
Finding Stories in Data
David Tarrant · @davetaz

What is data?

Your rights to use data to tell stories

How others tell stories with data

Open, big and personal data

Data discovery patterns

Session 1
Telling stories with data

Discussion

In your groups discuss – what is data for you?

Exercise

What is Open Data?

Definition of Open (OKF)

A piece of data or content is open if **anyone** is **free to use, reuse, and redistribute** it — subject only, at most, to the requirement to attribute and/or share-alike.

Your rights

Certain uses of content are prohibited under fair dealing (fair use in the US).

Typically these allow use of copyrighted material for:

- Commentary
- Search engines
- Criticism
- Parody
- News reporting
- Research
- Teaching
- Library archiving
- Scholarship

Fair dealing
Copyright act (Ch 63)

In deciding whether the use is a fair dealing, the following factors are considered:

- Purpose and character, including non-commercial, not for profit and educational usage
- Nature of the work or adaption
- Amount copied, relative to the whole work
- Affect on value of original work

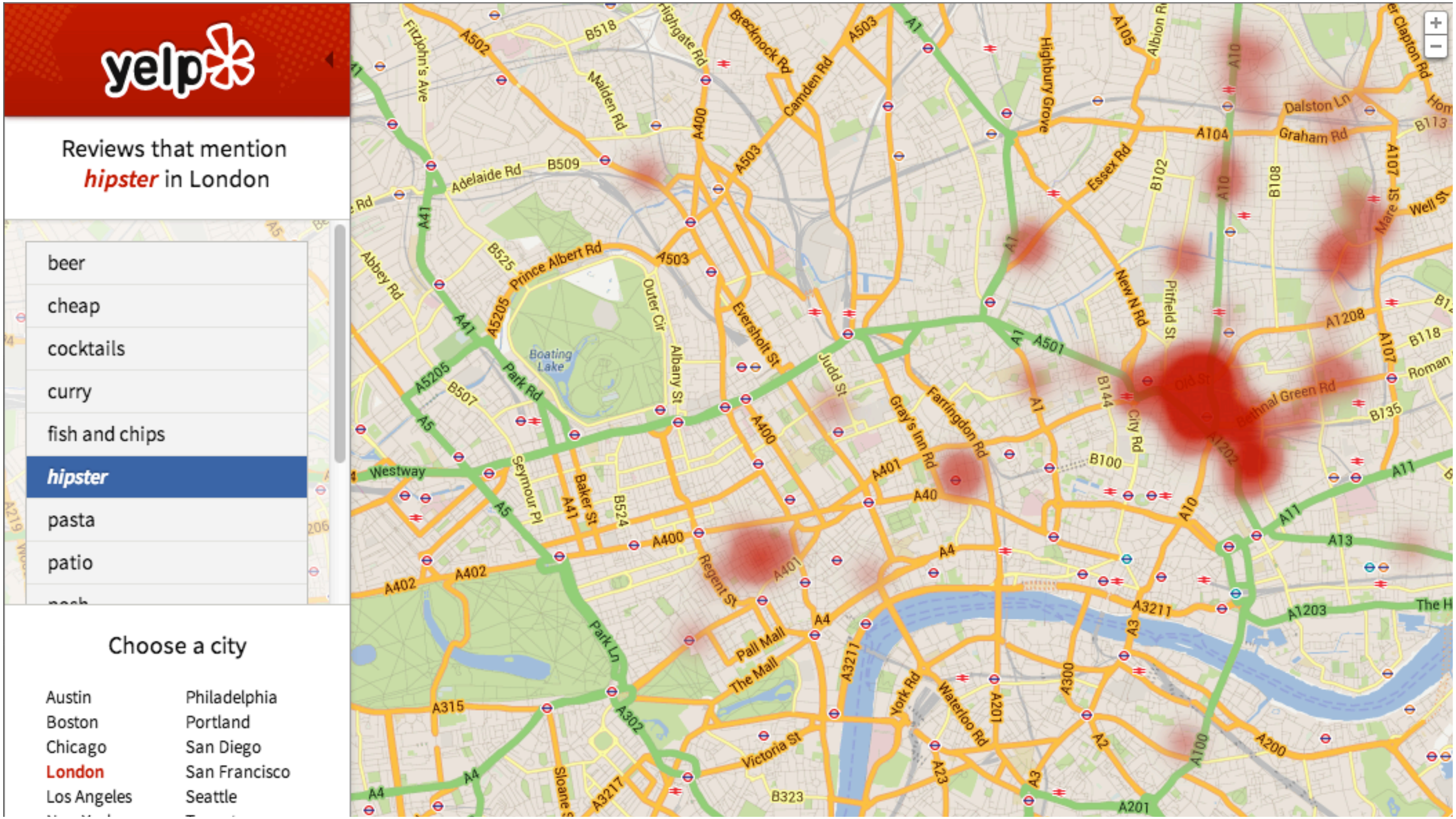
In the case of reporting, commentary and review then sufficient acknowledgement of the work is required.

http://en.wikipedia.org/wiki/Fair_dealing#Singapore

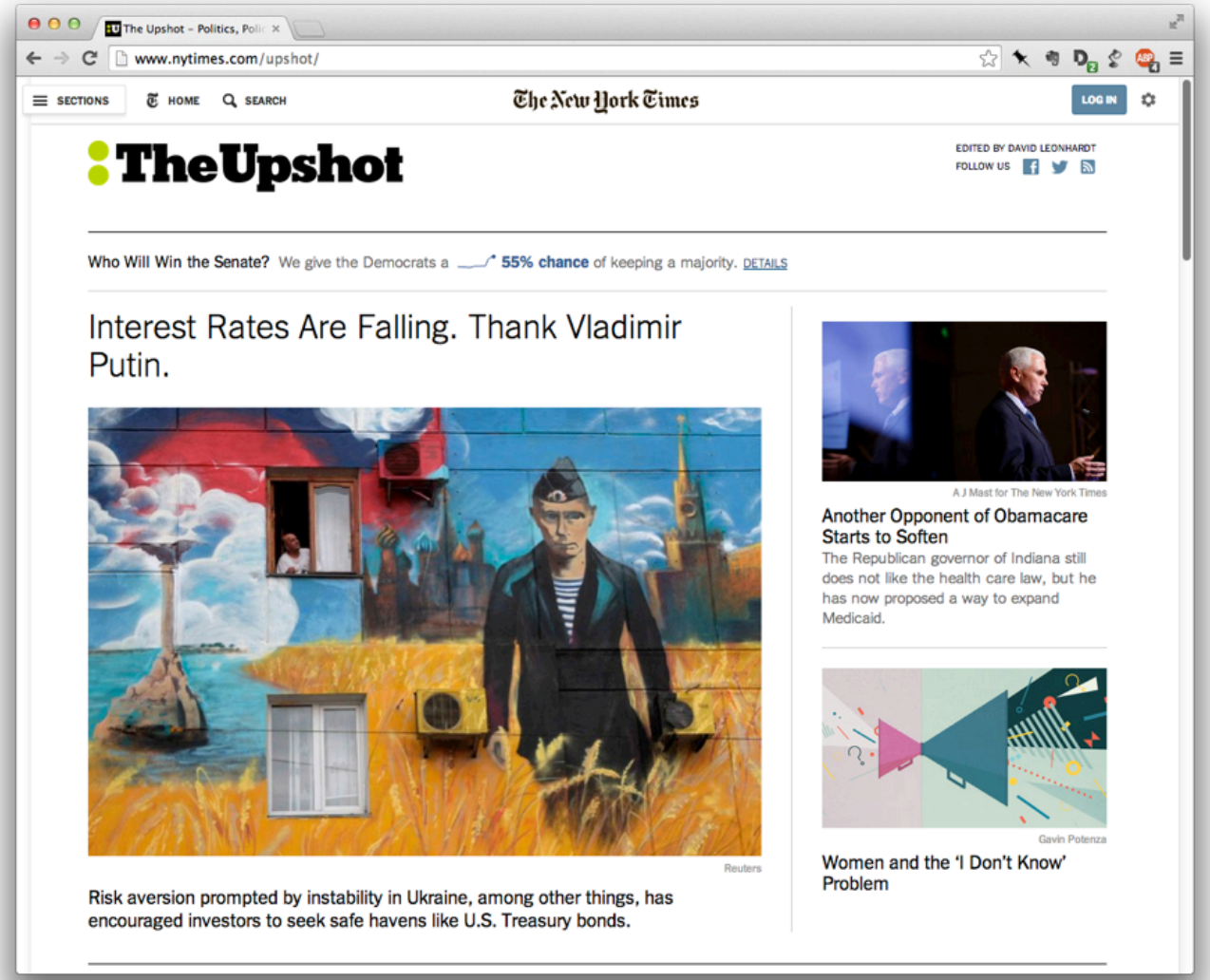
Be careful

Just because something is available in the public domain, does not mean that there is a public license.

Examples: Telling stories with data


yelp
Reviews that mention
hipster in London
beer
cheap
cocktails
curry
fish and chips
hipster
pasta
patio
Choose a city
Austin
Boston
Chicago
London
Los Angeles
Philadelphia
Portland
San Diego
San Francisco
Seattle


theguardian
News Sport Comment Culture Business Money Life & style
News Technology Airbnb
London's buy-to-let landlords look to move in on spare room website Airbnb
Data suggests investors with empty properties have carved out a huge presence on the site, leasing out their homes and flats
BBC
Sign in
News Sport Weather iPlayer TV Radio
HEALTH
England N. Ireland Scotland Wales Business Politics Health Education Sci/Environme
updated at 16:15
'miss winter target'
A&E tracker: The final week
Emergency departments, as cold weather, flu and other winter bugs lead to falls, chest infections or heart problems.
WATCH LIVE
7pm weekdays, wee
News
UK WORLD POLITICS BUSINESS SCIENCE TECHNOLOGY CULTURE
FRIDAY 07 FEBRUARY 2014 UK
Why is government website carrying fake jobs?


The Upshot – Politics, Polic
www.nytimes.com/upshot/
SECTIONS
HOME
SEARCH
The New York Times
LOG IN
TheUpshot
EDITED BY DAVID LEONHARDT
FOLLOW US
Who Will Win the Senate? We give the Democrats a 55% chance of keeping a majority. DETAILS
Interest Rates Are Falling. Thank Vladimir Putin.
Reuters
Risk aversion prompted by instability in Ukraine, among other things, has encouraged investors to seek safe havens like U.S. Treasury bonds.
A J Mast for The New York Times
Another Opponent of Obamacare Starts to Soften
The Republican governor of Indiana still does not like the health care law, but he has now proposed a way to expand Medicaid.
Gavin Potenza
Women and the 'I Don't Know' Problem

Ampp3d's Sunday Paper Review

Posted 2 days ago by **Federica Cocco** in HEALTH POLITICS

Flickr/David McDermott

Some of the most important stories of the weekend, in numbers.

3,700

According to the **Sunday Times 3,700 cancer patients waited more than 104 days for treatment in 2013**. 30% of trusts have been breaching a government target of 85% of patients receiving treatment within 62 days of an urgent GP referral.

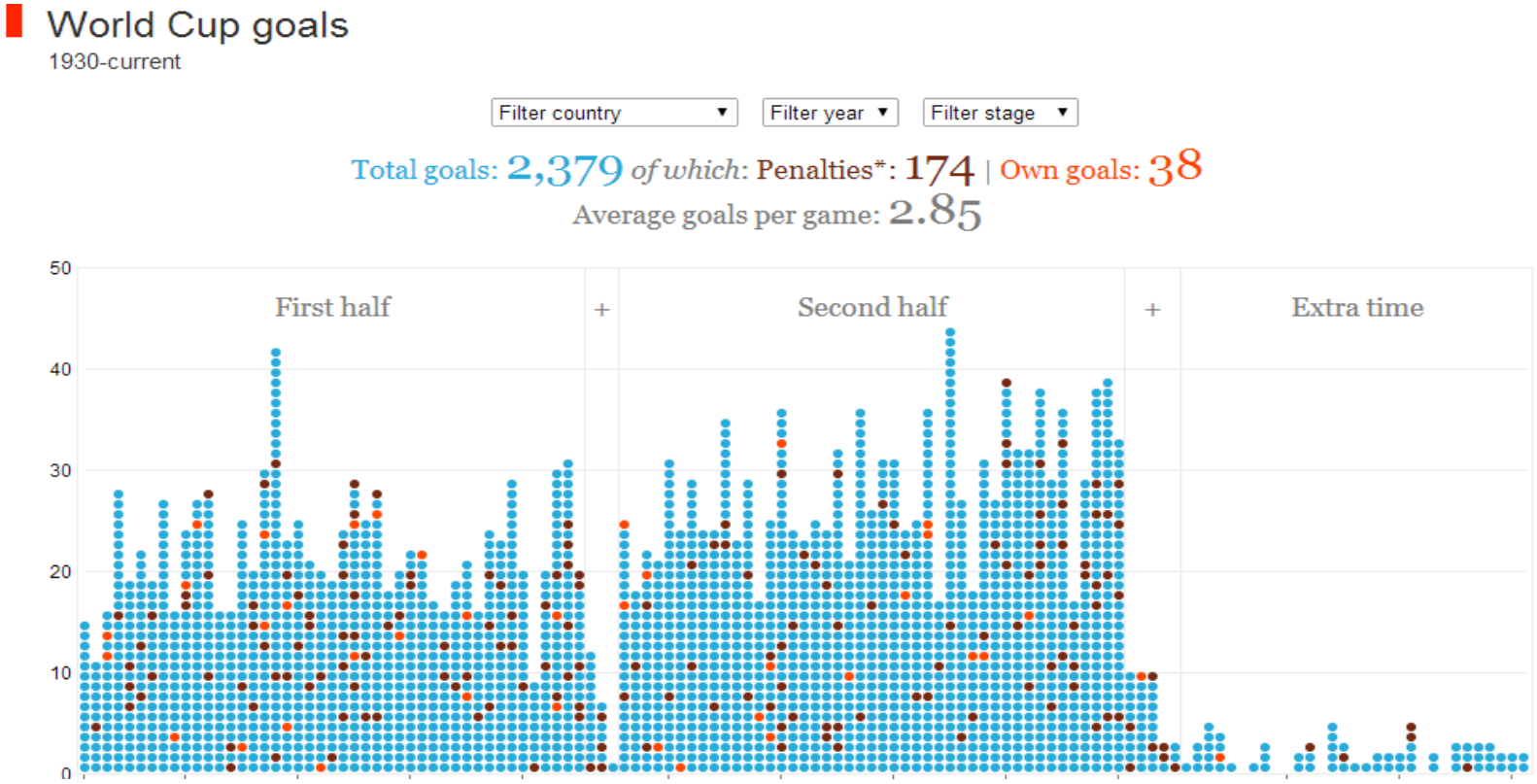
http://ampp3d.mirror.co.uk/

Data is a source

Two main methods of using data as a source

1. Story first – data used to enhance, fact check, dig deeper
2. Data first – story found/ presented through data analysis

Story, then data



Data – then story

For cops, no limit

Speeds reaching 90-130 mph are common among police

Even when there's no emergency, even when they're off duty

Punishment is rare, despite crashes and deaths

By Sally Kestin and John Maines | Staff writers

News: sources

Reactive	Proactive
UK National Statistics	FOI requests
Parliament	Surveys
Political groups	'Ideas journalism'
Businesses	Scraping

Thanks to David Ottewell Head of Data Journalism Trinity Mirror (Regionals) for permission in using this slide

Reactive news

Thanks to David Ottewell Head of Data Journalism Trinity Mirror (Regionals) for permission in using this slide

Proactive news

Thanks to David Ottewell Head of Data Journalism Trinity Mirror (Regionals) for permission in using this slide

Using data as a source ≠ must have visualisation

FINANCIAL TIMES

Welcome kcorrick(

ft.com/globaleconomy

Search

Home UK World Companies Markets Global Economy Lex Comment Manag

Economic Calendar Money Supply Americas China EU India Middle East UK US

September 18, 2013 2:51 pm

Arctic sea ice melting faster than expected, UN report finds

By Pilita Clark, Environment Correspondent

The Arctic's summer sea ice is set to nearly vanish in less than 40 years, according to the final draft of a sweeping UN climate change report that sharply revises past estimates of how fast the icy north is melting.

"A nearly ice-free Arctic Ocean in September before mid-century is likely," says the draft seen by the Financial Times of the first large-scale study in six years by the Intergovernmental Panel on Climate Change.

http://www.ft.com/cms/s/0/4b1a2f64-2048-11e3-9a9a-00144feab7de.html

Using data as a source ≠ (necessarily) big investigation

The EU could ban roaming charges completely this year

Posted 6 hours ago by **Anna Leach** in MONEY

"That call cost how much?!" Photo: Indi.ca on Flickr

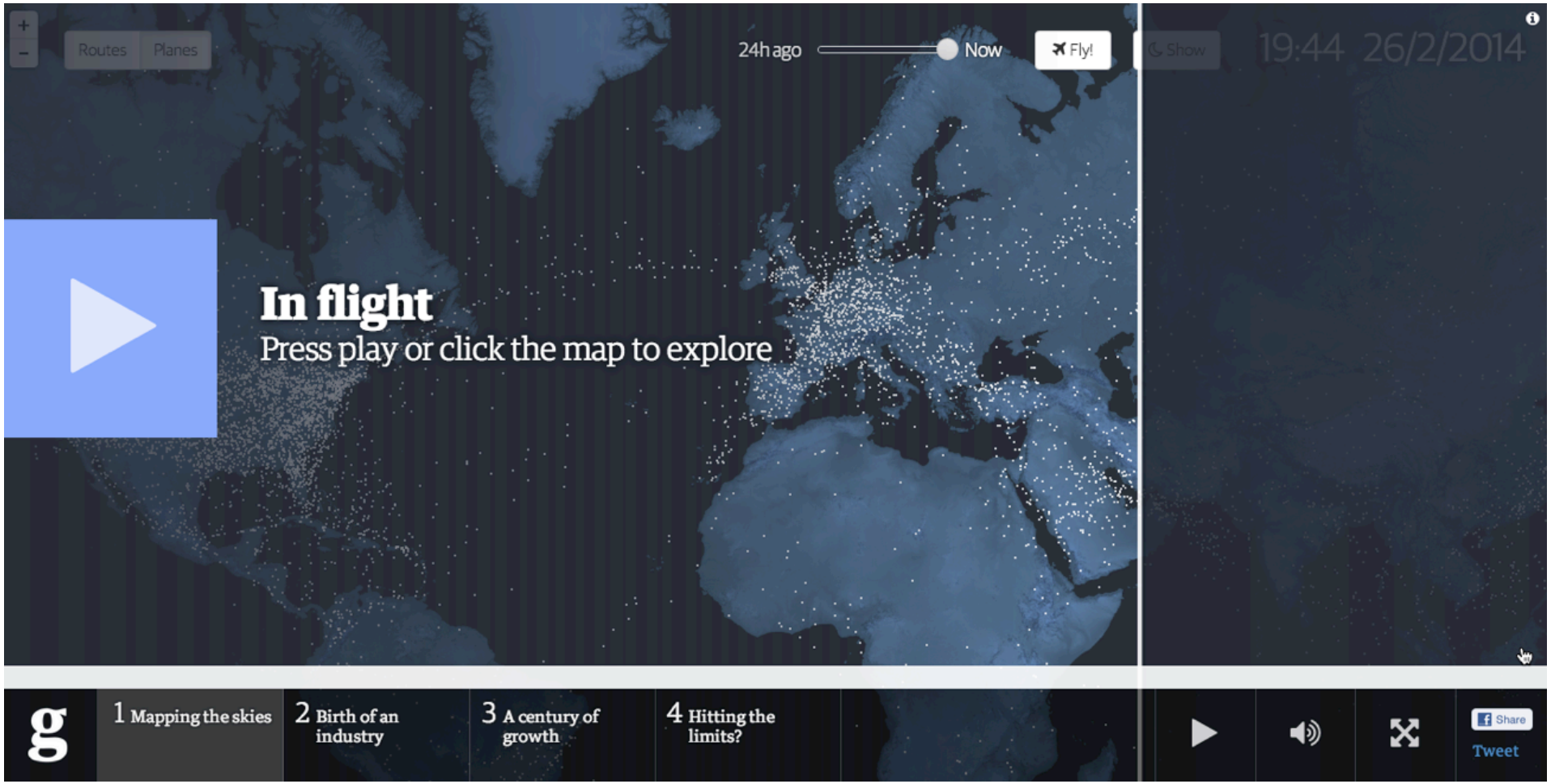
Here's how much that would might you on a 3 day holiday.

£60

That's our estimation anyway. But we do use our phones *a lot*.

http://ampp3d.mirror.co.uk/2014/02/26/the-eu-could-ban-roaming-charges-completely-this-year/

Data for education



http://aviation.live.kiln.it/

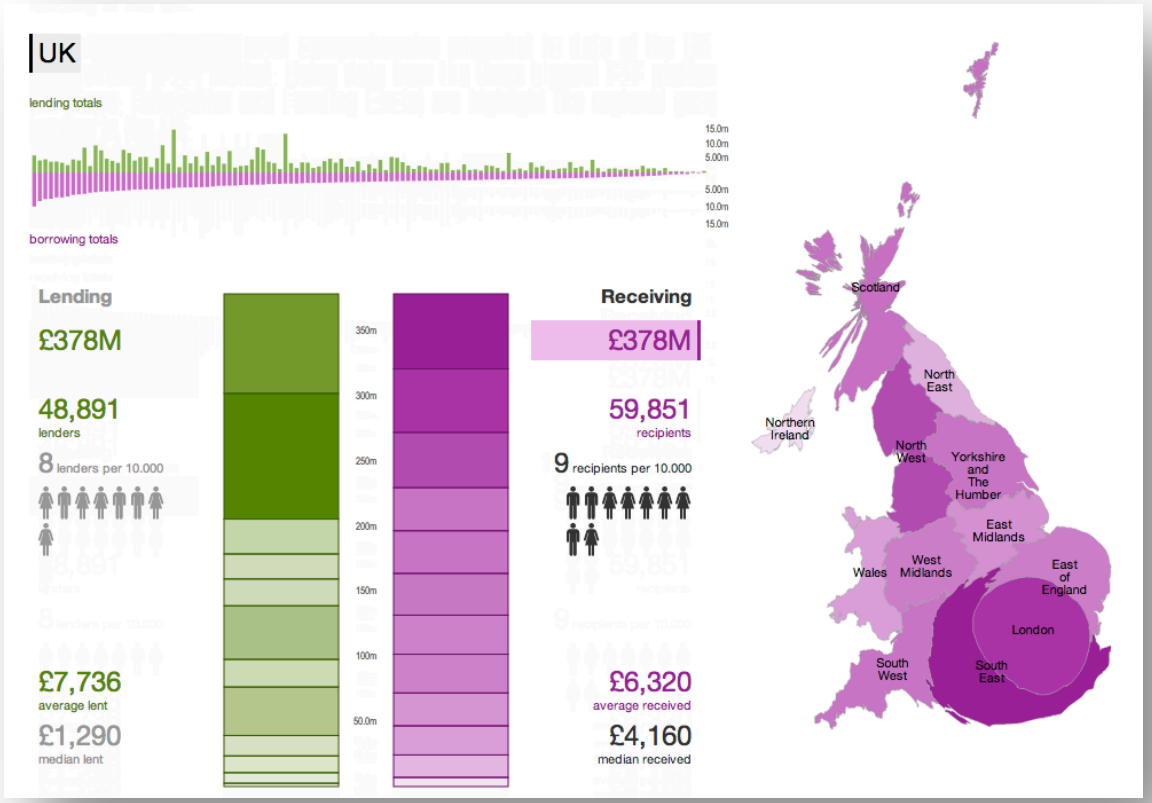
A Boom in New Housing

In spite of a recession and foreclosure crisis, the mayor presided over a boom in residential construction, encompassing everything from new aeries for the rich in Manhattan to disappearing vacant lots in the South Bronx. New York has added 40,000 new buildings since he took office, and the census counted an additional 170,000 housing units in 2010, up from 10 years earlier, more than any other city. Neighborhoods with the most growth: post-9/11 downtown; the West Side from Chelsea to Lincoln Square and Central Harlem in Manhattan; the Rockaways, Long Island City and Flushing, Queens; Williamsburg, Bushwick and Bedford-Stuyvesant, Brooklyn; the South Bronx.

http://www.nytimes.com/newsgraphics/2013/08/18/reshaping-new-york/

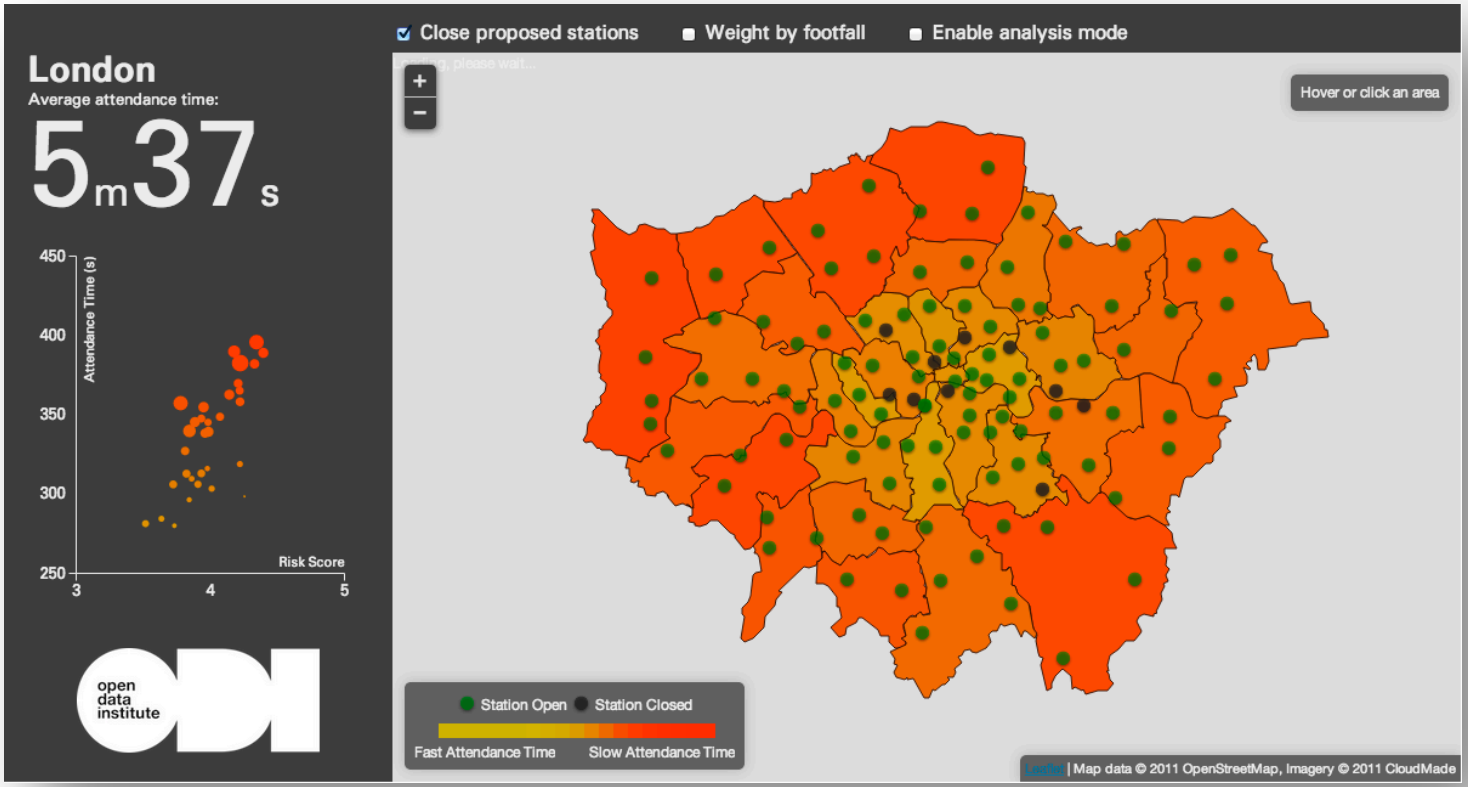
Self discovery

Show me the money



http://smtm.labs.theodi.org/

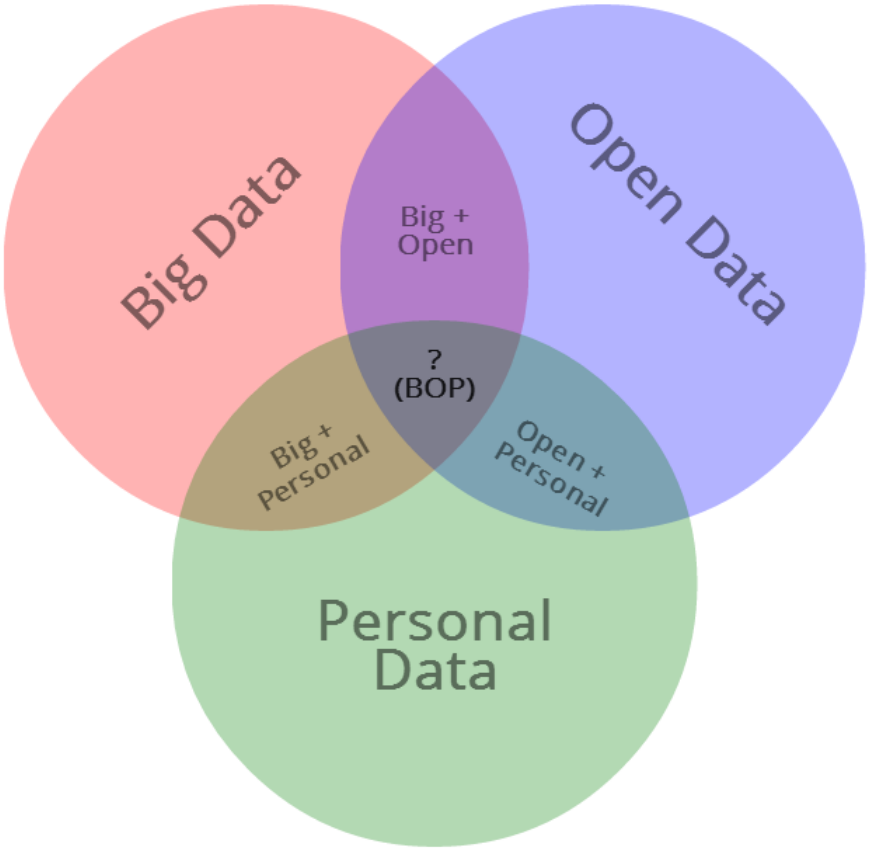
LFB Fire Station Closures



http://london-fire.labs.theodi.org/

Open big and personal data

Challenges and Risks



http://theodi.github.io/data-definitions/

Types of personal data

Open personal data

Data about people not a person

Available to anyone

Has been anonymised

e.g. number of people attending event, gender split, age ranges. (bigger numbers are better!)

Available personal data

Data about a person

Available to the person only!

Often known as MiData

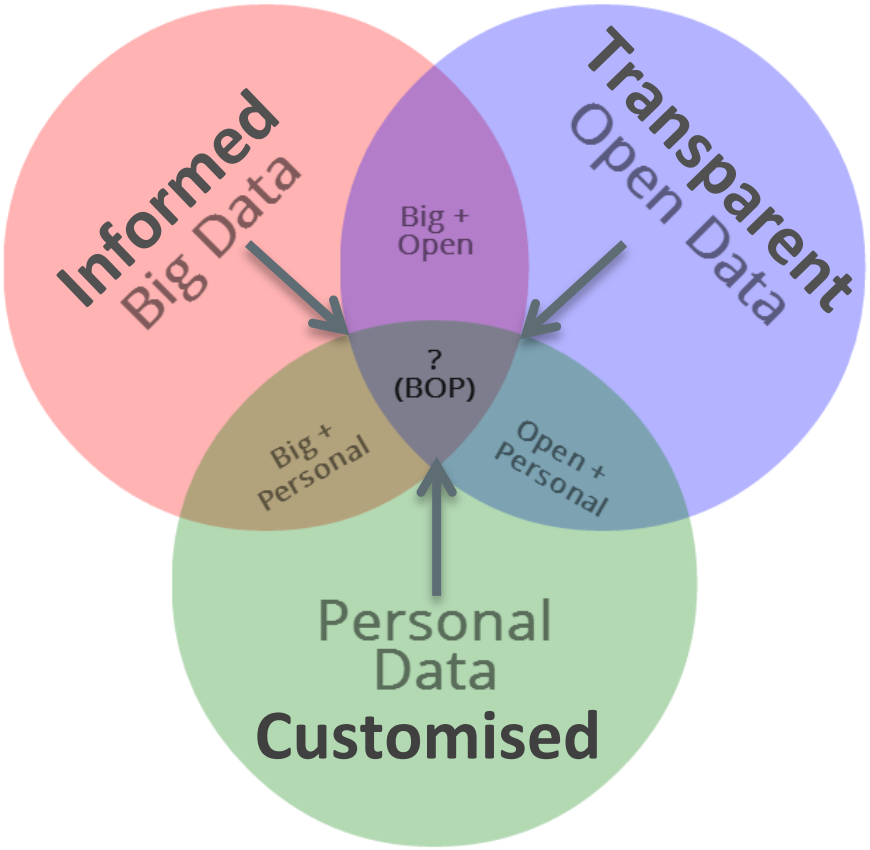
e.g. credit scores, energy and other consumption data.

Personal data

Data about a person which is neither open nor available.

Might belong to you or be collected by a company.

Opportunity



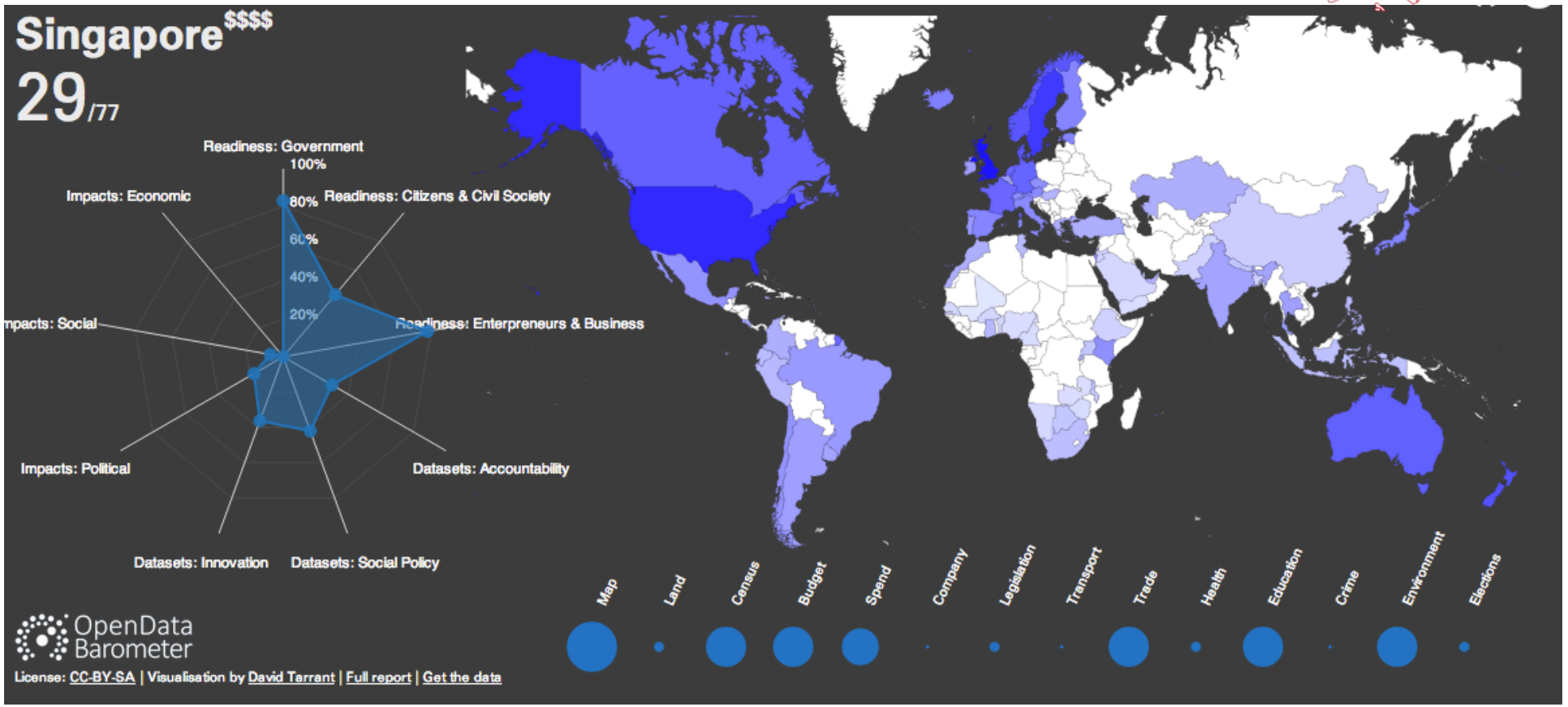
http://theodi.github.io/data-definitions/

Data discovery patterns

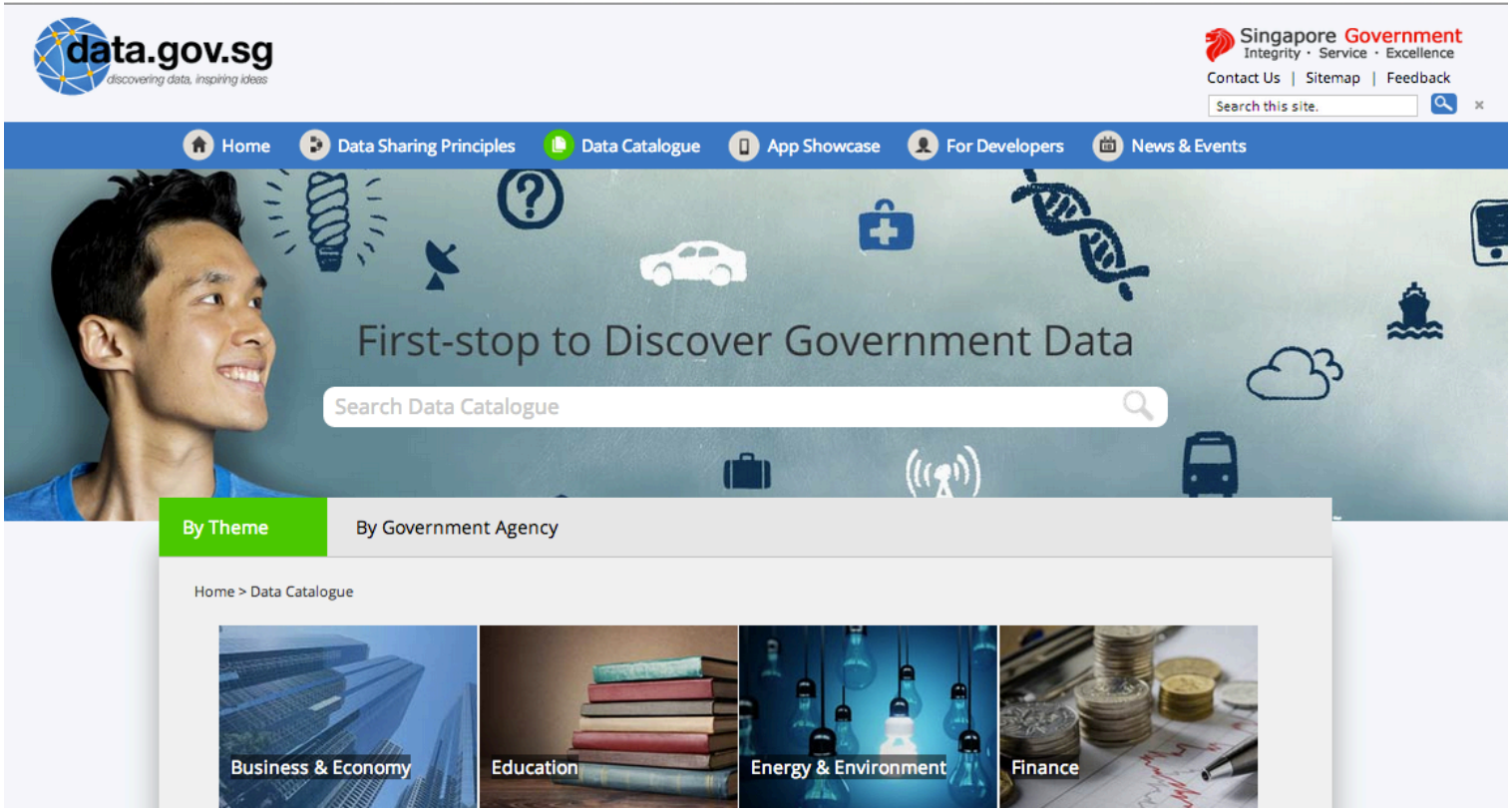
Finding data on the web **(of documents)**

- Government data
- Google advanced
- Aggregators and portals
- Scraping

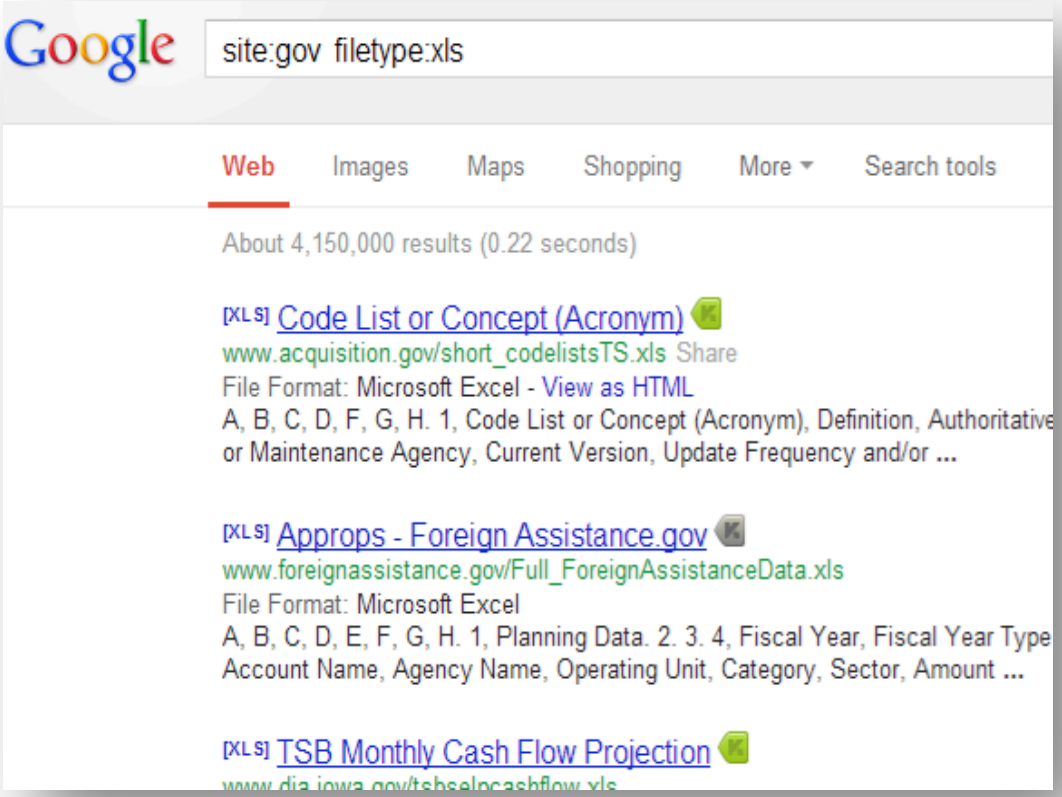
Government data



data.gov.XX



Google advanced



site: Get results only from certain sites or domains

link: Find pages that link to a certain page

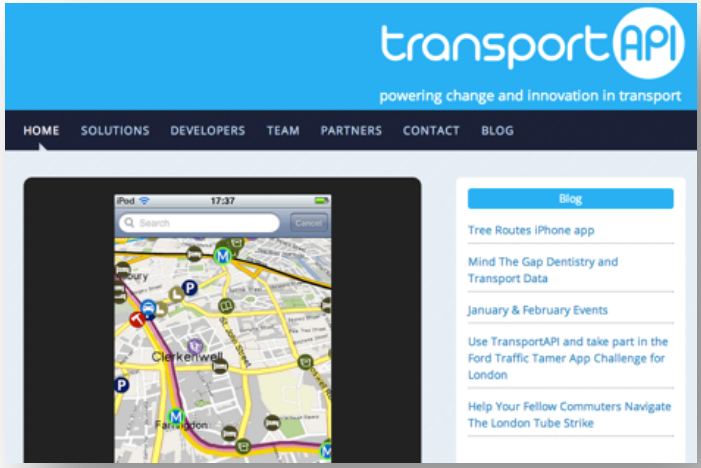
related: Find sites similar to one you already know

filetype: Find certain file types only

Aggregators and portals

Collect together data from across the web into one place.

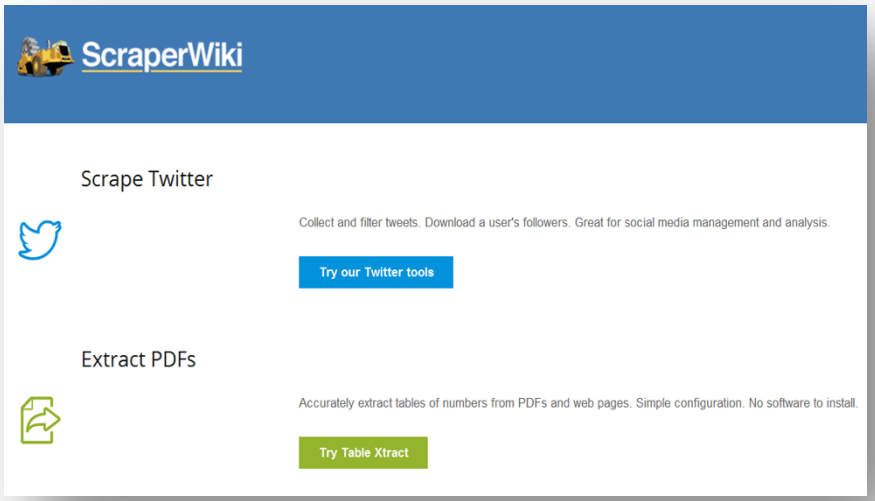
enigma.io



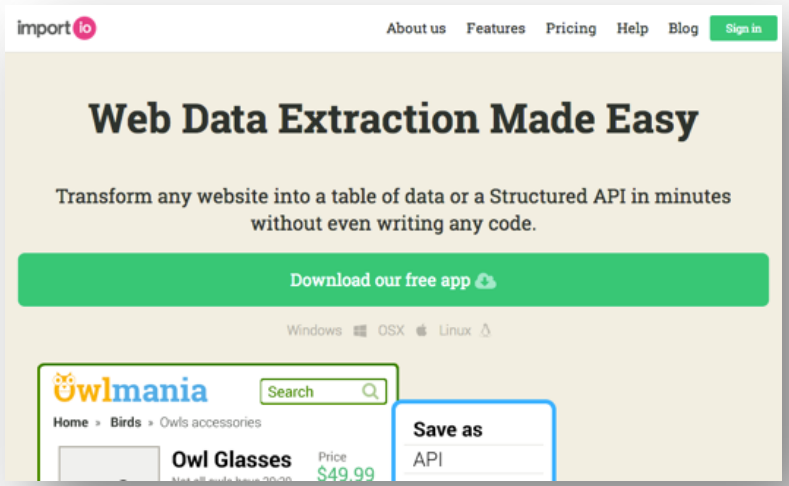
transportAPI

Scraping

If you can’t obtain usable data (csv, xls) then you may have to resort to scraping.



scraperwiki.com



import.io

Finding data on the web **(of data)**

1. Add random extensions (.xml, .json, .csv etc)
2. Look for alternative links (rss feeds etc)
3. Look for embedded data
4. Do some content negotiation
5. Spot the API
6. Scrape (or search google again)

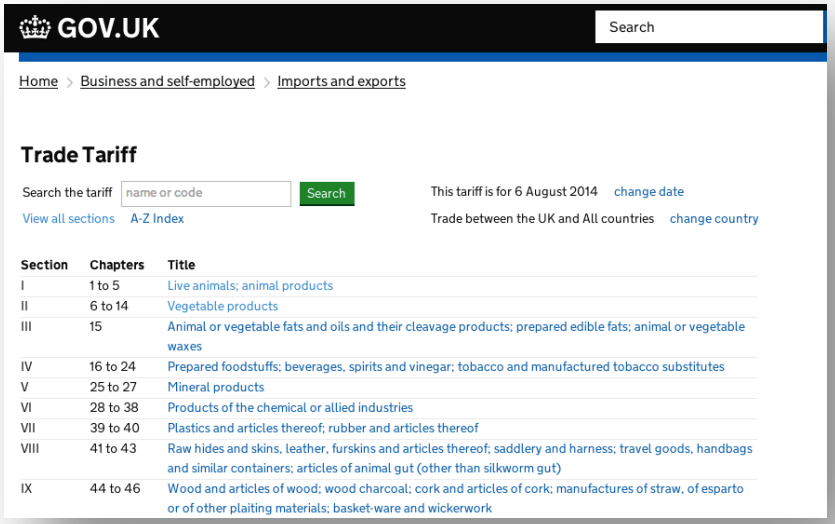
How the web should work, but people forgot that Tim put this in when he invented it!

Duck typed data

If it looks like a duck
and quacks like a duck,
then it's probably a duck.

Basically, keep an eye out for tables,
lists and other stuff that looks like data.

1. Adding random extensions

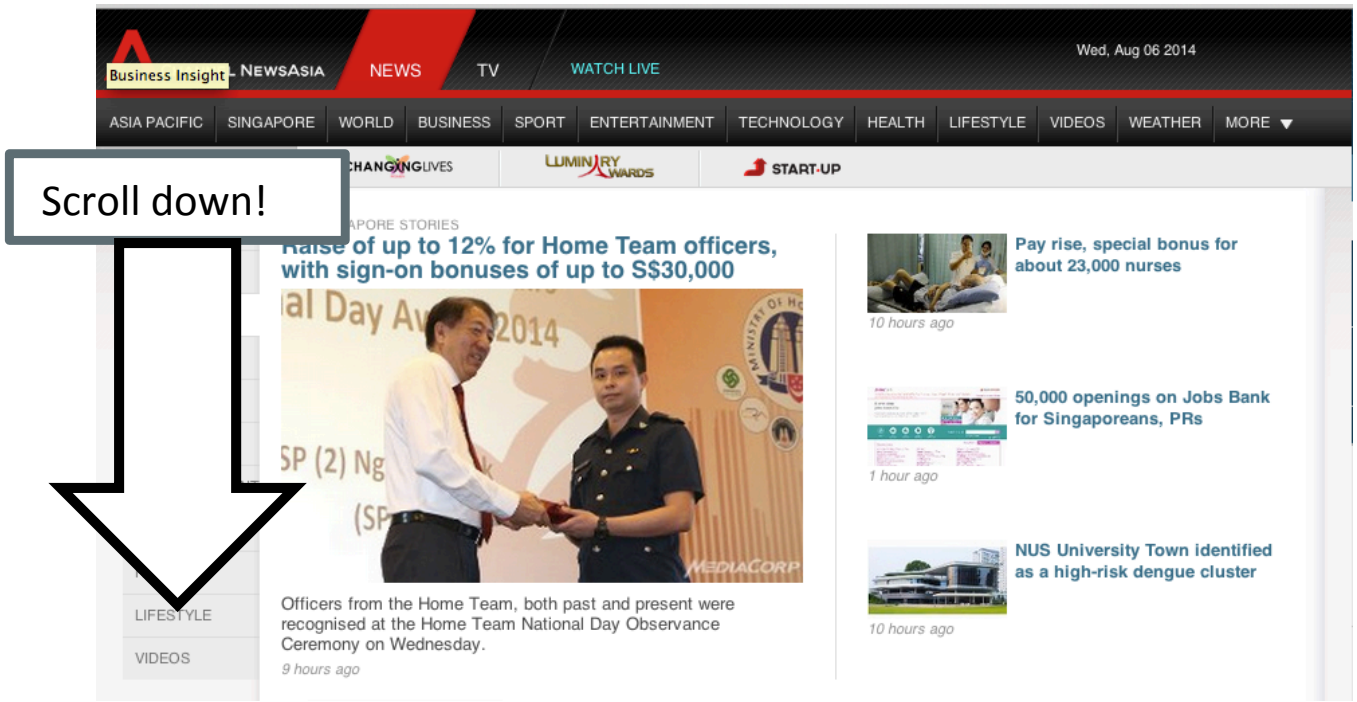


UK Trade Tariff

BBC Music and Programmes

Try using the following: .csv .json .xml .rss .rdf

2. Look for alternative links



2. Look for alternative links

CHANNEL NEWSASIA

MediaCorp News Group.
© 2014 MediaCorp Pte Ltd.
All Rights Reserved.

Terms and Conditions
Privacy Policy
About MediaCorp Pte Ltd

NEWS
Asia Pacific
Singapore
World
Business
Sport
Entertainment
Technology
Health
Lifestyle
Videos
Photos
Special Reports
Archives

TV
Live TV
TV Videos
TV Schedule

SERVICES
Weather

ADVERTISE WITH US
Online Advertising
Mobile Advertising
TV Advertising
Contact Sales

ABOUT US
About Channel NewsAsia
Our Logo
Our Coverage
Our Tagline
Presenters and Correspondents
Contact Us

GET OUR NEWS

RSS

Finding data on the web (c

Techniques 3-5 are not covered in this session. Please ask your trainer for more information if there is time.

1. json, .csv etc)
2. eeds etc)
3. Look for embedded data
4. Do some content negotiation
5. Spot the API
6. Scrape (or search google again)

Exercise

Find a data set using one of the routes we've just looked at.....

Ask yourself – (and discuss in groups)

Do you trust it? What makes it trustworthy?

A basic test with three questions:

1. Says who?
2. Compared to what?
3. Since when?

From the Statistical literacy guide
How to spot spin and inappropriate use of statistics

http://www.parliament.uk/briefing-papers/SN04446.pdf


Thank-you