Geography Policy Best-fit and Exact-fit

2023
National Records of Scotland (NRS)

Contents

Government Statistical Service (GSS) Geography Policy

The Government Statistical Service (GSS) Geography Policy (on the ONS Geography Portal) sets out the best practice for producing official statistics with a geographic breakdown in order to ensure outputs are accurate, consistent, and comparable in their use of geography.

Pillar 4 of the policy advises that statistics for all geographies should be built from aggregations of statistical building blocks. For Scotland, the statistical building block are

- Output Areas (produced by NRS)
- Data Zones (produced by the Scottish Government (SG))
- Intermediate Zones (produced by SG)

These building blocks are designed following each Census to contain a minimum number of residents and households to safeguard the confidentiality of any statistics released from them.
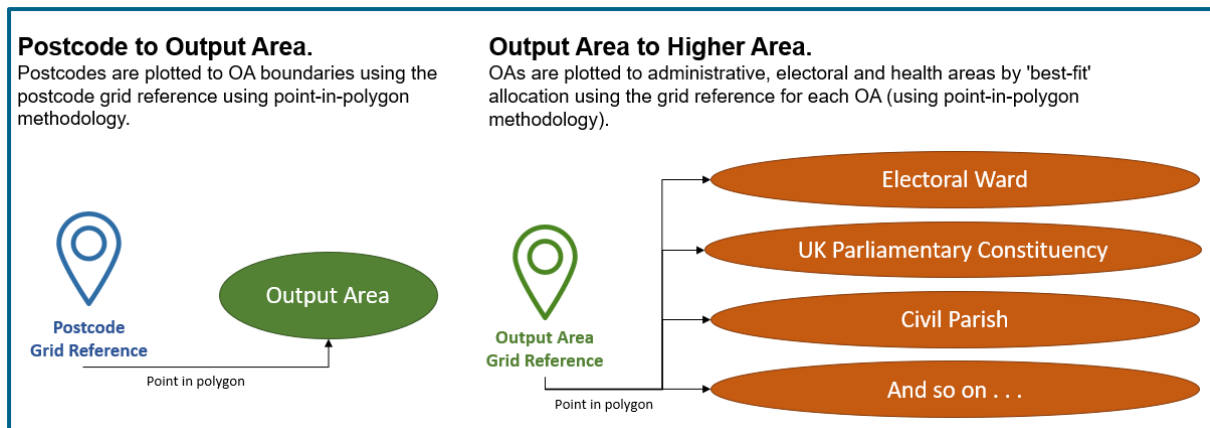
Building brick geographies

NRS produce two building brick geographies; postcodes and census output areas (which are built up from postcodes).

The Scottish Postcode Directory allocates postcodes directly to a number of higher geographies by using each postcode's 1-metre grid reference.
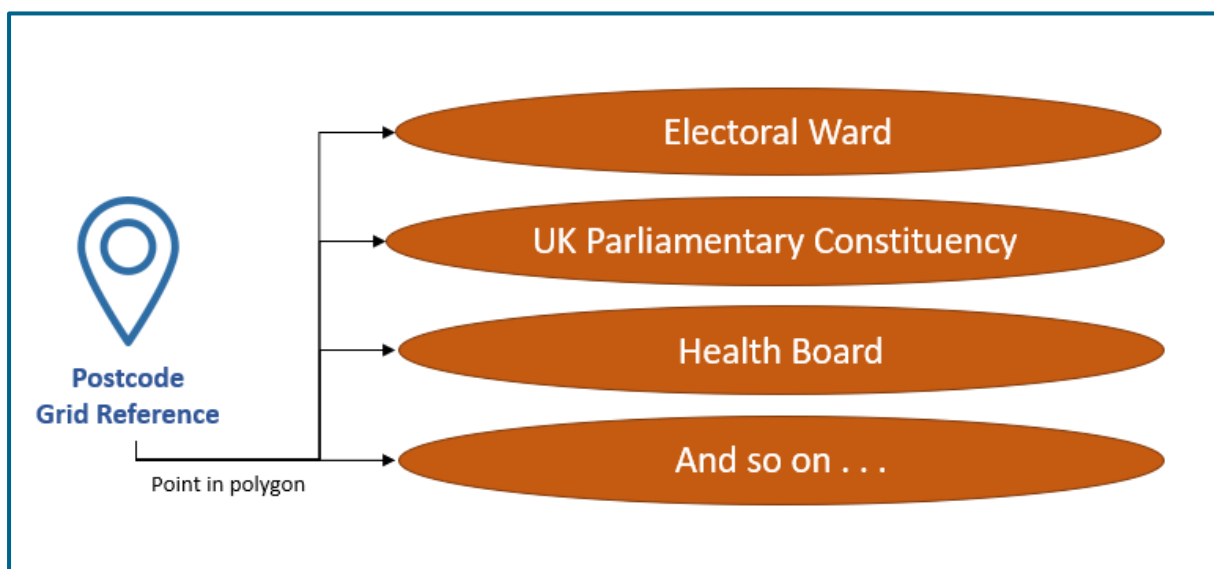
Whereas the Census Index assigns statistics to higher geographies via the grid reference of the census output area, which aligns with the GSS Geography Policy for National Statistics.

Postcode to Output Area.
Postcodes are plotted to OA boundaries using the postcode grid reference using point-in-polygon methodology.

Output Area to Higher Area.
OAs are plotted to administrative, electoral and health areas by 'best-fit' allocation using the grid reference for each OA (using point-in-polygon methodology).

Whilst the SPD method is more exact in allocating postcodes to a higher geography, it does not support the GSS Geography Policy for National Statistics.

SPD lookups should be used if you require the exact location of postcodes, and the associated higher area allocation, i.e., what Electoral ward am I in?

In each case the higher area assignment is carried out using the 'Point in Polygon' method (described in the next section), the only difference is the building brick geography being used. Using different building bricks will lead to slightly different statistics.

The example below is based on getting a postcode lookup to Electoral Ward. If you choose the postcode to higher area [SPD] assignment, the postcode EH7 5JB is in Leith Walk Ward. However, if you use the postcode to output area, output area to higher area [census index] assignment, EH7 5JB is within 2011 OA S00104375, and the centroid for that OA is in City Centre Ward.
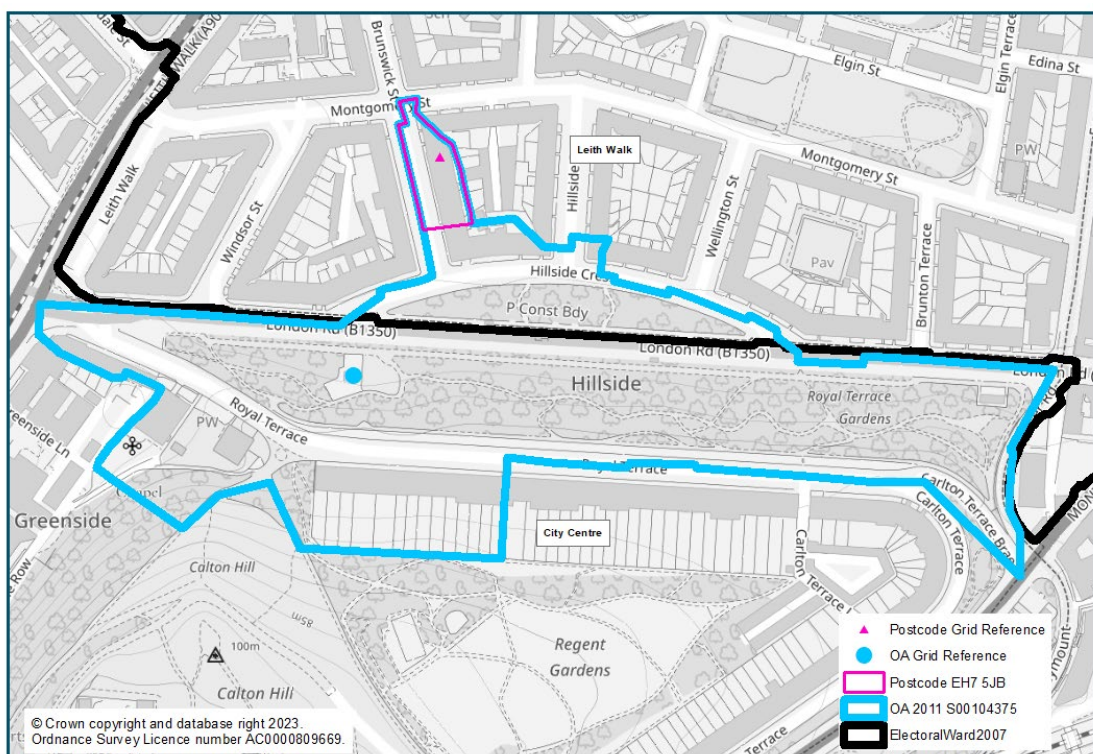
Point in Polygon method of assignment

In the diagram below we have points labelled 'A to G' and polygons labeled 'ward 1 to ward 4'.

The points are assigned their ward value using a 'Point in Polygon' method. This method is a spatial operation where points (i.e., grid references) are overlaid on the polygons (i.e., Electoral Ward boundaries) to determine which points are contained within the polygons.

In this example, the point in polygon results are:
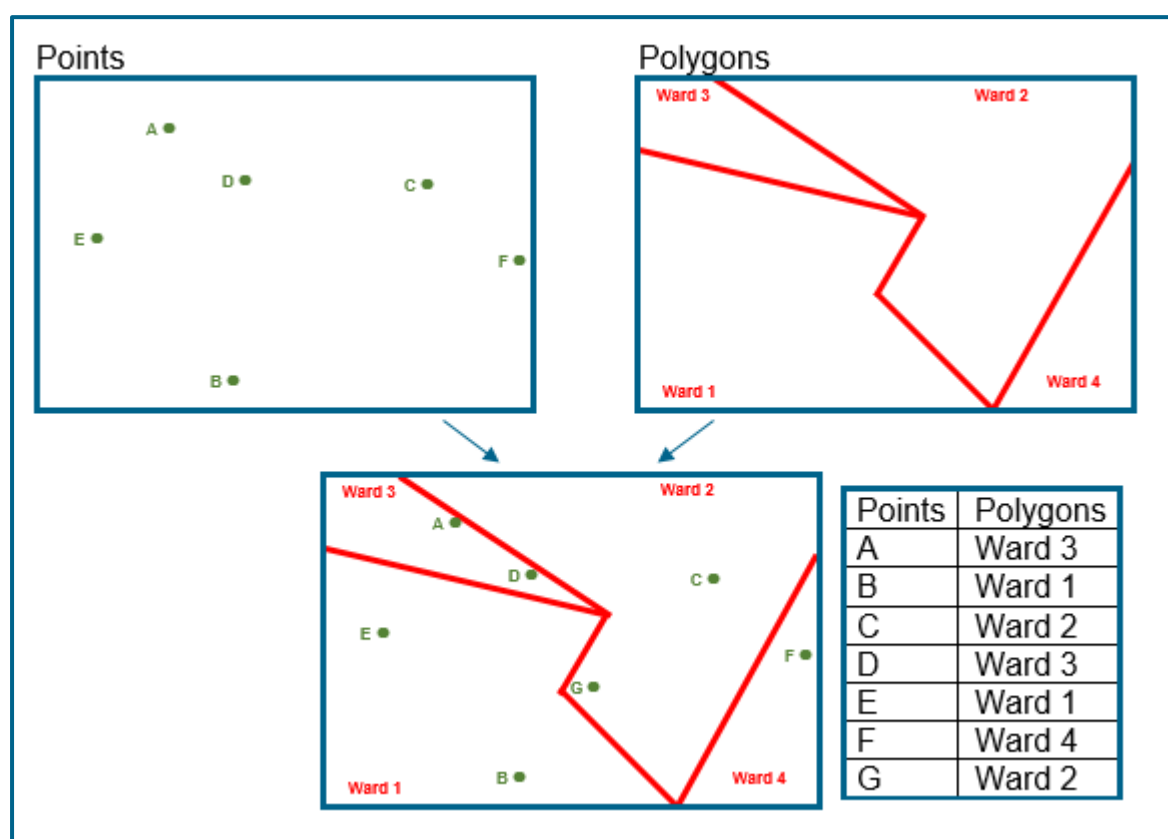
Note that postcode boundaries can straddle higher geography boundaries (i.e., have address content within the same postcode boundary, but in different higher geographies, for example wards). A postcode, and all the addresses contained within it, is assigned to one higher geography (i.e., ward) based on the geographic position of the grid reference of the postcode, not the polygon shape or size.

This is the same for Output Areas (OA), Data zones and Intermediate zones, the grid reference, (or centroid) would be used for allocation to higher geographies.

If you compare the electoral ward assignment from the OA against the postcode assignment, you will notice that three of the postcodes have been assigned to a different electoral ward. This illustrates the difference between referencing at postcode level and at OA level.

Output Area 2011	Postcode	Output Area 2011 to Electoral Ward 2007	Postcode to Electoral Ward 2007
S00106914	EH3 8BH	City Centre	Fountainbridge / Craiglockhart
S00106914	EH3 8BL	City Centre	Fountainbridge / Craiglockhart
S00106914	EH3 8BU	City Centre	Fountainbridge / Craiglockhart
S00106914	EH3 8BG	City Centre	City Centre
S00106914	EH3 8BJ	City Centre	City Centre
S00106914	EH3 8BP	City Centre	City Centre
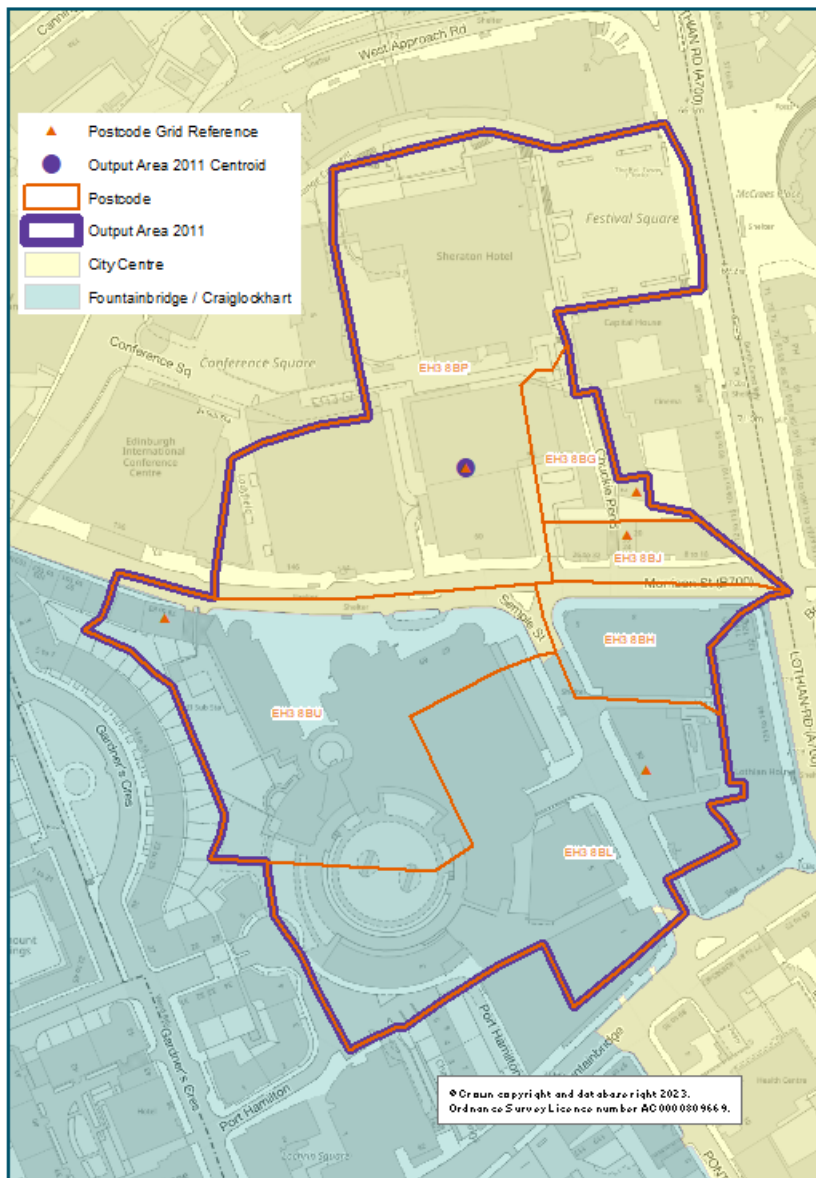
Neither are incorrect; the difference is in the building brick geography used, which, for census statistics, cannot be smaller than output areas.

In a worst case scenario where the same data is published in two different ways this may lead to disclosure of personal data. This is why GSS Geography Policy states that for statistical purposes we should use one method only. By convention, this is the output area based method.

Alignment to administrative geographies

Royal Mail defines postcode areas for sorting mail efficiently; the postcodes have no relationship with administrative or electoral boundaries. Royal Mail requires a stable geography in order to deliver its services, which these areas cannot provide.

NRS split postcodes for statistical reasons; they are not a feature of Royal Mail. Split postcodes are those with an 'A', 'B', or 'C' suffix and occur when:

- A postcode straddles two or more Council area boundaries. The most populous part of the postcode is identified by suffix A and the smaller parts by suffixes B, C, etc.

- a postcode straddles the Scottish/English border, and the Scottish postcode is allocated a suffix 'A'.

- an island and the mainland share a postcode, or a postcode contains property on more than one island.

By splitting the postcodes, NRS postcode boundaries are exact-fit to Council areas and the NRS Island dataset.

More information on postcodes can be found in the Postcode Background Information note within the Scottish Postcode Directory section of the NRS website.

Accuracy of best-fitted data

The difference between best-fit and exact-fit estimates will depend on the target geography. Generally, the bigger the target geography, the less difference there will be, meaning the best-fit allocation can be considered suitable for statistical analysis.

Best-fit

When postcodes or output areas are aggregated and assigned to a higher geography, they do not fit exactly into the boundary of the higher geography; instead, they form an approximation to the shape of the boundary of the higher geography.

An example of best-fit for 2011 Census Output Areas to 2022 Electoral Ward 'Inverclyde South' is shown below.

Exact-fit
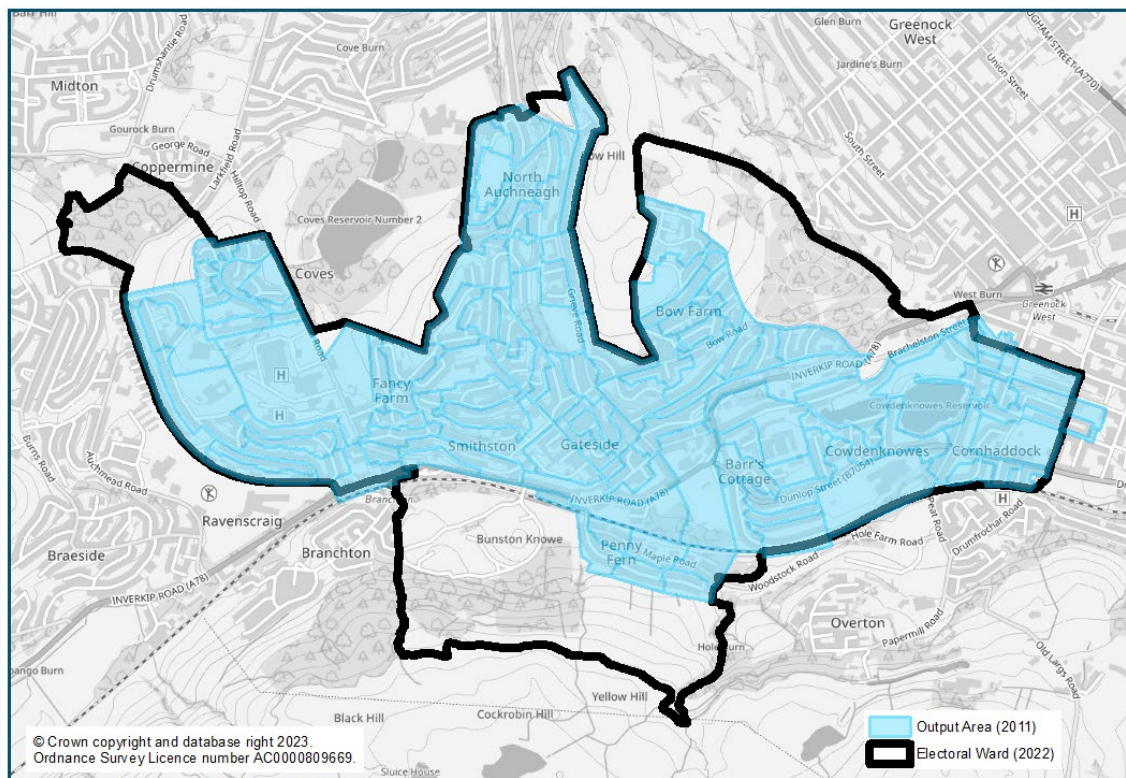
When postcodes or output areas are aggregated and assigned to a higher geography, they fit exactly into the boundary of the higher geography.

NRS policy of splitting postcodes at Council area boundaries allows for the creation of exact-fit data for postcode and output area to Council area. An example of exact-fit for postcodes to Dundee City Council area is shown below:

The following table shows exact-fit nesting of smaller geographical units to form larger units.

---

[1] The boundaries for Census geographies (Output Area, Data Zone, and Intermediate Zone) are exact-fit to Council area for the year that they are created. Council area boundaries are however subject to change in the interim period between Censuses and while the postcode boundaries are continually maintained and released every 6 months, the Census geographies are frozen.

Census Disclosure Control

If all statistics produced were published as exact-fit for a number of geographies, there could be overlap and confidential information could potentially be released about small populations in the overlap or 'sliver'.

For 1991 and 2001 Censuses, the risk of disclosure was addressed by adjusting cells in tables that had very small values and which could be disclosive. Small cell adjustment was not used as part of 2011 Census.

For 2011 the higher area assignment for OAs within slivers was changed. More information can be found in the 2011 Census Geography Background Information note within the 2011 Census Supporting Information section of the NRS website.

Cell Key Perturbation will be applied to Scotland's Census 2022 outputs. This means that small adjustments will be made automatically to cells in tables, including the Postcode household and population count. This is part of our Statistical Disclosure Control methodology, you can read more on the Scotland's Census website.

Output Area time-series comparisons

There are six main circumstances where change occurs between censuses.

- For some of the underlying postcodes that make up an Output Area (OA) the address content may have increased/decreased resulting in the shape of the postcode boundary changing. This can include postcode boundary changes requested by stakeholders.

- Brand new postcodes have been added as a result of new or re-developments,

- Some postcodes will have been deleted as a result of demolitions,

- Changes to Council area boundaries that have occurred between censuses,

- Changes in populations of postcodes may impact the shape of the OA. If the population has changed dramatically the OA may have been split or merged to ensure that they remain comparable in size and composition, and

- Differences in the location of the weighted centroid of an OA may have resulted in a different postcode being chosen as the master postcode for an OA, meaning its higher area assignments may also have changed.

The changes described above can result in differences between OAs produced at each Census.

Fitting to previous Census boundaries

Following the 2001 Census, the Scottish Government introduced Data Zones (DZ) as the key geography for small area statistics in Scotland, which are widely used across both public and private sector. Like OA they remain static between Censuses.

When 2011 Census outputs were made available, NRS produced a best-fit lookup of 2011 OA to 2001 DZ to allow users to produce statistics at DZ level while waiting on the redraw of DZ based on the 2011 OA.

If there has been little change in an area between Censuses (i.e., 2001 and 2011), generally the 2011 OAs best-fitted to 2001 DZ will produce a good fit, as shown in the example of 2001 data zone S02006212 below:
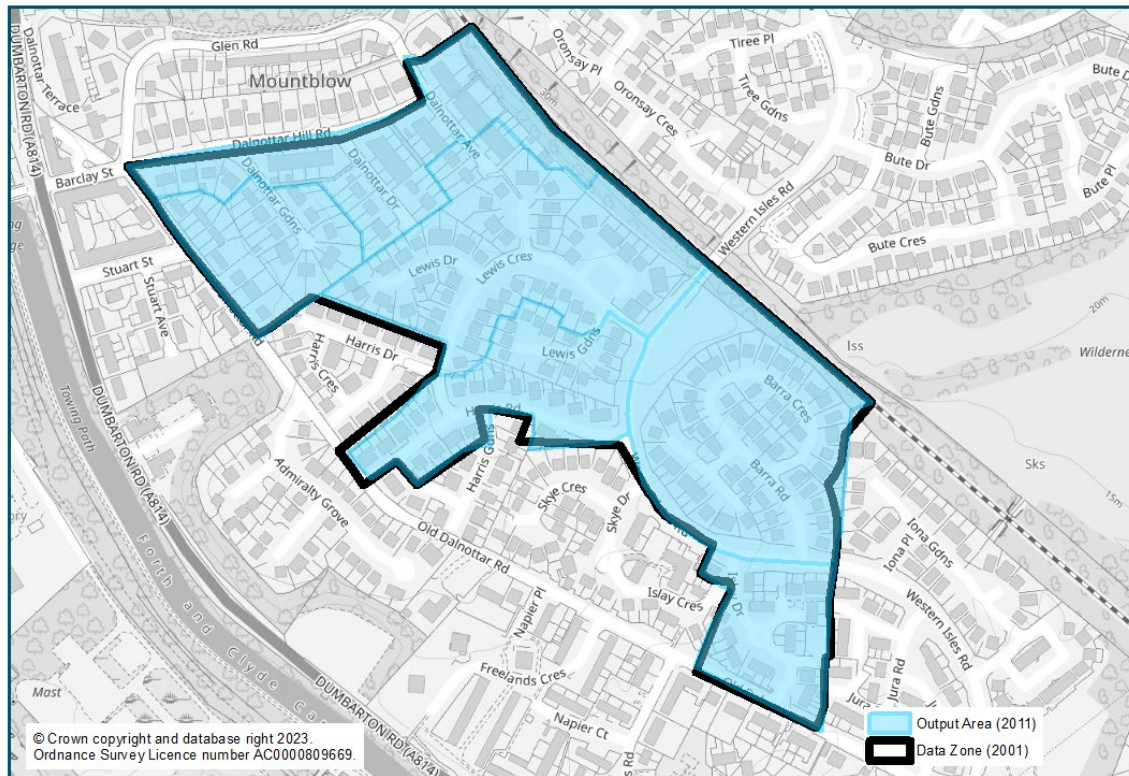
However, in some cases, caution should be taken when analysing best-fitted statistics, particularly for some areas within the smaller geographies such as 2001 DZ. For example, changes over time can mean that the best-fitted 2011 OAs do not accurately match the higher geography boundaries, such as in the example with a 2001 Data zone (S01001109) below.
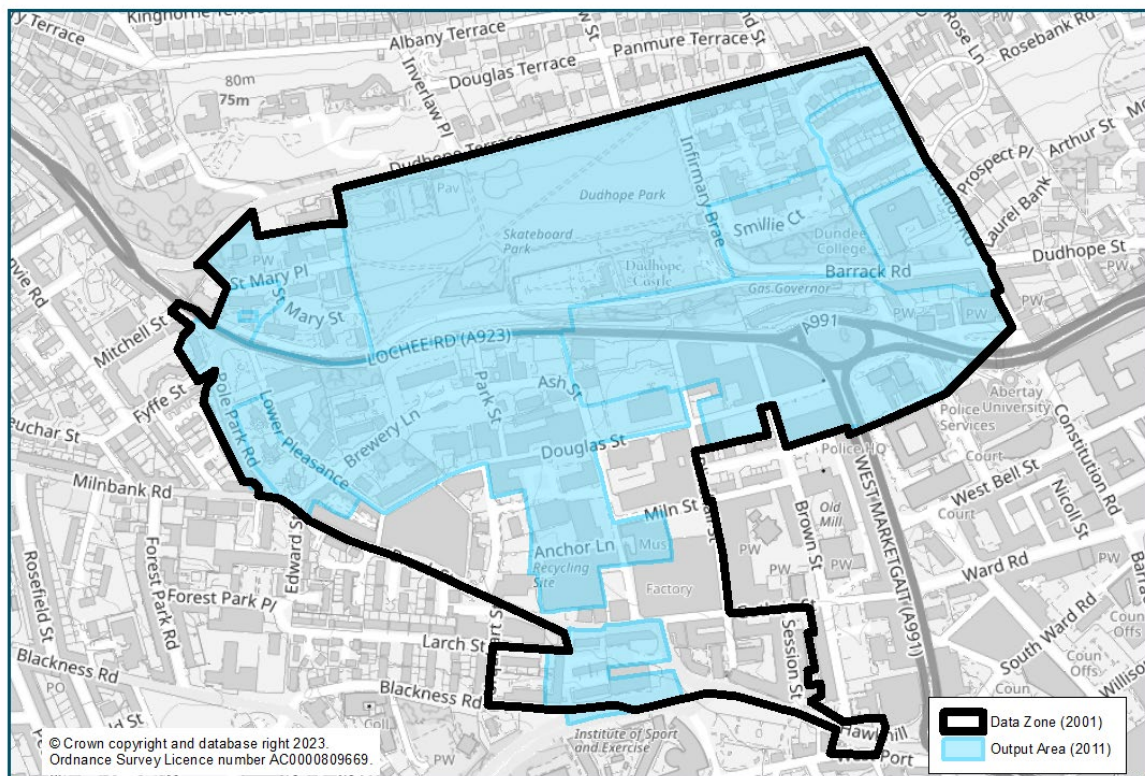
NRS Population and Migration branch produced a report '2001 Data zones: Population and Household Estimates' which can be found on the Small Area Population Estimates section of the NRS website.

The report details the work carried out to compare population and household estimates in the 2001 and 2011 Censuses. The 2011 population and household estimates used in this report have been created from 2011 postcode estimates. Building up from 2011 postcode estimates to 2001 data zones gives best population estimate for the 2001 Data zone geography. It also provides a more 'like with like' basis to compare population and household estimates over time for these areas.

Enquiries

Please contact our Geography Customer Services team if you need any further information.

E-mail: geographycustomerservices@nrscotland.gov.uk