Data visualization in reporting and storytelling

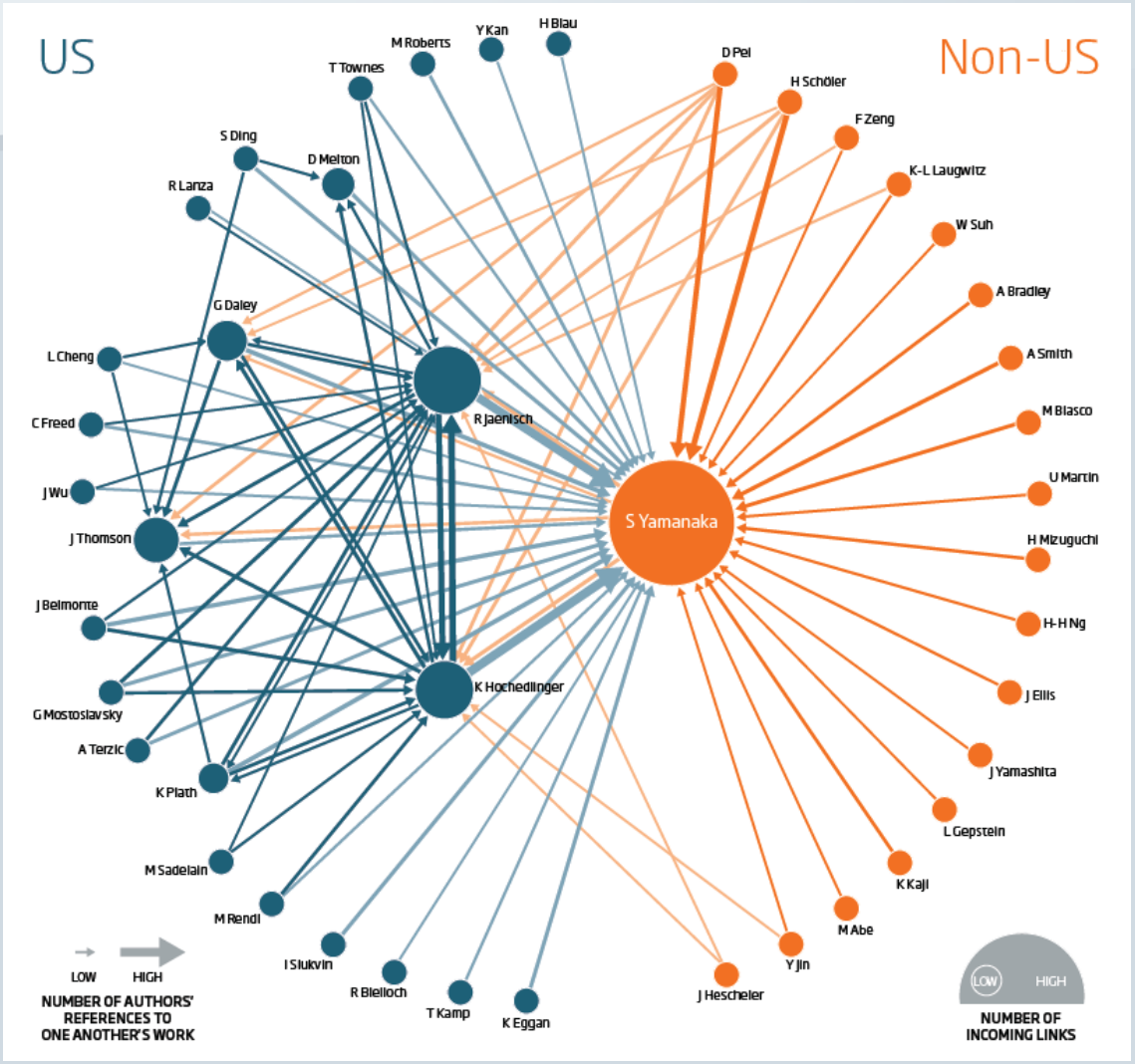
Peter Aldhous,
San Francisco Bureau Chief

From this ...

Analysis of iPS cell publications - data : iPS_all

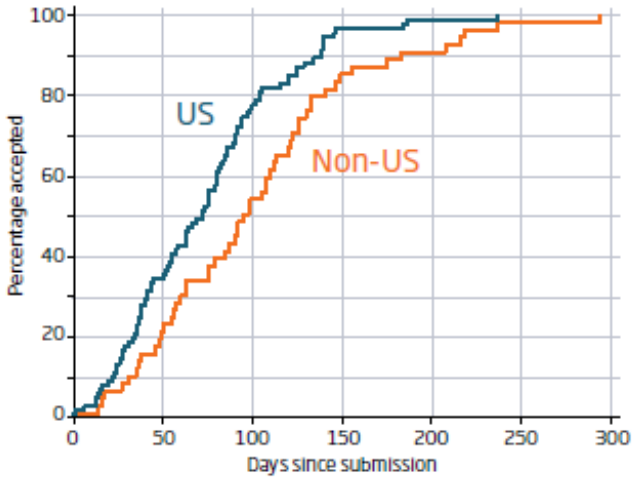
Title	Authors	Journal	Journal impact factor (2008)	Vol	Page	Author 1	Author 2	Author 3	Author 1 country	Institution
Electrophysiologic	Jiang, P; Rushing, SN;	Am. J. Physiol. Cell Physiol.	4.23	298	C486	R Li			US	Mount Sinai Sch Med
Induction and Isola	Taura, D; Sone, M; Hon	Arterioscl. Throm. Vas.	6.858	29	1100	M Sone			Japan	Kyoto U
Reprogramming of	Yakubov, E; Rechavi, G	Biochem. Bioph. Res. Co.	2.648	394	189	D Givol			Israel	Weizmann Inste
Generation of hum	Seifinejad, A; Taei, A; T	Biochem. Bioph. Res. Co.	2.648	391	329	H Baharvand	G Salekdeh		Iran	Royan Inste
Propagation of hur	Abraham, S; Sheridan,	Biochem. Bioph. Res. Co.	2.648	393	211	J Loring	R Rao		US	Salk Inste
In vitro pharmacol	Tanaka, T; Tohyama, S	Biochem. Bioph. Res. Co.	2.648	385	497	K Fukuda			Japan	Keio U
Gene targeting in I	Mitsui, K; Suzuki, K; Aiz	Biochem. Bioph. Res. Co.	2.648	388	711	K Mitani			Japan	Saitama Med U
The effects of carc	Yokoo, N; Baba, S; Kaic	Biochem. Bioph. Res. Co.	2.648	387	482	T Heike			Japan	Kyoto U
Differentiation of m	Morizane, R; Monkawa,	Biochem. Bioph. Res. Co.	2.648	390	1334	T Monkawa			Japan	Keio U
Generation of func	Ueda, T; Yamada, T; Hc	Biochem. Bioph. Res. Co.	2.648	391	38	T Yamara			Japan	Nara Med U
Multipotent adult g	Dressel, R; Guan, KM; I	Biol. Direct	3.724	4	31	R Dressel			Germany	U Göttingen
Evaluation of antia	Nakao, Y; Narazaki, G; I	Bioorg. Med. Chem. Lett.	2.531	18	2982	J Yamashita	Y Nakao		Japan	Kyoto U
Directed differentia	Grigoriadis, AE; Kennec	Blood	10.432	115	2769	A Grigoriadis			UK	Kings Coll
Generation of indu	Loh, YH; Agarwal, S; Pa	Blood	10.432	113	5476	G Daley			US	Harvard U
Definitive proof for	Okabe, M; Otsu, M; Ahr	Blood	10.432	114	1764	H Nakauchi			Japan	U Tokyo
Human-induced pl	Ye, ZH; Zhan, HC; Mali,	Blood	10.432	114	5473	L Cheng			US	Johns Hopkins U
AutoSOME: a clust	Newman, AM; Cooper, J	BMC Bioinformatics	3.781	11	117	J Cooper			US	UCSB
Regeneration and	Christen, B; Robles, V;	BMC Biology	4.734	8	5	J Belmonte			US	Salk Inste
Nanog Is the Gate	Silva, J; Nichols, J; Theu	Cell	31.253	138	722	A Smith	J Silva		UK	U Cambridge
Disease-specific in	Park, IH; Arora, N; Huo,	Cell	31.253	134	877	G Daley			US	Harvard U
Oct4-Induced Pluri	Kim, JB; Sebastiano, V;	Cell	31.253	136	411	H Schöler			Germany	MPI Münster
Role of the Murine	Sridharan, R; Tchieu, J;	Cell	31.253	136	364	K Plath			US	UCLA
Parkinson's Disea	Soldner, F; Hockemeye	Cell	31.253	136	964	R Jaenisch			US	MIT
Induction of pluripc	Takahashi, K; Tanabe,	Cell	31.253	131	861	S Yamanaka			Japan	Kyoto U
Induction of pluripc	Takahashi, K; Yamanak	Cell	31.253	126	663	S Yamanaka			Japan	Kyoto U
Role of MEF feede	Chen, MF; Sun, XR; Jia	Cell Biol. Int.	1.619	33	1268	J Ge			China	Sun Yat-sen U
Generation and ch	Gai, H; Leung, ELH; Co	Cell Biol. Int.	1.619	33	1184	Y Ma			US	Nevada Cancer Inste

… to this



What's the hold-up?

In a sample of 148 papers from high-profile journals, those from scientists outside the US took longer to be accepted for publication



Spreadsheets

Microsoft Excel

Open Office Calc

Google Documents spreadsheets

Database managers

Microsoft Access

MySQL

PostgreSQL

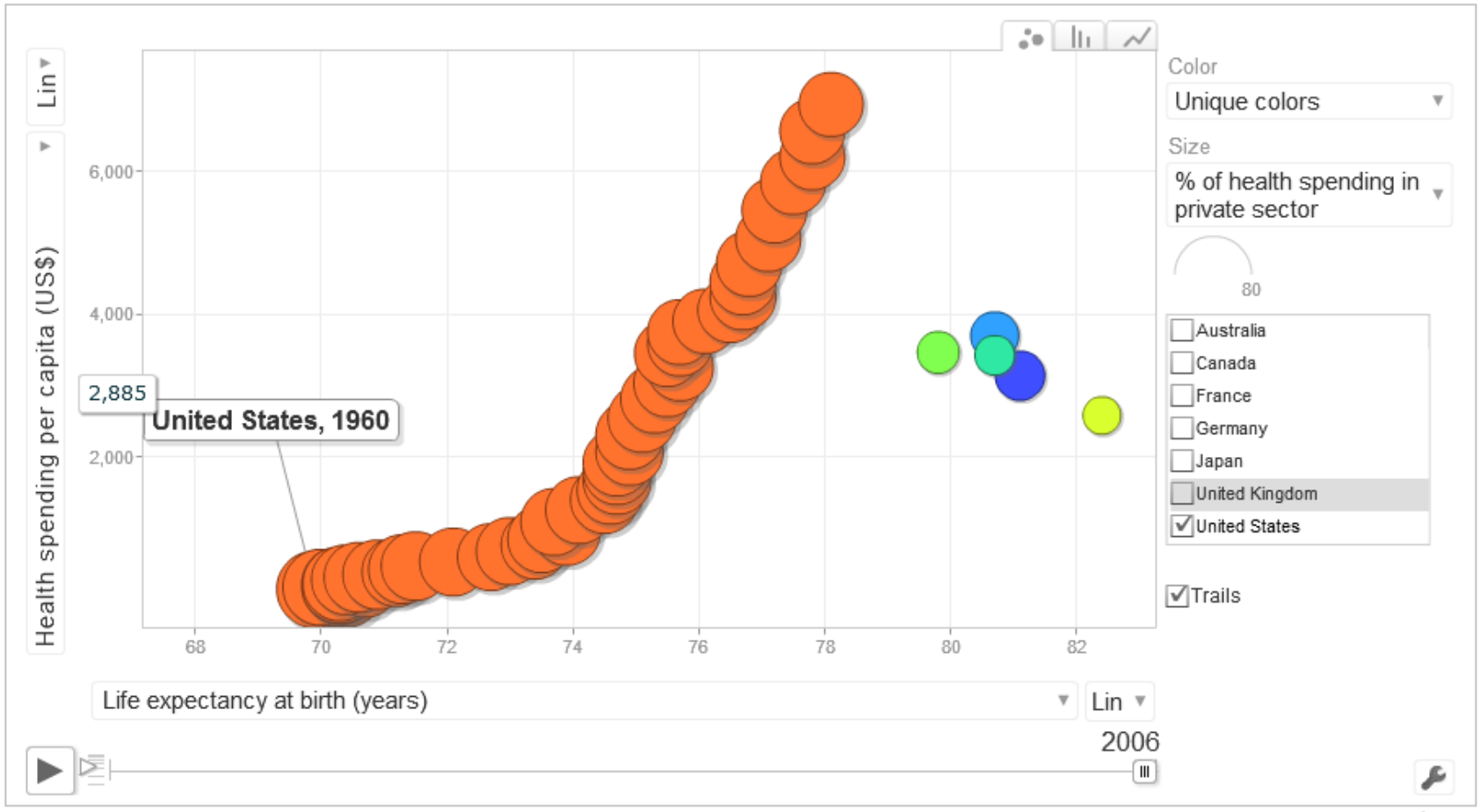
A rock star of data visualization

Watch the video

You can do it, too ...

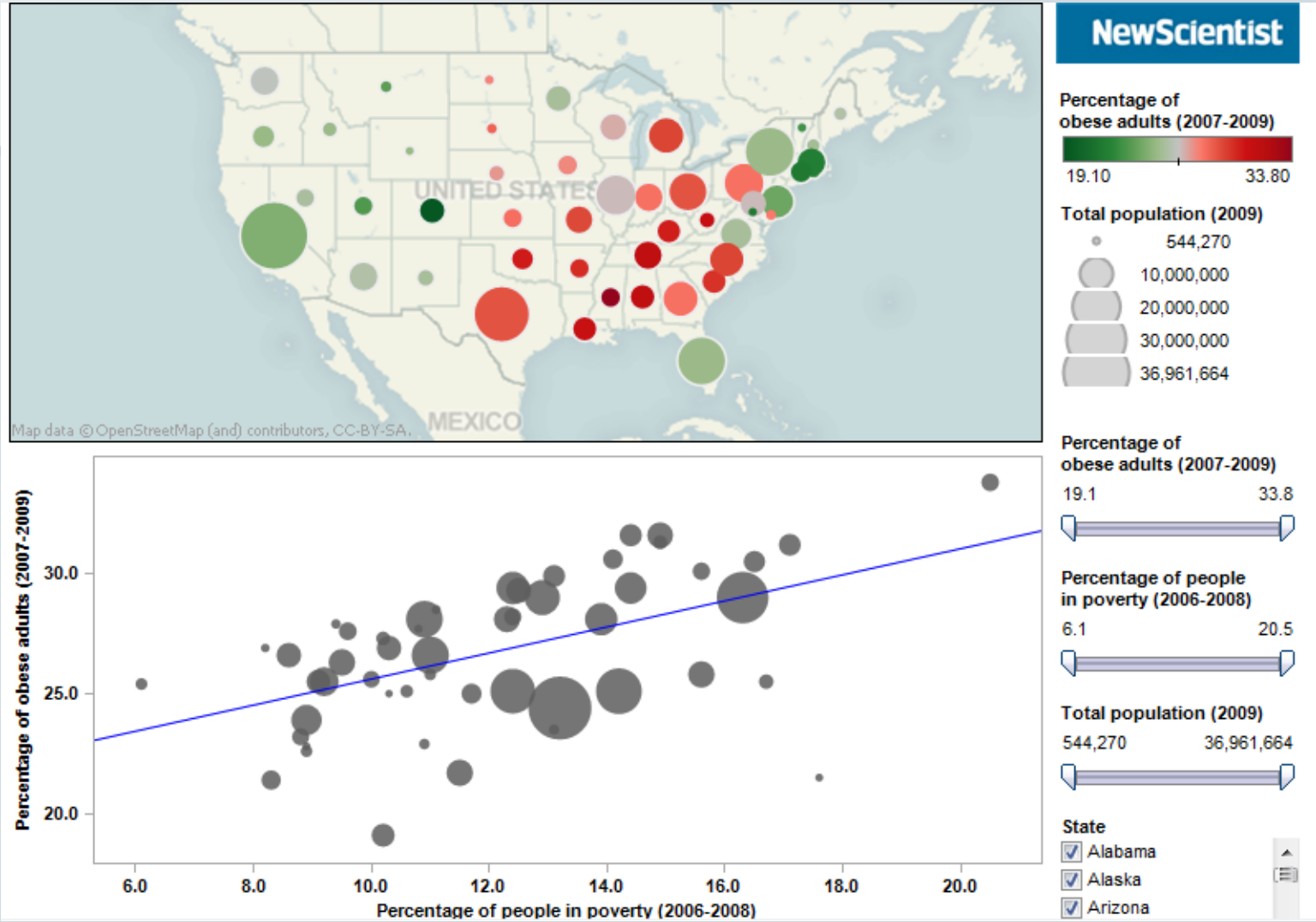
Healthcare costs and outcomes compared

Since the early 1980s, healthcare costs in the US have escalated compared to other developed countries, while life expectancy has lagged. Click the play button to start the animation; hover over the bubbles to identify each country. Explore other variables by clicking the small triangles on each axis and the legend. Data from OECD.



Explore the animation

Online interactive graphics made easy



Explore this graphic online

Free tools for online data visualization:

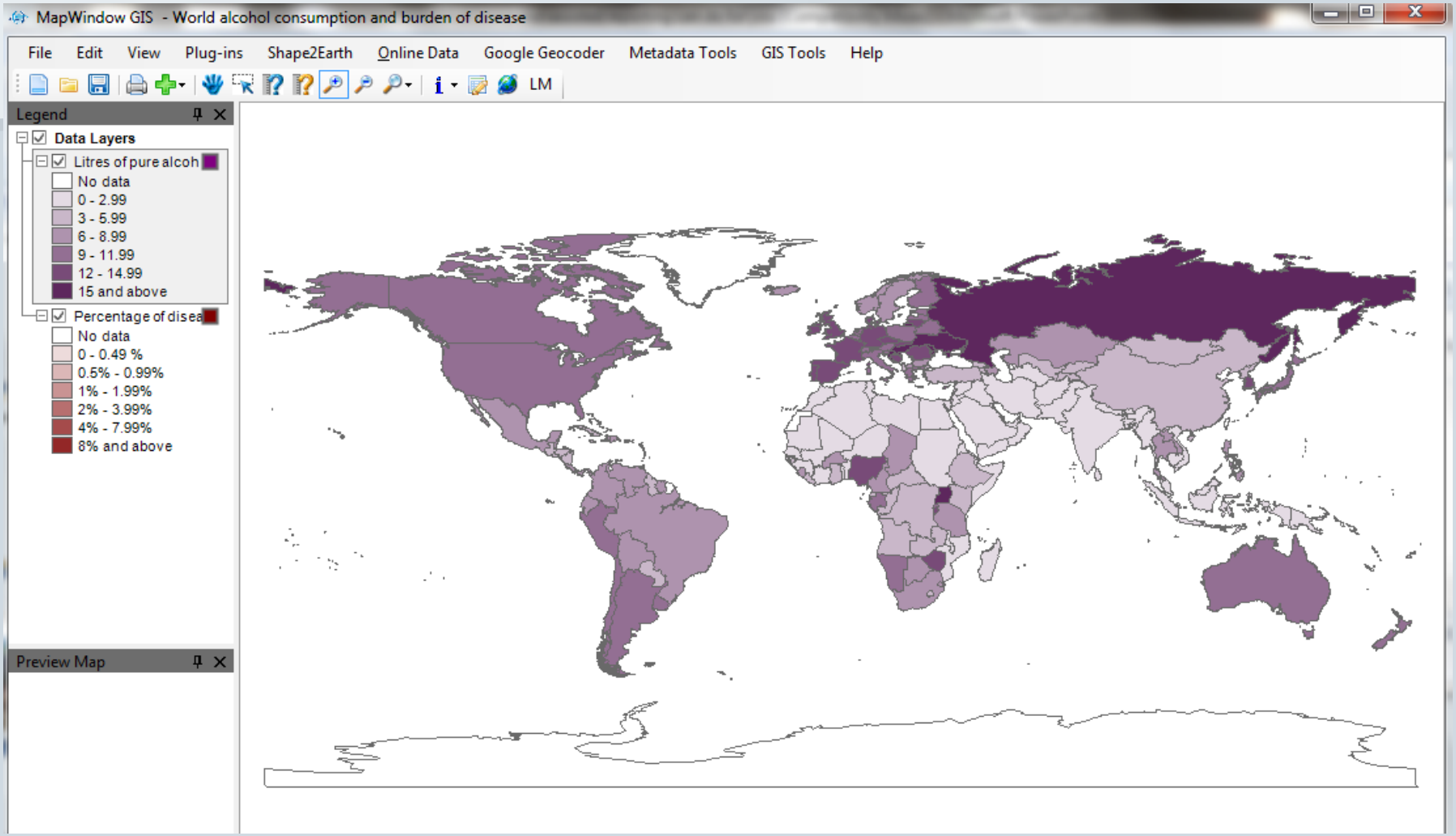
Tableau Public

Many Eyes

Google Documents Gadgets (includes Motion Charts)

Google Fusion Tables

Putting data onto maps: GIS ...



Dealing with booze's dark side

Smoking has become socially unacceptable. Now it's alcohol's turn

Alcohol can harm innocent bystanders

Andy Coghlan

HUMANITY's relationship with alcohol has never been easy. Now it is about to undergo as great a change as our attitude to tobacco, which has seen smoking plummet from the height of cool to the lowest of unpleasant habits.

That at least is the hope of the World Health Organization, which, between now and January, will be honing its draft of the first global strategy on reducing health damage from alcohol abuse, the fifth leading cause of premature death and disability worldwide.

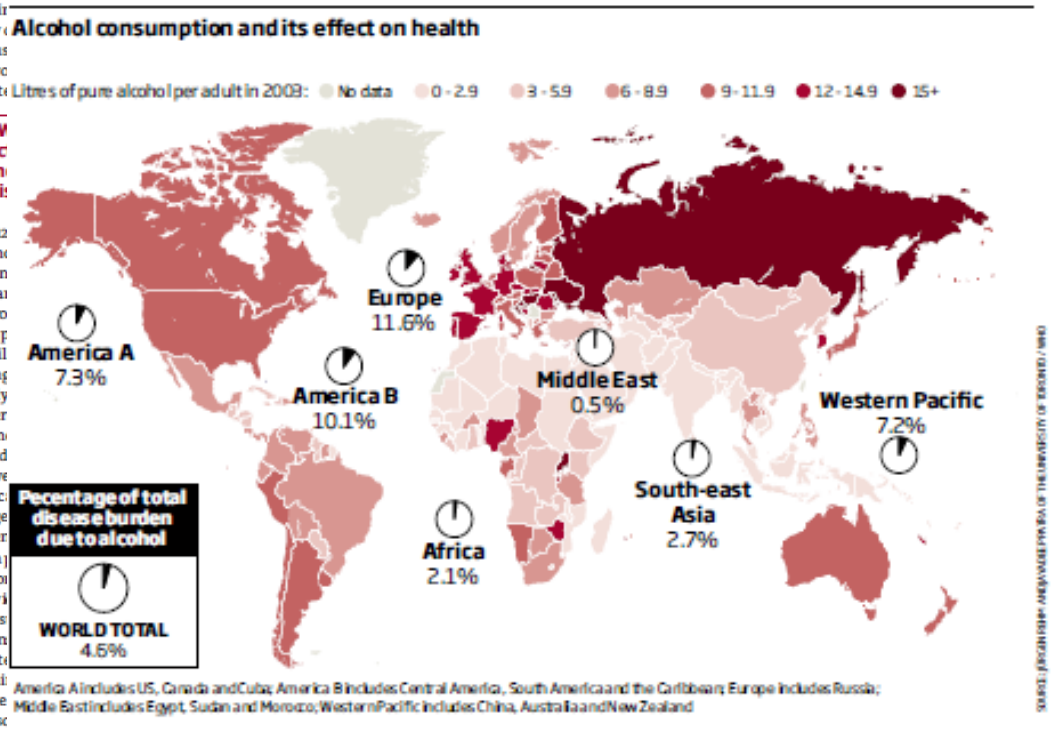
Unveiled last week in Geneva, Switzerland, the document is the culmination of talks between representatives from the WHO's 193 member states. "It is a landmark document," says Peter Anderson, a health consultant and adviser on alcohol to the WHO and the European Union.

Member states will be invited to ratify the finalised version of the document at the meeting of the World Health Assembly in May, but the document will not be legally binding. Its purpose instead is to raise awareness among governments about the importance of reducing alcohol abuse and to provide data that will persuade electorates that new laws are required – thereby emboldening governments to take action. The document will also present a menu of legal and governmental strategies that have been shown to work. "It will provide knowledge and awareness about the size of the problem, and advice about the most cost-effective policies," says Anderson.

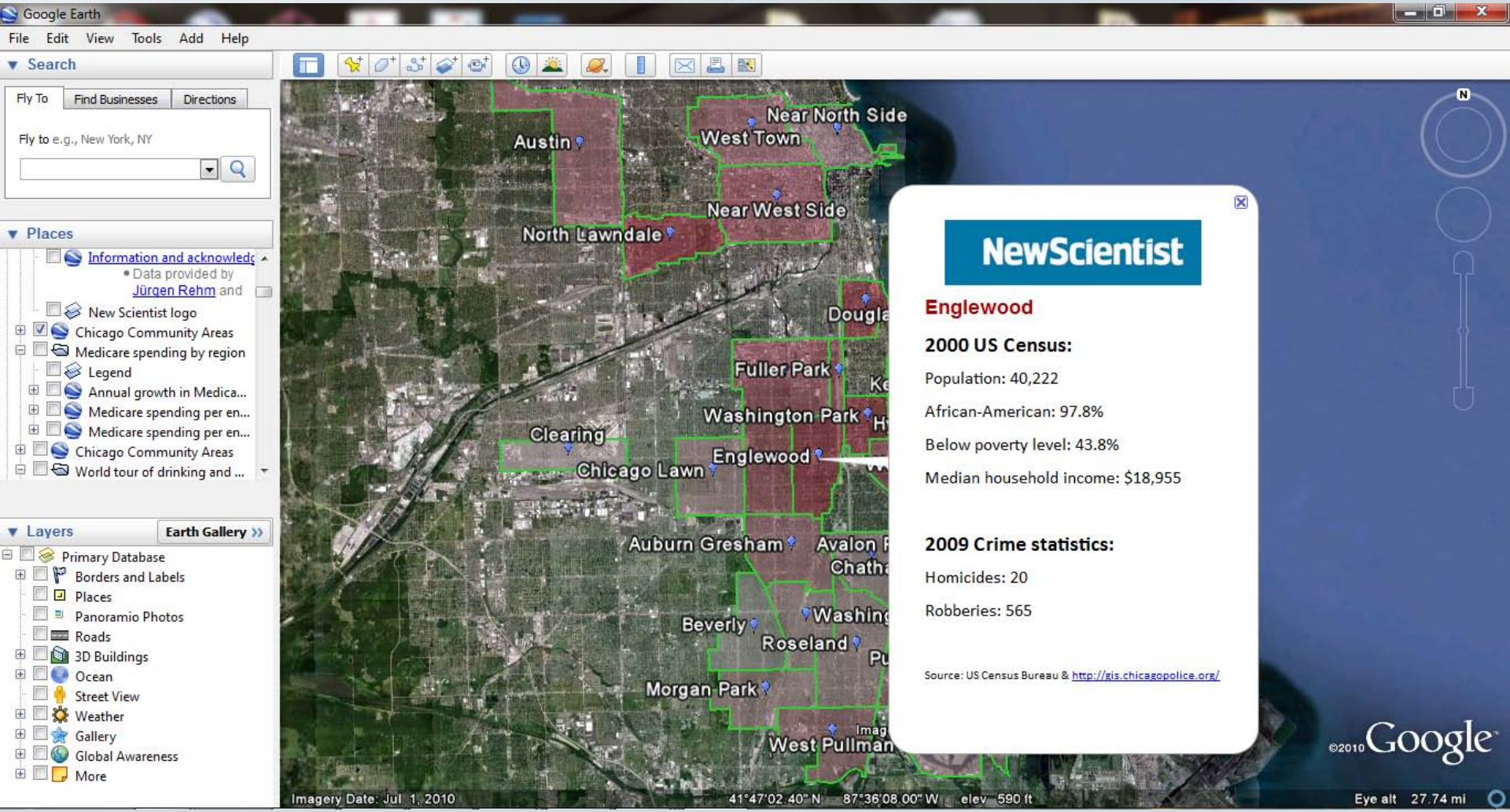
The impetus for action is founded on the growing realisation that alcohol doesn't just harm those who drink, combined with a better knowledge of intervention strategies. For example, in March the UK government's chief medical officer, Liam Donaldson, devoted a chapter of his 2008 annual report to "passive drinkin… heavy … To illus… the pro… report… countries, for example, the focus is likely to be on stopping young people from binge drinking, whereas introducing drink-driving laws may be a priority in rapidly developing countries,

"The W… effec… alcoh… to rais…

were 12… instan… that an… born a… syndro… 7000 p… 560 kil… driving… Sally… Univer… Zealan… WHO d… passive… public… stringe… challer… which… freedo… behavi… Pers… citizen… first st… any thi… people… To s… depend…



... and Google tools



NEIGHBOURHOODS THAT CAN KILL

The strain of inner-city life may send tumours into overdrive, reports Peter Aldhous. And cancer is just one disease fuelling US health inequalities

Diana Garmon-Spears knows breast cancer is more deadly for black women living in Chicago

DEEP down, Diana Garmon-Spears knew something was seriously wrong when she noticed a lump in her right breast, about the size of a peanut. "I ignored it, but then my breast started to deform. I started to form a mass of lumps all around," she recalls.

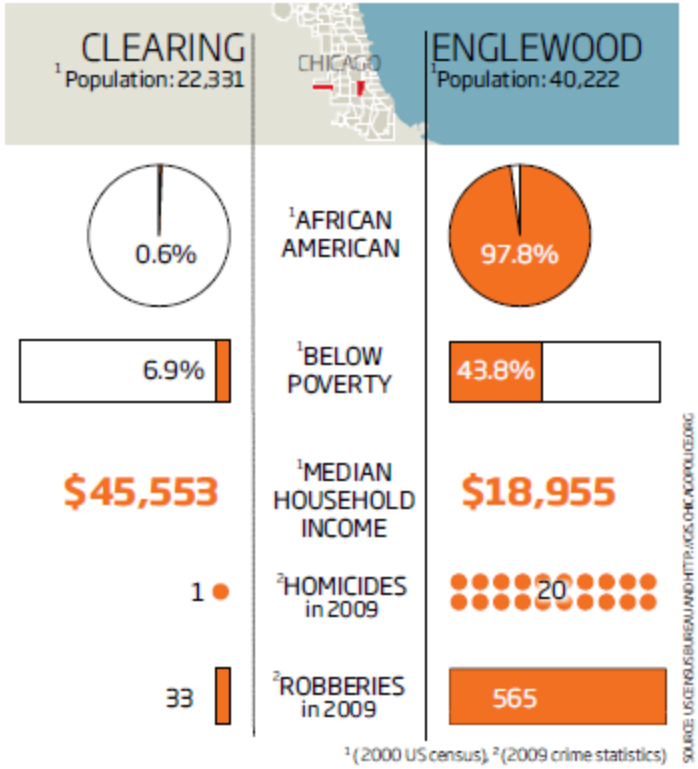
Two years later, she doesn't expect doctors to cure her cancer, and has been forced to give up her office work. "I don't think there's any employer who's going to understand you having chemo or radiation for 12 hours a day."

Asian African-American women living in Chicago, the dice are loaded against Garmon-Spears. Across the US, death rates among black women diagnosed with breast cancer are 37 percent higher than for whites, but in Chicago the difference is an astonishing 68 per cent (*Cancer Causes & Control*, vol 18, p 525). Something about this hearing motor pollution sending black women to an early grave.

Poor access to screening and therapy

The wrong side of town

There are stark social differences between a comfortable Chicago suburb like Clearing and an inner-city area like Englewood. Are these contributing to high death rates from breast cancer among black women living in the city?



[1] (2000 US census), [2] (2009 crime statistics)

SOURCE: US CENSUS BUREAU AND HTTP://CS.CHICAGOCRIME.ORG

NewScientist

Free GIS software:

Quantum GIS

MapWindow

Other free mapping tools:

Google Maps

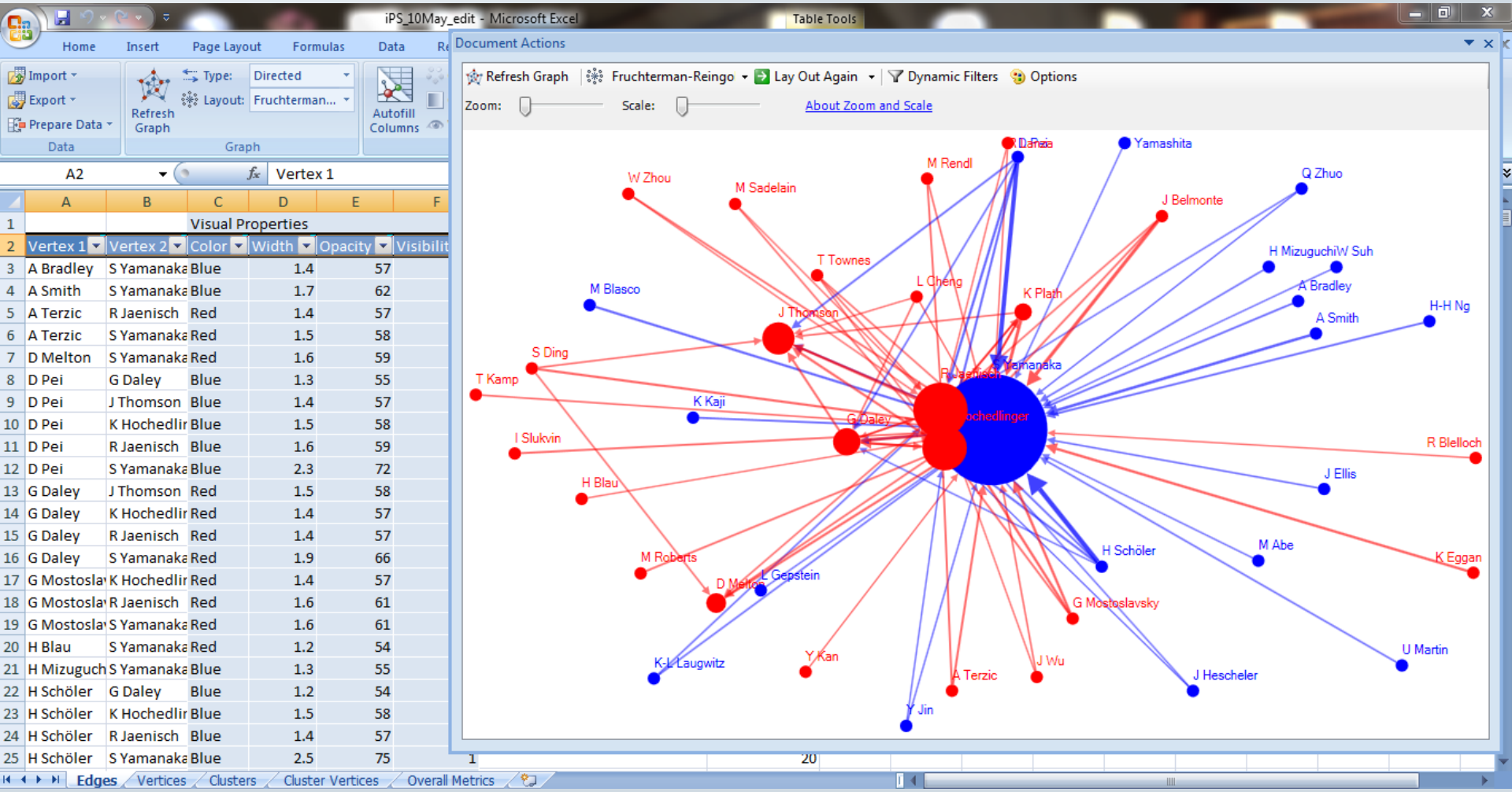
Google Earth

Google Fusion Tables

Geocommons

OpenHeatMap

Social network analysis

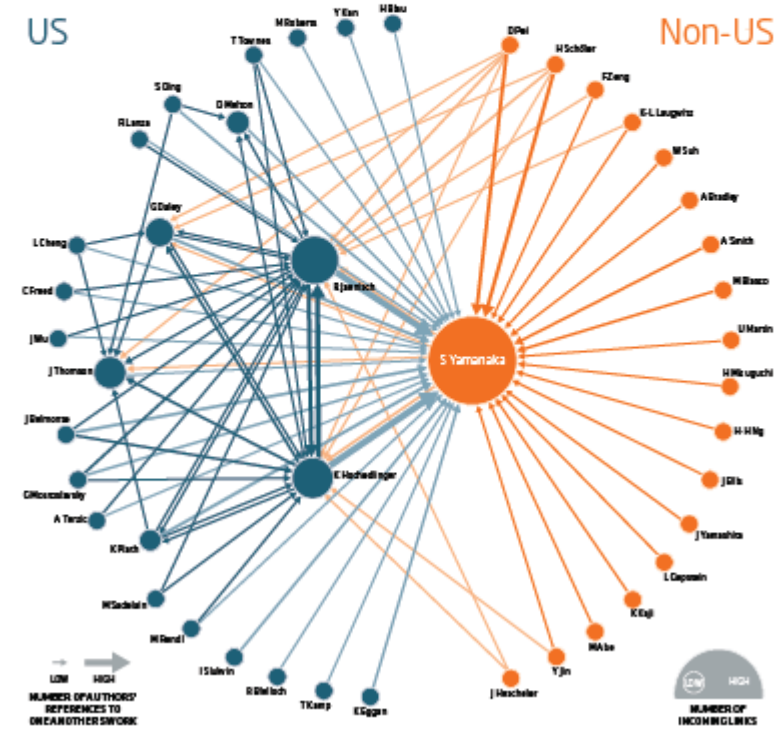


NewScientist

THE STEM CELL WARS

When a Nobel prize is up for grabs, do scientists across the globe compete on a level playing field? Peter Aldhous investigates

The most influential players in cellular reprogramming are revealed by recording how many times the scientists have referred to each other's work. Each link shows where one researcher cited another four or more times in papers in leading journals (for analysis, see "The strongest link", below right)



ALL'S fair in love and war, they say, but science is supposed to obey more noble ideals. New findings are submitted for publication, the studies are farmed out to experts for objective" peer review" and the best research appears promptly in the most prestigious journals.

Some stem cell biologists are crying foul, however. Last year, 14 researchers in this notoriously competitive field wrote to leading journals complaining of "unreasonable or obstructive reviews". The result, they claimed, is that "publication of truly original findings may be delayed or rejected".

Triggered by this protest, *New Scientist* scrutinised the dynamics of publication in the most exciting and competitive area of stem cell research, in which cells are "reprogrammed" to acquire the versatility of those of an early-stage embryo. In this fast-moving field, where a Nobel prize is arguably at stake, biologists are racing feverishly to publish their findings in top journals.

Our analysis of more than 200 research papers from 2006 onwards reveals that US-based scientists are enjoying a significant advantage, getting their papers published faster and in more prominent journals. The disparity is likely to spark debate when the International Society for Stem Cell Research (ISSCR) meets in San Francisco next week.

There are several plausible and reasonable explanations, but feelings are running high nonetheless. With two of the most delayed papers coming from a Japanese researcher who pioneered the field, and some of his rivals using controversial channels that give members of the US National Academy of Sciences an inside track to rapid publication, it is easy to see why.

The protest letter called for journals to publish the anonymised comments of researchers who act as reviewers of papers, to expose examples of potential obstruction. Just two of its signatories were from labs in the US. And when leaders of the protest talked to the media, unfair treatment of researchers outside the US was among the complaints. "There does seem to be this bias against groups from the rest of the world," Robin Lovell-Badge of the UK's National Institute for Medical Research in London told *New Scientist*.

Research on induced pluripotent stem (iPS) cells is the obvious place to look for biases in publication, given the high stakes involved. One of the signatories of the letter was the pioneer of cellular reprogramming – Shinya Yamanaka of Kyoto University in Japan. Less than four years after he first showed how to reprogram a mouse skin cell, Yamanaka is routinely mentioned as a candidate for a Nobel prize. He may be sharing that honour if other scientists make faster strides towards therapies based on cellular reprogramming. Our analysis of the citations between researchers reveals that Yamanaka is still the most influential figure in the field, but also shown that several well-connected US-based scientists are giving him a run for his money (see diagram, left, and "The strongest link", below).

New Scientist searched the Web of Science database for studies on iPS cells, recording the dates each was submitted, accepted for publication, and published. Advised by Matthew Strickland of Emory University in Atlanta, Georgia, whose research employs a branch of statistics called survival analysis, we found that papers submitted by authors outside the US took significantly longer to be accepted and published.

This difference was particularly clear for papers in 29 high-profile journals with an "impact factor" of 5 or more (see "What's the hold-up?", page 14). Impact factor is a measure of the frequency with which a journal's articles are cited in the scientific literature. We chose this cut-off score to focus on journals that received the protest letter, or those with similar prominence. "It's really very interesting," says Lovell-Badge. "I didn't think it would be possible get quantitative data."

So what might explain the pattern? Obstruction of papers could happen if a reviewer delays their comments, or makes many demands for changes. The journals with the greatest lag between US and non-US papers deny that their reviews are biased, and say that the former explanation can be ruled out. "The review process itself is quite short, and the majority of the intervening time is taken up by the authors performing revisions to address the criticisms that the reviewers raised," says Deborah Sweet, editor of *Cell Stem Cell*, which is the official journal of ISSCR – and is published by Elsevier, a sister company of the publisher of *New Scientist*.

Of papers published in leading journals:

78 per cent from US-based authors accepted within 100 days

54 per cent from authors elsewhere accepted within 100 days

THE STRONGEST LINK

Shinya Yamanaka of Kyoto University in Japan is the dominant scientist in cellular reprogramming, but he has stiff competition from a well-linked group of US-based researchers.

To map influence in the field, *New Scientist* constructed a social network diagram (left) based on citations, the references to each scientist's work by their peers. Citations are a measure of a researcher's impact and influence, and are sometimes used to help make decisions on promotions. They can also provide a snapshot of who's who in a field.

Assisted by Henri Schildt of Imperial College London, a specialist in citation analysis, we looked at references between 148 papers published in prominent journals since 2005 – drawing links where the authors cited one another's work four or more times.

Yamanaka's research is referred to by just about everyone. But there are no such links between other scientists outside of the US, and no links to them from any of the US-based researchers. In the US, there is a richer web of connections.

In large part, this reflects the greater number of papers in our sample from scientists in the US. But another tie links the best-connected researchers in the US: the Boston area. Rudolf Jaenisch at the Massachusetts Institute of Technology is Yamanaka's strongest rival, and two of the other main players – Konrad Hochedlinger and Kathrin Plath – used to work in his lab. Plath has moved to the University of California, Los Angeles, while Hochedlinger remains nearby at the Harvard Stem Cell Institute, which also hosts the labs of George Daley and Doug Melton.

The outsider is James Thomson at the University of Wisconsin-Madison, who first isolated human embryonic stem cells in 1998. He owes his prominence in this network to winning the race, in a tie with Yamanaka, to make human iPS cells.

Bigger, better?

Konrad Hochedlinger of the Harvard Stem Cell Institute, among the foremost in the field, suggests that leading US labs can deal with revisions more easily because they are often larger and better funded. When asked to run extra experiments, for example, he can quickly deploy junior scientists or hire a commercial lab to do them.

Could it also be that the US-based scientists tend to produce better work? There's no simple way to measure the calibre of studies in our sample – and if papers are rejected by leading journals and end up lower down the ›

Software for social network analysis:

NodeXL (free, extension to Excel 2007/2010)

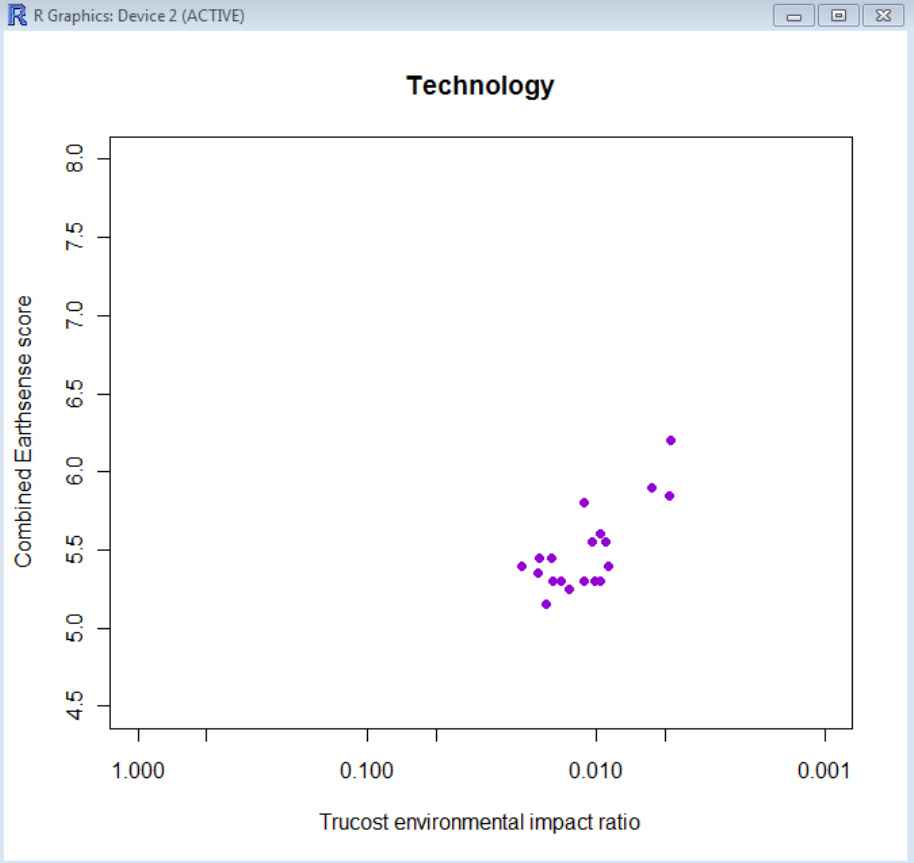
Gephi (free, also does other forms of analysis and visualization)

UCINET (free trial version for 60 days, then $250)

If you're feeling more ambitious ...

RGui

File Edit View Misc Packages Windows Help



R Console

```
> str (ICB_Technology)
'data.frame':   19 obs. of  15 variables:
 $ Ticker                : chr  "AAPL" "7751-TO" "DELL" "EMC" ...
 $ Industry.Parent.Group: Factor w/ 11 levels "Alcoholic Beverages",..: 2 2 2 2$
 $ ICB.Super.Sector      : Factor w/ 11 levels "Automobiles & Parts",..: 10 10 1$
 $ GICS.Industry.Group   : Factor w/ 15 levels "Automobiles & Components",..: 15$
 $ Sample.Size           : int  2000 1881 2023 1891 1847 1888 2043 1851 1974 189$
 $ Familiar.with.Company: num  0.97 0.95 0.96 0.22 0.98 0.98 0.91 0.33 0.81 0.9$
 $ Respondent.count      : int  1929 1774 1950 404 1814 1847 1852 599 1589 1873 $
 $ Customers             : num  0.46 0.61 0.63 0.1 0.9 0.76 0.62 0.2 0.46 0.94 .$
 $ Company.Earthsense    : num  6.1 5.6 5.9 5.2 5.6 5.8 5.5 5.4 5.6 6 ...
 $ Product.Earthsense    : num  5.5 5.3 5.3 5.4 6.8 5.3 5.4 5.2 5 5.7 ...
 $ Combined.Earthsense   : num  5.8 5.45 5.6 5.3 6.2 5.55 5.45 5.3 5.3 5.85 ...
 $ Trucost.EIR           : num  0.0113 0.0157 0.0096 0.0101 0.0047 0.0104 0.0176$
 $ Trucost.GHG           : num  0.0018 0.0027 0.001 0.0015 0.0008 0.0015 0.0054 $
 $ Trucost.Water         : num  0.0009 0.0011 0.0003 0.0008 0.0002 0.0004 0.0014$
 $ Trucost.Other         : num  0.0086 0.0119 0.0083 0.0078 0.0037 0.0085 0.0108$
> cor.test (ICB_Technology$Trucost.EIR, ICB_Technology$Combined.Earthsense, met$

        Spearman's rank correlation rho

data:  ICB_Technology$Trucost.EIR and ICB_Technology$Combined.Earthsense
S = 1751.465, p-value = 0.01791
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.536373

Warning message:
In cor.test.default(ICB_Technology$Trucost.EIR, ICB_Technology$Combined.Earthse$
  Cannot compute exact p-values with ties
>
```

Free software for statistical and graphical analysis:

R, plus RAndFriends for more user-friendly interfaces

NewScientist

SPECIAL INVESTIGATION

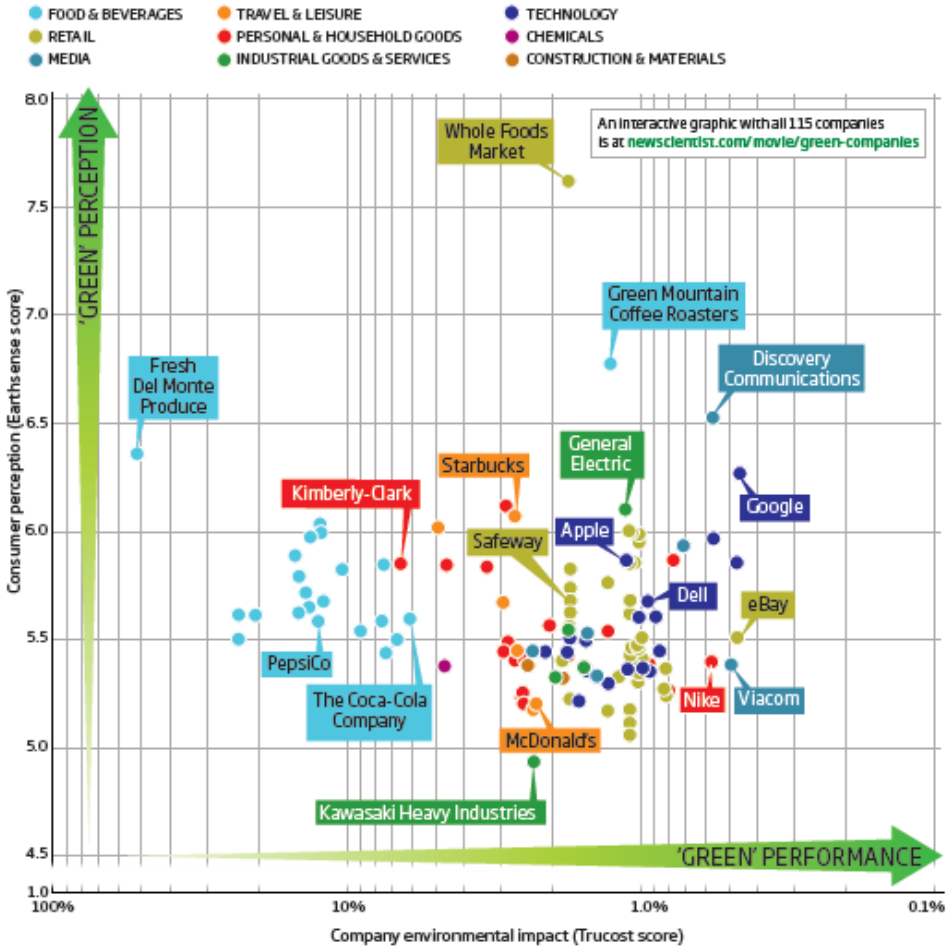
HEY, GREEN SPENDER

Do our ideas of which companies are eco-friendly live up to reality?
Peter Aldhous and Phil McKenna investigate

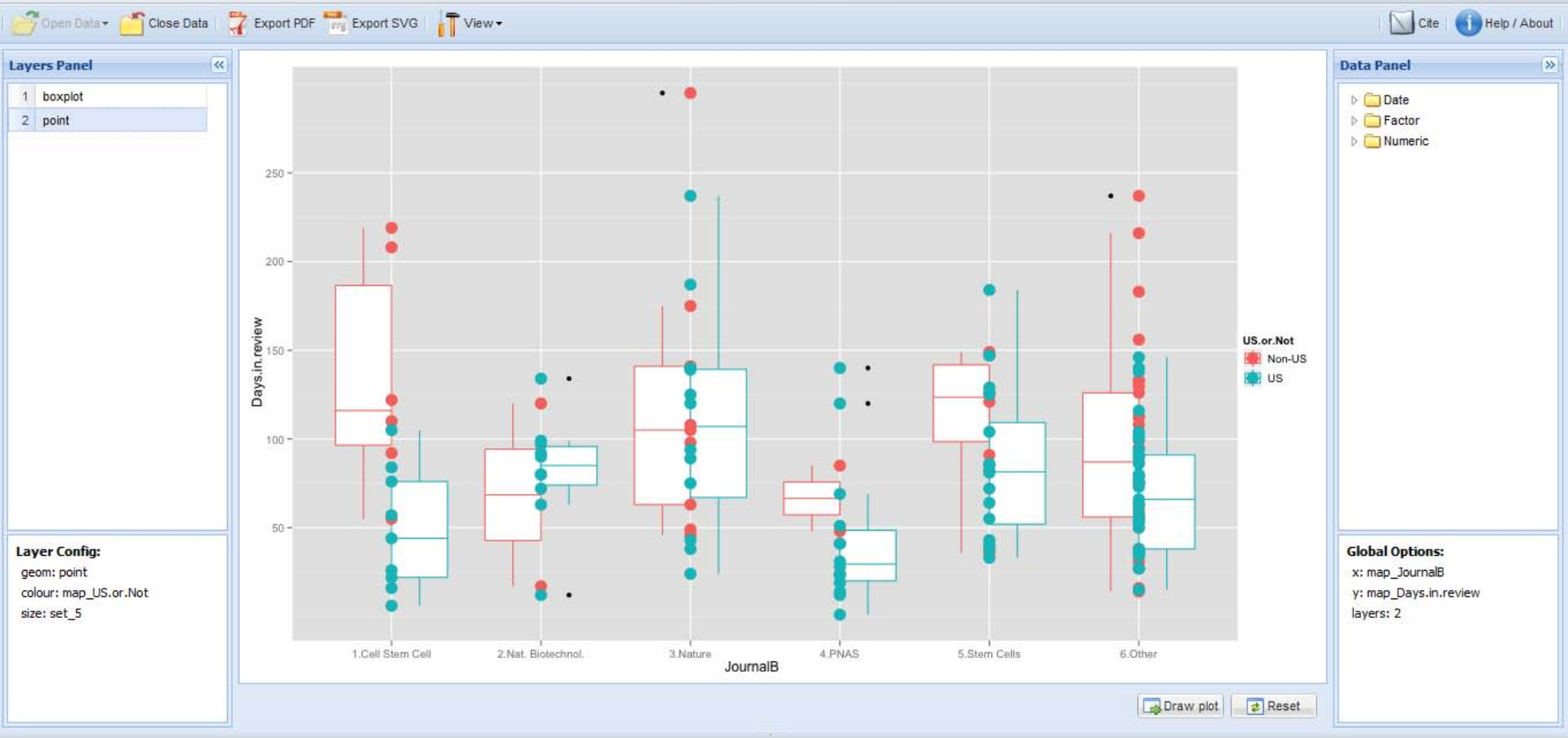
Consumer perception and environmental realities

There is very little correlation between how green a company is and how green it is perceived to be

Earthsense score – 2008 survey asking 30,000 US consumers to rate greenness of companies and products on a scale of 1 to 10
Trucost score – Estimated cost of the environmental impact of a company under a "polluter pays" system as a percentage of its annual revenue



Statistical graphics made easy



[Try it](#) with your own data

Data visualization in reporting and storytelling

Peter Aldhous,
San Francisco Bureau Chief