

Data, Technology, and Innovation in Government

Seminar 7: Data Journalism March 25, 2015

Nick Sinai
Walter Shorenstein Media and Democracy Fellow
Shorenstein Center for Media, Politics, and Public Policy
Harvard Kennedy School

Seminar Background

Personal Background

- Government: White House, Federal Communications Commission,
- Technology VC: Insight Venture Partners, Lehman Brothers, Polaris

Goals

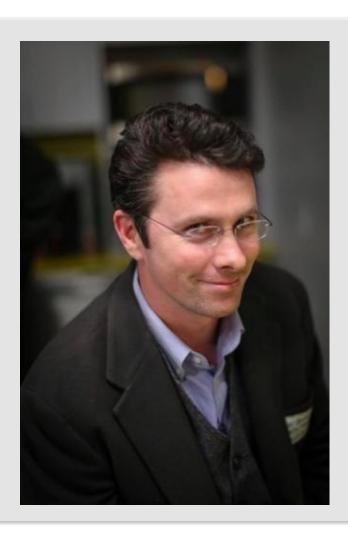
- Expose you to real projects, policy, and people
- Raise your Tech IQ
- Get credits toward your GSD degree

Logistics: Students get priority, seminar is not for credit

Seminar Series

- Introduction to Open Data
- Government Digital Services
- Privacy and Personal Data Portability
- Marketing U.S. Data: Data Jams, Datapaloozas, Hackathons, Prizes, & CDOs
- Government Spending Transparency
- Smart Cities/Internet of Things
- Data Journalism
- Freedom of Information Act
- Healthcare.gov: A Case Study
- Regulatory Modernization

Today's guest:



Alex Howard

- Founded blog "E Pluribus Unum," a top blog on government information technology, in 2013
- Columnist at TechRepublic
- Contributor to TechPresident

Past:

- Fellow at Columbia Tow Center for Digital Journalism
- Fellow at HKS Ash Center
- Washington Correspondent for Radar at O'Reilly Media

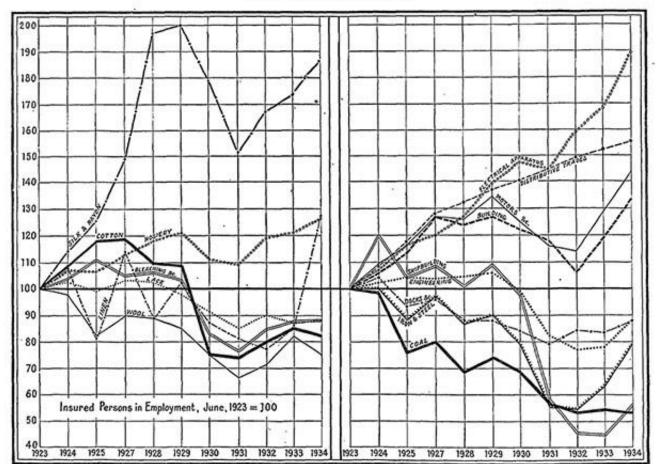


This John Snow knew something.

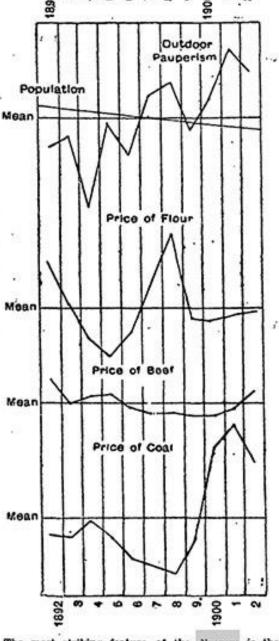


Newspapers have used data for centuries

ELEVEN YEARS' CHANGES IN EMPLOYMENT: CONTRACTING AND EXPANDING INDUSTRIES, 1923-1934



Source: The Guardian



The most striking feature of the diagram is the marked and rapid rise in the number of persons receiving outdoor relief. The lowest point, 1894, represents a weekly average of 395 persons in receipt of relief, while the highest, 1901, stands for 817 per week.

The Mileage of Congress

Published in The New York Tribune, Dec. 22, 1848

In the mid-1800s, Horace Greeley was the popular and controversial editor of The New York Tribune. For a few months starting in late 1848 he was also a congressman from New York. During that time, he produced this investigative data story, which accuses Abraham Lincoln, among many others, of taking too much money for mileage to and from the Capitol. Related Story *

PAGE

TEXT

would probably increase the House Excess to over

25 40 350 40 489 40

758 00 1288 00

, 59 20 1256 00

64 80

51 20

NEW-YORK TRIBUNE.

NEW-YORK, PRIDAY DEC. 29.

ILLISOIS.—The new Legislature of this State will soon assemble at Springfield, and its proceedings will be watched with anxiety, mainly with reference to the election of a United States Senster for the six years ensuing. Though a decided majority of the popular vote has just been given for the Whig and Free Soil Electoral Tinkets together, yet the Legislature is decidedly Loco-Foco, and will choose a Senator accordingly. It is said, however, that among the members of the majority are several decided Free Soil men, and and that the Whigs, by uniting with these upon Hom. Robert Smith or some other of the least exceptionable Cass-men, may secure the return of a practically and reliably Free Soil Senator. If so, we trust they will not miss the opportunity.

LAYE FRON Rise Jameiro.—The bark E. Corning, from Rio Junotro, arrived yesterday morning, buring left that port on the Sth November. The E. C. brings as intelligence of the arrival at Rio of the aspaniship California, Captain Forens, in the remarkable short passage from this port of 36 days. It will be remambered the California is one of the small steamers.

The Californian newspaper, published at San Francisco, says that the people of that territory are united, as one man, against the establishment or interdencing of Siavery.

An account of the opening of the Worcester and Maskua Railroad is crowded out this morning. It will appear in the Evening Edition.

Exview.—A review of Miss Lynch's Poems, intended for this morning's Tribuns, has been crowded web, but win appear in our Evening Edition.

The Mileage of Congress.

The following transcripts of the Mileage, or compensation of Members of the present Congress for travelling to and from Washington by reason of their attendance at the last Session, has been obtained and prepared for The Tribine, not without considerable delay and difficulty. The schedule of the House was open to us, and no objection made to its use; but the Secretary of the Senate refused to permit a transcript of the Mileage of that illutrious body to be taken, even on the requisition of a Member of the House, and we were constrained to apply to the Secretary of the Treasury, certify, ing that it was wanted as the basis of action in the House; and then our reporter was tossed not. Wherefore, we entreat you, men in Congress! to reform the Mileage at this present Session! HOUSE OF REPRESENTATIVES.

HOUSE (ctual No. of			Excess of	. I pess wempers byte contact or	
Names . N	ion by Post	20.00	Miner charged.	M.wees	Office Department list allows them, as	follows:
	Bouts."	charges	L charged.	800 40	Atkinson, Va\$14 40 Hampton,	I. J
mos Abbott, Mass	454	487	8389 60	329 00	Broadhead, Pa 7 30 J. R. Ingen	oll. Pa
Freen Adams, Ky	519	931	744 60	36 00	Brown, Pa 80 Levin, Pa	
Peorge Ashmun, M Arch'd Atkinson, V	ase. 363	408	326 40 224 00-	36 00		
rch'd Atkinson, V	998	290	337 90	46 40	SENATE.	
D. M. Berringer, P.	C 310	1122	897 60	368 40	Artest No. of	
Vash Barrow, Tes	ID. 004		691 00		Names. Raise by Post Males	-
Chomas H Bayly,	a. 197	300	240 00	82 40	Boste chirrent, (Margori.
Rich'd I T Reals	Va. LG	135	108 00	22.2		NO 20
denry Bedinger, Vi	65	149	119 20	67 20	Ches.Ashley(dead)Ark.2200	
illiam Belcher, me.		696	548 80	52 00	D. R. Atchison, Moinot down 72120 10 C. G. Athertoe, N. H., 447 540	96 00
K. H. Biocham, Mich	1 544	1121	896 80	461 60	C. G. Athertoe, N. H., 447 540	135 00
nabura Birdeall, N	Y. 296	390	472 00	235 20	Geo. E. Badger, N. C., 284 286	230 40
obn Blanchard, Pa	177	212	169 60	28 00	A. P. Bagby (out) Ala 1398	
. B. Bocock, Va	(not dow	n 256	204 80			956 40
obn M. Botts, Va	117	131	104 80	11 20	John Bell, Tenn 564 1122	97 60
W. Bowdon, Ala.	757	1148	918	312 80	Thos. H. Benton, Mo., 609 1670 13	336 00
ames B. Bowtin, M	0. 808	1508	1122 40	476 00	John M. Berrien, Ga., 662 760	608 00
inn Boyd, Ky		1300	1040 00	437 60	Solon Borland, Ark 1065 2360 1	O0 90
athan'l Boydon, N.	C 255	430	344 00	60 00		540 00
atasa i Boyood, A		130	104 00	32 00		336 00
asper E. Brady, Po						744 80
amuel A Bridges,	PE 100	189	151 20	7 20		159 20
tichard Brodhead,		190	152 00-			
Vm. G. Brown, Va.		330	264 00	98 40	J. C. Celhoun, S. C 531 923	738 40
harles Brown, Pa.	138	137	109 60-			130 00
libert G. Brown, Mi	se.1C47	2330	1864 00	1026 40	Lewis Case (out, Mich. 524 1081	964 90
lylett Buckner, Ky	611	987	789 60	300 60	John H. Clarke, R. L., 400 450	160 00
irmisteed Bart, 8.	C 548	740	592 00	153 60	John M. Clayton, Del., 115 120	96 00
hester Butler, Pa.	231	274	219 20	34 40	W. T. Colgaltt, Ga 1040	S32 00
C. C. Cabell, Fla	1069	1180	944 00	88 60	Thos. Corwin. Ohio 469 765	512 00
Richard S. Conby.	0 456	1053	842 40	477 60	J.J.Crittenden(out, Ky. 542 900	640 00
The W Catheres	-4 660	1806	1444 80	916 80		352 00
has W. Catheart,		40	32 00	6 49		584 80
lohn G. Chapman, I	Md. 32		32 00	216 00		164 80
Jucien B Chase, Te	mn. 730	1000	800 00			460 80
tea W. H. Clapp, M	e 545	600	480 00	44 00		
ranklin Clark, Me	588	651	520.80	9 60		320 00
Franklin Clark, Me Beverly L. Clark, K	y 688	1062	849 60	299 20		480 UO
L. Clingman, N. C	486	567	469 60	90 60	S. A. Douglas, Ill 884 1834 1	467 (31)
Howell Cobb, Ga	610	805	644 00	156 00	S. W. Downs, La 1190 2800 2	240 00
W. R. W. Cobb, Al	667	1300	1040 00	506 40	Jno Fairfield(dead)[not down] 600	90 00
Wm. M. Cocke, Ter	n 466	1104	883 20	510 40	A. Peich, Mich 526 1121	ry6 80
lacob Collamer, V		610	458 00	90 00	T. Pitzgerald, Mich [not down]1281 1	024 60
William Collins, N	Y 436	600	450 00	131 20	H. S. Foote, Miss 1010 2540 2	064 00
		557	445 60	194 40		360 00
H. S. Conger. N.Y.	313	430	344 00	17 60		453 60
Robt. B. Cranston.	H.1. 408			47 20		590 40
John W. Cristield,		209	167 20			281 60
John Crowell, O		885	708 00	465 60		
John H. Crozier, Te	EDD. 498	1071	856 80	458 40		496 00
J. D. Cummins, O. John R. J. Deniel,	laot dow	n 500	400 00		R M. T. Hunter, Va., 98 115	92 00
John R. J. Deniel.	N.C. 211	232	185 60	16 80		983 20
Mason C. Derling,	Wie. 943	1825	1460 00	704 00	Reverdy Johnson, Md. 40 42	33 (4)
John Dickey, Pa	253	360	288 00	85 60	Her. V. Johnson, Ga., 649 340	752 00
R. Dickinson, O	423	994	795 20	460 80	Her. V. Johnson, Ga., 642 240 W. R. King, Ala 918 1100 D. H. Lewis, (dead) Ala. 859 1090	980 00
James Dixon, Ct	336	375	300 00	31 20	D. H. Lewis, (dead) Ala. 859 1090	872 00
Rich'd S. Donnell, I	M.C. 348	443	354 40	75 (0		254 00
William Duer, N.Y.	373	554	443 20	104 80	J M. Mason, Va 74 143	144 40
Depiel Duncan, O.	368	996	796 60	504 00		557 60
		922	737 60	260 80	T. Metcalfe, Ky 493 697 J. W. Miller, N. J 294 282	225 60
Garnett Duncan, K	J 250				W. B S. Moor. Me. [not down 745	596 00
George G. Dunn, Is	d 621	1000	800 00	303 20		320 00
George G. Dunn, Is Geo. N. Eckert, Pa. Joseph E. Edsall, N	172	230	184 00	46 40		104 00
Joseph E. Edsall, N	.J. 254	317	253 60	50 40		124 00
Thos. O Edwards,	O., 375	773	616 80	316 80		
Elisha Embree, In-	4 712	1116	892 80 73 60	323 20		347 20
Alexander Evens	Md. 90	92	73 60	1 60	W. K. Sebastian, Ark. 1011 1900 1;	520 00
Alexander Evans, 1 Nathan Evans, Ohio	316	482	385 60	132 90	A. H Serier,(out) do not dwn 2125	700 00
Ismes I Person C	hio 400	790	638 00	238 40	P. Spruance, Del 132 150	20 00
James J. Parab, U.	10. 492	430		99 40	D. Sturgeon, Pa 195 300	240 00
James J. Paran, O. John W. Farelly, I W.S. Featherston, I	di 010		344 00	98 40 586 00	H. L. Turney, Tenn., 664 1207	¥65 60
W.S. Featherston, I	F186 310	1645	1316 00	266 (0)		992 00
Orlando B. Ficklin,	111. 805	1290	1032 00	388 00		180 00
David Pisher, Obio.	445	840	672 00	316 00		504 00
Thos. S. Flournoy,	Va. 214	325	260 00	88 80		20 00
Thos. S. Flournoy, John Freedley, Pa.	154	154	123 20		1.D. 101-1-1-1 - D. 1040 1000 (60 00
Richard French, K	y 514	785	628 00	216 80		190 40

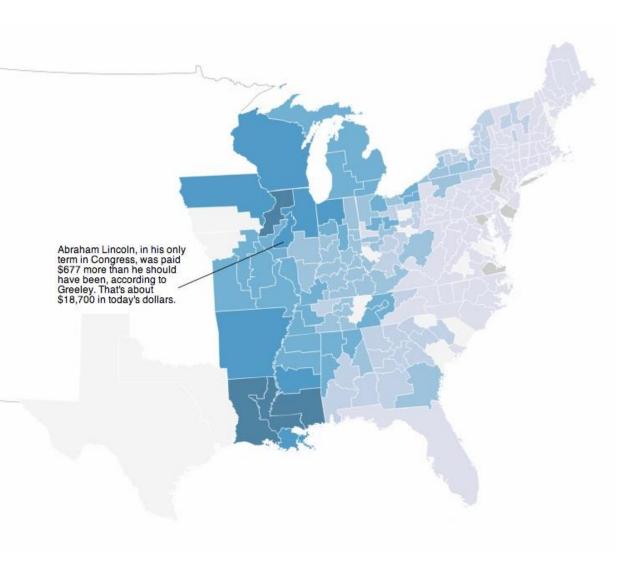
Horace Greeley New York Tribune 1848

Congressional Mileage Map

This map shows the "excess" mileage paid to each member of the 30th Congress, according to Greeley's story. Although the law that stood in 1848 specified only that the mileage would be paid by "the usually traveled route," Greeley argued that the postal route from each member's district to the Capitol ought to be the standard by which mileage was paid.

Higher excess mileage charged →

Notes: Mileage figures are based on original report in the Dec. 22, 1848 New York Tribune. We corrected arithmetic errors and rounded numbers to the nearest dollar. Sources: The New York Tribune, 30th Congressional District shapes by Jeffrey B. Lewis, Brandon DeVine, and Lincoln Pritcher with Kenneth C. Martis, U.C.L.A.



1960s: computer-assisted reporting (CAR)

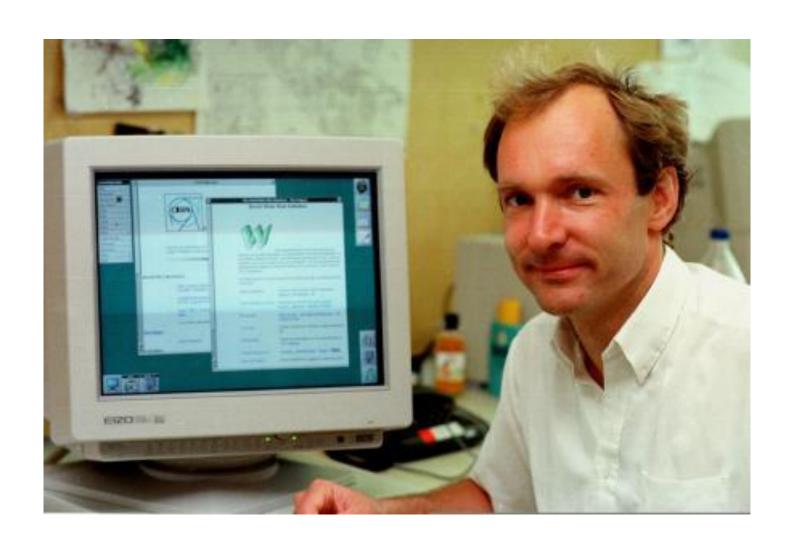


Bob Woodward, via Cliff1066

Traditional tools applying tech to journalism...

- Calculators and Graphs
- Mainframe and PCs
- Spreadsheets
- Databases
- Text and code editors
- Statistics
- Programming

In the 1990s, government and civil society spread the Internet globally



In the 2000s, mobile phones and social networking connected us ever more



In the 2010s, data creation exploded.



Image Credit: <u>Real Time Rome from Senseable.MIT.edu</u>

"Data-driven journalism is the future"



Source: <u>Tim Berners-Lee in the Guardian</u>

...combined with new tools & context...

- Online spreadsheets and wikis
- Data visualization tools
- Open source frameworks
- Code sharing
- Agile development
- Cloud storage and processing (EC2 & Heroku)
- More data and more access
- Privacy and security risks

2015: data journalism is the present

Gathering, cleaning, organizing, analyzing, visualizing and publishing data to support the creation of acts of journalism



Journalism in the public interest

Receive our top stories daily

Subscribe via e-mail

SIGN UP

zip-code

Home

Our Investigations

Tools & Data

Reporting Network

Blog

About Us



Docs on Pharma Payroll Have Blemished Records, Limited Credentials



Riundreds of doctors paid makes to promote their drups have been voPublics has freand. We name more stub believen

\$257.8 million in payouts since 2009 for speaking consulting and other duties. Road more.

This detabase only includes the service companies that have disclosed payments. Their 2009 sales totaled \$109 billion -

	Total	Quarterly
Et Lify	\$102.2 M	\$20.4 M
GlaxoSevite/Cine	\$95.0.14	\$17.4 M
AstroZeneca	\$20 A M	\$11.8 M
Plan	\$10.6 M	\$3.3 W
Marck	\$9.4 W	\$4.7 M
Johnson & Johnson	95.2 M	\$2.6 M

Look Up Your Doctor in Our "Dollars for Docs" Database

We compiled data from seven companies, covering \$257.8 million in payouts since 2009 for speaking, consulting and other duties. See if your doctor is in our data. Go to the database »

Doctors on Pharma Payroll Have Blemished Records, **Limited Credentials**

by Charles Ornstein, Tracy Weber, and Dan Nguyen, ProPublica, Oct. 19

Drug companies say they hire the most-respected doctors in their fields to teach about the benefits and risks of their drugs. But we've uncovered hundreds of doctors on company payrolls who had been accused of professional misconduct, were disciplined by state boards or lacked credentials. Their misconduct included inappropriately prescribing drugs, providing poor care or having sex with patients. Read more »

Lawsuits Say Pharma Illegally Paid Doctors to Push Their Drugs

Editor's Note

How Patients Can Use This Data Steal Our Dollars for Docs Widget

Trendy but not new

 The collection, protection and interrogation of data as a source, complementing traditional "shoe leather" investigative reporting relying on witnesses, experts and authorities

The Guardian

Cutswatch map: public sector cuts across the UK

Help the Guardian's reporting fearn plot the effects of the government's public sector cuts on local communities across the UK by participating in our Cutswatch project. Send us details about the cuts in your area or pubmit your accounts of how the cuts have affected you.
Find our more about how you can get involved.









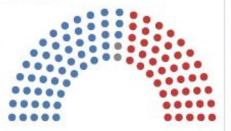


Search for lawmakers by name or state

Explore the 113th Senate

View members by party, link to member pages and see how frequently they vote with their party.





Recent key votes

SENATE | AGREED TO

Farm bill

\$956.4 billion package authorizes the end of billions of dollars in direct subsidy payments to ...

HOUSE | PASSED

Farm bill (>)

\$956.4 billion farm bill makes significant changes to the nation's farm support programs, but also ...

SENATE | AGREED TO

\$1.1 trillion spending bill (2)

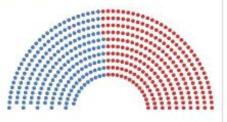
\$1.1 trillion omnibus spending package funds government activities through the end of fiscal year 2014.

View more votes: Senate | House

Explore the 113th House

View members by party, link to member pages and see how frequently they vote with their party.

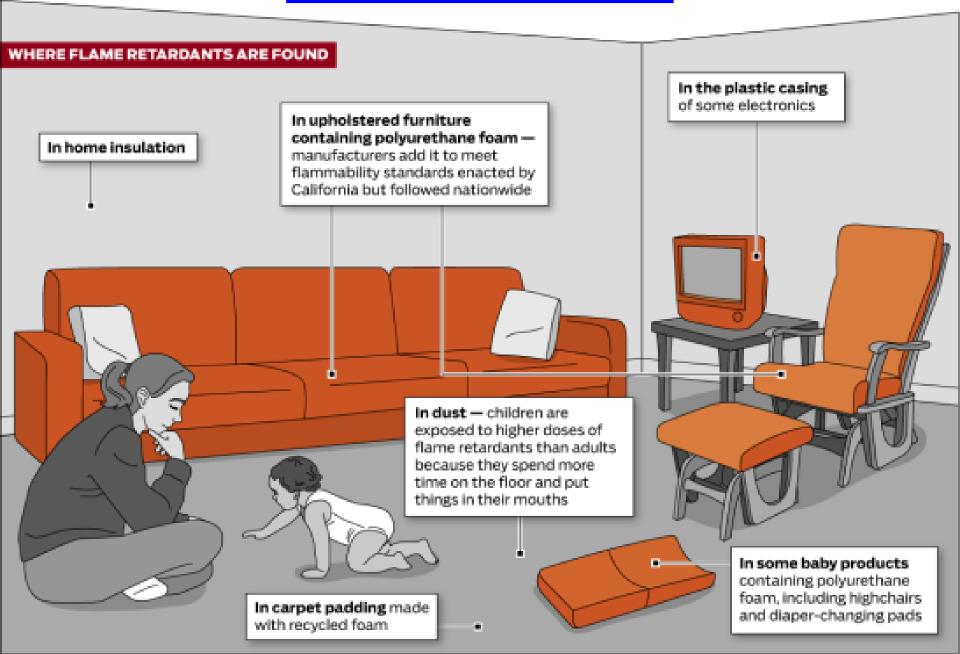
Explore the House (>)

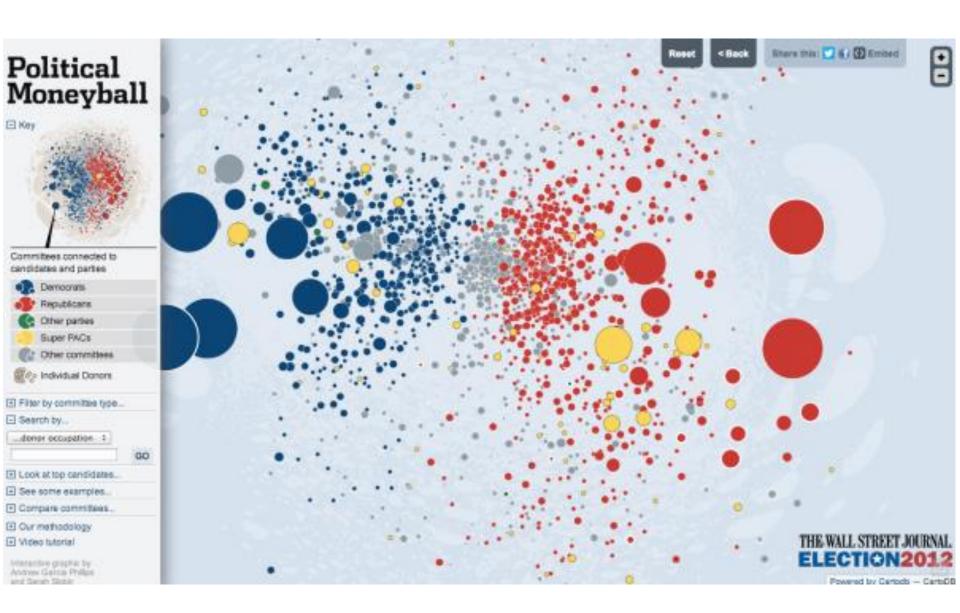


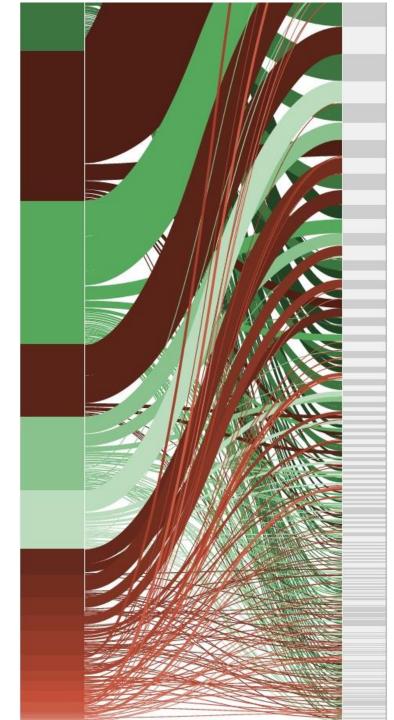
About the U.S. Congress Votes Database

The U.S. Congress Votes Database documents every vote and member of the House and Senate since 1991. Data is pulled from several sources, including the House clerk, the U.S. Senate and the Biographical Directory of the U.S. Congress. For the 112th Congress, users can analyze members and votes by various metrics, including caucus affiliations, 2010 margin of victory, and members endorsed by Tea Party Express or Freedom Works during the 2010 campaign. This site is a work in progress and will be expanded over time. If you have ideas on how to improve it, please e-mail us at politicsgraphics@washpost.com.

Chicago Tribune







Payments to

PayChex, Inc: \$779,429

Provides payroll, human resources and benefits outsourcing.

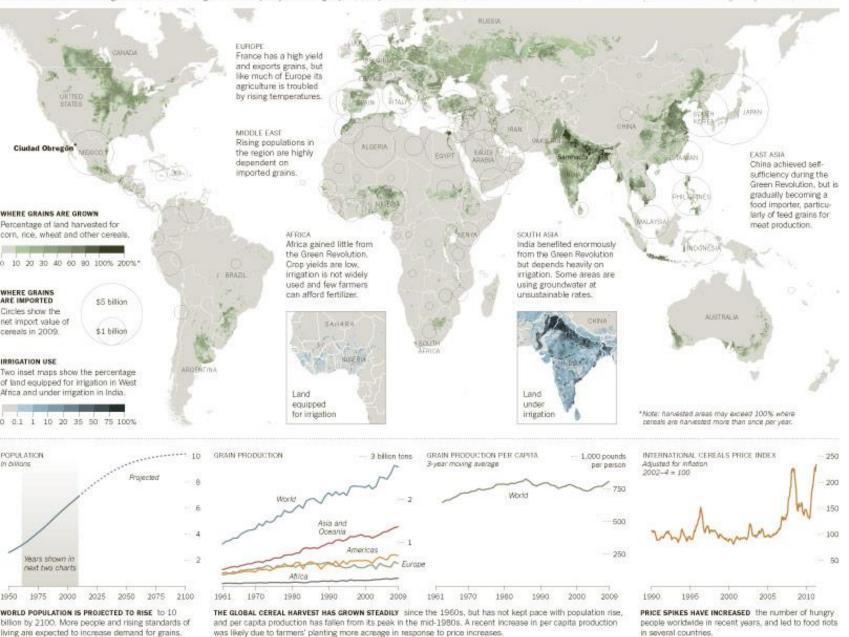
MAJORITY PAC	\$32,785
PRIORITIES USA ACTION	\$746,644

A tangled web

A Food System Under Strain

The New York Times

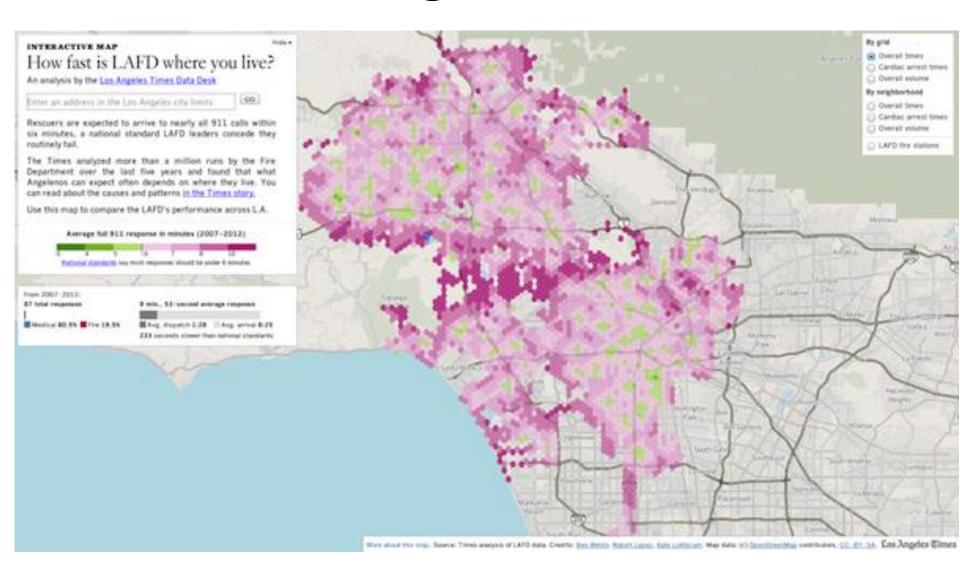
The Green Revolution that began in the 1940s raised global food output by introducing improved crop varieties and fertilizer to mi

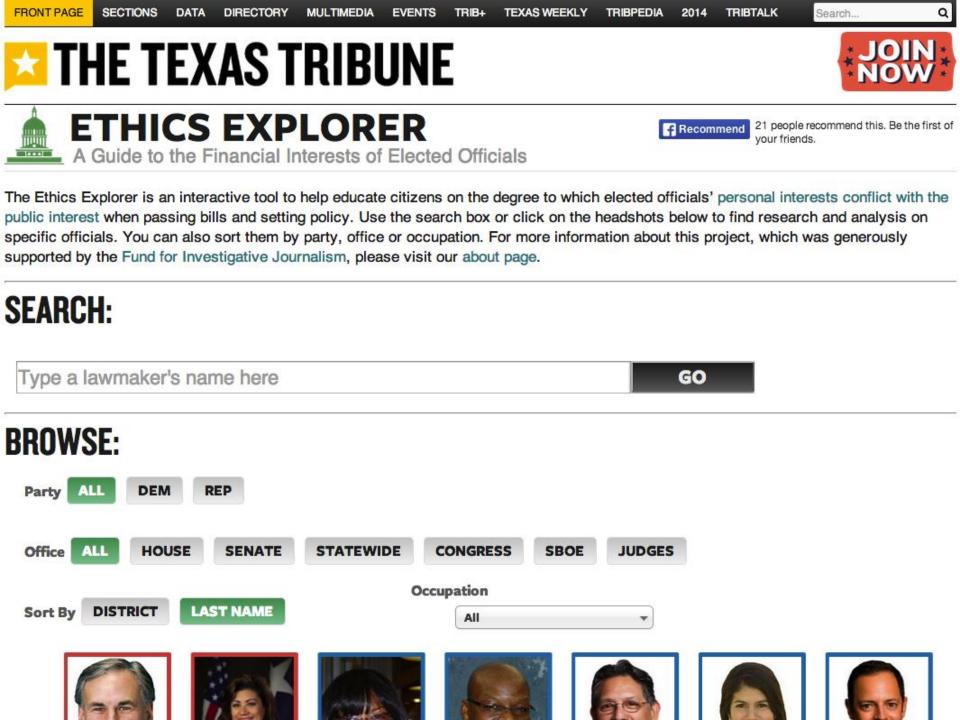


Sources: Chad Monthade, Arisona State Unite, Navin Rameniusts, McGill Unite, and Jonathan Poley, Unite, of Minnesota (harvested feed): International Trade Center that imports: Food and Apriculture Disparation (important, production, production, production, production)

in several countries.

Los Angeles Times



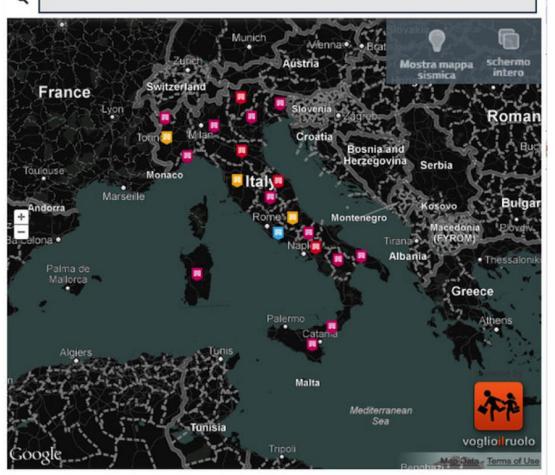


La tua scuola è sicura? Cercala sulla mappa

L'<u>inchiesta #scuolesicure</u> di Wired arriva a conclusione. Ecco le informazioni su rischio sismico, verifiche e relativi costi di tutti gli edifici scolastici italiani

09 novembre 2012 di Wired.it Staff

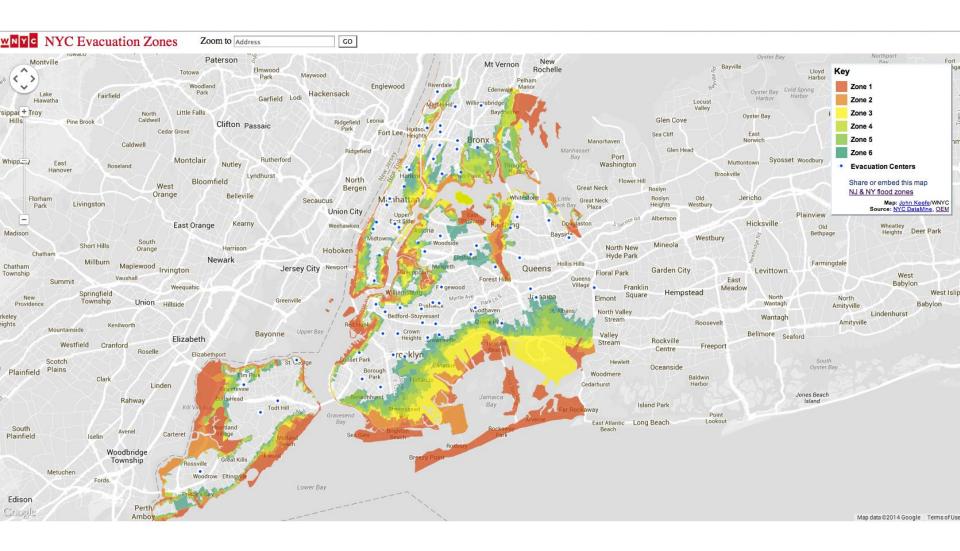
Q



La Nacion

LA NACIÓN





VozData







Últimas noticias

Edición impresa

Blogs

LN Data

Servicios

Guía LA NACION

Ianacion.com | VozData | Gastos Del Senado

Acerca | Preguntas frecuentes | Aviso legal | Contacto

Login

Gastos del Senado



¡Muchas gracias a todos por colaborar en el primer proyecto de Vozdata!

Ya completamos la primera etapa de 6657 documentos.

¡Lo logramos!

En breve, una vez chequeados, compartiremos la base en formato de datos abiertos.

Y el código.

¡Feliz 25 de mayo!

Liberá un documento
¡Ya revisamos 6657 de 6657 documentos!
Gastos del Senado procesados: \$ 54.186.544
Gastos del Senado procesados: \$ 54.186.544



Ranl	king de adjudicatarios	ver mas
1	Honorable Senado de la	\$ 29.739.287
2	Aerolíneas Argentinas S.A.	\$ 2.456.198
3	Derudder hnos. S.R.L.	\$ 1.480.058
4	Empresa General Urquiz	\$ 1.356.094
5	Nueva Chevallier S.A	\$ 1.213.723
6	Servicio de comedor - Ho	\$ 1.176.869
7	Vía Bariloche S.A.	\$ 1.044.005
8	CATA Internacional - Coo	\$ 852.234

Rank	king de rubros	ver mas
1	Pasajes aéreos, terrestre	\$ 16.142.305
2	Caja chica	\$ 10.995.448
3	Viáticos para agentes	\$ 8.410.171
4	Viáticos y gastos	\$ 4.765.822
5	Gastos de protocolo y cer	\$ 3.529.186
6	Otros N.E.P. (No especific	\$ 1.426.890
7	Canje de pasajes	\$ 818.758
8	Mantenimiento y Reparac	\$ 807.528

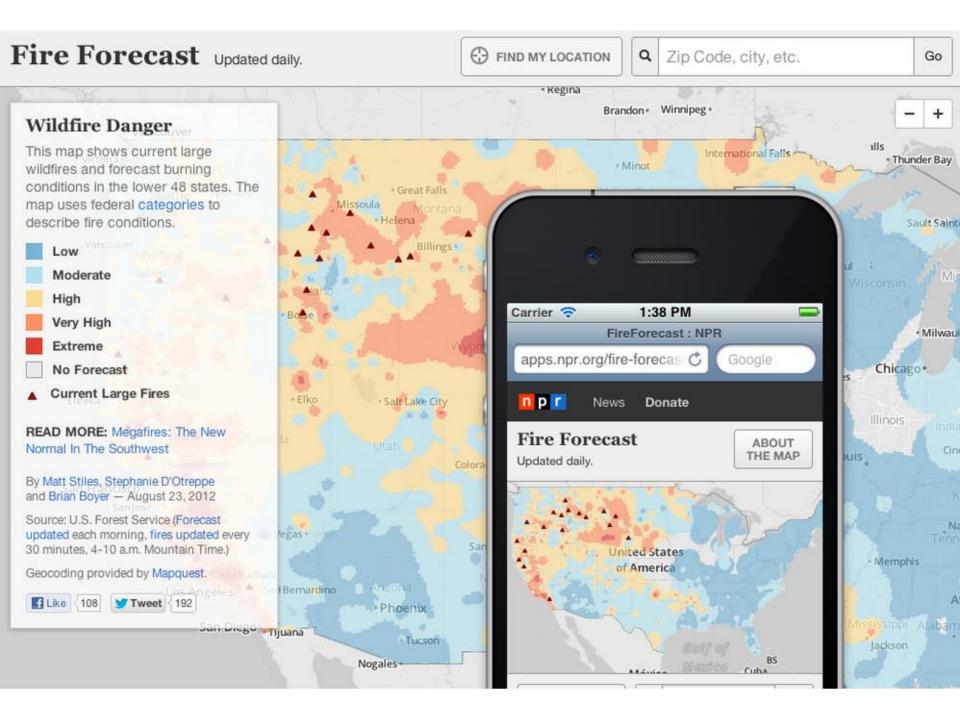
Ranking de usuarios Por cantidad de documentos revisados			
0	ArwenMawrn	1071	
Jeep IKA	Jeepika	961	
	chivacedo	695	
	jaelariadna	645	
×	alkymia	512	
	tessjackie	426	
TOP	Dativa	340	

Best practices?

Report it out

The US Embassy Dispatches Interactive Atlas Dossiers A-Z The map shows where the dispatches originated in the chosen time period. The Degree of classification (Color coded by degree) more dispatches, the larger the circle. The relative quantity of classified documents from a given location is indicated according to the color scale on the right. Click on a US Embassy location to see statistics for each representation, 2000 to 2010, by month 1966 to 2010, by year Total time period 2009 1966 1981 1986 2001 2010 1971 1976 1991 1996 2006

Show people something new about the world

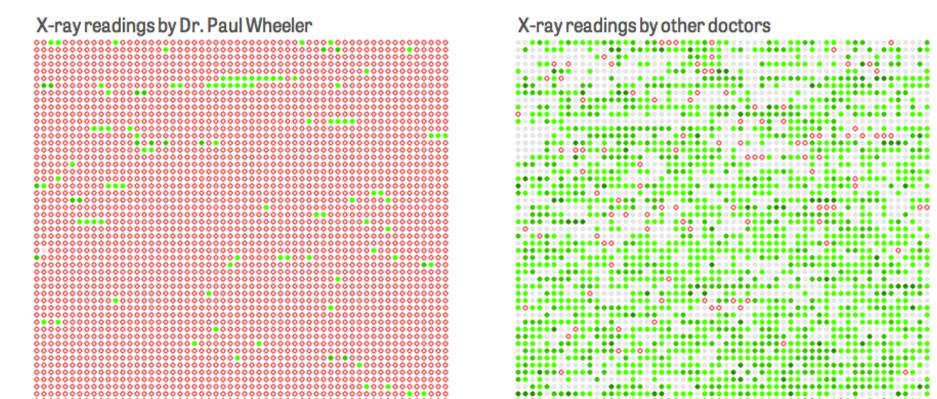


Tell a story

X-ray readings compared



For decades, Dr. Paul Wheeler has led a unit of radiologists at Johns Hopkins who often are enlisted by the coal industry to read X-rays in black lung benefits cases. The Center for Public Integrity identified more than 1,500 cases decided since 2000 in which Wheeler was involved, reading a total of more than 3,400 X-rays. In these cases, he never found a case of complicated black lung, and he read an X-ray as positive for the earlier stages of the disease in less than 4 percent of cases. Subtracting from these the cases in which he ultimately concluded another disease was more likely, this number drops to about 2 percent.



Storytelling still matters.

"We use these tools to find and tell stories.

We use them like we use a telephone.

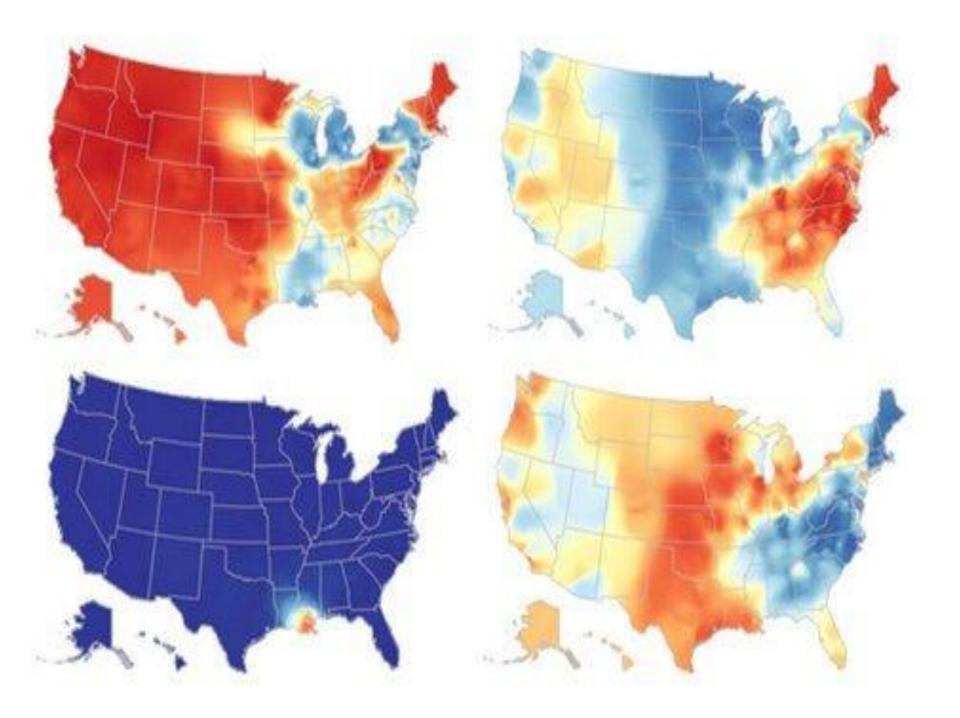
The story is still the thing."

Anthony DeBarros
 USA Today



Source: <u>Data Journalism and the Big Picture</u>

Make it personal



Understand the context for the data



Show your data



By Eric Sagara, Charles Ornstein, Tracy Weber, Ryann Grochowski Jones and Jeremy B. Merrill, ProPublica, Updated March 3, 2014

In recent years, drug companies have started releasing details of the payments they make to doctors and other health professionals for promotional talks, research and consulting. As of 2012, 15 companies published the information, most because of legal settlements. Use this tool to search for payments. Read more about the data »

Has Your Health Professional Received Drug Company Money?

Name	State		
	All States	+	SEARCH
Example searches: Klein, Duke University, Miami			More options +

With 2.1 million records, this database represents:



\$2.5 billion

in disclosed payments



10



~ 43%

of total market share

Payments in Your State

Click on a state to see payments made to practitioners and institutions there. See notes below.

State	Total
Alabama	\$32,352,175
Alaska	\$516,535
Arizona	\$38,739,643
Arkansas	\$15,599,178
California	\$298,077,233
Colorado	\$37,667,949
Connecticut	\$33,936,255
Delaware	\$4,263,545
District of Columbia	\$10,078,629
Florida	\$214,234,452
Georgia	\$72,421,234
Hawaii	\$3,879,724
Idaho	\$9,540,294
Illinois	\$82,167,799
Indiana	\$49,917,839

Company Disclosures

The totals listed here cover different time periods and spending categories, and aren't directly comparable. See notes below. See what each company discloses »

Company	Total Disclosed
AbbVie Disclosed: July to Dec., 2012	\$18M
Allergan Disclosed: July 2011 to Dec. 2012	Ranges*
AstraZeneca Disclosed: Jan. 2010 to Dec. 2012	\$275.3M
Cephalon Disclosed: Jan. 2009 to Dec. 2012	\$89.6M
Eli Lilly Disclosed: Jan. 2009 to Dec. 2012	\$608.8M
EMD Serono Disclosed: Jan. 2011 to Dec. 2012	\$4.4M

Our Stories



As Full Disclosure Nears, Doctors' Pay for Drug Talks Plummets

As transparency increases and blockbuster drugs lose patent protection, drug companies have dramatically scaled back payments to doctors for promotional talks. This fall, all drug and medical device companies will be required to report payments to doctors.

Freed of Disclosure Requirement, Drug Maker Pulls Doctor Payments Offline

Drugmaker Cephalon had been required to post its payments to doctors online as part of a lawsuit settlement. After its agreement expired, it removed them from its website.

Dollars for Docs Mints a Millionaire

New data show drugmakers' payments to hundreds of thousands of doctors, and some have made well over \$500,000.

Background Stories

Docs on Pharma Payroll Have Blemished Records, Limited Credentials

Med Schools Flunk at Keeping Faculty Off Pharma Speaking Circuit

Show your Work

Overview

Other Fore

Meet Leo, Our Senate Model

LEO was the <u>first computer</u> for ordinary businesses. We've named our Senate forecasting model after it. Our model, created by Amanda Cox and Josh Katz, combines polls with other information to predict how many Senate races Democrats and Republicans will win this year.

How does Leo interpret a poll?

LEO SAYS We focus on the margin between two major candidates, taking steps to make different polls directly comparable. We tweak polls that count registered voters instead of likely ones. We make further adjustments depending on who conducted the poll.

Share your code





Q Find a repository...

responsiveiframe

Responsive IFrames

Updated on May 13

WP-DS-NPR-API

Updated on May 9

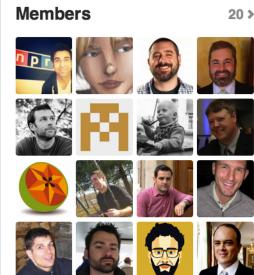
potluck

JavaScript ★ 10 🖟 4

PHP ★7 ₽5

JavaScript ★ 167 \$2 60

A social content aggregator prototype, built at NPR Digital Media during an NPR Serendipity hackathon with Node.js, Express.js, Node Bootstrap, Twitter Bootstrap and lots of love for Open Source Software + Journalism



Consider ethics

Questions

- Is the data clean?
- Is the data representative?
- What biases might be hidden in the data?
- Was the data legally obtained?
- Does the data contain personally identifiable information (PII)?

Collection

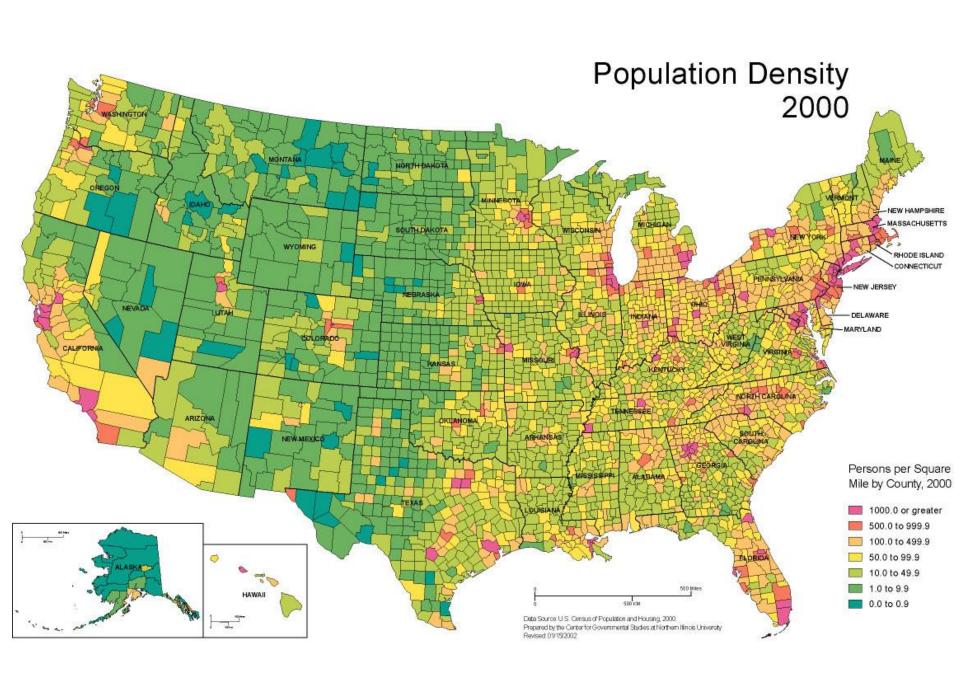
- Who gathered the data? How?
- Was it clear how data would be used?
- Can people opt-out of collection or usage?
- "Notice and consent" is not enough
- "Privacy by design" applies to news apps



FIRST DO NO HARM

Data Analysis & Numeracy

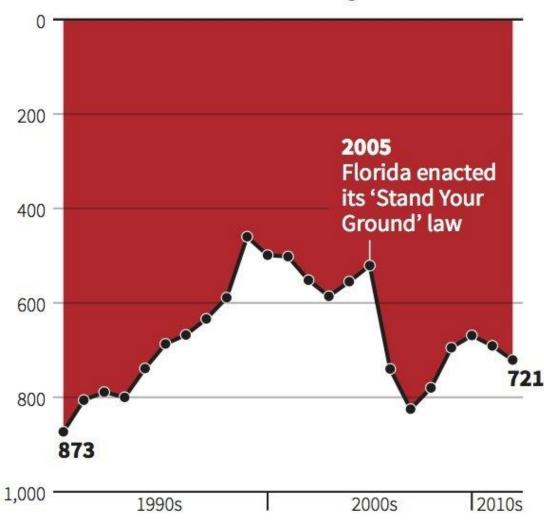
- N = ?
- Average vs Median
- Statistical significance?
- Correlation != causation
- Regression to the mean



Gun deaths in Florida

Presentation

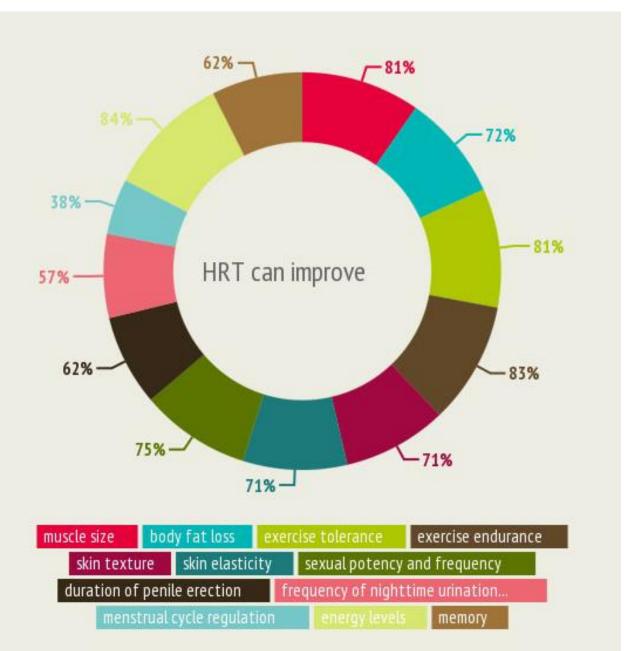
Number of murders committed using firearms



Source: Florida Department of Law Enforcement

Bad Data Viz

wtfviz.net



Present data with context, in context

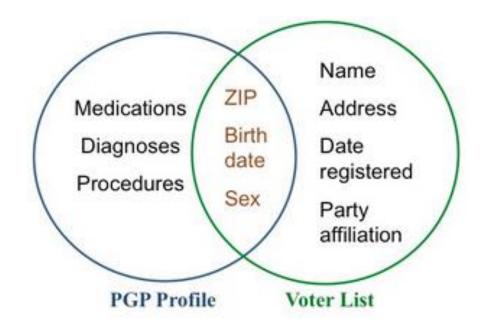


Be aware of de-anonymization risks



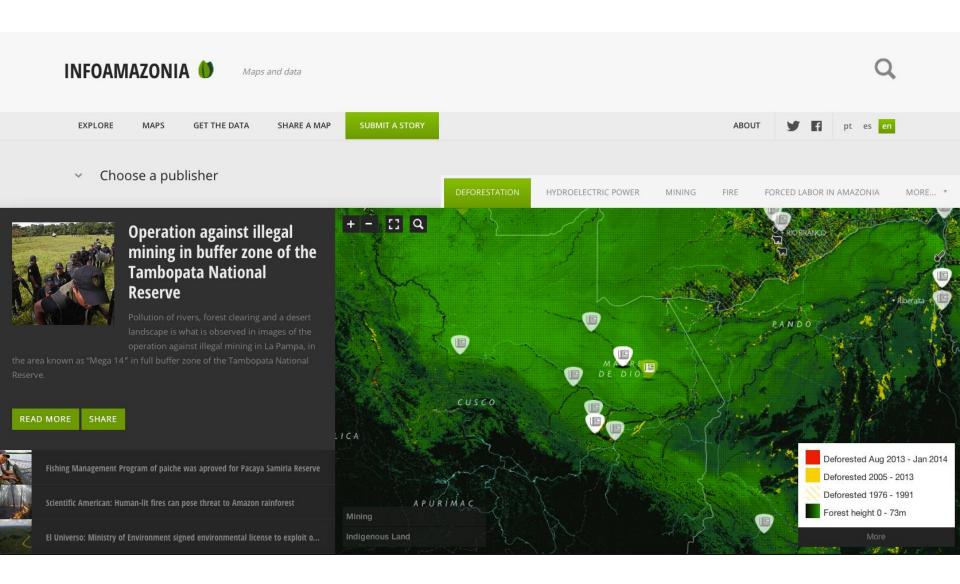
PGP Participants

Test your demographics



Emerging trends

geojournalism



Networked reporting of corruption

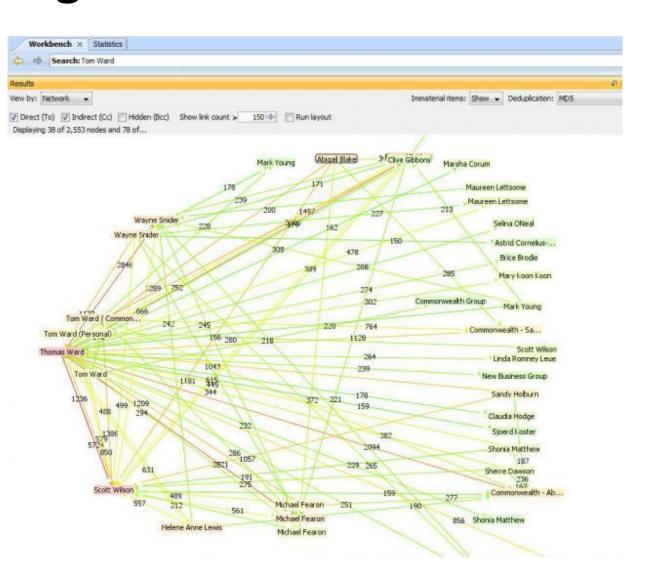


ICIJ: Offshore Leaks

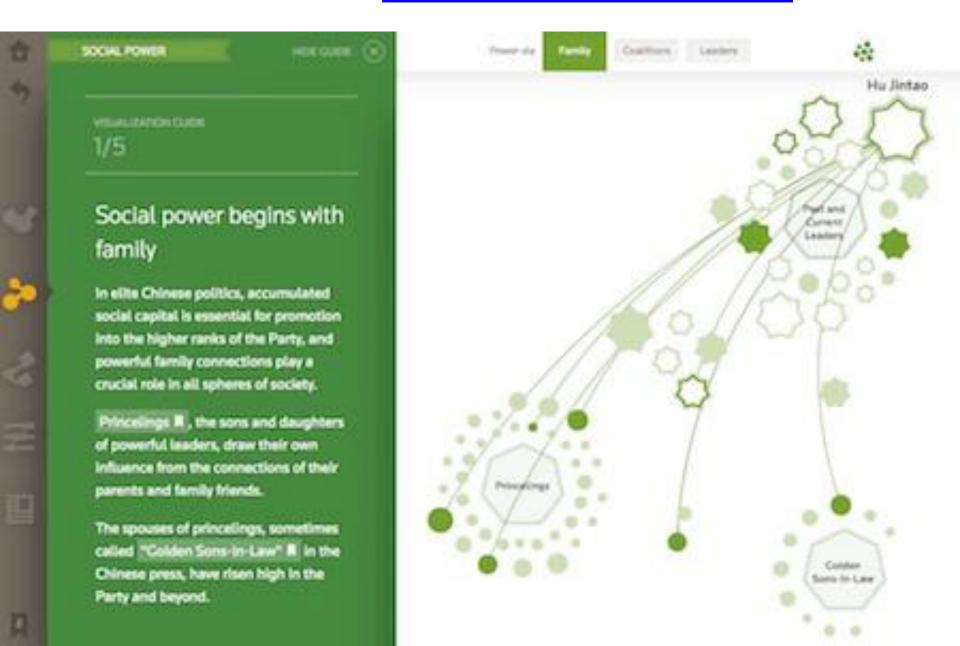
International Consortium of Investigative Journalists

Offshoring \$

80 journalists40 countries260 gigabytes2.5 million files



Reuters: Connected China



Create your data

"If Stage 1 of data journalism was "find and scrape data," then...

Stage 2 was "ask government agencies to release data" in easy to use formats.

Stage 3 is going to be "make your own data", and those sources of data are going to be automated and updated in real-time."

-Javaun Moradi, Mozilla

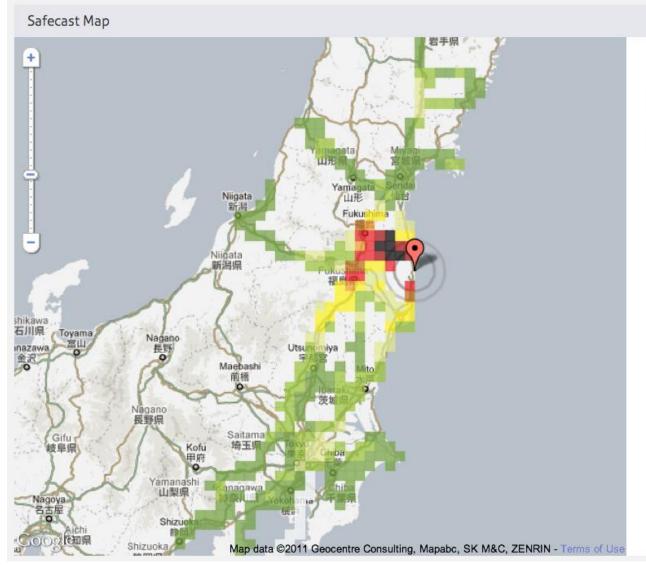
Safecast

open source

Geiger counter



Networked accountability



The Safecast Map depicts over 500,000 radiation data points collected by the Safecast team throughout Japan. For each square, numerous geiger readings have been collected and color-coded. For detailed maps showing all data points, please visit the Drive Maps page.

Interact with full data set

Individual drive maps

Want to help us collect more data?

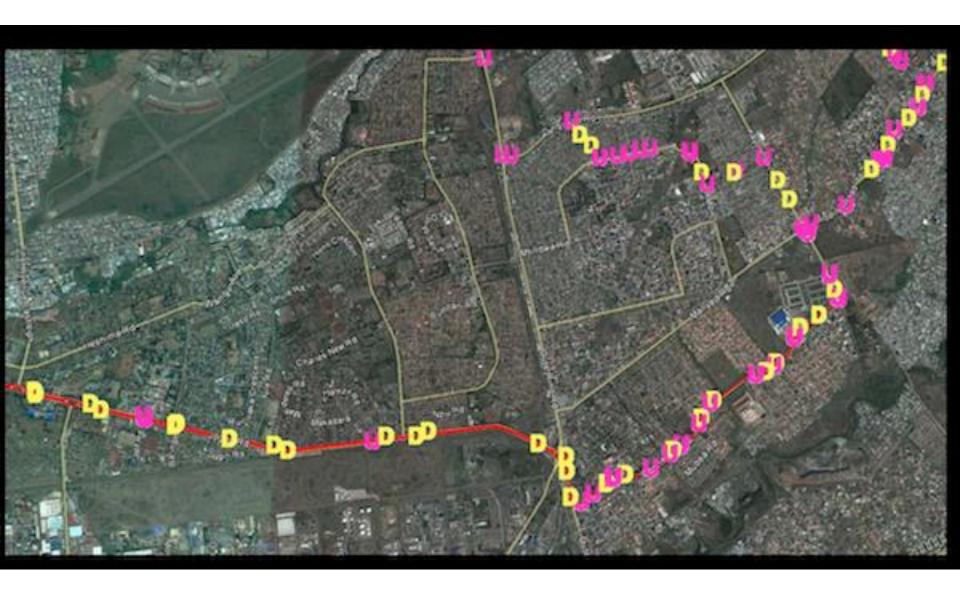








Bus route in Nairobi, Kenya



Sensor Journalism

SENSING AIR QUALITY AT THE OLYMPICS COUNTRYWIDE POLLUTION THE GOVERNMENT SOLUTION BACK TO MENU Location of testing (Bird's Nest) Particulate matter measurements in Beijing Click on photo to see data from selected date Particulate matter, or PM10, are inhalable coarse particles. with diameters smaller than 10 micrometers. The source of PM10 can be natural or man-made. Man-made causes include the burning of fossil fuels and construction projects. High concentration levels of PM10 in the air are a cause for concern because these inhalable particles can settle in the bronchi and lungs and cause health problems. 8 / 17 / 2008 **GOVERNMENTAL DATA** 700 micrograms per cubic meter 8 / 16 / 2008 400 8 / 15 / 2008 JULY AUGUST VHO auc quality target -----8 / 14 / 2008

photos taken at the time of sampling











Cicada Tracker

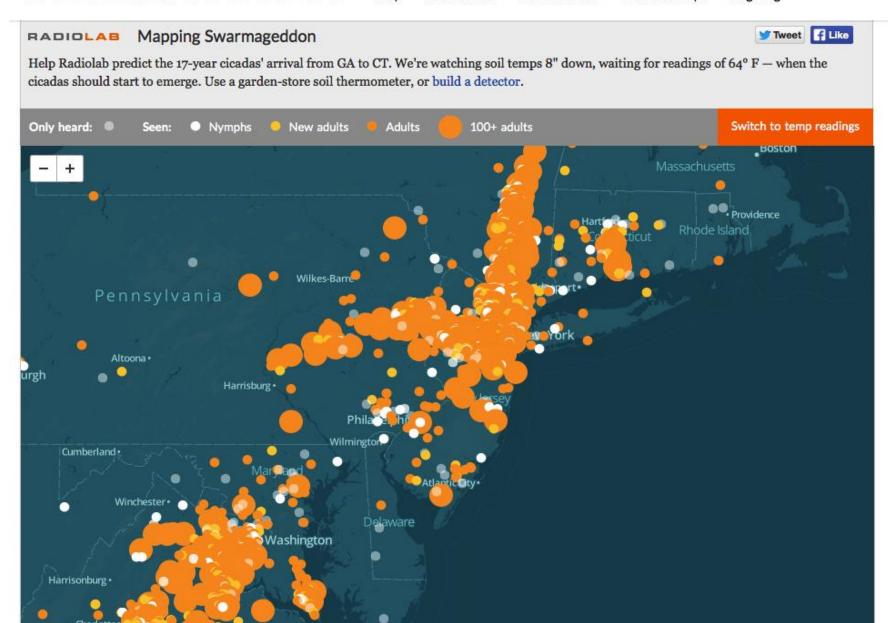
Man

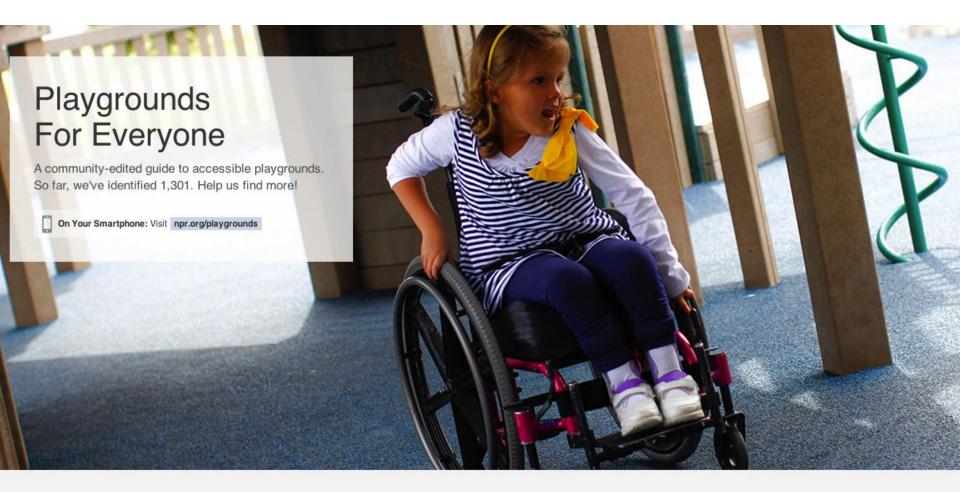
See Cicadas?

Build a Sensor

Decode Temps

Bug Blog





Find An Accessible Playground

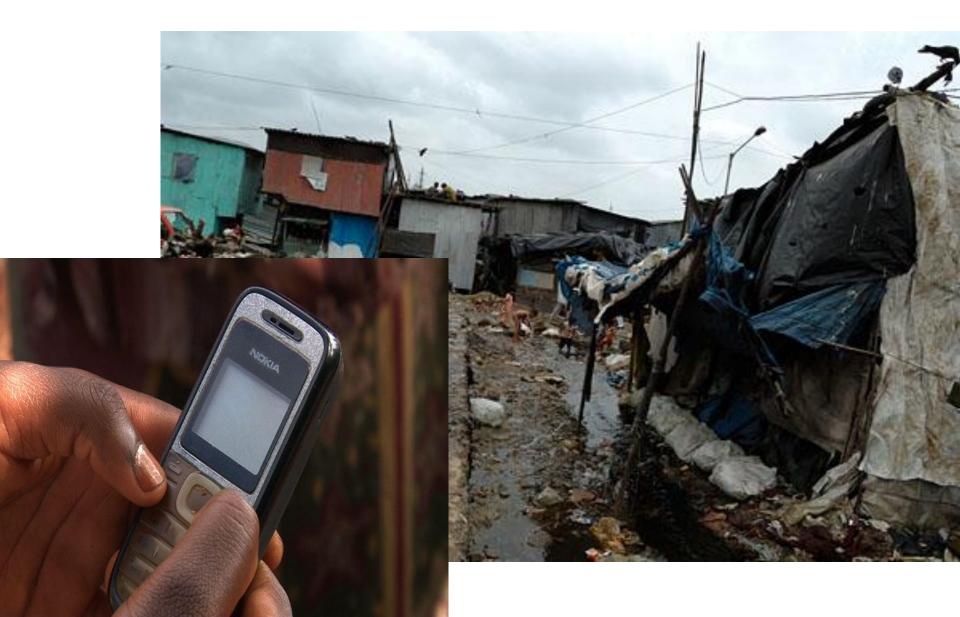
Address, City, State, or Zip Code

Q

OR

Find Playgrounds Nearby

Citizens as Sensors: Andhra Pradesh



Drones + data collection





Privacy challenges

Wikileaks Iraq war logs: every death mapped

The Wikileaks Iraq war logs provide us with a unique picture of every death in Iraq. These are those events mapped using Google Fusion tables

Download the data from the Datablog



Simon Rogers

guardian.co.uk, Saturday 23 October 2010 00.05 BST

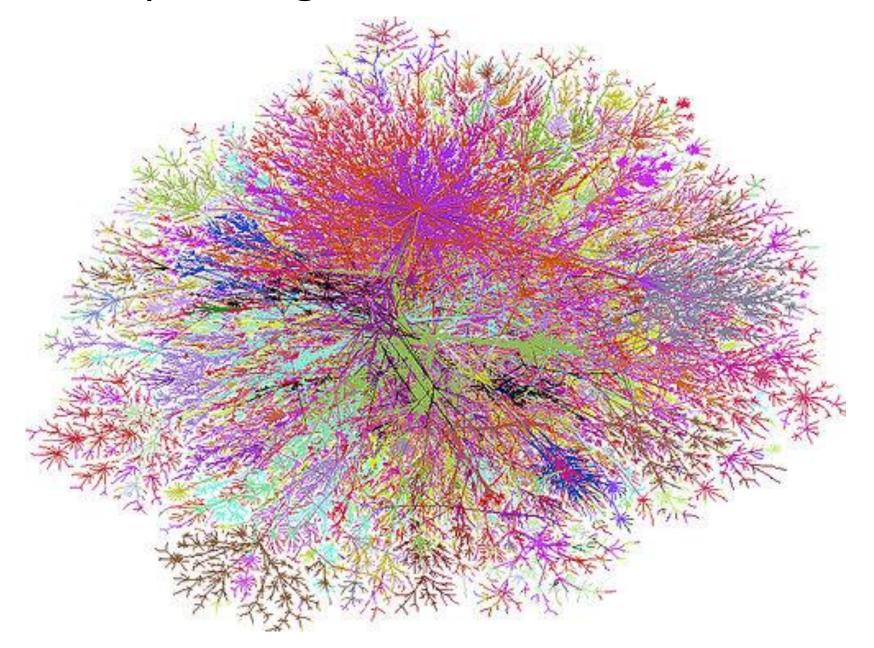






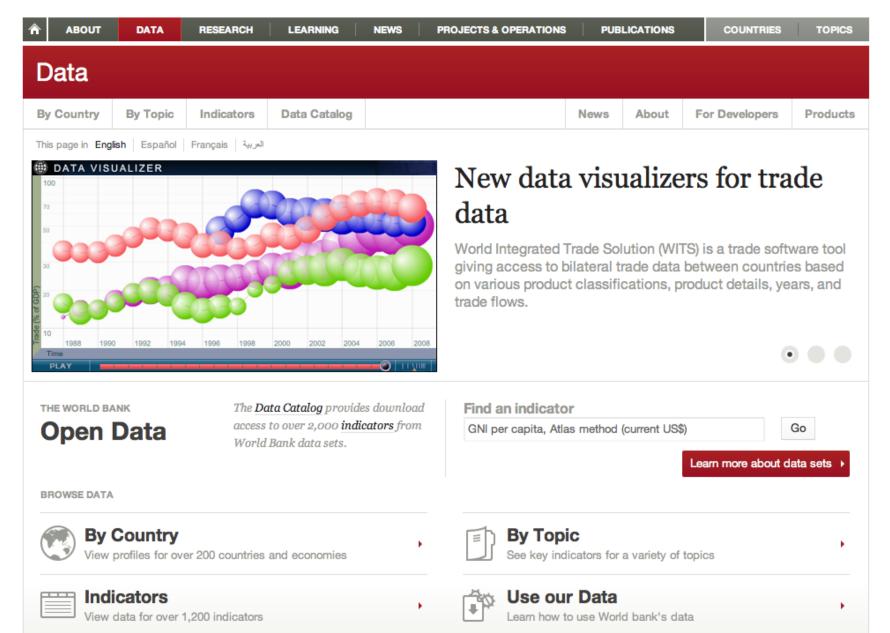
Open Data, FOIA & Press Freedom

An expanding number of data sources

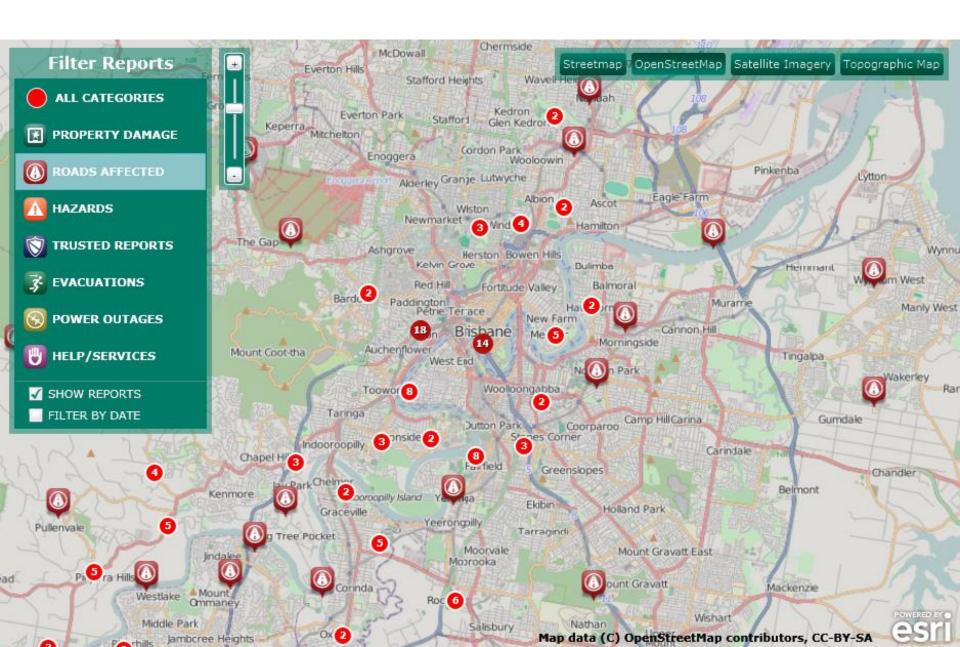








Social data and crisis data



Open government data platforms



Go

Enter your postcode to find tools and information about your local area using government data







Who ministers are meeting Business plans





FREEDOM OF THE PRESS WORLDWIDE IN 2014

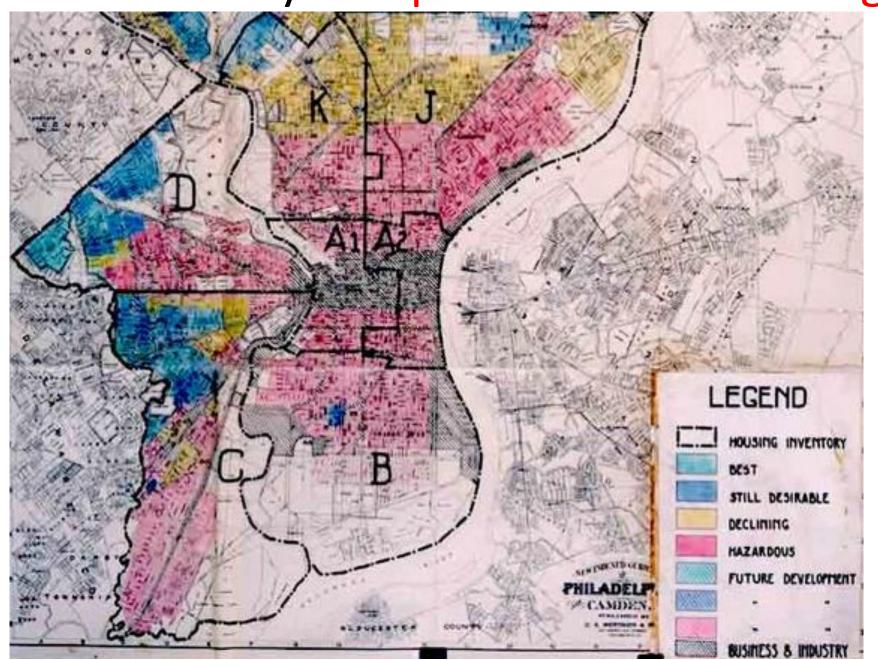
X-ray readings compared



For decades, Dr. Paul Wheeler has led a unit of radiologists at Johns Hopkins who often are enlisted by the coal industry to read X-rays in black lung benefits cases. The Center for Public Integrity identified more than 1,500 cases decided since 2000 in which Wheeler was involved, reading a total of more than 3,400 X-rays. In these cases, he never found a case of complicated black lung, and he read an X-ray as positive for the earlier stages of the disease in less than 4 percent of cases. Subtracting from these the cases in which he ultimately concluded another disease was more likely, this number drops to about 2 percent.

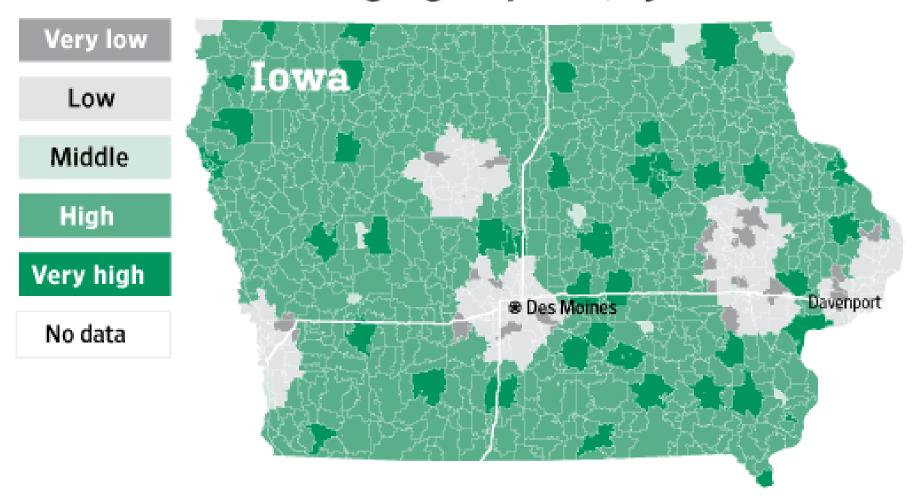


Accountability for "personalized redlining"



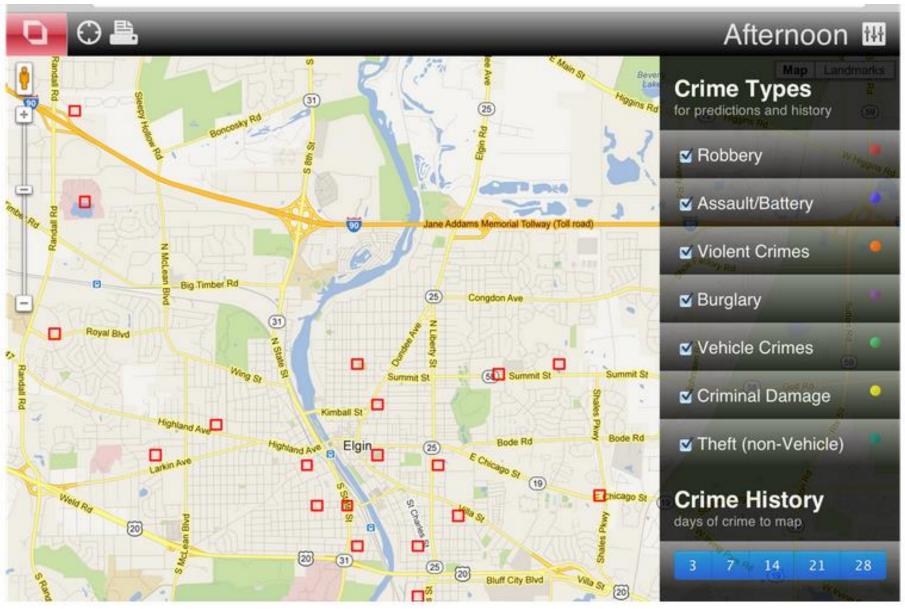
Transparency for geographic profiling

Likelihood of receiving higher prices, by ZIP code



WSJ: Websites vary prices, based upon user information

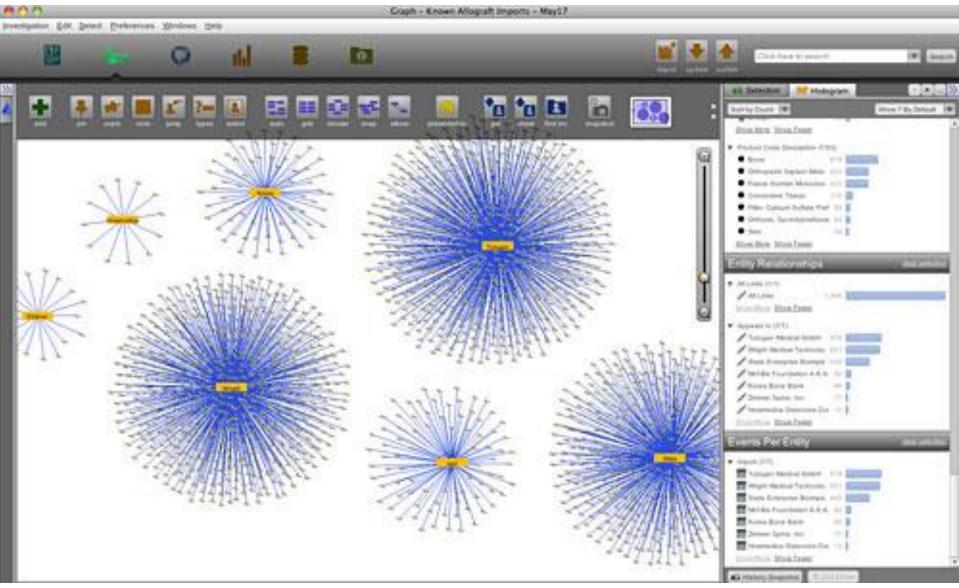
Monitoring predictive policing



Verge: Chicago crime and profiling

Geekwire: Predictive Policing

Investigating human tissue trafficking



ICIJ: The data behind skin and bone

Data + journalism + activism + responsive institutions = social change



The fun part: predictions, prognostications and recommendations!

1) Data will become even more of a strategic resource for media.



The ProPublica Data Store

institutions that have been disclosed by 15 pharmaceutical companies

ProPublica is making available the datasets that power our data journalism. The raw data we received as the result of a FOIA request is available for free, and datasets that reflect substantial cleaning and processing by our staff are available for a one-time fee. Journalists and academic researchers can purchase premium datasets, and interested commercial users can contact us for pricing, by clicking the "Purchase" button on any dataset. We also provide a pass-through link when a data download is available on another site. Related Story »

Premium Datasets (Purchase)

Cleaned up, categorized and often created from multiple sources, these Premium datasets are unique to ProPublica and sold for a nominal fee.

FOIA Data (Free)

Because ProPublica received these datasets from a FOIA request, we're posting the original, raw datasets free for download.

External Data

ProPublica frequently uses datasets that are free and available online. So instead of downloading copies from us, we send you straight to the source.

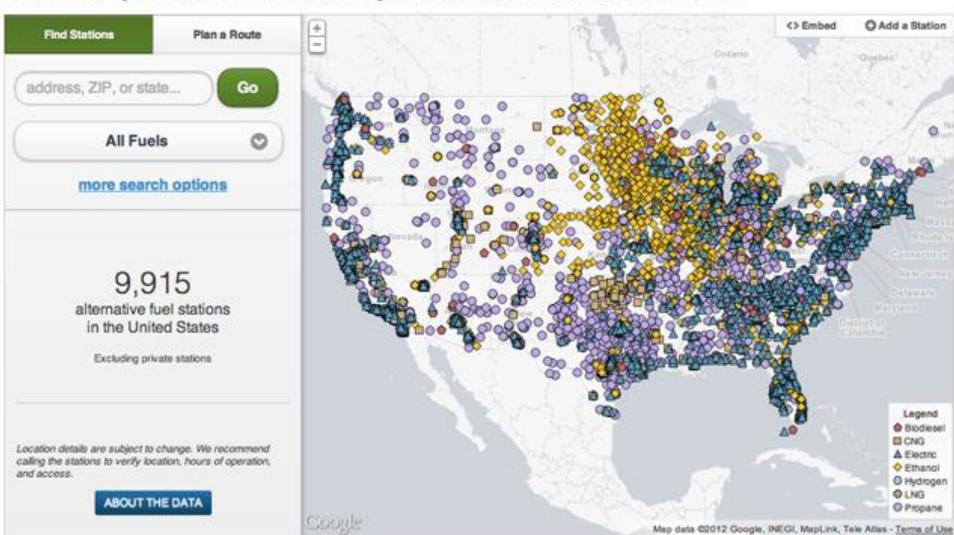
Try a Sample!

Health Datasets	Source	JOURN (\$)	ACAD (S)	
Premium: Dollars For Doctors Data (Per State)				
ProPublica's Dollars for Docs data. The data include about \$2 billion in payments to doctors, other medical providers and health care institutions that have been disclosed by 15 pharmaceutical companies from 2009 to 2011.	Pharmaceutical Company Disclosures	\$200	\$2,000	PURCHASE Try a Sample!
Size: Varies rows, Date Released: 03/2013				
Premium: Combined Dollars for Docs Dataset (National)				
ProPublica's Dollars for Docs data. The data include more than \$2 billion in payments to doctors, other medical providers and health care	Pharmaceutical Company Disclosures	\$1,000	\$10,000	PURCHASE

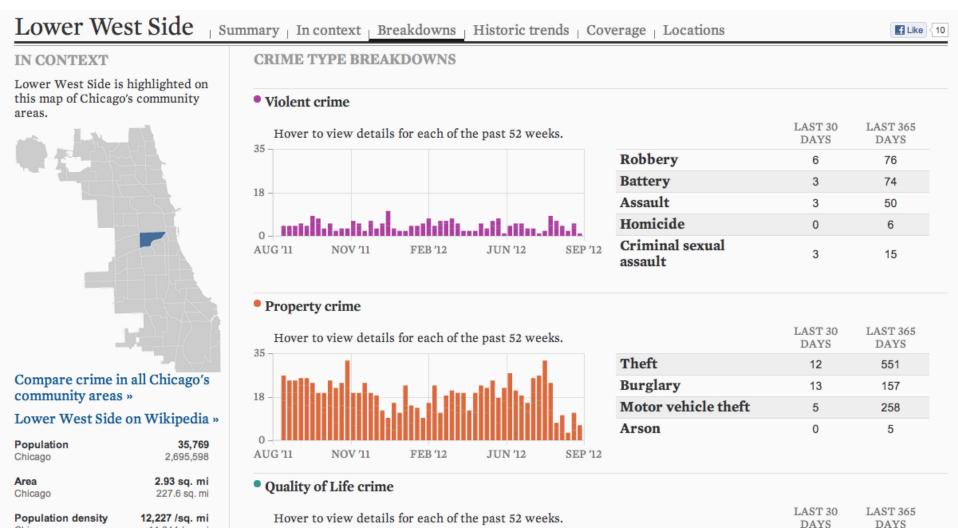
2) Better tools will emerge that democratize data skills.

Alternative Fueling Station Locator

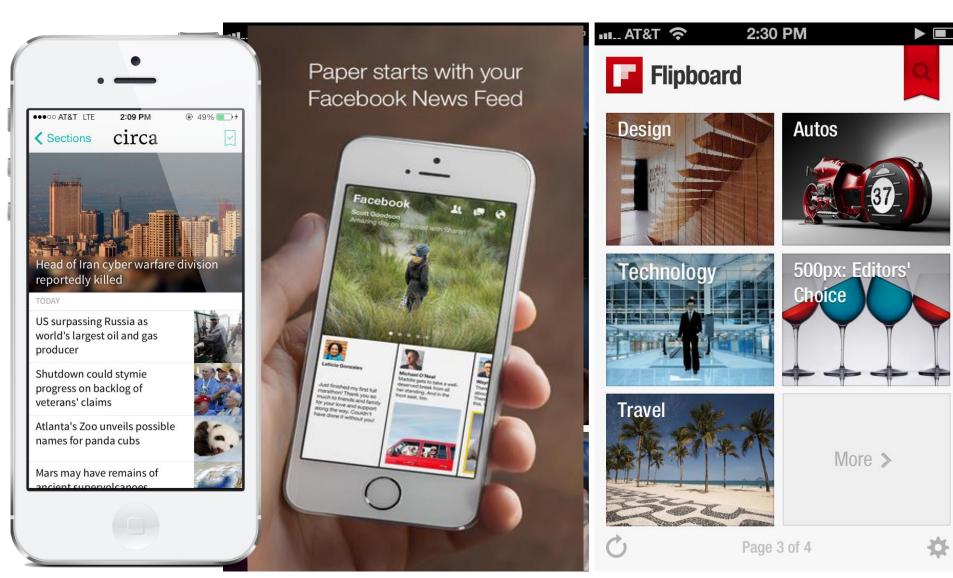
Find alternative fueling stations near an address or ZIP code or along a route in the United States. Enter a state to see a station count.



3) News apps will explode as a primary way people consume data journalism.



4) Being digital first means being datacentric and mobile-friendly.



5. Expect more robo-journalism. Human relationships and storytelling still matter.

The First News Report on the L.A. Earthquake Was Written by a Robot

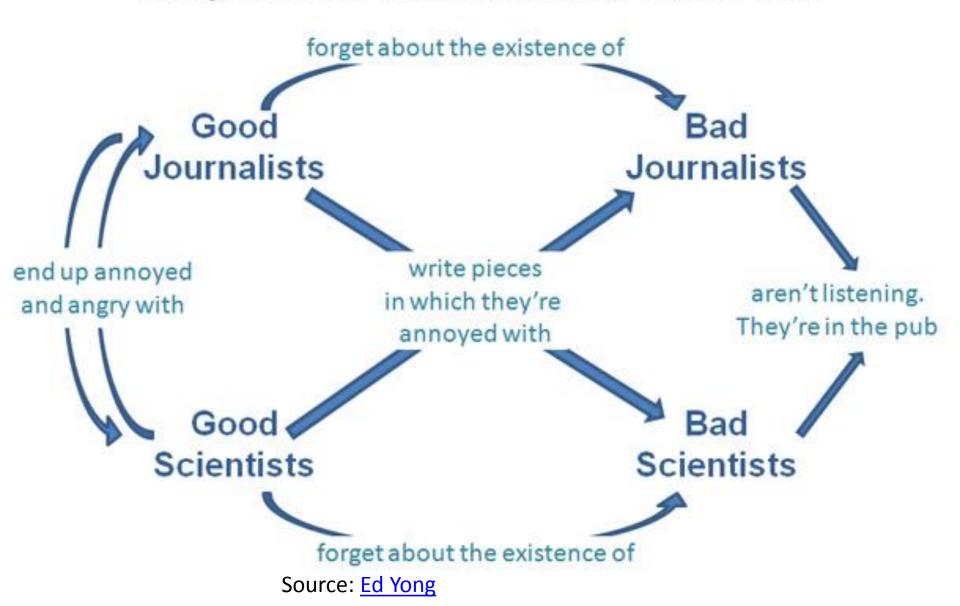
By Will Oremus



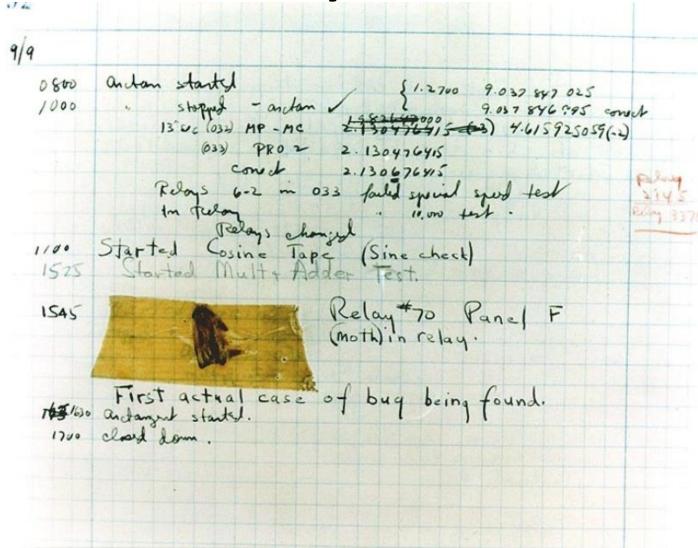


6) More journalists will need to study the

Every Scientists-versus-Journalists Debate Ever



7) There will be higher standards for accuracy and corrections.

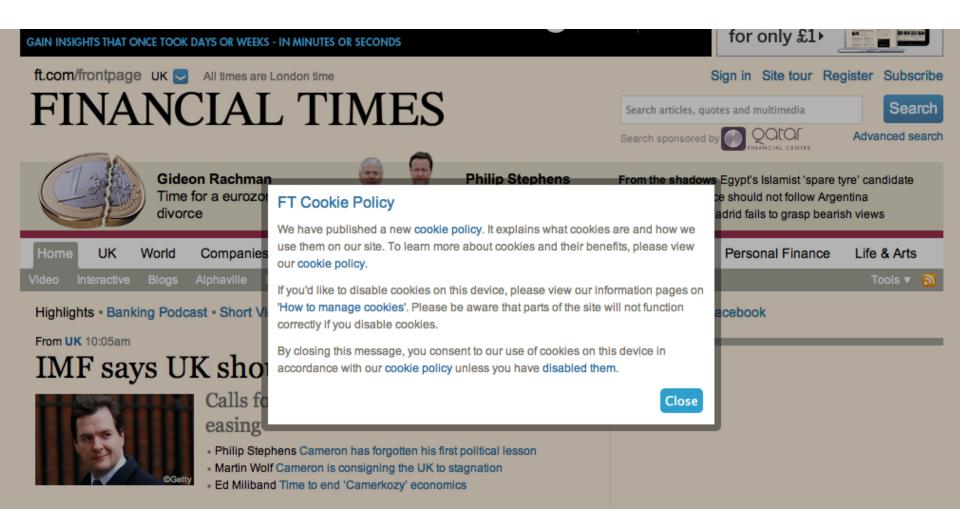


Source: <u>Jake Harris</u>

8) Competency in security and data protection will become more important.

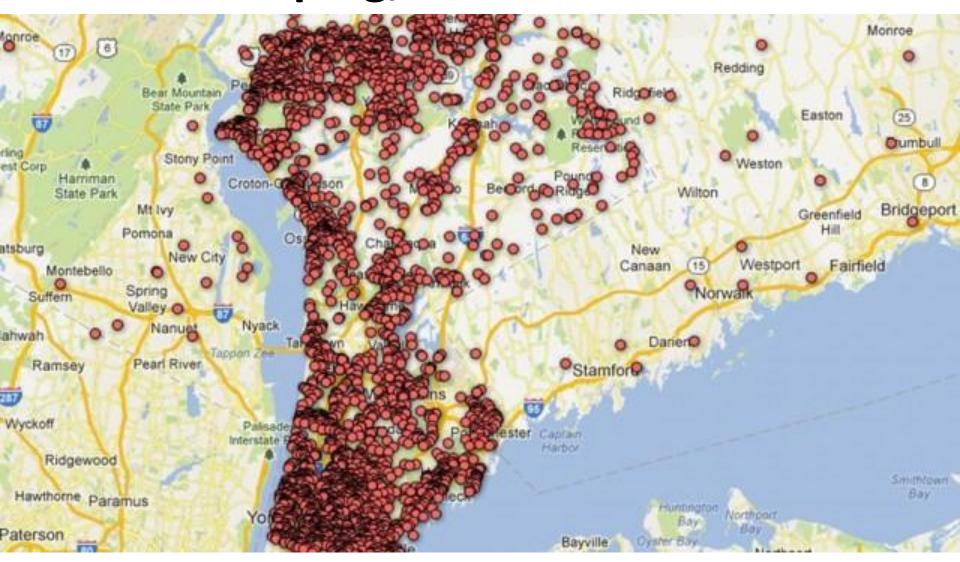


9) Demand for more transparency on reader data collection and use.



Source: <u>eConsultancy</u>

10) More conflicts over public records, data scraping, and ethics will arise.



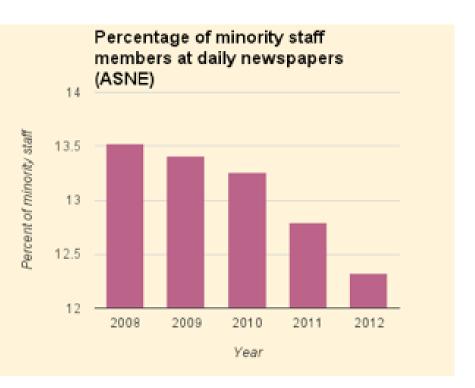
12) Data-driven personalization and predictive news in wearables.

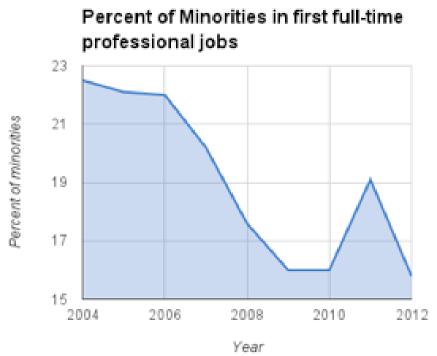
U.S. HOGSELINY

Secretary of the Treasury



13) More diverse newsrooms will produce better (data) journalism.

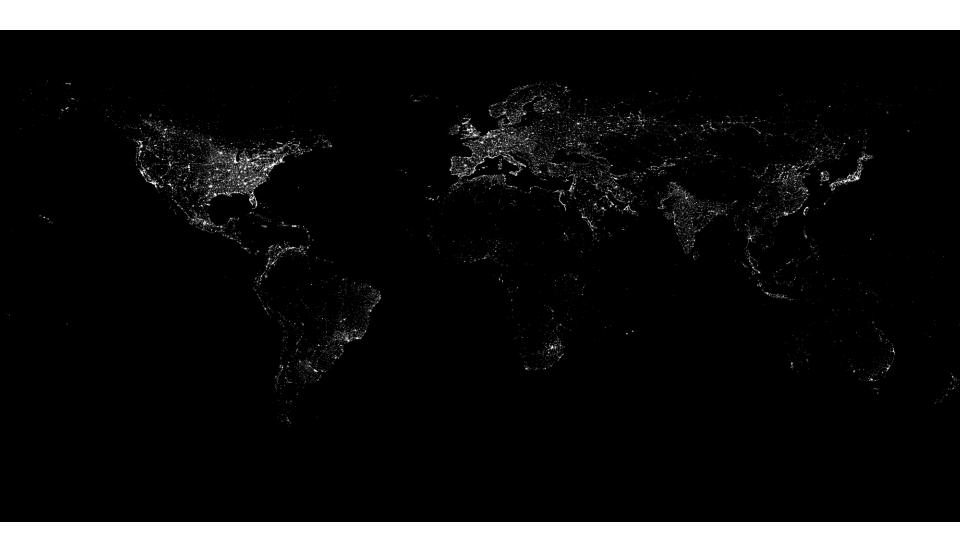




A 2013 ASNE survey of 68 online news organizations found that 63% of them had no minorities.

SOURCE: The Atlantic

14) Be mindful of data-ism and bad data. Embrace skepticism.



Homework

- Take a look at ProPublica's "Data Store" many sets are free!
 - https://projects.propublica.org/data-store/

Additional Resources

- Alex's blog: E Pluribus Unum
 - http://e-pluribusunum.org/
- Alex Howard: "Quantifying Journalism: Metrics, Data, and Computation" (video)
 - https://www.youtube.com/watch?v=JMDMClv0-So
- Stanford's "Journalism in the Age of Data" (video)
 - http://datajournalism.stanford.edu/
- World Economic Forum: "How data journalism can improve the world"
 - https://agenda.weforum.org/2015/01/how-data-journalism-can-improvethe-world/
- International Journalism Festival is happening 4/15-19 in Perugia, Italy
 - Check out program and speakers here: http://www.journalismfestival.com/