# Data Journalism

How do you get started?



By Marianne Bouchart

@Maid\_Marianne @GENinnovate
mbouchart@globaleditorsnetwork.org

## Introduction

Journalists have to acquire new skills to keep up with the amount of information now available online.

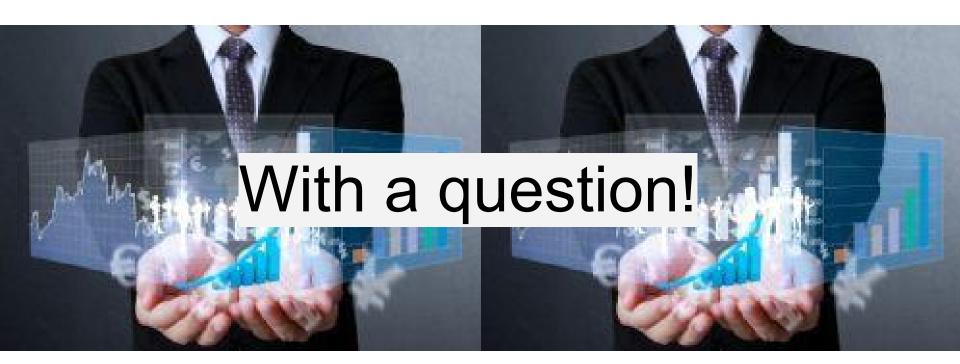
Being able to work with large datasets is one of them.



# First: How do I start?



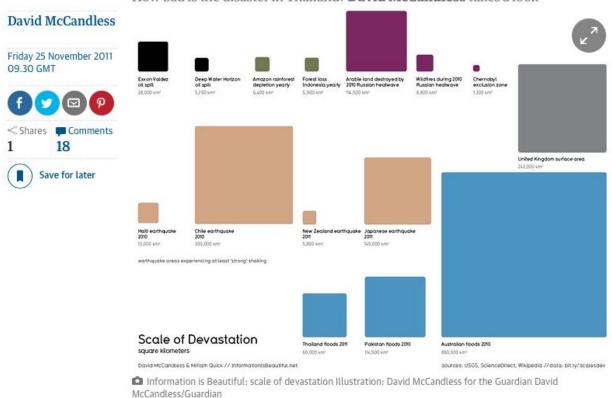
### First: How do I start?



# **Thailand** Datablog

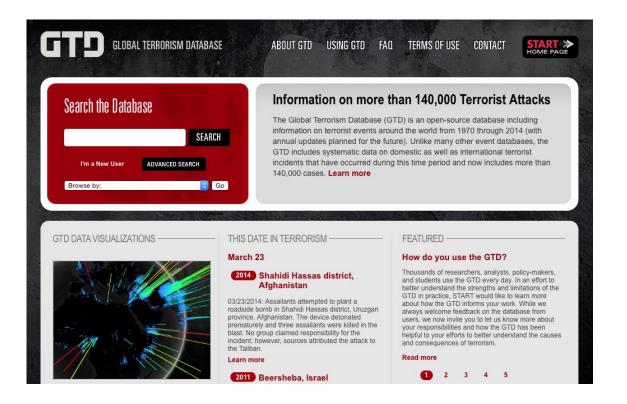
### Just how big are the Thailand floods?

How bad is the disaster in Thailand? David McCandless takes a look





### Global Terrorism Database



### RAND Database of Worldwide Terrorism Incidents



### **RDWTI Home**

### About RDWTI

**Database Scope** 

Research Coverage

Contact Us

### Search RDWTI

Search Instructions

Downloads

### **Research Publications**

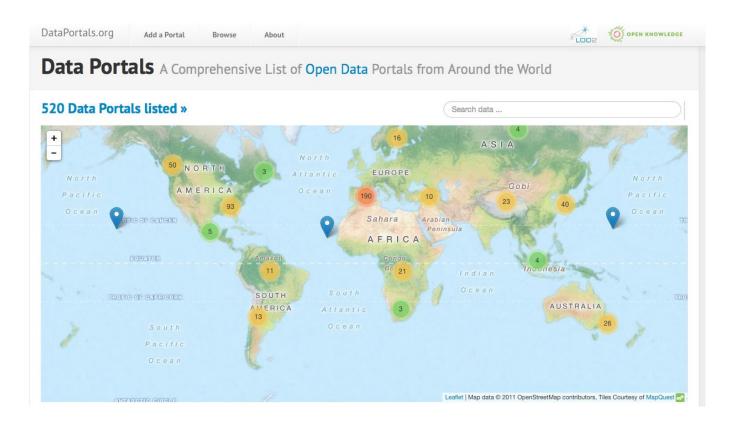
RAND > NSRD >

# RAND Database of Worldwide Terrorism Incidents

The RAND Database of Worldwide Terrorism Incidents (RDWTI) is a compilation of data from 1968 through 2009.

For 40 years, the RAND Corporation has been at the forefront of terrorism and counter-terrorism studies. In support of this work, RAND has developed and maintained a database of terrorism incidents stretching back to 1968, which provides comprehensive information on international and domestic terrorism. Over the years, many public and private sponsors have contributed to the maintenance of the RDWTI and its predecessors, the RAND Terrorism Chronology and the RAND-MIPT Terrorism Incident Database.

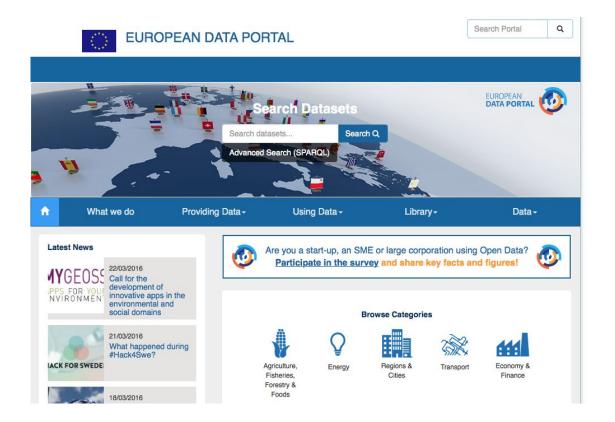
# DataPortals.org



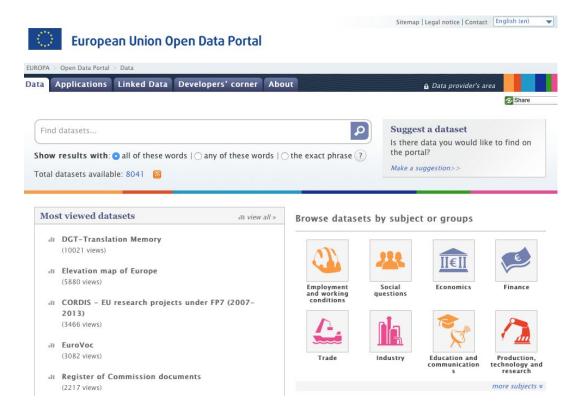
### FindTheData.com



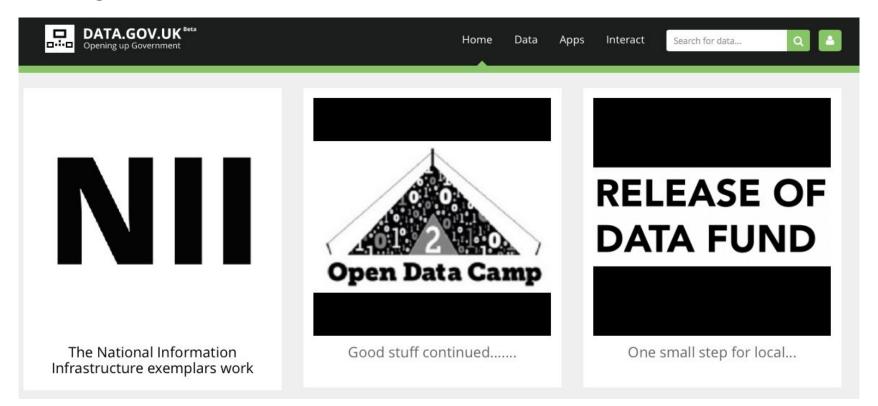
### **EU Data Portal**



# European Union Open Data Portal



# Data.gov.uk



# Data.gov



# Open Corporates



The largest open database of companies in the world



Browse all jurisidictions Search officers

"This is the kind of resource the (digital) single market needs"

Neelie Kroes, VP European Commission

### Wikileaks website



### Hillary Clinton Email Archive

A searchable archive for 30,322 emails & email attachments sent to and from Hillary Clinton's private email server while she was Secretary of State.

16 March 2016



### NSA targets world leaders for US geopolitical interests

NSA bugging: UN Secretary General Ban Ki-Moon's talks with German Chancellor Angela Merkel, conversation of Israel prime minister Netanyahu with Italian prime minister Berlusconi +more

23 February 2016



### EU military ops against refugee flows

Classified report about the first six month of Operation SOPHIA (EUNAVFOR MED), the EU military intervention against "refugee boats" in Libya and the Mediterranean.

17 February 2016



### The New Dirty War for Africa's uranium and mineral rights

A collection of documents that open up a corrupt multi-billion dollar war by Western and Chinese companies grab uranium and other mining rights in the Central African Republic (CAR) and escape paying for the environmental consequences.

5 February 2016



### Trade in Services Agreement

Secret documents relating to the secret Trade in Services Agreement, the largest 'trade deal' in history. TiSA is being negotiated by the US, EU and 22 other countries that account for two-thirds of global GDP.



### Saudi Cables

Over half a million searchable cables and other documents from the Saudi Foreign Ministry including secret communications from various Saudi Embassies around the world.



### Trans-Pacific Partnership Agreement

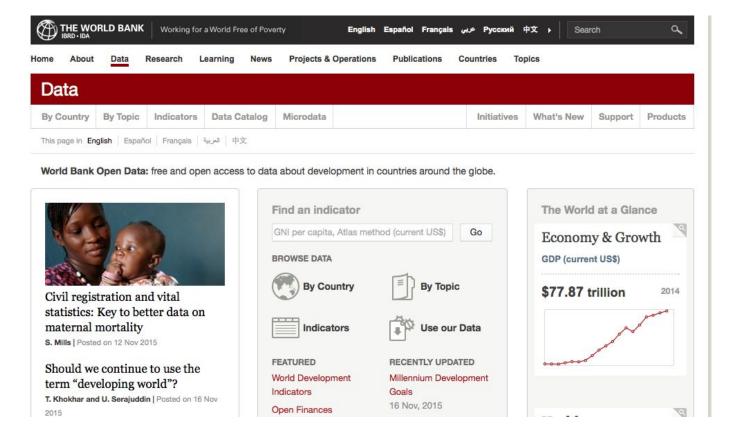
Documents relating to the Trans-Pacific Partnership Agreement during its negotiation period. The TPP encompasses 12 nations and over 40 percent of global GDP.



### SourceAmerica Tapes

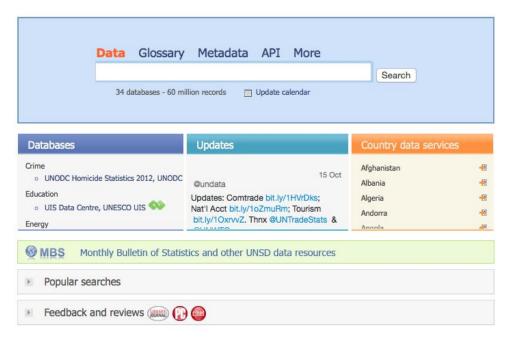
Thirty hours of secret tapes and transcripts, 13 Presidential Appointees embroiled in a multibillion dollar corruption scandal being probed by the US Dept. of Justice and four Inspector Generals.

### The World Bank

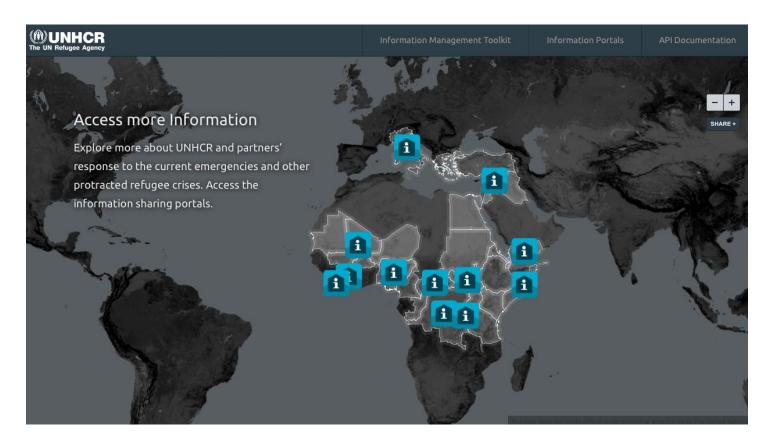


## The UN Data Portal

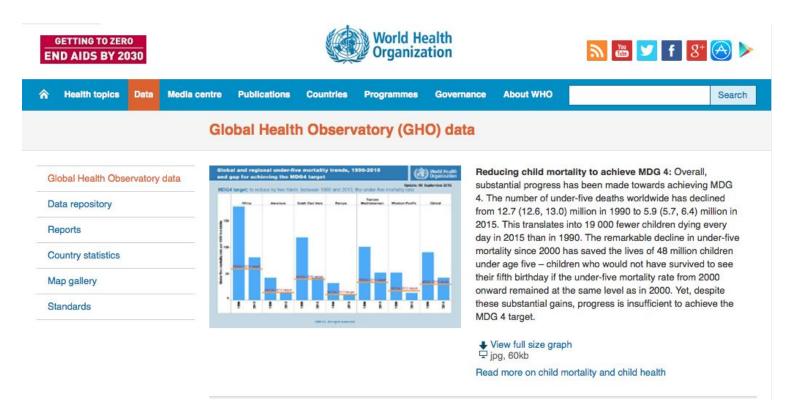




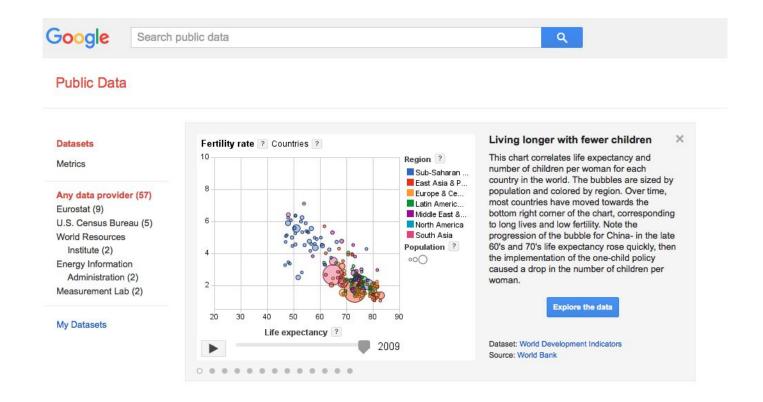
## The UNHCR Data Portal



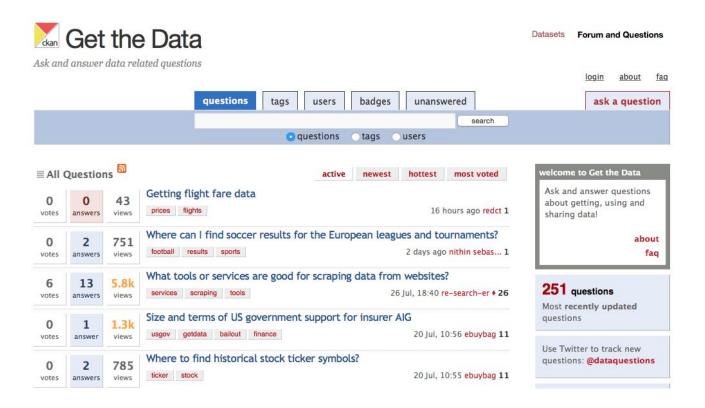
# The World Health Organization Data



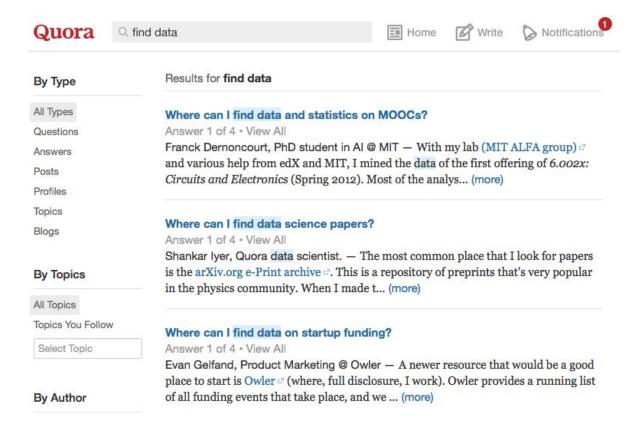
# Google Public Data Explorer



# GetTheData.org



### Quora



# Crowdsourcing using Google Forms

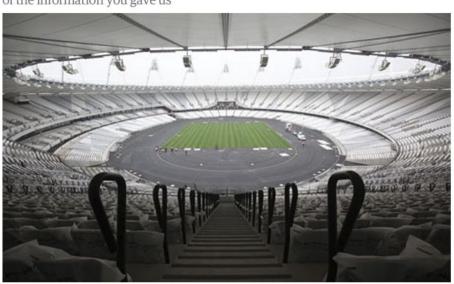
# Olympic tickets

London 2012 Olympics data

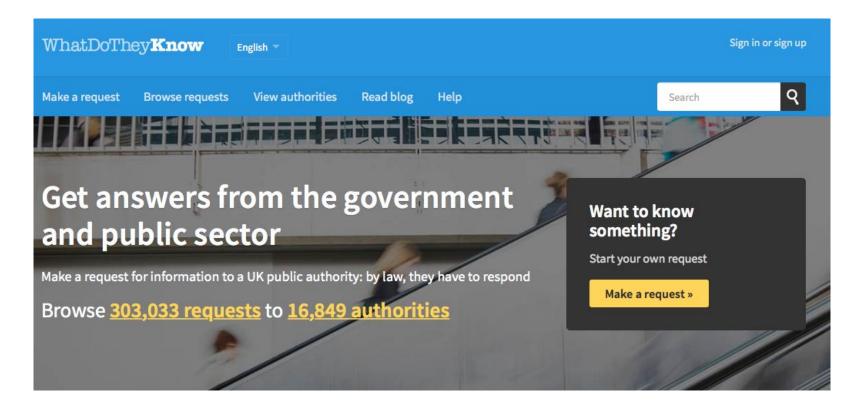
### How many Olympic tickets did you get? Here's our readers' results

We asked how you had fared in the London 2012 ticket ballot. Here's our analysis of the information you gave us





# WhatDoTheyKnow.com



# Find datasets directly on Google

### Google search operators:

- filetype:CSV and filetype:XLS for Excel spreadsheets
- filetype:shp for geo data
- filetype:MDB, filetype:SQL, filetype:DB for database extracts
- if you're so inclined, you can even look for filetype:pdf
- 'inurl:downloads filetype:xls
- site:agency.gov

### example:

site:adidas-group.com filetype:pdf

# Data Scraping with Google

One line magic formula in Google Spreadsheet to scrape data from HTML tables:

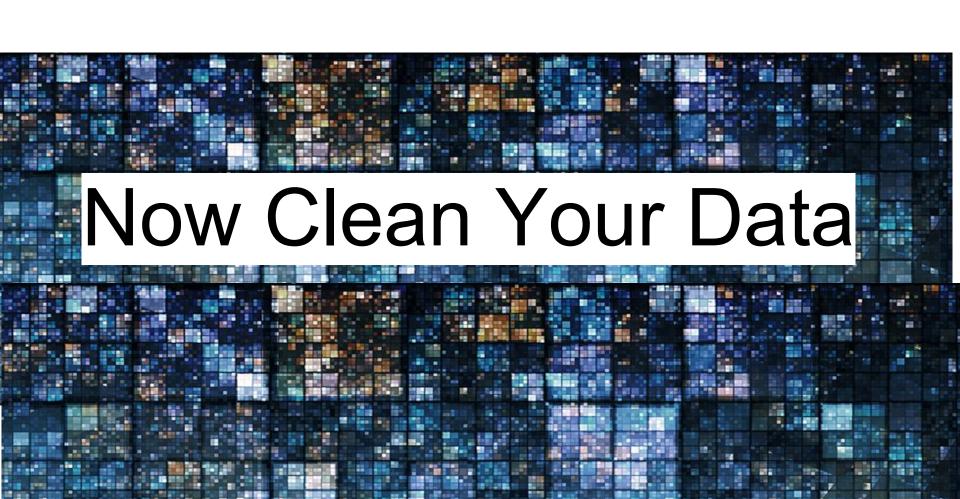
=importHTML("","table",N)

### example:

=ImportHtml("http://en.wikipedia.org/wiki/List\_of\_largest\_United\_Kingdom\_settlements\_by\_population", "table", 1)

### result:

https://docs.google.com/spreadsheets/d/1fPu-3wNjVnyB\_zF9x72u6X811exIQVxYzaUtdMxderk/edit#gid=0&vpid=A1



# Open Refine



### Welcome!

**Home** 

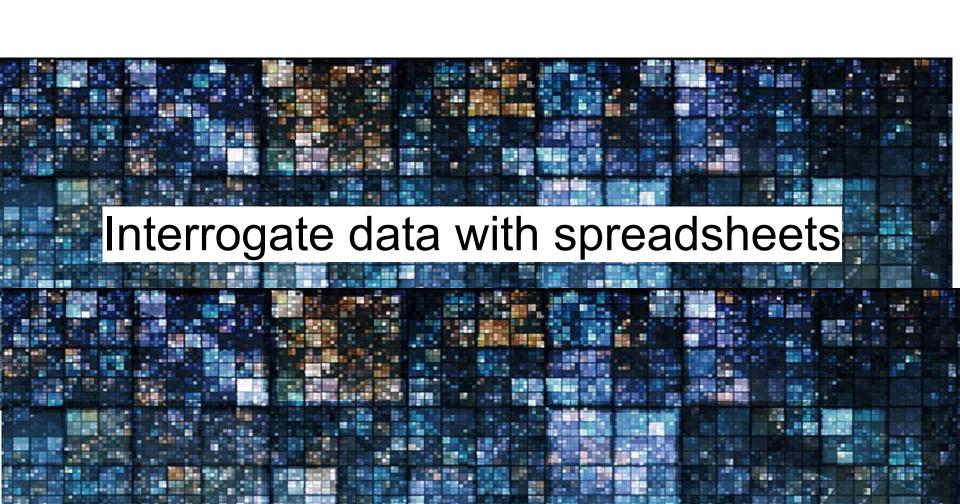
**Download** 

**Documentation** 

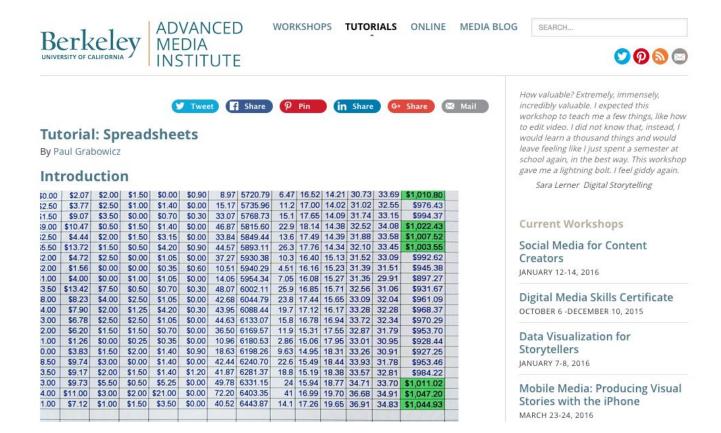
OpenRefine (formerly Google Refine) is a powerful tool for working with messy data: cleaning it; transforming it from one format into another; and extending it with web services and external data.

Please note that since October 2nd, 2012, Google is not actively supporting this project, which has now been rebranded to OpenRefine. Project development, documentation and promotion is now fully supported by volunteers. Find out more about the history of OpenRefine and how you can help the community.

Using Open Refine by Ruben Verborgh, Max De Wilde can be found <u>here</u> DataDrivenJournalism.net tutorial for Google Refine can be found <u>here</u>



# Berkeley's tutorial on Spreadsheets



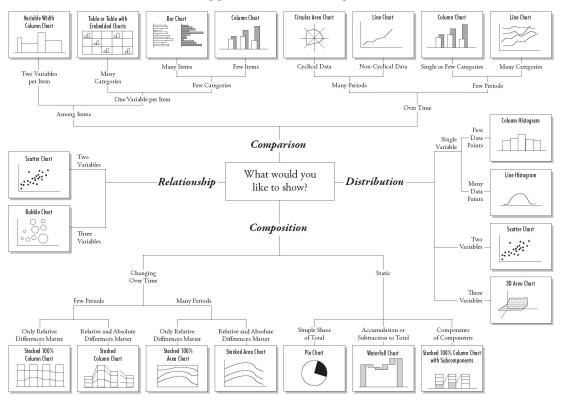
### CIJ's Data Journalism handbook



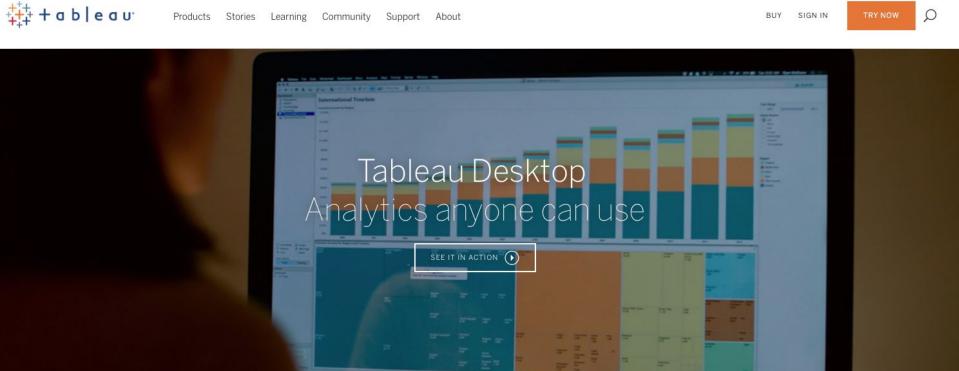
# How do I visualise data?

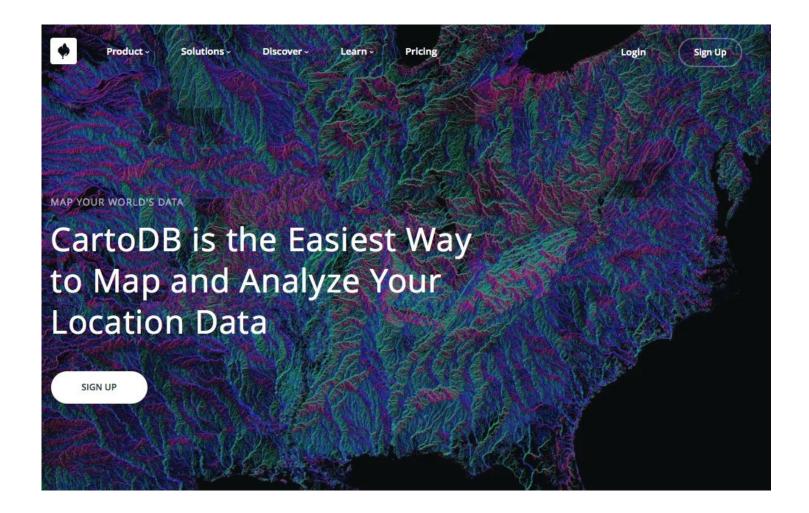


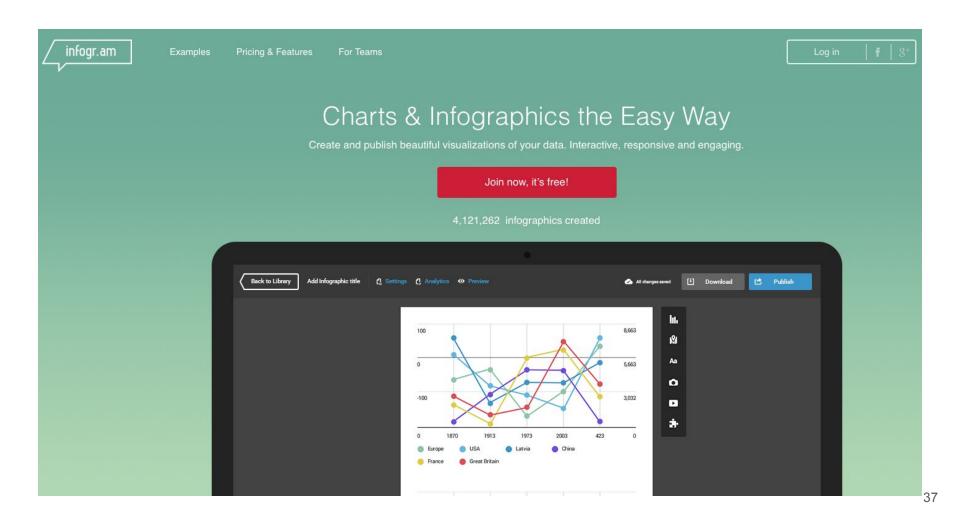
### Chart Suggestions—A Thought-Starter

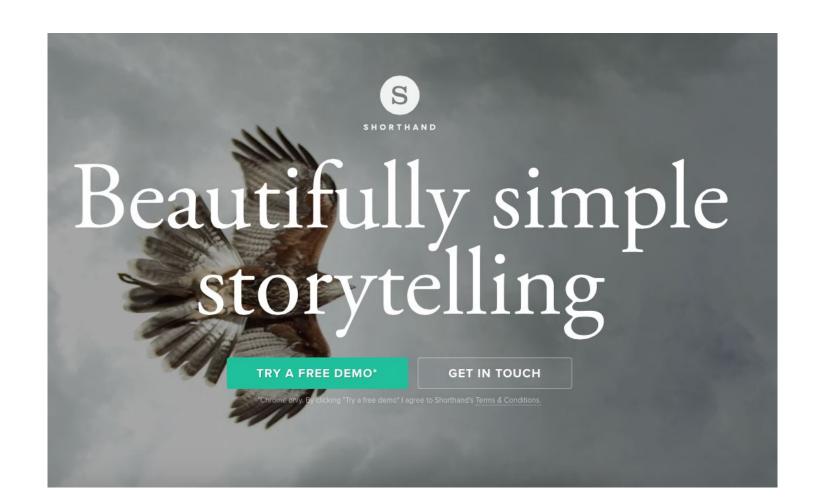


© 2006 A. Abela — a.v.abela@gmail.com















Course

About News & Analysis Events Featured Projects Resources Mailinglist

# Doing Journalism with Data: First Steps, Skills and Tools

A free online data journalism course with 5 leading experts.

View course information



Research

Admissions News & events People Careers, alumni & internships

### MOOC on data journalism opens for registration

Published by & James Hill at O March 2, 2016



# FUNDAMENTALS

The JMSC is offering a five-week course on the fundamentals of data journalism starting 4 April, in partnership with the Google News Lab and top news organisations in Asia.

Taught by a panel of experts on data journalism from New York City, San Francisco, Bangkok, Hong Kong and Shanghai, the course will introduce the latest tools in data acquisition, analysis, interpretation and visualisation. Students will also learn how to integrate these skills to tell compelling stories.

Regional newsroom partners for the course include IndiaSpend (India), Malaysiakini (Malaysia), KataData (Indonesia), Caixin (China) and DataLEADS (India).

Targeting journalists of all levels of experience, students and anyone with an interest in data journalism, the course will also feature case studies from newsrooms in Asia and data stories by leading international media organisations.





### Sign up to the DJA Newsletter at bit.ly/djaNL



### March Newsletter

Welcome to the Data Journalism Awards newsletter! You're receiving this because you've opted to get updates about THE global competition for data journalism.

This newsietter is curated by <u>Marianne Bouchart</u>, DJA manager working with <u>Simon Raggers</u>, DJA director, and <u>Paul Stelger</u>, president of the jury, it is sent twice a month with our pick of the latest in data journalism and visualisation.



The <u>Data Journalism Awards</u> is the first international competition recognising outstanding work in the field of data journalism worldwide. It is organised by the <u>Global Editors Network</u> and sponsored by the <u>Global Editors Network</u> and sponsored by the <u>Global Editors Network</u> and sponsored by the <u>Global Editors Network</u> and the <u>John S. and James L. Knight Foundation</u>. This year's competition is now on and you have until 10 April 2016 to submit a project via the <u>DJA 2016 website</u>.

As any awards competition go, you need to be in it to win it. So whether you are a large news organisation or a small startup, if you have created a compelling data visualisation, an interactive map or any data-driven project that occes innovation and flair, come and submit your work on the GEN Community platform.

Prizes for the DJA 2016 are worth 1,000€ each and will be awarded at a special ceremony at the Vienna City Hall during the fifth annual GEN Summit on 16 June 2016.

You can find all the information about this year's awards (including key dates, categories and rules) on the DJA 2016 homepage.

### Five things we liked this month

Each day brings a new data journalism or visualisation project. Here are some of our recent favourites:

### Where Trump could lose, and other things to watch for on Super Tuesday



By Ted Melinik, Kevin Schaul, Kevin Uhrmacher and Samuel Granados Publication: The Washington Post Country: U.S.A.

Nearly 600 delegates are up for grabs on Super Tuesday — the most for any day of the 2016 primary season. The bulk of them come from seven Southern contests, including the day's biggest prize: Texas. In this <u>data-driven article</u>, the team from the <u>Washington Post</u> lets you explore the data.

# Racism: The terror of the other

By Kai Biermann, Philip Faigle, Astrid Geisler, Karsten Poike-Majewski, Hannes Soltau, Julian Stahnke, Tilman Steffen and Sascha Venohr Publication: Zeit Online Country: Germany

There were many attacks on refugee shelters across Germany in the past year. 279, to be more precise, according to this investigation from Zeit Online. In this compelling series on racism, the team of editors investigated on these attacks to learn about their perpetrators.



By German Lopez Publication: Vox.com Country: U.S.A.

Among developed nations, the US is far and away the most violent — in large part due to the easy access many Americans have to firearms. This data-driven article shows what that violence looks like compared with the rest of the world, why it happens, and why it's such a tough problem to fix, using maps and graphs.



By Carla Pedret Publication: The Online Journalism Blog Country: U.K.

Podcasts are a great way to listen to stories on the move, be entertained, or keep up with developments in a particular field. However, have you ever thought about using them to learn data journalism? In <a href="mailto:the">this article</a>, Carla Pedret pulled together some of the best podcasts about data.



# Thank You!

@Maid\_Marianne @GENinnovate
mbouchart@globaleditorsnetwork.org