

# Automated zone design

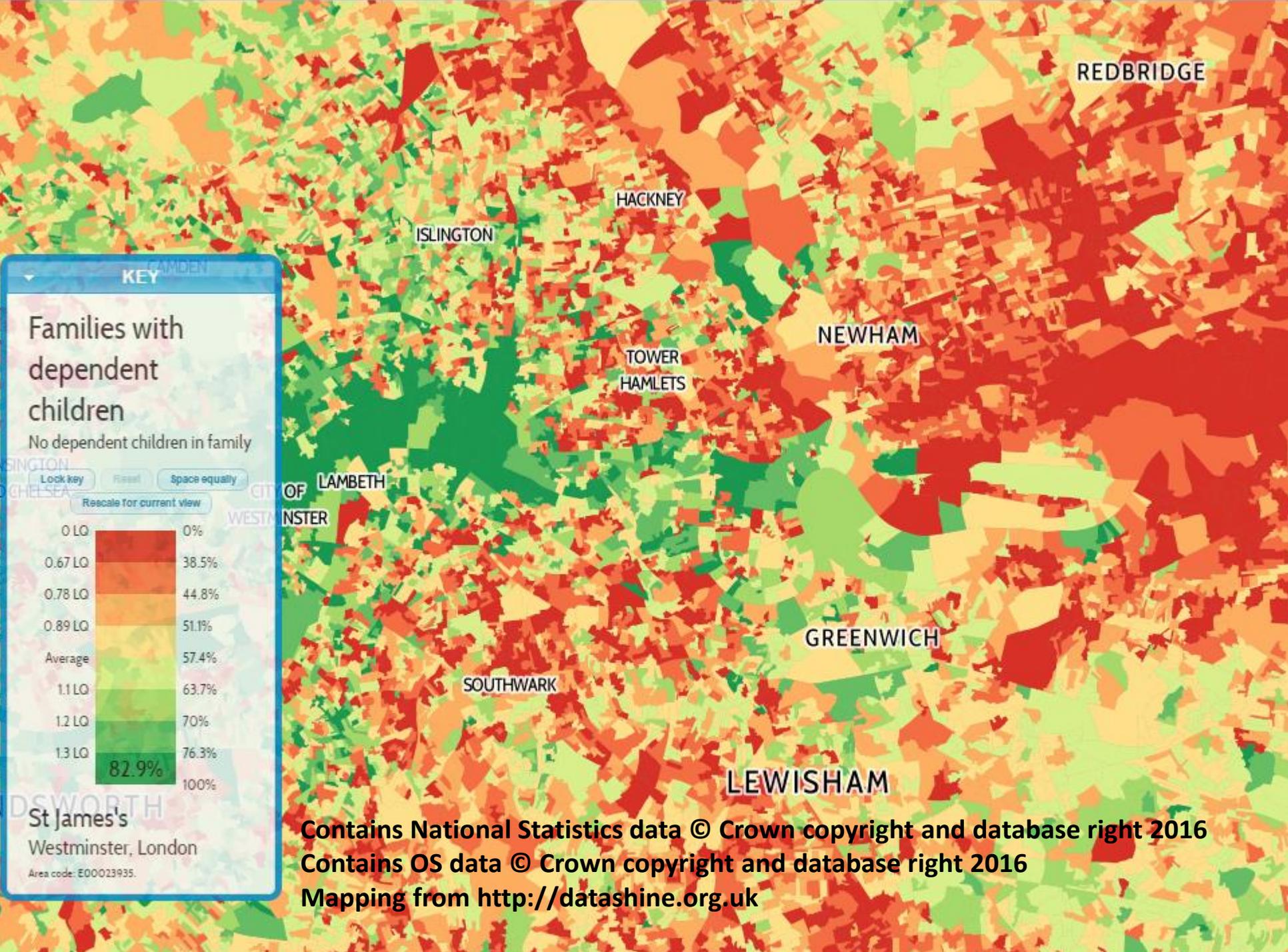
Professor David Martin

# What are zones?

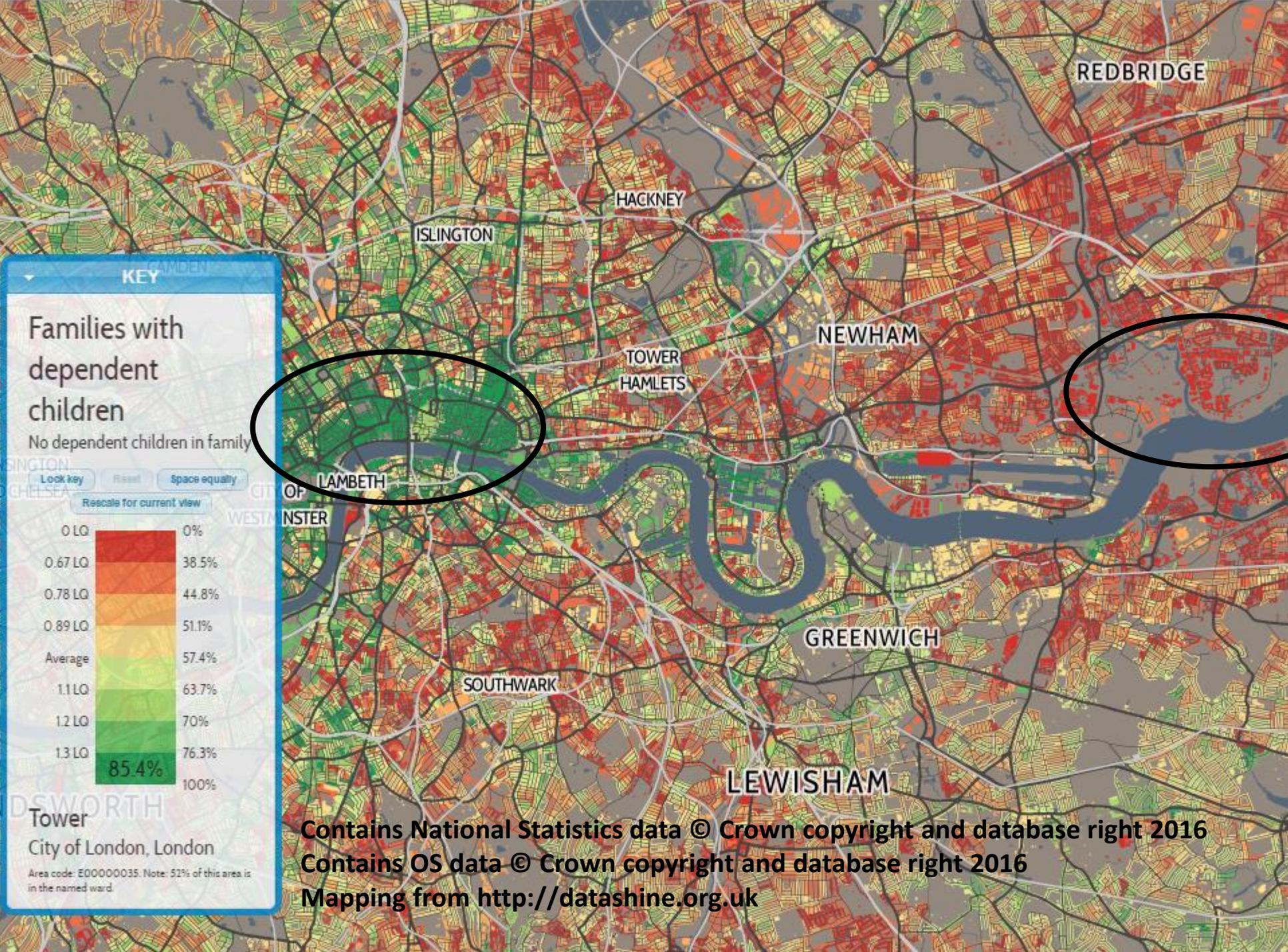
- Divisions of geographical space, usually defined in terms of polygons - often thought of as just shaded areas on a map
- Usually represented by a single polygon, although sometimes islands or separate parts
  - regions, counties, local authorities, wards, electoral districts, constituencies, states (US), communes (France), mesh blocks (Australia), postcode sectors, output areas (UK)

# Example – 2011 census output areas (England and Wales)

- Key characteristics
  - Mean population size 325 persons
  - Always having more than 100 persons and 40 households
  - Many based on 2001 Census Output Areas
  - Matching as far as possible to unit postcodes
  - Control over shape and social homogeneity
  - Used for the publication of small area census statistics



Contains National Statistics data © Crown copyright and database right 2016  
Contains OS data © Crown copyright and database right 2016  
Mapping from <http://datashine.org.uk>



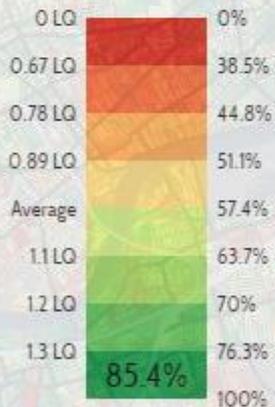
KEY

### Families with dependent children

No dependent children in family

Look key    Reset    Space equally

Rescale for current view



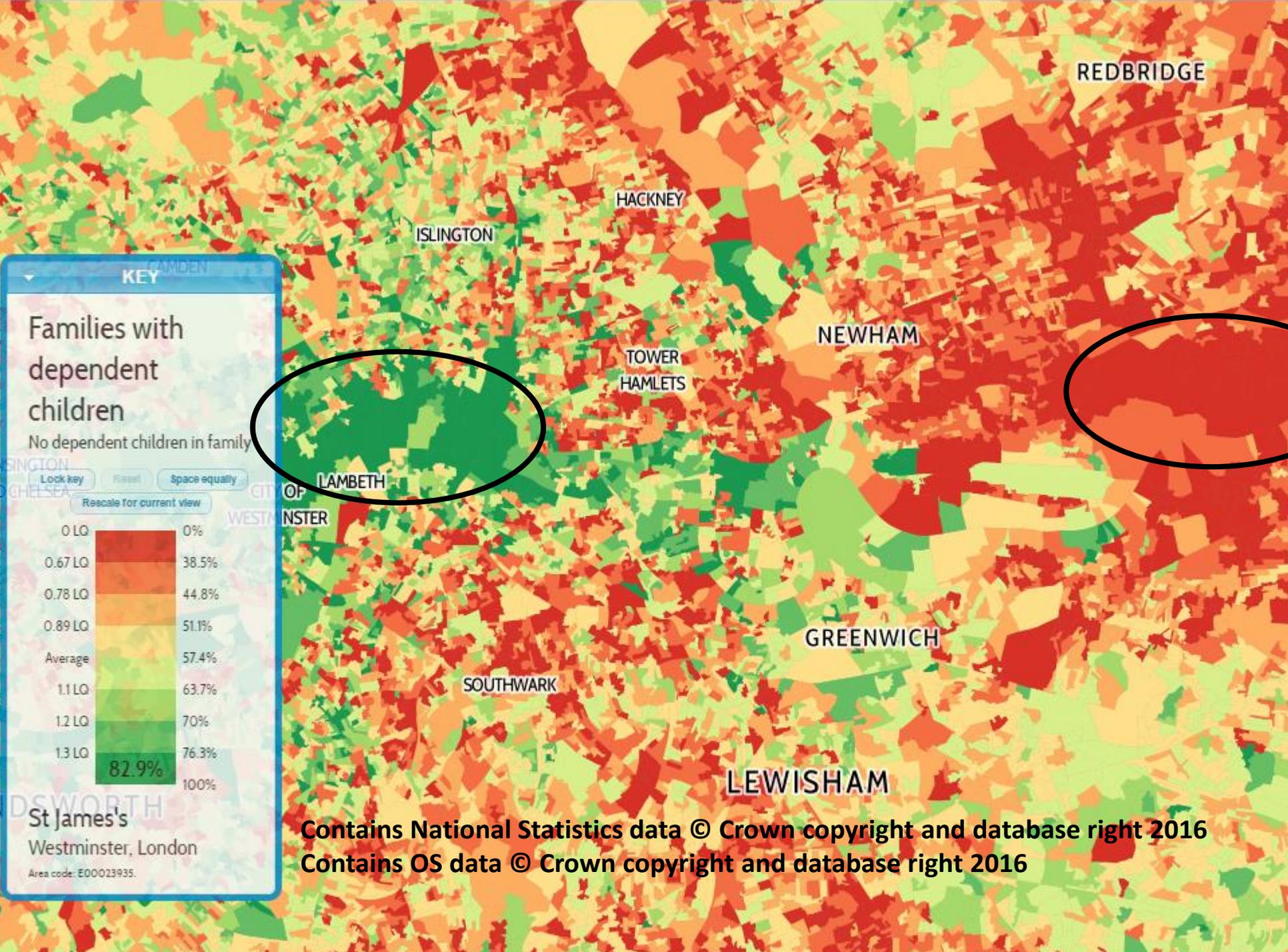
Tower  
City of London, London

Area code: E00000035. Note: 52% of this area is in the named ward.

Contains National Statistics data © Crown copyright and database right 2016

Contains OS data © Crown copyright and database right 2016

Mapping from <http://datashine.org.uk>



**KEY**

### Families with dependent children

No dependent children in family

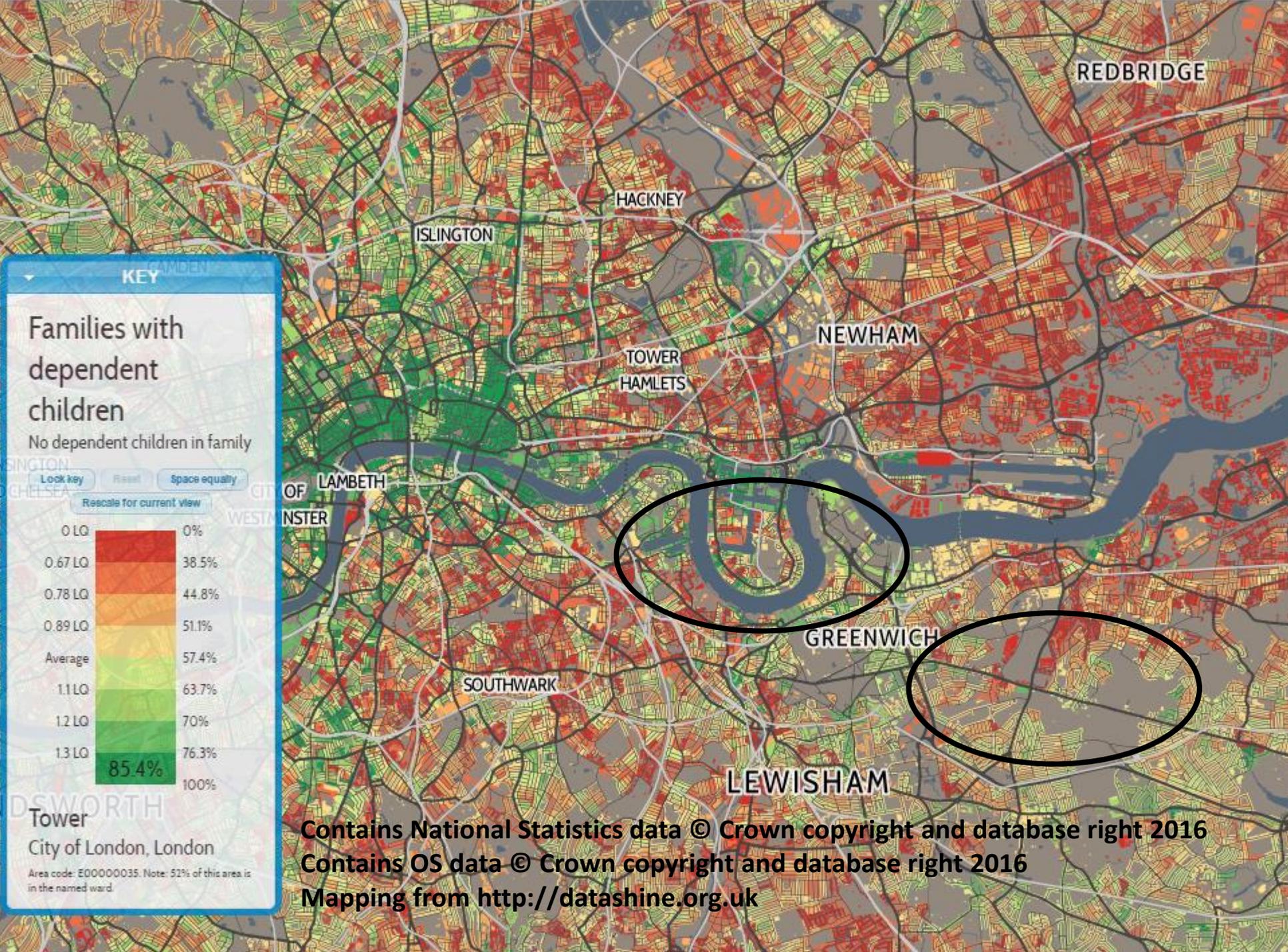
Lock key    Reset    Space equally

Rescale for current view

0 LO	0%
0.67 LO	38.5%
0.78 LO	44.8%
0.89 LO	51.1%
Average	57.4%
1.1 LO	63.7%
1.2 LO	70%
1.3 LO	76.3%
82.9%	100%

St James's  
Westminster, London  
Area code: E00023935

Contains National Statistics data © Crown copyright and database right 2016  
Contains OS data © Crown copyright and database right 2016



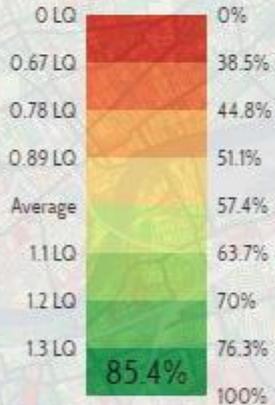
KEY

### Families with dependent children

No dependent children in family

Look key    Reset    Space equally

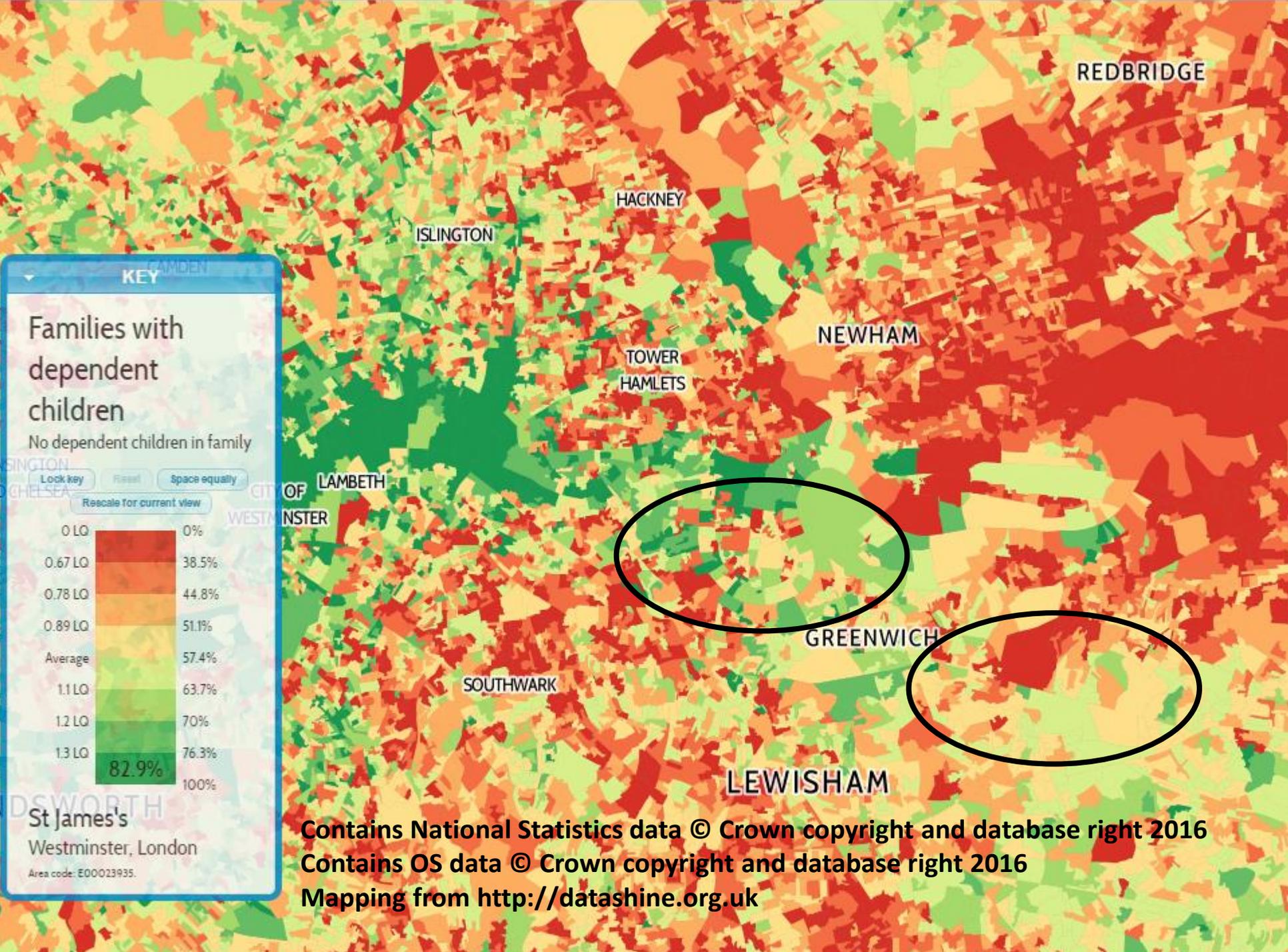
Rescale for current view



Tower  
City of London, London

Area code: E00000035. Note: 52% of this area is in the named ward.

Contains National Statistics data © Crown copyright and database right 2016  
Contains OS data © Crown copyright and database right 2016  
Mapping from <http://datashine.org.uk>



REDBRIDGE

HACKNEY

ISLINGTON

NEWHAM

TOWER  
HAMLETS

LAMBETH

WESTMINSTER

GREENWICH

SOUTHWARK

LEWISHAM

KEY

Families with dependent children

No dependent children in family

Lock key    Fit all    Space equally

Rescale for current view

0 LO	0%
0.67 LO	38.5%
0.78 LO	44.8%
0.89 LO	51.1%
Average	57.4%
1.1 LO	63.7%
1.2 LO	70%
1.3 LO	76.3%
82.9%	100%

St James's  
Westminster, London

Area code: E00023935.

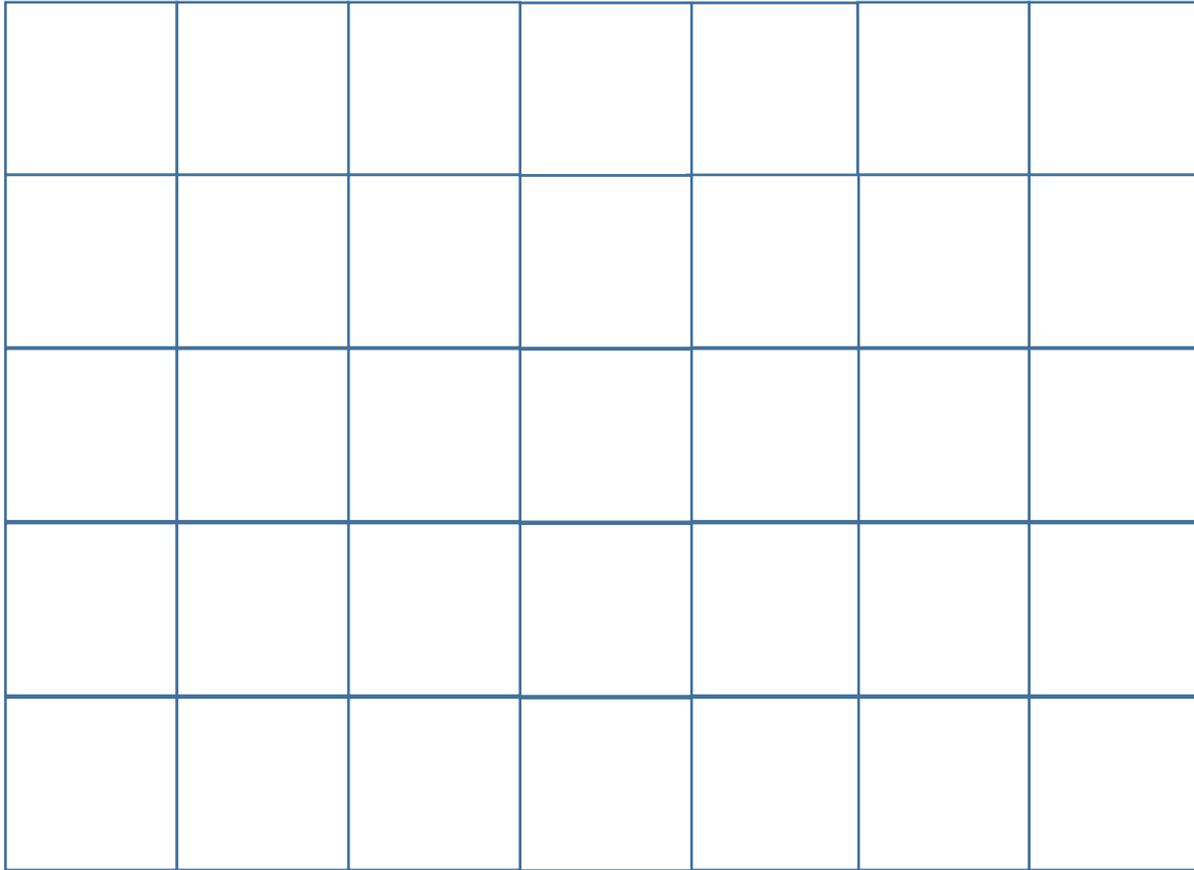
Contains National Statistics data © Crown copyright and database right 2016  
Contains OS data © Crown copyright and database right 2016  
Mapping from <http://datashine.org.uk>

# What is zone design?

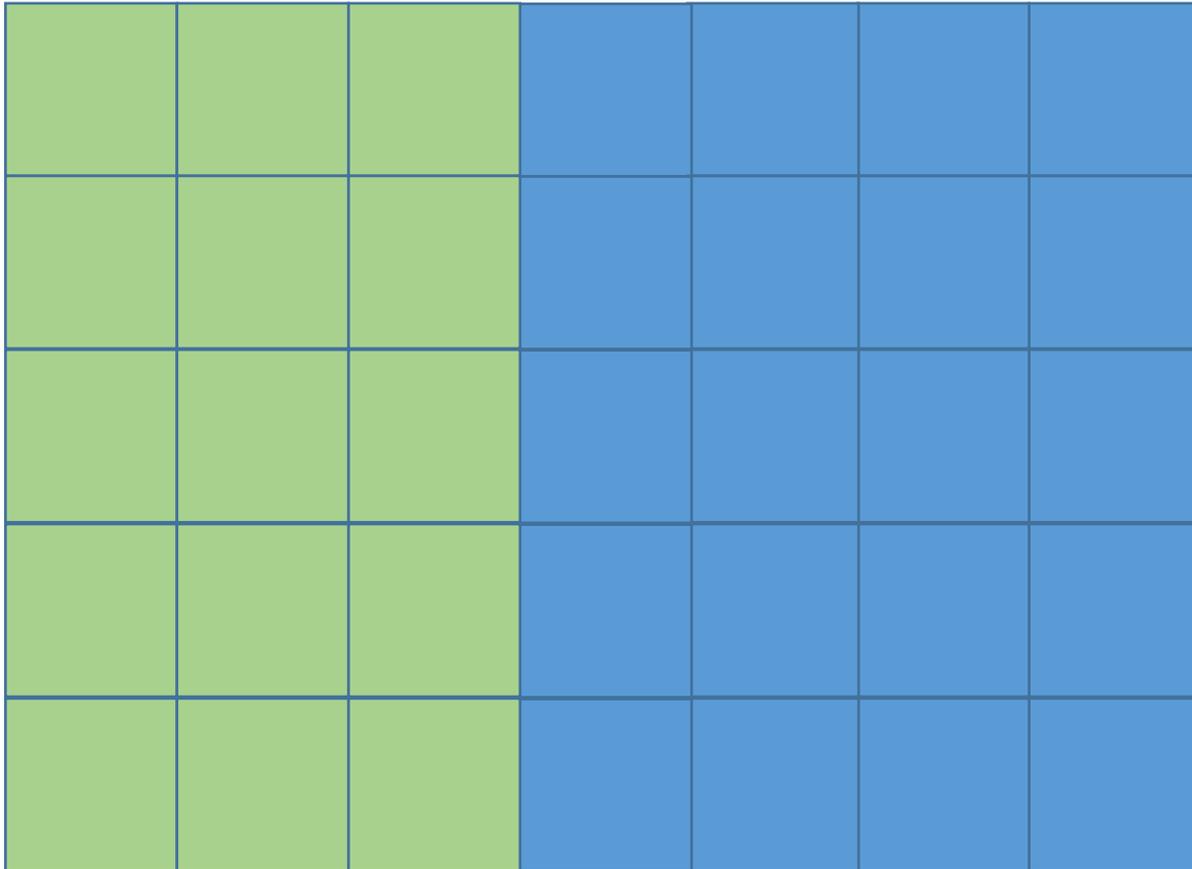
- Choice of the number and configuration of zones
- If used to count statistical units (persons, households), determines which units will be aggregated
- Disclosure control: ensuring sufficiently large populations
- Different combinations of historical, administrative processes or an algorithm
- May be result of very careful consideration or a relatively arbitrary process

# So why does it matter?

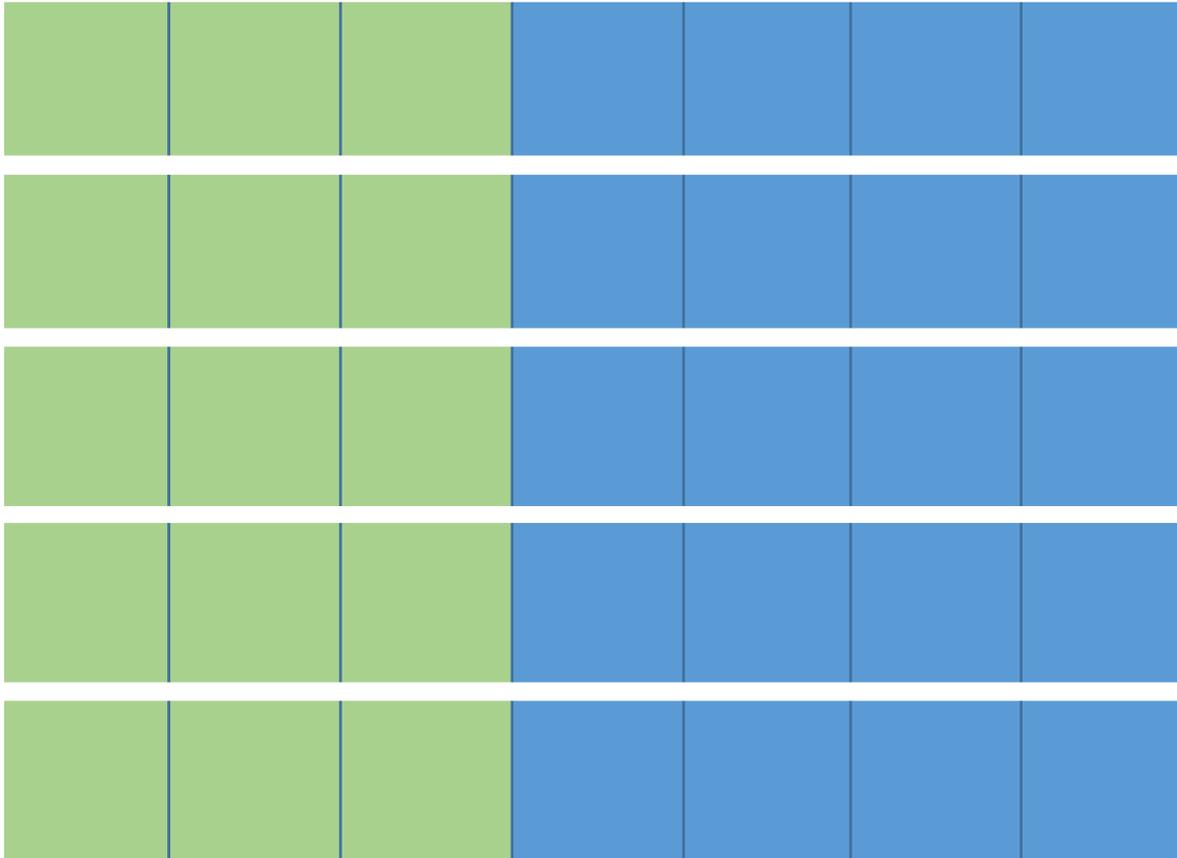
- Depending on the purpose, size and position of boundaries may matter in many different ways
- Geographers know this as the “Modifiable Areal Unit Problem” (Openshaw, 1984)
- Comprises “scale” and “aggregation” problems
- The same phenomenon when applied to the manipulation of electoral boundaries is known as Gerrymandering



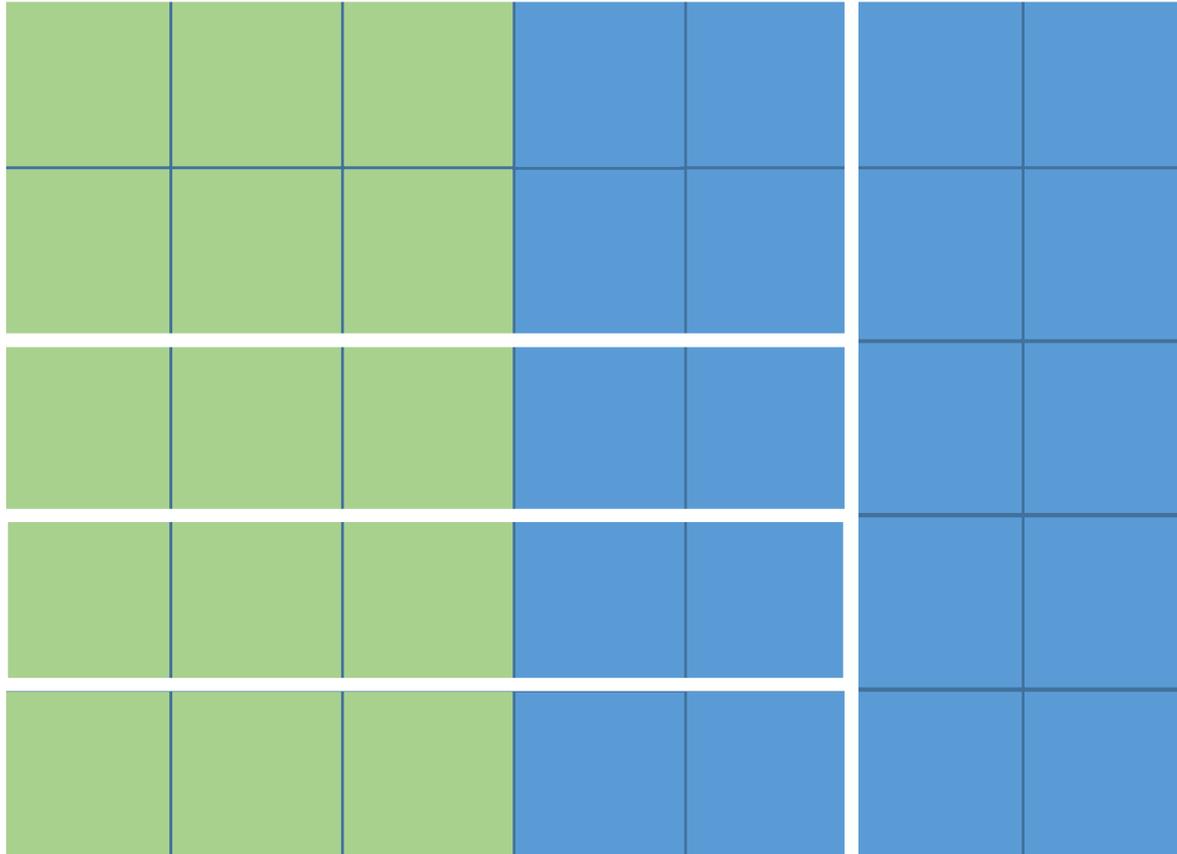
A vote is held in 35 (square) neighbourhoods



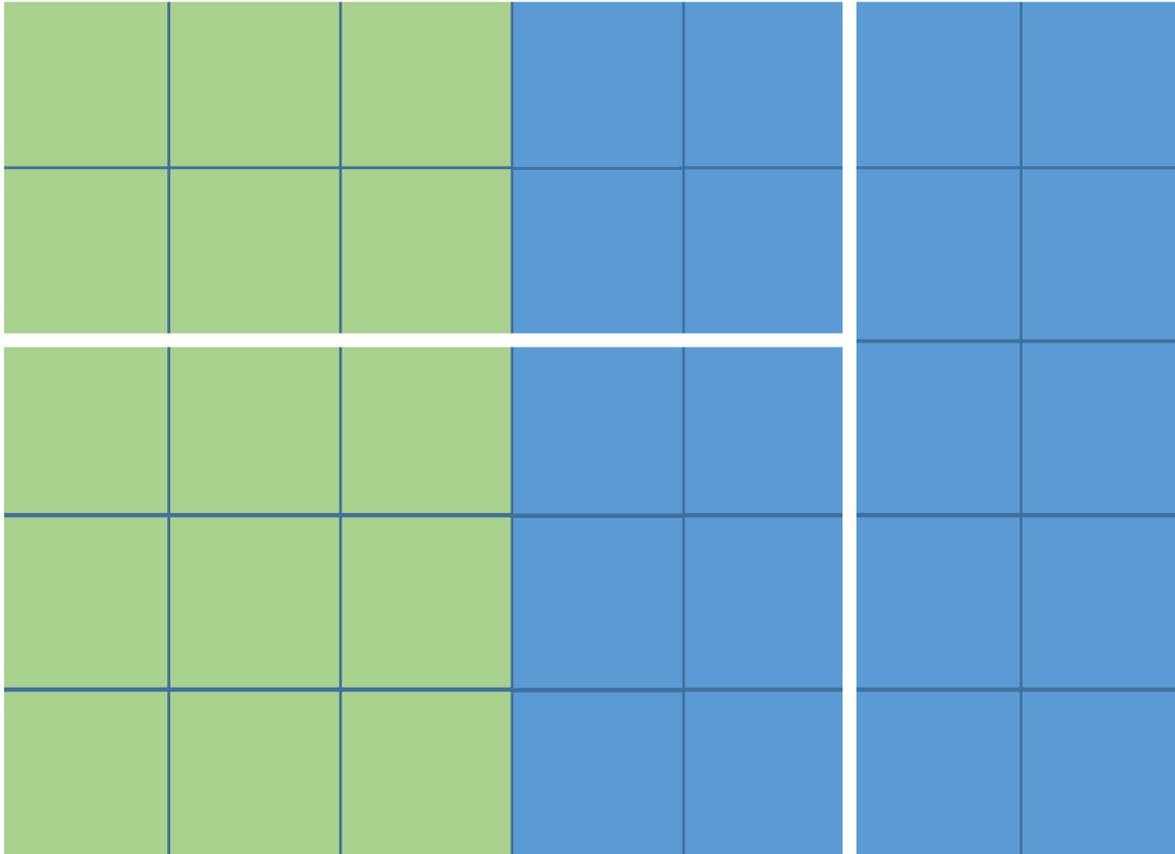
15 neighbourhoods vote green; 20 vote blue



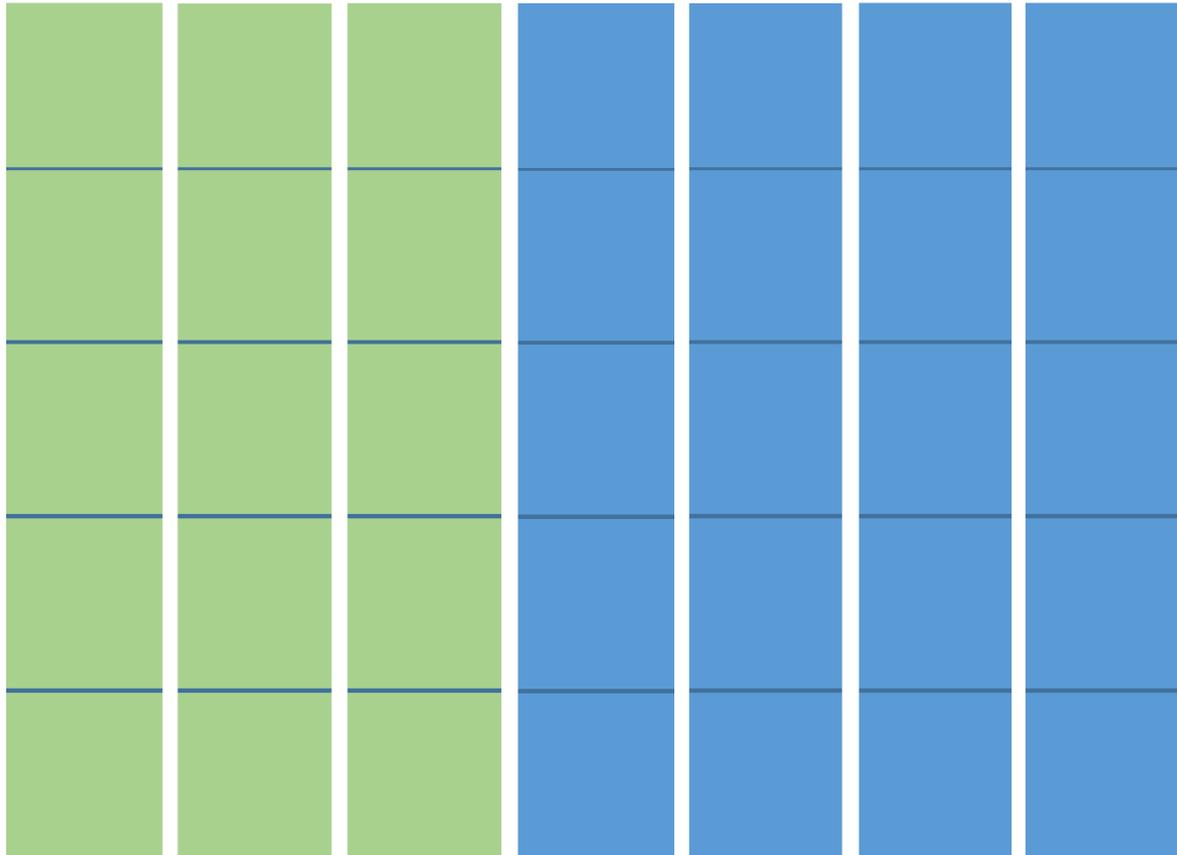
Arranged in 5 constituencies **blue** wins all 5



But with these 5 constituencies **green** wins 4 **blue** wins 1



With these 3 constituencies green wins 2, blue wins 1



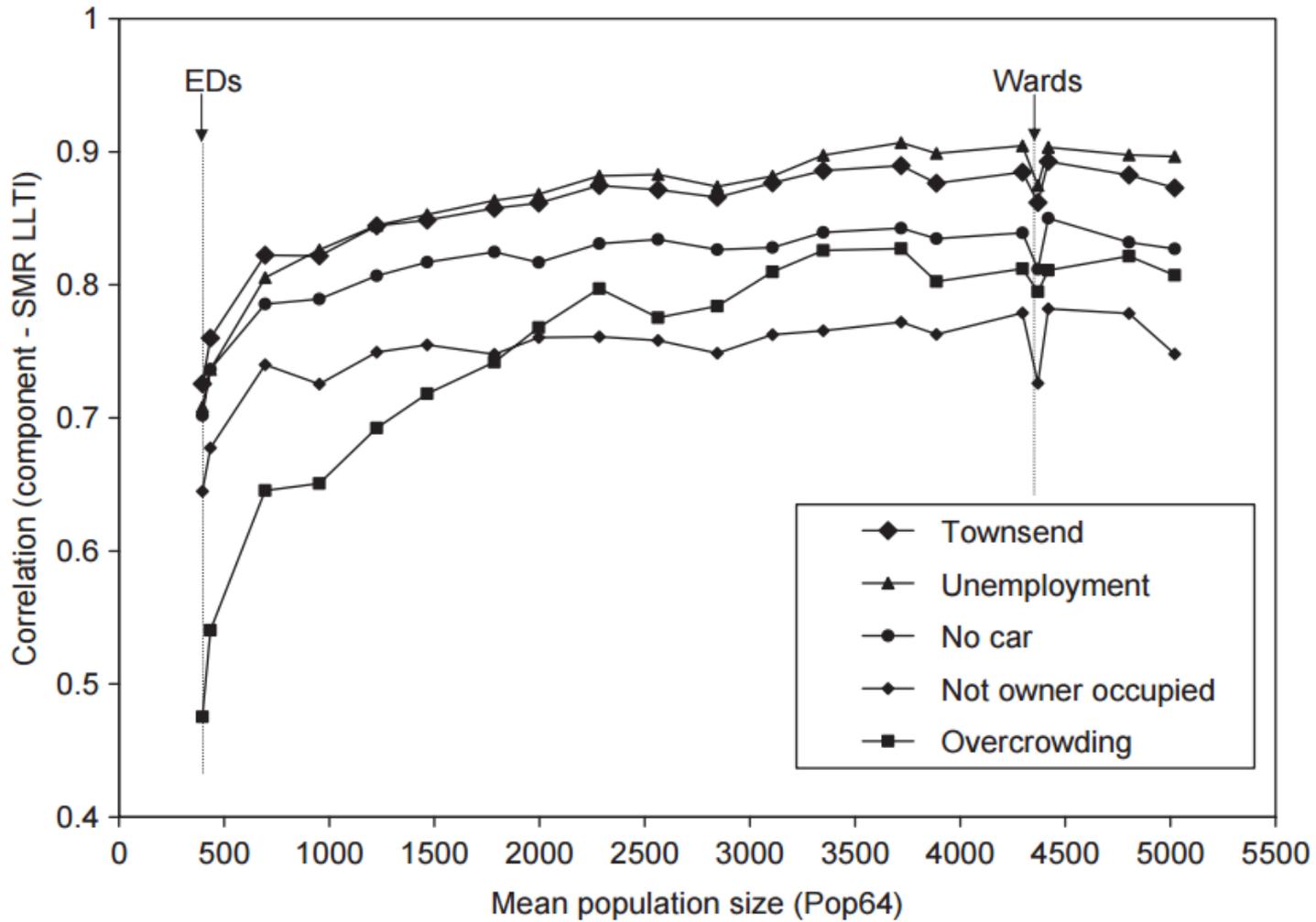
This 7 constituency solution reflects the exact proportion at the neighbourhood level, **green** wins 3, **blue** wins 4

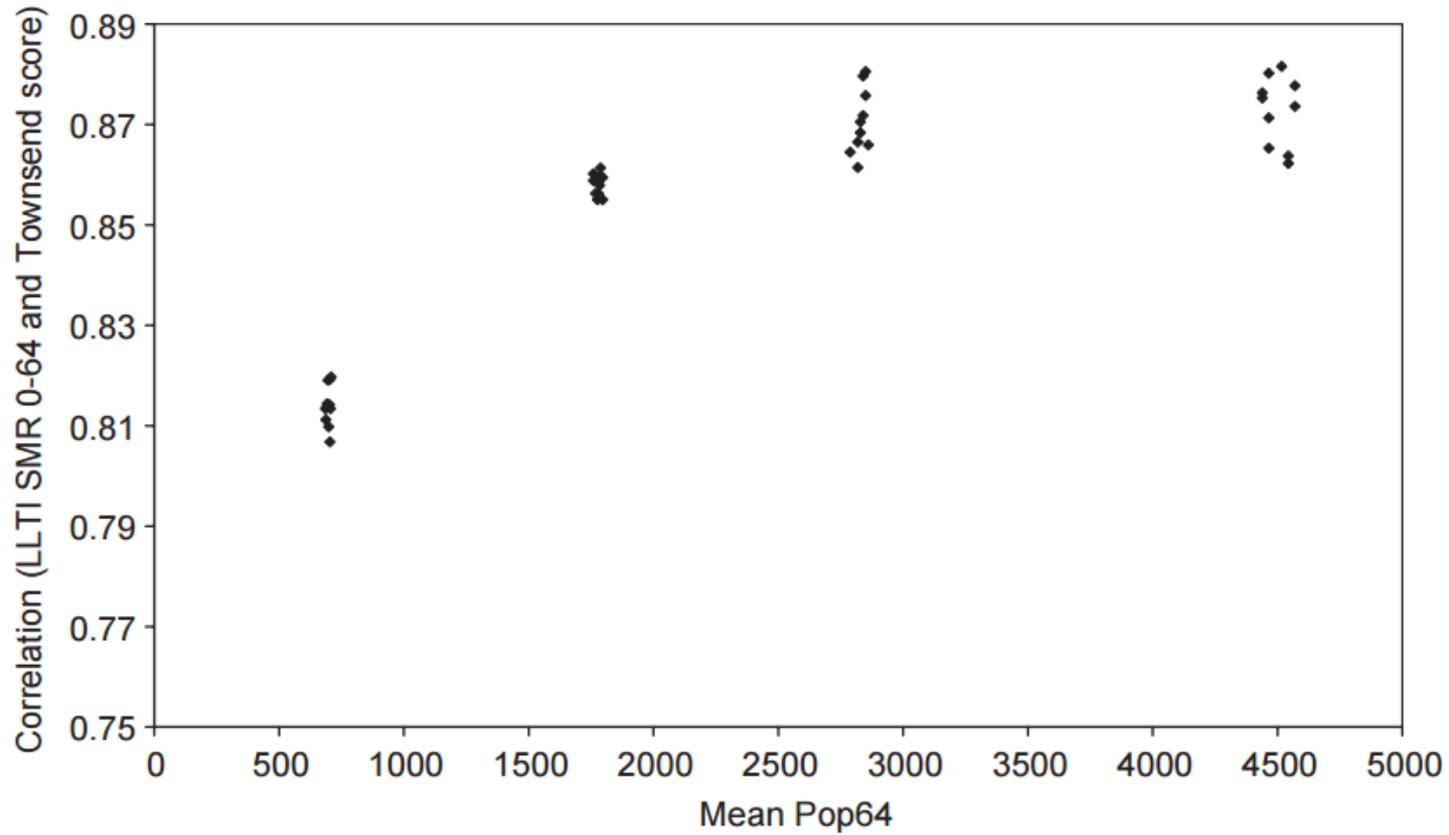
# Scale and aggregation problems

- Scale problem: how many constituencies
- Aggregation problem: which configuration of boundaries, at a given scale
- Gerrymandering and “postcode lottery” issues are real world consequences of zone design decisions
- Whether design of zones is actually a “problem” depends on the intended purpose

# Impact on statistical relationships

- Way in which counts are grouped may have a direct impact on measures such as election results
- Configuration of zone boundaries also affects observed relationships between variables and thus ecological associations
- Different relationships hold at different geographical scales, but also for different aggregations at the same scale





# Summary

- Zones used for many statistical and policy purposes
- Zone design can have big impacts on research and everyday life
- Researchers who use zone-based data need to understand the methods by which zones have been created
- Where appropriate, consider designing own zones appropriate to research objectives
- Particular significance in ensuring confidentiality of aggregated data

# References

- Cockings, S. and Martin, D. (2005) Zone design for environment and health studies using pre-aggregated data *Social Science and Medicine* 60, 2729-2742
- Openshaw, S. (1984) *The modifiable areal unit problem* Concepts and Techniques in Modern Geography No. 38 Geo Books, Norwich

For more information please visit  
[www.ncrm.ac.uk](http://www.ncrm.ac.uk)

