THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

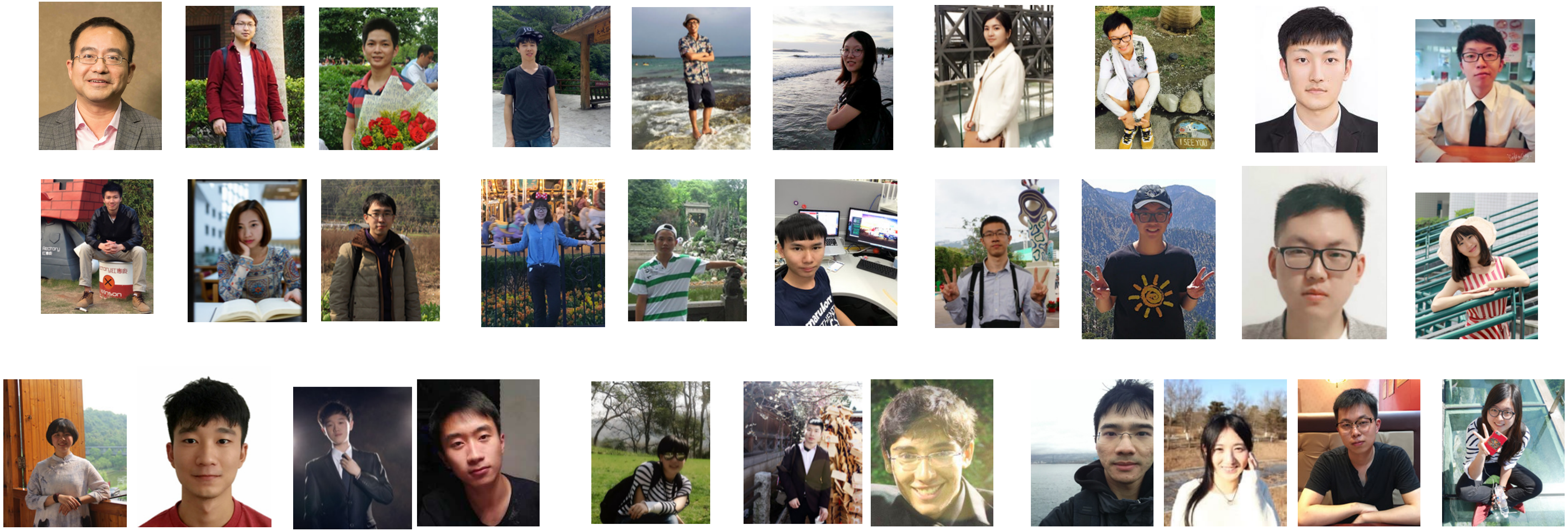
HKUST VIS

数据可视化与可解释性人工智能 （I）

Huamin Qu (屈华民)
HKUST (香港科技大学)

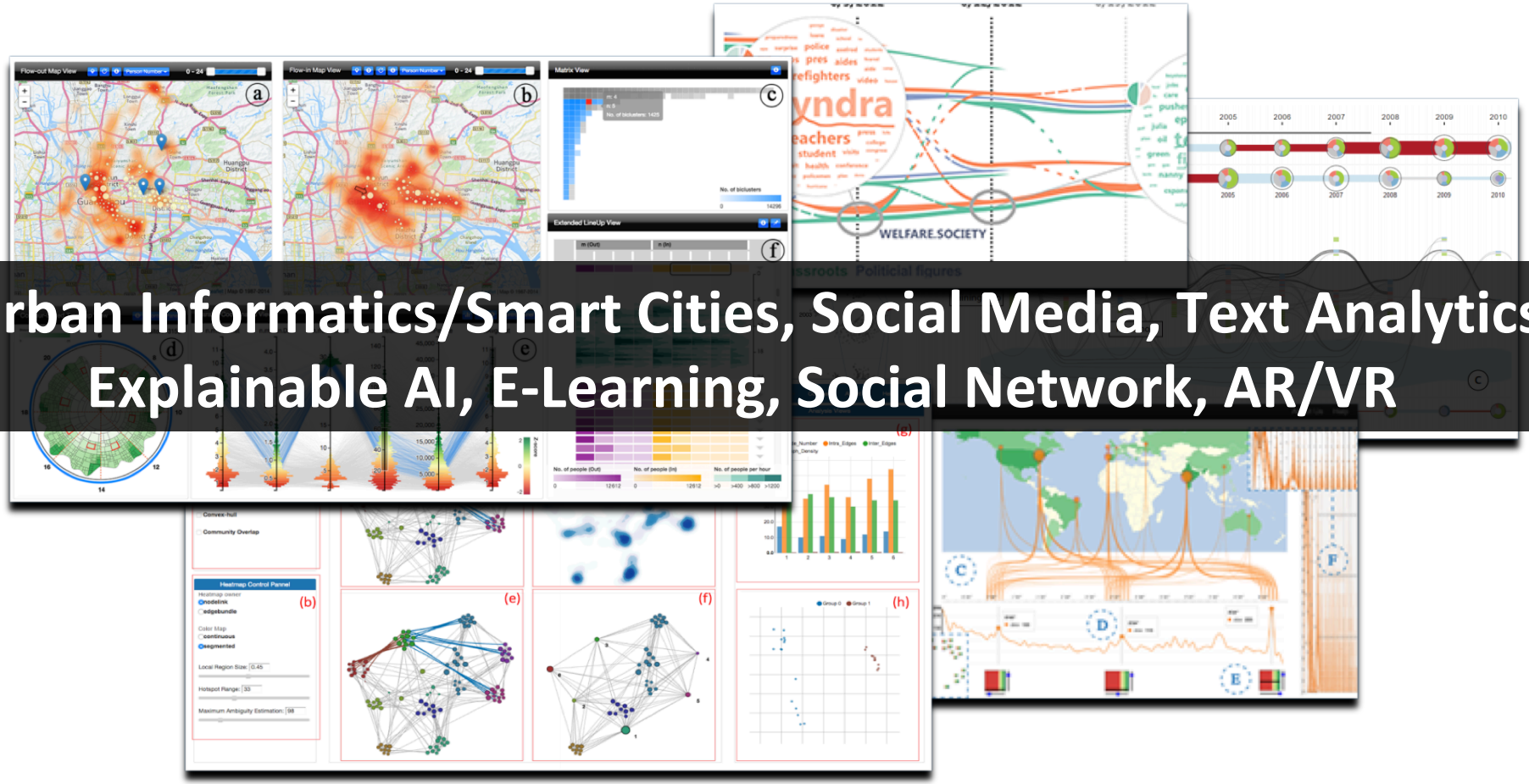
Hong Kong University of Science and Technology

HKUST VisLab



HKUST VisLab

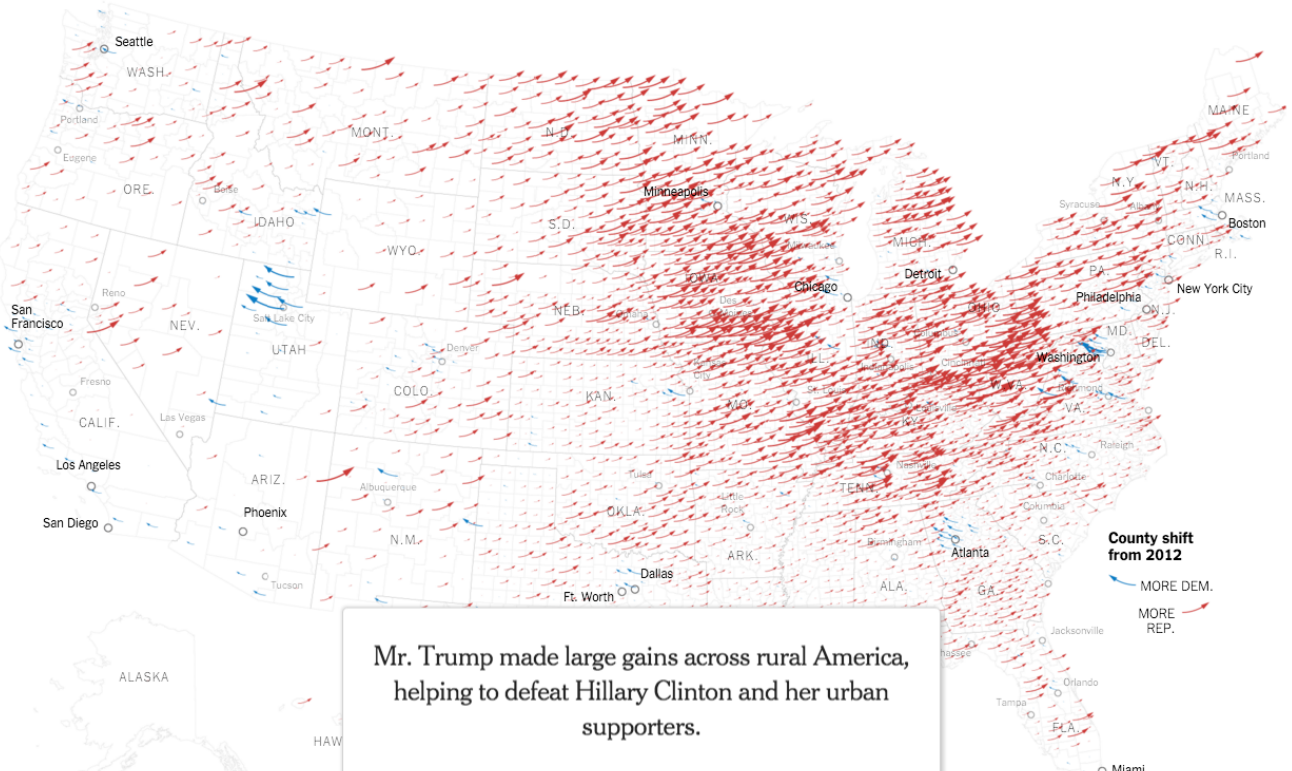
- **Awards**
  - 11 best paper/honorable mention awards
  - Distinguished Collaborator Award 2016 from Huawei Noah's Ark Lab
  - IBM Faculty Award 2009
  - 2014 Higher Education Scientific and Technological Progress Award presented by the Ministry of Education of China
  - HKICT Award and APICTA Award; IEEE VAST Challenge Award
  - Yelp Dataset Challenge Round 10 Grand Prize Award
- **Research**
  - Top 5 visualization group in the world based on the output in the top journal of the field
  - The largest visualization group in Asia and one of the largest in the world (30 members including 23 Ph.D)
  - Technologies adopted by Microsoft, IBM, Huawei, Tencent, Bosch, etc.
- **Alumni**
  - More than 70 (15 PhDs+17 Mphils+…)
  - Working in industry: Microsoft Research Asia, IBM Watson, Bosch Research USA, Siemens, Google, Airbnb, Visa Research, etc.
  - Working in academia: Zhejiang University, Tongji University, University of Electronic Science and Technology of China , Shenzhen University
  - Working in government: Office of the Government Chief Information Officer (OGCIO) Hong Kong



: Urban Informatics/Smart Cities, Social Media, Text Analytics
Explainable AI, E-Learning, Social Network, AR/VR

什么是数据可视化？

- Input: data Output: visual form Goal: insight



Anscombe's Quartet: Four datasets

Anscombes quarte

I		II		III		IV	
x	*y*	*x*	*y*	*x*	*y*	*x*	*y*
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

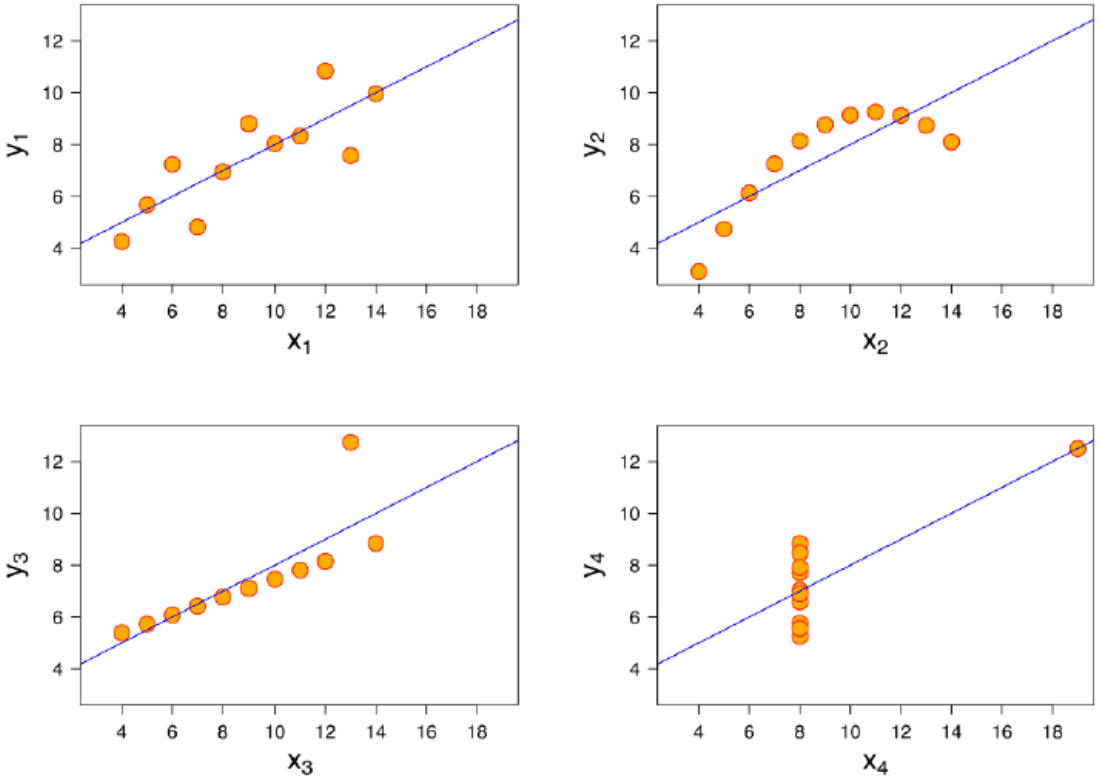
Table 1.1: Anscombe's quartet: four different datasets.

Anscombe's Quartet: Statistics

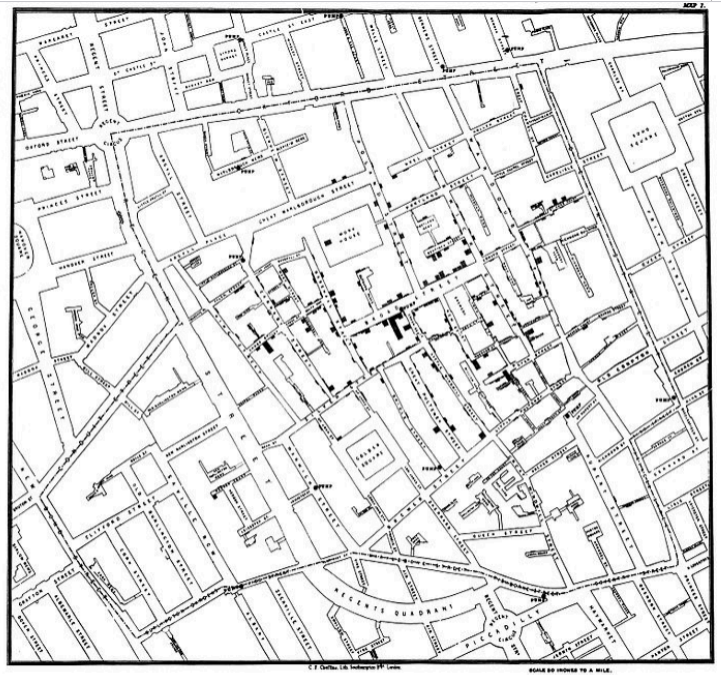
Property (in each set)	Value
Mean of x	9.0
Variance of x	10.0
Mean of y	7.50
Variance of y	3.75
Correlation between x and y	0.898
Linear regression line	$y = 0.5x + 3.0$

Table 1.2: Same statistics in Anscombe's quartet.

Anscombe's Quartet: Visualizations



数据可视化简史



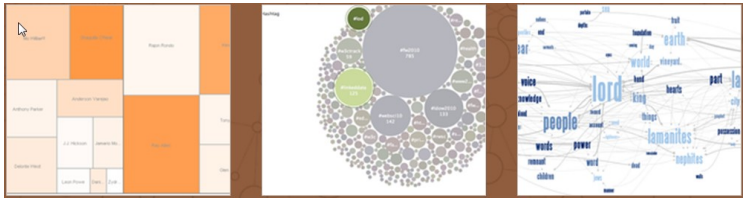
Original map by John Snow showing the clusters of cholera cases (indicated by stacked rectangles) in the London epidemic of 1854. The contaminated pump is located at the intersection of Broad Street and Cambridge Street (now Lexington Street), running into Little Windmill Street.


MARLBOROUGH STREET
POLAND STREET
D'ARBLAY ST
BERWICK STREET
WARDOUR STREET
DEAN STREET
Soho
Carnaby
Street
KINGLY STREET
BEAK STREET
GOLDEN
BRIDLE LANE
GREAT PULTENEY STREET
LEXINGTON STREET
HOPKINS ST
Coll
Sch
REGENT

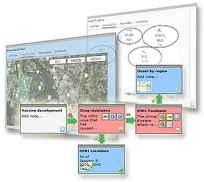
Subfields of Data Visualization

- Scientific Visualization (SciVis) – Spatial data

- Information Visualization (InfoVis)– Abstract data



- Visual Analytics (VAST) – Analytical reasoning



奥巴马的大数据计划

Office of Science and Technology Policy
Executive Office of the President
New Executive Office Building
Washington, DC 20502

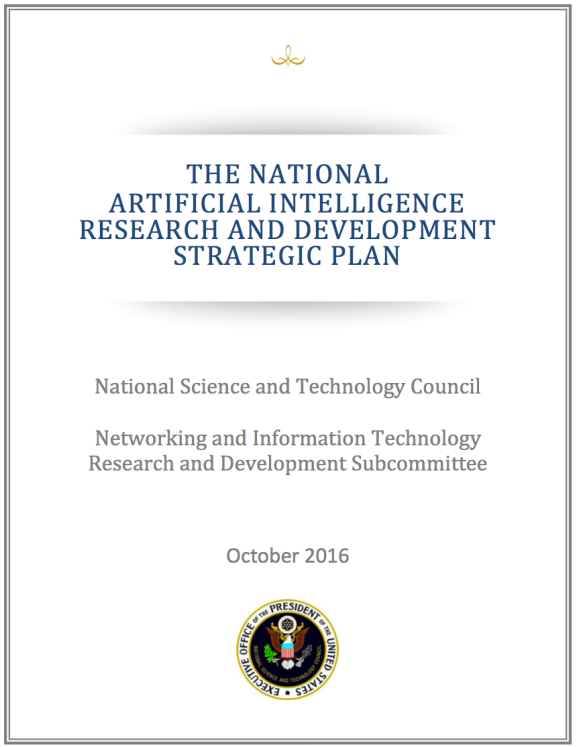
FOR IMMEDIATE RELEASE
March 29, 2012

Contact: Rick Weiss 202 456-6037 rweiss@ostp.eop.gov
Lisa-Joy Zgorski 703 292-8311 lisajoy@nsf.gov

OBAMA ADMINISTRATION UNVEILS "BIG DATA" INITIATIVE: ANNOUNCES $200 MILLION IN NEW R&D INVESTMENTS

- Issuing a $2 million award for a research training group to support training for undergraduates to use graphical and visualization techniques for complex data.

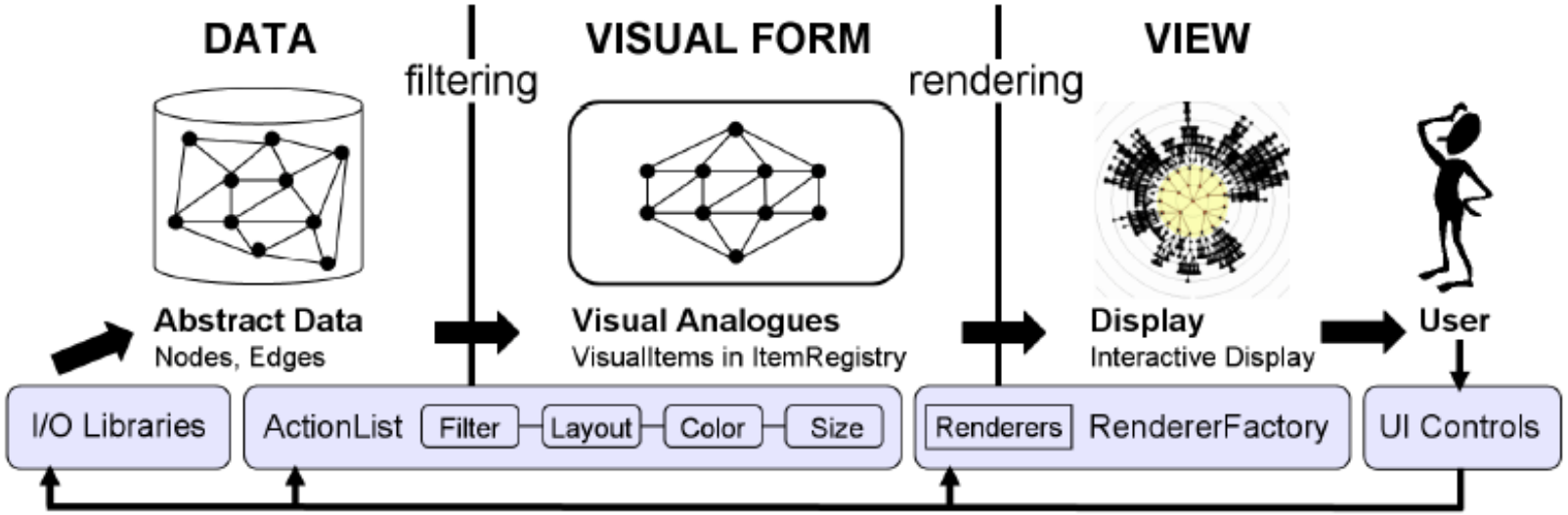
美国人工智能研发计划



Strategy 2: Developing Effective Methods for AI-human Collaboration

Better visualization and user interfaces are additional areas that need much greater development to help humans understand large-volume modern datasets and information coming from a variety of sources.

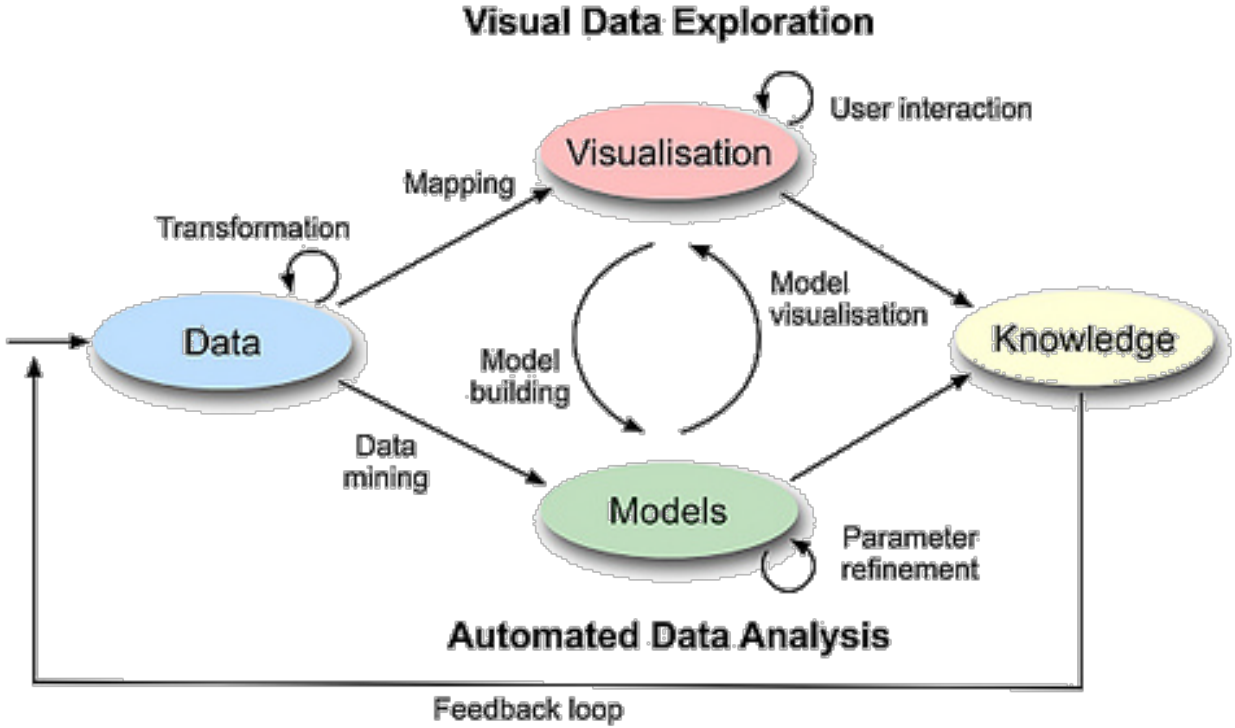
数据可视化的流程：90年代



Engineering part

Art part

数据可视化的流程：00年代



Courtesy of Denial Keim et al.

数据可视化与人工智能

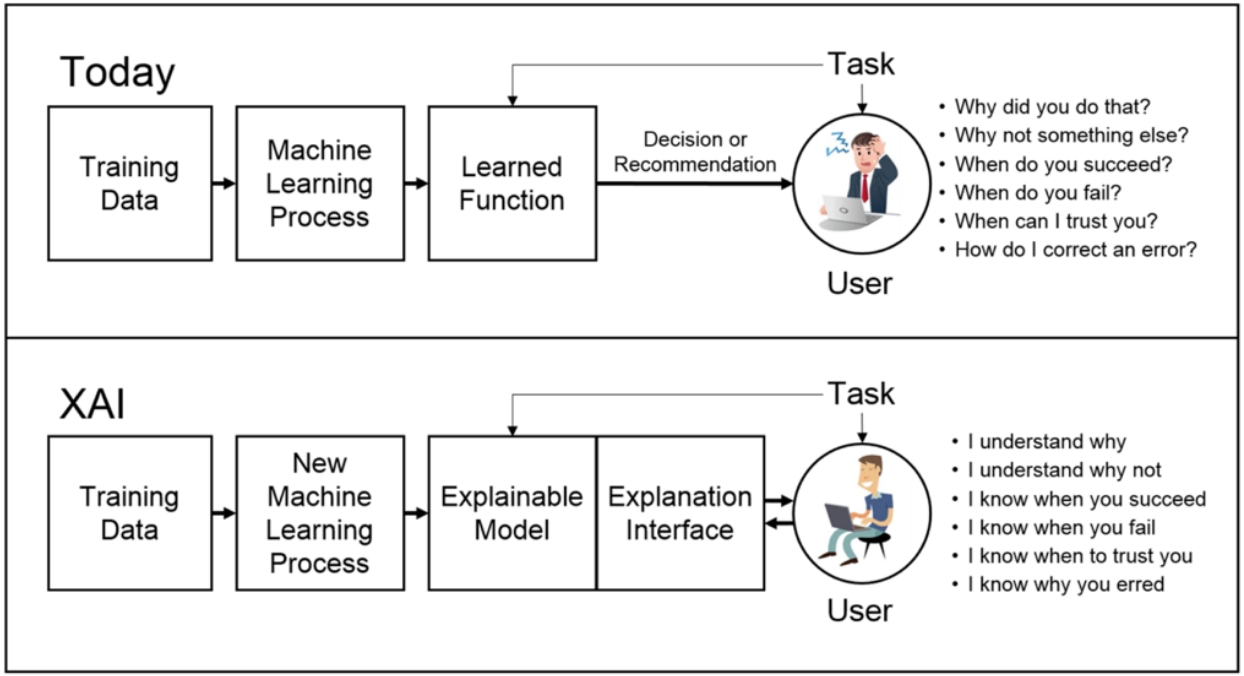


Figure 1: XAI Concept

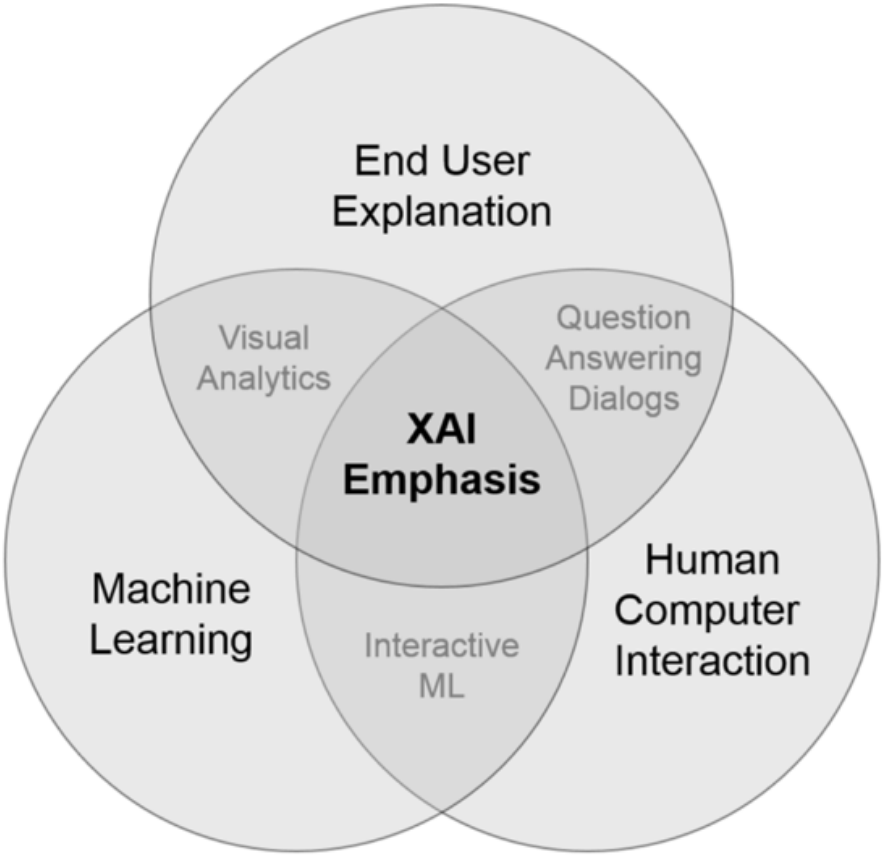
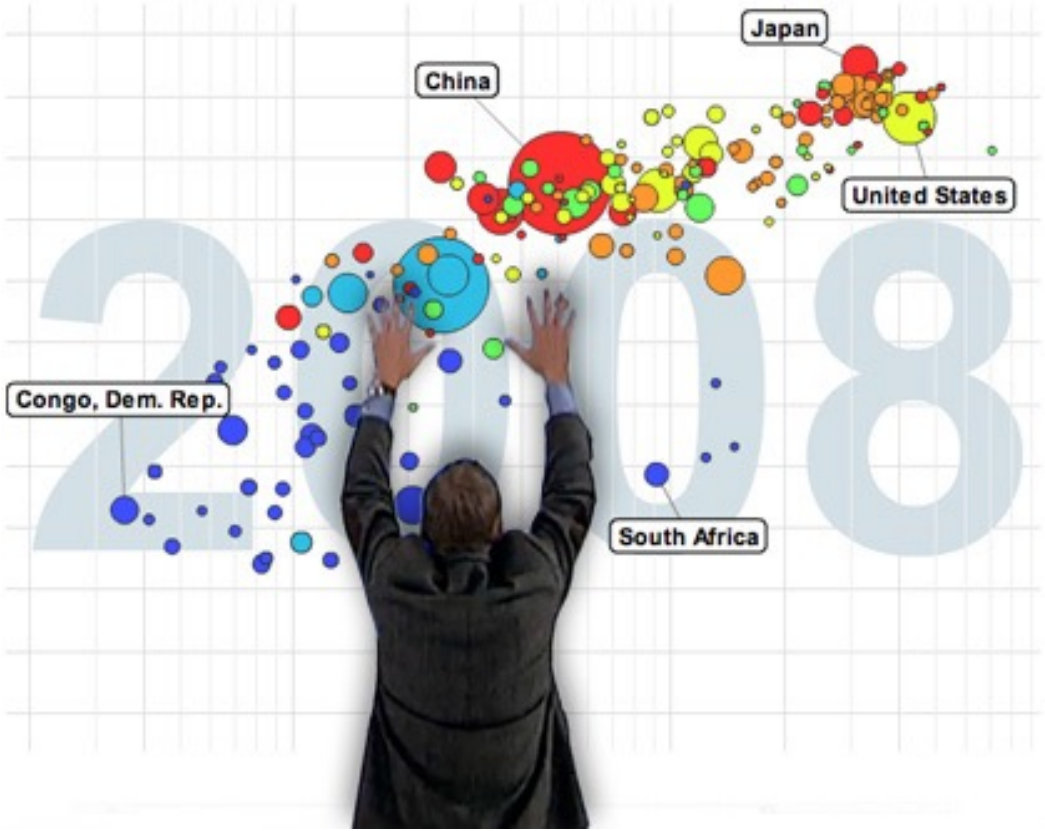


Figure 2: XAI Emphasis

GapMinder



WSJ

THE WALL STREET JOURNAL.

JOURNAL REPORTS: LEADERSHIP

Pictures Make Sense of Big Data

Visualization technology can turn data into pictures that are far more comprehensible

By DEBORAH GAGE

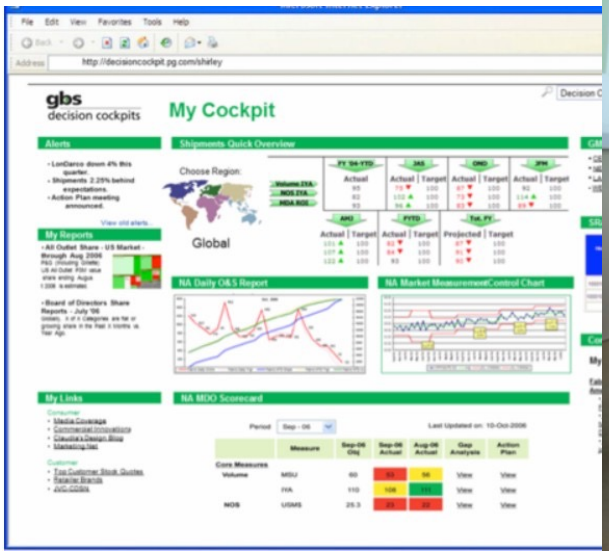
Updated Sept. 15, 2013 5:18 p.m. ET

Most people have trouble recalling strings of numbers that are longer than their phone numbers. So how do we begin to comprehend a hundred rows of data, let alone a thousand or a million or a billion rows?

HBR Blog Network

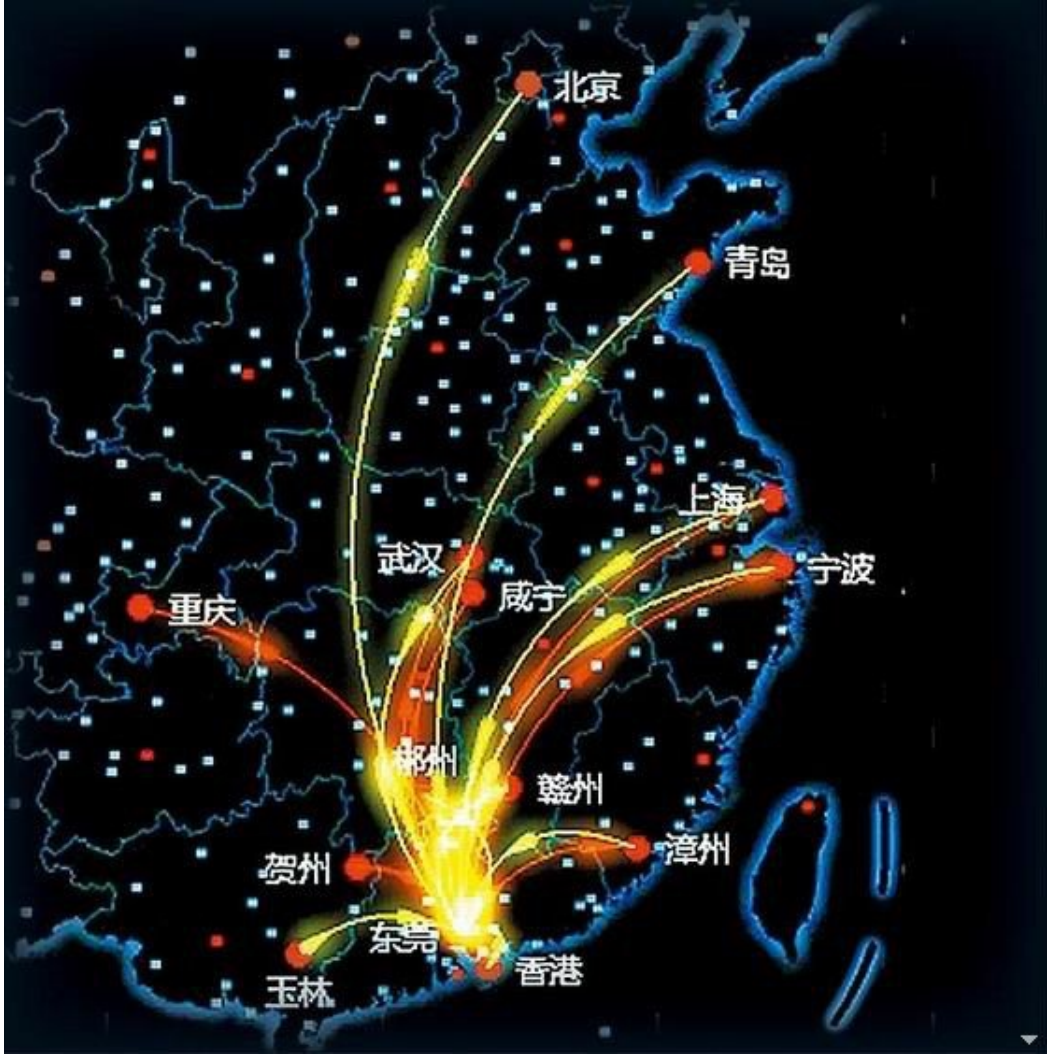
How P&G Presents Data to Decision-Makers

by Tom Davenport | 3:00 PM April 4, 2013



Make analytics easy.
For analysts, executives, IT, everyone.


CCTV 13
新 闻
北京
合肥
上海
数据说春节
新闻联播携手百度"数据说春节"


北京
青岛
上海
武汉
咸宁
宁波
重庆
郴州
赣州
漳州
贺州
东莞
香港
玉林

＋可视化

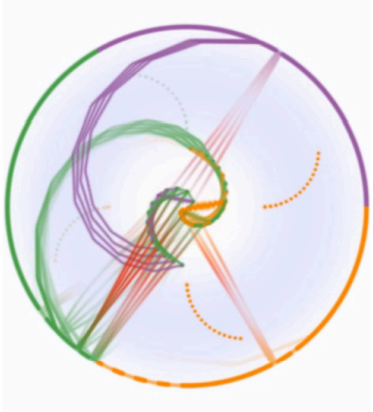
可视化

＋

Cognitive Science

Urban Informatics

Computer Graphics



Medical imaging

Human-Computer Interaction

Computational fluid dynamics

Machine learning

Social media

Graphical design/visual communication

Social science

可视化的标准

- 信：Accuracy
- 达：Intuitiveness, Efficiency, and Effectiveness
- 雅：Aesthetic

信 达 雅

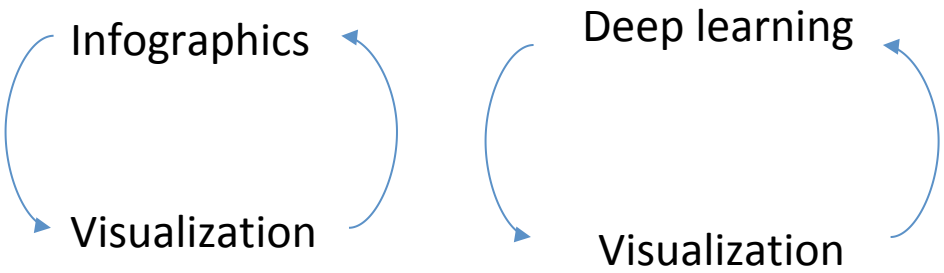
> *The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the quantities represented.*
>
> [Tufte, 1991]

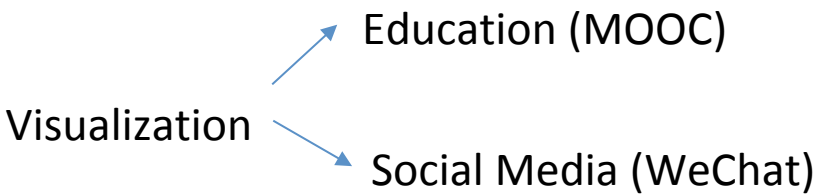
$$Lie\ Factor = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

where

$$size\ of\ effect = \frac{|\text{second value - first value}|}{\text{first value}}$$

Cognitive Science → Visualization

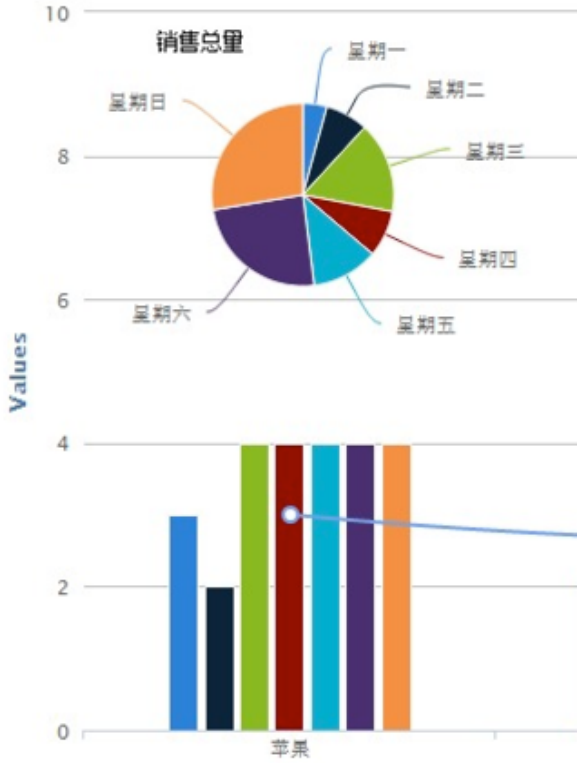




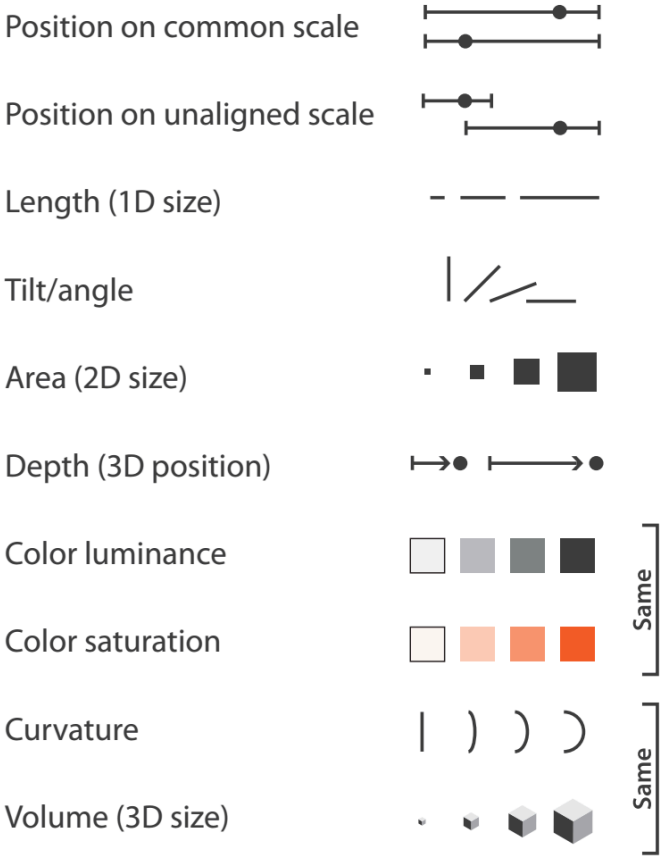
Cognitive Science

- Gestalt law
- Popout effect
- Visual encoding principles:
  - Effectiveness principles
  - Expressiveness principles

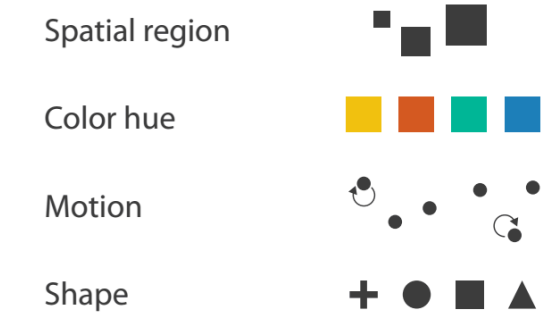
Which one is more effective?



Magnitude Channels: Ordered Attributes



Identity Channels: Categorical Attributes

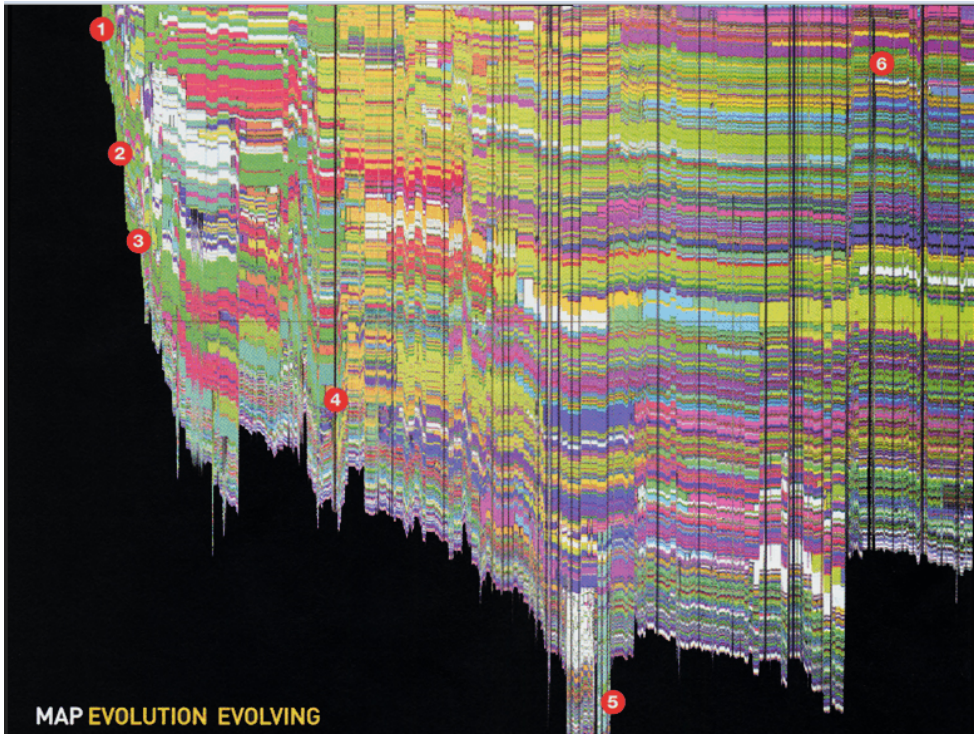


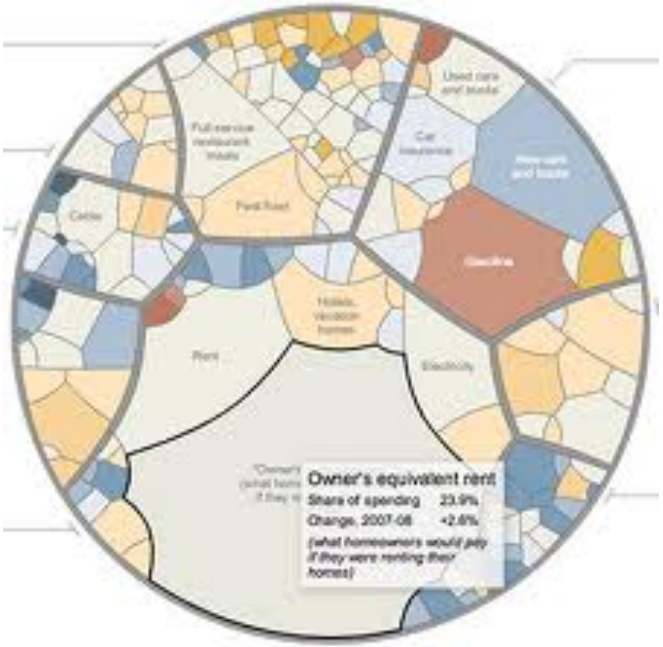
- effectiveness principle
  - encode most important attributes with highest ranked channels
- expressiveness principle
  - match channel and data characteristics

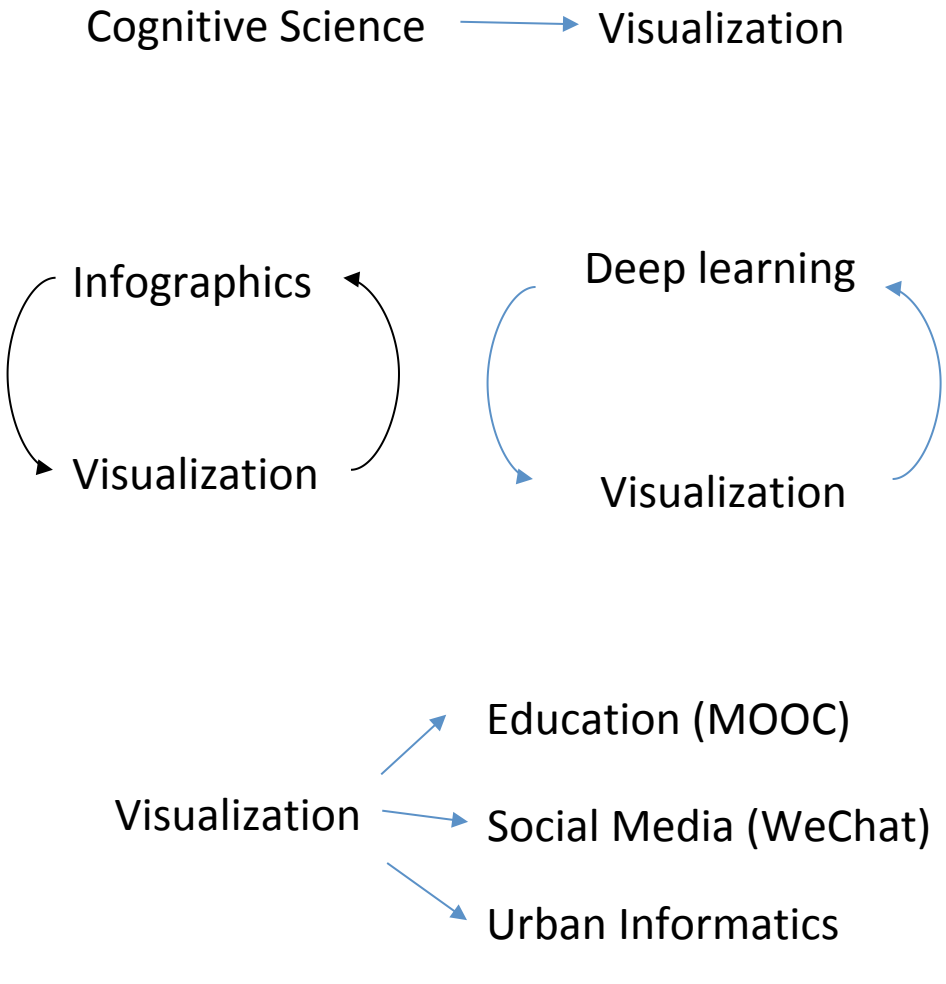
可视化的标准

- 信：Accuracy
- 达：Intuitiveness, Efficiency, and Effectiveness
- 雅：Aesthetic

信 达 雅


1
2
3
4
5
6
MAP EVOLUTION EVOLVING


Owner's equivalent rent
Share of spending 23.9%
Change, 2007-08 +2.6%
(what homeowners would pay if they were renting their homes)


Cognitive Science
Visualization
Infographics
Visualization
Deep learning
Visualization
Visualization
Education (MOOC)
Social Media (WeChat)
Urban Informatics

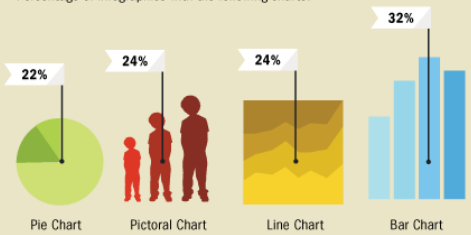
INFOGRAPHIC OF INFOGRAPHICS

Data visualization is a popular new way of sharing research. Here is a look at some of the visual devices, informational elements, and general trends found in the modern day infographic.

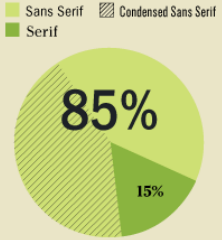
DESIGN

CHART STYLE

Percentage of infographics with the following charts:

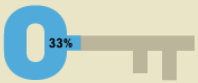


FONT



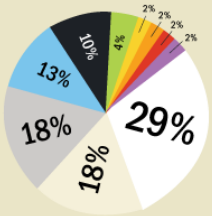
KEY INFO

Percentage of infographics with key:



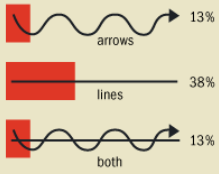
Average number of symbols per key: **5.1**

BASE COLOR



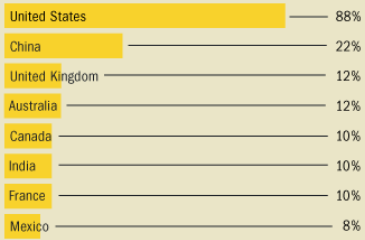
NAVIGATIONAL ICONOGRAPHY

Frequency of arrows & connecting lines in infographics:



CONTENT

COUNTRIES FEATURED



THEME

Relative popularity of different infographic themes:

SECTIONS

CREDITED SOURCES

Average number of sources per infographic: **2.29**

TITLE

Average number of words per infographic title: **4.36**

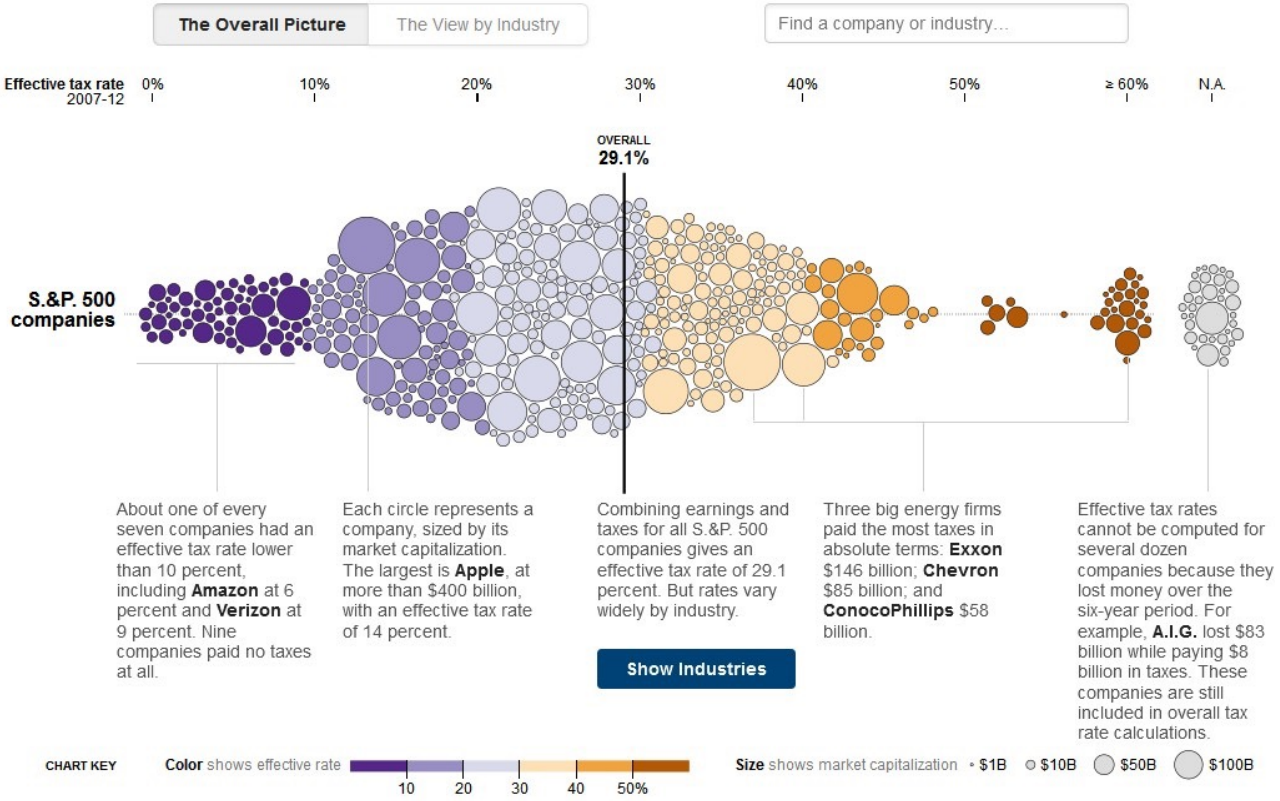
"RICHEST AND POOREST AMERICAN NEIGH

CONCEPT & DESIGN Ivan Cash **SOURCE** 49 infographics collected at random from www.good.is/infographics

The New York Times

2013: The Year in Interactive Storytelling

Multimedia Stories | Data Visualization | Explanatory Graphics | Breaking News | Visual and Interactive Features



IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 20, NO. 12, DECEMBER 2014

LoyalTracker: Visualizing Loyalty Dynamics in Search Engines

Conglei Shi, Yingcai Wu, *Member, IEEE*, Shixia Liu, *Senior Member, IEEE*, Hong Zhou and Huamin Qu, *Member, IEEE*

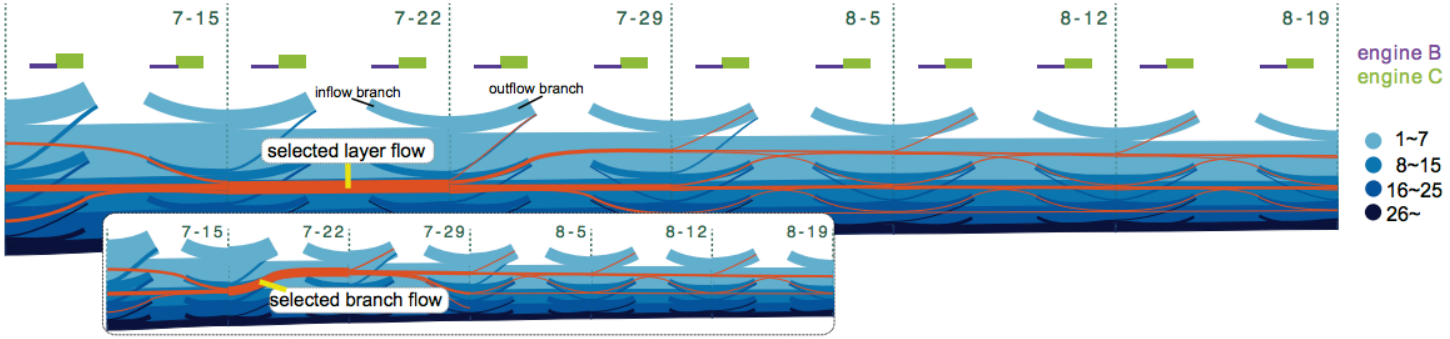
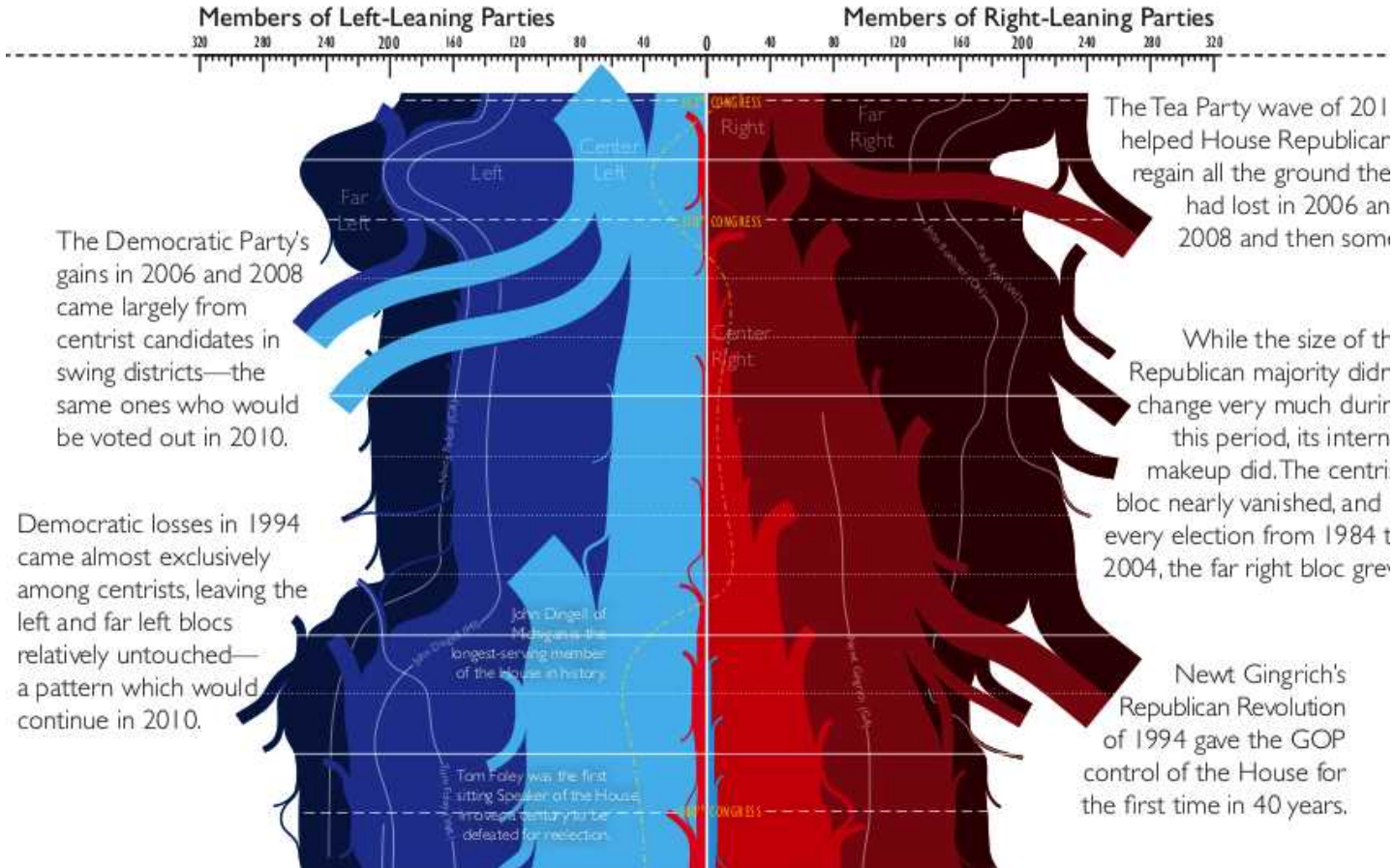


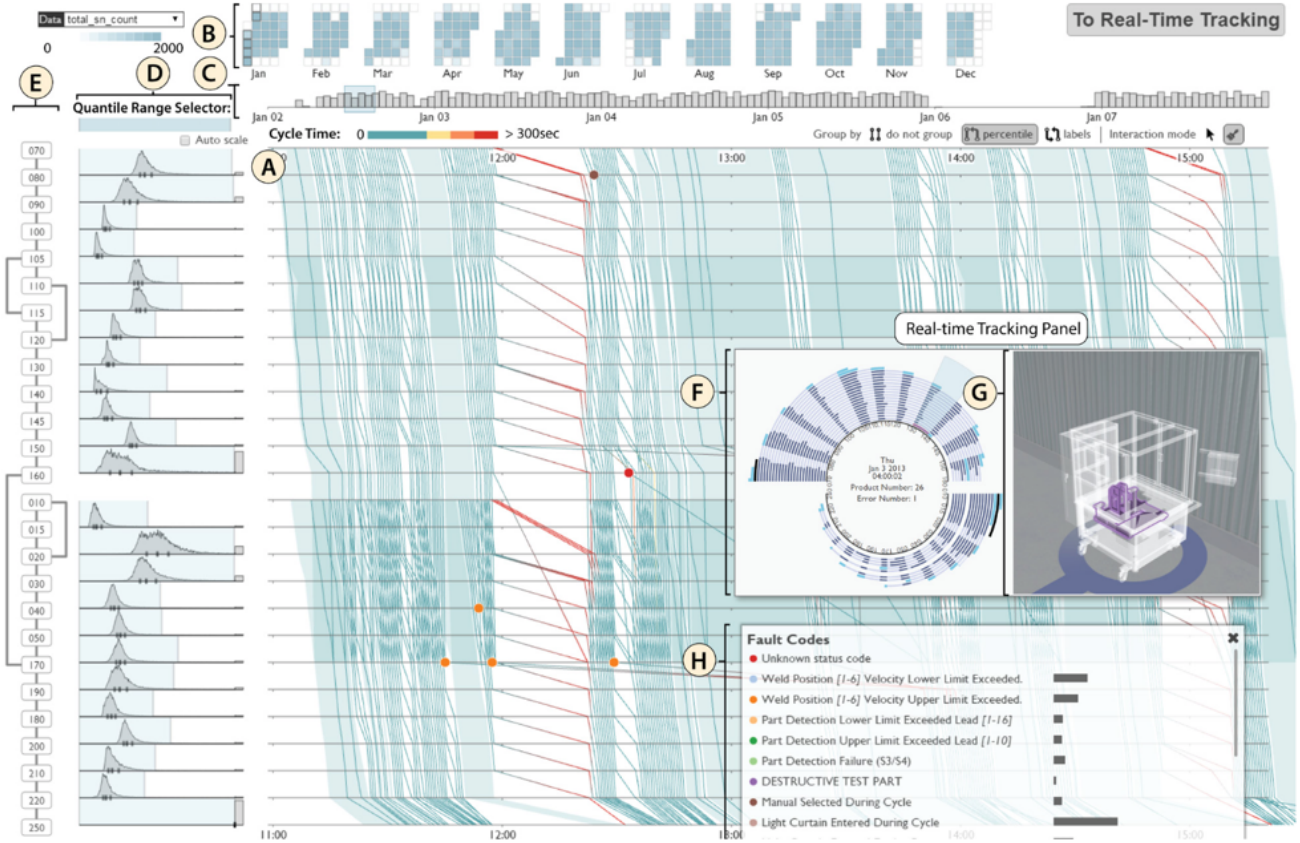
Fig. 1: LoyalTracker illustrates loyalty dynamics of the users using search engine **A**. Top and bottom show the same flow view that highlights two different flowing patterns of the users (in orange) selected from a layer flow (top) and a branch flow (bottom) across multiple loyalty categories (layers) over time. The switching histogram on the top shows a visual summary of switching behavior.

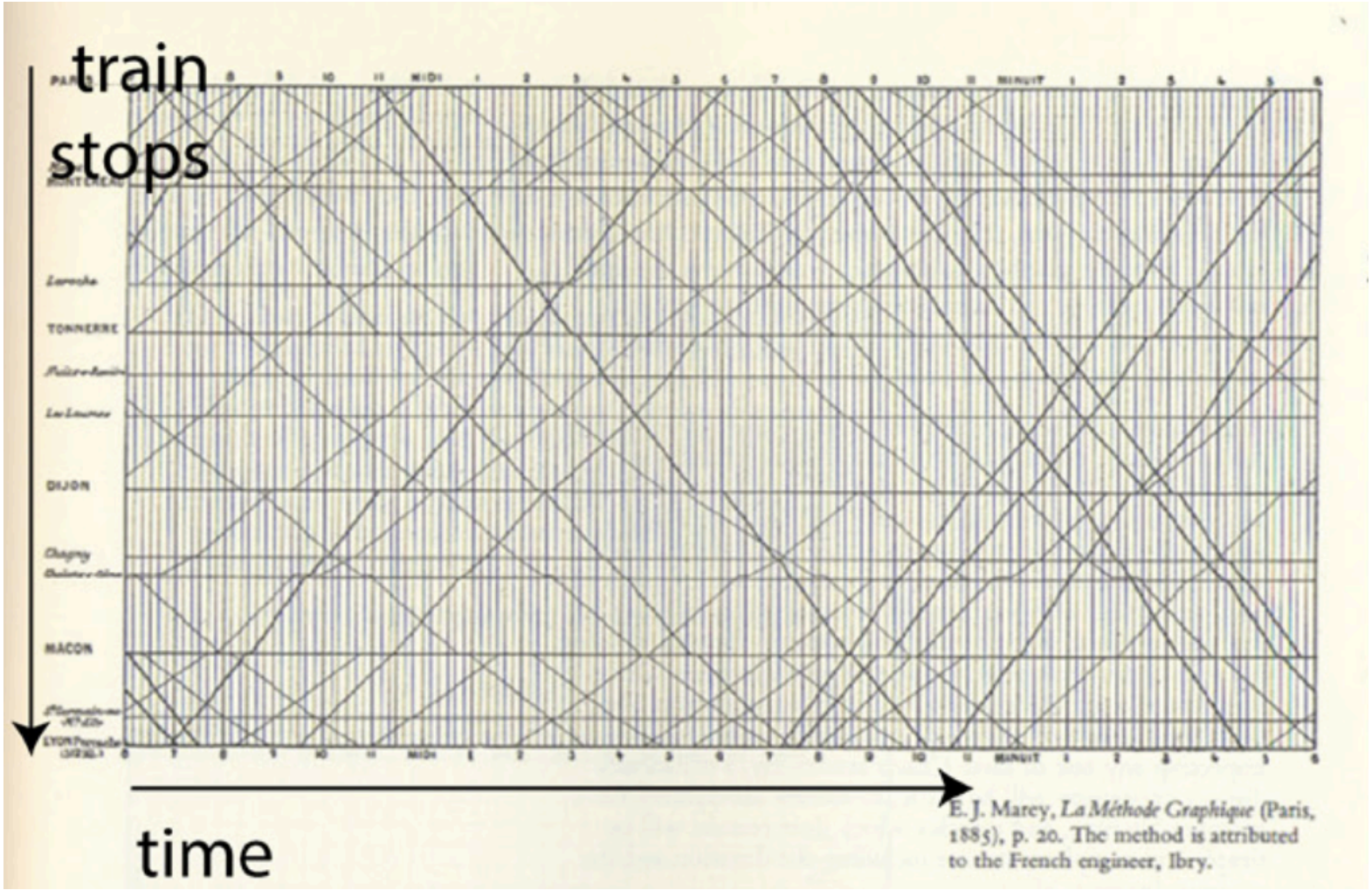
HOUSE


Members of Left-Leaning Parties
320 280 240 200 160 120 80 40 0
Members of Right-Leaning Parties
40 80 120 160 200 240 280 320
Far Left
Left
Center Left
Center Right
Right
Far Right
The Democratic Party's gains in 2006 and 2008 came largely from centrist candidates in swing districts—the same ones who would be voted out in 2010.
Democratic losses in 1994 came almost exclusively among centrists, leaving the left and far left blocs relatively untouched—a pattern which would continue in 2010.
John Dingell of Michigan is the longest-serving member of the House in history.
Tom Foley was the first sitting Speaker of the House in nearly a century to be defeated for reelection.
The Tea Party wave of 201 helped House Republicar regain all the ground the had lost in 2006 an 2008 and then som
While the size of th Republican majority didr change very much durir this period, its intern makeup did. The centri bloc nearly vanished, and every election from 1984 t 2004, the far right bloc gre
Newt Gingrich's Republican Revolution of 1994 gave the GOP control of the House for the first time in 40 years.

ViDX: Visual Diagnostics of Assembly Line Performance in Smart Factories

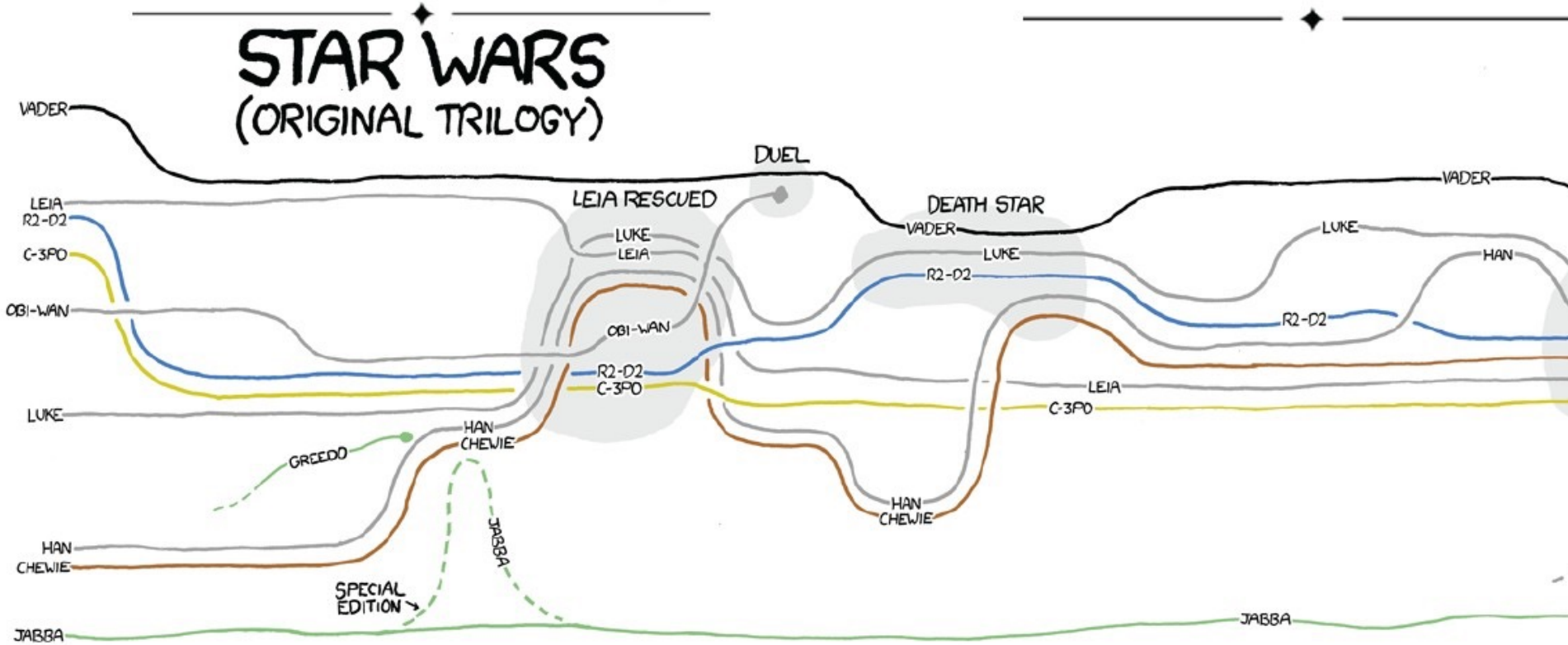
Panpan Xu, Honghui Mei, Liu Ren, and Wei Chen



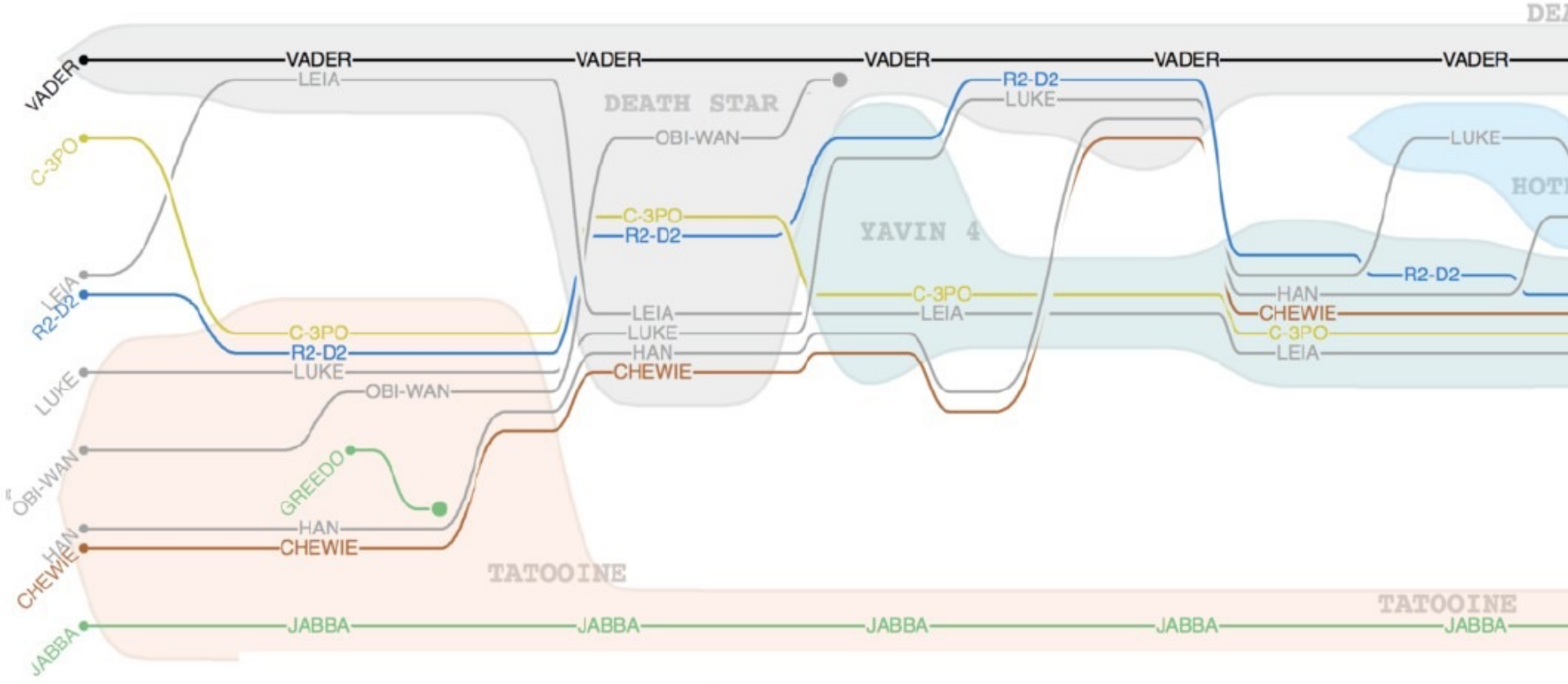


E. J. Marey, *La Méthode Graphique* (Paris, 1885), p. 20. The method is attributed to the French engineer, Ibry.

StoryLine


STAR WARS
(ORIGINAL TRILOGY)
VADER
DUEL
LEIA
R2-D2
C-3PO
OBI-WAN
LUKE
LEIA RESCUED
LUKE
LEIA
OBI-WAN
R2-D2
C-3PO
HAN
CHEWIE
GREEDO
HAN
CHEWIE
SPECIAL EDITION
JABBA
JABBA
DEATH STAR
VADER
LUKE
R2-D2
HAN
CHEWIE
VADER
LUKE
R2-D2
LEIA
C-3PO
LUKE
HAN
R2-D2
JABBA

StoryLine


VADER
VADER
VADER
VADER
VADER
VADER
LEIA
DEATH STAR
OBI-WAN
C-3PO
R2-D2
LUKE
C-3PO
R2-D2
YAVIN 4
LUKE
HOTH
R2-D2
LEIA
R2-D2
C-3PO
LEIA
HAN
CHEWIE
C-3PO
LEIA
C-3PO
R2-D2
LUKE
LEIA
LUKE
HAN
CHEWIE
LUKE
OBI-WAN
OBI-WAN
GREEDO
HAN
CHEWIE
HAN
CHEWIE
TATOOINE
TATOOINE
JABBA
JABBA
JABBA
JABBA
JABBA
JABBA

Data-Driven Guides: Supporting Expressive Design for Information Graphics

Nam Wook Kim, Eston Schweickart, Zhicheng Liu, Mira Dontcheva, Wilmot Li, Jovan Popovic, and Hanspeter Pfister

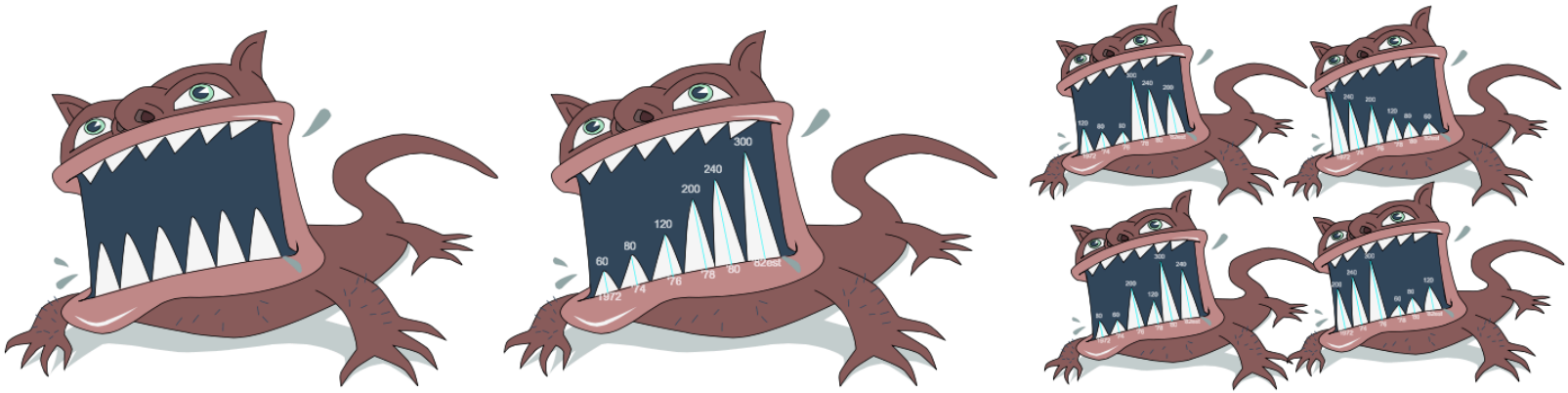
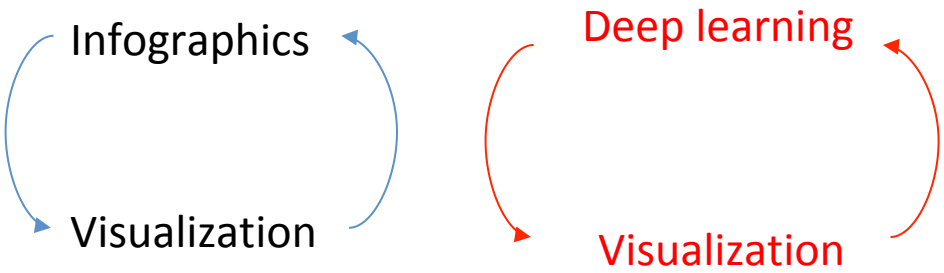


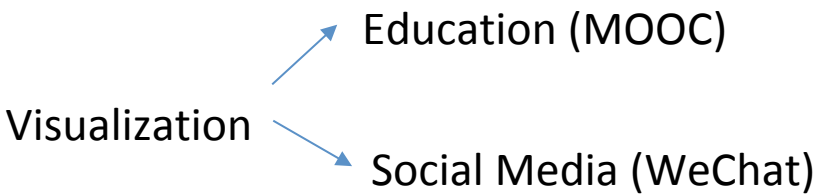
Fig. 1: Nigel Holmes' *Monstrous Costs* chart, recreated by importing a monster graphic (left) and retargeting the teeth of the monster with DDG (middle). Taking advantage of the data-binding capability of DDG, small multiples are easily created by copying the chart and changing the data for each cloned chart (right).

可视化的标准

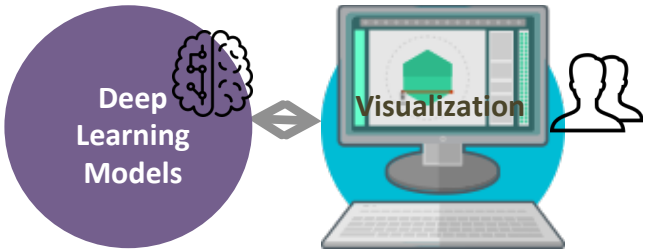
- 信：Accuracy
- 达：Intuitiveness, Efficiency, and Effectiveness
- 雅：Aesthetic

Cognitive Science → Visualization





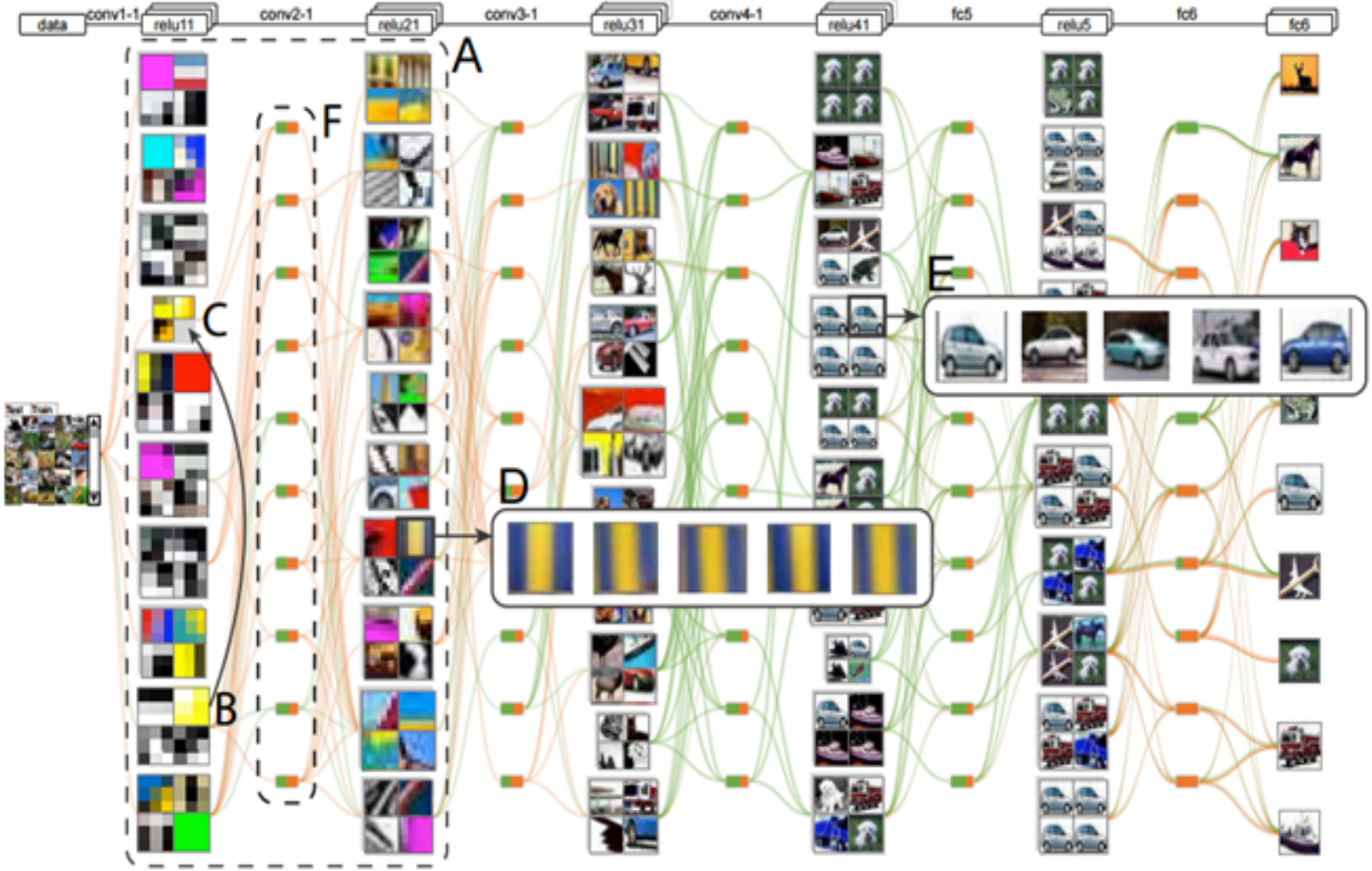
Keep Human in the AI loop



Visualization techniques to facilitate

- Debugging
- Understanding
- Performance improvement

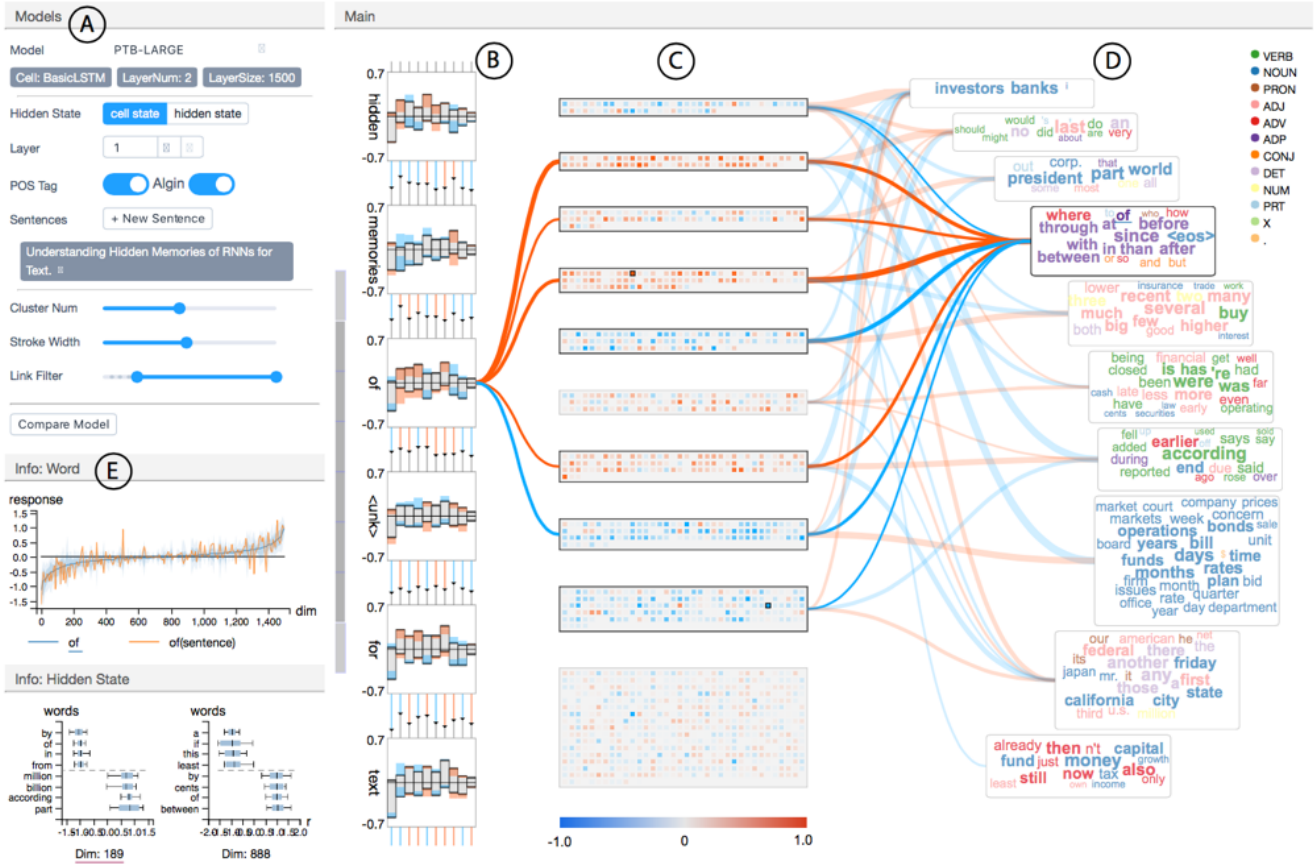
of deep learning models


data
conv1-1
relu11
conv2-1
relu21
conv3-1
relu31
conv4-1
relu41
fc5
relu5
fc6
fc6
A
B
C
D
E
F

Understanding Hidden Memories of Recurrent Neural Networks

Category: Research

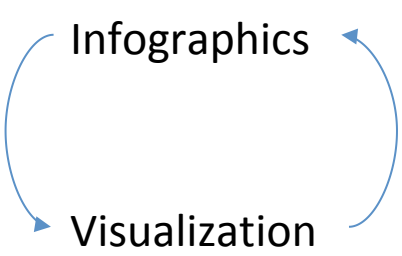
Paper Type: application/design study

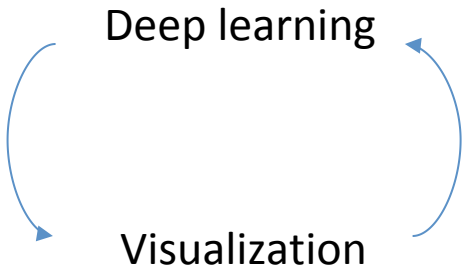


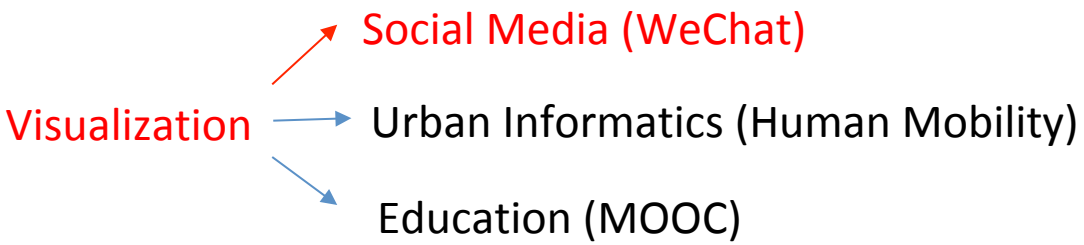
Research Focuses

Urban Informatics, **S**ocial Media, **T**ext Visualization
VisMOOCs, **I**mmersive Analytics, **S**ocial Network

Cognitive Science → Visualization







微信小秘密: 2016 年那些 10w+ 文章是怎么刷爆朋友圈的?

Original 2016-12-30 WeChat TechPower WeChat TechPower

Scan QR Code via We Chat
to follow Official Acco unt

告诉大家一个悲伤的消息，2016 的进度条即将告罄。在 2016 年最后一个工作日，我们还是踏实学点东西吧，和小编一起回头看看这一年发生的大事，又有哪些热点曾经刷爆了我们的朋友圈?

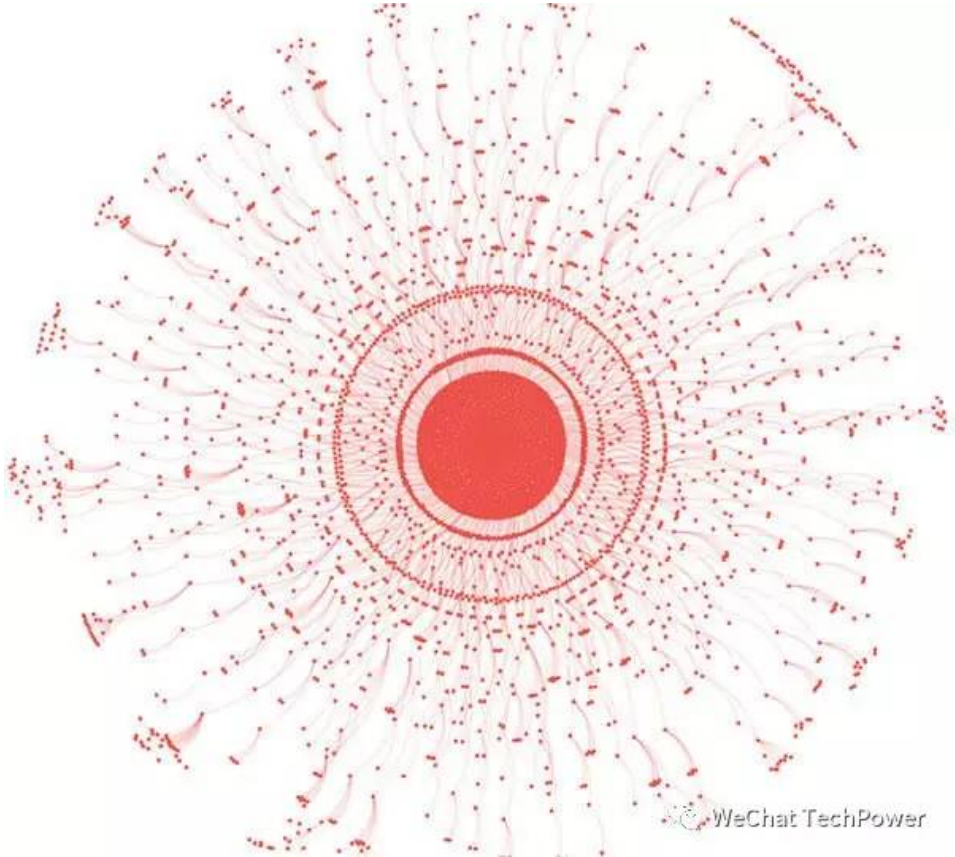
http://mp.weixin.qq.com/s?__biz=MzI5MDAwOTIzOQ==&mid=2650901442&idx=1&sn=ae2b21bb

【盘点】2016上半年最全的热点网络事件，你关注了几个？

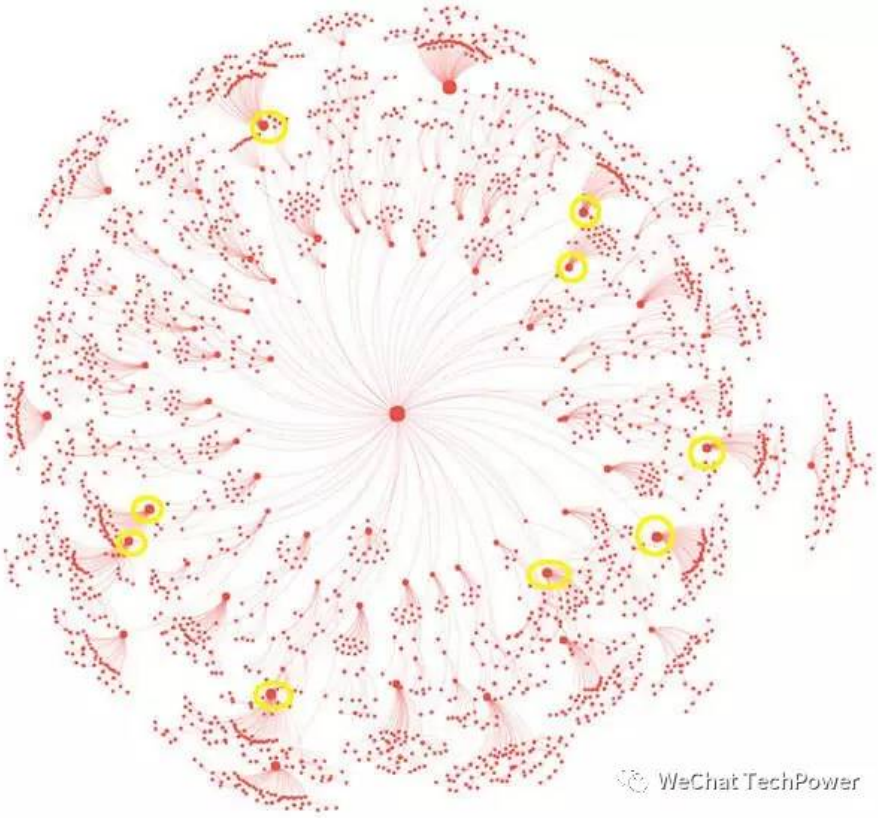
2016年已经过去一半了，下半年的第一天也快华丽丽的过去，回首看去，上半年发生了好多事情，今天我们就来盘点一下上半年发生的“大事情”，排名不分先后~

1、“2016中国第一网红”——papi酱

热点一：AlphaGo 大战李世石



热点二：iPhone7 发布会



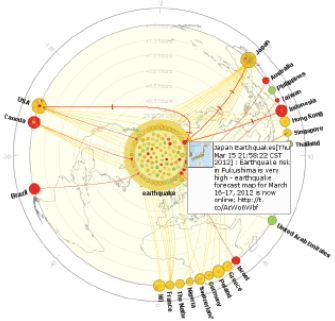
热点三：川普当选

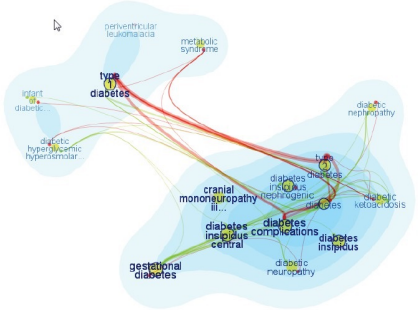
Some hot topics

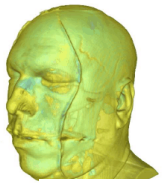
- VIS + Machine Learning: Interactive/Explainable Artificial Intelligence
- VIS + Storytelling: Narrative Visualization
- VIS + Augmented reality: Immersive Analytics

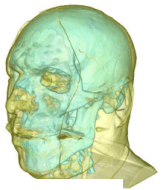
Thank You!

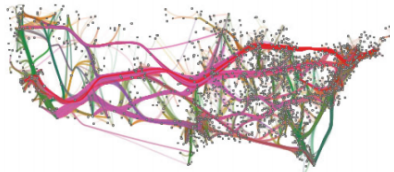

system
visualization
virtual
research
computer
graphics
interactive
interfaces
software
model
data
applications
environment
user
information
human
feat
method
process
simulation
object
based
approach
analysis
algorithm
rendering

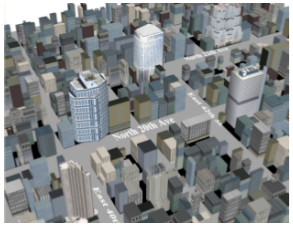


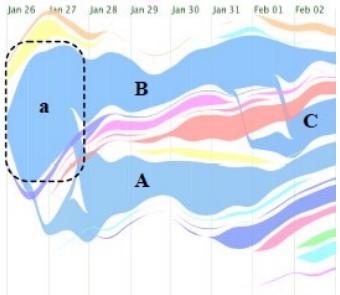

type 1 diabetes
metabolic syndrome
diabetic nephropathy
cranial mononeuropathy
diabetes insipidus
diabetes complications
diabetic ketoacidosis
gestational diabetes
diabetic neuropathy

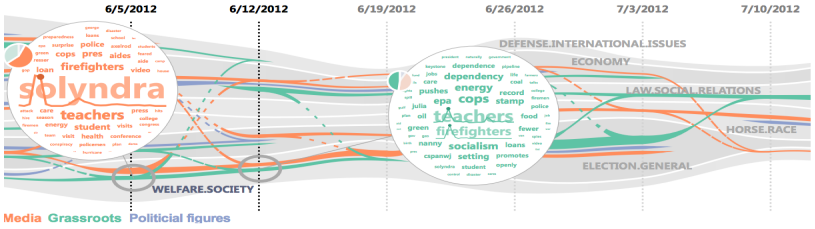









Jan 26
Jan 27
Jan 28
Jan 29
Jan 30
Jan 31
Feb 01
Feb 02
a
B
C
A


6/5/2012
6/12/2012
6/19/2012
6/26/2012
7/3/2012
7/10/2012
solyndra
police
cops
pres
firefighters
loan
teachers
energy
student
DEFENSE.INTERNATIONAL.ISSUES
ECONOMY
LAW.SOCIAL.RELATIONS
HORSE.RACE
ELECTION.GENERAL
WELFARE.SOCIETY
dependency
energy
cops
stamp
pushes
epa
tAachers
firefighters
socialism
setting
Media
Grassroots
Politicial figures


INNOVATING TODAY
IMAGINING TOMORROW
HKUST