数据可视化科普

GeekPlux

Node Party

17.8.19

GeekPlux

Xiang Gao
geekplux

Mediocre/Freelancer/Full Stack Developer, Full-time Learner. Student with Google Summer of Code 2017 **@freifunk**

Hire me
Hangzhou, China
geekplux@gmail.com
http://geekplux.com

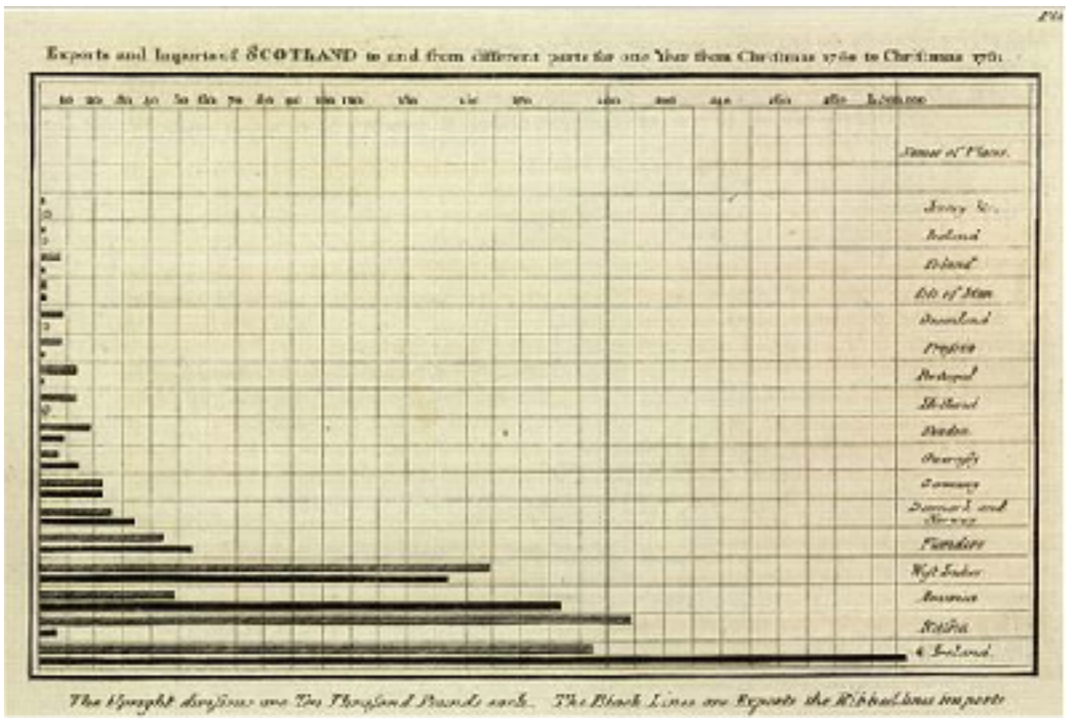
- 研究生 + 程序员
- 数据可视化 + 增强现实
- http://geekplux.com
- GitHub/geekplux

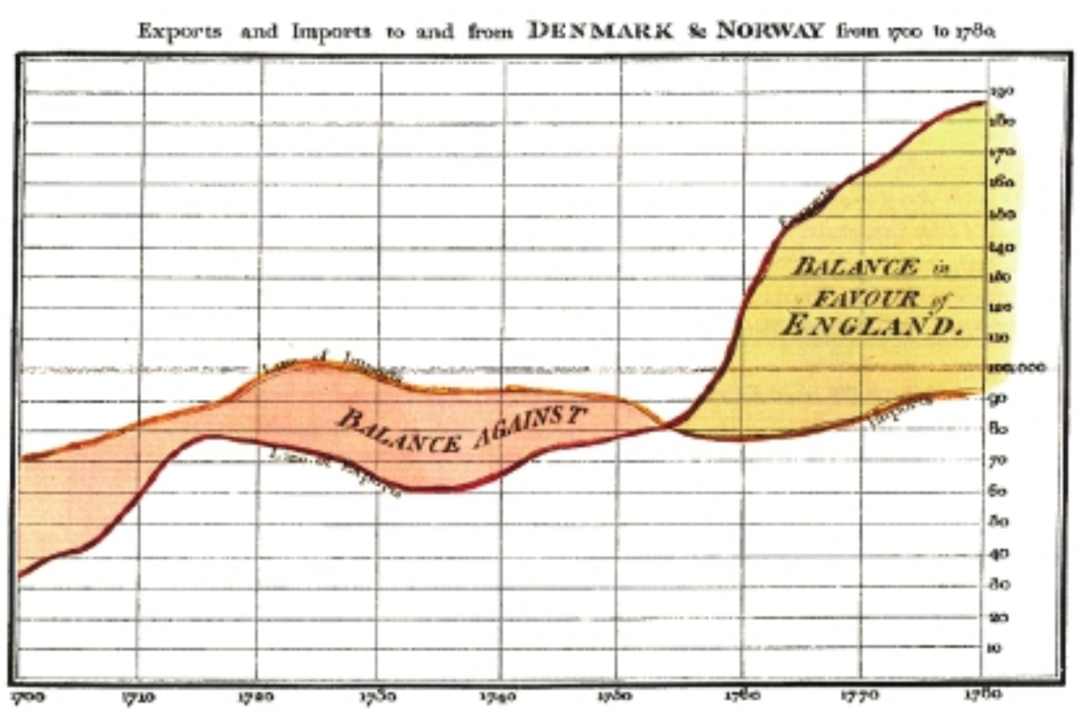
数据可视化

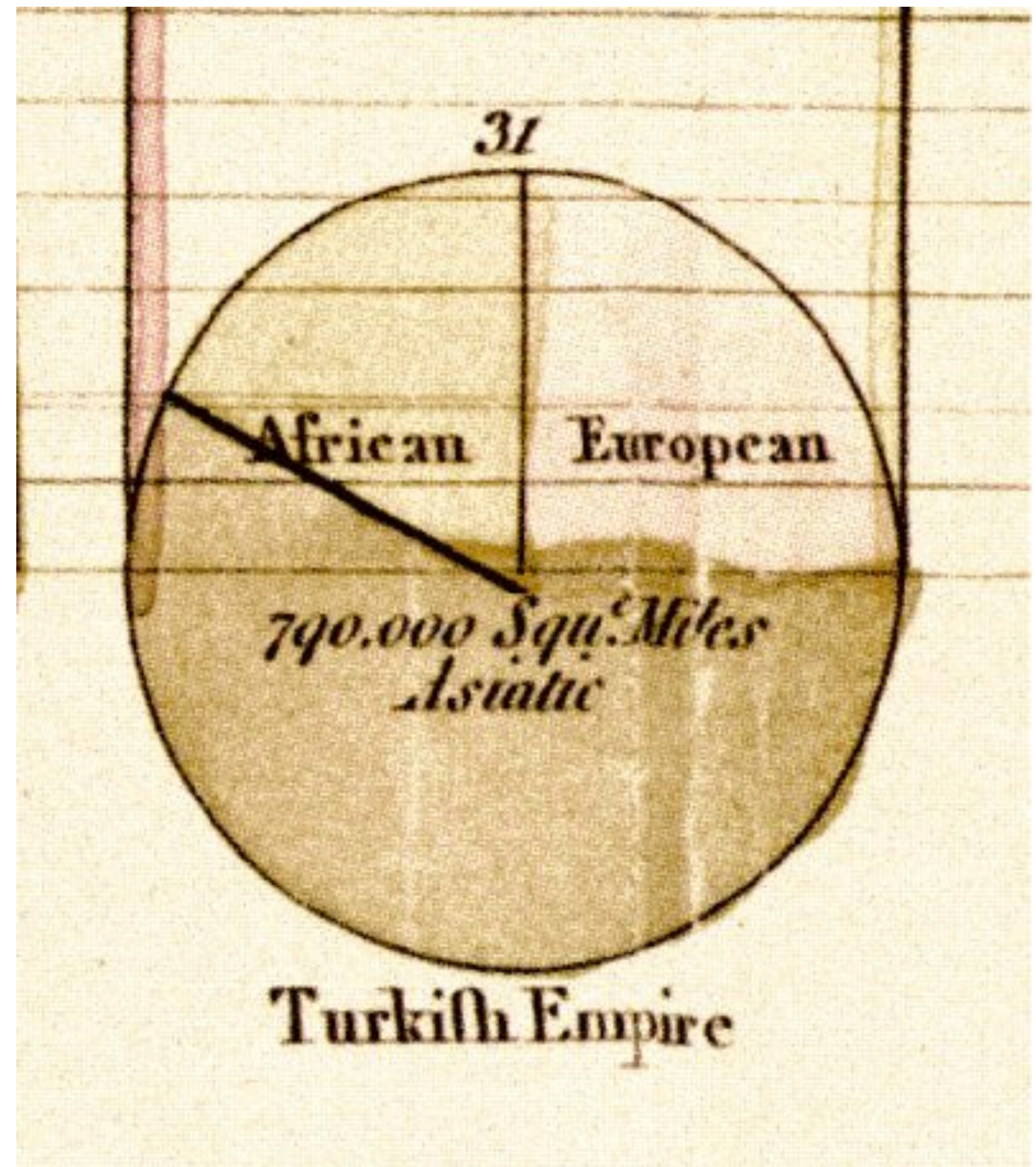
- 什么是数据可视化
- 数据可视化的分类
- 数据可视化的热点和趋势
- 数据可视化基础、工具

什么是数据可视化

“数据可视化不就是条形图、折线图？”



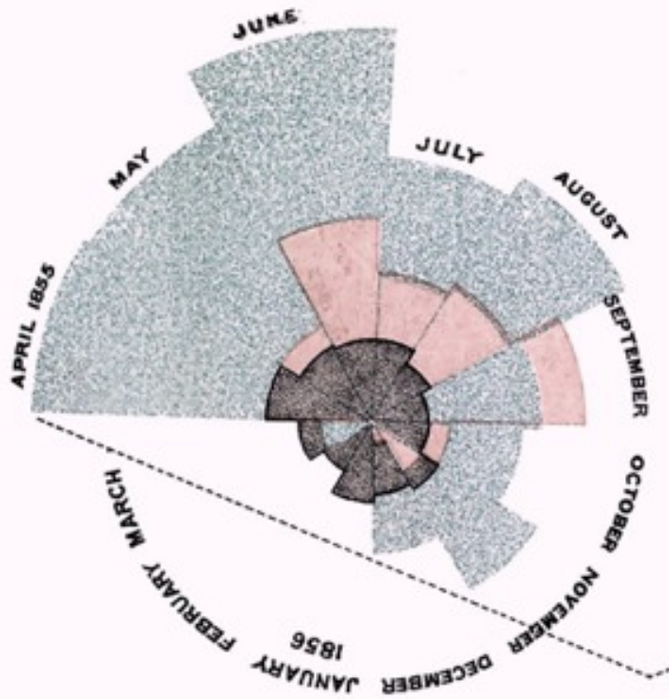

Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.
BALANCE AGAINST
BALANCE in FAVOUR of ENGLAND.


31
African
European
790.000 Squ. Miles
Asiatic
Turkish Empire

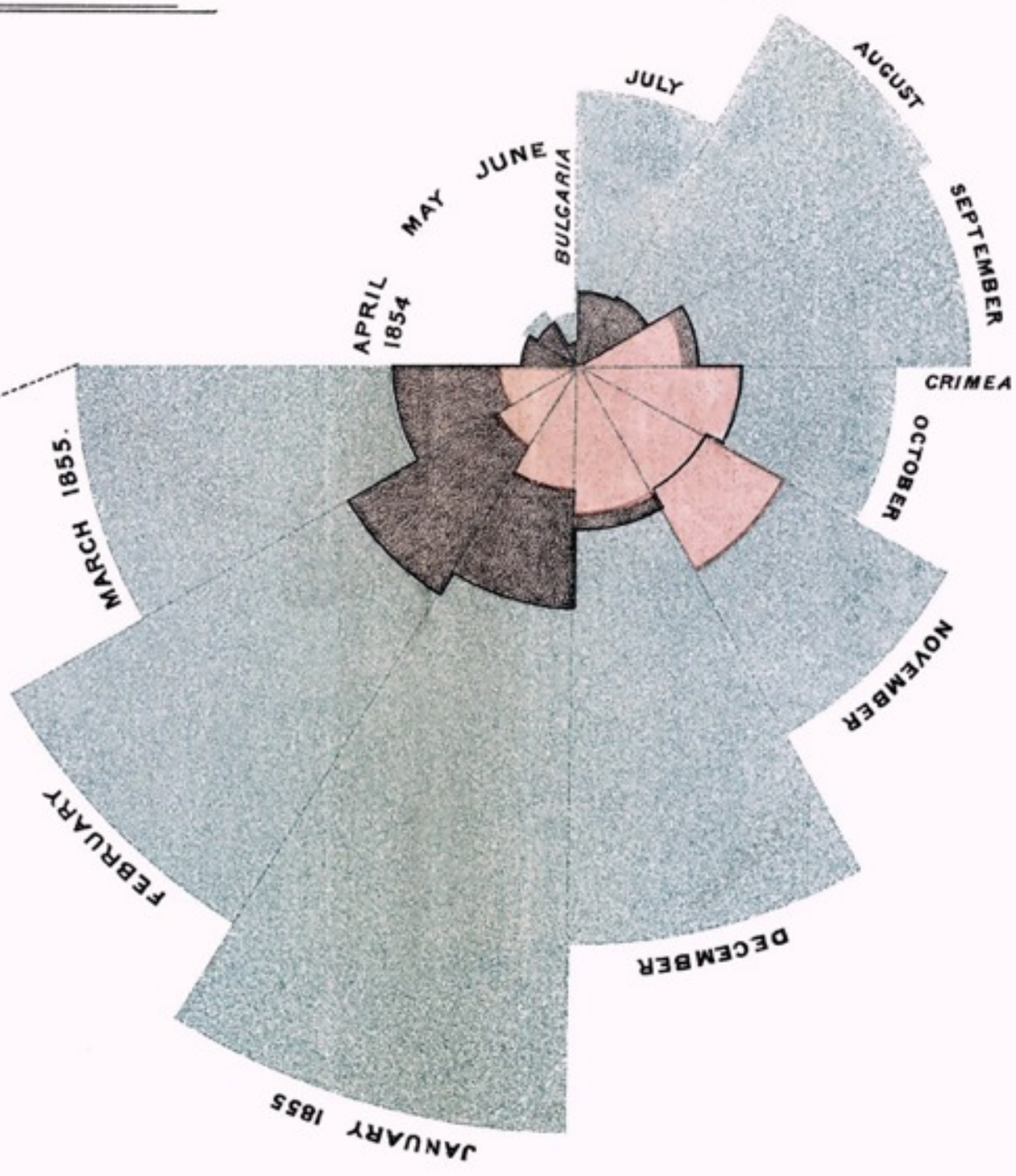
数据可视化可以更丰富

DIAGRAM of the CAUSES of MORTALITY
IN THE ARMY IN THE EAST.

2.
APRIL 1855 to MARCH 1856.



1.
APRIL 1854 to MARCH 1855.



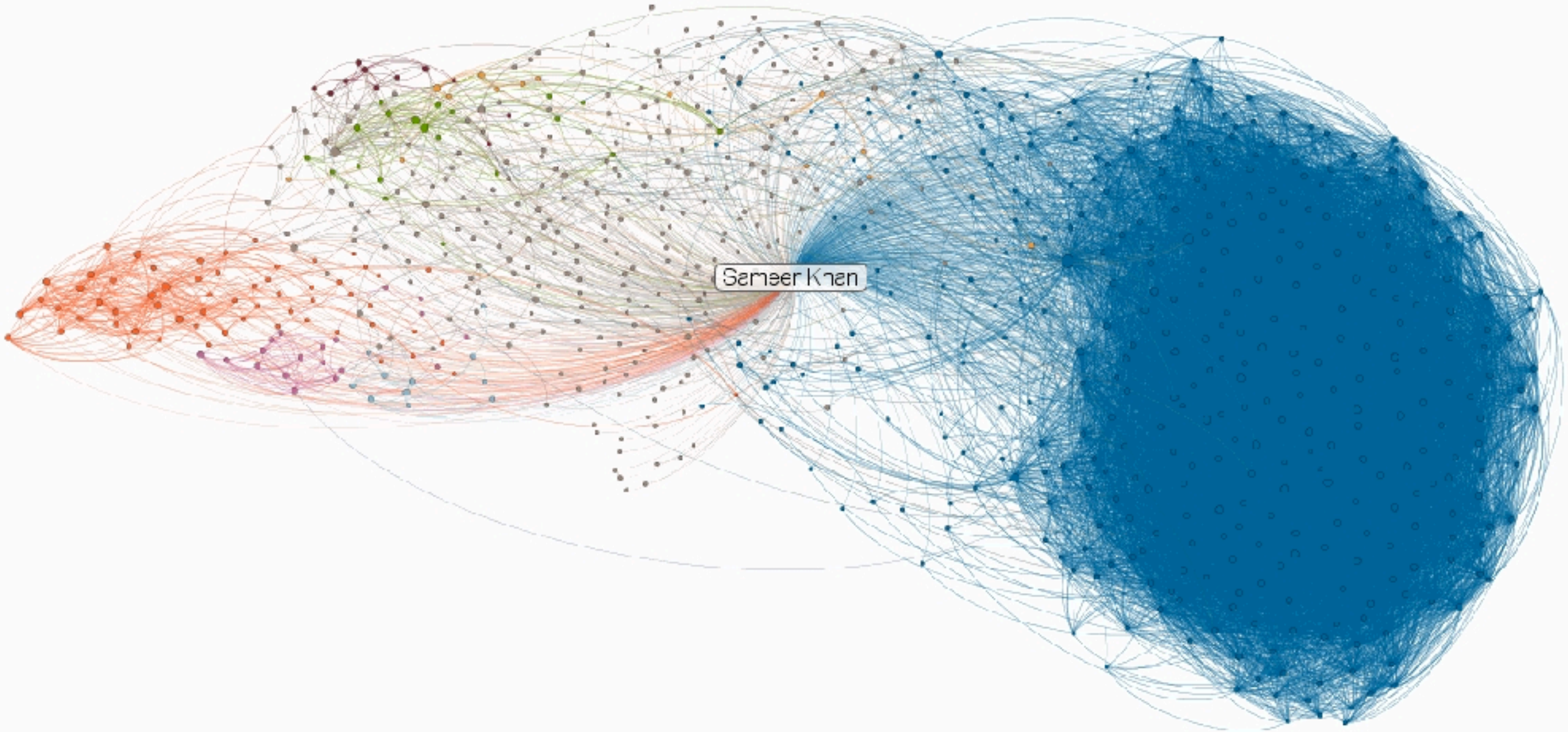
The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.
The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes.
The black line across the red triangle in Nov.r 1854 marks the boundary of the deaths from all other causes during the month.
In October 1854, & April 1855, the black area coincides with the red; in January & February 1856, the blue coincides with the black.
The entire areas may be compared by following the blue, the red & the black lines enclosing them.


facebook
December 2010

LinkedIn Maps

Sameer Khan's Professional Network

as of July 11, 2013



Presidential Election Results: Donald J. Trump Wins

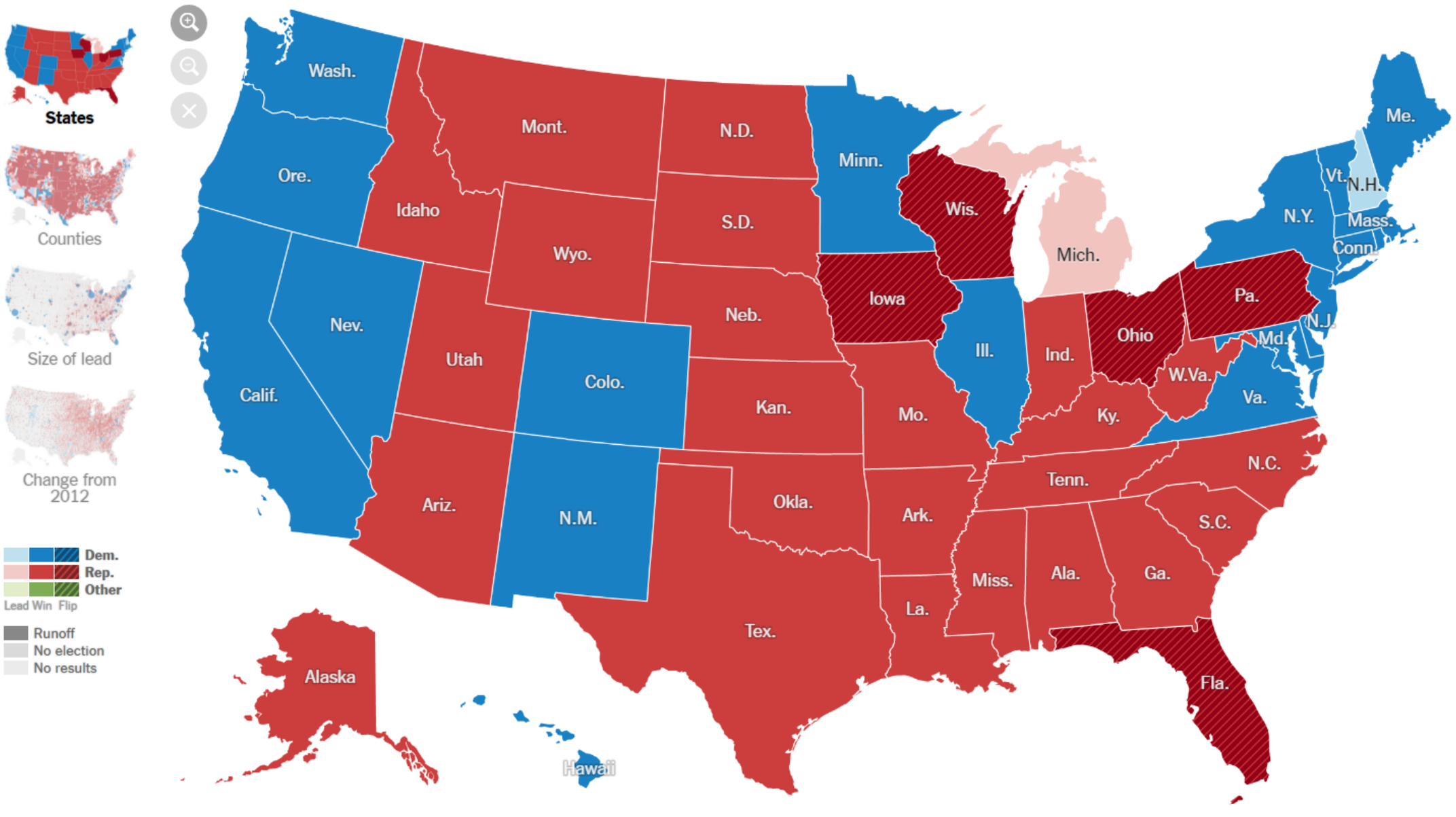
NOV. 11, 2016, 10:55 PM ET

Here's a look at how Trump reshaped the election map. As of Thursday morning, Michigan and New Hampshire are still too close to call. This page will be updated when a winner is called in those states.

228 Hillary Clinton — 60,605,174 votes (47.7%)

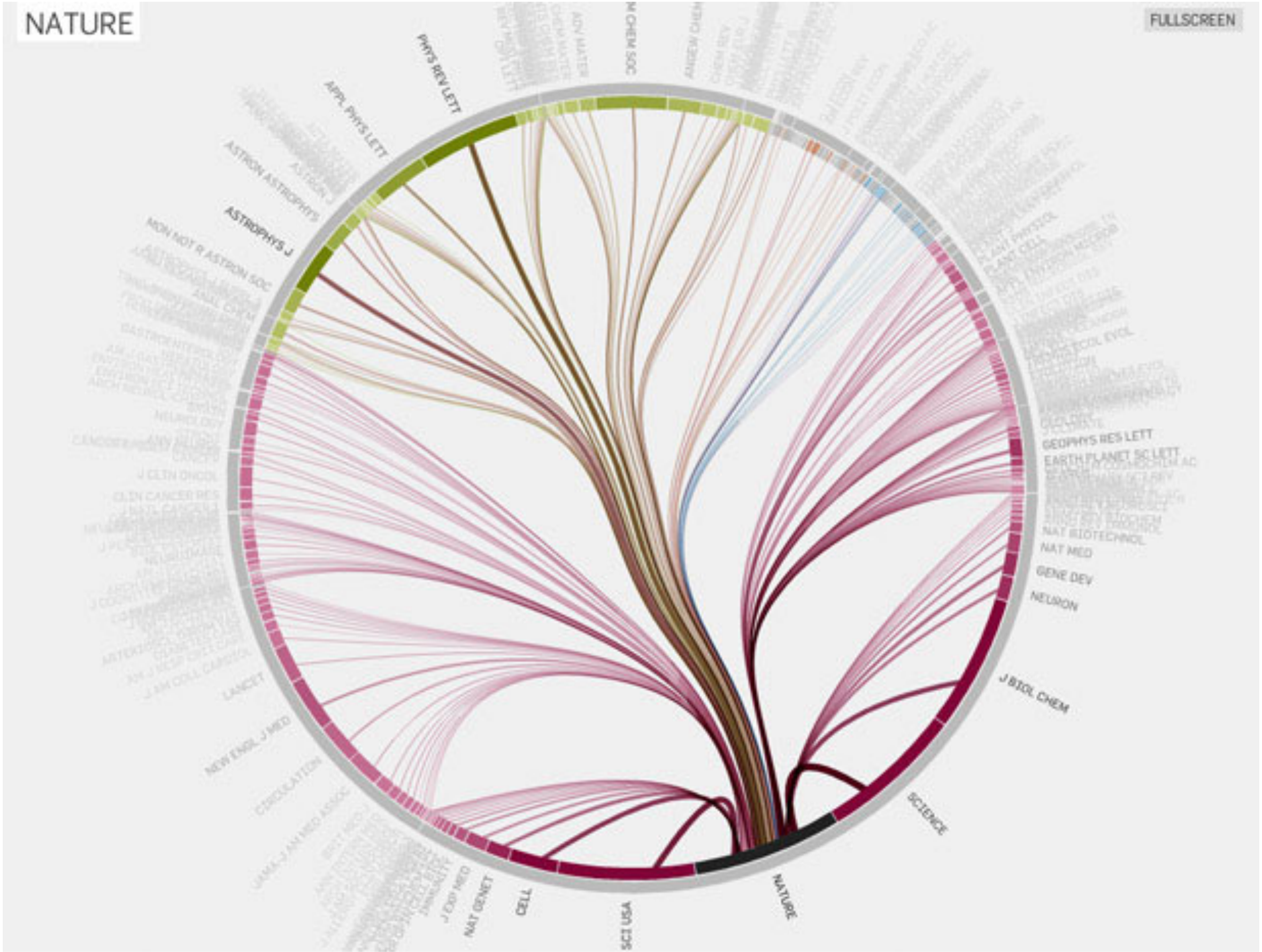
270 to win

Donald J. Trump **290** — 60,144,993 votes (47.4%)



NATURE

FULLSCREEN



双11全球狂欢节

2015 11.11 GLOBAL SHOPPING FESTIVAL

24:00:00

¥91,217,017,615

$14,341,847,366

GMV[1] Generated in US$ of 2015 11.11

无线成交占比：68%

Mobile GMV Percentage

（初步数据，未经审计。All data on the screen are preliminary and unaudited.）

全球热卖榜

Global Top Selling

手机 | 女装 | 男装 | 鞋子 | 运动户外

	品牌销售额排行	品牌销售量排行	商家销售额排行
1	小米	小米	小米
2	华为	华为	华为
3	苹果	苹果	苹果
4	锤子科技	锤子科技	锤子科技
5	魅族	魅族	魅族

单品销售排行 1 IPHONE6S 2 IPHONE6S 3 IPHONE6S

全国区域经济

Nationwide Consumption Patterns

湖北省

2015

2014

10:05

消费省/地区排行

1 北京 Beijing
2 北京 Beijing
3 北京 Beijing
4 北京 Beijing
5 北京 Beijing

消费县排行

1 江苏常州 Changzhou,Jiangsu
2 江苏常州 Changzhou,Jiangsu
3 江苏常州 Changzhou,Jiangsu
4 江苏常州 Changzhou,Jiangsu
5 江苏常州 Changzhou,Jiangsu

成交额(亿)

802
642
482
321
161

11:49:09 ¥571亿* (≈87.1 billion)

07:45:42 ¥65.6亿* (≈9.98 billion)

00:12:28 ¥100亿 (≈15 billion)

00:01:12 ¥10亿 (≈1 billion)

00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

菜鸟网络 – 仓配协同

Global Trading Trends

累计物流订单量

467,764,069

累计取用电子面单量

120,576,789

617 中转中心数量

176,000 网点数量

北京市 应收件量：89,945

正常 47

轻度繁忙 74

繁忙 99

全球交易热度

Global Transactions Patterns

232 个 已成交国家/地区总数

[1] All GMV referenced is settled through Alipay.

[2] 2014 11.11 GMV.

[3] 据comScore数据，2014年美国感恩节网购季（包括感恩节、黑色星期五、感恩节周末、网络星期一）PC成交额为约1.4亿美元。According to comScore, the total desktop GMV of 2014 Thanksgiving through Cyber Monday (Thanksgiving, Black Friday, Thanksgiving weekend, Cyber Monday) is US$ 6.56 billion.

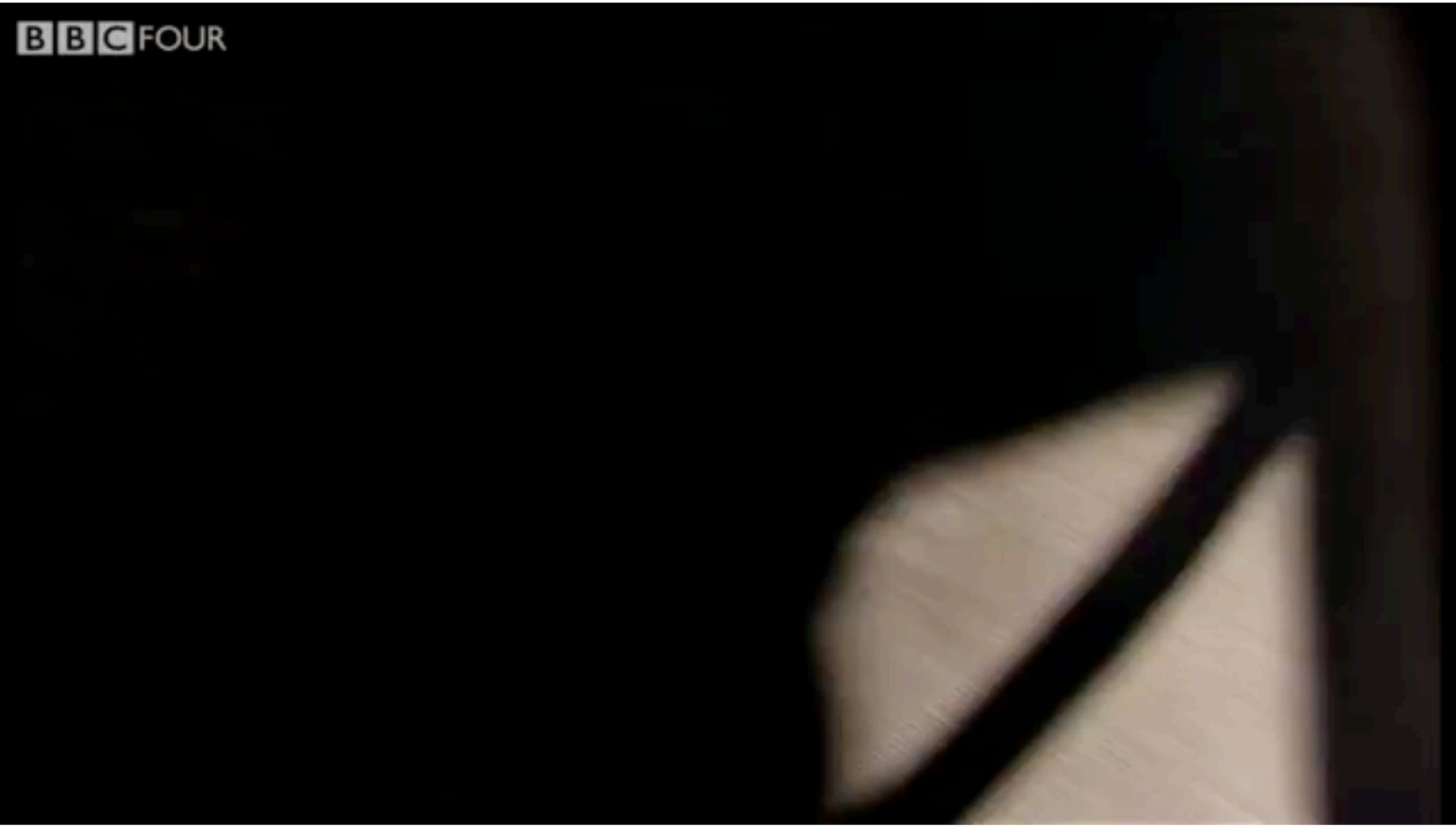
淘宝头条

…双十一50亿提前10分钟。

【new】截至到11日0：20分，天猫双十一成交额超过50亿，比2014年双十一…

“Data visualization is the creation and study of the visual representation of data”

–wikipedia



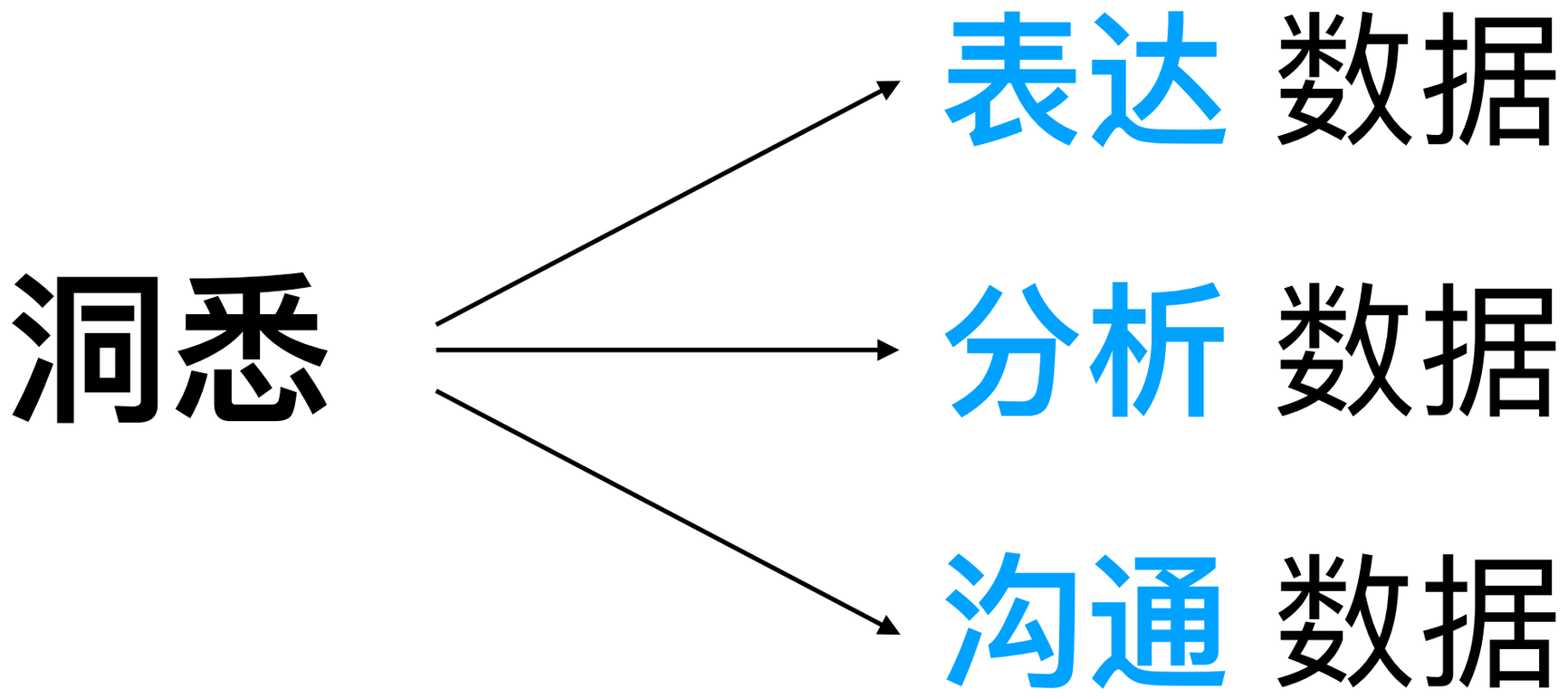
Hans Rosling's 200 Countries, 200 Years, 4 Minutes - The Joy of Stats - BBC Four

数据可视化

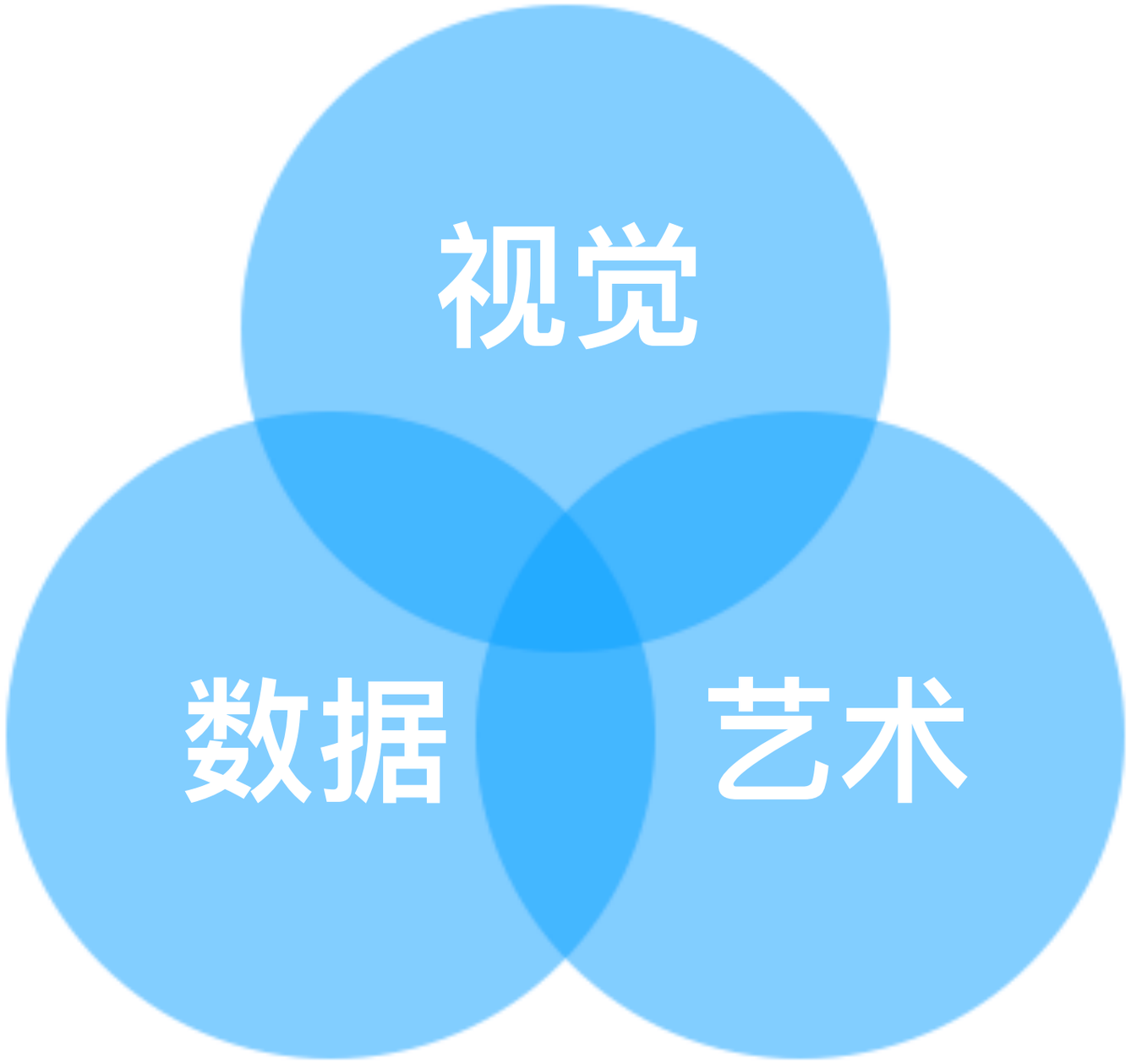
数据 → **视觉表达**

使我们更好地洞悉数据

数据可视化



三大元素

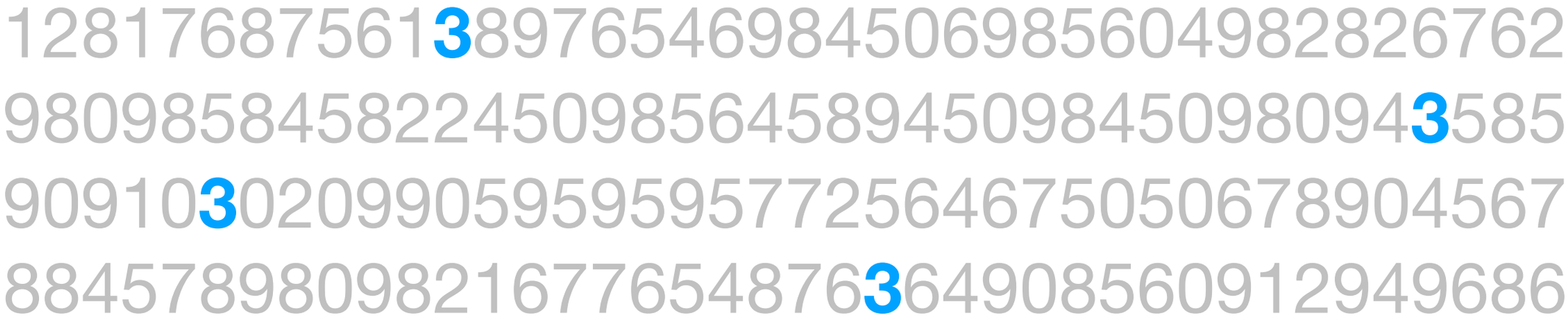


视觉与感知

图中有多少个 3

1281768756138976546984506985604982826762
9809858458224509856458945098450980943585
9091030209905959595772564675050678904567
8845789809821677654876364908560912949686

图中有多少个 3


128176875613897654698450698560498282676 2
9809858458224509856458945098450980943585
9091030209905959595772564675050678904567
8845789809821677654876364908560912949686

数据模式

Set A		Set B		Set C		Set D	
X	Y	X	Y	X	Y	X	Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.11	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Summary Statistics

$u_X = 9.0$ $\sigma_X = 3.317$

$u_Y = 7.5$ $\sigma_Y = 2.03$

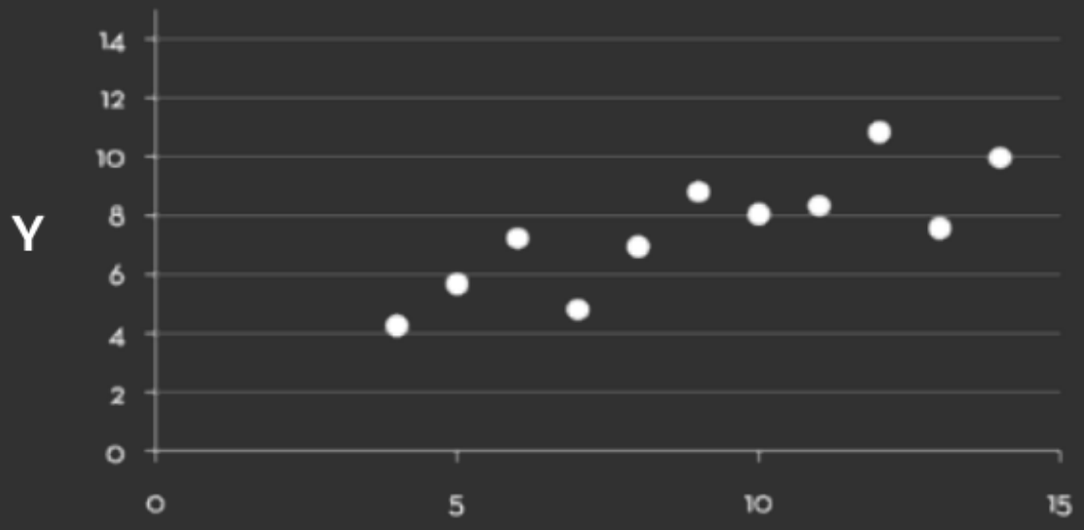
Linear Regression

$Y = 3 + 0.5\ X$

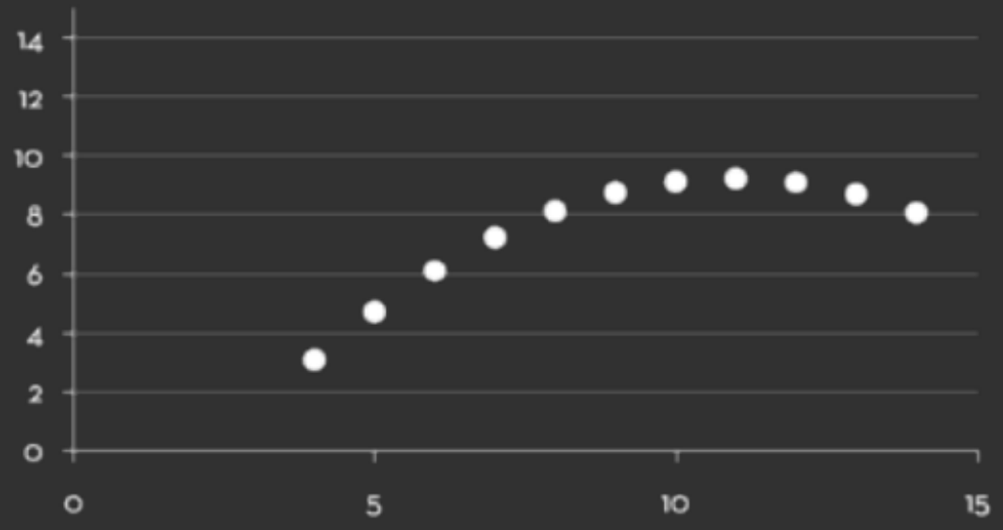
$R^2 = 0.67$

[Anscombe 73]

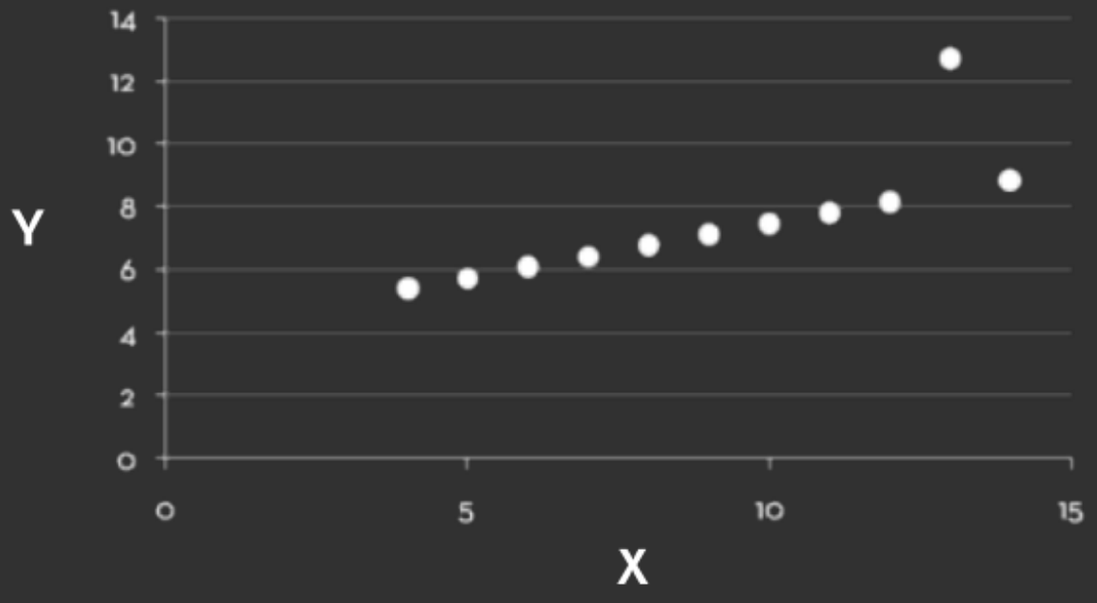
Set A



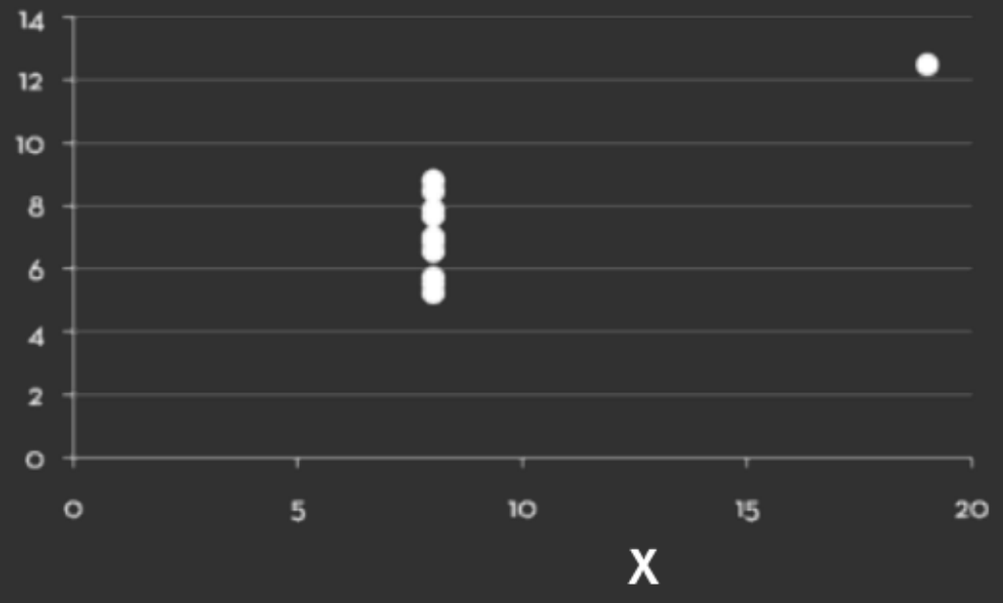
Set B



Set C



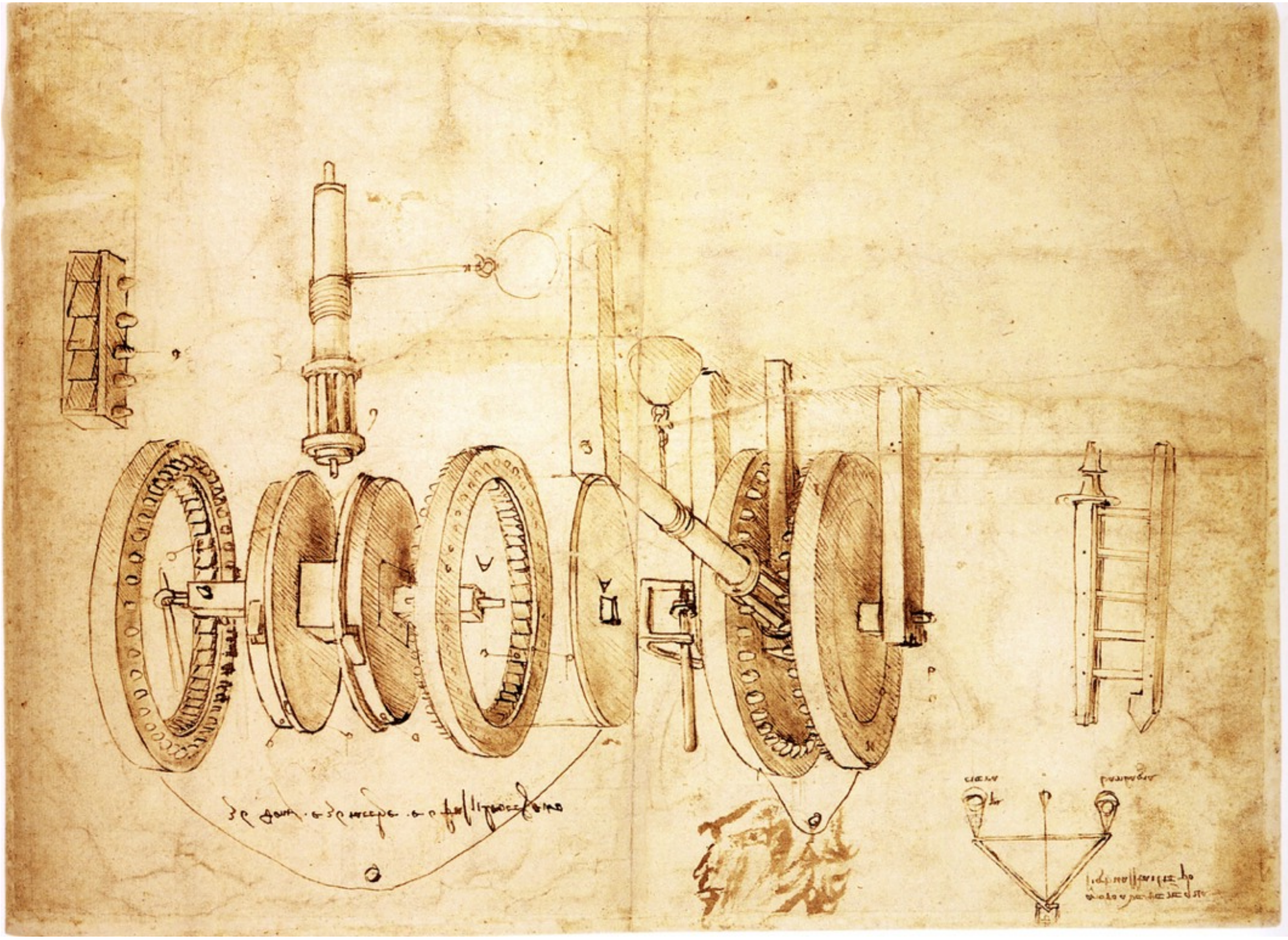
Set D

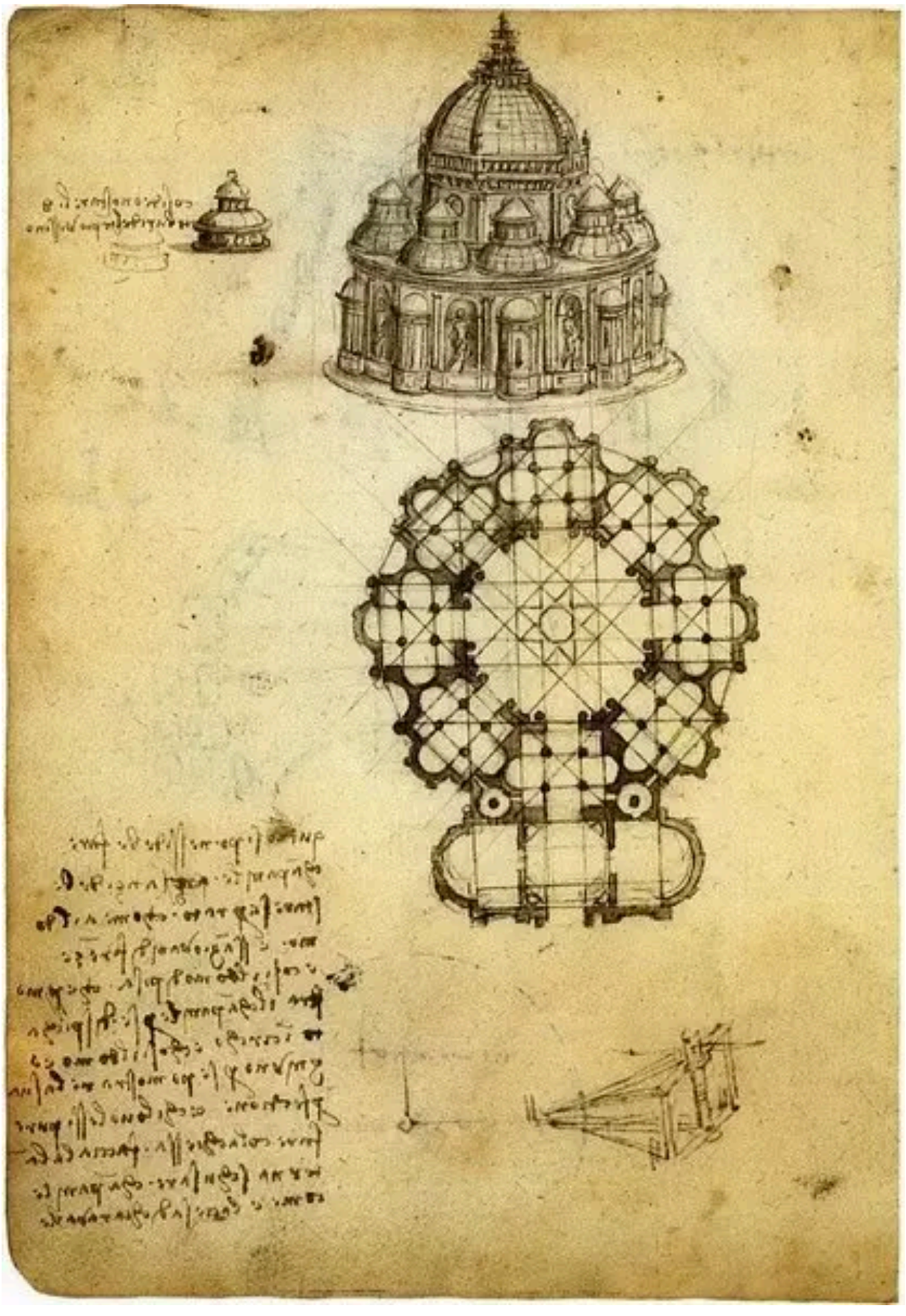


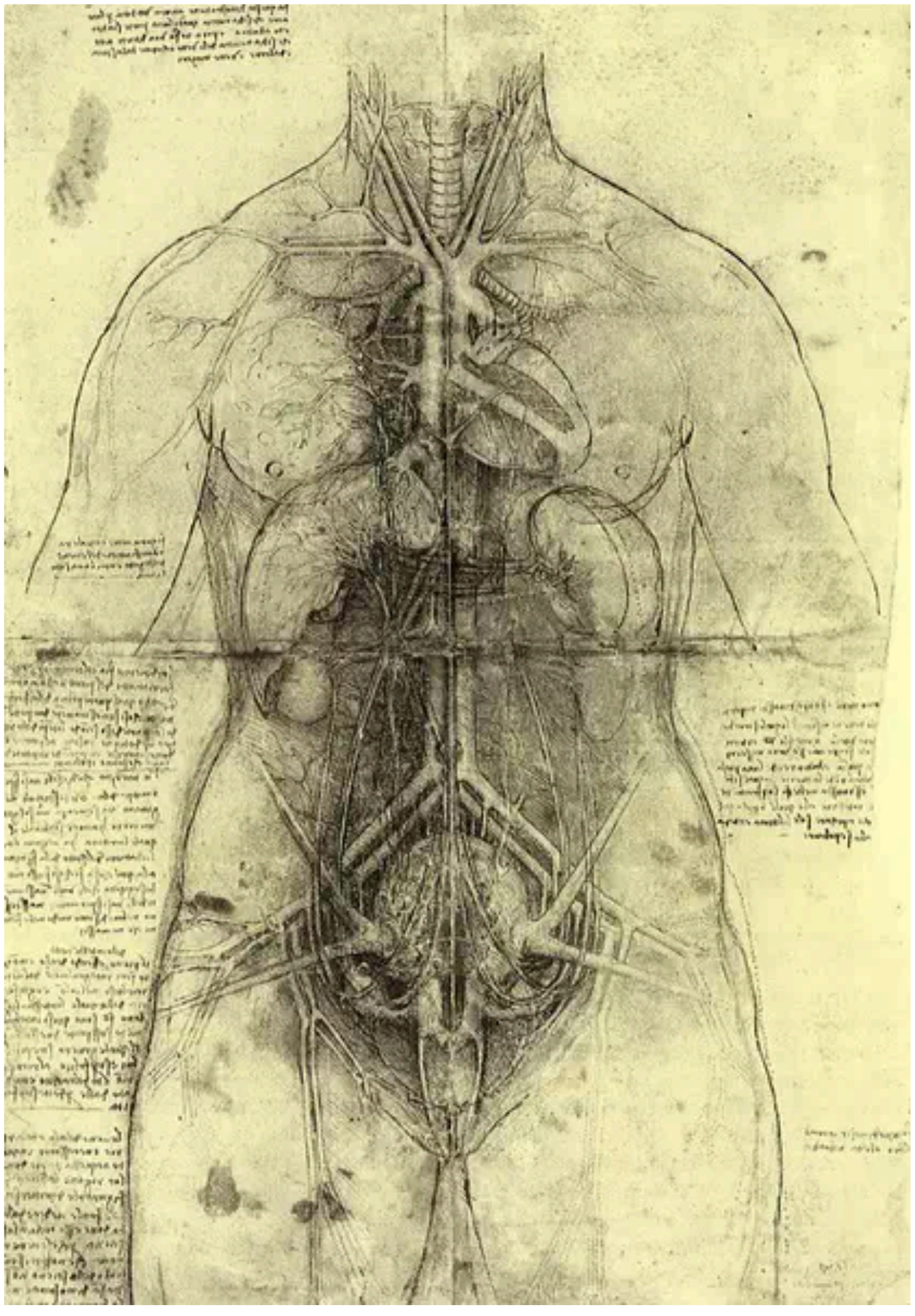
艺术

艺术

达·你永远不知道我多牛·**芬奇**







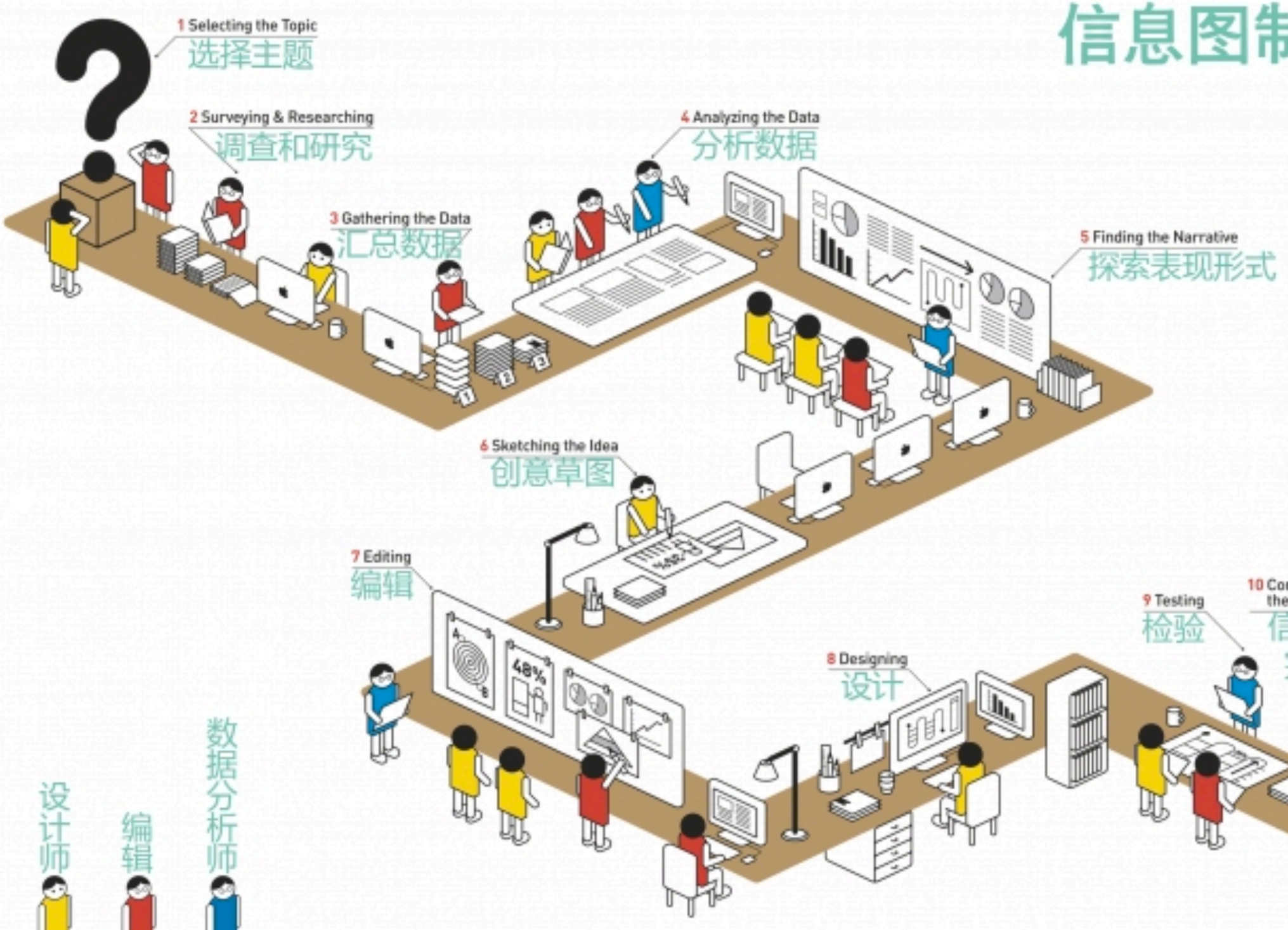
数据可视化的分类

数据可视化分类

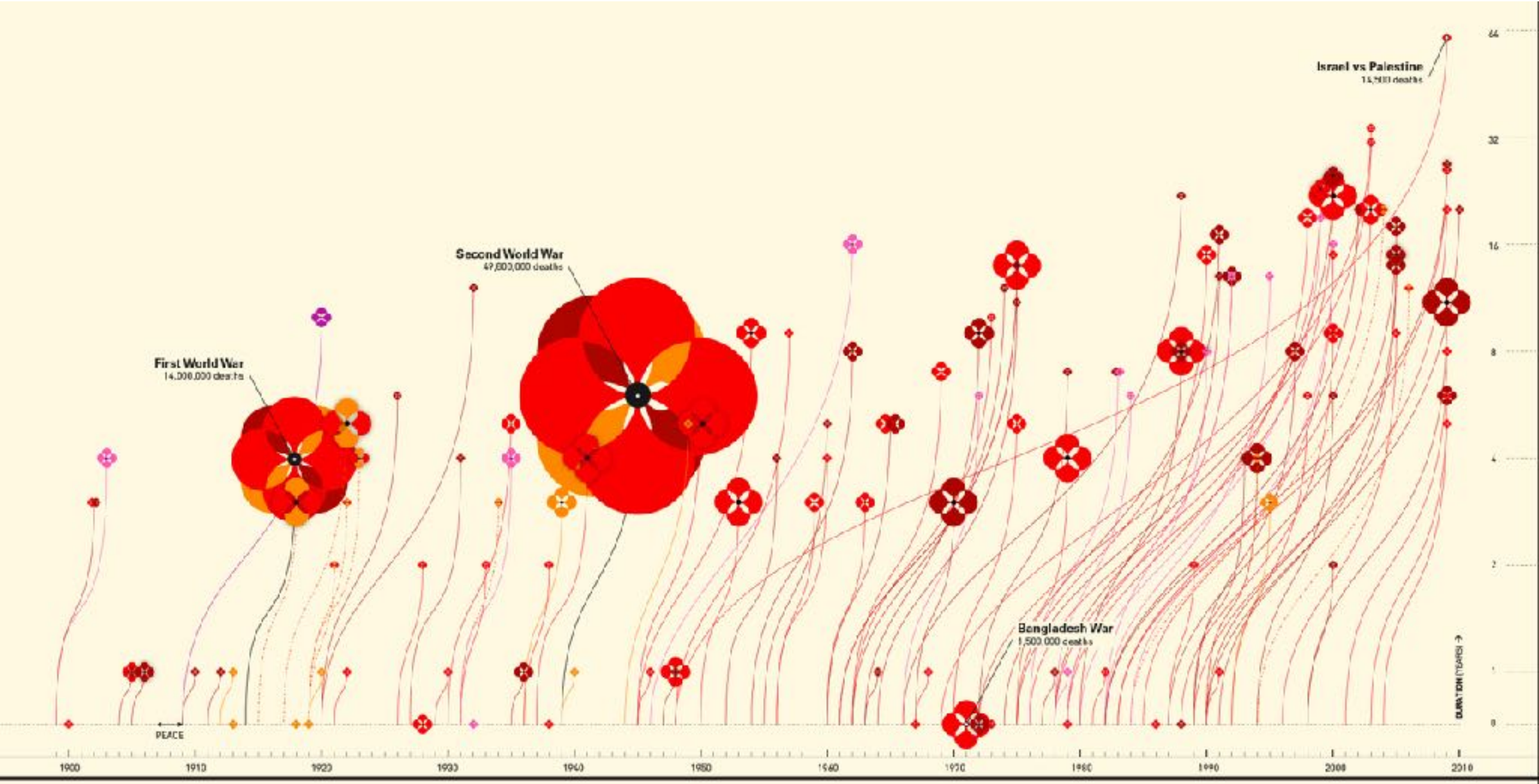
- 信息可视化 (Information visualization)
- 科学可视化 (Scientific visualization)
- 可视分析 (Visual Analytics)

信息可视化

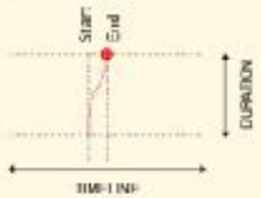
信息图制作流程


1 Selecting the Topic
选择主题
2 Surveying & Researching
调查和研究
3 Gathering the Data
汇总数据
4 Analyzing the Data
分析数据
5 Finding the Narrative
探索表现形式
6 Sketching the Idea
创意草图
7 Editing
编辑
8 Designing
设计
9 Testing
检验
10 Completing the Infographic
信息图完成
48%

设计师
编辑
数据分析师



POPPY DIAGRAM

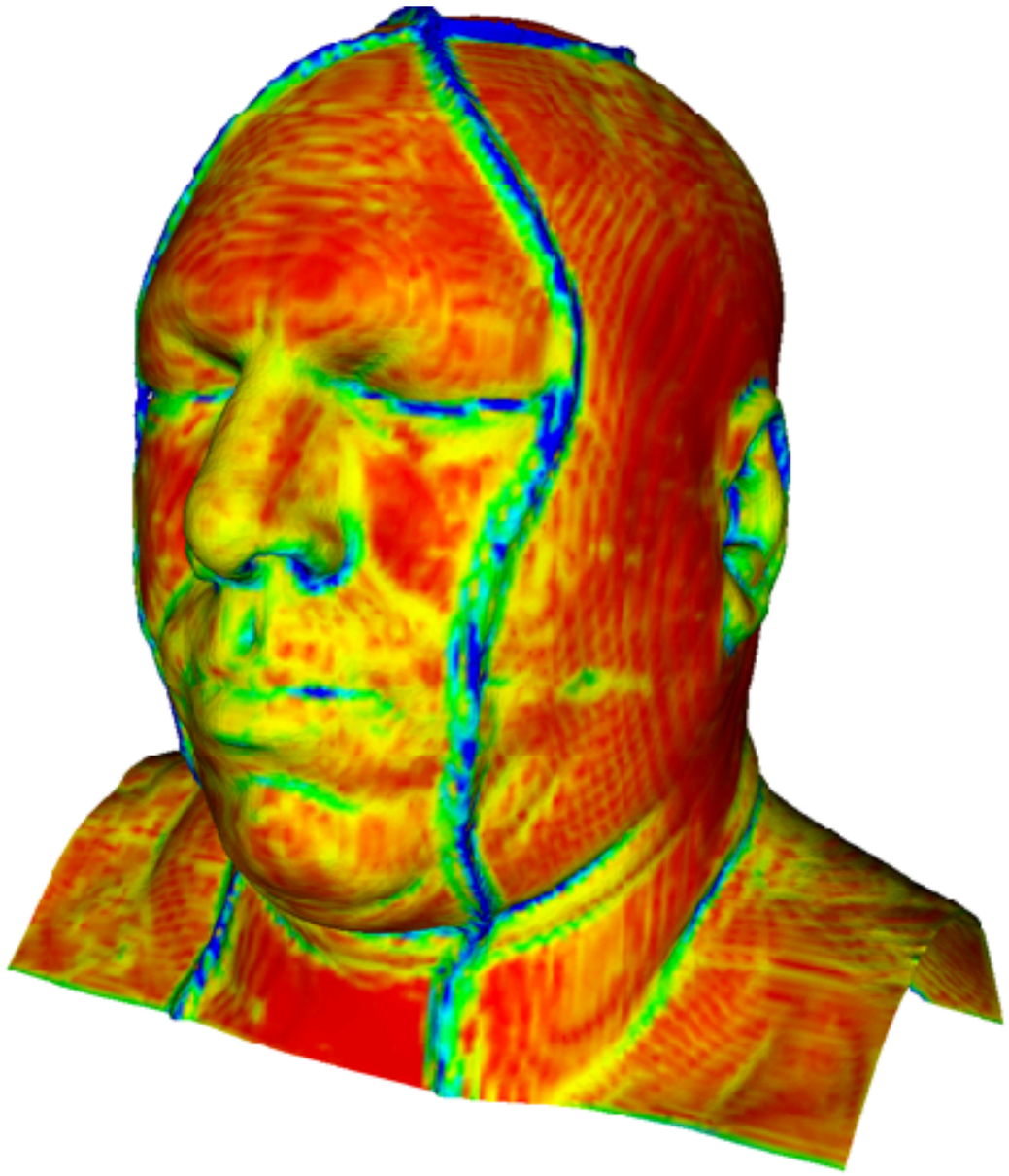


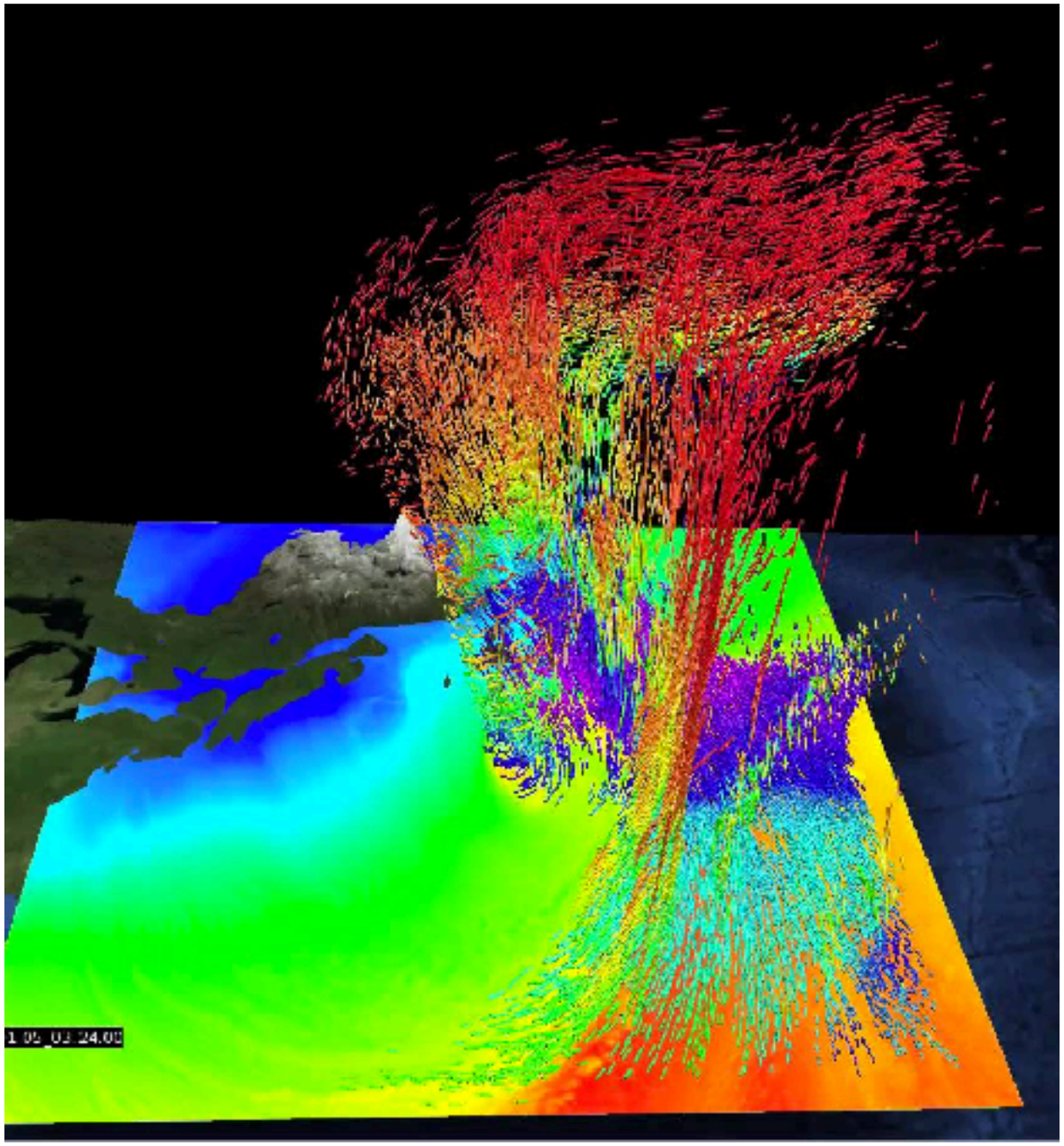
The remembrance poppy commemorates soldiers who have died in war. Each poppy in the diagram depicts a war of the last century (with more than 10,000 deaths). The stem grows from the year when the war started. The poppy flowers in the year the war ended. Its size shows the number of deaths.

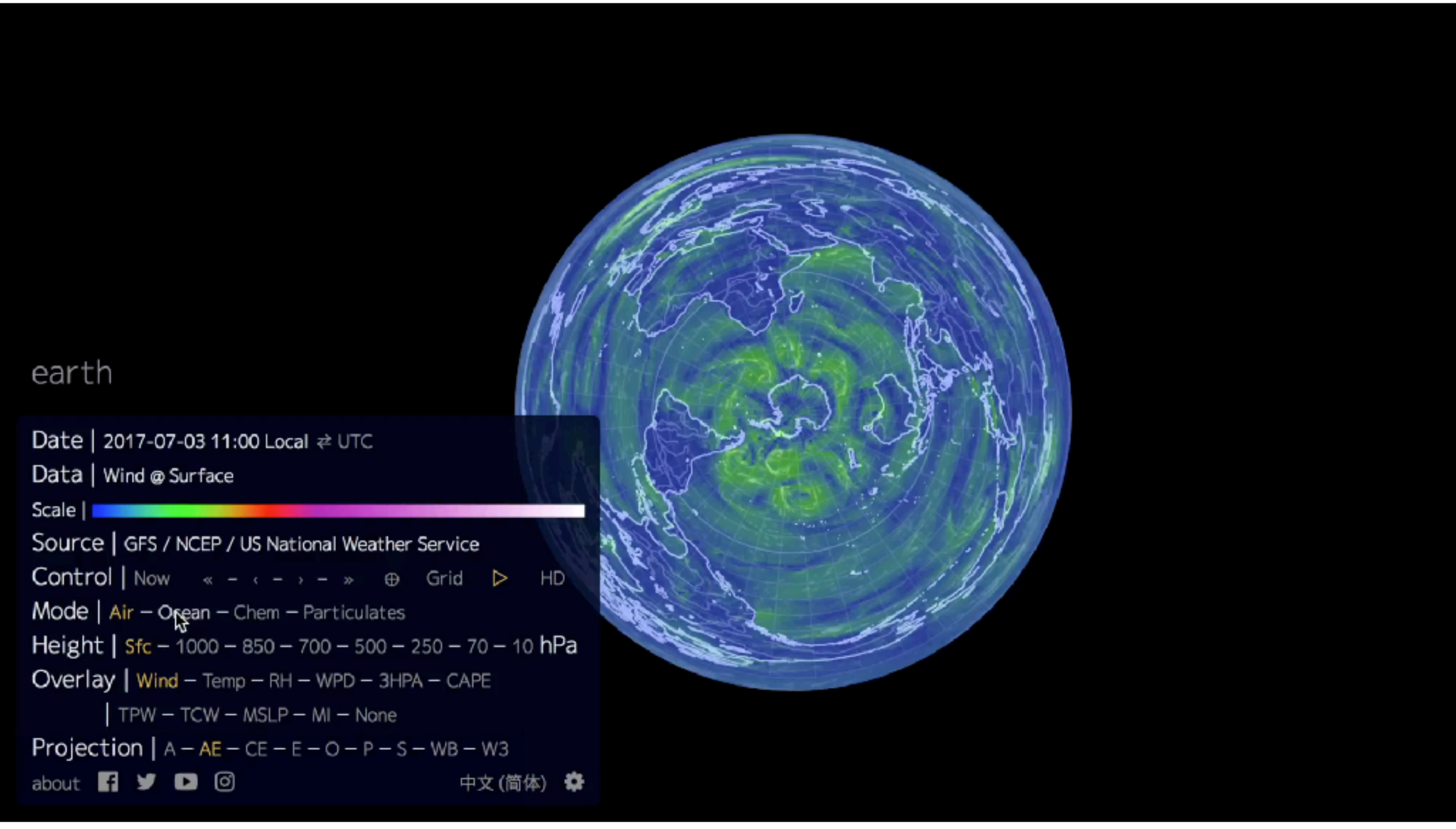
NUMBER of DEATHS IN THOUSANDS (POPPY'S SIZE)

REGIONS INVOLVED IN WARS (POPPY'S COLOUR)

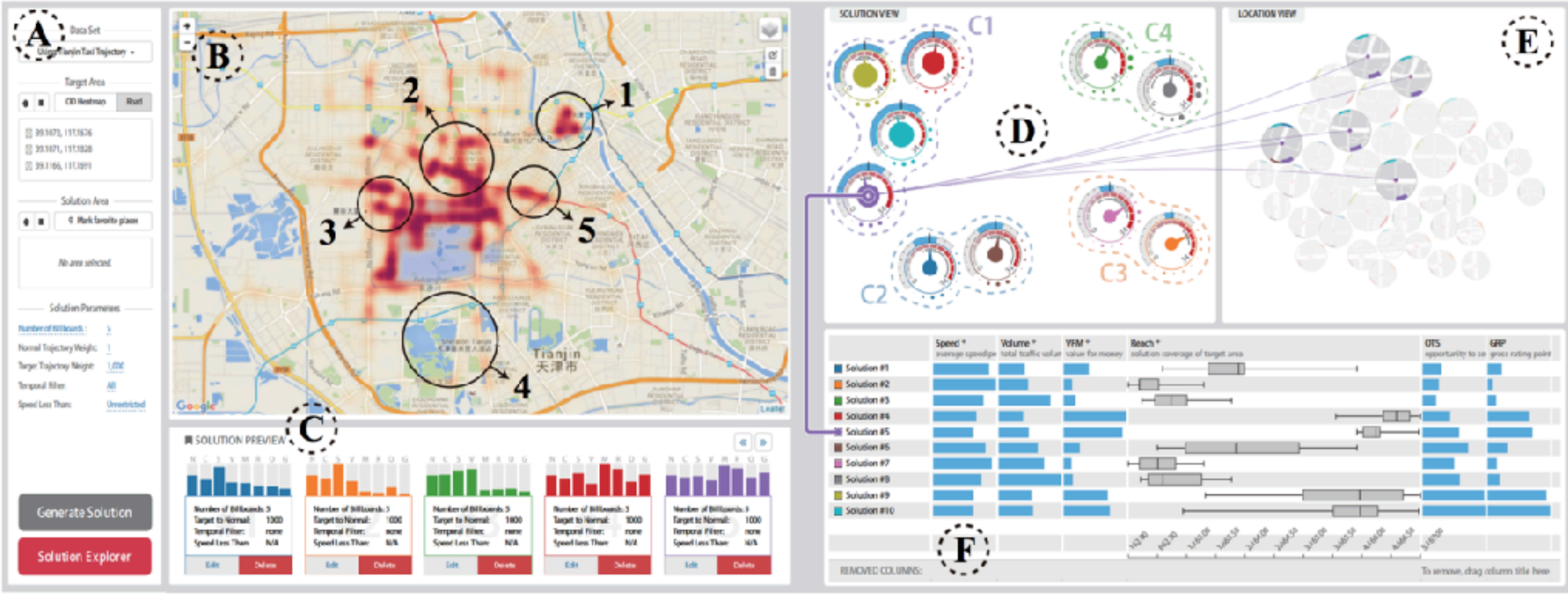
科学可视化



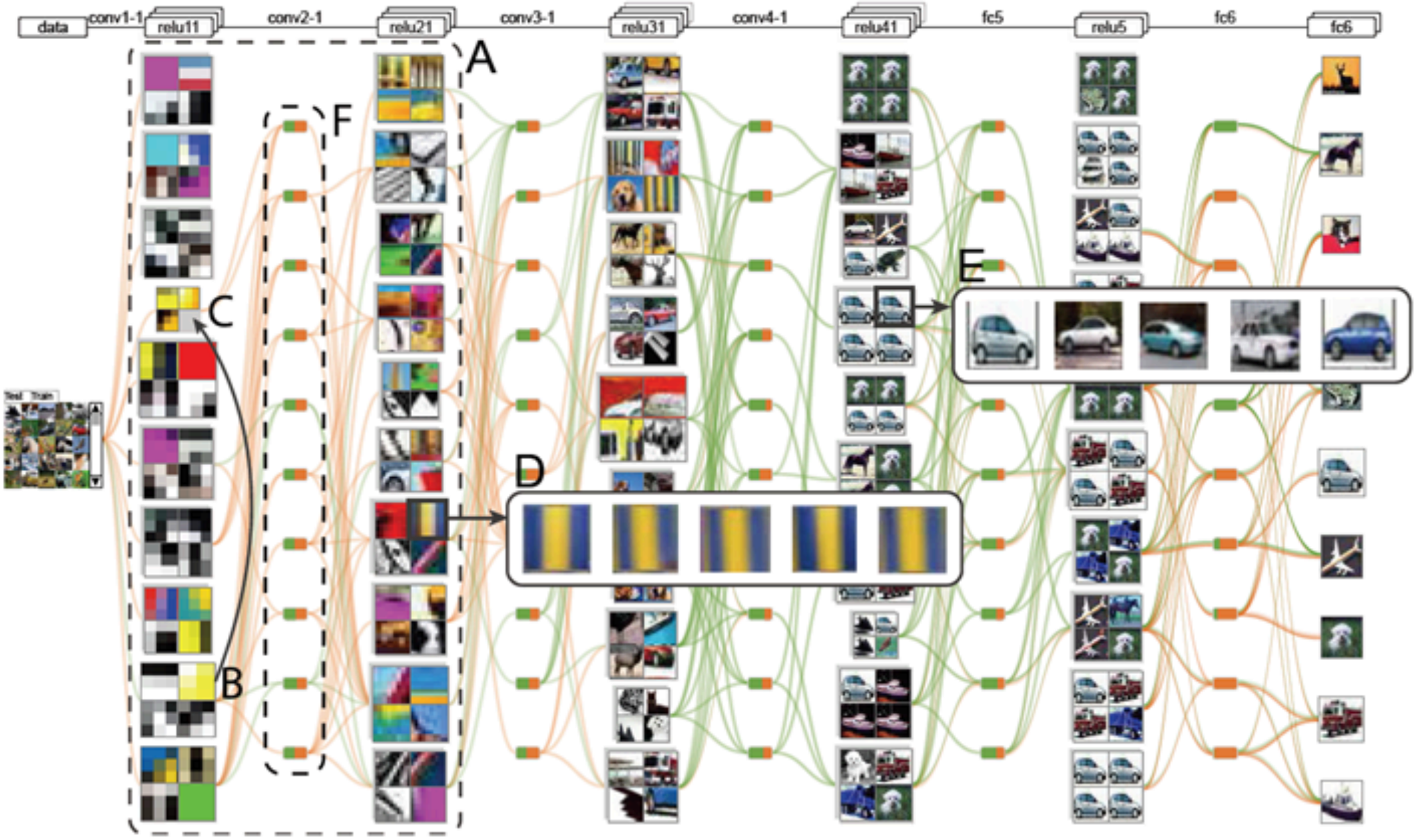



earth
Date | 2017-07-03 11:00 Local ⇄ UTC
Data | Wind @ Surface
Scale |
Source | GFS / NCEP / US National Weather Service
Control | Now « – ‹ – › – » ⊕ Grid ▷ HD
Mode | Air – Ocean – Chem – Particulates
Height | Sfc – 1000 – 850 – 700 – 500 – 250 – 70 – 10 hPa
Overlay | Wind – Temp – RH – WPD – 3HPA – CAPE
| TPW – TCW – MSLP – MI – None
Projection | A – AE – CE – E – O – P – S – WB – W3
about
中文 (简体)

可视分析



通过出租车轨迹数据
分析户外广告牌放置



帮助研究者更好的理解、判断和调整深度卷积神经网络

数据类型

- 时空数据 (Spatial Data Visualization)
- 地理数据 (Geographical Data Visualization)
- 时序数据 (Temporal Data Visualization)
- 高维数据 (High-Dimensional Data Visualization)
- 层次数据 (Hierarchical Data Visualization)
- 网络数据 (Network Data Visualization)
- 媒体数据 (Cross-Media Visualization)

为什么要做数据可视化

数据可视化的价值

呈现信息 → 讲故事
汇报总结、愿景

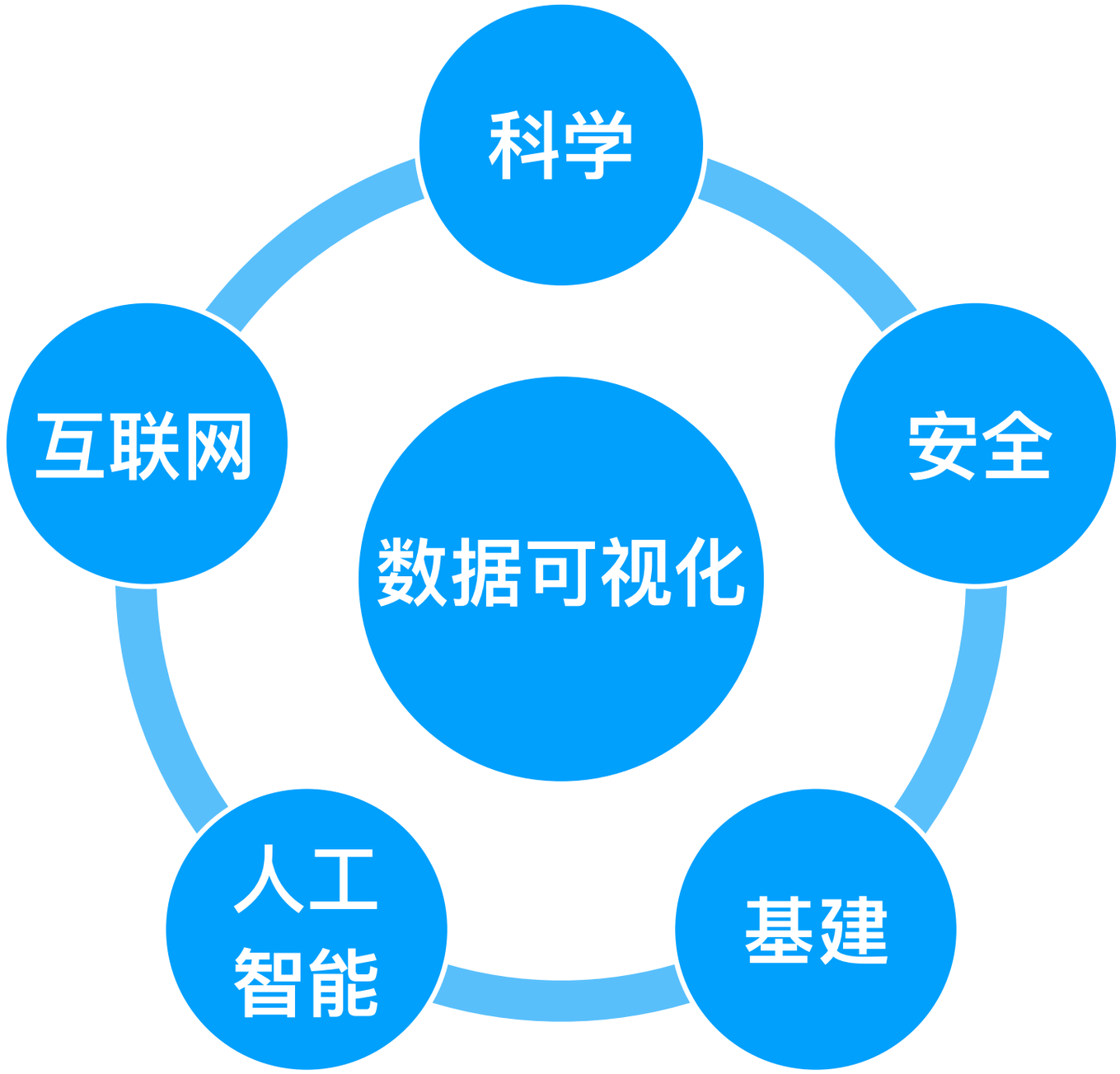
分析数据 → 扩展大脑内存
发现数据模式、特征
决策、预测

沟通数据 → 交互、共享、讨论
协作分析

数据可视化的

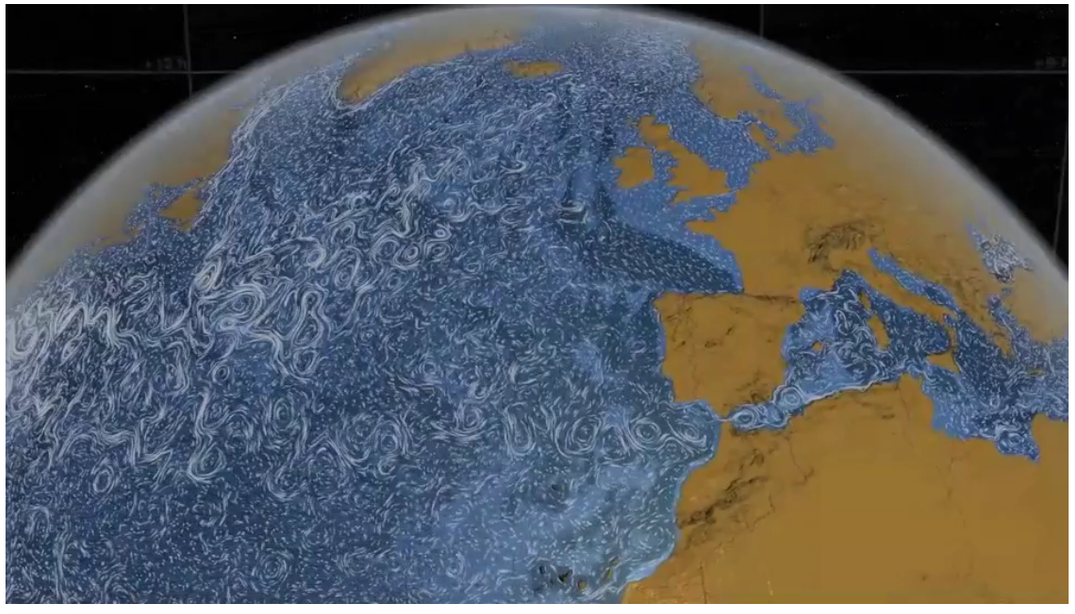
热点和趋势

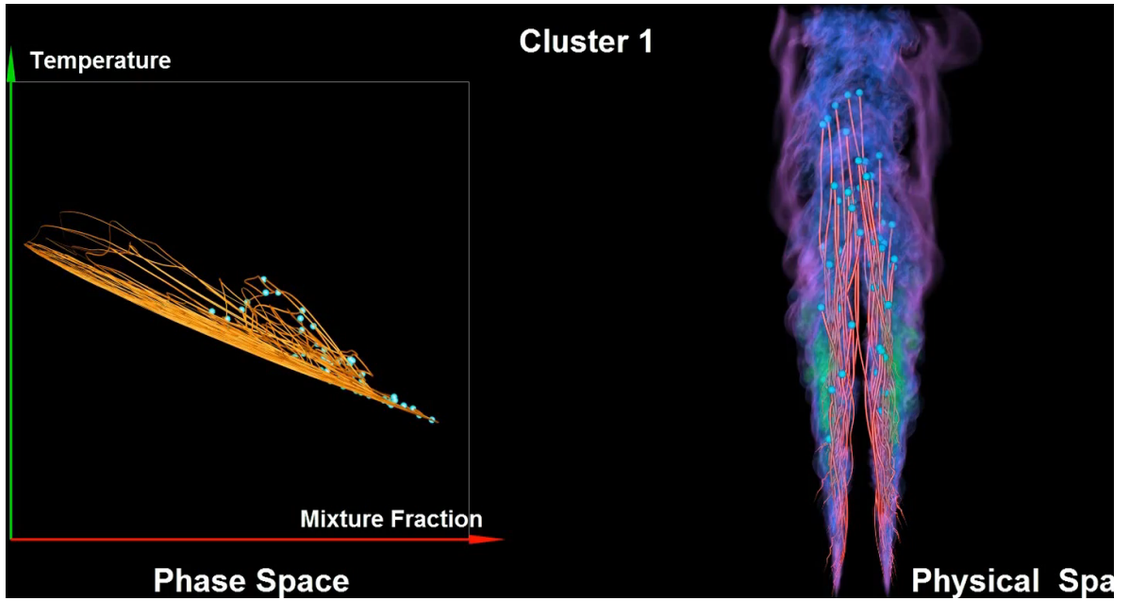
热点和趋势

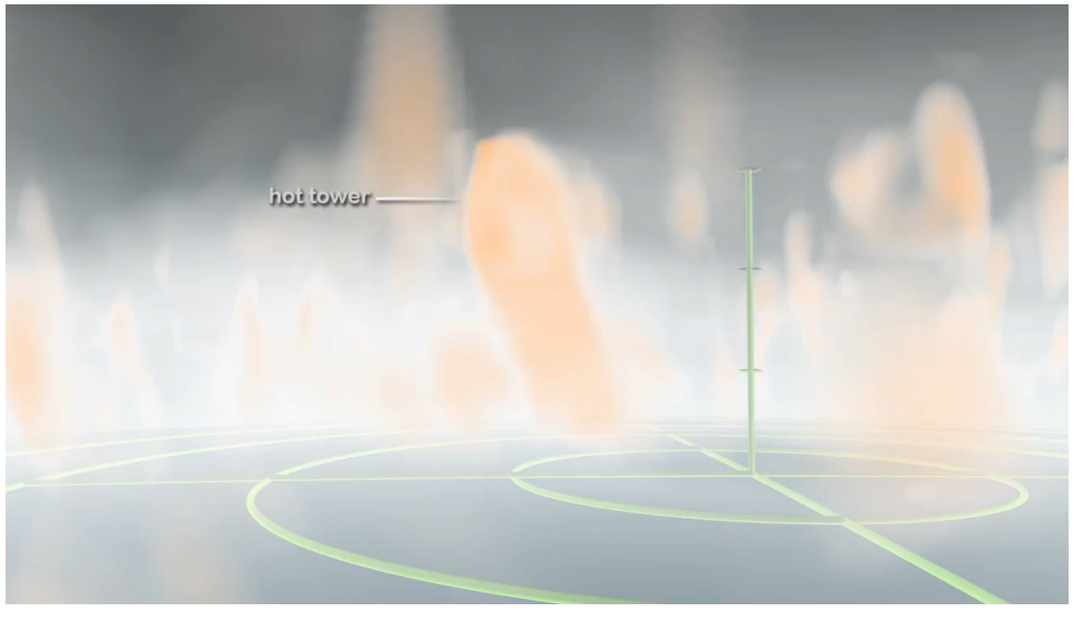


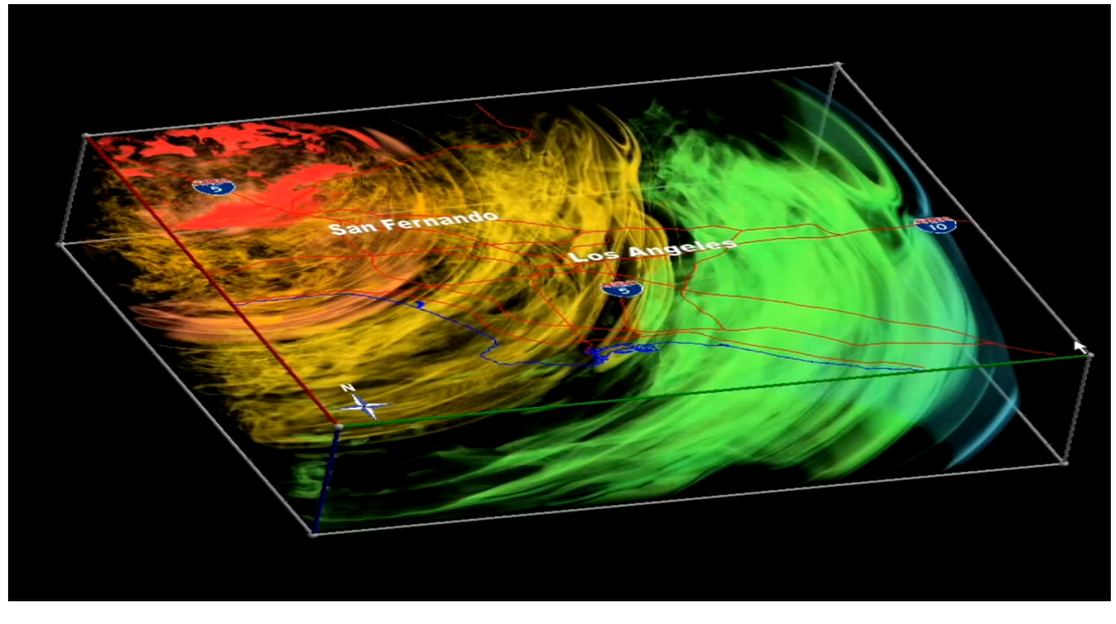
应用：科学

可视化是基础自然科学的必要手段，是科学大数据发展的必需









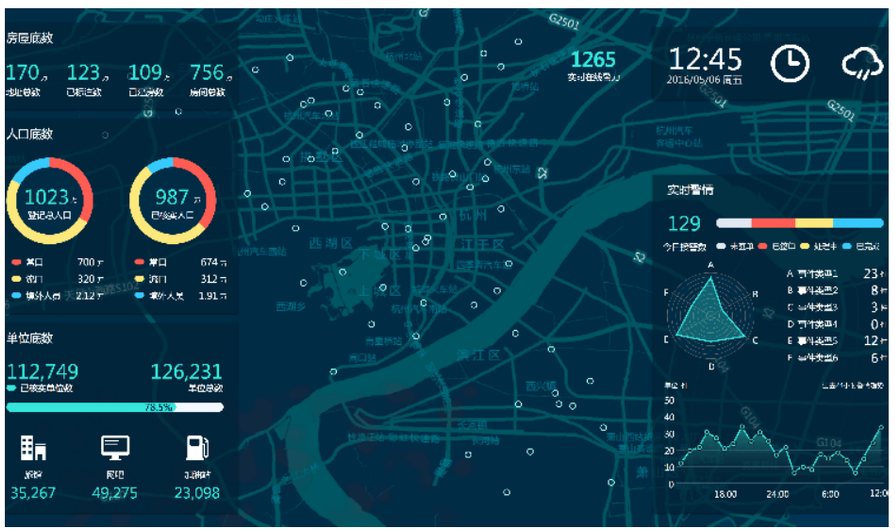
应用：安全

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

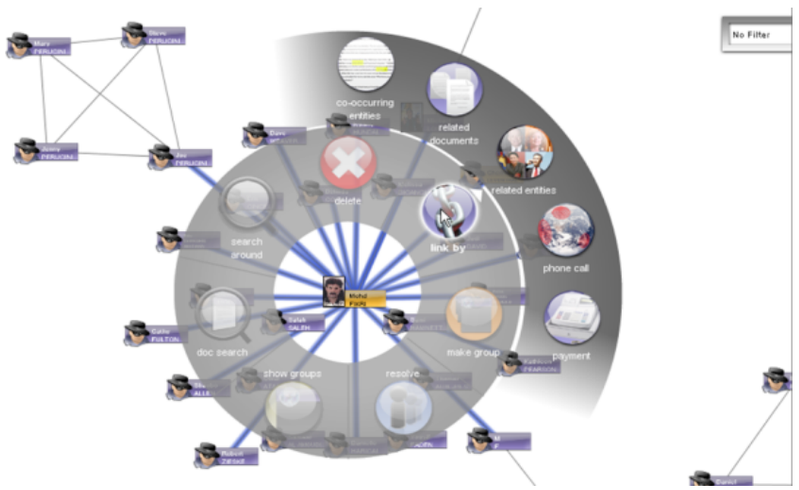
国土安全

网络安全

公共安全



金融安全

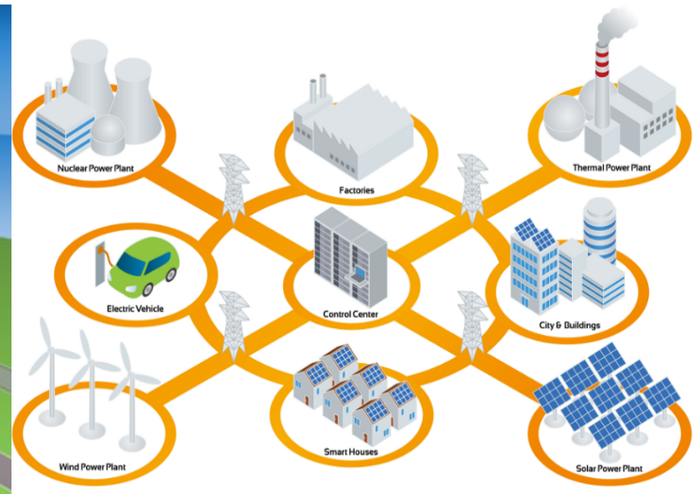


应用：基建

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

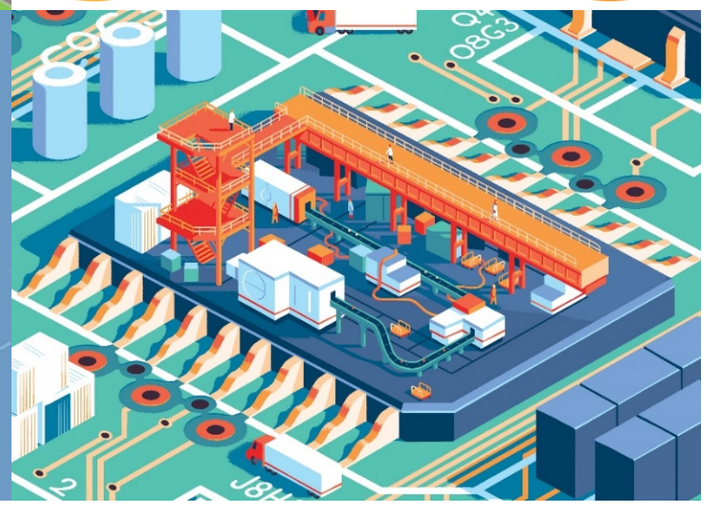
智能**交通**

智能**电网**



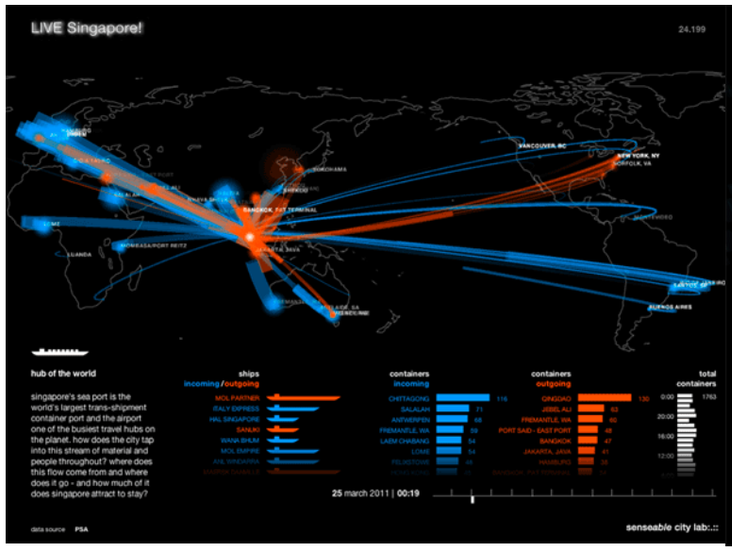
智能**物流**

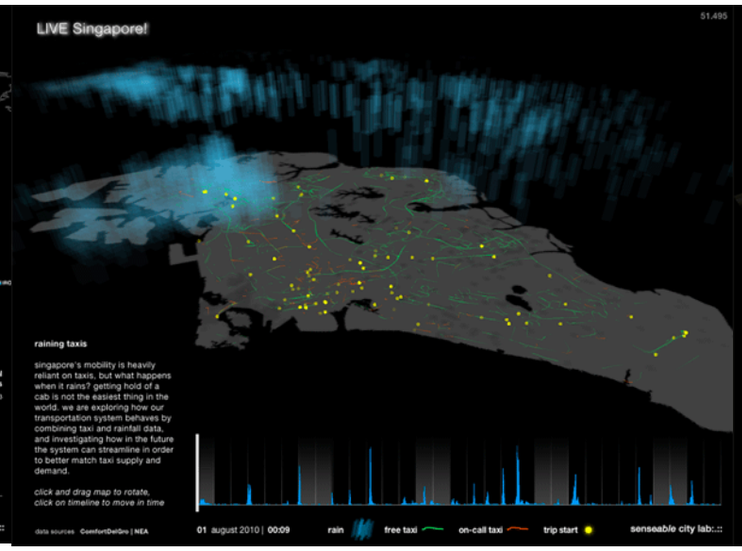
智能**制造**

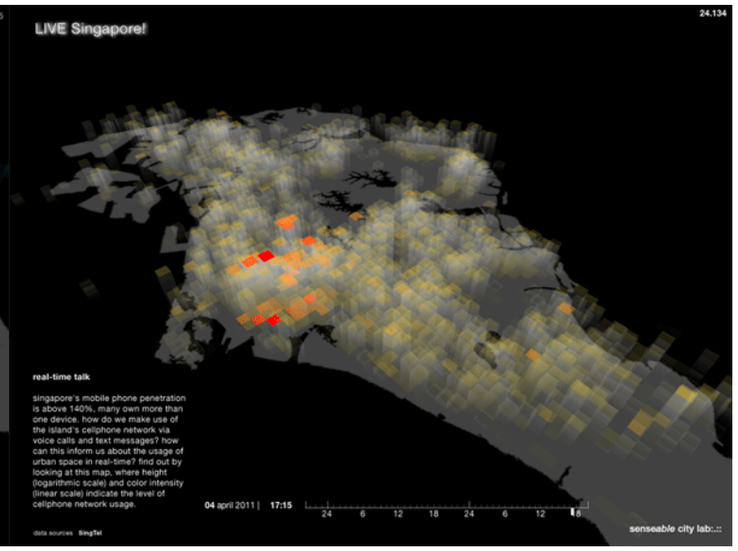


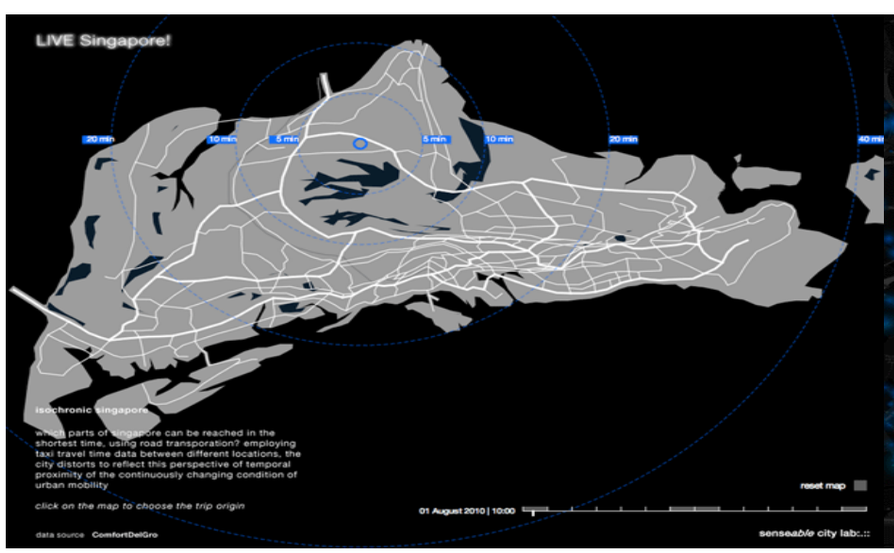
应用：基建

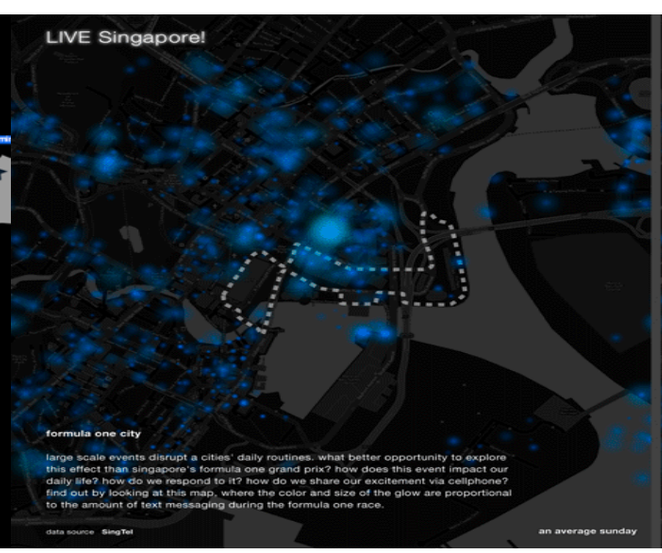
可视化是基于城市数据进行规划、理解、决策的敏捷分析途径

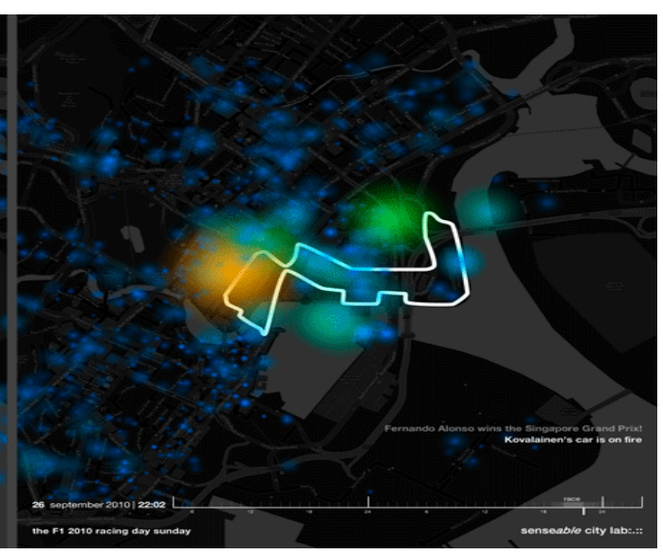




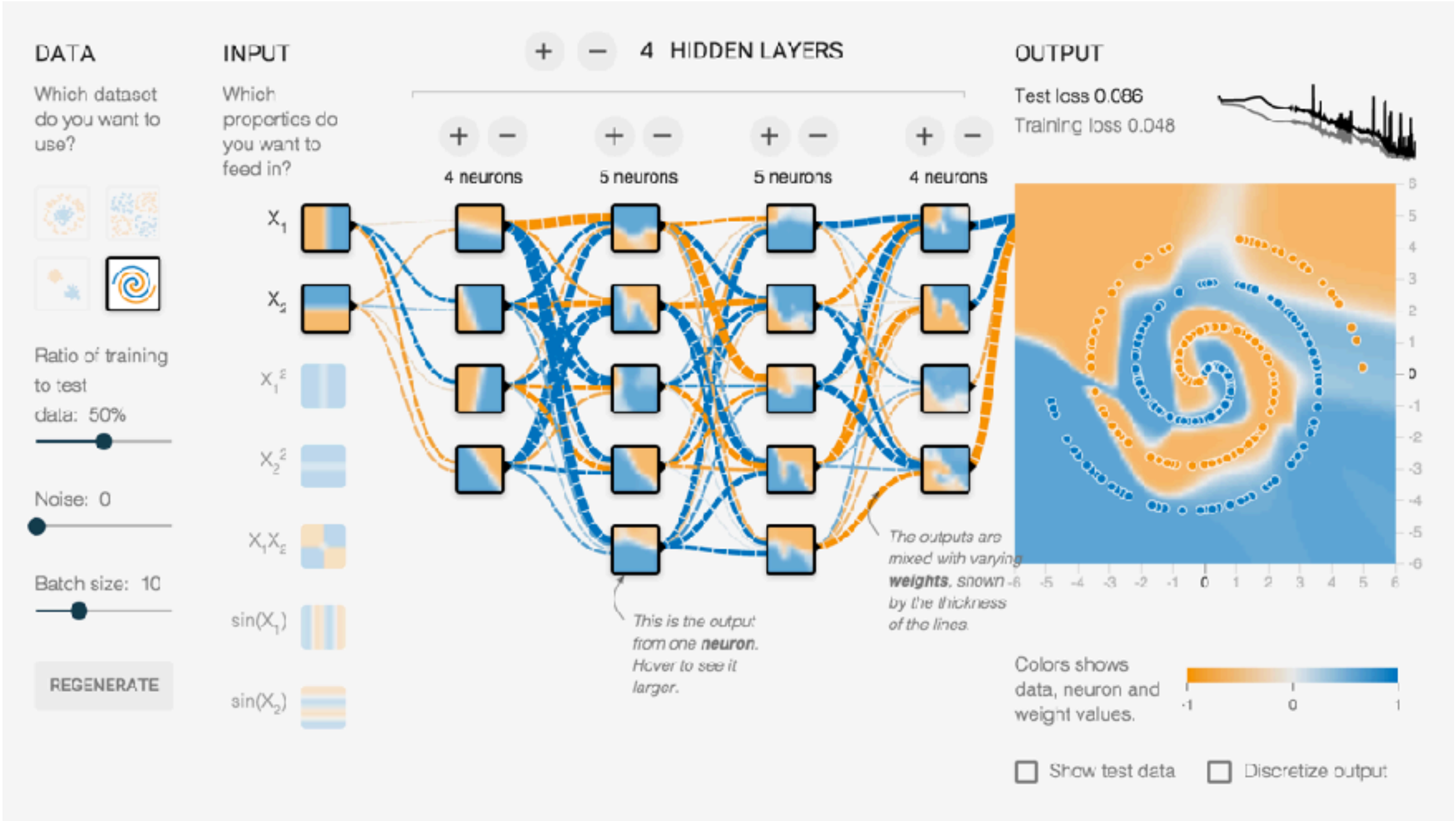




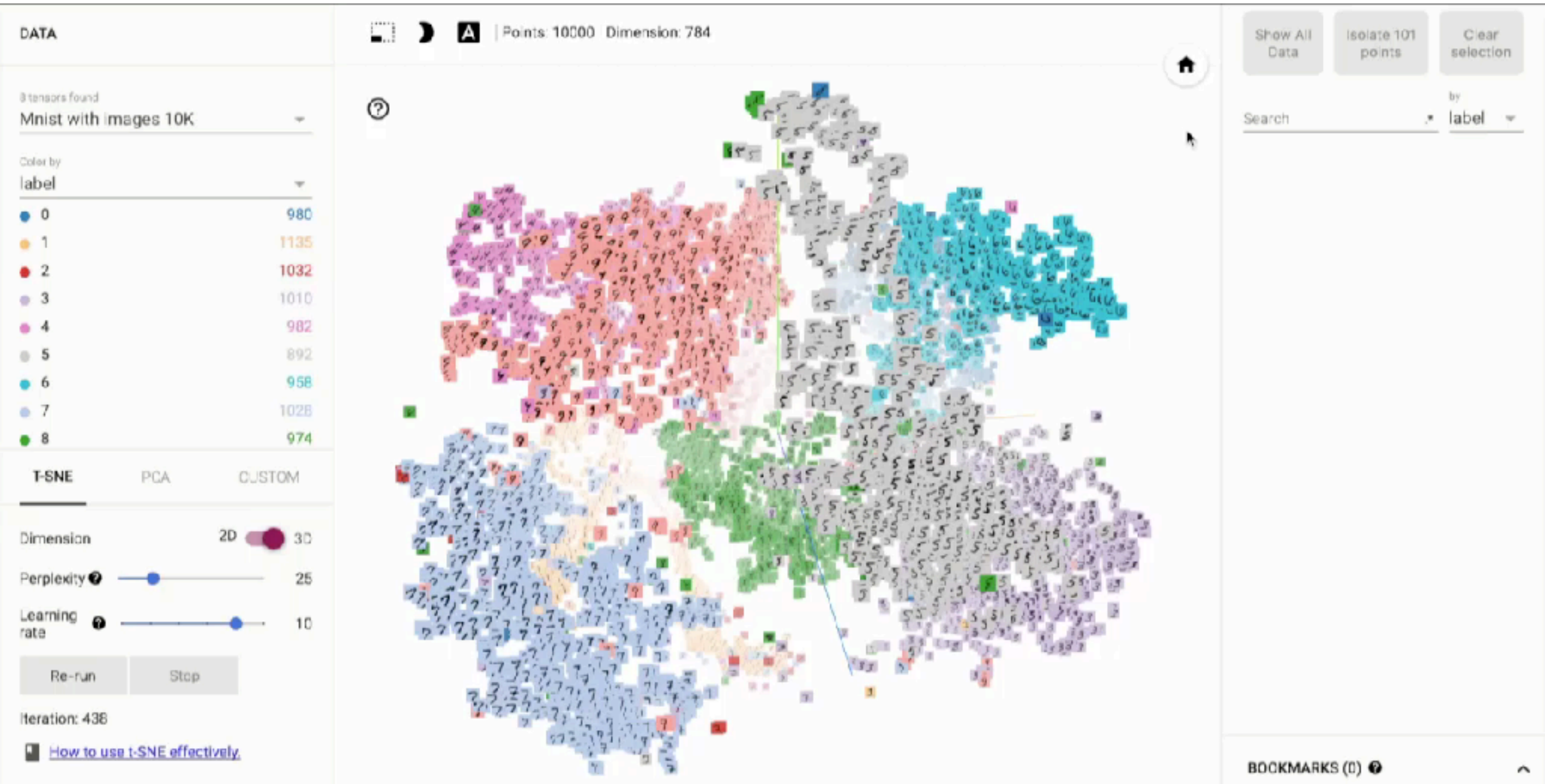




应用：人工智能


DATA
Which dataset do you want to use?
Ratio of training to test data: 50%
Noise: 0
Batch size: 10
REGENERATE
INPUT
Which properties do you want to feed in?
X1
X2
X1²
X2²
X1X2
sin(X1)
sin(X2)
4 HIDDEN LAYERS
4 neurons
5 neurons
5 neurons
4 neurons
This is the output from one neuron. Hover to see it larger.
The outputs are mixed with varying weights, shown by the thickness of the lines.
OUTPUT
Test loss 0.086
Training loss 0.048
Colors shows data, neuron and weight values.
Show test data
Discretize output

人工智能 - TensorBoard



应用：互联网

可视化是进行海量行为挖掘、分析、建模、理解的重要途径

Visual Analysis of Topic Competition on Social Media

Panpan Xu[3] *Yingcai Wu*[2] *Enxun Wei*[1,3] *Tai-Quan Peng*[4]
Shixia Liu[2] *Jonathan Z.H. Zhu*[5] *Huamin Qu*[1]

[1]*HKUST* [2]*MSRA* [3]*SJTU* [4]*NTU* [5]*CityU*

数据可视化基础

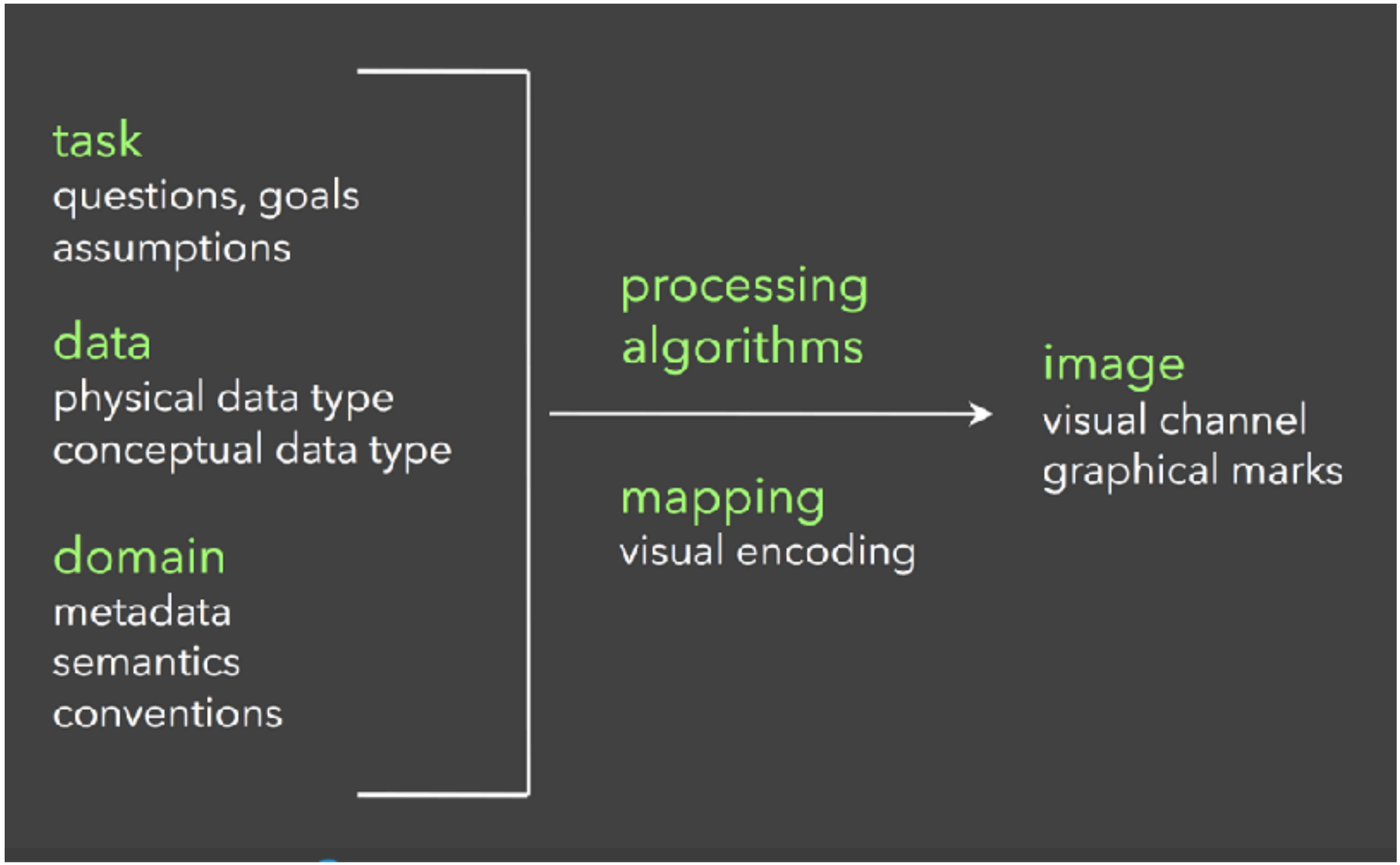
数据可视化基础

基础

- 可视化流程
- 数据模型
- 视觉编码

http://geekplux.com

数据可视化流程



http://geekplux.com/2017/01/01/basics-of-data-visualization-the-process-model.html

数据模型

- 类别型
- 有序型
- 数值型

id	类型	款式	尺码	销量	年增长
1	男款	上衣	L	50	10%
2	女款	上衣	S	35	5%
3	女款	裤子	M	40	20%
4	男款	上衣	XL	30	15%

http://geekplux.com/2017/01/02/basics-of-data-visualization-data-model.html

数据模型

- 类别型
- 有序型
- 数值型

id	类型	款式	尺码	销量	年增长
1	男款	上衣	L	50	10%
2	女款	上衣	S	35	5%
3	女款	裤子	M	40	20%
4	男款	上衣	XL	30	15%

http://geekplux.com/2017/01/02/basics-of-data-visualization-data-model.html

数据模型

- 类别型
- 有序型
- 数值型

id	类型	款式	尺码	销量	年增长
1	男款	上衣	L	50	10%
2	女款	上衣	S	35	5%
3	女款	裤子	M	40	20%
4	男款	上衣	XL	30	15%

http://geekplux.com/2017/01/02/basics-of-data-visualization-data-model.html

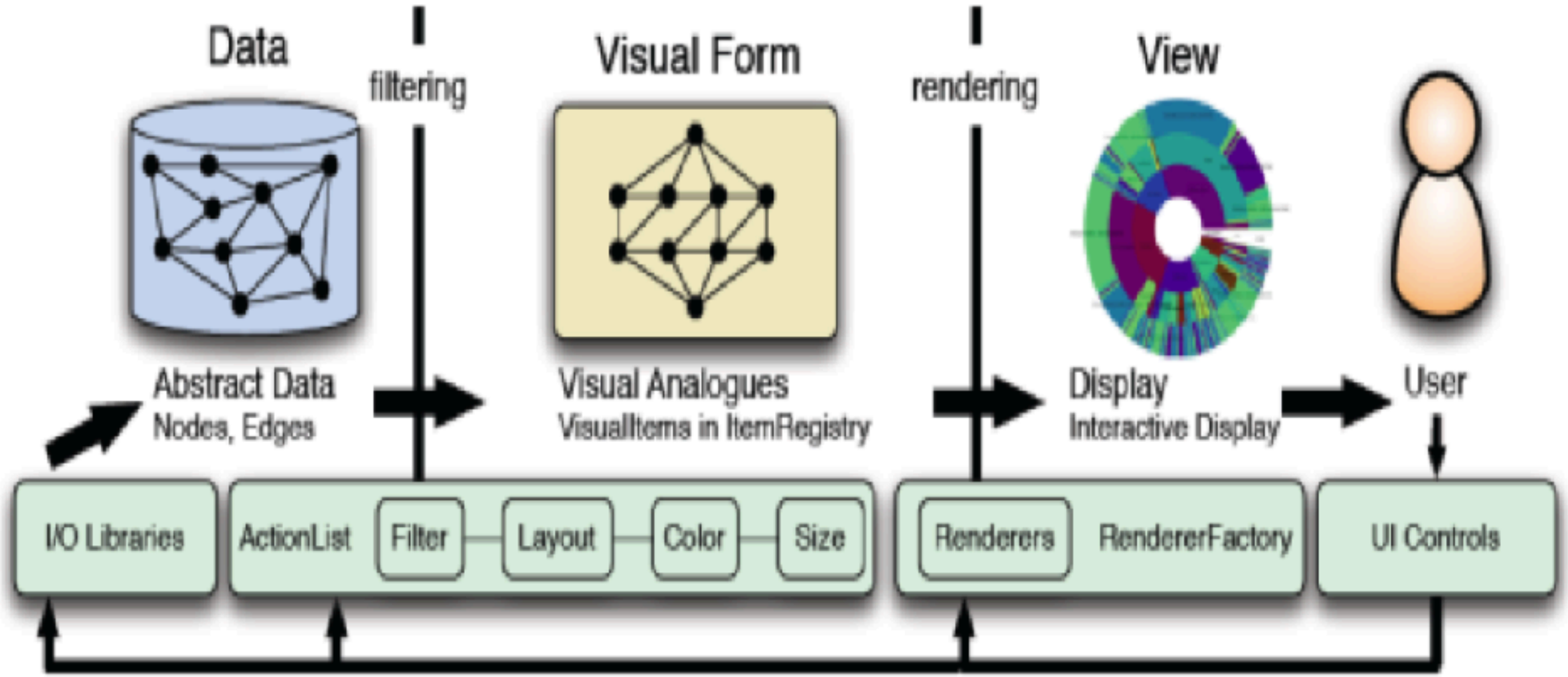
数据模型

- 类别型
- 有序型
- 数值型

id	类型	款式	尺码	销量	年增长
1	男款	上衣	L	50	10%
2	女款	上衣	S	35	5%
3	女款	裤子	M	40	20%
4	男款	上衣	XL	30	15%

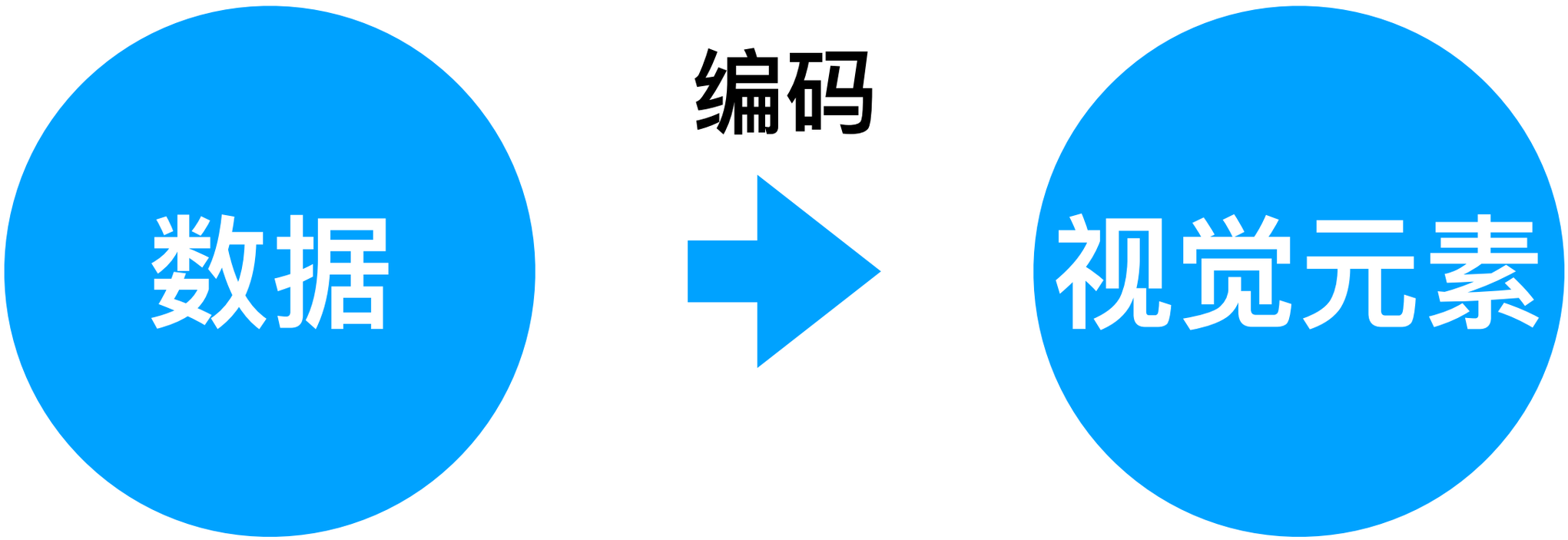
http://geekplux.com/2017/01/02/basics-of-data-visualization-data-model.html

视觉编码



http://geekplux.com/2017/01/03/basics-of-data-visualization-visual-encoding-principles.html

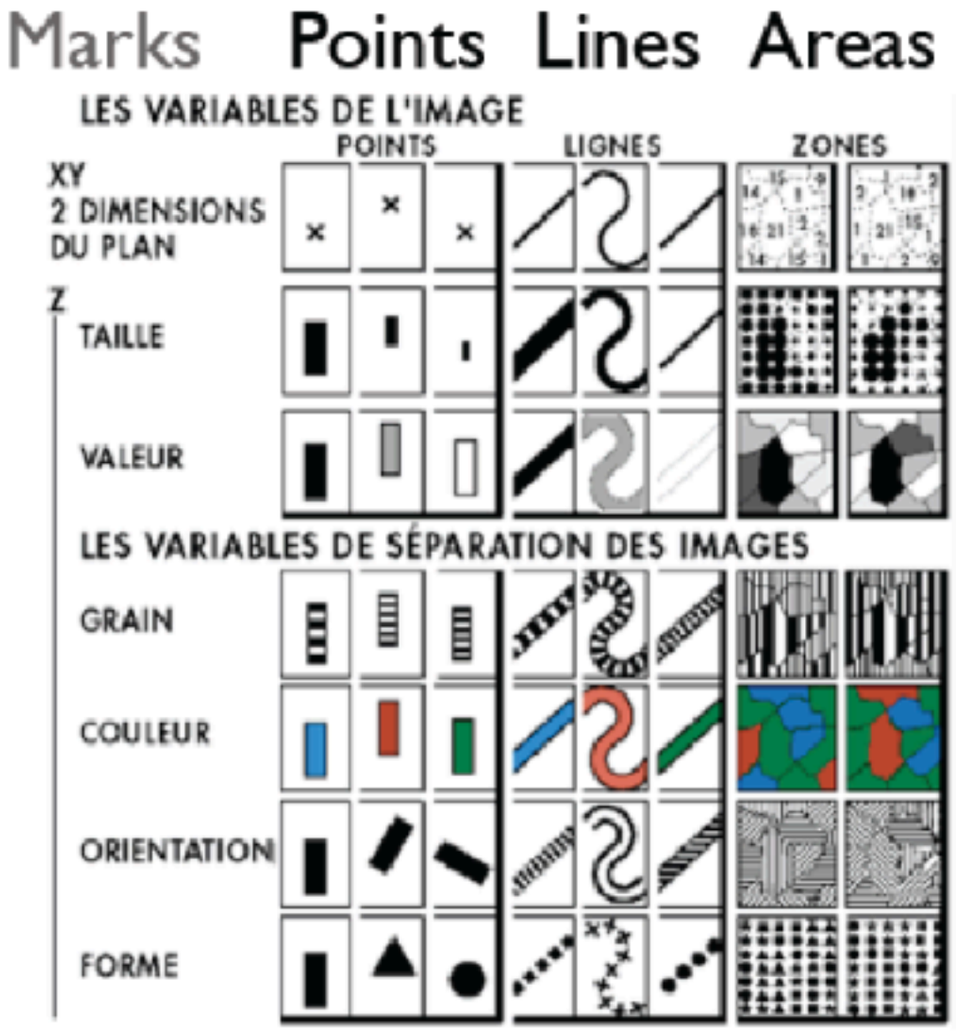
视觉编码



http://geekplux.com/2017/01/03/basics-of-data-visualization-visual-encoding-principles.html

视觉编码

- 位置 Position
- 尺寸 Size
- 数值 (Grey)Value
- 纹理 Texture
- 颜色 Color
- 方向 Orientation
- 形状 Shape



http://geekplux.com/2017/01/03/basics-of-data-visualization-visual-encoding-principles.html

视觉编码

- 位置
- 尺寸
- 数值
- 纹理
- 颜色
- 方向
- 形状

- 长度
- 面积
- 体积
- 透明度
- 模糊/聚焦
- 动画

http://geekplux.com/2017/01/03/basics-of-data-visualization-visual-encoding-principles.html

视觉编码

- 位置
- 尺寸
- 数值
- 纹理
- 颜色
- 方向
- 形状

类别型

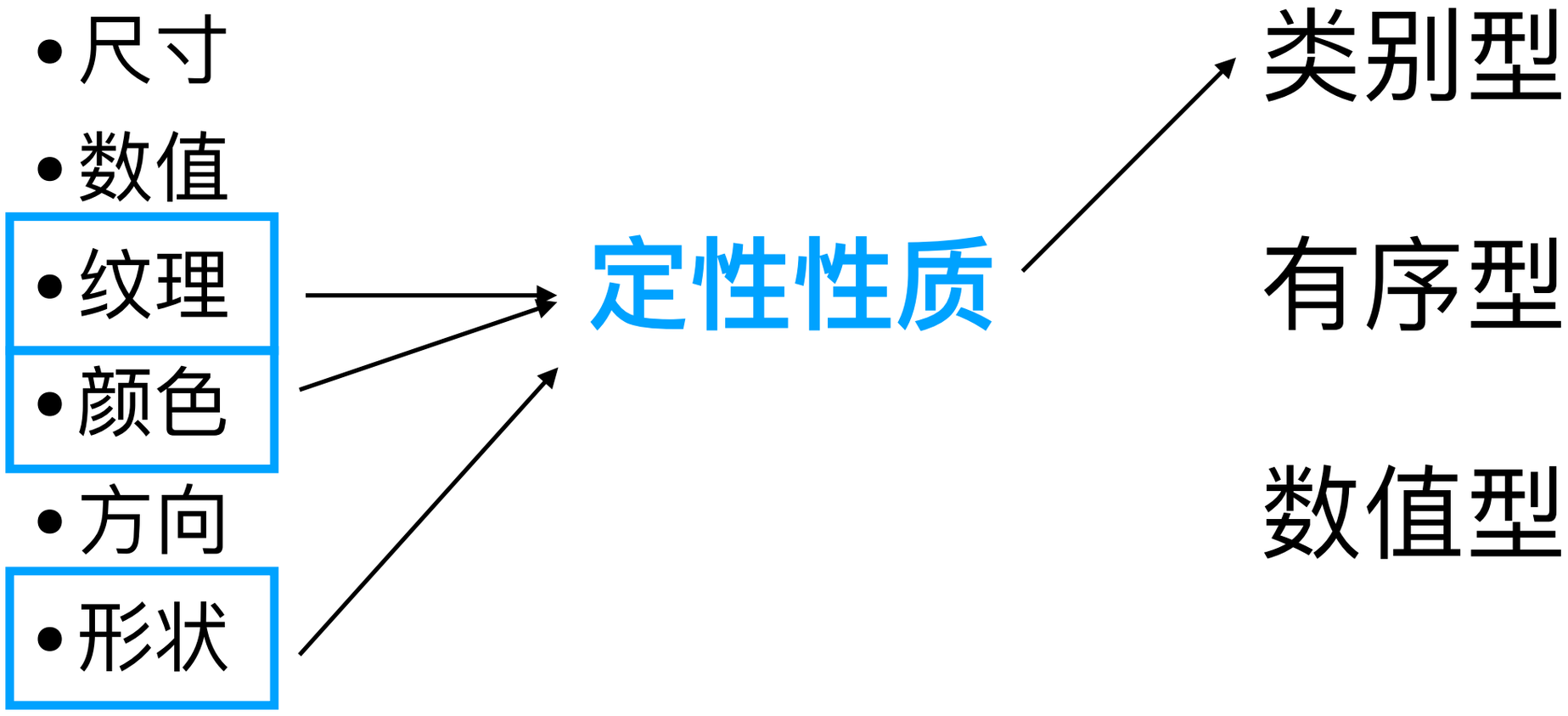
有序型

数值型

http://geekplux.com/2017/01/03/basics-of-data-visualization-visual-encoding-principles.html

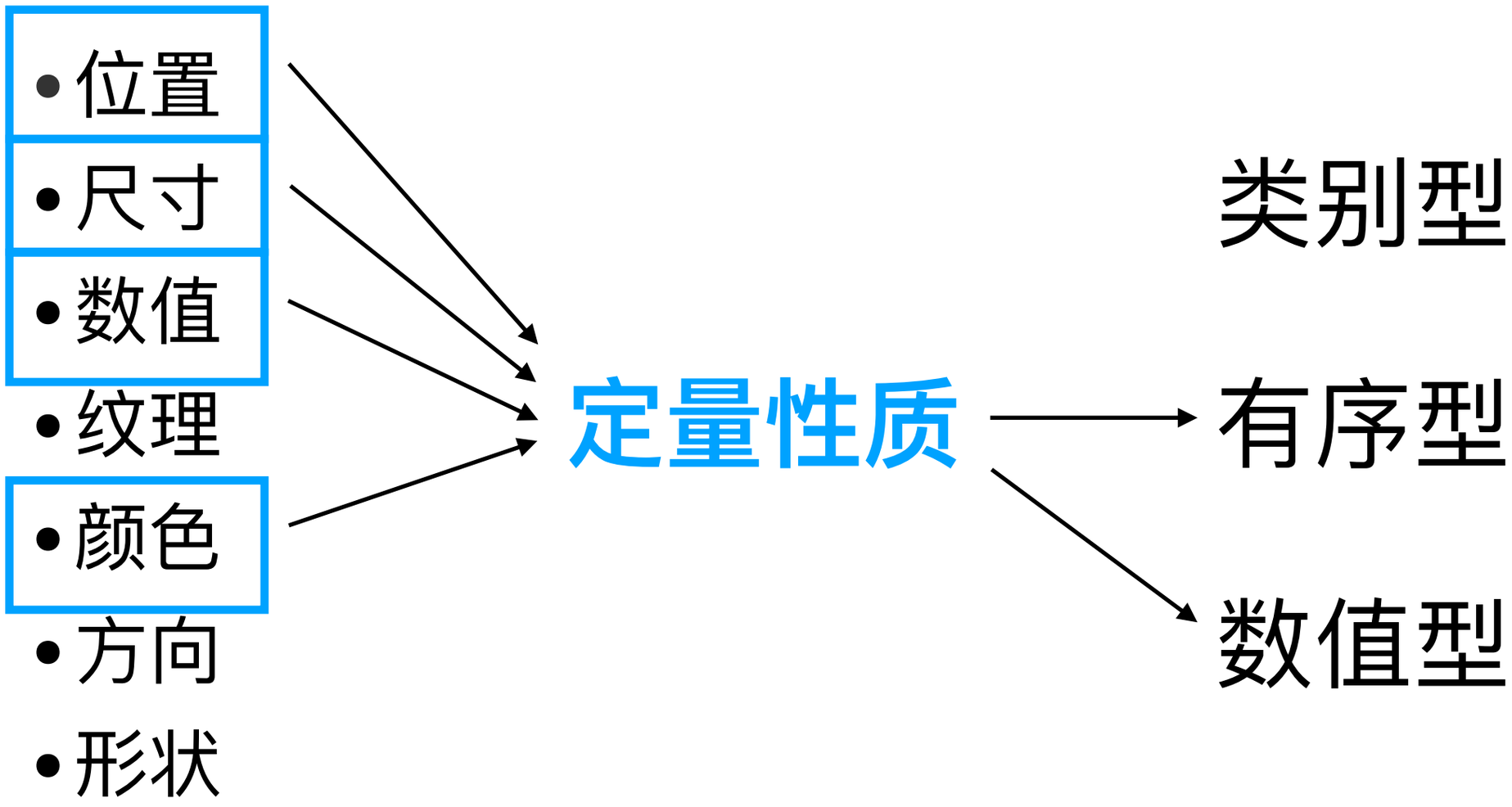
视觉编码

- 位置
- 尺寸
- 数值
- 纹理
- 颜色
- 方向
- 形状



http://geekplux.com/2017/01/03/basics-of-data-visualization-visual-encoding-principles.html

视觉编码



http://geekplux.com/2017/01/03/basics-of-data-visualization-visual-encoding-principles.html

视觉编码

property	marks	ordinal/nominal mapping	quantitative mapping
shape	glyph	○ □ + △ S U	
size	rectangle, circle, glyph, text		
orientation	rectangle, line, text		
color	rectangle, circle, line, glyph, y-bar, x-bar, text, gantt bar		min max

http://geekplux.com/2017/01/03/basics-of-data-visualization-visual-encoding-principles.html

视觉编码

- 色彩搭配
- 交互
- 美学因素
- 信息的密度
- 直观映射、隐喻

http://geekplux.com/2017/01/03/basics-of-data-visualization-visual-encoding-principles.html

数据可视化基础

https://antv.alipay.com/vis/doc/chart/classify/compare.html

数据可视化工具

数据可视化工具

语法级

Snap

配置级

Vega

地理相关

deck.gl

WebGL

Three.js

THANKS

GeekPlux
http://geekplux.com
geekplux@gmail.com