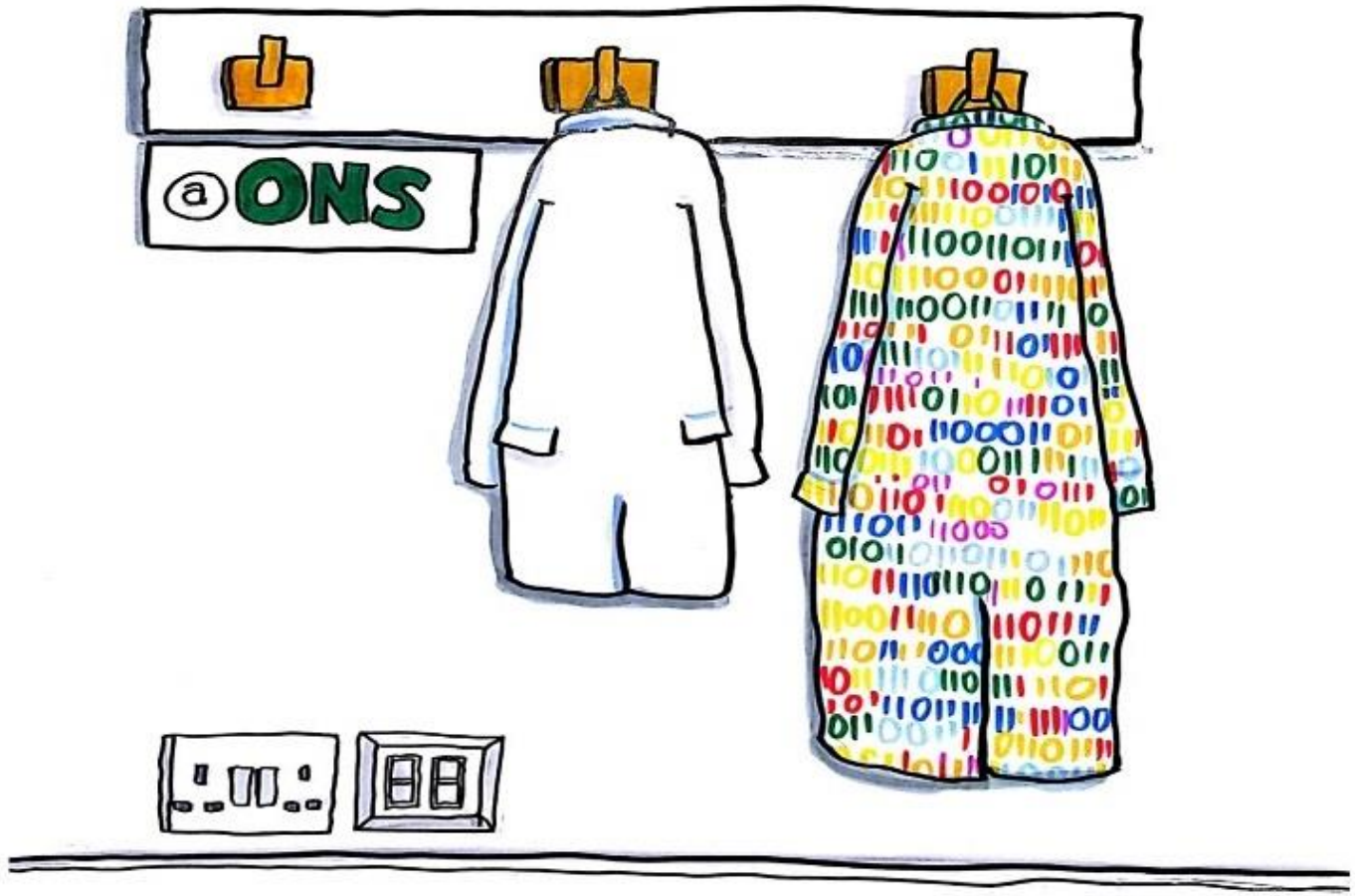
Data science for public good

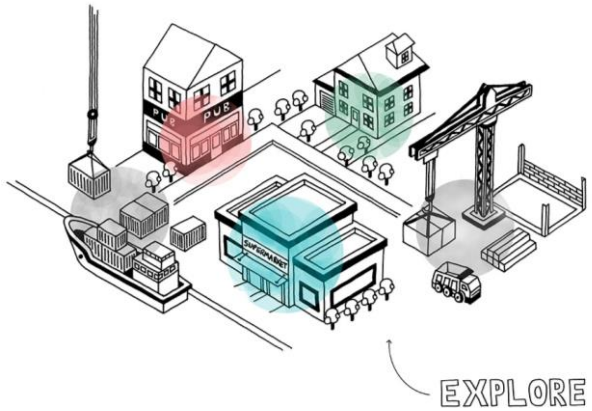
Tom Smith
ONS Data Science Campus
@_datasmith



Data Science Campus

web: datasciencecampus.ons.gov.uk
email: datasciencecampus@ons.gov.uk
twitter: @DataSciCampus

Office for National Statistics



Economy

GDP
Inflation
Labour market
+++

People

Population
Census
Incomes
+++

World

Trade
Sustainable Development Goals
+++

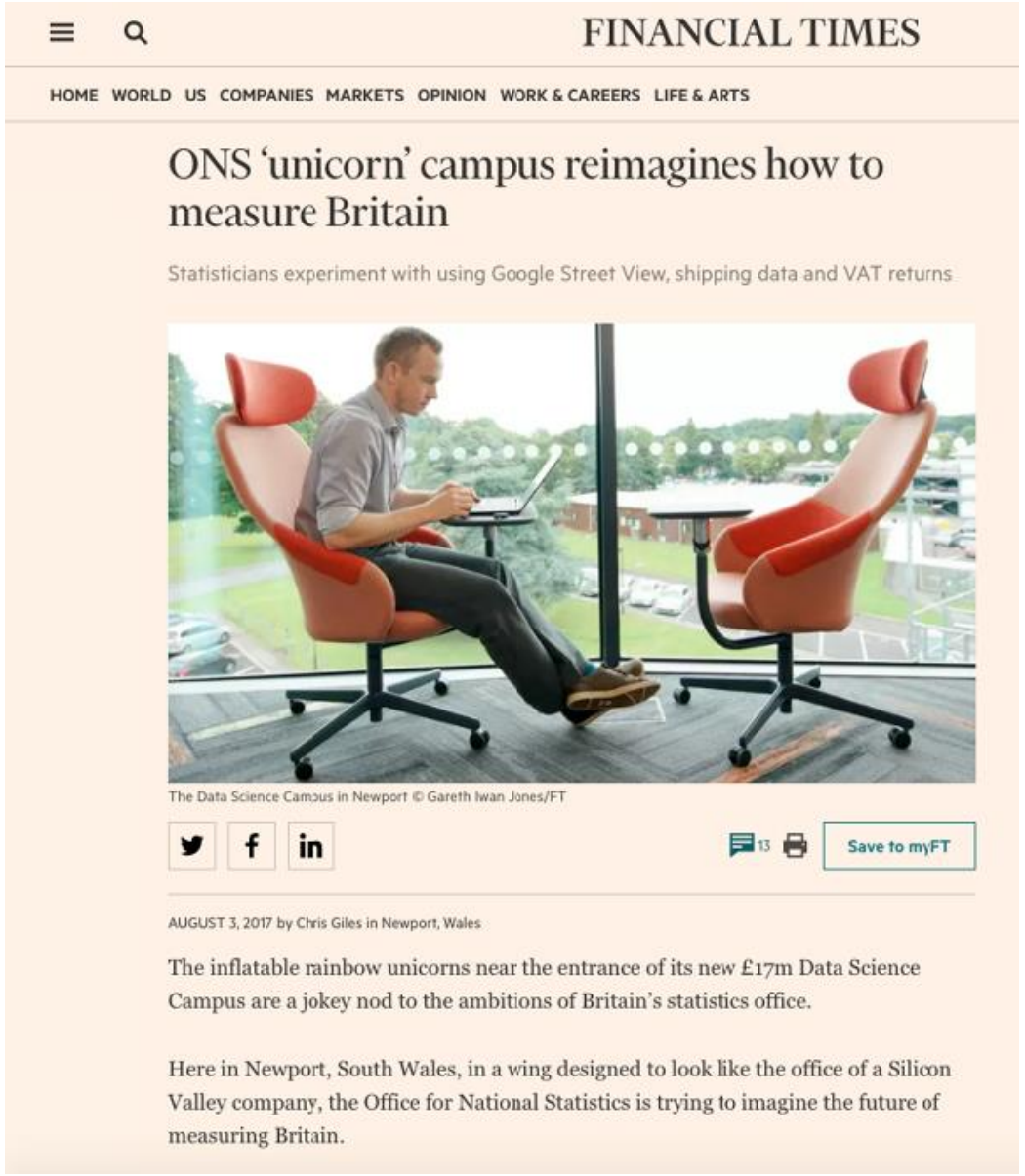
Data Science Campus creation

"Although **better use of [data]** has the potential to transform the provision of economic statistics, ONS will need to **build up its capability** to handle such data.

This will take some time and will require not only **recruitment of a cadre of data scientists** but also **active learning and experimentation**.

That can be facilitated through **collaboration with relevant partners** – in academia, the private and public sectors, and internationally."

Independent Review Economic Statistics
Professor Sir Charles Bean, 2016, p.11



FINANCIAL TIMES

HOME WORLD US COMPANIES MARKETS OPINION WORK & CAREERS LIFE & ARTS

ONS 'unicorn' campus reimagines how to measure Britain

Statisticians experiment with using Google Street View, shipping data and VAT returns

The Data Science Campus in Newport © Gareth Iwan Jones/FT

13 Save to myFT

AUGUST 3, 2017 by Chris Giles in Newport, Wales

The inflatable rainbow unicorns near the entrance of its new £17m Data Science Campus are a jokey nod to the ambitions of Britain's statistics office.

Here in Newport, South Wales, in a wing designed to look like the office of a Silicon Valley company, the Office for National Statistics is trying to imagine the future of measuring Britain.

1939 - London Transport workers manually examine over 4 million tickets to identify most and least popular routes

Gerry Cranham/Fox Photos/Hulton Archive/Getty Images

Transport for London

WiFi data collection

We are collecting WiFi data at this station to test how it can be used to improve our services, provide better travel information and help prioritise investment.

We will not identify individuals or monitor browsing activity.

We will collect data between Monday 21 November and Monday 19 December.

For more information visit: **tfl.gov.uk/privacy**

MAYOR OF LONDON

TRANSPORT FOR LONDON
EVERY JOURNEY MATTERS

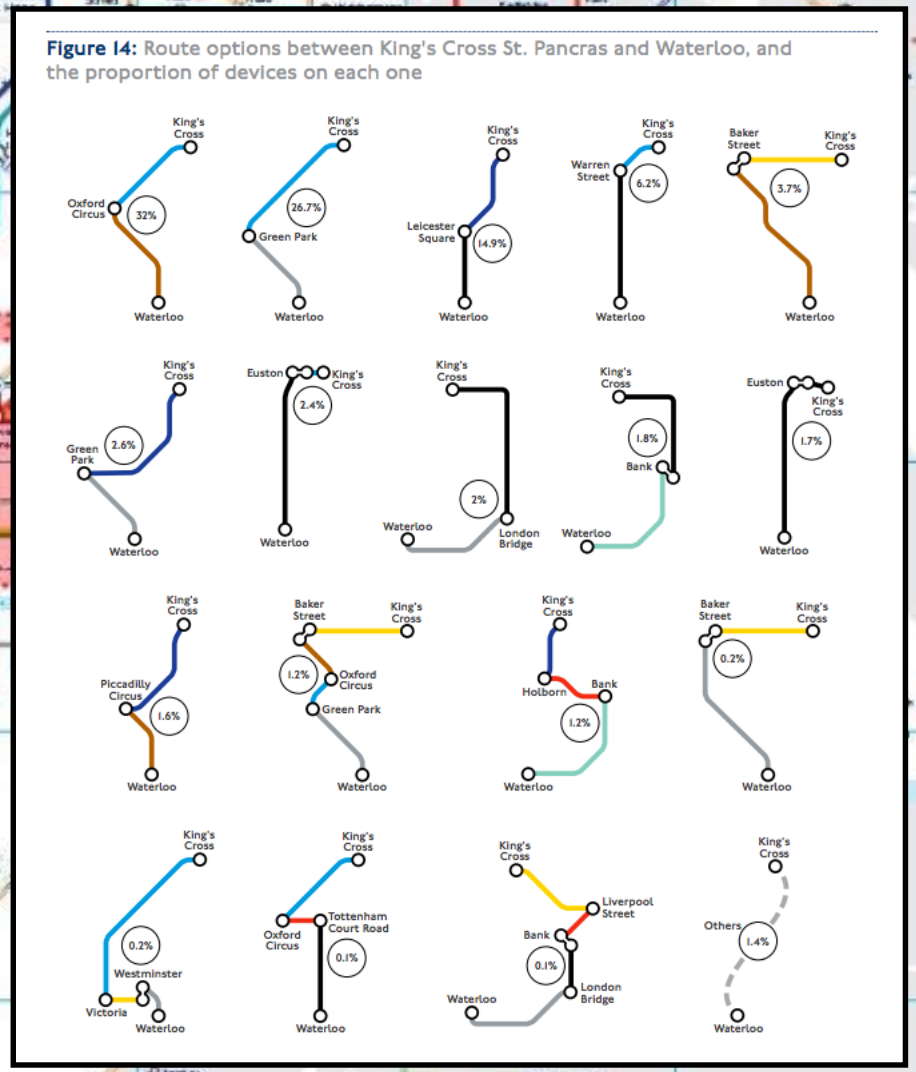


Figure 14: Route options between King's Cross St. Pancras and Waterloo, and the proportion of devices on each one

Transport for London 2016 pilot, assessing journeys by WiFi access

How Data Science helped identify potential savings of over £581m for the NHS

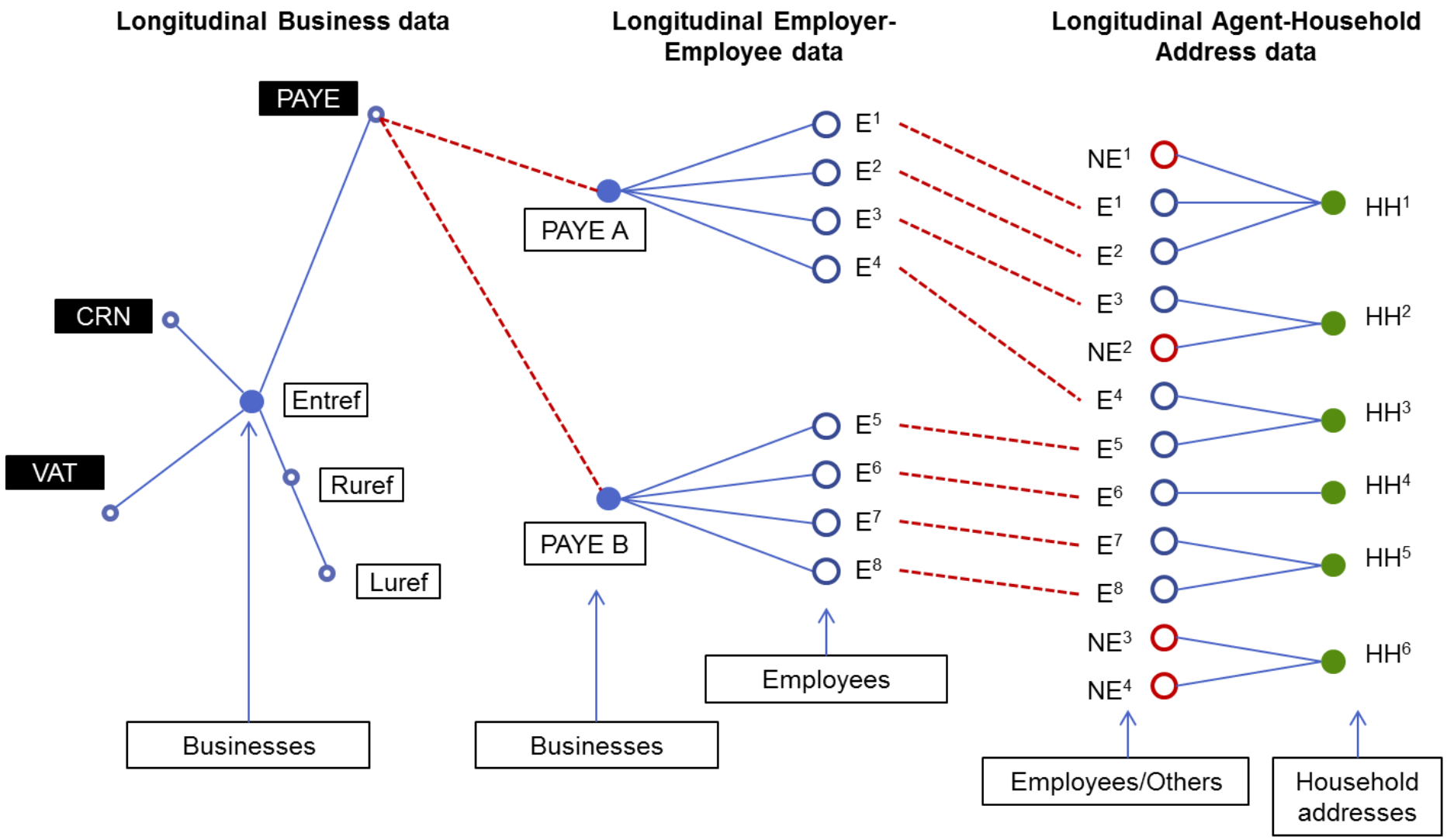
Abi Giles-Haigh, 31 January 2018 - Digital data and technology, People and Skills

Surveys should be the last resort

Focus on high quality surveys that tell us things we don’t already know

Linked administrative data is first prize

Linked administrative data sources (UK)



Using business tax data in GDP

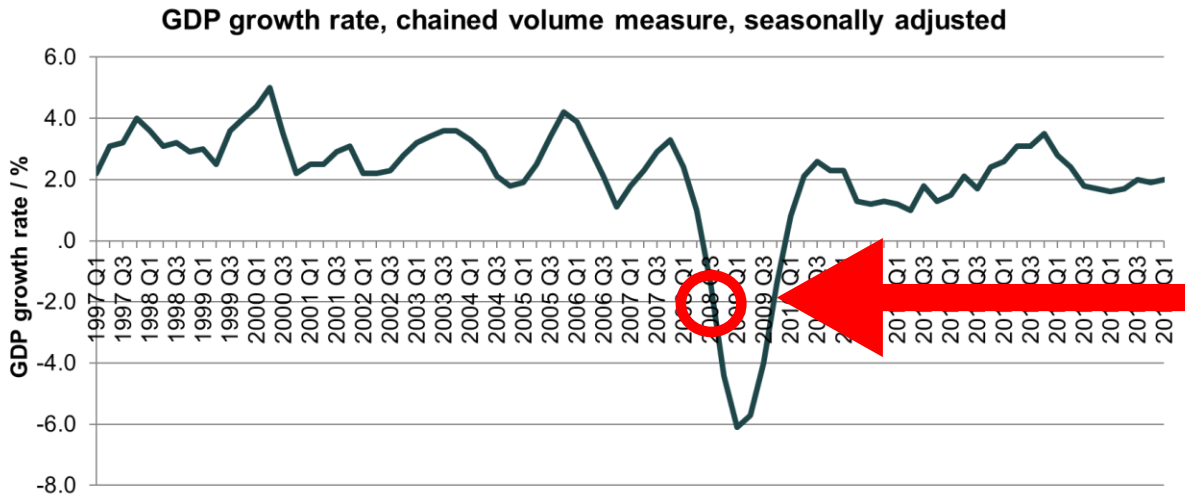


Fig 1. UK GDP Growth Rate

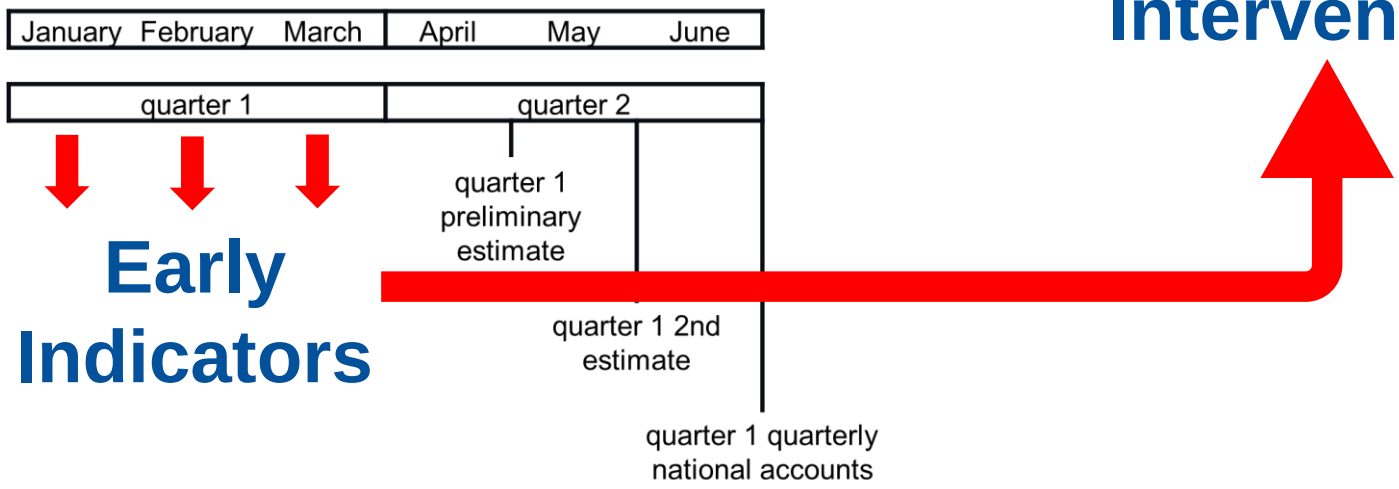


Fig 2. ONS National Accounts Publication Timetable

-6%

Change in UK GDP between first quarter of 2008 and second quarter of 2009

5 years

Length of time from 2008 for the UK economy to return to pre-recession size

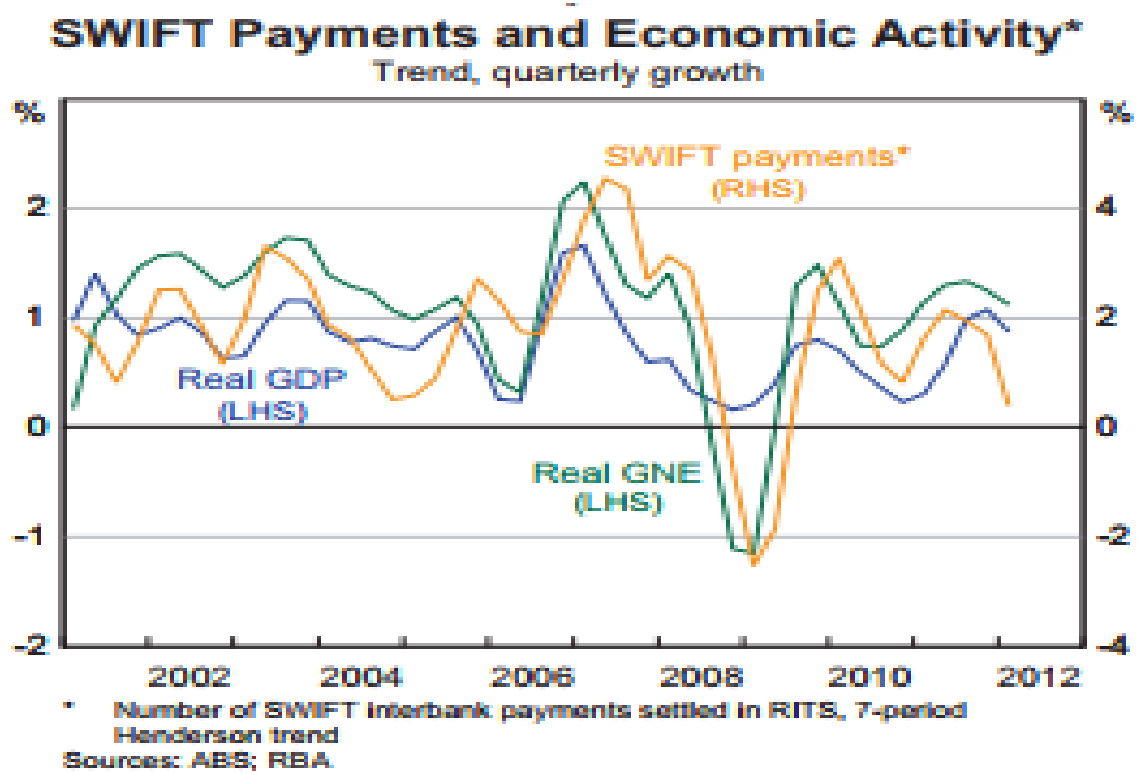
£12b

Estimated value for earlier identification of 2008 downturn

There’s a lot of data outside government …

Payments data for regional indicators

- Identifying rapid, local economic indicators - breakdowns by geography, industry, product, credit / debit card, on-line payment, international
- Collaboration with Barclays, 2-way secondments
- What can we learn about payments data?



Payments data for regional indicators

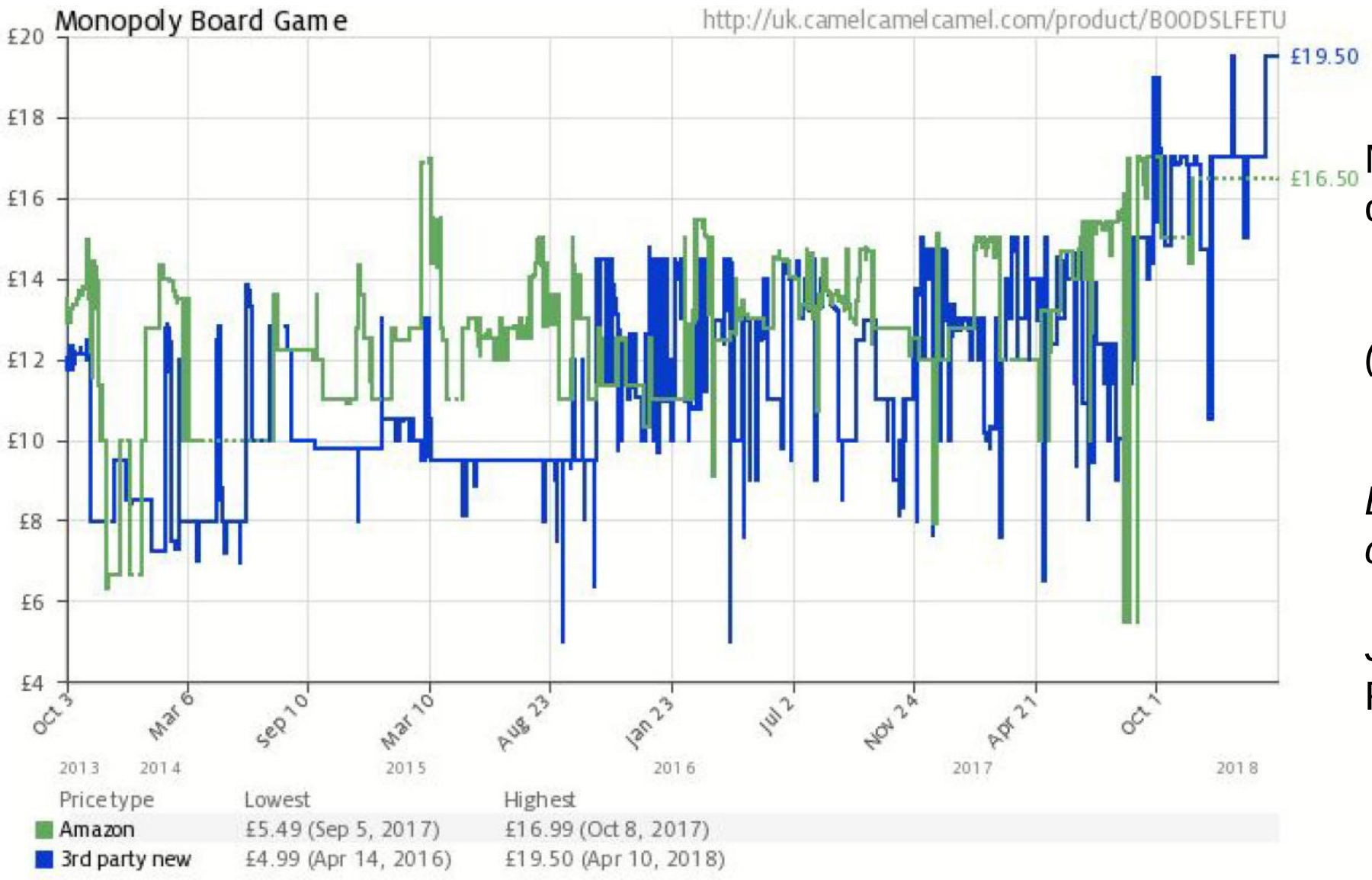
- Financial data held by banks
  - No sensitive or personally identifiable data shared
  - All outputs are aggregate and non-sensitive

- Hypotheses we are exploring include
  - Payments data as proxy for retail sales by sector & time (eg night time economy)
  - Payments data as proxy for private household consumption
  - Payments data can improve the accuracy of GDP nowcasting

- Data sources potentially available through secondments:

Consumer	Electronic payments	Business
Debit Card spend Credit card spend Personal Loans Mortgages Savings accounts Insurance	POS data ATM data Online gateway data (online purchases) Peer-to-peer	Merchant & Acquirer data Corporate Cards Business Bank products Corporate bank Products Investment bank products

Prices and volatility



Monopoly price fluctuation over 4 year period

High = £19.50

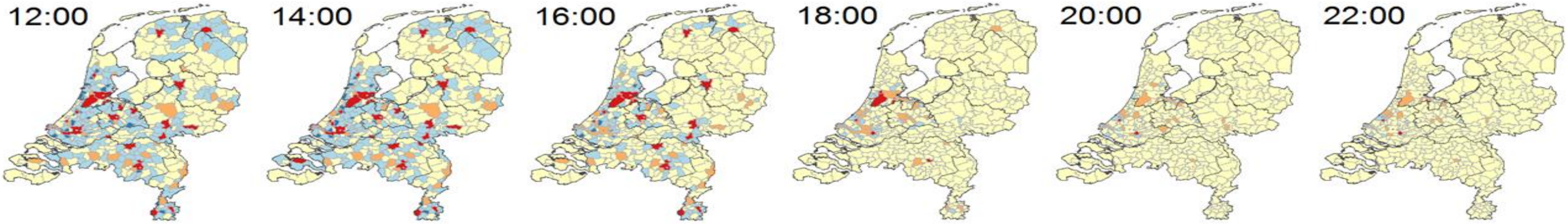
Low = £4.99

(Data from camelcamel)

Big Data is changing how consumer markets work

James Plunkett, 2017-18 Rybczynksi Prize Essay

Centraal Bureau voor de Statistiek



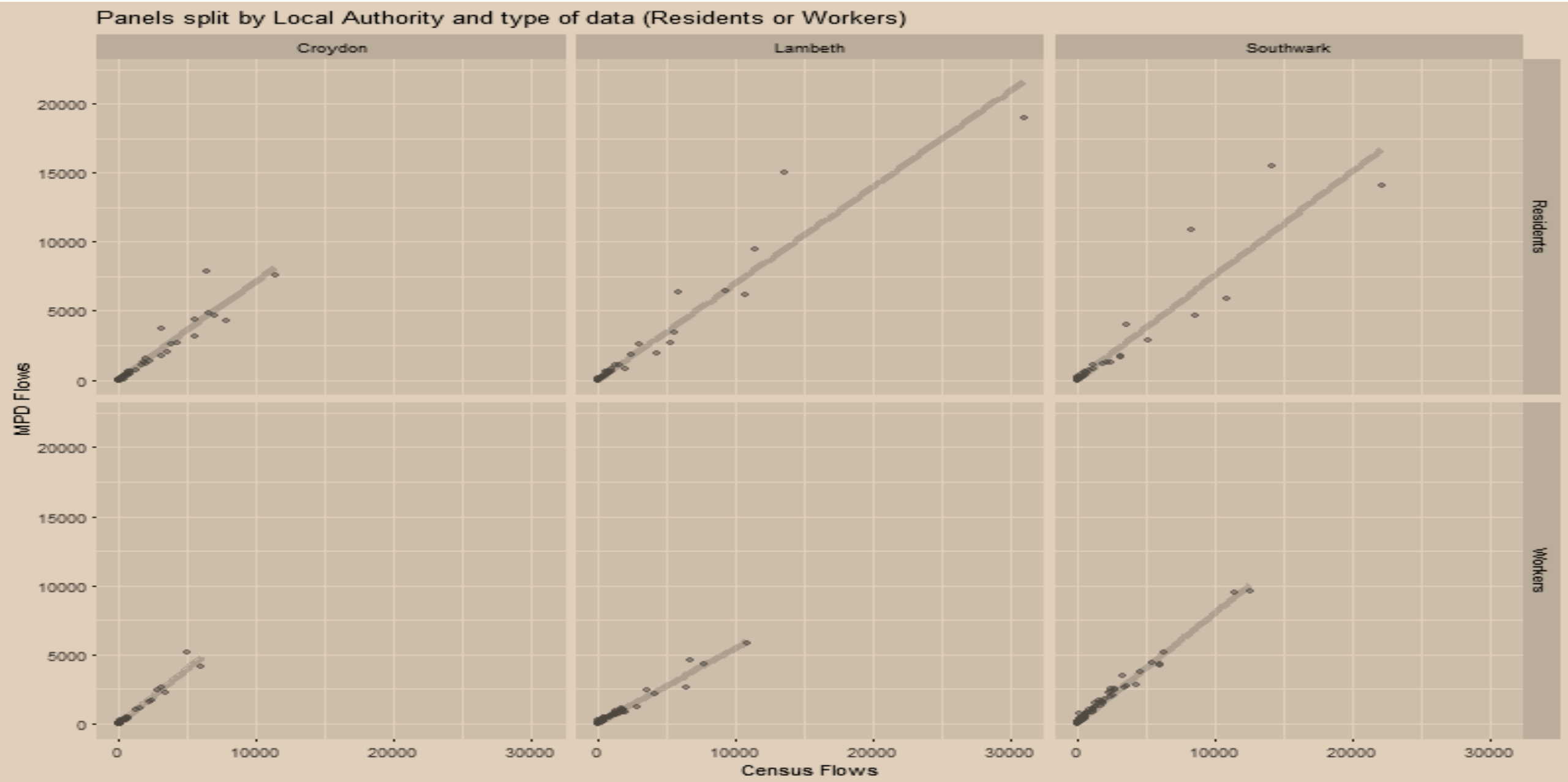
- Very sparsely populated
- Sparsely populated
- Normally populated
- Densely populated
- Very densely populated

Mobile phone data and daytime population

GDP and Road Traffic

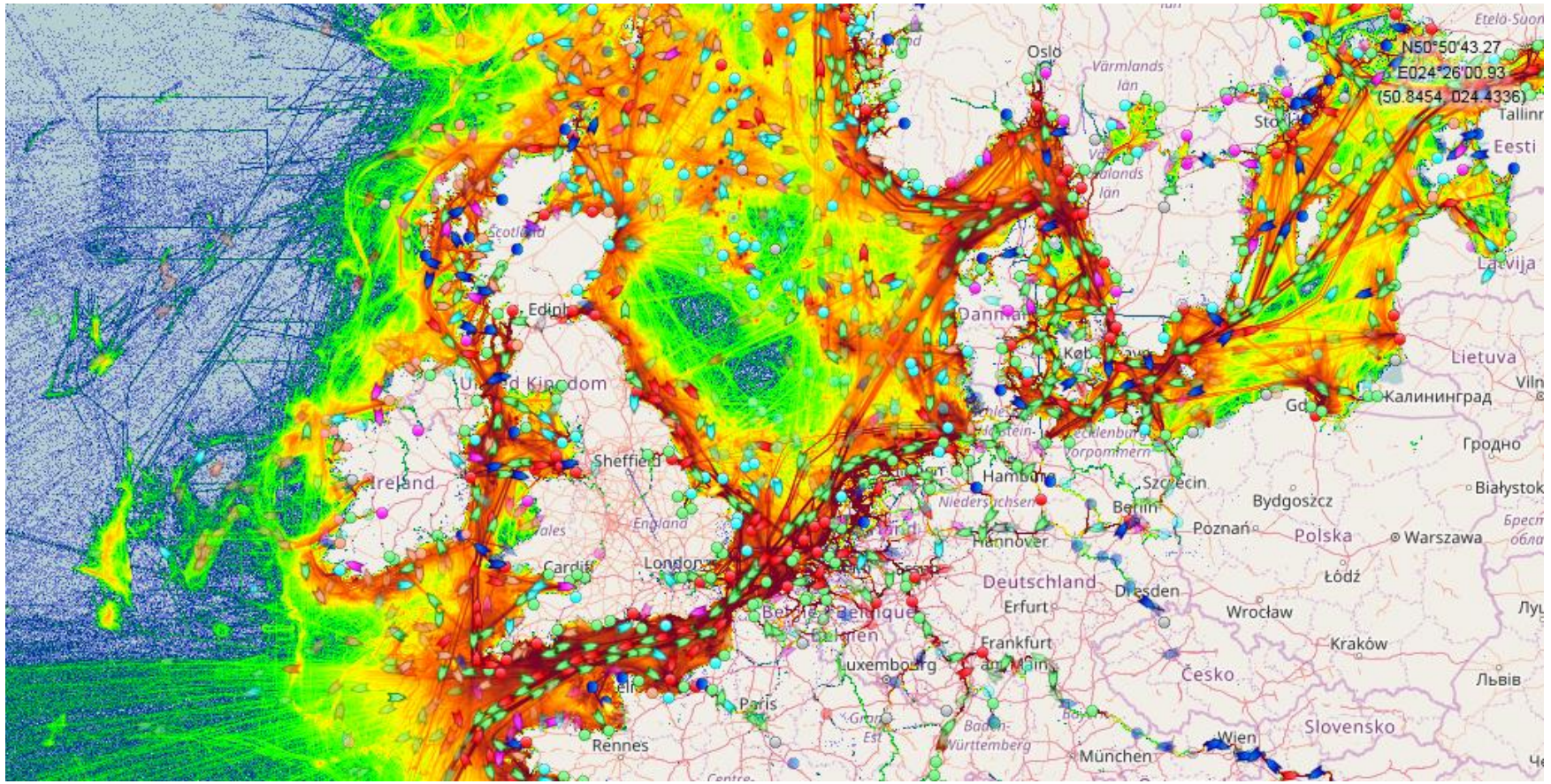
- Road Traffic predicts GDP publication
- Correlation
  - 82% from 2010-Q3 until 2014-Q4
  - 91% from 2011-Q2 until 2014-Q4

Travel to work, Census & mobile phones



Movement data - tracking ships to understand trade

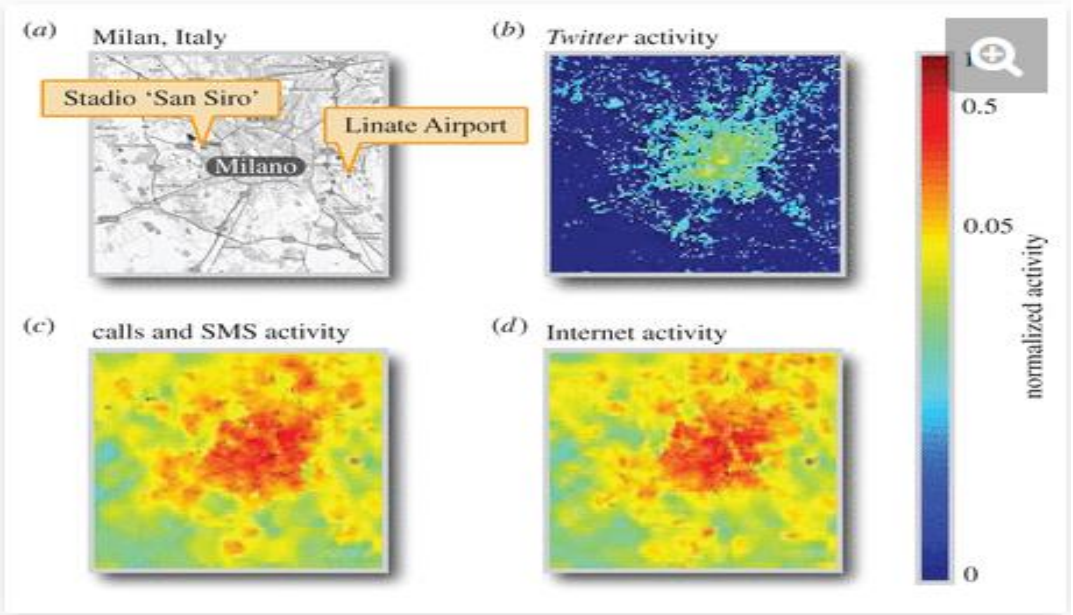
- Global GPS system for all shipping, rapid updates
- Leading indicator of trade? Ships in port? Inter-regional trade?



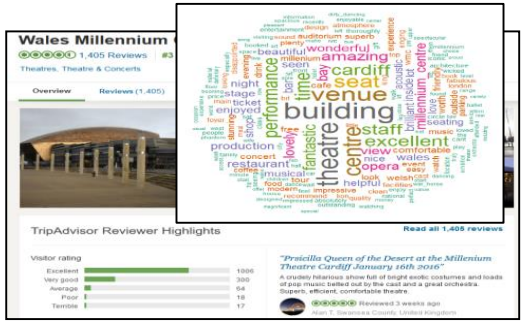
Estimating tourism levels through social media

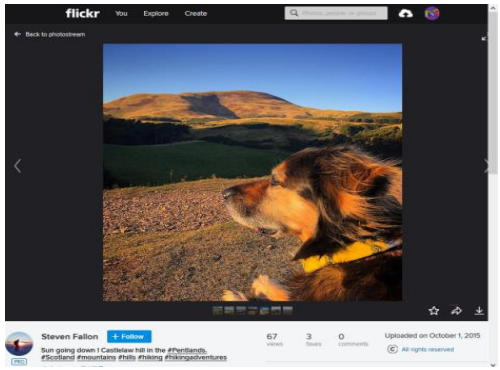
Research questions:

- Alternative data source for quality assurance of International Passenger Survey
- Nationality based under-representation
- Domestic travel trends
- Small area statistics, crowd size estimation
- Google analytics web journey

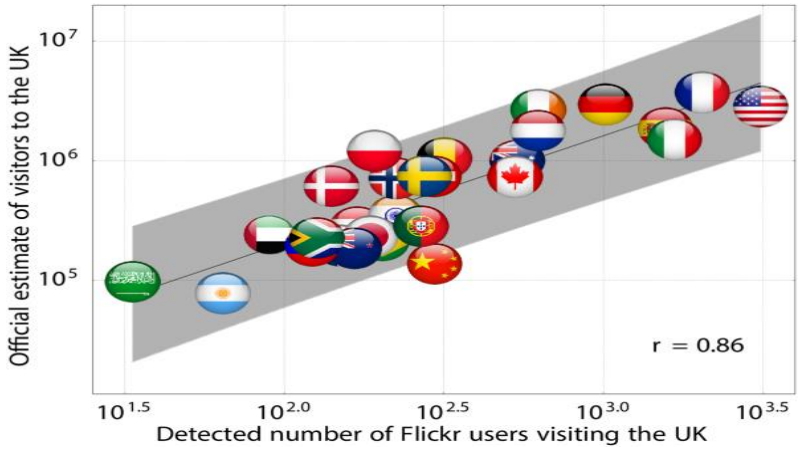


Visualisation of geo-located Flickr data





Machine Learning classification of photo tags

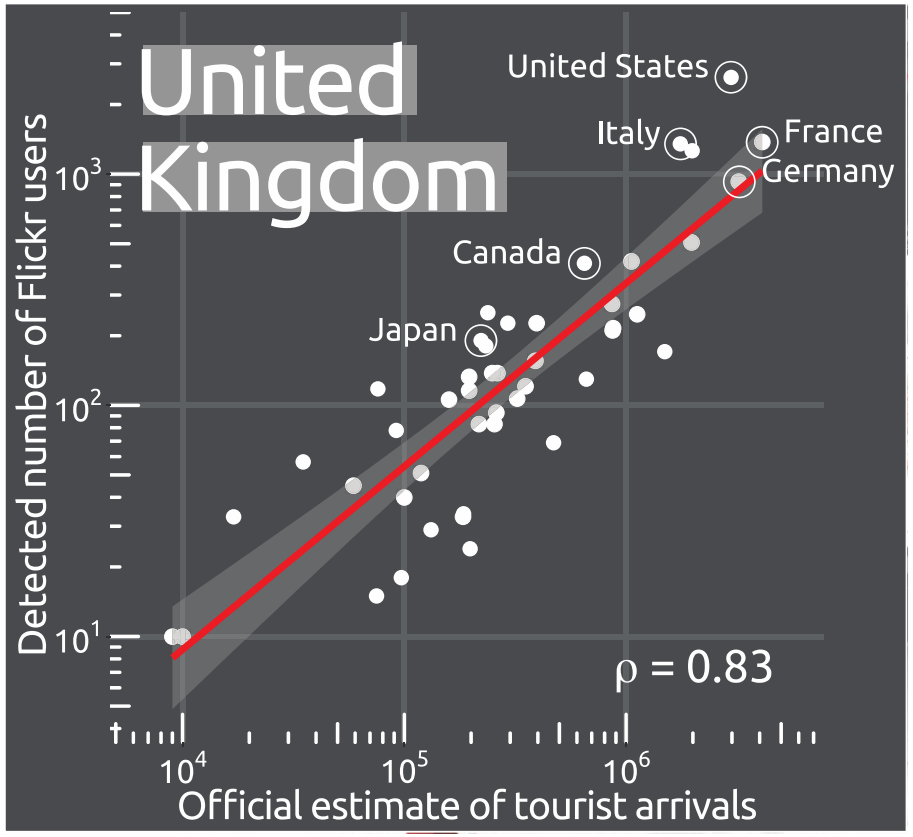


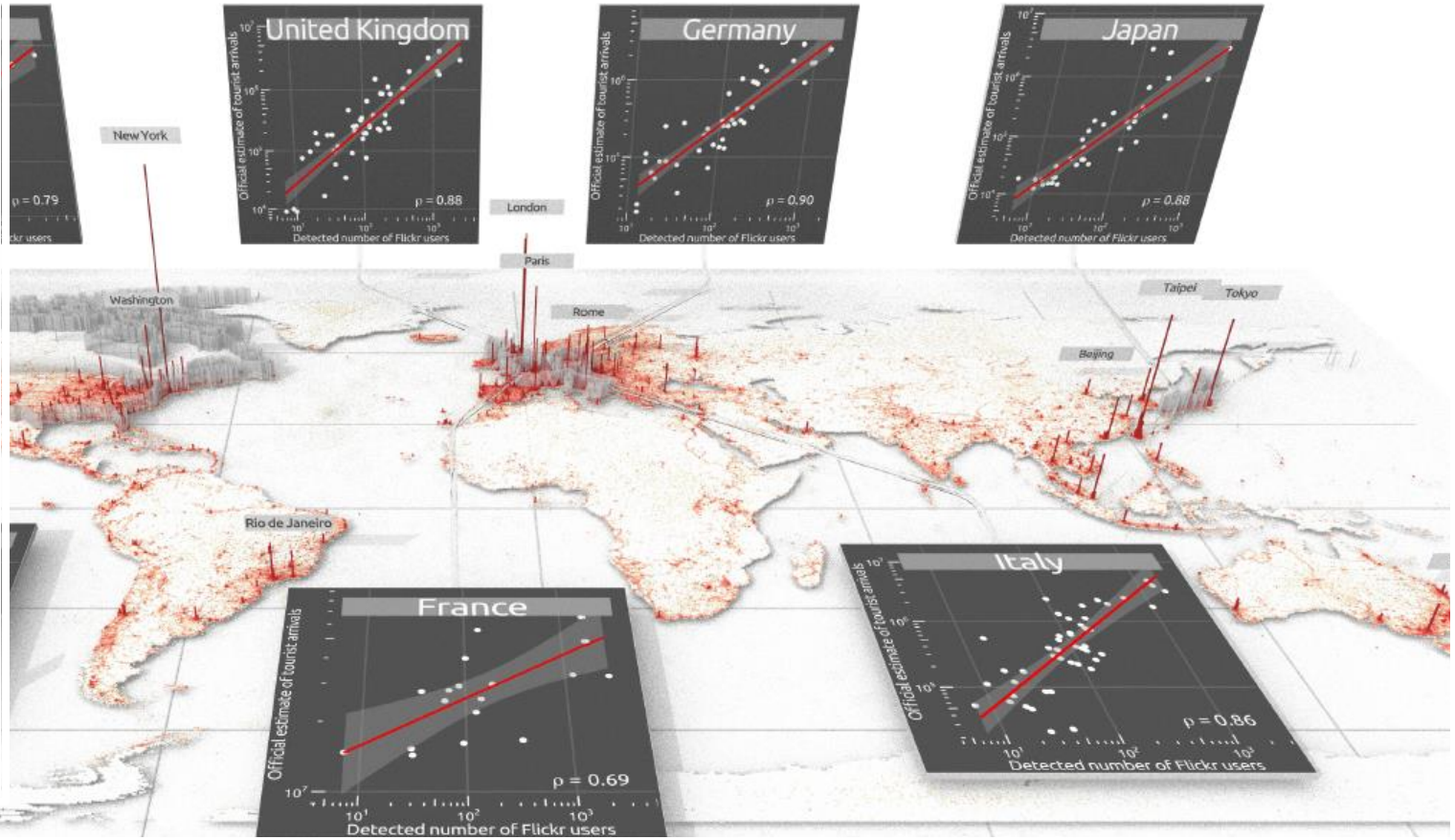
Sensing global tourism numbers with millions of publicly shared online photographs

Environment and Planning A
XX(X):1–4

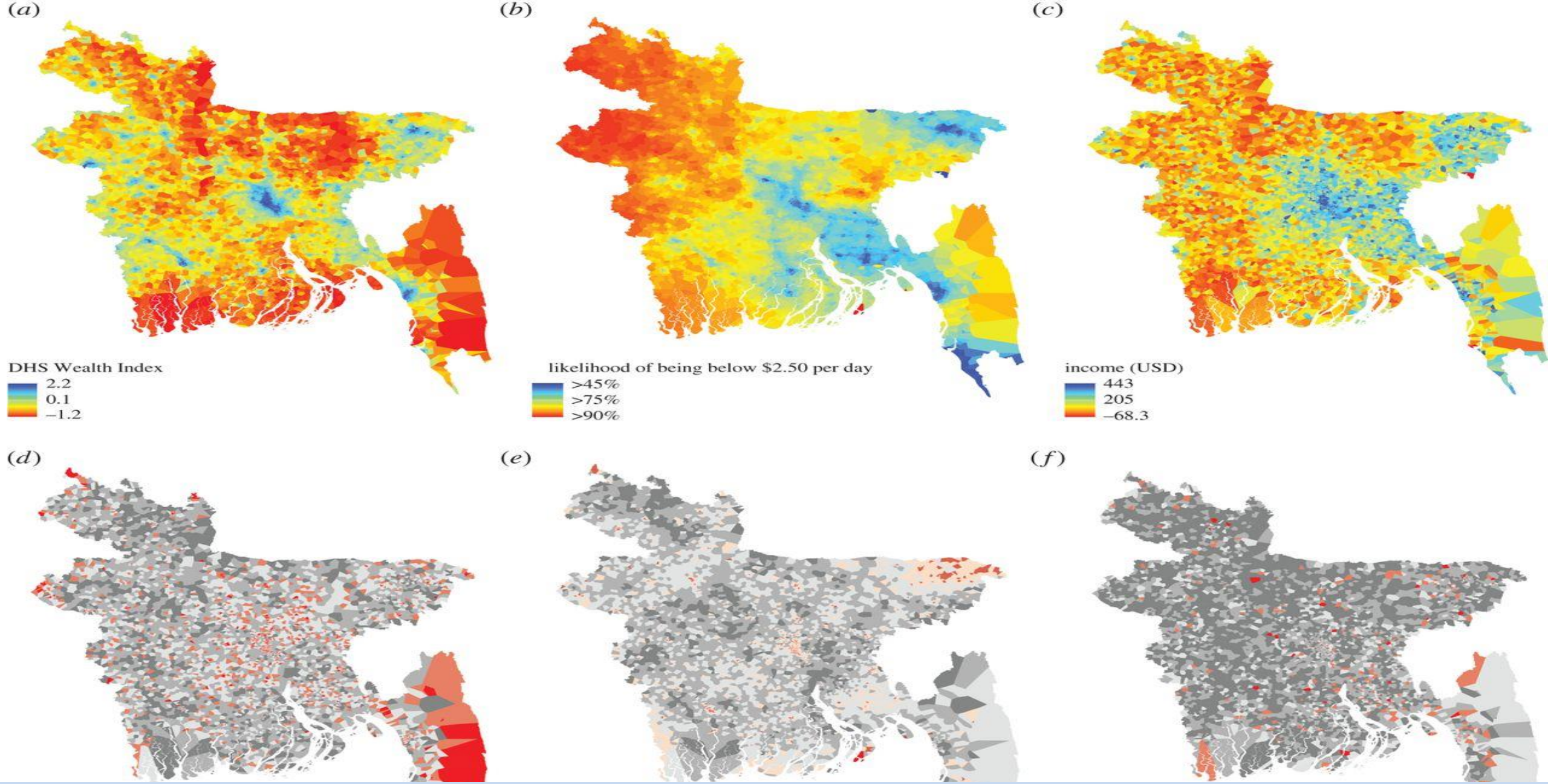
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Tobias Preis[1,2], Federico Botta[1], Lanthao Benedikt[3] and Helen Susannah Moat[1,2]



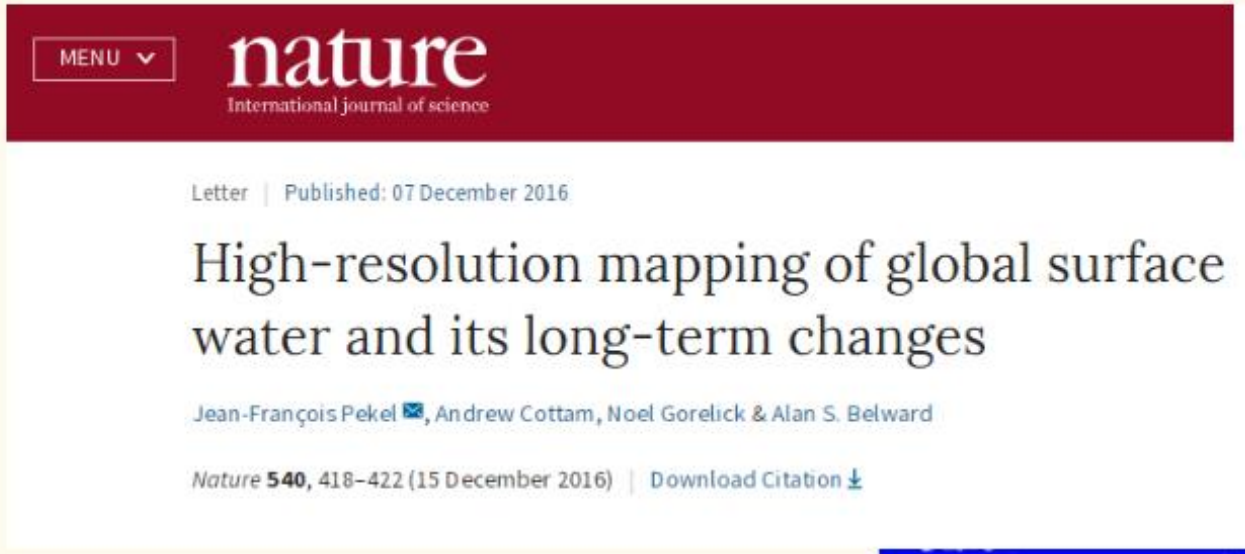


Preis et al (2018), Sensing global tourism numbers with millions of publicly shared online photographs



Mapping poverty using mobile phone and satellite data, Steele et al (Feb 2017), *Journal of the Royal Society Interface,* **14**:127. http://rsif.royalsocietypublishing.org/content/14/127/20160690

Sustainable Development Goal 6.6.1: Change in the extend of water-related ecosystems over time



In a recent study produced for the Office for National Statistics (ONS) Natural Capital Accounts, the UK's trees were estimated to remove 1.4 million tonnes of air pollutants in a single year. This would result in an annual saving of £1 billion in avoided health damage costs. In another study, London's 8.42 million trees have been estimated to remove 2,241 tonnes of pollution per year, which in addition to other services, is estimated to provide £132.7 million in annual benefits.

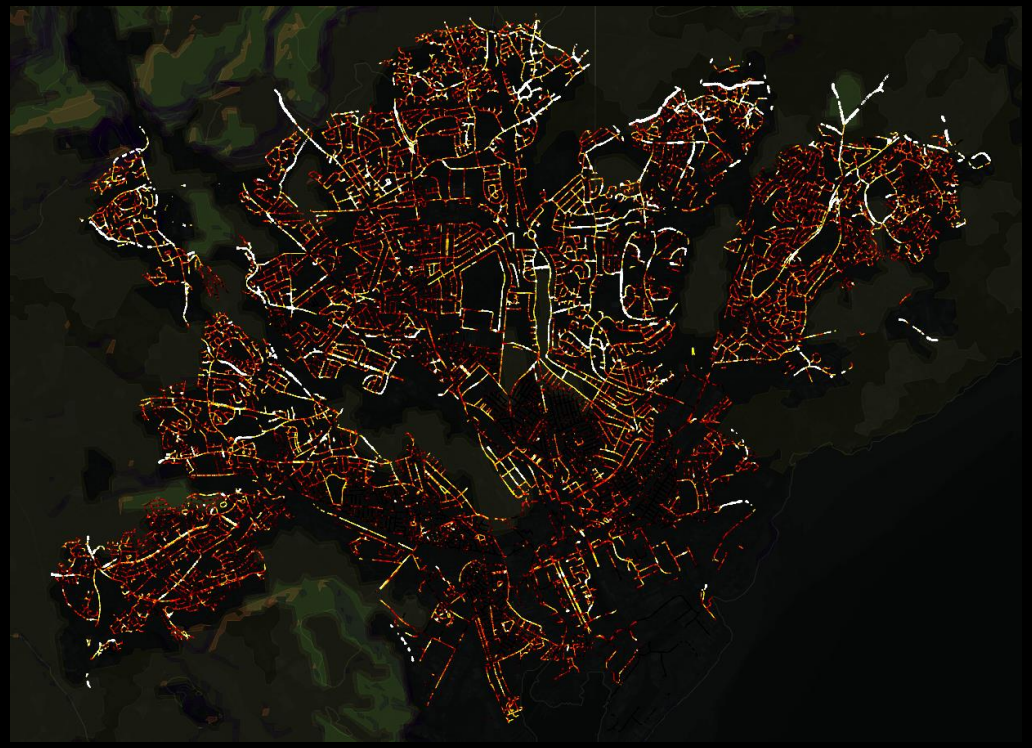
For Cardiff, the annual benefit is close to £8 million.

Aim: Generate a scalable, consistent, automated, **urban vegetation index**

Outcome: An end-to-end processing pipeline.

Making use of: **17 million images** from **Google Street View** for 112 cities in the UK.

… **OpenStreetMap** road network data
… Deep **image segmentation** methods



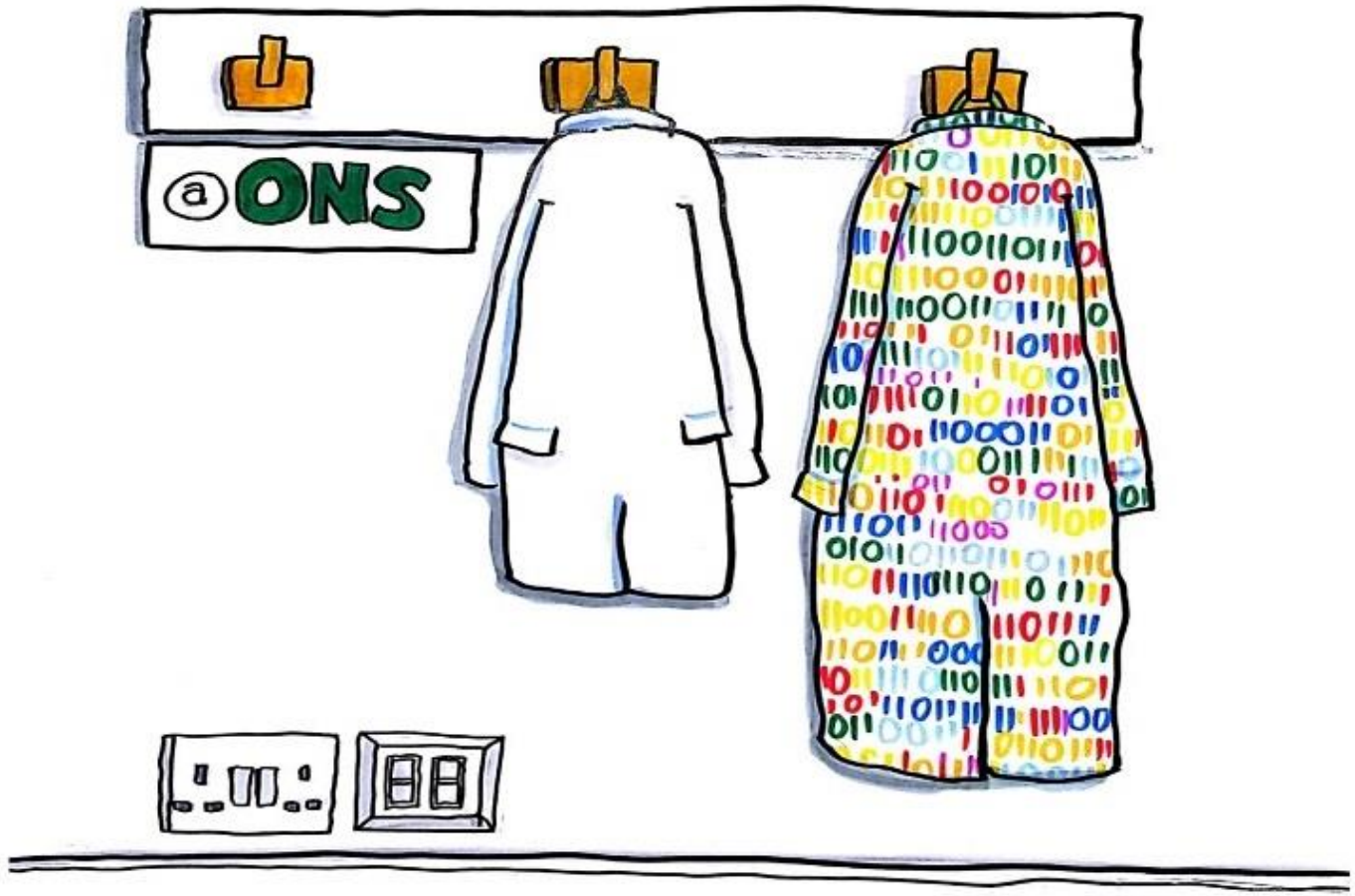
All data is biased

Surveys are not ground truth

Triangulation is key

Data science for public good

Tom Smith
ONS Data Science Campus
@_datasmith



Data Science Campus

web: datasciencecampus.ons.gov.uk
email: datasciencecampus@ons.gov.uk
twitter: @DataSciCampus