UNIVERSITY OF Southampton
School of Geography

Automated zone design:
Concepts, methods, applications

Samantha Cockings, University of Southampton

Statistics New Zealand
March 2011

UNIVERSITY OF Southampton
School of Geography

Presentation overview

- Acknowledgements: David Martin, Andy Harfoot, Economic and Social Research Council (ESRC), Office for National Statistics (ONS), Department for Transport (DfT)
- Automated zone design:
  - Concepts and history
  - Example applications
    - Census output areas
    - Workplace zones
  - Future directions
  - AZTool software

UNIVERSITY OF
Southampton
School of Geography

Automated zone design: Concepts and history

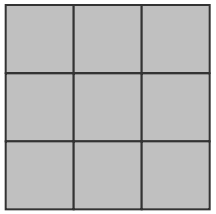
What is automated zone design?

- Iterative combination and re-combination of sets of building blocks in order to create output zones (tracts) which optimise a set of pre-specified design criteria
- Building blocks: the smallest elemental set of areas for which data are available and which make sense for the modelling of the specific phenomena e.g. parcels of land around addresses, post/zip codes, enumeration districts, meshblocks …
- Tracts: the set of output zones created following the automated zone design process e.g. output areas in E&W

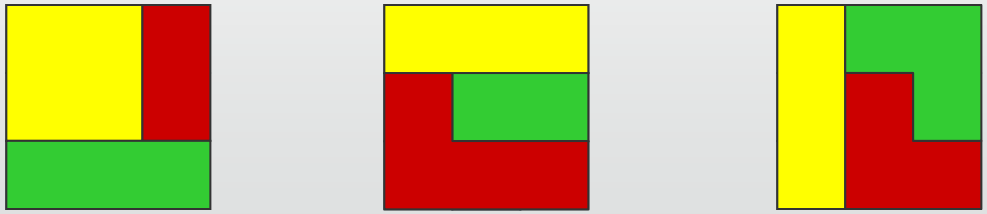
UNIVERSITY OF Southampton
School of Geography

Automated zone design

1. Create a set of building blocks



2. Aggregate the building blocks into larger zones using a set of design constraints and objectives/targets



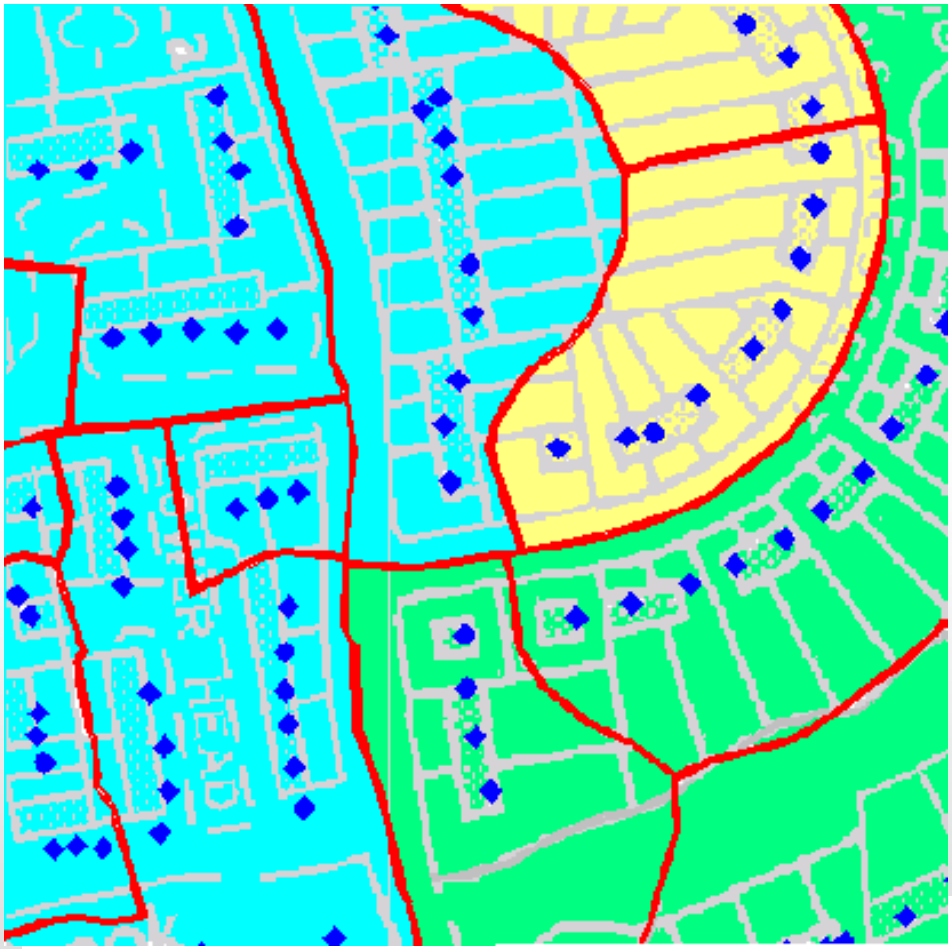
Typical design criteria

- 'Hard' constraints – must be met
  - Thresholds e.g. minimum population per tract
- 'Soft' constraints – are optimised (overall)
  - Targets e.g. population size per tract
  - Homogeneity (within zone) e.g. accommodation type
  - Shape e.g. compactness
- Some may be implemented as hard or soft
  - e.g. higher level geography (region)

Automated zone design (1)

UNIVERSITY OF Southampton
School of Geography

Initial random aggregation of building blocks into potential tracts



Slide courtesy of David Martin

Automated zone design (2)

UNIVERSITY OF Southampton
School of Geography

Choose one building block at random as candidate for swapping into a different tract

Slide courtesy of David Martin

Automated zone design (3)

UNIVERSITY OF Southampton
School of Geography

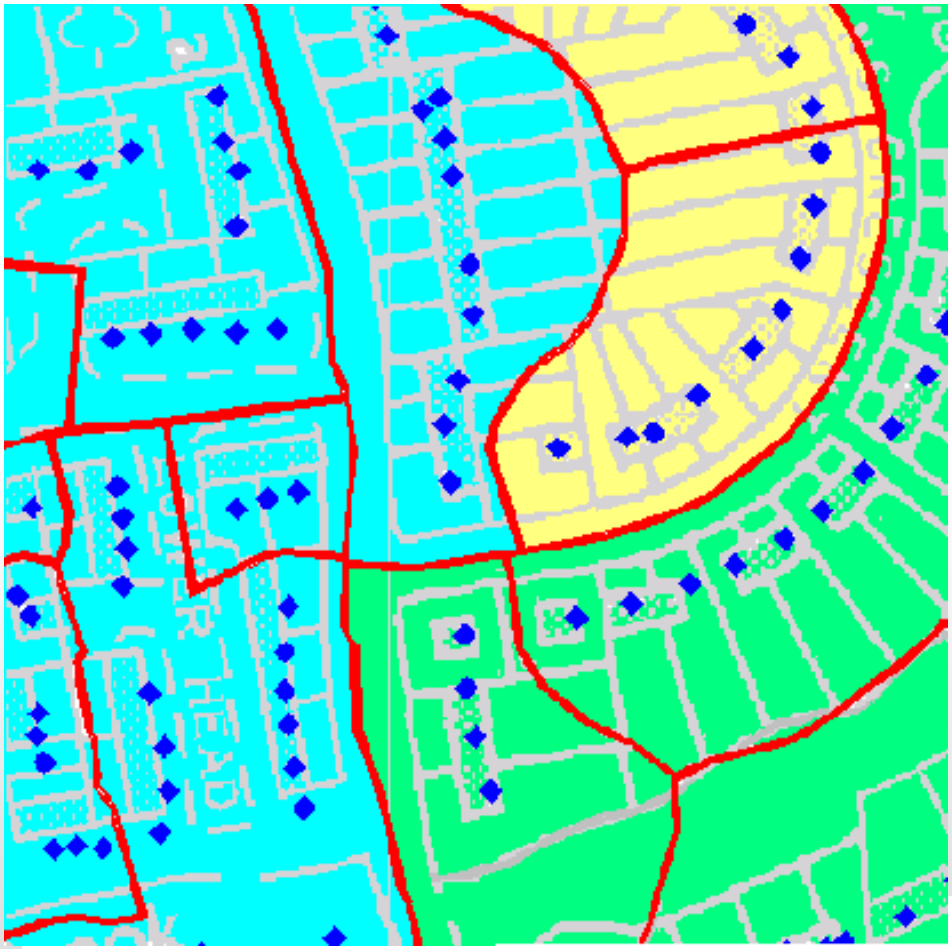
Make the swap and evaluate the impact on the overall solution

Slide courtesy of David Martin

Automated zone design (4)

UNIVERSITY OF Southampton
School of Geography

If swap does not result in an improvement, go back to the previous configuration

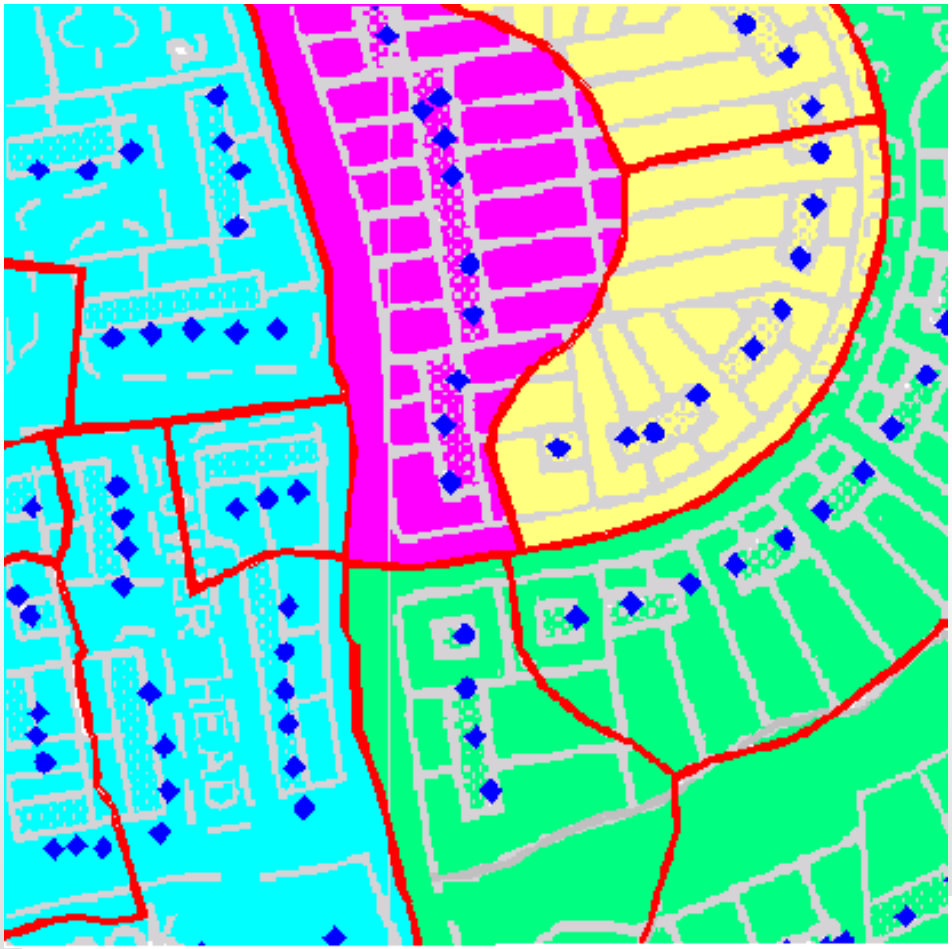


Slide courtesy of David Martin

Automated zone design (5)

UNIVERSITY OF Southampton
School of Geography

Choose another building block at random as candidate for swapping into another tract



Slide courtesy of David Martin

Automated zone design (6)

UNIVERSITY OF Southampton
School of Geography

If the swap results in an overall improvement, keep it as part of the solution and examine a new potential swap…

Slide courtesy of David Martin

University of Southampton
School of Geography

A brief history of automated zone design …

Year	Details
1977	Openshaw develops automated zoning procedure (AZP)
1995	AZP enhanced by Openshaw and Rao
1998-2003	Martin extends method and creates AZM software
2001-2003	ONS use Martin's algorithm to create 2001 Census output areas (OAs) + super output areas (SOAs)
2005	Cockings and Martin use AZM in health research
2006-	Cockings, Martin, Harfoot develop AZTool software
2008	Ang & Ralphs explore use of AZTool for creating Census output geographies in NZ
2008-10	Cockings et al research use of AZTool for OA maintenance
2009-11	Martin et al research use of AZTool for creating workplace zones (WZ)
2010-	ONS implementing AZTool for OA maintenance (and WZ creation?)

UNIVERSITY OF
Southampton
School of Geography

Automated zone design: Applications

UNIVERSITY OF
Southampton
School of Geography

Applications:
Census output geographies – creation and maintenance

UNIVERSITY OF Southampton
School of Geography

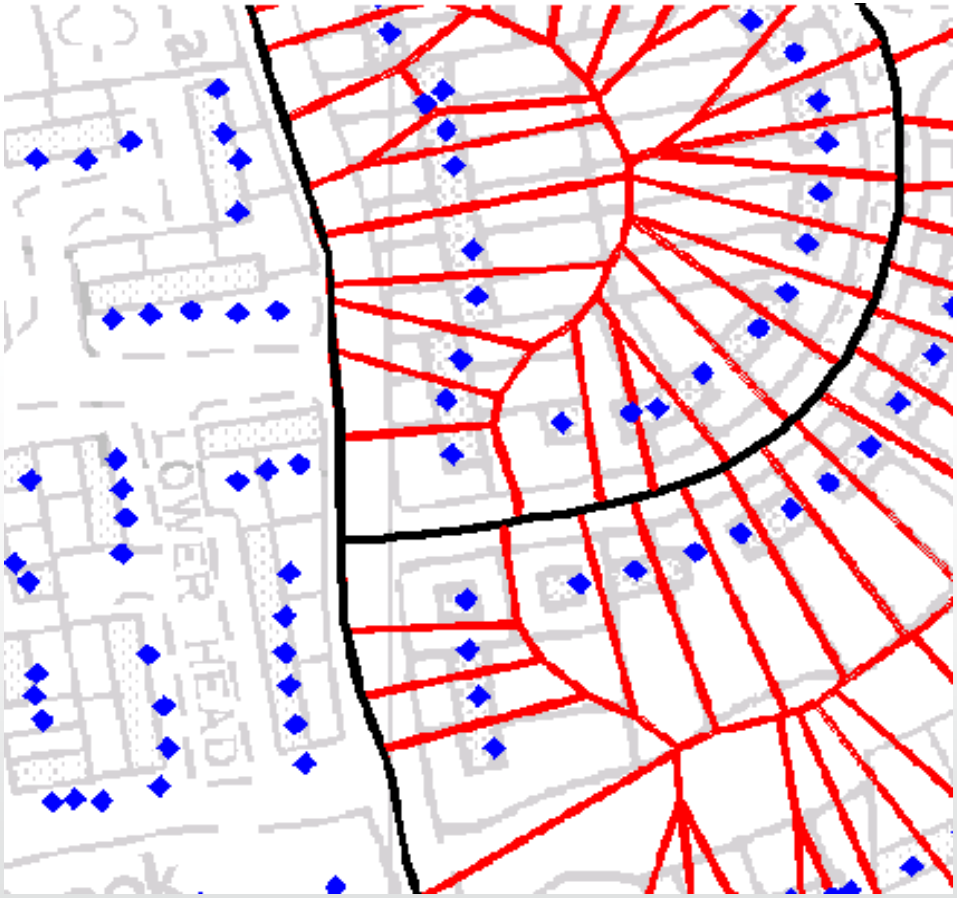
2001 Output Area zone design

- Building blocks: synthetic postcode polygons
  - Address-based thiessen polygons dissolved to create postcode polygons
  - Nested within wards/parishes
  - Respecting road centrelines where possible

UNIVERSITY OF Southampton
School of Geography

Address-based Thiessen polygons

- Thiessen polygons around individual ADDRESS-POINTS intersected with ED, ward, parish boundaries and road centrelines



Slide courtesy of David Martin

UNIVERSITY OF Southampton
School of Geography

Unit postcode building blocks

- Address polygon boundaries dissolved to form unit postcode polygon building blocks

Slide courtesy of David Martin

UNIVERSITY OF Southampton
School of Geography

2001 Output Area zone design

- Building blocks: synthetic postcode polygons
  - Address-based thiessen polygons dissolved to create postcode polygons
  - Nested within wards/parishes
  - Respecting road centrelines where possible
- Design criteria for OAs:
  - Population and household thresholds
  - Population size
  - Homogeneity of accommodation type and tenure
  - Shape

UNIVERSITY OF Southampton
School of Geography

2001 Output Areas (E&W)

- c. 175,000 output areas
- Mean 297 persons, 123 households
- Building blocks for "neighbourhood" geographies: Super Output Areas (LSOAs, MSOAs)
- Freely available digital boundary data

UNIVERSITY OF Southampton
School of Geography

Changes since 2001

- Population growth, especially migration
- More and smaller households
- Newly built properties
  - Greenfield/new land
  - Brownfield/in-filling
- Sub-division of existing properties
- Changing socio-economic characteristics of areas


© Samantha Cockings


BARRATT
BARRATT
BOVIS HOMES
BOVIS HOMES
BOVIS HOMES
Controlled ZONE
Mon - Sat
8 am - 8 pm
© Samantha Cockings


thecrescentapartments
www.crescentsouthampton.com
© Samantha Cockings


FOR SALE
© Samantha Cockings

UNIVERSITY OF Southampton
School of Geography

How much change by 2011?

- 2001-2005 threshold breaches, based on mid-year population estimates
- Output Areas:

	2005 below	2005 within	2005 above	2001 totals
2001 below	221	228	1	450
2001 within	147	173553	682	174382
2001 above	0	78	506	584
2005 totals	368	173859	1189	175416

0.1%
0.3%
0.2%
99.1%
0.7%

OAs: lower threshold = 100 people; upper threshold = 625 people (2 * target)
Population thresholds = 2.5 * household thresholds

University of Southampton
School of Geography

How much change by 2011?

- Lower Layer Super Output Areas:

	2005 below	2005 within	2005 above	2001 totals
2001 below	6	8	0	14
2001 within	34	34242	58	34334
2001 above	0	3	27	30
2005 totals	40	34253	85	34378

99.6%

LSOAs: lower threshold = 1000 people; upper threshold = 3000 (2 * target)
Population thresholds = 2.5 * household thresholds

University of Southampton
School of Geography

How much change by 2011?

- Middle Layer Super Output Areas:

	2005 below	2005 within	2005 above	2001 totals
2001 below	3	4	0	7
2001 within	8	7178	0	7186
2001 above	0	0	1	1
2005 totals	11	7182	1	7194

99.8%

MSOAs: lower threshold = 5000 people; upper threshold = 15000 (2 * target)
Population thresholds = 2.5 * household thresholds

UNIVERSITY OF Southampton
School of Geography

Key messages

- Most output areas (and LSOAs, MSOAs) unlikely to have breached thresholds by 2011
- BUT, changes clustered geographically, so could breach badly in some areas
- Some areas already known to be problematic in 2001

UNIVERSITY OF Southampton
School of Geography

What do we want from 2011 geogs?

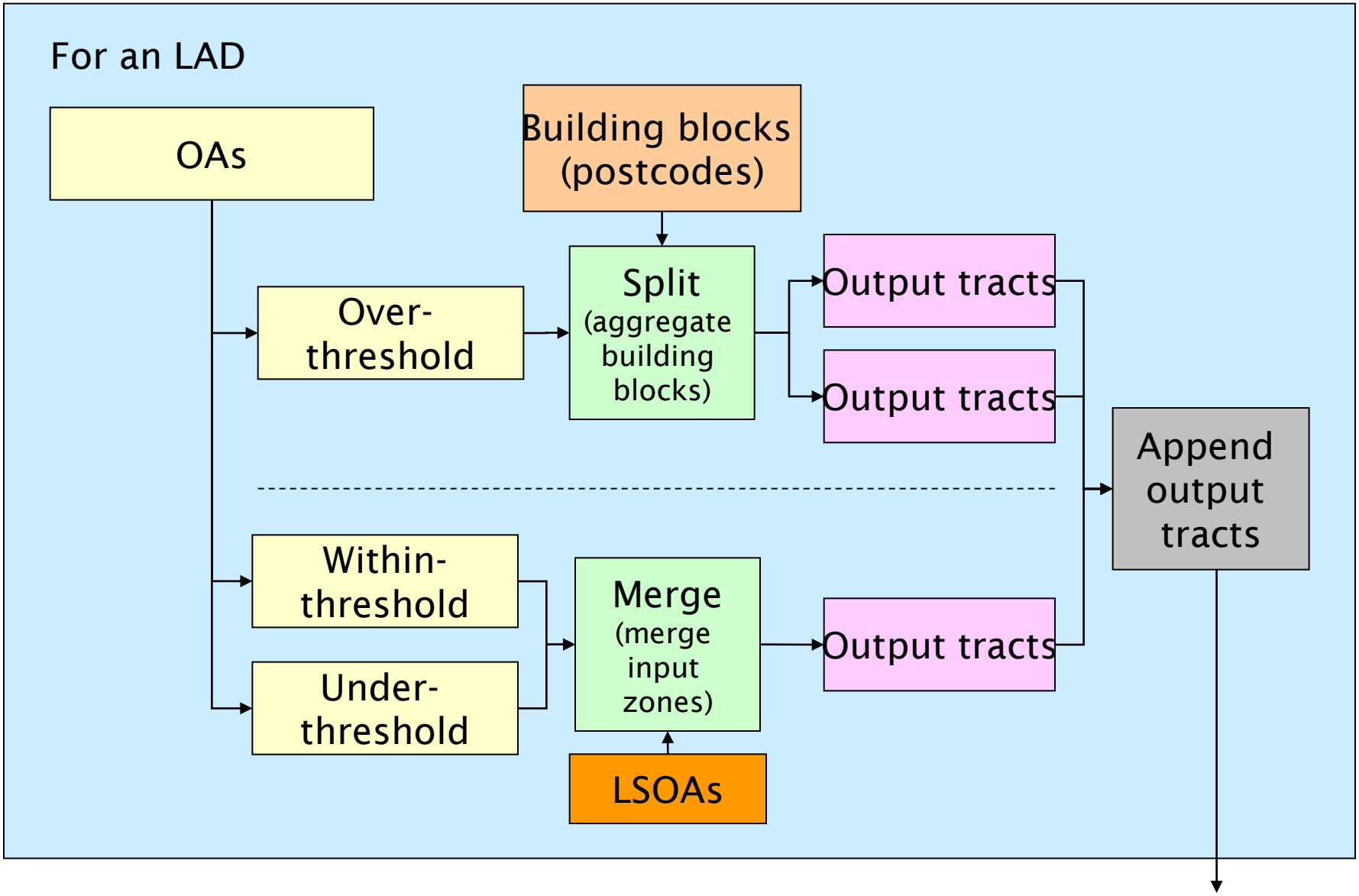
- Stable … but reflecting change … please!
- National Statistics Geography Consultations (2007, 2011)
  - Policy of minimal change where possible
  - Maintain 1:1 links where possible
  - Fix known problems e.g. Manchester, Westminster?
  - Consider “under-performing” OAs, LSOAs, MSOAs
  - Retain postcodes v use street blocks?
  - Better alignment with real-world features?
  - Alignment with wards/parishes?
  - Boundaries to remain free to end users where possible

UNIVERSITY OF Southampton
School of Geography

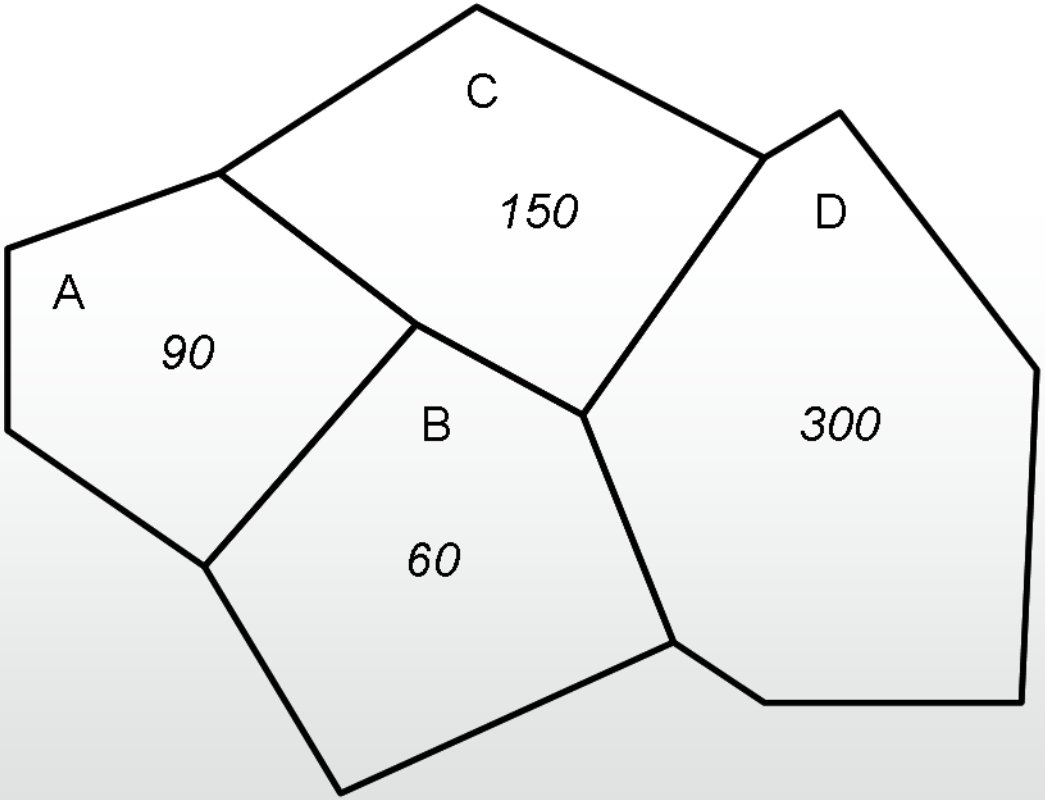
Census2011Geog project

- ESRC (2008-2010)
- Develop automated procedures for maintaining (splitting, merging, re-designing) the 2001 census output geographies in order to create 2011 output geographies for E&W
- Assess the implications of using different building blocks (e.g. postcodes, street blocks) for these maintenance procedures

Maintenance process: OAs

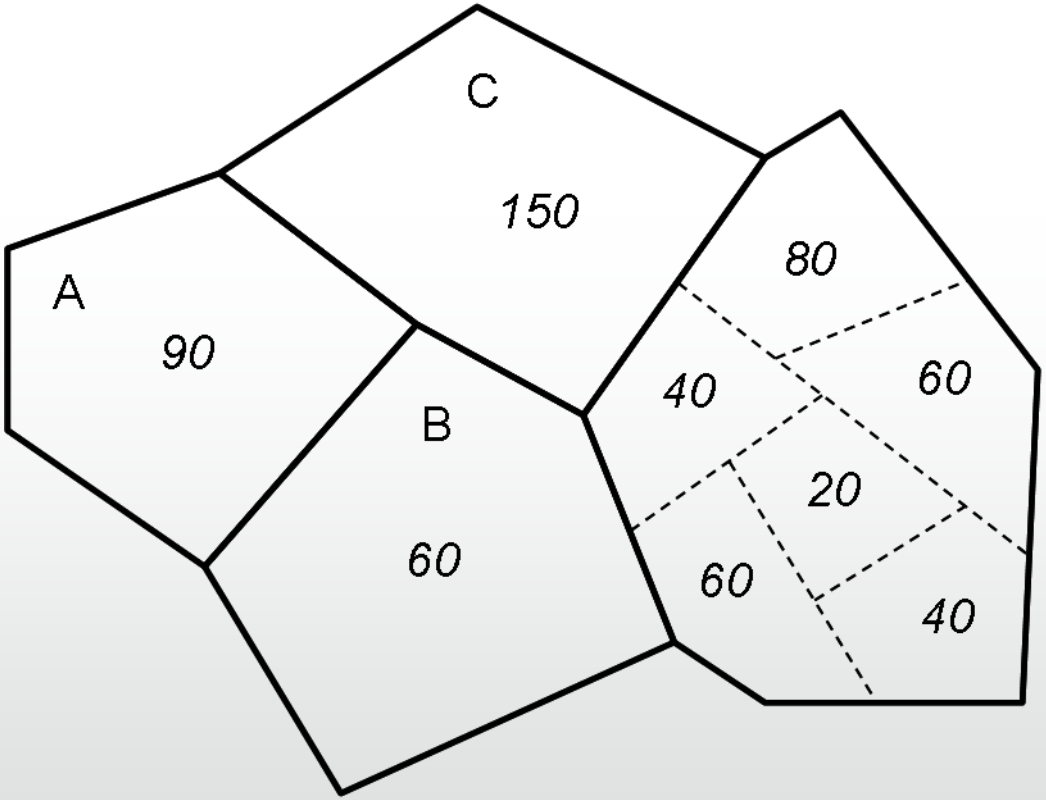


A simple example (1)



Lower threshold = 100
Upper threshold = 250

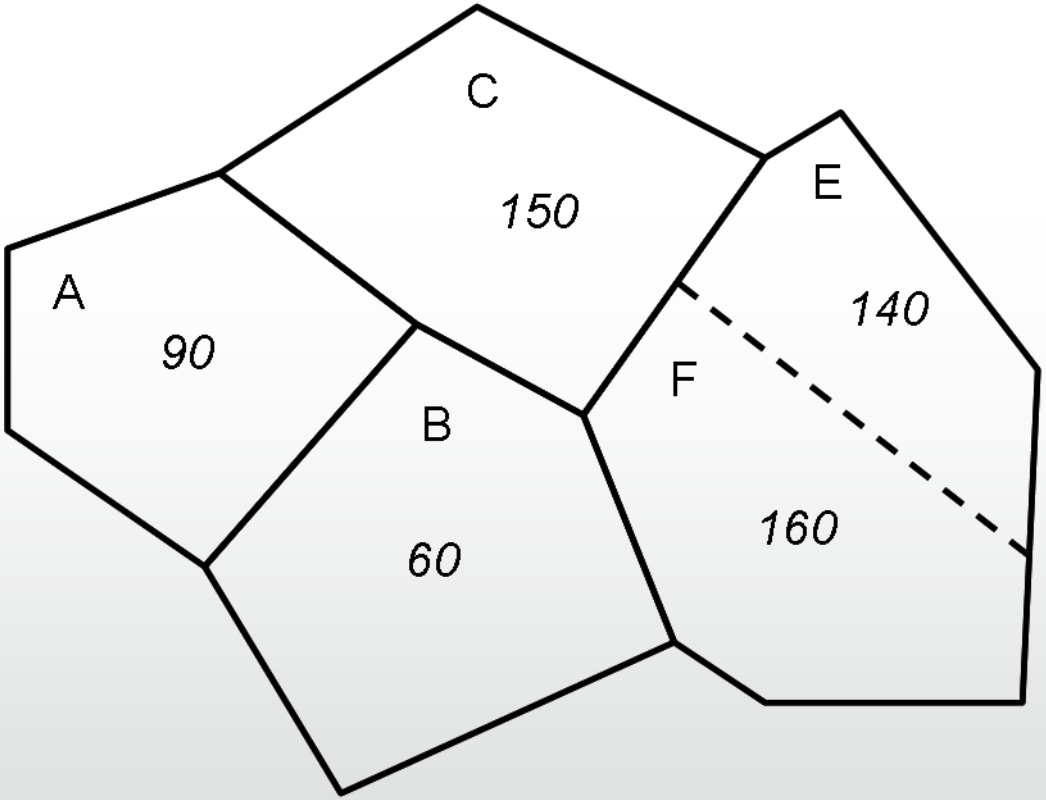
A simple example (2)



Lower threshold = 100
Upper threshold = 250

Create sub-threshold building blocks within over-threshold zone(s)

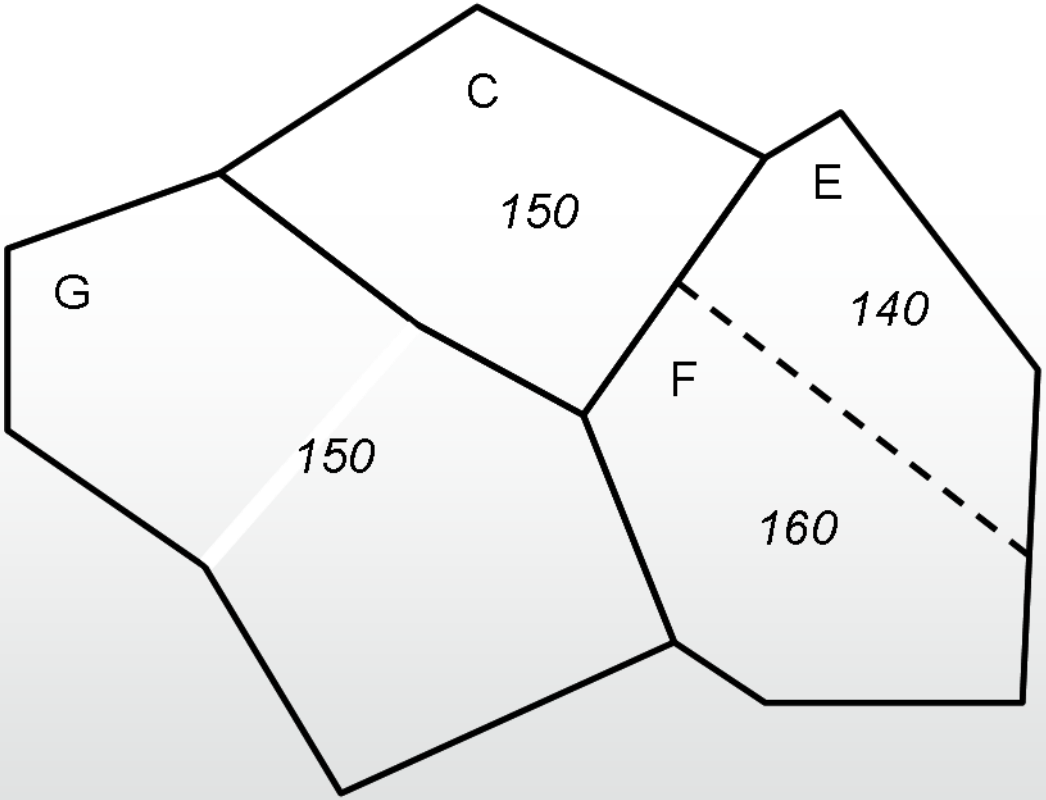
A simple example (3)



Lower threshold = 100
Upper threshold = 250

Aggregate building blocks into within-threshold tracts, optimising criteria (population, shape)

A simple example (4)



Lower threshold = 100
Upper threshold = 250

Merge under- or within-threshold zones into within-threshold tracts, optimising criteria (population, shape)

Methods

- 6 study areas (LADs of Camden, Isle of Anglesey, Lancaster, Liverpool, Manchester, Southampton)
- Create household level dataset for 2007/08: population, tenure, accommodation type
- Identify over-, within-, over-threshold OA/LSOA/MSOAs
- Create building blocks (postcodes) for over-threshold zones
- Use AZTool to split or merge zones as appropriate
- Evaluate effectiveness of maintenance procedures

Thresholds

	Population thresholds		*Household thresholds*	
Geography	*Lower*	*Upper*	*Lower*	*Upper*
OA	100	625	40	250
LSOA	1,000	3,000	400	1,200
MSOA	5,000	15,000	2,000	6,000

UNIVERSITY OF Southampton
School of Geography

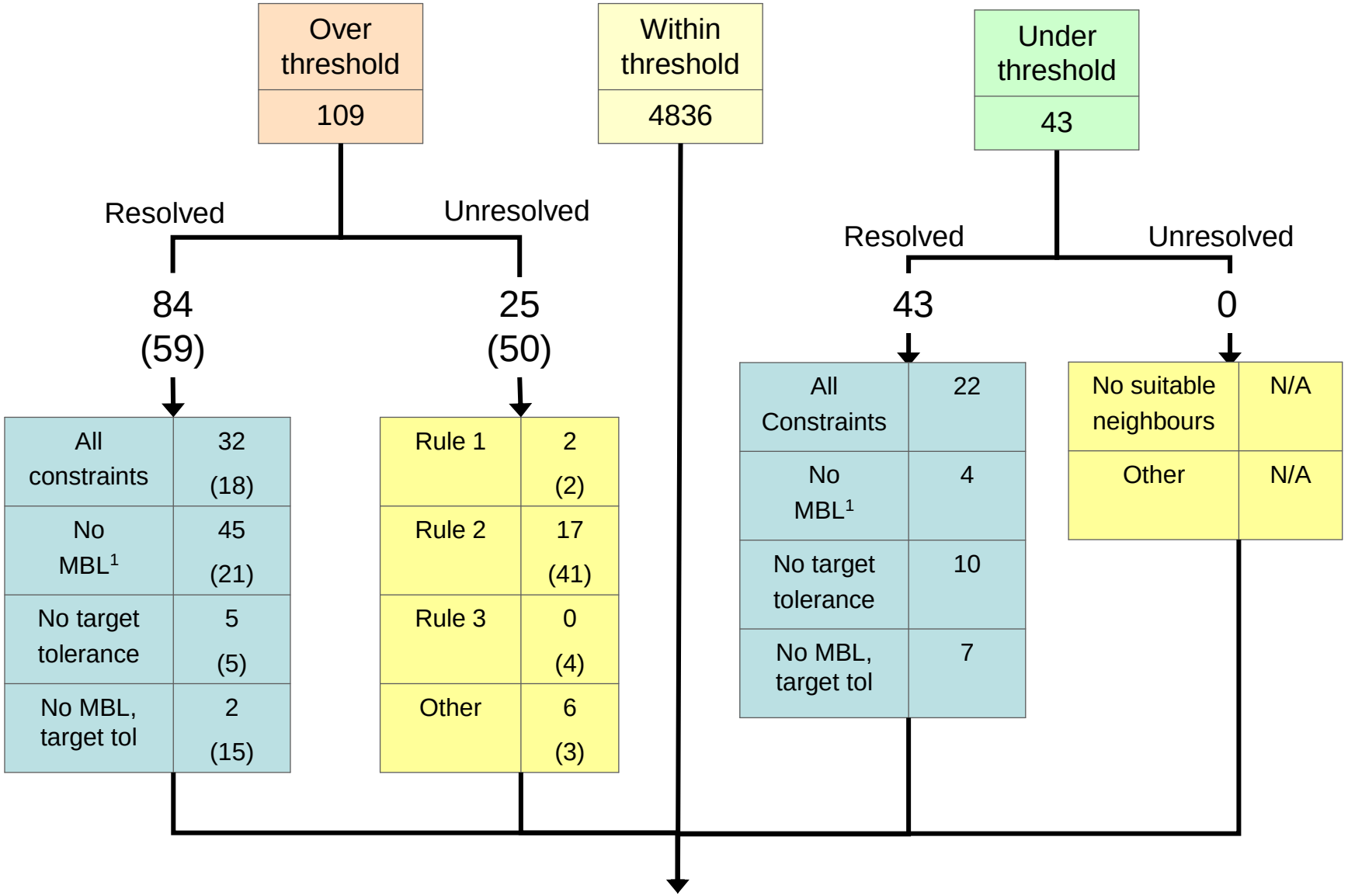
Design criteria

Constraint/criteria	*Details*	*Weighting*
Thresholds	As per Table 2	N/A
Target (households)	OA: 125; LSOA: 600; MSOA: 3,000	100
Homogeneity	IAC accommodation type and tenure	100
Shape	$Perimeter^2/Area$	100
Min boundary length	10% of total perimeter of shared boundaries	N/A
Regional constraint	Respect higher-level geogs (e.g. LSOA, MSOA)	N/A

Example: breached areas, Liverpool

Example: maintained zones, Liverpool

Results: resolved/unresolved zones



Results: statistical quality of zones

	Count	*Population*		*Households*		*Homogeneity*		*Shape*
		Mean	*Stdev*	*Mean*	*Stdev*	*Tenure*	*Acc type*	
2001	4988	290	102	125	16	0.18	0.29	37.83
2007	4988	315	141	128	44	0.16	0.26	37.83
Maintained	5074	309	129	126	30	0.16	0.26	37.79

Conclusions – OA maintenance

- Automated zone design (AZTool) can be used to maintain (split, merge, re-design) a set (or sub-set) of existing zones
- A viable alternative to completely re-designing zones or retaining all original zones
- Enables policy of stability, with change where needed
- Maintenance process more constrained than when designing from scratch: more manual intervention required
- Statistical qualities of maintained zones often not as good as original zones, but better than pre-maintenance ones

UNIVERSITY OF Southampton
School of Geography

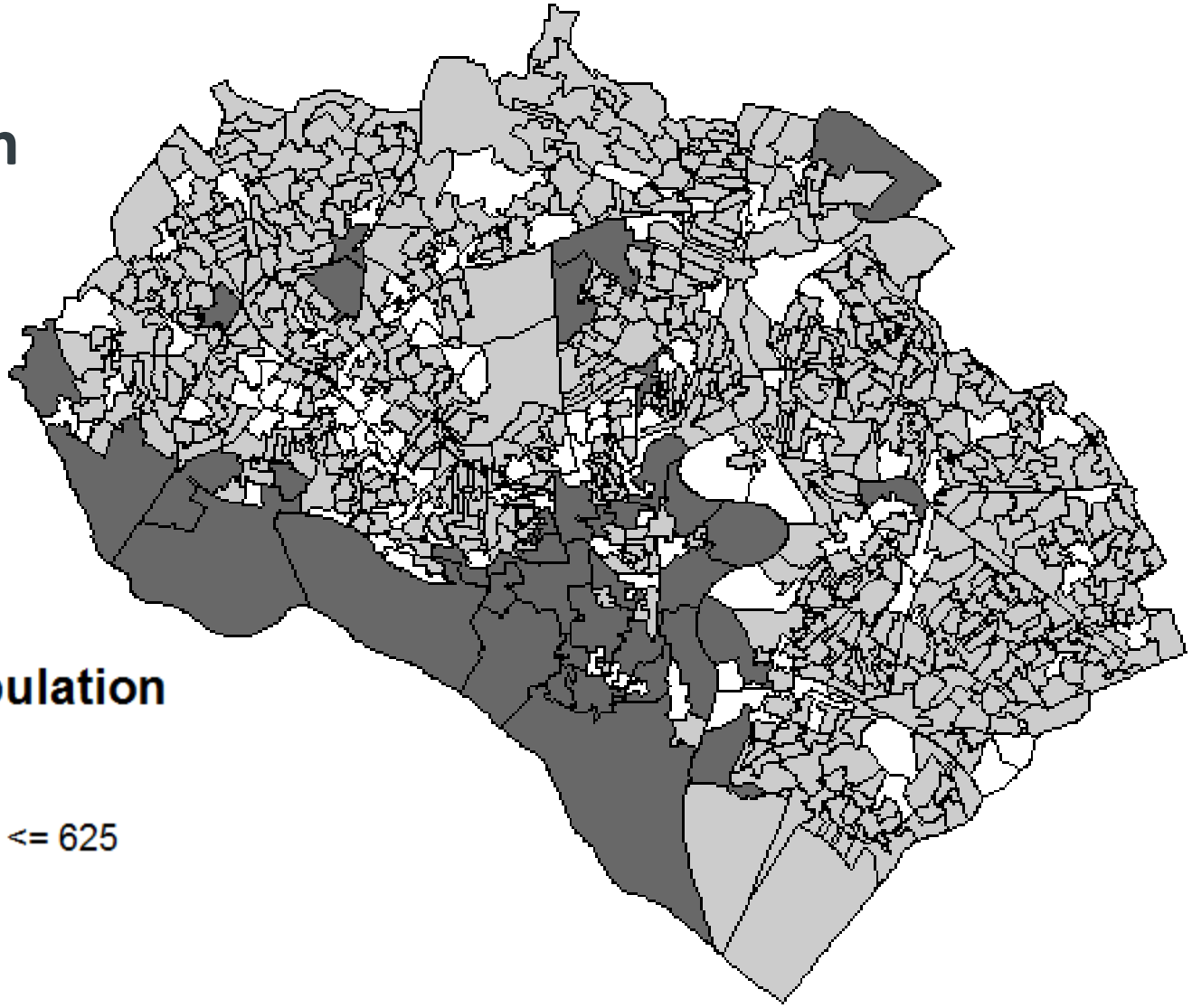
Applications: Workplace zones

The need for Workplace Zones (WZ)

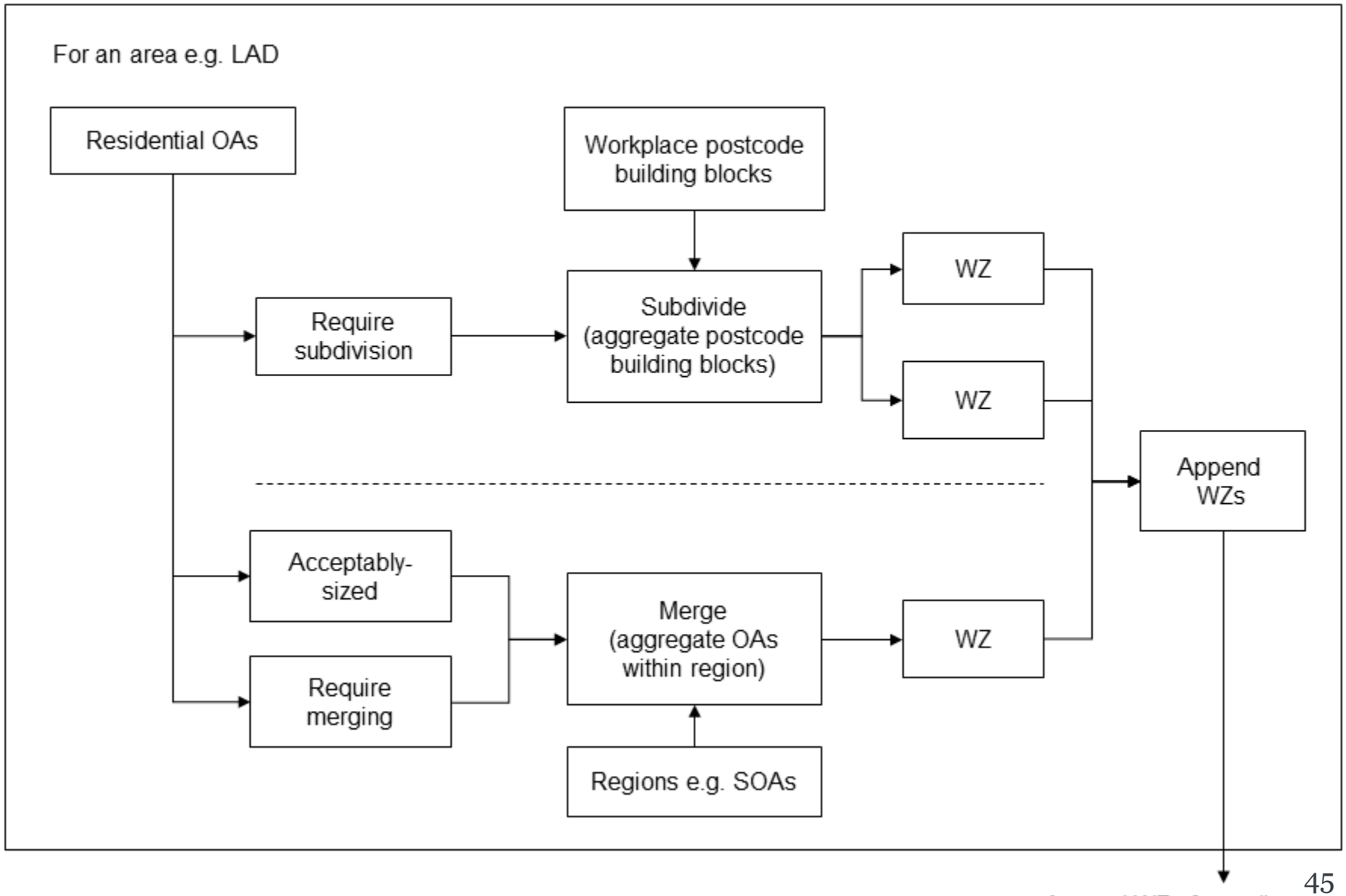
- Most output geographies are designed to reflect spatial distribution of residents and dwellings
- Distribution of workplace population/workplaces is very different to residents/dwellings
- Limits usefulness of published business-related data
  - Too many workplaces/workplace population in OAs in city centres, industrial/retail zones etc
  - Too few in residential suburban, rural areas etc
- Also restricts amount of data (variables) which can be released (statistical disclosure concerns)

The geographic nature of the problem …

Southampton

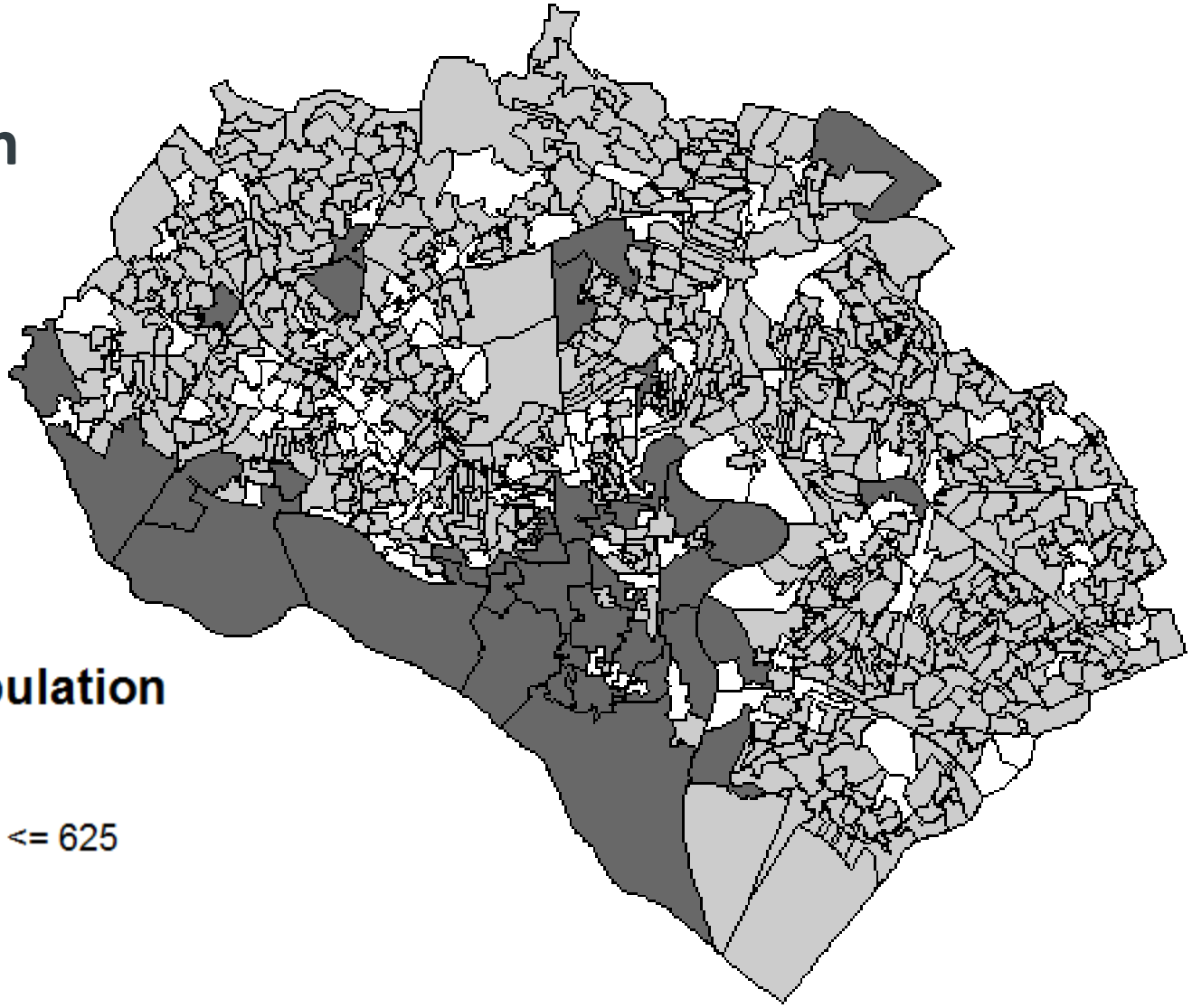


Proposed process



The geographic nature of the problem …

Southampton



Initial results

Southampton

Results: original (residential) OAs

		Workplace populations				Industry	Shape
	Count	Min	Max	Mean	SD	IAC	Mean
All study areas	2737	0	79956	292	2279	0.21	37.11
City of London	36	8	79956	8654	16326	0.02	32.54
Tower Hamlets	627	0	24257	251	1323	0.13	34.03
Nottingham	929	0	16787	185	879	0.17	38.09
Southampton	730	3	7550	152	539	0.22	37.44
Suffolk Coastal	415	2	3768	115	288	0.24	39.41

Results: new WZs

		Workplace populations				Industry	Shape
	Count	Min	Max	Mean	SD	IAC	Mean
All study areas	1164	6	15771	687	947	0.27	30.72
City of London	336	13	8557	930	957	0.19	21.25
Tower Hamlets	210	61	15771	743	1497	0.16	33.47
Nottingham	308	38	6716	559	661	0.23	32.38
Southampton	205	6	6132	541	606	0.25	36.57
Suffolk Coastal	105	121	3768	457	433	0.21	39.22
From splits	645	200	15771	927	1183	0.29	24.85
From mergers	452	200	1340	386	162	0.10	38.90

UNIVERSITY OF Southampton
School of Geography

Observations

- Patterns as expected: e.g. city centres get increase in number of zones; suburban residential areas see decrease
- Isolated sub-threshold OAs and densely populated postcode BBs result in some WZs remaining outwith thresholds
- Number of within-threshold WZs increases and population range decreases, at expense of increase in mean pop per zone
- WZs created by splits and mergers exhibit different statistical characteristics
- City of London and Tower Hamlets outliers in both study area and national context – dense concentration of workers in tower blocks: Canary Wharf, very few residents: City

Conclusions - WZs

- AZD techniques (AZTool) can be employed to create completely new output geographies from existing ones
- Novel approach which overcomes key disclosure control concerns by optimising output geography, thus maximising range and amount of data suitable for publication
- Key issues are now disclosure related rather than technical
- Methods and findings being evaluated by ONS for possible implementation following 2011 Census
- Generic approach, methods and software extensible to other countries and other applications

UNIVERSITY OF Southampton
School of Geography

Future directions

- Fine-tuning of AZTool: enhance GUI, add extra functions, alternative metrics/scores?
- Application within other countries
- Application to other problems
  - Political re-districting (controversial!)
  - Health
  - Collection geographies/workloads
  - Intervention/policy zones
- Beyond 2011/Censuses: flexible zonations

University of Southampton
School of Geography

Websites

- Census2011Geog project
  - http://census2011geog.census.ac.uk/
- AZTool software
  - http://www.geodata.soton.ac.uk/software/AZTool/
- ONS > 2011 Census > Output geography
  - http://www.ons.gov.uk/census/2011-census/produce-deliver-data/output-geog/index.html

Key references

- Ang L, Ralphs M, 2008, "Operations Research for New Geographies: Zone Design Tools for Census Output Geographies", Methodology Development Unit, Standards and Methods Group, Statistics New Zealand, Wellington, New Zealand, pp 40
- Cockings S, Harfoot A, Hornby D, 2009, "Towards 2011 output geographies: Exploring the need for, and challenges involved in, maintenance of the 2001 output geographies" *Population Trends* **138** 38 – 49
- Cockings, Harfoot A, Martin D and Hornby D 'Maintaining existing zoning systems using automated zone design techniques: methods for creating the 2011 Census output geographies for England and Wales', manuscript submitted to *Environment & Planning A*
- Cockings S, Martin D, 2005, "Zone design for environment and health studies using pre-aggregated data" *Social Science & Medicine* ***60*** 2729 – 2742
- Martin D, 2003, "Extending the automated zoning procedure to reconcile incompatible zoning systems" *International Journal of Geographic Information Science* **17** 181 - 196
- Martin D, Nolan A, Tranmer M, 2001, "The application of zone design methodology to the 2001 UK Census" *Environment & Planning A* **33** 1949 – 1962
- Openshaw S, 1977a, "A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modeling" *Transactions of the Institute of British Geographers, New Series* **2** 459 - 472
- Openshaw S, 1977b, "Algorithm 3: a procedure to generate pseudo-random aggregations of *N* zones into *M* zones, where *M* is less than *N*" *Environment & Planning A* **9** 1423 - 1428
- Openshaw S, Rao L, 1995 "Algorithms for re-engineering 1991 Census geography" *Environment & Planning A* **27** 425 - 446

University of Southampton
School of Geography

Acknowledgements and copyright

- Census2011Geog project funded by ESRC award number RES-348-25-0019

- Data*: MasterMap Address Layer 2 (March 2008):* Crown Copyright Ordnance Survey. Used under ONS PGA license GD272183 2009; *Household-level 2001 Census data, Postcode level mid-year population estimates and special population listings (mid-2007):* Access granted by ONS MicroData Release Panel, data accessed as Approved Researchers under secure setting conditions at ONS Titchfield; *MasterMap Integrated Transport Network Layer (December 2007), MasterMap Topography Layer (December 2007), Ordnance Survey Meridian 2 (October 2008):* Crown Copyright/database right 2009. An Ordnance Survey/EDINA supplied service; *2001 Census Output Area, Lower Layer Super Output Area, Middle Layer Super Output Area, Local Authority District boundaries*. Crown copyright 2003. Crown copyright material is reproduced with the permission of the Controller of HMSO. Data provided through EDINA UKBORDERS with the support of the ESRC and JISC; *National Statistics Postcode Directory (February 2008)*: Crown Copyright 2006. Source: National Statistics / Ordnance Survey. Extracts are Crown Copyright and may only be reproduced by permission. Data provided through EDINA UKBORDERS with the support of the ESRC and JISC; *Universities UK Student Residences List (March 2009)*: obtained from www.universitiesuk.ac.uk.

- WZ research funded by ONS, DfT; all data from ONS.

UNIVERSITY OF Southampton
School of Geography

Questions, discussion.

UNIVERSITY OF
Southampton
School of Geography

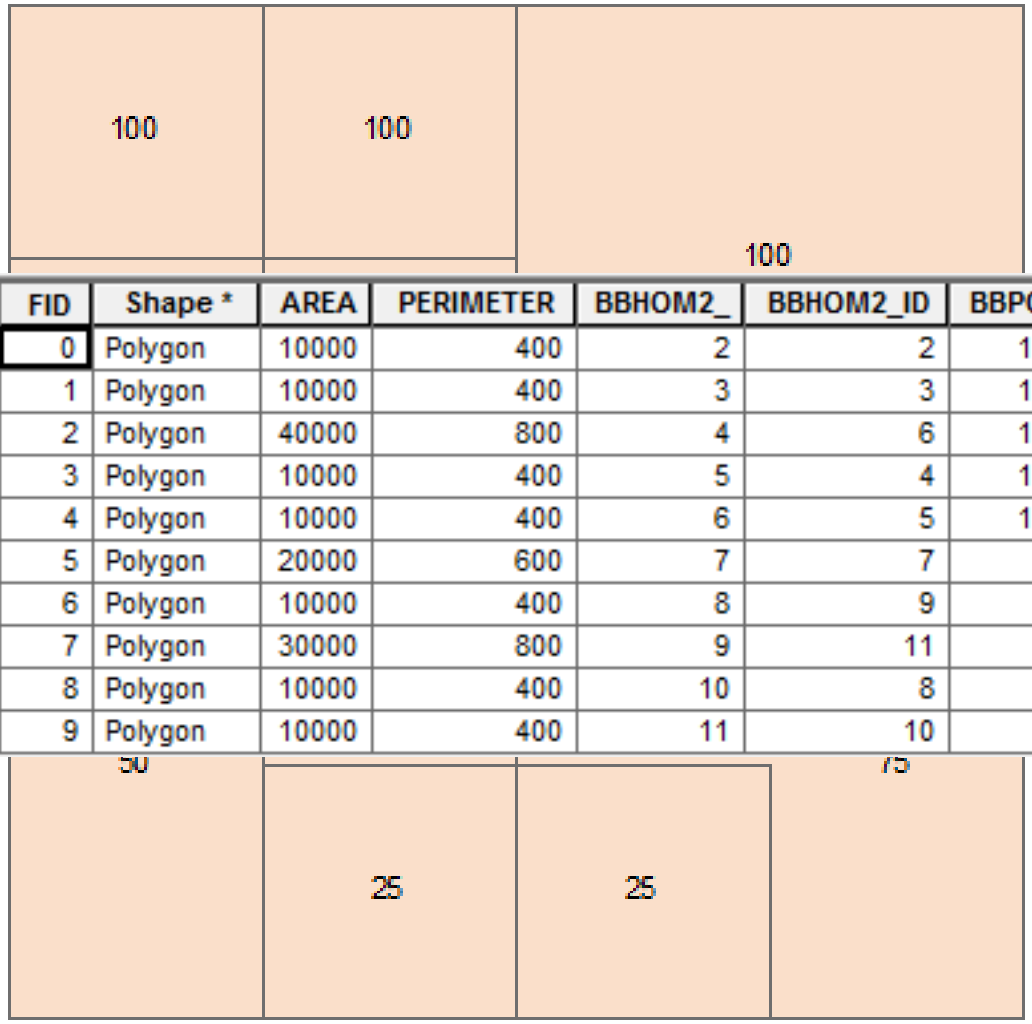
Automated zone design: AZTool software

Typical zone design process

University of Southampton
School of Geography

- Get data into input format for AZTool
  - Run AZTImporter
- Specify design constraints
  - Edit .xml parameter file
- Run AZTool
  - GUI or batch mode
- Inspect outputs and create new tracts
  - Log and tract composition files
  - Dissolve input zones on tract ID

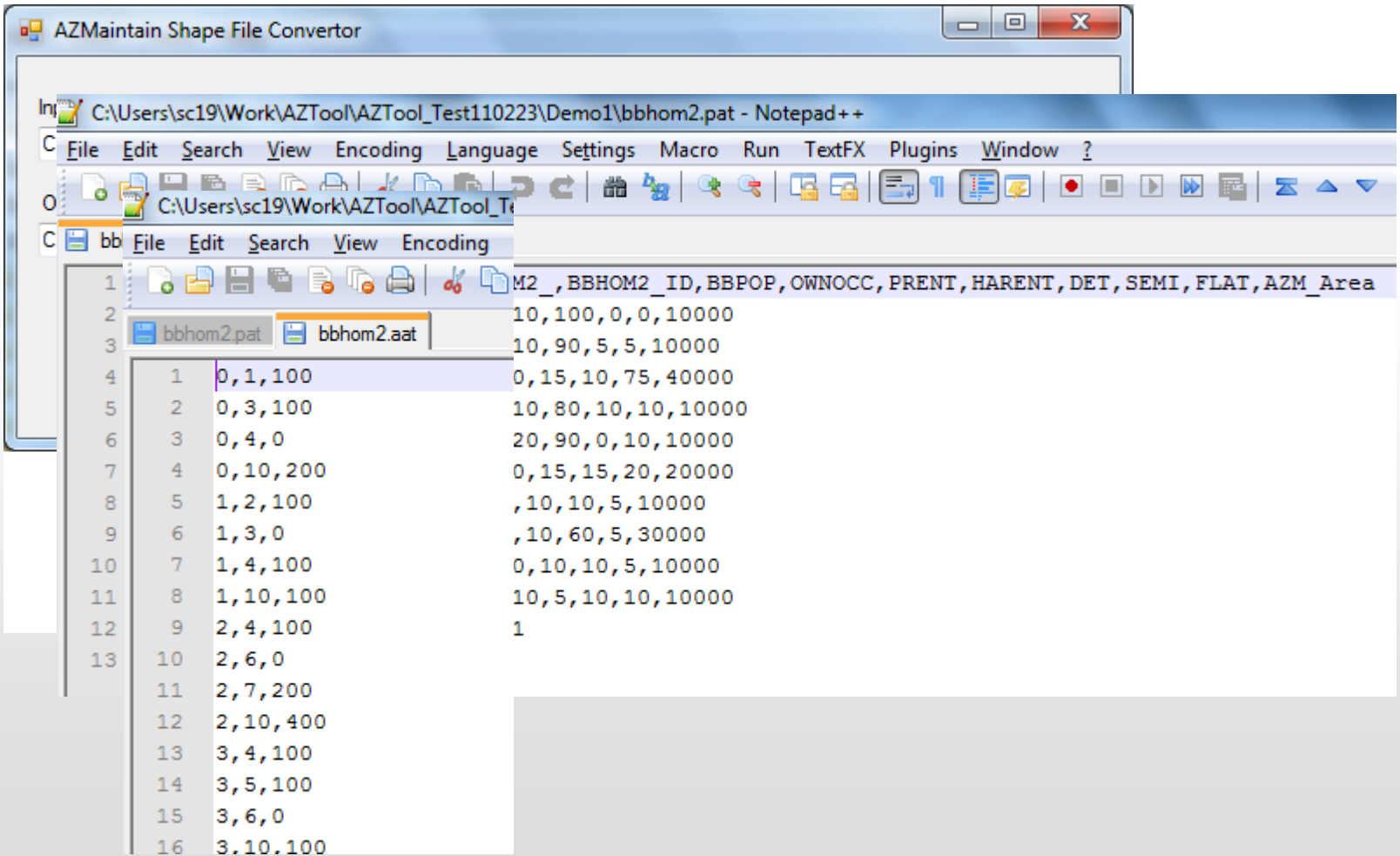
Original zones and attributes



FID	Shape *	AREA	PERIMETER	BBHOM2_	BBHOM2_ID	BBPOP	OWNOCC	PRENT	HARENT	DET	SEMI	FLAT
0	Polygon	10000	400	2	2	100	80	10	10	100	0	0
1	Polygon	10000	400	3	3	100	78	12	10	90	5	5
2	Polygon	40000	800	4	6	100	0	100	0	15	10	75
3	Polygon	10000	400	5	4	100	75	15	10	80	10	10
4	Polygon	10000	400	6	5	100	70	10	20	90	0	10
5	Polygon	20000	600	7	7	50	20	20	10	15	15	20
6	Polygon	10000	400	8	9	25	10	10	5	10	10	5
7	Polygon	30000	800	9	11	75	0	75	0	10	60	5
8	Polygon	10000	400	10	8	25	10	5	10	10	10	5
9	Polygon	10000	400	11	10	25	10	5	10	5	10	10

Create input data: .pat and .aat file

- Run AZTImporter if have shapefile



University of Southampton
School of Geography

Specify design criteria: .xml parameter file

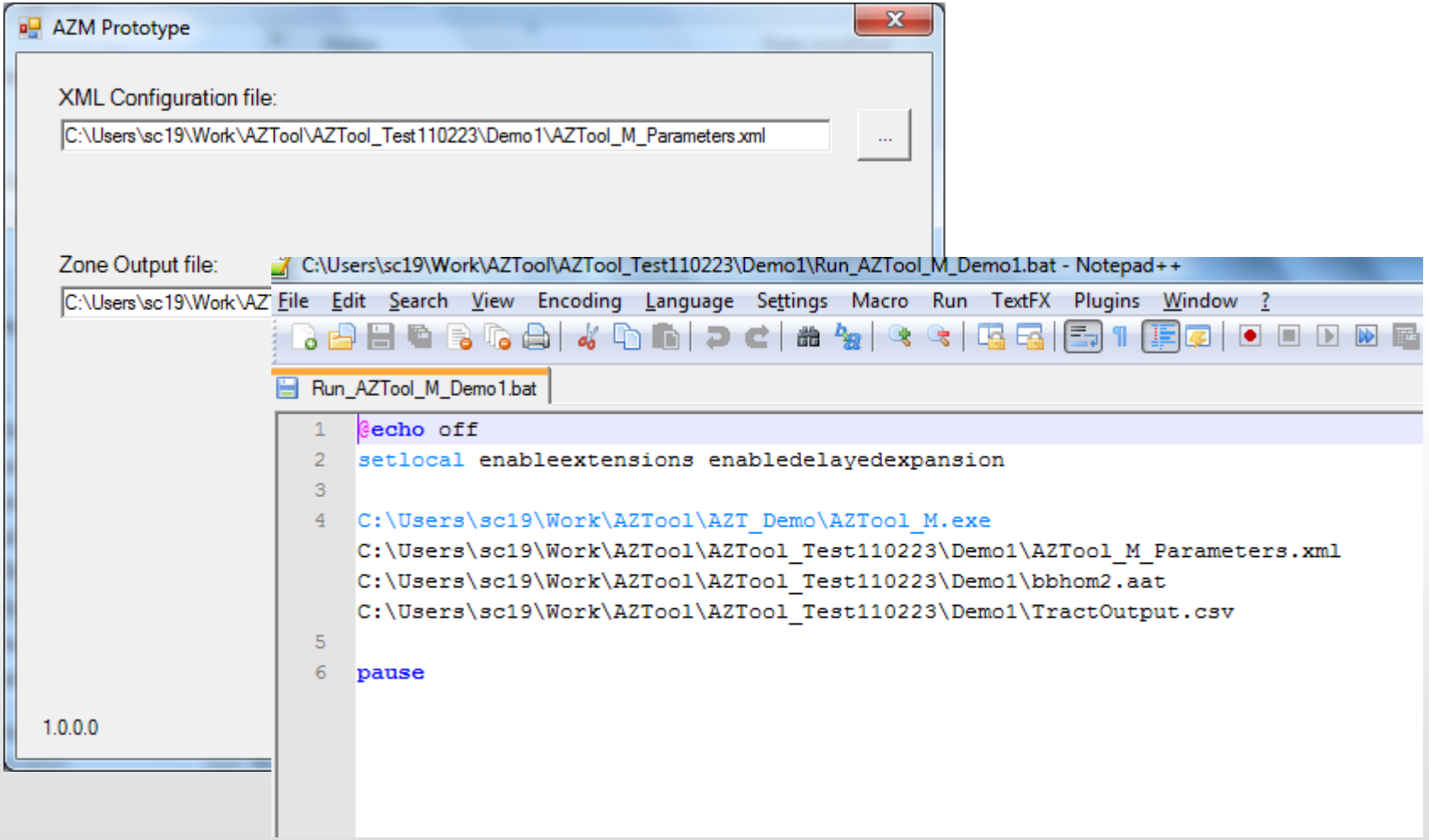
C:\Users\sc19\Work\AZTool\AZTool_Test110223\Demo1\AZTool_M_Parameters.xml - Notepad++

File Edit Search View Encoding Language Settings Macro Run TextFX Plugins Window ?

AZTool_M_Parameters.xml

```
1  <?xml version="1.0" encoding="utf-8"?>
2  <ProgramOptions xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xsd="http://www.
3    <InputPATFile>C:\Users\sc19\Work\AZTool\AZTool_Test110223\Demo1\bbhom2.pat</InputPATFile>
4    <InputAATFile>C:\Users\sc19\Work\AZTool\AZTool_Test110223\Demo1\bbhom2.aat</InputAATFile>
5    <Header>true</Header>
6    <IDIndex>1</IDIndex>
7    <RegionIndex>0</RegionIndex>
8    <RegionToUse>ALL</RegionToUse>
9    <RespectRegions>false</RespectRegions>
10   <TargThreshVars>
11     <TargetThresholdVar>
12       <Name>Population</Name>
13       <FileIndex>6</FileIndex>
14       <TargetSet>true</TargetSet>
15       <Target>300</Target>
```

Run AZTool (simple GUI or batch mode)



Outputs: log and tract composition files

TractOutput.log

```
1  XML Config file: C:\Users\sc19\Work\AZTool\AZTool_Test110223\Demo1\AZTool_M_Parameters.xml
2  PAT (AAT) file: C:\Users\sc19\Work\AZTool\AZTool_Test110223\Demo1\bbhom2.pat
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
```

C:\Users\sc19\Work\AZTool\AZTool_Test110223\Dem

File Edit Search View Encoding Language Set

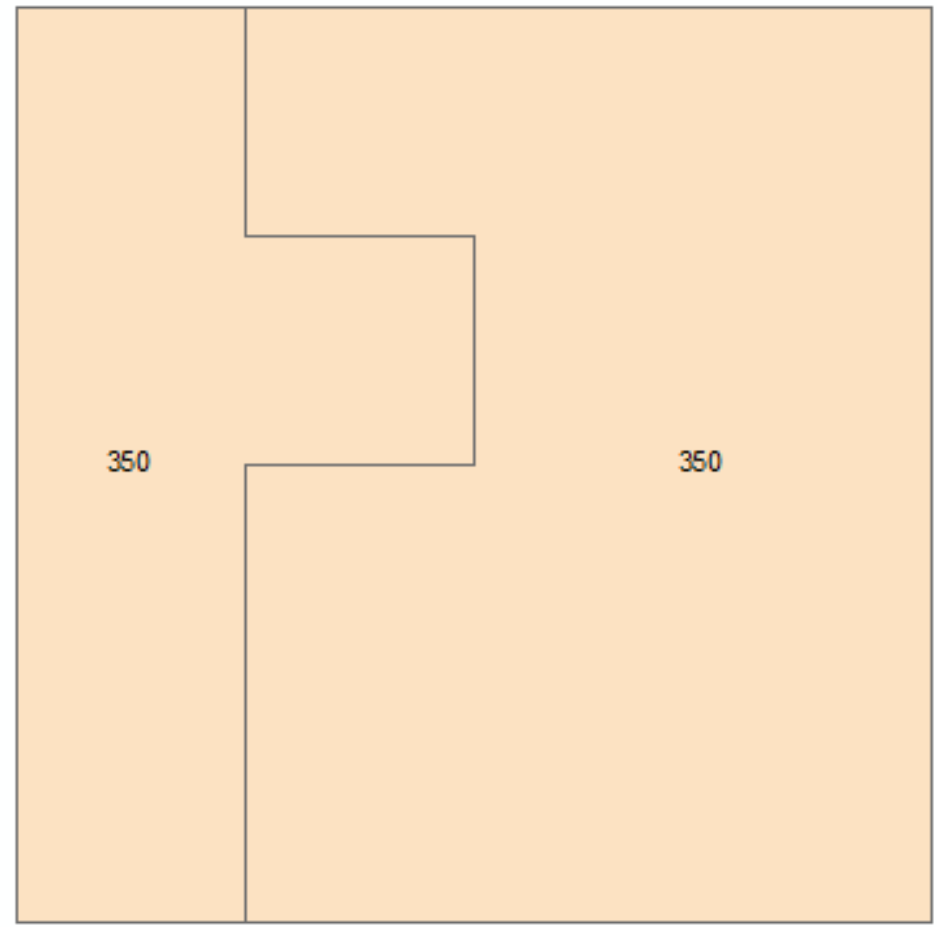
ol_Test110223\Demo1\TractOutput.csv

TractOutput.csv

```
1   BldBlID,TractID
2   8,2
3   9,2
4   2,2
5   7,2
6   1,2
7   6,2
8   0,3
9   3,3
10  4,3
11  5,3
12
```

count for target mean

Create tracts: aggregate BBs (dissolve on tract ID)



UNIVERSITY OF Southampton
School of Geography

s.cockings@soton.ac.uk