Data Zone 2001 and 2011 Matching – Quick Guide

Understanding Data Zones

What are data zones?

- Data zones are 'stable' small geographical areas. 'Stable' means that they do not often change.
- They are used to report some official statistics.
- A data zone contains about 500 to 1,000 people.
- They are made up from groups of census output areas.
- Data zones 'nest' within council area boundaries.
- In urban areas, data zones can be very small in size, whereas in rural areas they can cover large areas.

Why did they change?

- Data zones were introduced after the 2001 census.
- Housing demolitions since then meant that 220 data zones had an estimated population below 500 people.
- New housing developments meant that 929 data zones had an estimated population of 1,000 or more.
- Data Zones 2011 are made from the new 2011 census output areas and in collaboration with local authorities, NHS and others.

How do I know if a data zone is from 2001 or 2011?

- Data zones have a 9 digit unique identifier beginning with 'S01' followed by 6 numbers.
- Data Zones 2001 are labelled S01000001 to S01006505 (there were 6,505 Data Zones 2001).
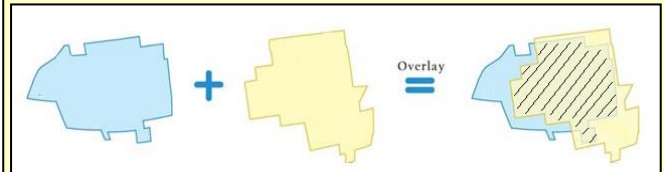- Data Zones 2011 are labelled S01006506 to S01013482 (there are 6,976 Data Zones 2011).

Matching Data Zones

There are two methods for identifying matches:

Population Matches
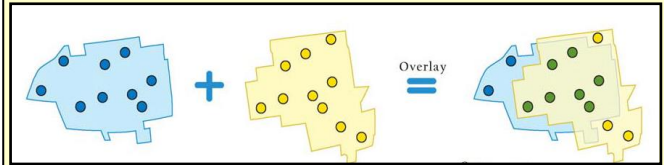
Area Matches

Of the 6,976 Data Zones 2011, there are 3,499 which have 1 to 1 population matches with data zones 2001, when using 2011 postcodes as the building block for both sets of data zones.

399 Data Zones 2011 have direct boundary matches to Data Zones 2001. Small variations can occur due to differences in coastlines, inland water, small changes to boundaries, small changes to the underlying output area geography or significant changes in rural areas that may not directly affect population.

Dependent on the situation, users will need to consider whether the population, area or a combination of both methods is most appropriate for their needs.

Using the Matching File – Do's and Don'ts

- Quality assurance of statistics and data sets
- Rough estimates of changes over time
- Informal guidance
- Identifying data zones that are unchanged

- Formal time-series analysis
- Creating new data sets based on statistics produced under Data Zones 2001
- Modelling or evidence-based policy development

- Data zones with 100% population matches may still mean that the boundaries have changed
- To complete time-series analysis, statistics should be recalculated based on Data Zones 2011

Comparing Population and Area of Data Zones 2001 and 2011
Methodology Report

1. Introduction

Data Zones are the key geography for small area statistics in Scotland and are widely used across the public and private sector. They are large enough that statistics can be presented accurately without fear of disclosure and small enough that they can be used to represent communities. Aggregations of data zones are often used to approximate a larger area of interest or a higher level geography at which statistics wouldn't normally be available. They have roughly standard populations to allow comparison between data zones. The boundaries remain static between the 10 year period between censuses to allow tracking of change over time.

Using 2011 Census data, a new version of data zones was created for 2011 and published in November 2014. At the request of the Office of the Chief Statistician and Performance (OCSP), a comparison between Data Zones 2001 and 2011 was carried out based on both population and area. The purpose of this document is to describe the method used for each comparison and describe appropriate uses for this analysis.

Each data zone is given a unique 9 digit identifier beginning with 'S01' followed by 6 numbers. Data Zones 2001 are labelled from S01000001 to S01006505 (6,505 individual data zones) and the Data Zones 2011 are labelled S01006506 to S01013482 (6,976 individual data zones).

2. Population Based Match

For the population based match between Data Zones 2011 and 2001, the percentage fit was assessed by comparing postcode (2011) based populations in a single 2011 data zone covered by associated unit(s) in the Data Zone 2001 geography. Populations were based on the 2011 set of postcodes, a snap shot of postcode units taken in January 2011 that were used as a basis for other Census products such as output areas. Postcodes were assigned to both Data Zones 2001 and 2011 and the results compared to assess the degree of fit between them.

For example, Figure 3.1 (left) shows the Data Zone (2011) S01006511 which has a population of 751 people - 723 of which are covered by Data Zone (2001) S01000007, 16 of which are covered by S01000023 and the remaining 12 covered by S01000011. In other words the fit between Data Zone S01006521 is 96.3% with 2001 Data Zone S01000007; 2.1% with S01000023; and 1.6% with S01000011.

3. Area Based Match

A similar process is used to assess the fit between Data Zones 2001 and 2011, but comparing the actual area covered by the data zones rather than just their population. This was achieved by intersecting the boundaries of in 2011 with those from 2001. The percent fit is the percent of the total area of a particular Data Zone 2011 covered by data zone(s) from 2001 (see figure 3.1 (right)).

Returning to our example, Data Zone S01006511 has a total area of about 40 hectares. 82.6% of this is covered by Data Zone S01000007, 11.5% by S01000011, and 4.9% by S01000023. The remaining 1% are comprised of small areas covering other data zones.

These small areas on the fringe of the boundary, called slivers, can be caused due to changes in the coastlines, internal waters and also due to small changes to boundaries.
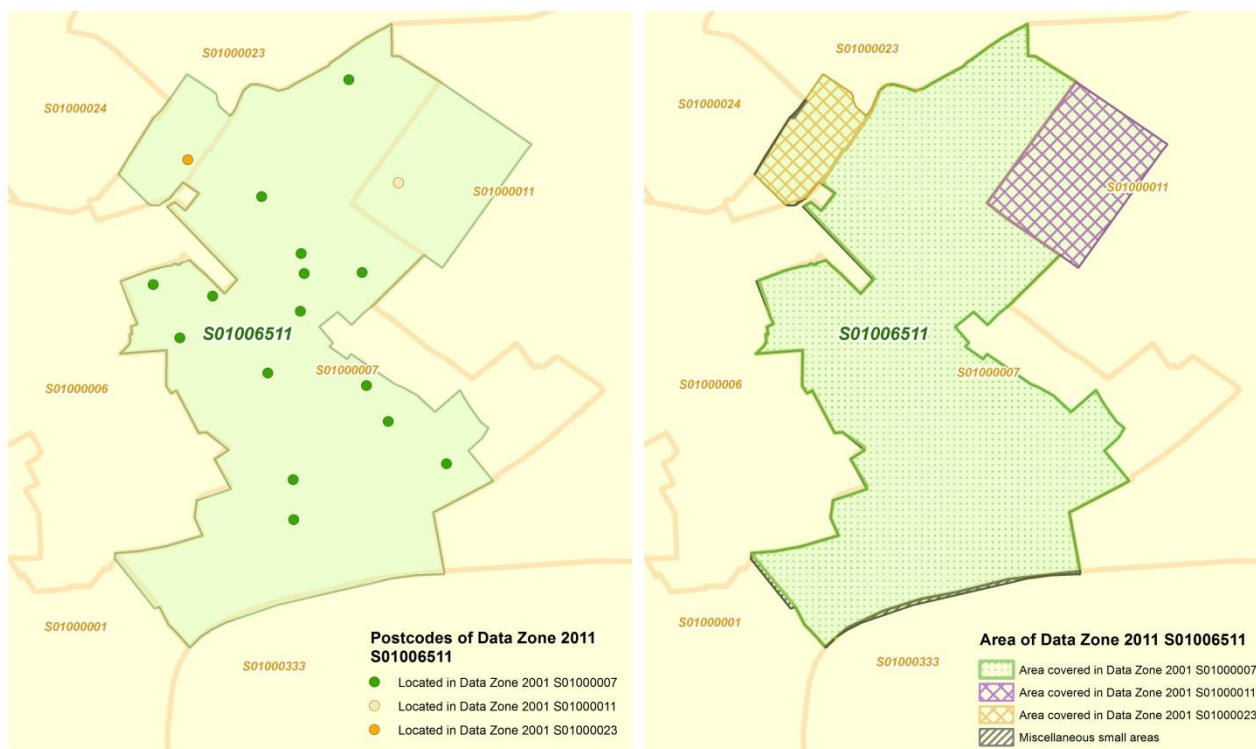

Figure 3.1: Population and area based matching methods

4. Output

Output for the match between Data Zones 2001 and 2011 is provided as a spreadsheet for both the population and area based comparison. Note that neither spreadsheet represents a one to one lookup between Data Zones 2001 and 2011. Where population is concerned, only 3,499 of the 6,976 data zones match directly with a single 2001 data zone. For the area based comparison, only 399 match directly to a single 2001 data zone.

5. Appropriate Uses

These files are produced for guidance only to help understand the link between 2001 and 2011 data zones. This should not be used to develop new data sets based on statistics that have been produced under a different time frame.

The main purpose of this information is to help producers of Official Statistics develop appropriate methods for quality assuring new statistical products and data users to understand the effects of the changes caused by new data zones.

To conduct time-series analysis it is recommended that you speak to the producer of the dataset to obtain a back-series for Data Zones 2011.

6. Further Information

For further information on Geography Policy in Scotland contact neighbourhood.statistics@scotland.gsi.gov.uk