香港浸會大學
HONG KONG BAPTIST UNIVERSITY

新聞系
DEPARTMENT OF JOURNALISM

Introduction to Data-driven Journalism

Lectured by: Dr. Xinzhi ZHANG

Research Assistant Professor, Department of Journalism

Hong Kong Baptist University

12 June, 2018

@CUCN Data-driven Journalism Workshop

Workshop Schedule: Day 1

Date	Time	Content	Venue	Remarks
D1	AM1	Introduction to data-driven Journalism	Hall	Public seminar
D1	AM2	Issues and cases of data-driven journalism	Hall	Public seminar
D1	PM1	News topic brainstorming, news sourcing, preliminary data collection	Classroom	Hands-on tutorial Team-up
	PM2	Data processing: concepts, measurements, coding	Classroom	Workshop + tutorial
	E	Data collection, drafting the outline and preparing texts	Self-organized	Initial draft done (the text part should be completed)

Workshop Schedule: Day 2

Date	Time	Content	Venue	Remarks
D2	AM1	Data analysis • Fundamentals of Excel • Descriptive statistics • Correlation relations of the variables • Pivot Table, OLAP	Classroom	Workshop + tutorial
	AM2	Data Visualization • Basic charts • Multi-dimensional data visualization	Classroom	Workshop + tutorial
	PM	Extended data types and visualization • Network • Map • Text mining	Classroom	Workshop + tutorial
	E	Data analysis and reporting writing	Self-organized	A complete story done

Workshop Schedule: Day 3

Date	Time	Content	Venue	Remarks
D3	AM	Project discussion and consultation	Classroom	Interactive session Camera-ready for presentation
	PM	Project presentation + feedbacks	Hall	Demonstration

Xinzhi ZHANG [張昕之], Ph.D.

- **Current Position**
  - **Research Assistant Professor**, Department of Journalism, Hong Kong Baptist University
  - **Associate Programme Director**, the Data and Media Communication Concentration (DMC)

- **Professional Experience**
  - **Research Assistant Professor**, Department of Journalism, Hong Kong Baptist University (2016 – Present)
  - **Lecturer**, School of Professional Education & Executive Development, The Hong Kong Polytechnic University (2014 – 16)
  - **Postdoctoral Fellow**, Department of Media and Communication, City University of Hong Kong (2013 – 14)

- **Education**
  - **Ph.D.** in Media and Communication, City University of Hong Kong (2009 – 13)
  - **M.A.** with Distinction, in Communication and New Media, City University of Hong Kong (2008 – 09)
  - **B.A.** in Broadcasting Journalism, Guangzhou University, China (2004 – 08)

Agenda

1. What is data-driven journalism
2. Data-driven investigation
3. Data-driven presentation
4. The intellectual origins of data-driven journalism
5. The flow of data-driven news production
6. Major players of data-driven journalism

How to present a story?

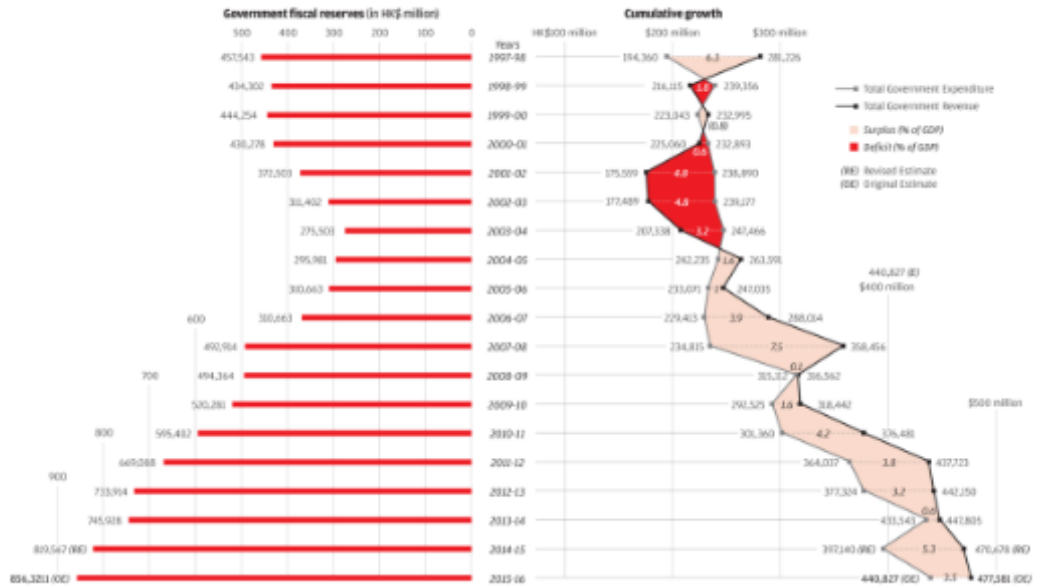
- “In photography, the smallest thing can be a great subject. The little human detail can become a leitmotiv” -- Henri Cartier-Bresson (1952). *The Decisive Moment*.
- The cartoon uses signs, especially symbols, to illustrate the meanings.
- How about data-driven journalism?

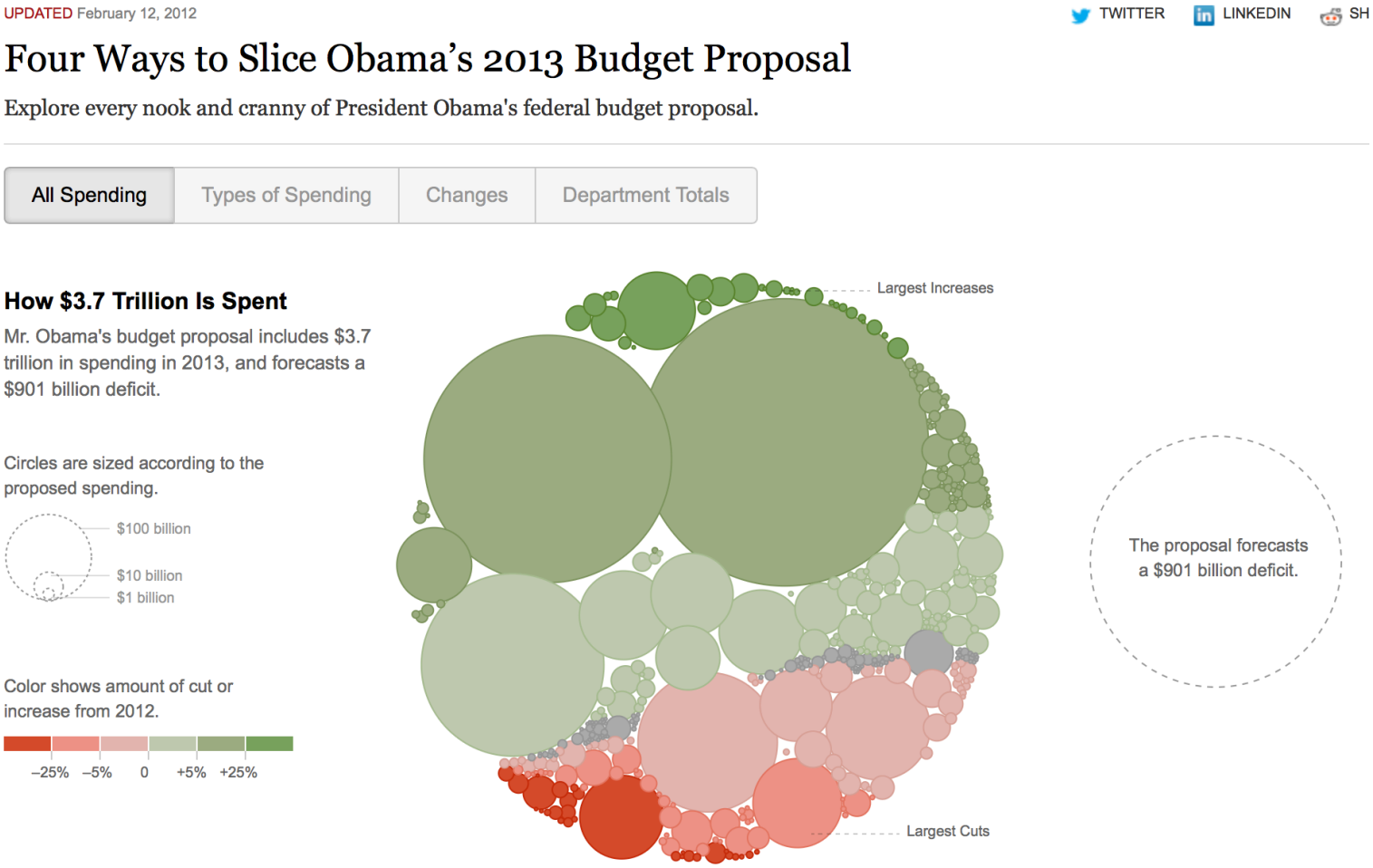
Where the money goes

Health and social welfare get some of the biggest increases in spending this year under the budget plans, with education and the environment also benefiting in the government's push to make Hong Kong "a better place with a brighter future for everyone". Even after the increases the government is predicting a surplus of some HK$37 billion for 2015-2016

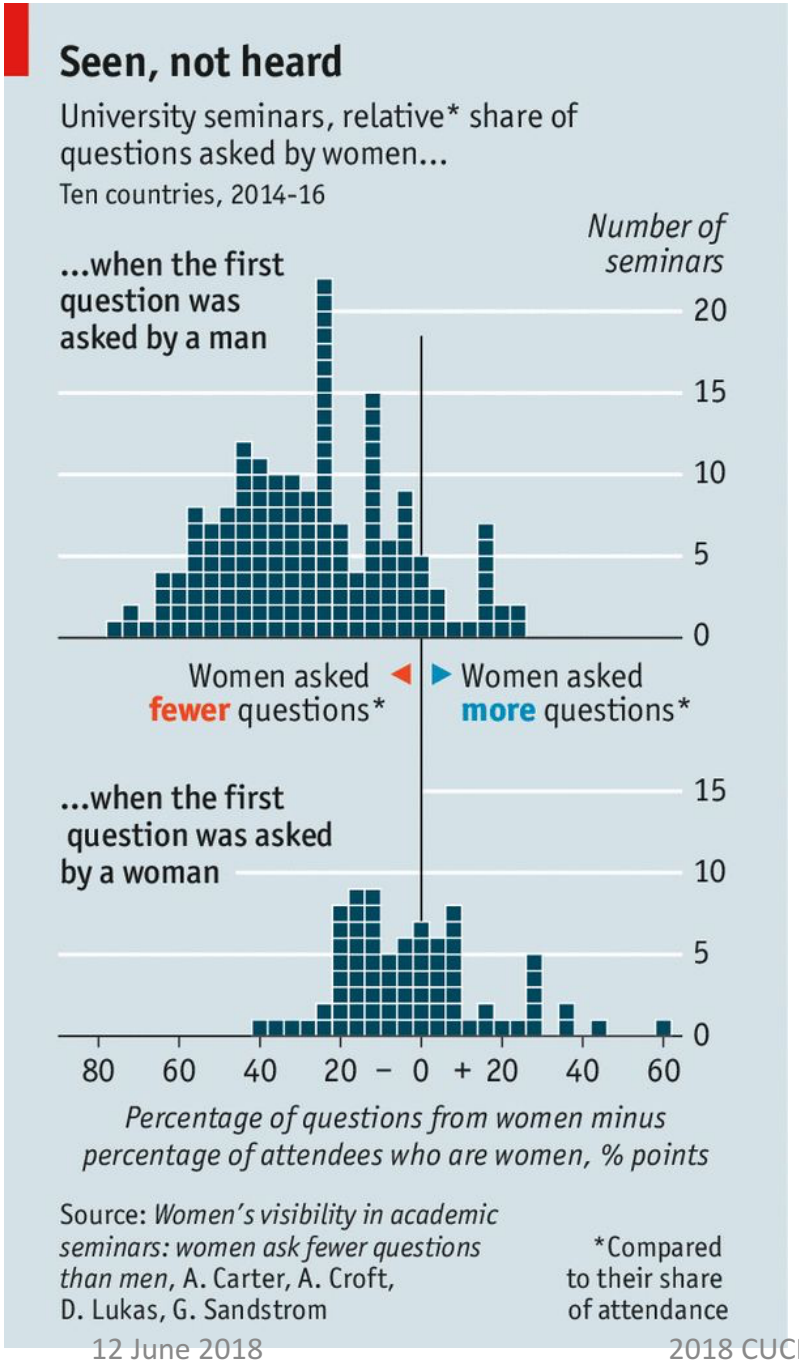
Financial Statistics

The surplus this year is HK$63.8 billion, six times higher than originally projected. The financial secretary said that by the end of March, Hong Kong will have HK$809.5 billion in its fiscal reserves, which have grown continuously since 2003



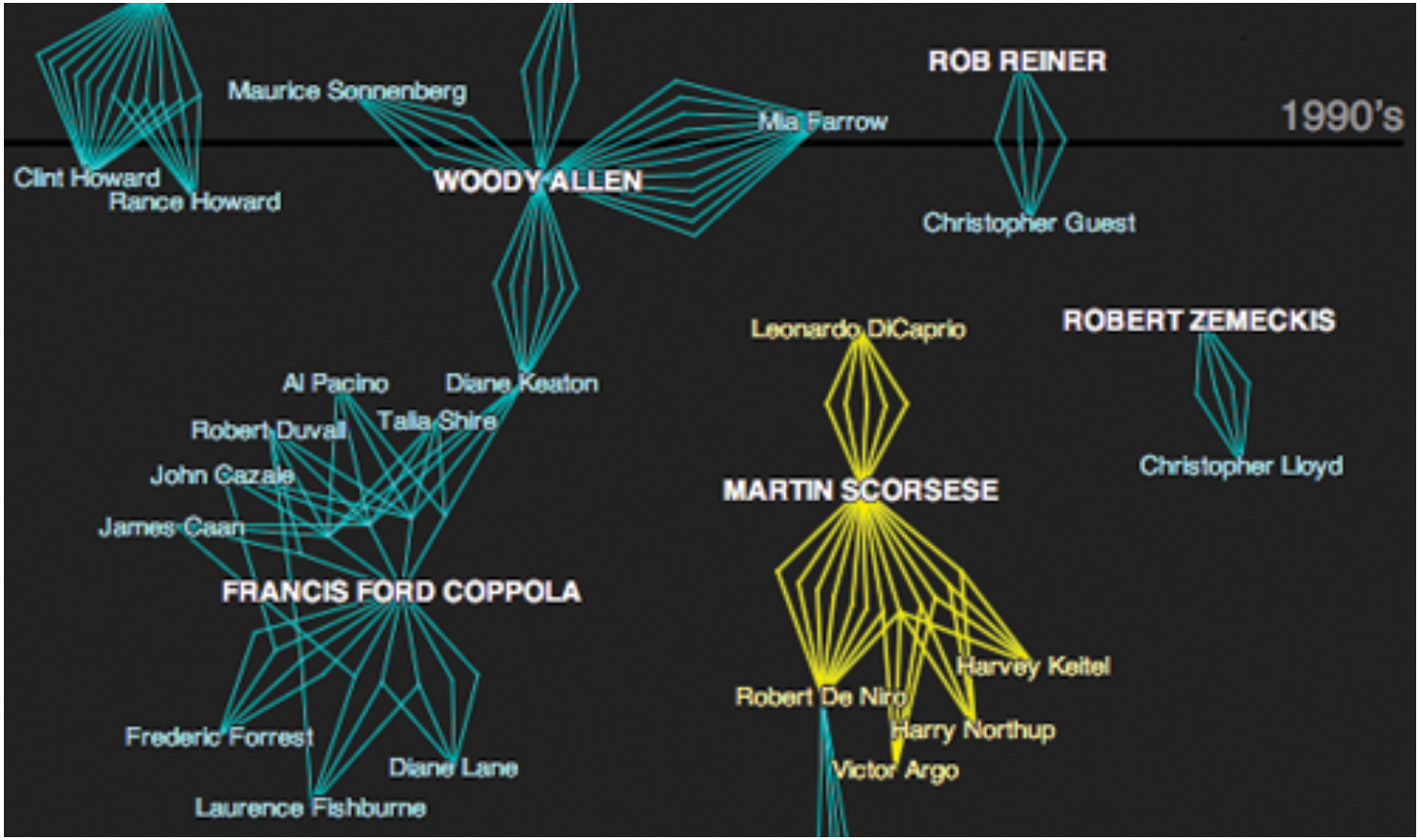


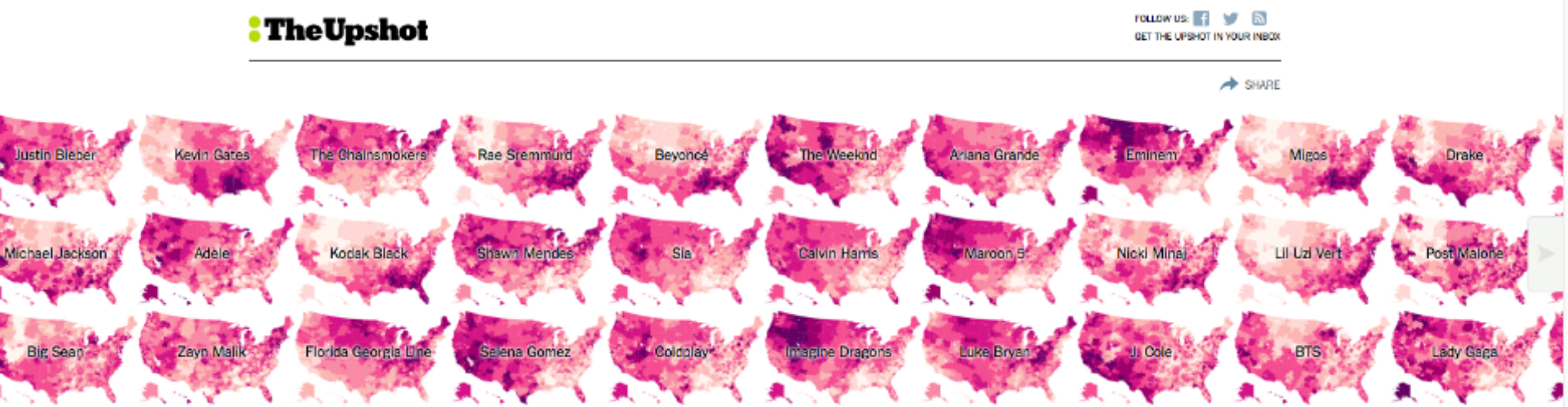
Obama's 2013 Budget Proposal – by New York Times [Link]



- *The Economist* 经济学人
- Observers counted the attendees, and the questions they asked, at 247 departmental talks and seminars in biology, psychology and philosophy that took place at 35 universities in ten countries.
- On average, half of each seminar's audience was female.
- Men, however, were over 2.5 times more likely to pose questions to the speakers—an action that may be viewed (rightly or wrongly) as a sign of greater competence.
- This male skew in question-asking was observable, however, only in those seminars in which a man asked the first question. When a woman did so, the gender split in question-asking was, on average, proportional to that of the audience. [Link]

- *The New York Times* (2013). *Constellations of Directors and Their Stars* [Link]
- "A long-running relationship between an actor and a director can indicate an artistic understanding, a functional routine or even a marketing strategy. Here, a selection of notable directors are shown with actors they cast in at least four films."





The Upshot

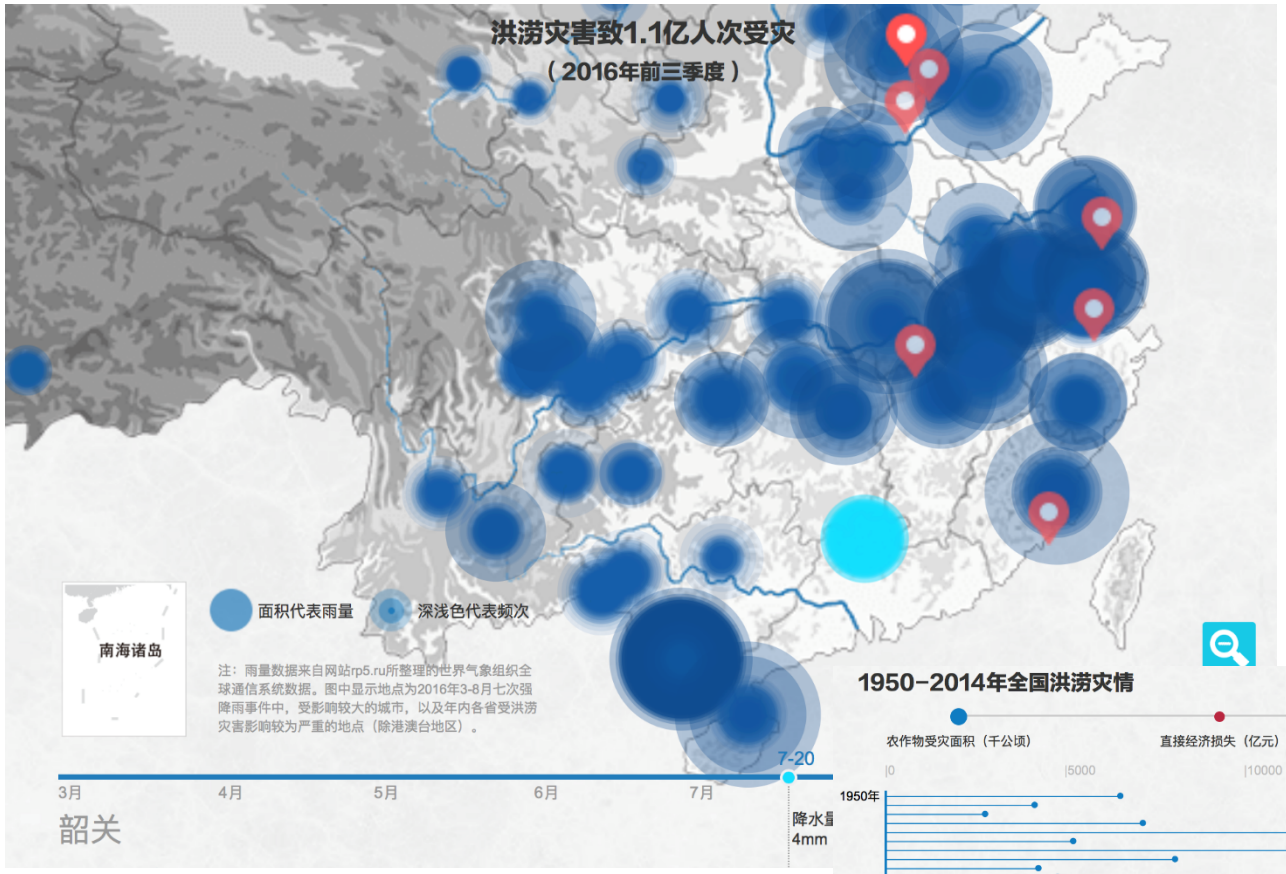
FOLLOW US:
GET THE UPSHOT IN YOUR INBOX

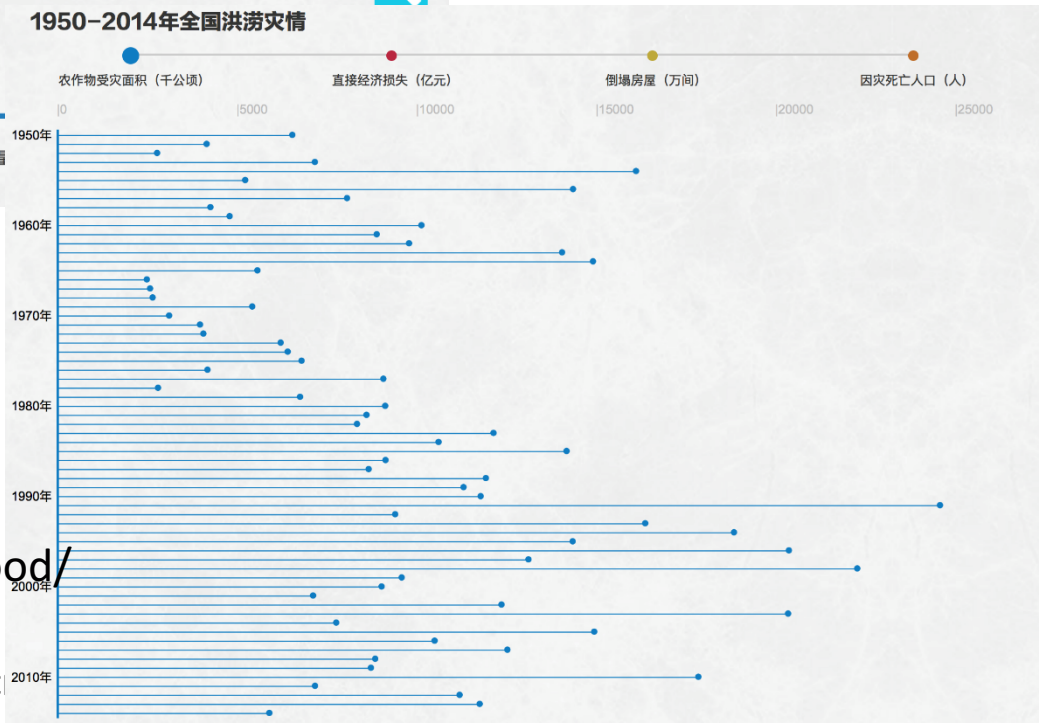
SHARE

What Music Do Americans Love the Most? 50 Detailed Fan Maps

By JOSH KATZ AUG. 7, 2017

YouTube has become a dominant force in the music industry in





http://datanews.caixin.com/mobile/flood/

What is data-driven journalism?

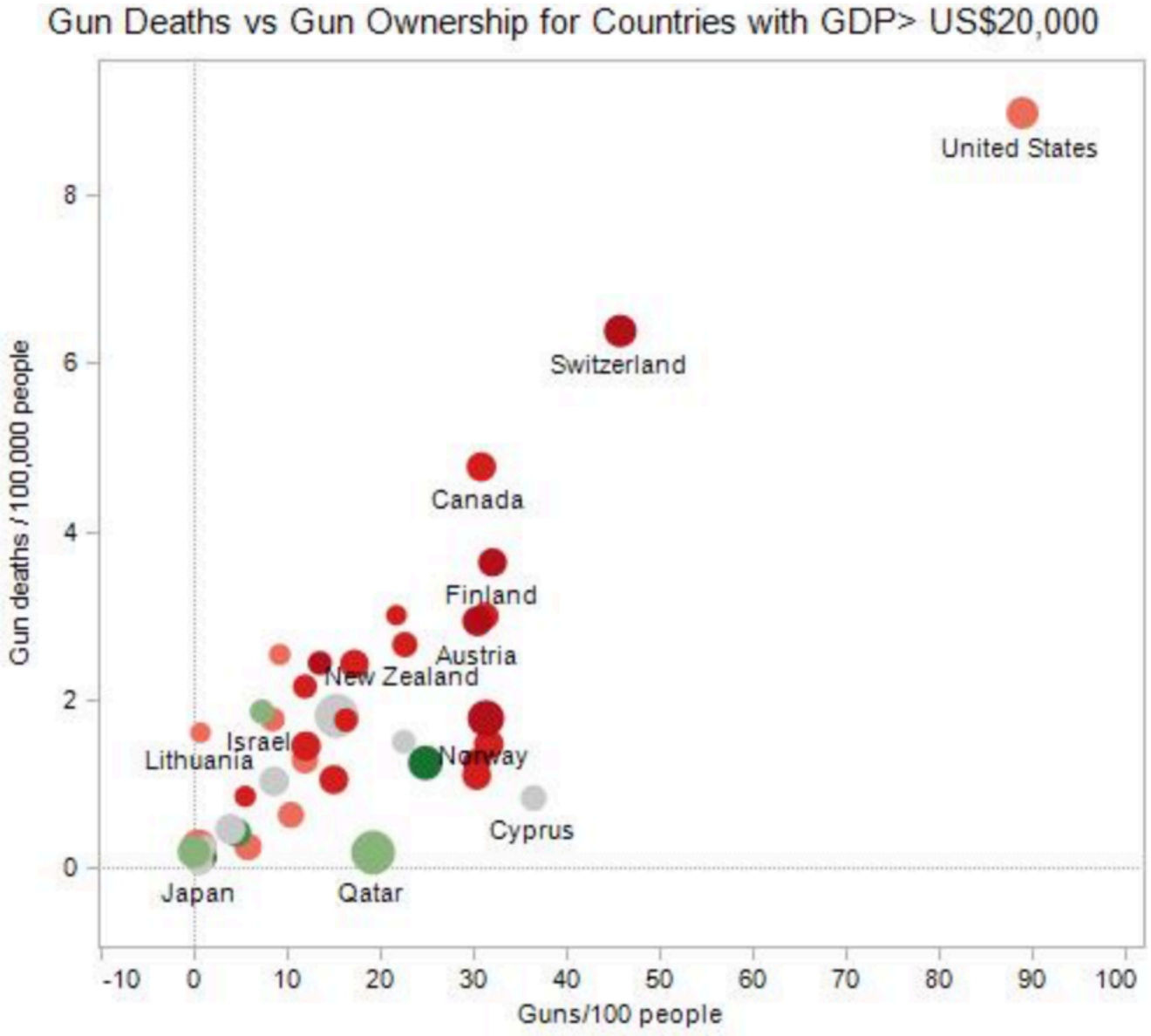
- Data-driven journalism (data journalism) is the journalistic practice of obtaining, reporting on, organizing, editing, and publishing data in the public interest via collaboration related to the techniques of statistics, computer science, visualization, designing, and news reporting (Coddington, 2015).
- Broadly defined, it can involve changing how stories are discovered, presented, aggregated, monetized, and archived. Computation can advance journalism by drawing on innovations in topic detection, video analysis, personalization, aggregation, visualization, and sense-making" (Cohen, Hamilton, Turner (2011). Computational Journalism)

Data science + Journalism = ?

- Data-driven investigation
  - identifying news sources
  - exploring the hidden stories
  - revealing patterns from the data: trend/outliers/correlations (Stray, 2017)
- Data-driven presentation
  - presentation and manifestation
  - data visualization and interactive functions
  - “有图有真相” “一图胜千言”
- Data-driven news filtering
- Data-driven post-product tracking

Data-driven investigating

- Philip Meyer (1967) at *Detroit Free Press*, the forerunner of computer-assisted reporting and precision journalism. [Link]
- *The People Beyond 12th street: A Survey of Attitudes of Detroit Negroes After the Riot of 1967*
- Not like popular belief, there was no correlation between economic status or educational levels and propensity to riot. Nor had the riots been the work of recent immigrants from the south. The main grievances were police brutality, overcrowded living conditions, poor housing, and lack of jobs.



https://www.kdnuggets.com/2012/12/new-poll-gun-violence-vs-gun-ownership.html

”Spotlight” [Link]

Data sources and data access

Table 3. Data source (multiple coding possible, *n* = 225).

Data source	%
Official institution	68.4
Other non-commercial organisation	41.8
Own source	20.4
Private company	20.4
Not indicated	7.1

Source: Loosen et al. (2017). page 10. Based on 225 DJA DDJ works

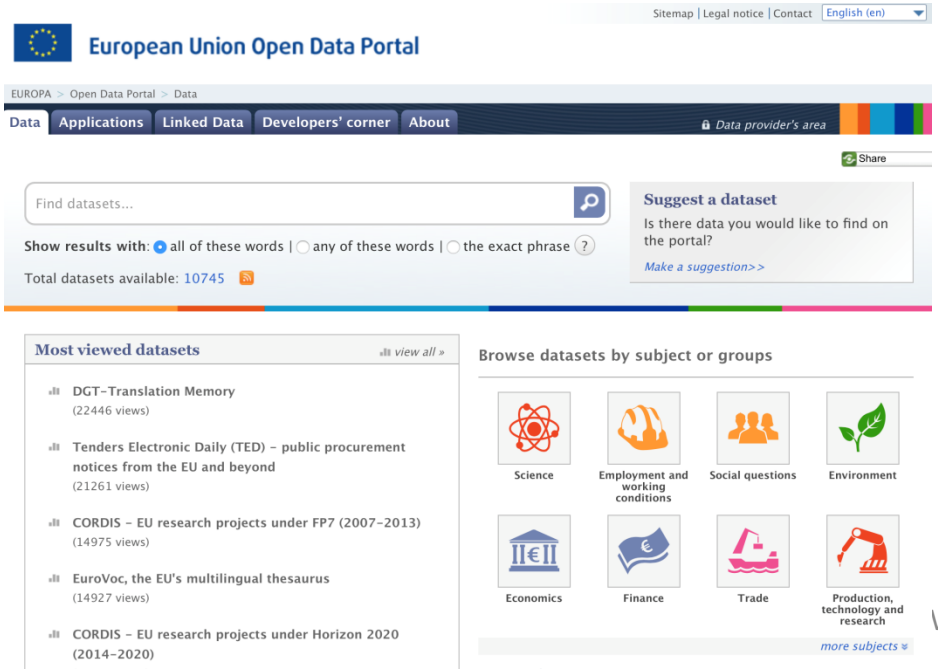
Table3. Data source: a sample of DDJ works of mainland China news agencies (n=137)

Data source	Freq.	%
No mention	38	27.7%
Public data	51	37.2%
Scholar research	3	2.2%
Company database	50	36.5%
Other news reports	35	25.5%

Source: Wang (2018). Unpublished Dissertation. Hong Kong Baptist University.


資料一線通 DATA.GOV.HK
主页 数据 历史性数据 应用 开发者
搜寻...
En 繁
主页 > 数据集
排序 按名称升序 提交
数据提供机构
政府统计处
康乐及文化事务署
保险业监理处
机电工程署
政府资讯科技总监办公室
屋宇署
卫生署
地政总署
香港惩教署
通讯事务管理局办公室
显示更多数据提供机构...
15岁及以上人口的教育程度分布
教育
教育局
提供15岁及以上人口的教育程度分布。
XLSX
2001年年度一般保险业务统计数字
财经
保险业监理处
2001年年度一般保险业务统计数字
XLS
2001年年度长期保险业务统计数字
财经
保险业监理处
2001年年度长期保险业务统计数字
XLS


DATA.GOV
DATA TOPICS IMPACT APPLICATIONS DEVELOPERS CONTACT
The home of the U.S. Government's open data
Here you will find data, tools, and resources to conduct research, develop web and mobile applications, design data visualizations, and more.
GET STARTED
SEARCH OVER 135,277 DATASETS
Credit Card Complaints
BROWSE TOPICS
Agriculture
Climate
Consumer
Ecosystems
Education
Energy
Finance


Sitemap | Legal notice | Contact
English (en)
European Union Open Data Portal
EUROPA > Open Data Portal > Data
Data
Applications
Linked Data
Developers' corner
About
Data provider's area
Share
Find datasets...
Show results with: all of these words | any of these words | the exact phrase
Total datasets available: 10745
Suggest a dataset
Is there data you would like to find on the portal?
Make a suggestion>>
Most viewed datasets
view all »
DGT-Translation Memory
(22446 views)
Tenders Electronic Daily (TED) – public procurement notices from the EU and beyond
(21261 views)
CORDIS – EU research projects under FP7 (2007-2013)
(14975 views)
EuroVoc, the EU's multilingual thesaurus
(14927 views)
CORDIS – EU research projects under Horizon 2020 (2014-2020)
Browse datasets by subject or groups
Science
Employment and working conditions
Social questions
Environment
Economics
Finance
Trade
Production, technology and research
more subjects


北京市政务数据资源网
www.bjdata.gov.cn
搜索
高级搜索>>
首页 数据 接口 定向数据 应用 工具 互动交流
网站统计
北京市旅游发展委员会
北京市旅游景区
游览舒适度指数
API接口上线
数据资源
最热下载
+更多
[教育科研] 高校 [3528]
[教育科研] 小学 [2505]
[教育科研] 中学 [2375]
[企业服务] 土地用途分区 [2055]
[交通服务] 轨道交通线路 [1183]
[旅游住宿] 星级饭店 [630]
最新数据
+更多
[市统计局] 主要农作物播种面... [2017-05-11]
[市统计局] 主要农业产品产量... [2017-05-11]
[市统计局] 种业生产 [2017-05-11]
[市统计局] 乡镇企业各业基本... [2017-05-11]
[市统计局] 乡镇企业出口供货... [2017-05-11]
[市统计局] 水产品生产 [2017-05-11]
APP应用
+更多
"逛逛博物馆"APP自助导览
逛逛博物馆手机APP是一款针对北京市政府数据开放的博物馆...
E上学
ESchool收录了最新一年部分小学与中学的入学关系，并...

Presentation

- “Unfounded” - The Global and Mail, 2017 [Data Journalism Award, 2017, Investigation of the Year] [Link]
- The cases of Caixin Media (DJA 2018 Best data journalism team) [Link]

Visualization technics

Table 5. Visualisation (multiple coding possible, $n = 225$).

Visualisation	%
Image	66.7
Simple static chart	60.0
Map	49.8
Table	31.6
Combined static chart	27.1
Animated visualisation	18.7
Other visualisation	3.1
No visualisation	0.9

Source: Loosen et al. (2017). page 12. Based on 225 DJA award-winning DDJ works

Table 6. Interactive functions (multiple coding possible, $n = 224$).

Interactive function	%
Zoom/details on demand	63.8
Filtering	52.7
Search	28.1
Personalisation	16.5
Gamified interaction	4.0
Other interactive feature	1.3
No interactive feature	17.0

Source: Loosen et al. (2017). page 13. Based on 225 DJA award-winning DDJ works

Table 4. Interaction functions (n=137)

Interaction operations	Frequency	%
None	126	92
Search/Select	10	7.3
Filter	3	2.2
Zoom in/out	0	0
Highlight	6	4.4
Navigation	4	2.9
Text input	3	2.2

Source: Wang (2018). Unpublished Dissertation. Hong Kong Baptist University.

The intellectual origins of data-driven journalism

Origins and domain knowledge	Terms and "Catchphrases"	Contributions to the DDJ
Social science research (theories and methods, especially quantitative research methods)	"Precision journalism" "How to lie with statistics" (n, M, SD)	Concepts, measurements, interpreting quantitative data; unidimensional analysis (pie chart, bar charts), bi-variate relationships (lines, scatterplots, bar charts)
Investigative reporting	"News labs" "The social/public goods"	FOIA, interviewing technics
The open-source software movement, open data, data scientists	"The Digital Robin Hood" "Hacks/hackers"	Data science and technologies, open source projects, open data applications, collaborative projects, accountability

The intellectual origins of data-driven journalism

- The National Institute for Computer-Assisted Reporting hosts the annual Philip Meyer Journalism Award, which "recognize excellent journalism done using social science research methods"
- An investigation often arises when a reporter perceives a difference between what is (the observed reality) and what should be (as articulated in law or policy). A high-impact investigative story looks at a situation where what is differs from what should be, and explains why (Broussard, 2015).
- *Scraping for Journalists* (Bradshaw, 2014) - Coders, designers, and full-stack applications in the newsrooms

"New" ways of storytelling

- Several terms
  - computer-assisted reporting
  - data-driven journalism
  - computational journalism
  - programmer journalism
  - open-source journalism
  - algorithmic journalism
  - robot journalism
  - automated journalism
  - …
- Coddington (2015): three types of quantitative journalism practice:
  - computer-assisted reporting
  - data-driven journalism
  - computational journalism

The flow of production

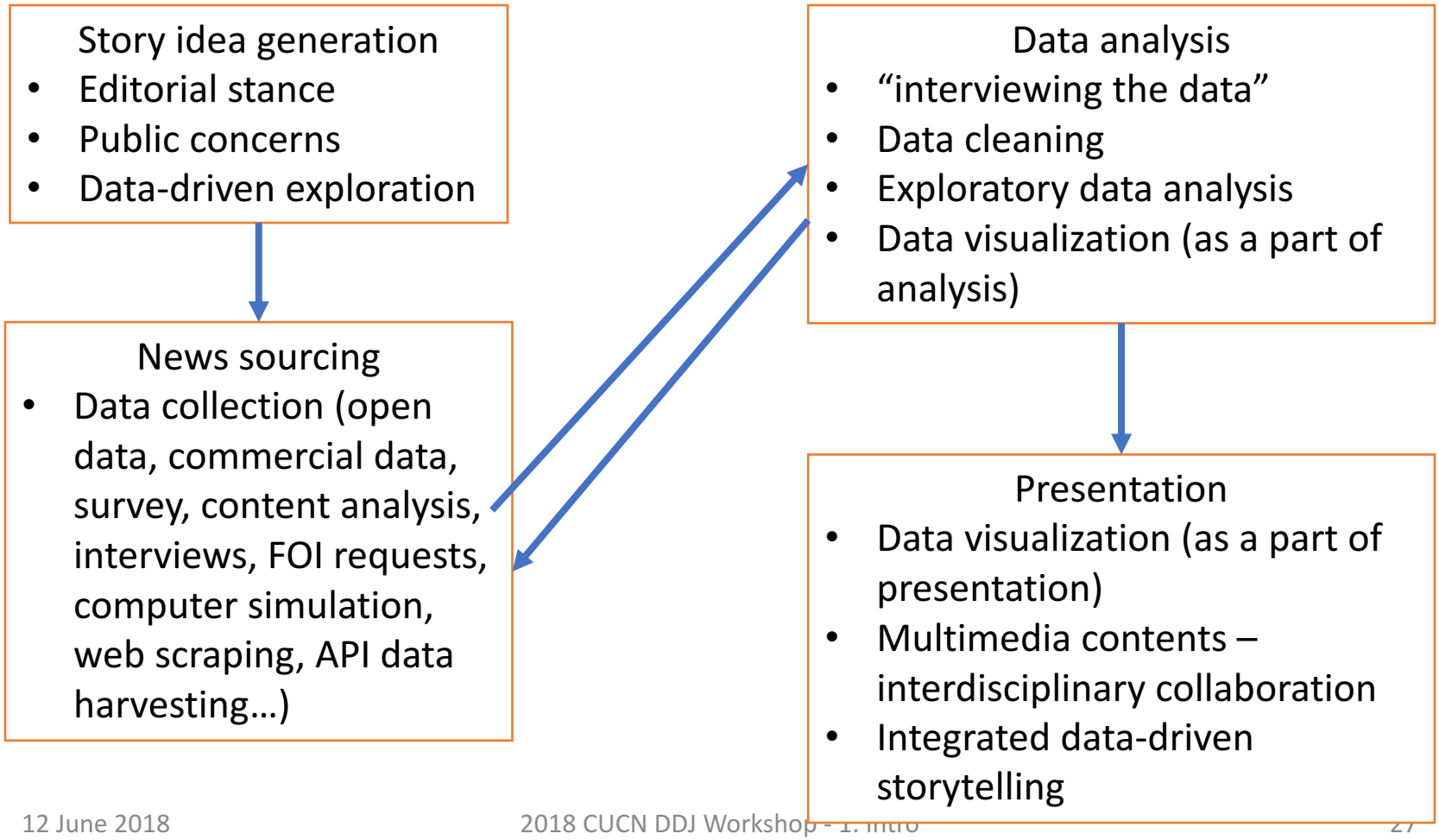


Table 6. Professional labelling from the bylines (n=137)

Professional labelling	Frequency	%
No mention	31	22.6
Content	21	15.3
Editor	84	61.3
Reporter	9	6.5
Designer	67	48.9
Supervisor	20	14.6
Programmer	5	3.6
Other position	5	3.6

Source: Wang (2018). Unpublished Dissertation. Hong Kong Baptist University.


BBC
NEWS
LABS


FT中文网
首頁
中國
全球
特別報導
數據新聞


TheUpshot


The
Economist


theguardian
browse all sections


ProPublica
ProPublica Illinois
Local Reporting Network
Data Store
PROPUBLICA
TOPICS
SERIES
NEWS APPS
GET INVOLVED
IMPACT
ABOUT


财新
数据可视化
实验室
Caixin Data
Visualization Lab


Los Angeles Times

Major Players

- Ausserhofer, J., Gutounig, R., Oppermann, M., Matiasek, S., & Goldgruber, E. (2017). The datafication of data journalism scholarship: Focal points, methods, and research propositions for the investigation of data-intensive newswork. Journalism. page 16.

- - The end of session -