

# 谁为可视化买单？

—从用户视角看可视化与可视分析研究

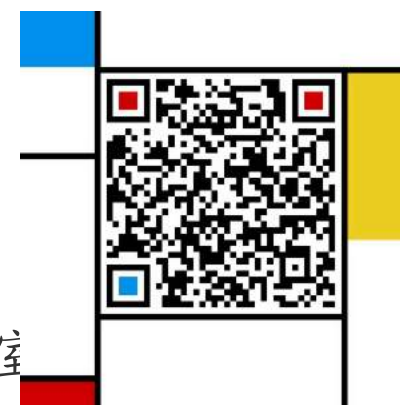
袁晓如

信息科学中心

机器感知与智能教育部重点实验室

北京大学 信息科学技术学院

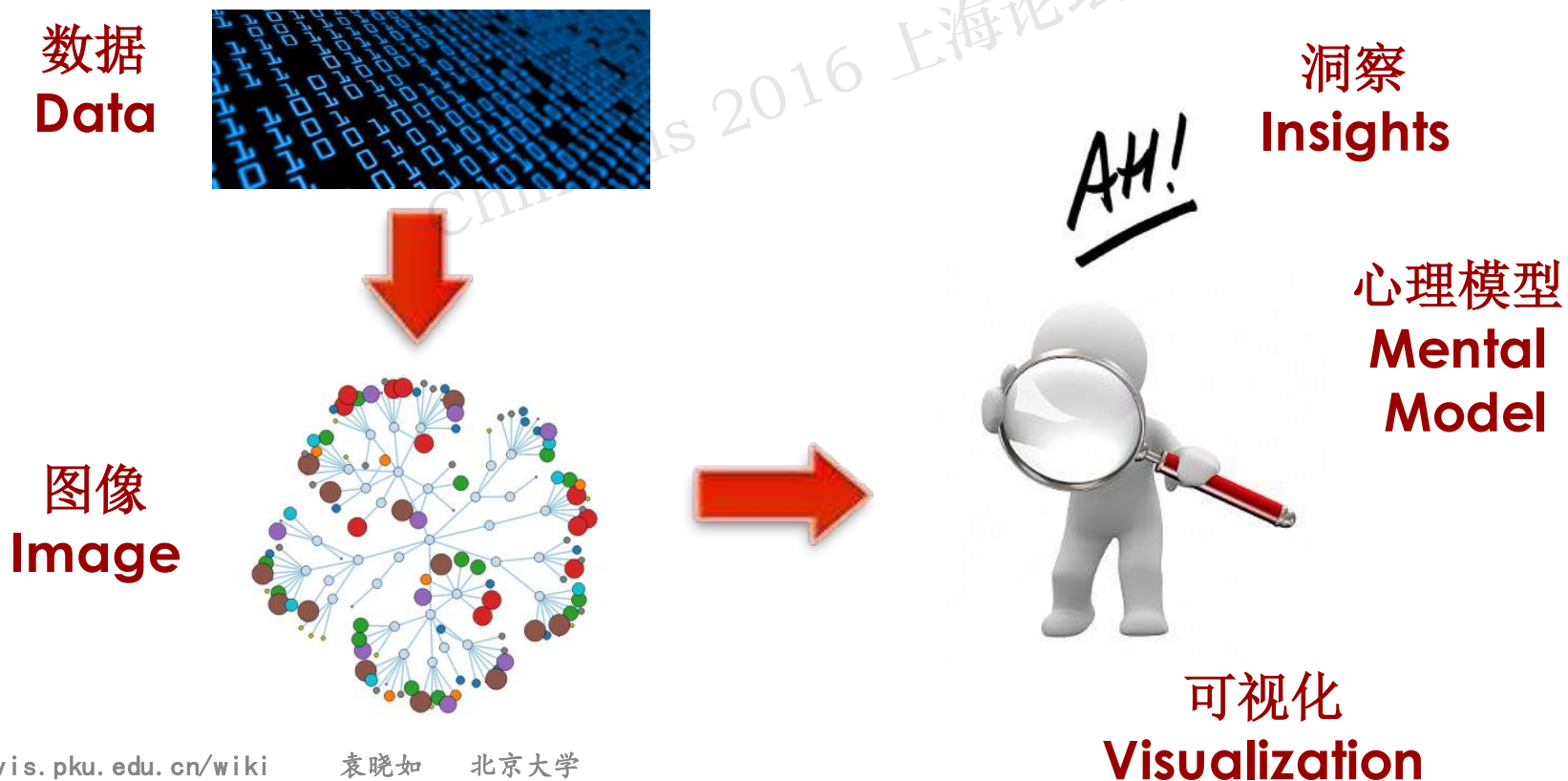
2016.06.05



@晓如微博

# 可视化/可视分析

- 可视化(Visualization) 对事物建立心理模型(mental model)或者心理图像(mental image)

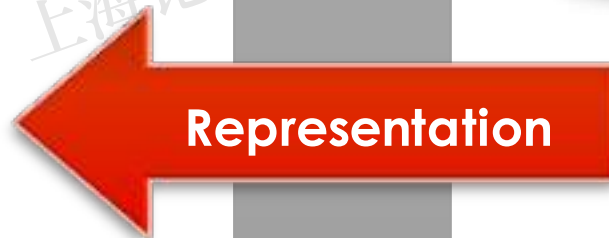


# 可视化双鸿沟模型 Double Gulf



Visualization User

Visualization Designer



Conceptual Model

Visualization

Data

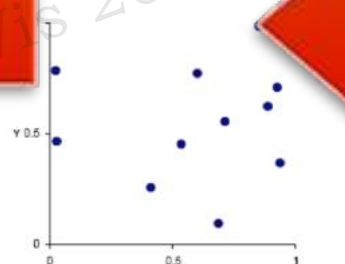
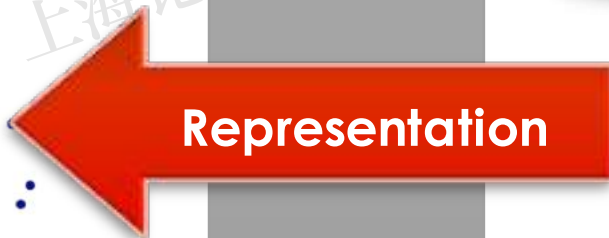


# 可视化双鸿沟模型 Double Gulf



Visualization User

Visualization Designer



X	Y
0.67	0.79
0.32	0.63
0.39	0.72
0.27	0.85
0.71	0.43
0.63	0.09
0.03	0.03
0.20	0.54
0.51	0.38
0.11	0.33
0.46	0.46

Conceptual Model

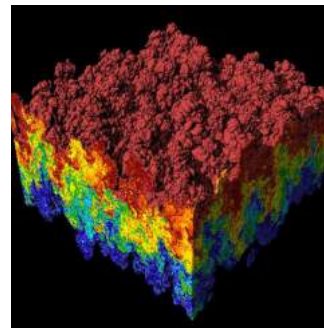
Visualization

Data



# 可视化为谁服务？

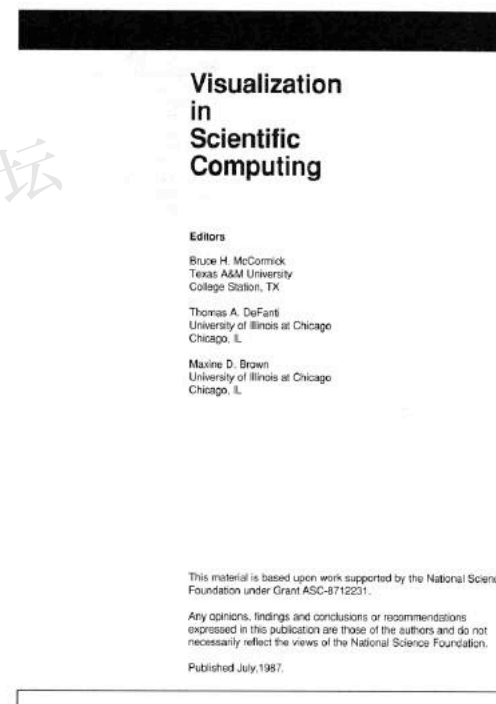




# 1987 美国NSF会议

- 1987 report by the National Science Foundation's Advisory Panel on Graphics, Image Processing, and Workstations

## “visualization in scientific computing”



“Visualization in Scientific Computing,” B.H. McCormick, T.A. DeFanti, and M.D. Brown, eds., Computer Graphics, Vol. 21, No. 6, Nov. 1987.

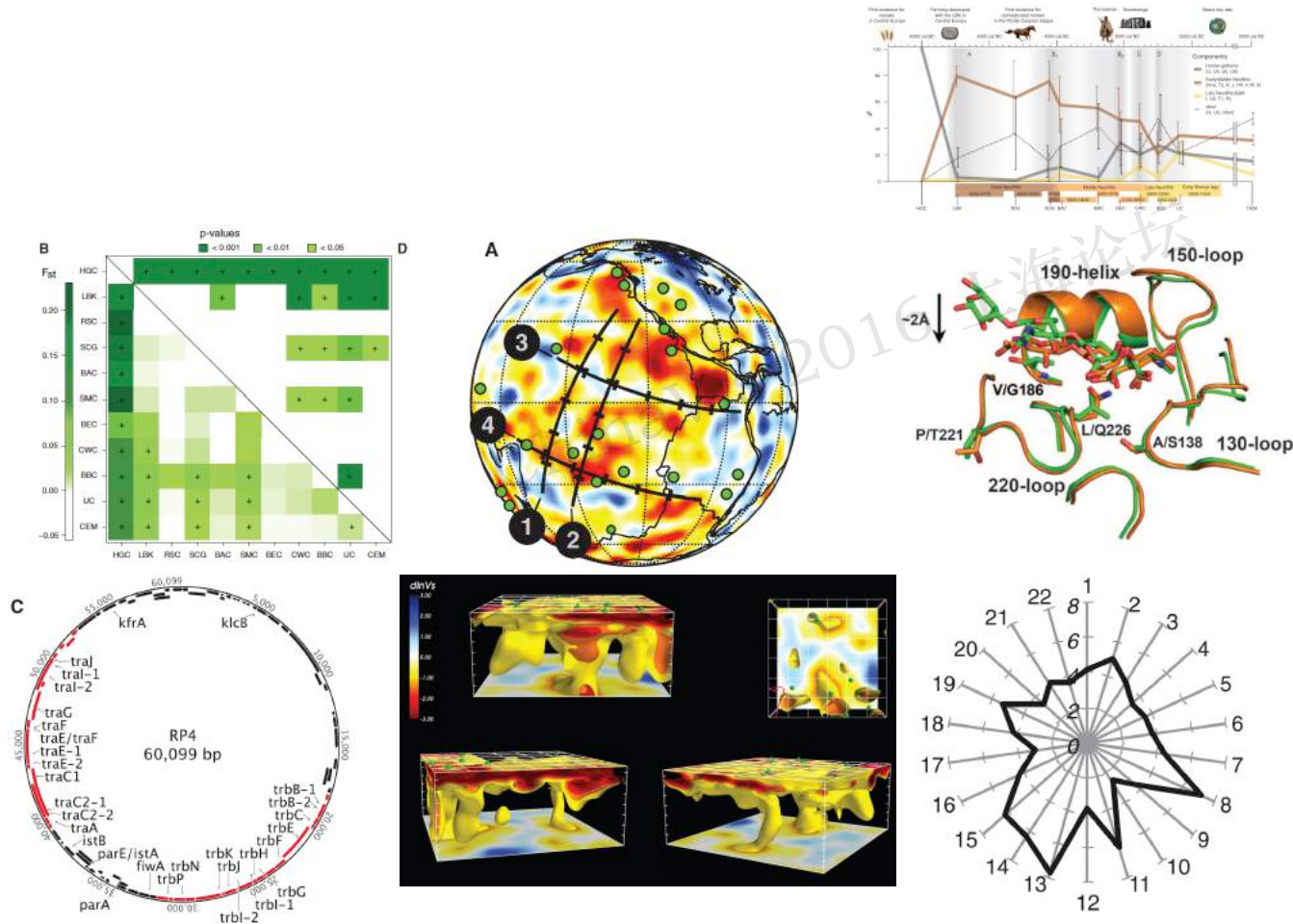


# 科学家



- 需求 / 层次与领域相关
- 需求:
  - 提供基本的数据展示以及基本的操作
  - 发掘隐含在三维空间与非空间属性之间的关系
  - 研究数学模型，比较模型间差异
  - 提出并测验假说

# 《科学》中的可视化



2013年10月11日

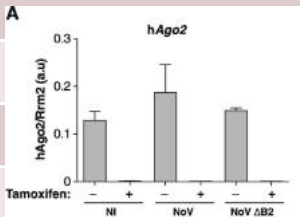


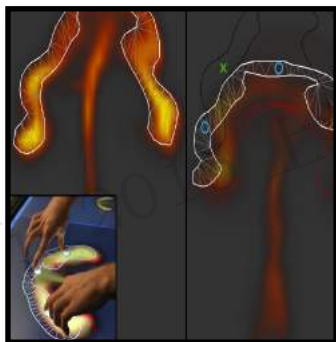
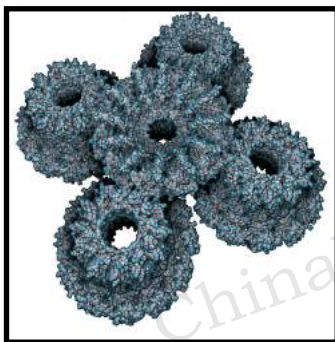
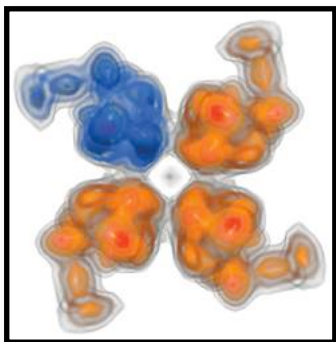
# 《科学》中的可视化

	序号	插图数	可视化形式
Article	1	6	Bar Char, Image, Graph
	2	6	Image, 2D Illustration, 3D Illustration, 2D Plot
Report	1	1	Bart Chart
	2	4	Illustration, Scatter Plot, Line Chart
	3	4	3D Illustration, 2D Illustration, 3D Geometry, Trajectory Drawing, Image
	4	4	Volume Rendering, Iso-contour, 3D surface, Heart Map, Isosurface, 2D plot
	5	4	Bar Chart, Plot, Illustration, Image, Multivariate Bar Chart
	6	4	Bar Chart, Star Glyph, Gene Bar
	7	4	Bar Chart, Pie Chart, Image, Scatter Plot
	8	4	3D structure
	9	2	Scatter Plot
	10	4	Bar Chart, Circular Plot, Image
	11	4	Bar Chart, Plot, Star Glyph, Small multiples
	12	4	Map, Matrix, Tree

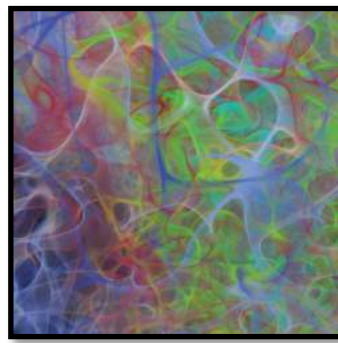
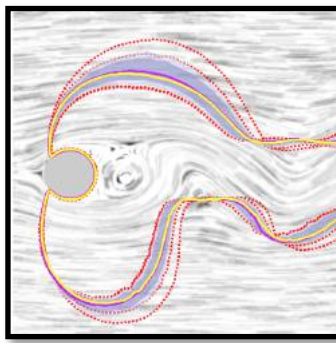
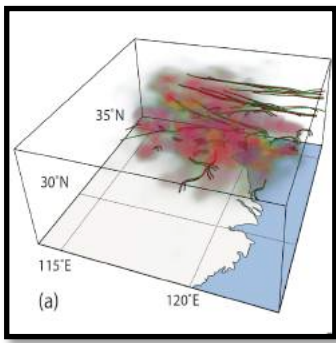
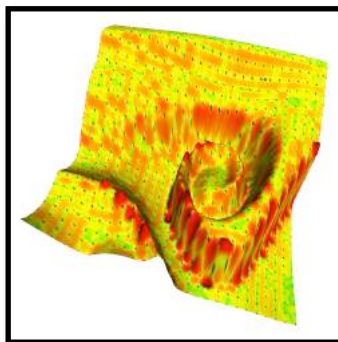


2013年10月11日





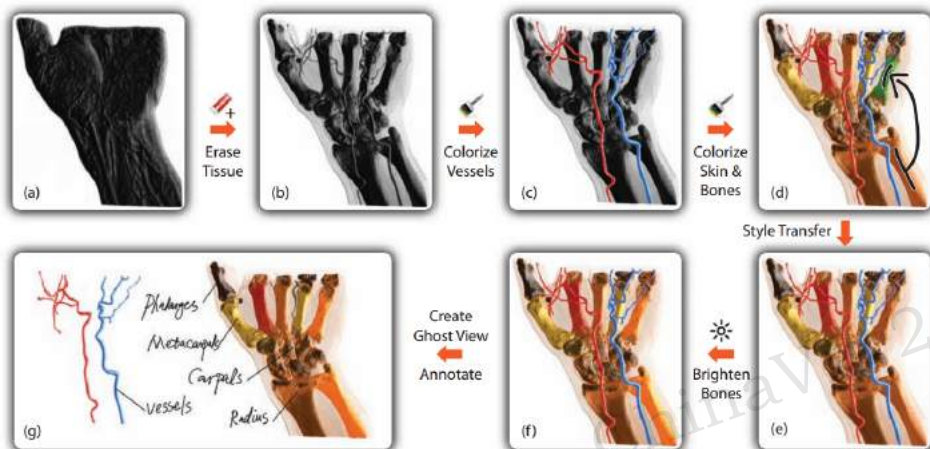
User	Serial Name	Quantity	Category	Year	Country	University
1	University of California, Berkeley	1	USA	2008	USA	University of California, Berkeley
2	University of California, Berkeley	1	USA	2008	USA	University of California, Berkeley
3	University of California, Berkeley	1	USA	2008	USA	University of California, Berkeley
4	University of California, Berkeley	1	USA	2008	USA	University of California, Berkeley
5	University of California, Berkeley	1	USA	2008	USA	University of California, Berkeley
6	University of California, Berkeley	1	USA	2008	USA	University of California, Berkeley
7	University of California, Berkeley	1	USA	2008	USA	University of California, Berkeley
8	University of California, Berkeley	1	USA	2008	USA	University of California, Berkeley
9	University of California, Berkeley	1	USA	2008	USA	University of California, Berkeley
10	University of California, Berkeley	1	USA	2008	USA	University of California, Berkeley







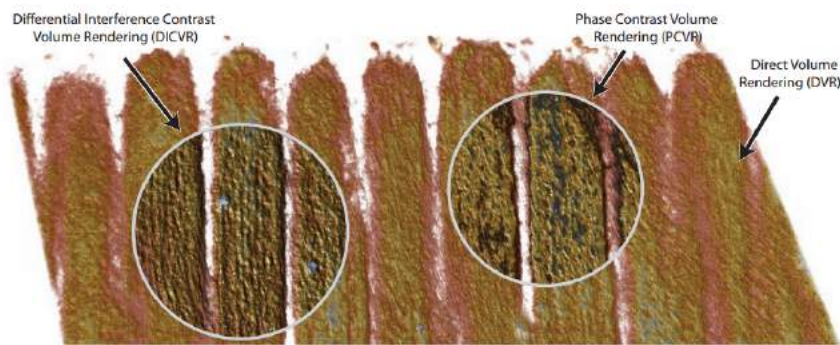
# 科学家：观察理解数据



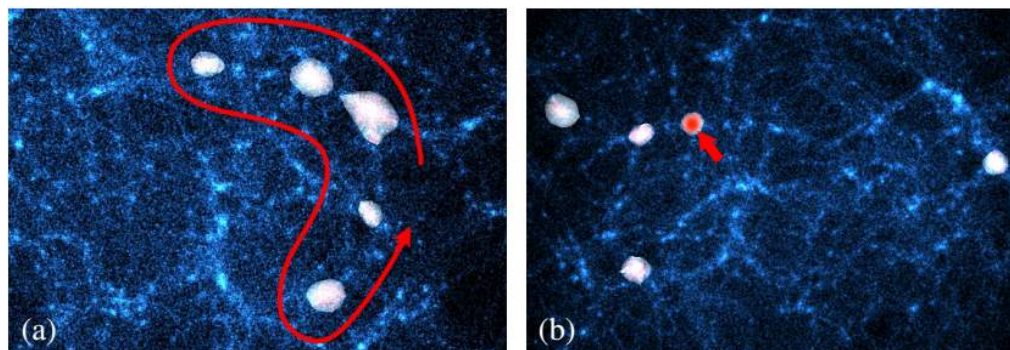
Guo et al. 2011.



NASA Scientific Visualization Studio. 2012.



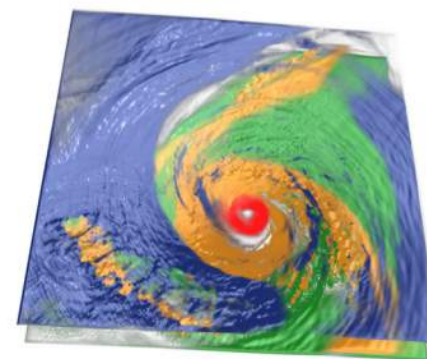
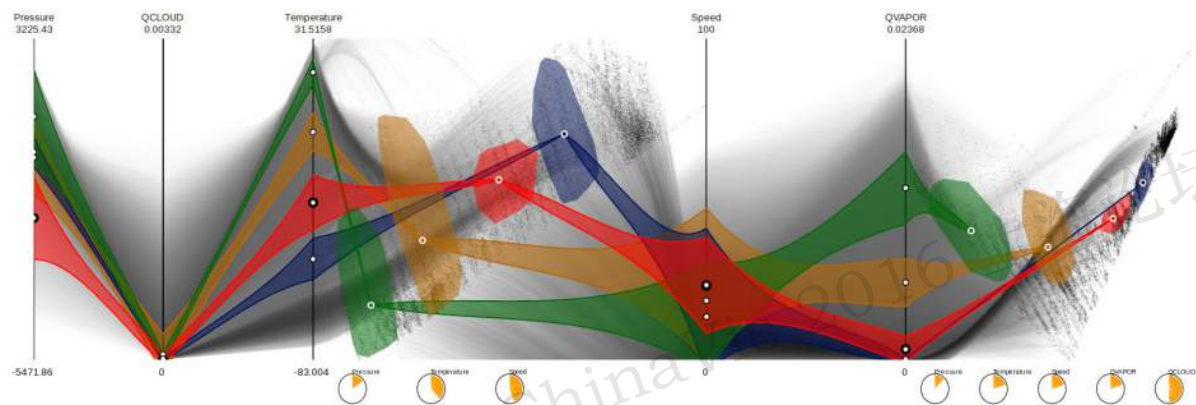
Guo et al. 2012.



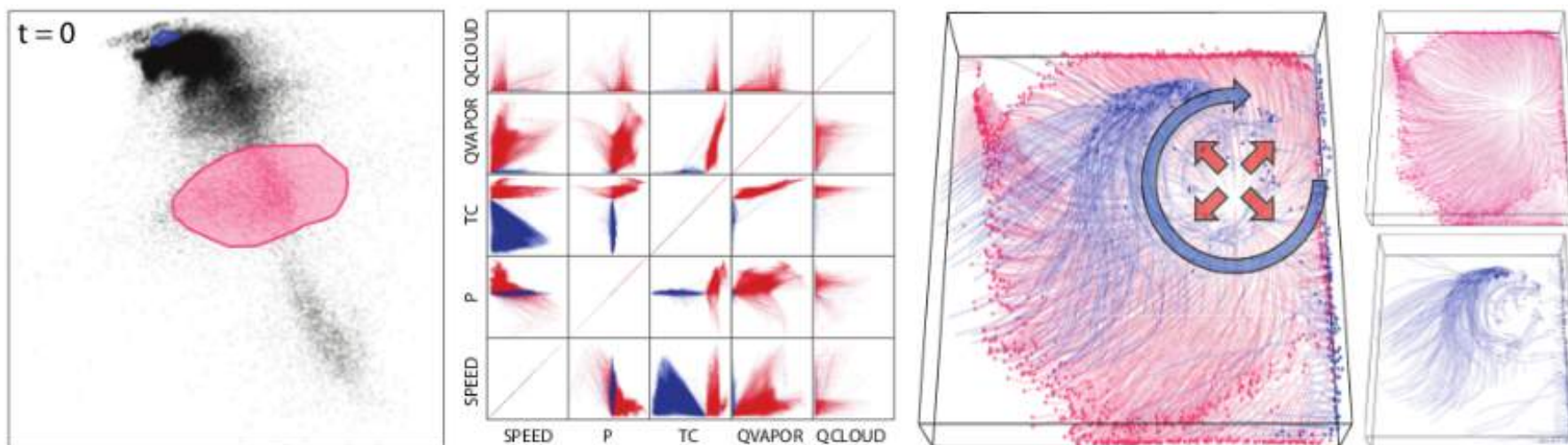
Yu et al. 2011.



# 科学家：关联非空间属性信息



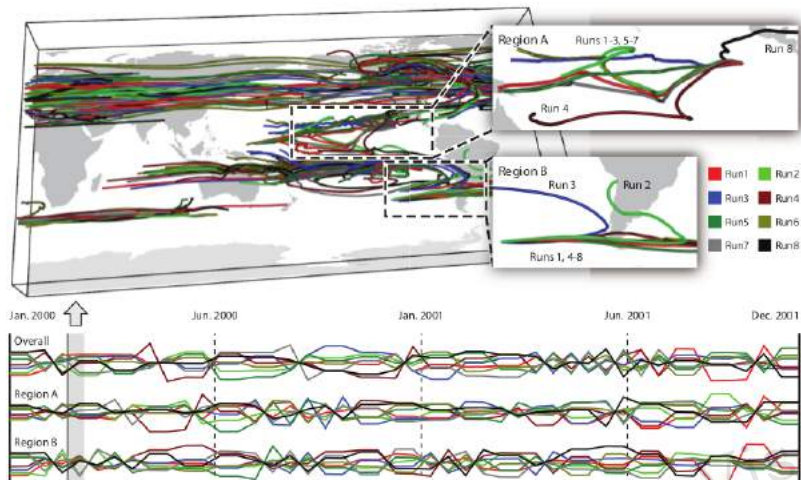
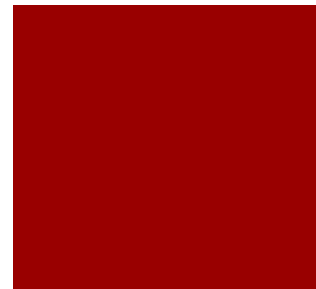
Guo et al. 2012.



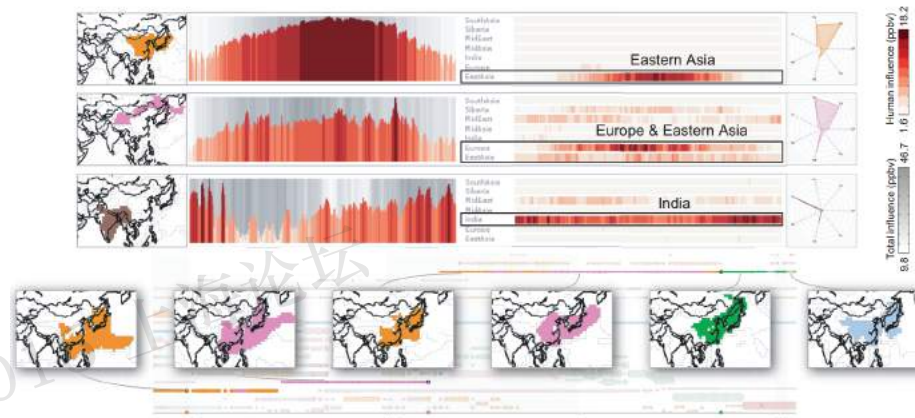
Guo et al. 2014.



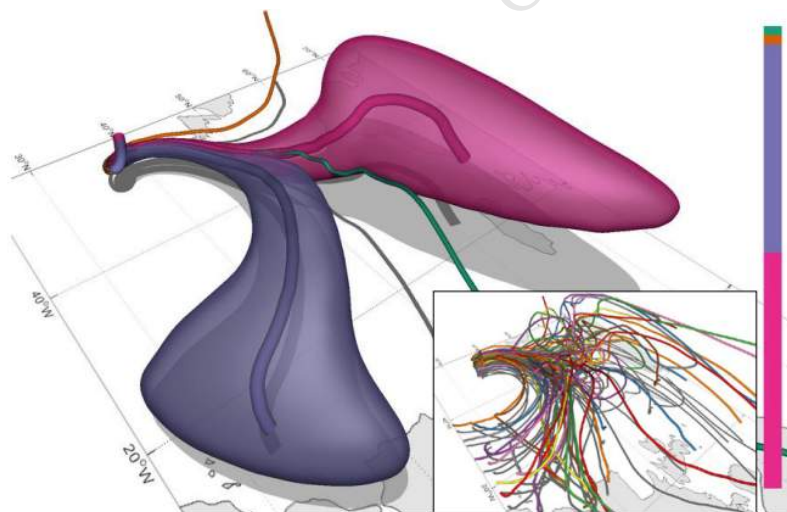
# 科学家：研究比较模型



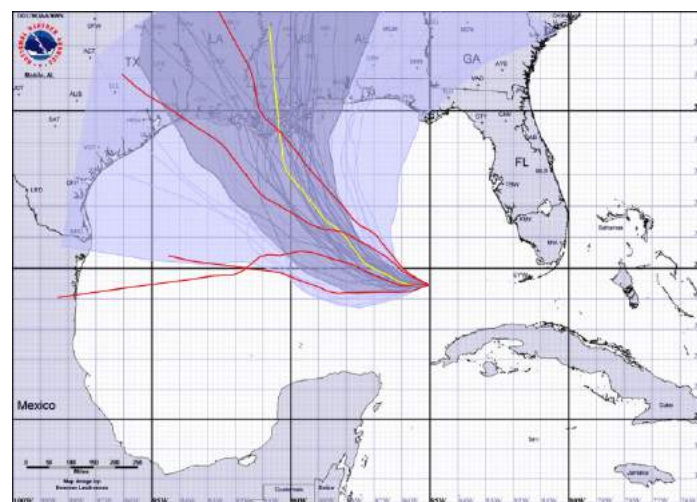
Guo et al. 2013.



Shu et al. 2016.



Ferstl et al. 2016.



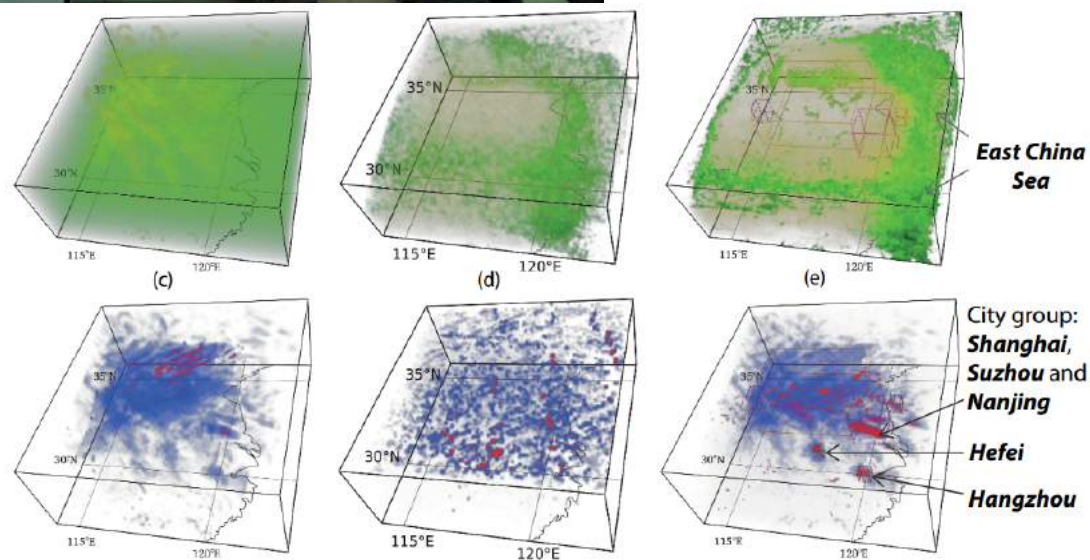
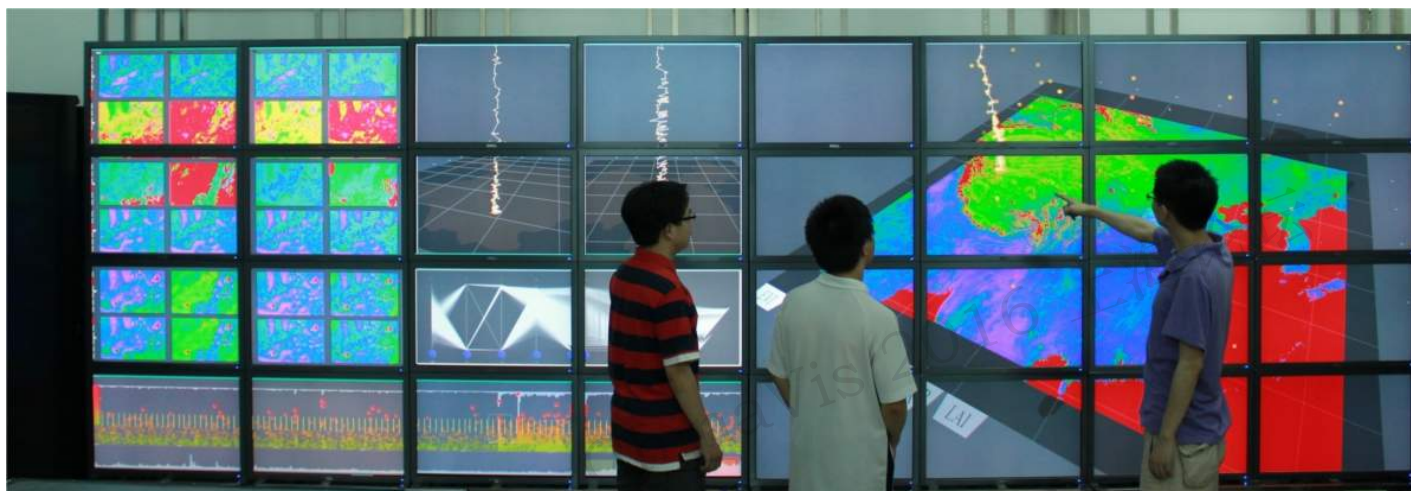
Mirzargar et al. 2014.





# 科学家：提出并测验假说

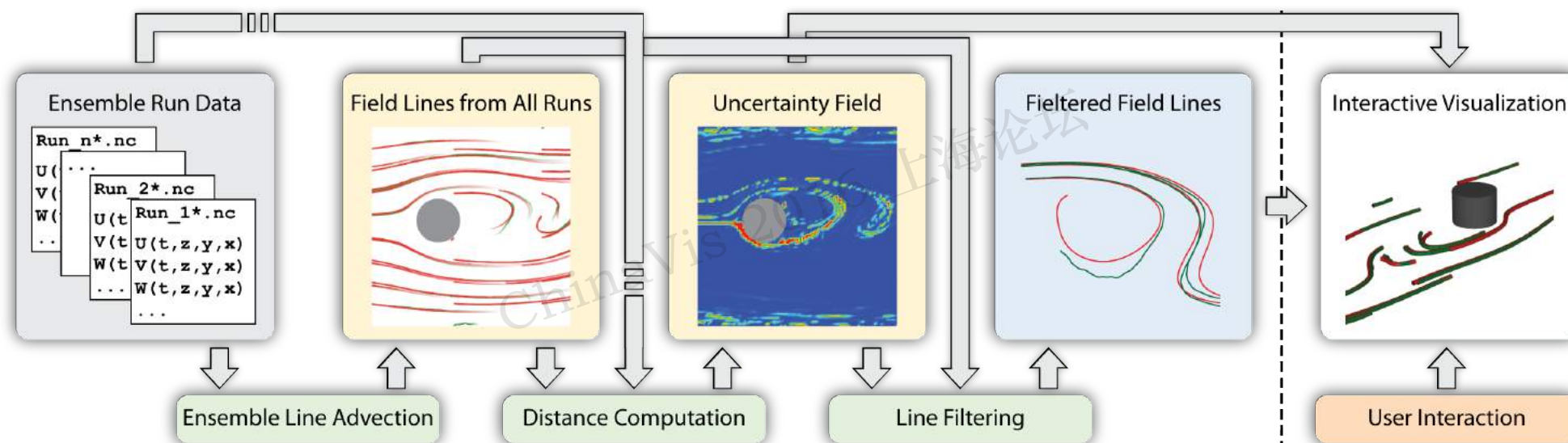
Yuan et al. 2013.



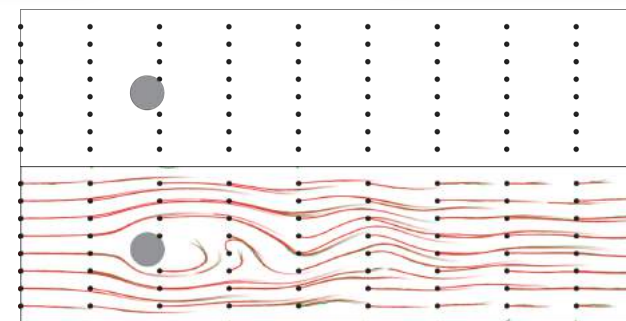
Liu et al. 2016.

# 集合模拟流场分析

## Couple Ensemble Flow Line Advection and Analysis (eFLAA)



- Ensemble data (large) **200GB** 输入数据
- Field line data (much larger than) **200TB** 中间结果
- Variation field (small) **1-10GB** 中间结果
- Filtered lines (even smaller) **100-200MB** 可视化结果



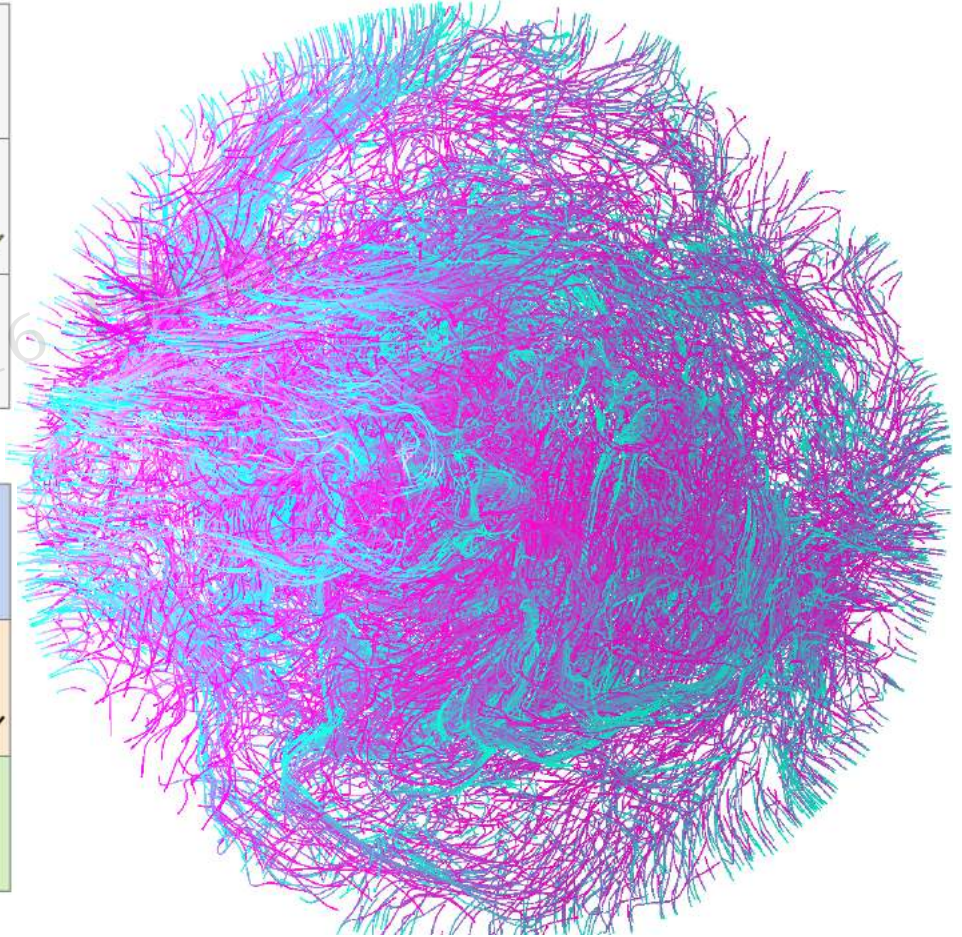
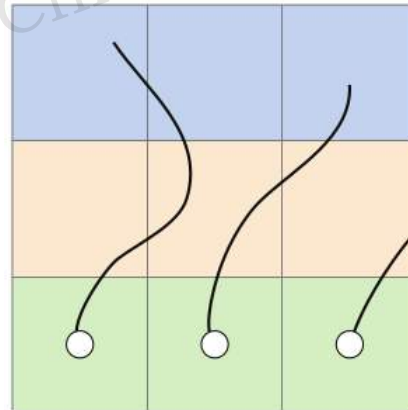
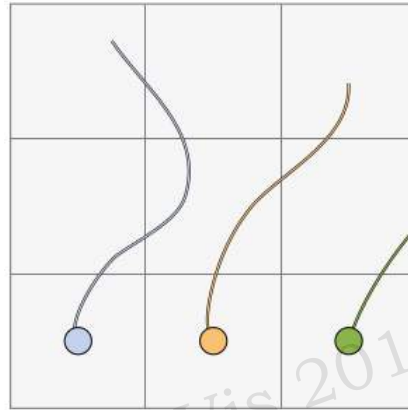
[Guo, Yuan, Huang and Zhu TVCG 2013 (SCIVis '13)]



# 并行粒子追踪问题

## Parallel Particle Tracing Problems

- Task-parallelism
  - Static partitioning over seeds
  - High I/O cost
  - Poor load-balancing
- Data-parallelism
  - Static partitioning over data
  - Requires loading all data
  - Poor load-balancing with naïve partitioning

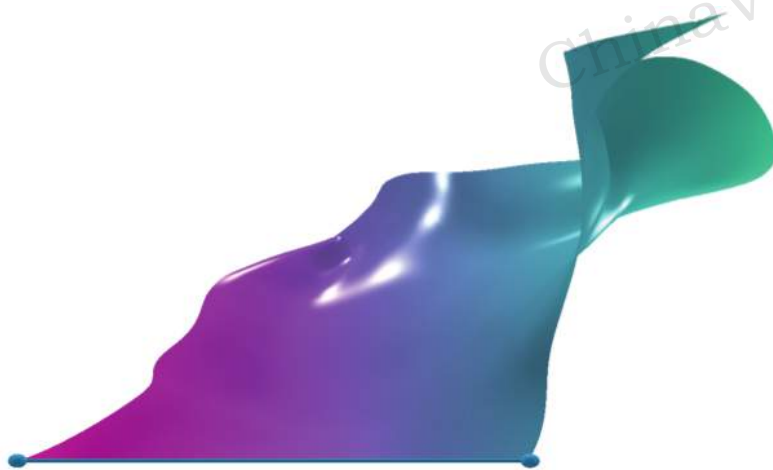


TB-scale turbulence simulation, developed by COE, PKU, ran on TH-1A

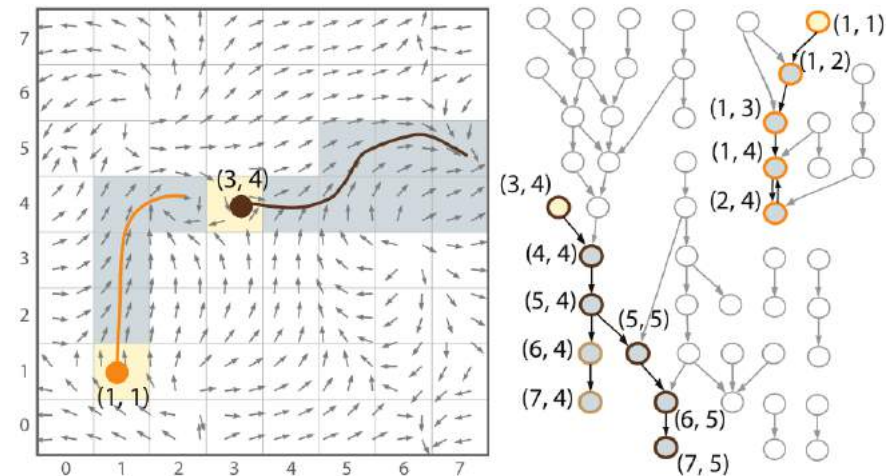
# 并行粒子追踪问题

## Advection-based Sparse Data Management for Visualizing Unsteady Flow

- Observations
  - Data is large, but the working set is usually small
  - Access pattern exists in flow field data
  - Coarse-grained partition is not good for load-balancing



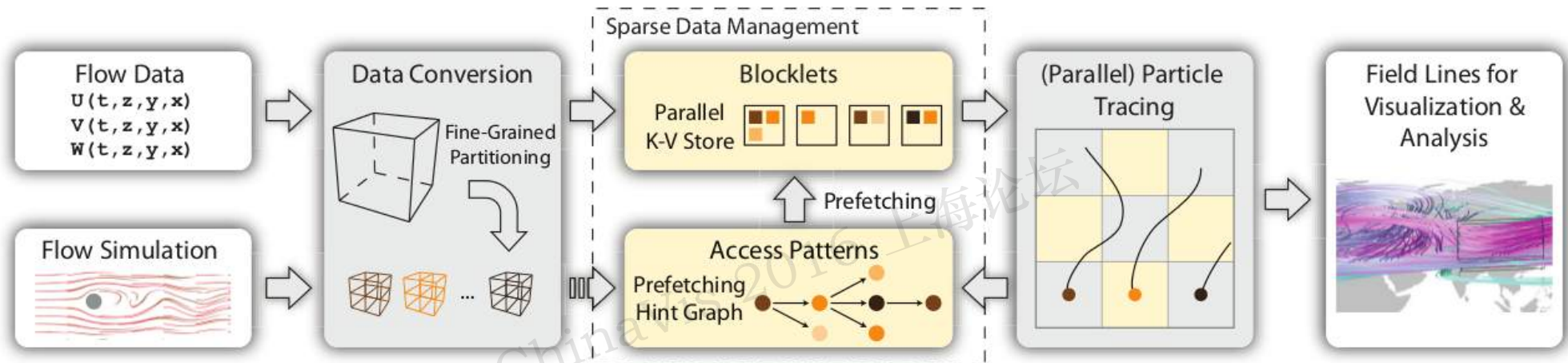
TB-scale turbulence data, only 1% data is actually accessed



Access patterns in flow field

## 并行粒子追踪问题

## Advection-based Sparse Data Management for Visualizing Unsteady Flow



- Our solutions

- On-demand data access with parallel key-value store
- High performance data prefetching
- Fine-grained data partitioning

- Benefits

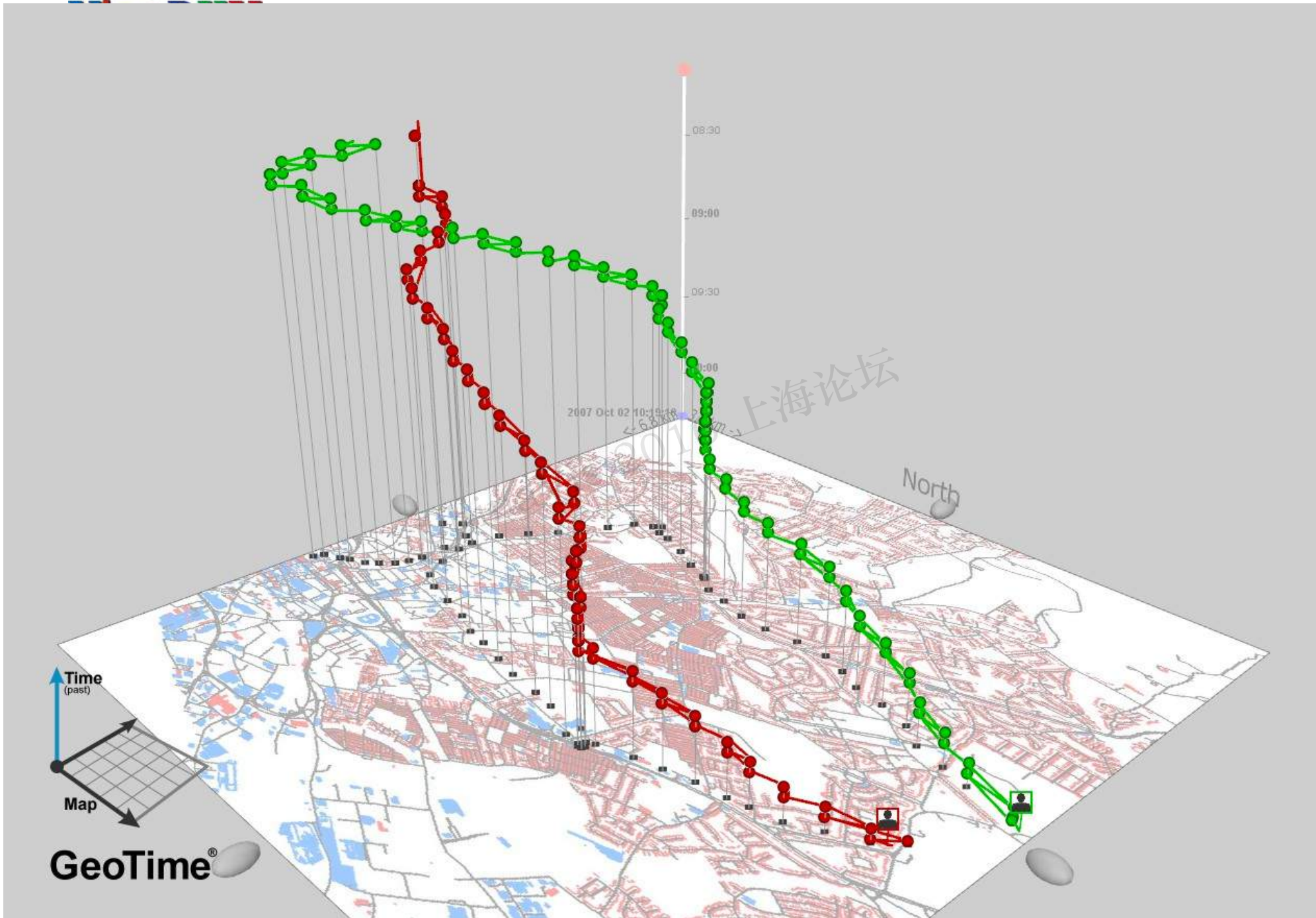
- Increase data analysis scale with any given resource limit
- Increase I/O and memory efficiency, and the scalability of task-parallel particle tracing



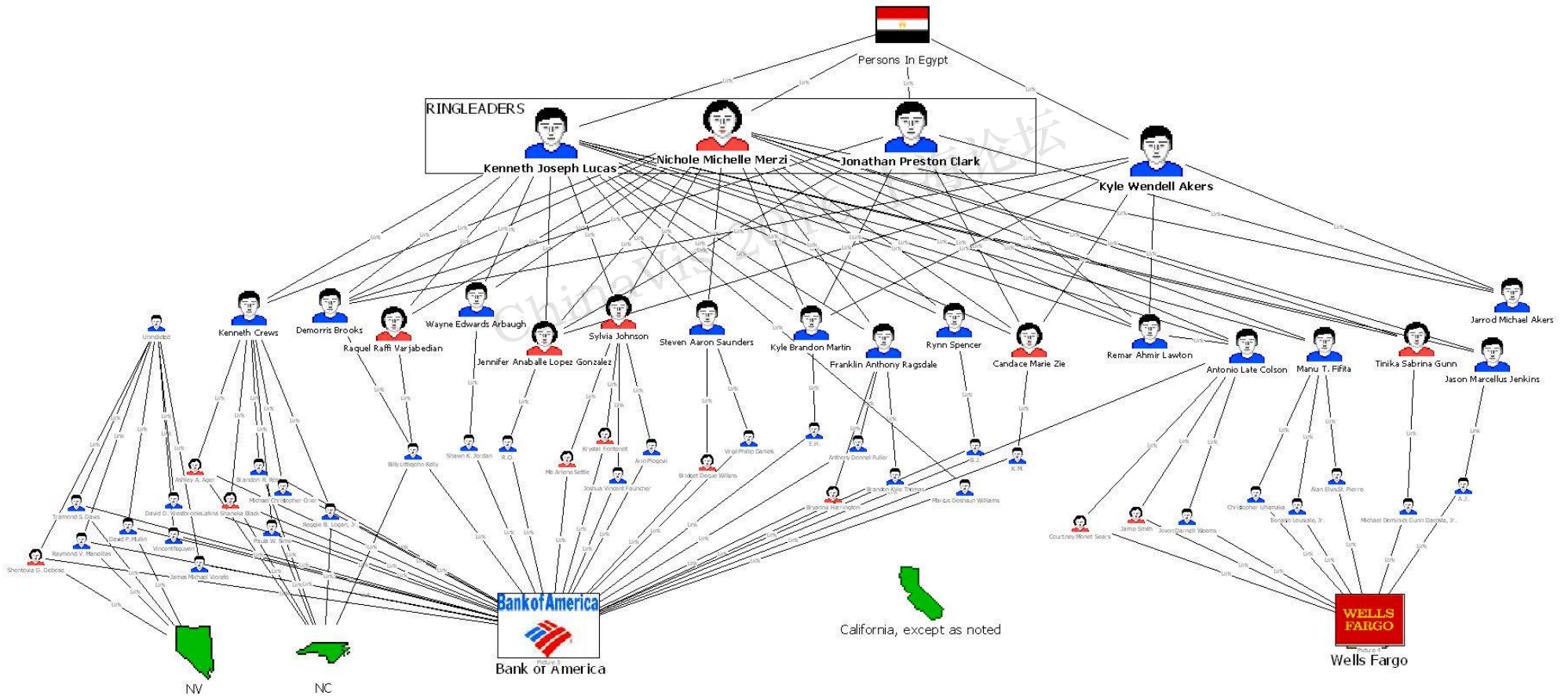
## 军事、安全领域专家

- 特点：对整体态势的理解与把握，破解案情，实时性的要求
- 需求：
  - 复杂、强大的分析功能，满足对时间、空间、关系等一系列维度的探索
  - 强大的关联与数据融合功能，通过对不同数据的整合理解，描绘出案情真相
  - 对细节的深入分析与把握
- 挑战
  - 异构数据分析，案情扑朔迷离，需要多重关联
  - 关键点隐藏在隐蔽的地方，需要深入的理解与挖掘



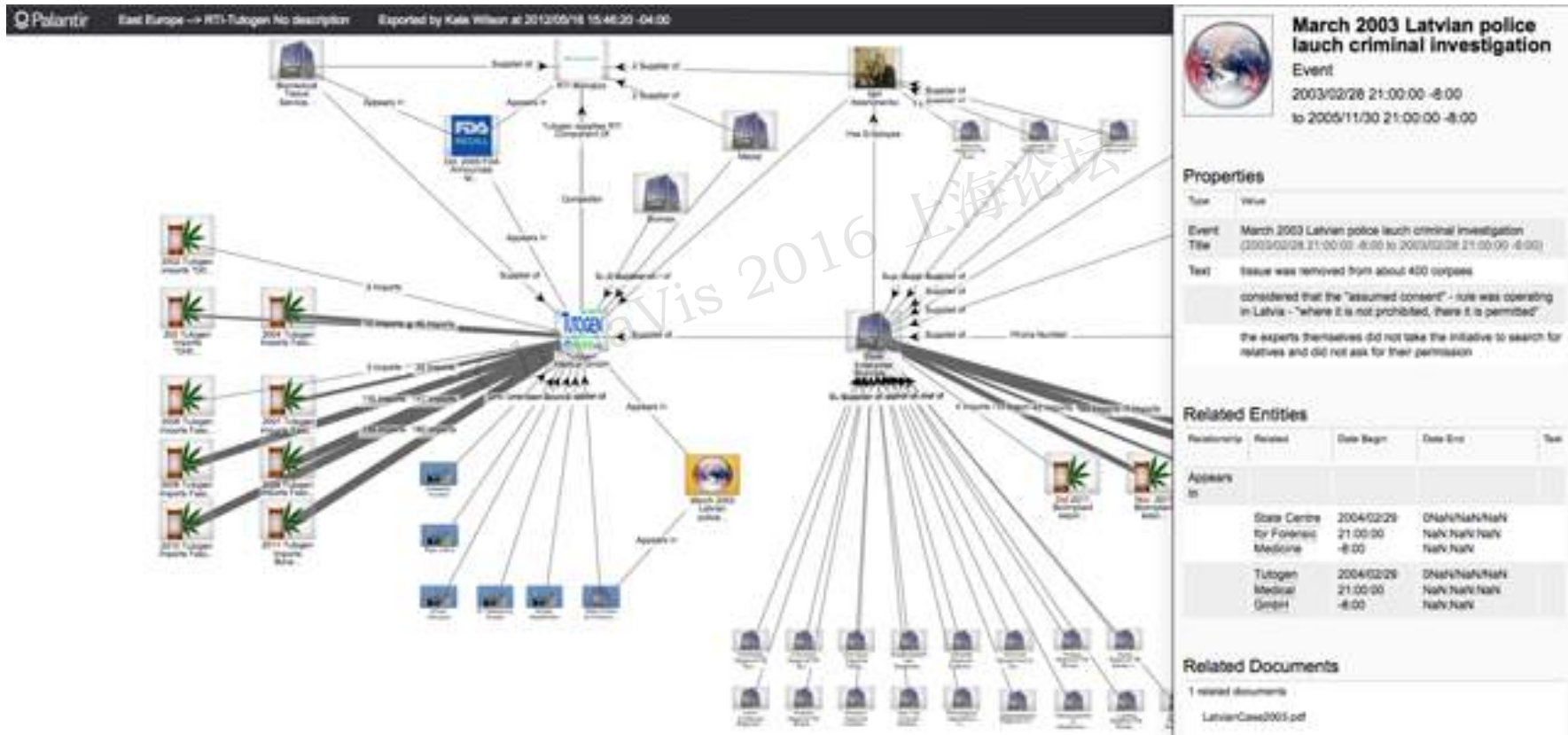


I2





# Palantir



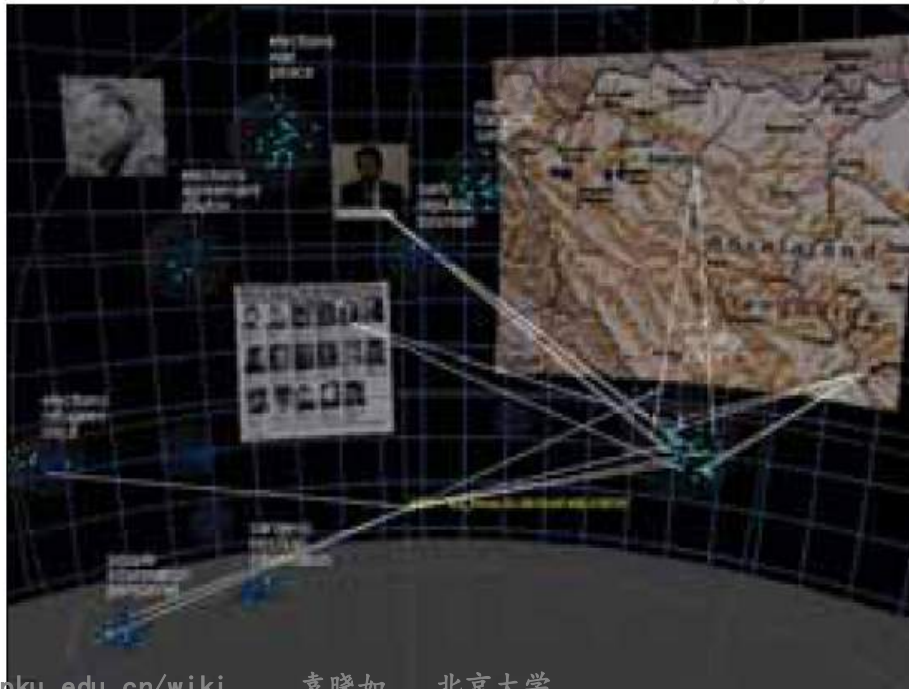
# OpenDNS





# Starlight Information Visualization System

*Starlight* is an experimental information visualization system currently under development at the Pacific Northwest National Laboratory for multiple U.S. Government clients. The *Starlight* software automatically organizes, characterizes, and integrates a variety of structured and unstructured information types, then generates easily interpretable, graphical representations of relationships among the data. This approach enables exciting and powerful new forms of information access, exploitation, and control.



## Key Features:

- Visualization-oriented GUI
- Advanced information model
- Sophisticated query tools
- Information extraction tools
- Integrated GIS
- XML-based
- Windows NT/2000/XP platform

## Principle Benefits:

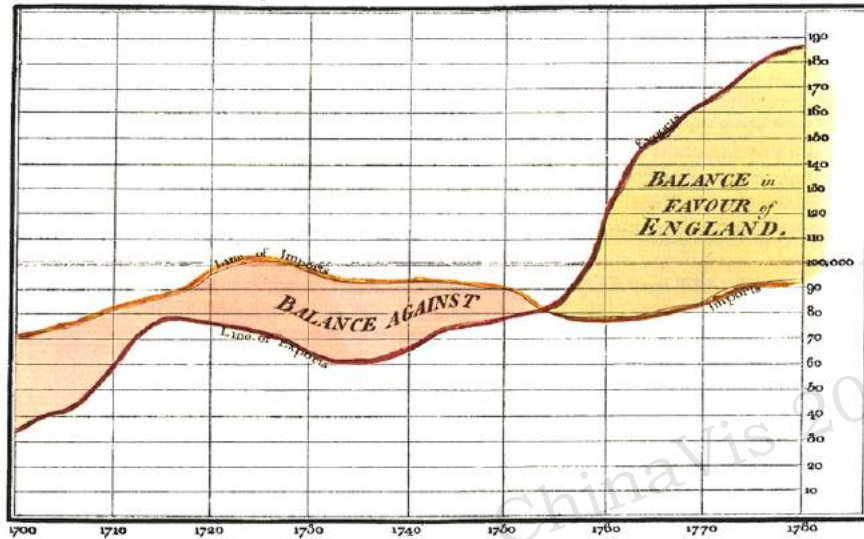
- Information Integration
- Complexity Management
- Holistic Analysis
- Workflow Continuity
- Accelerated Interpretation
- Improved Understanding

# 普通大众 - 非专业性可视化

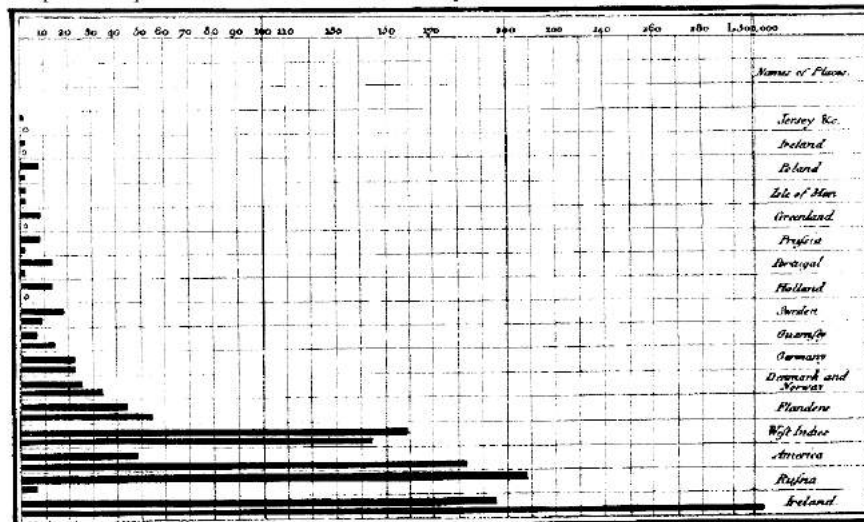
- 意义与目的：构建数据与普通用户的桥梁
- 展览 - 可视化是科普的重要工具
- 数据分析的兴趣爱好者 - 却缺少相应的数据分析工具
- 数据拥有者 - 社交媒体用户等

# 1786: William Playfair

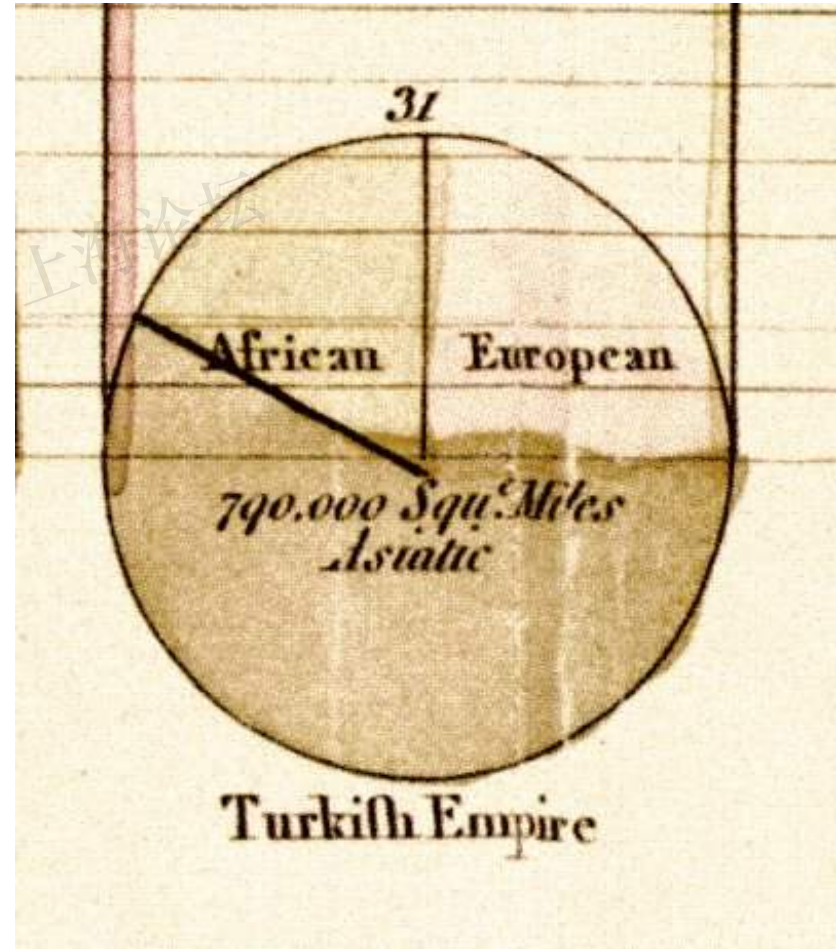
Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780



Exports and Imports of SCOTLAND to and from different parts for one Year from Christmas 1780 to Christmas 1781

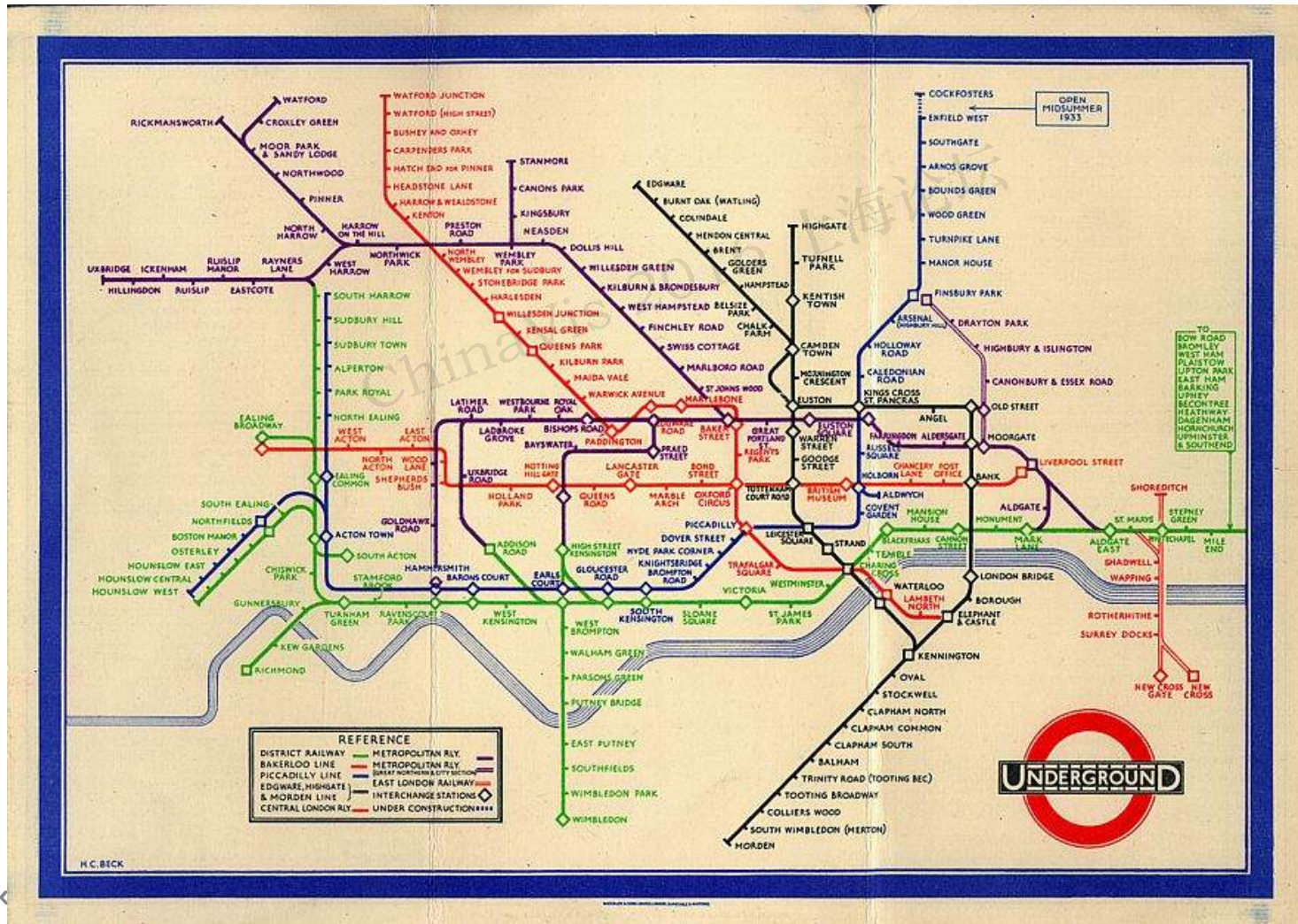
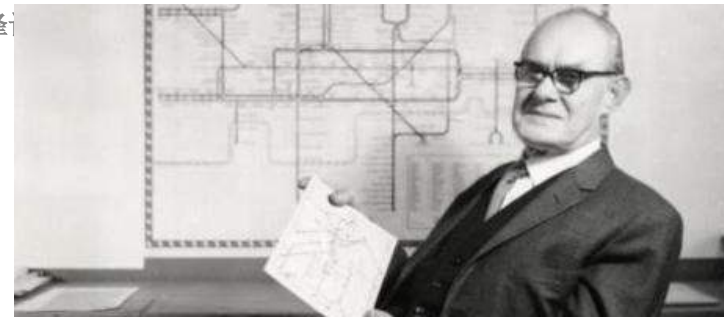


The Upright divisions are Ten Thousand Pounds each. The Black Lines are Exports the Ribbed Lines Imports.  
 Published in the Edinburgh Journal of 1781 by W. Playfair.  
 Math. comp. 1781. Printed in London.



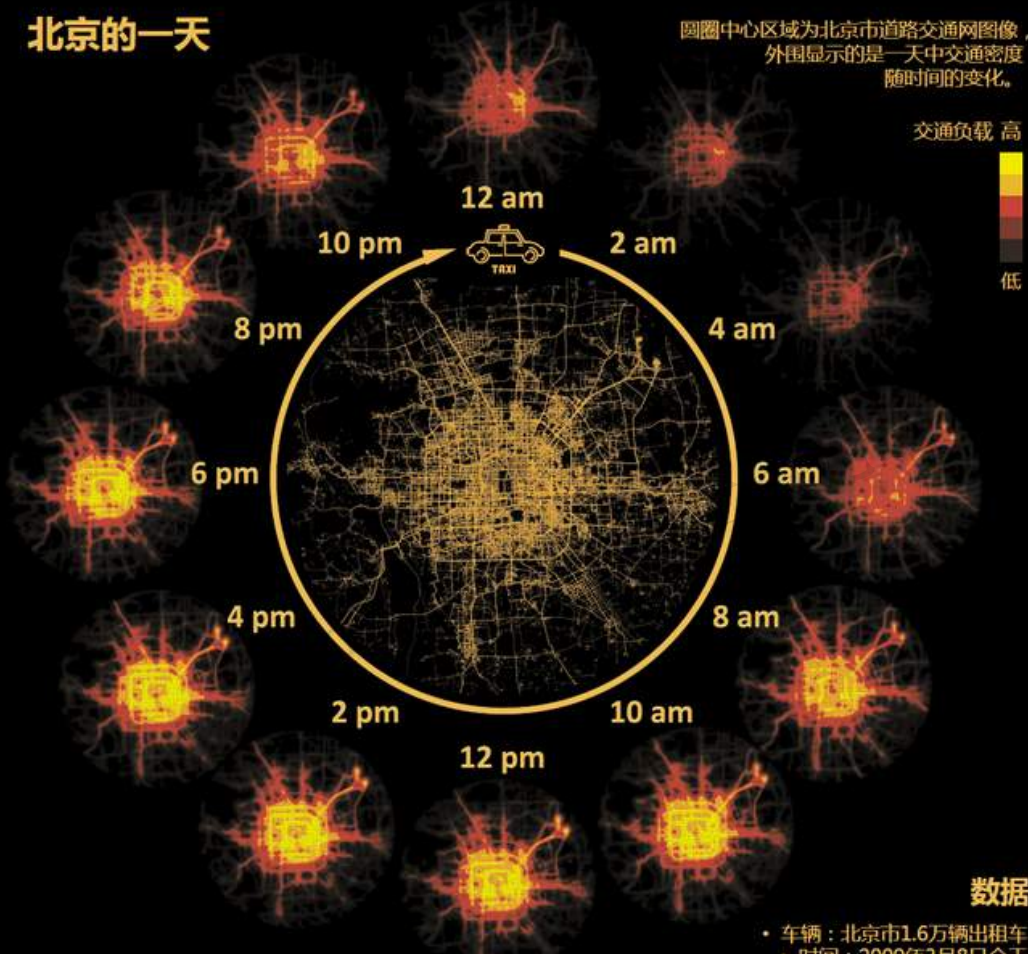


# 1933 伦敦地铁图

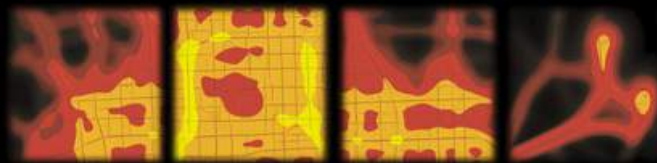


## 北京的一天

圆圈中心区域为北京市道路交通网图像，  
外围显示的是一天中交通密度  
随时间的变化。



## 北京各区域



中关村/北京大学

故宫/长安街

奥林匹克公园

首都机场

## 数据

- 车辆：北京市1.6万辆出租车
- 时间：2009年3月8日全天
  - 累计距离：176万公里
  - 累计时间：2700小时
  - 测量点数目：1600万
  - 数据来源：数据堂

## 联系方式

邮箱：xiaoru.yuan@pku.edu.cn

网站：http://vis.pku.edu.cn

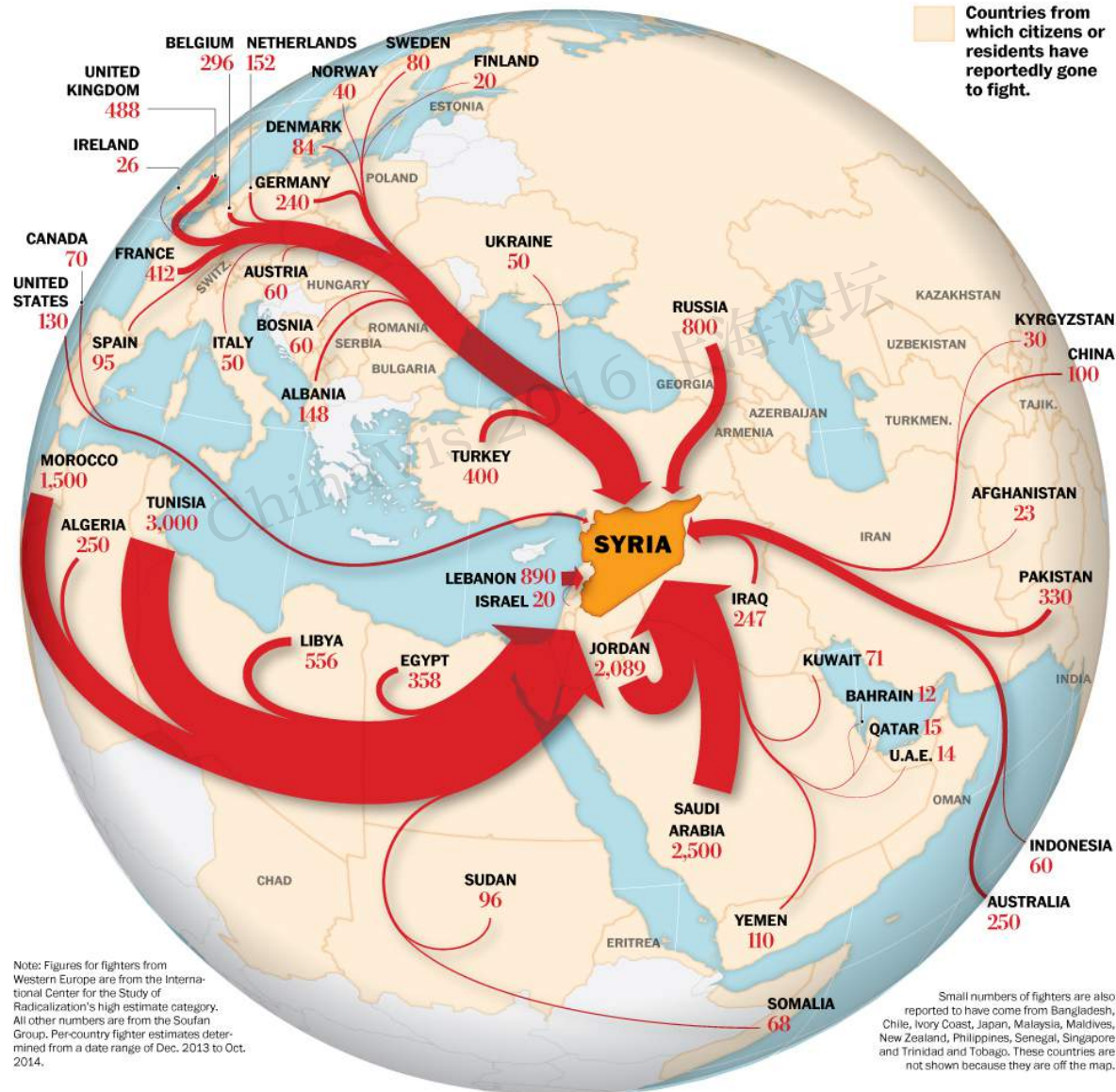


# 记者与媒体

- 需要用可视化来讲一个完整的故事
- 需求：
  - 直观易懂
  - 时效性高
  - 简单可交互
  - 讲述一个完整的故事，背后隐含着新闻的立意
- 挑战：
  - 如何设计合理的讲故事的流程
  - 实时性、快速构建一个数据新闻可视化系统



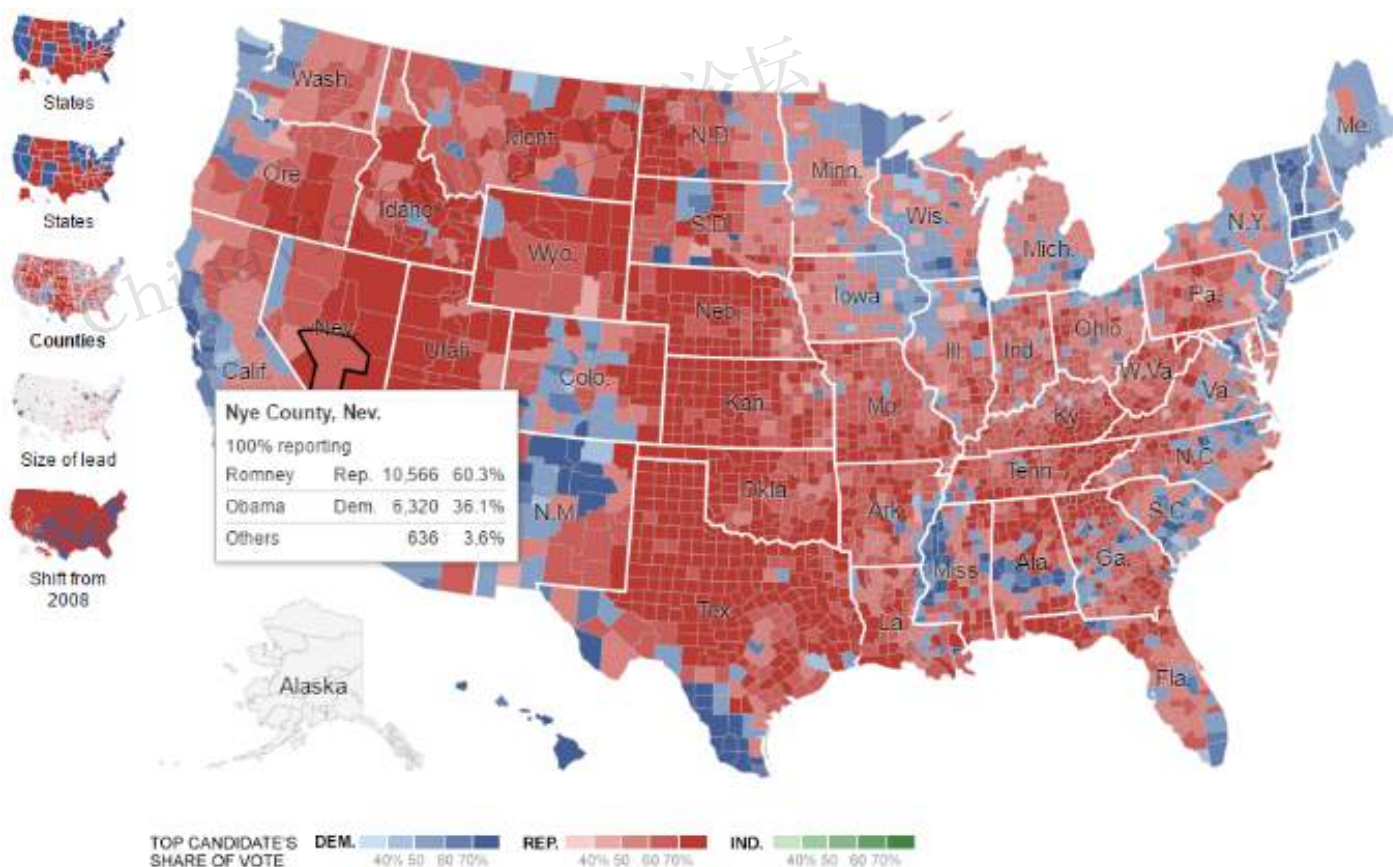
# 数据新闻 Foreign Fighters Flow to Syria





# 记者与媒体：数据新闻

- 政治形势可视化
  - 2012美国大选  
两党在各个地区的  
选票情况
  - [在线演示](#)

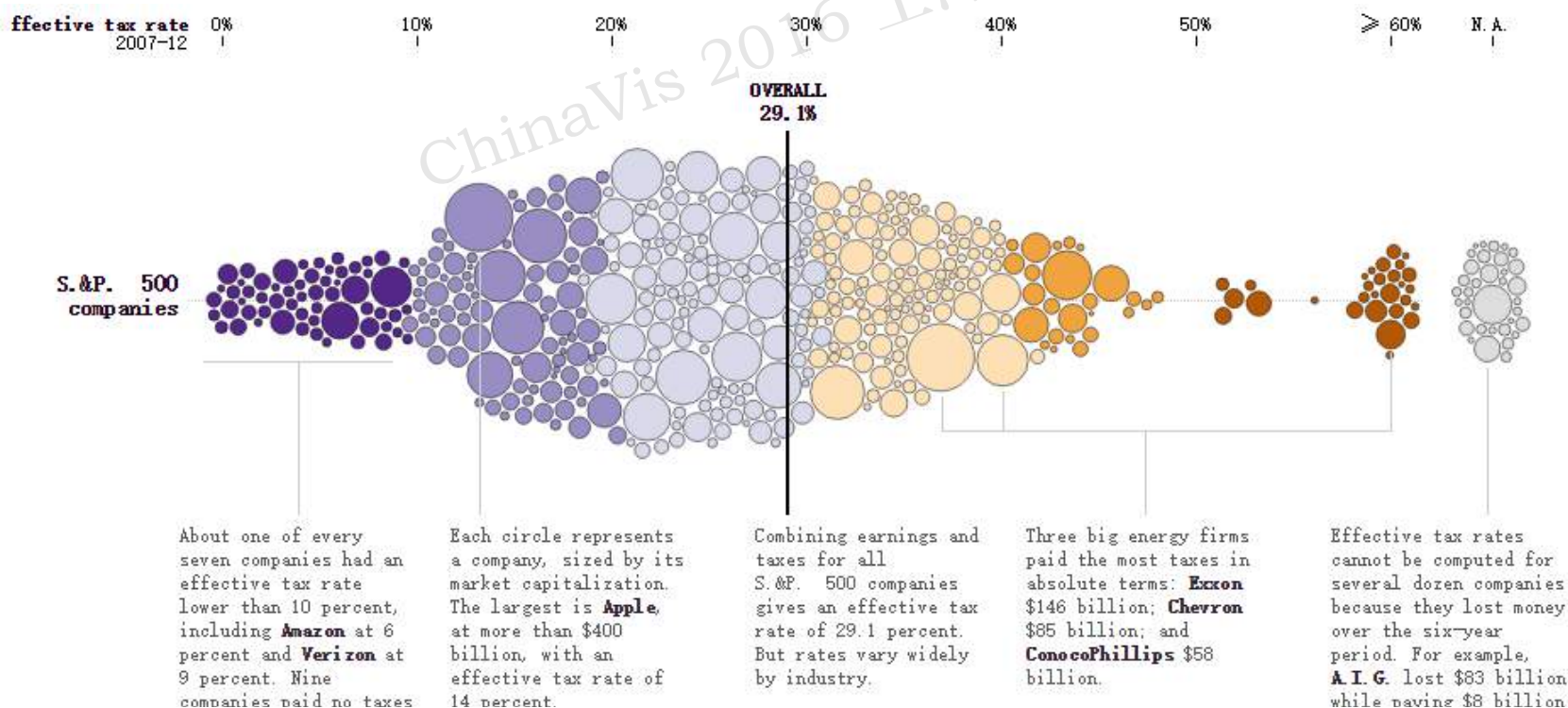




# 记者与媒体：新闻分析中的可视化



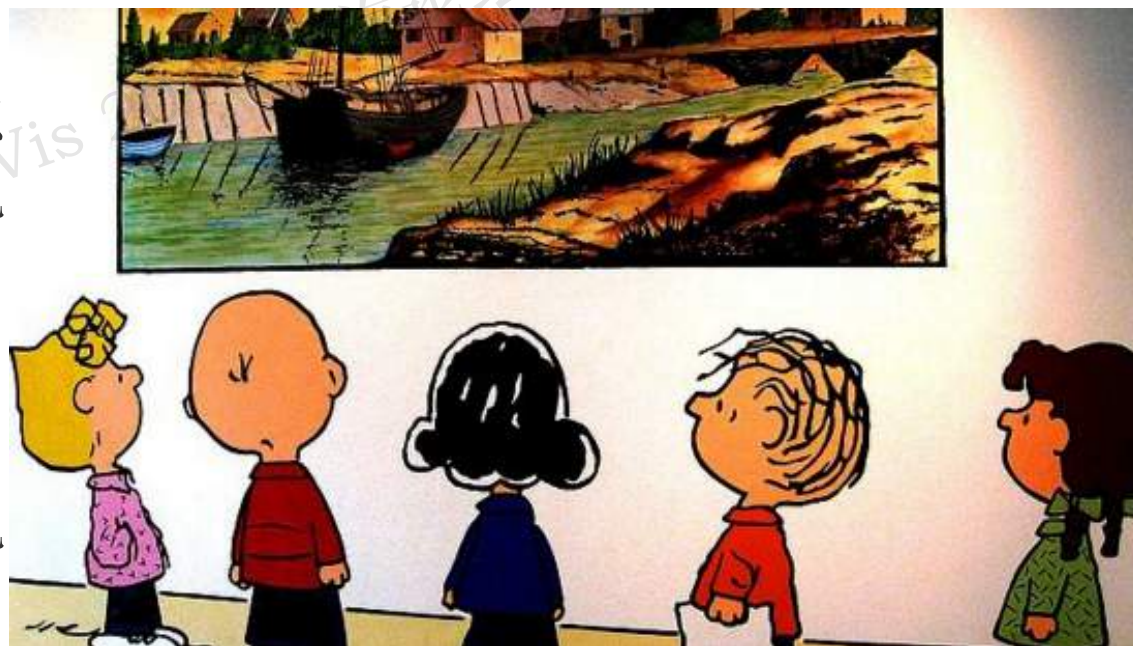
- 商业信息可视化
  - 五百家美国上市公司的纳税与盈利情况
  - [在线演示](#)





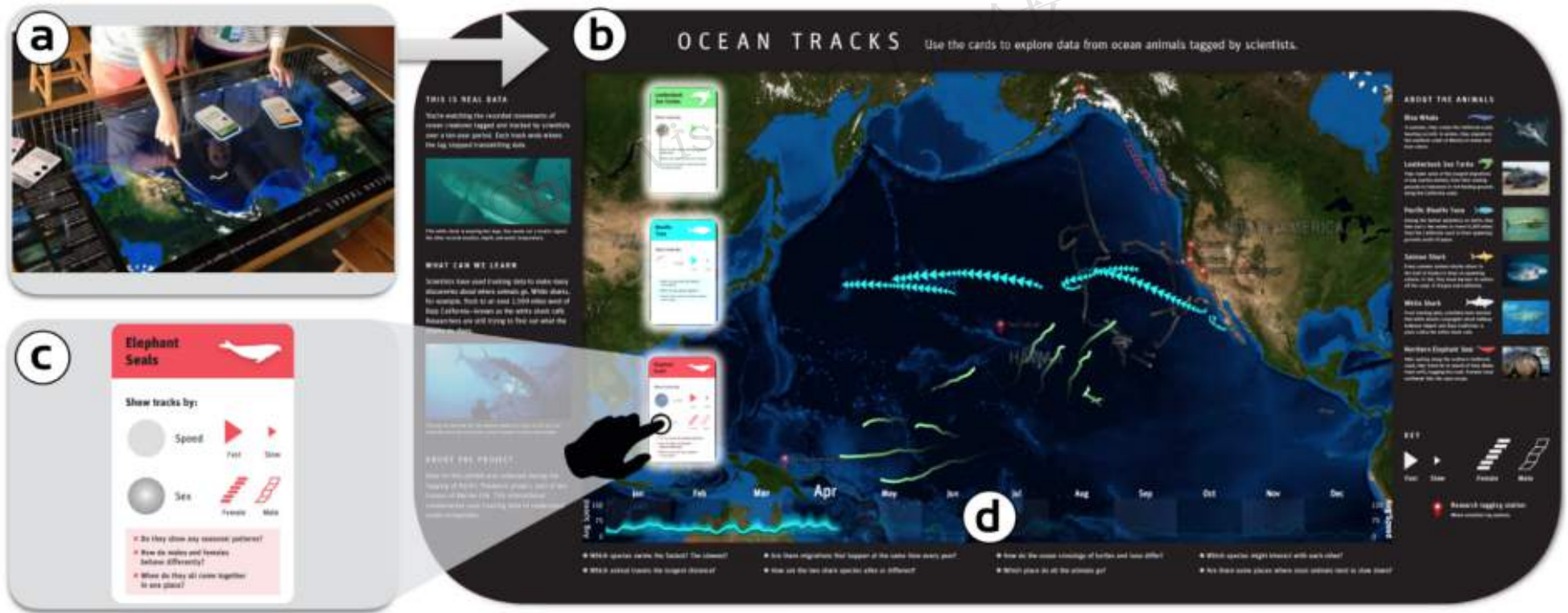
# 展览中的可视化 - 受众：普通用户

- 特点：科普作用，让用户实际地参与感知科学与历史
- 需求：
  - 让受众理解复杂的科学知识
  - 对广泛的知识范围进行可视化
- 挑战：
  - 交互性的设计
  - 如何让孩童会玩、理解可视化



# 展览中的可视化

- 受众：科普作用，让用户实际地参与感知科学与历史



# 个人用户 - 数据拥有者

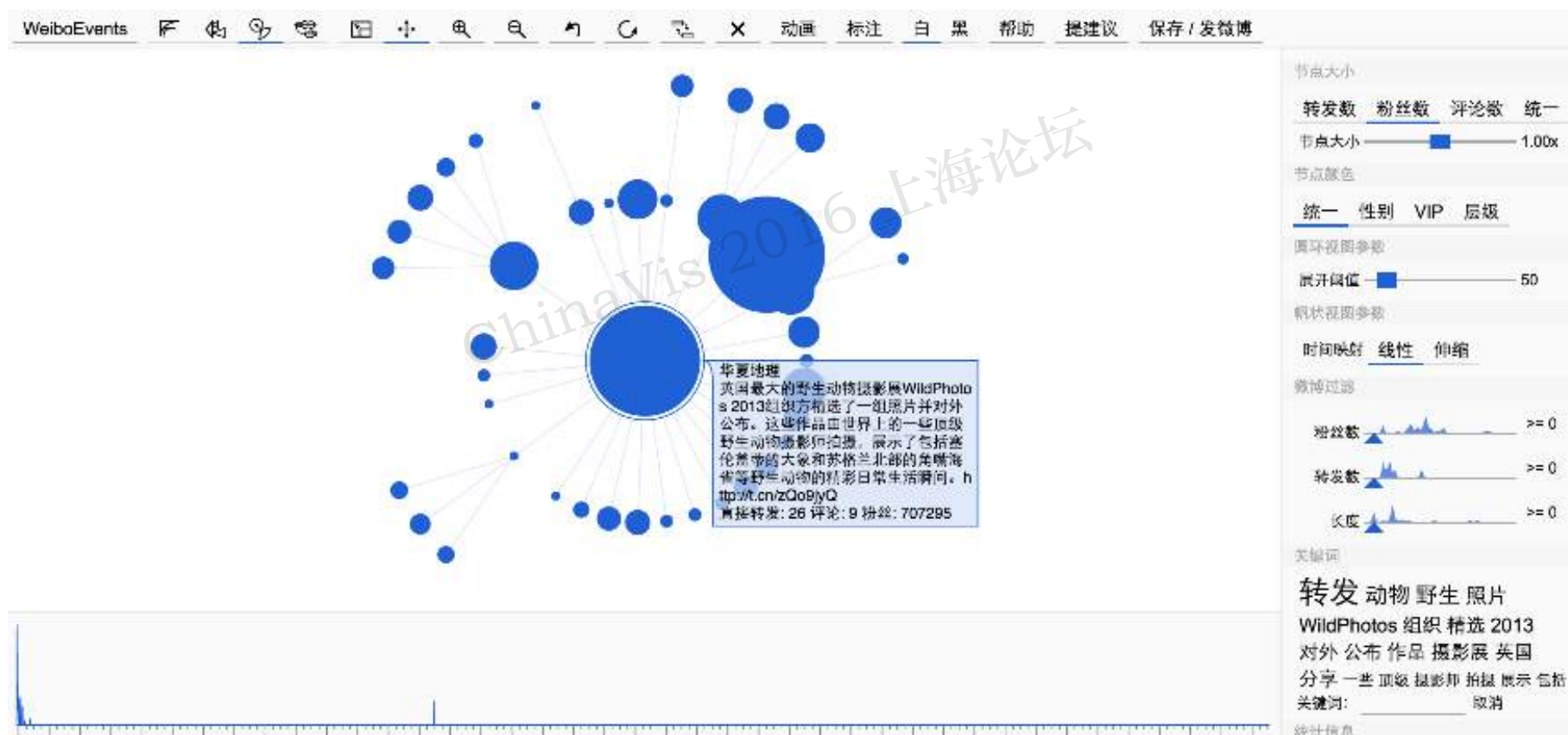
- 微博用户，推特用户
- 需求：他们对自己的信息，周围的信息十分关心，却无法得到有效的整理与总结
  - 单条微博传播模式的分析
  - 事件传播中关键用户的挖掘
  - 基于微博的个人轨迹分析
- 挑战
  - 对数据的实时分析
  - 针对个人定制化的可视化界面
  - 广泛传播与分享





# WeiboEvent

## 单条微博传播模式分析



<http://vis.pku.edu.cn/weibova/weiboevents/>

# WeiboEvent

## 单条微博传播模式分析

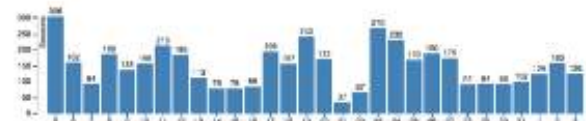
WeiboEvents 首页 使用说明 数据抓取 数据分享 联系我们 出版物 关于 English

### 最新分享

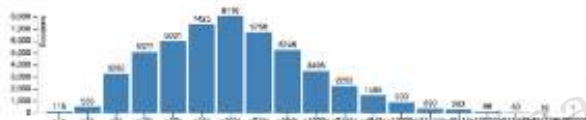
这里展示用户最新分享的微博可视化。1681 个用户创建了 15633 个可视化作品。使用次数: 83822, 用户数: 11213。

在我的内容里可以下载您之前保存的数据。

每日使用次数:



使用时长统计:



1 2 3 4 ... 311 312 313

木子林夕七采页: 分析 @贺卫方的微博。  
原始微博: @贺卫方: 【致校】今天回京, 看到满屏说梁静, 赶快看了一遍, 很感动。这篇演讲把令人恐怖的现状及根源都说清楚了, 有人指责说回避了某些更根本原因, 那多半是他们不能或不理解解文本意义而已。至于解决之道, 其实集团诉讼在不少国家发挥了重要作用, 可惜中国总有一种群体行动恐惧症, 司法又不独立, 难! 转发: 800 评论: 340 时间: Mar 1st, 2015 23:53 进入微博



Add tag

Jun 3rd, 2016 20:04

查看大图 打开

三岛来: 分析 @于路 的微博。  
原始微博: @于路: ...



最近30天活跃用户:

- @wangyirUC (110)
- @宋乐zz (70)
- @乌拉乌拉一号 (58)
- @文春莹 (40)
- @stellapink买太多 (37)
- @李家小彩影 (31)
- @没心没肺的小日子 (28)
- @圈圈棒棒糖 (25)
- @木子林夕七采页 (12)
- @阿拉丁斑马 (10)
- @梓墨牛 (10)
- @田大菊儿 (8)
- @一瓣牛奶草莓 (5)
- @tofoxbai (5)
- @cocoofish (5)
- @三岛来恩 (5)
- @阿布兵长\_Alwin (4)
- @萤火虫的silly2014 (4)
- @撒哈拉的猫一只 (4)
- @黄小远去远方 (4)

- 分享机制
  - 分享分析结果, 发表评论
  - 查看他人的最新分享, 进行进一步的分析探索

# Weibo Footprint

## 基于微博的个人轨迹分析

PKU Weibo Footprint Visual Analytics Home Publications Team PKU Weibo VA

### 北京大学 PKUVIS 微博可视分析系列工具之: PKU Weibo Footprint Visual Analytics

带有地理标签的社交媒体数据, 例如微博, 反映了人们带有社交属性的移动行为。这可能是记录一次旅行的轨迹, 或者分享一个特别的地方, 抑或只是每日在熟悉的工作场所、家户签到。这些带有地理标签的数据, 总结了每个人的行为轨迹。如果你也有发表过带有地理信息的微博, 那么, 来尝试一下吧!

在系统中, 你可以使用时间轴与空间轨迹, 点击进行时空穿越, 探索当年旅行的足迹, 或者某去的地方, 也可以邀请你的好友, 一同探索他们的轨迹。

Take a Try

VisPKU

北京大学微博可视分析系列

Copyright 2012-2016, Peking University. All rights reserved.  
© PKU Visualization and Visual Analytics Group 2009-2016

[http://vis.pku.edu.cn/weibova/weibo\\_footprint\\_vis/en/index.html](http://vis.pku.edu.cn/weibova/weibo_footprint_vis/en/index.html)

<http://vis.pku.edu.cn/wiki> 袁晓如 北京大学



# Weibo Footprint

## 基于微博的个人轨迹分析

- 数据：个人的携带地理标签的微博
- 目标：分析用户的轨迹



- (a) 省份级总览图。高亮的城市分别是上海和黑龙江
- (b) 城市级视图。该图展示的是上海市
- (c) 区域细节图。该图展示的张江区域的细节

## 数据分析兴趣爱好者

- 特征：对数据分析有兴趣，却不是专业的程序员或者数据分析专家
- 需求：直观、简单一上手的可视化工具
  - 所见即所得
  - 需要少量或者不需要编程知识
  - 通用数据分析功能
- 挑战：
  - 对通用性的数据进行探索
  - 提供完整、可用、完备的接口



# 数据分析兴趣爱好者：D3—数据驱动文档

- D3, JavaScript库, 协助用户在网页端快速进行数据可视化的开发
- 针对的用户群：基于网页端的数据可视分析人员
  - 拥有大量、类型丰富的数据
  - 具有一定的网页编程能力
  - 快速进行探索分析



# D3—数据驱动文档

- d3的特点：
  - 简易的语法
  - 详细的文档
  - 丰富的样例



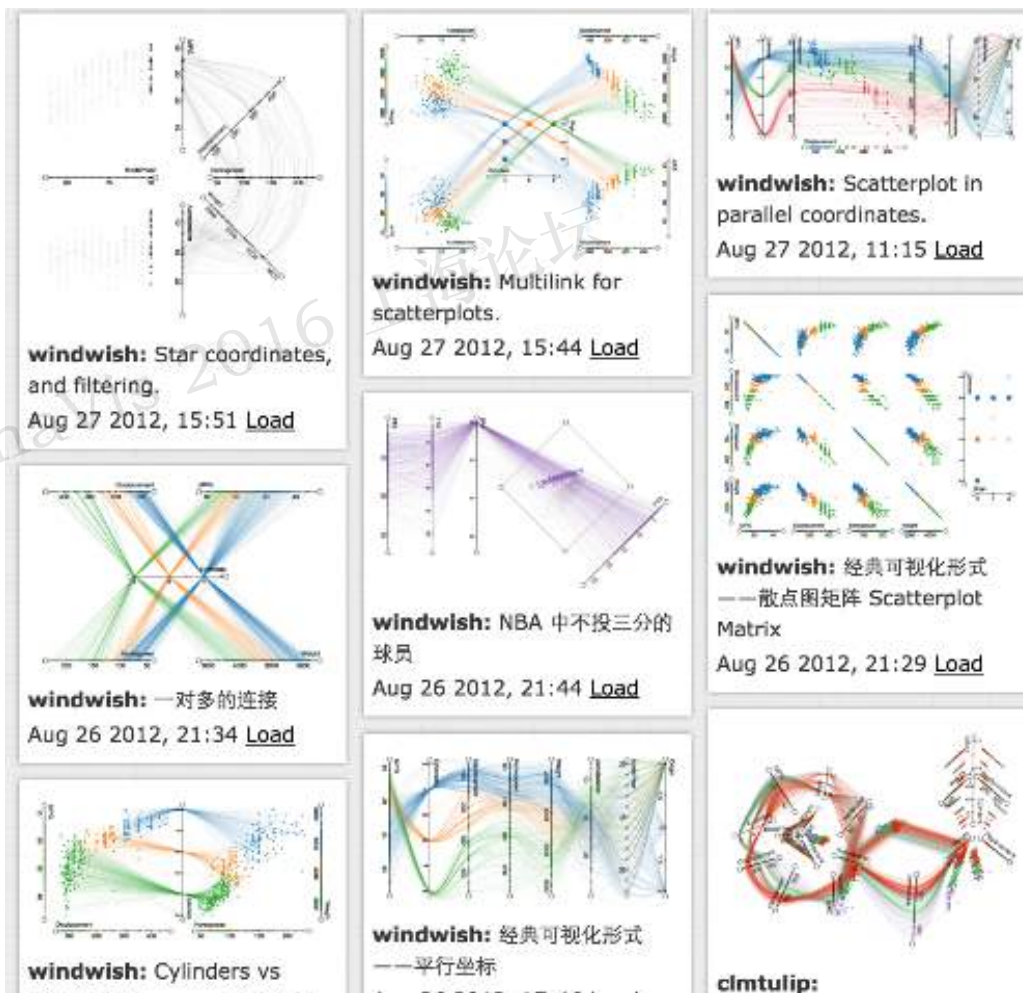
<https://d3js.org/>

# 数据分析兴趣爱好者： Jigsaw— 高维数据可视化分析工具

- Jigsaw: 高维数据可视化分析工具，支持用户交互式的分析探索高维数据蕴含的信息
- 针对的用户群：
  - 拥有的大量高维的数据
  - 快速进行数据分析探索
  - 可不具备编程能力

# Jigsaw—高维数据可视化分析工具

- Jigsaw的特点：
  - 无需编程
  - 交互式创建可视化
  - 多视图连接探索



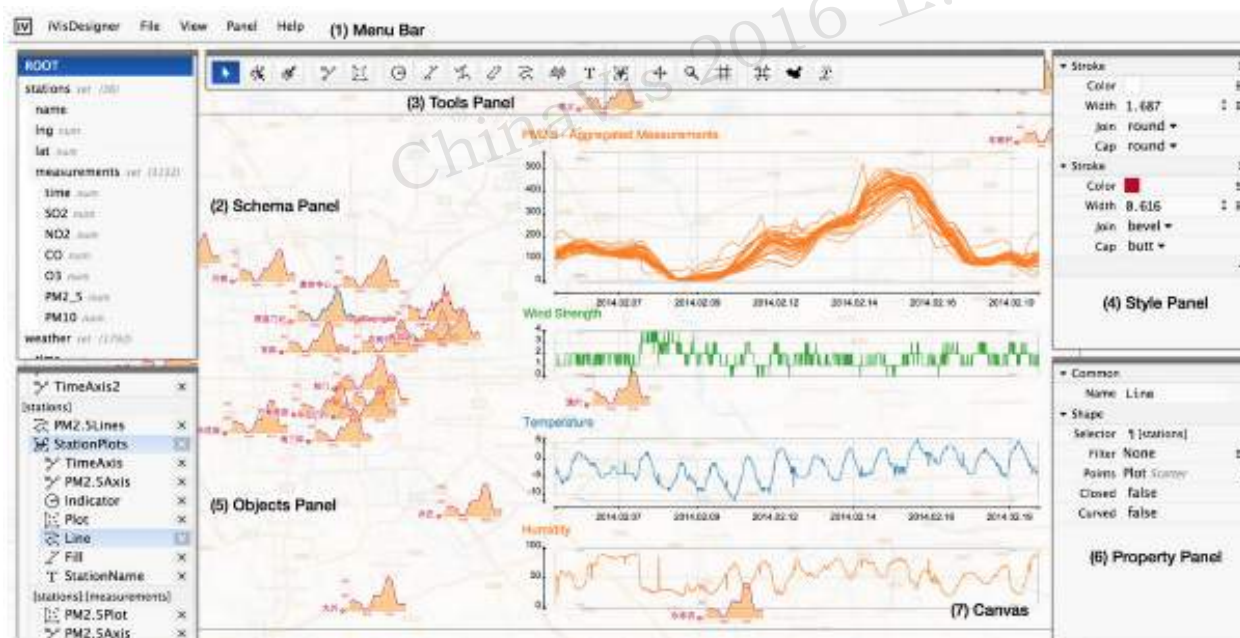


# 数据分析兴趣爱好者： iVisDesigner—可视化设计工具

- iVisDesigner, 可视化设计工具, 支持用户通过交互的方式, 快速创作可视化展示形式, 无需编程。
- 针对的用户群
  - 拥有大量、类型丰富的数据
  - 根据数据特点, 快速创作合适的可视化展示形式
  - 快速进行数据探索分析
  - 无需编程

# iVisDesigner—可视化设计工具

- iVisDesigner的特点：
  - 支持多种可视化展示形式，比如散点图，力导向图，时间轴等
  - 交互式创建可视化视图，无需编程



<https://donghaoren.org/ivisdesigner/>

# 可视化—发展趋势

- 数据对象从传统的单一数据来源扩展到多来源，多维度，多尺度等广泛的数据；
- 用户从少数专家用户扩展到广泛的不特定的群体；
- 研究更为强调方法的可扩展性和开发的简捷性；
- 革命性的新方法和原有普遍适用的可视化技术和工具结合；
- 面向领域和大众的可视化工具库和协同与众包的可视分析方式将被更广泛地引入到各种综合分析工具和平台，大幅度提升各类大数据分析的易用性和有效性；
- 沉浸式可视分析将在今后3-5年涌现 (New)
- 可视化即为服务，Visualization As Service



# 北京大学可视化与可视分析研究组



实验室主页 <http://vis.pku.edu.cn/>

实验室博客 <http://vis.pku.edu.cn/blog>

空间数据可视分析网站 <http://vis.pku.edu.cn/trajectoryvis/en/index.r>

公众号“可视分析”

