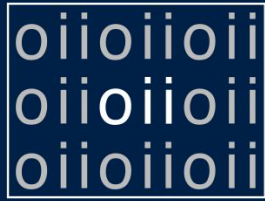




中国国际大数据大会
China International Big Data Summit

资料汇编

(仅供内部交流，请勿外传)



Oxford Internet Institute
University of Oxford



Big Data Big Decision

Ning Wang, PhD
Oxford Internet Institute
University of Oxford
www.oii.ox.ac.uk

07/02/12

“To be or not to be, that is the question.” - William Shakespeare



Decision Dilemma

US is preparing for a
outbreak of an unusual
disease, which is expected
to kill 600 people:

Choice1

A: 200 will be saved **72%**

B: 1/3 probability 600 saved,
2/3 no one saved.

Choise2

C: 400 will die

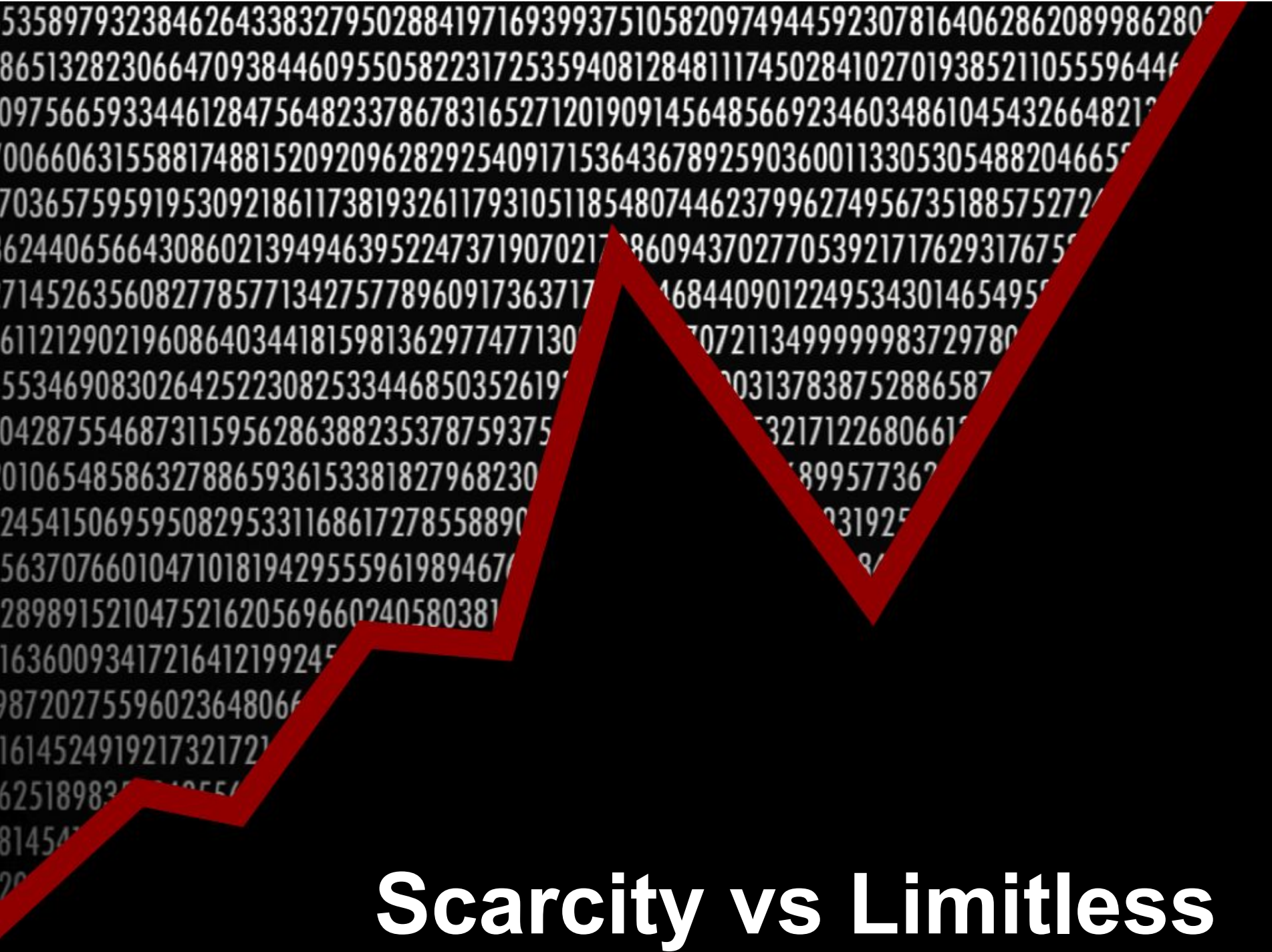
D: 1/3 probability 600 saved, **22%**
2/3 no one saved.



Daniel Kahneman

In decision theory Loss Aversion refers to people's tendency to strongly prefer avoiding losses to acquiring gains



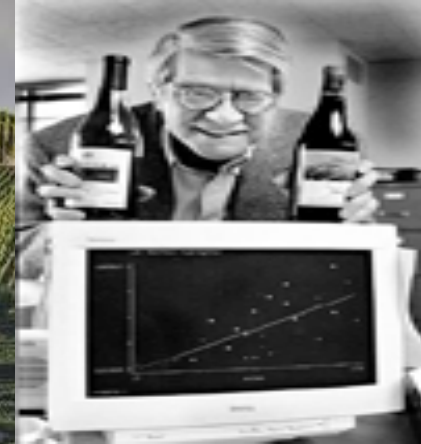


Scarcity vs Limitless



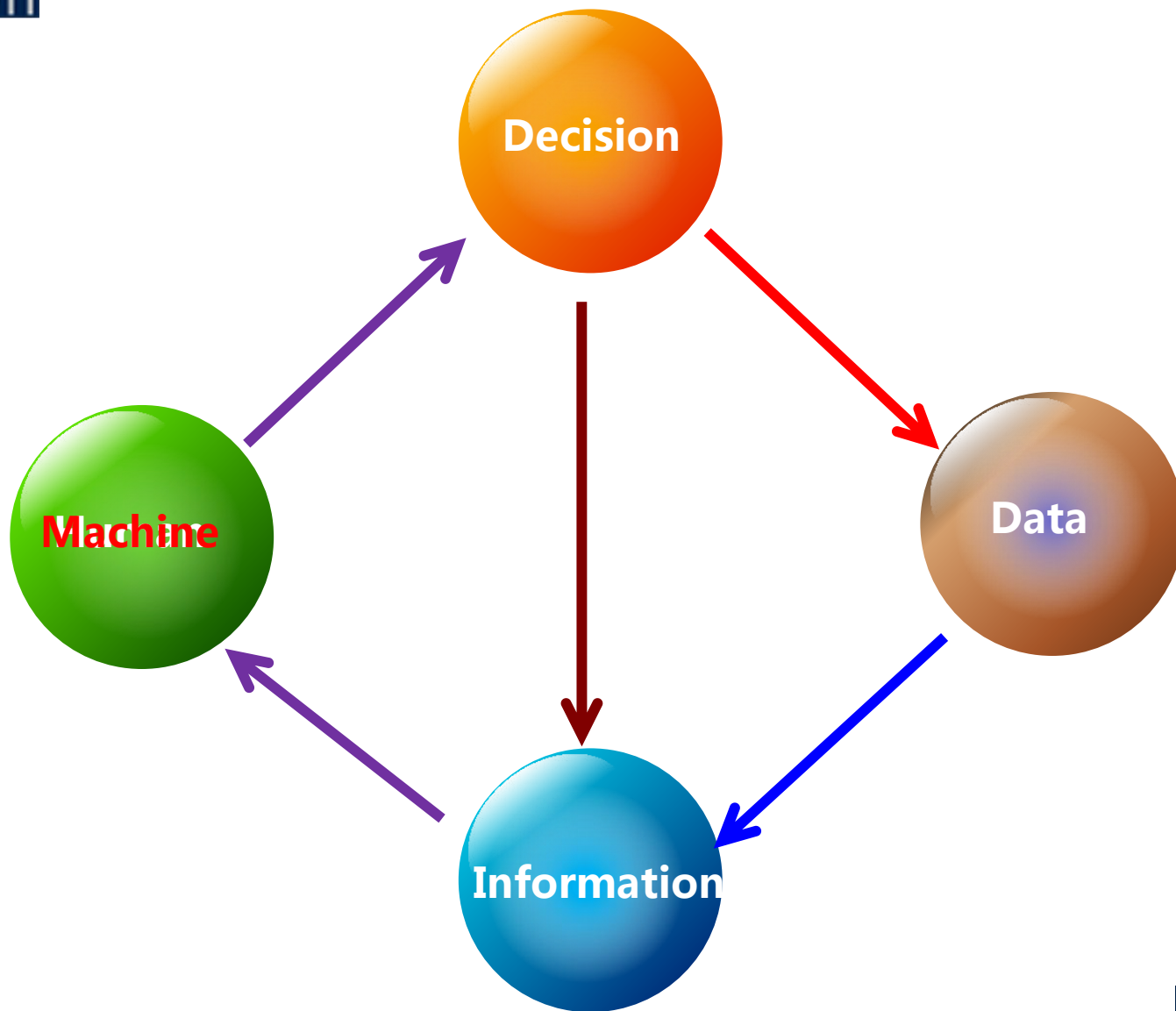
Monte Berner 750ml	Orlando ST Hugo Cab Sauv \$44.99 750ml	Ingoldby Cab Sauv \$22.99 750ml	Wolf Blass Presidents Cab Sauv \$32.99 750ml	Ingoldby Cab Sauv \$22.99 750ml	Peter Lehman Cab Sauv \$28.99 750ml	D'arenberg High Trellis Cab Sauv \$27.99 750ml	McGuigan Genus 4 Cab Sauv \$26.99 750ml	Brown Bros Cab Sauv \$24.99 750ml	Brown Myrah Cab Sauv \$23.99 750ml	Ingoldby Cab Sauv \$22.99 750ml	Ingoldby Cab Sauv \$22.99 750ml	Jim Barry Merle Wood Shiraz \$43.99 750ml	Wolf Blass Brown Label Shiraz \$39.99 750ml	Tatachi Vale Shiraz \$35.99 750ml
-----------------------	--	---------------------------------------	--	---------------------------------------	---	--	---	---	--	---------------------------------------	---------------------------------------	---	---	---





**Wine quality = 12.145 + 0.00117 winter rainfall +
0.0614 average growing season temperature -
0.00386 harvest rainfall**

Evolution of Decision Making



Breaking: Two Explosions in the White House and Barack Obama is injured

Reply Retweet Favorite More

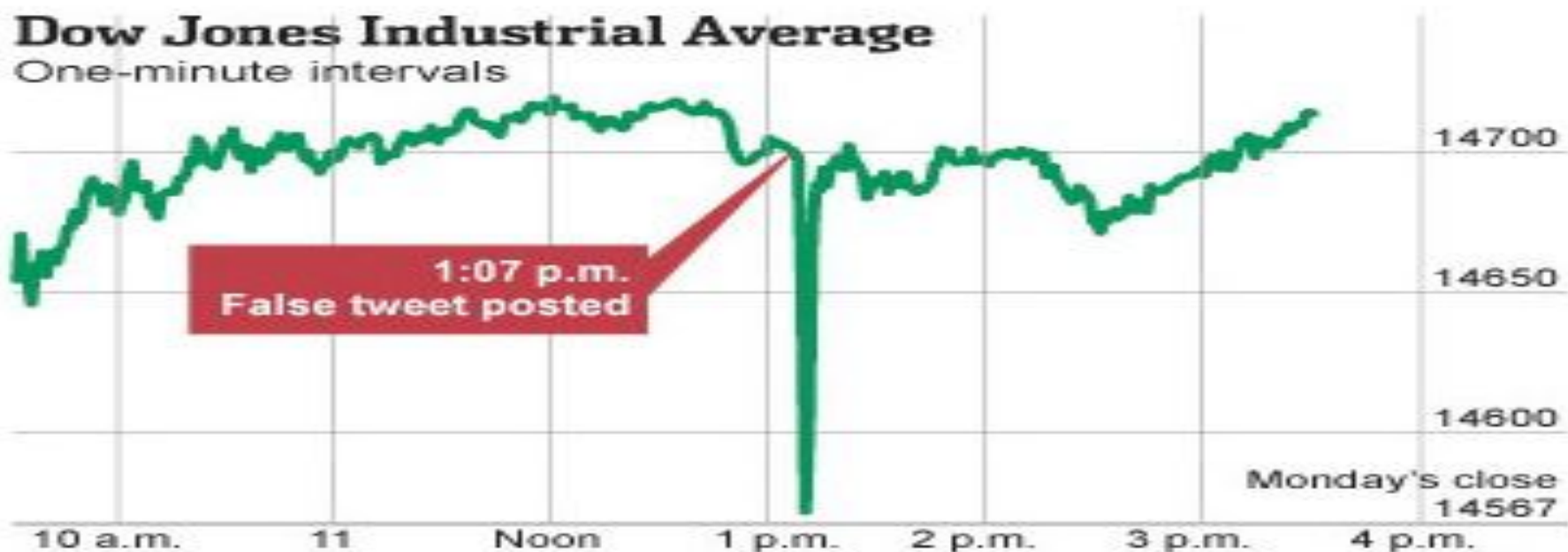
1,849
RETWEETS

82
FAVORITES



1:07 PM - 23 Apr 13

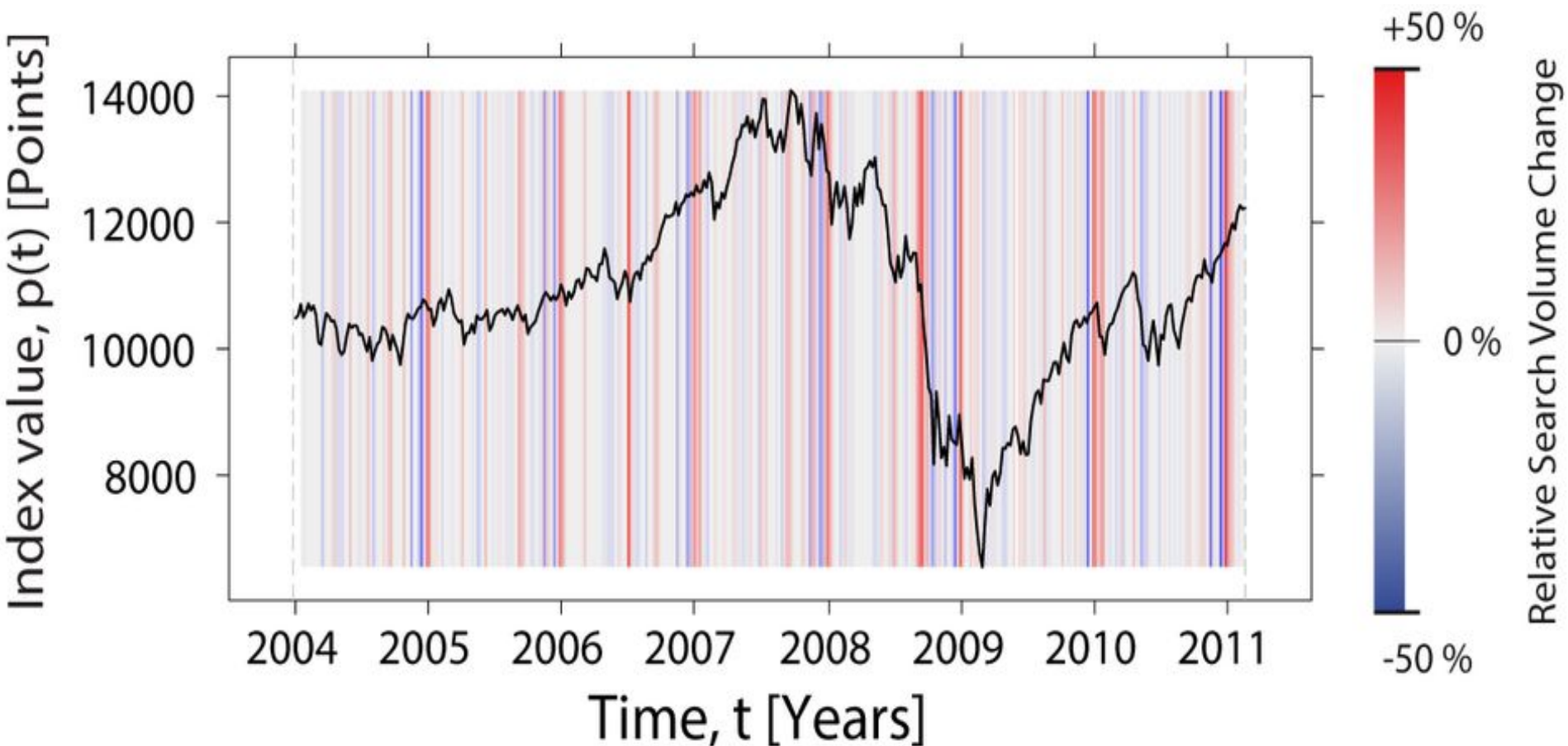
Dow Jones Industrial Average One-minute intervals



Source: FactSet

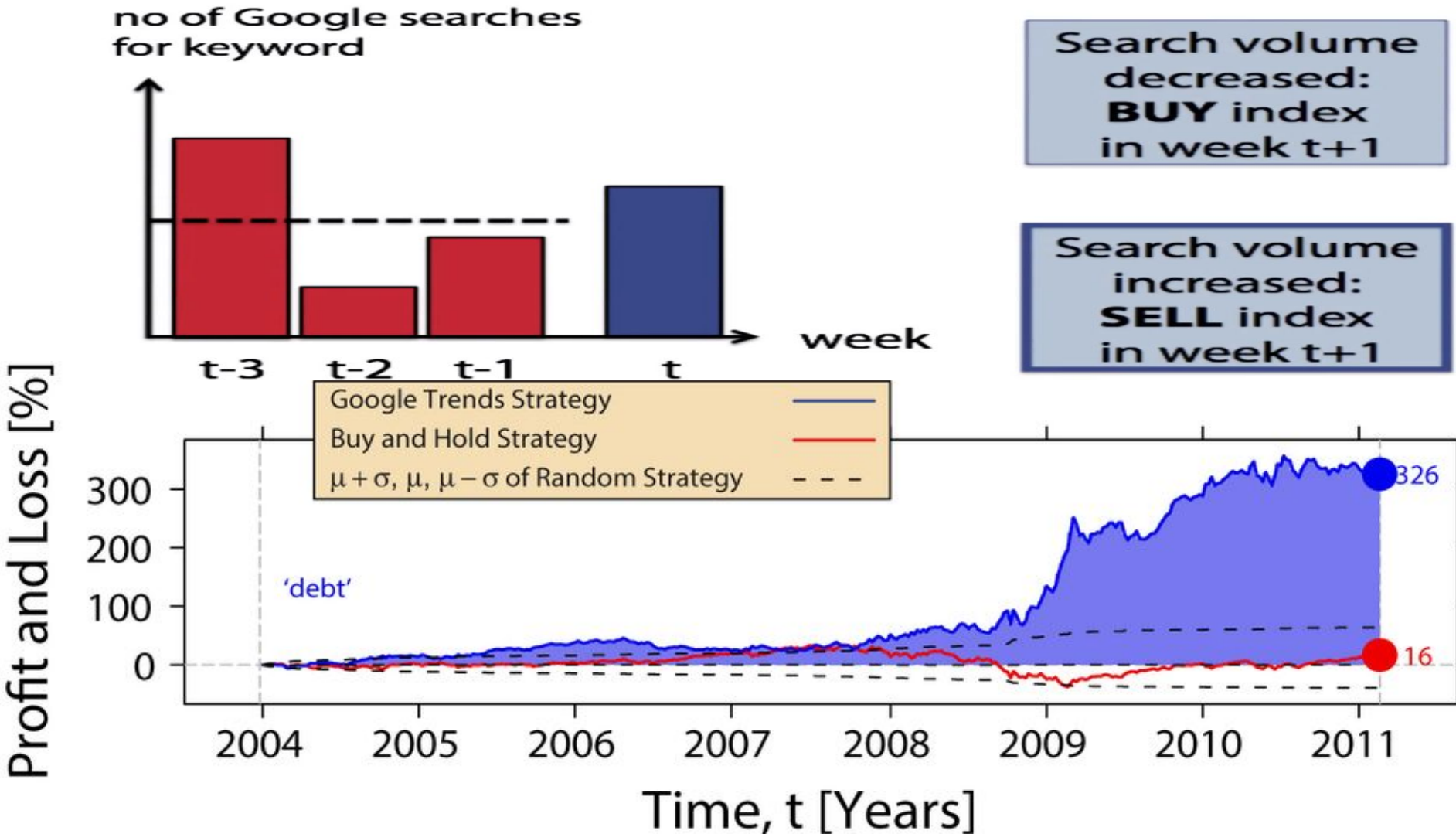


Google Trend Trading Strategy based on Keyword “Debt”

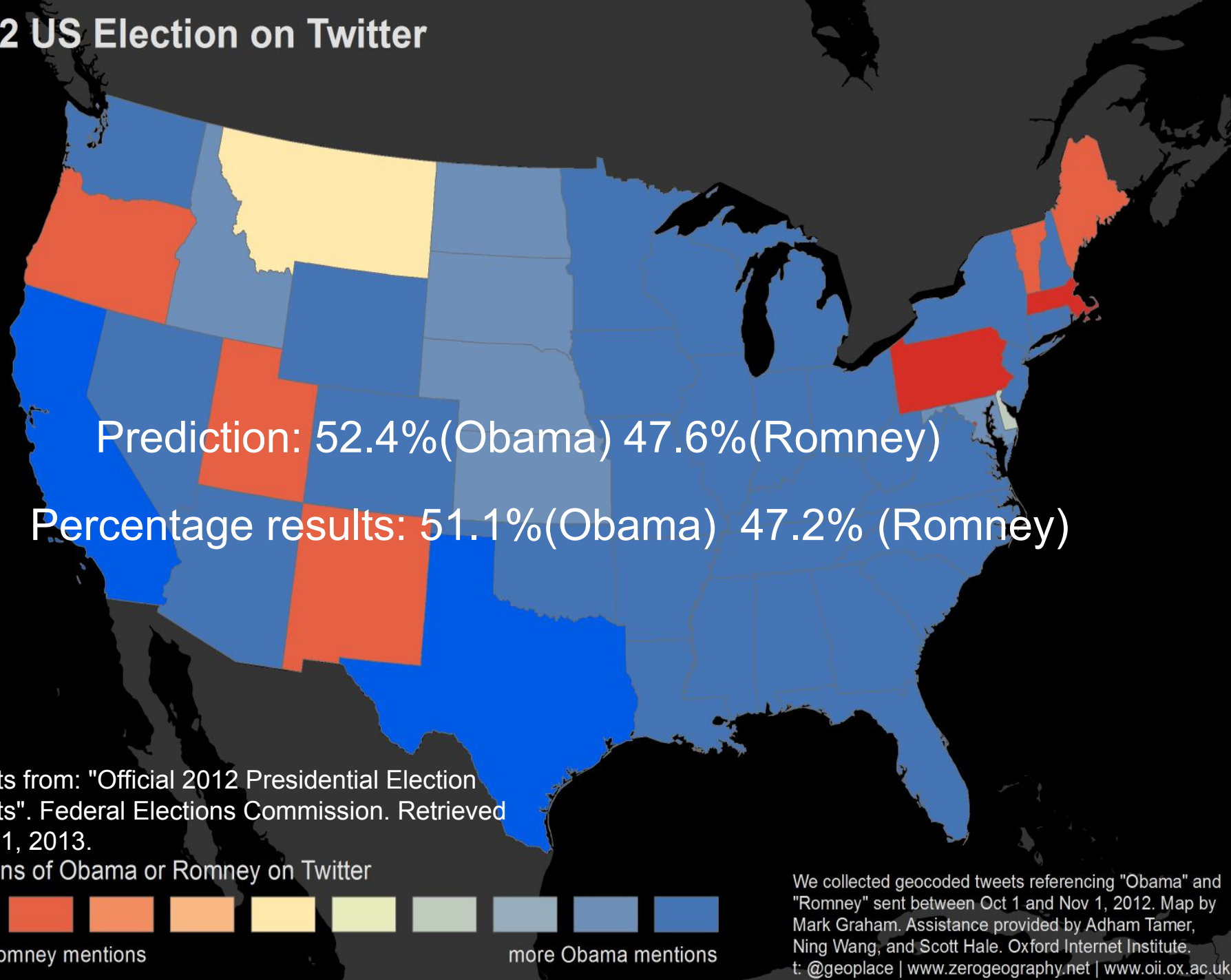




Google Trend Trading Strategy based on Keyword “Debt”



2012 US Election on Twitter



Results from: "Official 2012 Presidential Election Results". Federal Elections Commission. Retrieved May 21, 2013.

mentions of Obama or Romney on Twitter



more Romney mentions

more Obama mentions

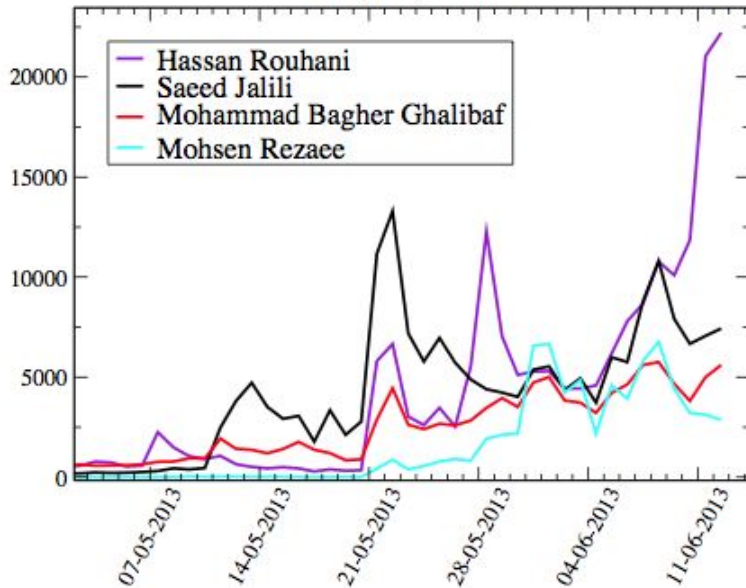
We collected geocoded tweets referencing "Obama" and "Romney" sent between Oct 1 and Nov 1, 2012. Map by Mark Graham. Assistance provided by Adham Tamer, Ning Wang, and Scott Hale. Oxford Internet Institute. t: @geoplace | www.zerogeography.net | www.oii.ox.ac.uk



2013 Iranian Presidential Election

Source: Yasseri, T., & Bright, J. (2013, December 10). Can electoral popularity be predicted using socially generated big data?

Daily Page View



Google Search Volume

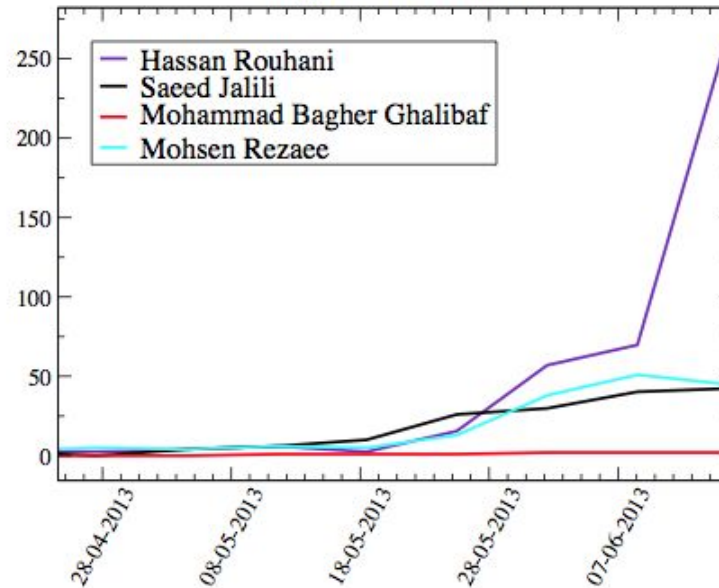


Figure 1: The time evolution of Wikipedia page views and Google search volume for the four leading candidates of Iranian presidential electin of 14 June 2013 are shown in the left and right diagrams respectively.

Candidate	Popular Vote	Percentage
Hassan Rouhani	18,613,329	50.88
Mohammad Bagher Ghalibaf	6,077,292	16.46
Saeed Jalili	4,168,946	11.31
Mohsen Rezaee	3,884,412	10.55

Table 1: Final results of the Iranian presidential election, 14 June 2013.



2010 UK General Election

Source: Yasseri, T., & Bright, J. (2013, December 10). Can electoral popularity be predicted using socially generated big data?

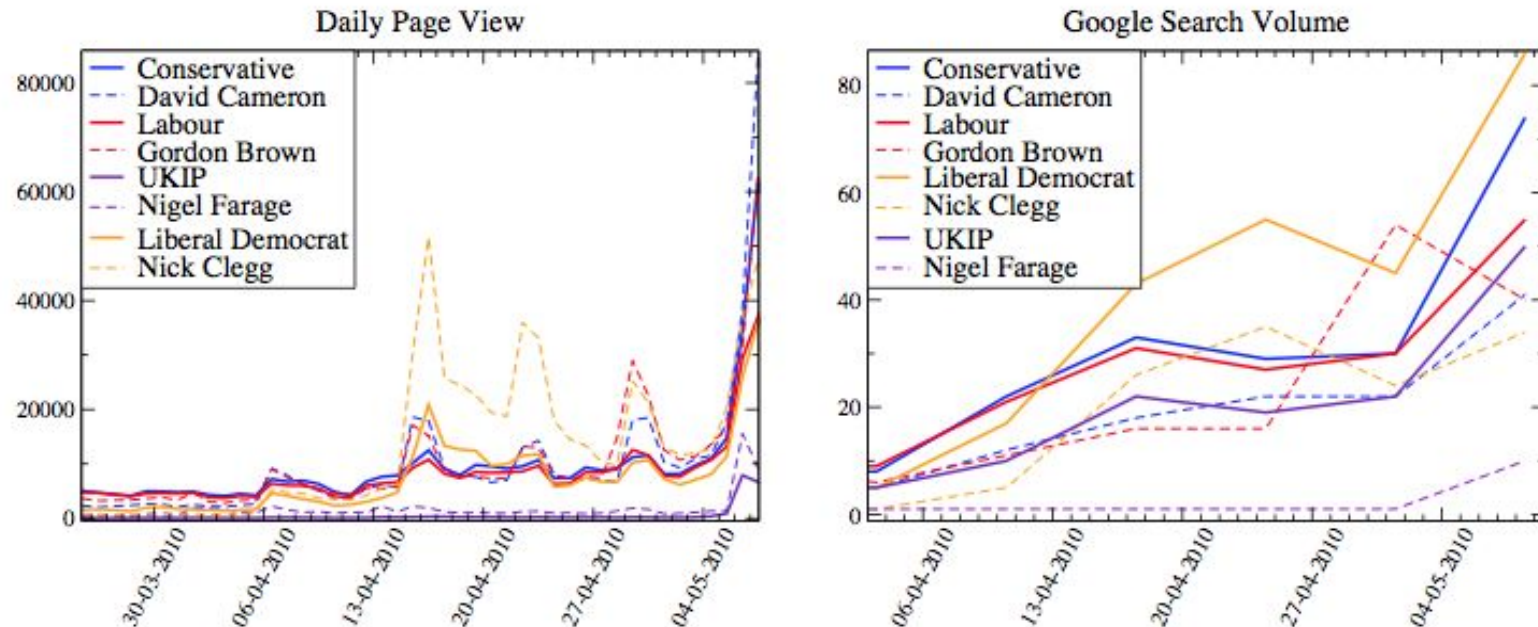


Figure 3: The time evolution of Wikipedia page views and Google search volume for the four leading parties and their leaders during the 6 May 2010 UK general election campaign are shown in the left and right diagrams respectively.

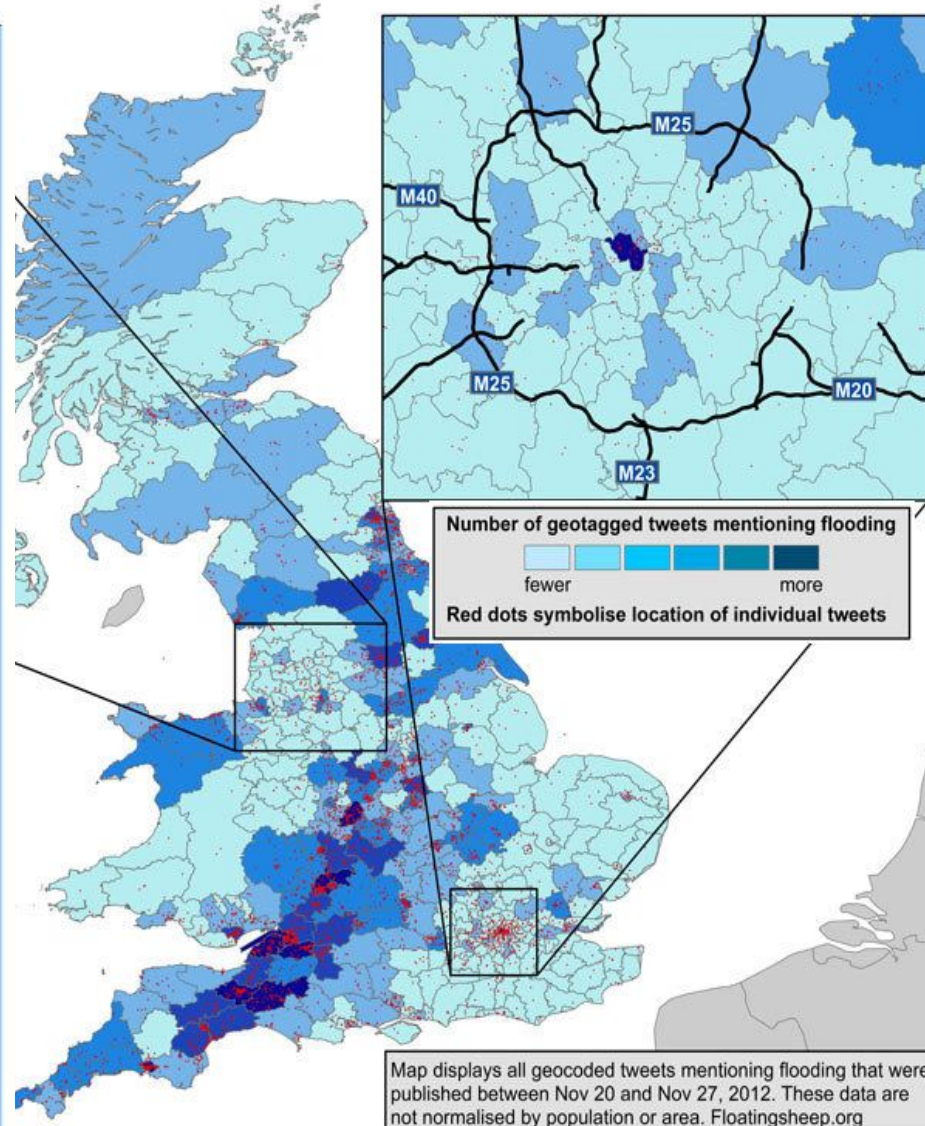
Party	Popular Vote	Percentage
Conservative	10,703,654	36.1
Labour	8,606,517	29.0
Liberal Democrat	6,836,248	23.0
UKIP	919,471	3.1

Table 3: Final results of the United Kingdom general election, 6 May 2010.

Rainfall vs 'flood' tweets

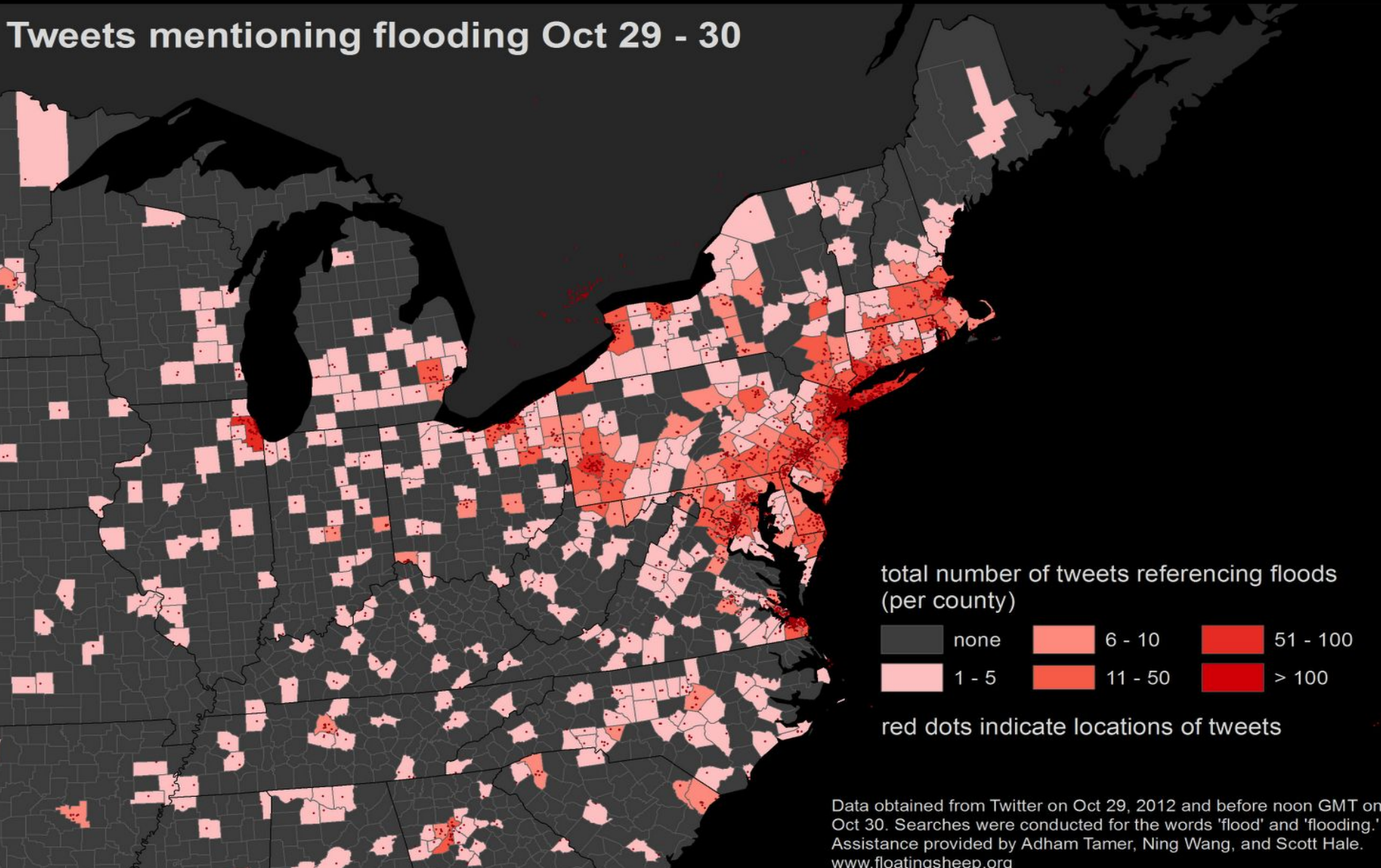


19-27 November 2012
Rainfall amount
8-day total



And in New York.....

Tweets mentioning flooding Oct 29 - 30



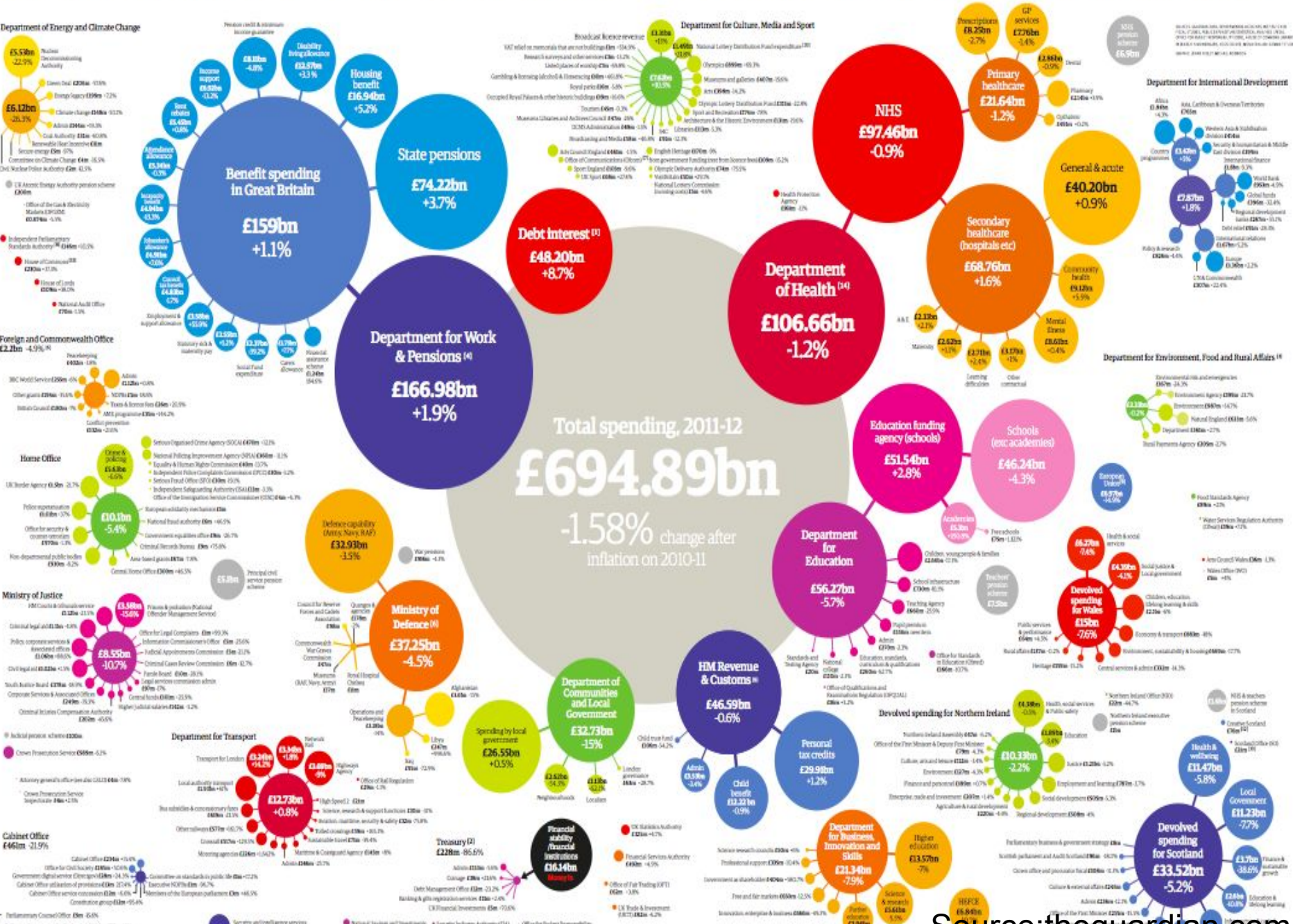
The Spatial Distribution of the Occupy Campaign

Links between cities are based on 'occupy*' hashtags co-used in the same tweets (i.e. #occupywallstreet for New York and #occupylsx for London).

Data aggregated for the period April 30 to May 30 2012.

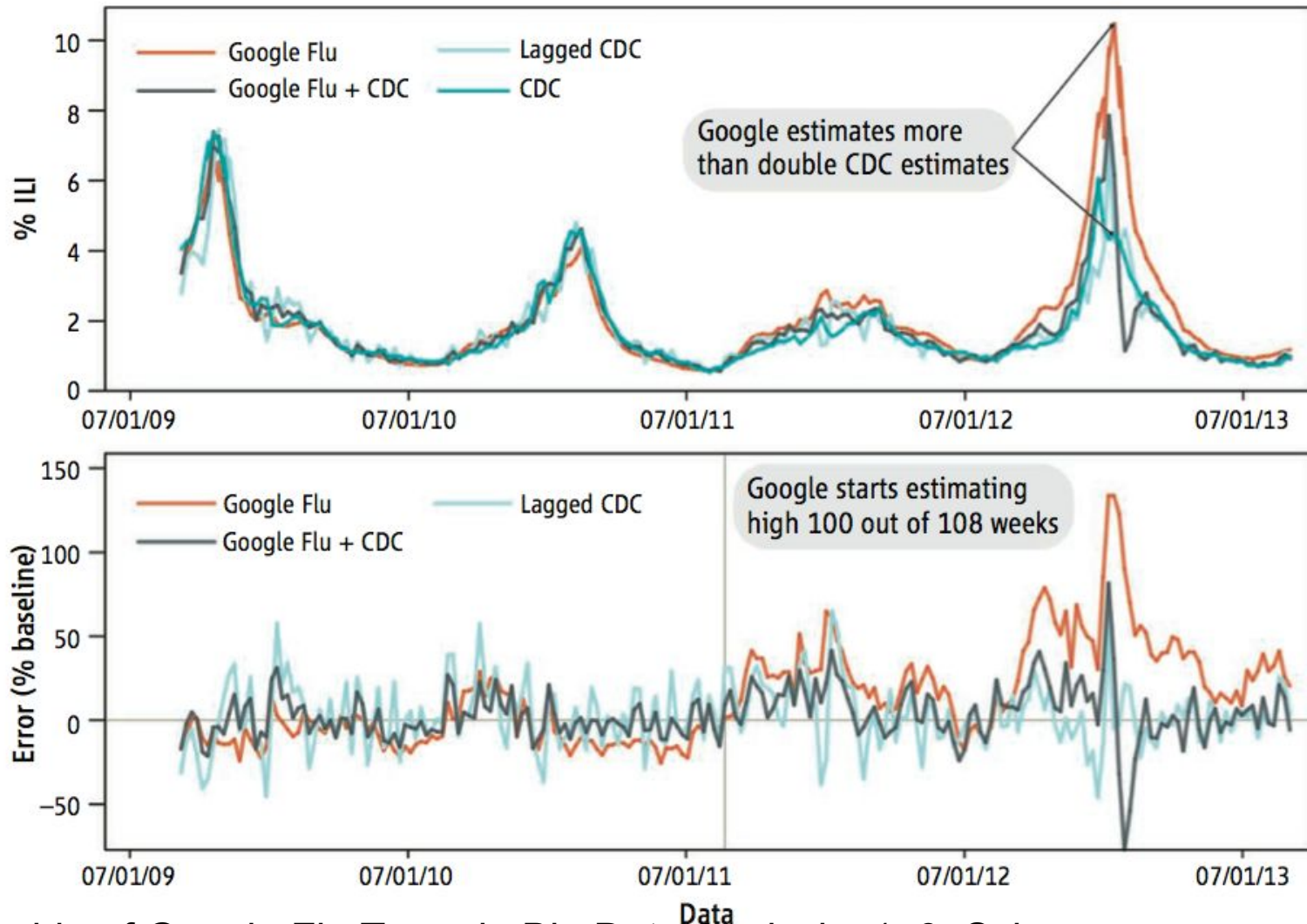


Where your taxes went this year - and where the cuts were made Public spending by the UK's central government departments, 2011-12



Source:theguardian.com

Data Misinterpretation

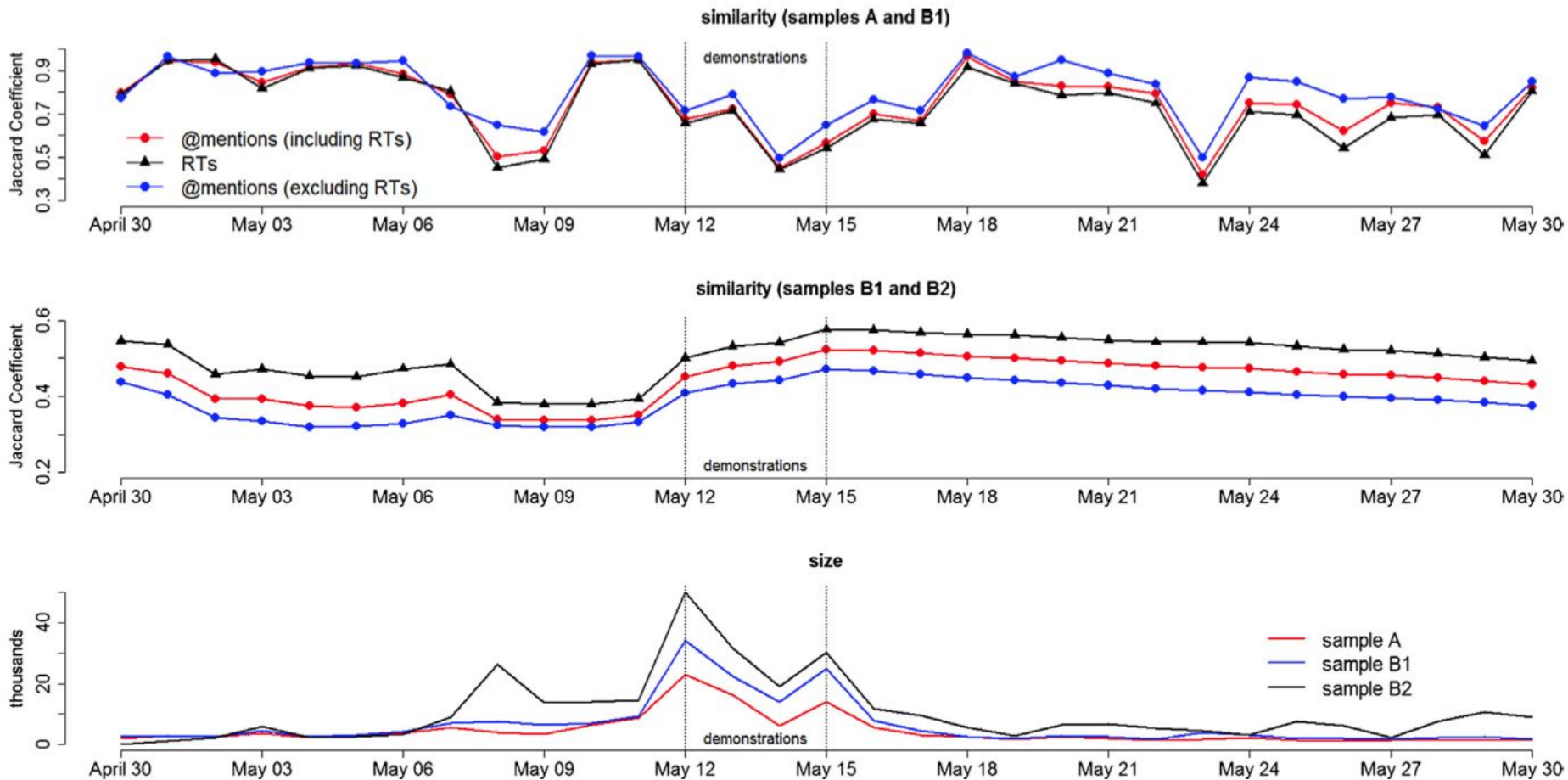


The Parable of Google Flu: Traps in Big Data Analysis, 1–3. Science (2014)



Bias in Large Datasets

Source: Ning Wang, et al. (2014). Assessing the bias in samples of large online networks. Social Networks, 38,



‘weak human + machine + better process was superior to a strong computer alone’ and, more remarkably, ‘superior to a strong human + machine + inferior process.’” - Erik Brynjolfsson

Thank you!