Office for National Statistics

Estimating population mobility using big data sources – the benefits and the challenges

Jane Naylor
Office for National Statistics

Overview

- Background
- 3 ONS Case studies:
  - Twitter
  - Mobile phone data
  - Google trends
- Benefits and Challenges
- How to manage the challenges

Data sources for official statistics

- Surveys – eg of businesses and households
- Census – every 10 years
- Administrative data – by-product of Government process
- Big Data?

  'Data that is difficult to collect, store or process within the conventional systems of statistical organizations. Either, their volume, velocity, structure or variety requires the adoption of new statistical software processing techniques and/or IT infrastructure to enable cost-effective insights to be made.'

  (UNECE, 2013)

ONS Big Data team

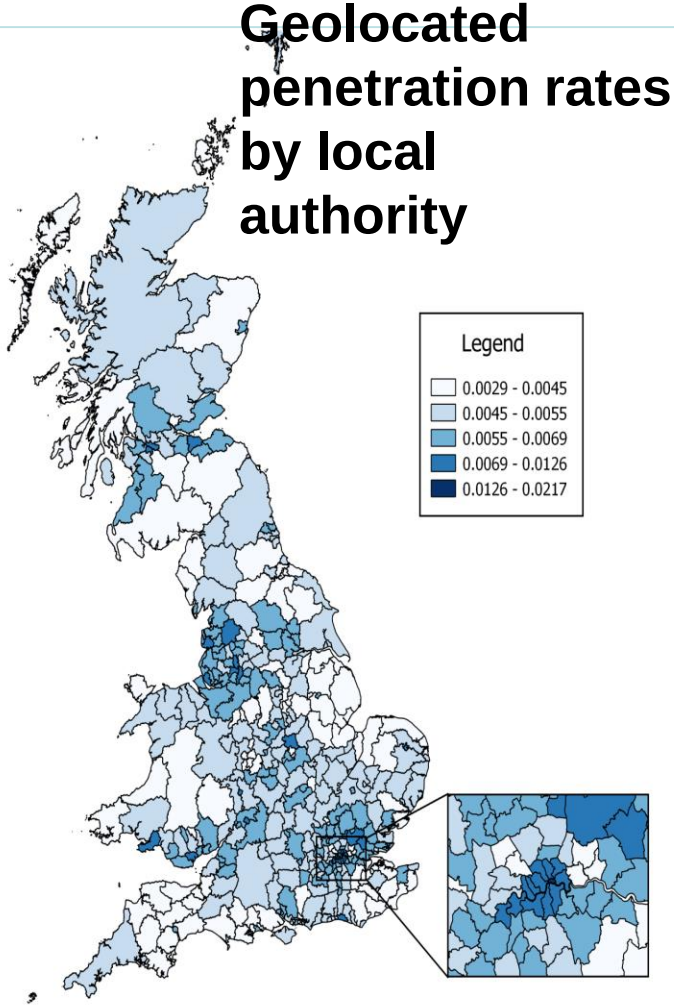
- Launched in January 2014
- Our goals:
  - developing an understanding within ONS
  - demonstrating the potential for using big data within official statistics
  - investigating the methodological and technological issues
- Approach: Collaborative working/partnerships and practical pilots

Twitter

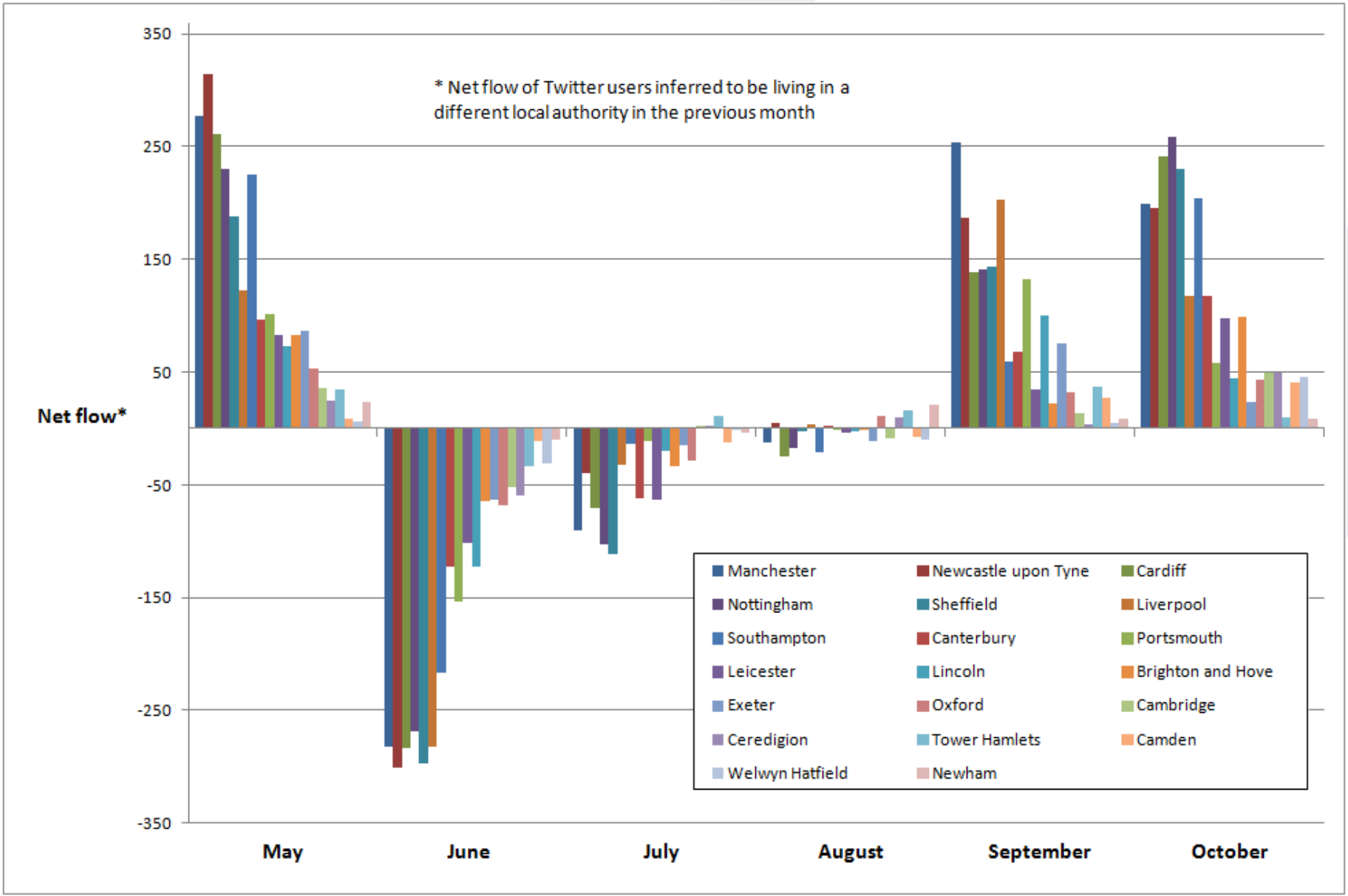
Rationale: Using geo-located Twitter to gain new insights mobility and migration

- 7 months of geo-located tweets within Great Britain (about 100 million data points)
- Methodology to infer place of usual residence:
  - Identify user ‘anchor points’ by clustering tweets using a DBSCAN algorithm
  - Identify residential anchor points using AddressBase and nearest neighbour analysis

Geolocated penetration rates by local authority

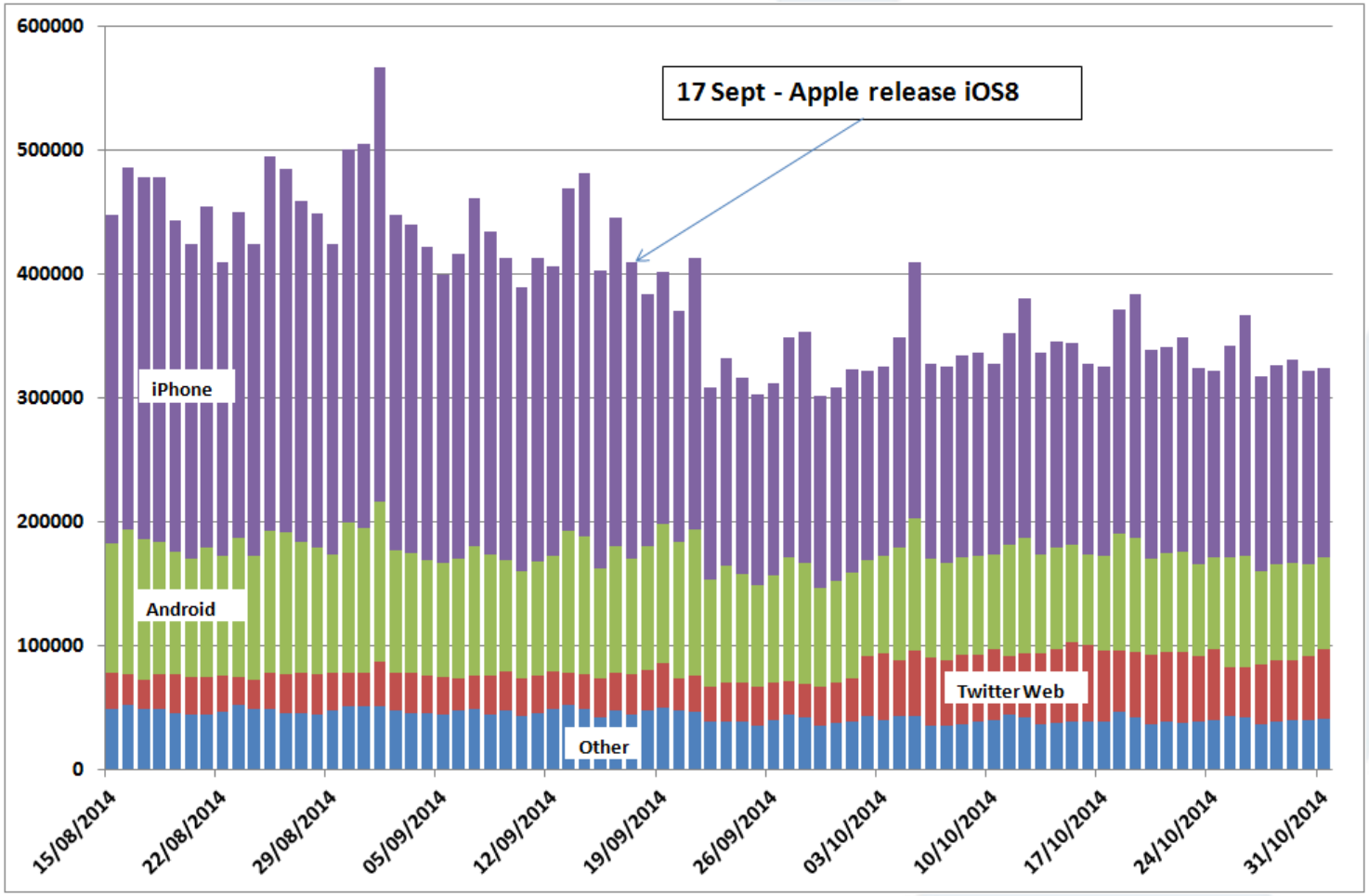


Use case: Student mobility

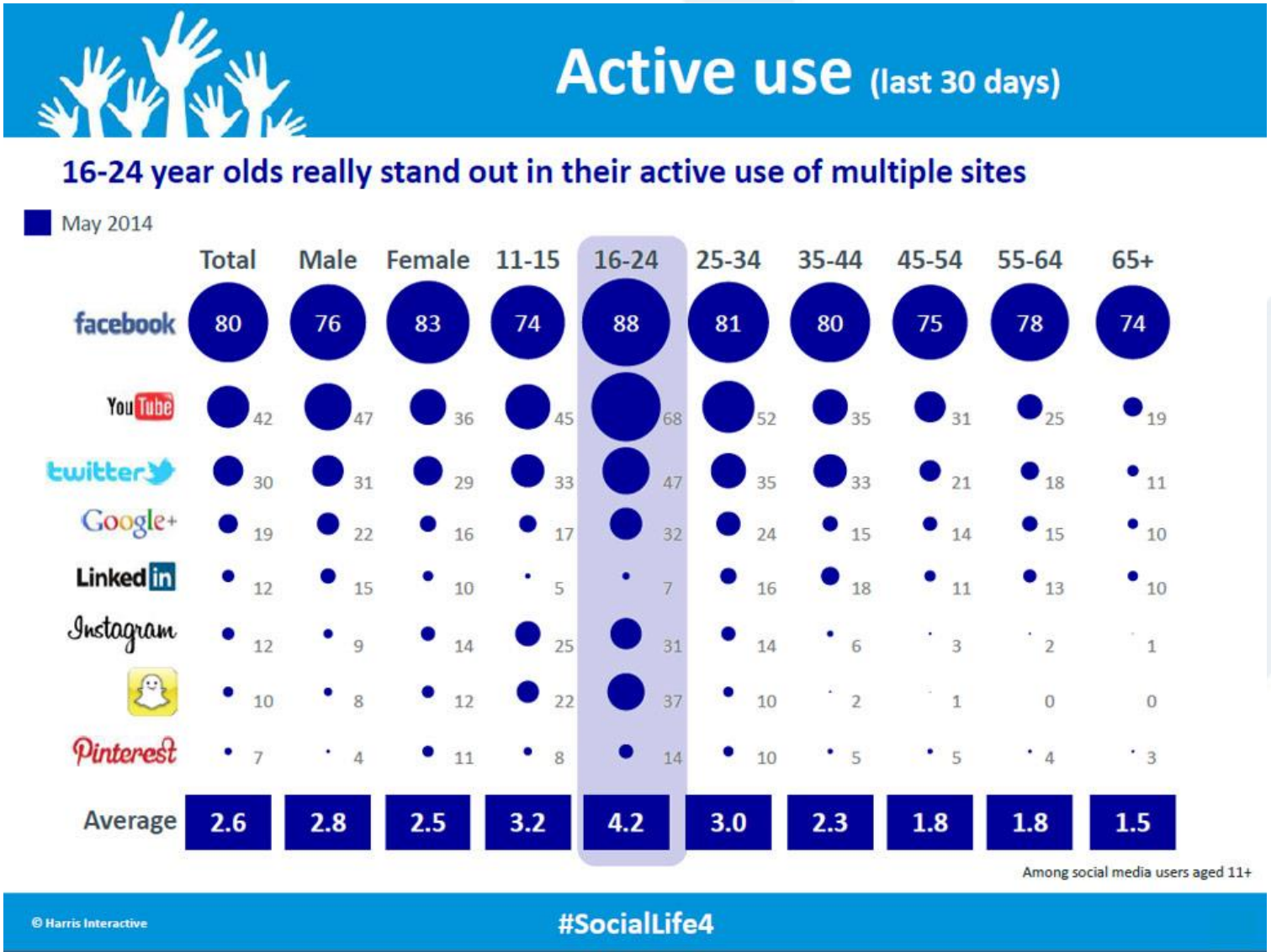


Data quality

Geo-located Twitter volumes by Device Type



Sample bias

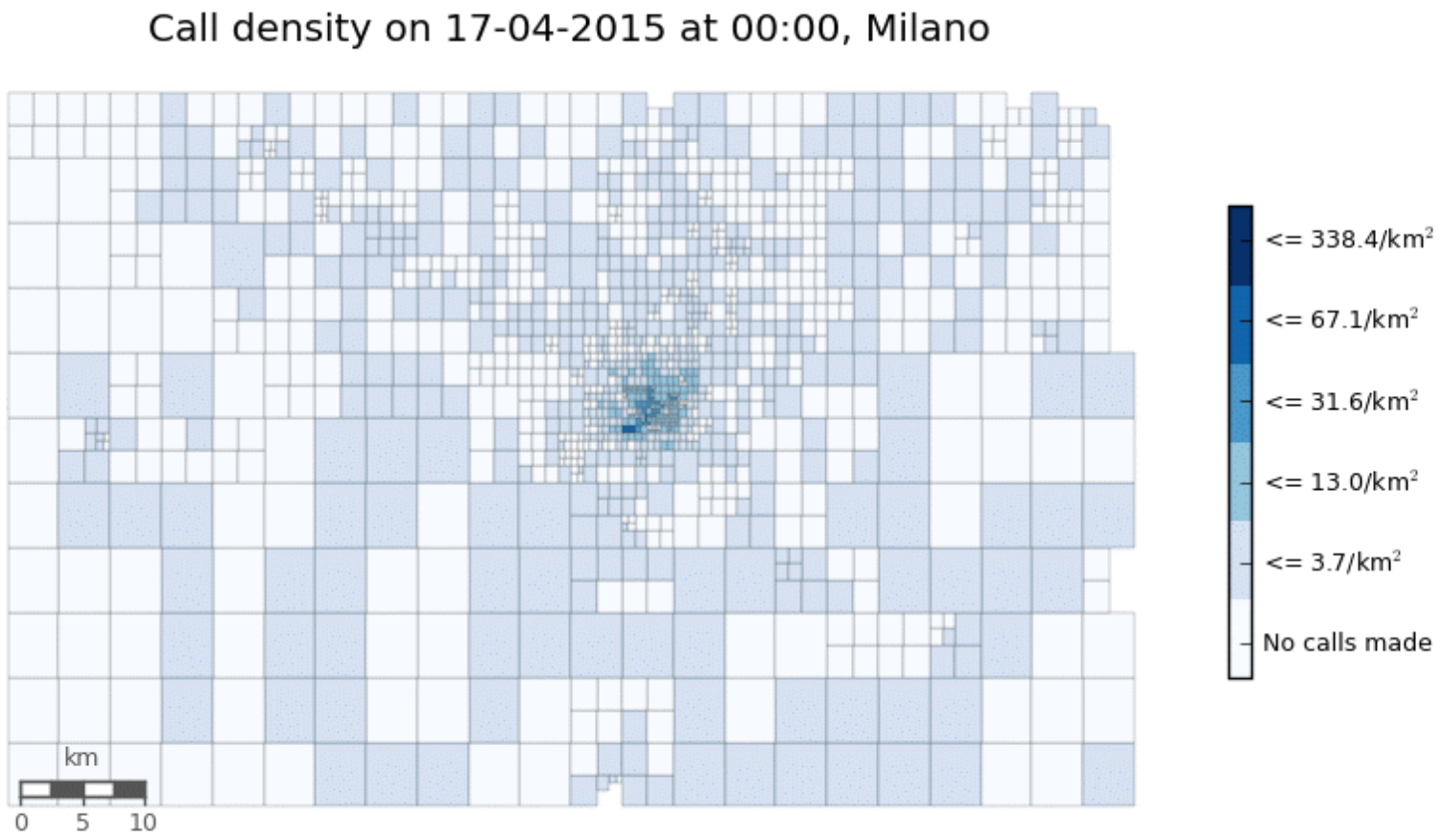

Active use (last 30 days)
16-24 year olds really stand out in their active use of multiple sites
May 2014
Total Male Female 11-15 16-24 25-34 35-44 45-54 55-64 65+
facebook 80 76 83 74 88 81 80 75 78 74
YouTube 42 47 36 45 68 52 35 31 25 19
twitter 30 31 29 33 47 35 33 21 18 11
Google+ 19 22 16 17 32 24 15 14 15 10
LinkedIn 12 15 10 5 7 16 18 11 13 10
Instagram 12 9 14 25 31 14 6 3 2 1
10 8 12 22 37 10 2 1 0 0
Pinterest 7 4 11 8 14 10 5 5 4 3
Average 2.6 2.8 2.5 3.2 4.2 3.0 2.3 1.8 1.8 1.5
Among social media users aged 11+
© Harris Interactive
#SocialLife4

Mobile Phones

Rational: Modelling population density and population flows, e.g. commuting statistics

- Conducted research into the potential uses for mobile phone data within official statistics
- Intelligence gathered through engagement with UK MNOs and public transport bodies

Mobile phone data – population density

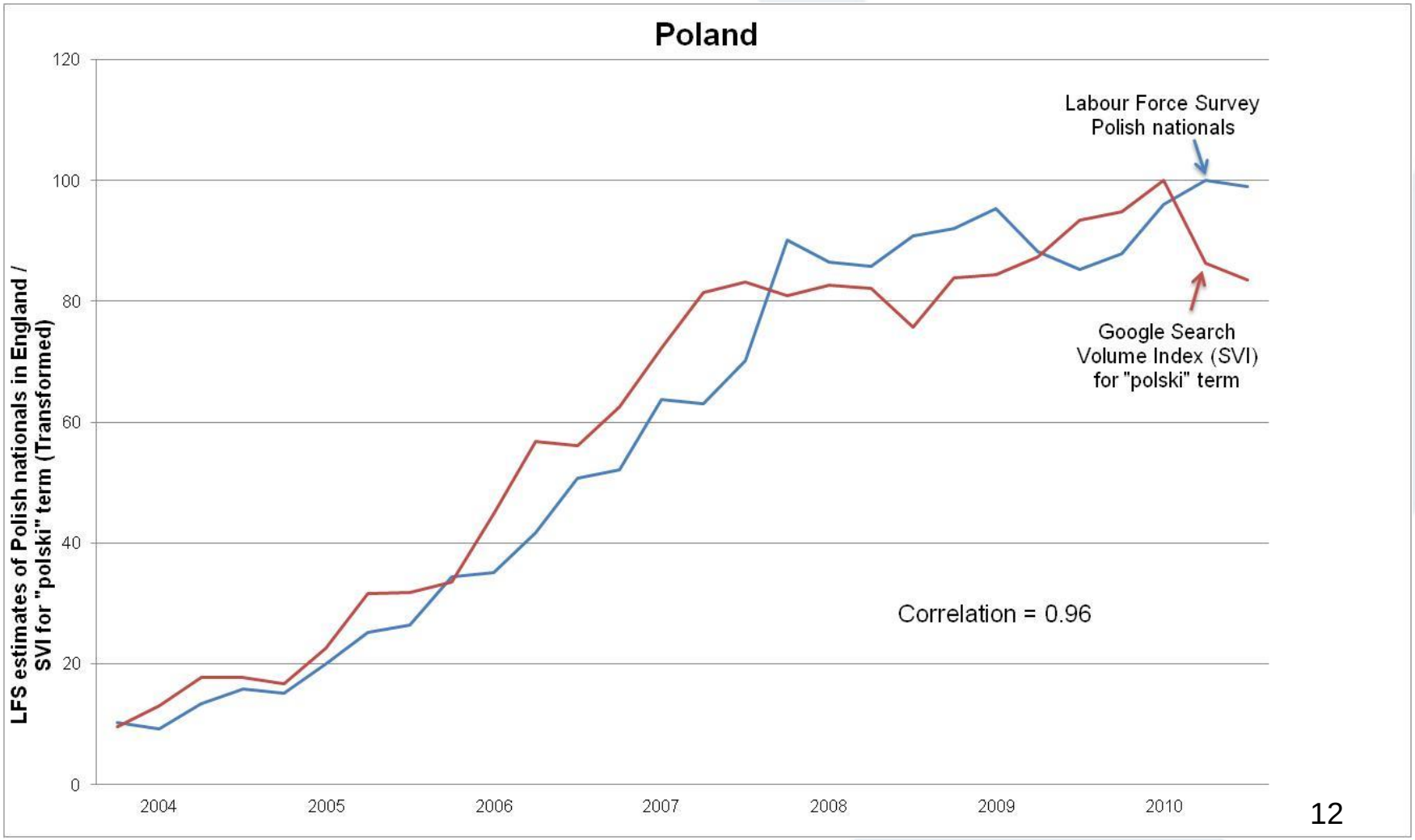


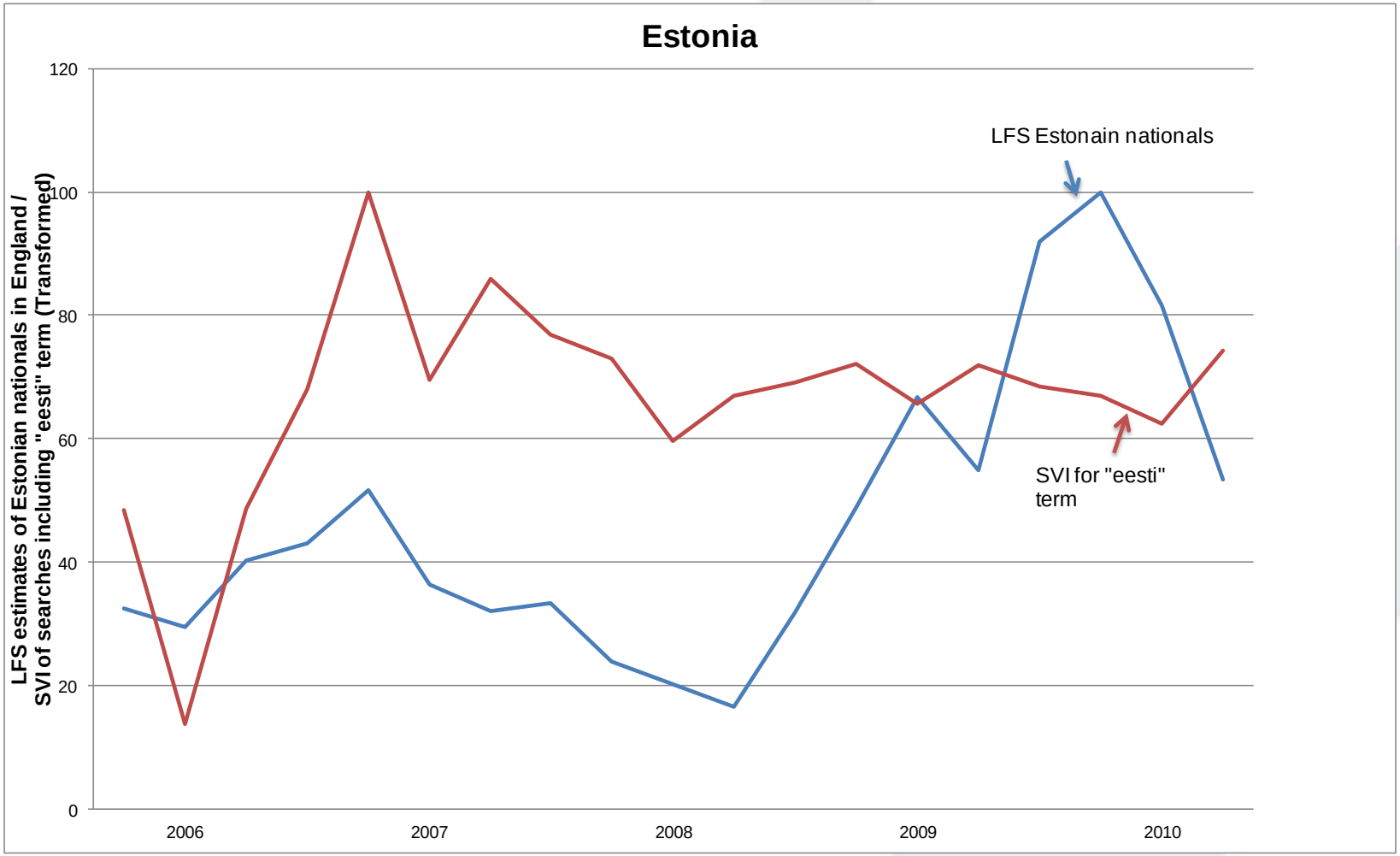
Source of the Dataset: TIM Big Data Challenge 2015, www.telecomitalia.com/bigdatachallenge

Google trends

Rational: Use of Google trends to inform on international migration

- Focus on EU expansion in 2004 and 2007
- Comparisons between Google SVI series and Labour Force Survey estimates
- Use of search term in native language


Poland
LFS estimates of Polish nationals in England / SVI for "polski" term (Transformed)
120
100
80
60
40
20
0
2004
2005
2006
2007
2008
2009
2010
Labour Force Survey
Polish nationals
Google Search
Volume Index (SVI)
for "polski" term
Correlation = 0.96


Estonia
LFS estimates of Estonian nationals in England / SVI of searches including "eesti" term (Transformed)
120
100
80
60
40
20
0
2006
2007
2008
2009
2010
LFS Estonain nationals
SVI for "eesti" term

Benefits

- Improve quality – more timely, more frequent, more relevant
- Reduce respondent burden
- Create efficiencies
- Produce new or complimentary outputs
- Improve operational processes

Challenges

- Bias/coverage
- Definitions
- Understanding and measuring uncertainty
- Data access and control
- Ethical and privacy issues

Managing the challenges

- Understand and explore the data
- Develop methods, estimation frameworks and quality frameworks for integrating survey, Census, admin and big data sources
- Ensure fit for purpose – experimental outputs, embrace new definitions, use for QA
- Legislation
- Establish ethical guidelines and independent advisory bodies, invest in public acceptability research
- Collaborative working – NSOs, academia, commercial sector

Contact

- Webpage: ONS.gov.uk
- Email: bigdata@ons.gov.uk
- Twitter: @naylor_jane