

信息可视化技术概述

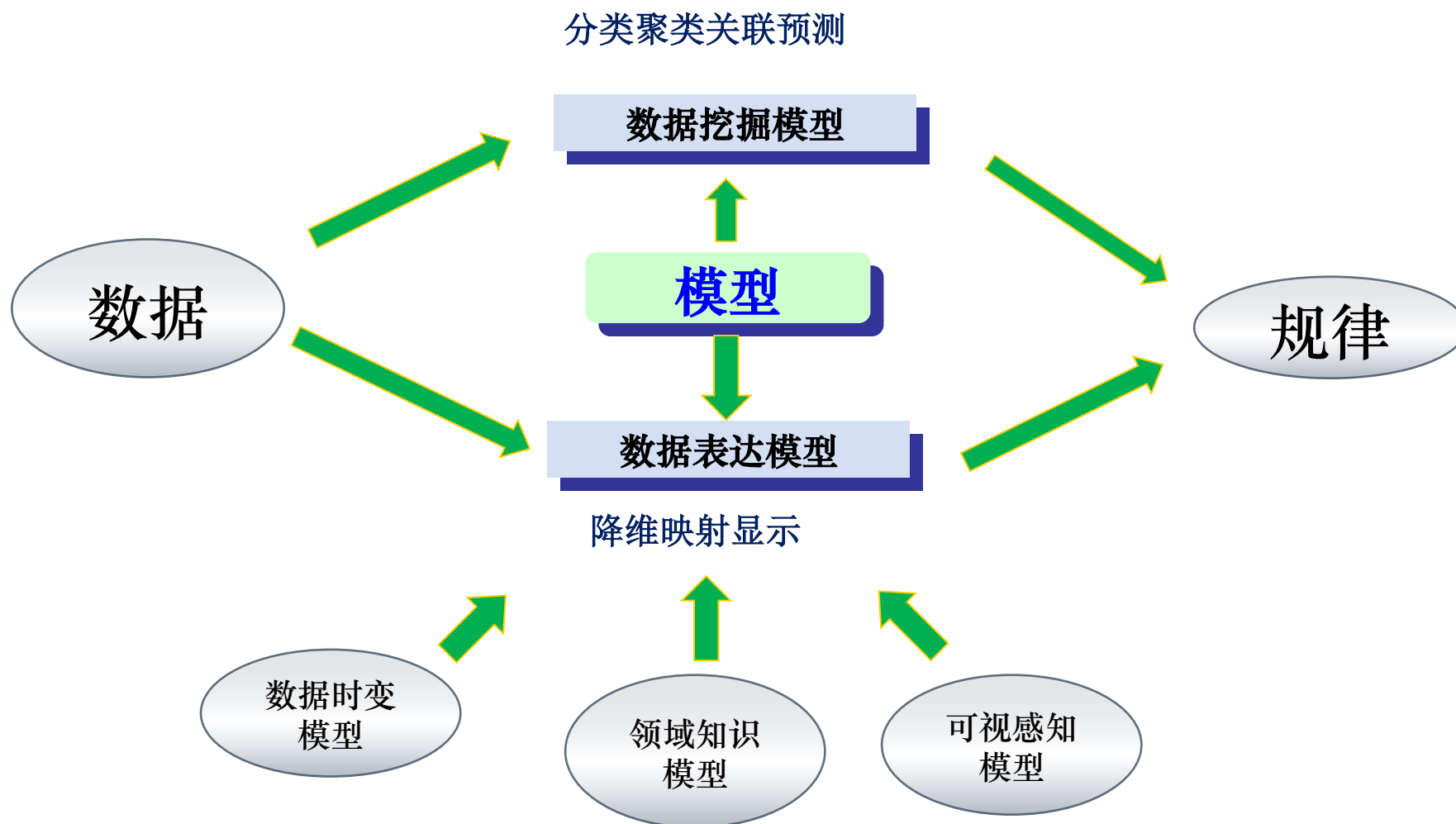
王长波 教授、博导

华东师范大学



信息可视化

1、数据分析的常见过程



1、数据分析的常见过程

第一步：明确分析目的



EIUSMOD TEMPOR

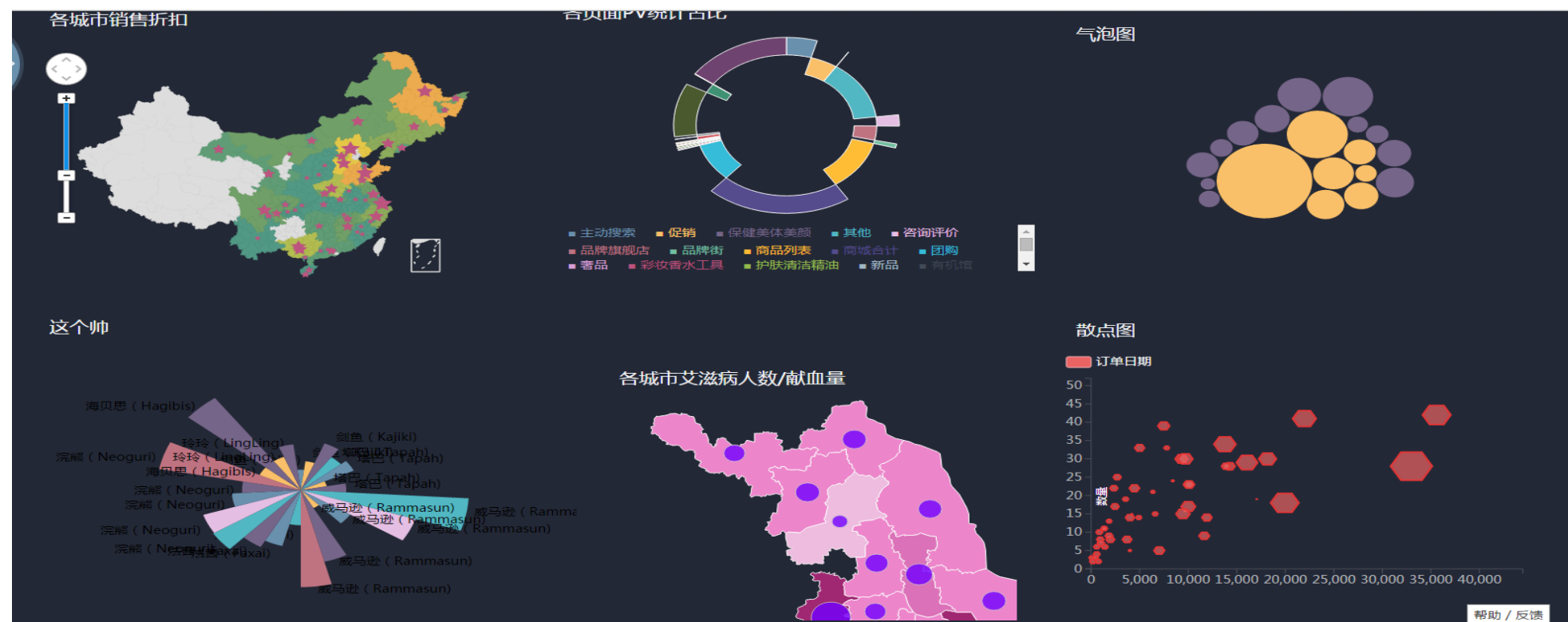
进行数据分析前要明确分析的主题和目的，需要通过数据分析展示什么样的成果，数据需求直接源于最后的分析结果

EIUSMOD TEMPOR

对于观测、跟踪数据进行分析时，强调实时性、变化性。对于强调数据呈现度的数据进行分析时，进行交互、检索的设计等。不同的目的决定了不同的图形表现形式。大数据魔镜作为专业的图表可视化软件，可根据不同的数据分析需求，满足不同企业和个人的分析需求。

1、数据分析的常见过程

第二步：收集整理数据



准确的找到所需要的数据

保证数据的可靠性,

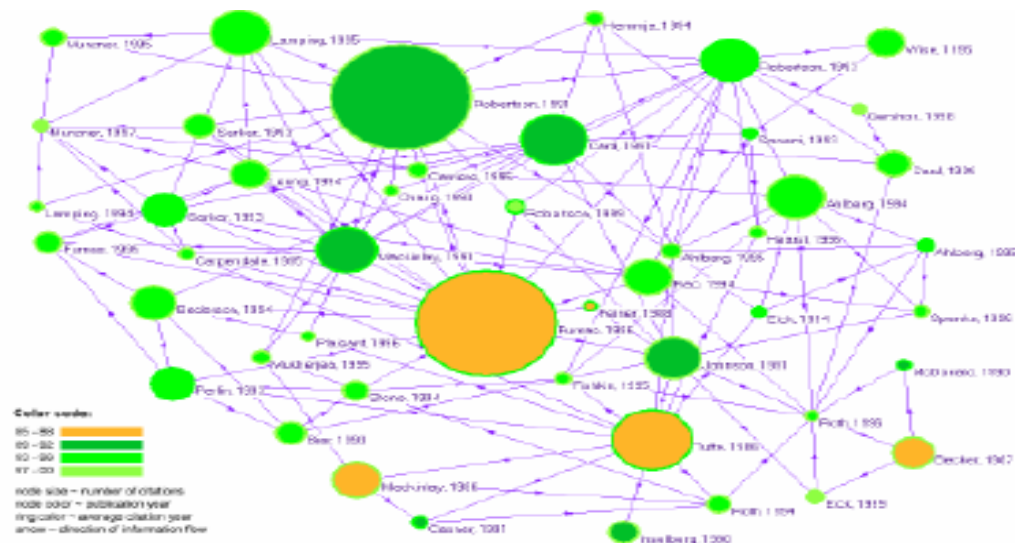
数据要丰富充盈

1、数据分析的常见过程

第三步：数据降维与清洗

“

- 数据处理技术主要降低数据的维数
 - ⊙ 子集技术 (包括取样和查询)
 - ⊙ 分割技术 (分割技术是基于属性的取值的)
 - ⊙ 聚合技术 (通过绘制柱状图和饼图等进行可视化)
- 支持分析的存储和索引技术



1、数据分析的常见过程

第四步：数据挖掘与分析

(1)分类

- 根据数据的特点将其划分进某一个类别。



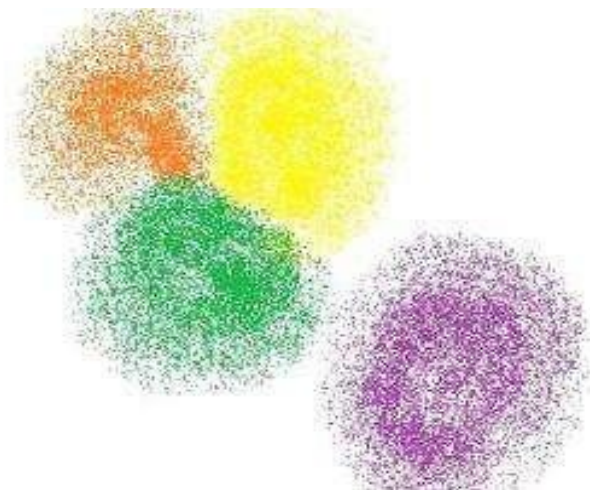
例如，判断学生的性格类型

(2)聚类

- 其目的在于将相似的数据对象归为同一类别。
- 与分类不同的是，聚类要划分的类别是未知的。



例如，将相近的进行聚类，按类别存储与分析

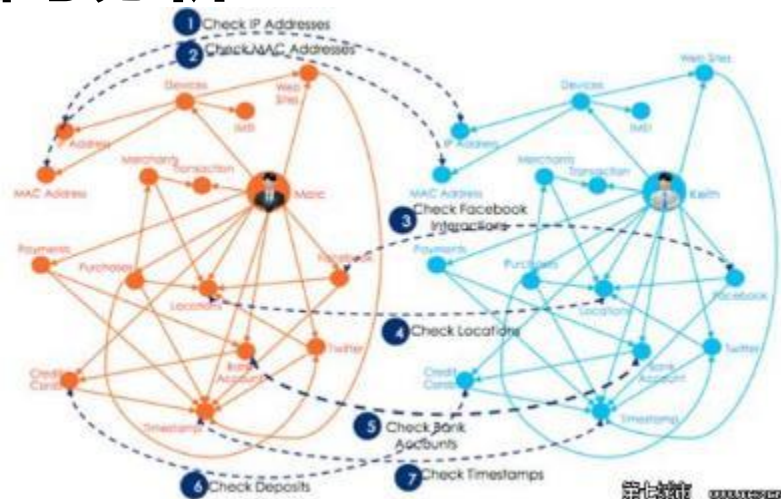


1、数据分析的常见过程

第四步：数据挖掘与分析

(3)关联

- 其目的在于发现数据对象之间的关联或关系。



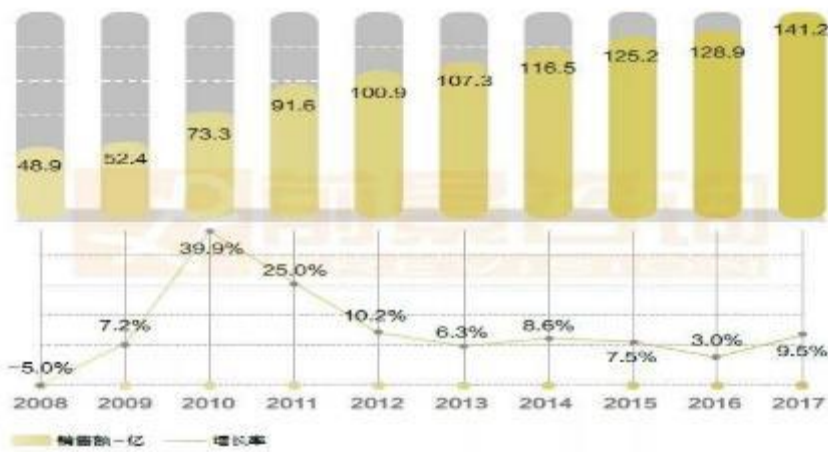
作用



例如，发现学生同时选修的课程，即这样的课程具有关联性

例如，根据学生以往的表现，预测学生未来的成绩。

作用



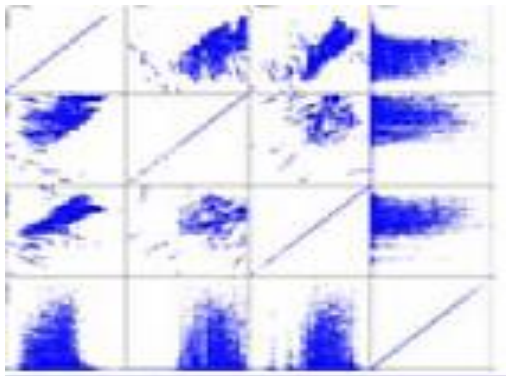
(4)预测

- 回归：其目的在于为数据对象赋予一个数值。

1、数据分析的常见过程

第五步：数据可视分析结果

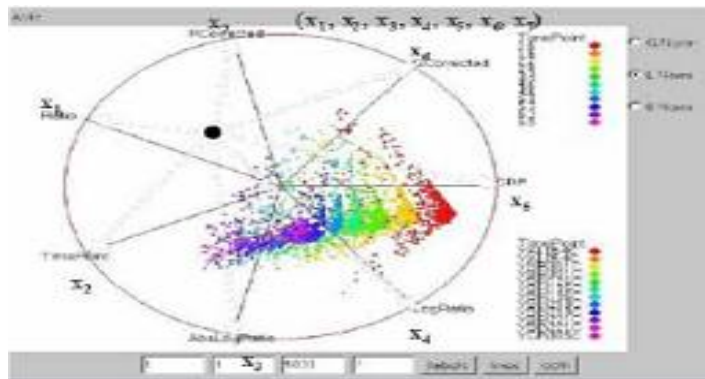
丰富灵活的前端展示，完备的数据生态系统秒级渲染



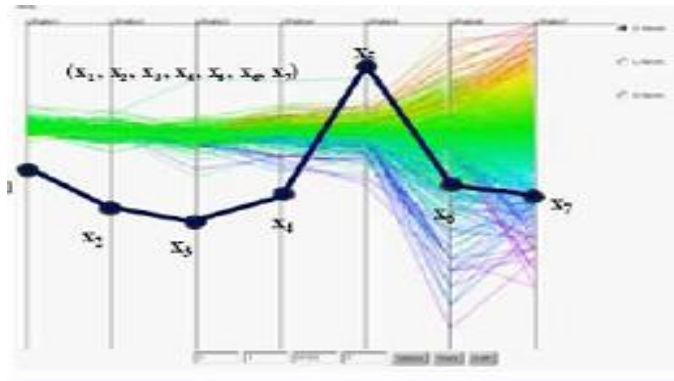
魔镜
www.moojnn.com



超快的分析速度和卓越的分析性能



500多种可视化分析效果，最数据研究的可视化分析平台





二、可视分析

信息可视分析技术

可视分析的基本方法

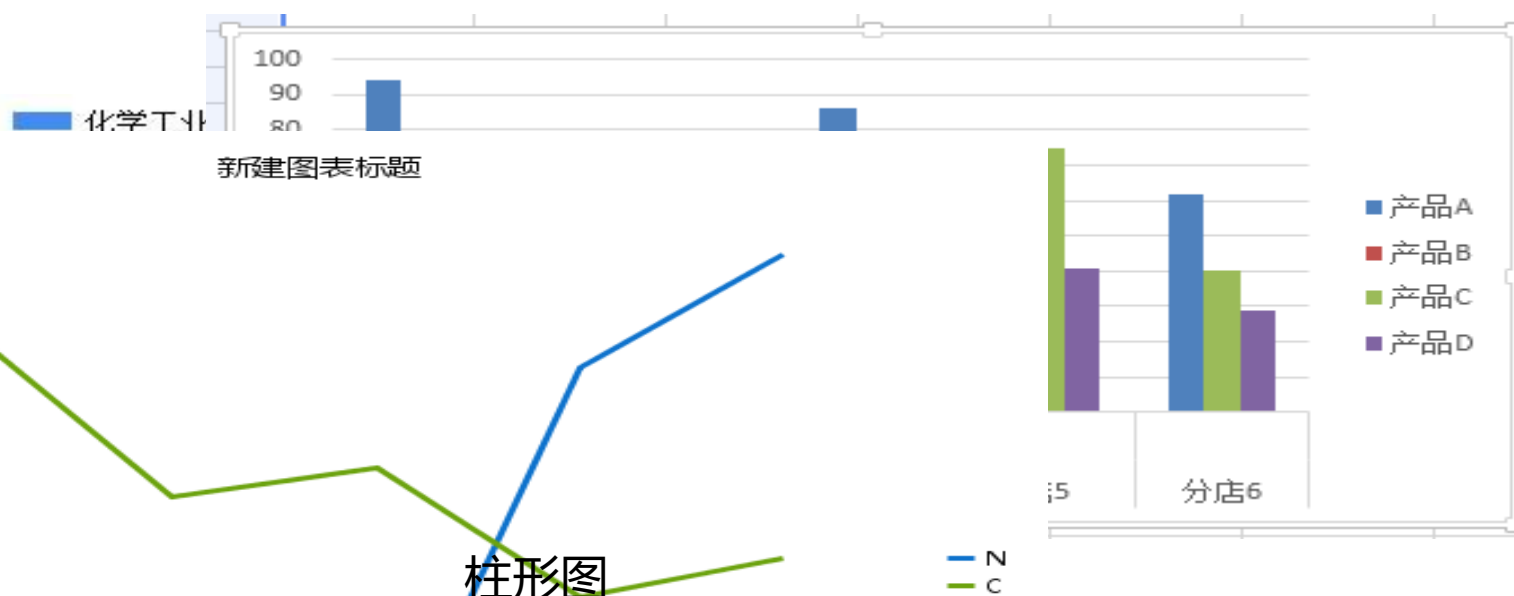
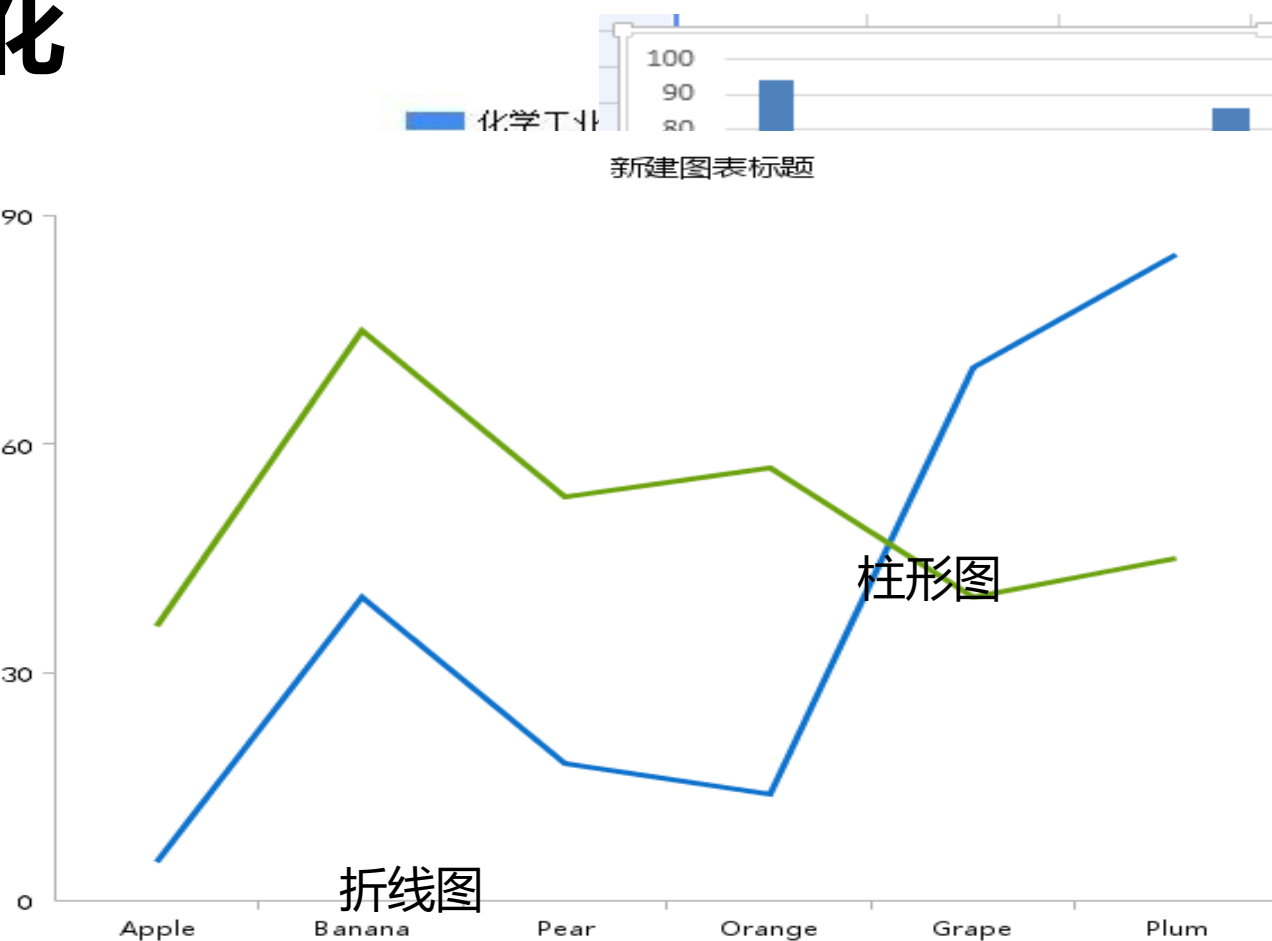
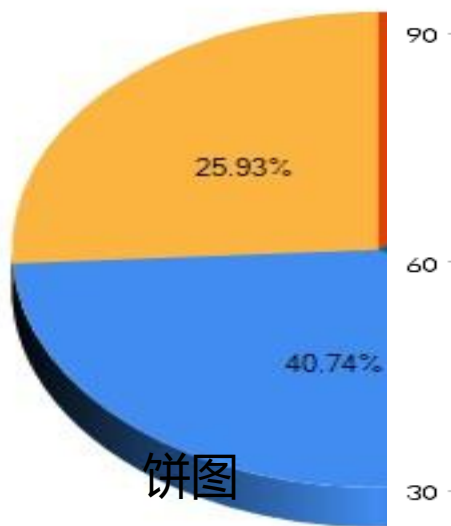
□将事物图形化



利用图形表示事物，传递信息。

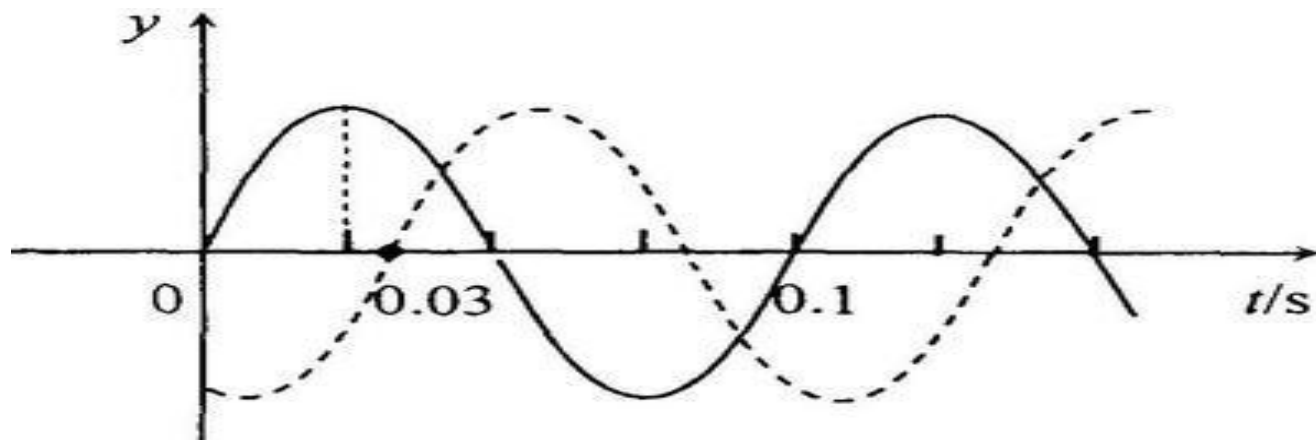
可视分析的基本方法

将数值图形化

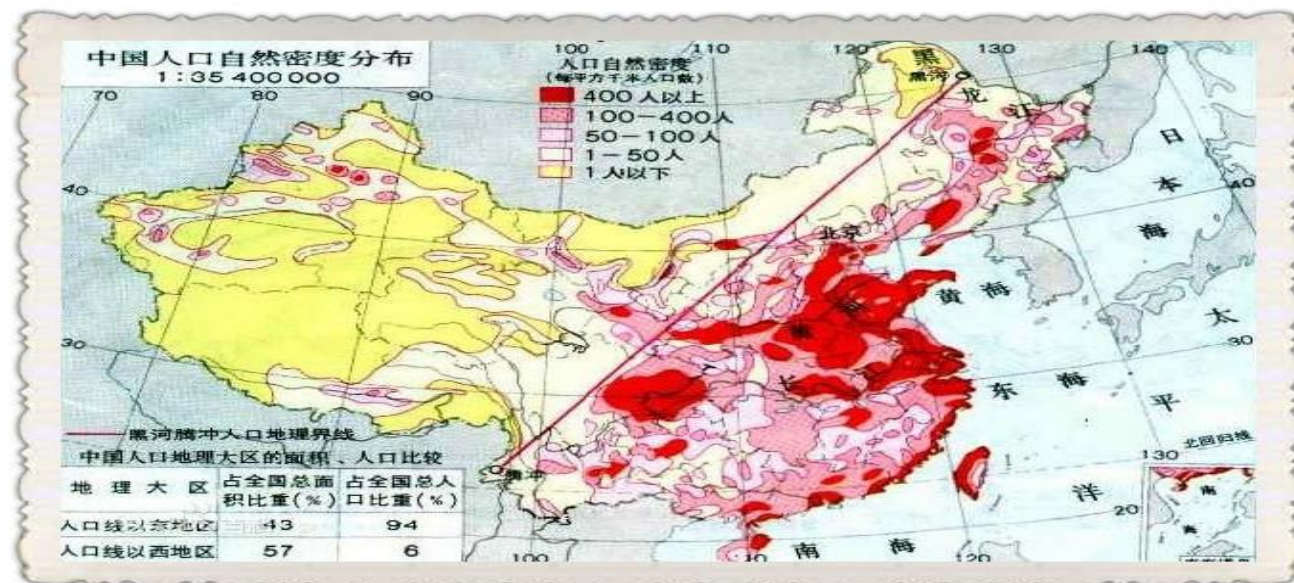


可视分析的基本方法

●将时空可视化

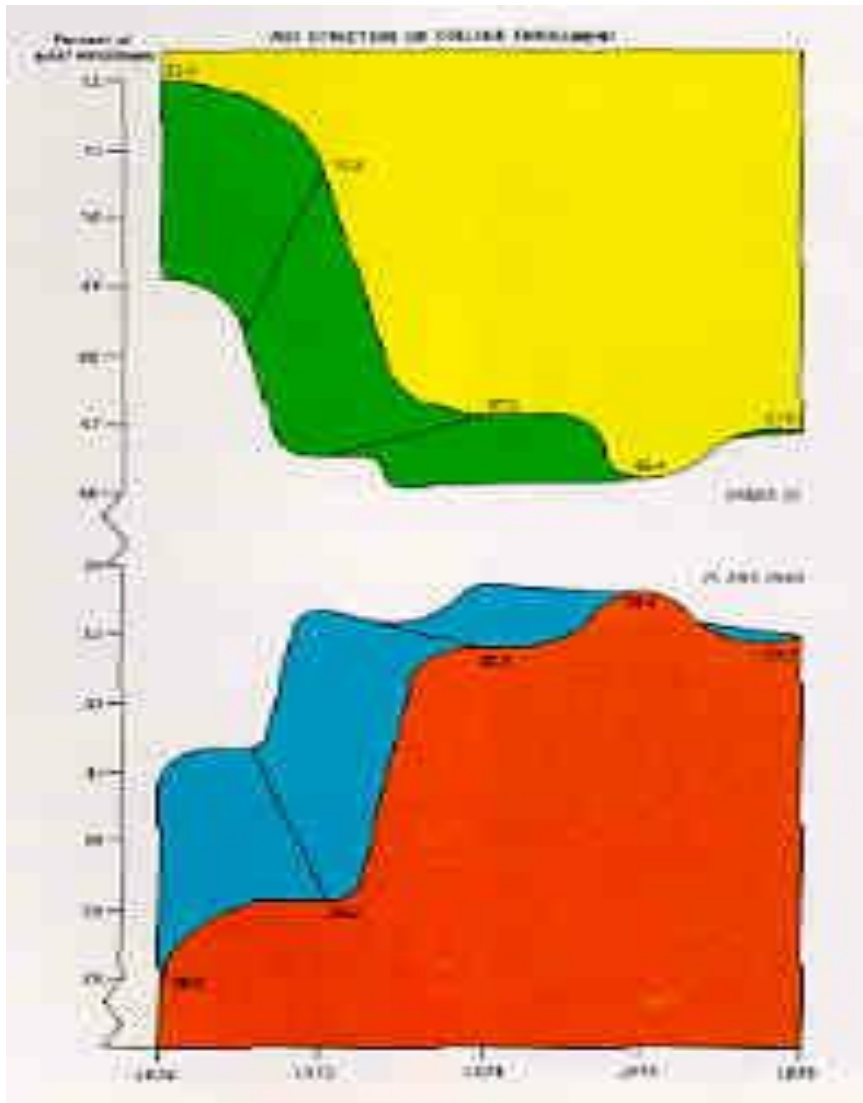


时间



空间

Good and Bad Visualizations



Bad visualization - As a substitute for substance, one can try lots of color, 3D effects, or disguised redundancy. This graph uses all three techniques, to display just **five** numbers. Note the clever use of mirror-imaging -- the top series is just (100 - the bottom series) and the interesting use of curved lines, front and back to avoid the appearance that there's a lot less here than meets the eye.

La France des pains

French breads / Frankreich und seine Brotsorten / La Francia del pan / I differenti tipi di pane francesi



- 1 Pain pie
- 2 Bata Michen
- 3 Pain de Morlain
- 4 Pain Chapou
- 5 Bonimare
- 6 Mirand
- 7 Pain saumon
- 8 Mense
- 9 Gâtrot
- 10 Pain de Cherbourg
- 11 Pain bateau
- 12 La couronne mirale
- 13 Le Tourton
- 14 La gâche
- 15 Pain pennais
- 16 Pain brié
- 17 Pain à soupe
- 18 Pain de mie
- 19 Le magret
- 20 Pain régence
- 21 La faluche
- 22 Pain polka
- 23 Pain arrichant
- 24 Petit pain rhaître
- 25 Petit pain chéme
- 26 Petit pain porroet
- 27 Petit pain souverain
- 28 Petit pain empereur
- 29 Petit pain mirand
- 30 Pain de fantaisie
- 31 Pain marchand de vin
- 32 Le benoît
- 33 Pain saucisson
- 34 Pain fendu
- 35 Pain boulot
- 36 Pain aux noix
- 37 Pain cordon
- 38 Pain rhaître
- 39 Sabot
- 40 Pain Graham
- 41 Pumpernickel
- 42 Pain tressé et pain torté
- 43 Le fer à cheval
- 44 Pain collier
- 45 La foude
- 46 Le pain cordé
- 47 Le pain de seigle
- 48 Pain chemin de fer
- 49 La couronne
- 50 La couronne de Bugry
- 51 Le pain vandois
- 52 La couronne boodelaise
- 53 La soufflée
- 54 Le mellel
- 55 Le sedr
- 56 La manode
- 57 La fougasse
- 58 Le pain bouilli
- 59 La rousse
- 60 Le gascou ou l'agenais
- 61 La moute
- 62 Le tignolet
- 63 La flambade, flambé, flambèche
- 64 Le rouda
- 65 Le quatre-banés
- 66 Le pain-manteau
- 67 L'échaudé
- 68 Le Pain de Lodève
- 69 Le phosix, le pain viennois
- 70 Le charleston
- 71 Le cavaille
- 72 Le pain coiffé
- 73 Le beaucaire
- 74 Le pain scié
- 75 Le pain d'Aix
- 76 La tête d'Aix
- 77 Le charleston nigois
- 78 La michette
- 79 La main de Nîmes et le monde-dessus
- 80 La coupette

Les pains régionaux par Lionel Poilâne.

Lionel Poilâne's regional bread.

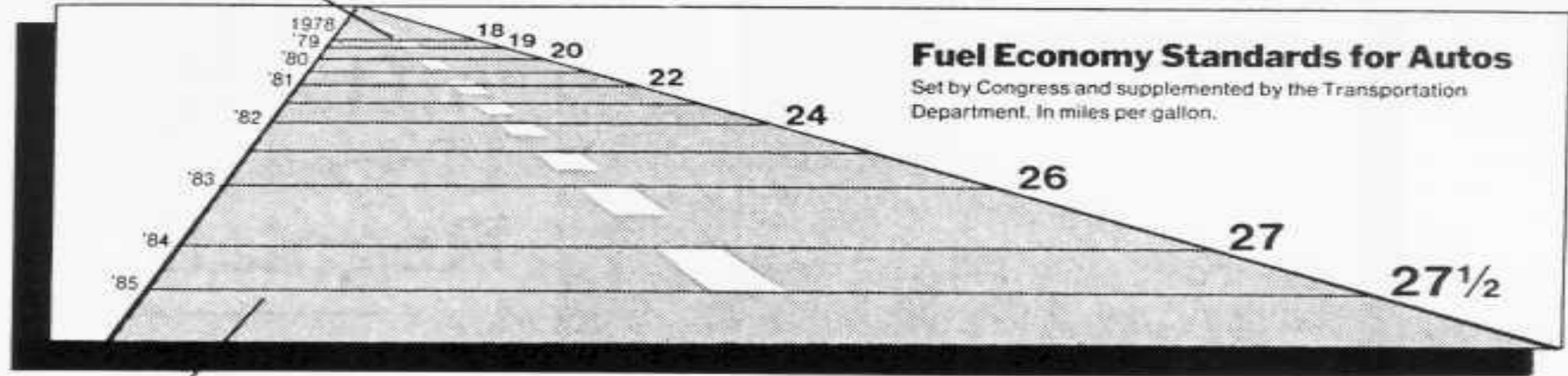
Regionale Brotspezialitäten von Lionel Poilâne.

El pan regional por Lionel Poilâne.

I pani regionali di Lionel Poilâne.

Bread map of France

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.

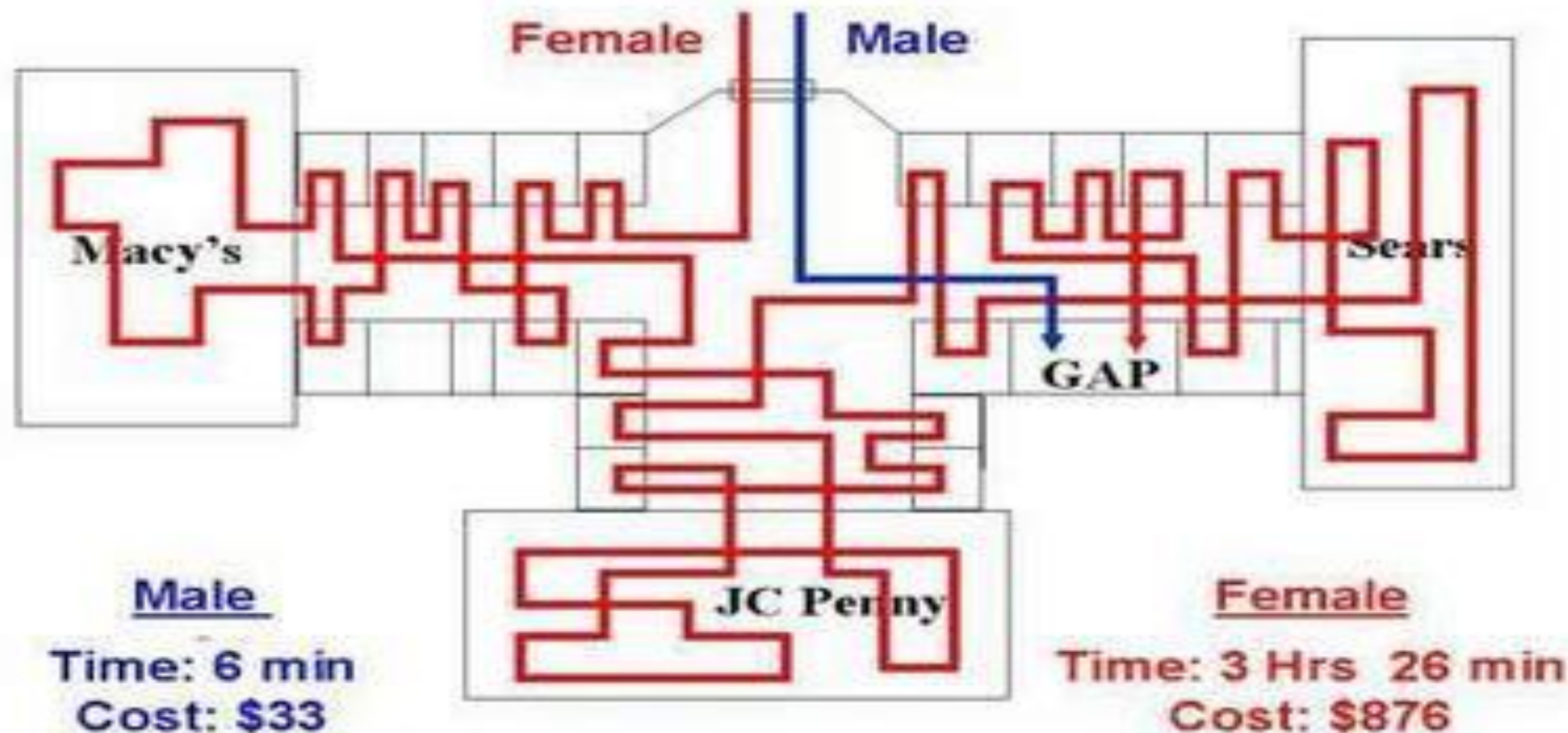


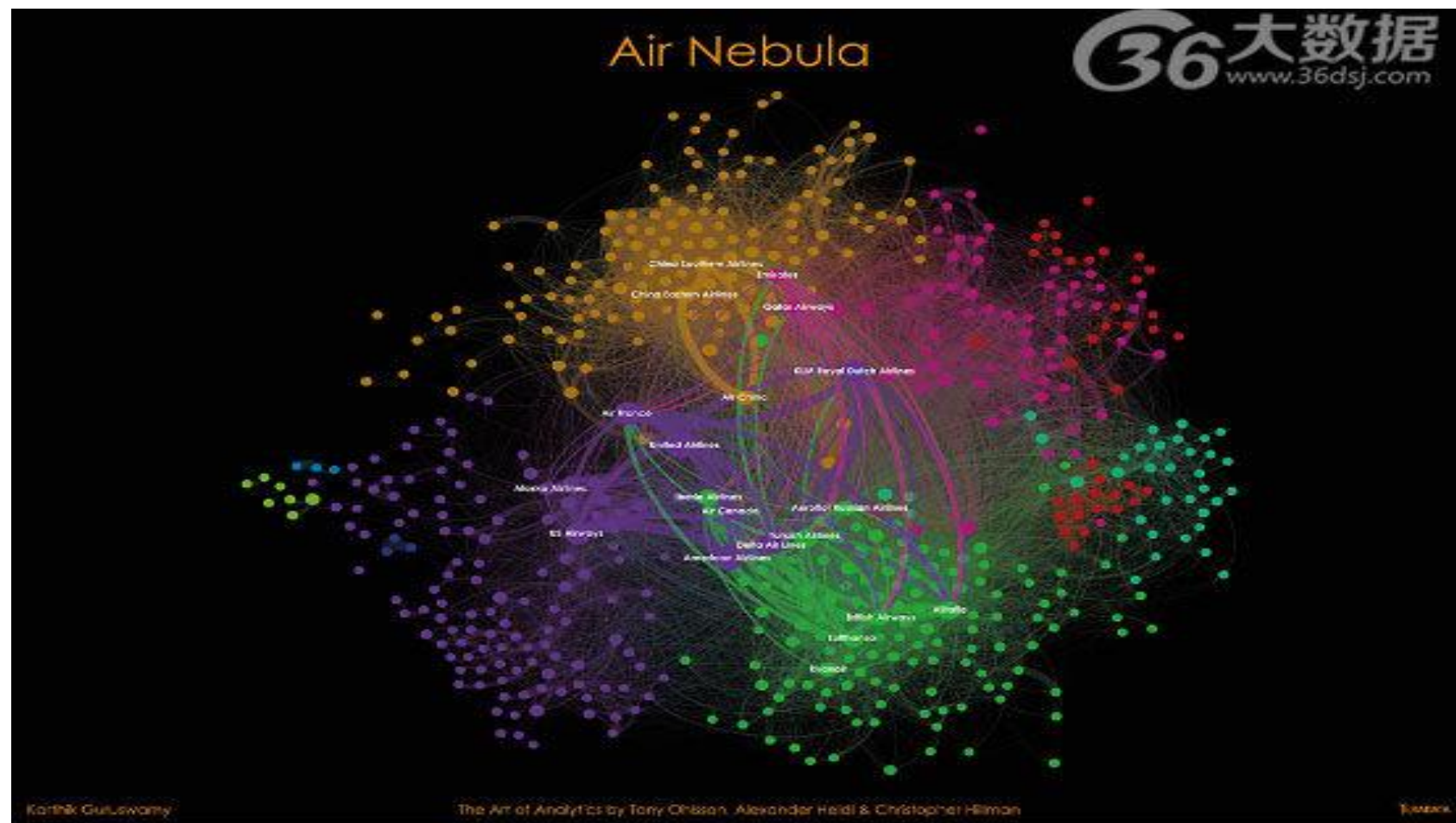
This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

New York Times, August 9, 1978, p. D-2.

Bad visualization - Mandated fuel economy standards set by the US Dept of Transportation. The standard required an increase in mileage from 18 to 27.5, an increase of **53%**. The increase here is **783%**, for a lie factor = $(783/53) = 14.8$!

Mission: Go to Gap, Buy a Pair of Pants





航线星云

■ Solid Dem. ■ Leans Dem ■ Toss-up ■ Leans GOP ■ Solid GOP

244

TOSS-UPS

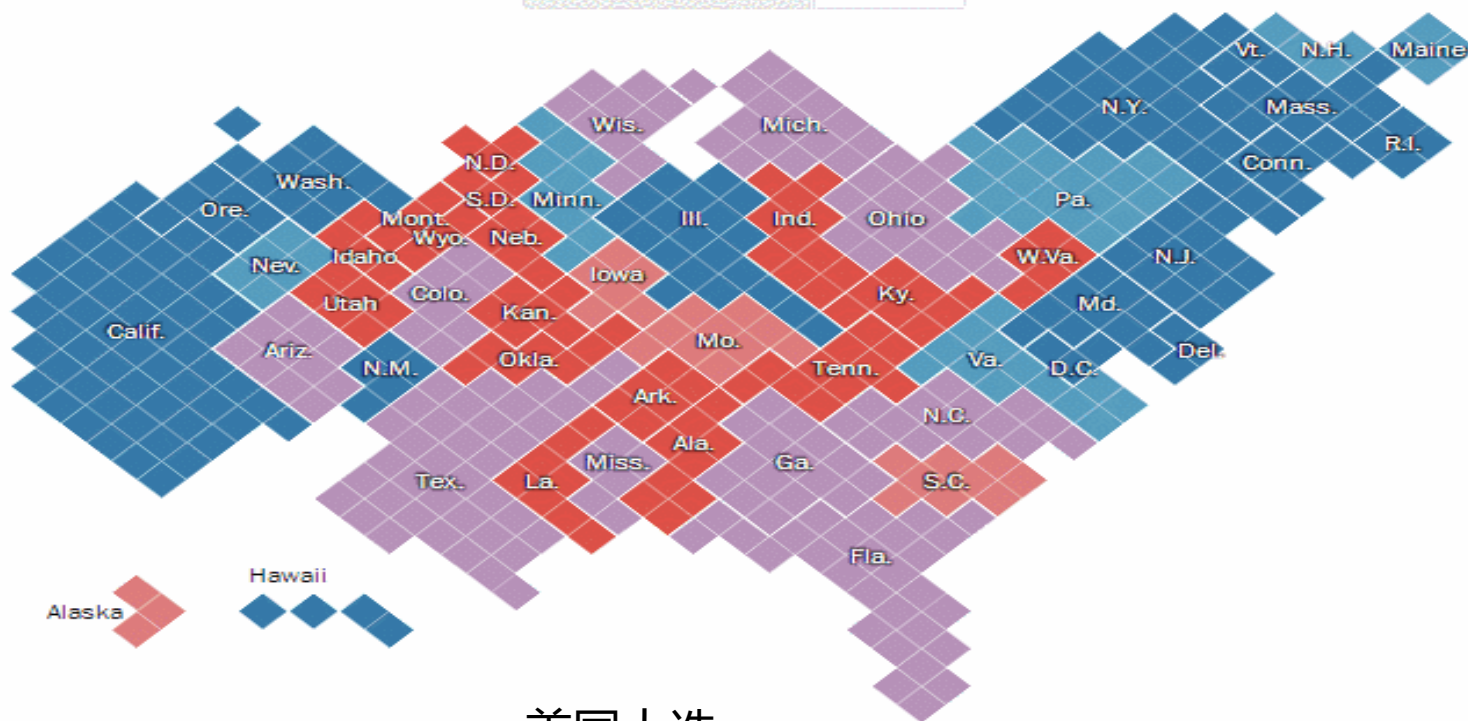
168

126

solid and leaning



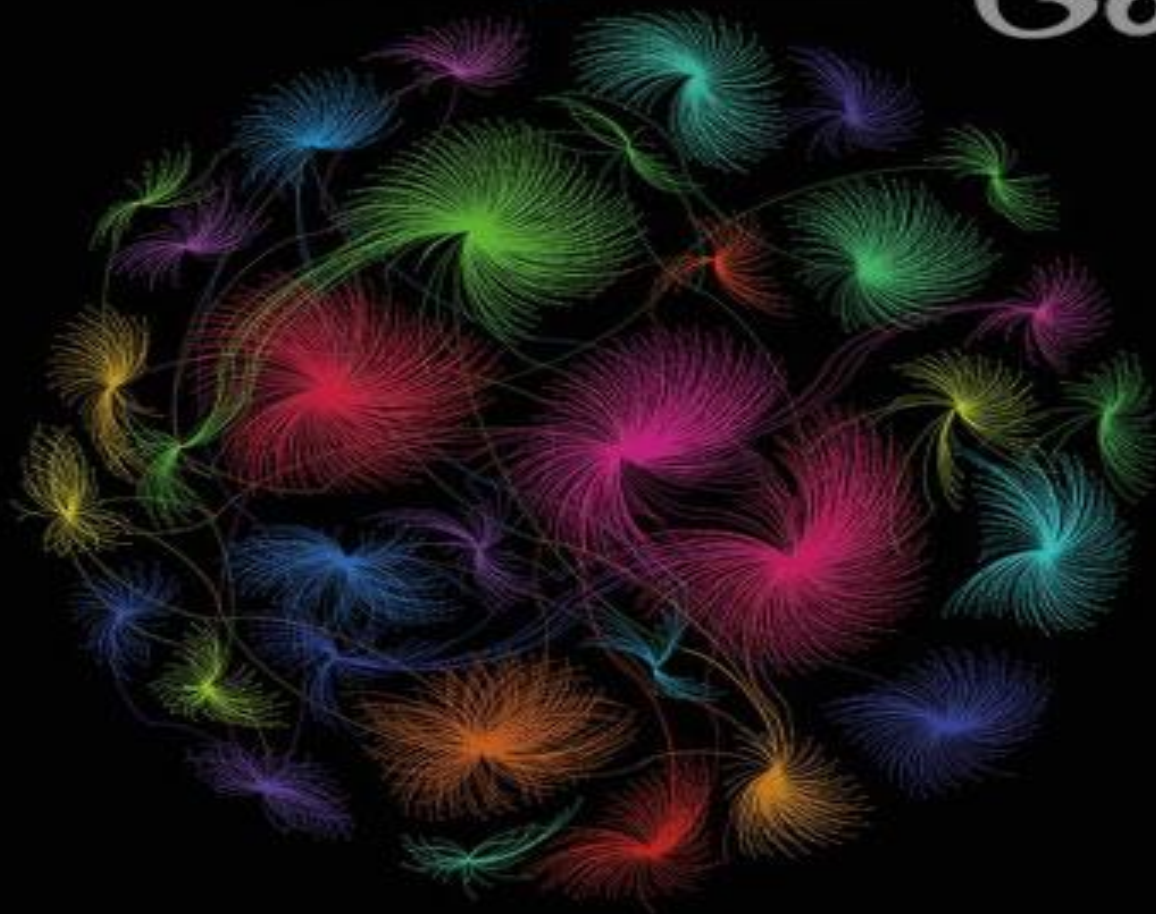
Electoral votes Map



美国大选

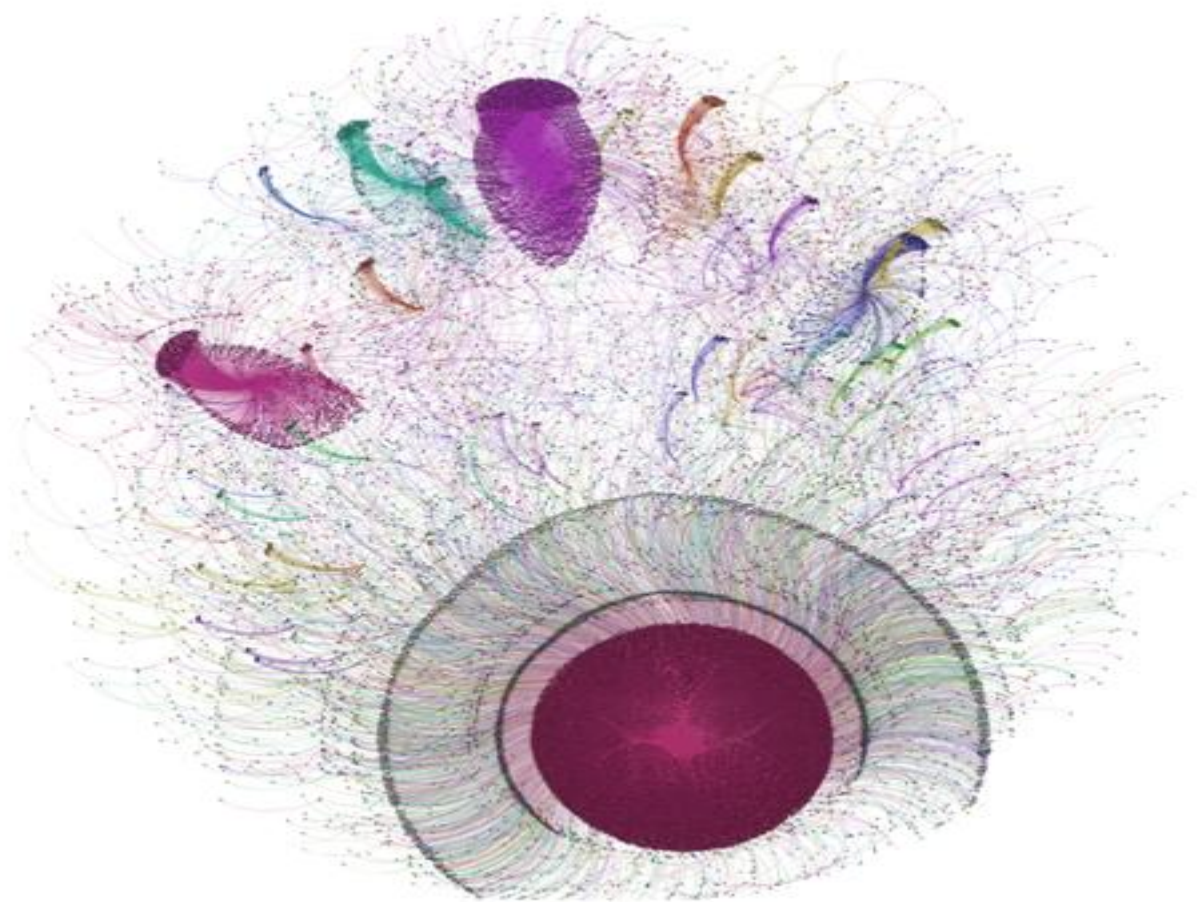
Funding Fountains

36大数据
www.36dsj.com

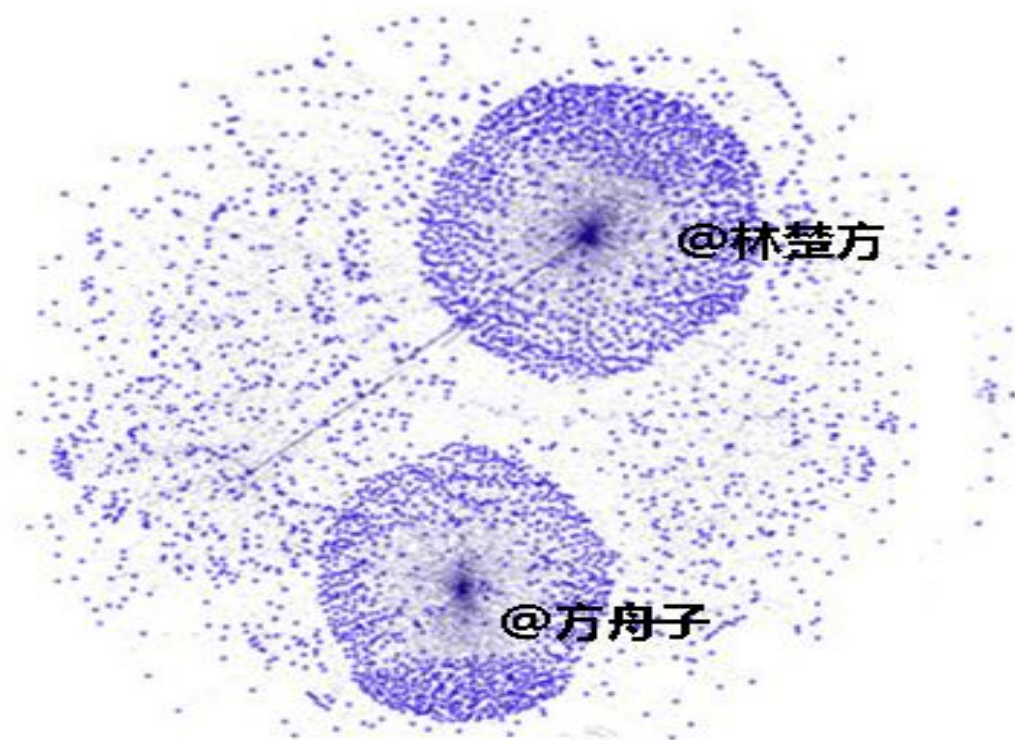


资金喷泉

大型企业之间的资金
走向



蒲公英转发



双星对战

微博转发可视化

如何从数据到可视化效果的例子

1. 输入数据表格

Example Data

– 16 Variables

– 98 Records

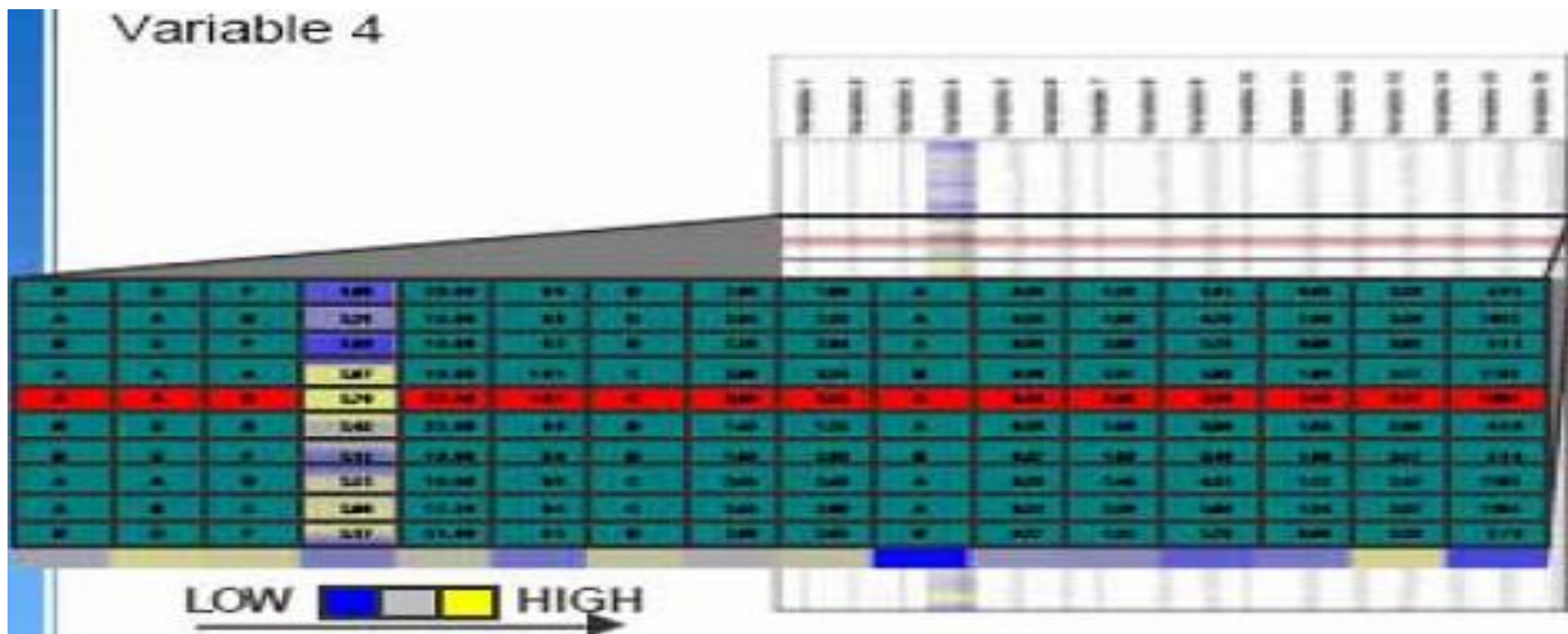
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	

如何从数据到可视化效果的例子

2. 详细的数据表格

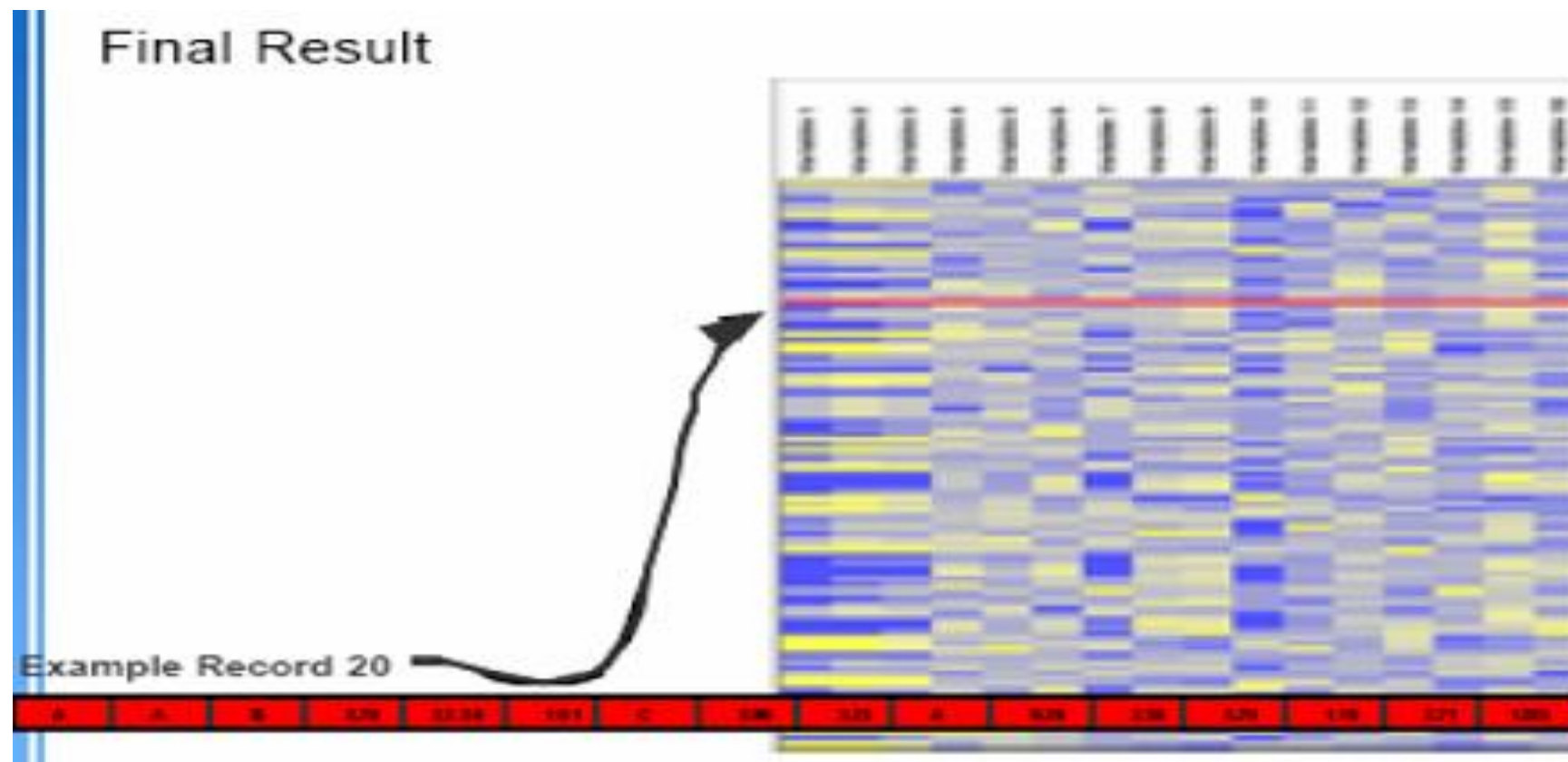
[illegible]

3. 对每个变量要求着色



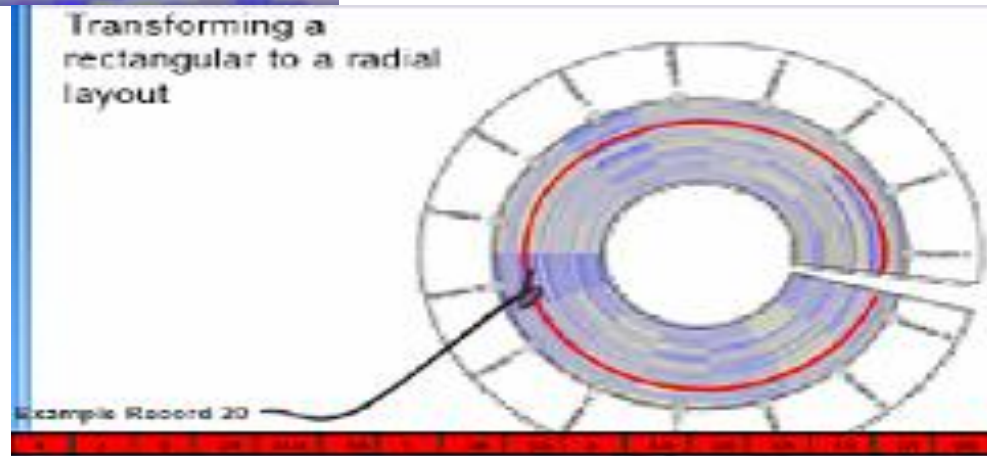
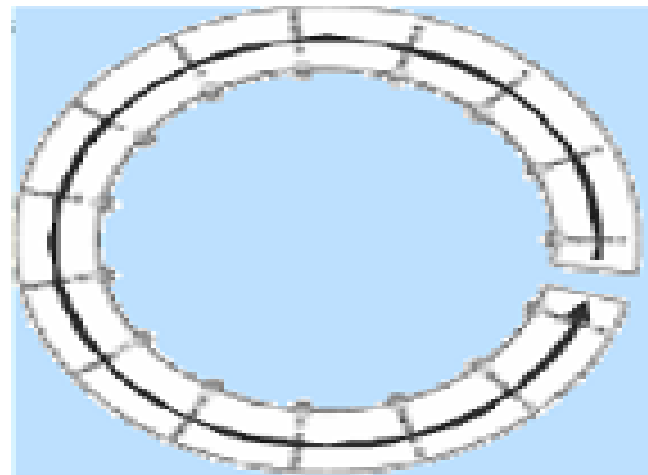
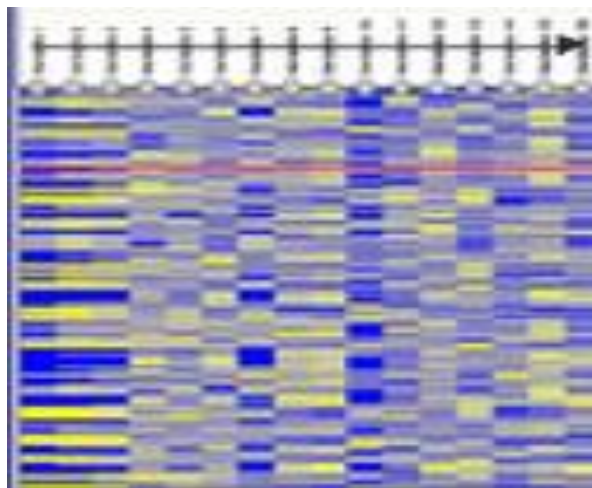
如何从数据到可视化效果的例子

5. 最后的结果



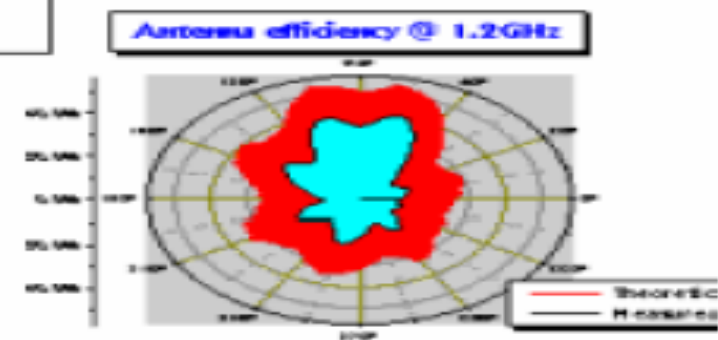
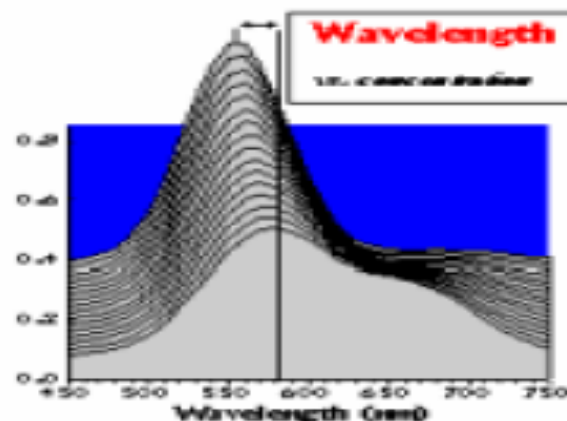
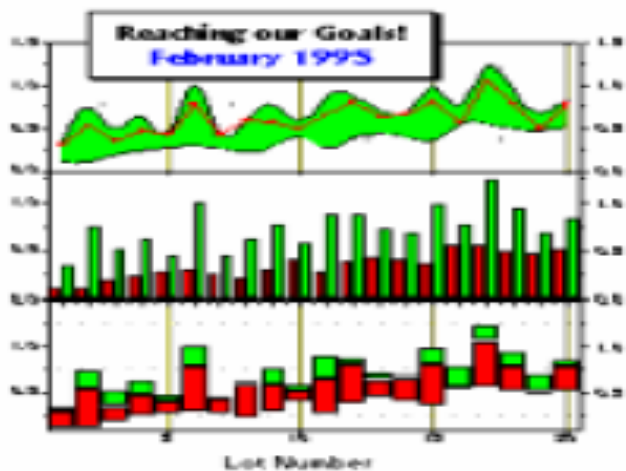
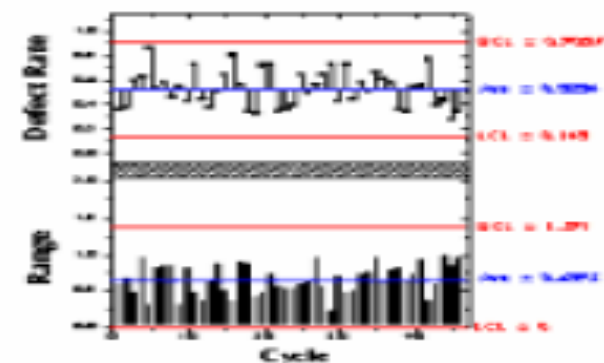
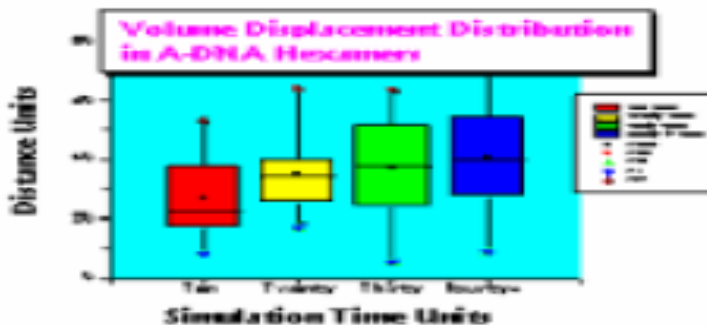
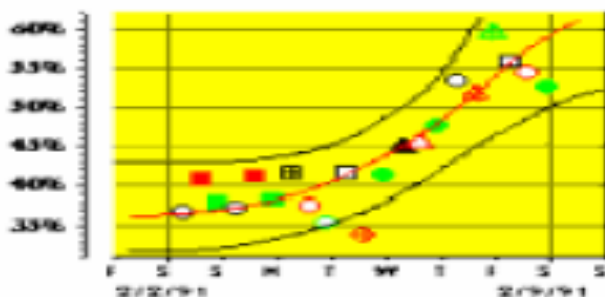
如何从数据到可视化效果的例子

5. 最后的结果



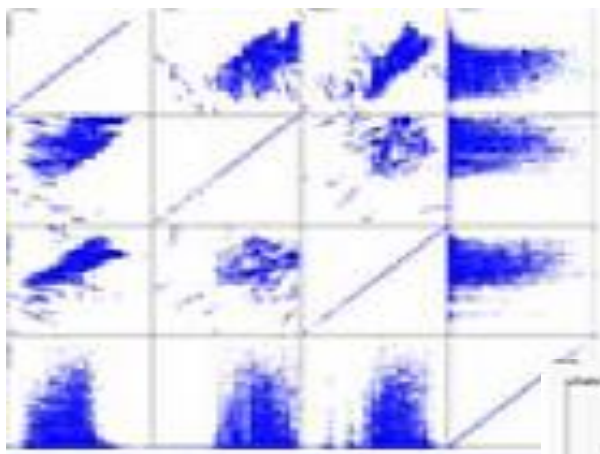
可视化分析技术

- 传统的可视化图表

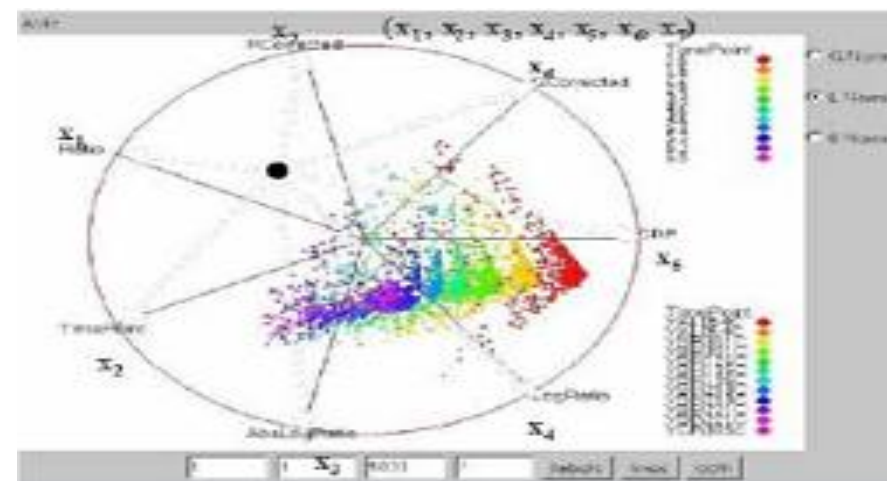


高级可视化技术

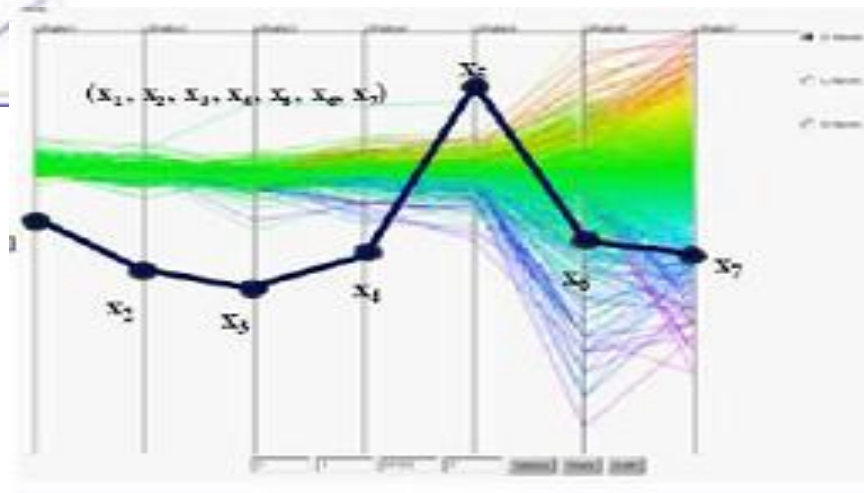
- (1) 多维信息的表达



散点矩阵



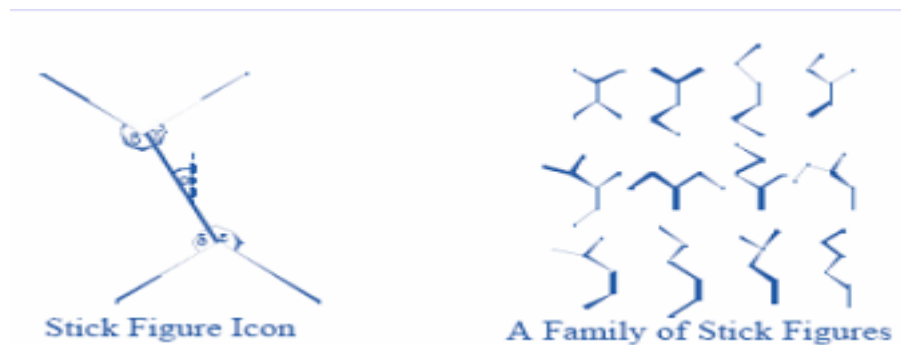
径向图



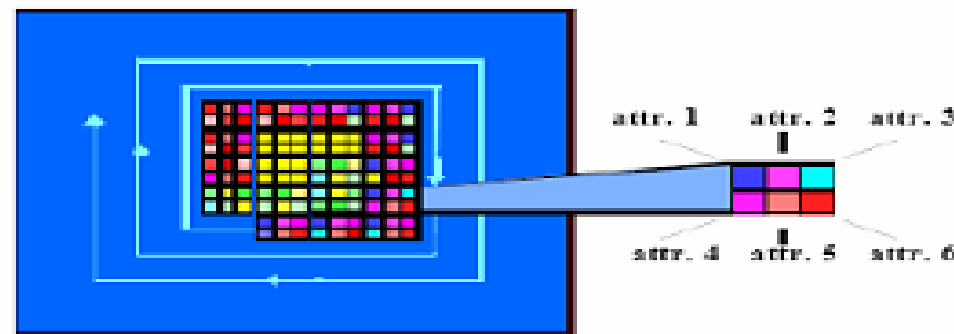
平行坐标

高级可视化技术

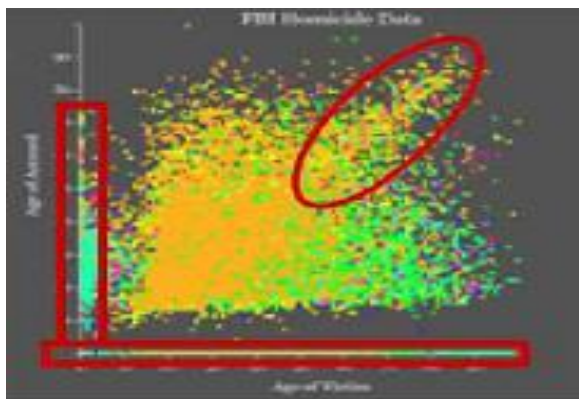
□ (1) 多维信息的表达



图标显示



颜色图标



密集颜色图标



地图图标

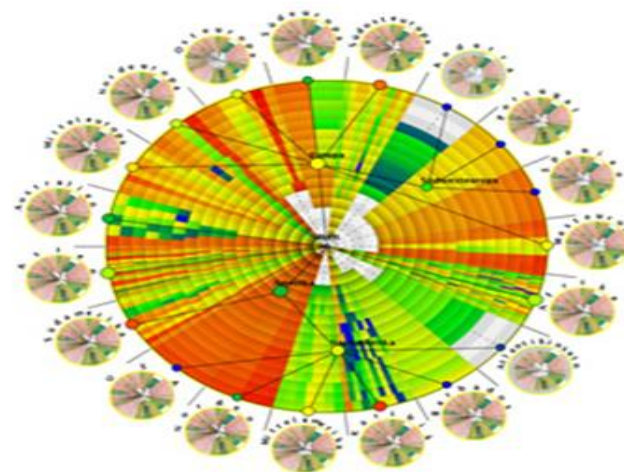
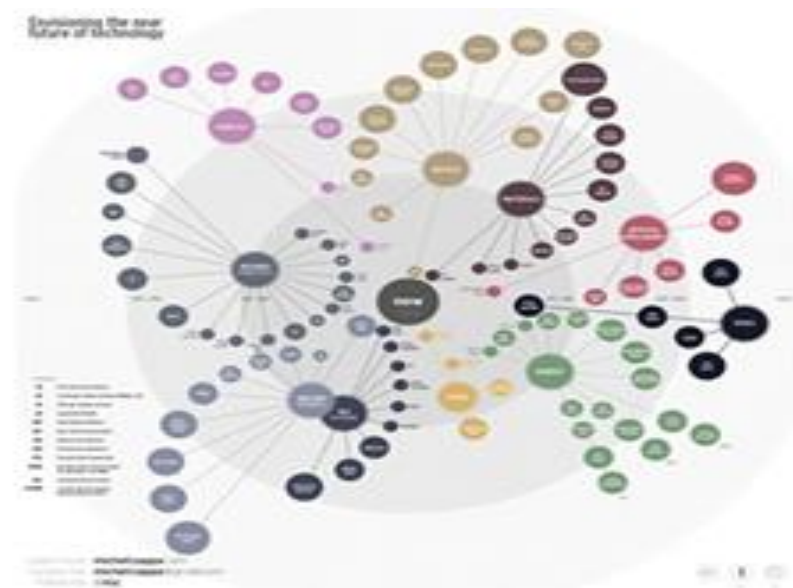
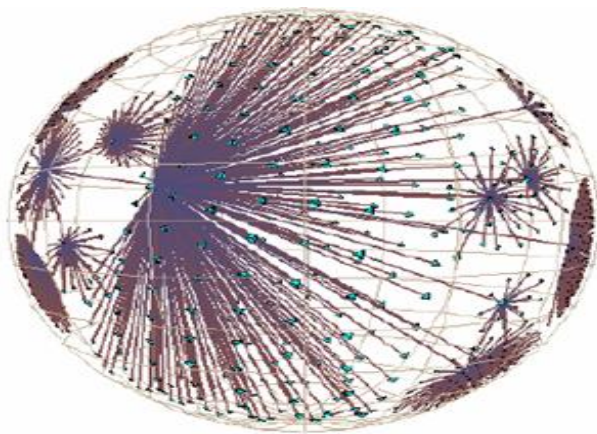
高级可视化技术

□ (2) 层次的视图显示

- 滤波

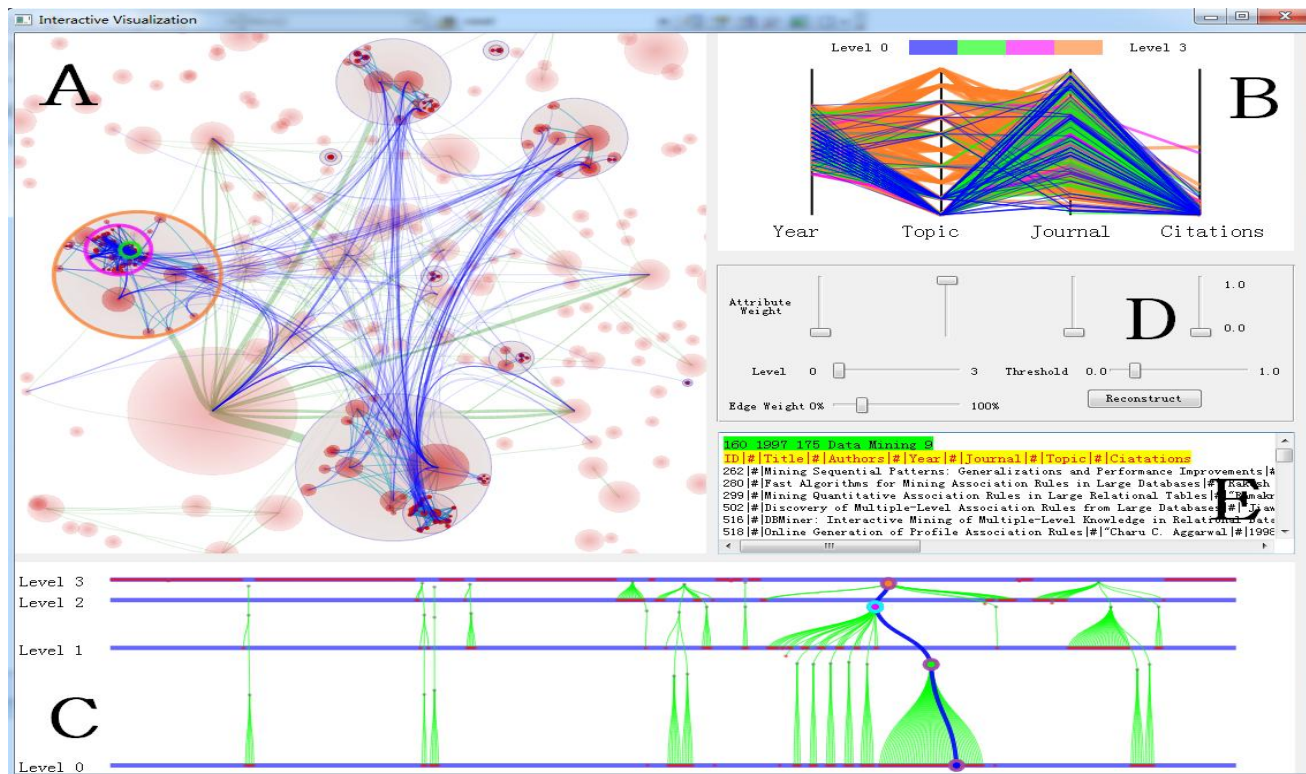
- 扭曲

- 叠加



高级可视化技术

□ (2) 层次的视图显示



全局与局部

高级可视化技术

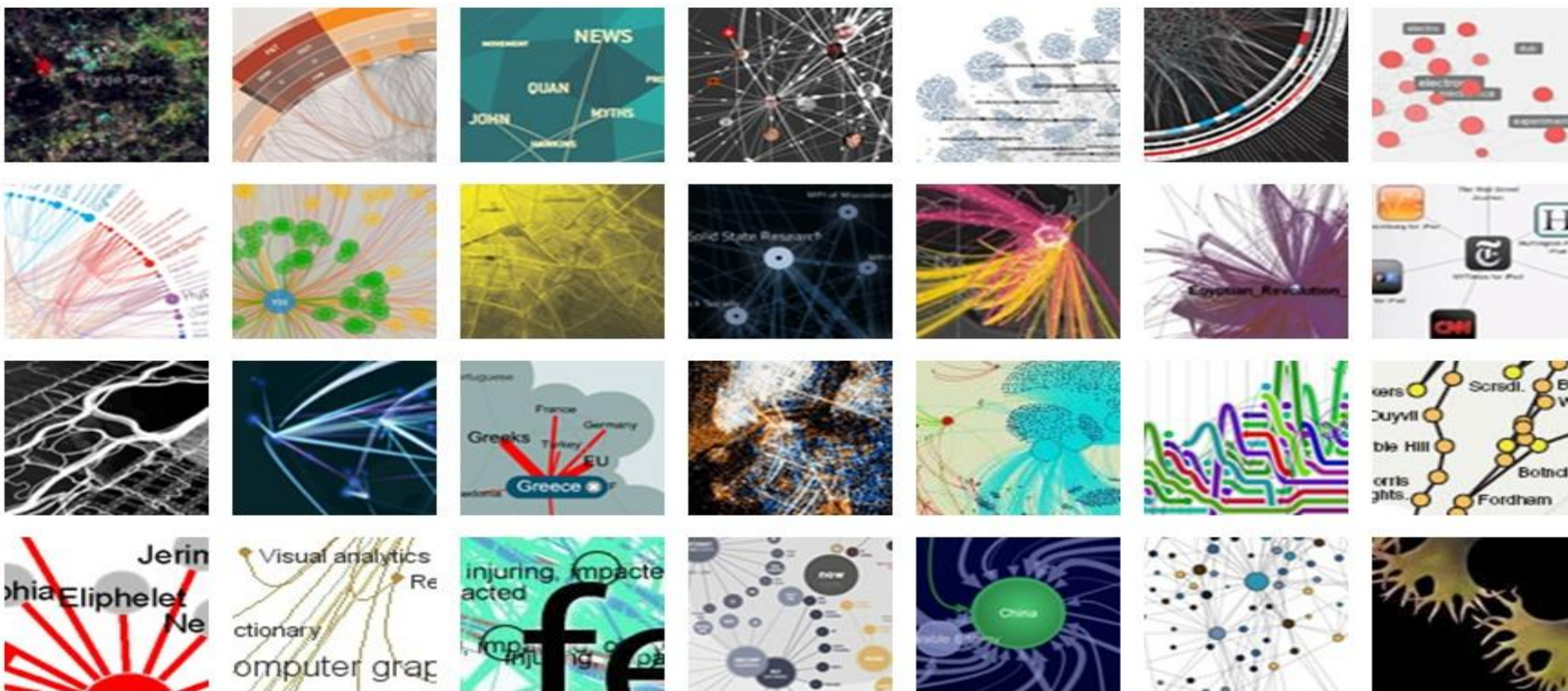
□ (3) 多机的交互协同



大屏幕可视化系统

可视化分析技术

- 新型的可视化方法(<http://www.visualcomplexity.com/vc/>)





三. 分析软件

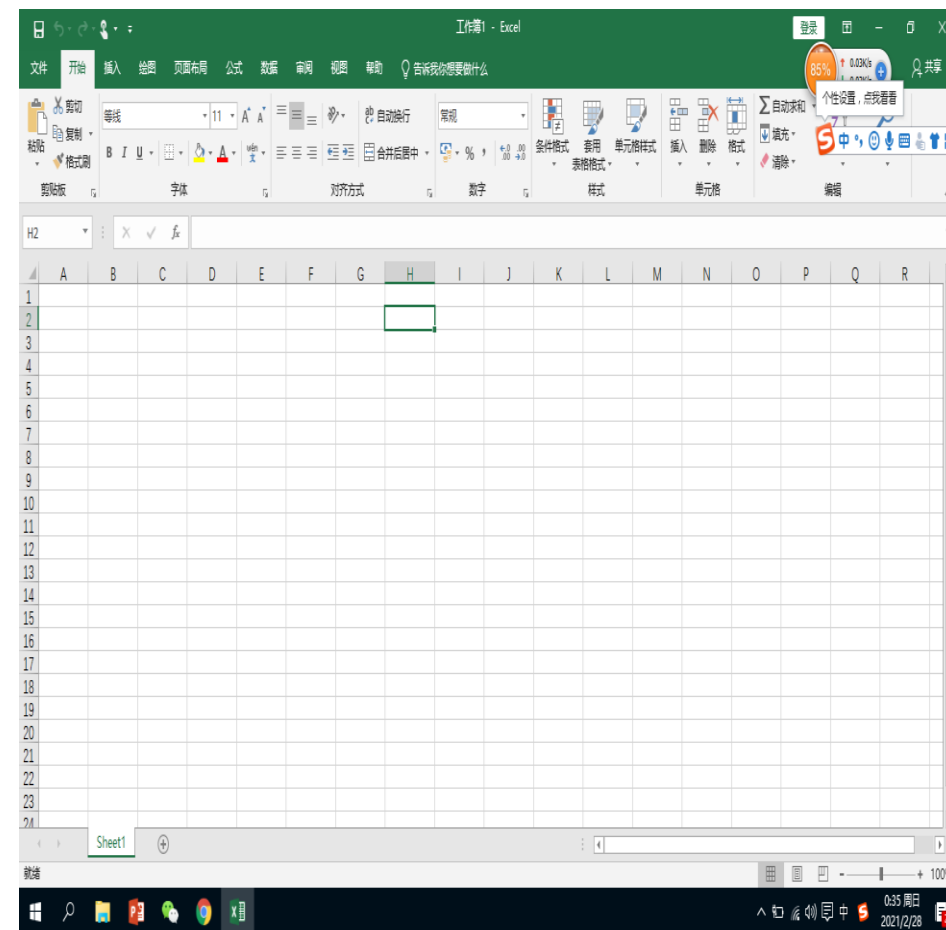
常用数据分析与可视化软件



数据分析软件

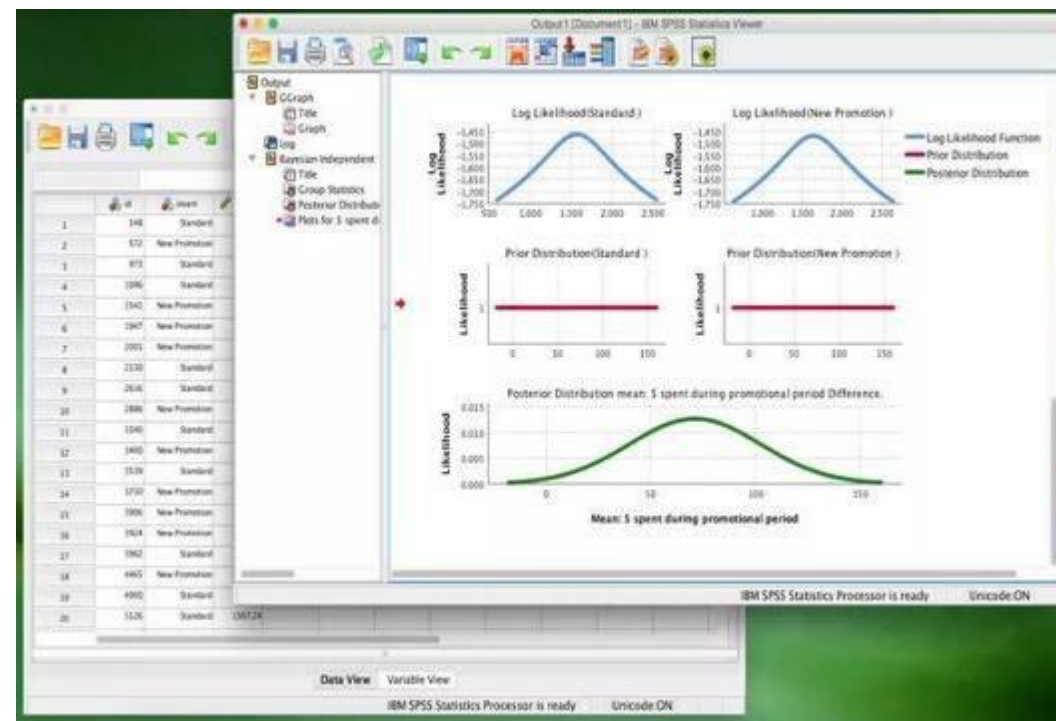
■ Excel图表工具

- Excel满足了绝大部分办公制表的需求，同时也拥有相当优秀的数据处理能力。其自带的ToolPak(分析工具库)和Solver(规划求解加载项)可以完成基本描述统计、方差分析、统计检验、傅立叶分析、线性回归分析和线性规划求解工作。
- Excel也提供较为常用的统计图形绘制功能。这些功能涵盖了基本的统计分析手段，同时也提供相当友好的操作界面，对于具备基本统计学理论的用户来说是十分容易上手的，但处理的数据量较小。



■ SPSS

- SPSS是一个专业的统计分析软件。除了基本的统计分析功能之外，还提供非线性回归、聚类分析(Clustering)、主成份分析(PCA)和基本的时序分析。
- SPSS在某种程度上可以进行简单的数据挖掘工作，比如K-Means聚类，SPSS Modeler的建模功能非常强大且智能化，同时还可以通过其自身的CLEF框架和Java开发新的建模插件，扩展性相当好。



■ R语言

■R 是一个有着统计分析功能及强大作图功能的软件系统，是由奥克兰大学统计学系的Ross 和 Robert共同创立。

■R是主要用于统计分析、绘图的语言，通过对R语言的学习，将有助于未来的生物信息领域的学习。

■注意！R语言适合统计分析，不适合做文本分析，太慢！



数据分析软件

■ python语言

■ Python是一种面向对象、解释型计算机程序设计语言。主要优点如下：

- ◆简单：Python是一种代表简单主义思想的语言。
- ◆易学：Python极其容易上手，因为Python有极其简单的说明文档。
- ◆速度快：运行速度非常快。
- ◆免费、开源：Python是FLOSS（自由/开放源码软件）之一。
- ◆面向对象：Python既支持面向过程的编程也支持面向对象的编程。
- ◆可扩展性、可嵌入性以及丰富的函数库。

数据分析软件



Tableau是用于可视分析数据的商业智能工具。用户可以创建和分发交互式 and 可共享的仪表板，以图形和图表的形式描绘数据的趋势，变化和密度。该软件允许数据混合和实时协作。

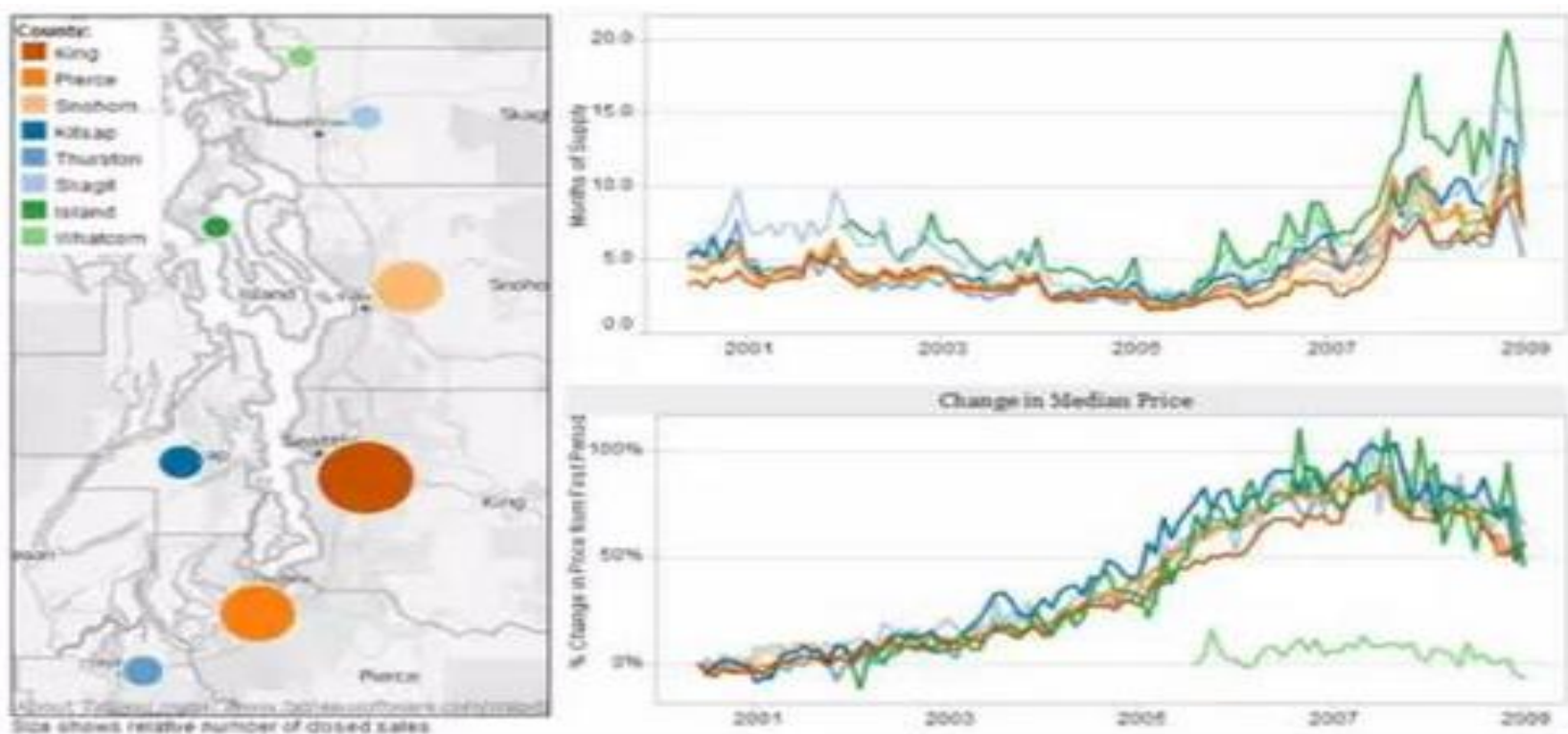
SAS (Statistical Analysis System)是一个模块化、集成化的大型应用软件系统。它由数十个专用模块构成，功能包括数据访问、数据储存及管理、应用开发、图形处理、数据分析、报告编制、运筹学方法、计量经济学与预测等等。



数据分析软件

□ Tableau

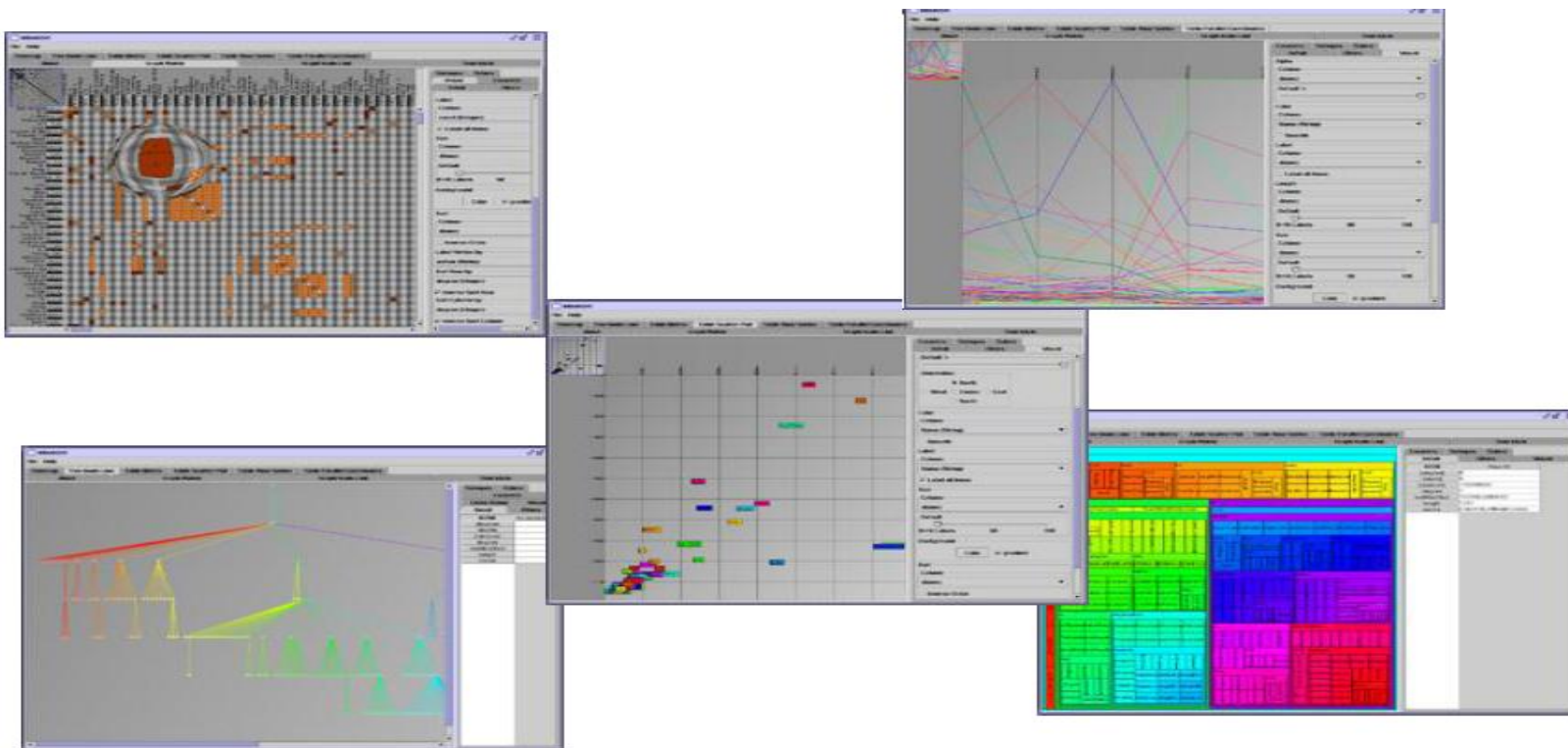
- 对数据做深入分析，不需要编程
- 将数据上传到服务器；可直接分享到社交媒体



数据分析软件

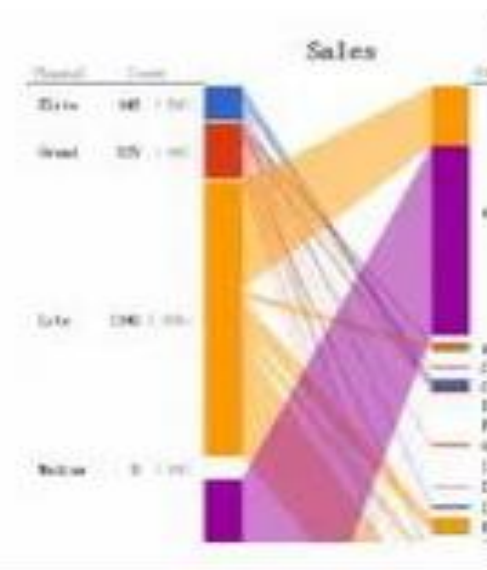
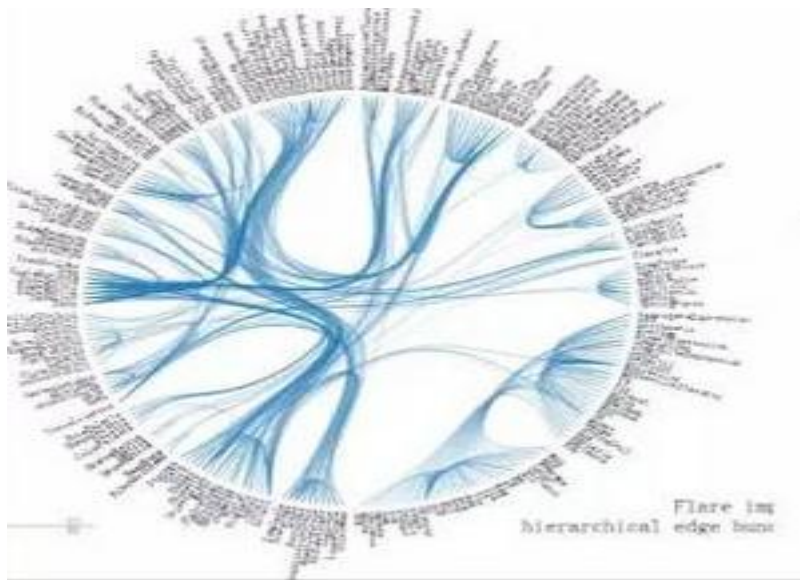
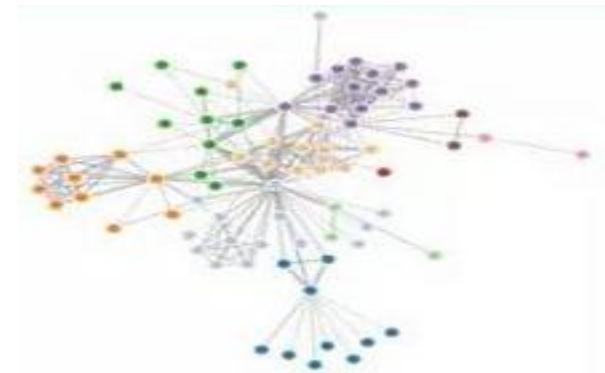
□ InfoVis工具包

- Java组件集，颜色管理，表姐，动态查询等附加功能
- 多视图：时序图、平行坐标、散点图、树图等



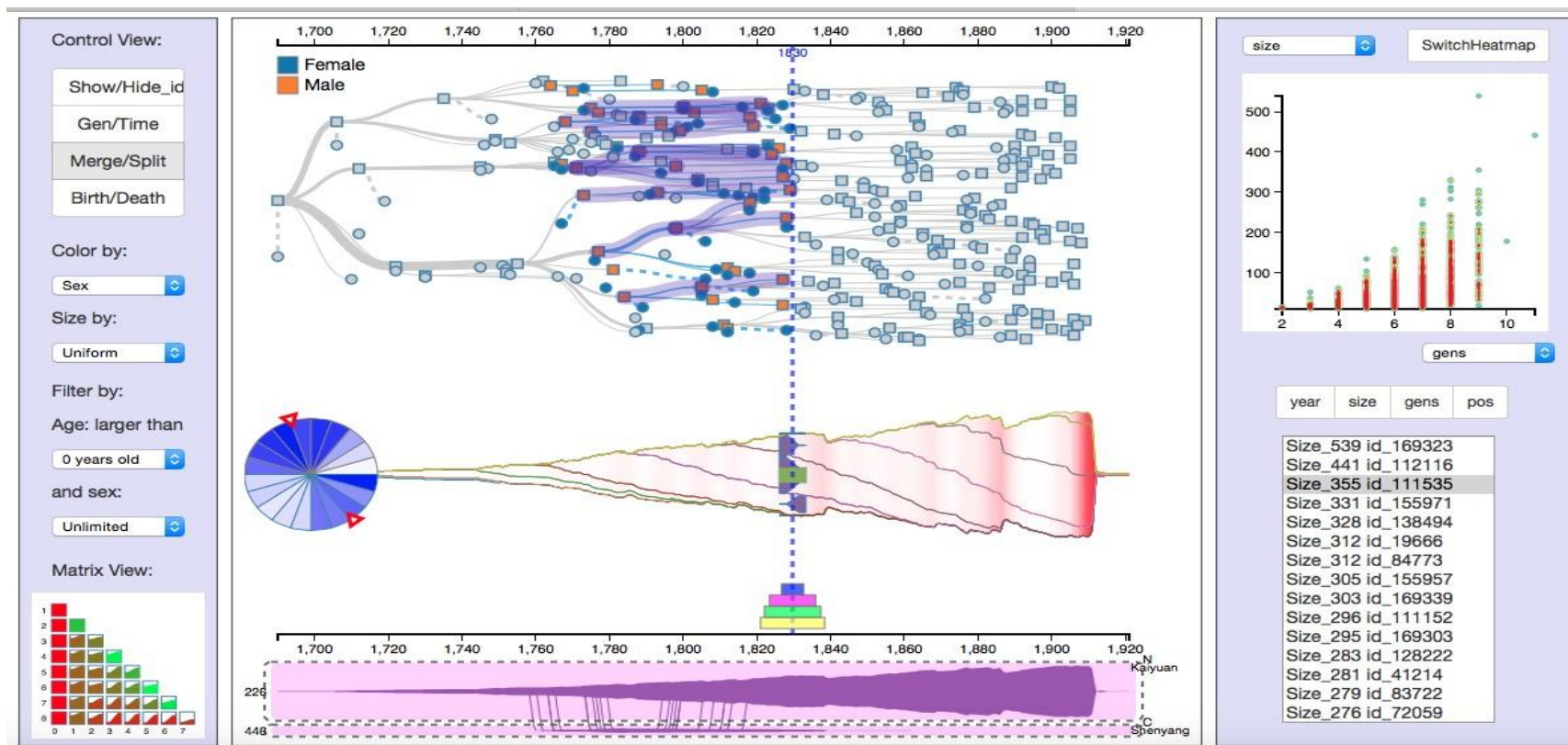
数据分析软件

- D3.js:
 - 在线可视化的工具
 - 使用HTML、SVG和CSS生成图表效果
 - 可视化组件与数据驱动决策结合



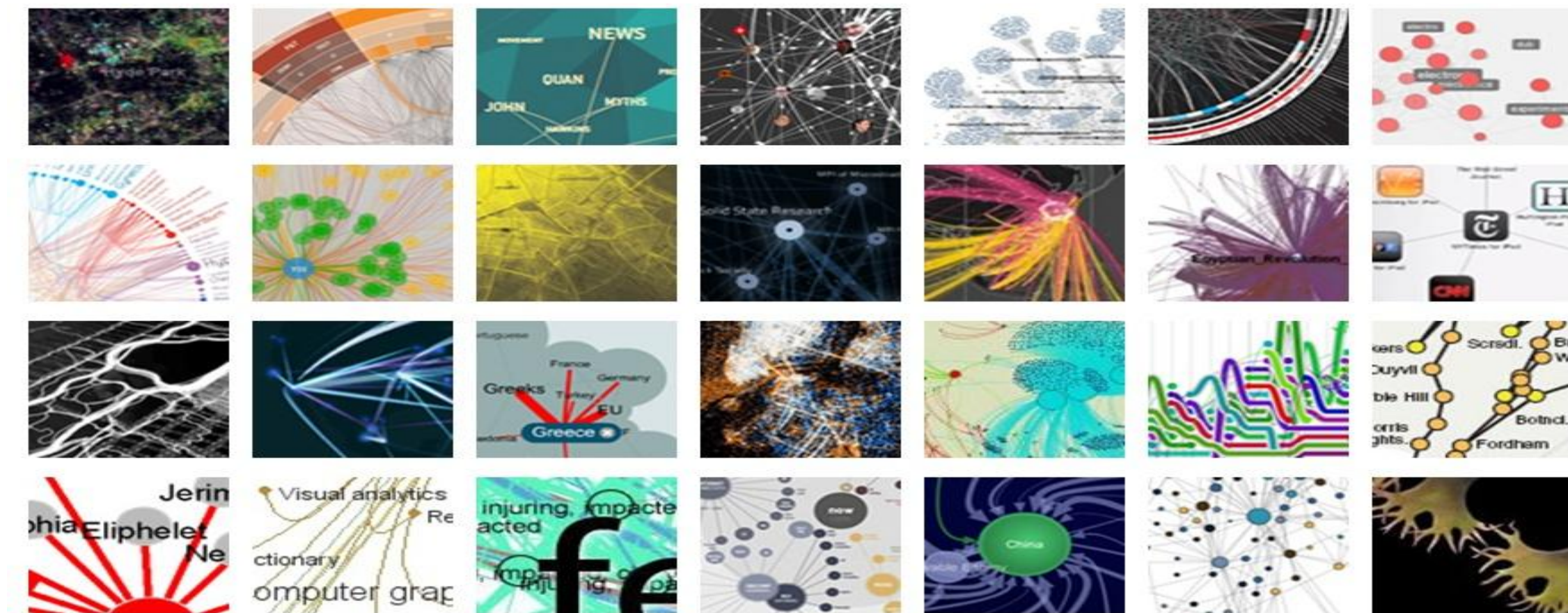
数据分析软件

- 综合定制的数据分析系统
 - ⊙ 个性化的分析方法/人机交互



数据分析软件

其他示例: <http://www.visualcomplexity.com/vc/>



开放式讨论

- 谈谈你接触或想到的一个信息可视化需求？
- 介绍你认为好的数据可视化效果或系统？
- 结合你熟悉的某个应用谈谈大数据可视分析体会。

THANKS

王长波

E-mail: cbwang@dase.ecnu.edu.cn

Phone: 021-62235578

华东师范大学数据科学与工程学院



END