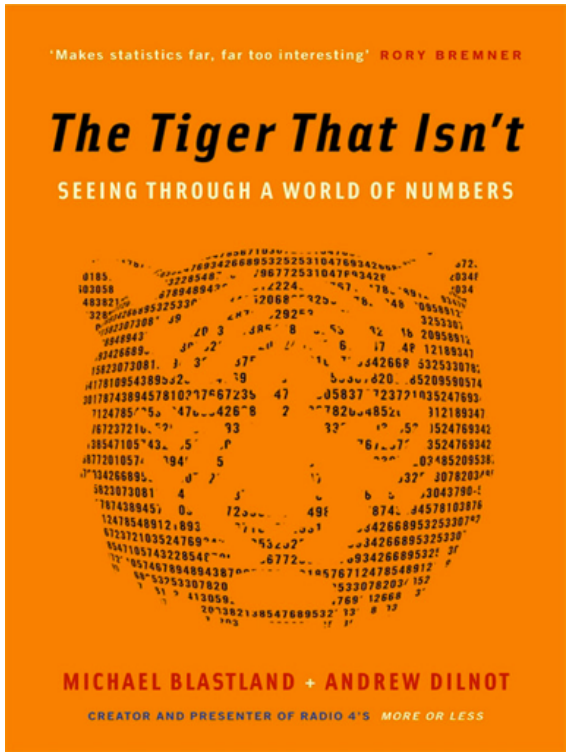

Finding Stories in Data
David Tarrant · @davetaz

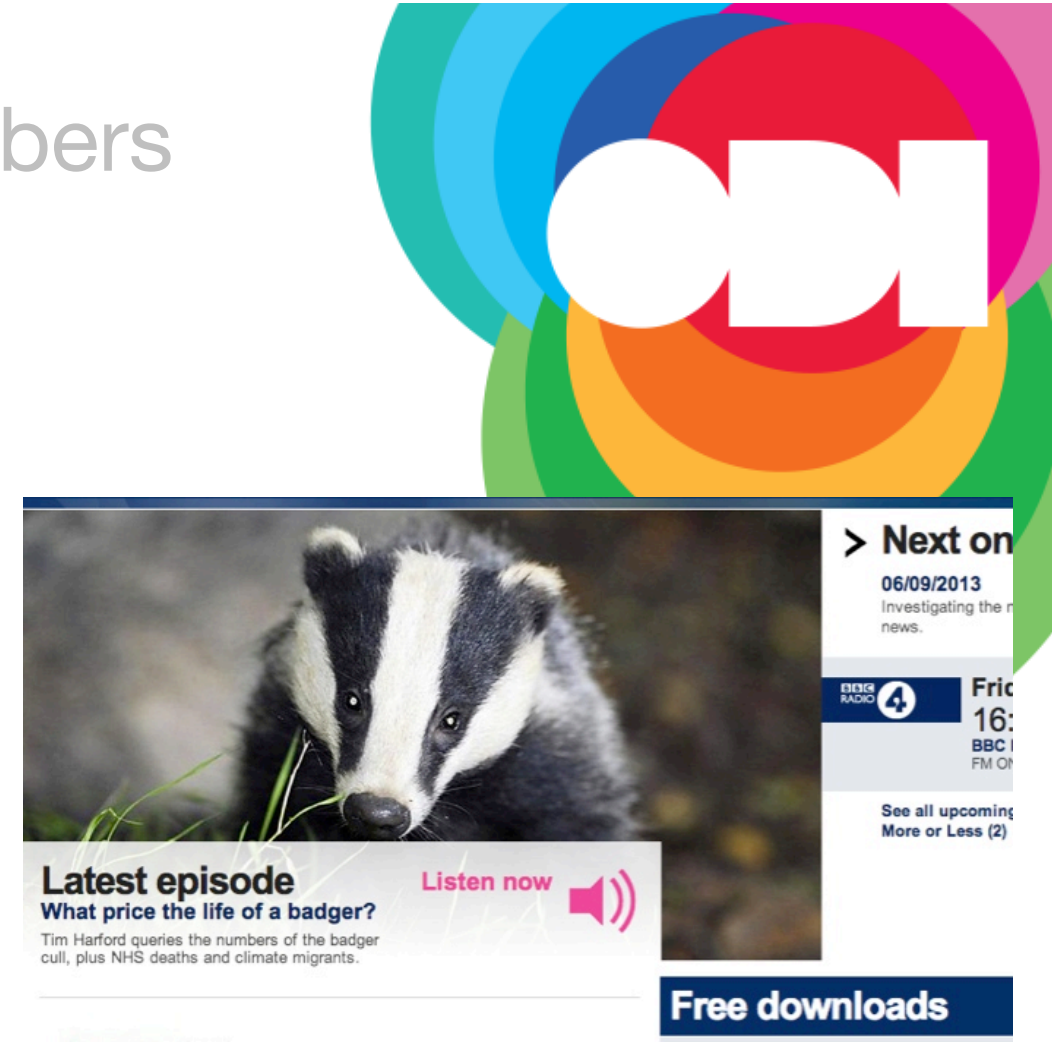
Misleading statistics
Enriching data and using pivot tables

Session 3
Statistical analysis of data

Misleading statistics

12 problems with numbers





1. Counting

Flickr: mattbrittaine

2. Big numbers

£300m

boost for childcare

1,000,000

new places

£1.15

per week

per child

3. Chance

Random events cluster

They do not evenly distribute

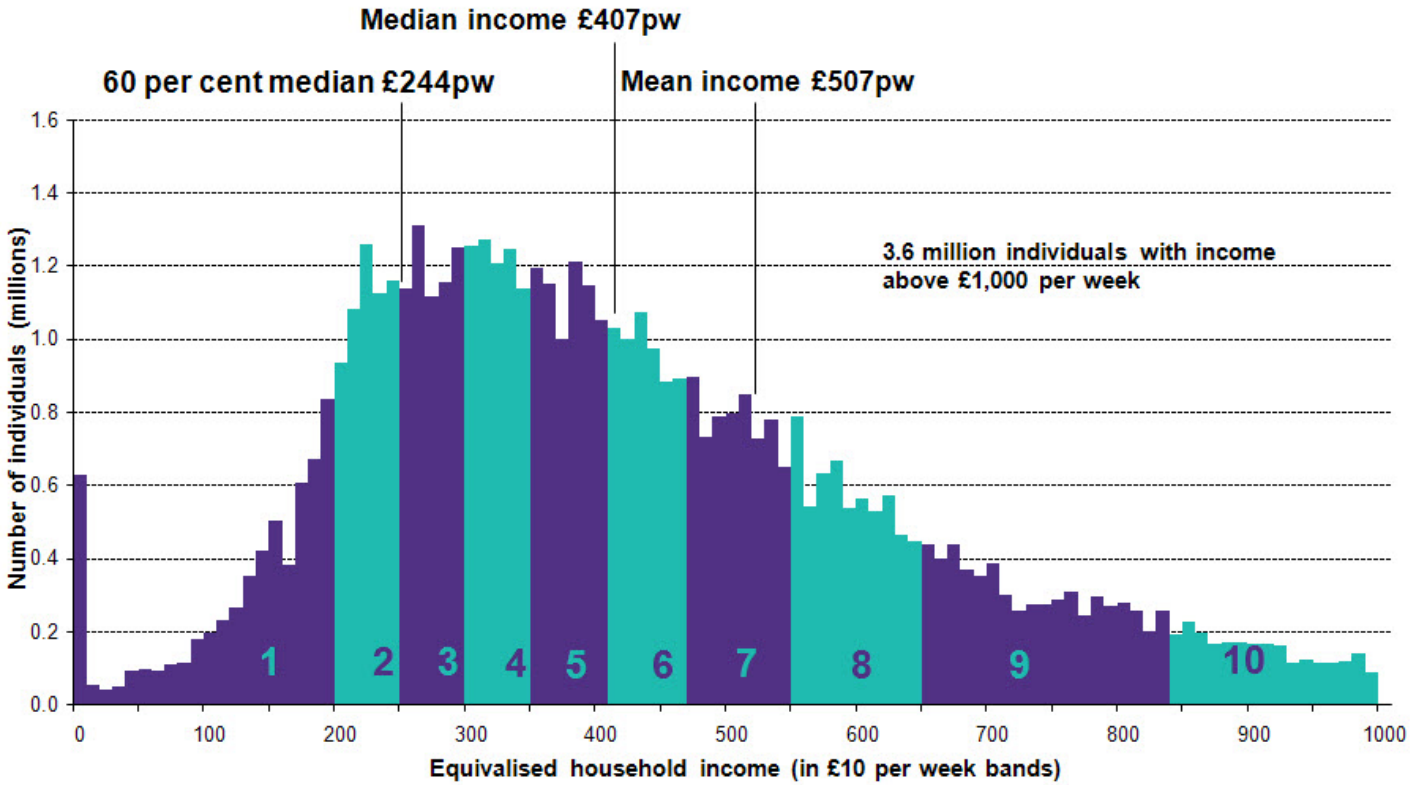
4. Fluctuation

Speed cameras installed on dangerous roads reactively

If you install on the top of a high wave, then the next wave is almost certainly going to be lower

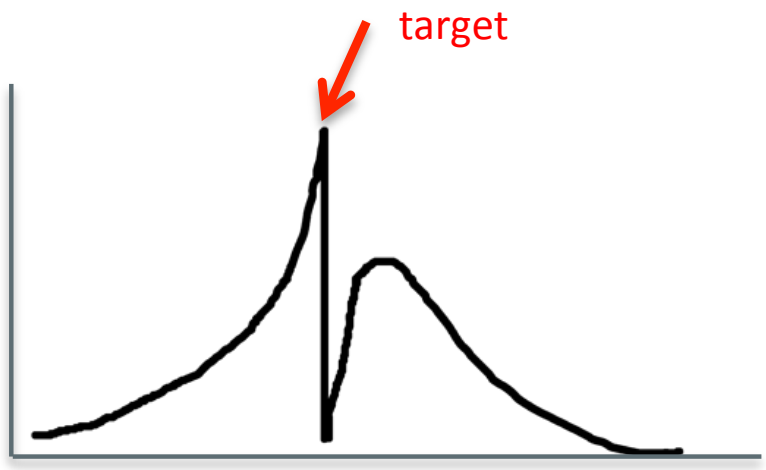
Very much related to correlation

5. Averages



6. Targets

Look at one aspect to measure an entire service.



Encourage gamification

7. Risk


SECTIONS
OTTAWA CITIZEN
HOME
NEWS
NATIONAL
LOCAL NEWS
Combined va
seizure risk ir
ELIZABETH PAYNE
More from
Published on: June 9, 2014 | Last Upda
CANCER RESEARCH UK
Donate
HOME
MENU
SEARCH
Home › About us › Cancer News › News report › Global cancer incidence predicted to increase by 75 per cent by 2030
Global cancer incidence predicted to increase by 75 per cent by 2030
News report
31 May 2012
In collaboration with the Press Association
The number of worldwide cancer cases is set increase by 75 per cent in the next two decades, according to researchers in France.
The scientists predict cancer cases will increase from 12.7 million in 2008 to 22.2 million by 2030.
Recent news
Investing in cancer research boosts economy as well as

8. Sampling

Between 2,000 and 5 million cases of norovirus in winter 2007/08.

Based upon 2,000 confirmed cases extrapolated from BMJ report based on sample size of …. **1**

The Telegraph

Home | News | World | Sport | World Cup | Finance | Comment | Cultu

Politics | Investigations | Obits | Education | Earth | Science | Defence

HOME » NEWS » UK NEWS

GPs urge millions hit by bug to stay at home

PHONEY NICKLE

The NHS advises symptoms

BBC NEWS

LIVE LI BBC NEWS CHANNEL

News Front Page
World
UK
England
Northern Ireland
Scotland
Wales
Business
Politics
Health
Medical notes
Education
Science & Environment
Technology

Last Updated: Friday, 11 January 2008, 12:02 GMT

E-mail this to a friend Printable version

Vomiting bug 'hits three million'

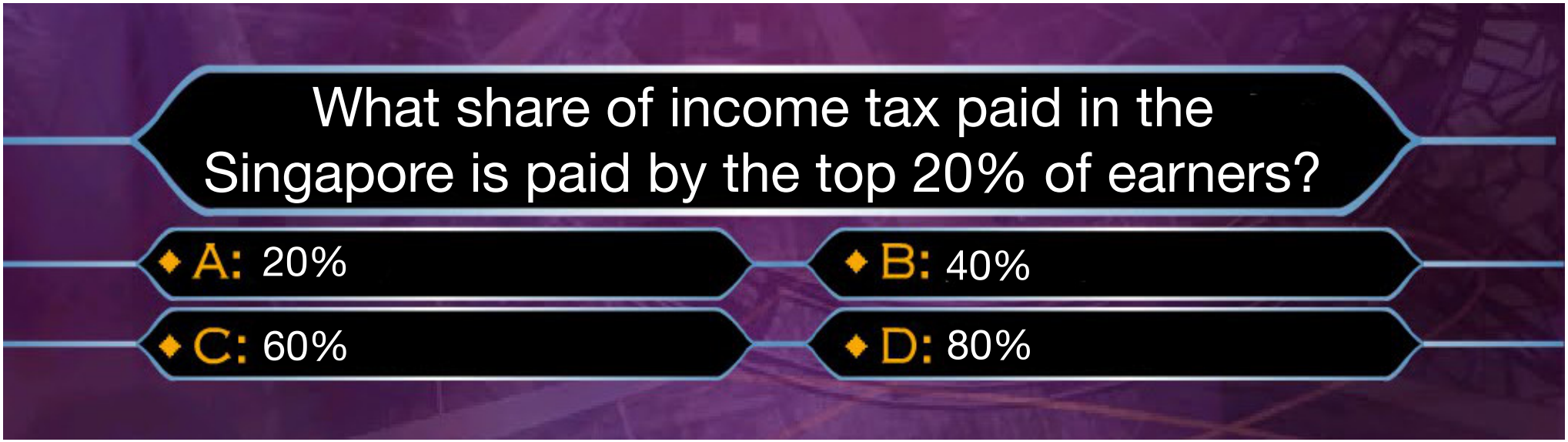
Almost three million people have been affected by the norovirus stomach bug so far this winter, figures suggest.

Surveillance from the Health Protection Agency shows cases in England and Wales are double those seen last year.

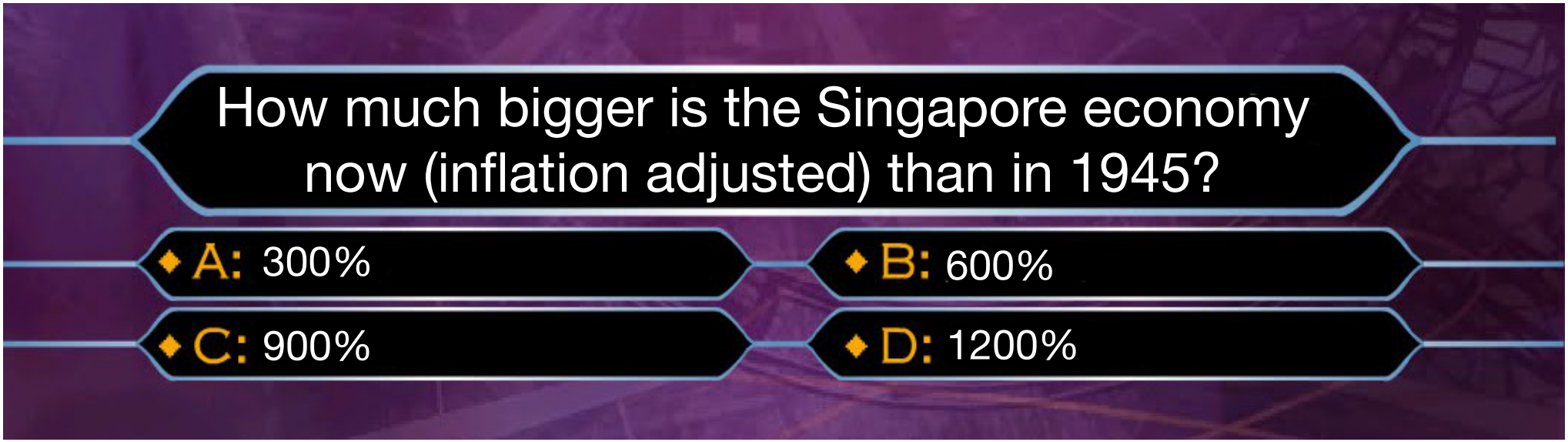
Doctors advise people to stay at home for 48 hours after

Norovirus causes sudden vomiting and diarrhoea

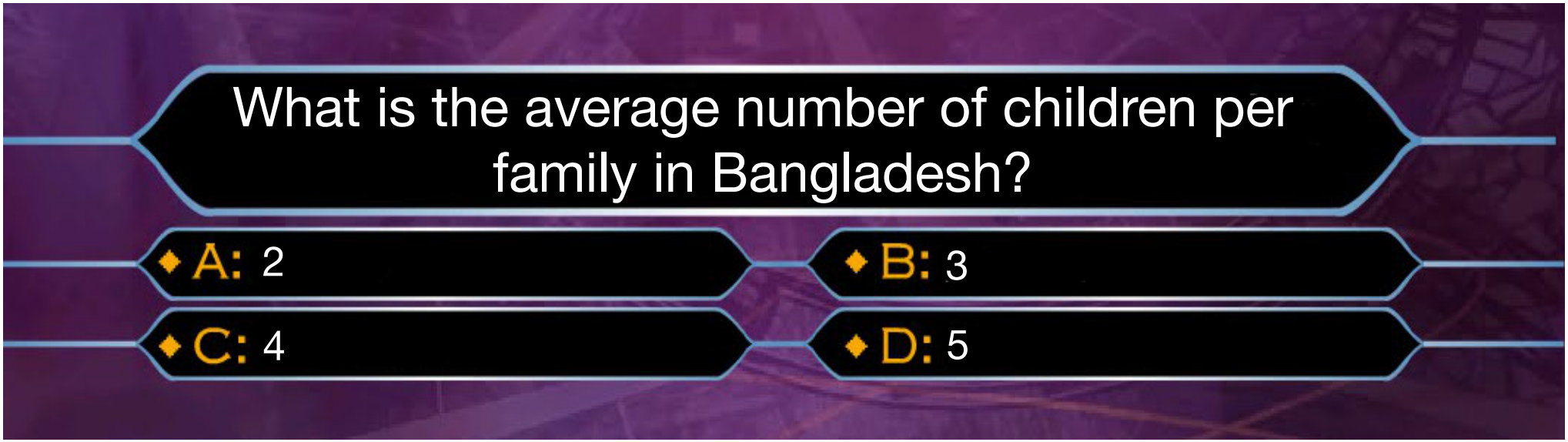
9. Data (known unknowns)



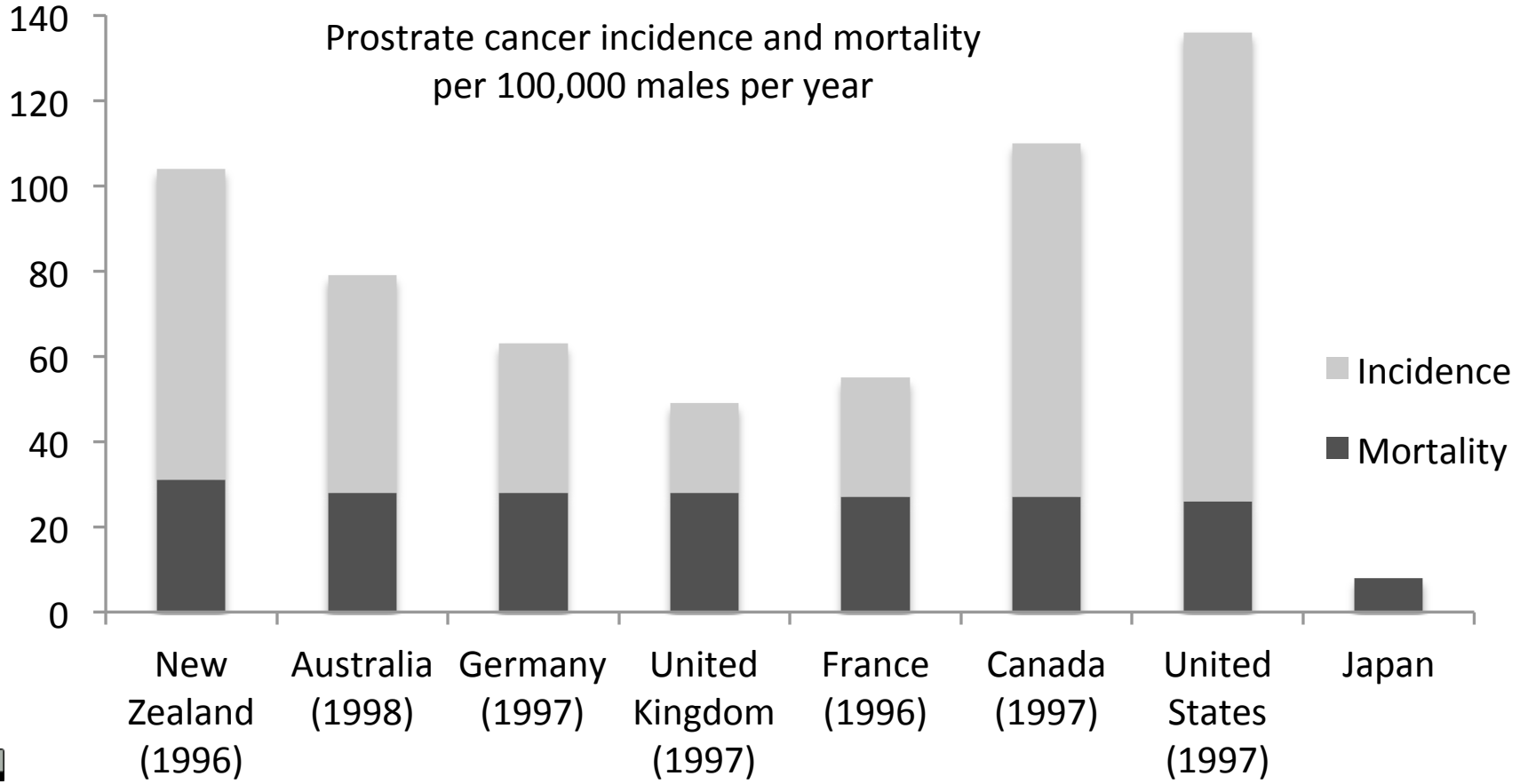
9. Data (known unknowns)



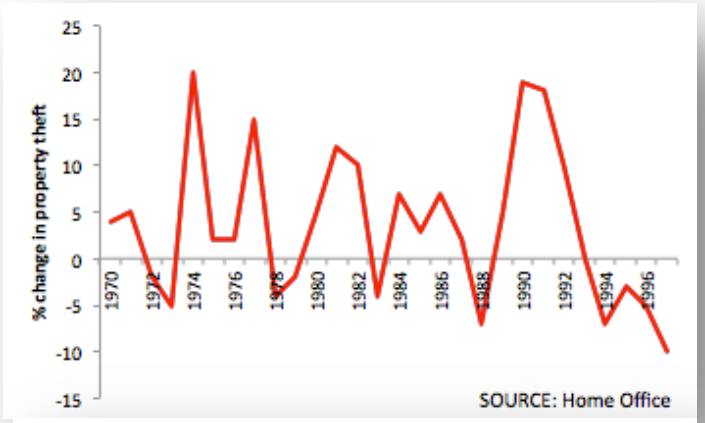
9. Data (known unknowns)

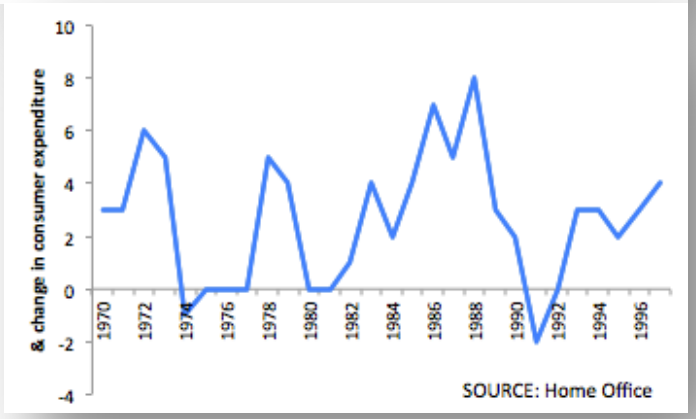


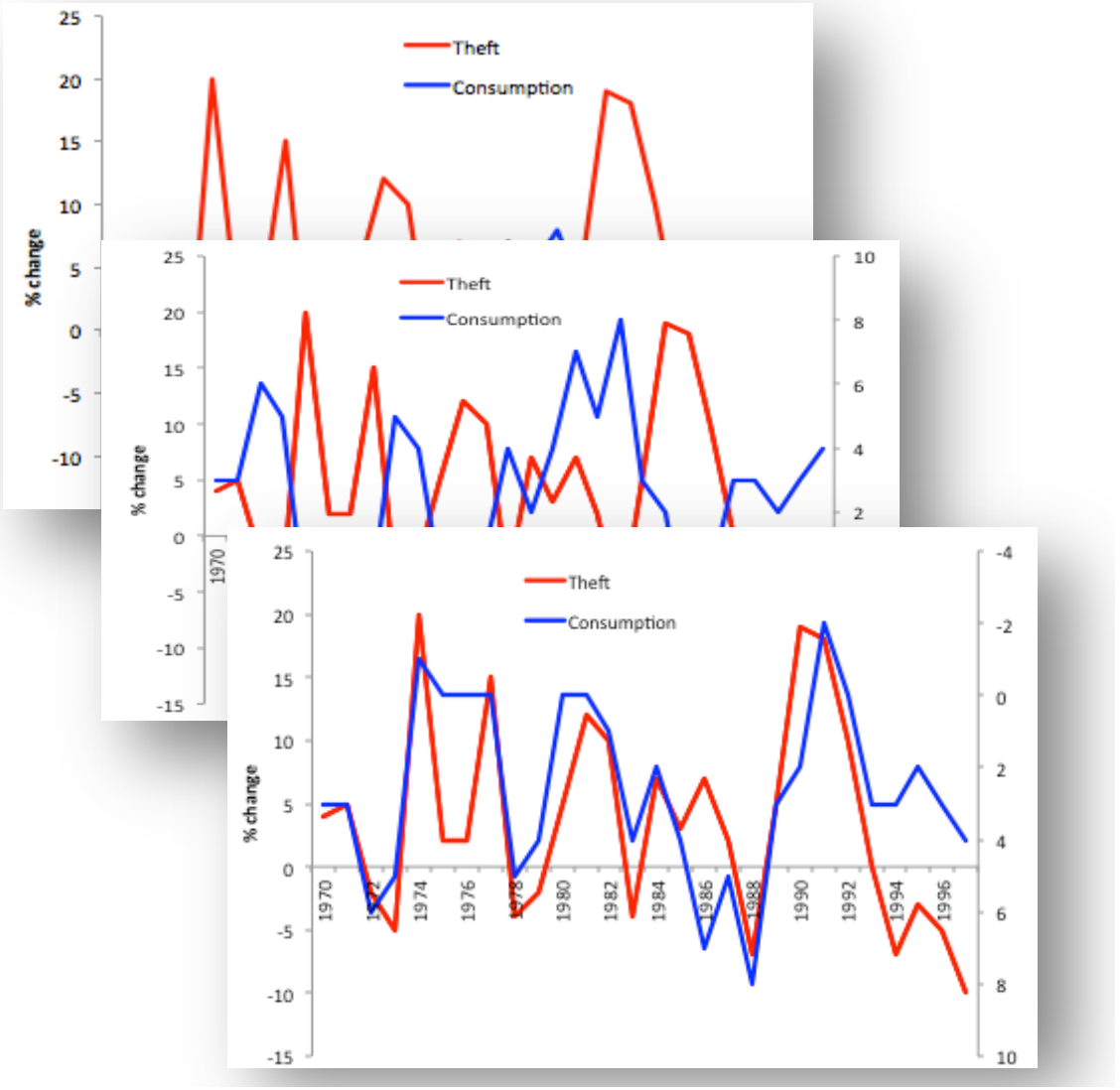
10: Comparison



11. Correlation







12. Percentages

Know the difference between a **percentage** and a **percentage point**.

VAT increased from 17.5% to 20% on January 2011.

This is a rise of 2.5 percentage points **not** a rise of 2.5%.

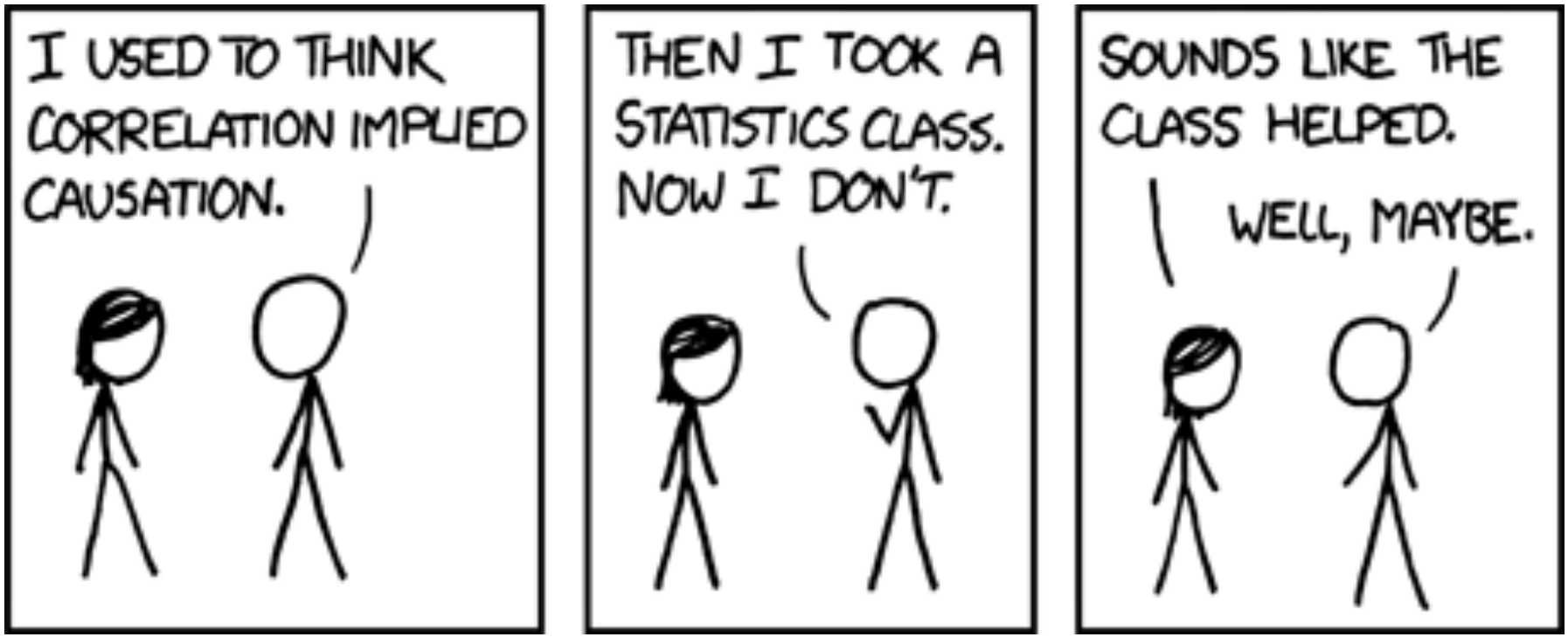
How much would a rise in 2.5% actually be?

From: Stories and Statistics by Frank Swain

5 ideas for finding a story

1. **Create a ranking**
   Sort columns to explore top and bottom property prices
2. **Compare groups**
   Newly built properties to established residential building
3. **Use sum/average/min/max functions**
   Find the median price of a property
4. **Use pivot tables to compare groups**
   Compare the prices of different geographies in the UK
5. **Look for anomalies in the data**
   e.g. in the “date” or “postcode” columns

XKCD



Prepare

Prepare

2.1 **CLEAN**

2.2 **TRANSFORM**

2.3 **COMBINE**

2.4 **ENRICH**

2.5 **ANALYSE**

Enriching data and using pivot tables

opencorporates

The Open Database Of The Corporate World

Aggregator/Enabler

We have information on 70,597,888 companies

SEARCH

search companies search officers

Filter by jurisdiction

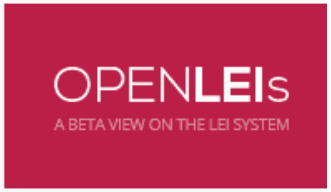
1,298	Abu Dhabi (UAE)
144,755	Alaska (US)
40,157	Albania
899,455	Arizona (US)
46,537	Aruba
165,582	Bahamas
99,185	Bahrain
88,563	Bangladesh

Just released: OpenCorporates API v0.3

Corporate network data, financial accounts, complex filters, and more. Read more

Announcing Open LEIs

Today, OpenCorporates announces a new sister website, Open LEIs, a user-friendly interface on the emerging Global Legal Entity Identifier System. Read more



New! Just added: Open corporate network data

Read more about this important new feature

Exercise

Enriching a dataset containing company names (e.g. transactions) with company data from OpenCorporates

Sense-checking

The best way to sense-check is to get a second pair of eyes to help you.

Any stories of common mistakes you'd like to share?


Thank-you