



Finding Stories in Data

<http://training.theodi.org/Malaysia>

David Tarrant · @davetaz

Improve understanding of how to source,
analyse and visualise data to discover insight
and tell stories.

Identify a number of data driven stories.

Understand the stages in creating an open data story.

Create your own open data story using analysis tools.

Visualise your findings using an interactive graphic.

Examples: Telling stories with data

The Upshot - Politics, Poli...
www.nytimes.com/upshot/

SECTIONS HOME SEARCH

The New York Times

LOG IN

EDITED BY DAVID LEONHARDT
FOLLOW US

The Upshot

Who Will Win the Senate? We give the Democrats a 55% chance of keeping a majority. [DETAILS](#)

Interest Rates Are Falling. Thank Vladimir Putin.



Risk aversion prompted by instability in Ukraine, among other things, has encouraged investors to seek safe havens like U.S. Treasury bonds.

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.



A.J. Mast for The New York Times

Women and the 'I Don't Know' Problem



Gavin Potenza

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

World Cup goals

1930-current

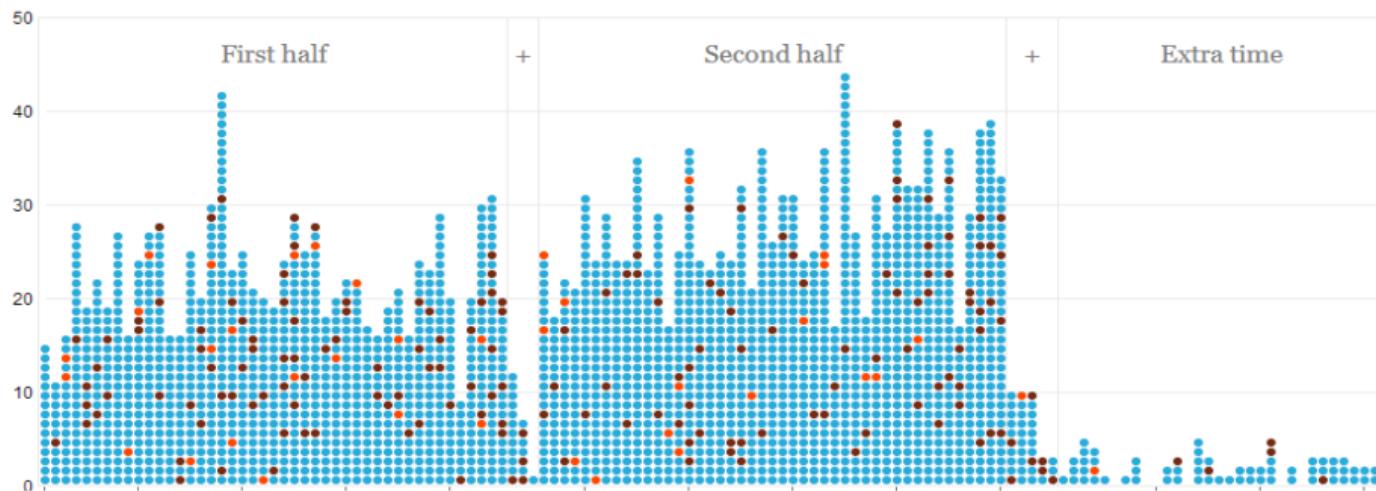
Filter country ▾

Filter year ▾

Filter stage ▾

Total goals: 2,379 of which: Penalties*: 174 | Own goals: 38

Average goals per game: 2.85



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

UK National Statistics

Parliament

Political groups

Businesses

Proactive

FOI requests

Surveys

'Ideas journalism'

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 Manage

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

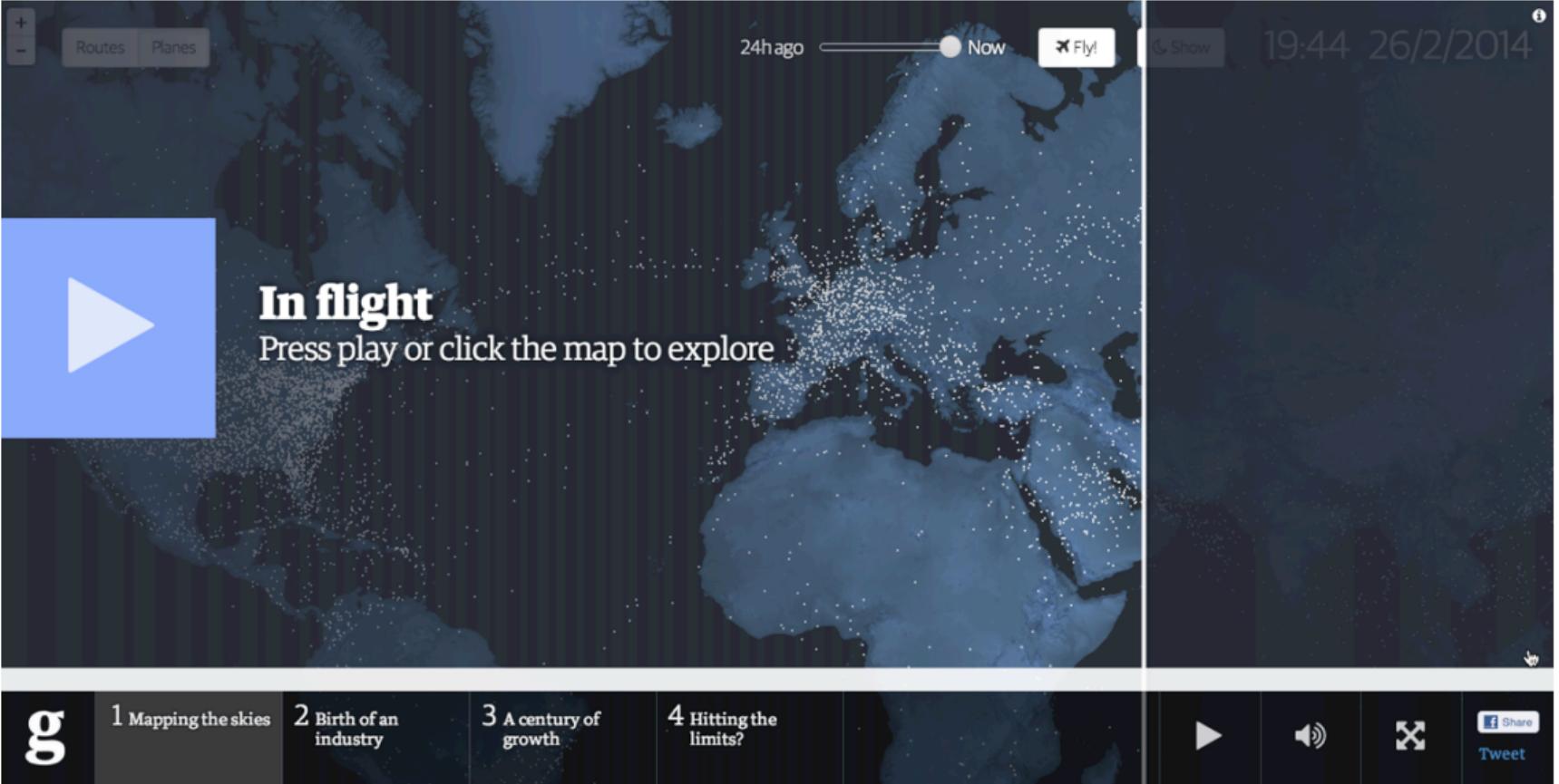
That's our estimation anyway. But we do use
our phones *a lot*.



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

Data for education



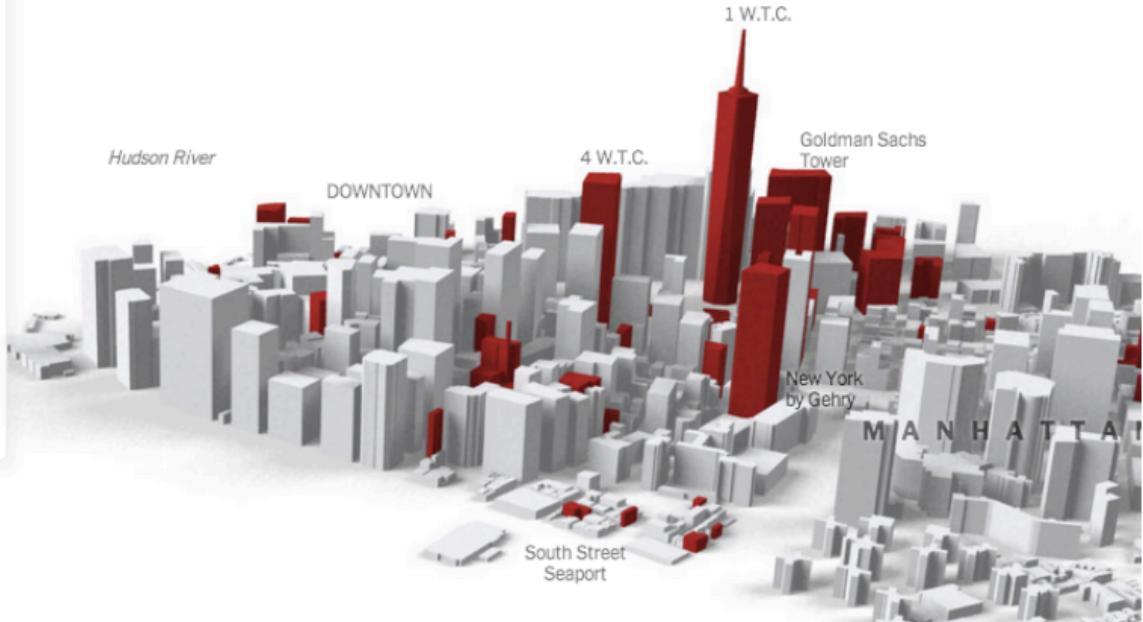


A Boom in New Housing

In spite of a recession and foreclosure crisis, the mayor presided over a boom in residential construction, encompassing everything from new towers 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.



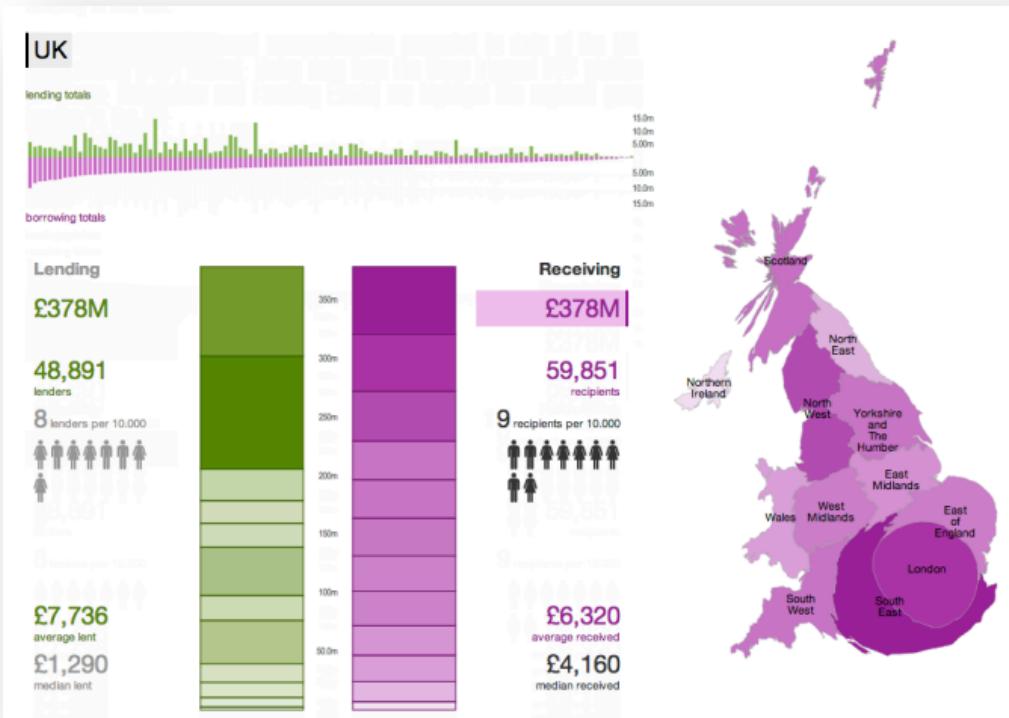
New buildings constructed during Mr. Bloomberg's tenure as mayor.



Self discovery



Show me the money



LFB Fire Station Closures



Data discovery patterns

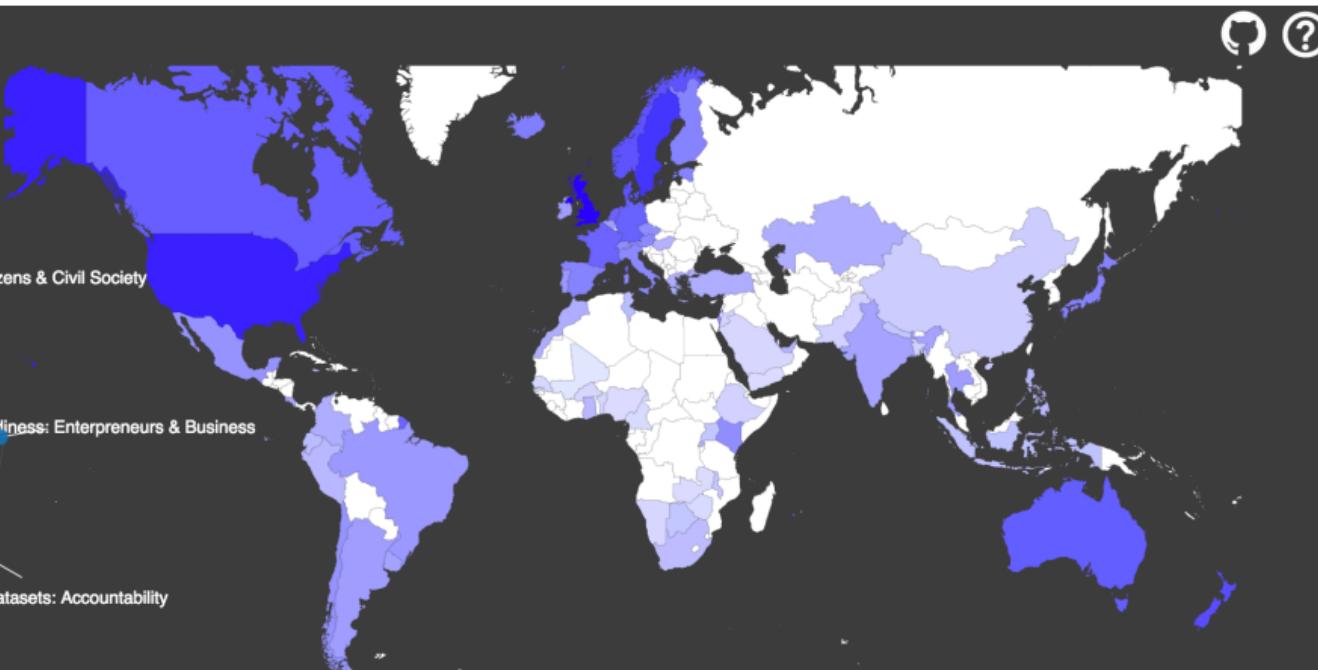
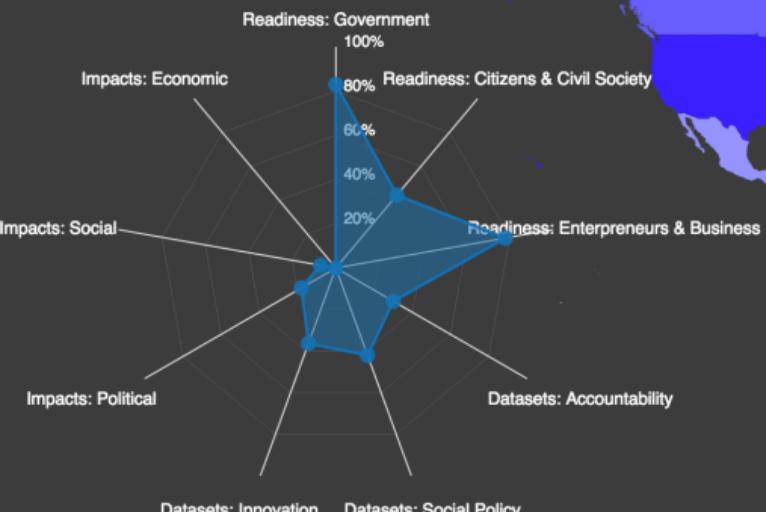
Finding data on the web (of documents)

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



Government data

Singapore **\$\$\$\$**
29_{/77}



[HOME](#)[ABOUT US](#)[CATALOGUE](#)[INFOGRAPHICS](#)[MOBILE APPS](#)[CONTACT](#)

DATASETS INFO

117 Datasets
11 Ministries
10 Sectors

RESOURCES

Malaysia Directory
Malaysian Open Source Centre
Open Government Data

MOBILE APPS

myHealth | myJakim | KPDNKK
SPAD | SELAWAT | myMAHTAS
6 Android Mobile Apps
4 iOS Mobile Apps

QUICK LINKS

Malaysia Informative Data Centre
Malaysian Population Quick Info
Tourism Malaysia - Facts & Figures

[Latest Datasets](#)[Top Publishers](#)[Feedback](#)

Aplikasi Pemetaan Belia Malaysia

- 1 Info belia berdasarkan domain
by : Ministry of Youth and Sports

Last Updated
2014-06-26 11:30:19

Aplikasi Pemetaan Belia Malaysia

- 2 Statistik Sikap Belia Malaysia mengikut
by : Ministry of Youth and Sports

Last Updated
2014-06-26 11:27:30

Latest News and Events

Currently, there are no latest event.

Google advanced

Google site:gov filetype:xls

Web Images Maps Shopping More Search tools

About 4,150,000 results (0.22 seconds)

[\[XLS\]](#) Code List or Concept (Acronym)

www.acquisition.gov/short_codelistsTS.xls Share

File Format: Microsoft Excel - [View as HTML](#)

A, B, C, D, F, G, H, 1, Code List or Concept (Acronym), Definition, Authoritative or Maintenance Agency, Current Version, Update Frequency and/or ...

[\[XLS\]](#) Approps - Foreign Assistance.gov

www.foreignassistance.gov/Full_ForeignAssistanceData.xls

File Format: Microsoft Excel

A, B, C, D, E, F, G, H, 1, Planning Data, 2, 3, 4, Fiscal Year, Fiscal Year Type, Account Name, Agency Name, Operating Unit, Category, Sector, Amount ...

[\[XLS\]](#) TSB Monthly Cash Flow Projection

www.dia.iowa.gov/tsb/cashflow.xls

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.

The screenshot shows the homepage of enigma.io. At the top, there is a navigation bar with links for SOLUTIONS, PLANS, ABOUT, PRESS, CAREERS, Create Account, and Sign In. Below the navigation bar is a large, detailed historical map of London. Overlaid on the map is the text "Navigate the world of public data". Below the map is a search bar with the placeholder "Companies, people, locations..." and a blue "SEARCH" button. At the bottom of the page, there is a call-to-action: "Quickly search and analyze billions of public records published by governments, companies and organizations."

enigma.io

The screenshot shows the homepage of transportAPI. At the top, there is a header with the transportAPI logo and the tagline "powering change and innovation in transport". Below the header is a navigation bar with links for HOME, SOLUTIONS, DEVELOPERS, TEAM, PARTNERS, CONTACT, and BLOG. The main content area features a large image of a mobile device displaying a map application with traffic information. To the right of the map is a sidebar titled "Blog" which includes links to "Tree Routes iPhone app", "Mind The Gap Dentistry and Transport Data", "January & February Events", "Use TransportAPI and take part in the Ford Traffic Tamer App Challenge for London", and "Help Your Fellow Commuters Navigate The London Tube Strike".

transportAPI



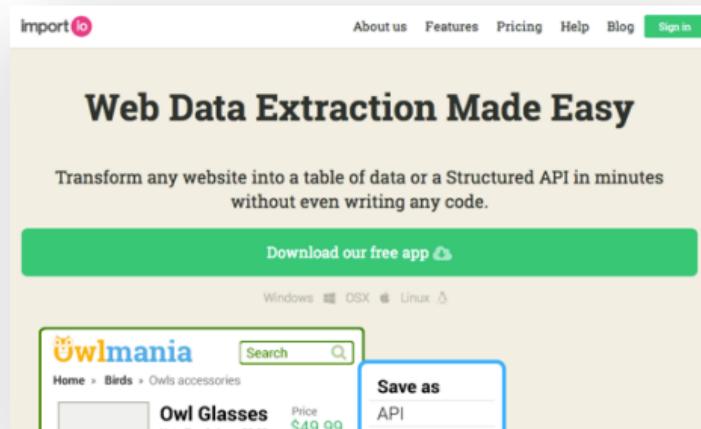
Scraping

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



The ScraperWiki homepage features a blue header with the logo 'ScraperWiki' and a yellow bee icon. Below the header, there are two main sections: 'Scrape Twitter' and 'Extract PDFs'. The 'Scrape Twitter' section includes a Twitter icon, a brief description, and a 'Try our Twitter tools' button. The 'Extract PDFs' section includes a document icon, a brief description, and a 'Try Table Xtract' button.

scraperwiki.com



The import.io homepage has a white header with the logo 'import.io' and navigation links for 'About us', 'Features', 'Pricing', 'Help', 'Blog', and 'Sign in'. The main content area features the heading 'Web Data Extraction Made Easy' and a subtext about transforming websites into tables or APIs. It includes a green 'Download our free app' button and a screenshot of the software interface showing a search bar and a 'Save as API' button over a product listing for 'Owl Glasses'.

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)

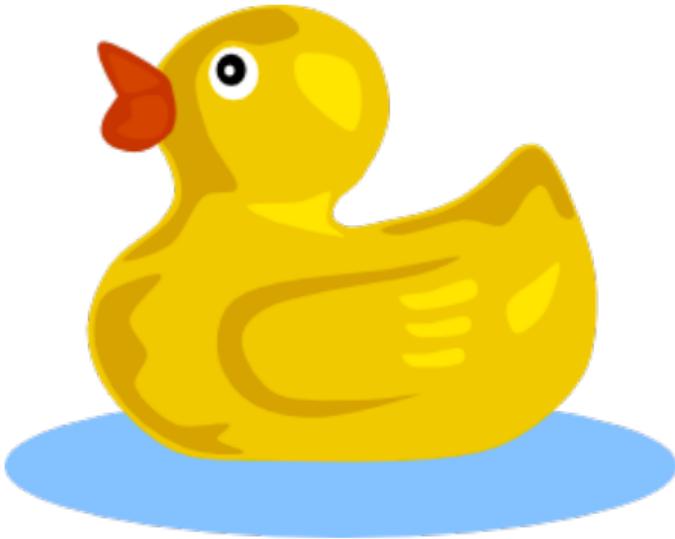


How the web should work,
but people forgot that Tim
put this in when he
invented it!

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

GOV.UK

Search

Home > Business and self-employed > Imports and exports

Trade Tariff

Search the tariff name or code

This tariff is for 6 August 2014 [change date](#)

View all sections [A-Z Index](#)

Trade between the UK and All countries [change country](#)

Section	Chapters	Title
I	1 to 5	Live animals; animal products
II	6 to 14	Vegetable products
III	15	Animal or vegetable fats and oils and their cleavage products; prepared edible fats; animal or vegetable waxes
IV	16 to 24	Prepared foodstuffs; beverages, spirits and vinegar; tobacco and manufactured tobacco substitutes
V	25 to 27	Mineral products
VI	28 to 38	Products of the chemical or allied industries
VII	39 to 40	Plastics and articles thereof; rubber and articles thereof
VIII	41 to 43	Raw hides and skins; leather, furskins and articles thereof; saddlery and harness; travel goods, handbags and similar containers; articles of animal gut (other than silkworm gut)
IX	44 to 46	Wood and articles of wood; wood charcoal; cork and articles of cork; manufactures of straw, of esparto or of other plaiting materials; basket-ware and wickerwork

UK Trade Tariff

one

DOCTOR WHO

Home Episodes Clips Galleries Latest News Characters Monsters Fun and Games More

It's Tomorrow... Get the Latest on the Launch!

This programme will be available shortly after broadcast

The Day of the Doctor

SATURDAY 19:00
BBC THREE
All upcoming
(0 NEW AND 1 REPEAT)

BBC Music and Programmes

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



2. Look for alternative links



Business Insight - NEWSASIA NEWS TV WATCH LIVE

Wed, Aug 06 2014

ASIA PACIFIC | SINGAPORE | WORLD | BUSINESS | SPORT | ENTERTAINMENT | TECHNOLOGY | HEALTH | LIFESTYLE | VIDEOS | WEATHER | MORE ▾

CHANGINGLIVES LUMINARY AWARDS START-UP

SCROLL DOWN!

SINGAPORE STORIES

Raise of up to 12% for Home Team officers, with sign-on bonuses of up to S\$30,000

National Day Award 2014

SP (2) Ng (SP) [REDACTED]

MEDIA CORP

Officers from the Home Team, both past and present were recognised at the Home Team National Day Observance Ceremony on Wednesday.

9 hours ago

Pay rise, special bonus for about 23,000 nurses

10 hours ago

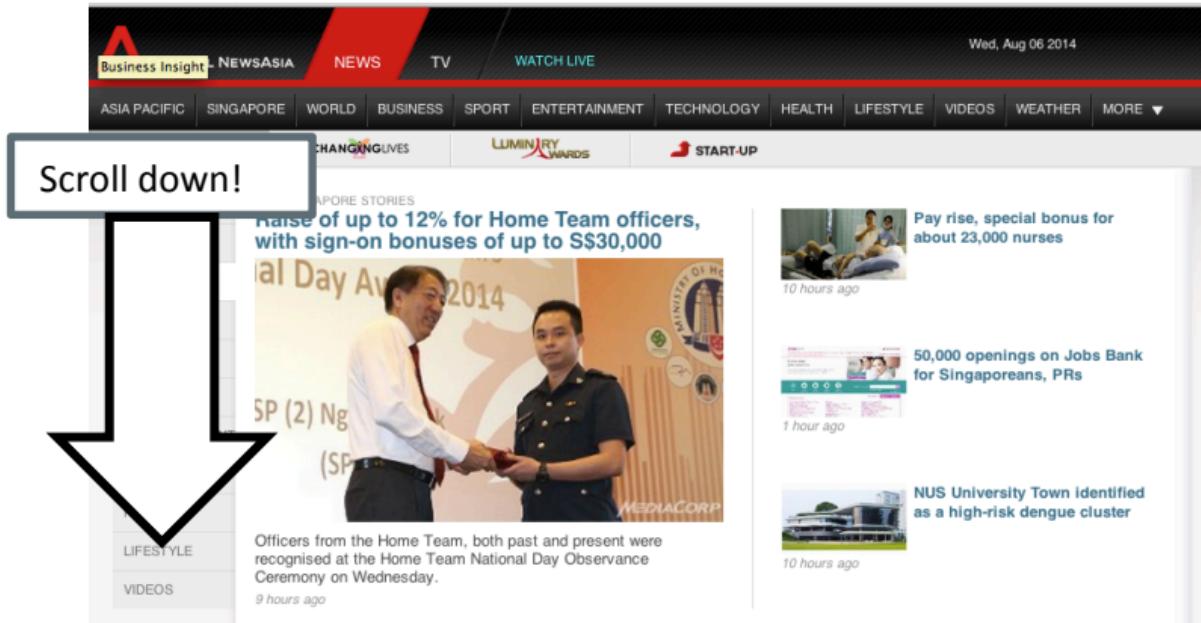
50,000 openings on Jobs Bank for Singaporeans, PRs

1 hour ago

NUS University Town identified as a high-risk dengue cluster

10 hours ago

LIFESTYLE VIDEOS



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

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

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



Exercise: Find a story

<http://bit.ly/odi-stories3>



[http://en.wikipedia.org/wiki/
File:Stöwer_Titanic.jpg](http://en.wikipedia.org/wiki/File:Stöwer_Titanic.jpg)

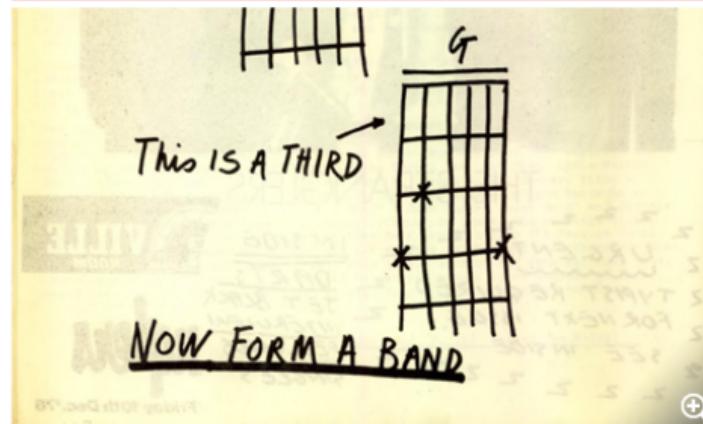
Exercise

1. Go through (some of) the Checklist for Exploring Data
2. Create a pivot table
3. Five ideas for finding a story
 - Choose **one story** you want to tell
 - Create a headline
 - (Bonus: Create a chart, e.g. with *datawrapper.de*)

Anyone can do it. Data journalism is the new punk

Can anyone be a data journalist? **Simon Rogers** on what we can learn from a 1977 diagram

- Another view: [What data can and cannot do](#) by Jonathan Gray



Page two of [Sideburns](#), January 1977

● This is a chord... this is another... this is a third. NOW FORM A BAND



Posted by
Simon Rogers
Thursday 24 May 2012
13.00 BST
[theguardian.com](#)
[Jump to comments \(8\)](#)



[Article history](#)

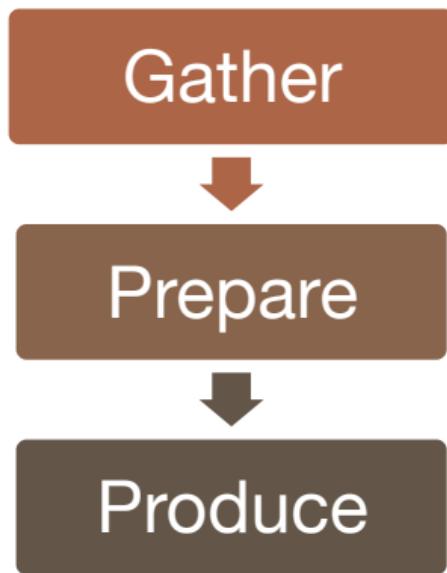
Media

Data journalism · Open journalism

Technology

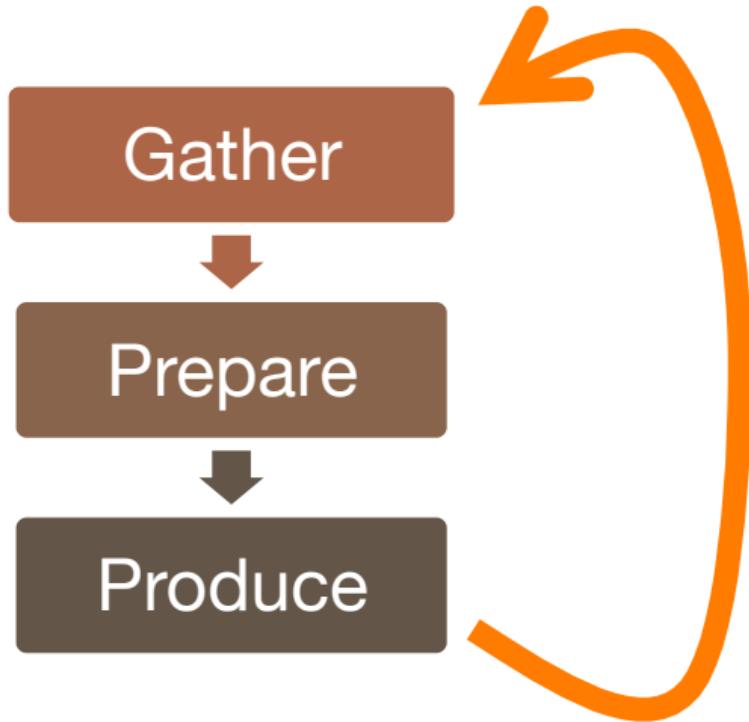
The data percolator

Data percolation: A model of data preparation and analysis

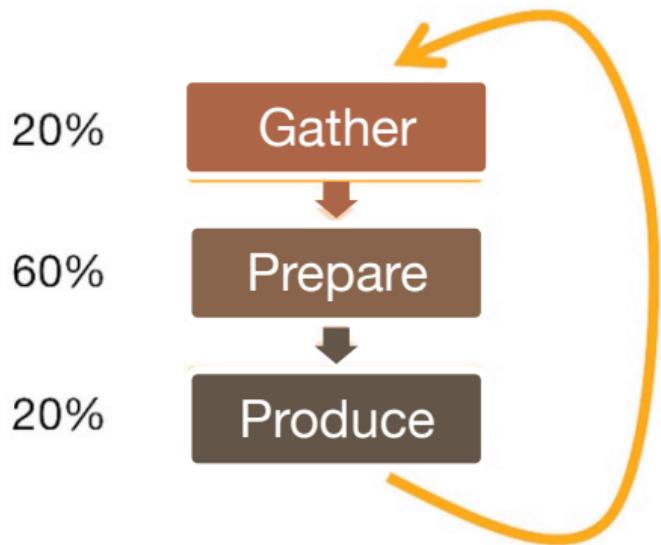


See also: the Data Journalism Handbook

Data percolation: A model of data preparation and analysis

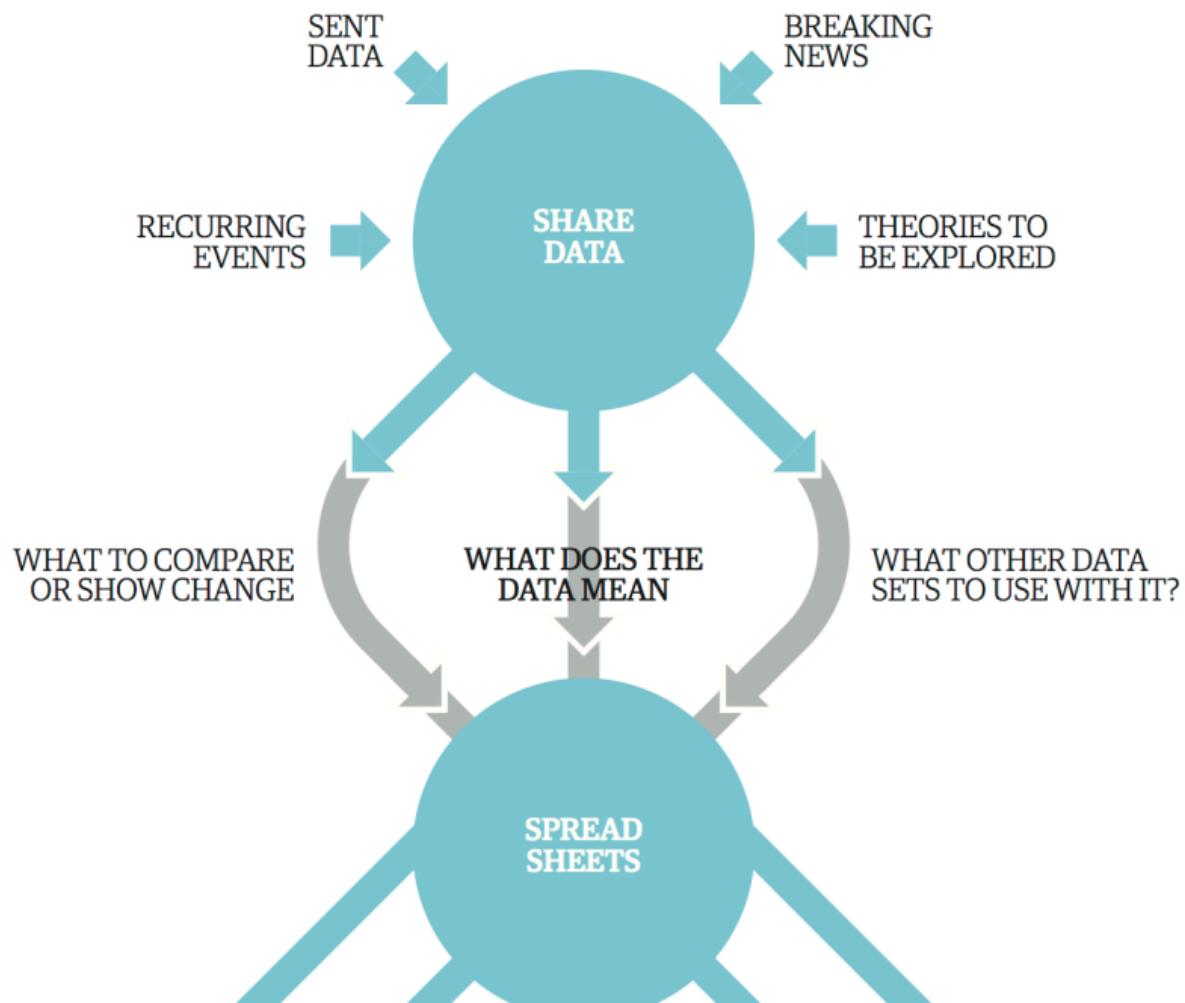


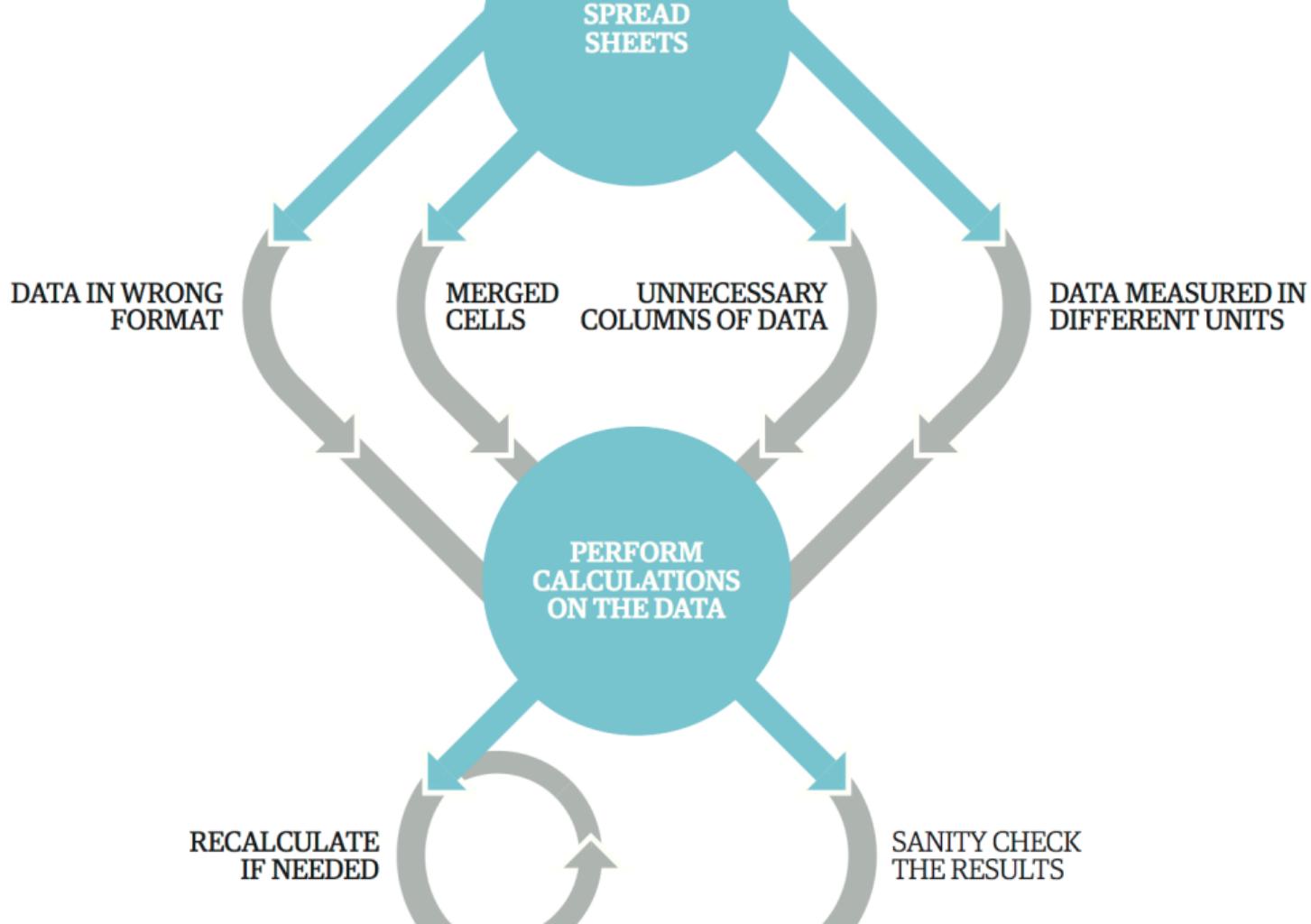
How should I budget my time?

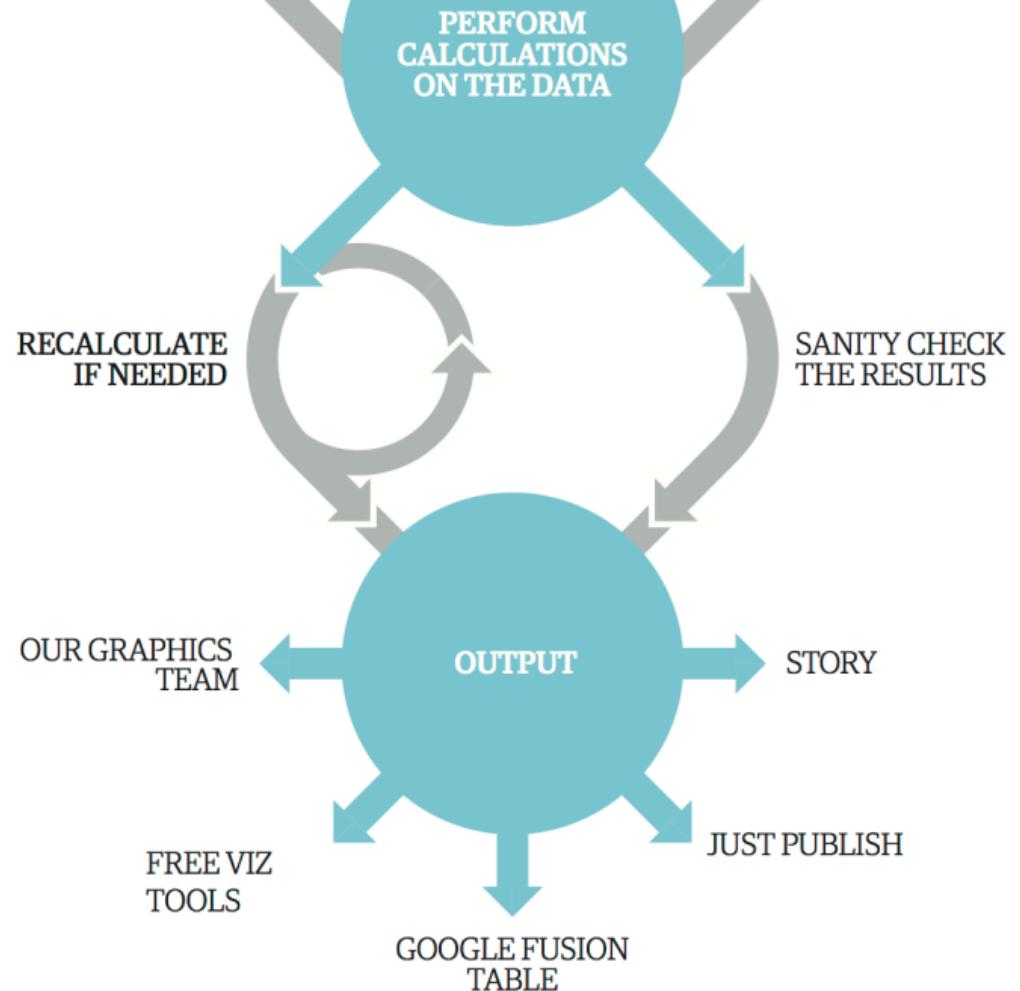


Finding a story is a creative process.

Let it percolate!







How should I budget my time?



- 1.1 **FIND** reliable data sources
- 1.2 Understand your **RIGHTS**
- 1.3 Visualise and **UNDERSTAND** your data
- 2.1 **CLEAN** your data
- 2.2 **TRANSFORM** it where useful
- 2.3 **COMBINE** it with other data sets
- 3.1 **REDUCE** and find the story
- 3.2 Think and understand the **CONTEXT**
- 3.3 Do your results pass a **SENSE-CHECK**?

Time planning

Gather

Produce

Prepare

2.1 CLEAN

2.2 TRANSFORM

2.3 COMBINE

2.4 ENRICH

2.5 ANALYSE

Tools

Excel

Refine

Datawapper.ie

Plot.ly

CartoDB

d3.js



Exercises

Choose one or more of the following to try:

- Data cleaning in Refine
- Enriching data to discover insight (in Refine)
- Creating interactive graphics in Plot.ly
- Creating interactive maps in cartoDB.
- Building your own interactive website and d3.js visualisation.





Thank You

<http://training.theodi.org/Malaysia>

David Tarrant · @davetaz