An introduction to 2021 Census geography datasets

James Crone
EDINA
The UK Census

A population census is held every 10 years by the devolved UK statistical agencies.

Ask questions of the whole country.

A mix of Household and Individual Questions.

Some difference in the questions asked by the nations, reflecting local preferences.

Submission of paper form or online completion.
Processed Output Census Data

- Information submitted by census form on census night is processed into aggregate output tables by the national statistics agencies.
- Producing Univariate and Multivariate tables of census statistics.
- Census Statistics tables are output as counts of people or counts of households.
- Data output at different levels of small area output geography.
- The Smallest level of output geography across the UK is the **Census Output Area**.
- Output Areas are a synthetic geography especially created for the publication (output) of census statistics.
Census Output Areas

Each OA has a minimum of 40 households and 100 residents to a maximum of 250 households and 625 residents.

OAs are constructed using an automated zone building system.

A design goal when building OAs is that each OA should contain a homogenous population.

To ease comparisons with the previous 2011 census, a further goal is to minimise change between 2011 and 2021.
Changes to Output Areas

In the 10 years since the 2011 census there has been both an increase in the size of the UK population and changes to population distribution and population density as where people live has changed.

So that Census Output Areas continue to meet their design criteria in terms of the numbers of residents / households and the homogeneity of population characteristics there needs to be adjustment to the Output Areas.
Minor adjustment to Output Areas

Across England and Wales as a whole ~95% of 2011 OAs are the same in 2021. Of the rest:
Some 2011 OAs merged into 2021 OA.
Some 2011 OAs split into 2 or more 2021 OAs.
Other more complex types of change in a smaller number of cases.
At the time of the 2011 census there is a single Census Output Area. At the time of the 2021 census a new housing development has been built on a former brownfield site so the 2011 OA has been split. In addition to meet the design goal of maximizing the homogeneity of the underlying population the region at the top left containing high-rise residential housing now forms an OA on its own.
Super Output Areas

Statistical Disclosure Control is there to ensure that individuals cannot be identified from census statistics. In some cases, given the size of Output Areas and the small associated populations there is a risk of individuals being identifiable.

To prevent this, not all census statistics are released at the most detailed Output Area level but instead are released at much larger Super Output Area zones level.

2 levels of Super Output Areas:
- Lower Layer (LSOA): 400 – 1200 HHs / 1000 – 3000 people
- Middle Layer (MSOA): 2000 – 6000 HHs / 5000 – 15000 people

LSOAs are created by merging OAs
MSOAs are created by merging LSOAs
OAs; LSOAs and MSOAs nest within Local Authorites.
Some national differences

Updated Output Areas in due time will be available for England, Wales and Scotland. In Northern Ireland Output Areas are known as Small Areas.

England and Wales have Lower and Middle Super Output Areas. Scotland has Lower and Middle Super Output Areas as well but in 2011 at least they are called Datazones and Intermediate Geographies.

Scottish Datazones are equivalent to LSOAs.
Scottish Intermediate Geographies are equivalent to MSOAs.
Northern Ireland only has 1 layer of Super Output Areas.
Census Output Geographies

**Statistical**: Output Area; Lower / Middle SOA

Producing Census data at different geographies enables the census data to be compared with other non-census data which may be collected and output at different geographies:

**Administrative**: Local Authorities

**Electoral**: Electoral Ward; Parliamentary Constituencies

**Health**: NHS Regions; Integrated Care Boards
Geospatial Data

Two forms of geospatial data – **Vector data** consisting of **Points**; Lines and **Polygons** and Raster data consisting of gridded data.

All geospatial data is referenced to a **Spatial Reference System**. British National Grid is an example of an SRS.

**Geographic Identifiers** are alphanumeric codes that unambiguously identify specific instances of a given output geography.

**Geographic Look-up Tables** relate one or more geographies to one another.
GSS Codes

Instances of UK Census Geography small areas are assigned a Geographic Identifier that uses the UK Government Statistical Service (GSS) system of codes.

GSS Codes are 9 digits in length. First 3 digits identify the entity (the type of geography). Remaining 6 digits identify the specific instance within that type.

The ONS maintain a Register of Geographic Codes which lists each of the entities.

So E00126679 has Entity Code of E00, this is instance 126679 of English Output Areas geography.

So E07000118 has Entity Code of E07, this is instance 000118 of English Non-Metro District geography.

GSS Codes were first used in the 2011 UK Census. Before 2011, in the 1991 and 2001 Census, Census codes were hierarchical in nature. GSS Codes are non-hierarchical – it’s not possible purely from the code itself to determine what LSOA, MSOA or LAD, an OA with code E00126679 falls within. Determining this requires the use of a geographic look-up table.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Entity code</td>
<td>Entity name</td>
<td>Entity abbreviation</td>
<td>Entity theme</td>
</tr>
<tr>
<td>2</td>
<td>England</td>
<td>County</td>
<td>CTRY</td>
<td>Administrative</td>
</tr>
<tr>
<td>3</td>
<td>E02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>E03</td>
<td>Output Areas</td>
<td>OA</td>
<td>Statistical Building Block</td>
</tr>
<tr>
<td>5</td>
<td>E04</td>
<td>Super Output Areas, Lower Layer</td>
<td>LSOA</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>E05</td>
<td>Super Output Areas, Middle Layer</td>
<td>MSOA</td>
<td>Statistical Building Block</td>
</tr>
<tr>
<td>7</td>
<td>E06</td>
<td>Civil Parishes</td>
<td>PAR</td>
<td>Administrative</td>
</tr>
<tr>
<td>8</td>
<td>E07</td>
<td>Electoral Wards/Divisions</td>
<td>WD</td>
<td>Administrative/Electoral</td>
</tr>
<tr>
<td>9</td>
<td>E08</td>
<td>Unitary Authorities</td>
<td>UA</td>
<td>Administrative</td>
</tr>
<tr>
<td>10</td>
<td>E09</td>
<td>Non-metropolitan Districts</td>
<td>NMD</td>
<td>Administrative</td>
</tr>
<tr>
<td>11</td>
<td>E10</td>
<td>Metropolitan Districts</td>
<td>MD</td>
<td>Administrative</td>
</tr>
</tbody>
</table>
Census Boundary Datasets

CensusBoundaries describe the spatial footprint of an instance of a given geography and consist of 1 or more polygons formed from a series of geographic coordinates located within a spatial reference system.

Census Boundaries are provided in different GIS Data Formats. Examples include Shapefiles and MapInfo TAB files. Shapefiles are ubiquitous.

Different flavors of polygonal geometry are available:

Boundaries may occupy the Extent of the Realm or be clipped to the Mean High Water mark. Boundaries clipped to the Mean High Water Mark provide a more familiar geography.

Generalized boundaries have polygons whose features are made up from fewer geographic coordinates resulting in smaller file sizes. Smaller file sizes mean that the data is quicker to load and display in mapping applications. However for carrying out (spatial) analysis un-generalized (full-resolution) boundaries should be used.
Census Centroid Datasets

A Centroid simplifies the spatial footprint of a census geography feature from a polygon to a simple vector point geometry or simply an x/y coordinate. There are 2 forms of centroid:

**Geometrically Weighted Centroid** (GWC): the location of the point is defined based on the spatial average of the coordinates that define the polygon.

**Population Weighted Centroid** (PWC): the location of the point is defined based on the underlying population distribution within the polygon.

Within Census data, Population Weighted Centroids are the norm.

Centroids can be used for simple analysis, for example as a proxy for a spatial footprint within travel time or network analysis.

Mostly though Centroids are used for georeferencing purposes to relate a given geography to 1 or more other geographies where the nature of the polygonal boundaries is such that there would be an imperfect nesting and cases of overlap etc.
Census Look-Up Table Datasets

Census Look-Up Tables relate one or more Census geographies to one another. Enabling census data to be shown at those geographies.

Some examples:
2021 OA : 2021 LSOA : 2021 MSOA : 2022 LAD
2021 OA : Care Board : NHS Region
2011 OA : 2021 OA

Postcode Directories are a special form of geographic Look-Up Table relating postcodes to other geographies, including Census Geographies

Access to Census datasets through the UK Data Service

The UK Data Service:

• brings together current and historic census datasets from the different UK national statistic agencies
• adds value to those datasets with the specific needs of the UKDS user community in mind
• provides alternative access mechanisms to those provided by the national statistic agencies themselves
• supports the use of census data through online training materials; webinars and hands-on user workshops
Plans for release of 2021 census geography datasets through online UK Data Service applications

The Boundary Data Selector allows one to make sub-selections from any boundary dataset held within the UKDS boundary database and the download of those boundaries in different geospatial data formats.

Easy Download provides simple access to downloads of full boundary datasets without any sub-selecting; centroids and geographic look-up tables.

The Postcode Data Selector allows one to make sub-selections from current and historic ONS postcode directories held by UKDS. Postcode directories include lookups to many standard UK geographies.

The UKDS team at EDINA are in the process of adding the new census geography datasets to UKDS applications.

New census geography datasets: boundaries; centroids and LUTs expected to be added to applications in the next few months following the building of derived products; database/application updates and UI tweaks.

Following user feedback we plan to add dataset preview functionality to the Boundary Data Selector and enable boundaries to be output in additional geospatial data formats like GeoPackage.
Use of boundaries for the visualization and analysis of census data.

Join a table of aggregate census statistics to digital boundaries for the same geography.

Boundaries now have associated with them all of census statistics.

Contains OS data © Crown copyright and database right 2022
Choropleth Maps

Polygons are shaded in proportion to the measurement of the statistical variable being displayed on the map.

Choropleth maps provide an easy way to visualize how a measurement varies across a geographic area or to show the level of variability within a region.

Contains public sector information licensed under the Open Government Licence v3.0
Choropleth mapping of census variables at different levels

Disclosure control might mean that a given aggregate census statistic is not available at all levels of geography – therefore it may not be possible to map the statistic at most detailed OA level.

Where a census statistic is available across multiple levels of geography:

Mapping the data at Local Authority level uncovers general patterns and smooths the data

Mapping the data at a more detailed LSOA / Output Area level provides more detail but also adds more noise
Problems of Choropleths

Choropleth maps imply that the population is distributed uniformly across the extent of the polygonal census zone.

Choropleth maps are also subject to the Modifiable Areal Unit Problem (MAUP) especially when making comparisons over time where the census boundaries may have changed between each census.
Masked Choropleth Maps

Masked Choropleth Maps modify traditional Choropleth maps using additional land use datasets to present a more realistic distribution of the variable.

DataShine is a form of this where a layer of Ordnance Survey buildings is used to mask the census areas.

This helps with the problem of the choropleth implying that population is uniformly distributed across polygons.
Cartograms

A cartogram is a special form of map projection where polygon areas are drawn in proportion to the variable being mapped rather than the land area of the polygon.

Different types of Cartogram:
- Non-contiguous Cartograms
- Contiguous Cartograms
- Dorling Cartograms

Cartograms help avoid the problems of census areas with large populations but covering a small area being hidden by census areas with small populations but covering much larger areas.
Census Cartograms in the wild

People and Places: A 21st-century atlas of the UK
Danny Dorling and Bethan Thomas

The QGIS Desktop GIS has a plugin which can be used to create these kind of Cartograms.
Further Information

UKDS Data Impact blog post written by Professor David Martin talking about the development of UK Census Output Geography:
https://blog.ukdataservice.ac.uk/david-martin-profile-1/

Videos and information about AZTool – the automated zone design tool used to build the Census OAs developed by Professor David Martin and colleagues at the University of Southampton:
http://aztool.geodata.soton.ac.uk/supporting_materials/

Online ONS Geography and Statistics Training course which provides awareness of geography and statistics (including census statistics) and use of data in QGIS; R and Python:
https://onsgeo.github.io/geospatial-training/
Thank you.
James Crone

James.Crone@ed.ac.uk

Please fill in the evaluation form for this webinar.