



Open data driven policy making

AN2: Data and analysis

Please fill in the register
bit.ly/csl-register

Welcome

Commissioned by



Civil Service
Learning

Delivered by



Dr David Tarrant

The Open Data Institute

Lucy Knight

Local Government



Aim

To equip policy makers with the knowledge and skills they need to effectively use open data in policy making



Introductions

Who are you?

What is your role in relation to data driven policy making?

What would you like to get out of this training?



Today

Introduction to open data driven policy making

Benefits, caveats and risks of using data in policy making

Planning a data driven policy process

Data driven policy making in practice



Data driven policy making



Introduction to data driven policy making

Outcomes

1. Define data, big data and open data.
2. Describe how data is used to inform policy making in different fields.
3. Identify the benefits of using data in policy making.
4. Assess the risks, caveats and limitations of using data in policy making.



Exercise

What is data?

In as few words as possible, define ‘data’.

You can use an example as an answer if you wish.

One twist: you cannot use the word **information** in your answer.

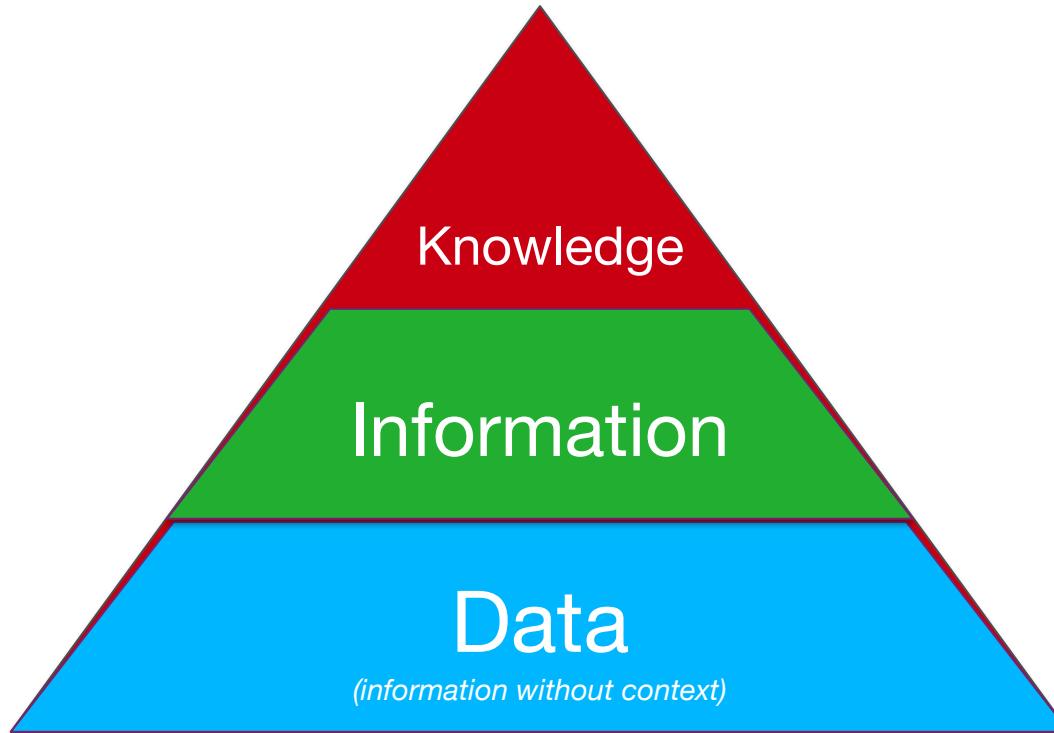


What is data?

- A collection of facts, information and statistics that can be analysed to develop new knowledge.
- A collection of numbers assigned as values to quantitative variables and / or characters assigned as values to qualitative variables.
- The lowest level of abstraction from which information and then knowledge are derived.

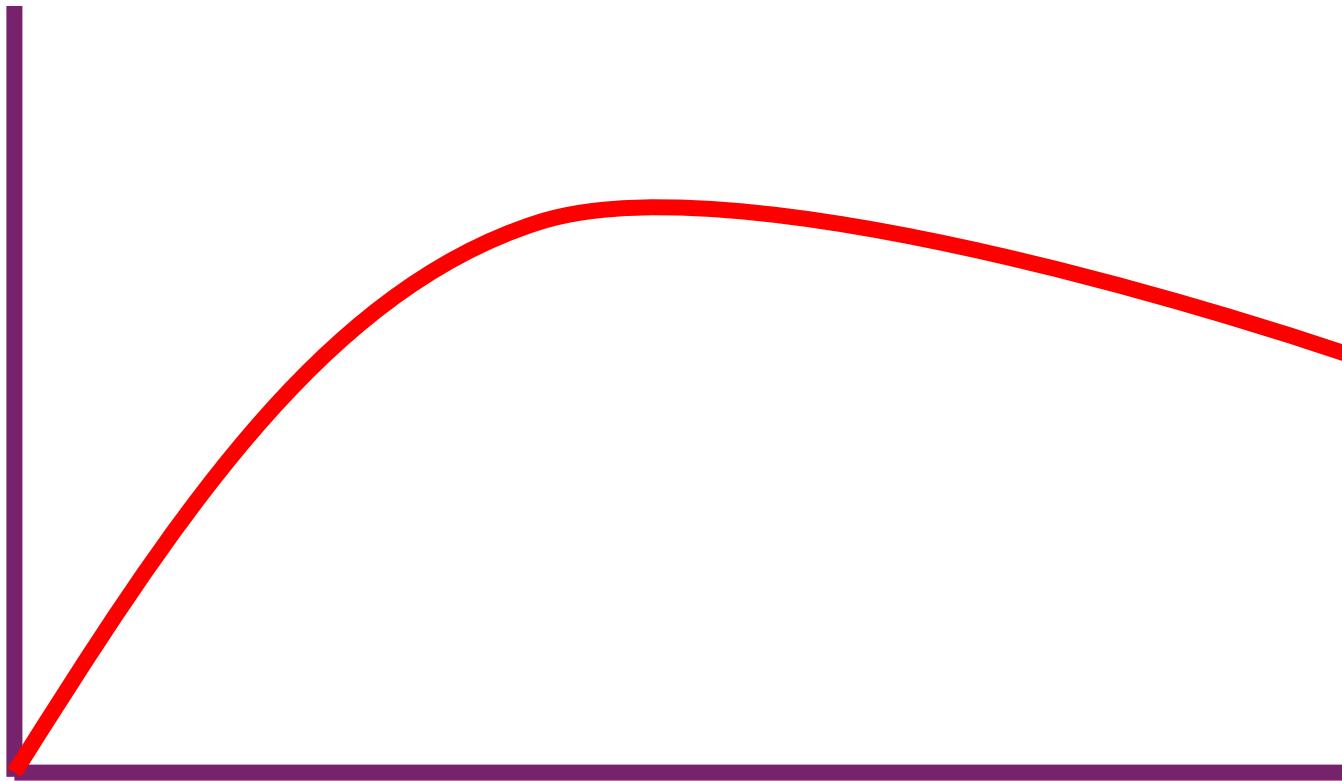


How is data different from information?





Civil Service
Learning





What did you see?

Data: What did you see?

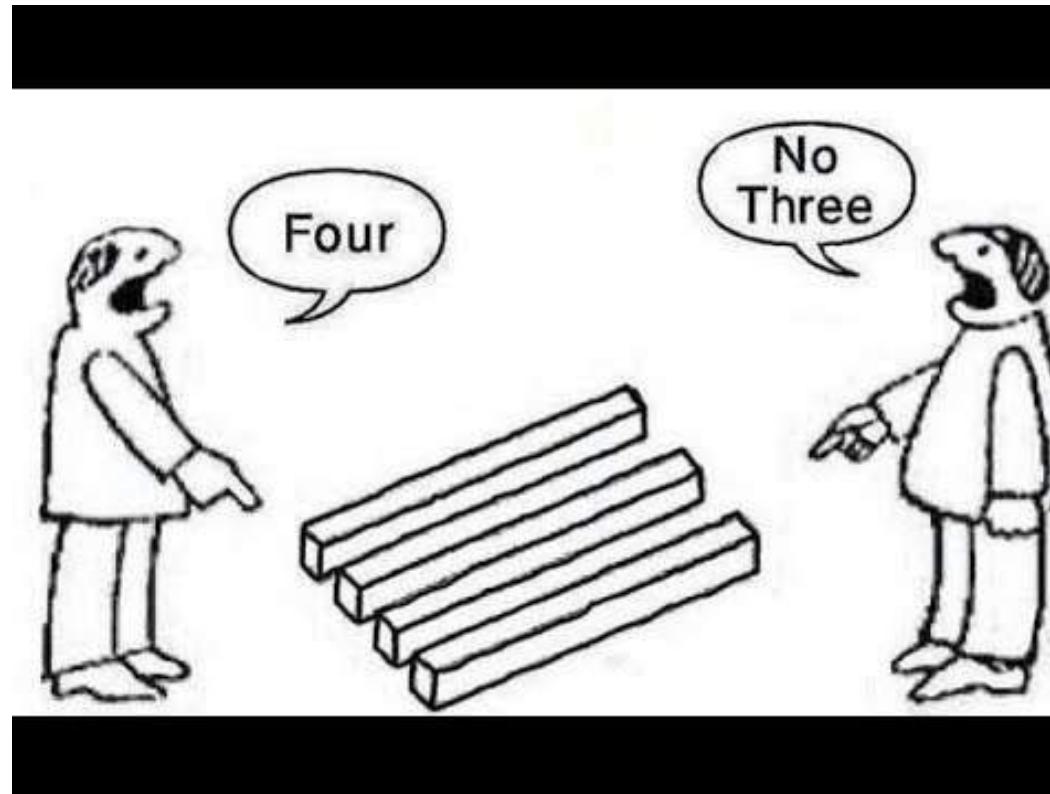
Information: How did you interpret it?

Knowledge: Did it mean (or relate to) anything for you?



Seeing what isn't there

Even counting is difficult (more on that later)





Exercise

What is open data?

In as few words as possible, what is ‘open data’?



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.



Civil Service
Learning

Data that anyone can access, use and share.

The Open Data Institute



Open data is data that is published in an open format, is machine readable and is published under a license that allows for free reuse.



Open data

The important points are:

- licensed openly (e.g. Open Government Licence)
- free to use, not always free to access (currently government open data must be available at no cost)
- users must be able to modify and redistribute the data

Data.gov.uk includes:

- published in open format
- machine readable



Case study: LIDAR

 **DATA.GOV.UK**^{Beta}
Opening up Government

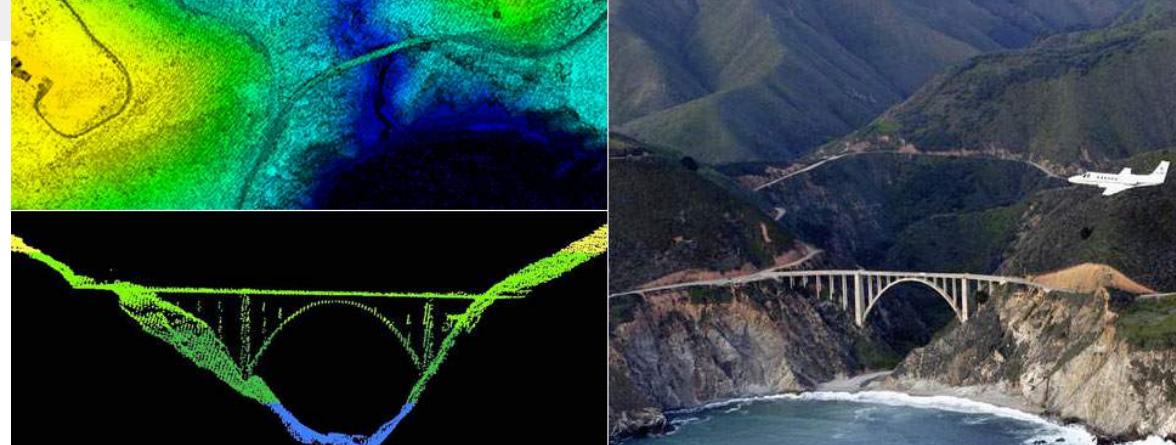
Home

Datasets Map Search Data Requests Publishers Data API Organograms Site Analytics

Home / Datasets / LIDAR Composite DTM - 50cm

LIDAR Composite DTM - 50cm

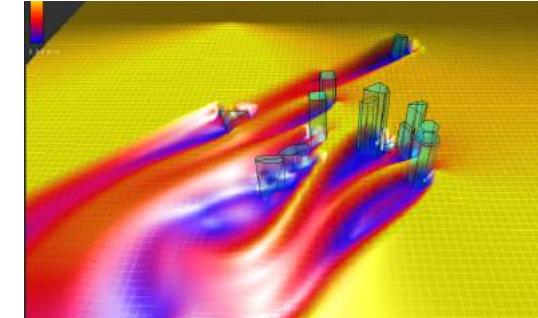
Published by Environment Agency. Licensed under  Open Government Licence.
Openness rating: ★★★☆☆





Impact of LIDAR data

Wind modelling



Archology



Educational games (Minecraft)





Video

Q: What types and uses of data do Emily and David focus on?



Civil Service
Learning



Review

What types of data do David and Emily focus on?



EMILY focusses on open data and data that is already showing benefit regardless of how big or small it is. Such as LIDAR, Transport, Environmental, Energy and other non-personal data.

DAVID focusses on Big Data and Exhaust Data that come from social interactions and from people. This brings into question lots of ethical and data protection issues and the Big Data Hubris.





Big data is changing the world

But what is big data?

Collect some ideas together on some post-it notes.
Do you have any categories emerging?



Big data is changing the world

Whenever you work with big data you must eliminate the effect of any of these aspects on your result.





Big data is changing the world

nature

Vol 457 | 19 February 2009 | doi:10.1038/nature07634

LETTERS

Detecting influenza epidemics using search engine query data

Jeremy Ginsberg¹, Matthew H. Mohebbi¹, Rajan S. Patel¹, Lynnette Brammer², Mark S. Smolinski¹ & Larry Brilliant¹

Seasonal influenza epidemics are a major public health concern, causing tens of millions of respiratory illnesses and 250,000 to 500,000 deaths worldwide each year¹. In addition to seasonal influenza, a new strain of influenza virus against which no previous immunity exists and that demonstrates human-to-human trans-

By aggregating historical logs of online web search queries submitted between 2003 and 2008, we computed a time series of weekly counts for 50 million of the most common search queries in the United States. Separate aggregate weekly counts were kept for every query in each state. No information about the identity of any user was retained. Each



97%

Google's claimed accuracy when compared to Centre for Disease Control data.

google.org Flu Trends

[Google.org home](#)

[Dengue Trends](#)

[Flu Trends](#)

[Home](#)

[United States](#)

[Washington](#)

[Download data](#)

[How does this work?](#)

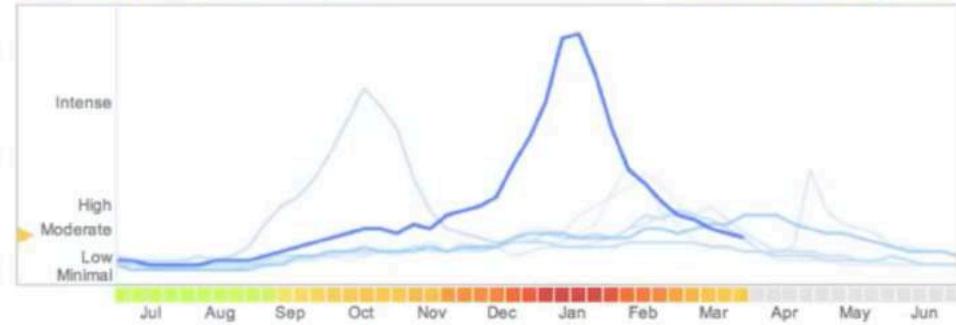
[FAQ](#)

Explore flu trends - United States

We've found that certain search terms are good indicators of flu activity. Google Flu Trends uses aggregated Google search data to estimate flu activity. [Learn more »](#)

[United States > Washington](#)

● 2012-2013 ● Past years ▾



[States](#) | [Cities](#) (Experimental)





Google

fever aching joints headache

All Shopping Images News Videos More Settings Tools

About 369,000 results (0.52 seconds)

WebMD Symptom Checker helps you find the most common medical conditions indicated by the symptoms **chills, fever, headache** and **joint aches** including Lyme disease, Acute sinusitis, and Aseptic meningitis (adult). ... Osteoarthritis happens when the cartilage in your **joints** breaks down causing pain, stiffness, and swelling.

Chills, Fever, Headache and Joint aches: Common Related Medical ...
[symptomchecker.webmd.com/multiple-symptoms?...chills%7Cfever%7Cheadache%7Cjoints%7C...](http://symptomchecker.webmd.com/multiple-symptoms?...chills%7Cfever%7Cheadache%7Cjoints%7C)

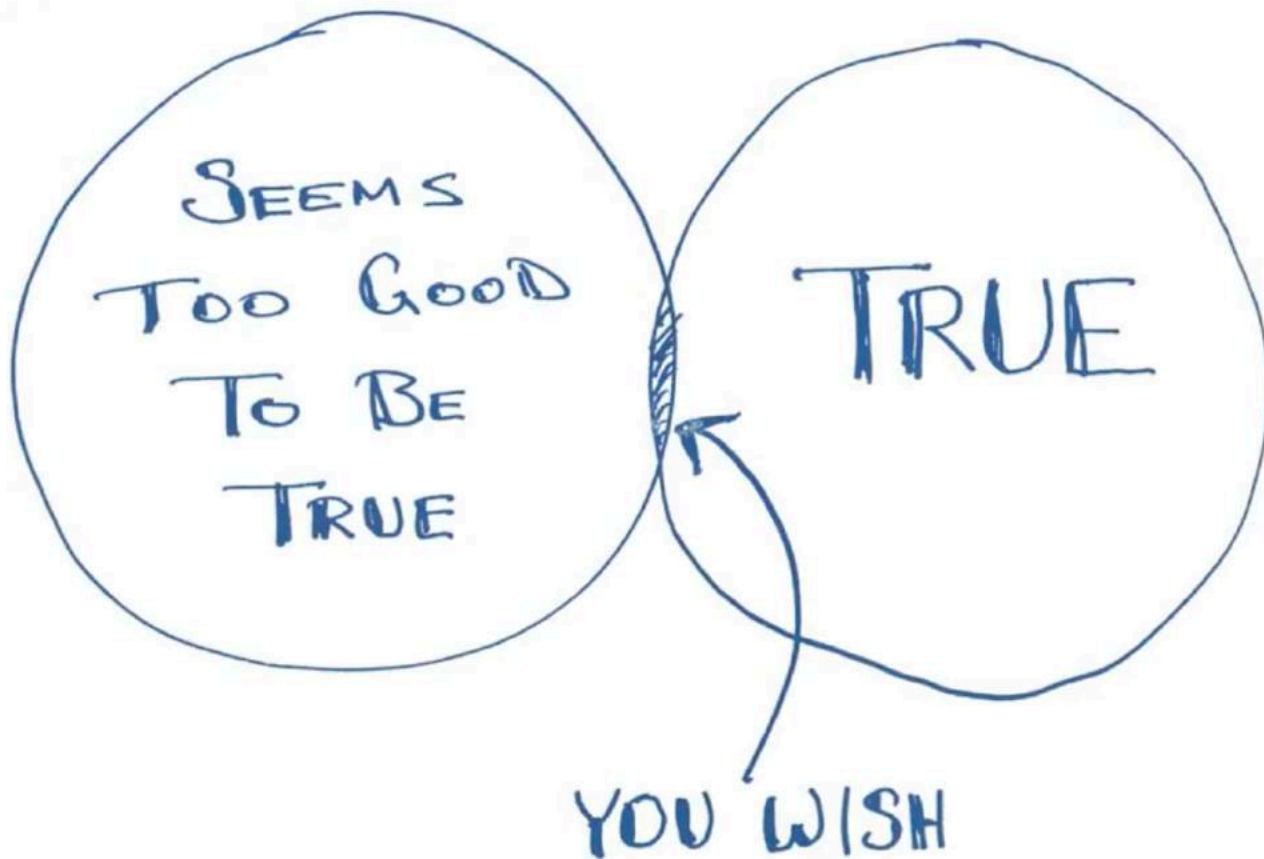
About this result • Feedback



VERACITY
Uncertainty of data

People making flu-related Google searches may know very little about how to diagnose flu?

Does a search for flu mean they have flu or just an interest in it?





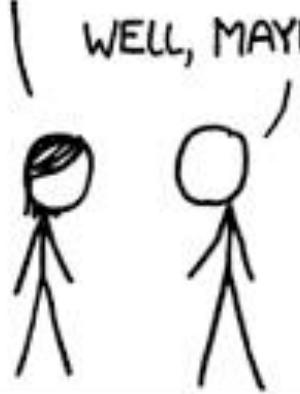
I USED TO THINK
CORRELATION IMPLIED
CAUSATION.



THEN I TOOK A
STATISTICS CLASS.
NOW I DON'T.



SOUNDS LIKE THE
CLASS HELPED.
WELL, MAYBE.



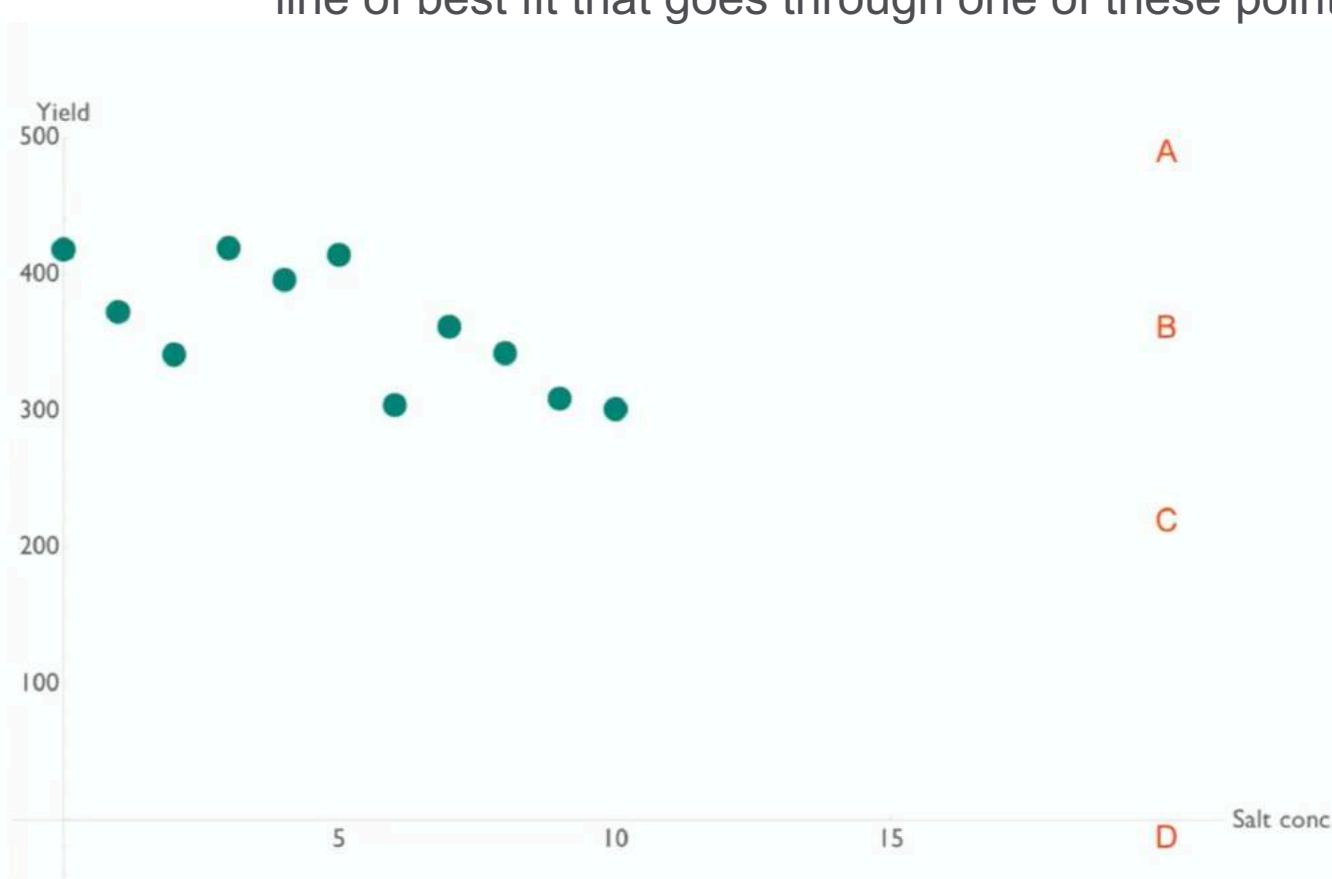
<https://xkcd.com/552/>

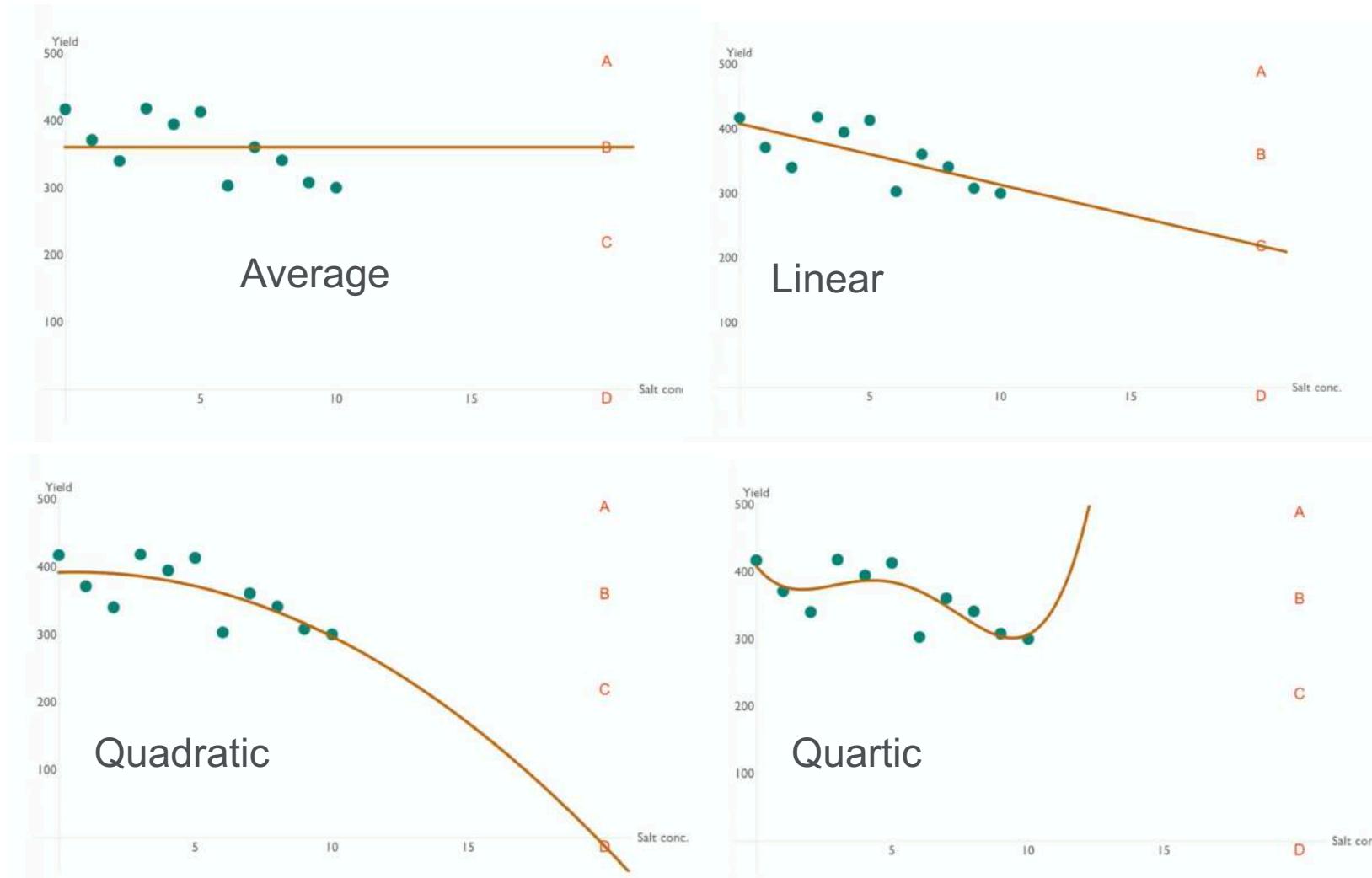


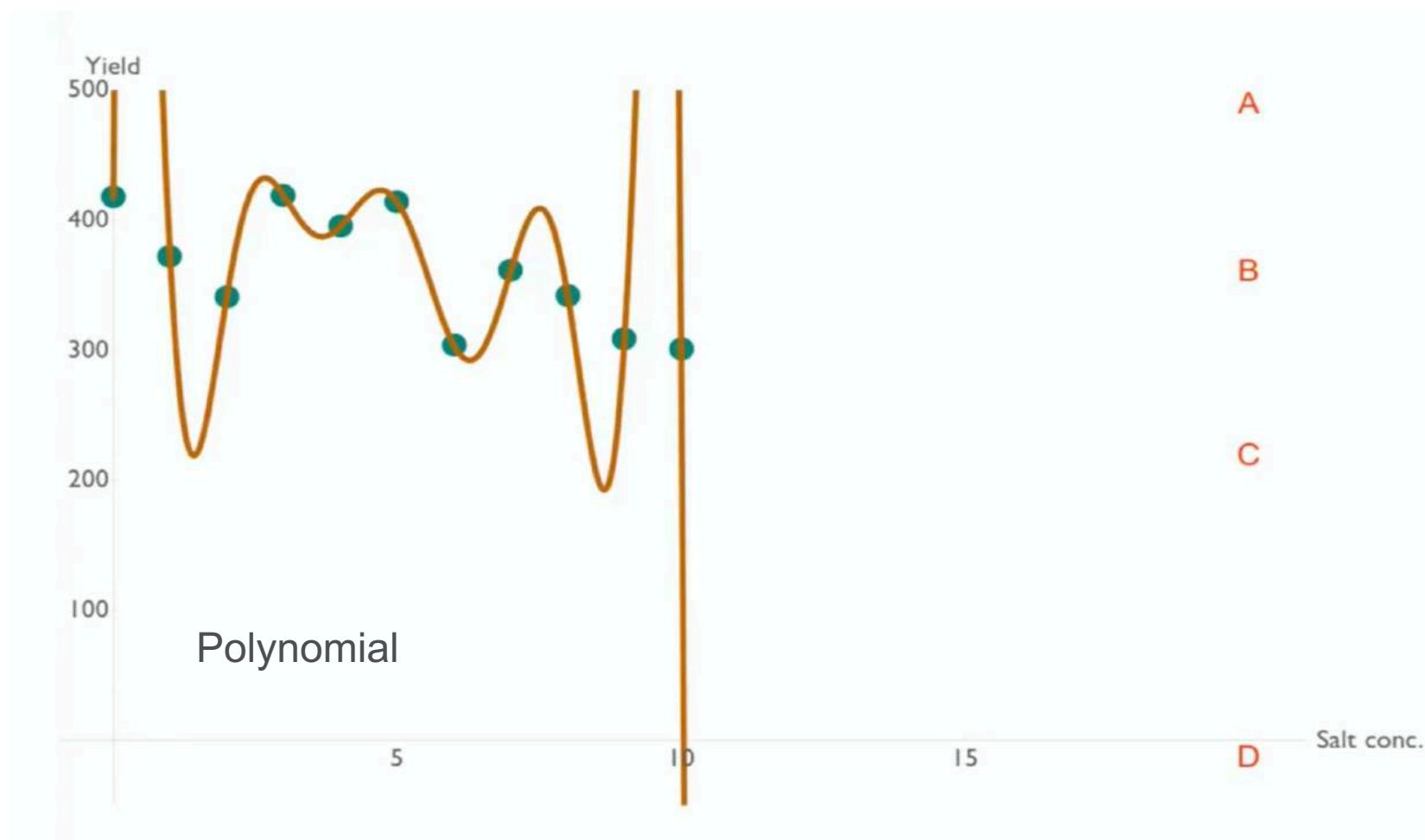
Overfitting

Plotted here are 10 points in a series that have been generated using a function.

If another 10 points were plotted of the same function which point (A,B,C or D) would they tend towards. Add a line of best fit that goes through one of these points.









Why did the flu trends example fail?

- Big Data Hubris (complement not supplement)
- Huge veracity problem!
- Used nth degree polynomial (overfitting)
- Solely relied on this data over scientific method

For more on big data search YouTube for “calling bullshit on big data”

Excellent lecture series from the University of Washington





Averages

You each have a set of cards containing salaries for the players in the red bulls soccer team.

Can you work out the average salary please?

33,750	33,750	33,750	33,750	44,000
44,000	44,000	45,566	65,000	95,000
103,500	112,495	138,188	141,666	181,500
185,000	190,000	194,375	195,000	205,000
292,500	301,999	4,600,000	5,600,000	



Averages

33,750	33,750	33,750	33,750	44,000
44,000	44,000	45,566	65,000	95,000
103,500	112,495	138,188	141,666	181,500
185,000	190,000	194,375	195,000	205,000
292,500	301,999	4,600,000	5,600,000	

What problems does this dataset have when working out the average?

How might we solve these problems?

Is the middle value (with data points ordered) better than the average?

Outliers



\$ 33,750.00	\$ 33,750.00
\$ 44,000.00	\$ 33,750.00
\$ 138,188.00	\$ 33,750.00
\$ 45,566.67	\$ 33,750.00
\$ 44,000.00	\$ 44,000.00
\$ 141,666.67	\$ 44,000.00
\$ 292,500.00	\$ 44,000.00
\$ 5,600,000.00	\$ 44,000.04
\$ 103,500.00	\$ 45,566.67
\$ 190,000.00	\$ 65,000.00
\$ 65,000.00	\$ 95,000.00
\$ 33,750.00	\$ 103,500.00
\$ 195,000.00	\$ 112,495.50
\$ 44,000.04	\$ 138,188.00
\$ 4,600,000.00	\$ 141,666.67
\$ 194,375.00	\$ 181,500.00
\$ 33,750.00	\$ 185,000.00
\$ 112,495.50	\$ 190,000.00
\$ 95,000.00	\$ 194,375.00
\$ 301,999.00	\$ 195,000.00
\$ 181,500.00	\$ 205,000.00
\$ 33,750.00	\$ 292,500.00
\$ 185,000.00	\$ 301,999.00
\$ 205,000.00	\$ 4,600,000.00
\$ 44,000.00	\$ 5,600,000.00

Average (mean): \$518,311.64

Median: \$112,495.50

16% of 25 = 4

16% trimmed mean: \$128,109.09
=trimmean(RANGE,0.16)



Averages

33,750	33,750	33,750	33,750	44,000
44,000	44,000	45,566	65,000	95,000
103,500	112,495	138,188	141,666	181,500
185,000	190,000	194,375	195,000	205,000
292,500	301,999	4,600,000	5,600,000	

Ensure your cards are in order.

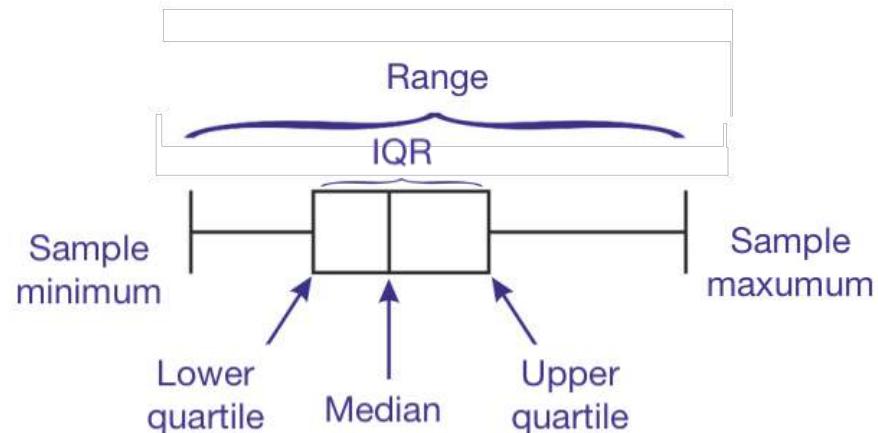
- Find the value $\frac{1}{4}$ of the way through
- Find the value $\frac{3}{4}$ of the way through



Five-number summary

The five-number summary

1. the sample minimum (smallest value)
2. the lower quartile (value $\frac{1}{4}$ through the list)
3. the median (middle value in the list)
4. the upper quartile (value $\frac{3}{4}$ through the list)
5. the sample maximum (largest value)





Five-number summary in Tableau

The screenshot shows the Tableau software interface with three main panels:

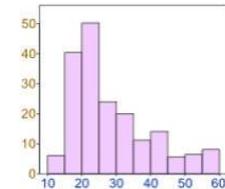
- Connect:** On the left, a dark sidebar with options like "To a File" (Microsoft Excel, Text file, JSON file, PDF file, Spatial file, Statistical file), "To a Server" (OData), and "More...". It also includes a message about working with big data and a "Upgrade Now" button.
- Open:** The central panel, titled "Open", displays four items:
 - A blank white square labeled "LFB_All_Data".
 - A blank white square labeled "LFB_Trial1".
 - A map of London boroughs labeled "London_boroughs".
 - A scatter plot labeled "Greater Manche...".
- Discover:** On the right, a sidebar with links to "How-to Videos", "Overview", "Intro to the Interface", "Chart Types", and "More how-to videos...". It also features a "Viz of the Day" section titled "THE POTTERVERSE - FAMILY TREE" with links to "Blog - Step and jump into Tableau Public 2018.1", "Sample Data Sets", "Live Training", and "Current Status".



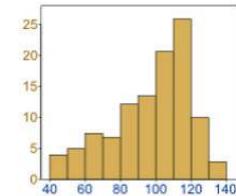
The distribution of data

Another important aspect to consider is the distribution of data. It is always a good practice to know the distribution of your data before analysing it further. Certain analyses require certain distributions.

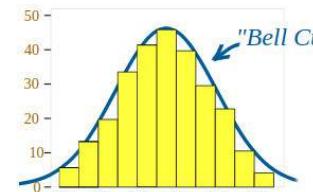
The examples here show different types of distributions of data.



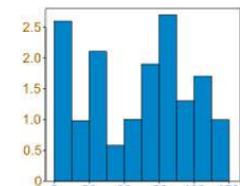
Positively skewed distribution



Negatively skewed distribution



Normal distribution



Non-normal distribution



Machine learning and prediction

Each table has a set of “Top Trump” cards relating to properties in two cities.

Build a decision tree to sort them into “New York” and “San Francisco”.

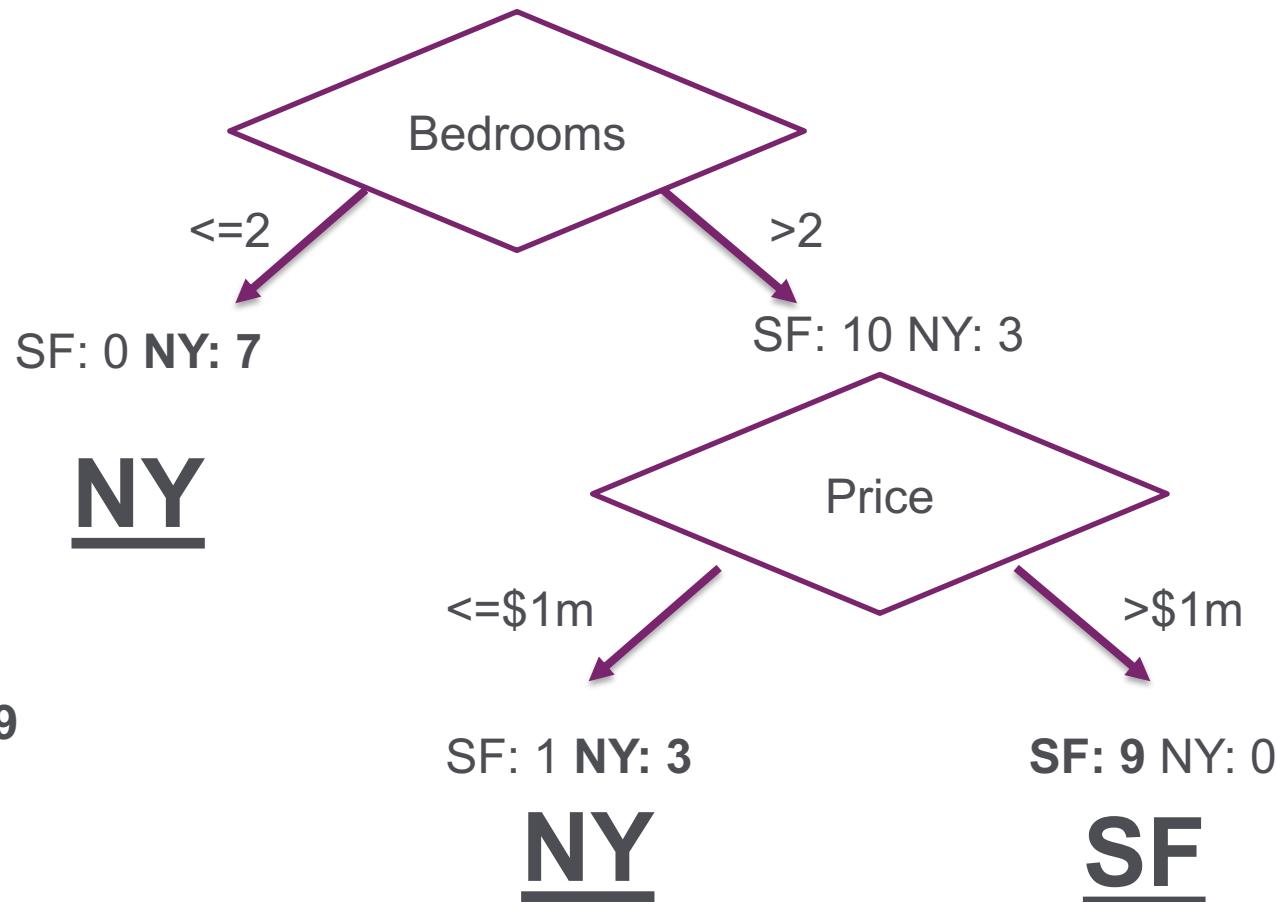
You cannot use the name of the city to sort them.



95%

Confidence

Example decision tree





Approaches



Data first



Knowledge first



Assumption

What share of income tax paid in the UK is paid by the top 1% of earners?

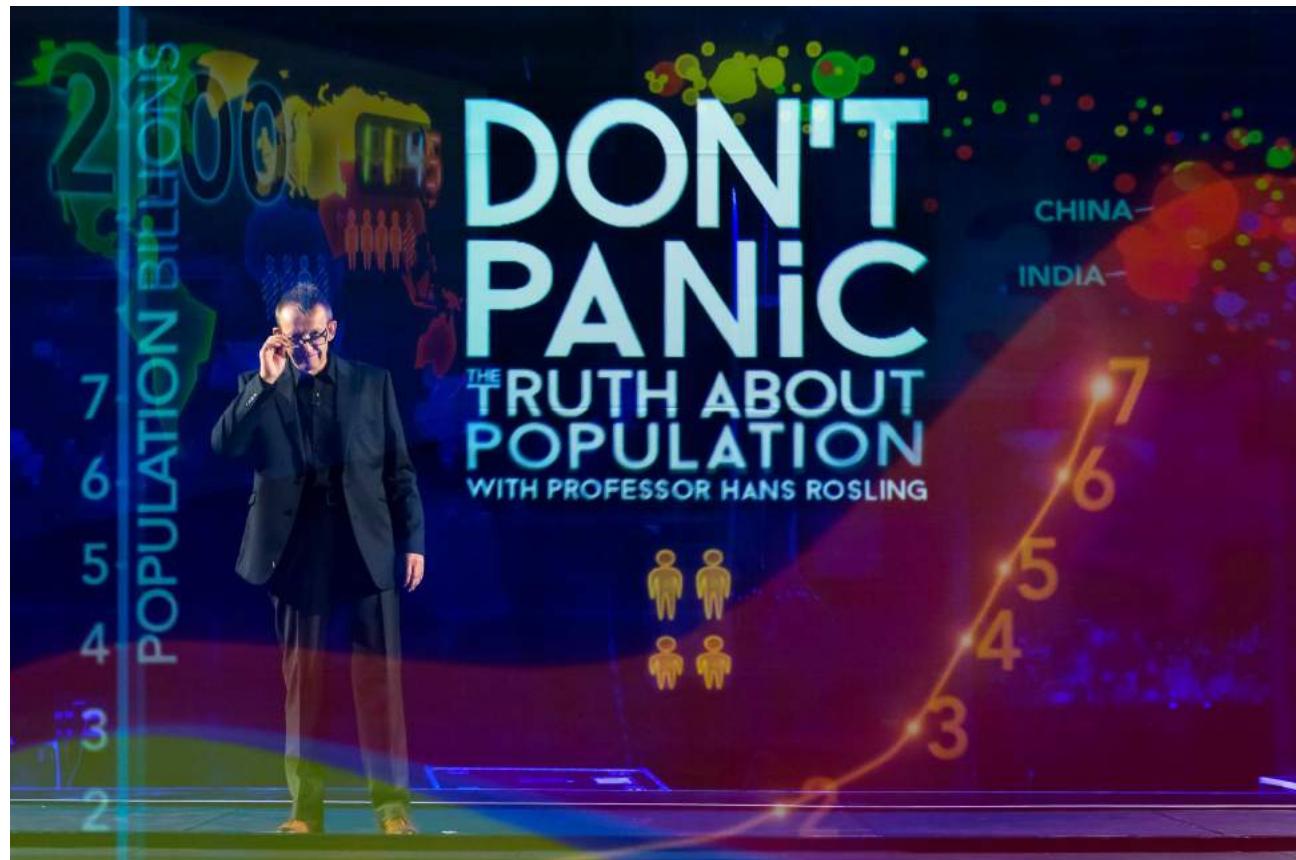
- ◆ A: 5%
- ◆ C: 14%
- ◆ B: 9%
- ◆ D: 17%



Assumption

What is the average number of children per family in Bangladesh?

- ◆ A: 2
- ◆ B: 3
- ◆ C: 4
- ◆ D: 5



gapminder.org



Data analysis of the houses dataset in Tableau

The screenshot shows the Tableau Public interface with the following details:

- Top Bar:** Standard toolbar with various icons.
- Left Panel (Data Source):**
 - Dimensions:** Target, Year Built (selected), Zip, Measure Names.
 - Measures:** Bath, Beds, Elevation, Index, Price, Price Per Sqft, Sqft, Latitude (generated), Longitude (generated), Number of Records, Measure Values.
- Middle Panel:**
 - Pages:** Shows 'Sheet 1'.
 - Filters:** Empty.
 - Marks:** Automatic, with options for Color, Size, Text, Detail, and Tooltip.
- Sheet 1 View:** A blank canvas with three 'Drop field here' placeholder areas.
- Bottom Bar:** Includes tabs for Data Source, Sheet 1 (highlighted), and other sheet icons.



Options?

Grow the decision tree until...

1. every classification is perfect?
2. confidence level is above 80%?
3. it is too big to process the data in reasonable time?
4. Until the evaluation and training set have the same confidence?
5. You have used all principal components?



Civil Service
Learning

A visual introduction to machine learning

r2d3.us





Considerations

Combine the data with existing knowledge but make sure the existing knowledge is correct!

Grow the tree around the principal components (using PCA). Stop when only minor improvements are achieved on the training set. Don't overfit!



Civil Service
Learning

GoCompare

ALL Details EXACTLY the same

Hi John!

Adjust cover

49 car quotes found, 4 telematics quotes included.

Sort: Annually

£538.26

Total Excess: £150

[View details >](#)



Legal
Assistance
+£30.99



Breakdown
Cover
+£43.99



Personal
Accident
✓



Windscreen
✓



Courtesy Car
✓

M&S BANK



Legal
Assistance
+£26.29



Breakdown
Cover
+£31.54



Personal
Accident
✓



Windscreen
✓



Courtesy Car
✓

£571.66

Total Excess: £150

[View details >](#)



Legal
Assistance
+£30.99



Breakdown
Cover
+£43.99



Personal
Accident
✓



Windscreen
✓



Courtesy Car
✓

£573.39

Total Excess: £150

[View details >](#)



Legal
Assistance
+£30.99



Breakdown
Cover
+£43.99



Personal
Accident
✓



Windscreen
✓



Courtesy Car
✓

£577.87

Total Excess: £150

[View details >](#)



Civil Service
Learning

GoCompare

Car: Ford Fiesta Ghia 2002-2008 1.6 Petrol
Profession: Insurance Director
Address: Milford Haven (PPI Company :P)

Adjust cover

37 car quotes found, 5 telematics quotes included.

Always resided in UK
Date of birth: 01/01/1980
No claims and license: 16 years
Car kept on drive

Hi Mohammed!

Sort: Annually

 LLOYDS BANK	 Legal Assistance +£30.99	 Breakdown Cover +£43.99	 Personal Accident ✓	 Windscreen ✓	 Courtesy Car ✓	£1,446.32 Total Excess: £150	View details >
 HALIFAX	 Legal Assistance +£30.99	 Breakdown Cover +£43.99	 Personal Accident ✓	 Windscreen ✓	 Courtesy Car ✓	£1,458.04 Total Excess: £150	View details >
 BANK OF SCOTLAND Decisions well made	 Legal Assistance +£30.99	 Breakdown Cover +£43.99	 Personal Accident ✓	 Windscreen ✓	 Courtesy Car ✓	£1,459.64 Total Excess: £150	View details >
M&S BANK	 Legal Assistance ✓	 Breakdown Cover ✓	 Personal Accident ✓	 Windscreen ✓	 Courtesy Car ✓	£1,476.70 Total Excess: £150	View details >



Good or bad ideas?

1. a tool that analyses the sentiment of a user's tweets, assesses whether they are suicidal and alerts friends	4. a risk-assessment tool that uses AI to advise on prison sentences based upon criminal profile analysis
2. automatic pricing algorithm for taxi firm which responds to surges in demand	5. using energy efficiency data and winter fuel allowance data to target efficiency advice
3. using performance data to advise on how to save money in the emergency services	6. publishing genomes of 100,000 individuals for use in public health



Automated Inference on Criminality using Face Images

Xiaolin Wu

McMaster University
Shanghai Jiao Tong University

xwu510@gmail.com

Xi Zhang

Shanghai Jiao Tong University
zhangxi_19930818@sjtu.edu.cn



“Unlike a human examiner/judge, a computer vision algorithm or classifier has absolutely no subjective baggages [sic], having no emotions, no biases whatsoever due to past experience, race, religion, political doctrine, gender, age, etc., no mental fatigue, no preconditioning of a bad sleep or meal. The automated inference on criminality eliminates the variable of meta-accuracy (the competence of the human judge/examiner) all together.”

—Wu & Zhang(2017)



“Unlike a human examiner/judge, computer vision algorithm or classifier has absolutely no subjective baggages [sic]—no emotions, no biases whatsoever due to experience, race, religion, political affiliation, gender, age, etc., no mental fatigue, no conditioning of a bad sleep or meal. That is, it is not subject to the variable of meta-accuracy (the ‘confidence’ of the human judge/examiner) all together.”

**Utter Bull...
—Wu & Zhang(2017)**

Example from “calling bullshit” lecture series from the University of Washington



Good or bad ideas?

1. a tool that analyses the sentiment of a user's tweets, assesses whether they are suicidal and alerts friends	4. a risk-assessment tool that uses AI to advise on prison sentences based upon criminal profile analysis
2. automatic pricing algorithm for taxi firm which responds to surges in demand	5. using energy efficiency data and winter fuel allowance data to target efficiency advice
3. using performance data to advise on how to save money in the emergency services	6. publishing genomes of 100,000 individuals for use in public health



Review

What is open data and what are the general benefits of open data?

What impact could data have on policy making?

In which situations should it be used/not used?

How to you spot red herrings in data analysis?



Civil Service
Learning

Course materials

<http://training.theodi.org/csl-an2/>



Afternoon Session

Outcomes

6. List the stages in carrying out data analysis for policy making.
7. Create a plan for carrying out data analysis for policy making.
8. Describe a number of ways to effectively present the results of data analysis.
9. Carry out a simple data analysis using a number of tools.
10. Create an interactive data visualisation.
11. Communicate the results of a data analysis to decision makers.
12. Review the role of open data in policy making



Data analysis practical

You have been tasked with saving money in the local fire service.

You must propose a solution that saves money while minimising the impact on service delivery.

Which factor do you think is the most significant in analysing performance to identify savings (top of the decision tree)

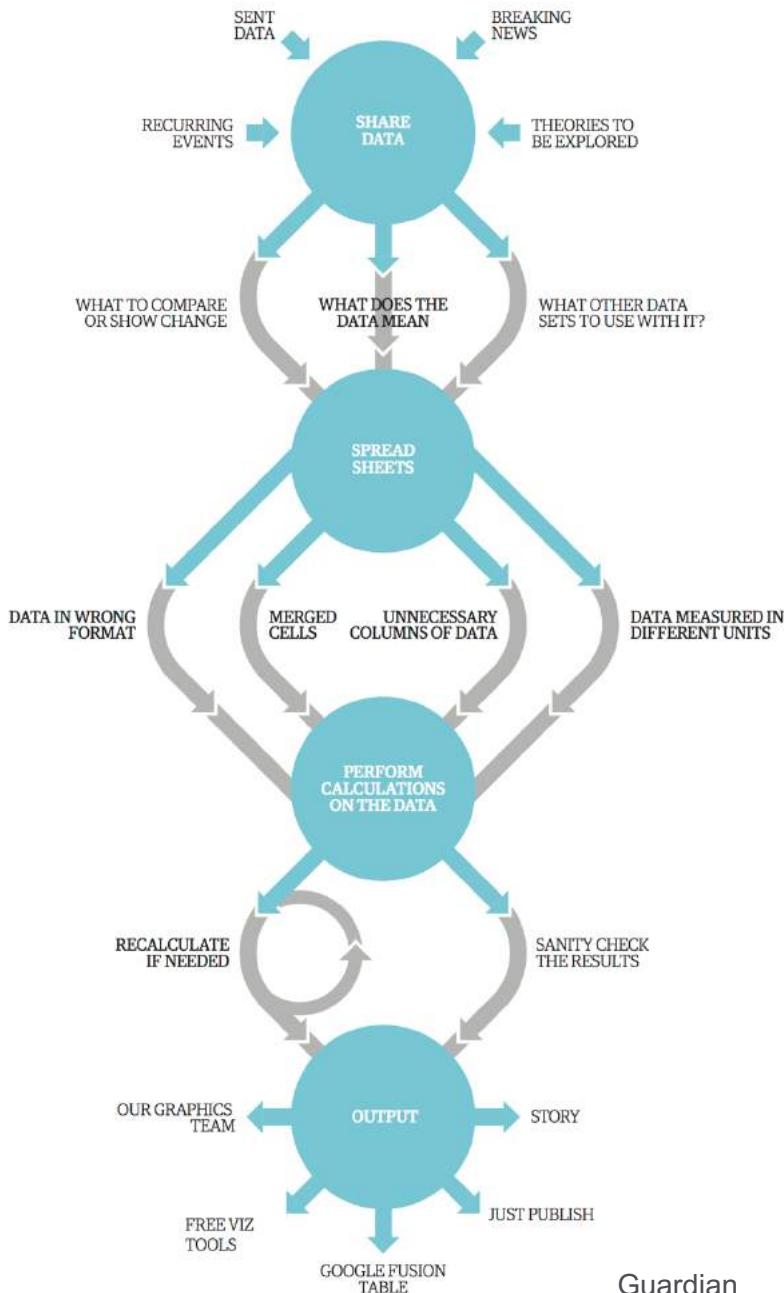


Paul Hudson



Using data in analysis

1. Collect data.
2. Establish context of data.
3. Clean the data.
4. Translate, filter and merge data.
5. Initial evaluation.
6. Go back to start?
7. Secondary evaluation and analysis.
8. Sanity check.
9. Create output.





Data analysis practical

Modelling the impact of London Fire Station closures.

3. using performance data to advice on how to save money in the emergency services



Paul Hudson



Stage 1: collect data

Make a space on your group desk

Put a yellow post-it note in the middle and draw a fire on it. Write a time on it between 4 minutes and 7 minutes.

Scatter a series of all different colour (including yellow) post-it around this one to keep it in the middle. Write a time between 4 and 7 minutes on each one.

Stage 2: establish context

Each post-it note represents an incident that has been responded to by the London fire brigade.

Each colour represents a different fire stations from which the appliance (fire engine) was sent



Stage 3: clean the data

Given the context does your data make sense?

Stage 4: Translate, filter and merge data

Nothing to do...



Stage 5: initial evaluation

How do we calculate the impact of closing the YELLOW fire station (and specifically the middle incident)?

What is your design?



Stage 8: Sanity check

Does your design overfit (or use nth degree polynomial?)

Can your design really model the future or just show how the past was?



What has been the impact of closing 10 fire stations in London?



Stage 1: collect data

<http://bit.ly/LFAData>

Stage 2: establish context

- What does this dataset tell us?
- What doesn't it tell us?
- What more do we need to know to answer our question?
- Do we need more data?



Stage 3: clean the data

Google refine government IT contracts [Permalink](#)

Facet / Filter Undo / Redo : [Reset All](#) [Remove All](#)

512 matching rows (5200 total)

Show as: rows records Show: 5 10 25 50 rows

#	Contract ID	Contractor Name	Type of Contract	Date of Award	Start Date	End Date
70.	3038	CGI FEDERAL INCORPORATED	FFP	10/03/2008	10/03/2008	11/03/2008
71.	3039	CGI FEDERAL INCORPORATED	FFP	01/08/2008	01/09/2008	09/09/2008
72.	3040	CGI FEDERAL INCORPORATED	FFP	01/08/2008	01/09/2008	09/09/2008
73.	3041	INTERNATIONAL BUSINESS MACHINES CORPORATION	FFP	03/17/2008	03/23/2008	11/03/2008
74.	3042	CGI FEDERAL INCORPORATED	FFP	04/21/2008	04/21/2008	09/21/2008
75.	3043	SOLUTIONS ENGINEERING CORP	FFP	11/01/2008	11/01/2008	10/11/2008
76.	3044	EVERGREEN INFORMATION TECHNOLOGY	FFP	11/20/2008	11/20/2008	01/11/2009
77.	7946	INTERNATIONAL BUSINESS MACHINES CORPORATION	FFP	10/01/2008	10/01/2008	09/10/2008
78.	7947	THE NEWBERRY	FFP	10/01/2008	12/01/2008	04/10/2009

Type of Contract change invert result
815 choices Sort by: name count Cluster

- FFAA: Fiscal/Financial Agent Agreement 1
- FFIP 1
- FFP 312** will include
- FFP 1
- FFP 1
- FFP (OPS) 2
- FFP (F&E) 1
- FFP (Power Supply Retrofit) Old # DTFA01-02-D00004 1
- FFP BPA 1
- FFP CPAF CPIF 1



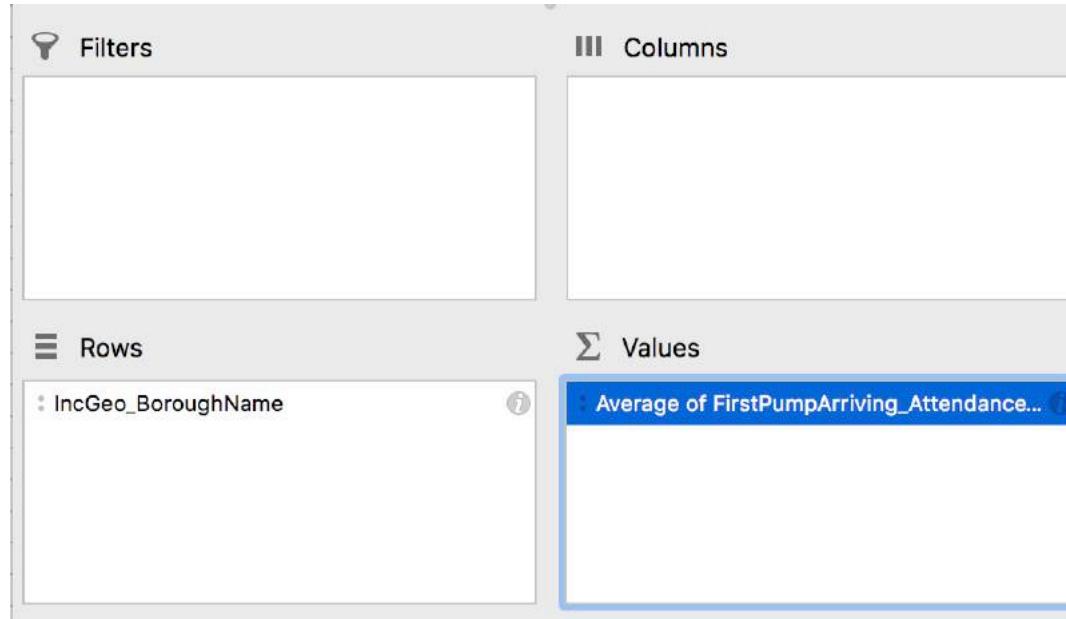
Stage 4: translate, filter and merge data

- Remove records for before the closures (7 Jan 2013).
- Ensure that none of the closed stations have records entered after this date.
 - Belsize, Bow, Clerkenwell, Downham, Kingsland, Knightsbridge, Silvertown, Southwark, Westminster and Woolwich.



Stage 5: initial analysis

- Work out the mean attendance time for the first appliance to arrive across London.
- Work out the mean attendance time per borough (using a pivot table).





Stage 5: initial analysis (part 2)

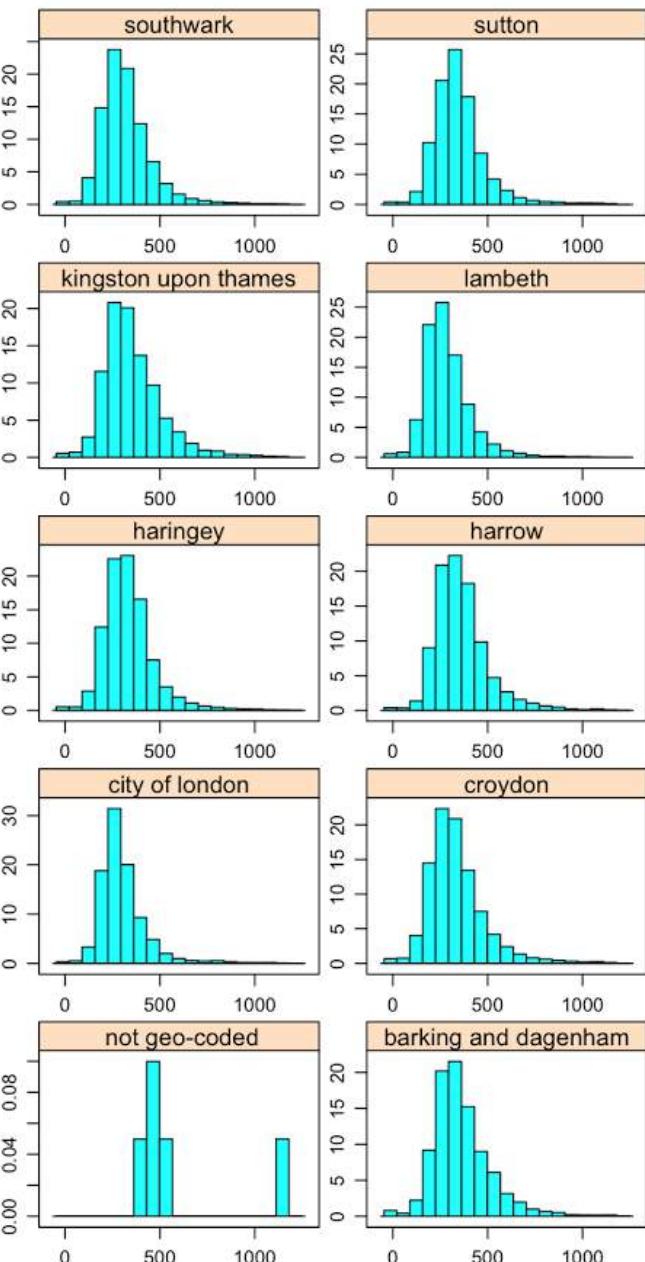
- Copy the average attendance time from Column B to Column C.
- Hide column B.
- Create column D from column C by dividing all values by 86400 (seconds in a year).
- Hide column C.
- Format the cells of column D and select mm:ss from custom data types.
- In a new cell at the top insert the value (=360/86400).
- Format that as per column D to turn it to a time.
- Use conditional formatting to highlight rows where the response time is greater than this new cell value (6 minutes).



Stage 6: secondary evaluation and analysis in R

R is an advanced statistical library that is much more powerful than excel and can handle vast quantities of data.

Follow the exercise at <http://bit.ly/LFA-R>
to find out more.

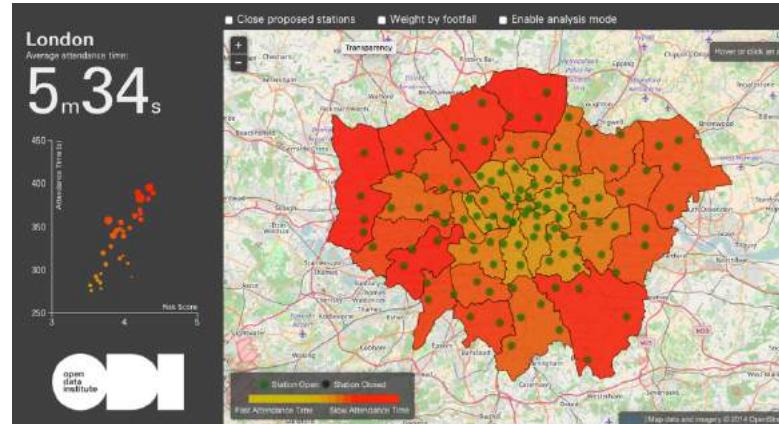




Stage 7: sanity check

- Download the dataset prior to closures and look at differences in response times?
- Was the ODI impact analysis tool accurate?

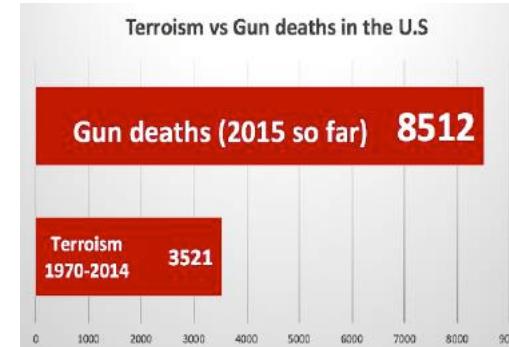
<http://london-fire.labs.theodi.org/explore/>





Stage 8: creating output

- What is your headline?
- If you want to communicate one message (or headline), keep any visualisations simple and effective.
- If you want people to be able to explore the data for themselves and create their own story, then it is possible to create something interactive.

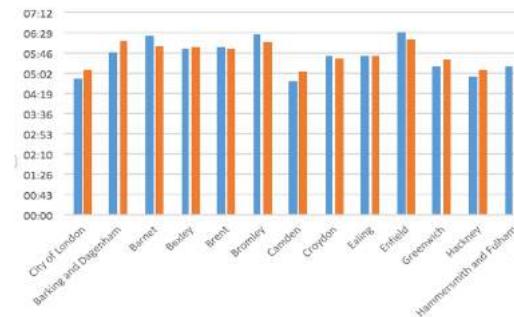




Stage 8: creating output

There are many options for output. Consider the best one for your purpose.

- A map can be ideal for such geographic datasets (search for London Data Store borough Excel KML).
- A chart showing a comparison of before an after statistics can be easily generated in excel.
- An highly interactive dashboard can be created with dataseedapp.com.



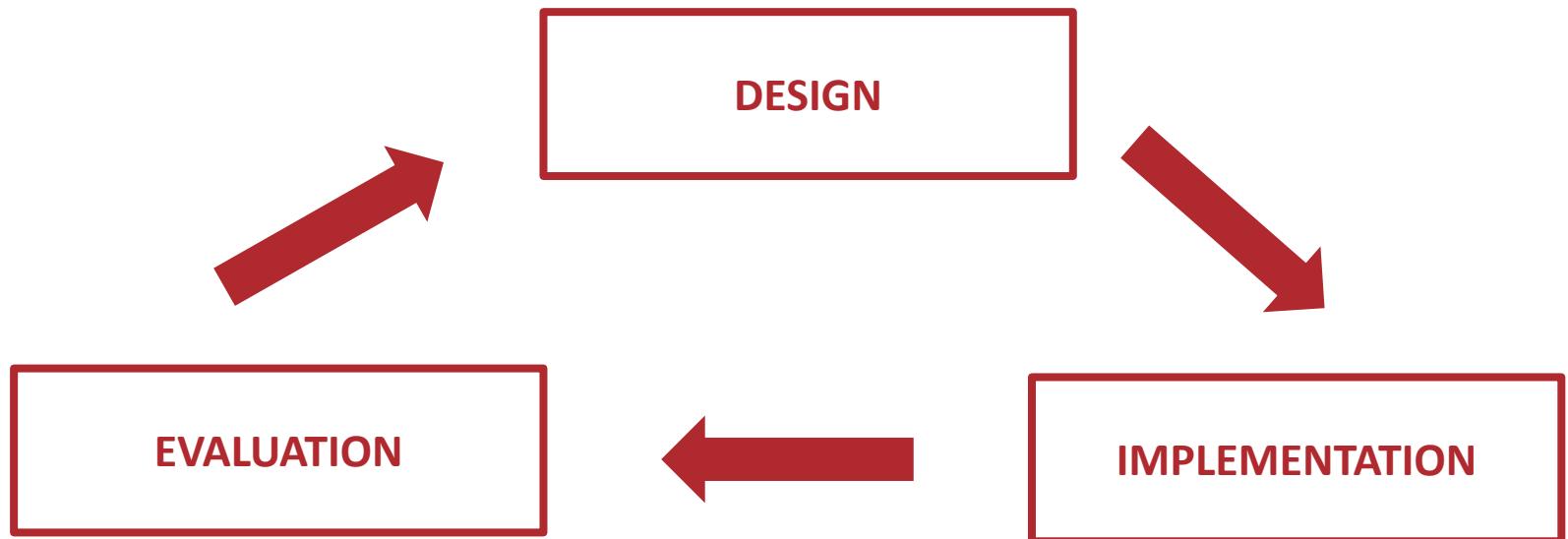


Civil Service
Learning

Open data in policy cycles



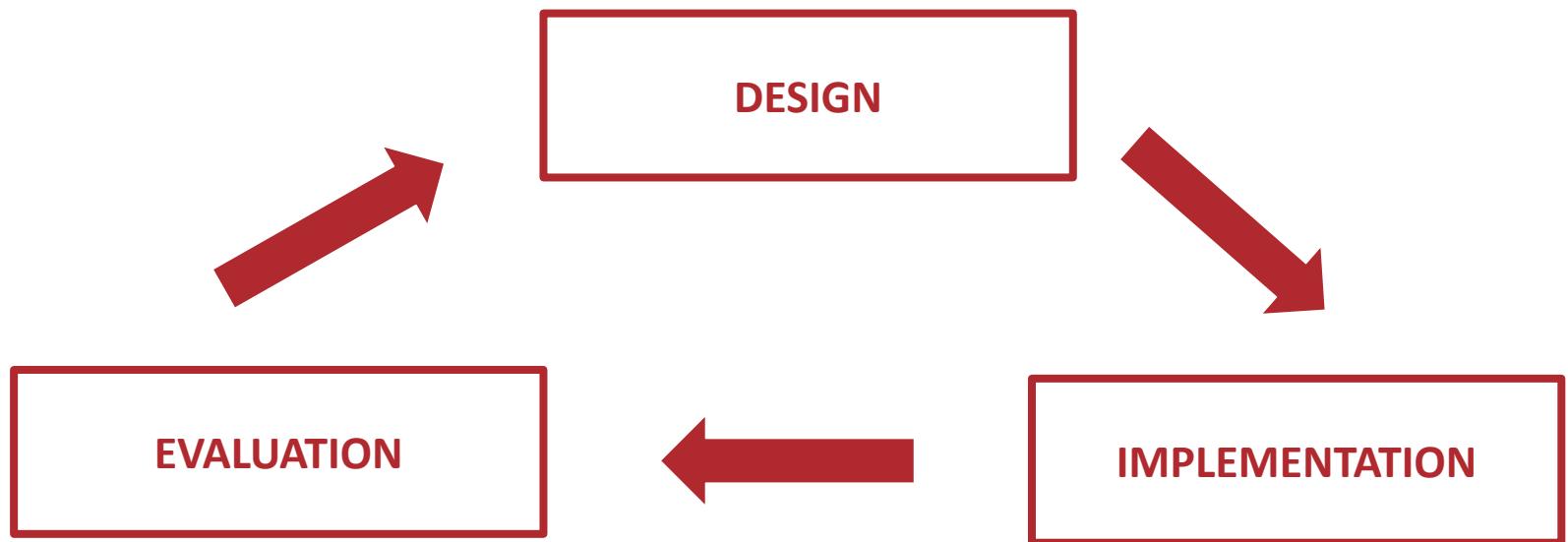
Public policy cycle





Question

What role can open data play in each stage of the policy cycle?





The city of Xalapa's waste management problem

Rubbish collection was frustrating the public as there was no clarity how the service operated.

The main problems:

- Where is the rubbish collected from?
- When is the rubbish collection?
- What day is the rubbish collected?
- What time is the rubbish collected?

Create a policy cycle/s

What role can open data play?





1. Put a GPS on every garbage truck (**design**)
2. Collected the data (**implementation**)
3. Opened the data (**implementation**)
4. Engaged community in design (**design**)
5. Analyzed the maps for new routes (**evaluation & design**)
6. Implemented new routes (**implementation**)
7. Evaluated the impact, repeating 1-3 (**evaluation**)
8. Civil society organizations organized a Hackathon (**design**)
9. Developed an app (exact location, timetables, citizens reports) (**implementation**)
10. Evaluated the solution (**evaluation**)





1. Put a GPS on every garbage truck (**design**)
2. Collected the data (**implementation**)
3. Opened the data (**implementation**)
4. Engaged community in design (**design**)
5. Analyzed the maps for new routes (**evaluation & design**)
6. Implemented new routes (**implementation**)
7. Evaluated the impact, repeating 1-3 (**evaluation**)
8. Civil society organizations organized a Hackathon (**design**)
9. Developed an app (exact location, timetables, citizens reports) (**implementation**)
10. Evaluated the solution (**evaluation**)

Is open data
just an input
in this cycle?

If not what
does this
change?



1. Input data for evidence based policy making

Data is drawn from a number of sources and is analysed to inform policy making.

Examples include:

- environmental impact of third runway at Heathrow
- impact of London fire station closures
- how to regulate peer to peer lending



flickr: lucianf



2. Output data for transparency and to encourage citizen interaction

Data is published as part of a transparency or other open government agenda. There is no immediate desire for the data to have any other impact.

Examples include:

- government spending data
- planning application data
- LIDAR data





Now we've used data as an input,
what about output and as a tool?



3. Tool to change behaviour

Where the data is the catalyst for change required by the policy.

Examples include:

- plastic bag usage data
- waste and emissions data
- pay gap data
- mobile coverage data
- broadband speed data



US Fish and Wildlife Service



Review

In your group identify a policy problem (1 sentence)

Using what you have learnt today, create list of key considerations and actions you need to take to solve this problem

What is the main thing you have learnt today that changes your thinking on how to solve this problem?



Review

How do the topics introduced so far affect you?

What is your main takeaway so far?



2. Output data for transparency and to encourage citizen interaction

The [ODI Open Data Certificates](#) will help you publish high quality, usable open data.

A certification is a mark of quality to let users know they can rely on the quality of your work.

Getting your data certified helps to establish trust in the quality of your data, showing compliance with the best publishing practices and gaining a broader user base.

With ODI Open Data Certificates, you can also get free advice on improving the quality of your data. Getting your Open Data Certificate is simple, fast and free.





3. Tool to change behaviour

Going beyond the publication of high quality data. Investment can also be made to bring about change in other ways.

The Open Data Challenge Series (ODI and Nesta) was a £500,000 series of challenges where the winning startups in each category were awarded a cash prize.

Independent analysis by PwC estimated the return of investment to be between

4-10x



**OPEN DATA
CHALLENGE**
SERIES



FOOD
OPEN DATA
CHALLENGE
SERIES

HOUSING
OPEN DATA
CHALLENGE
SERIES

**CRIME
AND JUSTICE**
OPEN DATA
CHALLENGE
SERIES

EDUCATION
OPEN DATA
CHALLENGE
SERIES

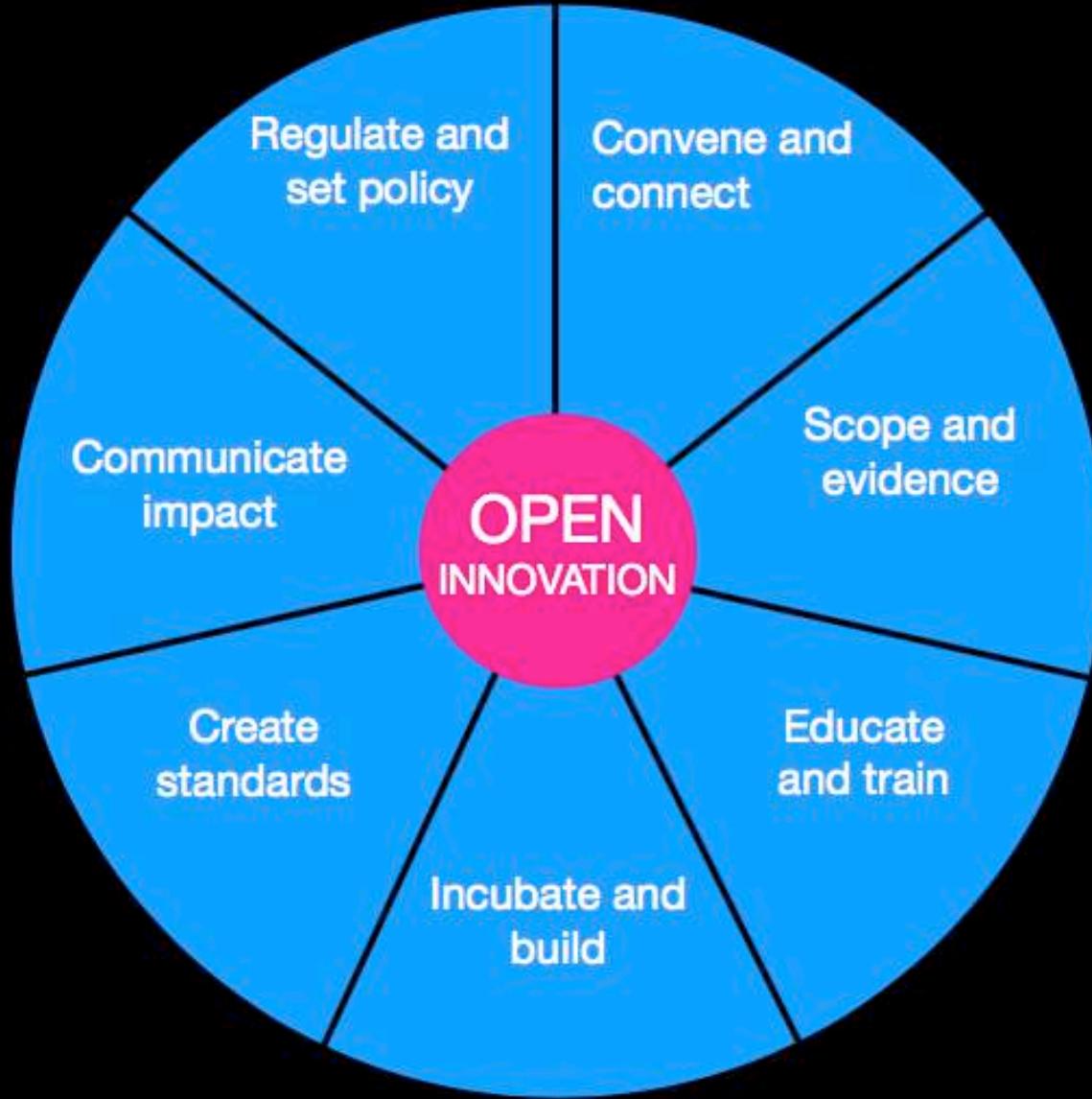
*Heritage
and culture*
OPEN DATA
CHALLENGE
SERIES

NESTA and OPEN DATA INSTITUTE in partnership

**ENERGY
+ environment**
OPEN DATA
CHALLENGE
SERIES

Nesta...

ODI





Review

London Fire Stations analysis was a story started by a participant in an ODI Training course like this one. What is your story? Who and what would you need to help you?

If I'm a policy maker, what do I now need to go and do?

What do I need to go and do next?

What is your main takeaway from the course overall?

What are you going to apply back in the workplace?



Civil Service
Learning

Thank you

We hope you enjoyed this
experience brought to you by

Delivered by



Now over to you! What are you going to do differently?



Exercise

How can open data be used in policy making?

How many areas / ways can you think of?

1. Input data for evidence based policy making.
2. Output data for transparency and to encourage citizen interaction.
3. Tool to change behaviour.



1. Input data for evidence based policy making

Data is drawn from a number of sources and is analysed to inform policy making.

Examples include:

- environmental impact of third runway at Heathrow
- impact of London fire station closures
- how to regulate peer to peer lending



flickr: lucianf



2. Output data for transparency and to encourage citizen interaction

Data is published as part of a transparency or other open government agenda. There is no immediate desire for the data to have any other impact.

Examples include:

- government spending data
- planning application data
- LIDAR data





3. Tool to change behaviour

Where the data is the catalyst for change required by the policy.

Examples include:

- plastic bag usage data
- waste and emissions data
- pay gap data
- mobile coverage data
- broadband speed data



US Fish and Wildlife Service



Exercise

What are the benefits of using data in policy making?

What are the risks of using data in policy making?

When should data not be used in policy making?

What are the challenges in using data in policy making? (especially when relying on third parties)





1. Input data for evidence based policy making

Key advantage

Evidence based policy making is less likely to fail or be rebutted.

Key disadvantage

The data may be complex and not appropriate for the analysis. The data may not back the popular view. Additionally the cost of acquiring and processing the data might be high.

A decision might have already been made and senior figures are looking for the analysis to support their decision.

Excessive use of data could lead to fears of a big brother state (e.g. Samaritan's radar).



flickr: lucianf



2. Output data for transparency and to encourage interaction

Key advantage

Reducing information asymmetry. Leading from transparency to trust in the open democratic process.

Key disadvantage

Repeat publication of such data is costly for little perceived gain.

Privacy plays a big factor in what can be released.

Prior agreements on release of data can be hard to overcome (Public Sector Mapping Agreement).

Trust in the quality of the data.

Potential misuse of the data.

5 stages of data grief.





3. Tool to change behaviour

Key advantage

Encourages others to implement policies for themselves.
Lowers cost of action.

Key disadvantage

Humans are not perfect economic agents and will often rage against the machine.

Businesses gamify the targets imposed to the detriment of the public.

In the early 2000s in the US, hospitals were incentivised to have low mortality rates. Redding Medical Centre gamified this by giving heart bypasses (a very expensive operation) to perfectly healthy people, thus increasing profits and bonuses based on a comparatively low mortality rate for their delivery of this operation.



US Fish and Wildlife Service



External impact?

What could be the impact of external factors on the success of your policy?

- As an input.
- As an output.
- As a tool to change behaviour.

How would you manage this external impact?





1. Input data for evidence based policy making

Impact

Costly to collect data whose quality and usefulness may vary.

There may already be a supported stance where the data will make no difference to the opinion (no change of behaviour).

The data modelling may be closed, thus introducing new risks.



flickr: lucianf



2. Output data for transparency and to encourage citizen interaction

Impact

Maintaining the open publication of commercially held data is difficult.

There are other factors ranging from IT to outdated security policies, to a lack of standardisation that play a big factor in data being made available.

Even if the data is published, it might not be in a usable format.





3. Tool to change behaviour

Impact

The data has no impact, due to the way it is communicated.

This is the case with the £10bn smart meter rollout.



US Fish and Wildlife Service



Review

How do the topics introduced so far affect you?

What is your main takeaway so far?