Office for National Statistics
Swyddfa Ystadegau Gwladol

Item VIII: New Data Sources and Recent Innovations in Producing National and International Accounts

Data Science, Big Data and Economic Statistics

Grateful for inputs from Tom Smith, Data Science Campus

Economic Commission for Europe
Conference of European Statisticians
Group of Experts on National Accounts

Salle XI, Palais des Nations, Geneva, Switzerland
25th May 2018

Sanjiv Mahajan
Head of International Strategy and Coordination
Office for National Statistics (UK)
Sanjiv.mahajan@ons.gov.uk

Data Science, Big Data and Economic Statistics

Overview

- Creation of the UK Data Science Campus
  - Purpose and mission
  - What we do – delivery and capability
- Specific projects
  - Project 1. Identifying business growth characteristics using website (text) data
  - Project 2. Payments data for regional indicators
  - Project 3. Superfast indicators of GDP growth
  - Project 4. Mapping the urban forest
  - Project 5. Internet traffic indicators
- Any questions?

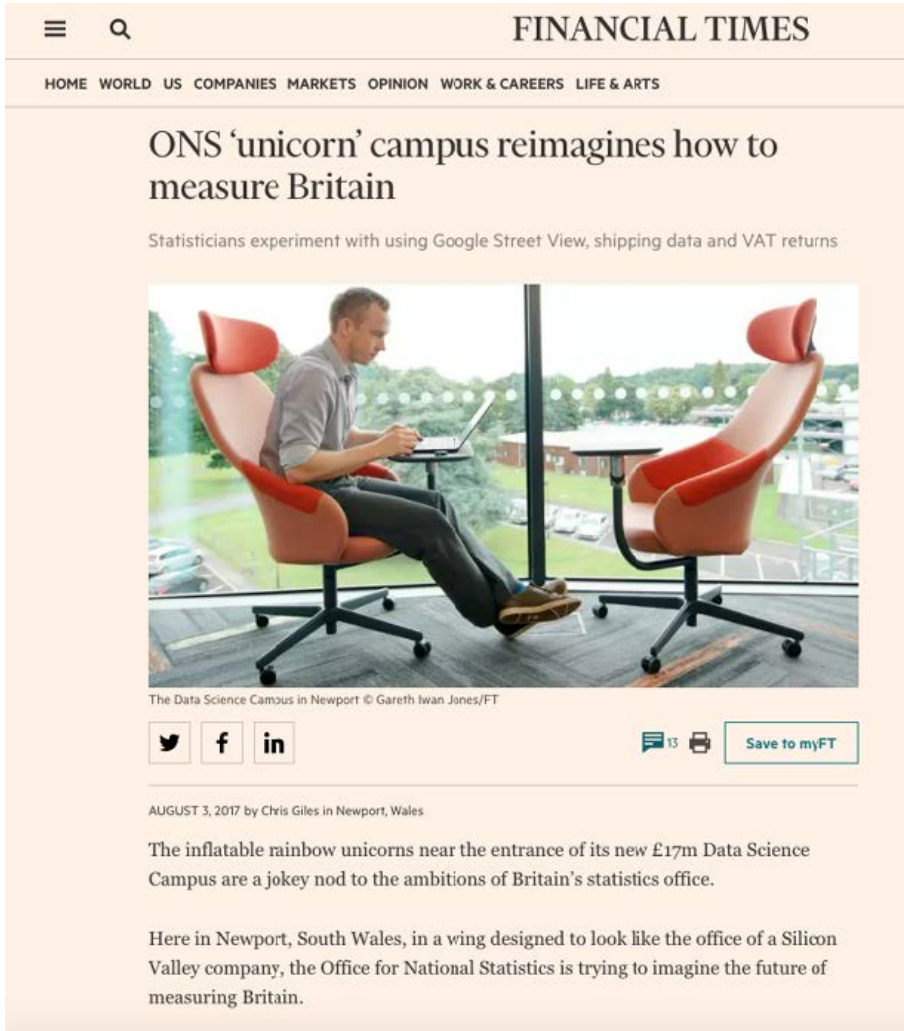
Creation of the UK Data Science Campus

"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

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.

UK Data Science Campus

Purpose

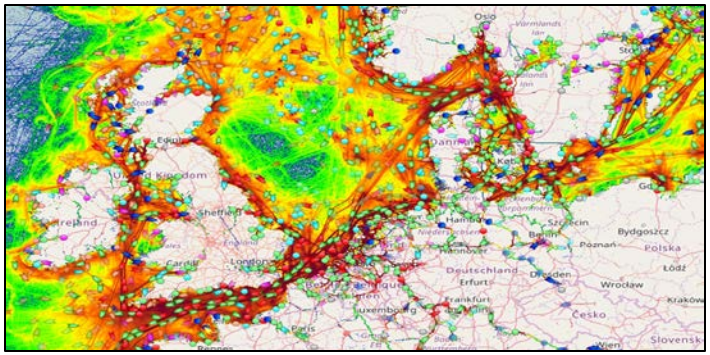
We apply data science, and build skills, for public good across the UK and internationally.

Mission

We work at the frontier of data science and Artificial Intelligence - building skills and applying tools, methods and practices - to create new understanding and improve decision-making for public good.

What we do – delivery and capability

Data science projects



New data sources, e.g. satellite images, text, big data, Internet of Things, social media.

New techniques – machine learning, neural networks, network, text & image analysis, big data processing etc.

Short, exploratory research – innovation and risk.

Building capability

Cross-government training & train-the-trainers

Apprenticeships in Data Analytics.

MSc Data Analytics for Government.

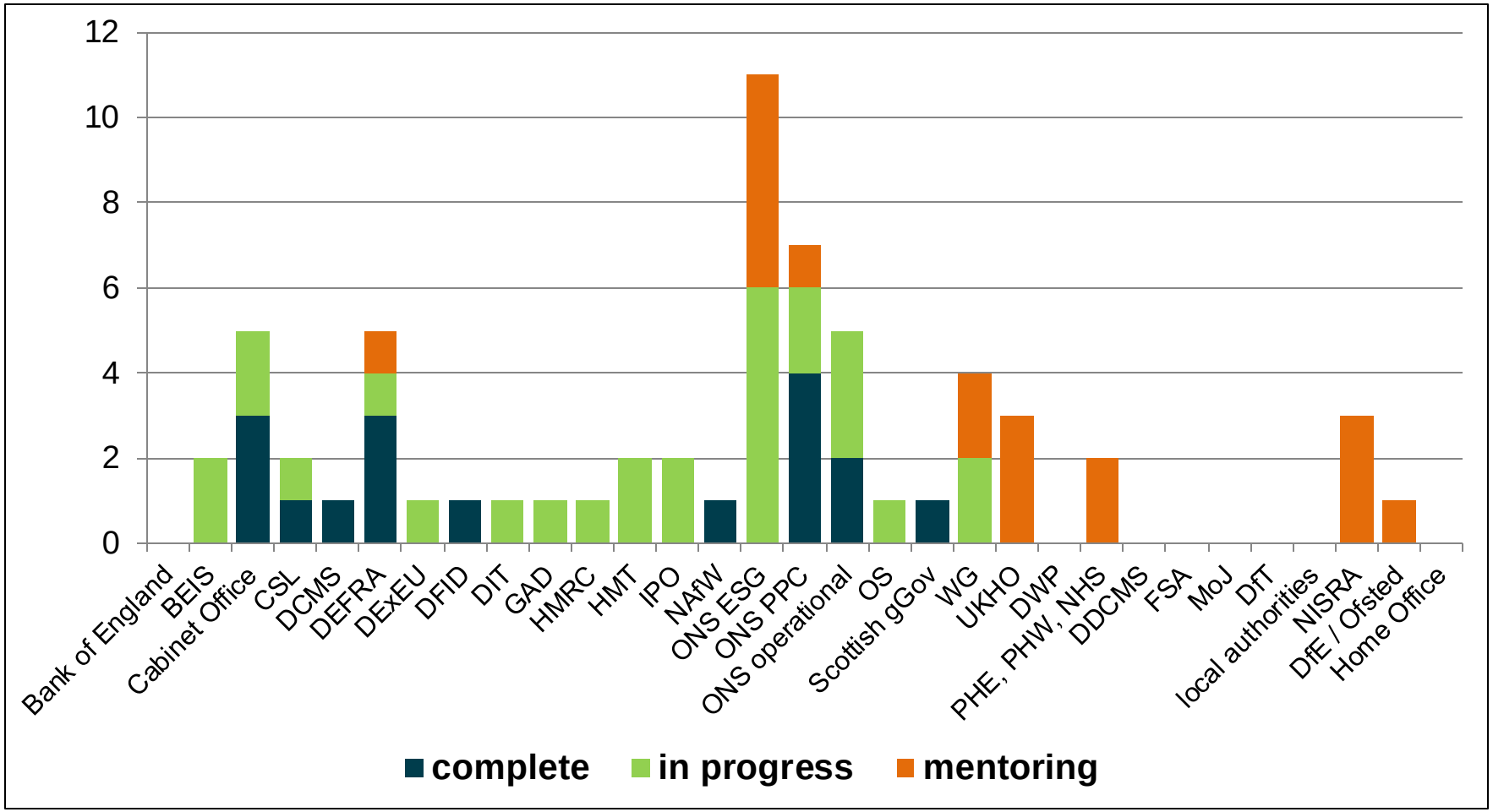
Continuous Professional Development.

Data Science Accelerator and ONS Data Science Academy mentoring.

STEM Ambassadors.

Co-funded & co-supervised PhD placements and programmes.

Projects and mentoring with 21 government departments



Project 1. Identifying business growth characteristics using website (text) data

GlassAI

.co.uk websites

30k

Job Vacancies - Vacancies advertised on the organisation's website.

News Articles - News published on the organisation's website.

Bios - Mention's of the organisation in people's bios.

Mentions - Number of mentions of the organisation on other websites. Number of mentions of other organisations on the organisation's website.

Office for National Statistics

Inter Departmental Business Register

20k

Matched on name and address

Department for Business, Energy & Industrial Strategy

High Growth Flag

7k

Active in 2013, high growth in 2016

Initial results suggest non-traditional data sources can provide features for modelling high growth companies

Active Website (7k)

- High growth companies are more likely to have an active website, 9% of matched dataset are high growth compared to 3% of the IDBR.

Network (6K)

High growth companies are more likely to:

- mention 14 or more other organisations on their website.
- be mentioned by 4 or more other organisations on their websites.
- be mentioned in bios on other organisations' websites.
- mention 3 or more other organisations in bios on their website.

Bios (2.5k)

- High growth companies are more likely to have 9 or more bios on their website.

Health Warning!

These findings are from initial investigation of the GlassAI data. The dataset has not been controlled for factors such as size of the organization. Further analysis is needed to validate these findings.

Next steps

Topics (using NLP)

- Derive topics from news articles or titles.
- Derive topics from job descriptions or job titles.
- Derive topics from bios.

Sectors

- ONS Big Data Team is running a project identifying how representative a website is the “official" activity of that company.
- Initial topic analysis of the website descriptions derived topics relating to sectors.
- Further analysis of how these topics relate to NACE Rev. 2 would be interesting.

Machine Learning

- Apply machine learning algorithms to find the most indicative features.

Social Media

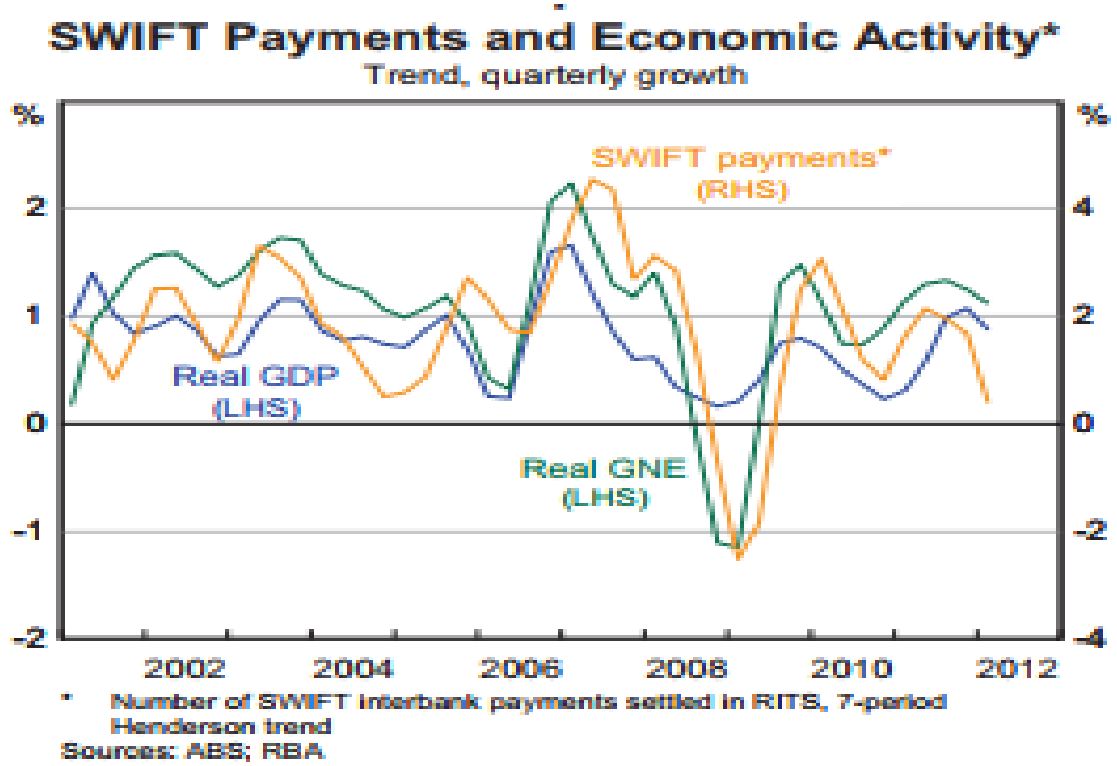
- Test features from Brandwatch.

Other Sources

- Consider other non-traditional data sources.

Project 2. Payments data for regional indicators

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



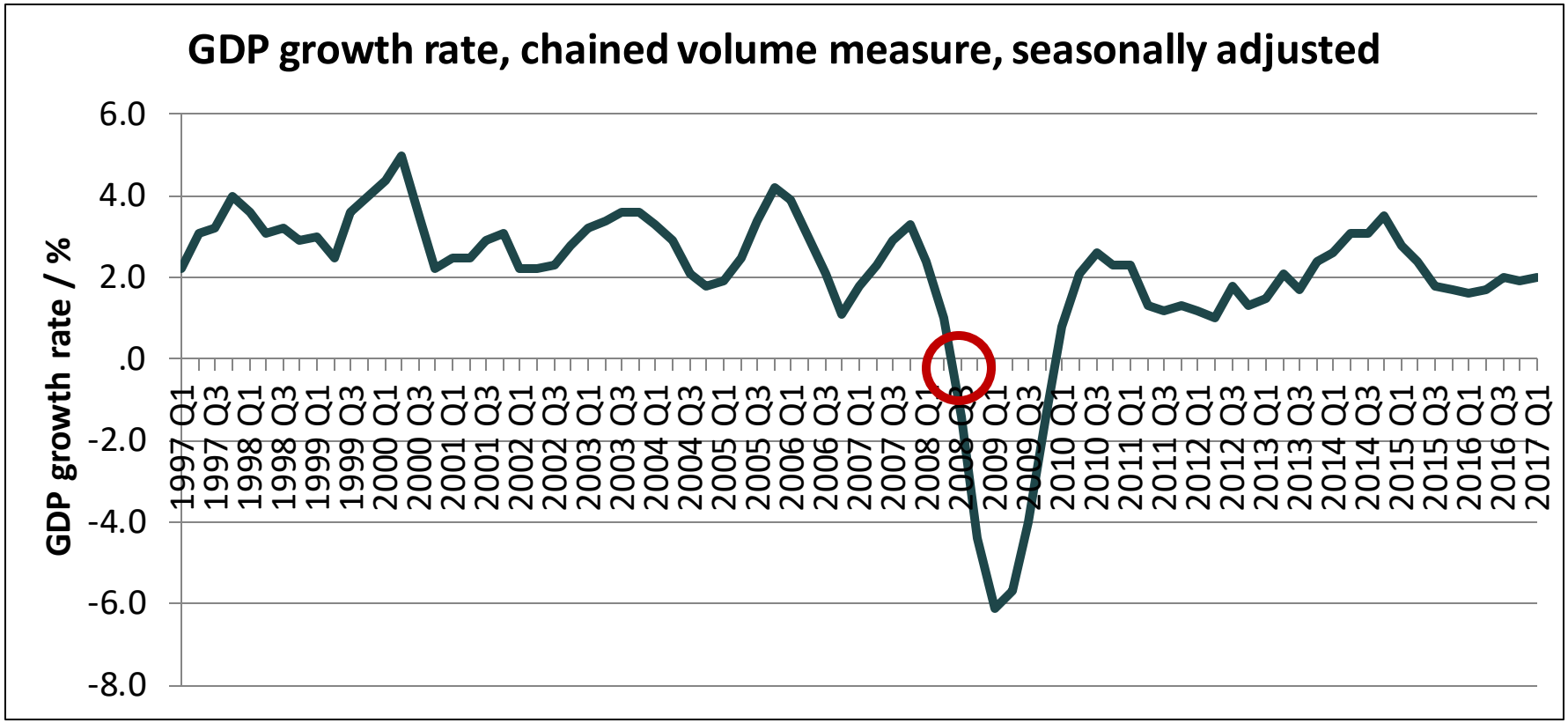
Project 2. Payments data for regional indicators

- Financial data held by banks:
  - **No sensitive or personally identifiable data shared**
  - **All outputs are aggregate and non-sensitive**
- Hypotheses being explored include:
  - Payments data can be used as a proxy for retail sales.
  - Payments data can be used as a proxy for private household consumption.
  - Payments data can improve the accuracy of GDP nowcasting.
- Possible data sources:

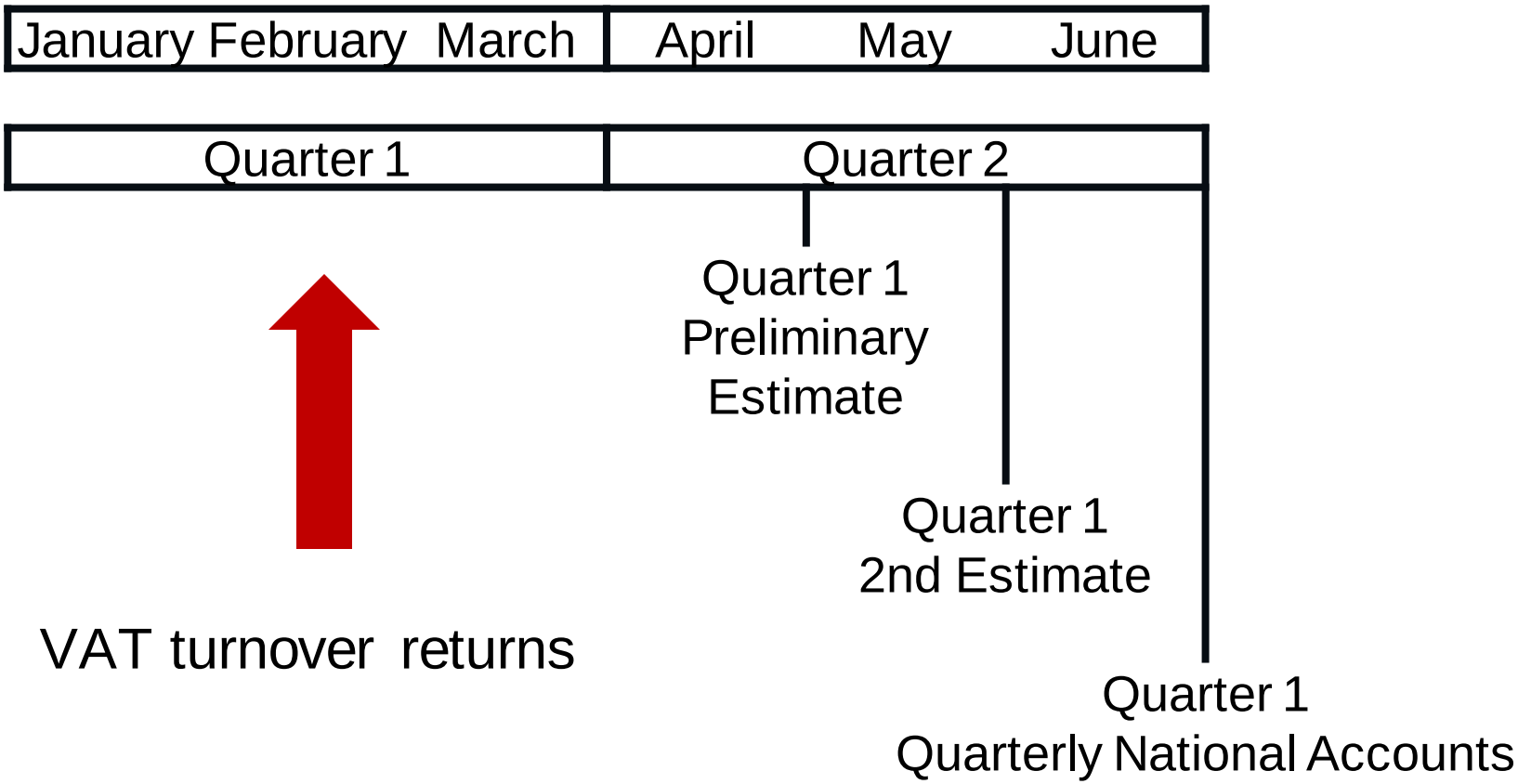
Consumer	Electronic payments	Business
Debit Card spend Credit Card spend Personal Loans Mortgages Savings accounts Insurance	POS data ATM data On-line gateway data (online purchases) Peer-to-peer	Merchant and acquirer data Corporate cards Business bank products Corporate bank products Investment bank products

Project 3. Superfast indicators of GDP growth

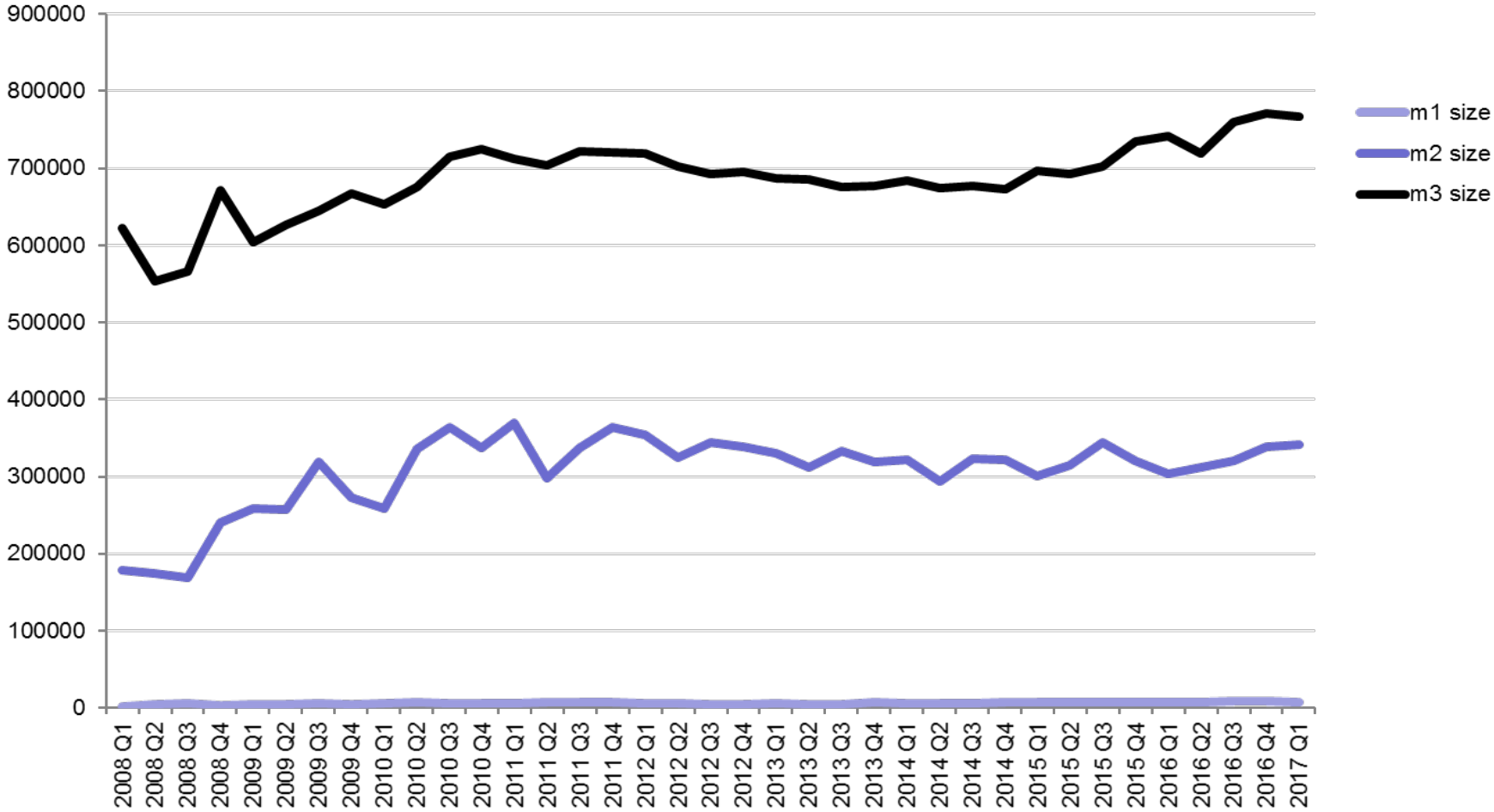
How early can we identify negative GDP growth?



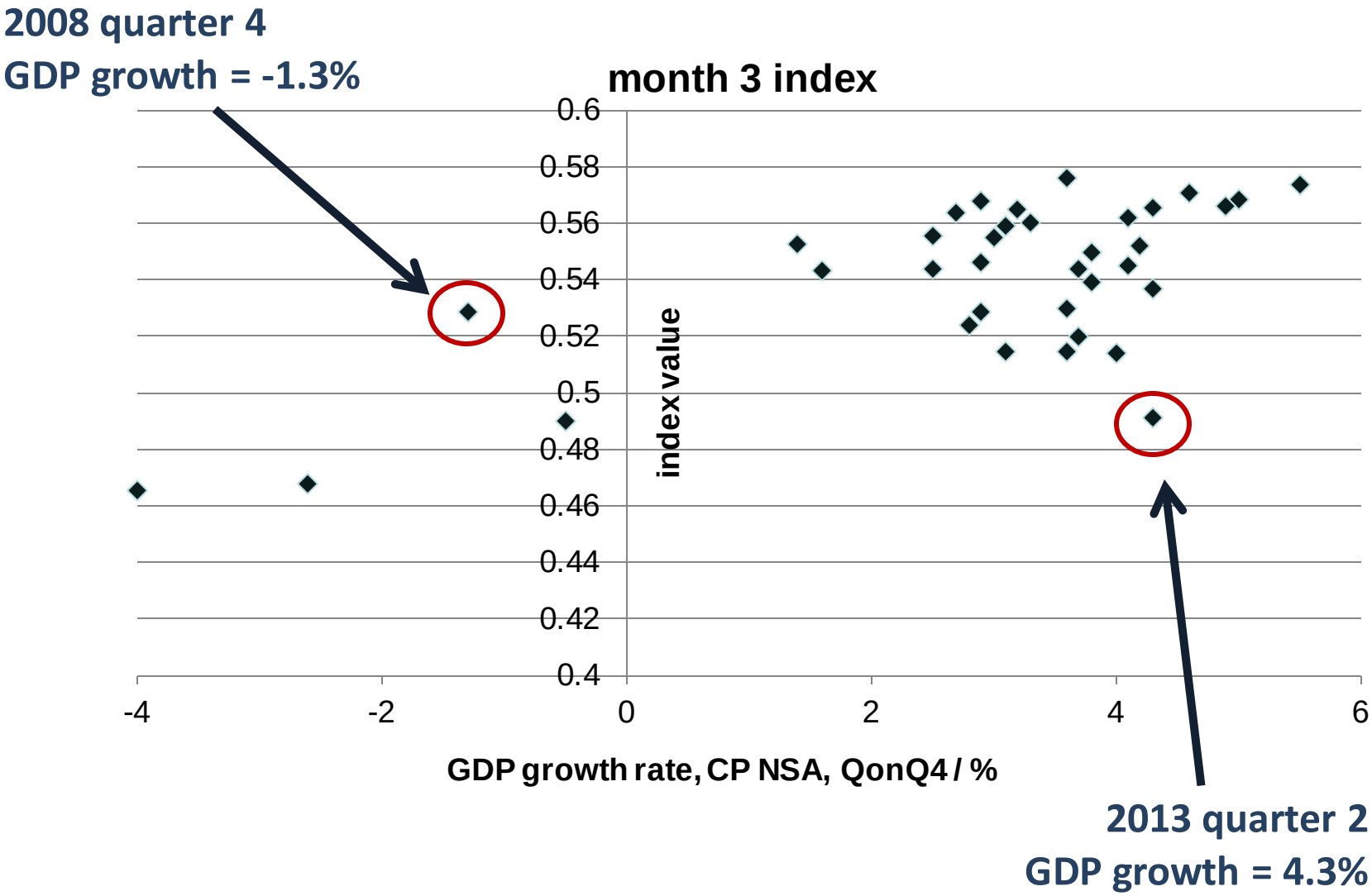
Superfast indicators of GDP growth



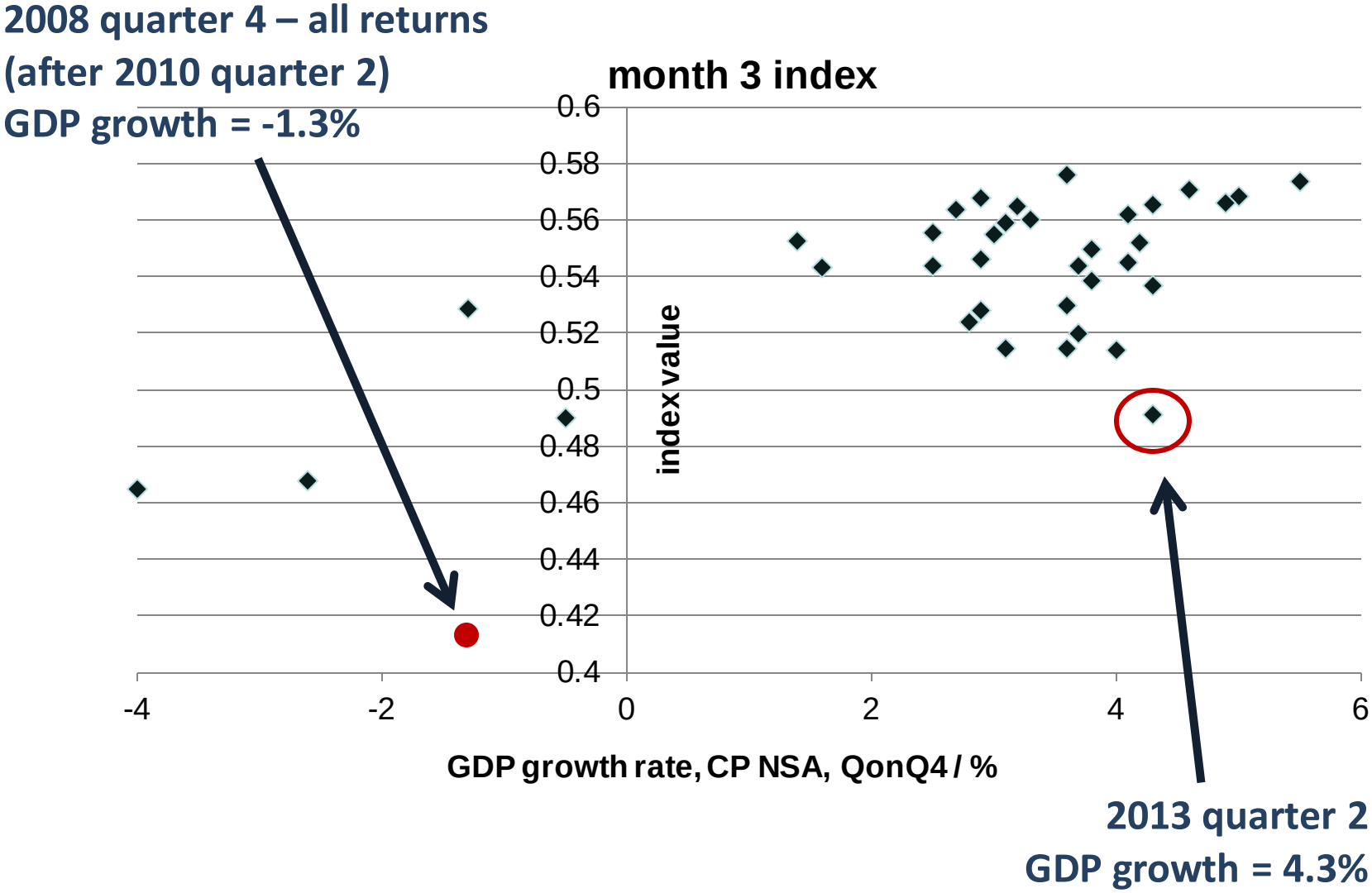
Sample size


900000
800000
700000
600000
500000
400000
300000
200000
100000
0
2008 Q1
2008 Q2
2008 Q3
2008 Q4
2009 Q1
2009 Q2
2009 Q3
2009 Q4
2010 Q1
2010 Q2
2010 Q3
2010 Q4
2011 Q1
2011 Q2
2011 Q3
2011 Q4
2012 Q1
2012 Q2
2012 Q3
2012 Q4
2013 Q1
2013 Q2
2013 Q3
2013 Q4
2014 Q1
2014 Q2
2014 Q3
2014 Q4
2015 Q1
2015 Q2
2015 Q3
2015 Q4
2016 Q1
2016 Q2
2016 Q3
2016 Q4
2017 Q1
m1 size
m2 size
m3 size

Superfast GDP indicator - results



Superfast GDP indicator - results

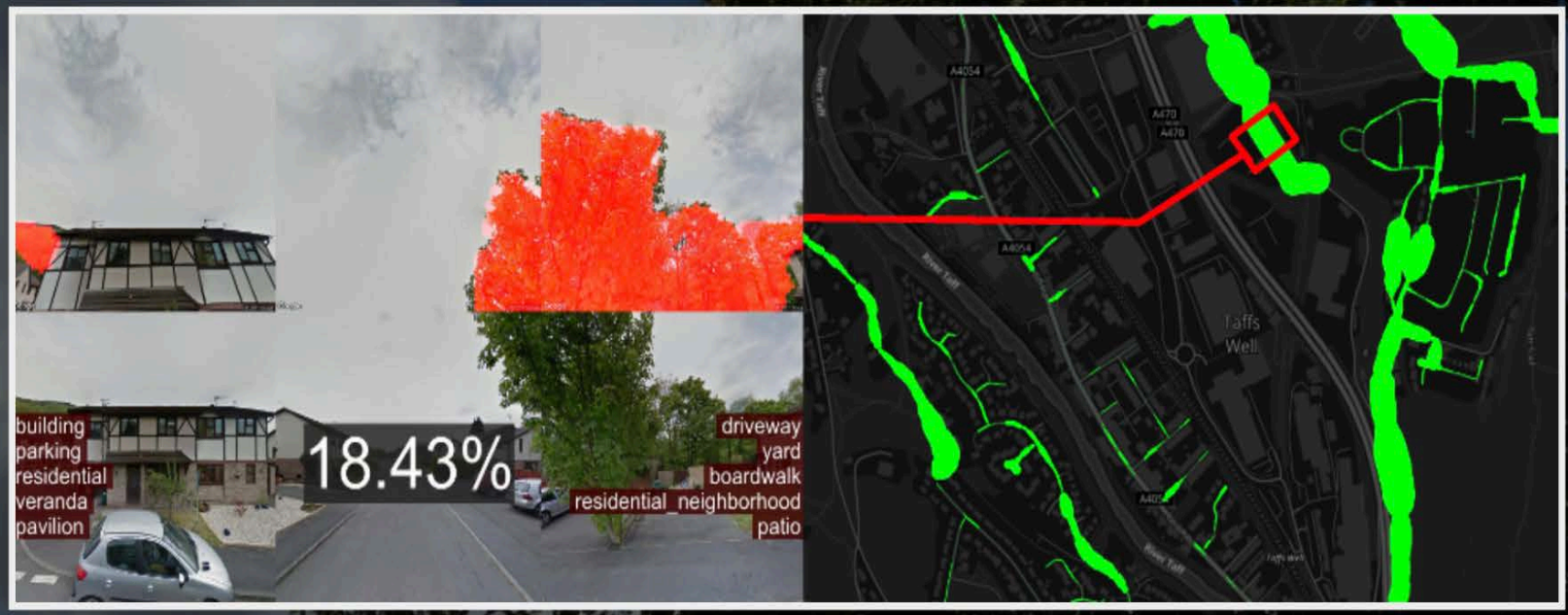


Superfast GDP summary

- Goal achieved: simple, useful.
- However, did not identify the 2008 downturn before other methods – company early estimates of VAT are optimistic.
- Publish as classification not index value - date to be determined.
- Other analysis:
  - Outlier detection – not significant.
  - Deflation – not significant.
  - Seasonal adjustment – high complexity.
  - Births and deaths – under investigation.

Project 4. Mapping the urban forest

Our approach...



Makes use of:

1. Google streetview imagery
2. OpenStreetMap road network data

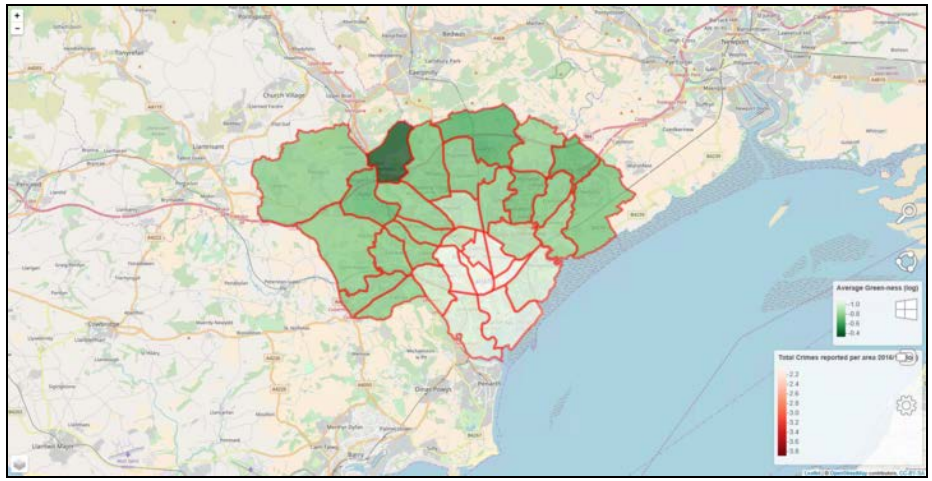
Mapping the urban forest – indicators from images

- Analysing images to improve data on local environment.
- £1bn value trees in urban areas (air pollution, health, wellbeing).
- Poor data at local level on tree & urban greenery.

Liverpool, Edinburgh

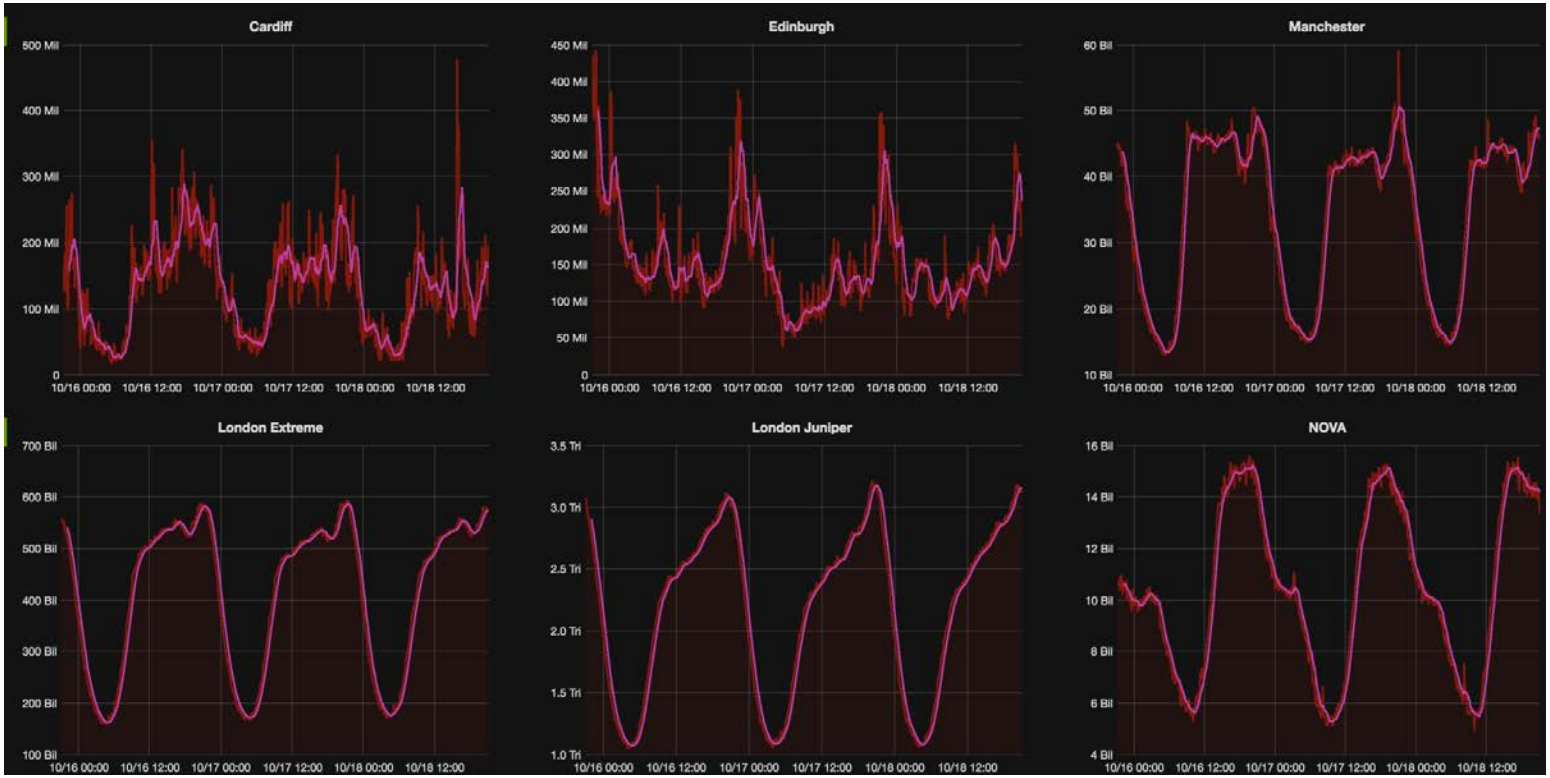
National Tree Map, Blue Sky

Cardiff



Project 5. Internet traffic indicators

- UK most internet-dependent economy in the world (BCG, 2012), 8.3% GDP, more than twice G20 average.
- Increasing rapidly (e.g. predicted to double between 2010 and 2016).
- Internet traffic data from LINX.

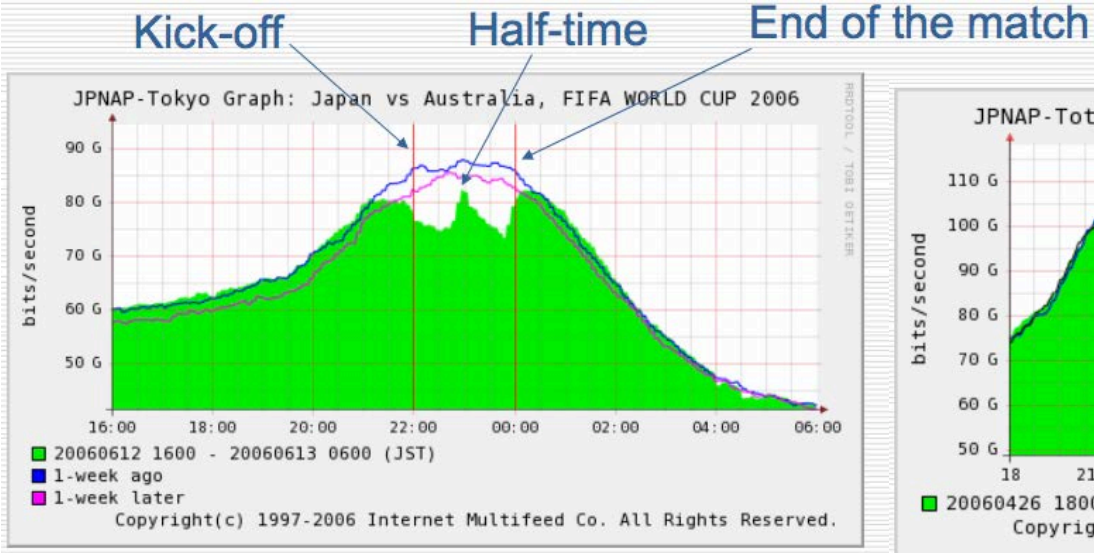


Internet traffic indicators

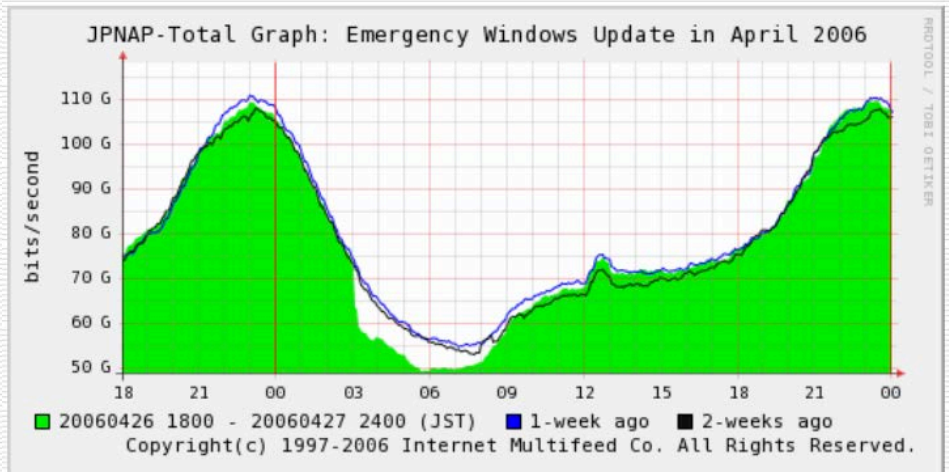
As with road traffic, can we:

- observe patterns in intra-day economic activity?
- longer term economic growth in internet activity?

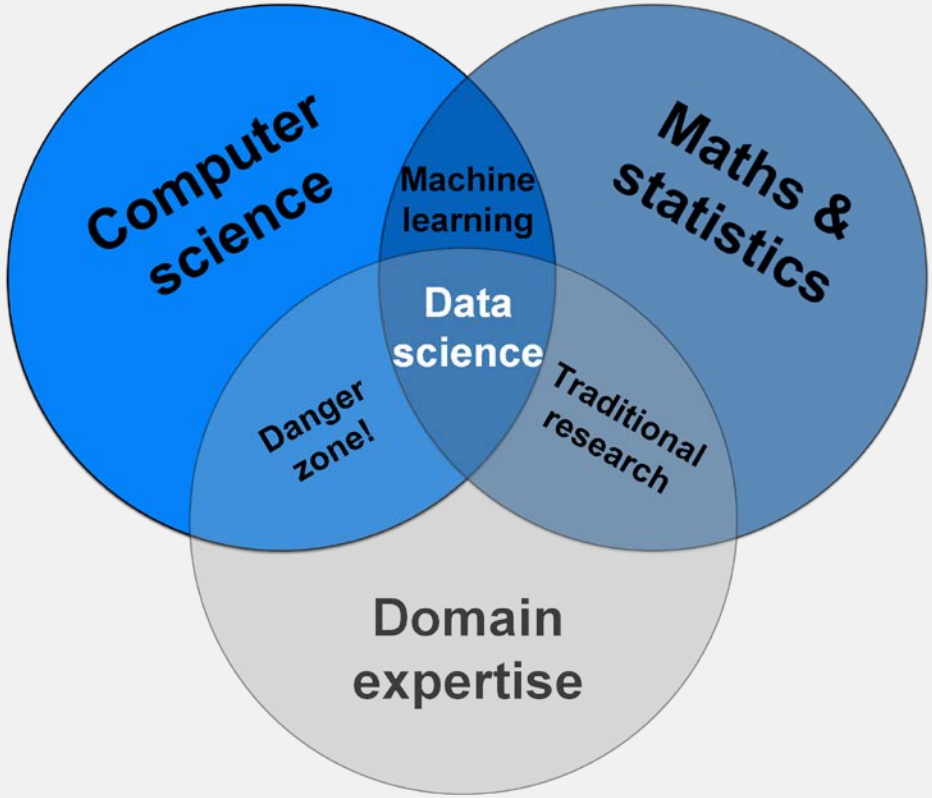
Japan v Australia World Cup football



Windows update



BUILDING CAPABILITY – THE DATA SCIENCE CAMPUS



Office for National Statistics

Data Science Campus

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