

How to Rezone

A method for redistributing census populations
to create bespoke geographies

January 2014



Contents

Contents	2
Introduction	3
Limitations of this method	4
Extract the zip file	6
Part 1: Select Postcodes.....	7
Part 2: Create Gazetteer File	9
Part 3: Rezone Census tables using Gazetteer File	11
Notes	12

Introduction

Census data are published by the Office for National Statistics (ONS) in England & Wales and National Records for Scotland (NRS) in Scotland. These data are published at a number of geographies from the national level right down to output area. Output areas themselves are built from postcodes and typically contain around 125 households in England and Wales but are somewhat smaller in Scotland.

Rezoning is necessary when data are required for non-standard geographies; areas for which ONS/NRS do not publish data. In some cases these areas can be created quite simply by aggregating existing geographies. In other cases the boundary of the area in question does not align with any existing geography. In these cases an output area, may sit partly within and partly without the new zone and decisions must be made of how best to distribute data to the new area. There are a number of ways to approach this task and this document outlines a method using postcode population estimates.

Methodology Outline

The first stage of the process is to identify those postcodes which fall within the new zone. This is achieved using GIS (Geographic Information System) and the result should be a list of postcodes in CSV format.

The postcode CSV file is used in SASPAC along with a specially designed System File to produce a list of Output Areas and the proportion of the population of each that falls within the new zone (a Gazetteer file).

The Gazetteer file is then used in SASPAC to rezone the census data so that statistics for the new zone are available.

Process



Limitations of this method

Inaccurate allocation of populations

This method distributes all variables in the same way: An output area that which is found to lie half inside the new zone and half outside will have each variable halved. The issue here is that variables are likely to be distributed in a variety of ways within an output area and so a characteristic may be attributed to the new zone when no one with that characteristic resides within the boundary.

An example of this is as follows. An output area contains 300 people. It is bisected by the boundary of a new zone so that half of the population is within the new zone and half is outside. Of the total population 200 are female and 100 are male. All of the men live in the part of the output area within the new zone while the female population is split so that 50 live within the new zone and 150 live outside. The results of the rezone are shown below alongside the actual population of the area.

	Actual population of area	Rezoned population of area
Males	100	50
Females	50	100

Because the rezone distributes half of the population of each variable the result is inaccurate when compared to the actual population of the area.

Inconsistent Totals

A second issue is inconsistencies in totals which arises from the rounding of data in SASPAC and so is not specific to this method of rezoning. Census data deals in whole integers, each number representing a member of the population or a household. This means that it is necessary to round, both up and down, when a rezoning task takes place. As each number is rezoned and rounded in isolation of all others this can mean that variables within a table no longer add up to the totals they should.

An example is an output area that has 99 males and 201 females giving a total population of 300. Again, the output area is divided into two. The total population in the new zone is 150. The number of men calculated is 49.5 which is rounded to 50. The calculated number of women is 100.5, rounded to 101. As a result the calculation 'males' plus 'females' equals 151, not 150. This same process is shown in the table below.

	Total population	Half of population	Rezoned population (rounded)
Males	99	49.5	50
Females	201	100.5	101
Total	300	150	150

Mitigation

There are two factors which mitigate this. Firstly, inaccuracies are unlikely to be as significant as those shown in the above example, populations will tend to be distributed more evenly within an output area than is the case in this example.

Secondly, the original base census data is not intended to be an accurate count of a population but rather an estimate of the characteristics. The underlying uncertainty in the estimates means that problems of distribution are less significant than they may appear. The only way to truly mitigate this issue would be to rezone each variable independently but this is not possible with the information available and would be massively time consuming to attempt if the data were available.

Note on Detailed Characteristics

This method uses Output Areas as its base unit. Detailed Characteristics data are not available at output area level and so rezoning using this method is not possible. It would therefore be necessary to use MSOA data rather than OA. The MSOA geography is relatively large (containing an average of 7,200 people) and so issues of accuracy and data distribution would be magnified. As a result the rezoning of Detailed Characteristics using this method is not recommended.

Download Rezone.zip

If you have not already downloaded the Rezone.zip file click here:

<http://data.london.gov.uk/datastorefiles/datafiles/demographics/2011-census/Rezone.zip>

The rezone.zip file (60Mb) file contains the files and folders that you will need to undertake the rezoning task.

If clicking the link fails to download the zip file copy the link into your web browser directly.

Extract the zip file

Next, extract the contents of Rezone.zip.

1. Right click on the file Rezone.zip
2. Click Extract All
3. Extract the file to C:\SASPAC
4. Click Next

Note: This is necessary to ensure that the file paths referred to in these instructions are correct.

5. A folder named Rezone has been created in your SASPAC folder.
6. The following folders are now in the Rezone folder:
 - Shape files
 - Tab files
 - Scaling Factors

Part 1: Select Postcodes

The first stage of the process is to identify the postcodes that sit within the new zone using GIS. This process requires the new zone to have already been mapped as a boundary file in your GIS software.

Using ESRI Arc GIS

1. Open Arc GIS and a blank map
2. Add the boundary of the new zone to the table of contents
3. Add the postcode point data for your region.
(C:\SASPAC\Rezone\Shape Files\[region name].shp)

NOTE: It is necessary to use the Shape files provided as the zone identifiers in these shape files are consistent with those in the SASPAC system file that will be used in stage 2 of this process. For further information see end note on postcodes.ⁱ

4. Select those postcodes that lie within the new zone boundary
 - a. Go to Selection
 - b. Select by location
 - c. Selection method: select features from
 - d. Target layer: [region name] i.e. postcode data
 - e. Source layer: [name of your shape file]
 - f. Spatial selection method: Target layer(s) features are within the Source layer feature
 - g. OK
5. Export the selected postcodes to a CSV file
 - a. Right click on the postcode layer in the Table of Contents
 - b. Open Attribute Table
 - c. From the dropdown list in the top left corner choose Export
 - d. For the Export dropdown list choose: Selected records
 - e. Click on the browse button
 - f. Change the file type to Text file
 - g. Navigate to C:\SASPAC\Rezone\
 - h. Name the file Postcodes for Rezone.CSV
 - i. Click Save
 - j. Do you want to add the new table to the current map?: Click No
6. Edit the CSV file
 - a. Open the CSV file in Excel
 - b. Delete columns A C and D
 - c. Delete the first row
 - d. Save the file and exit Excel.

Using MapInfo

1. Open MapInfo and open the .tab file for the new zone in a map window using 'File|Open...'
2. Go to 'File|Open...' and open the postcode point data for your region in the Current Mapper window. (C:\SASPAC\Rezone\Tab Files\[region name].tab)

NOTE: It is necessary to use the Tab files provided as the zone identifiers in these files are consistent with those in the SASPAC system file that will be used in stage 2 of this process. For further information see end note on postcodes.ⁱⁱ

3. Select those postcodes that lie within the new zone boundary
 - a. Go to 'Query|SQL Select...'
 - b. In the 'from tables' window add the two tables currently open from the 'Tables' drop-down box (making sure to select the region table first and the new zone second)
 - c. This should automatically create the statement '[Region name].Obj Within [New zone].Obj'
 - d. In the 'Select Columns' window add the '[Region].Postcode' column from the 'Columns' dropdown box.
 - e. Now select 'OK' (with the 'Browse Results' tick box checked)
4. Export the selected postcodes to a CSV file
 - a. A new query window will appear and present a list of all the unit postcodes contained within the new zone
 - b. Select 'Table|Export...' from the main menu bar
 - c. Select the name of the query ('Query[X]') just undertaken from the list and 'Export'
 - d. Browse to C:\SASPAC\Rezone\
 - e. Name the file 'Postcodes for Rezone' and save in a '.csv' format
 - f. Click Save
 - g. Leave the File Character Set to 'Windows US & W.Europe ("ANSI") and **do not** check the 'Use First Line for Column Titles'
 - h. Select 'OK'

Part 2: Create Gazetteer File

The second stage of the process is to use the list of postcodes that has just been output from GIS to create a Gazetteer File. This is done in SASPAC.

1. Open a new Export CSV task in SASPAC
2. Input the scaling factors System File:
C:\SASPAC\Rezone\scaling factors\SCALING FACTORS.sys
3. Click Select Areas and choose the Use Mapshore tab
 - a. Click Open Zoneslist File
 - b. Select C:\SASPAC\Rezone\Postcodes_for_Rezone.CSV
 - c. Choose Add All and then Include Selected Zones
 - d. Close the Select Areas window

Note: A maximum of 4,000 zones can be input this way.

4. Choose to Output the file to C:\SASPAC\Rezone\Gazetteer file.csv
5. In the Selected Variable Name(s) box type: zlabel sfactor
6. Click OK and when prompted to include zoneid choose No, then run the task.

Note: Depending on the number of zones that you have input this process may take some time

7. Click view in Excel, delete the first row
8. Insert a column at the beginning so that your data starts in B1
9. In cell A1 type a zone id for your new area

Note: Zone ids must be alpha-numeric and contain no spaces

10. Copy the zone id down all of the used rows in the CSV file
11. Save the file and close Excel

Edit the file path

1. Navigate to C:\SASPAC\Rezone\ in Windows Explorer
2. Change the file extension of the Gazetteer file from csv to gaz:
Gazetteer file.csv to Gazetteer file.gaz

In SASPAC

1. Go to File and Open and choose Open Gazetteer File
2. Navigate to C:\SASPAC\Rezone\Gazetteer file.GAZ
3. Press Enter twice to create two blank rows
4. Return to the first row and type your zone id followed by a space and then a name for your new area.
5. Save the file and close

Part 3: Rezone Census tables using Gazetteer File

This stage of the process must be re-run for each dataset that is being rezoned (Key Statistics, Quick Statistics or Local Characteristics) – this example uses Key Statistics.

1. Open a new Create New Zones task in SASPAC
(File > New Task > Create New Zones...)
2. Choose the dataset you wish to rezone by inputting a System File (in this case Key Statistics). The geographic level of the System File MUST be output areas.
3. Click Output to... and navigate to c:\SASPAC\Rezone\
4. Save your file as 2011 Key Statistics for [name of zone].sys
5. Click save
6. In the System File Label type:
 - a. KS2011

(For Quick Statistics type QS2011, for Local characteristics type LC2011)

7. Click on Gazetteer File: 'Input from...'
 - a. Navigate to c:\SASPAC\Rezone\
 - b. Choose Gazetteer file.gaz and open

Note: The following values are dependent on the length of your zone id. For example if you gave your zone the identifier area1, then your id would be 5 characters long (x=5).

- c. Input the following values in the boxes provided

Existing Zone	(x+2)	to	(x+10)
New Zone	1	to	x
Scale Factor	x+12	to	(x+21)

- d. Tick the With Labels box
- e. Save and Close

8. Click OK and run the task

Note: If you are rezoning Key Statistics the task will complete with warnings because SASPAC will have recalculated the percentages in these tables.

9. A new System file(s) containing Census tables for your new zone(s) has now been created ready for data extraction.

Notes

ⁱ UK postcodes contain up to seven alpha-numeric characters. Some also contain embedded spaces. In order to use postcodes as zone identifiers in SASPAC these spaces must be removed. In both the GIS Shape files and the SASPAC system files the spaces have been replaced with the character 'x'. For example

Postcode	Becomes
LE3 0JE	LE3x0JE
E3 2PP	E3xx2PP
NW10 5SN	NW105SN

In addition, some postcodes were split by ONS when creating the output areas. As a result these postcodes need to have multiple entries in the system file. However, each zone must have a unique zone identifier and so where multiples exist they are identified using xb, xc, xd, etc.

For example postcode E1 6PU is split five times and so appears in the GIS Shape file and SASPAC System file five times as:

E1xx6PU
E1xx6PUxb
E1xx6PUxc
E1xx6PUxd
E1xx6PUxe

More Information

For assistance please contact the SASPAC helpdesk:

info@saspac.org
0207 983 5523

ⁱⁱ UK postcodes contain up to seven alpha-numeric characters. Some also contain embedded spaces. In order to use postcodes as zone identifiers in SASPAC these spaces must be removed. In both the GIS Shape files and the SASPAC system files the spaces have been replaced with the character 'x'. For example

Postcode	Becomes
LE3 0JE	LE3x0JE
E3 2PP	E3xx2PP
NW10 5SN	NW105SN

In addition, some postcodes were split by ONS when creating the output areas. As a result these postcodes need to have multiple entries in the system file. However, each zone must have a unique zone identifier and so where multiples exist they are identified using xb, xc, xd, etc.

For example postcode E1 6PU is split five times and so appears in the GIS Shape file and SASPAC System file five times as:

E1xx6PU
E1xx6PUxb

E1xx6PUxc
E1xx6PUxd
E1xx6PUxe

More Information

For assistance please contact the SASPAC helpdesk:
info@saspac.org
0207 983 5523