UK Data Service

Census 2021

Developing an area profile using multivariate data

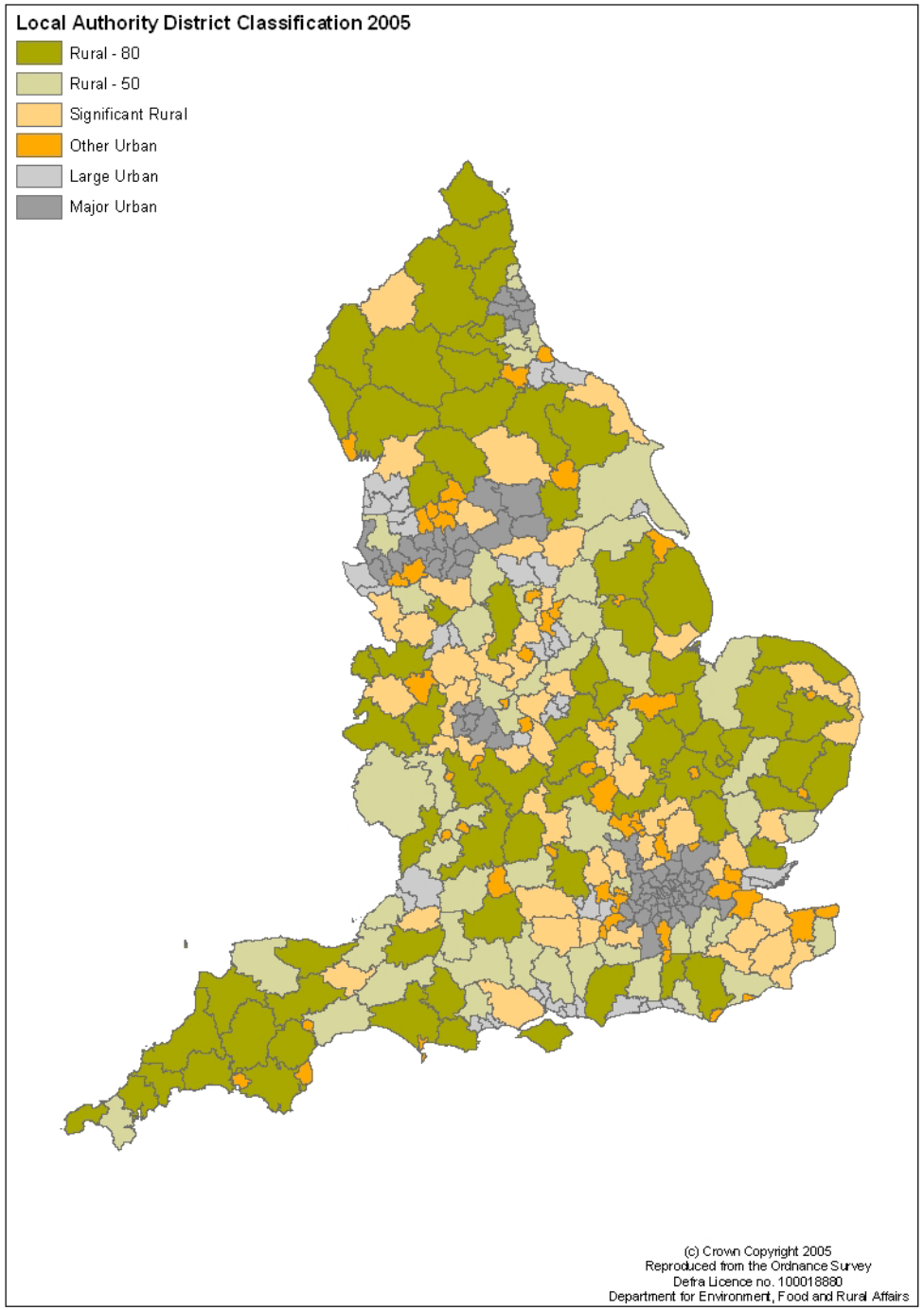
Economic and Social Research Council

UK Data Service

Census 2021

Developing an area profile using multivariate data

Nigel de Noronha, UK Data Service



UKRI Economic and Social Research Council

Agenda

Sources of information

Selecting your geography

Sources of data

Statistical disclosure control

Defined tables

Create a custom dataset

UK Data Service

Sources of information

Guidance on geography, measurement and variables

https://www.ons.gov.uk/census/census2021dictionary

Census methodology

- Response rate 97% with more than 88% in all local authorities
- Comprehensive QA process including invitation to all LAs to take part
- Estimates rather than counts
- Coverage survey used to estimate missing and double counting
- Students counted at their term-time address but likely to exclude some overseas students
- Temporary outflow from urban areas

Office for National Statistics (ONS), released 2 November 2022, ONS website, methodology, Quality and methodology information (QMI) for Census 2021

Quality information

Comparison of sex and age by single year to 2020 mid-year estimates and admin-based population estimates. Checking students, fertility and mortality rates, LSOA comparisons and exploration of outliers, communal establishments, migration flows, community QA (country of birth, ethnicity and religion), discussion with local area authority about the analysis

Individual level comparison of census and census coverage survey with the statistical population dataset showing incorrect exclusions and inclusions with age/sex breakdown, local authority level analysis grouped in the report, communal establishments

Individual level comparison of census and census coverage survey with administrative data with age and sex breakdown, regional statistics, LA matching considering military personnel, communal establishments, students

Local authority comparison between mid-year population estimates with census 2021, differences based on age, sex and geography, largest differences in London, large student populations and special population such as military personnel

Office for National Statistics (ONS), released 7 November 2022. *Quality assuring the local authority census population estimates, England and Wales*

Office for National Statistics (ONS), released 28 February 2023. *Understanding quality of the Statistical Population Dataset in England and Wales*

Office for National Statistics (ONS), released 1 March 2023. *Understanding quality of linked administrative data sources in England and Wales*

Office for National Statistics (ONS), released 28 February 2023, ONS website, article, Reconciliation of mid-year population estimates with Census 2021, England and Wales

UK Data Service

Selecting your geography

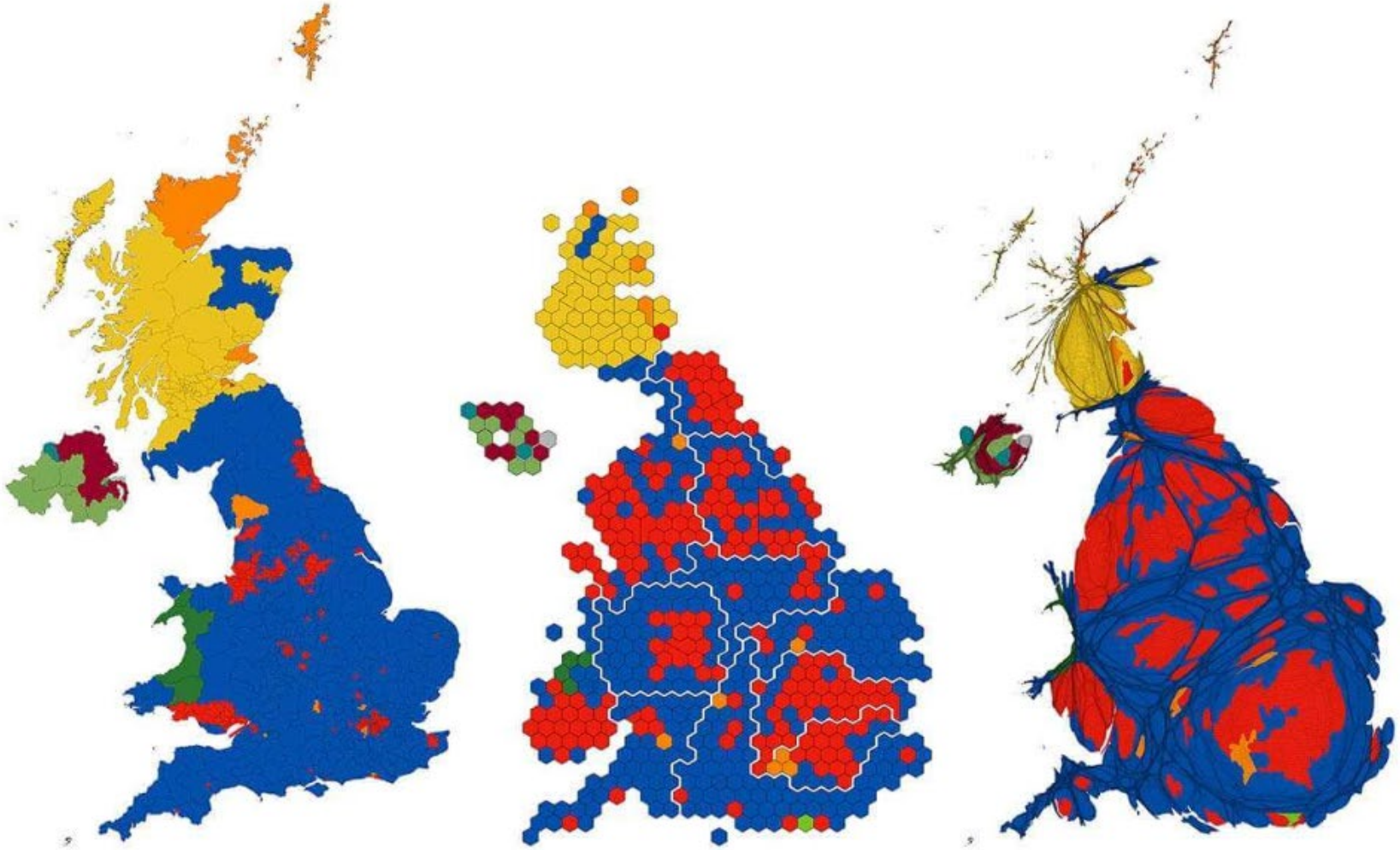
Administrative

Trafford wards and localities

From 04 May 2023

Electoral

2019 UK general election results



Physical space

Equal size constituencies

Cartogram – based on population size

Statistical

INTRODUCED in 2001

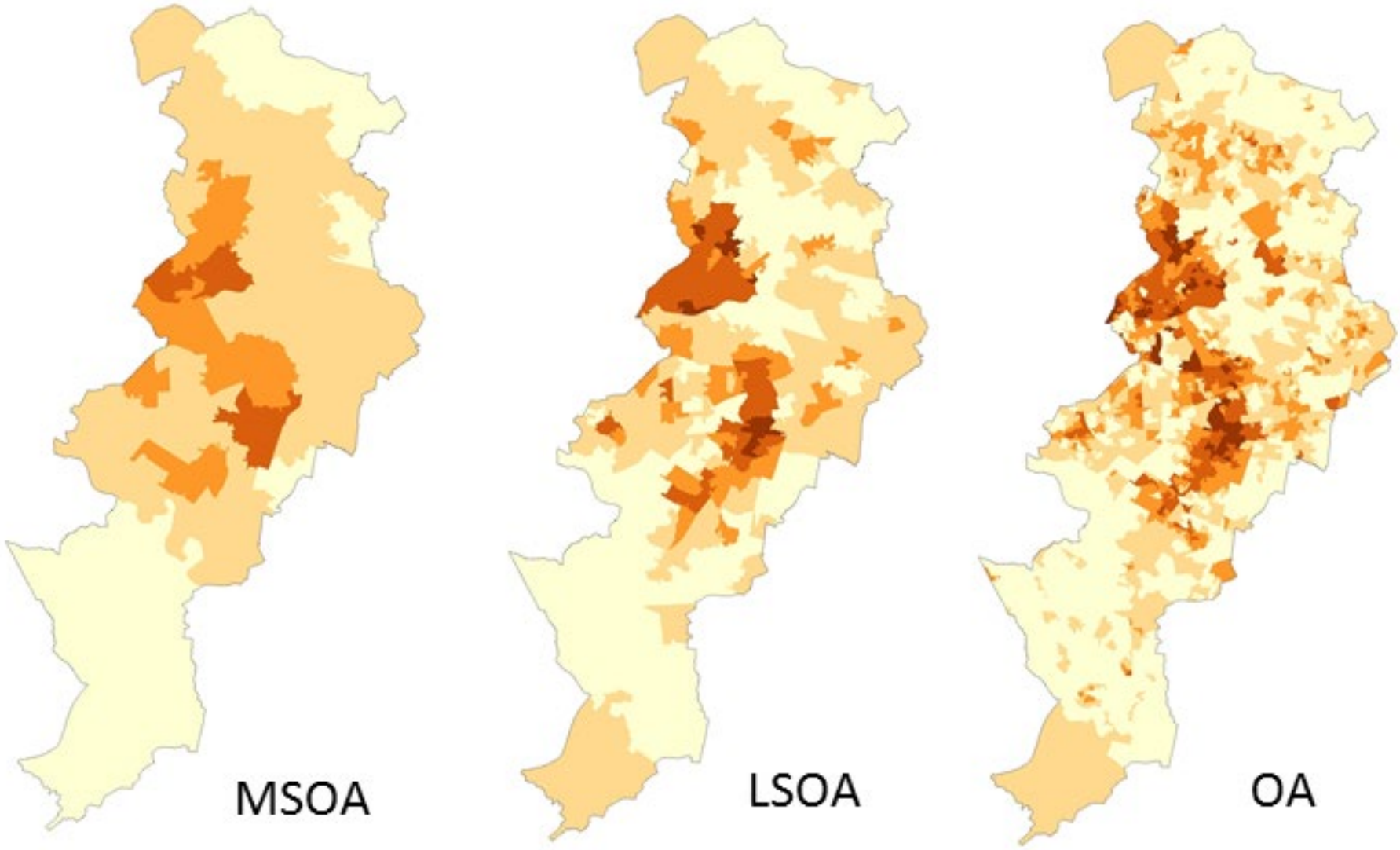
Designed to produce more homogeneous population by matching characteristics e.g. tenure, property type. Over time aim to to minimise changes. All boundaries constrained within local authorities

- Output area minimum 40 households and 100 residents, average target 125 households, maximum 250 households and 625 people

Built into larger super output areas

- lower layer super output areas (LSOAs) minimum 400 households and 1,000 people, maximum 1,200 households and 3,000 people – used in published statistics e.g. IMD, recorded crime
- Mid layer super output areas (MSOAs) minimum 2,000 households and 5,000 people, maximum 6,000 households and 15,000 people – used in published statistics e.g. educational attainment, Covid cases

Statistical geographies



Private rented sector in Manchester
Source: 2011 census

UK Data Service

Sources of data

Getting the data

UK Data Service, NOMIS and ONS provide topic summaries for individual variables.

For multivariate data the UK Data Service and NOMIS present a defined set of tables. They have taken the decision to present tables with minimum suppression. As a result, they use less detailed categories in many cases.

ONS interface allows you to select variables to create a custom dataset and select the number of categories. The interface is quite fast and informs you of levels of suppression due to statistical disclosure control.

Summary

Source	Data selection	Geographical selection	Data format	Dynamic tables
NOMIS	List of tables	Yes	Table(/s)	No
ONS	List of pre-defined tables with filters Build a custom dataset	Yes	Individual data holds each category combination and no of observations	Yes
UKDS	List of tables with filters	No	Individual data holds each category combination and no of observations	No

Our experiences?

I have mainly used the ONS interface as it supports the level of detail I need. Finding the right table in NOMIS and UKDS is time-consuming.

If you know the table UKDS may provide what you need. There are plans to develop this interface to offer geographical selection, improved filtering and more detailed categories.

Both ONS and UKDS data needs to be prepared as we will demonstrate later.

NOMIS provides tabular data formatted so that it can be used directly, but once you move to two variables with geographical breakdown there may be one produced for each category of one of the variables.

I will be using the ONS interface in this session.

UK Data Service

Statistical disclosure control

Making sure that respondents cannot be identified

Protecting confidentiality of respondents to census 2021 by dynamically

- Swapping records between areas – based on analysis of uniqueness mostly within local authorities
- Cell key method for each table changes values by adding or subtracting one or two to the counts leaving overall totals unchanged (with consistent results)
- Disclosure rules for tables – tested for likely disclosure risk and subjected to penetration testing

Office for National Statistics (ONS), released 9 March 2023, ONS website, methodology, Protecting personal data in Census 2021 result

Practical impact

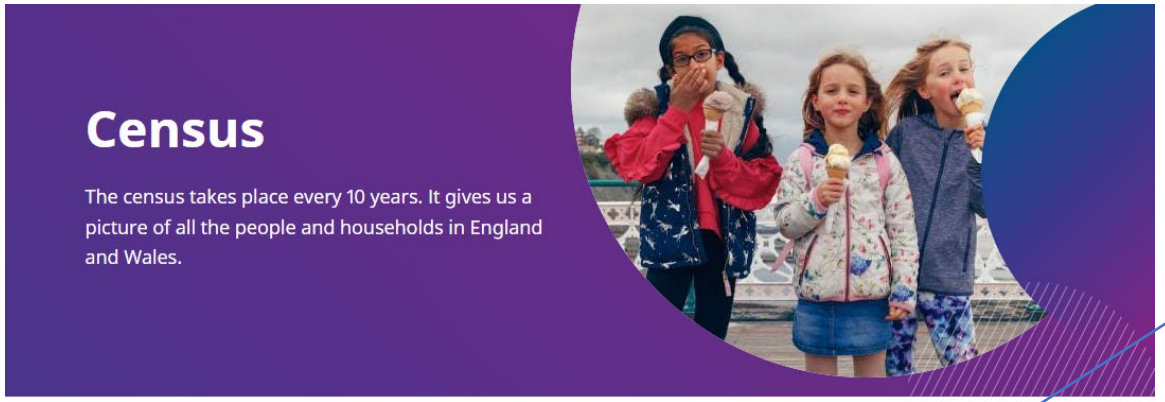
- Low frequency categories may lead to lots of empty cells e.g. when using ethnicity, the low number of gypsy / traveller and Roma are unevenly spread
- Increasing the number of areas with empty cells will increase the likelihood of their suppression

Strategy to address:

- Use selected geography
- Where areas are suppressed use a higher-level geography to cover the gaps

Census data from ONS

From the ONS front page select census 2021 releases and the following page is displayed



About the census

Find out what the census is and why it's important for all of us.

— About the census

Census 2021 data

Find data for Census 2021.

— Get census data

— Create a custom dataset

— Bulk data

Census releases

See what we've published, and our plans for the future.

— Release calendar

— First results and timeline

Geography

Find Census 2021 data for different areas.

— Find facts and figures about areas

— View census data on a map

— Build a custom area profile

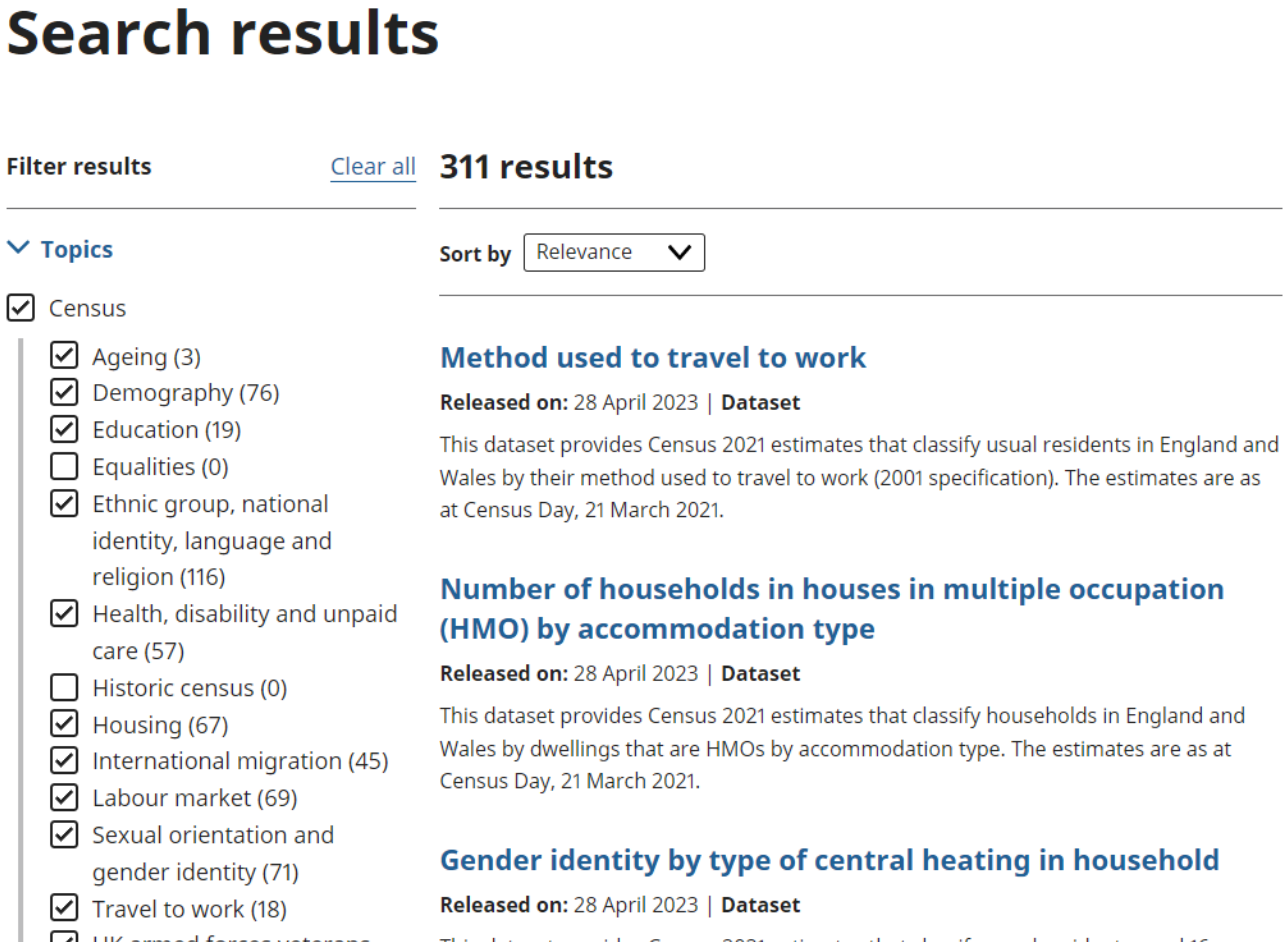
Selecting Get census data takes you to links to **standard tables**

Selecting Create a custom dataset takes you to an interface to select variables you want (the **flexible table builder**)

Selecting Bulk data takes you to links to download **topic summaries** (univariate data)

Standard (defined) tables

This option comes from Get census data on the first page and provides a search page



You can refine your results using the filters on the left, then select your area type and coverage. The variable categories are fixed for some of the tables.

Some of the data has been used for ONS reports prior to the application of the dynamic statistical disclosure control so will suppress more data. The majority will apply dynamic statistical disclosure control.

Build a custom dataset

This option comes from Create a custom dataset on the first page and generates a dialogue outlined below

Select population type

- ◯ Households
  Either one usual resident living alone or a group of people who share cooking and living facilities, where that group includes at least one usual resident.
- ◯ Household reference persons
  A person who serves as a reference point, mainly based on economic activity, to characterise a whole household.
- ◯ Usual residents in communal establishments
  A usual resident who lives in a place that provides managed full-time or part-time supervision of residential accommodation such as a university hall of residence, care home or prison.
- ◯ Usual residents in households
  A person who usually lives in England or Wales, and in a household.
- ◯ Usual residents
  The main population base for census statistics. It includes people usually living in England and Wales and excludes non-UK born short-term residents and visitors.

Continue

Selecting household variables gives access to household related variables such as household composition, tenure and deprivation.

In this example we will select *Usual residents in households* to explore housing deprivation by race and migration.

Housing deprivation by household composition and tenure

Household composition	Own outright	Own with mortgage or loan	Social rented	Private rented
single person under 66	1,226 (3%)	1,004 (2%)	2,926 (4%)	5,966 (9%)
couple no children	1,253 (1%)	1,060 (1%)	517 (2%)	2,905 (4%)
other family structure	584 (14%)	747 (19%)	1,680 (27%)	2,422 (23%)
other household	3,672 (13%)	4,601 (14%)	4,026 (24%)	17,891 (25%)
couple dependent children	13,998 (17%)	47,546 (9%)	48,021 (34%)	43,469 (24%)
lone parent dependent children	2,016 (16%)	7,136 (13%)	33,925 (27%)	19,900 (21%)
other household dependent children	16,235 (40%)	29,042 (43%)	21,586 (60%)	22,696 (62%)
couple non-dependent children	5,150 (5%)	5,664 (6%)	3,294 (11%)	2,077 (14%)
Lone parent non-dependent children	2,702 (7%)	1,930 (8%)	4,596 (12%)	2,098 (14%)
Single pensioner	2,260 (3%)	84 (2%)	1,924 (5%)	495 (5%)
Pensioner family	1,709 (1%)	65 (1%)	279 (2%)	160 (3%)

Greater Manchester, 2021 census

Housing deprivation by tenure

Ward	owns outright	Owns with mortgage / loam	Social rented	Private rented
Bramhall North	81 (2%)	190 (3%)	71 (16%)	110 (9%)
Bramhall South and Woodford	84 (2%)	127 (2%)	51 (19%)	53 (7%)
Bredbury and Woodley	219 (5%)	311 (5%)	238 (20%)	214 (13%)
Bredbury Green and Romiley	201 (4%)	280 (5%)	503 (19%)	117 (8%)
Brinnington and Central	205 (11%)	474 (11%)	1307 (17%)	608 (16%)
Cheadle and Gatley	222 (4%)	411 (5%)	193 (22%)	239 (12%)
Cheadle Hulme North	119 (3%)	328 (5%)	192 (15%)	172 (11%)
Cheadle Hulme South	91 (2%)	210 (3%)	123 (15%)	117 (7%)
Davenport and Cale Green	204 (7%)	435 (7%)	725 (19%)	293 (10%)
Edgeley and Cheadle Heath	207 (7%)	421 (7%)	419 (21%)	500 (16%)
Hazel Grove	153 (3%)	272 (4%)	229 (29%)	206 (11%)
Heald Green	144 (3%)	523 (8%)	127 (24%)	198 (13%)
Heatons North	141 (3%)	259 (4%)	203 (20%)	344 (17%)
Heatons South	193 (4%)	430 (6%)	181 (22%)	209 (11%)
Manor (Stockport)	198 (6%)	400 (6%)	346 (21%)	277 (11%)
Marple North	115 (2%)	190 (3%)	102 (19%)	66 (7%)
Marple South and High Lane	126 (3%)	231 (5%)	157 (17%)	125 (12%)
Offerton	116 (3%)	322 (5%)	380 (15%)	173 (12%)
Reddish North	219 (6%)	600 (11%)	405 (18%)	524 (15%)
Reddish South	203 (5%)	389 (7%)	295 (20%)	395 (16%)
Stepping Hill	112 (3%)	231 (4%)	29 (7%)	144 (9%)

UK Data Service

Demonstration

Building a custom dataset

You can always get what you want?

If these methods don’t work, you can either

- Consider commissioning a dataset from ONS
- Explore the microdata

Some notes on the data

Geography

- Households, household reference persons and usual residents– to Output Area level
- Residents in communal establishments and usual residents in households – to MSOA level

Statistical disclosure control

- There will be more limitations on the number of categories available as you look at smaller areas. Potentially significant issue with ethnic categories and household composition from my experience
- Develop a phased approach to answering the question to overcome the effects – high level geography, tables

Some notes on the data

Process

- Plan out what you want to find out (open-ended exploring might be interesting but you are more likely to waste time – I do)
- Document the steps you take
- Try to separate out calculations so that you do not overwrite the original data e.g. use formulae in Excel

Next steps

- Try it out for yourself
- Contact me if you have queries at nigel.denoronha@manchester.ac.uk

UK Data Service

Questions and discussion

nigel.denoronha@manchester.ac.uk

https://ukdataservice.ac.uk