Consumer Data Research Centre

An ESRC Data Investment

Quantitative Geography and the SmartCensus

Paul Longley
Professor of Geographic Information Science, UCL

Consumer Data Research Centre

An ESRC Data Investment

Consumer Data Research Centre

An ESRC Data Investment

Consumer Data Research Centre

An ESRC Data Investment

Free Movement of Data (1): Private/public sectors

- Big Data
  - Social survey response rates
- Real share of *consumer* data:
  - forms
  - Time and space: 24 hour cities
- Public versus private sector
  - The value of open platforms
- Exhaust; metadata and provenance?
  - Linkage
  - Shared research challenges

Consumer Data Research Centre

An ESRC Data Investment

In 1992 all the pollsters went wrong doing the same thing. [In 2015] they went wrong doing different things. Some firms contact people via telephone, others online, and they ask different questions [or use different] statistical methods.

Joe Twyman, YouGov

UCL

Consumer Data Research Centre
An ESRC Data Investment
YOUR M&S
Sainsbury's
Boots
Demographics USER GROUP
Making Data Work
TESCO
John Lewis
The co-operative
British Gas
e·on
Camelot
WHITBREAD
BARCLAYS
DIXONS RETAIL
BRINGING LIFE TO TECHNOLOGY
LDC
gsk GlaxoSmithKline
Argos
SHOP DIRECT
goHenry
Experian
Thomas Cook
CACI
movement strategies
easyJet
AXA PPP HEALTHCARE

EMERGENCY
DOOR RELEASE

Consumer Data Research Centre

An ESRC Data Investment

Co-Production/Analytics

Data Provider

Data Science Projects

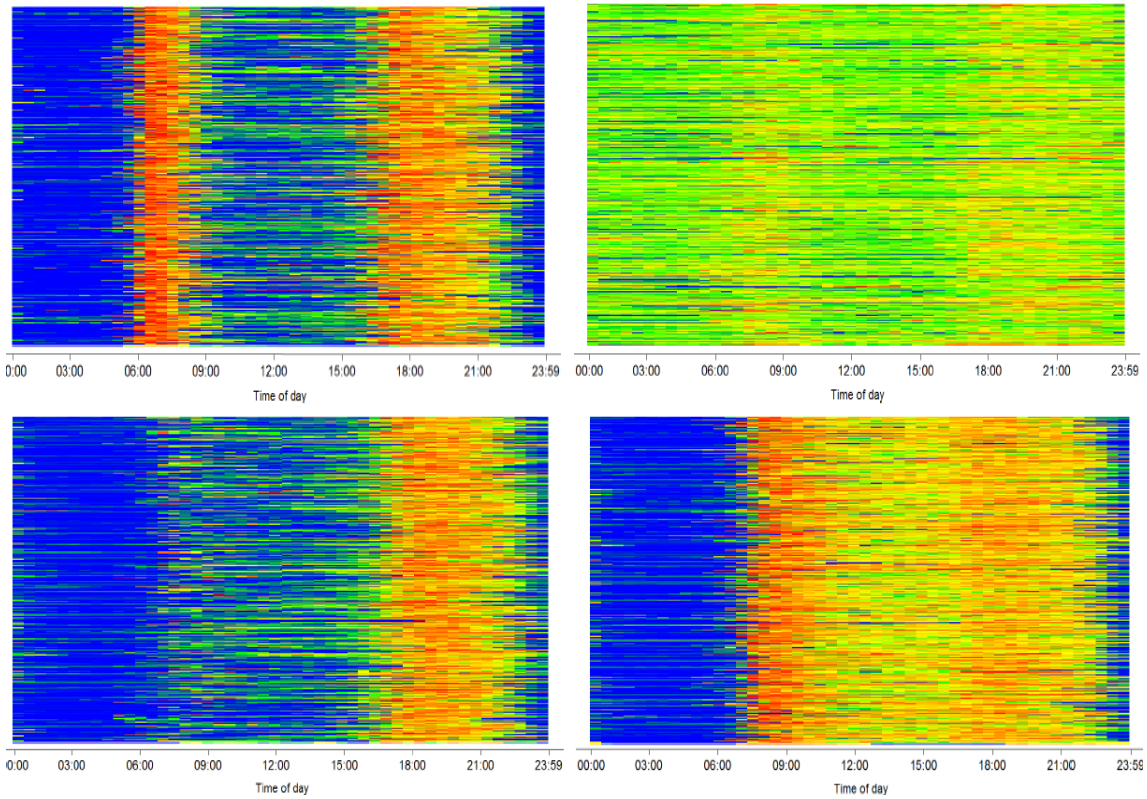
PhD Projects

MSc Projects

Consumer Data Research Centre

An ESRC Data Investment

Domestic Energy Providers



Timeline

- 2014 – CDRC Dissertation student project
- 2015 – Second dissertation in the same research field
- 2015 – Co-funding of a PhD with the CDRC
- 2015 – nationwide data for the CDRC secure data service

- CDRC Dissertation project (Nicholas Samson, 2014)
  - Grouped smart meters into 4 key time profiles
  - Identified key demographic characteristics of each profile

Consumer Data Research Centre

An ESRC Data Investment

Co-Production/Analytics

Data Provider

Data Science Projects

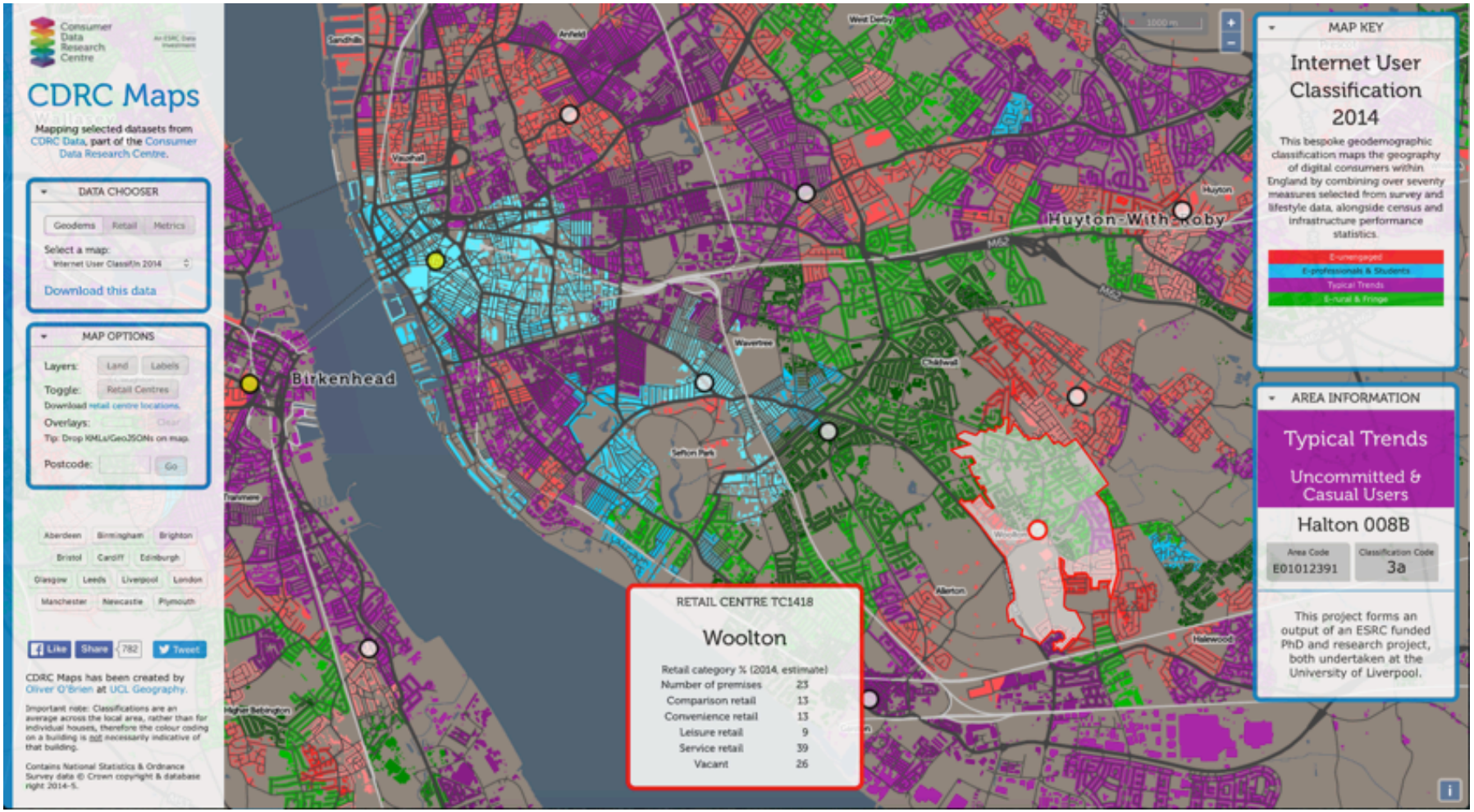
PhD Projects

MSc Projects

Consumer Data Research Centre

An ESRC Data Investment

Internet User Classification



Consumer Data Research Centre

An ESRC Data Investment

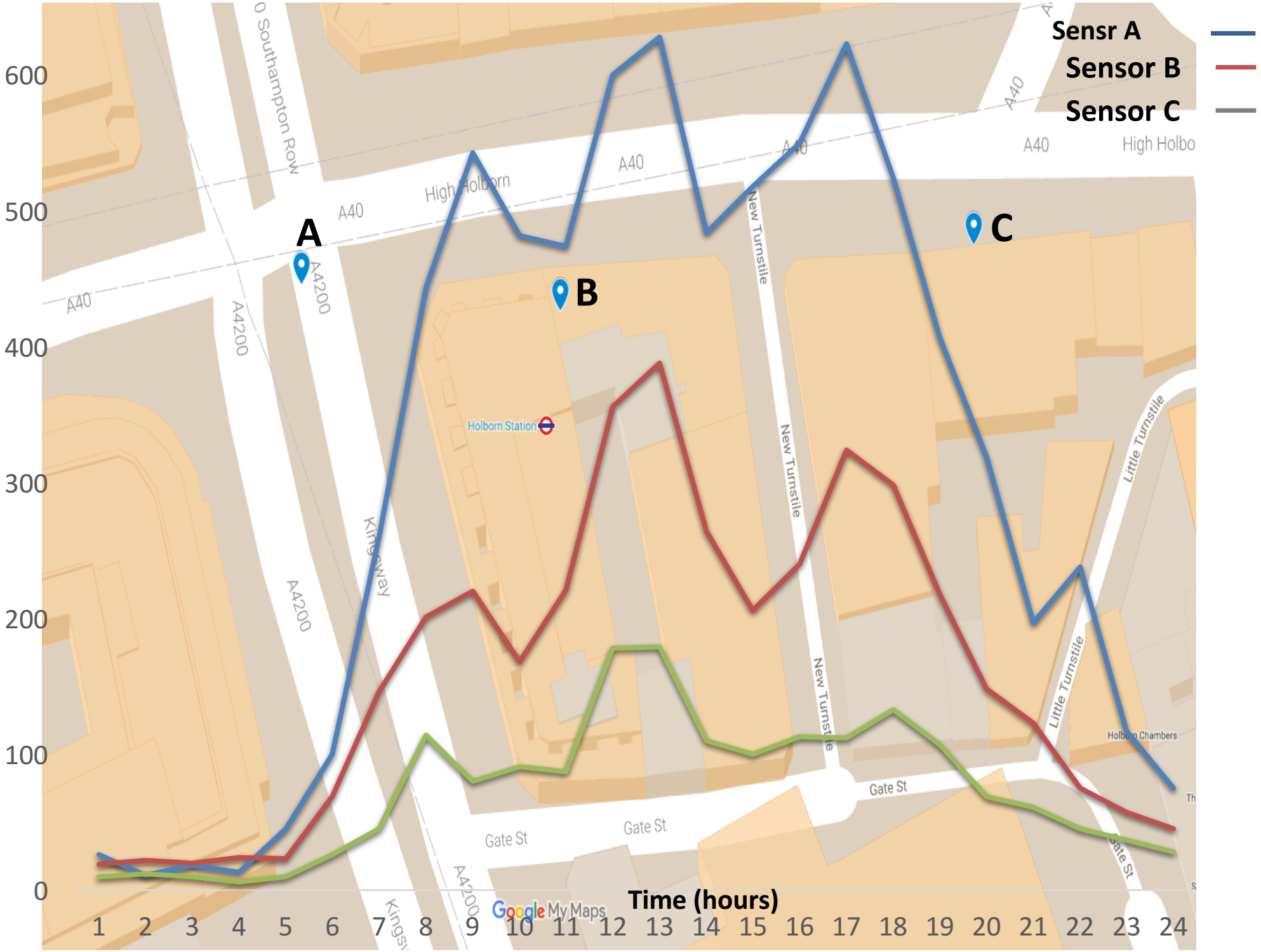
Co-Production/Analytics

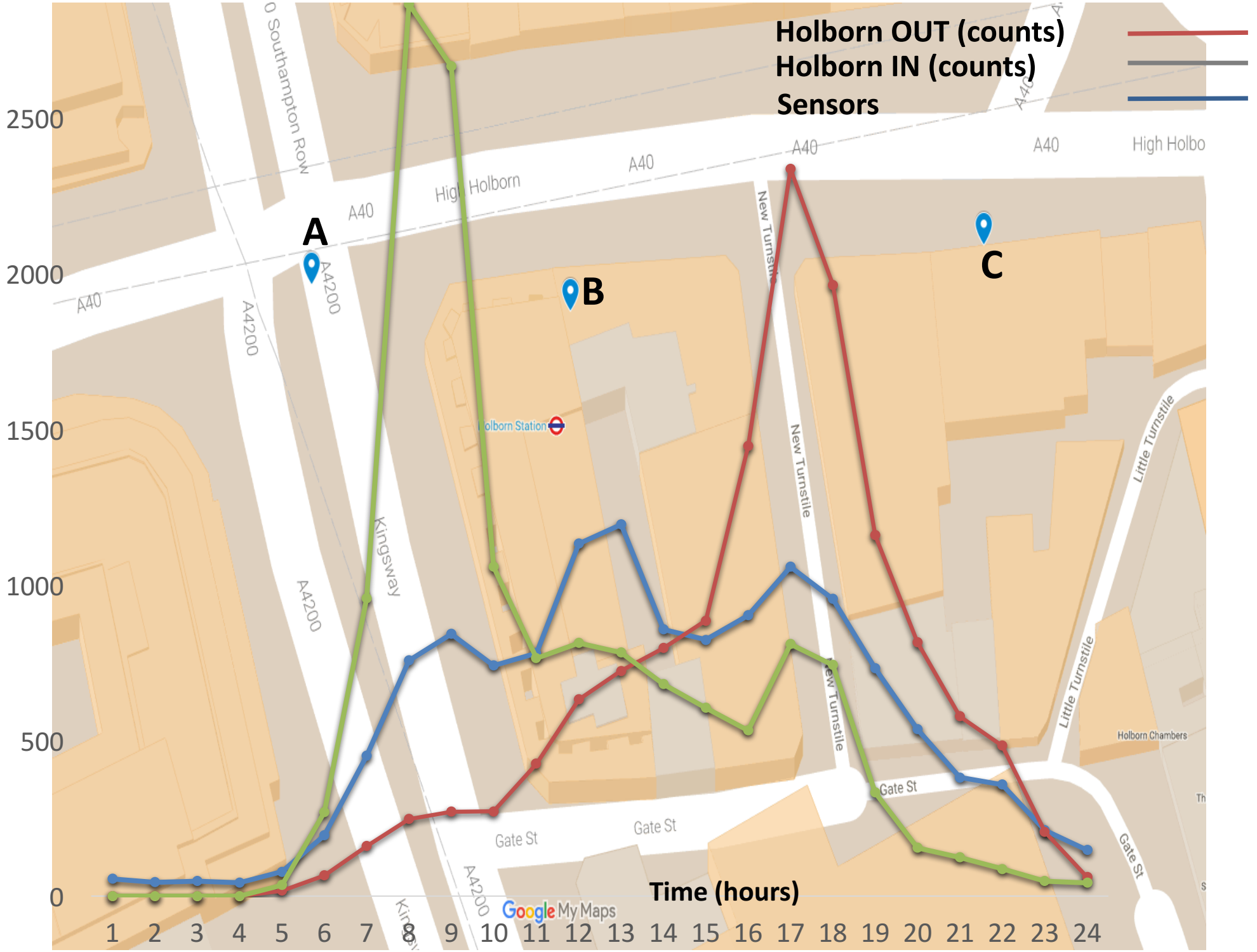
Data Provider

Data Science Projects

PhD Projects

MSc Projects


Sensr A
Sensor B
Sensor C
600
500
400
300
200
100
0
A
B
C
A40
A4200
High Holborn
Holborn Station
New Turnstile
Little Turnstile
Gate St
Kingsway
Southampton Row
Holborn Chambers
Google My Maps
Time (hours)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24


Holborn OUT (counts)
Holborn IN (counts)
Sensors
2500
2000
1500
1000
500
0
A
B
C
A40
A4200
Kingsway
New Turnstile
Little Turnstile
Gate St
Holborn Chambers
Time (hours)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Google My Maps

Consumer Data Research Centre

An ESRC Data Investment

Free Movement of Data (2): Open Data and Access

- Economic case fro Open platforms, e.g. TfL
- Apportioning the cost of infrastructure creation and maintenance, e.g. NMOs
- Disclosure control and new modes of access, e.g. EthnicityEstimator
- Implications for the practice of science

Consumer Data Research Centre

An ESRC Data Investment

Free Movement of Data (3): Policy Issues

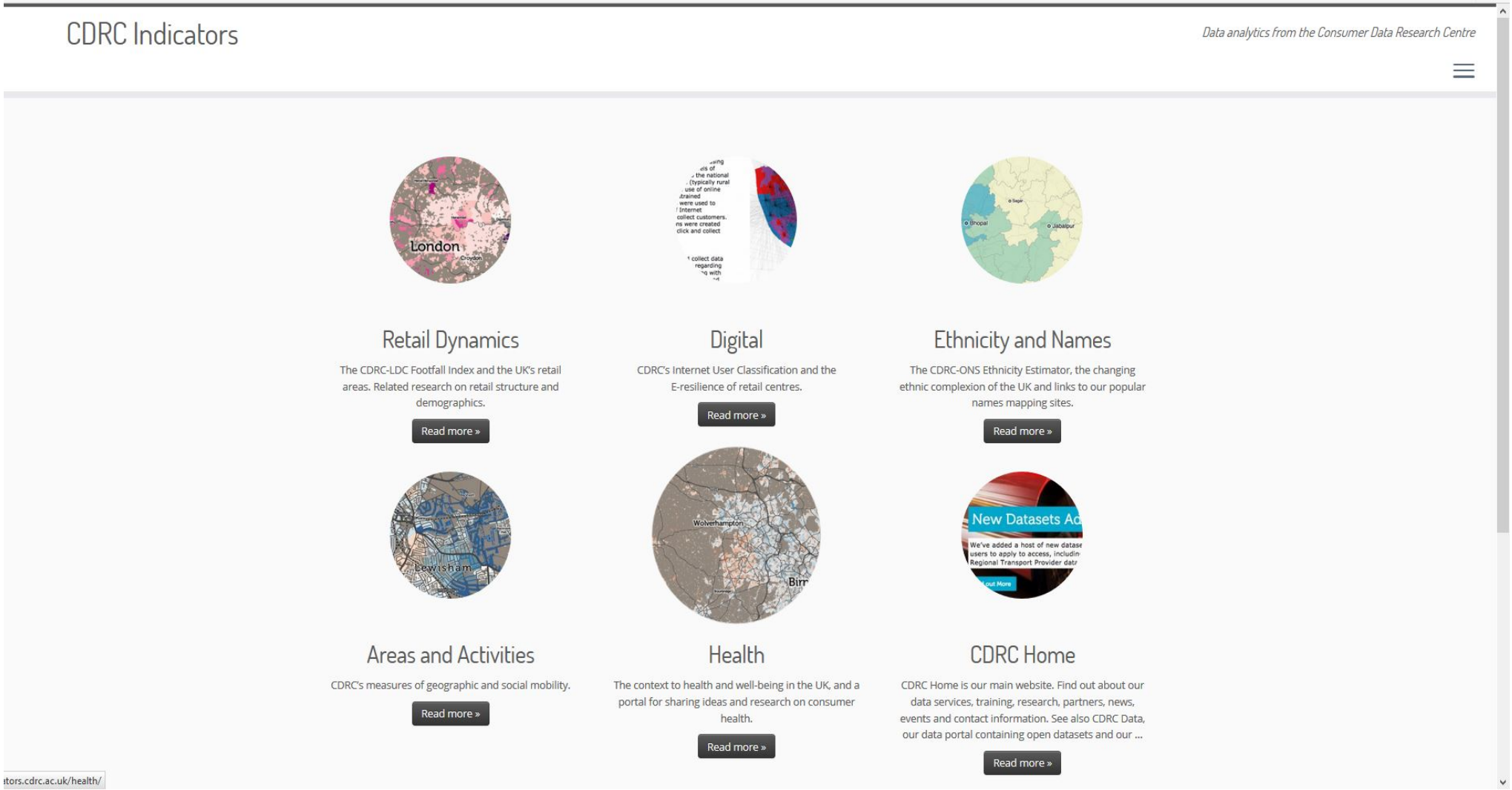
- Public and private goods: infrastructure and industrial strategy

Consumer Data Research Centre

An ESRC Data Investment

Next Steps

- CDRC research indicators

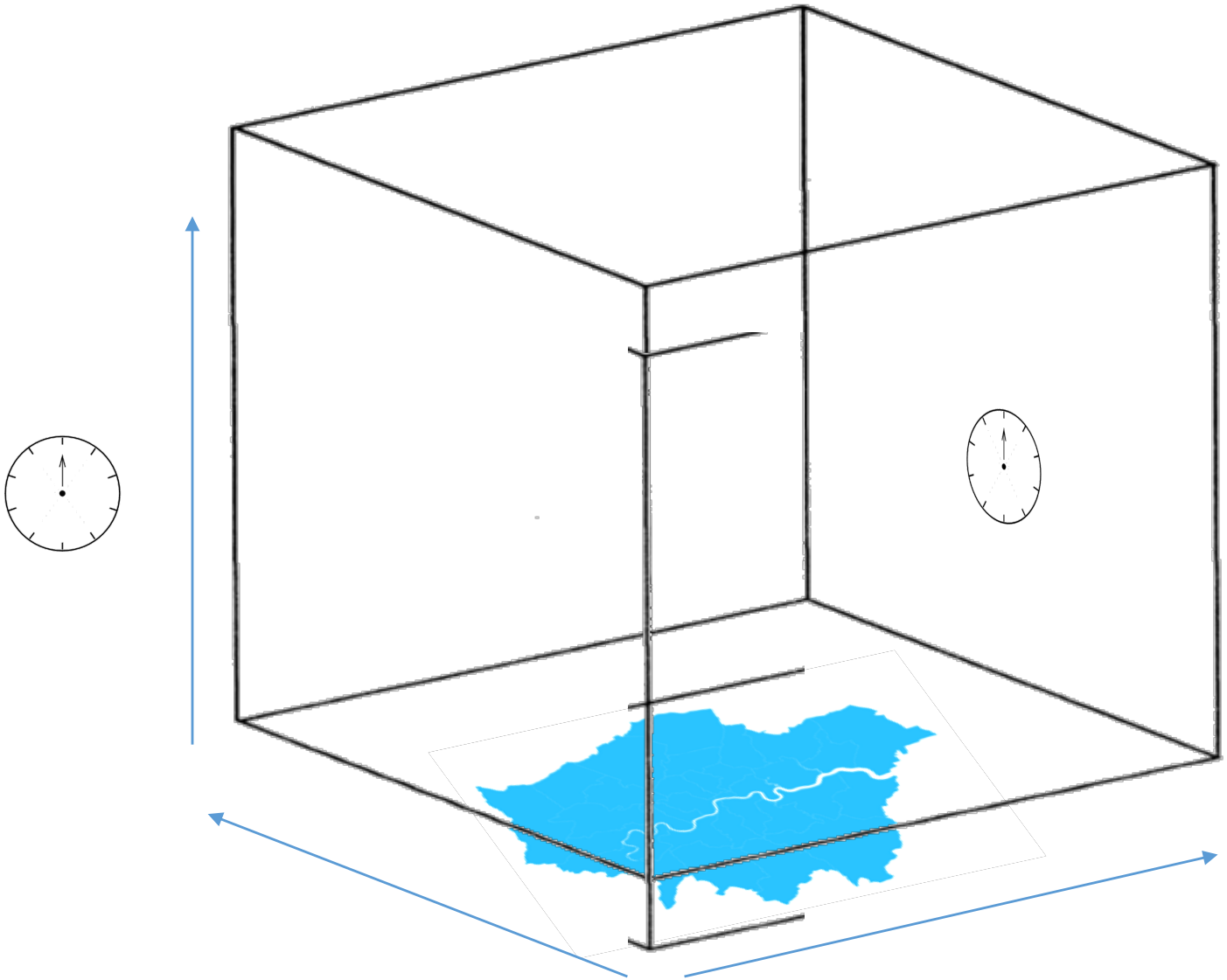


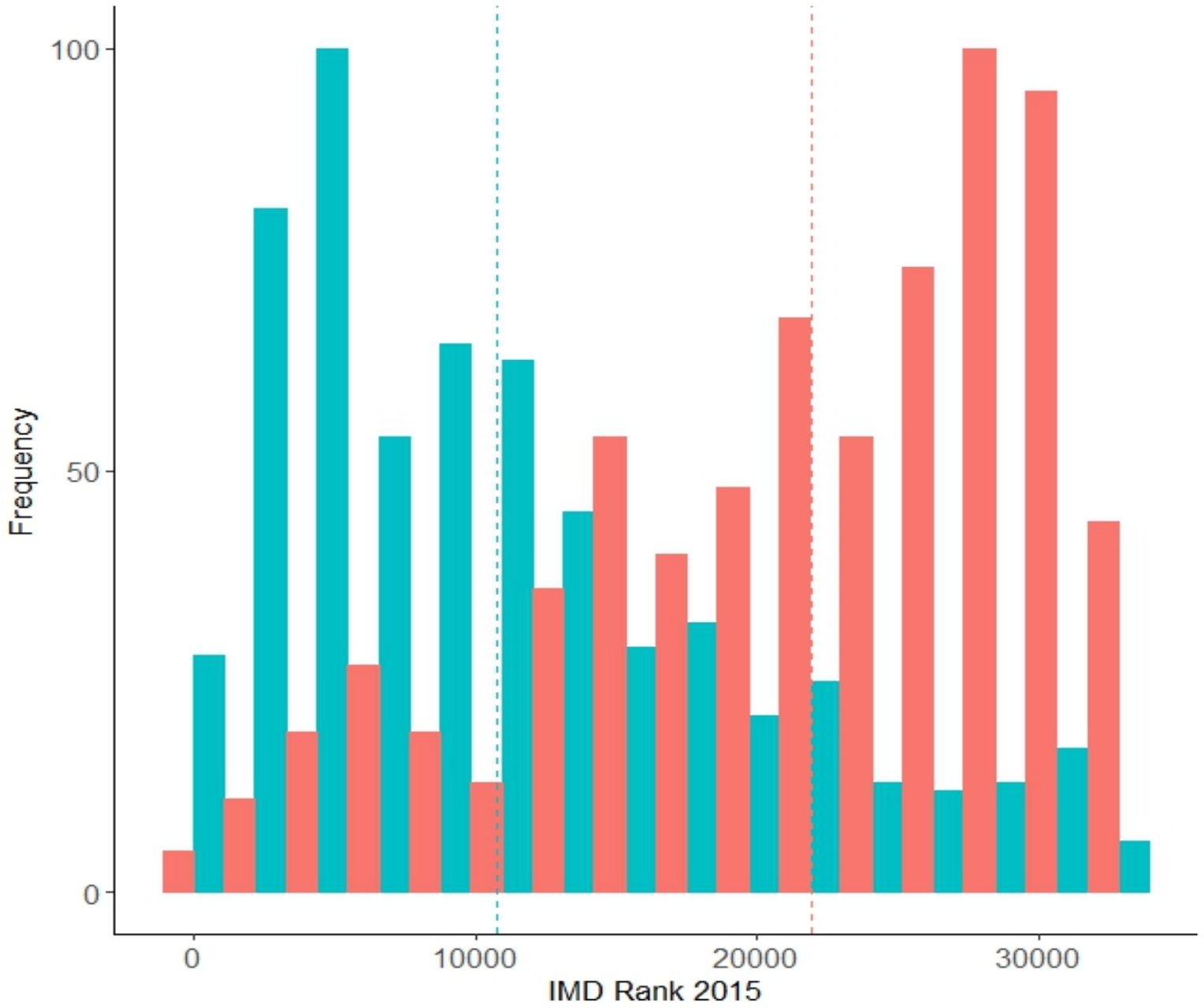
Consumer Data Research Centre

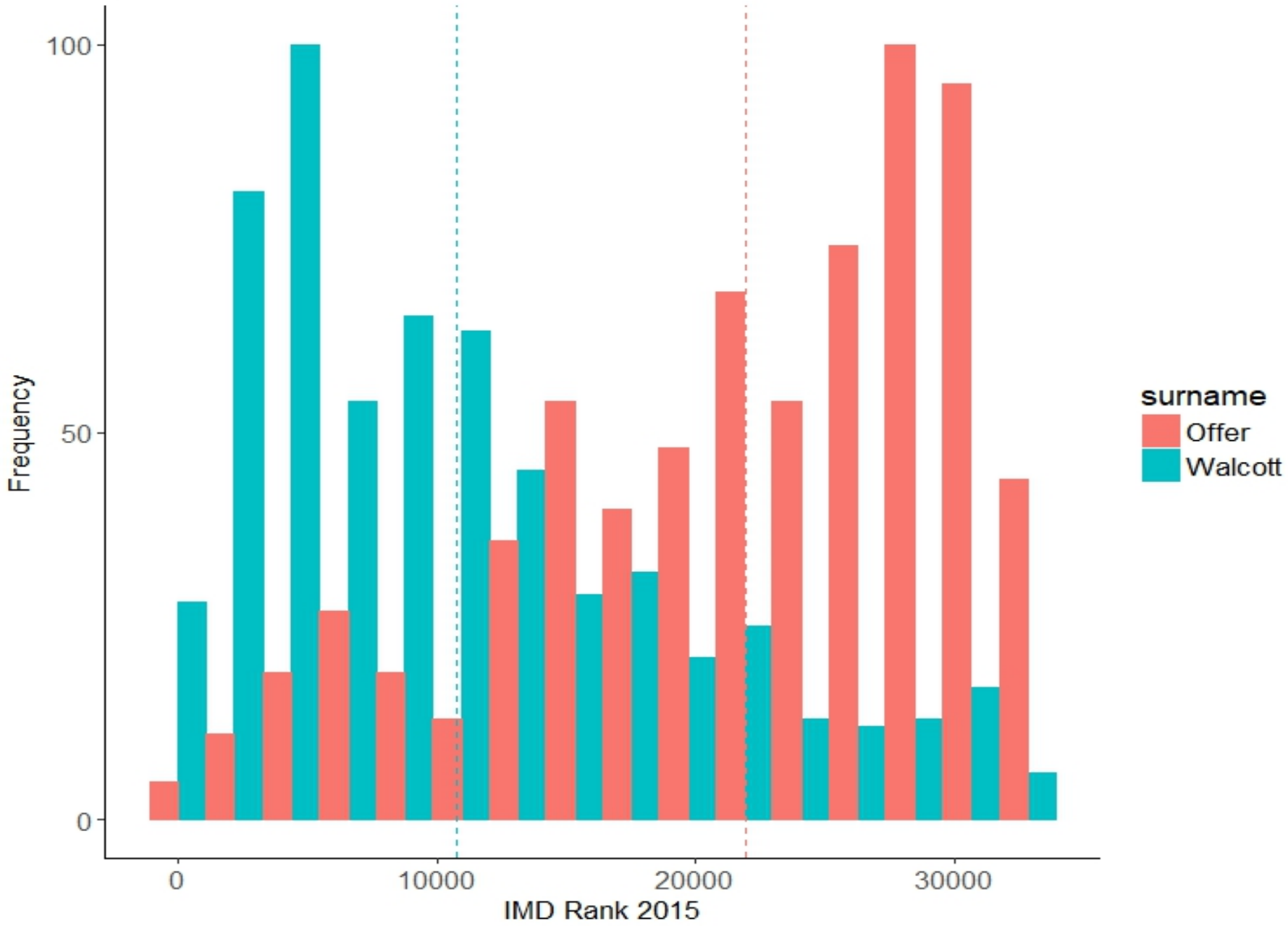
An ESRC Data Investment

Free Movement of Data (3): Policy Issues

- Public and private goods: infrastructure and industrial strategy
  - Competitive economies / competition
  - No open market for data
  - New silos: A, A, A, F, M
- UK Digital Economies Act
  - Passporting for research?
- Linkage:
    - Indicators, e.g. health, social mobility, carbon footprints
  - Preservation (of what?); curation
  - Research exemptions




100
50
0
Frequency
0
10000
20000
30000
IMD Rank 2015


100
50
0
Frequency
0
10000
20000
30000
IMD Rank 2015
surname
Offer
Walcott

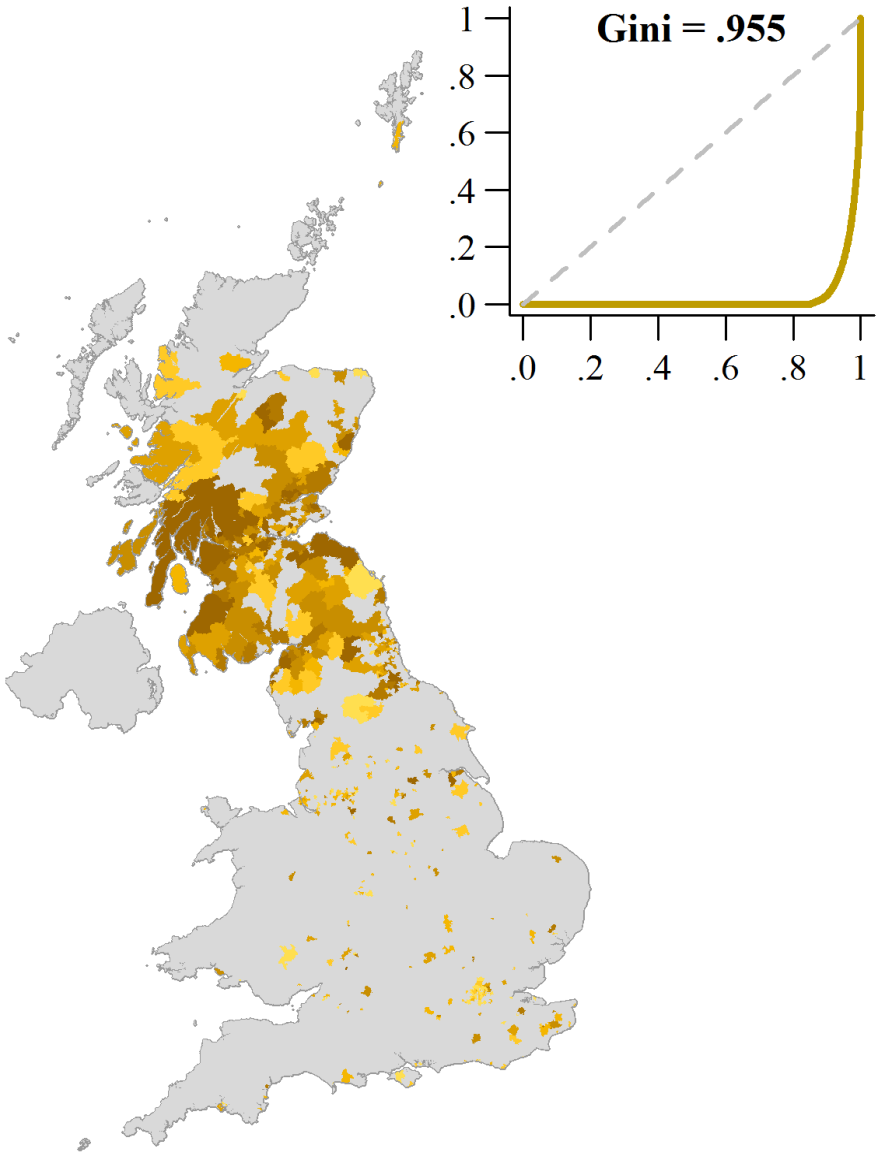
: ‘Surnames and Social Mobility’,Gregory Clark and Neil Cummins

http://www.econ.ucdavis.edu/faculty/gclark/ecn110a/readings/Surname%20Mobility%202013.pdf

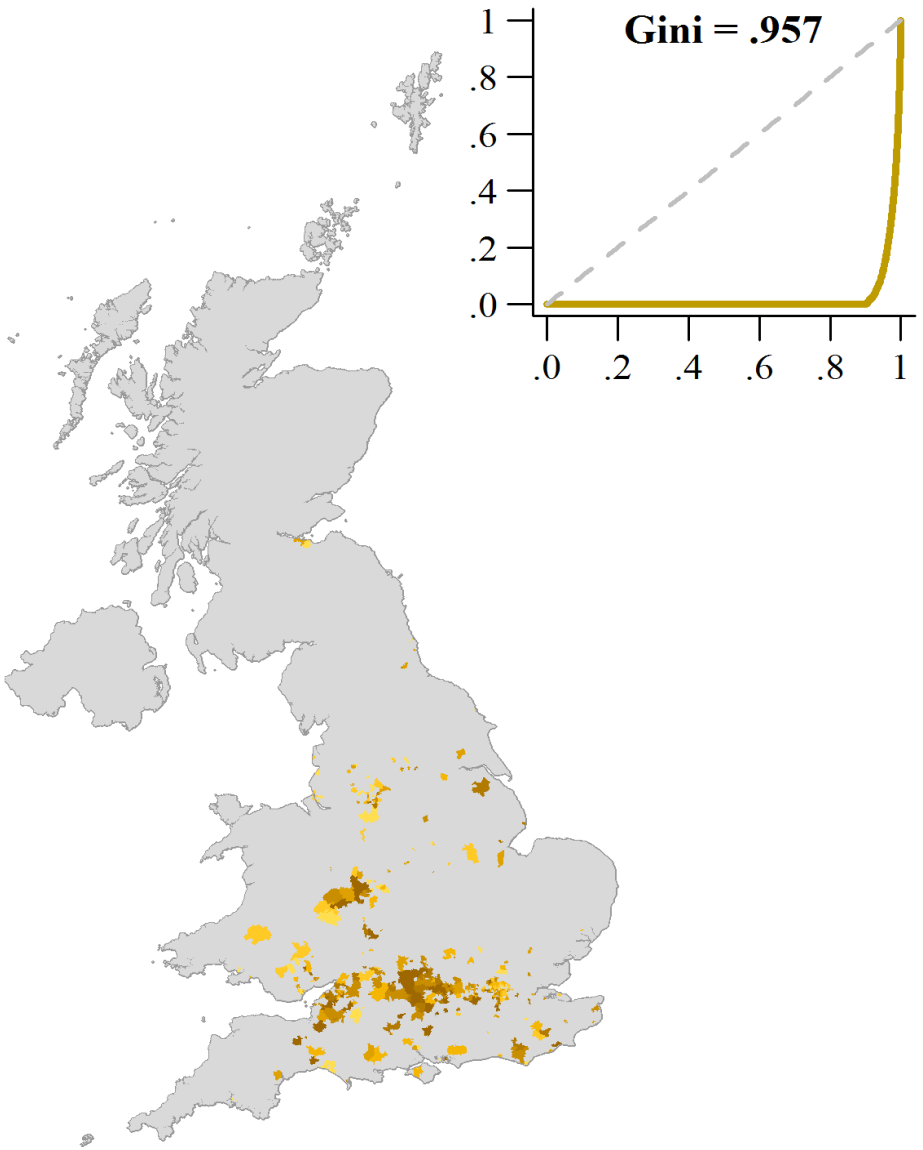
‘using educational status in England 1170-2012 as an example, … the true underlying [intergenerational] correlation of social status is in the range 0.75-0.85. Social status is more strongly inherited even than height.’

This ‘stems from the nature of inheritance of characteristics within families. Strong forces of familial culture, social connections, and genetics must connect the generations. There really are quasi-physical “Laws of Inheritance.”’

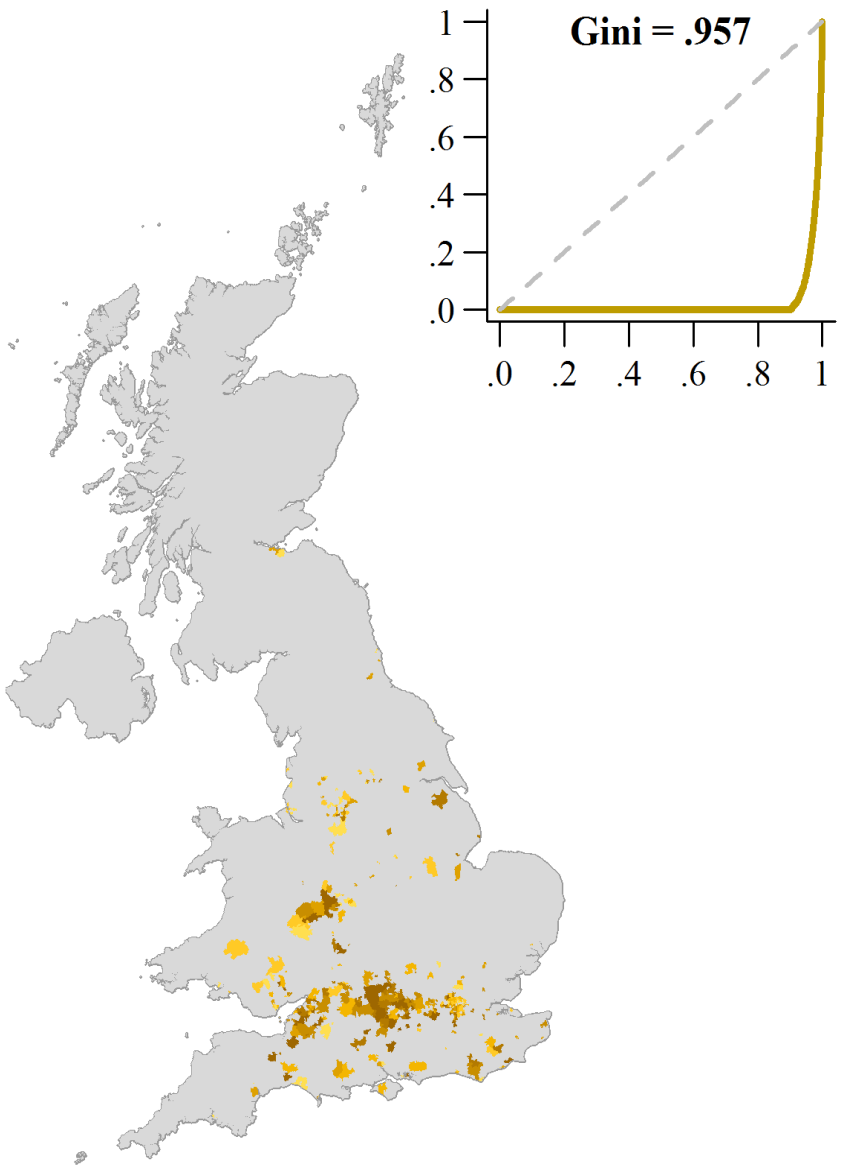
Blair (1881)



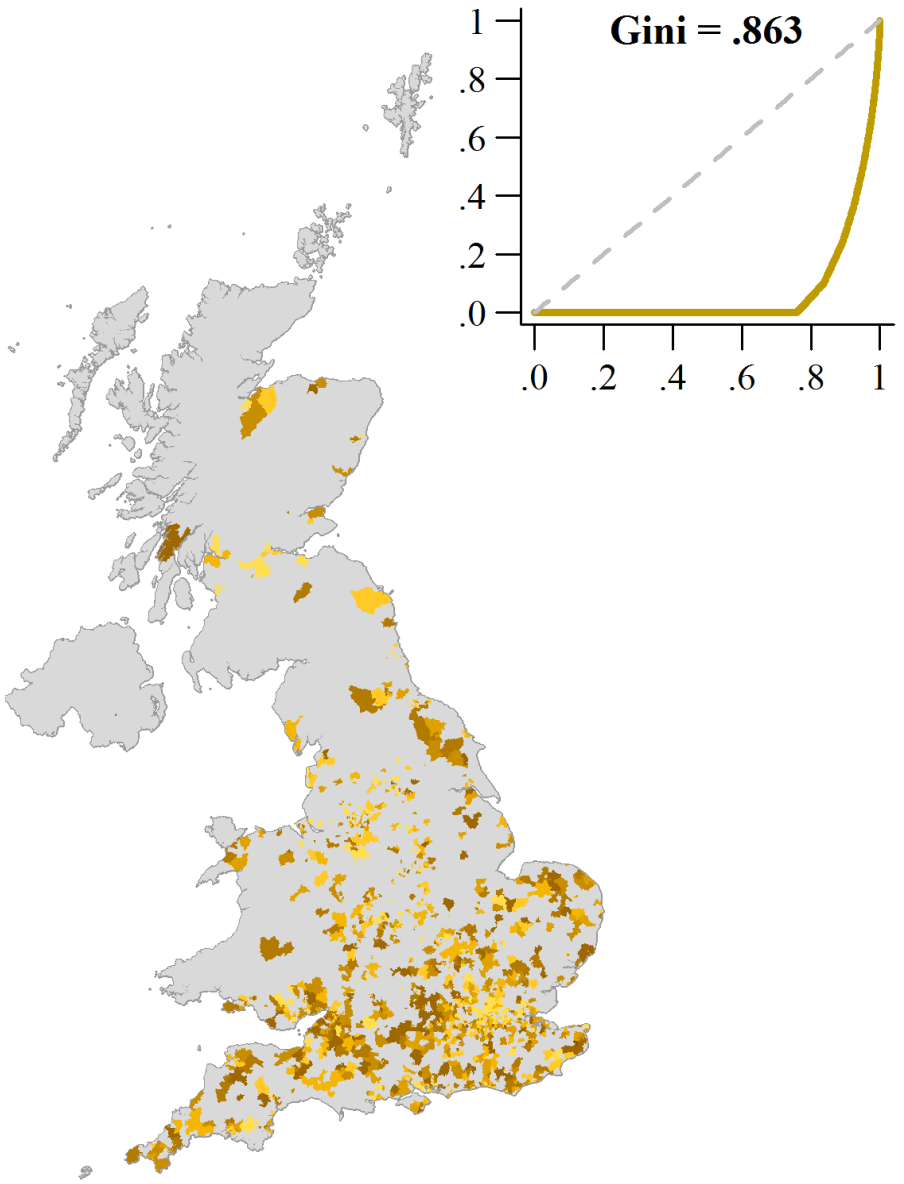
Thatcher (1881)

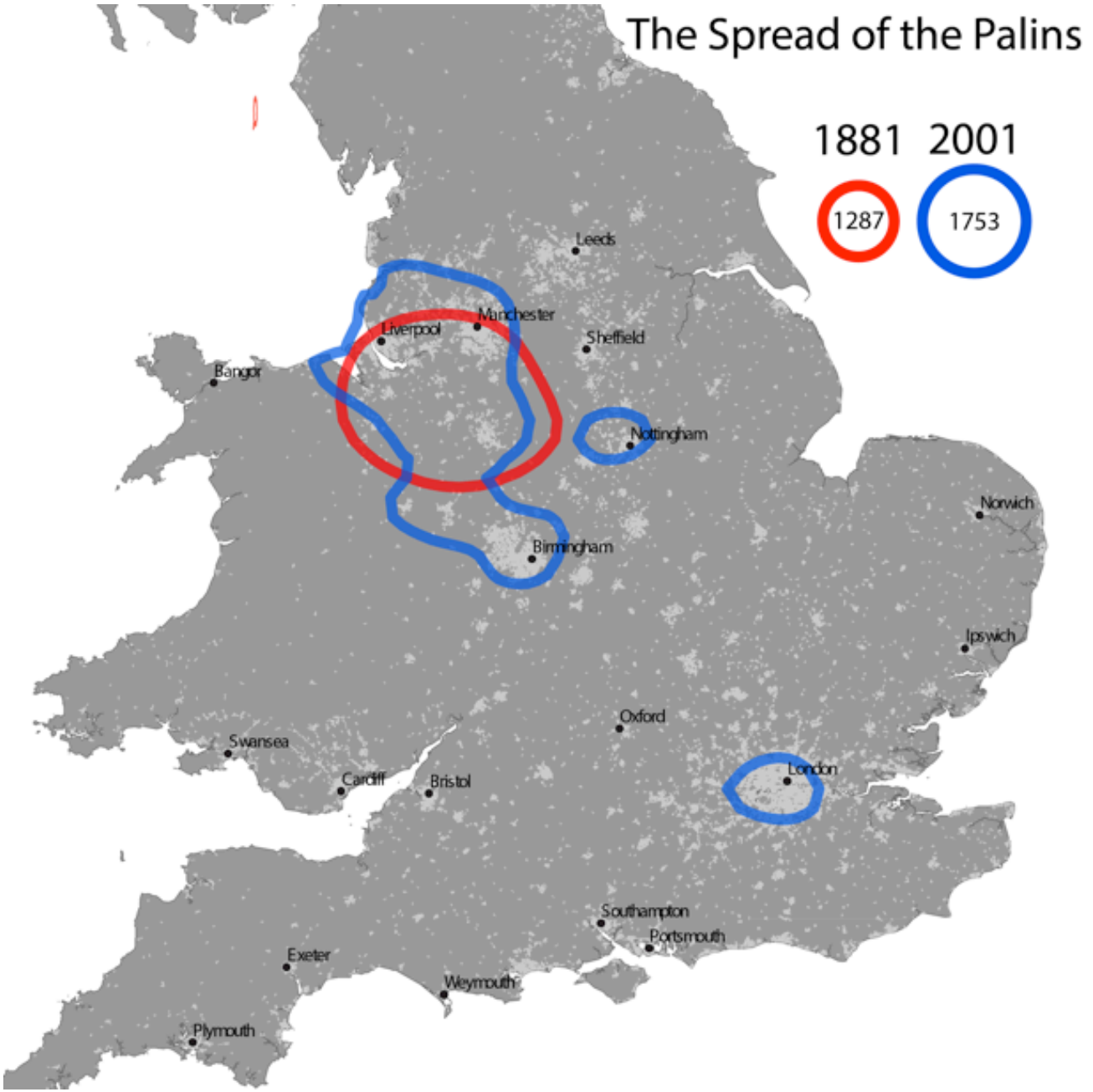


Thatcher (1881)

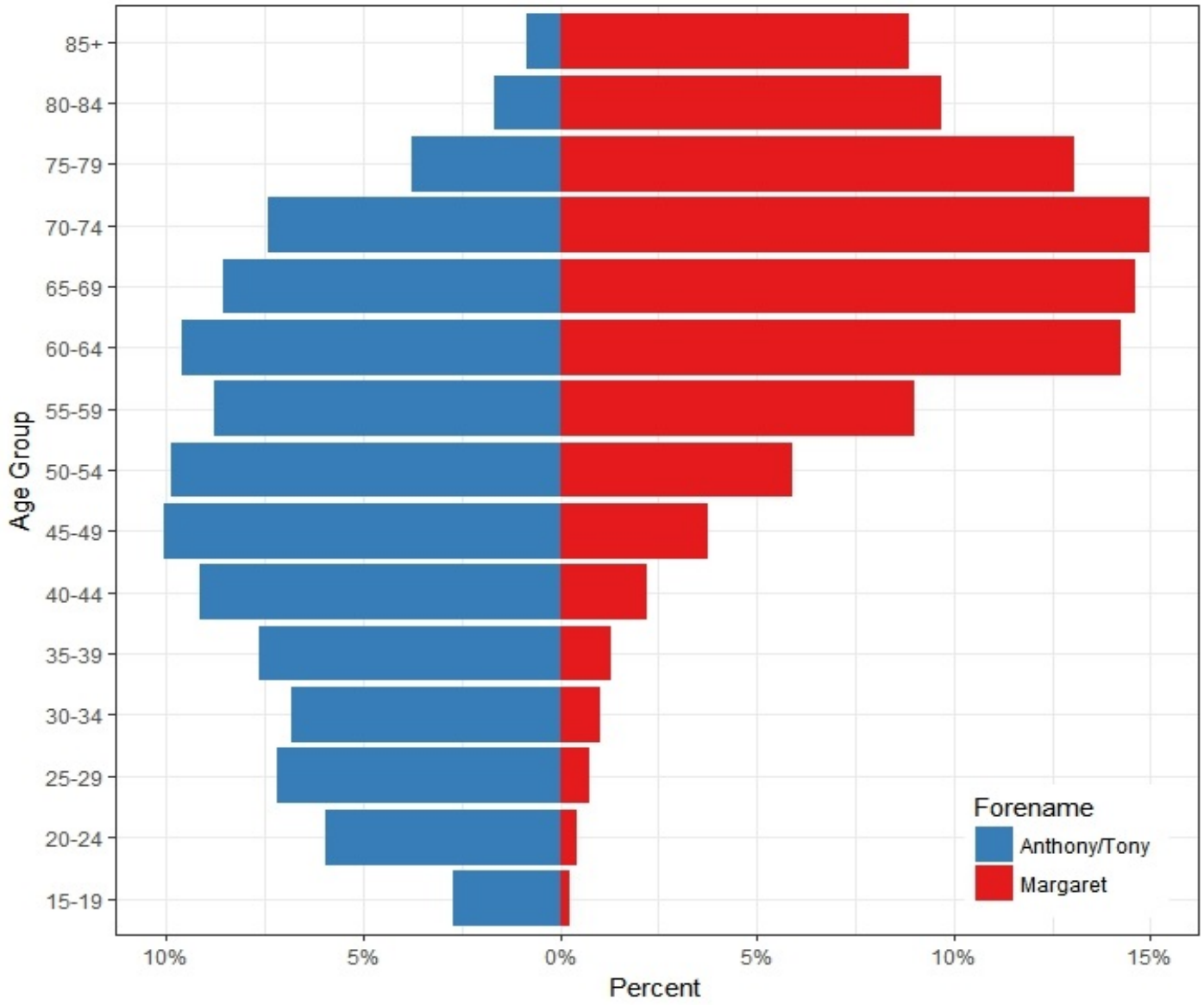


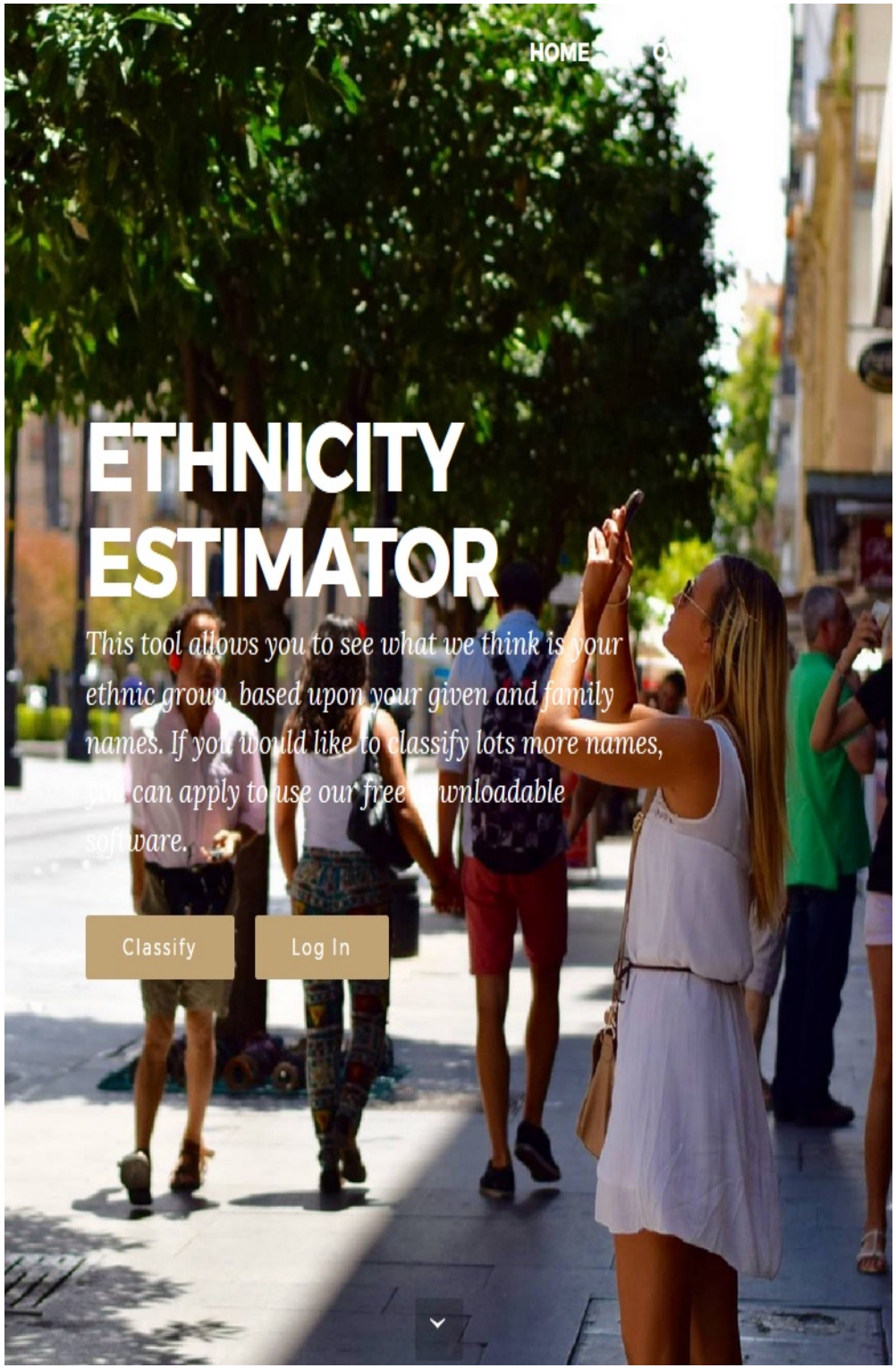
Thatcher (2016)





Courtesy:
James
Cheshire


85+
80-84
75-79
70-74
65-69
60-64
55-59
50-54
45-49
40-44
35-39
30-34
25-29
20-24
15-19
Age Group
10%
5%
0%
5%
10%
15%
Percent
Forename
Anthony/Tony
Margaret



HOME OVERVIEW

Classify a name

This Website is part of a project devised by the Consumer Research Data Centre (CDRC) in association with the UK Office for National Statistics (ONS). The CDRC is funded by the Economic and Social Research Council (ESRC)

Results from searches and user feedback are used to improve our ongoing research.

Type a Forename and a Surname in any style you want: upper, lower case; any special character. You need to fill at least one of the two fields.

Forename: Muhammad

Surname: Gill

Age (optional): 0

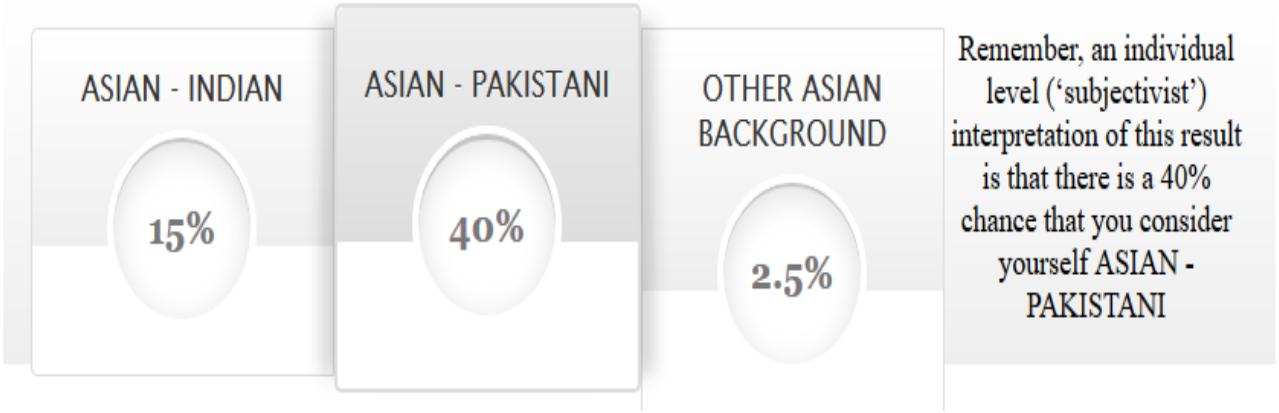
Gender (optional): ---

CLASSIFY

About Ethnicity Estimator

Our classifier is based on analysis of combined names data from consumer sources assembled by CDRC to survey and survey data for England and Wales held securely by ONS. **The classifier does not use or contain any personal data** Predicting ethnicity cannot be an exact science – not least because some people share identical names despite having quite different ethnicities. We acknowledge this by reporting the three most likely estimated ethnicities for a name based on a likelihood score ranging from 0 to 1. For example, the result **"White British (35%); Asian – Pakistani (25%); Asian – Indian (12.5)"** can be understood as 'We think that there is a 25% chance that you describe yourself as White British, a 35% chance that you describe yourself as Pakistani and an 12.5% chance that you describe yourself as Indian (An alternative, frequentist, interpretation of the same numbers is that 35% of people who have your precise name describe themselves as White British, and so on. The numbers do not sum to 100% if there is a residual possibility of membership of one or more other ethnic groups and we do not report very small percentages). We can get things wrong and people themselves may see things differently. So we would really appreciate your feedback using the drop down menus provided in the next page. We have used a particular classification of ethnicity that is widely used in the UK by statistical agencies – but if you think this is unhelpful in any way please let us know this, too.

The ethnicity for Muhammad Gill has been estimated as:



Do you agree? Feedback will help us in our future research, so please let us know what you think, using the drop down boxes below.

Do we got it right? Yes

I consider myself to be: WHITE - BRITISH

Other feedback (max 500 characters including spaces):

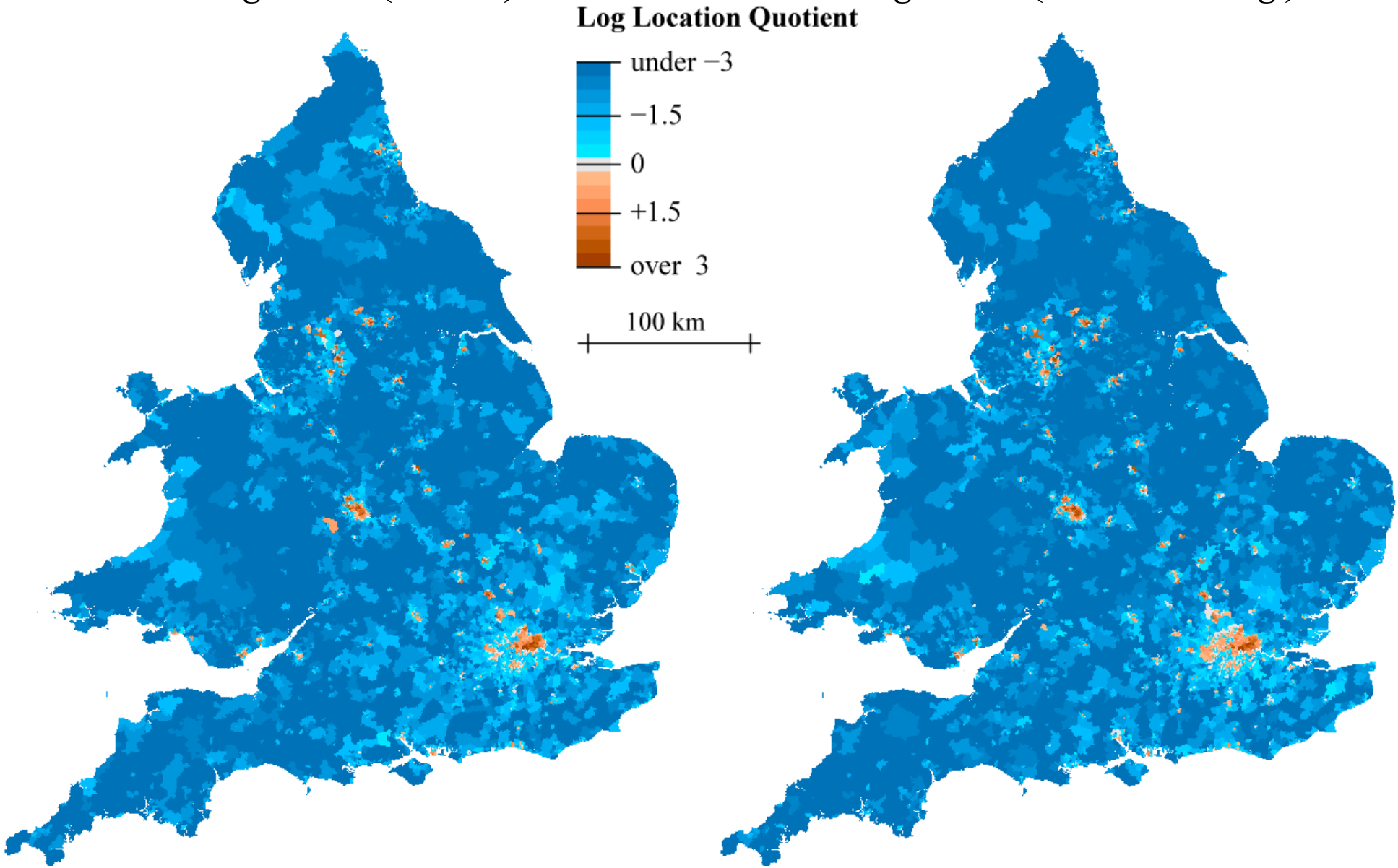
Send

Home

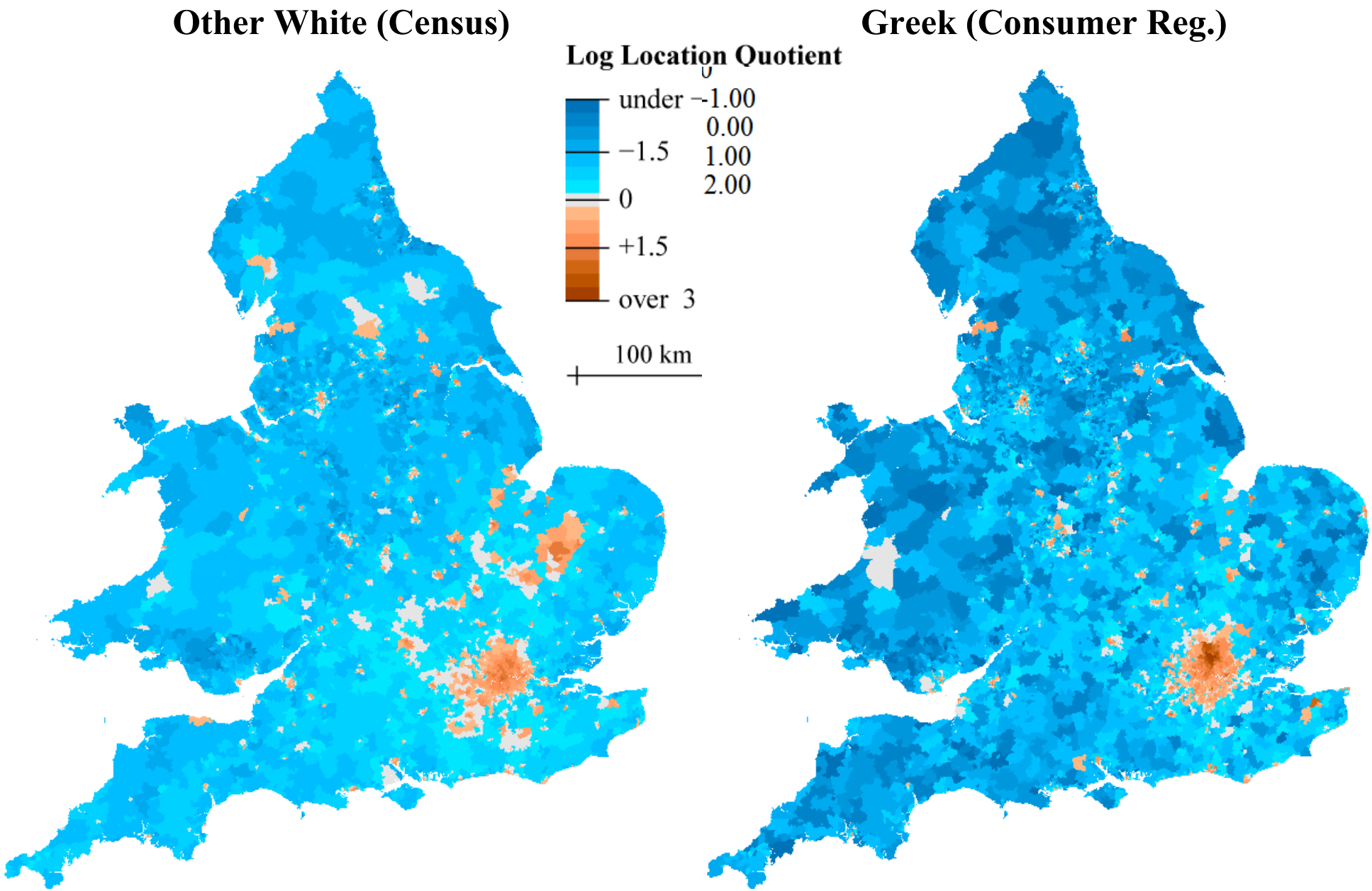
SPATIAL DISTRIBUTION OF ETHNIC GROUPS

Bangladeshi (Census)

Bangladeshi (Consumer Reg.)



SPATIAL DISTRIBUTION OF ETHNIC GROUPS



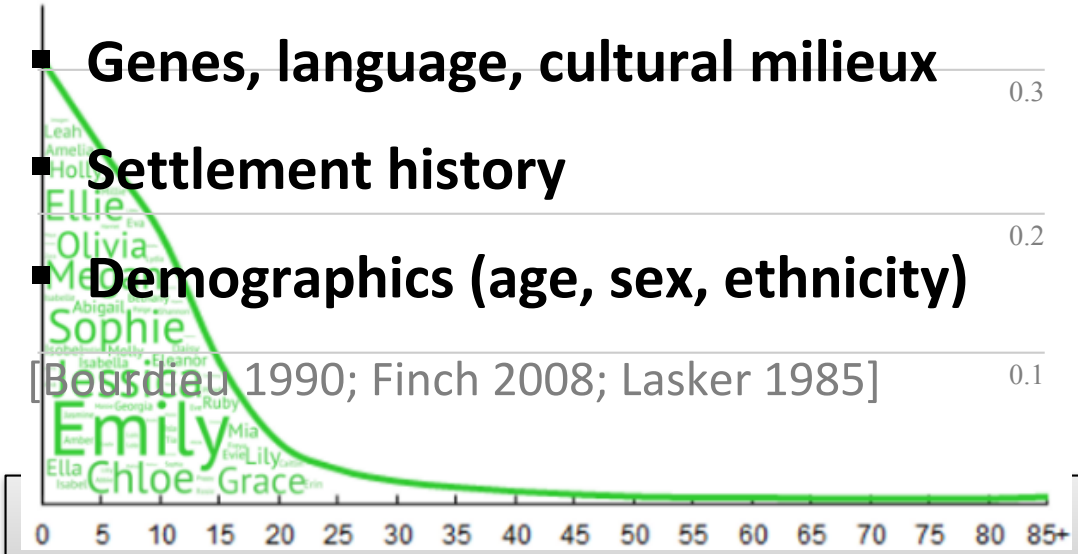
Population characteristics revealed by names

- Inheritance of surnames
- Regional name origins
- Periodical forename fashions
- Widely collected ‘Big Data’

Markers of

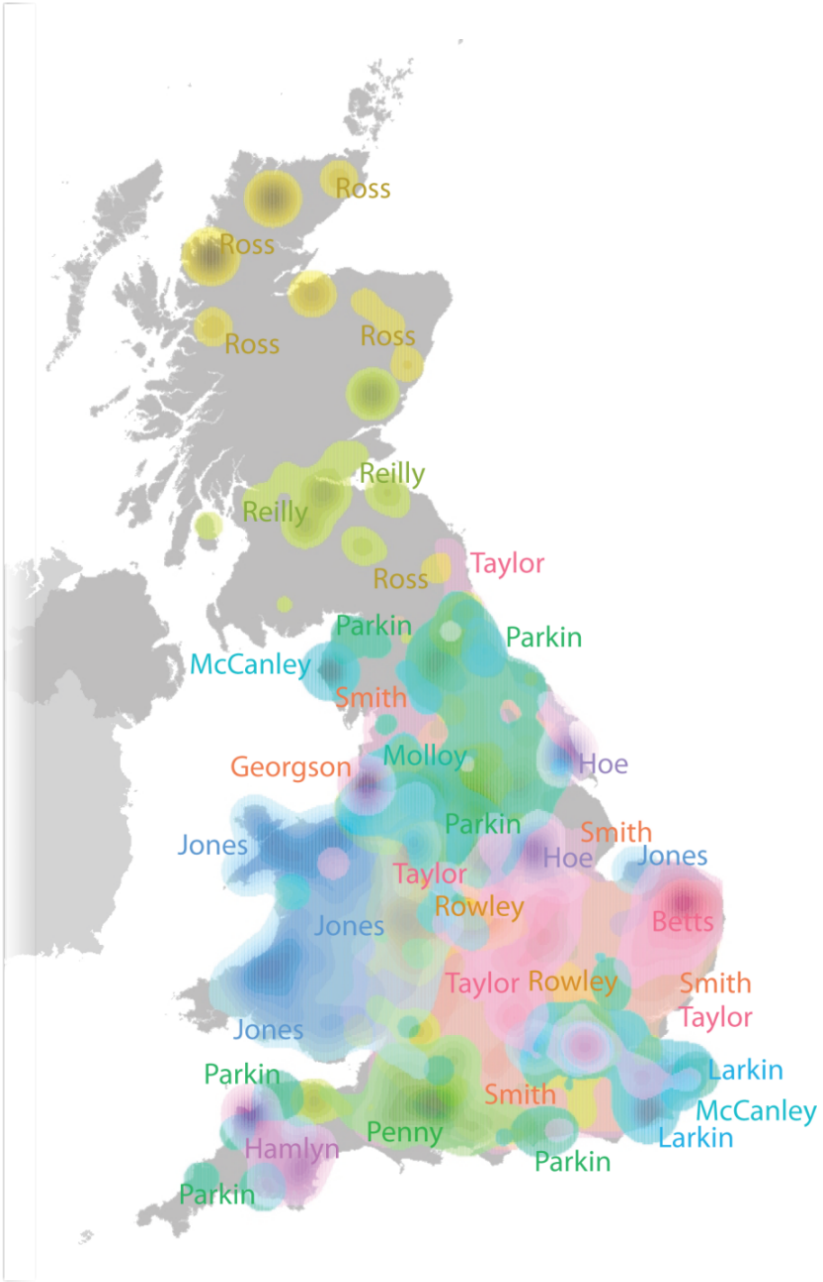
- Genes, language, cultural milieux
- Settlement history
- Demographics (age, sex, ethnicity)

[Bourdieu 1990; Finch 2008; Lasker 1985]



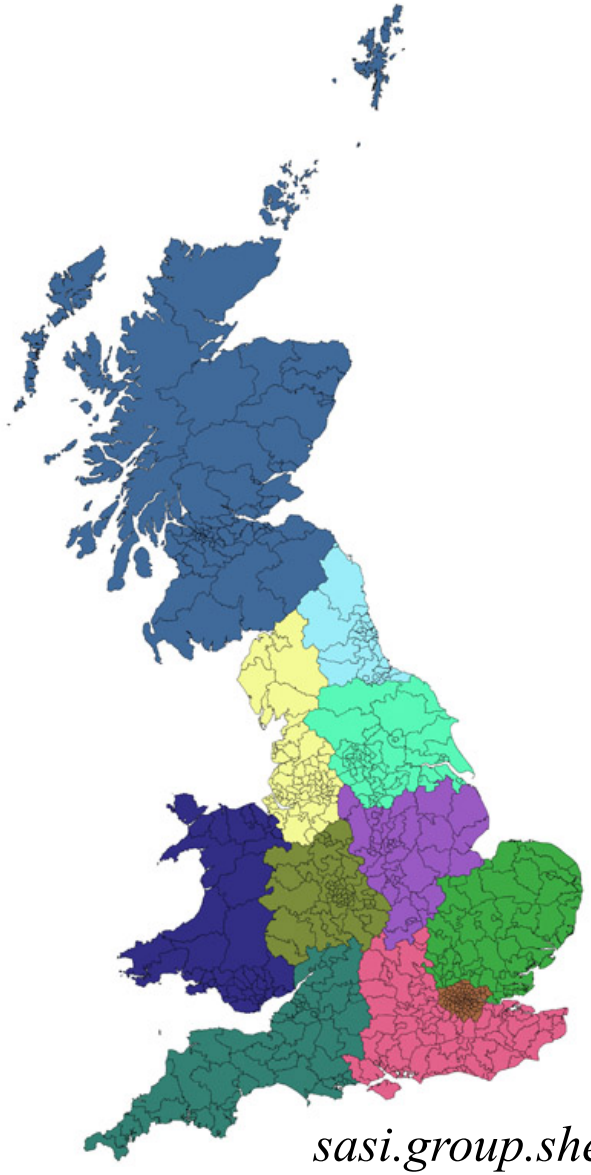
Consumer Registers 2007, 2013, 2014-2016

Lansley & Longley 2016 *J Retail Cons Serv*



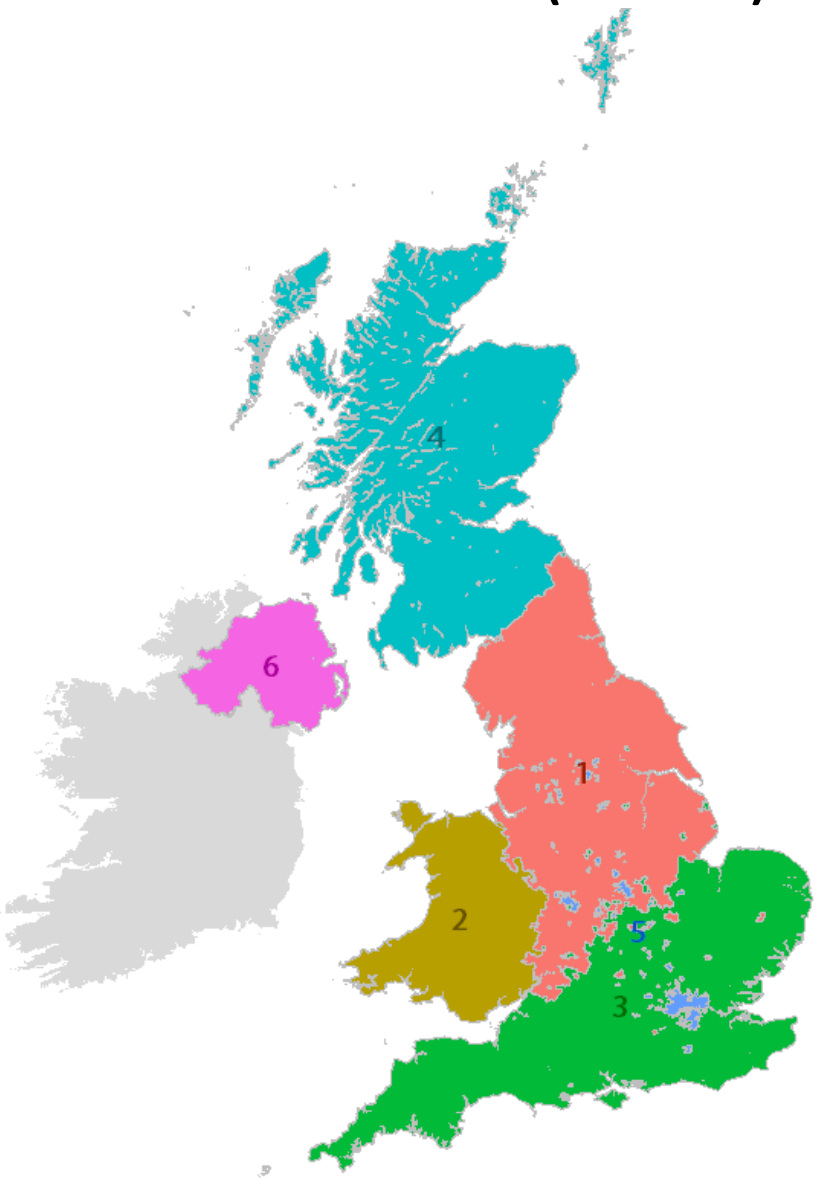
REGIONALISATION

based on jurisdictions



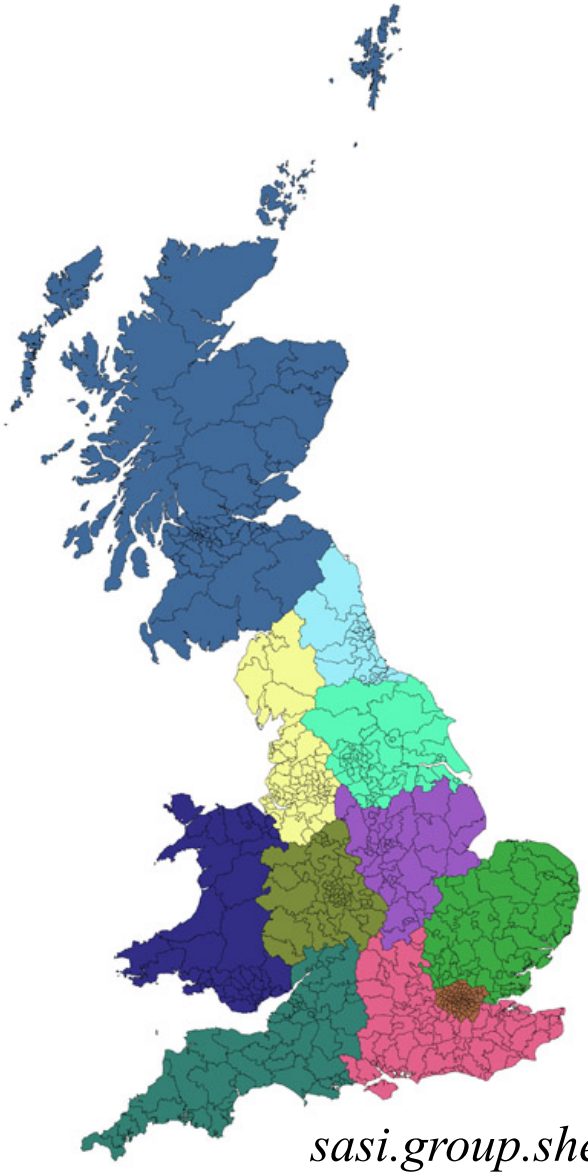
sasi.group.shef.ac.uk

based on surnames (CR 2011)



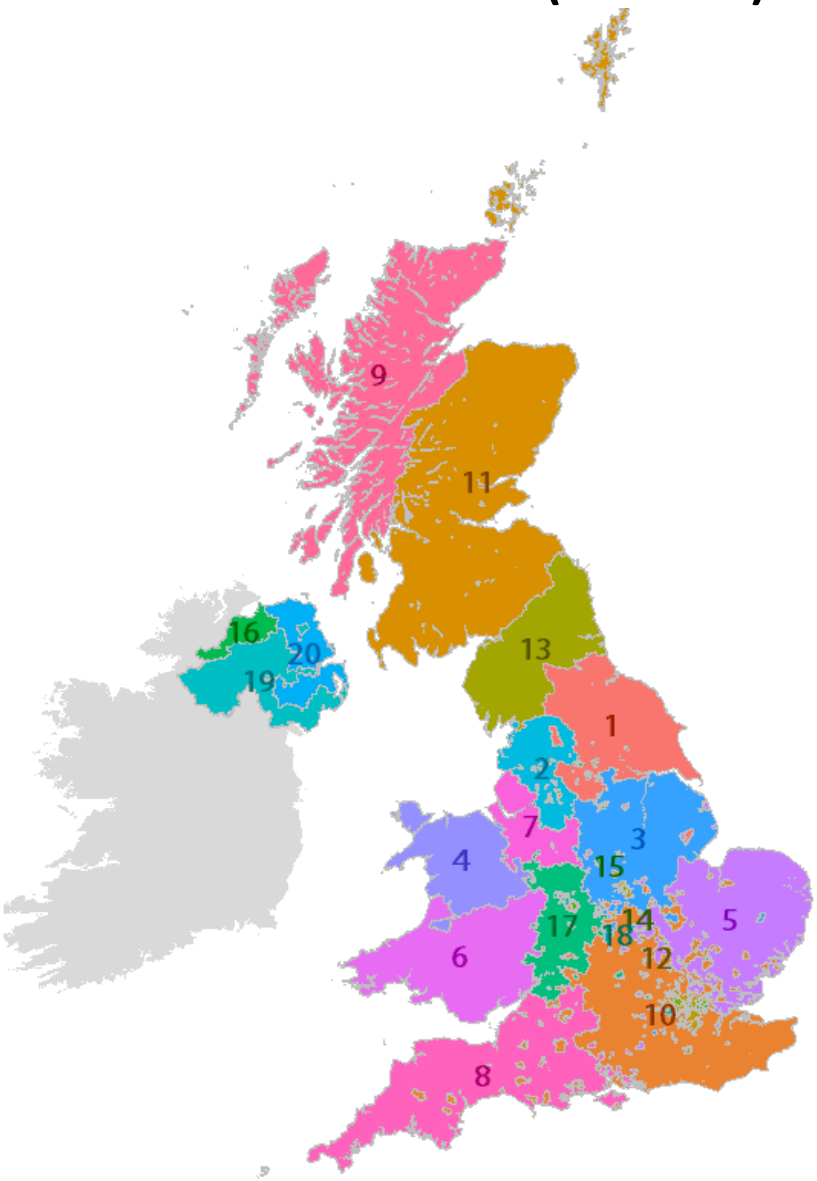
REGIONALISATION

based on jurisdictions



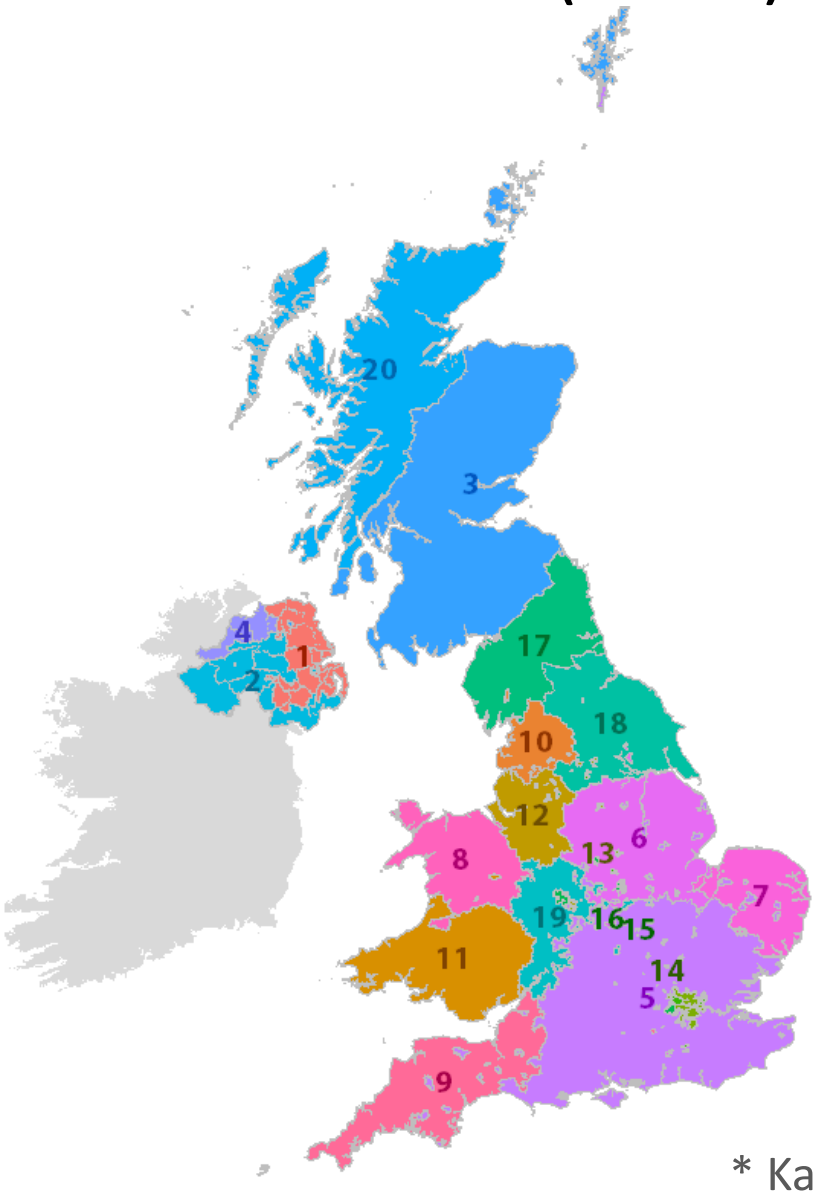
sasi.group.shef.ac.uk

based on surnames (CR 2011)

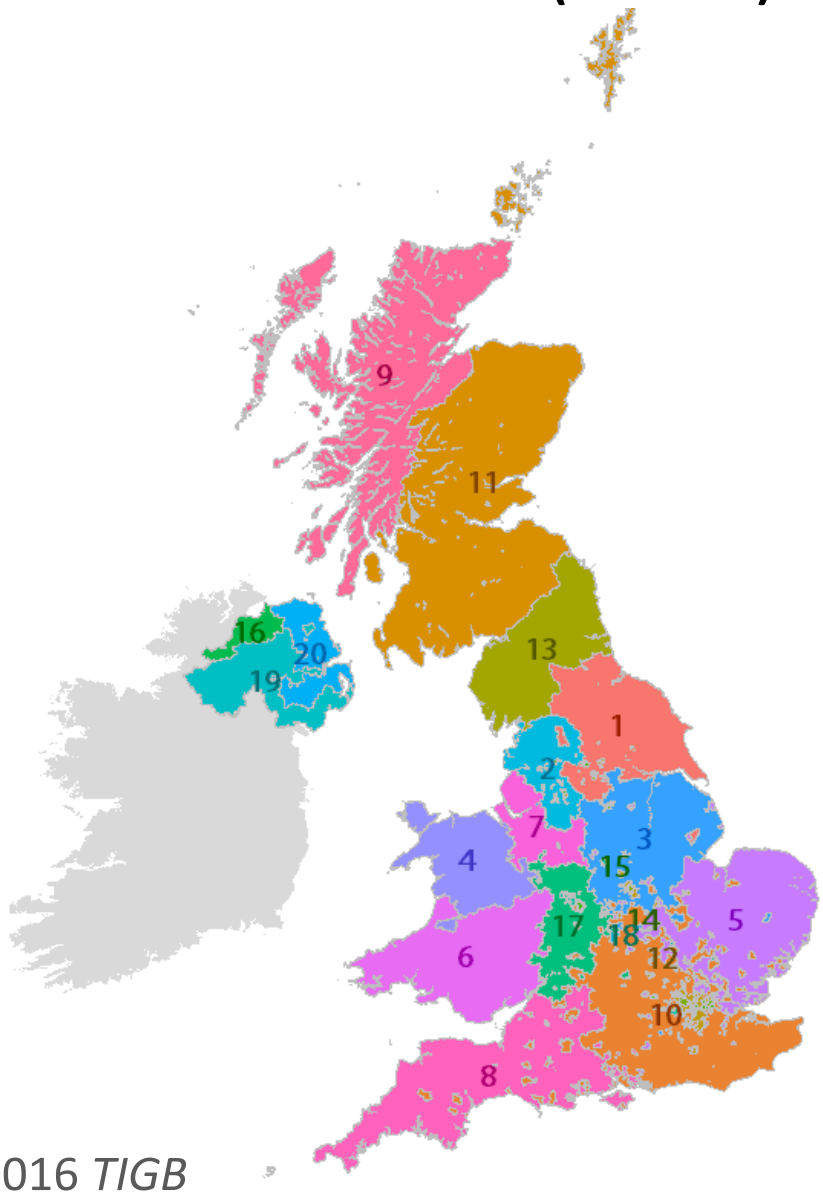


‘TEMPORAL’ REGIONAL INTEGRITY*

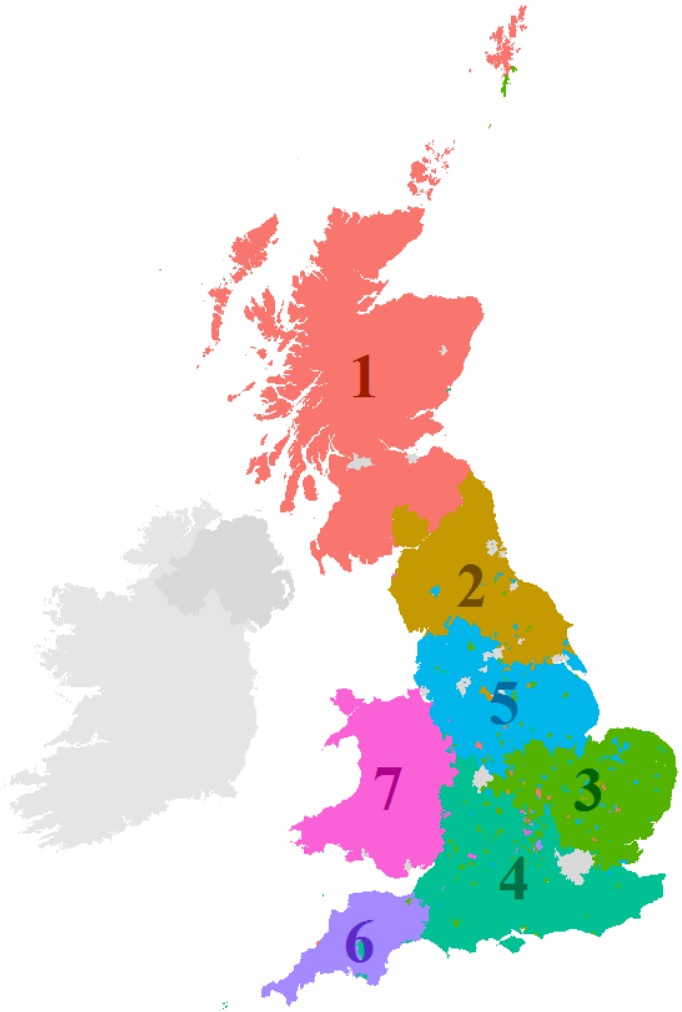
based on surnames (CR 2007)

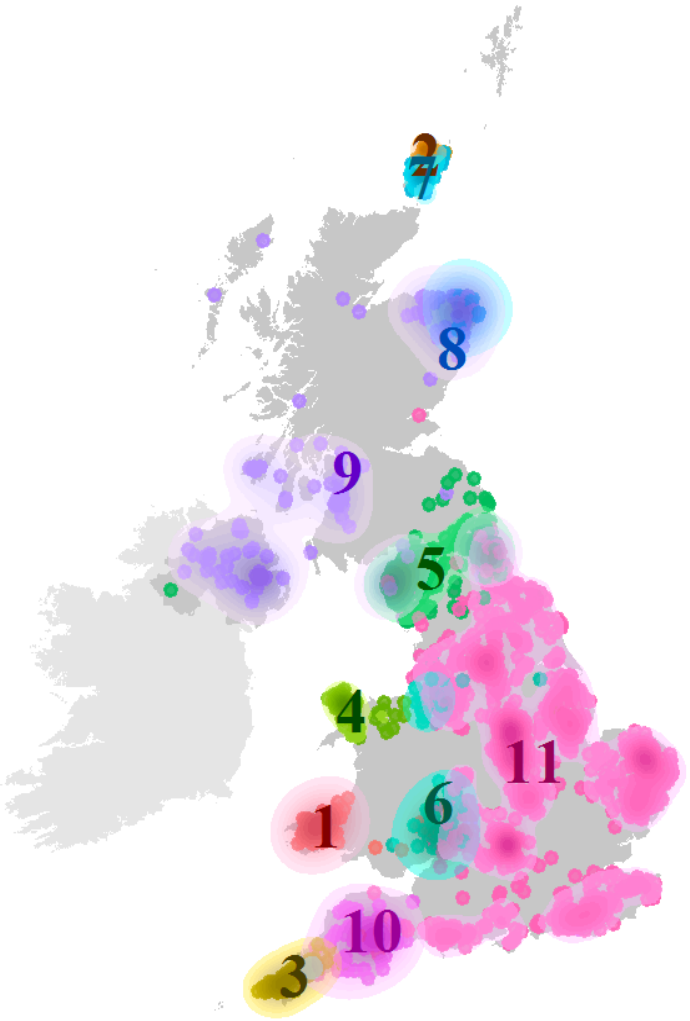


based on surnames (CR 2011)



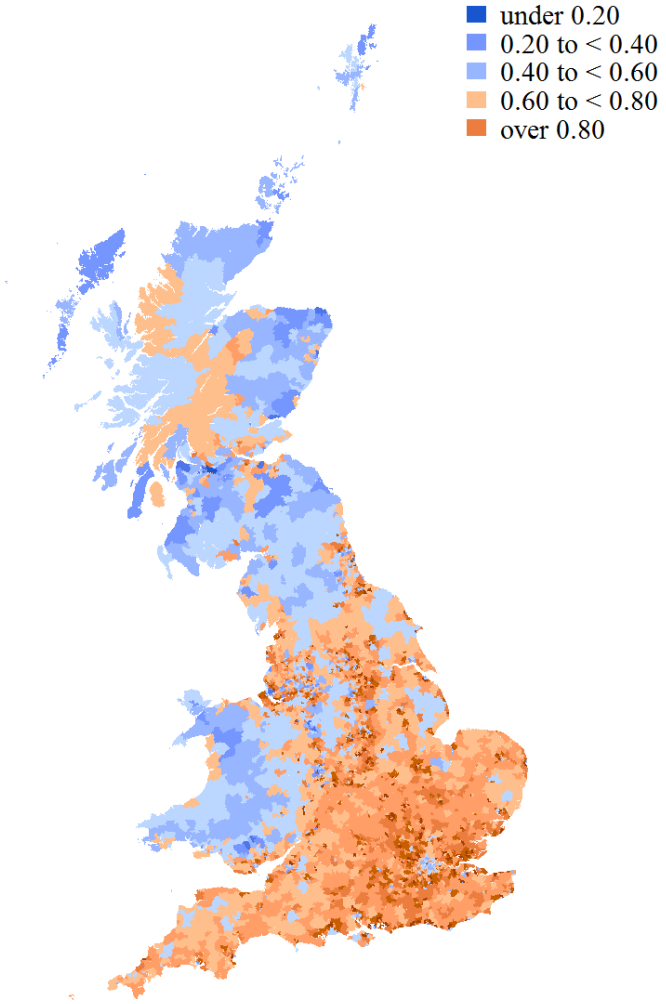
* Kandt et al 2016 *TIGB*


1
2
5
7
3
4
6

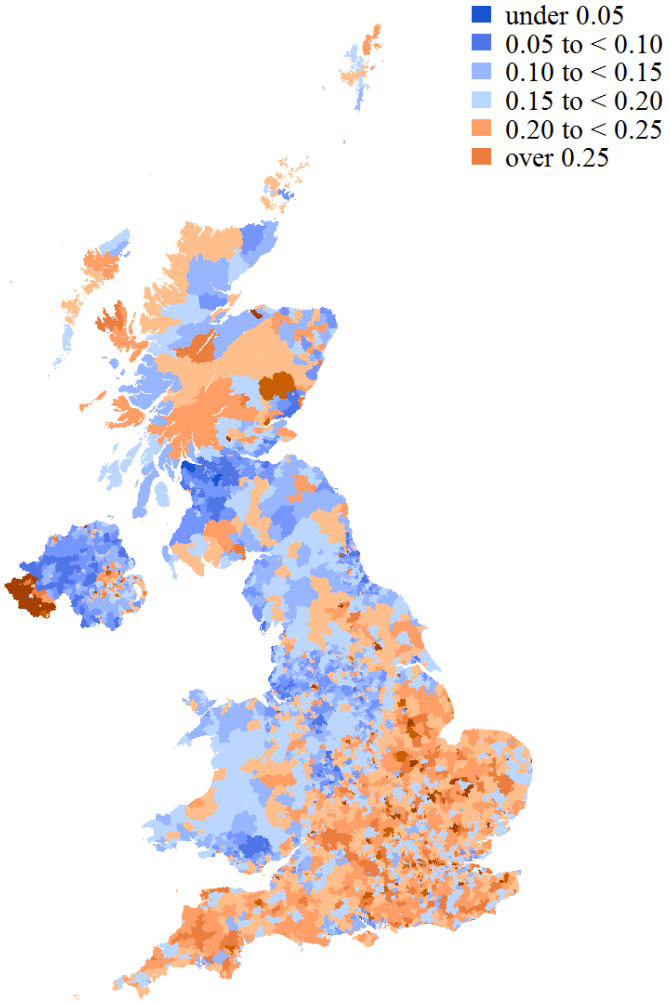

7
8
9
5
4
11
1
6
10
3

PROPORTION OF NEWLY IMPORTED SURNAMES

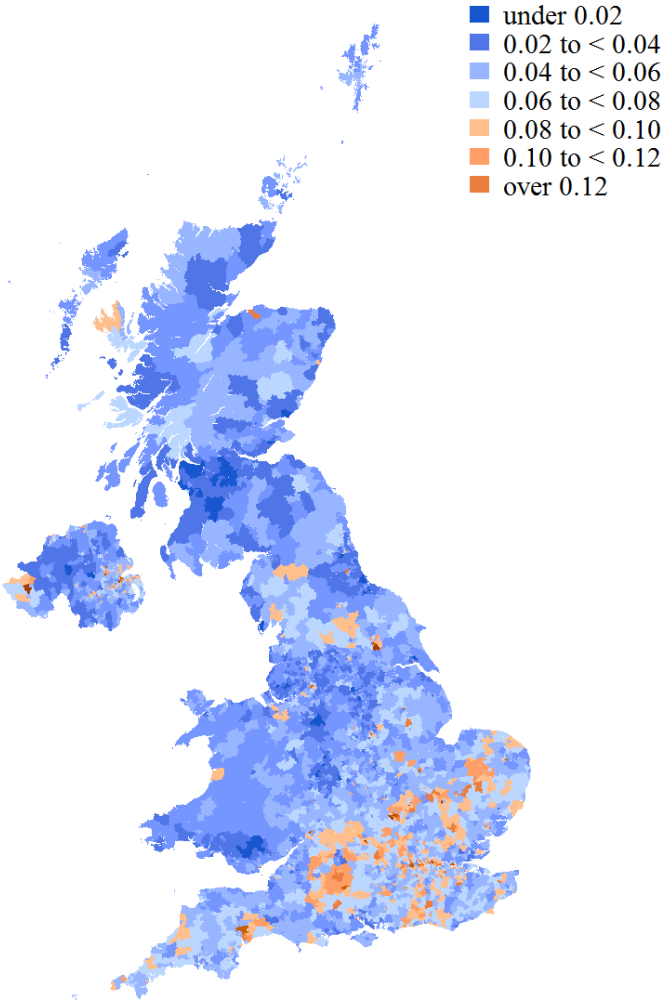
1881-2007

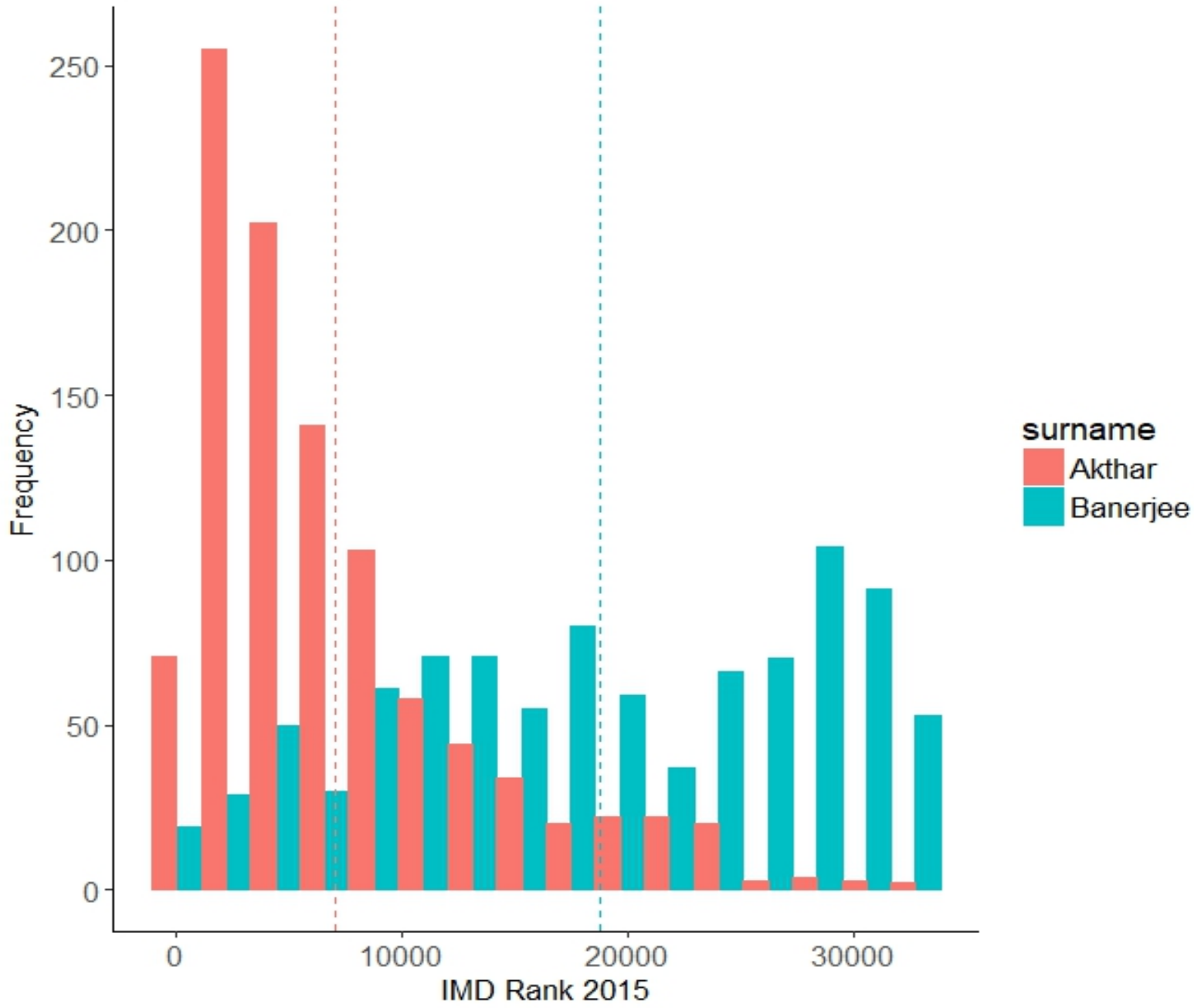


2007-2013



2013-2016



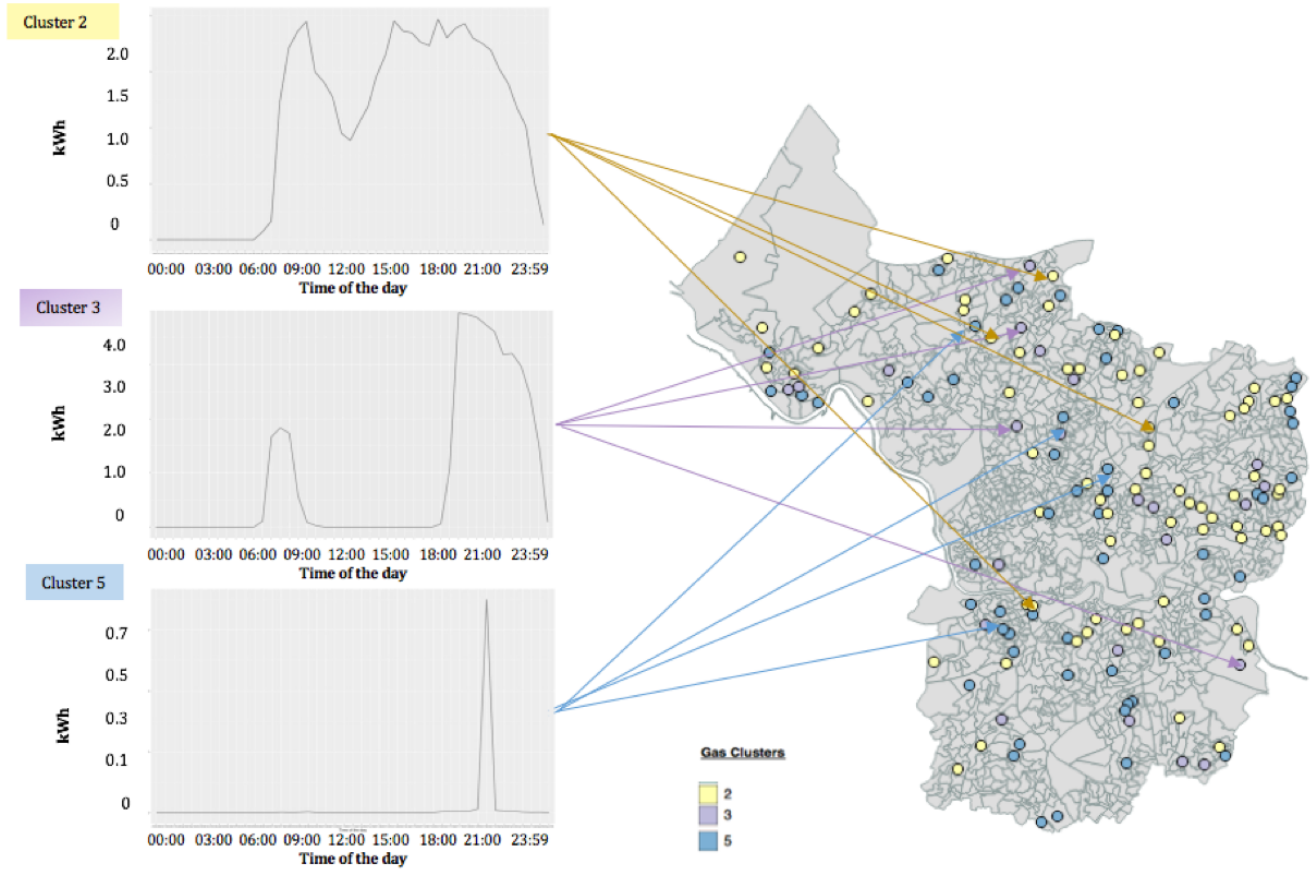

250
200
150
100
50
0
Frequency
0
10000
20000
30000
IMD Rank 2015
surname
Akthar
Banerjee

Consumer Data Research Centre

An ESRC Data Investment

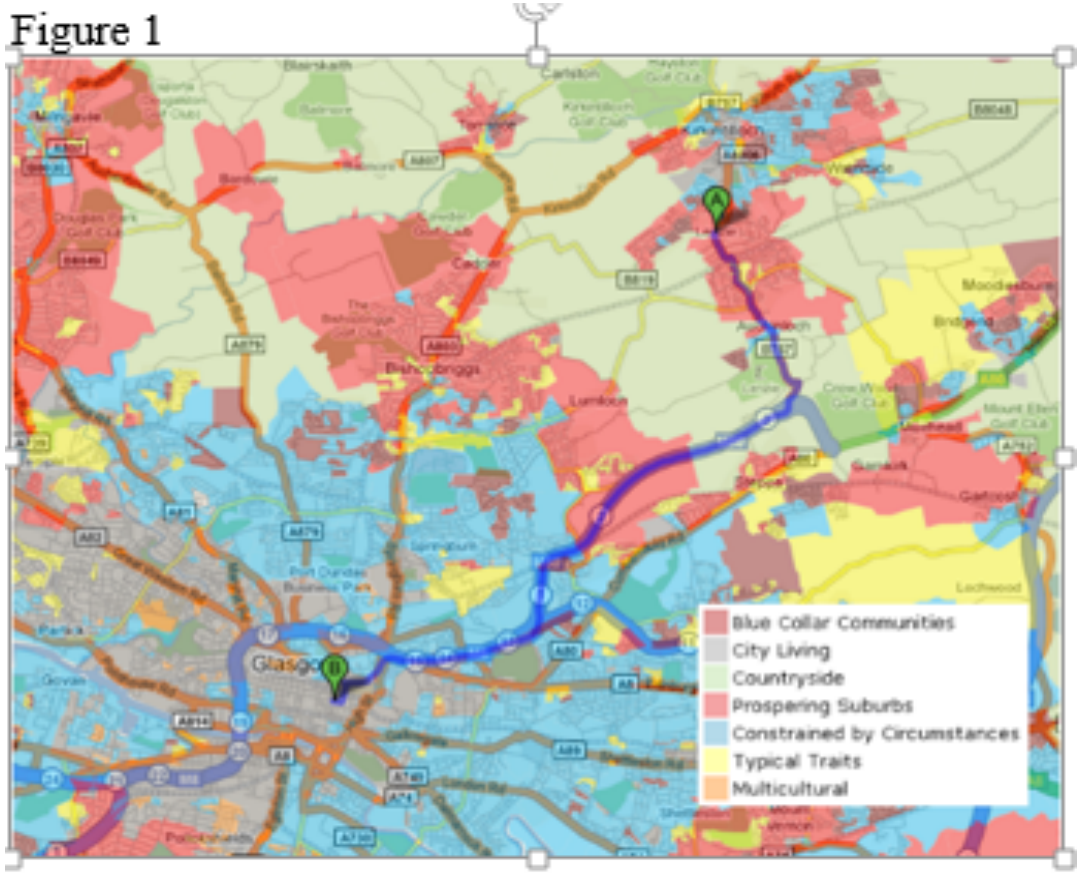
Profiling Customers

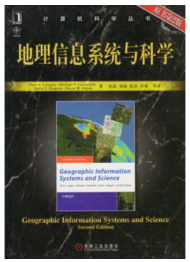
City of Bristol

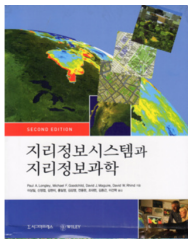


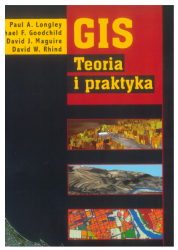
Regents Park
Euston Square
A501
Gower Pl
Gordon St
Tavistock St
Warren Street
Grafton Way
University St
Whitfield St
Huntley St
A400

Figure 1

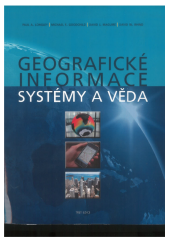

Blue Collar Communities
City Living
Countryside
Prospering Suburbs
Constrained by Circumstances
Typical Traits
Multicultural

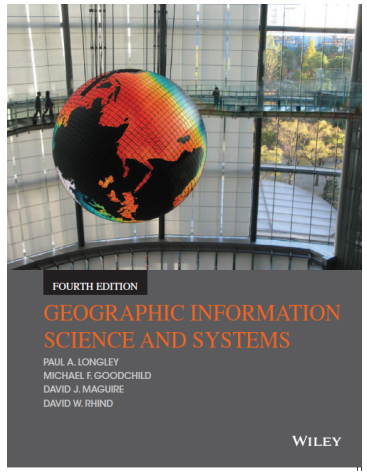

地理信息系统与科学
Geographic Information Systems and Science

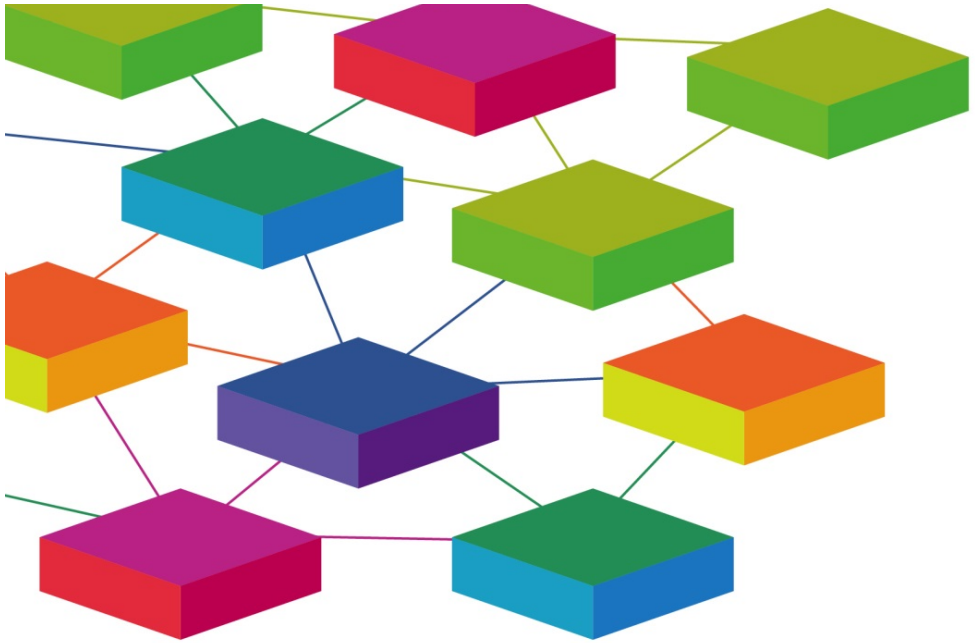

지리정보시스템과
지리정보과학


GIS
Teoria i praktyka

SISTEMAS E CIÊNCIA DA
INFORMAÇÃO
GEOGRÁFICA


GEOGRAFICKÉ
INFORMACE
SYSTÉMY A VĚDA


FOURTH EDITION
GEOGRAPHIC INFORMATION SCIENCE AND SYSTEMS
PAUL A. LONGLEY
MICHAEL F. GOODCHILD
DAVID J. MAGUIRE
DAVID W. RHIND
WILEY



www.cdrc.ac.uk