

Bringing (Web) Databases to the Masses

Alon Halevy

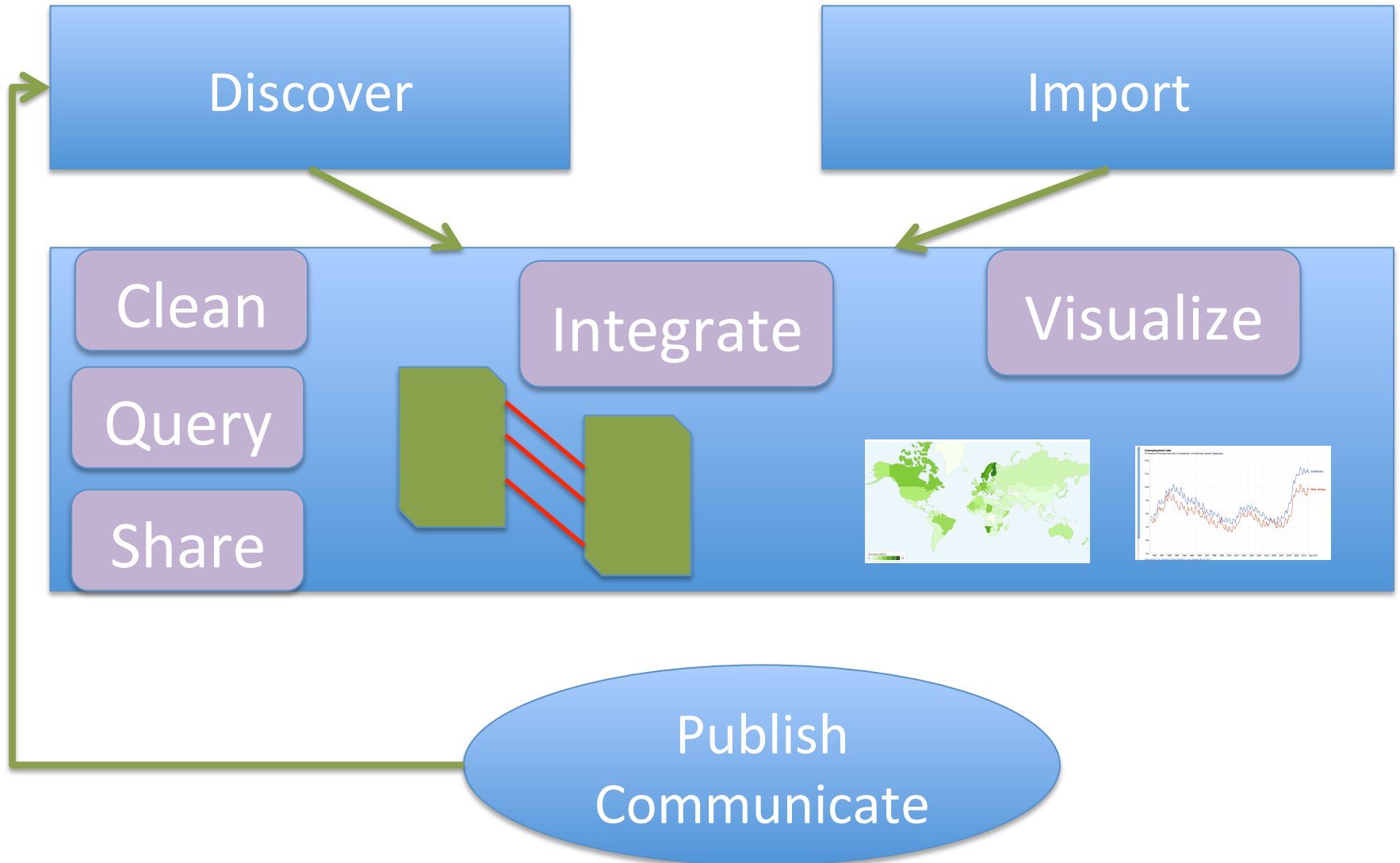
Google

January 11, 2012

Structured Data and the Web

- A huge amount of structured data on the Web
 - Reference data, hobbies, products, coffee, ...
 - *Reacting to existing data*
- A platform for getting more data out:
 - Government data, crime, water conditions, ...
 - *Being proactive*
- New kinds of data collection and management
 - Collaboration, crowd-sourcing, real-time data, crisis response, ...
 - *Inventing a bright future*

Goal: Structured Data Ecosystem



Challenges

- Finding high-quality structured data
- Getting data out of silos
- Extracting data from web pages
- Databases are hard to use
- Integrating heterogeneous data
- Publishing data is cumbersome

Outline

- Google Fusion Tables:
 - A Database management service for the Web
- WebTables:
 - Discovering a (structured) needle in an (unstructured) haystack

Fusion Tables

google.com/fusiontables

- Goal: an easy-to-use database system that is integrated with the Web.
- Key features:
 - **Easy** upload (CSV, KML, spreadsheets)
 - **Sharing** (even outside your company)
 - **Visualizations** front and center
 - Easy **publishing**
- Goal 2: a data cloud -- discover others' data and combine with yours.

Coffee Consumption Per Capita

International Coffee
Organization































Get link

Share

File View Edit Visualize Merge

Showing all rows [options](#)

1 - 100 of 182 [Next »](#)

Country ▾	Coffee Consumption			
Finland	12			
Solomon Islands	11.8			
Norway	9.9			
Namibia	9			
Denmark	8.7			
Netherlands	8.4			
Sweden	8.2			
Switzerland	 	7.9		
Bahrain	6.8			
Belgium	6.8			
Luxembourg	6.8			
Brunei	6.6			
Canada	6.5			
Germany	6.4			
Chad	6.4			

Rwanda 2010 Cup of Excellence Winners Cup of Excellence

merge

place ▾ latitude ▾

1 -1.891708

2 -1.819305

3 -1.766459

4 -2.226917

5 -2.011286

6 -1.996361

7 -1.776754

8 -2.209764

9 -1.813814

11 -1.721505

12 -2.743473

13 -1.700571

14 -2.562393

15 -1.727338

16 -1.75136

17 -2.551932

18 -2.059664

19 -2.521578

Share this table ✕

Invite people ?

☒ **As viewers**

Can see and comment on the data

☐ **As collaborators**

Can also edit the data

☐ **As owners**

Can also invite people to view or collaborate

Separate email addresses with commas.

☒ **Send email invitations**

Visibility options ?

☒ **Public**

Everyone on the internet can find and access. No sign-in required.

☐ **Unlisted**

Anyone who has the link can access. No sign-in required.

☐ **Private**

Only people explicitly granted permission can access. Sign-in required.

Viewers (0)

Collaborators (0)

Owners (1)

halevy@google.com

Coffee Consumption Per Capita

International Coffee Organization

Get link Share

File View Edit Visualize Merge

Showing all rows

1 - 100 of 182 [Next »](#)

Country ▾




































Finland
Solomon Islands
Norway
Namibia
Denmark
Netherlands

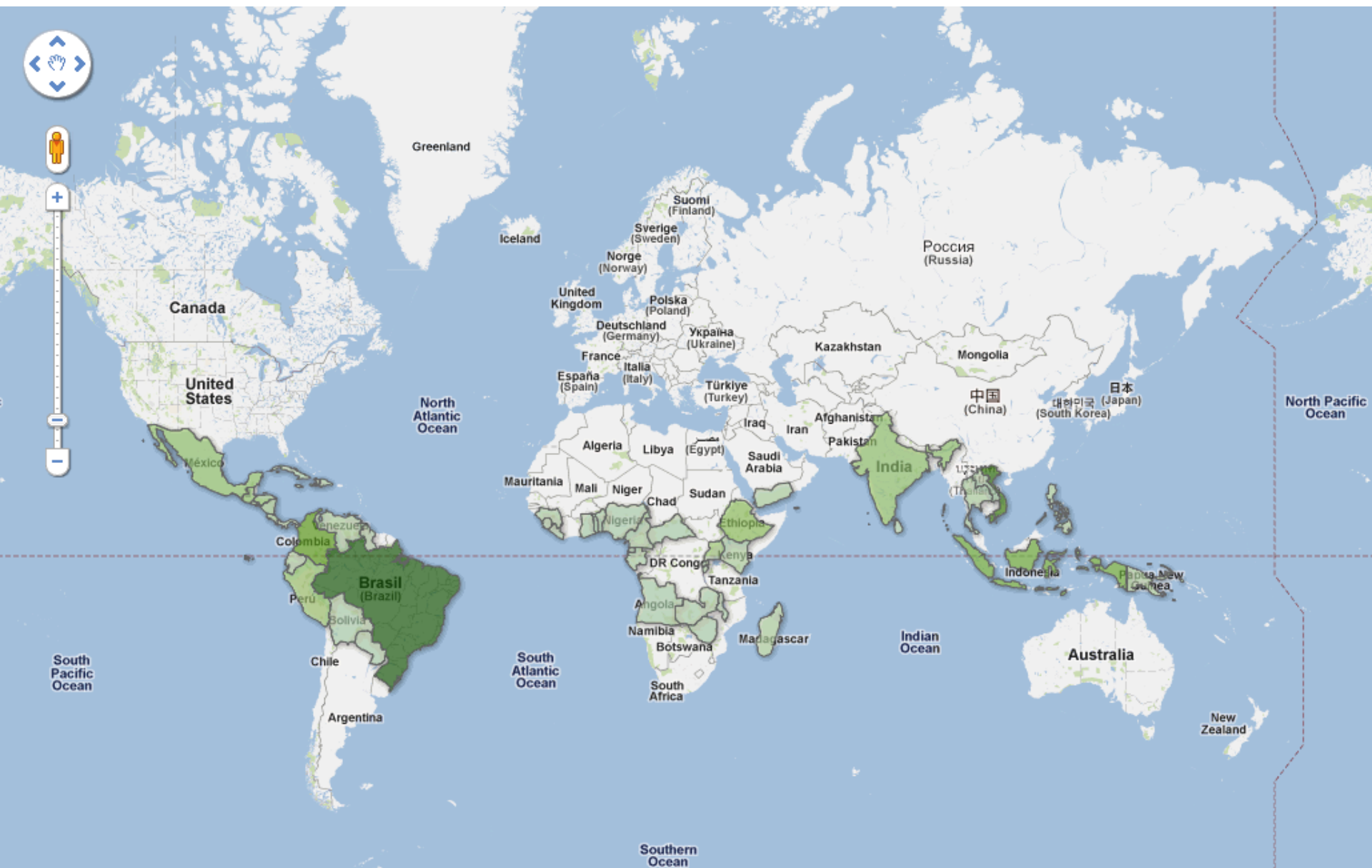
- Table
- Map
- Intensity map
- Line
- Bar
- Pie
- Scatter
- Timeline (date, text, number)

Coffee Consumptiic

Order ▾

Finland	12	1		
Solomon Islands	11.8	2		
Norway	9.9	3		
Namibia	9	4		
Denmark	8.7	5		
Netherlands	8.4	6		
Sweden	8.2	7		
Switzerland	7.9	8		
Bahrain	6.8	9		
Belgium	6.8	10		
Luxembourg	6.8	11		
Brunei	6.6	12		
Canada	6.5	13		
Germany	6.4	14		
Chad	6.1	15		
Austria	6.1	16		
Samoa	6	17		
Italy	5.9	18		



DATA BLOG

Facts are sacred

Wikileaks Iraq war logs: every death mapped

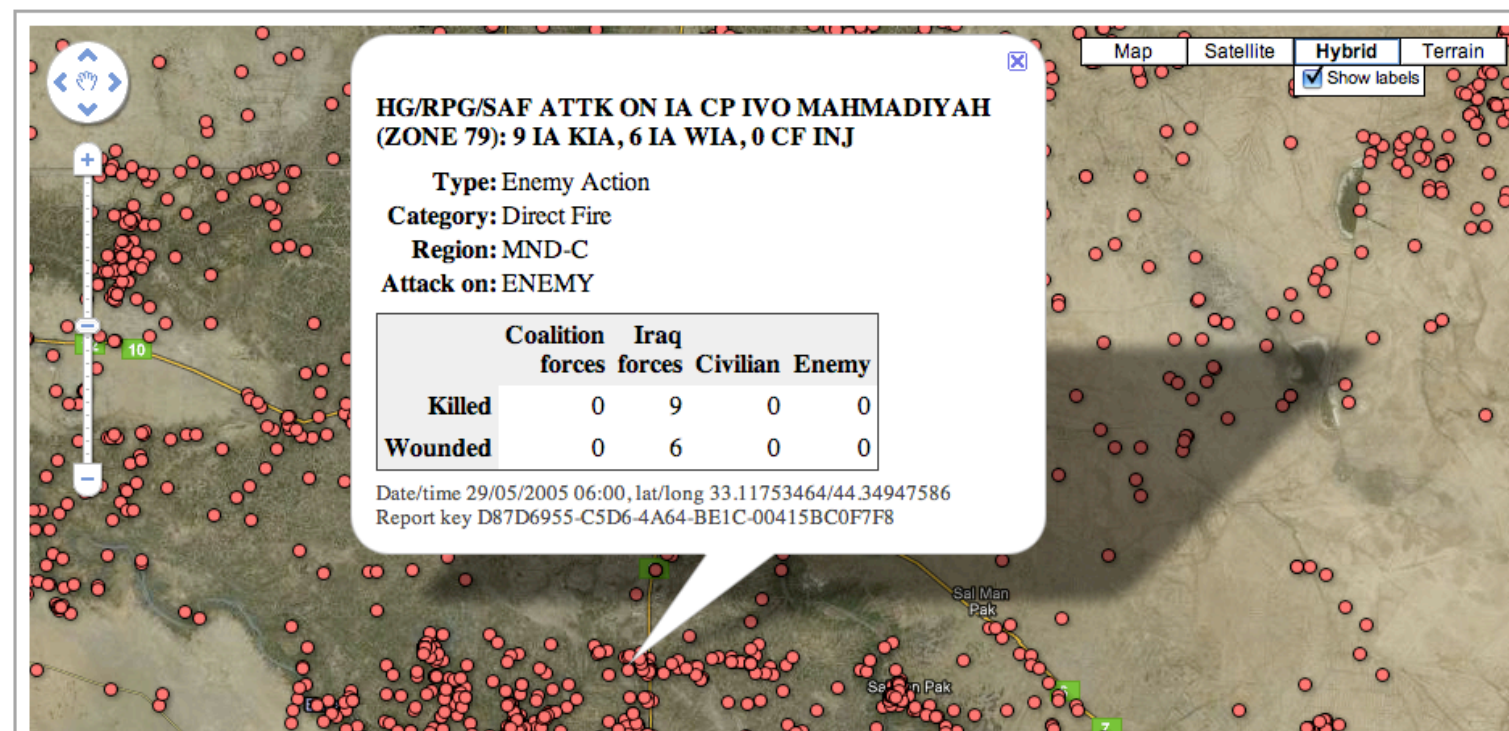
The Wikileaks Iraq war logs provide us with a unique picture of every death in Iraq. These are those events mapped using Google Fusion tables

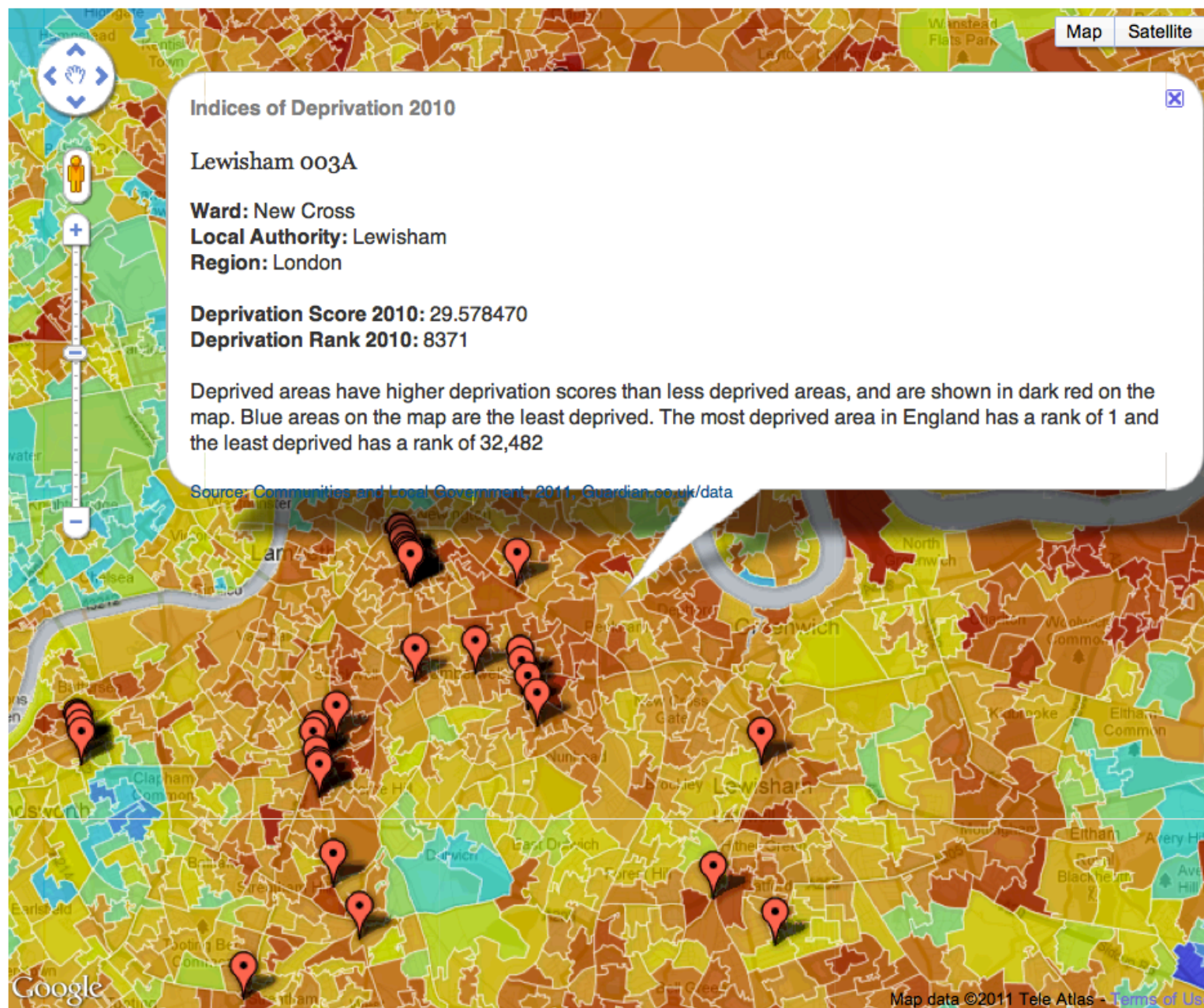
- [Download the data from the Datablog](#)

[Tweet](#) 1,706[Share](#) 8952[Comments \(85\)](#)

Simon Rogers

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





America in 2010

The latest census report underscores a decades-long demographic shift, as people and jobs have migrated away from the northern and central U.S. toward the Sunbelt. There's also a transformation under way in the country's racial and ethnic makeup as the Hispanic population surges. See change in population and total population by race across the U.S.

Hispanic

White

Black

Asian

Percent Change



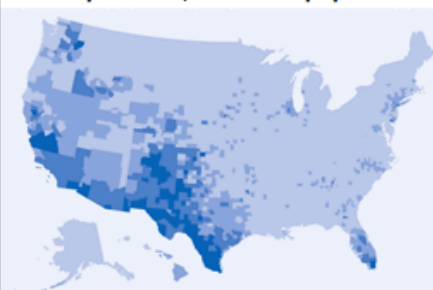
0% 100%

Show:

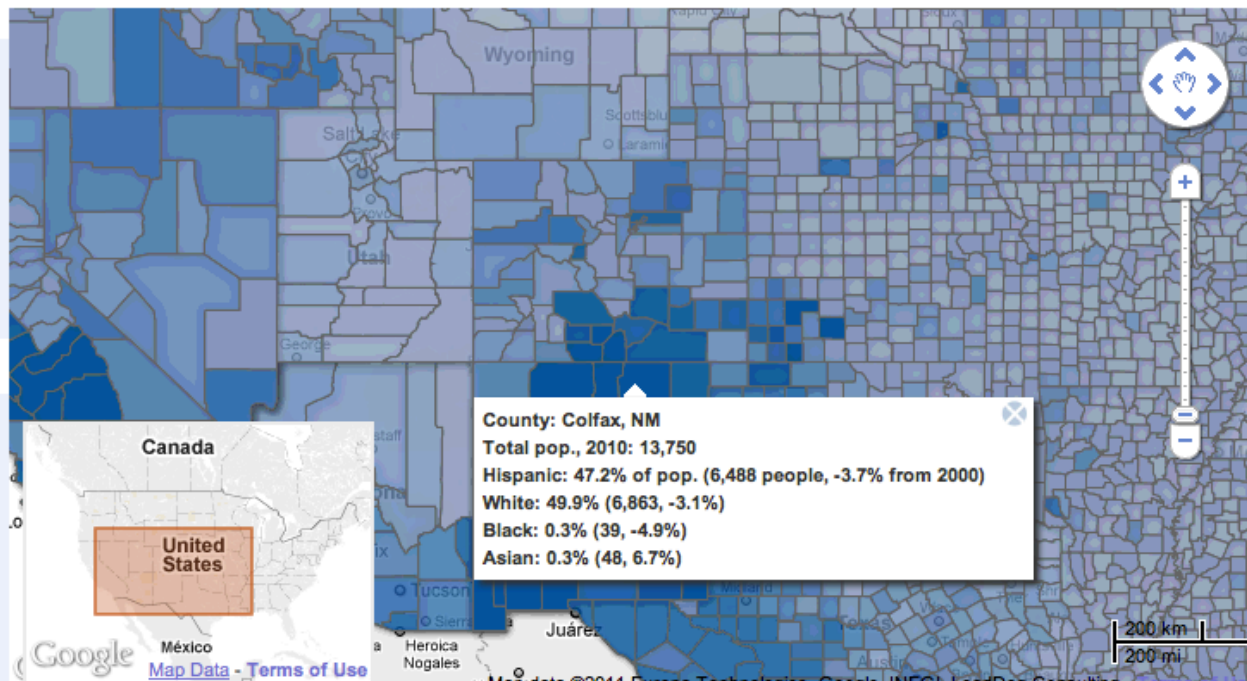
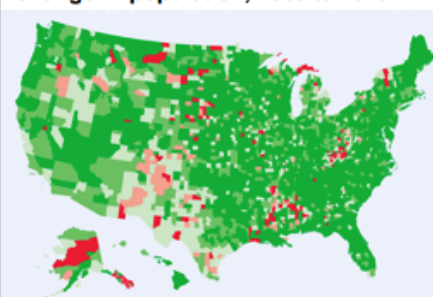
- ☒ Hispanic pop. (%)
- ☐ Hispanic pop. (% chg)

Latinos now constitute 16% of the nation's total population, with the largest proportion living in the Mexico border areas of Texas, Arizona, New Mexico and California.

Racial presence, % of total population



Change in population, 2000 to 2010



Email Like 16 people like this. Be the first of your friends.

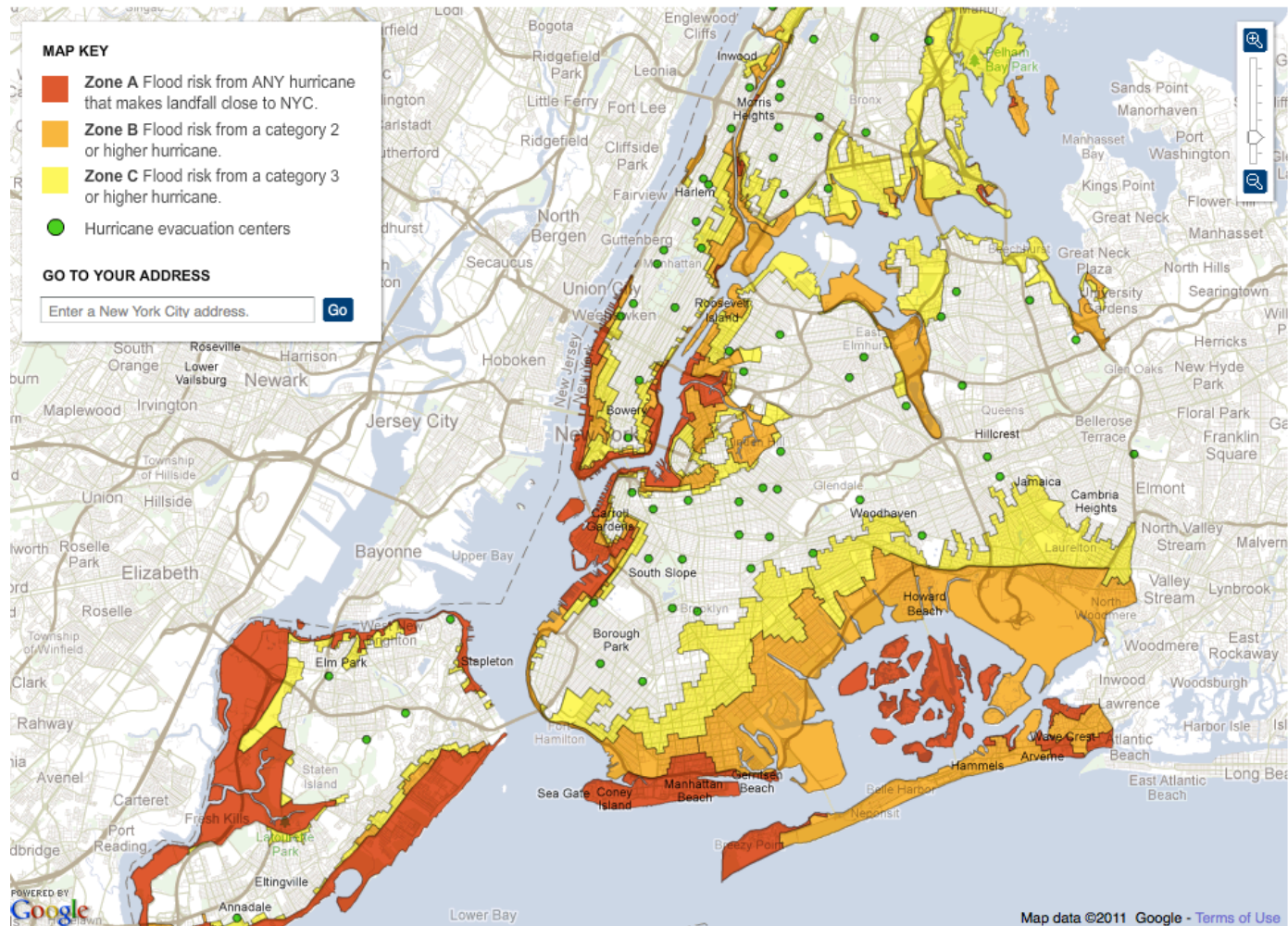
Share:

THE WALL STREET JOURNAL.

Published: August 26, 2011

New York City Hurricane Evacuation Zones

There are three evacuation zones in New York City that are based on the strength of the hurricane making land mandatory evacuation of Zone A, plus areas of the Rockaways that are in Zone B, by 5 p.m. on Saturday. [Related](#)

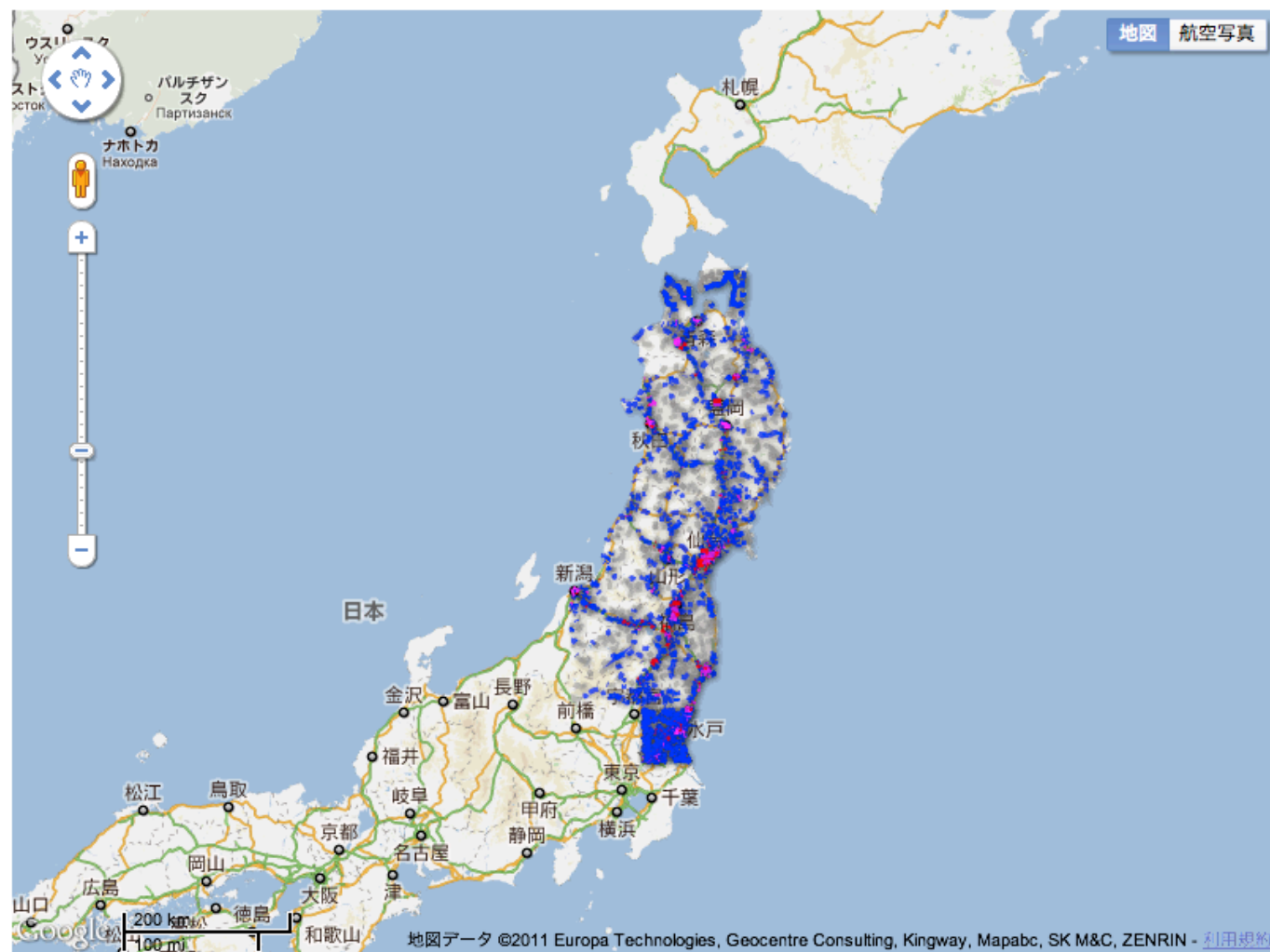


[東日本大震災](#) > 自動車通行実績情報マップ

下記マップ中において、前日の午前 6 時～午前 10 時に渋滞が発生していた道路を**赤色**、同時間帯に混雑が発生していた道路を**桜色**、上記以外の道路で前日の 0 時～24 時に通行実績のあった道路を**青色**、通行実績情報がなかった道路を灰色で表示しています。（最終更新日時：2011/08/17 08:14 JST）

住所を入力して検索:

検索



提供：本田技研工業（株）、パイオニア（株）

Iowa Caucuses

+You Web Images Videos Maps News Shopping Gmail More ▾

Sign in 



Politics & Elections

U.S. Edition ▾



Home

Trends

Results

On the Ground

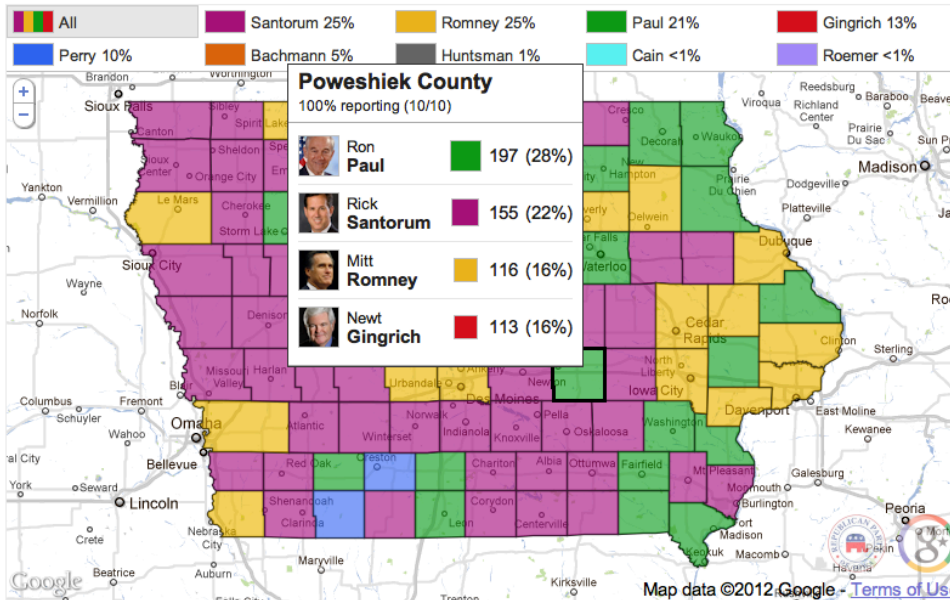
Candidates

Michele Bachmann (R)
Newt Gingrich (R)
Jon Huntsman (R)
Ron Paul (R)
Rick Perry (R)
Mitt Romney (R)
Rick Santorum (R)
Barack Obama (D)

Issues

Economy
Immigration
Healthcare
Government Spending
Jobs & Unemployment
National Security
Social Issues

Election 2012 Results



More Election Resources

Politics on Google+



Add Google Politics & Elections to your circles for the latest political news and trends.

YouTube POLITICS



Watch hot political videos and see how the candidates stack up on YouTube.

2012 Political Calendar



Track the whereabouts of all the 2012 presidential candidates and other key events.

 PBS NEWSHOUR

Politics & Elections Toolkit



Organize staff, reach voters, get a scoop: use Google's tools for a leg up in politics.

Global earthquake activity since 1973 and nuclear power plant locations



☒ Enable earthquake heatmap

[Terms of Use](#)

maptd.com

Crowd Sourcing

otrobache.com ^(SPAIN)

BACHES

SABER MÁS

Que nosotros sepamos,
hay **1258 baches** en España



Municipios cerca

pozuelo de alarcón

madrid

マドリッド

majadahonda

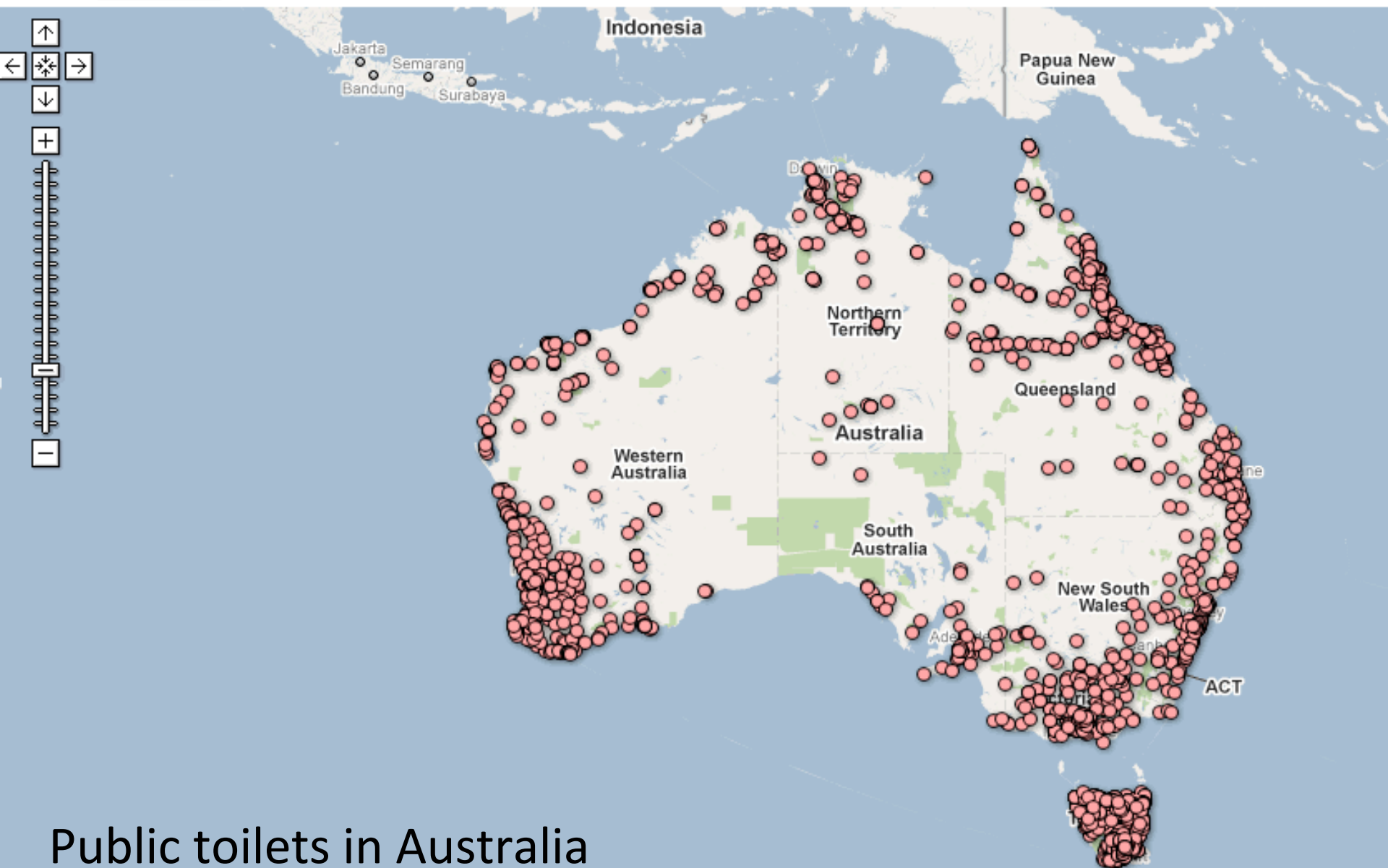
las rozas de madri

[Otros lugares](#) →

+ Reporta un nuevo bache

A GIS in the Cloud

- That's not what we set out to do, really.
- Challenges:
 - Trickle: show only a small number of features (points, polygons) from a large data set
 - Need to thin polygons, clip to the window
 - Style features on the fly
 - All in less than 100ms



Public toilets in Australia

Mid-flight Observations

- Journalists are cool.
- API and platform are important.
- Need to make joins much easier to do.
- Annotations haven't caught on yet:
 - Too vague for the users
- Need to think about supporting better commonly occurring user tasks

And the Credit Goes to...

- Hector Gonzalez
- Jayant Madhavan
- Sree Balakrishnan
- Heidi Lam
- Hongrae Lee
- Warren Shen
- Anno Langen
- Rebecca Shapley
- Anish Das Sarma
- Boulos Harb
- Fei Wu
- Cong Yu
- Spiros Papadimitriou
- Chris Olston
- Nitin Gupta

Outline

- ✓ Google Fusion Tables:
 - A Database management service for the Web
- *WebTables:*
 - *Discovering a (structured) needle in an (unstructured) haystack*

Discovery = incentive to publish

Main Points

- Structured data comes in many forms
 - We're focusing on HTML tables for now.
- Discovering good tables is like finding a needle in the haystack.
- Interesting ways of recovering table semantics:
 - Leveraging open information extraction
 - Detecting related tables

Tables on the Web

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
African-Americans						Artists		Explorers of the US					Inventors			US Presidents				US Symbols			US States		



[President's Day Activities](#)

[EnchantedLearning.com](#)

The Presidents of the United States of America

[In the order in which they served](#)

[Alphabetical order](#)

[Short table of Data](#)

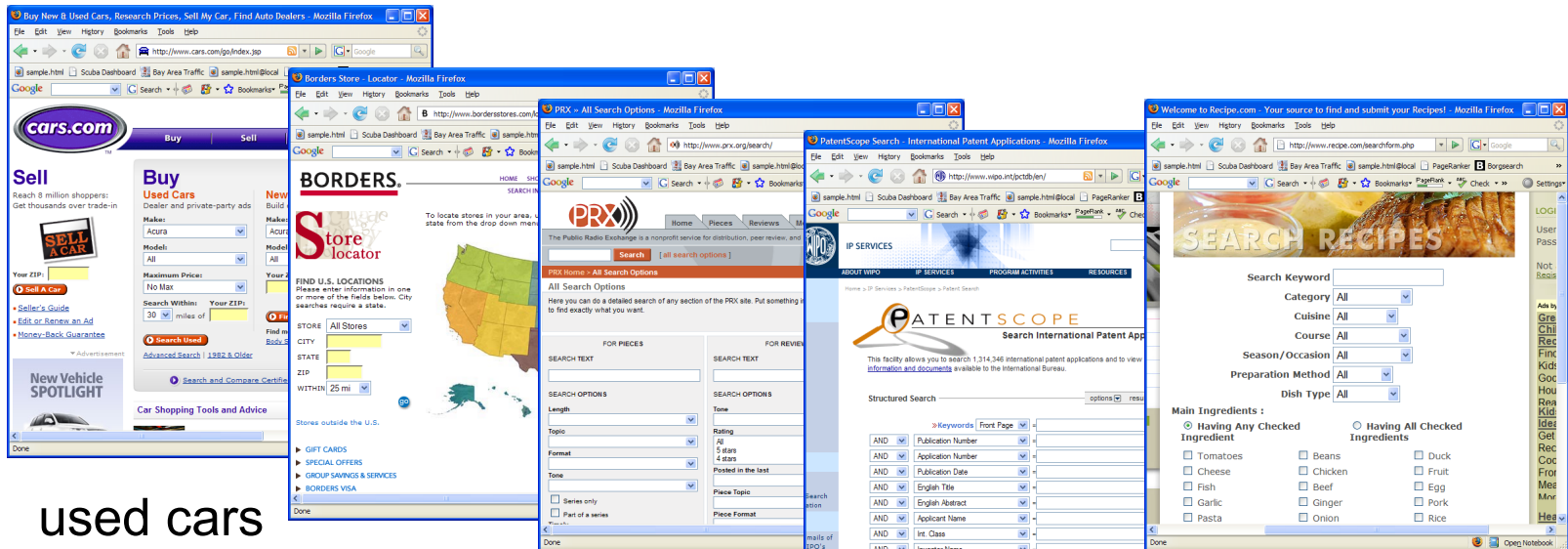


[Abraham Lincoln](#)

The President and Vice-President are elected every four years. They must be at least 35 years of age, they must be native-born citizens of the United States, and they must have been residents of the U.S. for at least 14 years. (Also, a person cannot be elected to a third term as President.)

President	Party	Term as President	Vice-President
1. George Washington (1732-1799)	None, Federalist	1789-1797	John Adams
2. John Adams (1735-1826)	Federalist	1797-1801	Thomas Jefferson
3. Thomas Jefferson (1743-1826)	Democratic-Republican	1801-1809	Aaron Burr, George Clinton
4. James Madison (1751-1836)	Democratic-Republican	1809-1817	George Clinton, Elbridge Gerry
5. James Monroe (1758-1831)	Democratic-Republican	1817-1825	Daniel Tompkins
6. John Quincy Adams (1767-1848)	Democratic-Republican	1825-1829	John Calhoun
7. Andrew Jackson (1767-1845)	Democrat	1829-1837	John Calhoun, Martin van Buren
8. Martin van Buren (1782-1862)	Democrat	1837-1841	Richard Johnson
9. William H. Harrison (1773-1841)	Whig	1841	John Tyler
10. John Tyler (1790-1862)	Whig	1841-1845	
11. James K. Polk (1795-1849)	Democrat	1845-1849	George Dallas
12. Zachary Taylor (1784-1850)	Whig	1849-1850	Millard Fillmore
13. Millard Fillmore (1800-1874)	Whig	1850-1853	
14. Franklin Pierce (1804-1869)	Democrat	1853-1857	William King
15. James Buchanan (1791-1868)	Democrat	1857-1861	John Breckinridge

The Deep Web



used cars

store locations

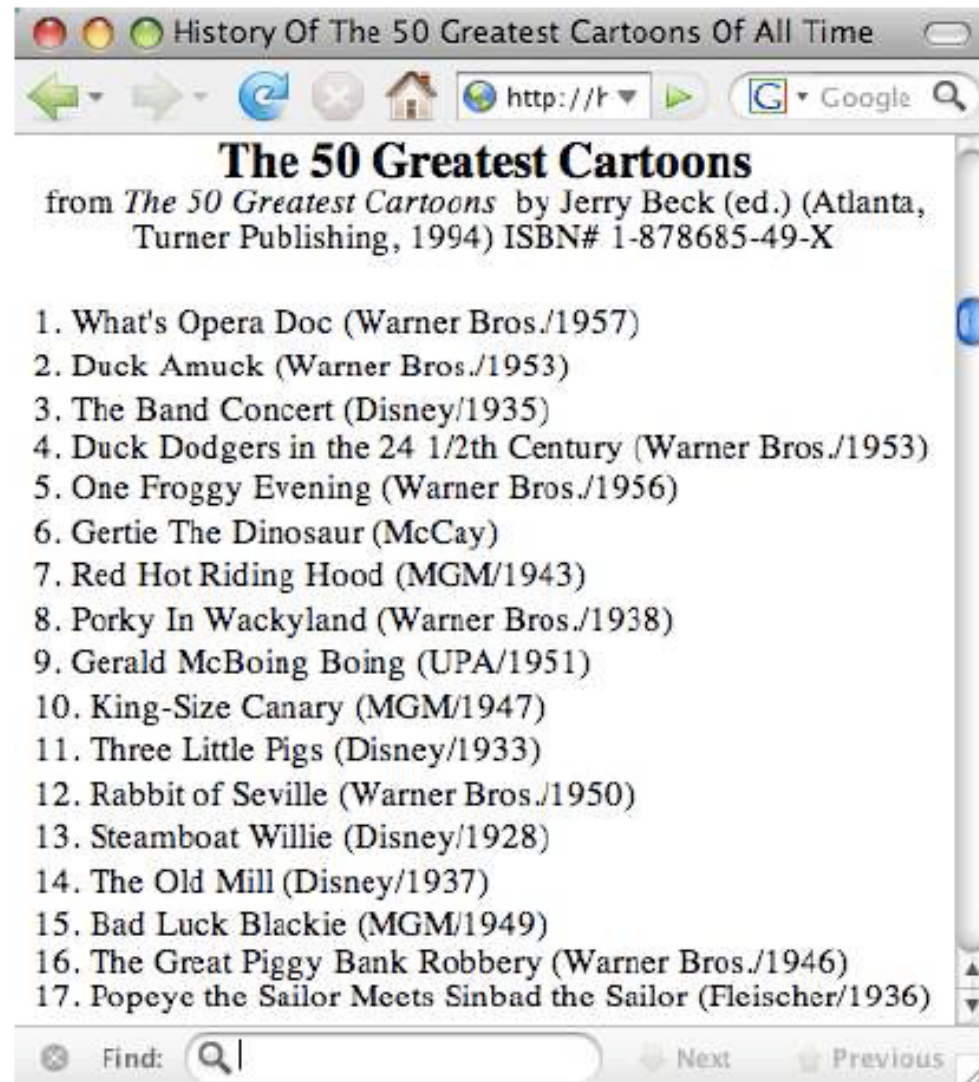
radio stations

patents

recipes

See “Google’s Deep Web Crawl”, VLDB 2008

HTML Lists



Annual Competition Results:

2010 London



The week of 23-25 June 2010, London, England was the focus of the eyes of the specialty coffee world, both online and at the host event, Caffè Culture, held in association with the Specialty Coffee Association of Europe (SCAE).

It was another epic event as Mike Phillips of Chicago, USA, walked away with the 2010 crown. 53 national barista champions in total were cheered on by an audience of over one thousand people, with a total internet audience of 29,600 (unique IP address visitors) over the course of the 3-day event via our GoLive streaming video project.

2010 – London, United Kingdom

(click on name for video)

- 1st place: [Michael Phillips](#) (United States)
- 2nd Place: [Raul Rodas](#) (Guatemala)
- 3rd Place: [Scottie Callaghan](#) (Australia)
- 4th Place: [Colin Harmon](#) (Ireland)
- 5th Place: [Soren Stiller Markussen](#) (Denmark)
- 6th Place: [Stefanos Domatiotis](#) (Greece)

[Full results](#)

2009 Atlanta



The event that marked the 10 year celebration of WBC presented in Atlanta, Georgia, April 16-19th. Featuring the first time WBC Espresso Bar with interactive presentations and sampling of competition quality beverages.

Over 8000 fans, from the stage audience to the daily online blog and our live-stream followers, were all witness to Gwilym Davies from the United Kingdom, as he claimed the title of 2009 World Barista Champion.

Held in conjunction with "The Event" SCAA 22nd Annual Exposition.

2009 – Atlanta, Georgia, United States

(click on name for video)

- 1st place: [Gwilym Davies](#) (United Kingdom)
- 2nd Place: [Sammy Piccolo](#) (Canada)
- 3rd Place: [Mike Phillips](#) (United States)
- 4th Place: [Colin Harmon](#) (Ireland)
- 5th Place: [Lee Jong Hoon](#) (Korea)
- 6th Place: [Attila Molnar](#) (Hungary)

[Full results](#)

The Needle in the Haystack

Check-In



Introductions (1 Viewing)

Glad you found OTz, now come on in and introduce yourself. Introducing yourself is the first step to finding a cure for boredom.

Sub-Forums: [Get To Know the Mods](#)

995

Threads: 2,011
Posts: 45,772



Forum How-to's (2 Viewing)

Learn how to use BBCode, upload a picture, change your signature and many other tips and tricks for forum use.

995

Threads: 97
Posts: 1,457



Board Rules & FAQ's

Here you will find announcements and frequently asked questions about new features on the forum etc. If there are any new topics here, please read them as they are important.

995

Threads: 219
Posts: 5,865



Guest Area

A place where guests (unregistered) members can post. Its the ONLY place they can post. Helpful if your having problems registering or you wanna flame away because you got slapped in the face with a ban stick.

995

Threads: 86
Posts: 1,681



Feedback & Suggestions

Got an idea or think something could be better? Let us know. We're always looking to improve, help make OTz the place to be.

995

Threads: 591
Posts: 9,542



Testing

Do all your testing of pics, sigs, etc. here. It might get deleted some day, it might not. Depends on how lazy the staff feels.

995

Threads: 179
Posts: 1,435

Customer Reviews


Average Customer Rating

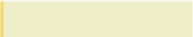
★★★★★ [\(210 customer reviews\)](#)

[5 star:](#)  (156)

[4 star:](#)  (37)

[3 star:](#)  (7)

[2 star:](#)  (4)

[1 star:](#)  (6)

Google Inc.



Type	Public
Traded as	NASDAQ: GOOG  FWB: GGQ1 
Industry	Internet Computer software
Founded	Menlo Park, California (4 September 1998) ^{[1][2]}
Founder(s)	Sergey Brin Larry Page
Headquarters	1600 Amphitheatre Parkway, Mountain View, California, United States
Area served	Worldwide
Key people	Larry Page (Co-Founder and CEO) Eric Schmidt (Executive Chairman) Sergey Brin (Co-Founder)
Products	See list of Google products .
Revenue	 US\$ 29.321 billion (2010)
Operating income	 US\$ 10.381 billion (2010)
Profit	 US\$8.505 billion (2010)
Total assets	 \$57.851 billion (2010)
Total equity	 US\$46.241 billion (2010)
Employees	28,768 (As of 2011-06-30) ^[3]

Vertical Tables

Coffee



A cup of black coffee

Type	Hot
Country of origin	Ethiopia
Introduced	Approx. 15th century (beverage)
Color	Dark brown, beige, black, light brown

Semantics Embedded in Surrounding Text

The following table lists the total coffee production of each [coffee exporting country](#) in the year [2006](#)¹

Country	60 kilogram bags	Kilograms	Pounds
Brazil	42,512,000	2,550,720,000	5,611,584,000
Vietnam	15,000,000	900,000,000	1,980,000,000
Colombia	11,600,000	696,000,000	1,531,200,000
Indonesia	6,850,000	411,000,000	904,200,000
Ethiopia	5,500,000	330,000,000	726,000,000
India	5,005,000	300,300,000	660,660,000
Mexico	4,500,000	270,000,000	594,000,000
Guatemala	4,000,000	240,000,000	528,000,000
Peru	3,500,000	210,000,000	462,000,000
Honduras	2,700,000	162,000,000	356,400,000
Uganda	2,500,000	150,000,000	330,000,000
Ivory Coast	2,350,000	141,000,000	310,200,000
Costa Rica	1,808,000	108,480,000	238,656,000



Alphabetical

▶ ALPHABETICAL

A			
ALLEN, Mike	(253) 851-2236	Police	Traffic Officer
ALTMAN, Mort	(253) 851-2236	Police	COPS Volunteer
ANDREWS, Cindy	(253) 851-6170	Planning	Planning Assistant
APPLETON, Emily	(253) 853-7620	Engineering	Senior Engineer
AUCLAIR, Jaci	(253) 853-7624	Finance	Finance Technician
B			
BELBECK, Angela	(253) 853-7611	Administration	City Attorney
BENAK, Jennifer	(253) 851-7808	Court	Lead Court Clerk
BOWER, Dick	(253) 851-6137	Bldg-Fire Safety	Bldg / Fire Safety Director
BREKKE, Laurelyn	(253) 853-7638	Administration	Executive Assistant
BRUNSON, Raquel	(253) 851-2236	Police	Police Officer
BUSEY, Kelly	(253) 853-2422	Police	Sergeant
C			
CABACUNGAN, Mike	(253) 853-7573	Police	Detective
CHAPMAN, Garret	(253) 851-2236	Police	Police Officer
CHUNN, Jami	(253) 853-7625	Engineering	Engineering Assistant
COLBERG, Stacy	(253) 851-7808	Court	Court Administrator
COLBERG, William	(253) 853-2421	Police	Lieutenant
COX, Sharon	(253) 853-2424	Police	Police Officer
D			
DAHM, Gary	(253) 851-2236	Police	Police Officer
DAVIS, Mike	(253) 853-2420	Police	Chief of Police
DENNIS, Chet	(253) 851-2236	Police	Police Officer
DOLAN, Tom	(253) 853-7615	Planning	Planning Director
DOUGIL, Matt	(253) 853-2424	Police	Police Officer

It's Complicated

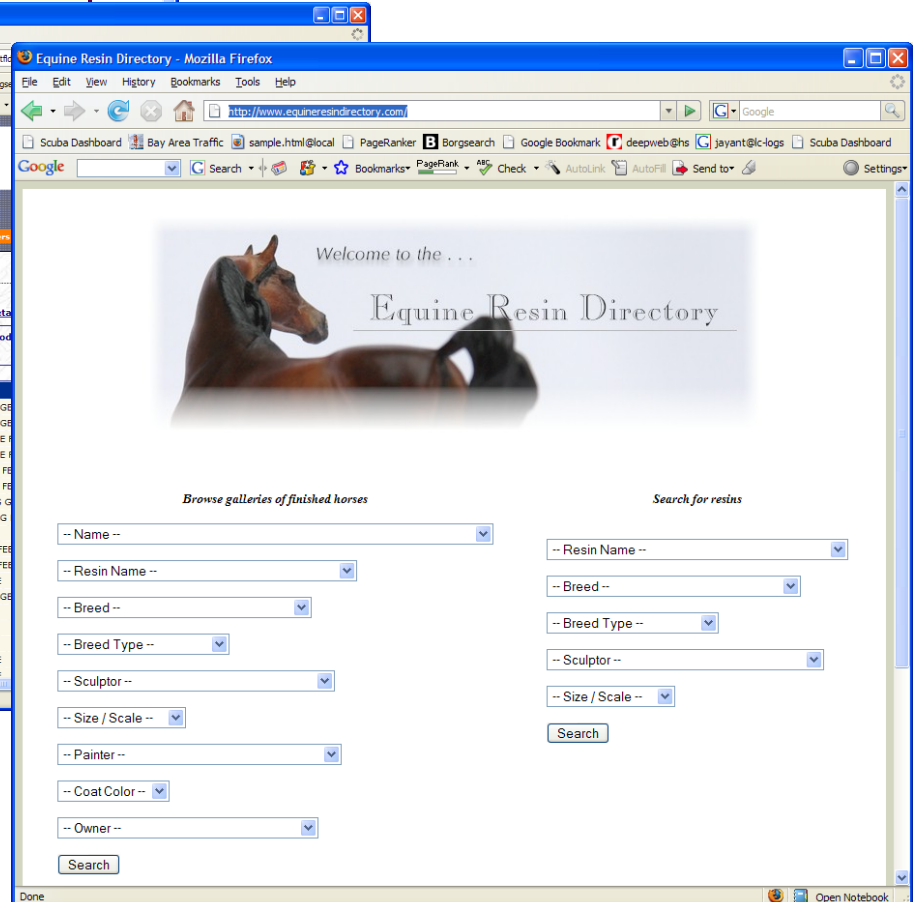
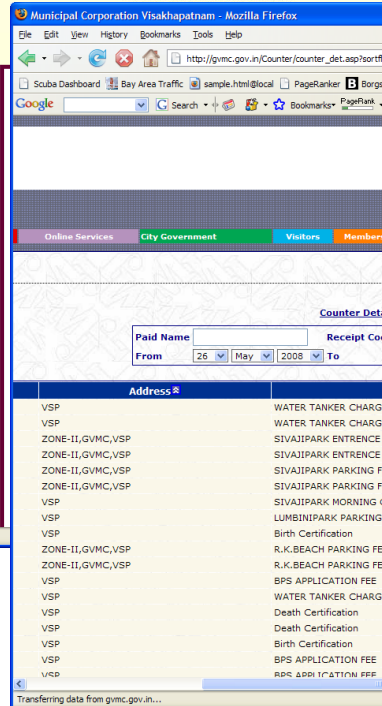
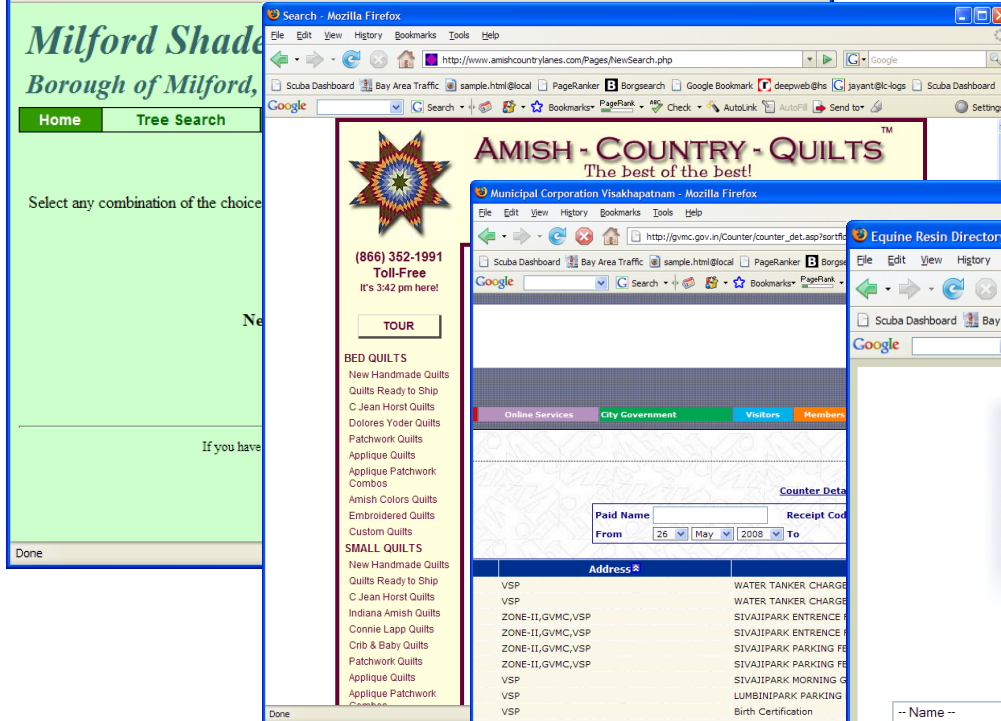
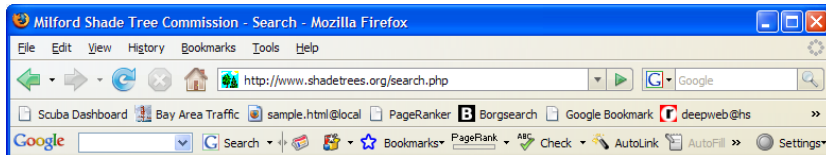
Largest Asian American Ethnic Groups, 2000 Census

Ethnic Group	Asian alone		Asian & at least One Other Race (i.e., Filipino-White)	Total Population, Alone or in Any Combination
	Single Ethnicity	Two or More Asian Ethnicities (i.e., Chinese- Vietnamese)		
Chinese	2,314,537	130,826	289,478	2,734,841
Filipino	1,850,314	57,811	456,690	2,364,815
Asian Indian	1,678,765	40,013	180,821	1,899,599
Korean	1,076,872	22,550	129,005	1,228,427
Vietnamese	1,122,528	47,144	54,064	1,223,736
Japanese	796,700	55,537	296,695	1,148,932
Cambodian	171,937	11,832	22,283	206,052
Pakistani	153,533	11,095	39,681	204,309
Laotian	168,707	10,396	19,100	198,203
Hmong	169,428	5,284	11,598	186,310
Thai	112,989	7,929	29,365	150,293
Taiwanese	118,048	14,096	12,651	144,795
Indonesian	39,757	4,429	18,887	63,073
Bangladeshi	41,280	5,625	10,507	57,412

Data is About *Everything*

Tree Search

Amish quilts



Parking tickets in India

Horses

WebTables: Exploring the Relational Web

[Cafarella et al., VLDB 2008, WebDB 08]

- In corpus of 14B raw tables, we estimate 154M are “good” relations
 - Single-table databases; Schema = attr labels + types
 - Largest corpus of databases & schemas we know of
- The Webtables system:
 - Recovers good relations from crawl and enables search
 - Builds novel apps on the recovered data

Searching Tables is Tricky

[Tweak Document Search, Cafarella 08]

- Consider new cues in ranking:
 - Hits on left column
 - Hits on schema (where there is one)
 - Number of rows, columns
 - Hits on table body
 - Size of table relative to page
- ~25% increase in good results in top-10 results (compared to filtering google results for tables)

Recovering Table Semantics

[Venetis et al., VLDB 2011]

Trees at Lake Elkhorn

As a recreational facility in Columbia, Lake Elkhorn has a fine collection of both native and non-native trees, many of which are linked to its [ArborTag](#) which describes the distinguishing features of the species.

[Hazel Alder](#)

[Green Ash](#)

[White Ash](#)

[Baldcypress](#)

[Beech](#)

[River Birch](#)

[Boxelder](#)

Red Cedar

[Black Cherry](#)

Sweet Cherry

[Crab Apple](#)

Alnus serrulata

Fraxinus pennsylvanica

Fraxinus americana

Taxodium distichum

Fagus grandifolia

Betula nigra

Acer negundo

Juniperus virginiana

Prunus serotina

Prunus avium

Malus coronaria

Some tree species exhibit elongate dead areas of bark on diseased
existence of targeted defenses resulting from a shared evolutionary history in
an exotic North American species such as Green Ash are more susceptible to
species when planted at the same site.

Recovering Binary Relationships

[Crapemyrtle](#)

[Dogwood](#)

[Korean Dogwood](#)

Redosier Dogwood

White Fringetree

Lagerstroemia indica

Cornus florida

Cornus kousa

Cornus sericea

Chionanthus virginicus

Dogwood Tree Articles



Flowering Dogwood Tree Facts

Flowering **dogwoods** are known by the name *Cornus florida* or simply "**dogwood**". They are...

***Recovering semantics: better search
and quality filter***

Finding Related Tables

- Important for a compelling table search service.
- Tables can be related in various ways:
 - Joinable, unionable, fragments of one uber table
- Turns out to also be a signal for quality:
 - If table T is related to T' that is high in the search result ranking, T could also be good and relevant.

1 - 100

As of Monday,
27.12.2010 ▼

Rankings by Country:

All Countries ▼

Additional Standings:

Top 100 ▼

Rank	Name & Nationality	Points	Position Moved	Tournaments Played
1	Nadal, Rafael (ESP)	12,450	0	20
2	Federer, Roger (SUI)	9,145	0	21
3	Djokovic, Novak (SRB)	6,240	0	21
4	Murray, Andy (GBR)	5,760	0	19
5	Soderling, Robin (SWE)	5,580	0	24
6	Berdych, Tomas (CZE)	3,955	0	26
7	Ferrer, David (ESP)	3,735	0	24
8	Roddick, Andy (USA)	3,665	0	21
9	Verdasco, Fernando (ESP)	3,240	0	25
10	Youzhny, Mikhail (RUS)	2,920	0	24

101 - 200

As of Monday,
27.12.2010 ▼

Rankings by Country:

All Countries ▼

Additional Standings:

101-200 ▼

Rank	Name & Nationality	Points	Position Moved	Tournaments Played
101	Gil, Frederico (POR)	551	0	29
102	Phau, Bjorn (GER)	551	0	31
103	Beck, Karol (SVK)	549	0	26
104	Brands, Daniel (GER)	541	0	28
105	Falla, Alejandro (COL)	540	0	23
106	Dimitrov, Grigor (BUL)	536	0	20
107	Bolelli, Simone (ITA)	532	0	29
108	Devvarman, Somdev (IND)	526	0	27
109	Darcis, Steve (BEL)	521	0	23
110	Zeballos, Horacio (ARG)	517	0	32

1 - 100

As of Monday,
27.12.2010 ▼

Rankings by Country:

All Countries ▼

Additional Standings:

Top 100 ▼




Rank	Name & Nationality	Points	Position Moved	Tournaments Played
1	Nadal, Rafael (ESP)	12,450	0	20
2	Federer, Roger (SUI)	9,145	0	21
3	Djokovic, Novak (SRB)	6,240	0	21
4	Murray, Andy (GBR)	5,760	0	19
5	Soderling, Robin (SWE)	5,580	0	24
6	Berdych, Tomas (CZE)	3,955	0	26
7	Ferrer, David (ESP)	3,735	0	24
8	Roddick, Andy (USA)	3,665	0	21
9	Verdasco, Fernando (ESP)	3,240	0	25
10	Youzhny, Mikhail (RUS)	2,920	0	24

2010 Men's Tennis ATP Rankings

Year: 2010 ▼

Type: ATP Rankings ▼

Men's Singles Rankings

RK	PLAYER	COUNTRY	MOVEMENT	POINTS
1	Rafael Nadal		↔	0 12450
2	Roger Federer		↔	0 9145
3	Novak Djokovic		↔	0 6035
4	Andy Murray		↑	1 5760
5	Robin Soderling		↓	1 5580
6	Tomas Berdych		↔	0 3955
7	David Ferrer		↔	0 3735
8	Andy Roddick		↔	0 3665
9	Fernando Verdasco		↔	0 3240
10	Mikhail Youzhny		↔	0 2920

Related Tables

2009 Men's Tennis ATP Rankings

Year: 2009 ▼
Type: ATP Rankings ▼

Men's Singles Rankings

RK	PLAYER	COUNTRY	MOVEMENT	POINTS
1	Roger Federer		↔ 0	10550
2	Rafael Nadal		↔ 0	9205
3	Novak Djokovic		↔ 0	8310
4	Andy Murray		↔ 0	7030
5	Juan Martin del Potro		↔ 0	6785
6	Nikolay Davydenko		↔ 0	4930
7	Andy Roddick		↔ 0	4410
8	Robin Soderling		↔ 0	3410
9	Fernando Verdasco		↔ 0	3300
10	Jo-Wilfried Tsonga		↔ 0	2875

2010 Men's Tennis ATP Rankings

Year: 2010 ▼
Type: ATP Rankings ▼

Men's Singles Rankings

RK	PLAYER	COUNTRY	MOVEMENT	POINTS
1	Rafael Nadal		↔ 0	12450
2	Roger Federer		↔ 0	9145
3	Novak Djokovic		↔ 0	6035
4	Andy Murray		↑ 1	5760
5	Robin Soderling		↓ 1	5580
6	Tomas Berdych		↔ 0	3955
7	David Ferrer		↔ 0	3735
8	Andy Roddick		↔ 0	3665
9	Fernando Verdasco		↔ 0	3240
10	Mikhail Youzhny		↔ 0	2920

Conclusions

- Fusion Tables: helping get the ecosystem started.
- Search for structured data sets:
 - Much more to do!
 - Unify with other search
 - Manually created ontologies vs. extracted ones?
- Can we get the crowds to help?
 - Resolving heterogeneity
 - Create new data sets
- Can we help domain-specific expert communities?

A Few References

- Communications of the ACM: Feb 2011
- Deep web: VLDB 2008
- WebTables: VLDB 2008, 2009, 2011
- Fusion Tables: SIGMOD 2010, SOCC 2010
 - google.com/fusiontables
- Principles of Data Integration (Doan, Halevy, Ives): Morgan Kaufmann, 2012.
- The Infinite Emotions of Coffee (Halevy): next month!