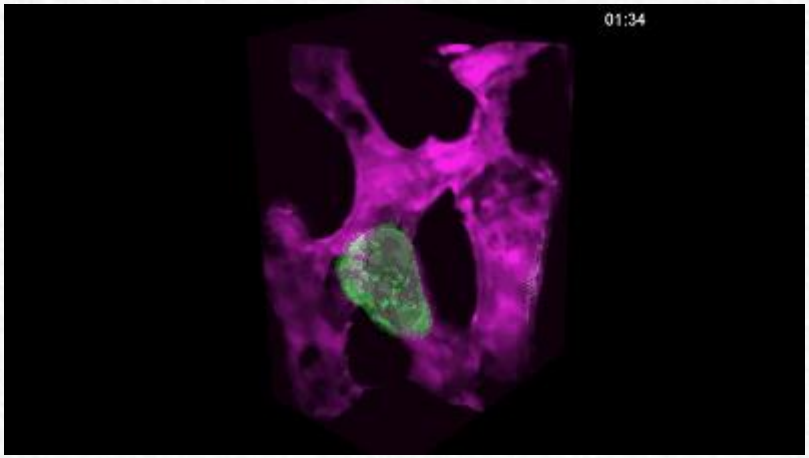
可视化重要应用：大科学

可视化成为基础自然科学研究的**必要手段**，是科学大数据发展的**必需**

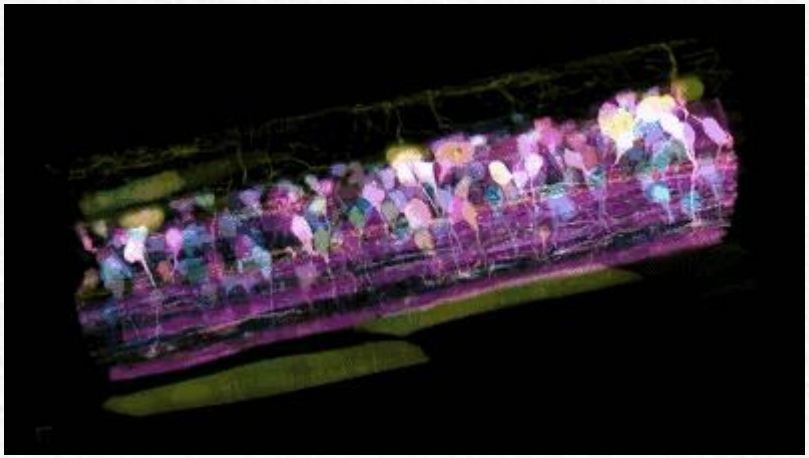


海洋

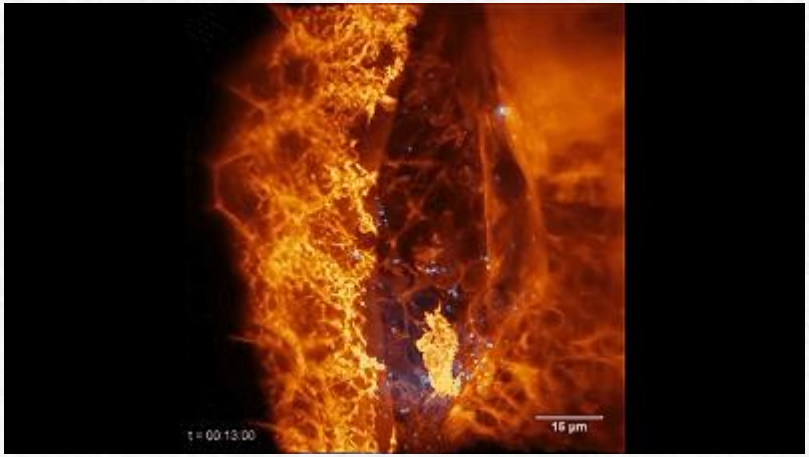
活组织细胞运动三维影像可视化



乳腺癌细胞转移



脊髓神经回路



免疫组织穿梭

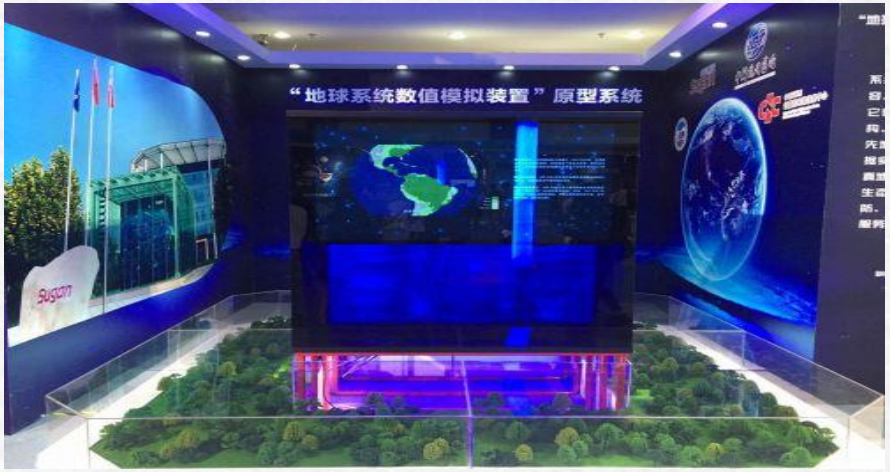
免疫组织穿梭（换个角度）

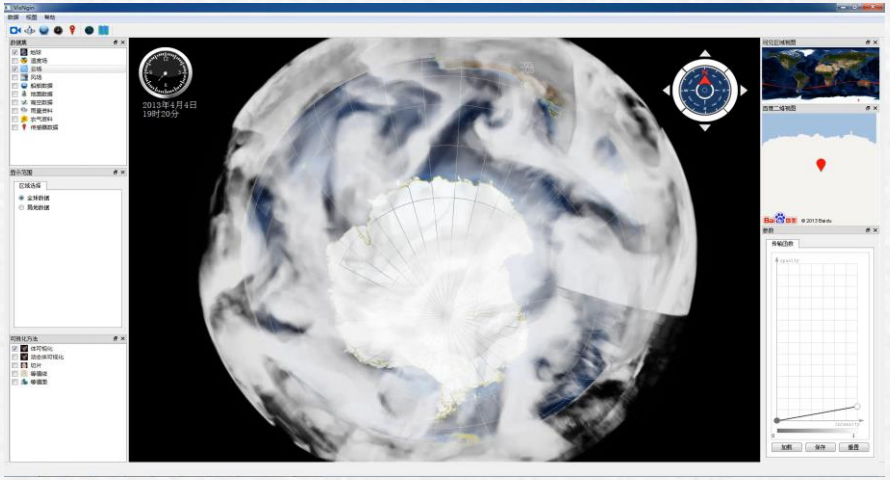
Liu T, Upadhyayula S, Milkie D E, et al. Observing the Cell in Its Native State: Imaging Subcellular Dynamics in Multicellular Organisms. Science, 2018.

可视化重要应用：大科学

“地球系统数值模拟装置”

国家十三五重大科学装置项目，拟投资13亿，其中可视化部分7000万。



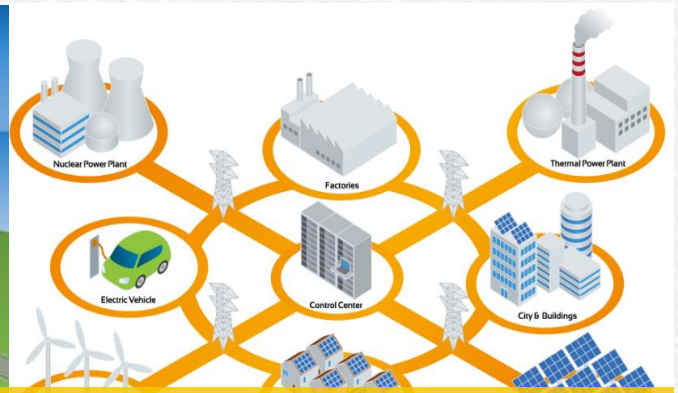


国家卫星气象中心全球大气数据可视化平台
（浙江大学）

可视化重要应用：大工程

可视化是对大工程仿真、实测、融合、预测、测试等不同环节产生的信息进行综合理解与分析的**必要手段**

智能交通

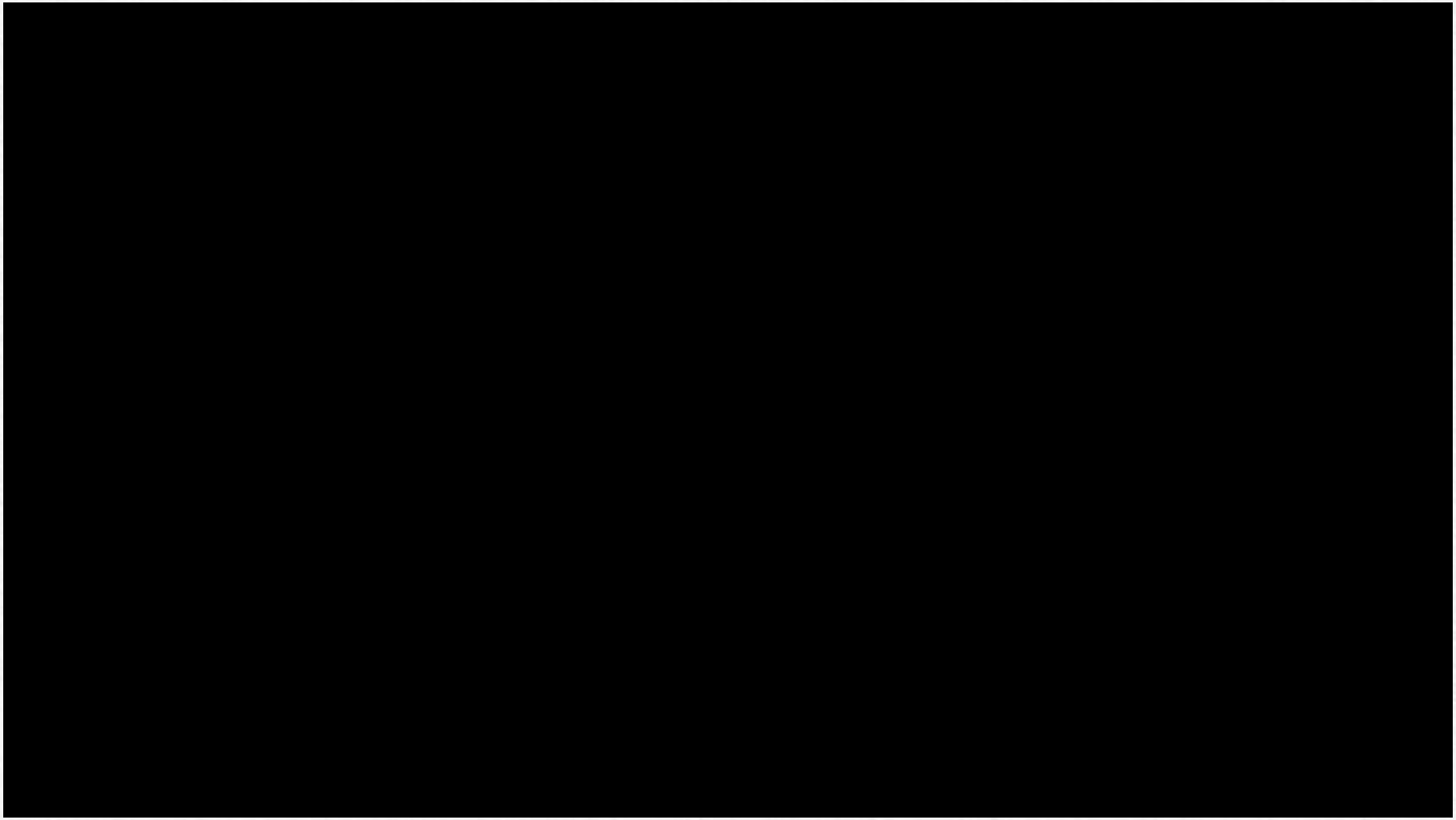


智能电网

智能物流

智能制造

可视化重要应用：大工程



博世智能工厂数据可视化分析

（美国博世研究院，浙江大学）

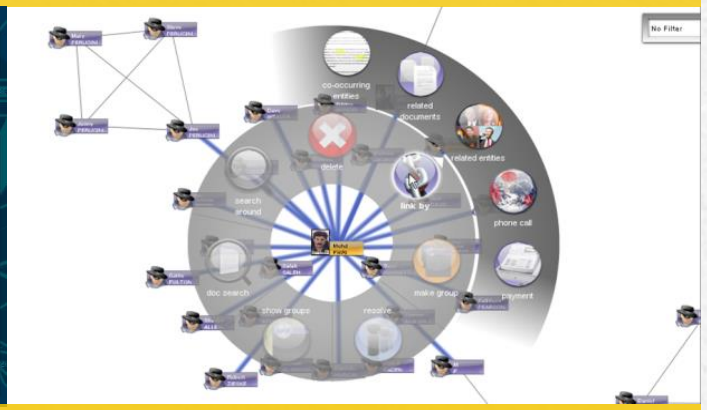
可视化重要应用：大安全

可视化是面向与人博弈任务的智能分析的最主要的**交互界面**

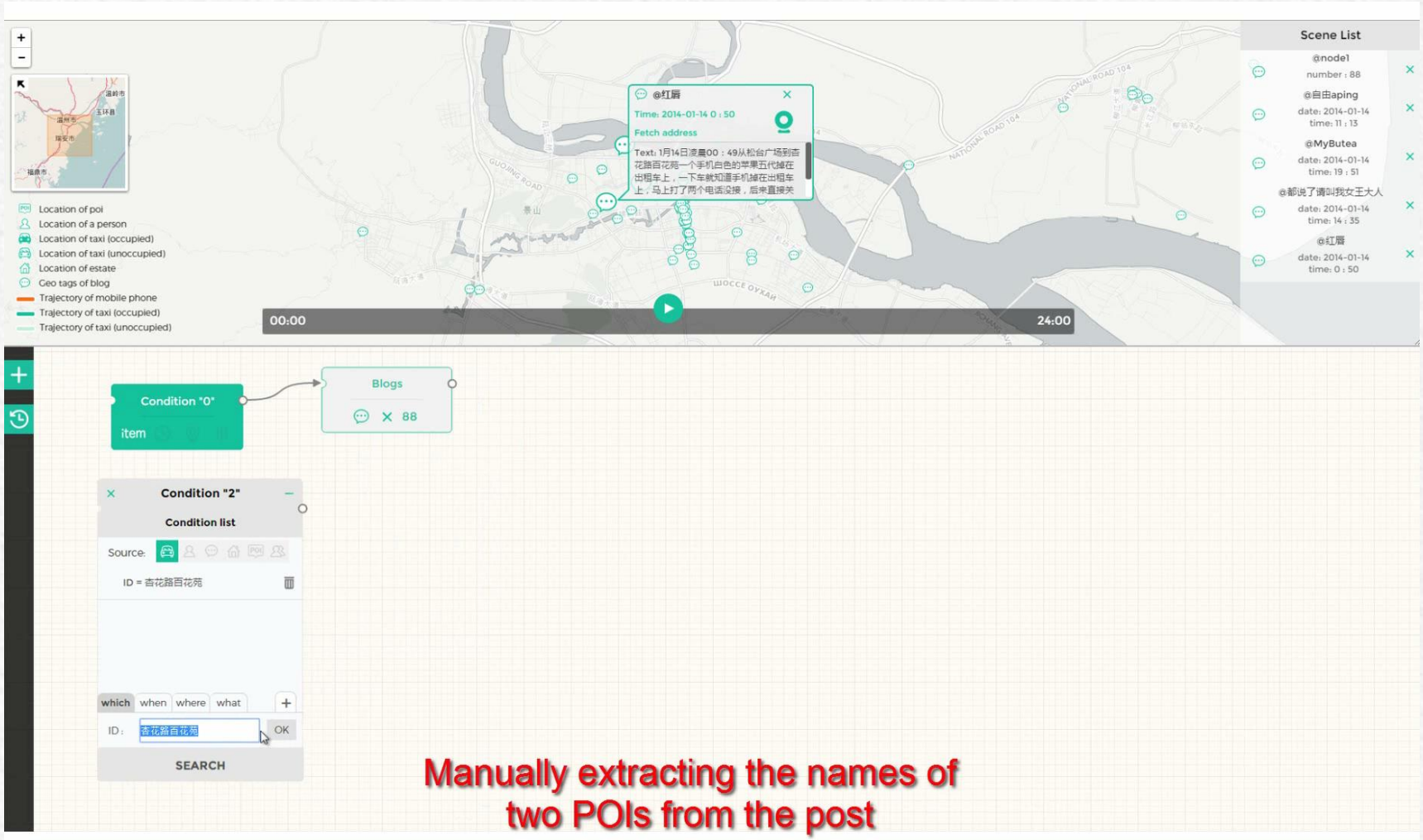
国土安全

网络安全

公共安全

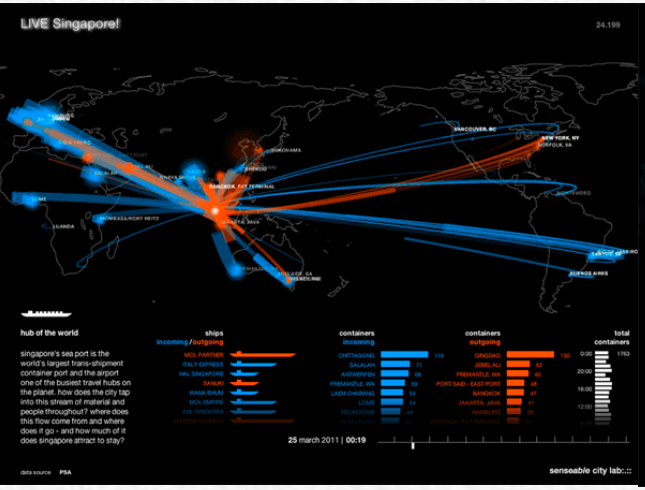


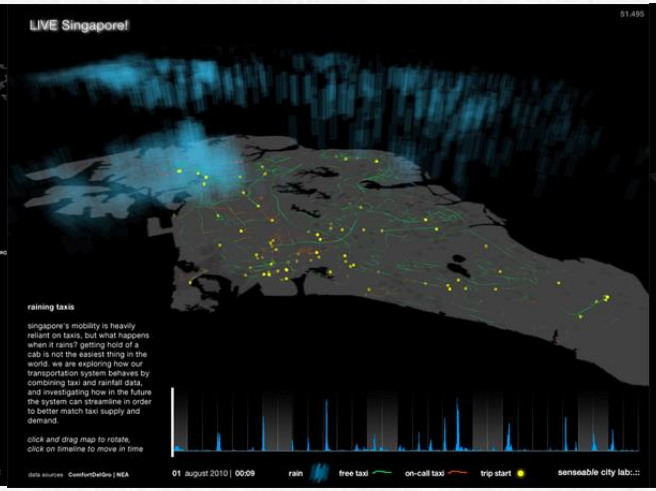
金融安全

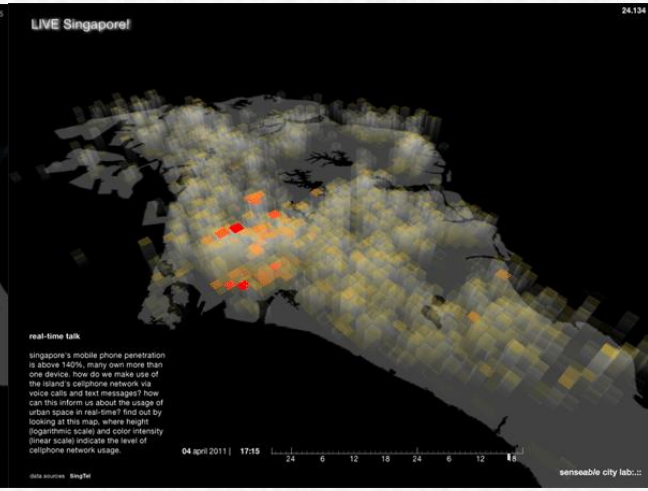


基于定位、网络和社交数据的时空城市数据可视查询（浙江大学）

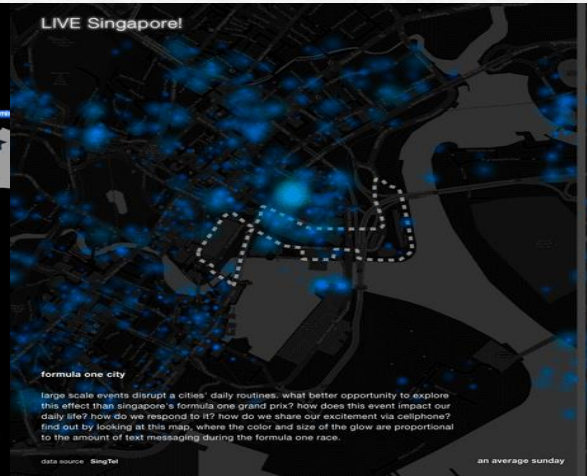
可视化重要应用：物联网与智慧城市

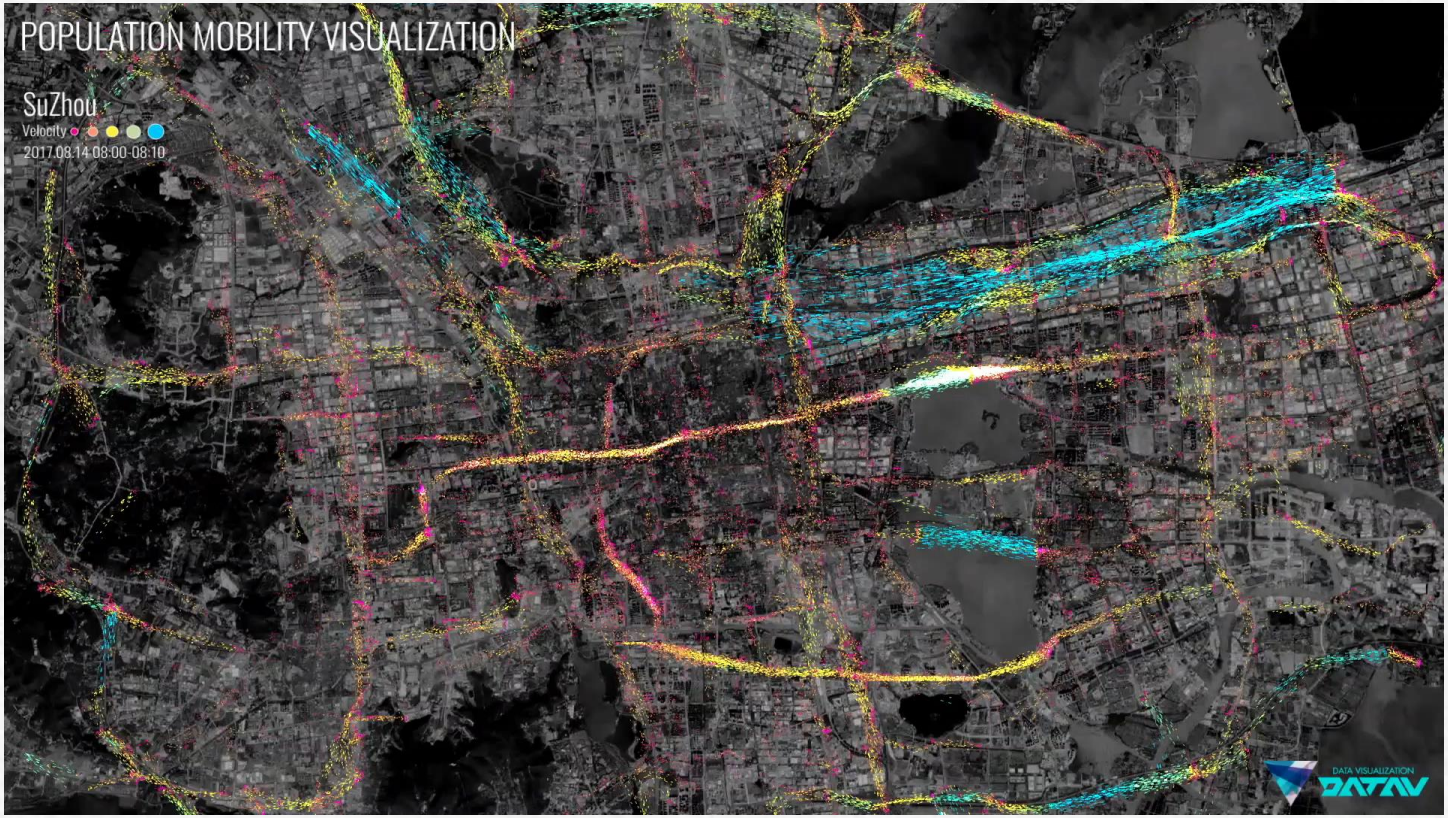






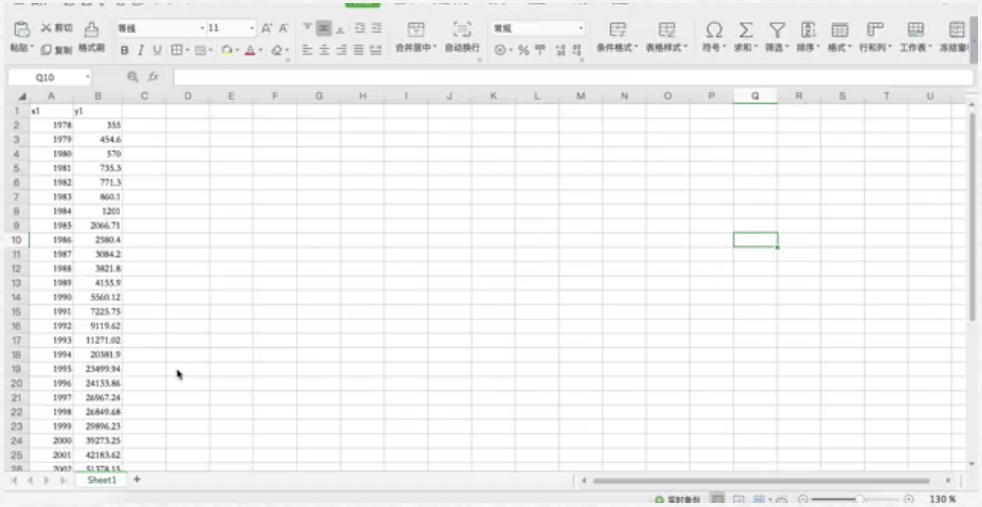
可视化是基于CPS数据进行规划、理解、决策的**敏捷分析**途径



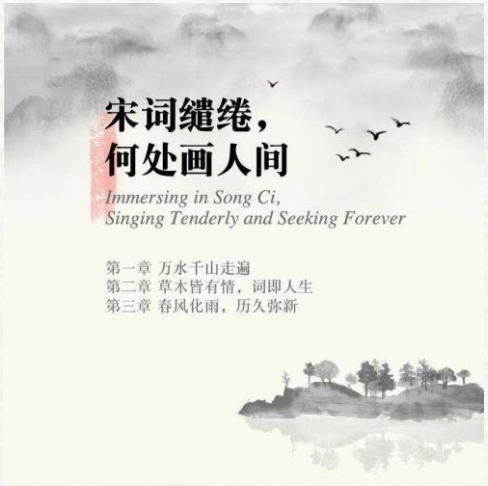

POPULATION MOBILITY VISUALIZATION
SuZhou
Velocity
2017.08.14 08:00-08:10
DATA VISUALIZATION
DATAV

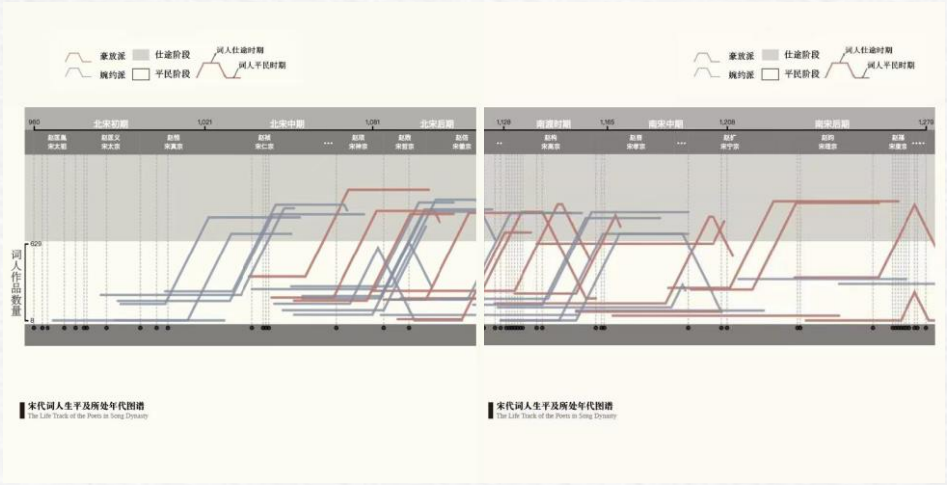
阿里-浙大：苏州城市大脑

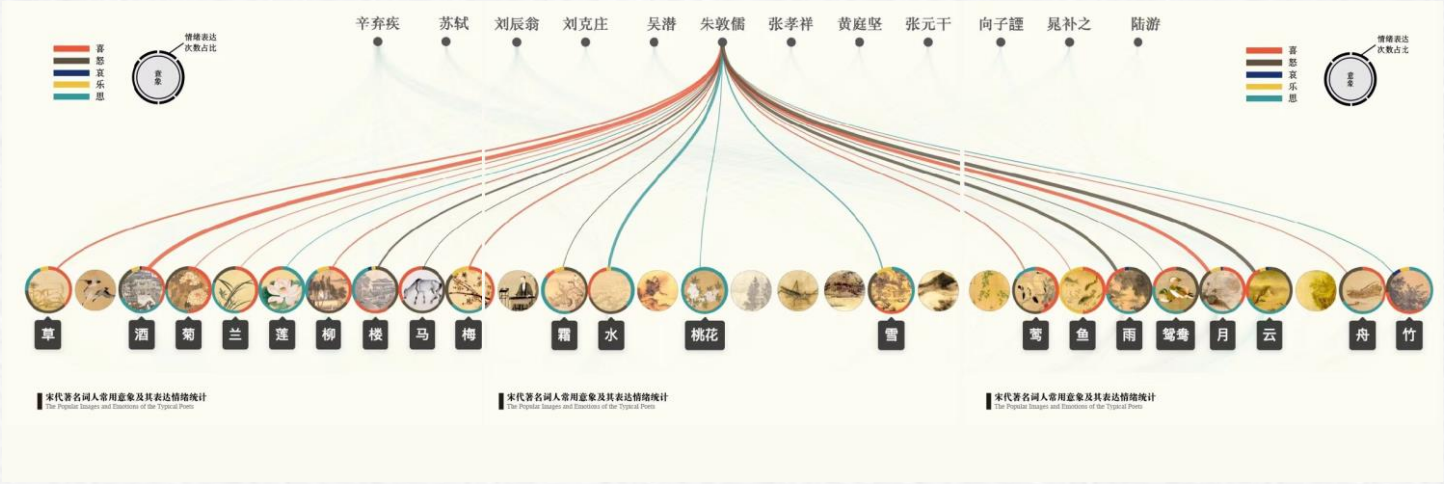
智能新闻与短视频（新华智云）



浙大-新华网制作的数据新闻：宋词文化可视化





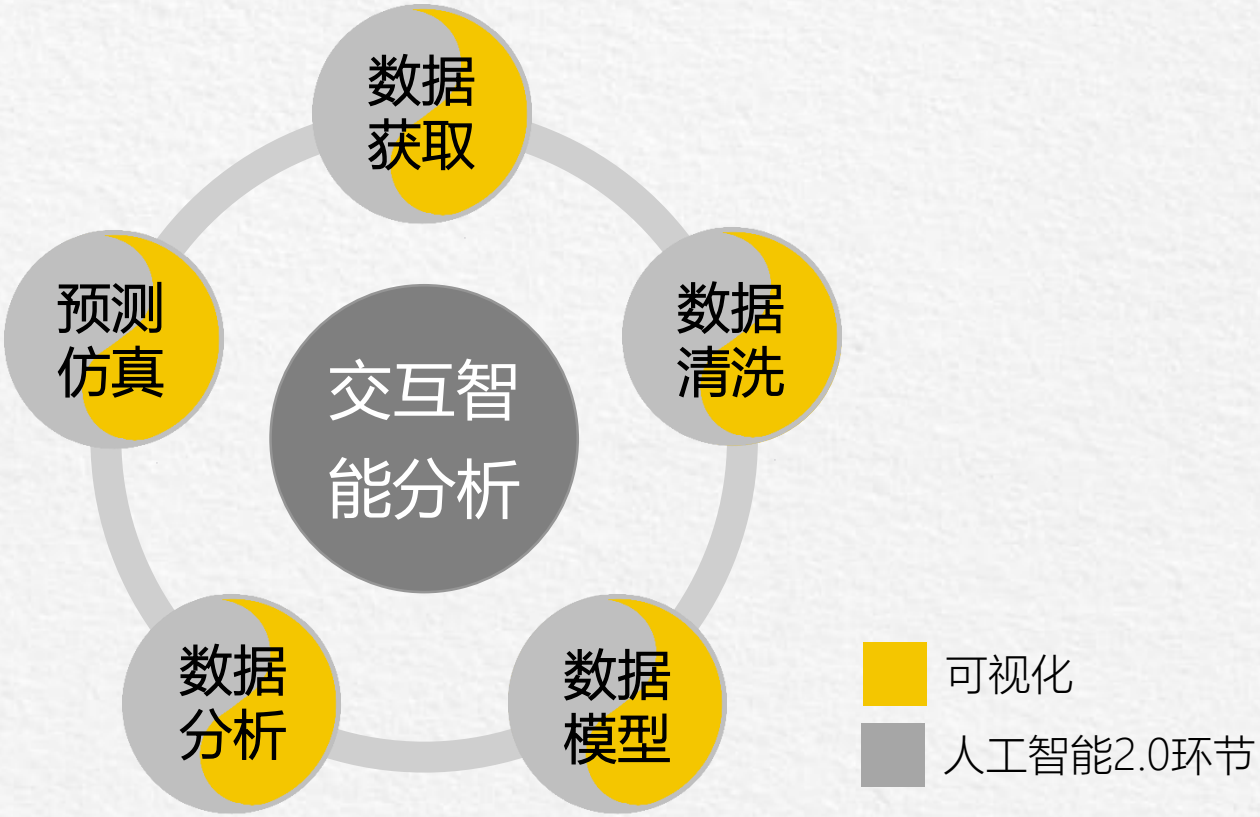


互动版（pc端支持更多交互）：http://fms.news.cn/swf/2018_sjxw/quansongci/index.html#/
图文版链接 http://www.xinhuanet.com/video/sjxw/2018-09/05/c_129947285.htm
微博链接 http://t.cn/Rsfnshp

大数据可视化
展望

智能交互 2.0

人工智能2.0的每个环节与**可视化**融合



基于预计算的表格型数据快速查询

Scenario 3
Taxi trajectory dataset

面向科学研究的智能交互计算

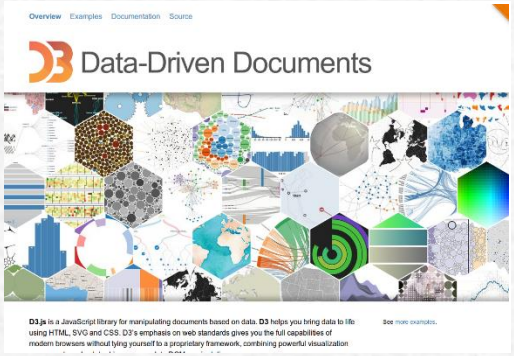
科学研究中的**计算、数据、分析和人智**的交互融合

Usage Scenario

Dataset: Bitcoin trading networks(207689 nodes, 547500 links)

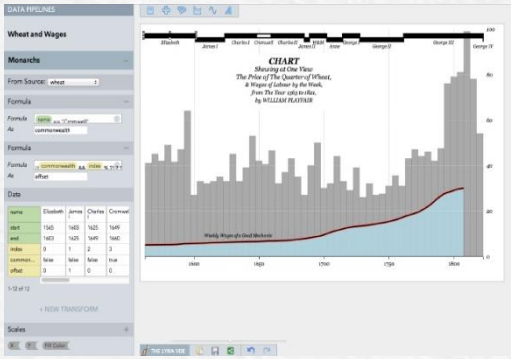
面向行业的通用可视化软件与组件库研发

以数据模态或应用场景区分的可视化**标准件**与**通用系统软件**



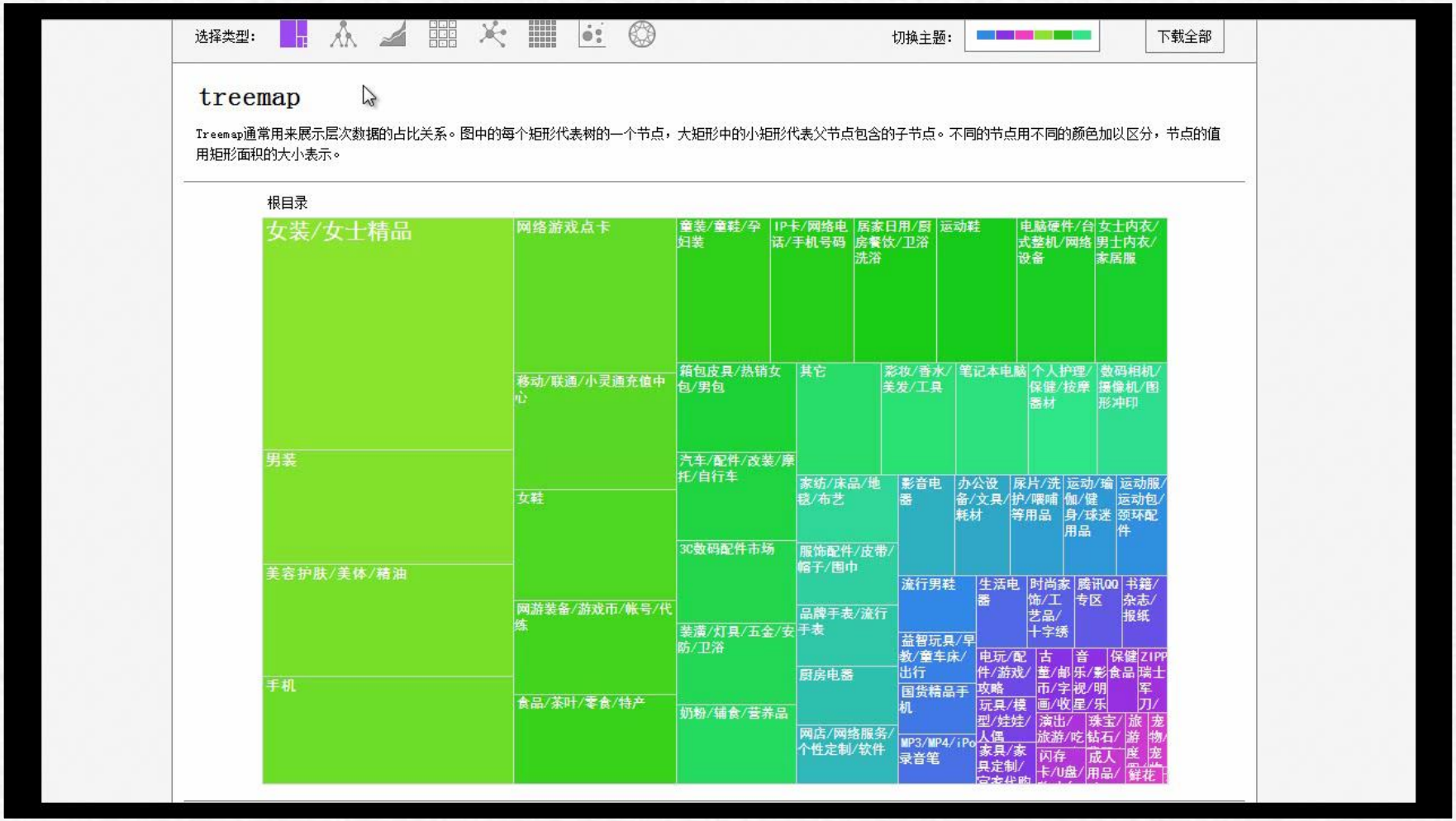
软件

编程



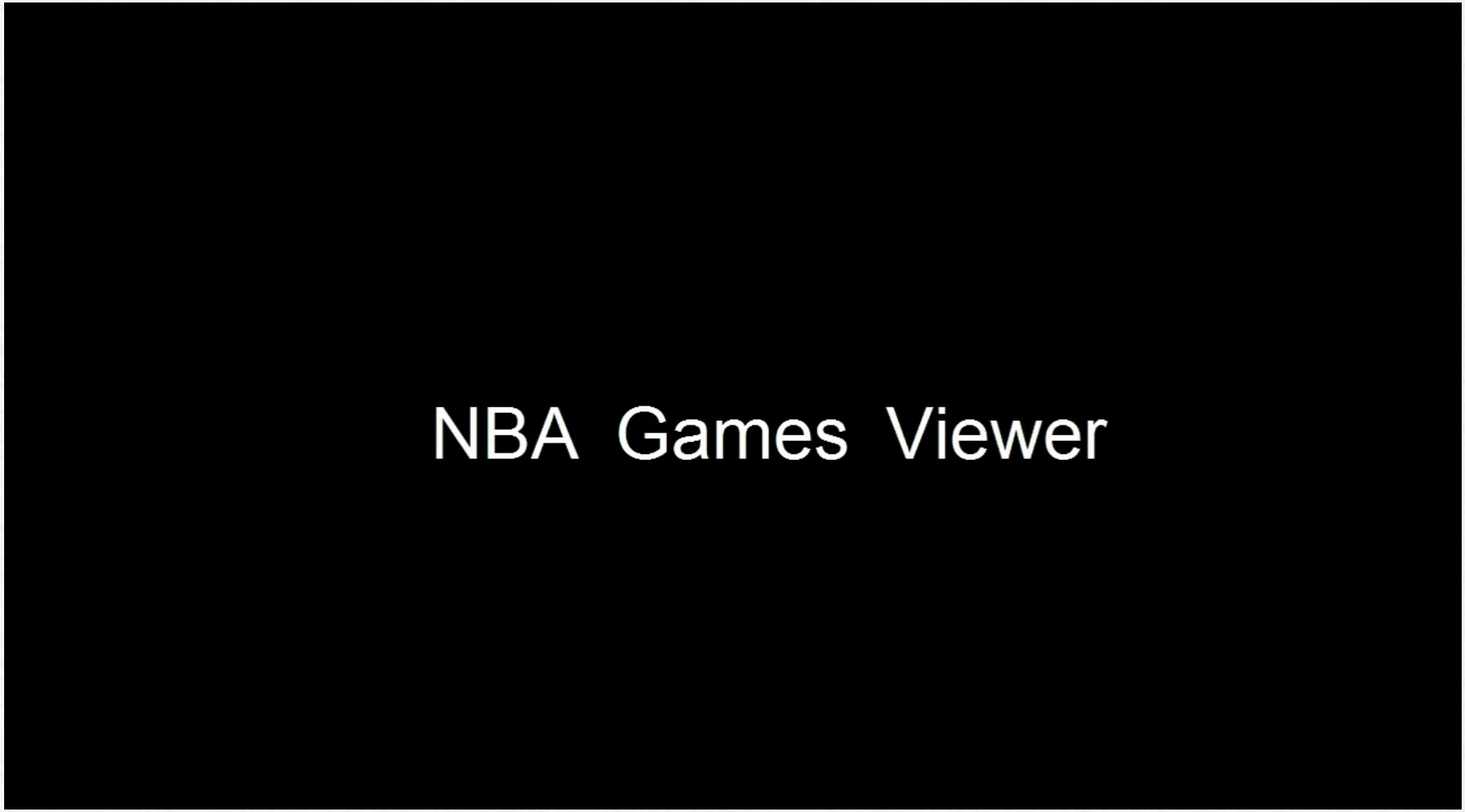
交互开发环境

可视化组件库



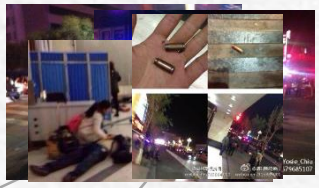
DataV 1.0（阿里巴巴, 浙江大学）

NBA比赛进程可视化

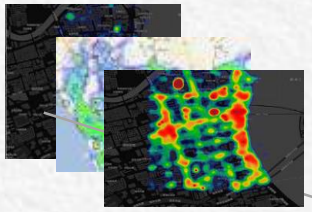


真实应用场景下的态势感知与临场决策（沉浸式）环境

社交数据



网络影像



手机数据

情报数据

新闻报导

监控视频

交通数据

- 信息时空断裂
- 数据稀疏不关联不互补
- 低效的应急分析预测能力
- 信息呈现不到位不直观

真实应用场景下的态势感知与临场决策（沉浸式）环境

大屏拼接沉浸式环境（浙江大学）

陈为
chenwei@cad.zju.edu.cn
浙江大学CAD&CG国家重点实验室