

Using QGIS to map 2021/22 census data from the UK Data Service

James Crone
with Nadia Kennar and Gill Meadows



Aims for the session today

- Use the open source QGIS desktop GIS application to map 2021/22 census data downloaded from the UK Data Service
- Create a univariate choropleth map
- Create a bivariate choropleth map
- Create an area cartogram

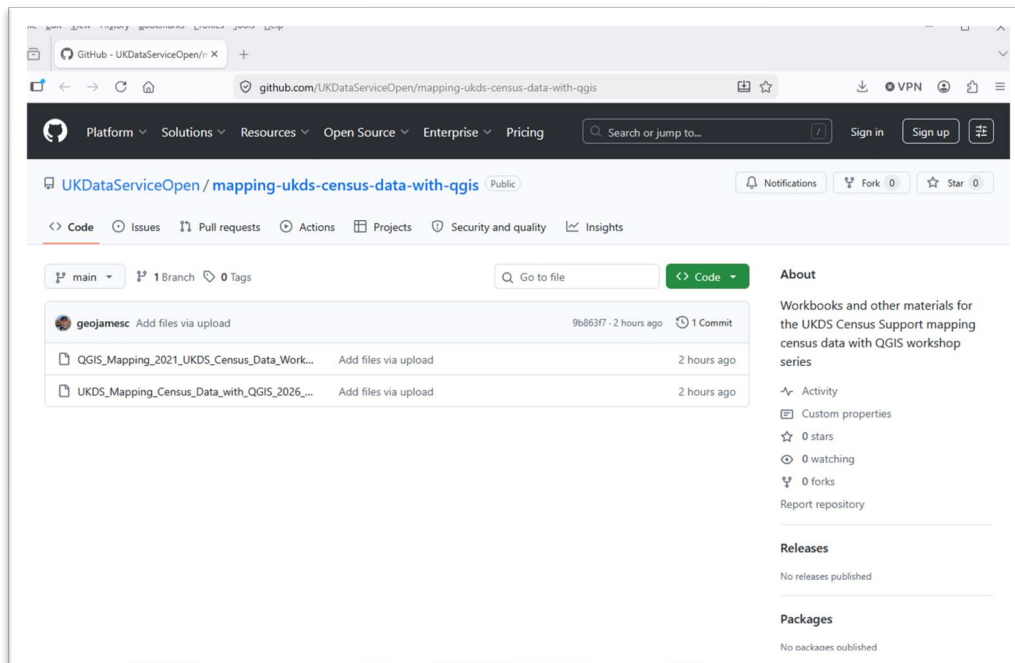
Along the way you will:

- Learn how to download 2021 census aggregate data and 2021 census boundary data from the UK Data Service
- Learn how to carry out data transformations and data processing on the downloaded census aggregate data to prepare it for mapping
- Learn about some of the issues that we need to be aware of when mapping census data
- Learn how to use the open source QGIS desktop GIS application to create maps and other geographic visualisations from census data
- Learn about resources within the UK Data Service to help you with using and mapping census data

QGIS Workbook and Data Pack

- A PDF workbook which includes all of the steps demonstrated today and an accompanying data pack is available. To use the workbook you need to have installed on your computer the QGIS desktop GIS application and have access to a spreadsheet application like LibreOffice Calc or Excel.
- To download the PDF workbook and the data pack go to this UKDS GitHub repository:

<https://github.com/UKDataServiceOpen/mapping-ukds-census-data-with-qgis>



Introducing the decennial UK census

The image shows the front cover of the 'Household Questionnaire' for the 2011 Census. It features the Royal Coat of Arms, the text 'Household Questionnaire England Office for National Statistics', and the '2011 Census' logo. There are instructions on how to complete the form online at www.census.gov.uk or by post. A 'Declaration' section is also visible, along with a signature line and a date field. A large 'H1' logo is at the bottom right.

This is page 6 of the census form, titled 'Household questions - continued'. It contains questions H7 through H10. Question H7 asks about the type of accommodation (detached, semi-detached, terraced, flat, etc.). Question H8 asks if the accommodation is self-contained. Question H9 asks for the number of rooms available for use only by the household, with a list of rooms to count (bathrooms, toilets, etc.). Question H10 asks for the number of bedrooms. There are also questions H11 and H12 about central heating and household ownership/renting.

This is page 10 of the census form, titled 'Person 1 - continued'. It contains questions P2 through P12. Question P2 asks for the main job. Question P3 asks if the person is an employee, self-employed, or an employer. Question P4 asks for the full and specific job title. Question P5 asks for a brief description of the main job. Question P6 asks if the person supervises any employees. Question P7 asks for the main activity of the employer or business. Question P8 asks for the name of the organization. Question P9 asks if the person had a job last week. Question P10 asks for the address of the workplace. Question P11 asks how the person usually travels to work. Question P12 asks for the number of hours usually worked per week.

- 1971
- 1981
- 1991
- 2001
- 2011
- 2021 / 2022

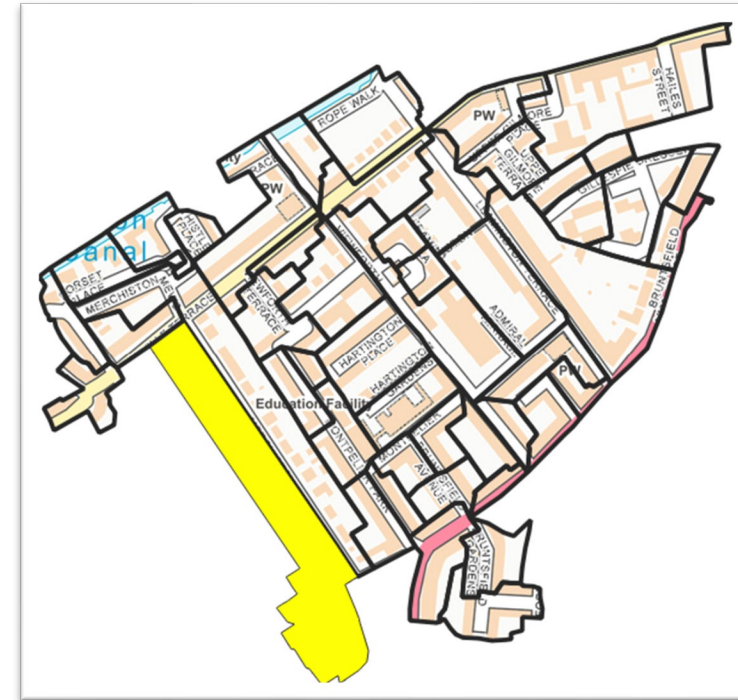
The census form is normally submitted online

Some topics covered by the UK census

- Education – highest level of qualification; numbers of students.
- Housing – accommodation or tenure type.
- Health – good or bad health.
- Labour market – occupation; unemployment.
- Language – proficiency in English language.
- Transport – method of travel to work and distance travelled.

Output census data

- Information submitted by census form on census night is processed into **census aggregate** data
- Data output as counts of *people* or *households*
- Data output at different types of *small area* geography
- Smallest level of geography is the Census Output Area (as here)
- Output Areas are a synthetic geography especially created for the publication of census data

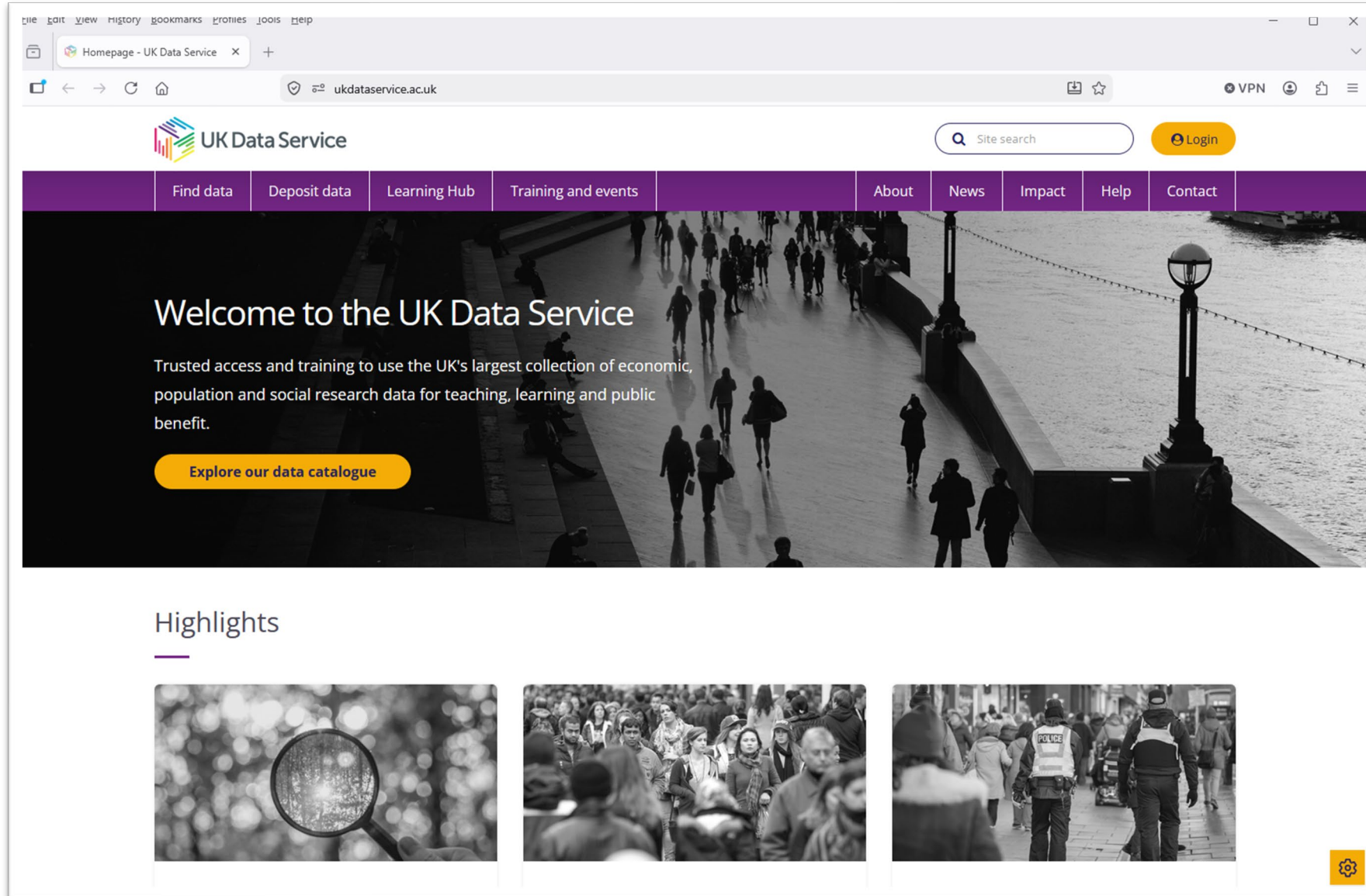


	fid	oocode	totpop	hhdspace	male	female	detached	flatpur
21	31	S00104890	138	66	63	75	0	64
22	24	S00106918	110	69	60	50	0	67
23	25	S00104904	73	54	29	44	0	53
24	26	S00104922	113	41	53	60	9	6
25	27	S00104908	153	53	75	78	3	39
26	4	S00104910	87	57	38	49	0	56

Univariate vs Multivariate census aggregate data

- Univariate census data contains information about a single census variable giving insight into a single census topic
- Multivariate census data is richer by combining different census variables and allows you to look at the relationships between the different census variables
- Take a *univariate* e.g. **method used to travel to work** census dataset (on foot; by car; by rail) and add 1 or more extra data dimensions such as occupation giving a *multivariate travel to work* census dataset.
- Use this multivariate dataset to understand how people working in different occupations travel to work

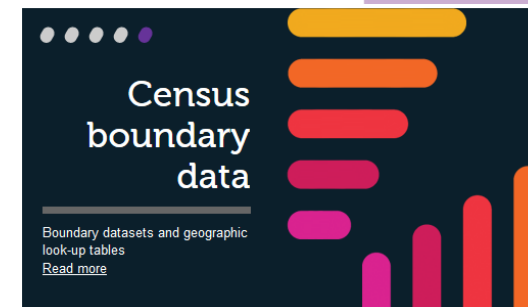
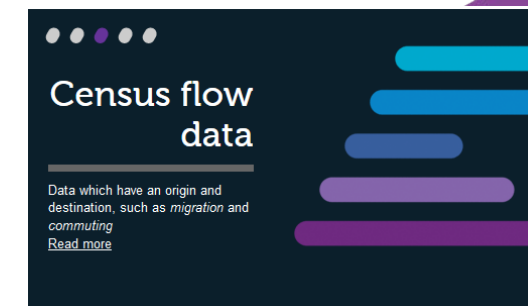
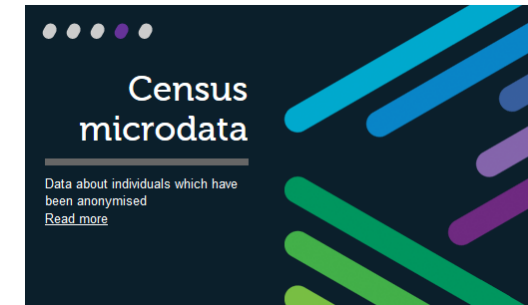
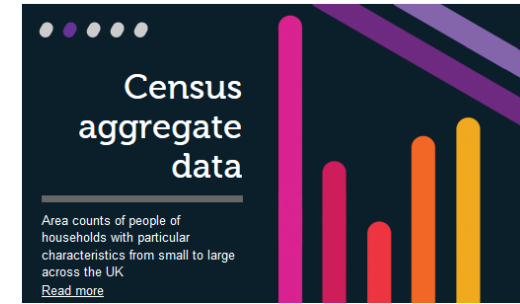
The UK Data Service



The screenshot shows the UK Data Service homepage in a web browser. The browser's address bar displays 'ukdataservice.ac.uk'. The page features a purple navigation bar with the following menu items: Find data, Deposit data, Learning Hub, Training and events, About, News, Impact, Help, and Contact. A search bar and a yellow 'Login' button are positioned in the top right. The main content area has a dark background with a photograph of a busy pedestrian walkway. The text reads: 'Welcome to the UK Data Service' followed by 'Trusted access and training to use the UK's largest collection of economic, population and social research data for teaching, learning and public benefit.' Below this is a yellow button labeled 'Explore our data catalogue'. A 'Highlights' section is located below the main banner, featuring three image thumbnails: a magnifying glass over data points, a crowd of people, and police officers. A small gear icon is visible in the bottom right corner of the highlights section.

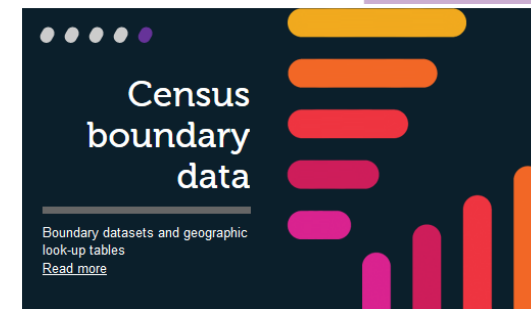
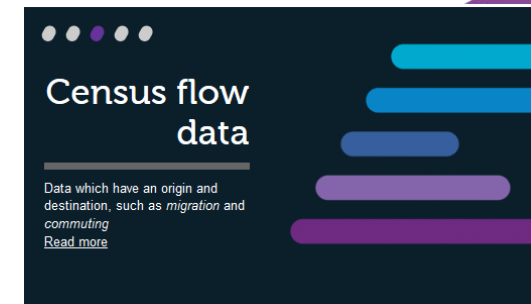
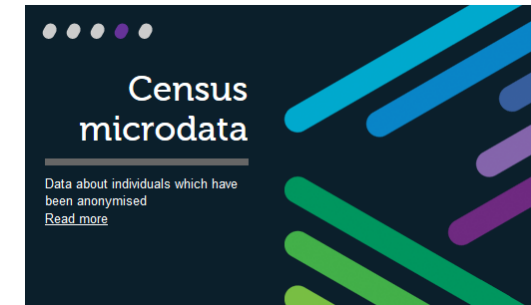
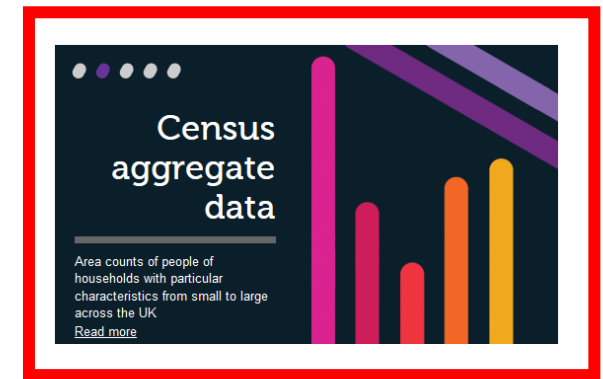
UK Data Service Census Support

- The UK Data Service Census Support team is responsible for building applications that provide access to census data from 1971 onwards and for supporting the onward use of this census data
 - Census *aggregate data*
 - Census *microdata*
 - Census *flow data*
 - Census *boundary data*
- Today we will look at and use census aggregate data and census boundary data downloaded from the UK Data Service



UKDS census aggregate tools

- **CKAN** provides access to bulk census aggregate data but no ability to filter and extract subsets
- **Data Explorer** provides the ability to extract subsets from census aggregate data
- Census aggregate data from earlier census years before 2021/2 is in the process of being migrated from legacy UK Data Service applications like **InFuse** into the Data Explorer platform



Demo 1: Downloading 2021 census aggregate data using the UK Data Service Data Explorer

- Use the UK Data Service Data Explorer to download 2021 univariate census aggregate data for electoral wards in Leeds providing information on the accommodation type that people are living in.

Prepare downloaded census data for mapping

- Before we can open the census aggregate data downloaded from Data Explorer in QGIS and use it to create maps and other geographic visualisations we need to do some data manipulations on this downloaded data:
 - Transform data from Long to Wide form
 - Normalise census variables
 - Add a categorical summary variable
- The census aggregate data downloaded as a CSV file from Data Explorer can be manipulated in a spreadsheet application like LibreOffice Calc or Excel

Transform data from Long to Wide form

- How a tabular dataset is organized into rows and columns is known as that datasets *shape*
- A datasets shape may be *long* or *wide*
- The UK Data Service Data Explorer provides census aggregate data in long form
- In this long form the different census variables for a given small area census output geography instance are spread over separate rows
- In order to be able to map the census aggregate data in a GIS like QGIS we need to change the shape of the data from long to wide form so that all of the different census variables for a given small area census output geography instances are provided in a single row
- The process of changing the shape of the data from long to wide is done by grouping (aggregating) the different census variables by the small area census output geography instances. This can be done by constructing a *pivot table* in a spreadsheet application like LibreOffice Calc or Excel

	I	L	S
1	GEOG_AREA	Accommodation type (8 categories)	OBS_VALUE
2	E05011384	Detached	3211
3	E05011384	Semi-detached	3612
4	E05011384	Terraced	2353
5	E05011384	In a purpose-built block of flats or tenement	495
6	E05011384	Part of a converted or shared house, including bedsits	25
7	E05011384	Part of another converted building, for example, former school, church or warehouse	12
8	E05011384	In a commercial building, for example, in an office building, hotel or over a shop	30
9	E05011384	A caravan or other mobile or temporary structure	3
10	E05011385	Detached	622
11	E05011385	Semi-detached	3197
12	E05011385	Terraced	4362
13	E05011385	In a purpose-built block of flats or tenement	2613
14	E05011385	Part of a converted or shared house, including bedsits	569
15	E05011385	Part of another converted building, for example, former school, church or warehouse	335
16	E05011385	In a commercial building, for example, in an office building, hotel or over a shop	110
17	E05011385	A caravan or other mobile or temporary structure	10
18	E05011387	Detached	649
19	E05011387	Semi-detached	4537
20	E05011387	Terraced	3490
21	E05011387	In a purpose-built block of flats or tenement	1840
22	E05011387	Part of a converted or shared house, including bedsits	140
23	E05011387	Part of another converted building, for example, former school, church or warehouse	75
24	E05011387	In a commercial building, for example, in an office building, hotel or over a shop	72
25	E05011387	A caravan or other mobile or temporary structure	8

I	WdGeoID	caravan	detached	commb	flat	sharedh	schchwh	semidet	terrace	total
2	E05011384	3	3211	30	495	25	12	3612	2353	9741
3	E05011385	10	622	110	2613	569	335	3197	4362	11818
4	E05011387	8	649	72	1840	140	75	4537	3490	10811

Normalise census variables

- Avoid mapping raw counts of census variables - people or households
- Raw counts of a given census variable will naturally be higher or lower in small areas where the population is higher or lower
- Instead the counts should be normalised (standardised) so we can tell if these raw counts are high or low given the size of the overall population
- Normalising the data makes the counts in one small area comparable with the value in other small areas
- Option 1 – Normalise the census variable being mapped by **dividing it by the total geographic area** in the small area. This expresses the variable as a density.
- Option 2 – Normalise the census variable being mapped **by dividing it by the total population size** (people or households) in the small area. This expresses the variable as a proportion or percentage.
- Option 2 where we *express the census variable as a proportion of the total number of people or households* in the small area is usually the approach that should be used to normalise census aggregate data prior to mapping it

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	WdGeoID	rentfree	othprivrent	othsocrent	owsoutright	mortgage	privrent	socrent	sharedowner	total	rentfree_prop	othprivrent_prop	othsocrent_prop	owsoutright_prop	mortgage_prop	privrent_prop	socrent_prop	sharedowner_prop	dominant_ten_type
2	E05011384	11	166	281	3251	4020	1109	840	63	9741	0.112924751	1.704137152	2.884714095	33.37439688	41.26886357	11.38486808	8.623344626	0.646750847	mortgage
3	E05011385	37	191	949	2102	2451	3353	2686	47	11816	0.313134733	1.616452268	8.031482735	17.78943805	20.74306026	28.37677725	22.73188896	0.397765741	privrent

Add a categorical summary variable

- The census aggregate data downloaded from Data Explorer on accommodation type will for each electoral ward tell us across a range of different types of accommodation type the number of households that live in that accommodation type
- To the data add a categorical summary variable that looks across all of the different accommodation types to find the accommodation type in which the most households live in – the *dominant* accommodation type

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	WdGeoID	rentfree	othprivrent	othsocrent	owsoutright	mortgage	privrent	socrent	sharedowner	total	rentfree_prop	othprivrent_prop	othsocrent_prop	owsoutright_prop	mortgage_prop	privrent_prop	socrent_prop	sharedowner_prop	dominant_ten_type
2	E05011384	11	166	281	3251	4020	1109	840	63	9741	0.112924751	1.704137152	2.884714095	33.37439688	41.26886357	11.38486808	8.623344626	0.646750847	mortgage
3	E05011385	37	191	949	2102	2451	3353	2686	47	11816	0.313134733	1.616452268	8.031482735	17.78943805	20.74306026	28.37677725	22.73188896	0.397765741	privrent

Demo 2: Prepare downloaded census aggregate data for mapping

- Spreadsheet applications can be used to carry out data manipulations on census aggregate data provided as tabular data arranged in rows and columns
- Microsoft Excel
- LibreOffice Calc
- Here we will open the census aggregate data for Leeds that we downloaded from Data Explorer as a CSV file and manipulate the data in LibreOffice Calc to prepare the data for mapping in QGIS later:
 - Transform data from long to wide form
 - Normalise census variables
 - Add a categorical summary variable

Geospatial data and GIS

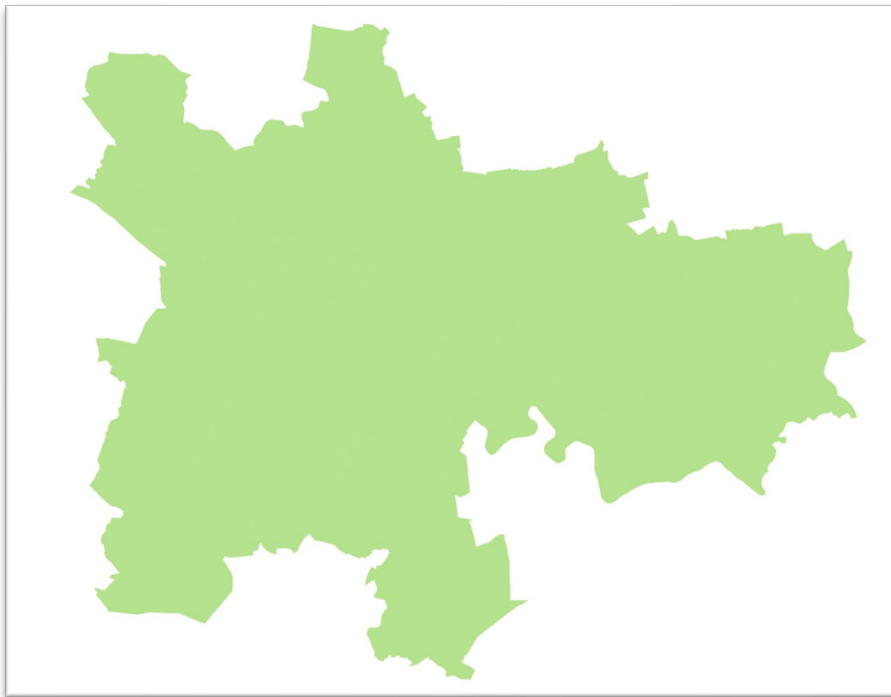
- Geospatial data is used to model some aspect of the world.
- Vector and Raster data
- Spatial reference systems
- Spatial relationships between geospatial features
- Geospatial data formats
- Geospatial standards
- GIS: Geographical Information System (Science)
- GIS Software applications
- ArcGIS proprietary family of GIS applications
- QGIS is an Open Source desktop GIS application

Census output geographies

- Census aggregate data is output for a range of different small area geographies
- This includes geographies that exist for electoral or administrative purposes e.g. electoral wards or local authorities
- As well as geographies created specifically for the output of census and other small area statistics e.g. census output areas and lower/middle layer super output areas
- Together these census output geographies form a census output geography hierarchy
- Statistical disclosure control means that not all census topics and variables are provided at all census output geographies

Census boundary data

- A census boundary dataset is a geospatial dataset that through a set of polygonal boundaries describes the shape and location of instances of small area census output geographies



Council Area boundaries

Source: Office for National Statistics licensed under the Open Government Licence v.3.0
Contains OS data © Crown copyright and database right 2026

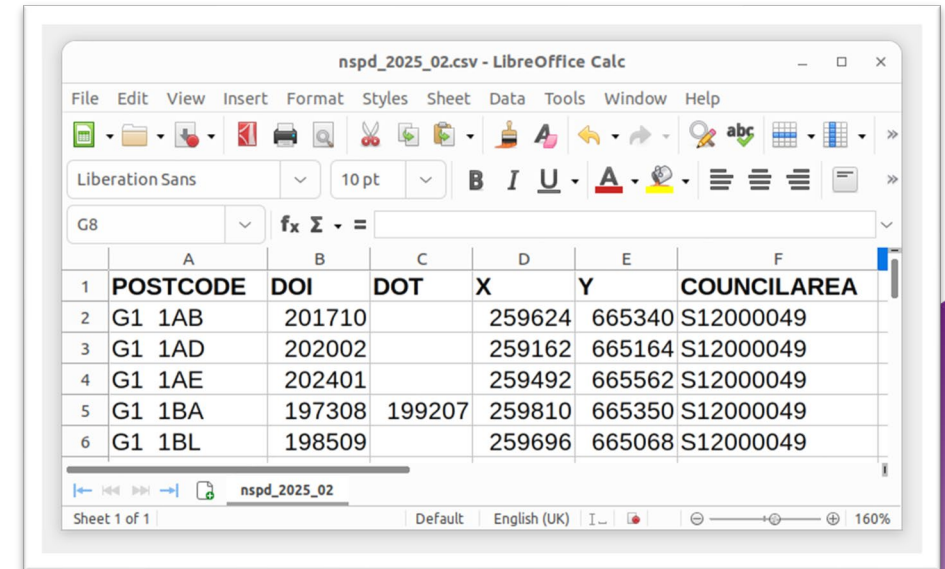


Census Output Area boundaries

Source: Office for National Statistics licensed under the Open Government Licence v.3.0
Contains OS data © Crown copyright and database right 2026

Geographic look up tables

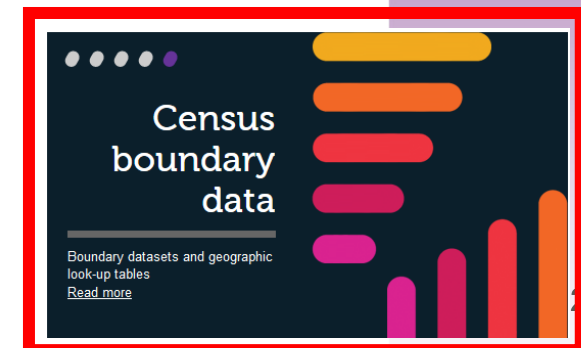
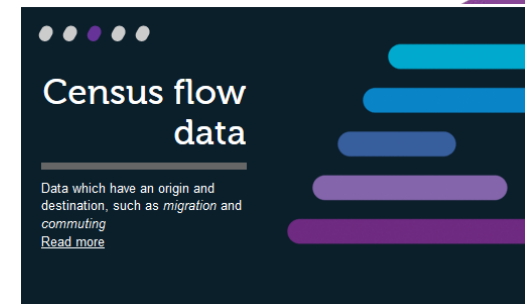
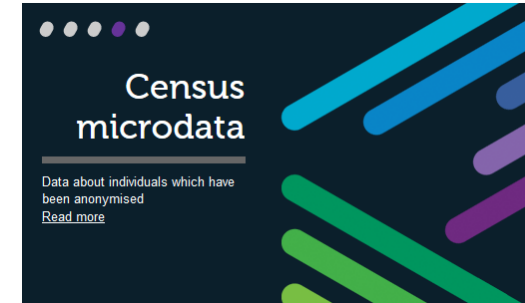
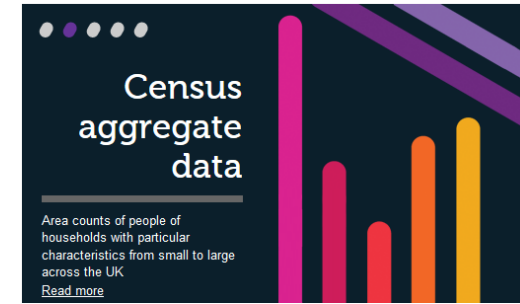
- A geographic look up table is a dataset which provides information on the spatial relationships between 2 or more types of geography
- A geographic look up table can be used to convert data from one type of geography to another facilitating *georeferencing* and *geodata linkage*
- Geographic look up tables are produced for the census providing lookups between different census geographies
- ONS postcode directories are a type of geographic look up table that also include census geographies



	A	B	C	D	E	F
1	POSTCODE	DOI	DOT	X	Y	COUNCILAREA
2	G1 1AB	201710		259624	665340	S12000049
3	G1 1AD	202002		259162	665164	S12000049
4	G1 1AE	202401		259492	665562	S12000049
5	G1 1BA	197308	199207	259810	665350	S12000049
6	G1 1BL	198509		259696	665068	S12000049

UKDS Census boundary data download tools

- Easy Download
- Boundary Data Selector
- Postcode Directory Download
- Postcode Data Selector



Demo 3: Downloading census boundary data using the UKDS Boundary Data Selector

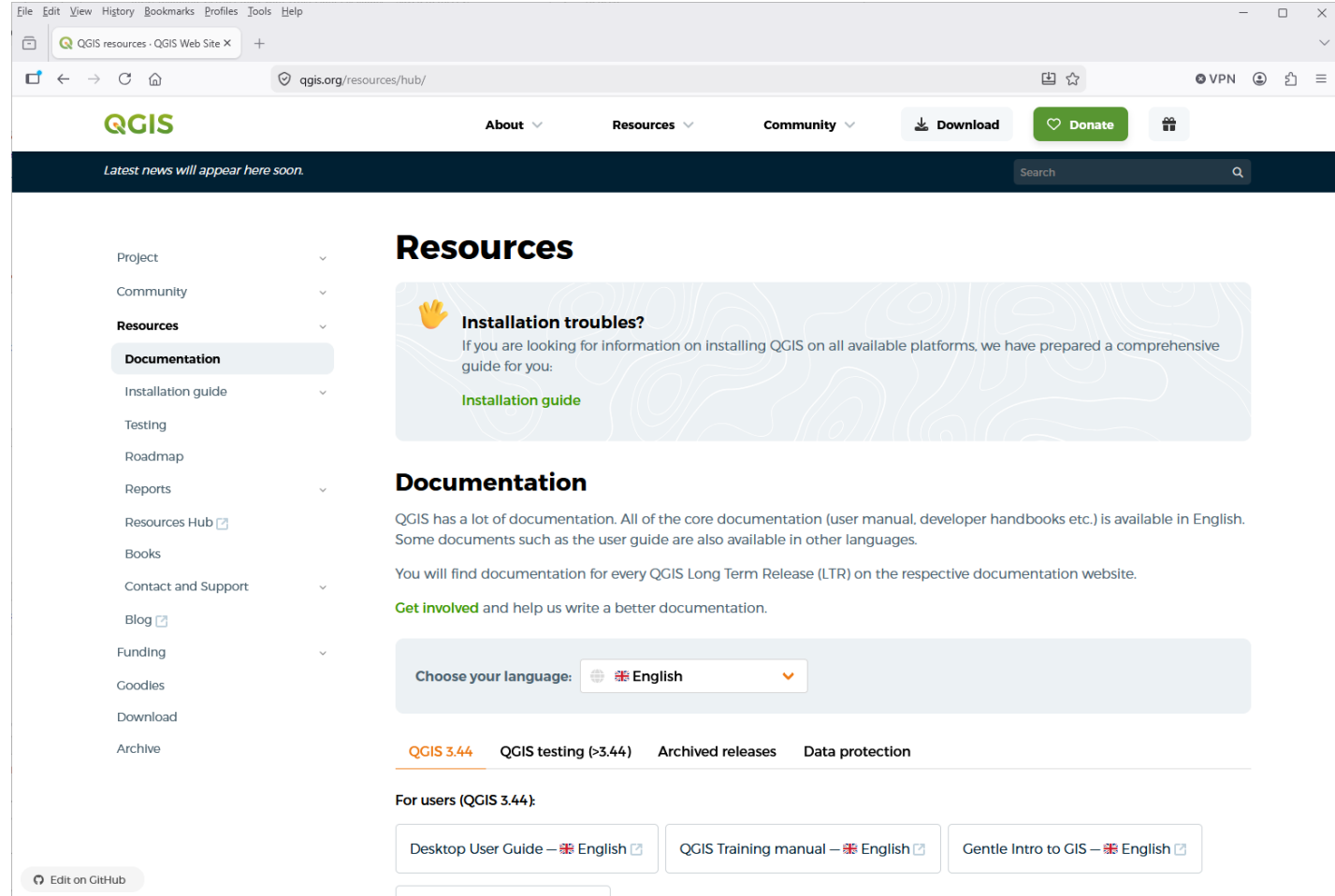
- Use the UK Data Service Boundary Data Selector to download census boundary data of electoral wards for Leeds

Demo 4: Downloading map data from the Ordnance Survey

- Download from the Ordnance Survey background mapping that when added to QGIS will provide context to our census data

Demo 5: introduction to the QGIS desktop GIS

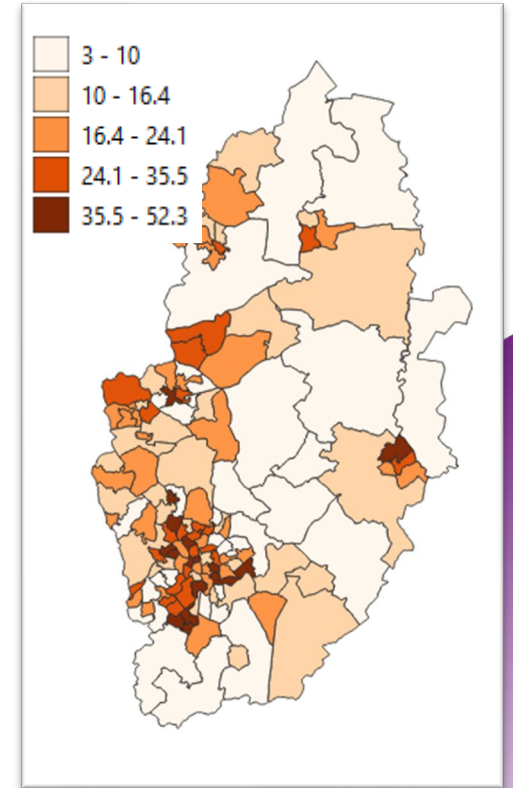
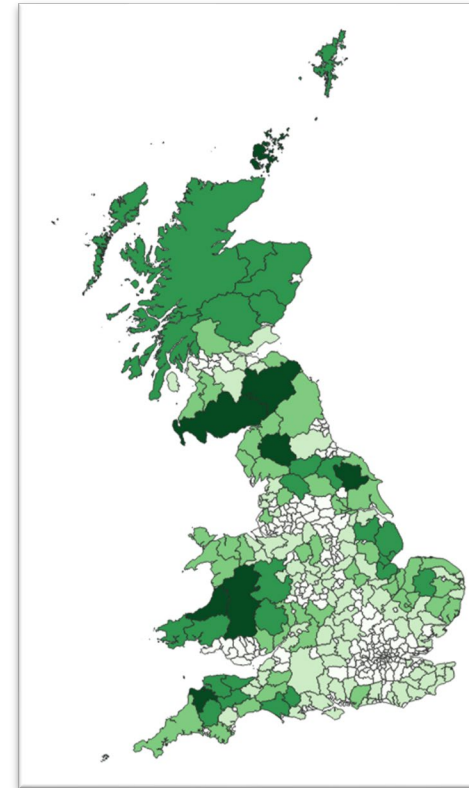
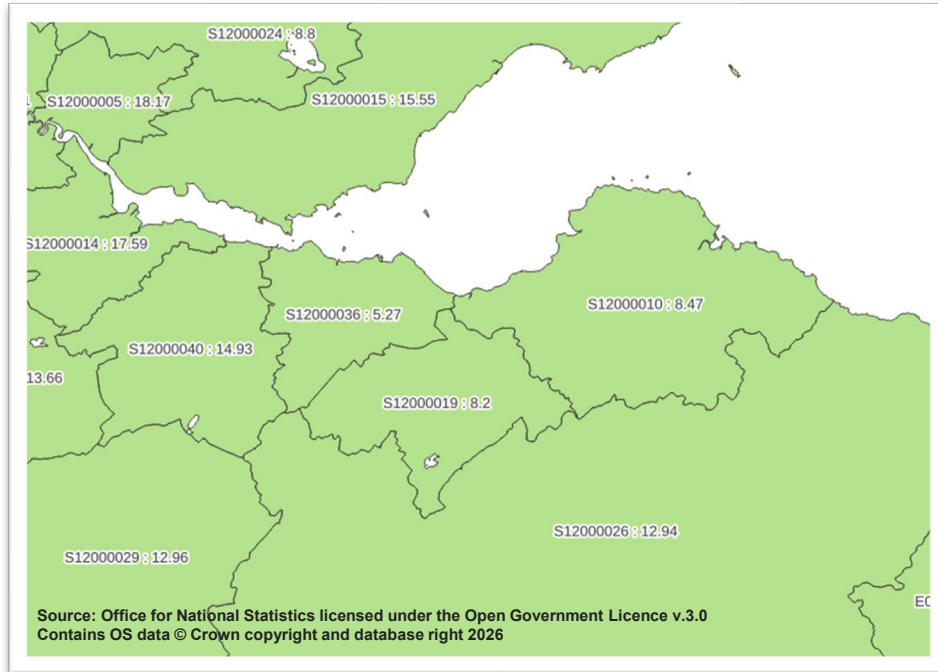
- In this demo we will now start QGIS and open the data we downloaded from the UKDS and the Ordnance Survey
- The Online QGIS Hub has lots of well-written and helpful documentation which it is worth spending time with to learn all about using QGIS



A Short Break

- Having downloaded from the UK Data Service census aggregate data and census boundary data and background map data from the Ordnance Survey, carried out data manipulations on the census data and opened all of our data in QGIS in the rest of the session we will get on with using GIS to make maps and other geographical data visualisations from the census data.
- See you back soon to continue our mapping census data with QGIS journey!

The choropleth map



- Polygon features within a digital boundary dataset 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.

Limitations of the choropleth map

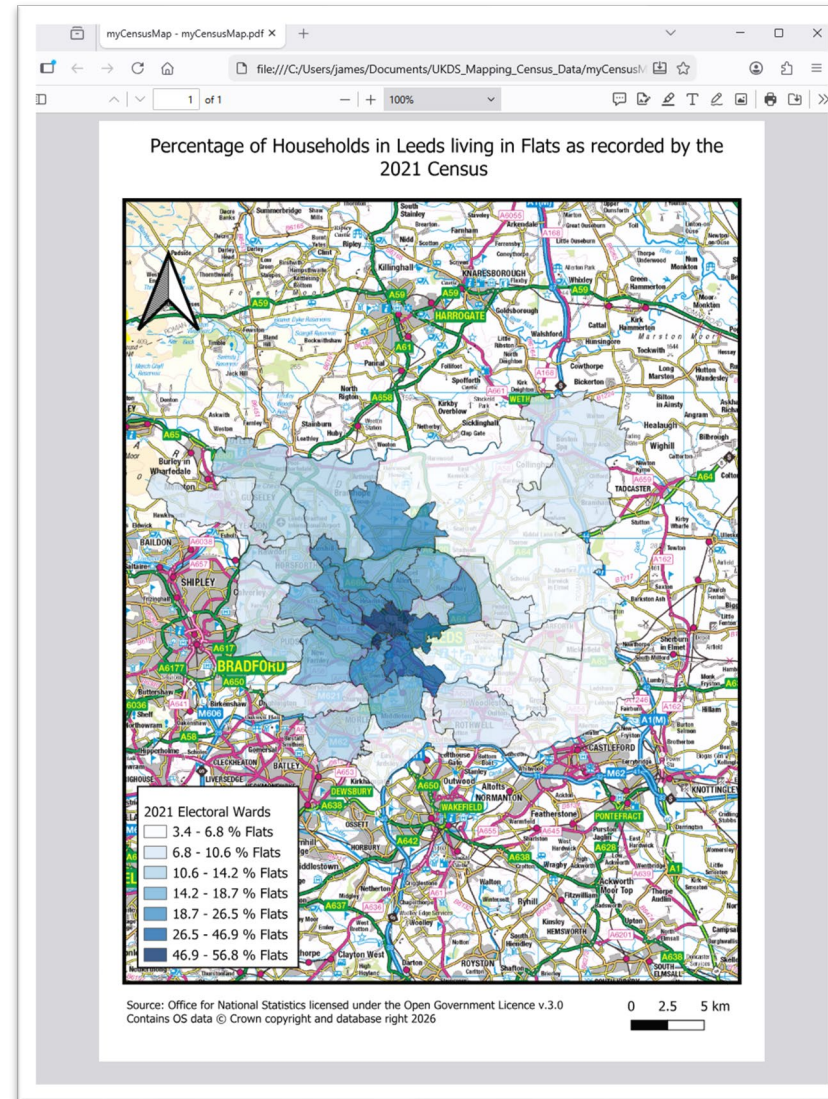
- Choropleth maps imply that the underlying population is distributed uniformly across the extent of the polygon
- Reality says otherwise
- However as long, as we are aware of this limitation, the choropleth map still provides a powerful means of exploring and communicating the geographic distribution of census variables



Source: Office for National Statistics licensed under the Open Government Licence v.3.0
Contains OS data © Crown copyright and database right 2026

Components of a choropleth map

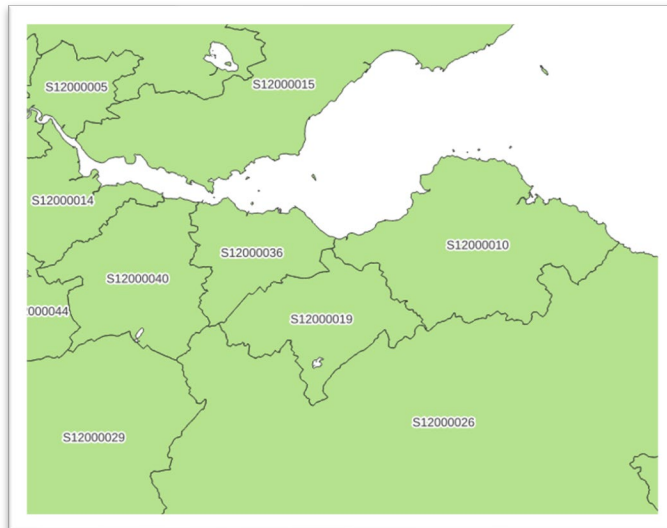
- Mapped census variable
- Census geography boundaries
- Census variable linked to the census boundaries
- Choropleth map classification
- Choropleth colour ramp
- Data attribution statements



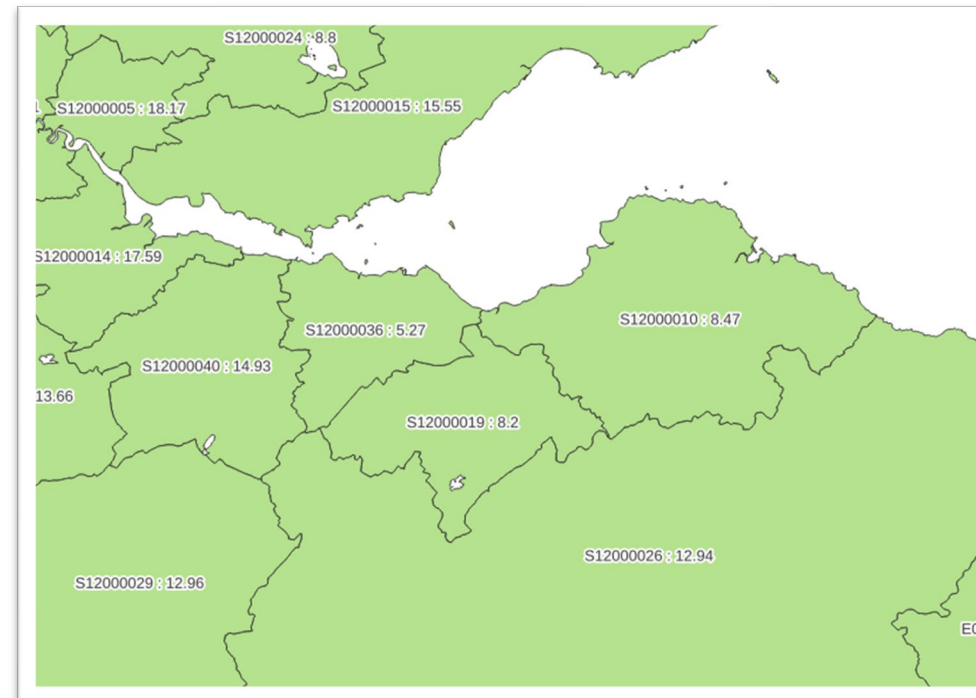
Linking census statistics to census boundaries

- Different tables of census statistics are output at different geographies and hence have different geographic identifiers (GeoIDs) associated with them

	GEOID	GEONAME	M	manur_m	manur_m_pcmt
174	S12000036	Edinburgh, City of	121562	6407	5.270561524
175	S12000038	Renfrewshire	41219	5691	13.80673961
176	S12000039	West Dunbartonshire	20402	2644	12.95951371
177	S12000040	West Lothian	44709	6677	14.93435321
178	S12000041	Angus	28723	4446	15.47888451



Source: Office for National Statistics licensed under the Open Government Licence v.3.0
Contains OS data © Crown copyright and database right 2026



Source: Office for National Statistics licensed under the Open Government Licence v.3.0
Contains OS data © Crown copyright and database right 2026

Demo 6: linking census statistics to census boundaries in QGIS

- Using the Geographic Identifier (GeoID) code for electoral wards that is present in both the census aggregate data downloaded from Data Explorer and the census boundary data downloaded from Boundary Data Selector we will *link* the two separate datasets together by way of a **Table Join** in QGIS

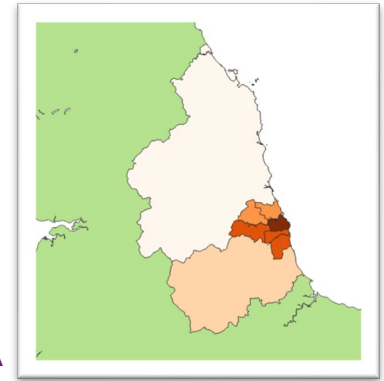
Consideration of choropleth mapping choices

- Choice of census output geography
- Choice of choropleth map classification method
- Choice of choropleth colour ramp

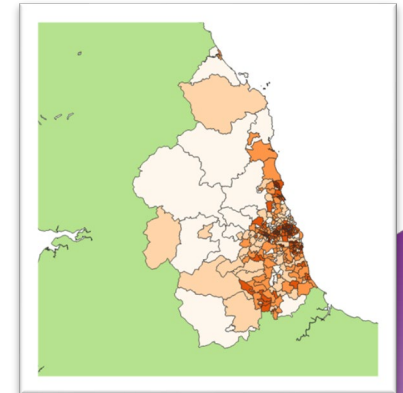
Choice of census output geography

- UK census data is available at different levels of output geography.
- Disclosure control means that not all census variables are available at all levels
- Mapping / analysing data at different levels of output geography may lead to different insights
- Noise in the data when presented at census Output Area level is smoothed when presented at Lower and Middle layer Super Output Area levels

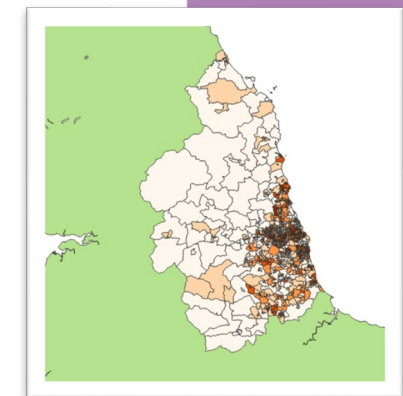
MSOA



LSOA



Census OA



Numerical data classification

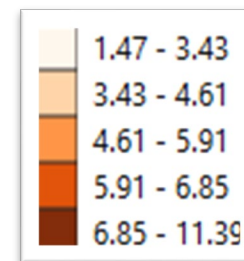
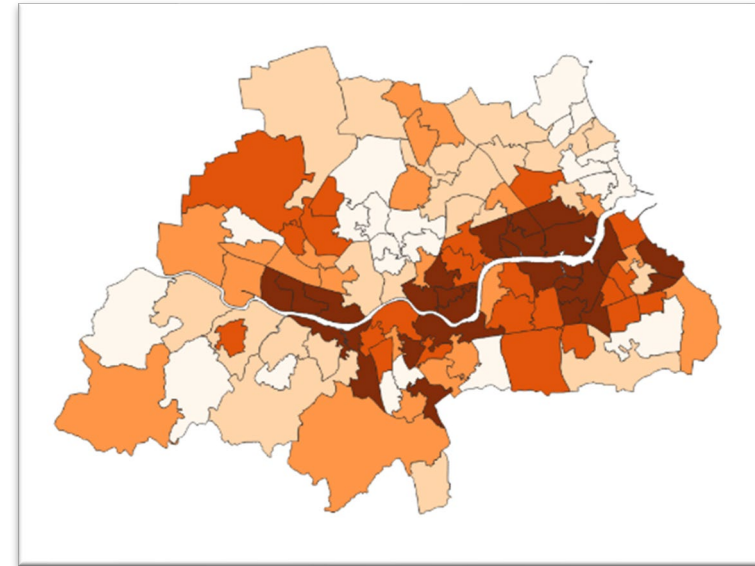
- A choropleth map classification is a special set of numerical data classification methods that is used to classify small area statistical data
- We can take the full range of data values held in any numerical dataset and classify them into a set of data buckets referred to as data classes
- Each data class has a minimum and maximum value referred to as the limits or ranges of that data class
- This classification of the numeric data into a set number of data classes simplifies the data and helps us to recognise and explore patterns present in the data
- Each polygon in a choropleth map is shaded a different colour depending on which data class the value of the underlying census data variable falls within

Data Value	Class
3	3 to 20
4	3 to 20
5	3 to 20
18	3 to 20
20	3 to 20
22	21 to 38
30	21 to 38
32	21 to 38
39	39 to 56
53	39 to 56
57	57 to 74
60	57 to 74
61	57 to 74
62	57 to 74
64	57 to 74
64	57 to 74
67	57 to 74
77	75 +
81	75 +
91	75 +
93	75 +

A simple **equal interval** classification of some numerical data

Choropleth map data classification methods

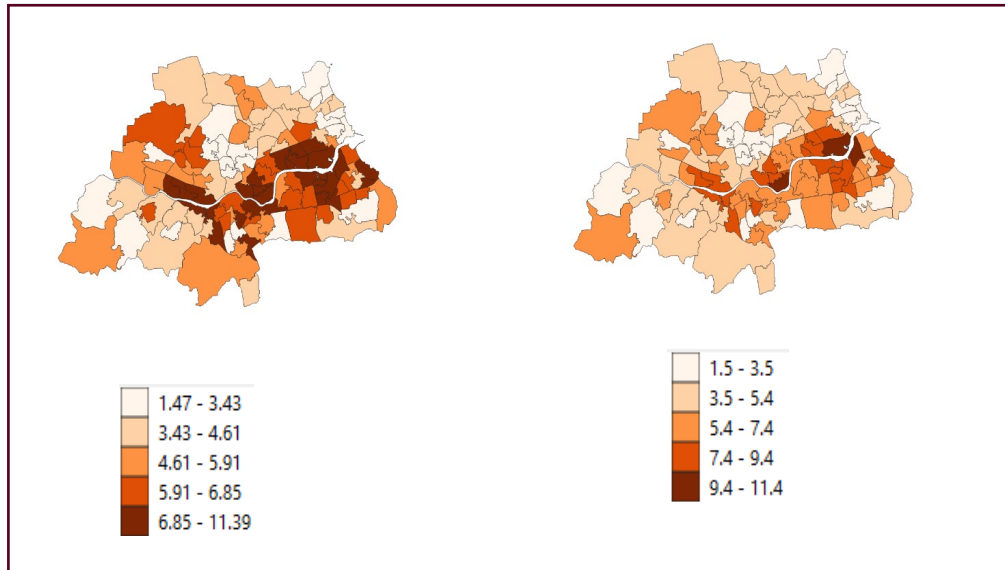
- Equal Count (Quantile)
- Equal Interval
- Fixed Interval
- Logarithmic Scale
- Natural Breaks (Jenks)
- Pretty Breaks
- Standard Deviation
- User Defined



Source: Office for National Statistics licensed under the Open Government Licence v.3.0
Contains OS data © Crown copyright and database right 2026

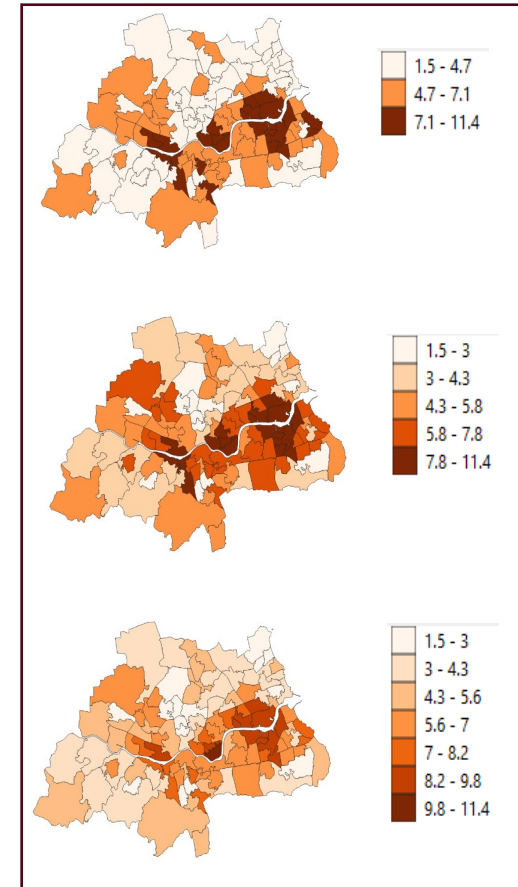
Same dataset – different choropleth maps

- Applied to the same dataset using different choropleth map classification methods or selecting a different number of data classes will produce a different map.
- **No classification method is right or wrong.**
- The choice of choropleth map classification method that is selected should be based on understanding the characteristics of the census variable being mapped with the aim of avoiding constructing a misleading map to mislead



Here in each map the **same number of choropleth map classes** are used but a **different choropleth map classification method** is used in each map

Source: Office for National Statistics licensed under the Open Government Licence v.3.0
Contains OS data © Crown copyright and database right 2026



Here the **same choropleth map classification method** is used but the **number of map classes used varies**

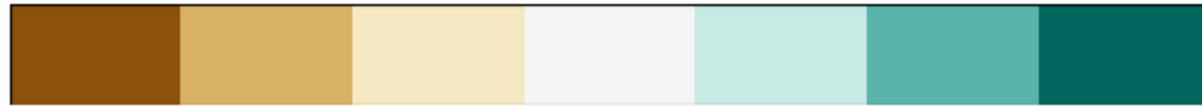
Source: Office for National Statistics licensed under the Open Government Licence v.3.0
Contains OS data © Crown copyright and database right 2026

Choice of choropleth colour ramp

Sequential



Diverging



Qualitative



Demo 7: Creating categorical and choropleth maps in QGIS

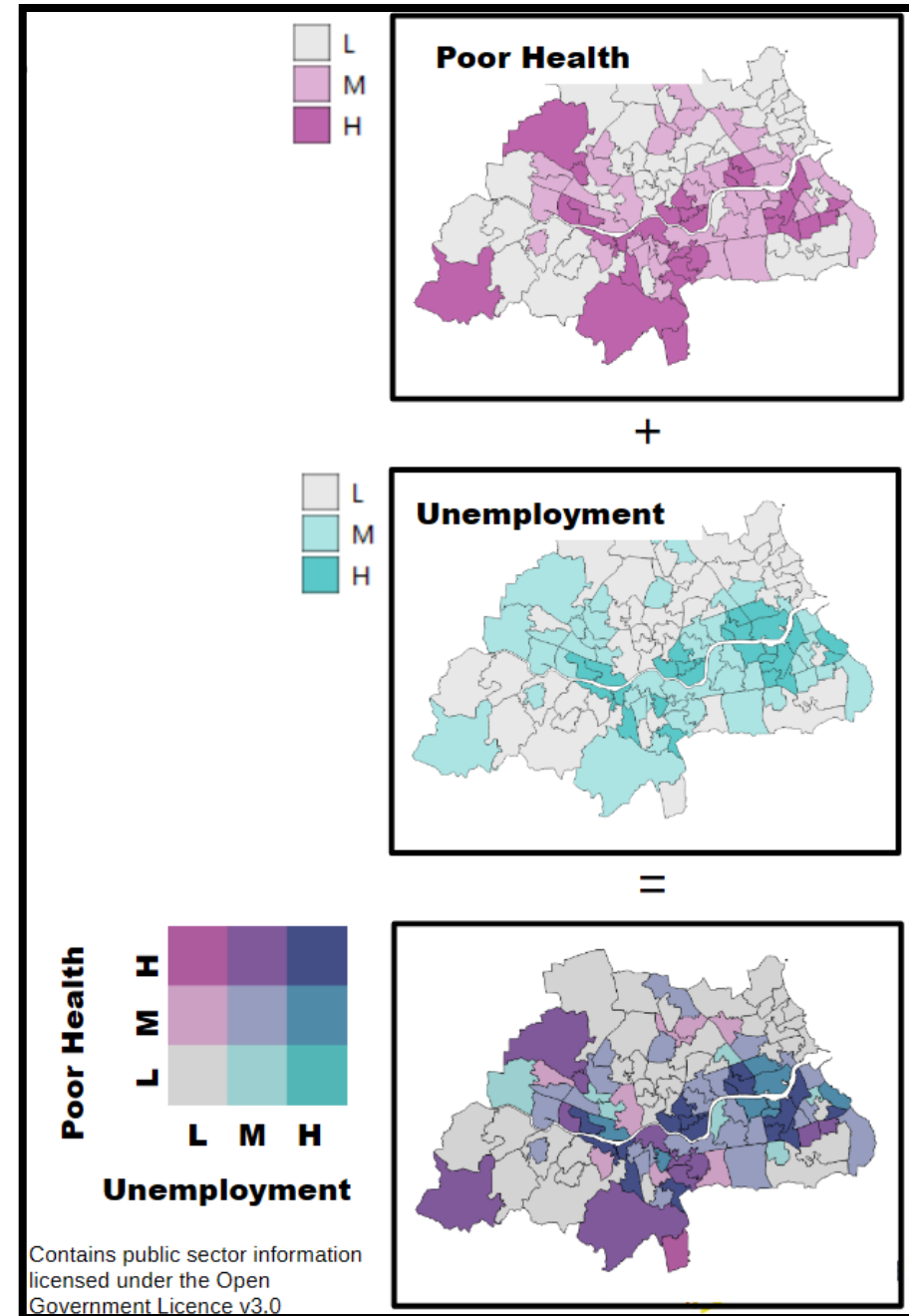
- Use QGIS to create a categorical map and a univariate choropleth map from 2021 census data downloaded from the UK Data Service displaying information about the type of accommodation that people in Leeds are living in.

Other types of census map

- Bivariate choropleth maps
- Cartograms
- Dasymetric and masked choropleth maps

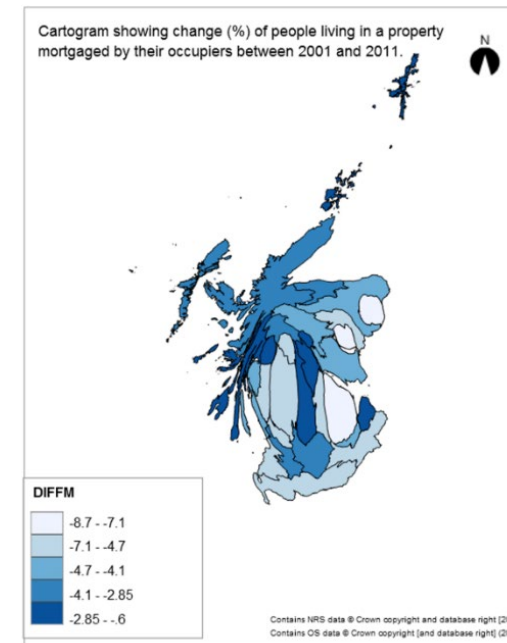
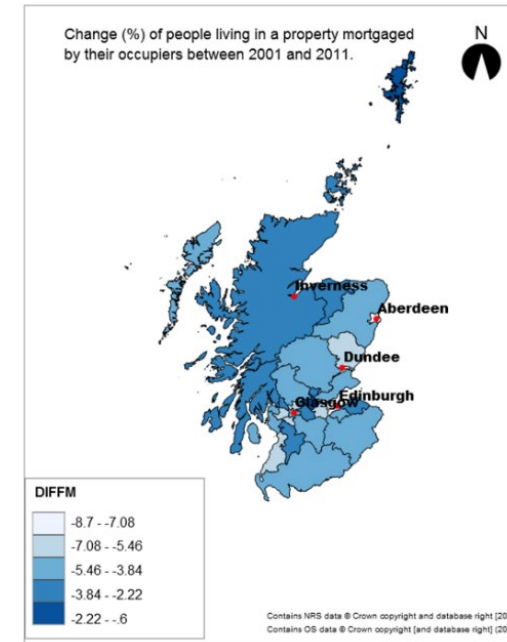
Bivariate choropleth maps

- Most choropleth maps display a single variable and are known as univariate choropleth maps
- Bivariate choropleth maps combine data from two variables
- For Newcastle and Tyneside combine into a single map variables from the census on poor health and unemployment.



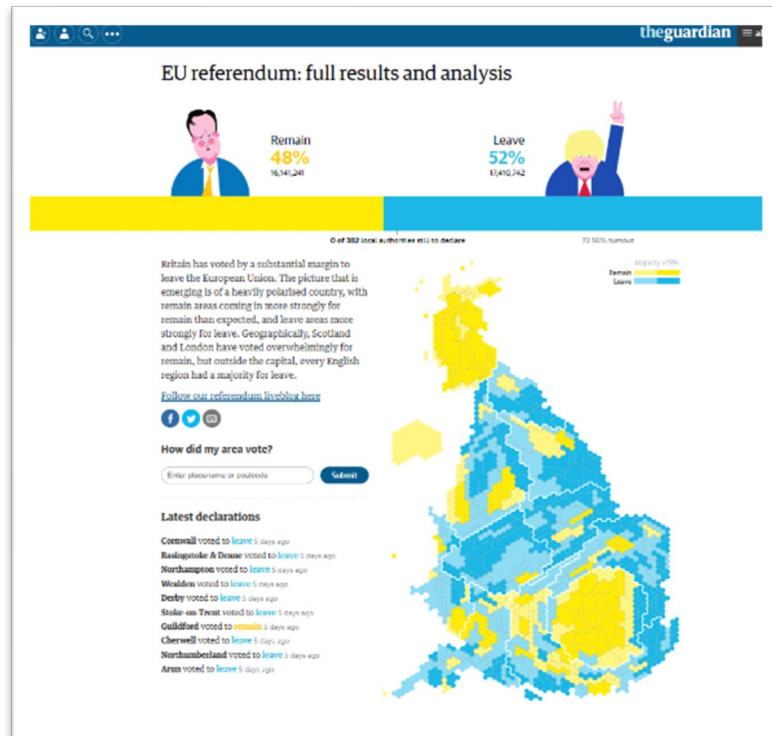
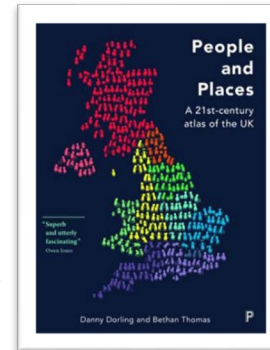
Cartograms

- 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



Cartograms in the wild

- 2016 Brexit referendum cartograms
- Census cartograms

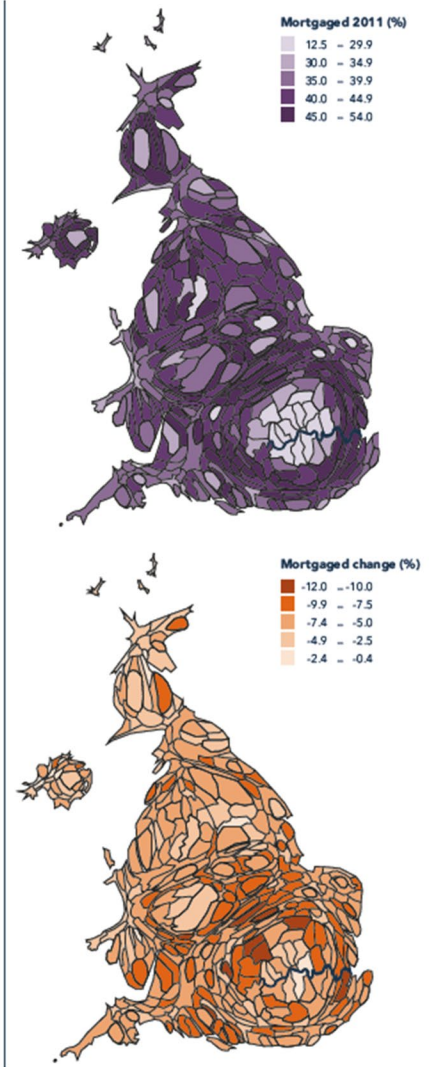
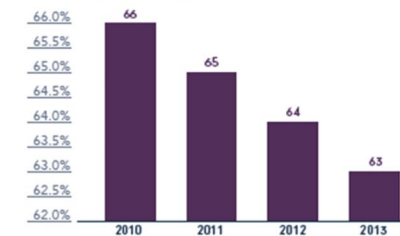


MORTGAGED

Having a mortgage here includes having a loan that, even if you pay it all off, will only ensure you have part ownership of a property. This is called *shared ownership*, a combination of buying and renting, and it is becoming increasingly popular because the cost of property in many areas is now so high. The number of people living in a home on which they had a mortgage fell from 27.0 million in 2011 to 25.0 million in 2013. That fall of 2 million people occurred at a time when the overall population was rising, especially of those in middle age who might be expected to have mortgages. The largest and most dramatic falls were recorded in Luton (-12.0%); Slough (-10.9%); Enfield (-10.8%); Milton Keynes (-10.7%); Brent (-10.6%); Eastleigh (-10.4%); Medway (-10.2%); Redditch (-10.1%); and Harrow (-10.0%). Nowhere experienced a rise, with the slowest declines being in Southwark (-0.4%); Falkirk (-0.6%); Strabane (-1.3%); and Barnsley (-1.5%).

By 2011 there were only 10 areas where more than half the population lived in a mortgaged property, and even then in each case it was only just a majority: East Renfrewshire (54.0%); Wokingham (52.3%); Blaby (52.1%); Carrickfergus (51.7%); East Dunbartonshire and Newtownabbey (both 51.0%); Rochford and South Northamptonshire (both 50.9%); Hart (50.5%); and Castlereagh (50.3%). By 2011 less than a fifth of the population lived in a mortgaged property in the Isles of Scilly (12.5%); Westminster (13.7%); Kensington & Chelsea (15.0%); Camden (16.1%); Tower Hamlets (16.5%); Hackney (17.7%); Islington (18.3%); and the City of London (18.8%). Nationally the proportion of dwellings that are both owned outright and mortgaged by their occupiers has fallen rapidly since 2010, from just under two-thirds of all dwellings to less than 63.5% in 2013. As yet there is no sign of a slowdown in this fall, and it is all due to the decline in the number of those with a mortgage.

Owner occupied dwellings, UK %



Demo 8: Creating a bivariate choropleth map and an area cartogram in QGIS

- Use QGIS to create a bivariate choropleth map from 2021 census data downloaded from the UK Data Service which displays both accommodation type and tenure type on the same map.
- Use QGIS to create an area cartogram from 2021 census data downloaded from the UK Data Service.

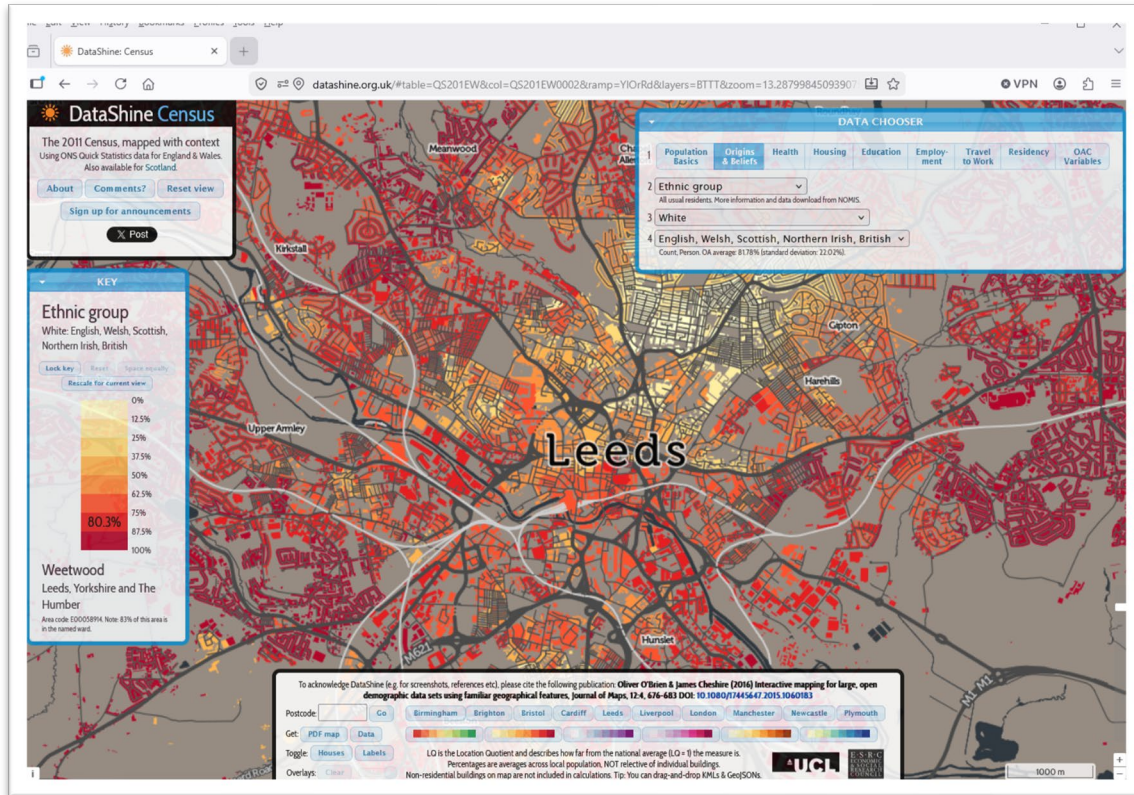
Dasymetric and masked choropleth maps

- A dasymetric map modifies the traditional choropleth map by using additional land use datasets to display a more realistic distribution of the mapped variable
- Within each census boundary polygon describing a given small area, the associated census aggregate statistic is spatially redistributed to different land use regions within the small area rather than being output across the entire small area
- A simpler form of dasymetric map simply *masks* the choropleth map using a buildings layer such that only the buildings within each small area are shaded
- Either of these approaches can help with the problem of the traditional choropleth map that implies that population is uniformly distributed across the small area
- Creating a dasymetric or masked choropleth map from UK census data can be done in QGIS but is a reasonably involved process

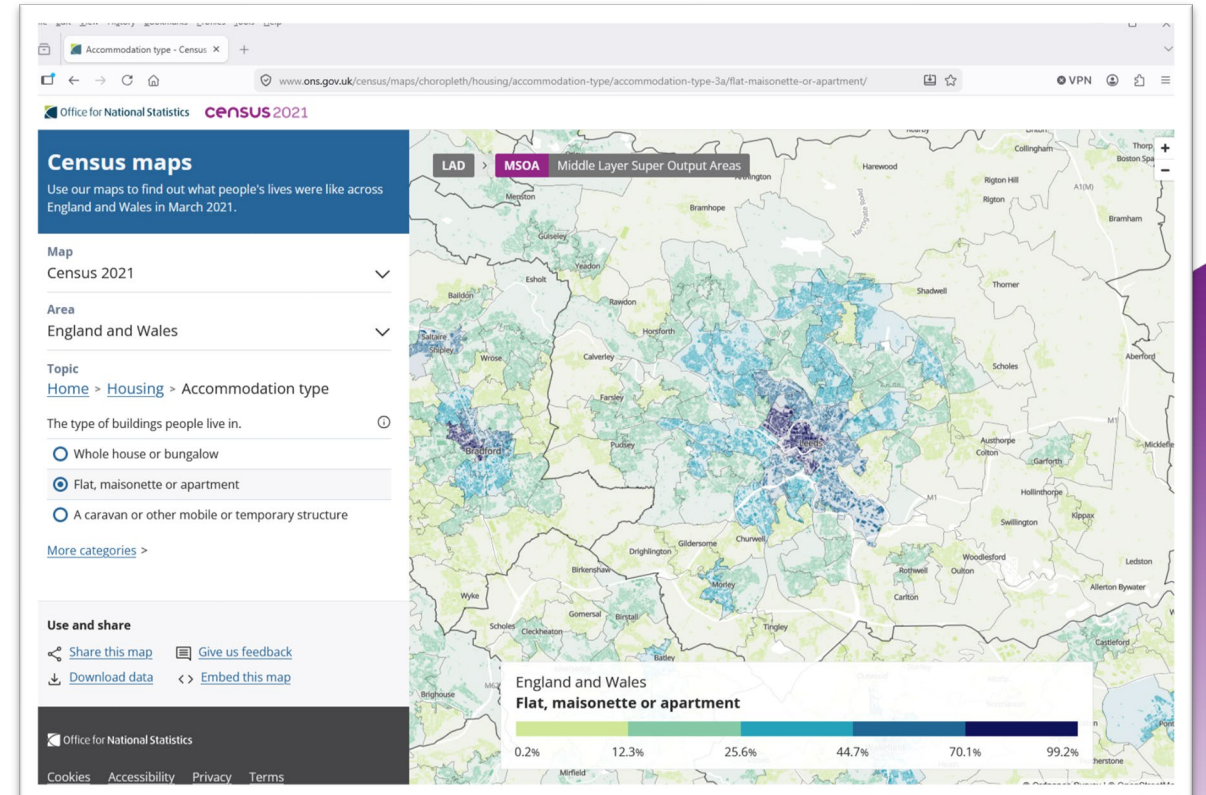


Source: Office for National Statistics licensed under the Open Government Licence v.3.0
Contains OS data © Crown copyright and database right 2026

DataShine and ONS Census maps



<https://datashine.org.uk>



<https://www.ons.gov.uk/census/maps/>

Helping you with census data

- UK Data Service training programme
- UK Data Service census learning modules
- UK Data Service helpdesk
- UK Data Service census data drop-in sessions:
 - Come along and discuss any questions you might have about what was shown today or any other questions you might have about census data with myself and other colleagues from the UK Data Service



Thank you.

James Crone

james.crone@ed.ac.uk

