

**Creating big data  
journalism impact with  
tiny resources**

# WalesOnline Datastore

**Wales**Online

14°C  
Sunny  
5 day forecast »



HOME

NEWS

SPORT

WHAT'S ON

LIFESTYLE

IN YOUR AREA

BUY, SELL & TELL

Search



Hot Topic: Karina Menzies - British & Irish Lions - Cardiff 10K - Carwyn Jones - Go Green

Follow Us: Facebook

Twitter

Newsletters



## Behind The Numbers

'Behind The Numbers' is all about the explosion in 'data': information about our hospitals and schools, crime and the way it is policed, business and sport, arts and culture.

Like 0

Tweet 1

How healthy are you?  
Find out how you  
compare to the rest  
of Wales with our  
interactive quiz »

Health Check Wales 06 Jun 2013



## Business Events Wales »

### Forthcoming Events

Letters of Credit  
Orion Suite, Newport  
Newport Meet & Greet  
Roughmound School, SA  
A Foundation Course  
Cooper Court, Swans

Listings for all  
business,  
networking and  
hospitality events  
in the Principality

- Using statistics to find stories for print and online
- Trying to present them in ways that are accessible

# Tools for data journalism:

- Excel or Open Office Calc
- PDF reader and a program to open word documents.
- Datasets
- A place to write stories

# Tools for data journalism:

- Excel or Open Office Calc
- PDF reader and a program to open word documents.
- Datasets
- A place to write stories

You may also need:

- A Google account
- Something to visualise data
  - with – Google Charts, Google Fusion Tables, Tableau, Infogr.am, Openheatmap etc.
- Google Refine

# Sources of data

- Government statistics
- Written answers
- Searching
- FOI
- Crowdsourcing

The screenshot shows the DATA.GOV.UK website interface. At the top, there is a navigation bar with the HM Government logo and a 'Log in or sign up' link. Below this is a green header with the DATA.GOV.UK logo and the tagline 'Opening up Government'. A secondary navigation bar contains links for Home, Data, Participate, Apps, Location, Linked Data, Library, Lab, and About. The main content area features a carousel of featured items: 'WEATHER DATA DOWNLOAD' (Over 5000 UK sites), 'Data driven enterprises' (Blog by Professor Brian Collins), 'Better Open Data' (Blog by Tim Davies), 'OPEN DATA CASE STUDIES', and 'OPEN DATA WHITE PAPER' (Unleashing the potential, Read the UK Government's vision for the future of open data).

National Statistics releases for  
17 August 2012

#### Turnover and Orders in Production and Services Industries - June 2012

Release time: 9:30am  
Business and Energy | Office for National Statistics | GB  
Monthly Turnover, Exports, and Orders data for Production and Services industries.



#### Statistical Press Release - Mortgages: Actions for Possession - April - June 2012

Release time: 9:30am  
Crime and Justice | Northern Ireland Statistics and Research Agency | Northern Ireland  
This statistical press release provides statistics for writs and originating summonses issued, cases disposed and orders made in respect of mortgages in the Chancery Division of the Northern Ireland High Court.



#### Key Welsh Economic Statistics - August 2012

Release time: 9:30am  
Economy | Welsh Government | Wales  
Presents a summary of the latest economic and labour market statistics for Wales.



#### Drug-related deaths in Scotland - 2011

Release time: 9:30am  
Health and Social Care | National Records of Scotland | Scotland  
The number of, and trends in, drug-related deaths in Scotland, broken down by, for example, age, sex, Health Board and Council areas and whether particular types of drug were found in the body.



The key to getting to grips with  
data  
journalism is remembering it is  
just like all other types of  
journalism...

What is the story?

# Things that might make a story:

- o The biggest, the highest, the most
- o The smallest, the fewest
- o Fastest growing
- o Location
- o Successes
- o Failures
- o Why?

# It's all about the headline

[TITLE: Unemployment - September 2013

DATE: Wednesday, September 11

COVERAGE: UK

OF POTENTIAL INTEREST TO: all newsrooms

PREPARED BY: Claire Miller (02920243614, [claire.miller@mediawales.co.uk](mailto:claire.miller@mediawales.co.uk))

URL:

<http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-276583>

## Headlines

In the three months to July:

Wales - Unemployment fell 7,000 compared to the last quarter, with unemployment rate now at 8%

North East - Unemployment was highest and employment lowest in UK, rising 5,000 in the quarter to 136,000.

North West - highest levels of economic inactivity, seeing a rise of 25,000 across the quarter and 82,000 compared to a year before - falls in employment and rises in unemployment - biggest quarterly rise in UK.

Yorkshire and the Humber - Unemployment stayed fairly steady at 245,000, a small rise of around 1,000 people.

West Midlands - unemployment rose by 7,000 continuing the recent trend, with the unemployment rate at 9.8% closing on one in 10 out of work.

South East - Unemployment was lowest and employment highest in UK

London - Unemployment fell 7,000 across the quarter to 359,500.

- Send out daily bulletins to newsdesks across England and Wales
- Try to highlight interesting data and possible stories for the different regional



# Making use of FOI efficiently

Two ways:

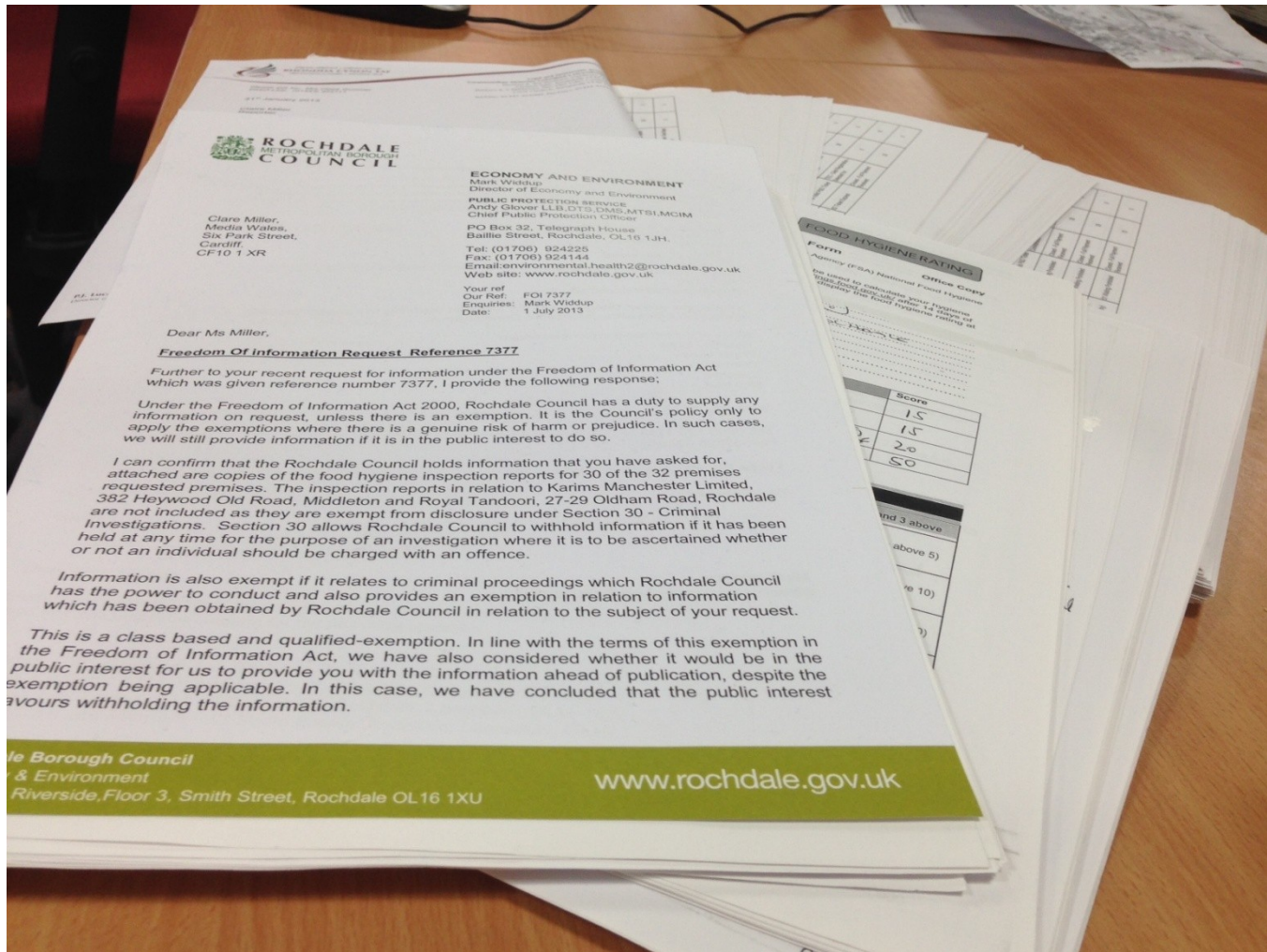
- Ask for a database
- Build your own (but make sure it's straightforward)

**DON'T**

Ask for lots of documents, you're never going to get through.

# Making use of FOI efficiently

is is all the documents I'm busy not transcribing



# Making use of FOI efficiently

Envisage what data will look like, then ask for it.

Be specific – do you want a number per year per area? Per month?

Have the spreadsheet you want to build in

	A	B	C	D	E
1	Other Care Authority	2009/10	2010/11	2011/12	Where Placed
2	Bath & North East Somerset		X	X	Monmouthshire
3	Birmingham City Council		X		Monmouthshire
4	Blaenau Gwent County Borough Council	X	X	X	Monmouthshire
5	Bolton Council			X	Monmouthshire
6	Bridgend County Borough Council		X		Monmouthshire
7	Bristol City Council			X	Monmouthshire
8	Caerphilly County Borough Council		X	X	Monmouthshire
9	Cardiff Council		X	X	Monmouthshire
10	Doncaster Council			X	Monmouthshire
11	Gloucestershire County Council			X	Monmouthshire
12	Herefordshire Council	X			Monmouthshire
13	Lewisham Council			X	Monmouthshire
14	Merthyr Tydfil County Borough Council		X	X	Monmouthshire
15	Milton Keynes Council	X		X	Monmouthshire
16	Neath Port Talbot County Borough Council	X			Monmouthshire
17	Newport City Council	X	X		Monmouthshire
18	Pembrokeshire County Council			X	Monmouthshire
19	Powys County Council	X		X	Monmouthshire
20	Rhonda Cynnon Taff County Borough Council	X	X	X	Monmouthshire
21	Solihull Metropolitan Borough Council	X			Monmouthshire
22	Surrey County Council			X	Monmouthshire
23	Swansea City Council		X	X	Monmouthshire
24	Torfaen County Borough Council	X	X	X	Monmouthshire
25	Vale Of Glamorgan Council			X	Monmouthshire
26	Worcestershire County Council		X		Monmouthshire
27	Blaenau Gwent Borough Council		1		Pembrokeshire
28	Bridgend County Borough	4	1	3	Pembrokeshire
29	Caerphilly County Borough Council		1		Pembrokeshire
30	Cardiff Council		3	1	Pembrokeshire
31	Carmarthen County Council	7	4	2	Pembrokeshire

# Data is messy

If you cannot understand or read the data, it is very hard to get a story from it.

Google Refine Children Placed with Welsh Councils.csv Permalink

Facet / Filter Undo / Redo 7 415 rows Extensions: Freebase

Refresh Reset All Remove All Show as: rows records Show: 5 10 25 50 rows < first < previous 1 - 10 next > last >

Other Care Authority change All Other Care Authority 2009/10 2010/11 2011/12 Where Placed

209 choices Sort by: n

(blank) 1  
Anglesey 2  
Barnet 1  
Barnsley 1  
Bath 1  
Bath & North East So  
Bedford 1  
Bedfordshire 1  
Birmingham 4  
Birmingham City Cour  
Blackburn 1

### Cluster & Edit column "Other Care Authority"

This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. Find out more ...

Method: nearest neighbor Distance Function: PPM Radius: 2 Block Chars: 6 36 clusters found

Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value
8	9	<ul style="list-style-type: none"><li>London Borough of Lewisham (2 rows)</li><li>London Borough of Haringey (1 rows)</li><li>London Borough of Hounslow (1 rows)</li><li>London Borough of Southwark (1 rows)</li><li>London Borough of Newham (1 rows)</li><li>London Borough of Lambeth (1 rows)</li><li>London Borough of Camden (1 rows)</li><li>London Borough of Hillingdon (1 rows)</li></ul>	<input type="checkbox"/>	London Borough of Lewisham
4	4	<ul style="list-style-type: none"><li>Thameside Metropolitan Boroug (1 rows)</li><li>Solihull Metropolitan Borough Council (1 rows)</li><li>Sandwell Metropolitan Borough (1 rows)</li><li>Oldham Metropolitan Borough Co (1 rows)</li></ul>	<input type="checkbox"/>	Thameside Metropolitan Boroug
4	6	<ul style="list-style-type: none"><li>Oxfordshire (3 rows)</li><li>Staffordshire (1 rows)</li><li>Hertfordshire (1 rows)</li><li>Bedfordshire (1 rows)</li></ul>	<input type="checkbox"/>	Oxfordshire
4	4	<ul style="list-style-type: none"><li>London Borough of Haringey (1 rows)</li><li>London Borough of Hounslow (1 rows)</li></ul>	<input type="checkbox"/>	London Borough of Haringey

Select All Deselect All Merge Selected & Re-Cluster Merge Selected & Close Close

# Choices in Cluster: 2 - 8

# Rows in Cluster: 2 - 9

Average Length of Choices: 8 - 33

Length Variance of Choices: 0 - 7.32

# Getting the story

H	I	J	K
Number of placements	2009/10	2010/11	2011/12
Blaenau Gwent	11	16	19
Bridgend	49	147	89
Caerphilly	65	69	89
Cardiff	11	25	33
Carmarthenshire	117	157	207
Ceredigion	15	17	28
Conwy		Refused	
Denbighshire		Refused	
Flintshire	64	33	42
Gwynedd			
Isle of Anglesey	75	60	89
Merthyr Tydfil	25	15	15
Monmouthshire	56	53	44
Neath Port Talbot	109	107	48
Newport	39	41	72
Pembrokeshire	42	32	22
Powys	211	188	180
Rhondda Cynon Taf	34	52	52
Swansea	No historic numbers		
Torfaen	33	21	17
Vale of Glamorgan	5	3	6
Wrexham	67	89	67
<b>Wales</b>	<b>1028</b>	<b>1125</b>	<b>1119</b>

Once you have a spreadsheet, you can start looking for stories

- Sorting
- Adding calculations – rates per 10,000
- Using pivot tables
- Merging with other datasets

# Things that might make a story:

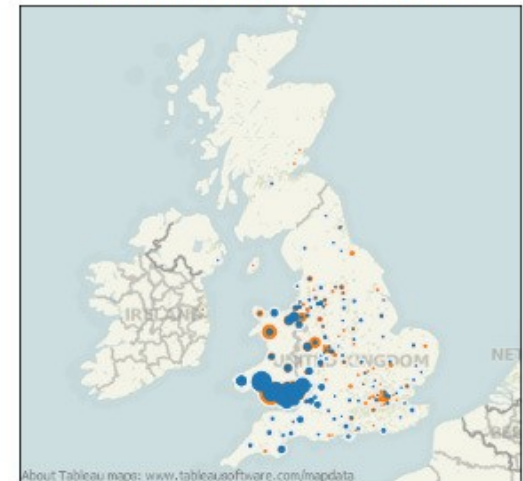
- o The biggest, the highest, the most
- o The smallest, the fewest
- o Fastest growing
- o Location
- o Successes
- o Failures
- o Why?

# Getting the story

	A	B	C	D	E	F	G
1	Council	Local Authorities	Latitude	Longitude	Placed by or Placed With	Year	Placements
2	Powys	Leeds	53.801279	-1.548567	placed with Welsh Council	2009/10	3
3	Powys	Luton	51.878671	-0.420026	placed with Welsh Council	2009/10	1
4	Flintshire	Wrexham	53.04304	-2.992494	placed with Welsh Council	2010/11	8
5	Powys	Oxfordshire	51.761206	-1.246467	placed with Welsh Council	2011/12	1
6	Powys	Wrexham	53.04304	-2.992494	placed with Welsh Council	2010/11	2
7	Cardiff	Merthyr Tydfil	51.74873	-3.381646	placed with Welsh Council	2010/11	2
8	Powys	Walsall	52.586214	-1.982919	placed with Welsh Council	2010/11	2
9	Powys	Wiltshire	51.246271	-1.992213	placed with Welsh Council	2009/10	1
10	Caerphilly	Rhondda Cynon Taf	51.674498	-3.468708	placed with Welsh Council	2010/11	27

Tidying up the spreadsheet saves time later when making visualisations

Placements of Children in Care



Note: not all Welsh councils gave figures and/or locations for where children were placed or which councils placed children in their local authority area.

placed by Welsh Council placed with Welsh Council

Select options below to explore the figures in detail

**Council**

- (All)
- Bridgend
- Caerphilly
- Cardiff
- Carmarthenshire
- Ceredigion
- Denbighshire
- Flintshire
- Isle of Anglesey
- Merthyr Tydfil
- Monmouthshire
- Neath Port Talbot
- Newport
- Pembrokeshire
- Powys
- Rhondda Cynon Taf
- Swansea
- Torfaen

**Placements**

- 1
- 20
- 40
- 80
- 83

**Placed by or Placed With**

- (All)
- placed by Welsh Council
- placed with Welsh Council

**Year**

- (All)
- 2009/10
- 2010/11
- 2011/12

Share Download

1,504 views

See more by this author

Learn About Tableau

# Visualisations

Use online tools for quick impact – Batchgeo, Datawrapper, infogr.am, Google charts

batchgeo  
Mapping should be this easy

make maps features pro pricing support blog

pro sign in

Paste your location data below to generate a free map:

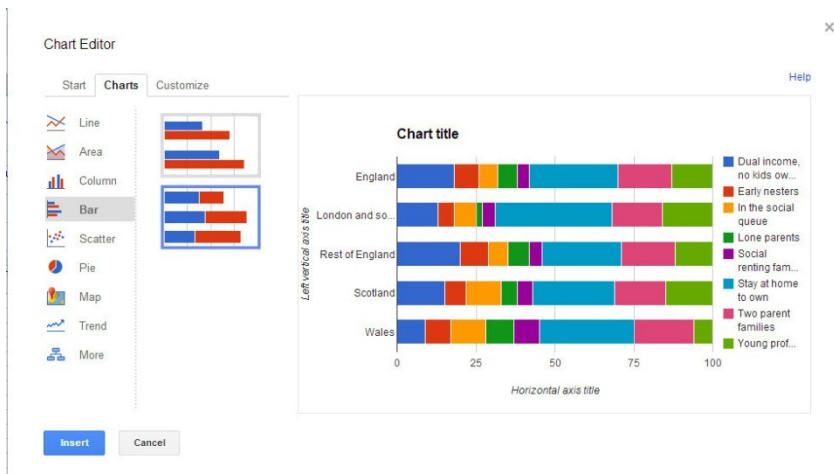
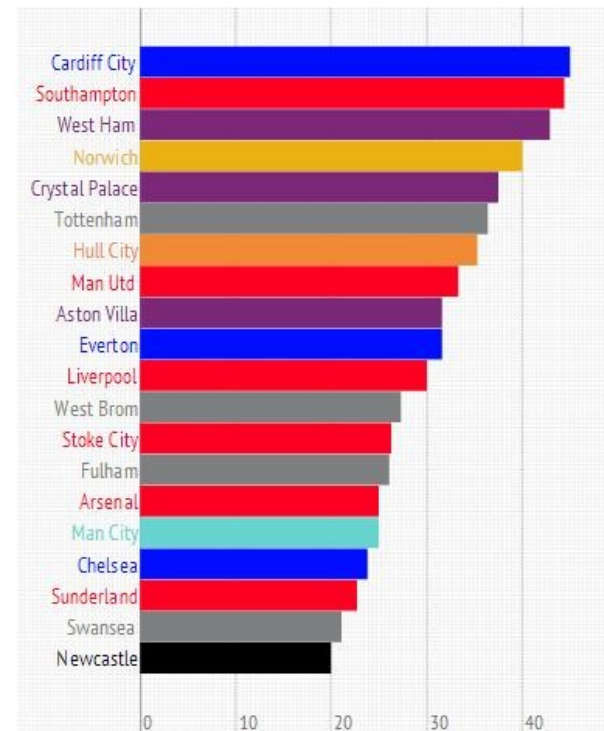
Address	Accepting NHS?	Telephone Number
A Jones Dental Surgery, 45 Barry Road, Barry, Val...	3 week wait	01446 732 231
Advance Dental Care, Lakin Drive, Barry, Vale of G...		01446 740 056
Daly & Associates Dental Surgery, 77 Tynewydd ...		01446 734640
Holton Road Dental Surgery, 157 Holton Road, Bar...		01446 720222

(Tab-delimited is the format most used. You can also start by using our Spreadsheet Template or hit "Map Now" and try it out with our example data.)

Validate & Set Options or Map Now

## League of Englishness

Clubs ranked by percentage of English players used this season





# Visualisations

Build resources in free tools to save time –  
Tableau Public, Google Fusion Tables



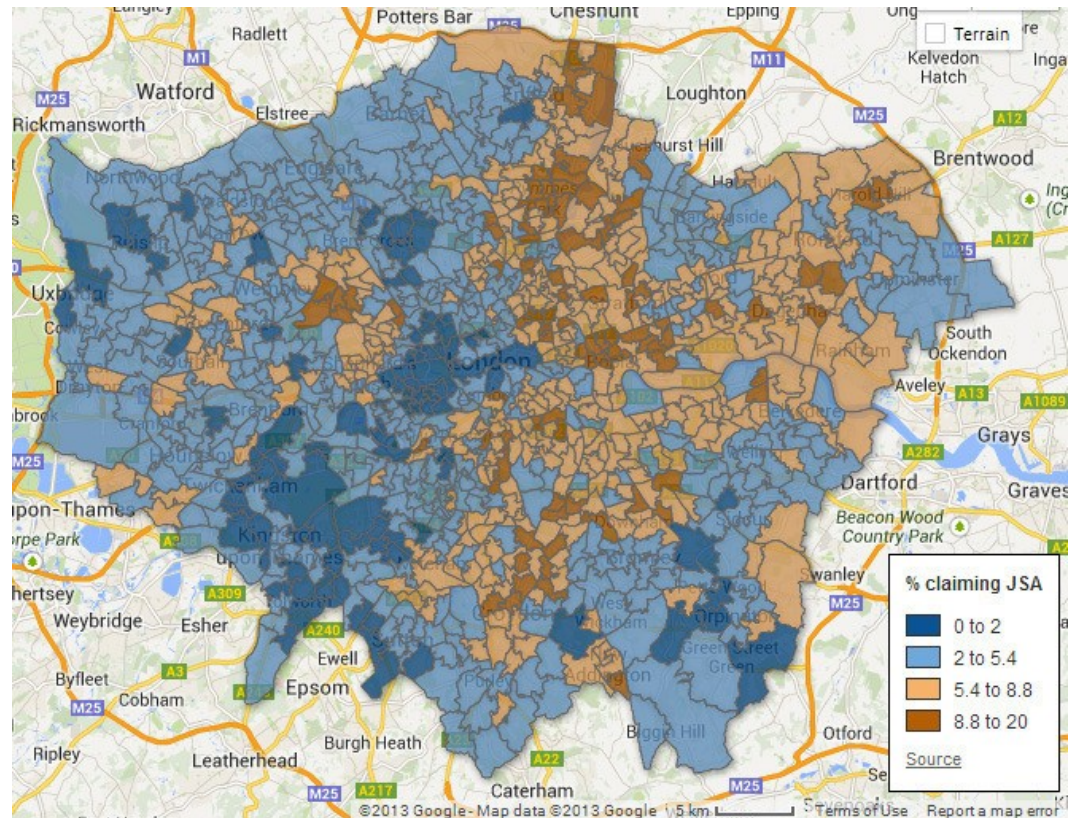
Tableau Public –  
great for quick  
maps with  
options for  
filtering.

Build once then  
update as you  
update your  
spreadsheet by  
connecting the  
visualisation to

# Visualisations

Build resources in free tools to save time -  
Tableau Public, Google Fusion Tables

Set up  
shapefiles in  
Google Fusion  
Tables ready to  
merge with  
datasets for  
quick heatmaps



# Visualisations

Make use of the features of the tools you have

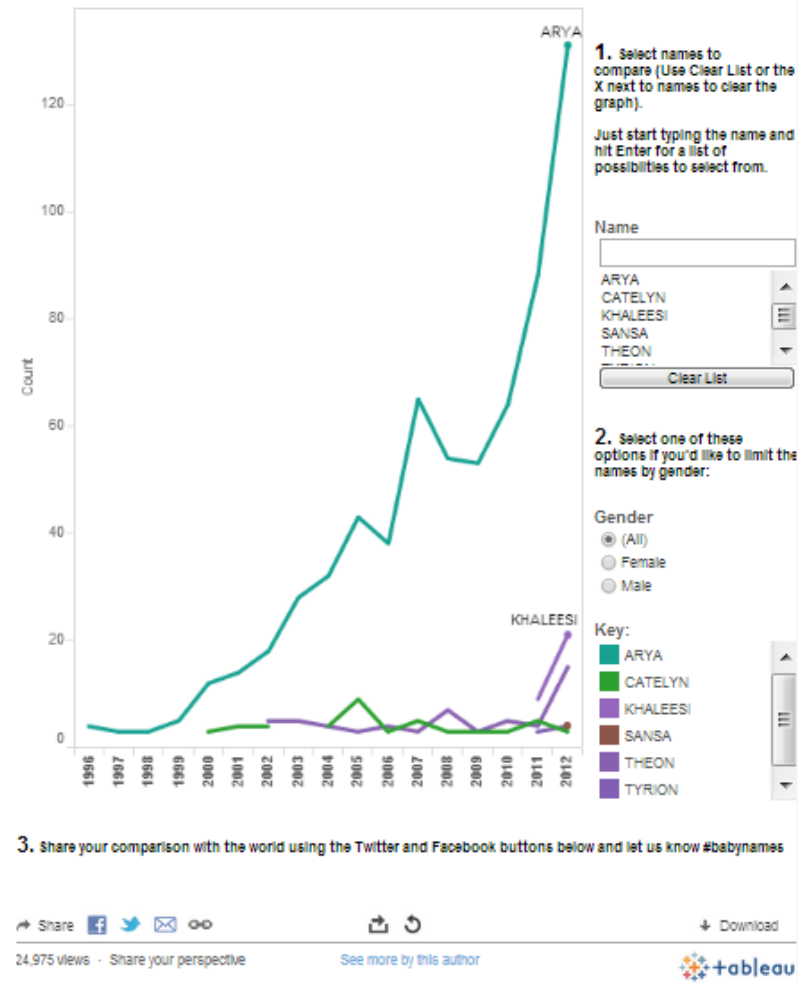
Baby names social sharing interactive made in Tableau

Uses filters and search options as well as facility to share views

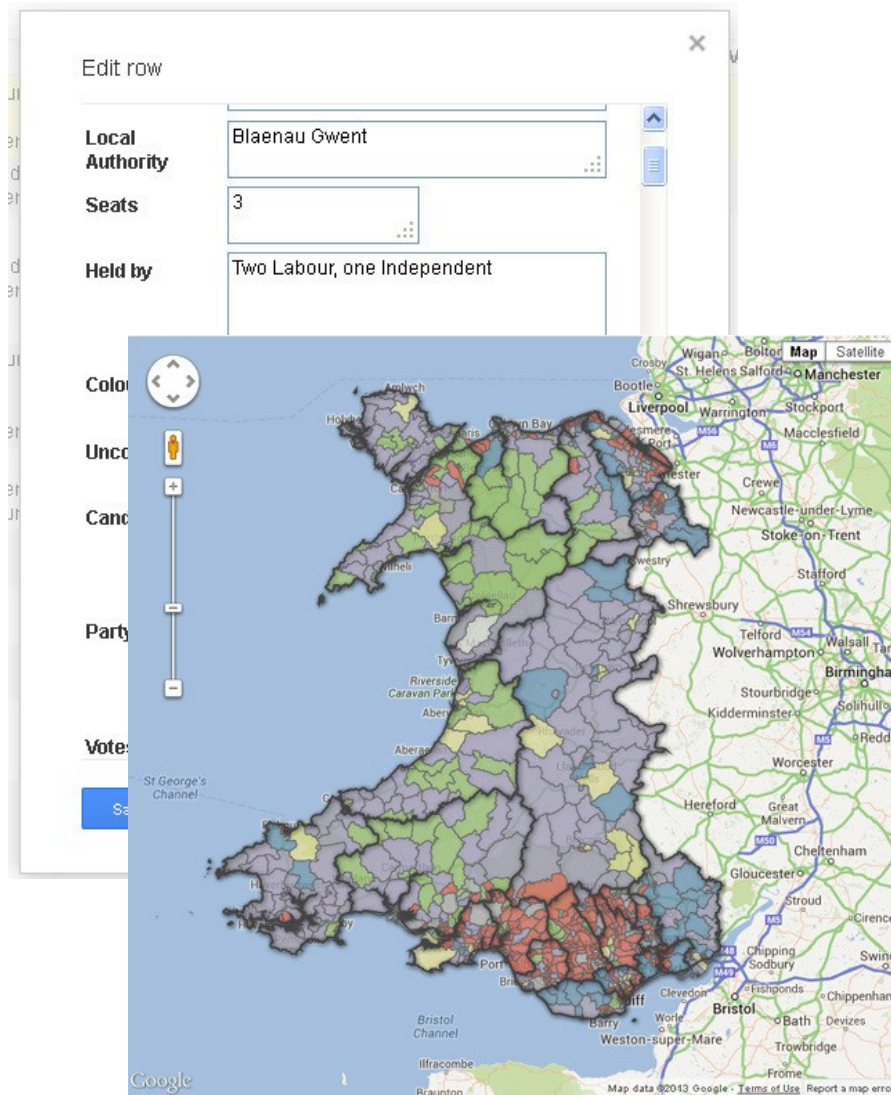
## Most Popular Baby Names in England and Wales

How many babies born since 1996 share your name and is it becoming more or less popular.

Use our simple comparison tool to find out the fortunes of the names of your friends, family and favourite celebrities.



# Visualisations



Make use of the features of the tools you have

Election map using Google Fusion Tables.

Updating the table updates the colours on the map on the pages it is embedded on.



# Getting Started with Data Journalism

*Writing data  
stories in any size*

[http://leanpub.com/  
newsroom-  
datajournalism](http://leanpub.com/newsroom-datajournalism)