

RGS - 21th October 2013

**The Census and future provision of
population statistics in England and Wales
– Public consultation**

Alistair Calder & Andy Teague
Beyond 2011

The Beyond 2011 Programme

- Census – every 10 years for over 200 years
- Review a normal part of the census cycle but the need greater than ever

Rapidly changing society

Evolving user requirements

Technological advances

Improved data sources

DRIVERS : Cost, **efficiency**, opportunity, burden

CRITERIA : Cost v social and economic **benefit**

of outputs, privacy, public acceptability, risk etc

- Government proposes to Parliament (with NS advice)
- Beyond 2011 findings will be published in 2014

Beyond 2011 : Full range of approaches assessed

Census options

Full Census (long form to everyone)

Rolling Census (over 5/10 year period)

Short Form (everyone), Long form (Sample)

Short Form + Annual Survey (US model)

Administrative data options

Aggregate analysis

Partial (1%?) linkage

Full (100%) linkage

Survey option(s)

Address register + Survey

Beyond 2011 : Two potential approaches

A census once a decade
-like that conducted in 2011,
but **primarily online**

A census based on
administrative data
and large annual surveys



The two approaches

What they are

Beyond 2011 : An online census – what it is

A census once a decade
-like that conducted in 2011,
but **primarily online**

A compulsory questionnaire for every household (and communal)

Majority of responses online (and mix of ways to complete)

1% survey to adjust for those who don't respond

Administrative data to check the quality

Population estimates produced annually using births, deaths, etc

Questions and topics similar to 2011 Census – but will consult later

Beyond 2011 : Administrative data and surveys – what it is

NHS Patient Register
DWP/HMRC Customer Information System
Electoral roll (> 17 yrs)
School Census (5-15 yrs)
Higher Education Statistics Agency data (Students)
Birth and Death registrations

**NO PERSONAL DATA HELD – ALL NAMES &
ADDRESSES AND DATES OF BIRTH
ANONYMISED**

**A census based on
administrative data
and large annual surveys**

Re-use of admin sources to produce annual population estimates
Anonymous data from eg NHS, DWP, HMRC, DfE, HESA
Annual compulsory 1% survey to adjust for error in the admin sources
Annual compulsory 4% survey to collect characteristics information
Majority of responses online (and mix of ways to complete)
Questions and topics similar to 2011 Census – but will consult later

Beyond 2011 : An online census – what you get

A census once a decade
-like that conducted in 2011,
but **primarily online**

Huge richness of data

Data for very small areas and very small populations

Detailed cross tabulations – nearly 6 billion cells

Continuity – tradition

A benchmark – a definitive snapshot of the nation - certainty

Data that is (or might be) out of date most of the time

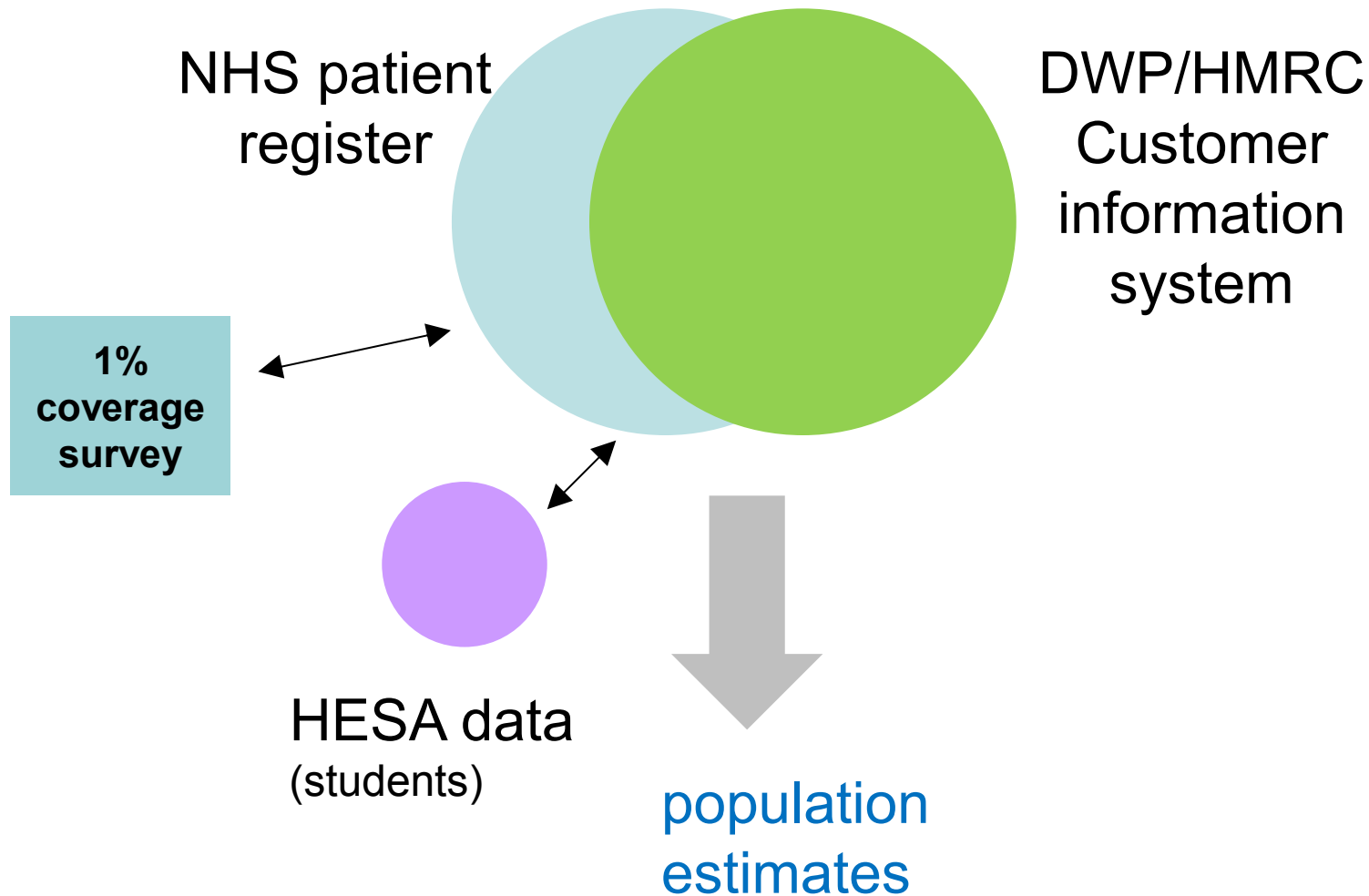
An illusion of knowledge (some of the time) but the best we've got

Beyond 2011 : The admin data approach – what you get

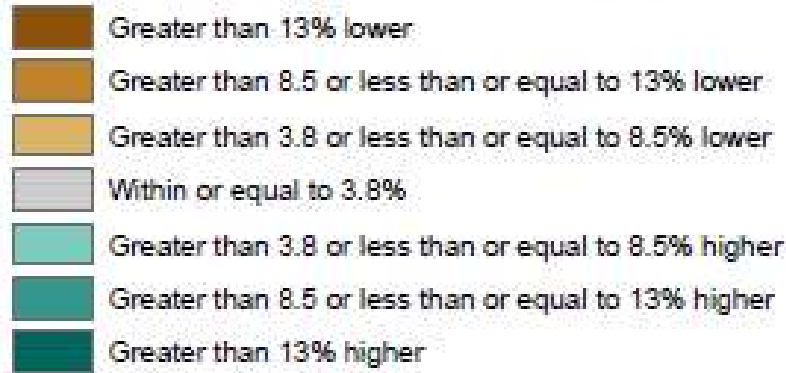
**A census based on
administrative data
and large annual surveys**

**Administrative data and survey approach
population estimates**

Administrative data and survey option



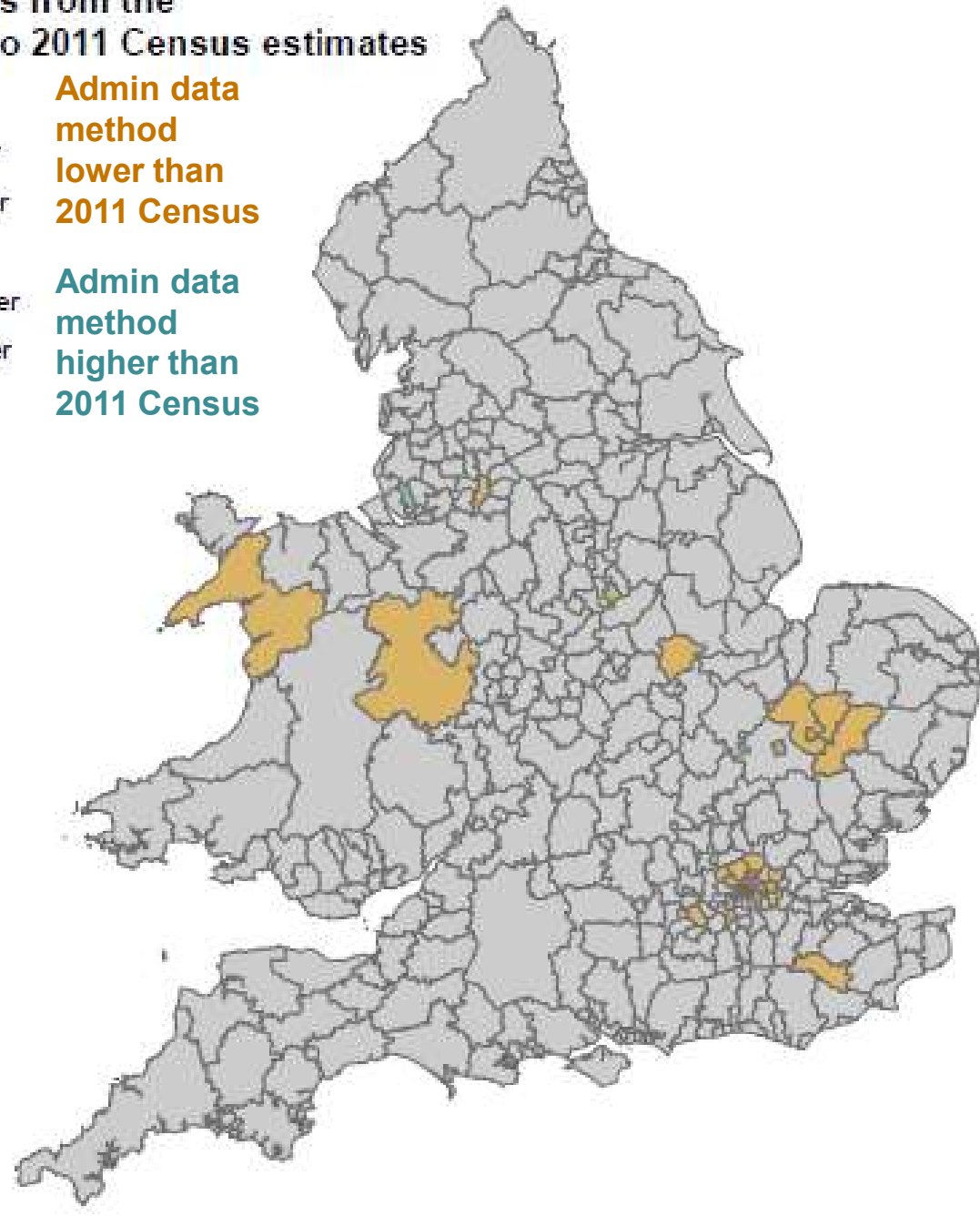
Percentage difference between LA totals from the administrative data method compared to 2011 Census estimates



Admin data method lower than 2011 Census

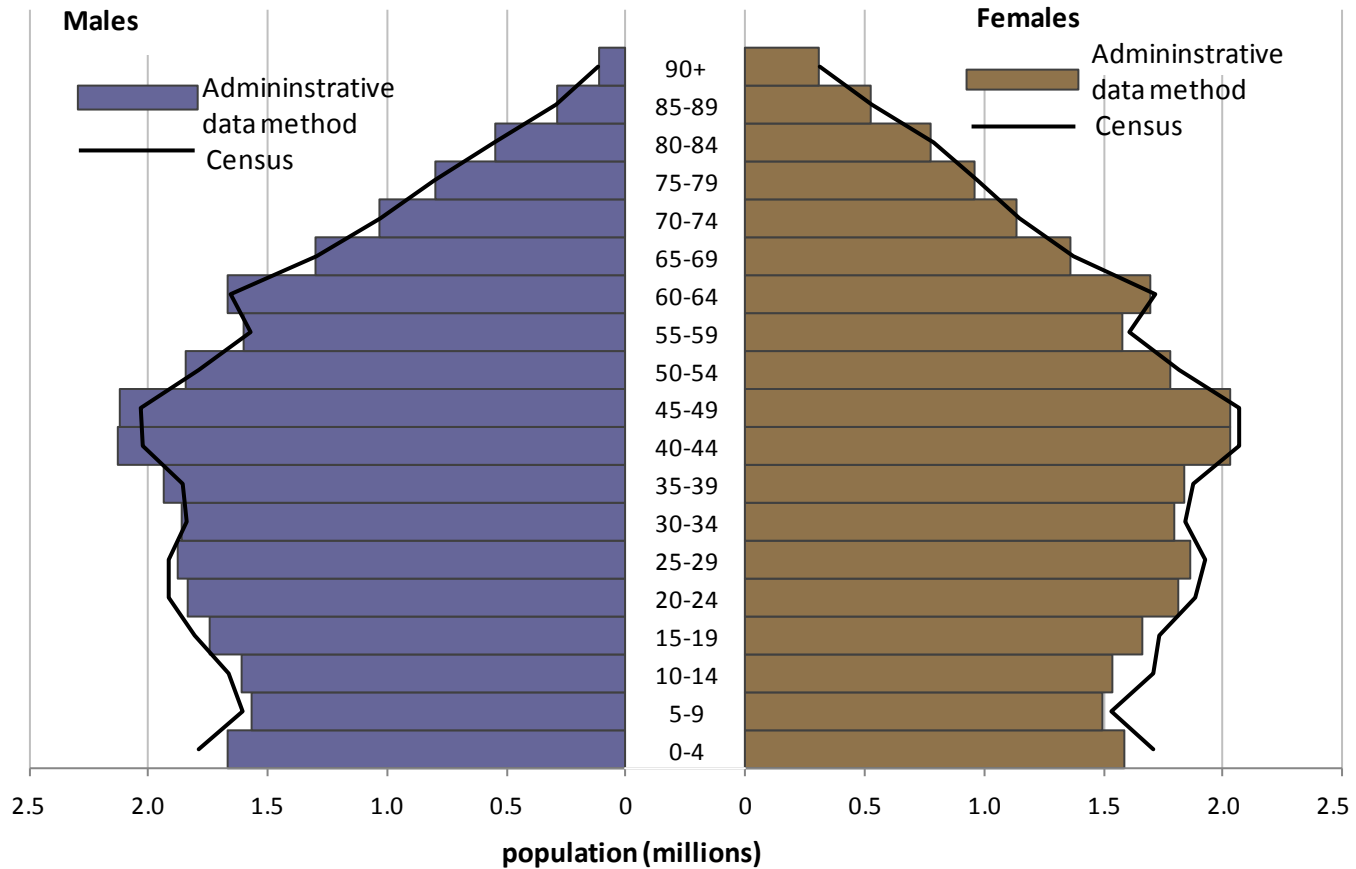
Admin data method higher than 2011 Census

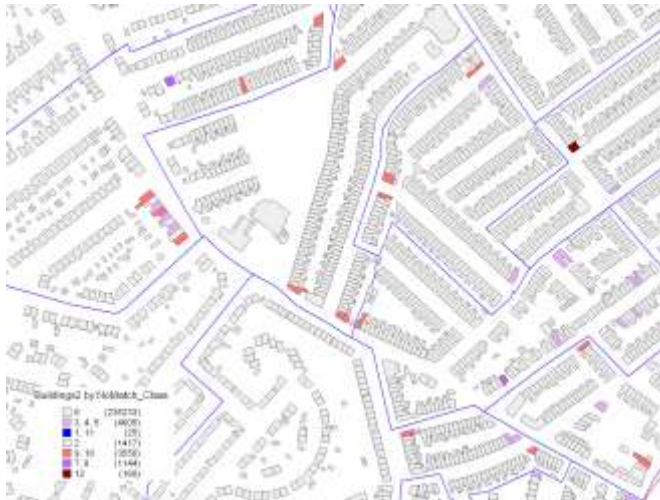
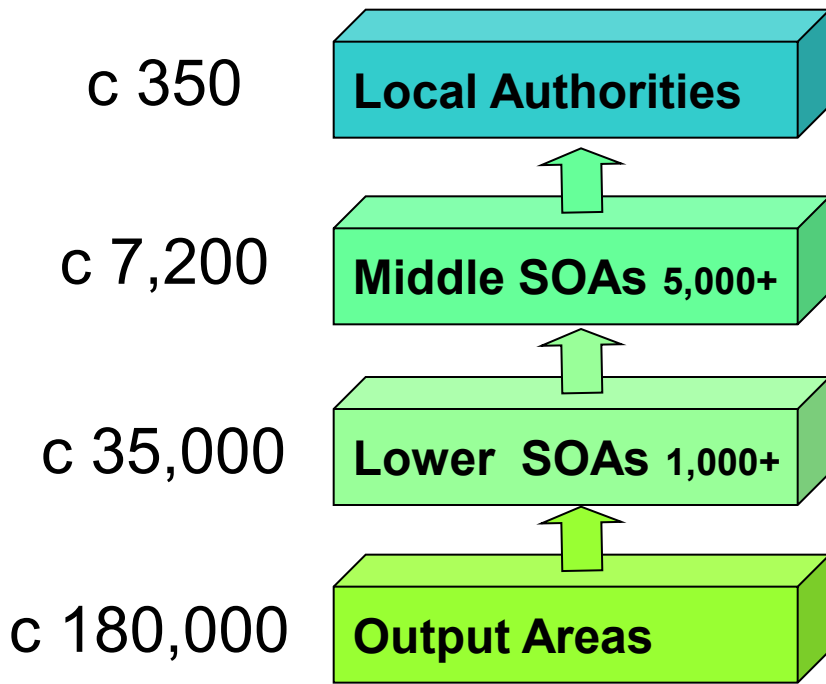
Greater London



Population Pyramids using admin data

Administrative data method population pyramid with Census
comparison: England & Wales





Super Output Areas (SOAs)

Background

Super Output Areas were designed to improve the reporting of small area statistics and are built up from groups of Output Areas. Statistics for Lower Layer Super Output Areas and Middle Layer Super Output Areas were originally released in 2004 for England and Wales. Scotland also released statistics for Data Zones (equivalent to LSOAs) in 2004 and Intermediate Geographies (equivalent to MSOAs) in 2005. Northern Ireland introduced Lower Layer SOAs in 2005 but do not have a Middle Layer SOA geography.

2011 Super Output Areas

Maintaining stability as far as possible was key for the 2011 Census. LSOAs and MSOAs created following the 2001 Census continue to exist unless a significant population change occurred between 2001 and 2011, and household minimum and maximum thresholds were breached. Simplistically, where populations have become too big, the LSOA/MSOA has been split into two or more areas; where populations have become too small the LSOA/MSOA has been merged with an adjacent one. Responses to the Output Geography Consultation from December 2009 to March 2010 were also considered in the redesign of OAs and SOAs. However, the total changes across the output area hierarchy were no more than five per cent overall.

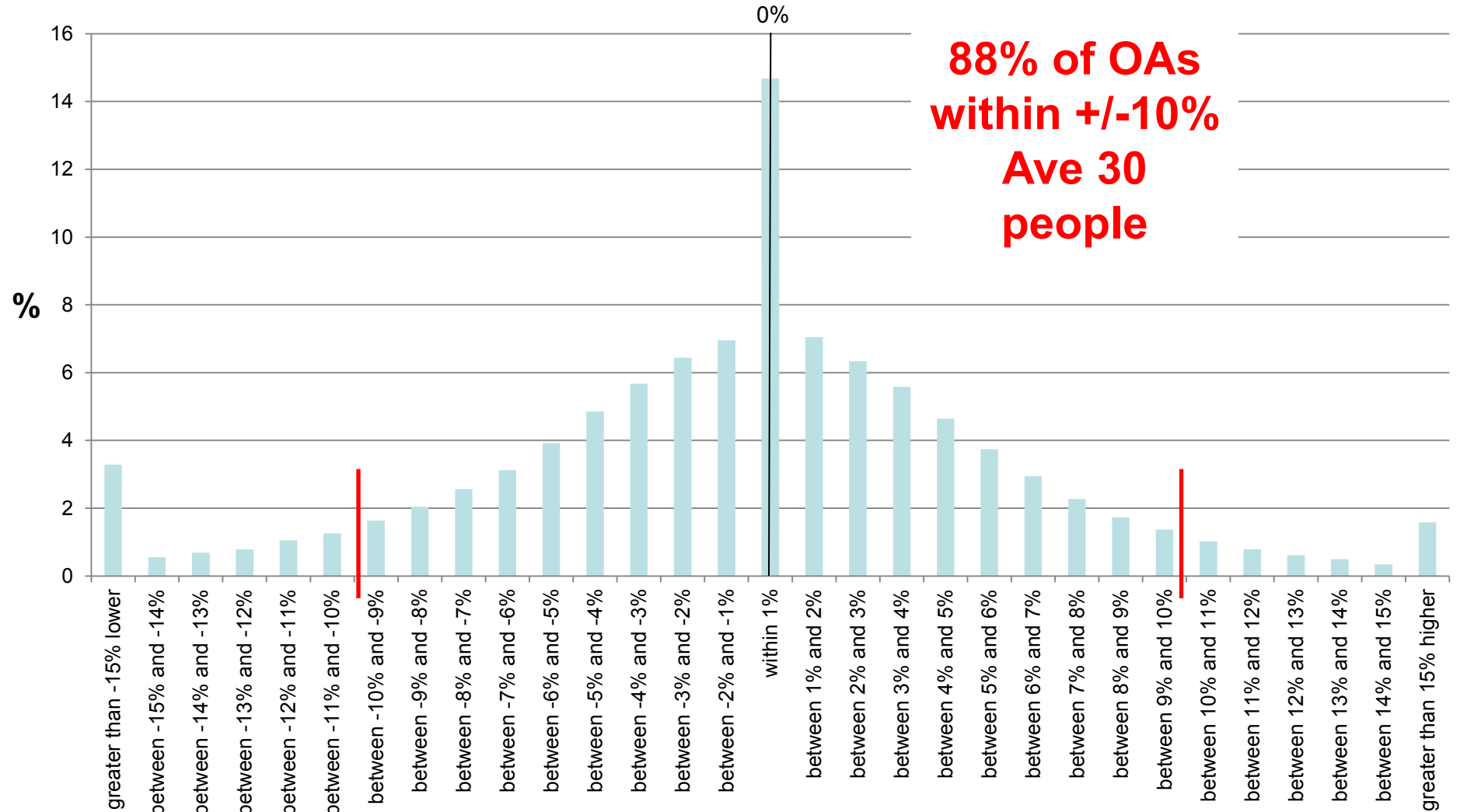
Population and household minimum and maximum thresholds for SOAs in England and Wales:

Geography	Minimum population	Maximum population	Minimum number of households	Maximum number of households
LSOAs	1,000	3,000	400	1,200
MSOAs	5,000	15,000	2,000	6,000

The total of 2011 LSOAs and MSOAs for England and Wales:

Geography	England	Wales
LSOAs	32,844	1,909
MSOAs	6,791	410

Percentage difference between administrative data approach and Census estimates – Output Area level



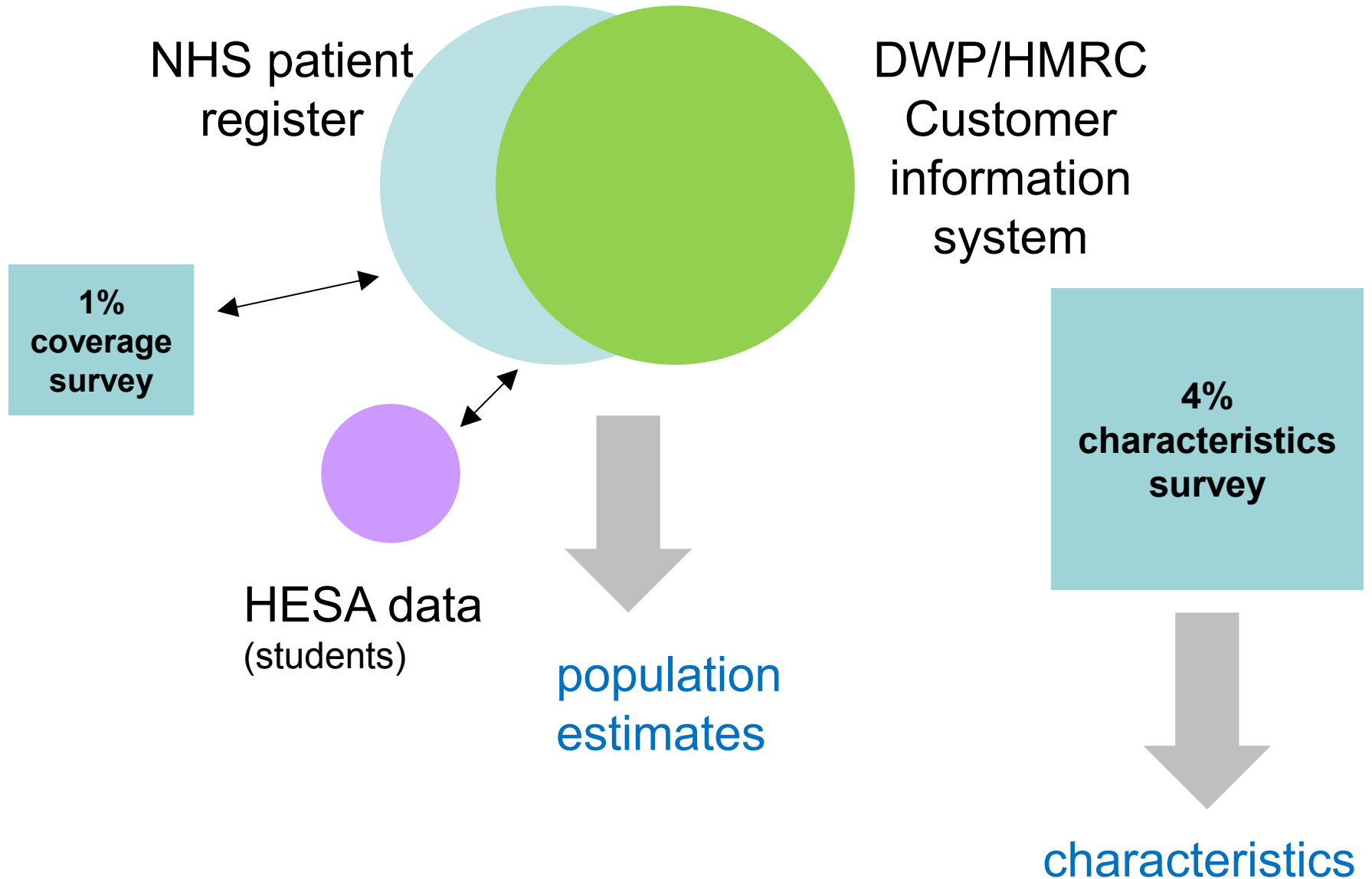
**88% of OAs
within +/-10%
Ave 30
people**

Note that central bar covers double the range of the other bars

What you get – population estimates

- Annual population estimates – for all geographic levels – down to output areas
- Annual estimates age & sex – for all levels down to LSOA
 - OA level currently unproven but all the signs are that this will be possible
- Research is ongoing – hope to be able to say more in November

Administrative data and survey option



Beyond 2011 : The admin data & survey approach – what you get

**A census based on
administrative data
and large annual surveys**

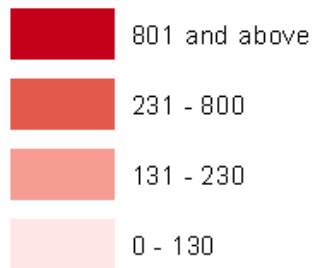
**Administrative data and survey approach
characteristics**

How survey works

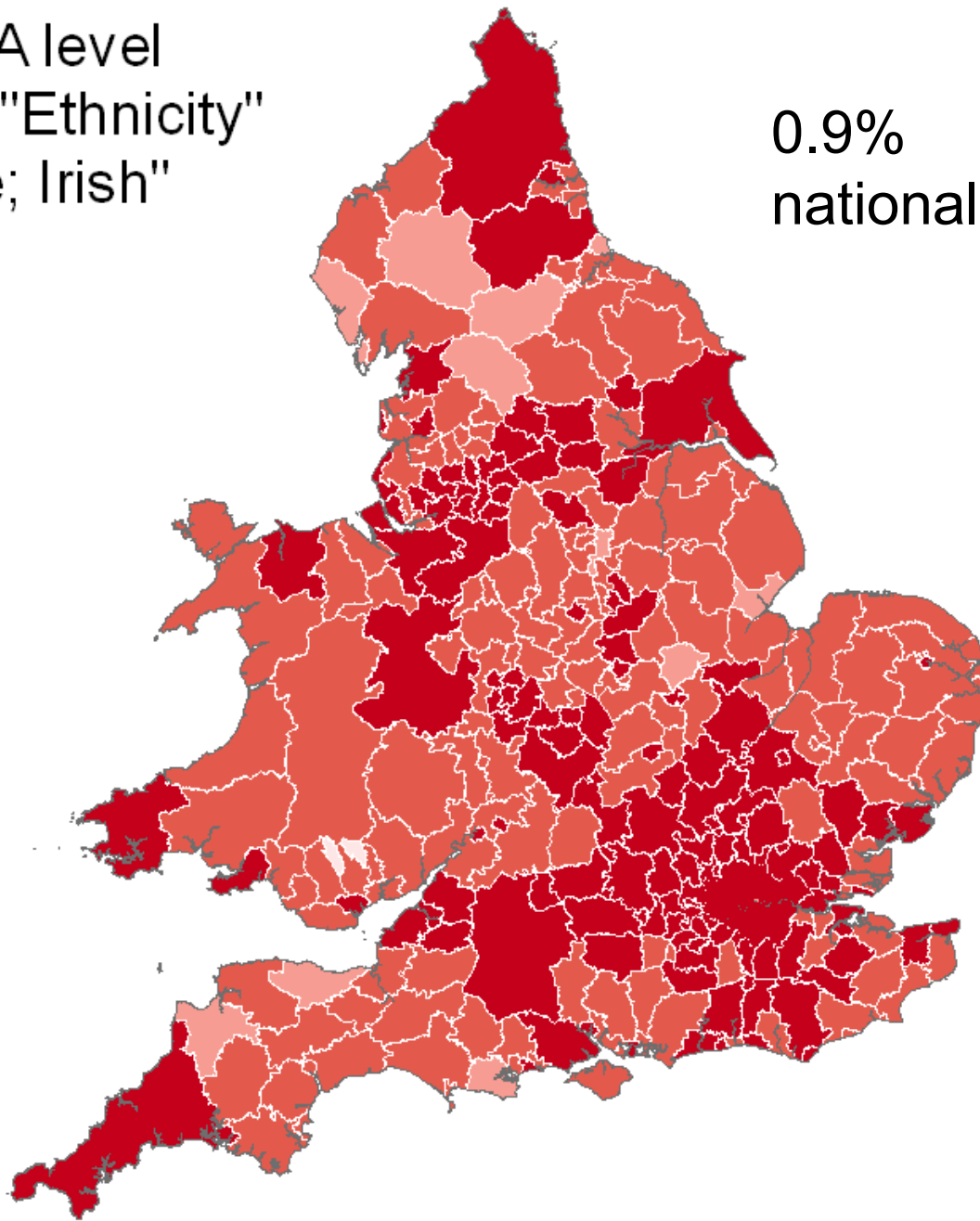
- Most characteristics not currently covered by an administrative source
- Need **compulsory 4% survey**
- Reliable statistics could be produced for characteristics representing:
 - **800** or more people using 1 year's data (4%)
 - **230** or more people using 3 years' data (12%)
 - **130** or more people using 5 years' data (20%)
- Statistics for smaller populations would be produced but lower accuracy (CI > 40%)

2011 Census LA level populations for the "Ethnicity" category "White; Irish"

0.9%
nationally

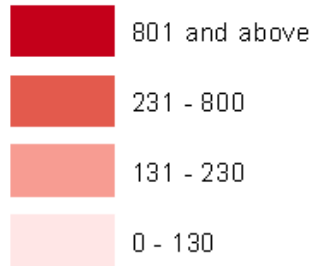


Greater London

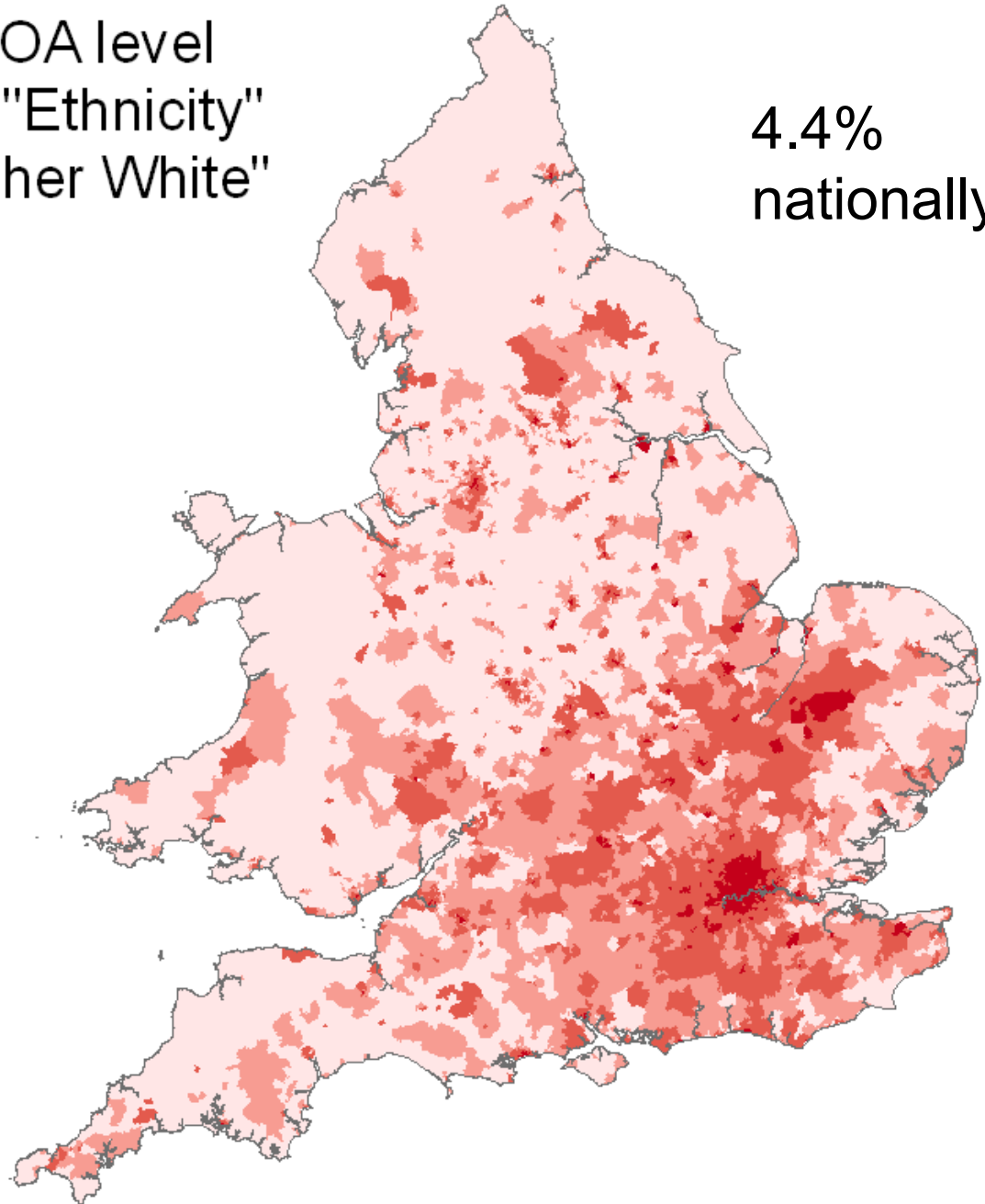
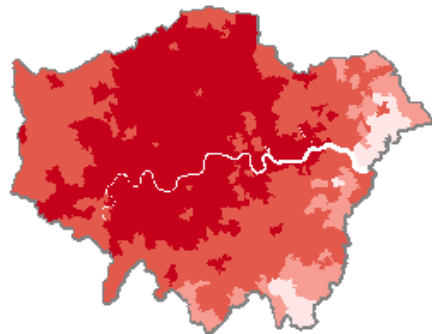


2011 Census MSOA level populations for the "Ethnicity" category "White; Other White"

4.4% nationally



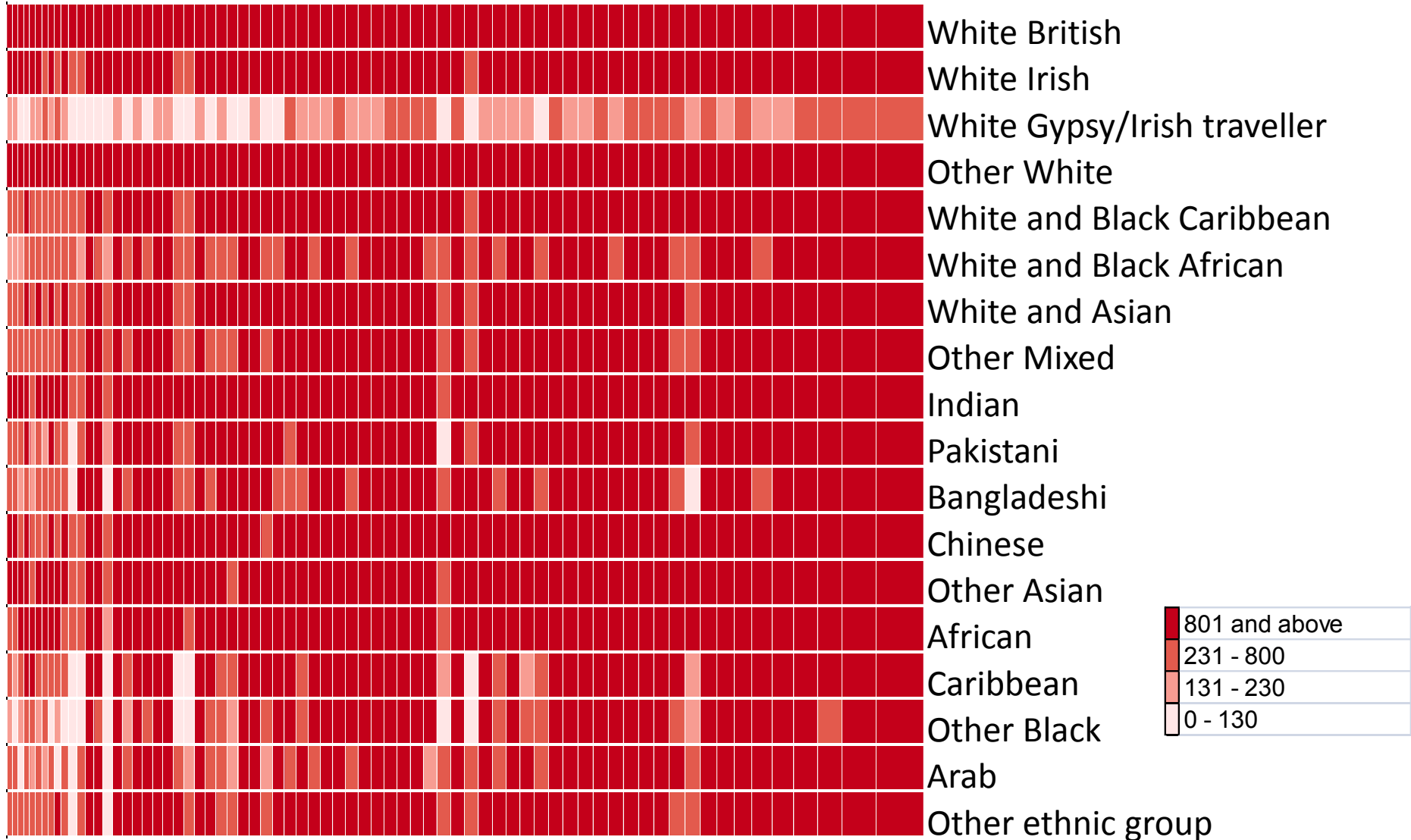
Greater London



2011 Census LA level populations (71 Major urban) for Ethnicity

Smallest

Largest



Simulated survey LA level population estimates (71 Major Urban) for Ethnicity

Smallest

One year sample

Largest



Simulated survey LA level population estimates (71 Major Urban) for Ethnicity

Smallest

Three year sample

Largest

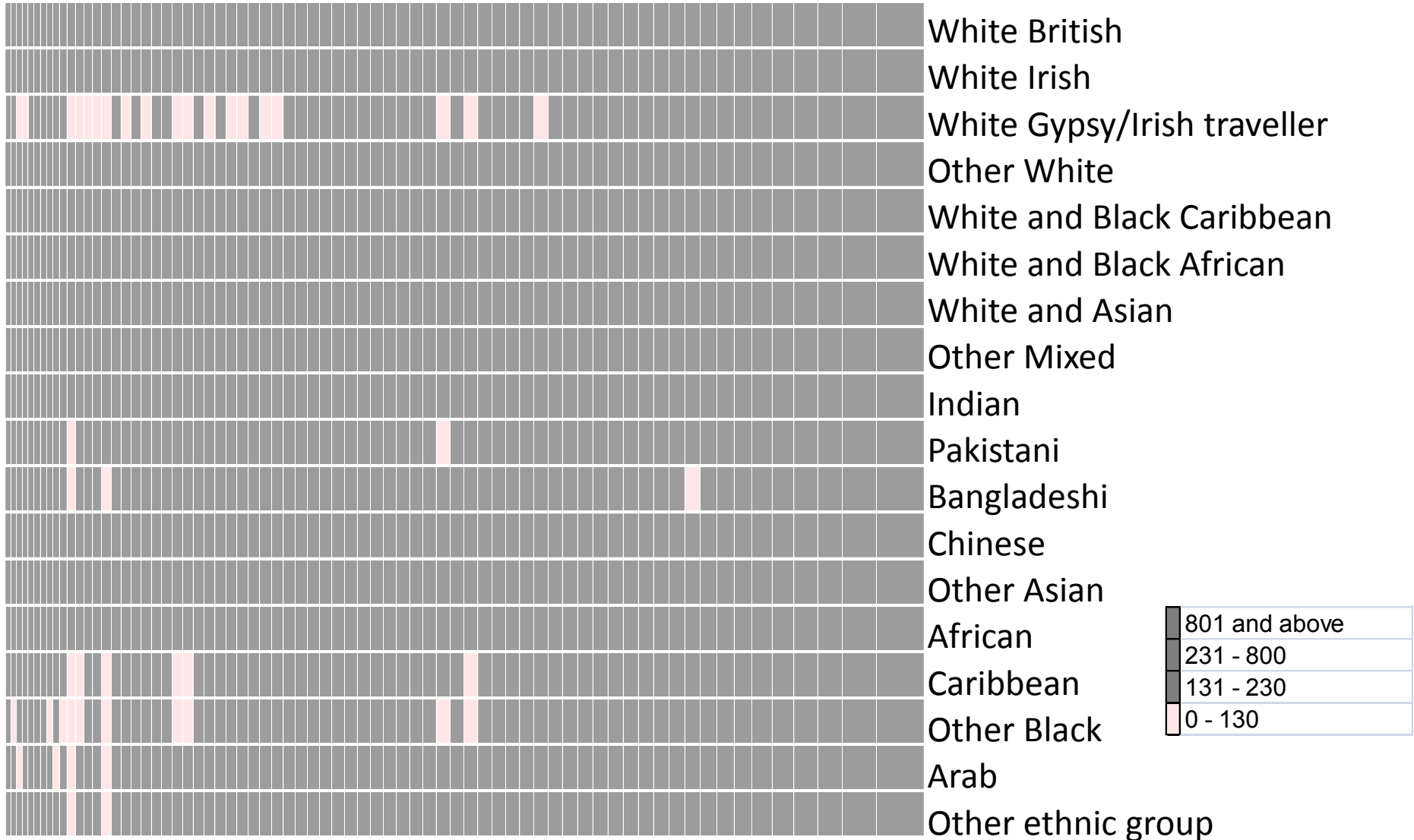


Simulated survey LA level population estimates (71 Major Urban) for Ethnicity

Smallest

Five year sample

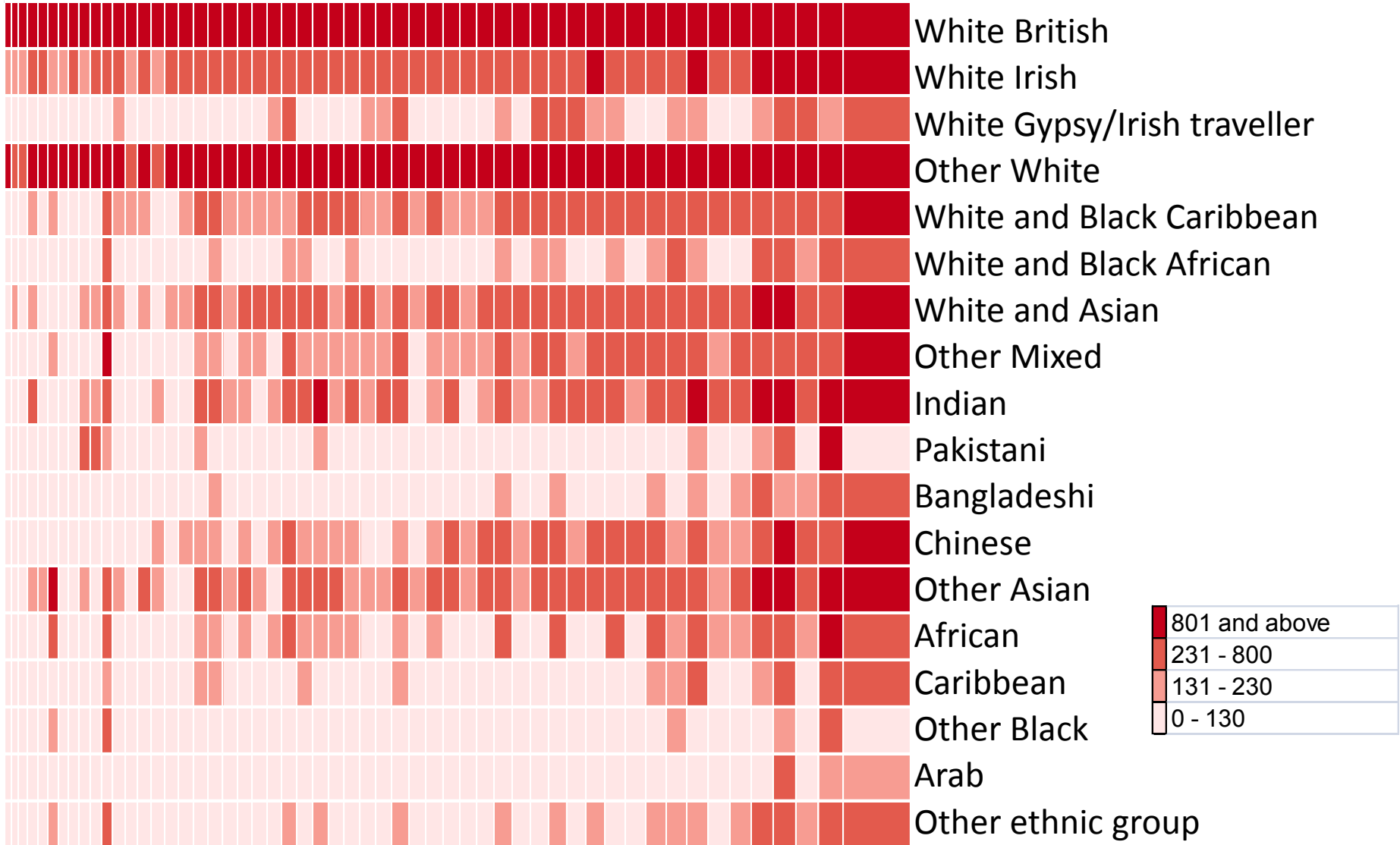
Largest



2011 Census LA level populations (55 Most rural) for Ethnicity

Smallest

Largest

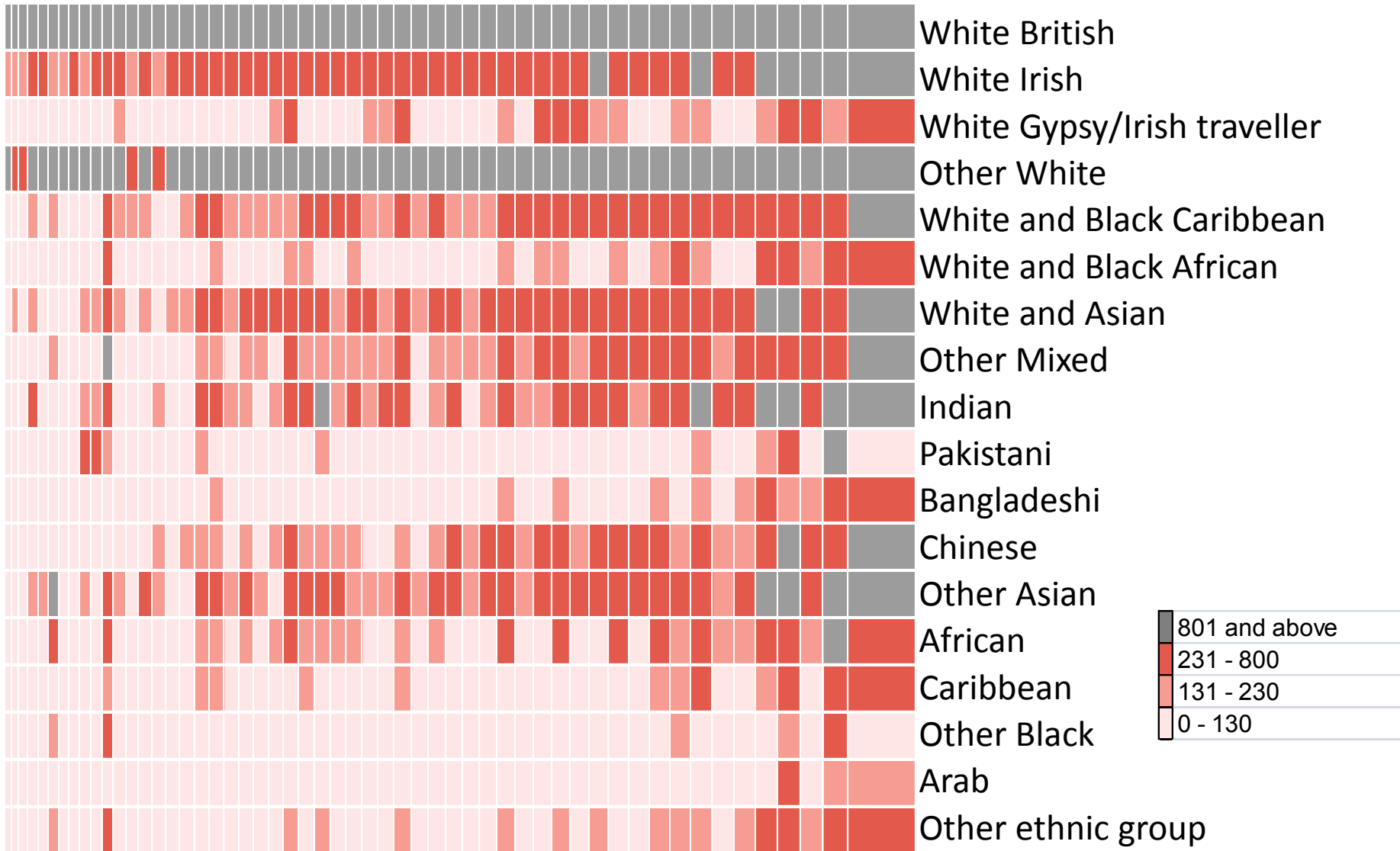


Simulated survey LA level population estimates (55 Most rural) for Ethnicity

Smallest

One year sample

Largest

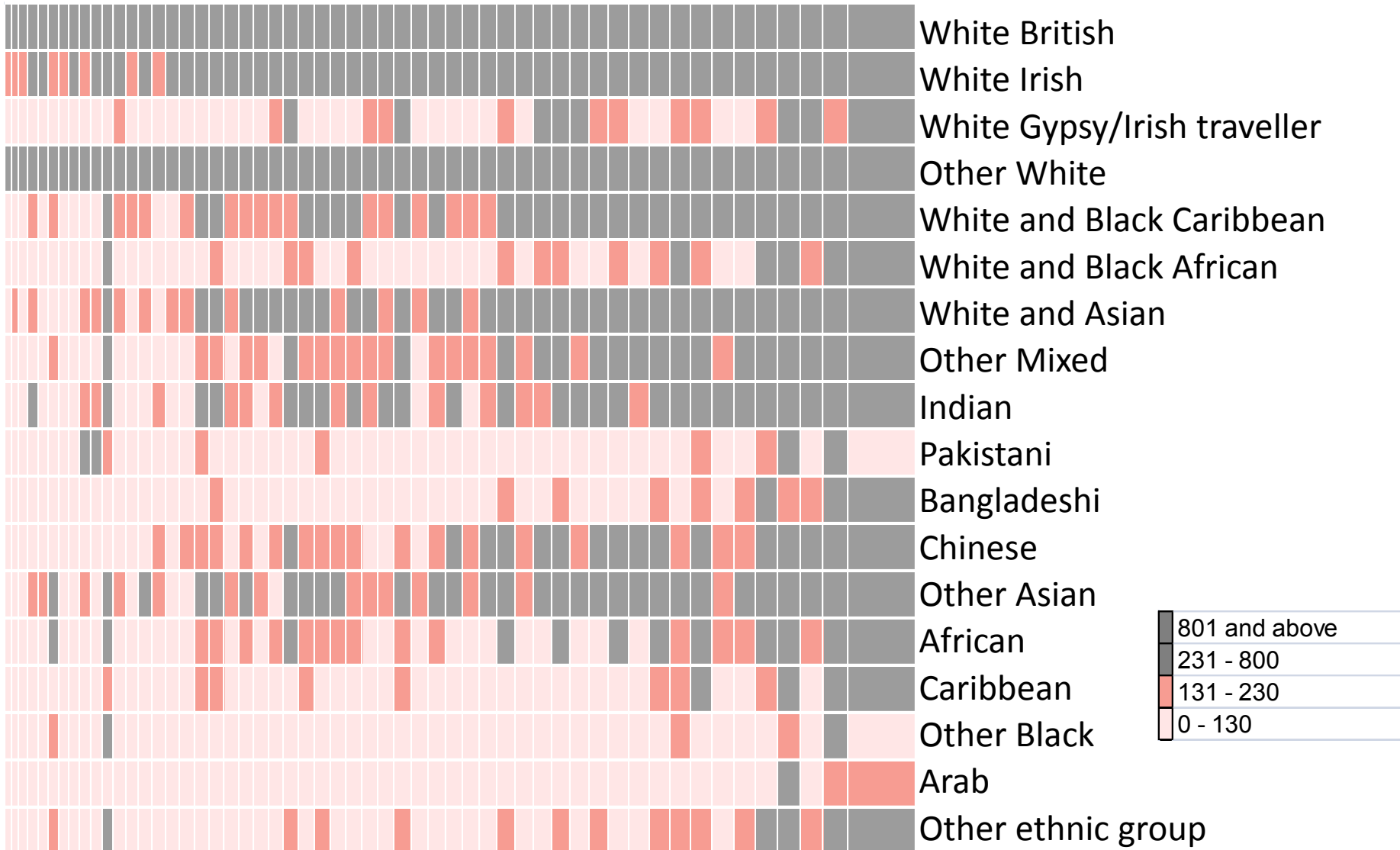


Simulated survey LA level population estimates (55 Most rural) for Ethnicity

Smallest

Three year sample

Largest

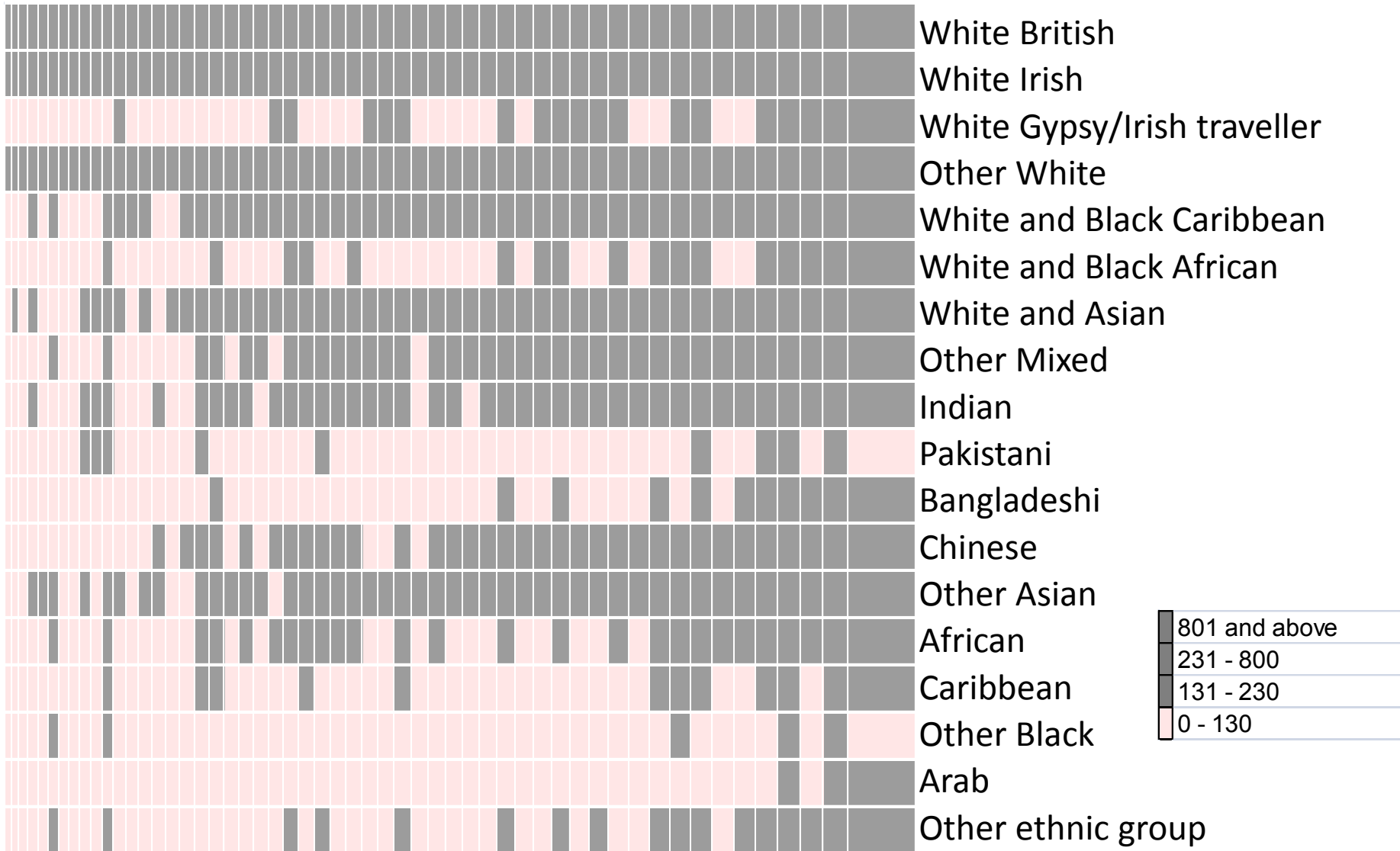


Simulated survey LA level population estimates (55 Most rural) for Ethnicity

Smallest

Five year sample

Largest



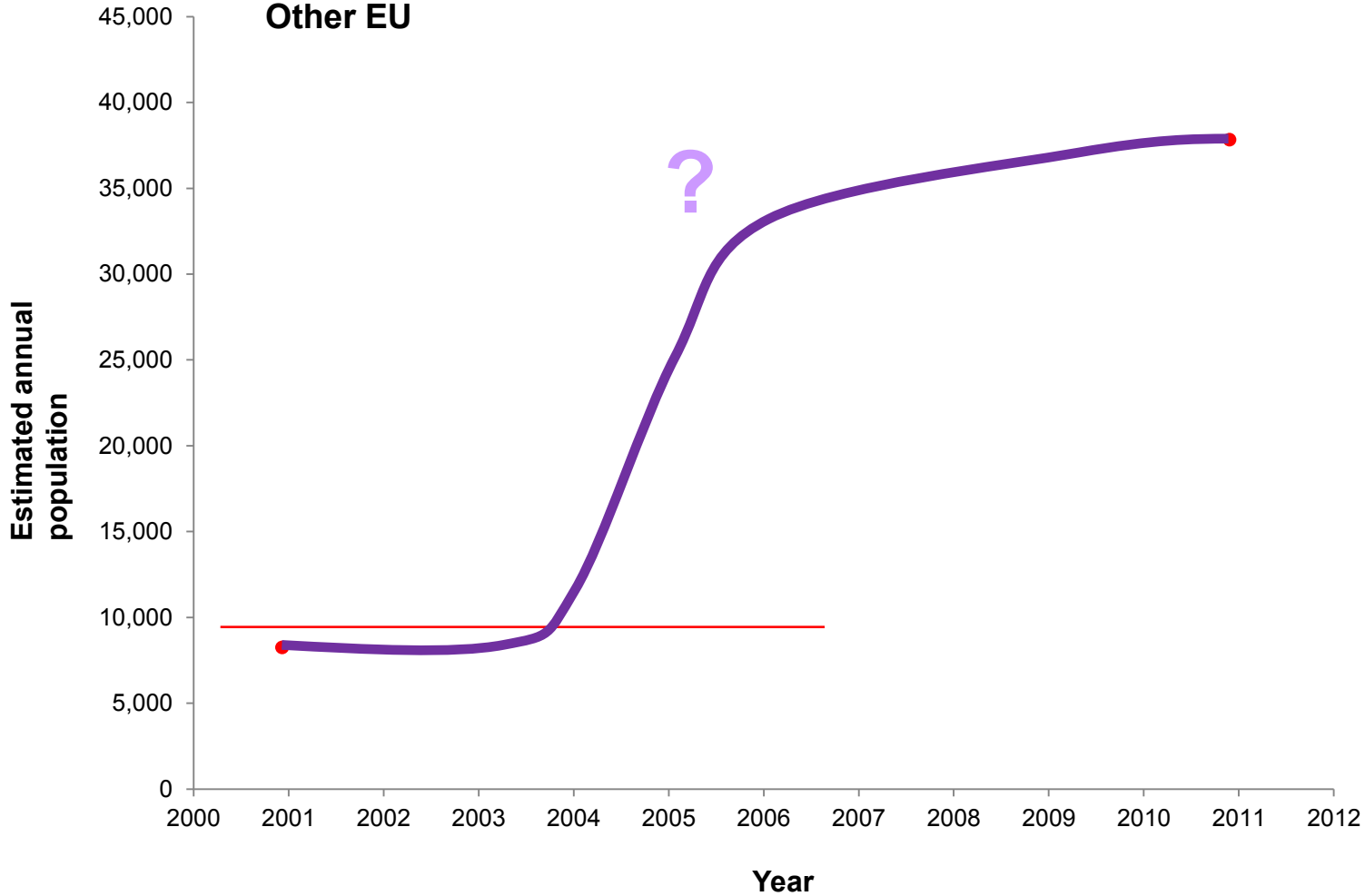


Administrative data and survey approach
spotting change
over time

Administrative data and survey approach

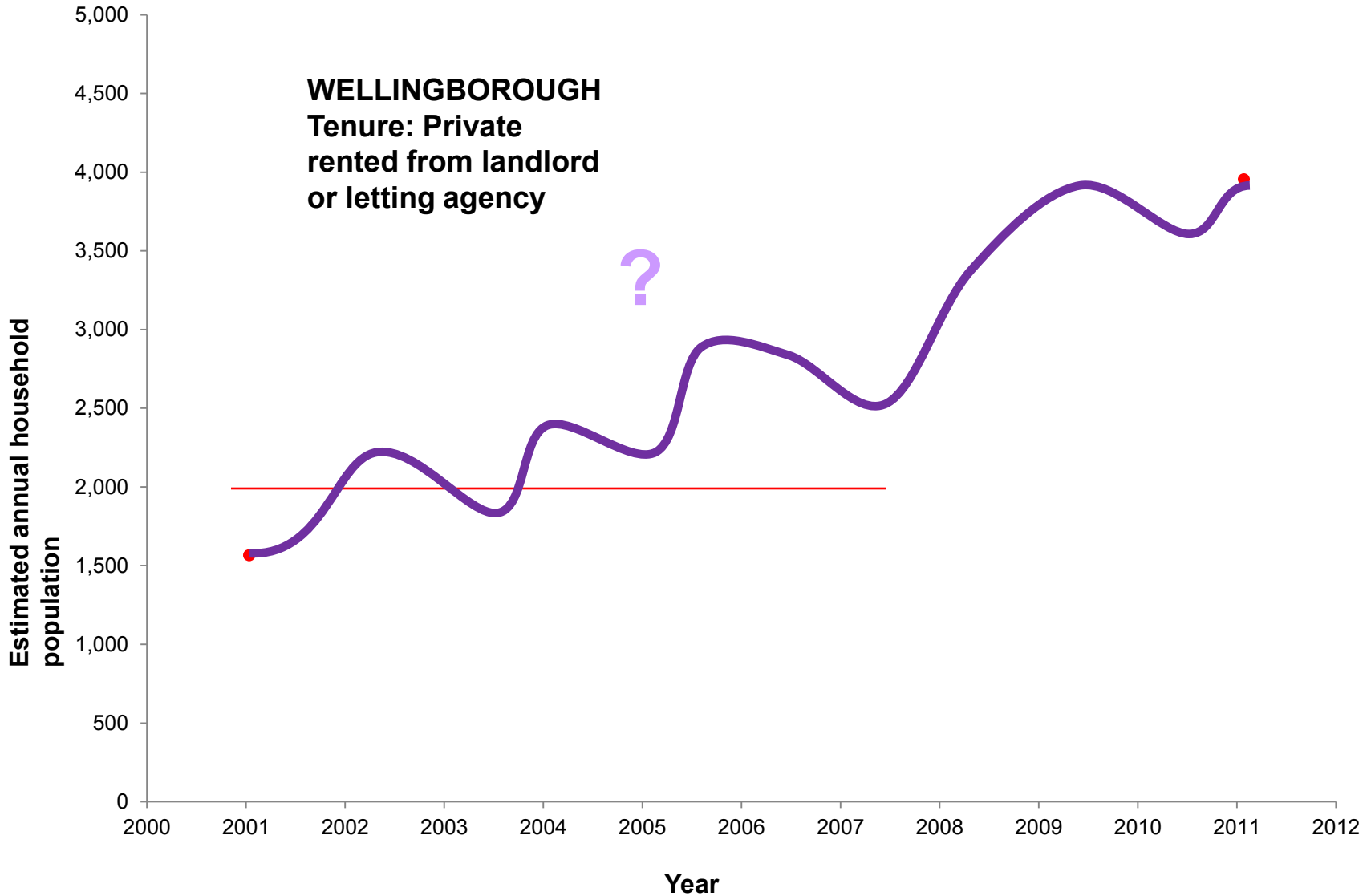
spotting change over time

HARINGEY
Country of Birth:
Other EU



Administrative data and survey approach

spotting change over time



Administrative data and survey approach

What you get - characteristics

Box F: Statistics possible using survey data				
Area type	Average number of residents	1 year's data (800 threshold)	3 years' data (230 threshold)	5 years' data (130 threshold)
LA	160,000	Detailed cross-tabulations (c 200 cells)	Detailed cross-tabulations (c 500 cells)	Very detailed cross-tabulations (c 1000 cells)
MSOA	7,800	Some single variable statistics (c 10 cells)	Very simple cross-tabulations (c 30 cells)	Simple cross-tabulations (c 50 cells)
LSOA	1,600	Not available	Some single variable statistics (c 5 cells)	Some single variable statistics (c 10 cells)
OA	300	Not available	Not available	Not available

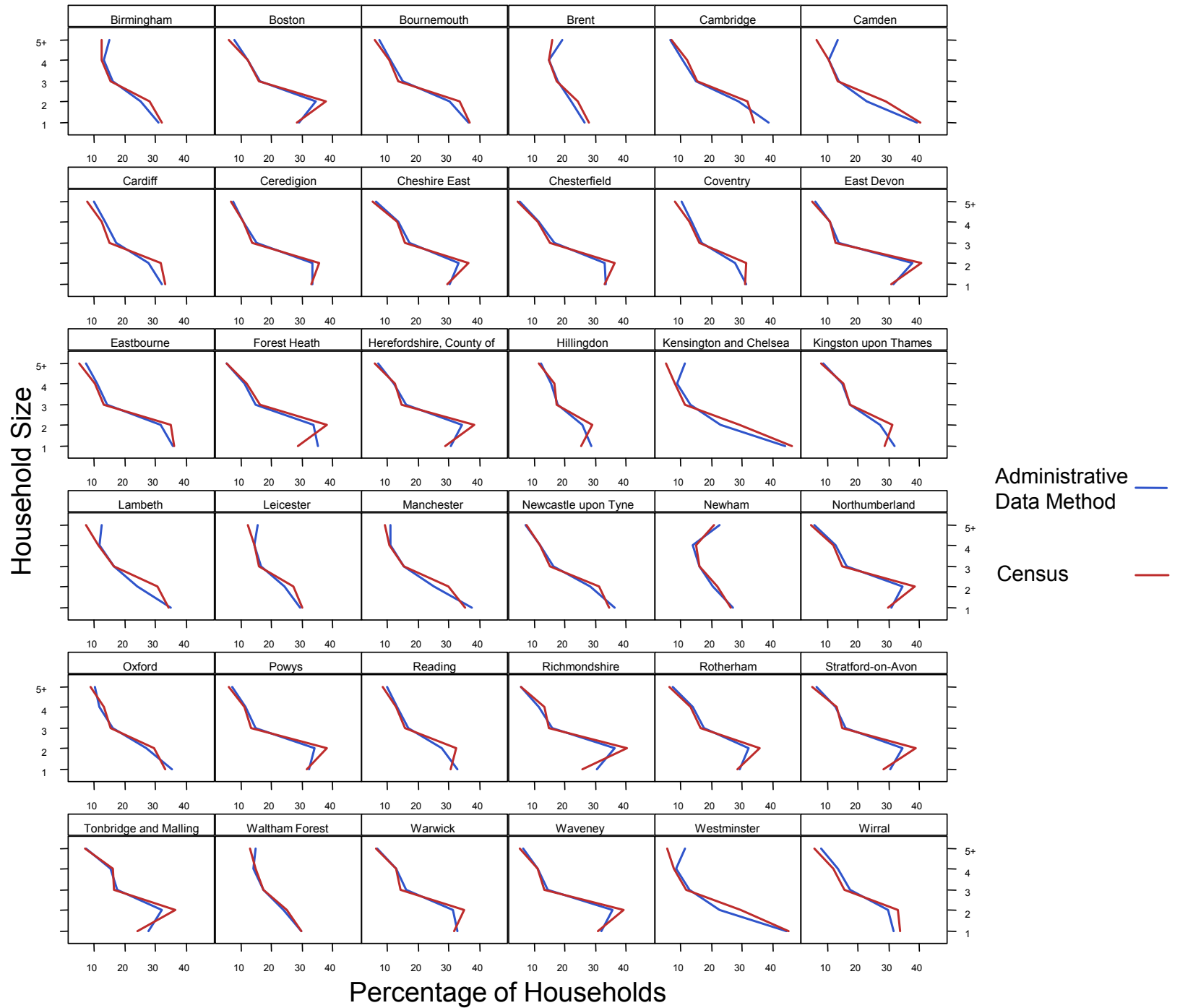
What you get - characteristics

Box G: Statistics available for high priority topics				
	Level of geography, and years of survey data needed			
Topic	LA	MSOA	LSOA	OA
Ethnicity	1 for large groups 3 for most others 5 for very small groups	3 and 5	5	Not available
Economic activity	1	3	5	Not available
Marital status	1	1 for large categories 3 for others	3 for large categories 5 for others	Not available
Tenure	1	3	3 for large categories 5 for others	Not available
Household composition	1	1 for large household types 3 for others	5	Not available
Accommodation type	1	3	5	Not available



Administrative data and survey approach
household size
& composition

Comparison of Percentage of Households of Each Size for selected LAs



Admin data - the potential ??

Key advantage – broad coverage sources allow statistics at lowest geographies ? – every year ?

Household composition	– various
Income	– HMRC / DWP
Economic status	– HMRC / DWP
Health status (index?)	– HSCIC / NHS Wales
Qualifications	– Census / DfE / BIS
Industry of employer	– HMRC
Carers	– HMRC / DWP (limited)
Ethnicity	– NHS (quality?) – HMRC / DWP (limited)

(Full list in paper M12)



The two approaches

Advantages and disadvantages

Beyond 2011 : Advantages and disadvantages

A census once a decade
-like that conducted in 2011,
but **primarily online**

RISKS

- Increasingly difficult to get high response
- Other methods of completion required for some households
- Considered an invasion of privacy by some

STRENGTHS

- A rich set of statistics for a range of geographies, a wide range of topics, small populations, detailed cross tabulations
- Proven ability to deliver - proven and tested
- High degree of continuity
- A single, high quality snapshot of the nation

WEAKNESSES

- Only every 10 years (except for LA population estimates) – reduces usefulness
- Costs more than the admin data option - £625m per decade - £1.10 per person per year
- Build-up and run-down challenging
- A burden on all households

OPPORTUNITIES

- Online completion will be cheaper and more efficient

Beyond 2011 : Advantages and disadvantages

STRENGTHS

- Continually updated – statistics on an annual basis
- Changes and trends identified more quickly
- Less expensive - £460m a decade – 80p per person per year
- Reduced burden on households

WEAKNESSES

- Will never produce the detail provided by the census
- Data combined for several years – makes date to which it refers more complex
- Loss of a single historical record (options to store more – but not yet developed)
- Requires new legislation

OPPORTUNITIES

- Use of admin data can be extended over time
- Potential to be more flexible in questions
- New opportunities for historic research (in 2121)

A census based on administrative data and large annual surveys

RISKS

- New and untested methods – other countries have taken decades
- Some discontinuities
- Requires access to admin data
- Survey response will be challenging here too
- Requires public acceptance of use of admin data

Consultation - Planned events

Public consultation 23rd October – 13th December 2013

- London, Manchester public consultation events
- Census Research Users Conference
- Demographics User Group
- IDSG
- Cardiff public consultation event
- RGS – small area data - academics event
- Academy of Social Sciences
- CLIP – LGA – further LA events?? (focussed benefits meeting)
- British Academy Event
- Genealogy forum
- Equality and Diversity forum
- Meetings of our Privacy & Equality Groups
- HoC event
- Voluntary sector
- Health Event
- London, Manchester events
- (International review panel, Research conference) etc etc etc



The case for small area data

The case for small area data

- We have not decided – and are open to debate
- But this is a difficult decision

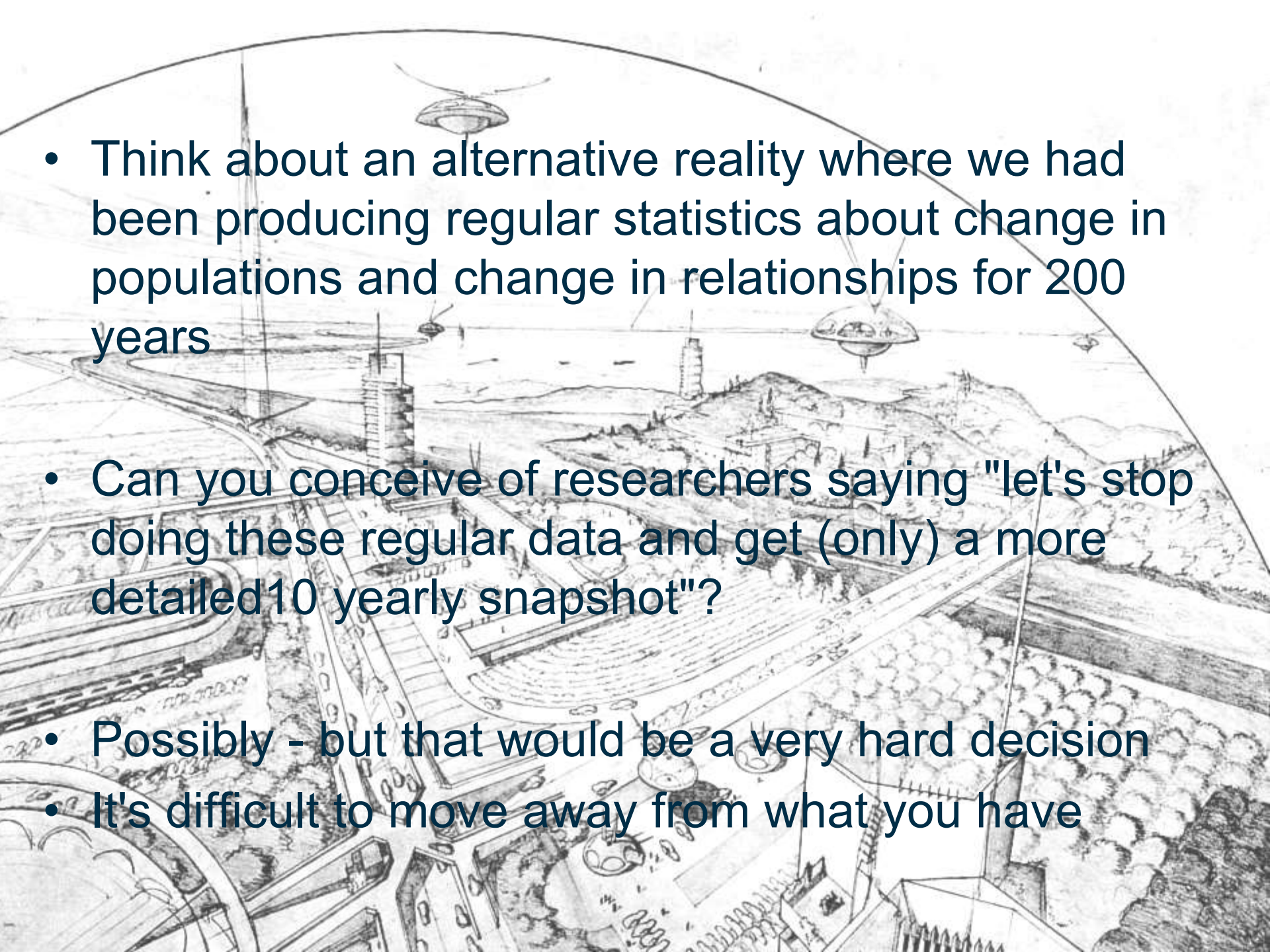
risks & opportunities

- The case for small area data is not yet made
- Welcome today's discussion - & case studies

Case studies

- Don't tell us that relationships between things matter
We will deliver relationships LA / MSOA
(Unless they are very subtle differences that make a big difference)
- Don't tell us that the spatial distribution of things matter –
we will deliver info on spatial distributions to MSOA / LSOA
(Unless the distribution of small numbers really matter or small areas matter)
- Tell us where 'the ecological fallacy' is really an issue

And remember we have to make a case in terms of impact

- 
- Think about an alternative reality where we had been producing regular statistics about change in populations and change in relationships for 200 years
 - Can you conceive of researchers saying "let's stop doing these regular data and get (only) a more detailed 10 yearly snapshot"?
 - Possibly - but that would be a very hard decision
 - It's difficult to move away from what you have

*you don't know what you've
got 'til it's gone*

Joni Mitchell & Neil Storer

*you only get one shot, do not
miss your chance ... this
opportunity comes once in a
lifetime yo*

Eminem

beyond2011@ons.gov.uk