

Building a custom dataset in census 2021

Using the ONS tools



[24/05/2023]

Copyright © Year University of Manchester. Created by Organisation, UK Data Service.

Version No. [001]



Table of contents

Accessing the data	3
Type of observations.....	3
Geographical scope	5
Selecting variables	6
Preparing the data	8
Missing values	8
Category changes.....	8
Category order	8
Analysing the data	10

Accessing the data

Type of observations

The data is available from the ONS census pages at [Create a custom dataset - Office for National Statistics \(ons.gov.uk\)](#). The first choice you need to make is the unit of observation.

Create a custom dataset

We group Census 2021 data together based on who or what the information is about, for example, people or households. We make population types from these groups or subsets of them. For example, people who are usually resident in England or Wales make up the population type usual residents. [Read about the measurements we used for Census 2021 data.](#)

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.

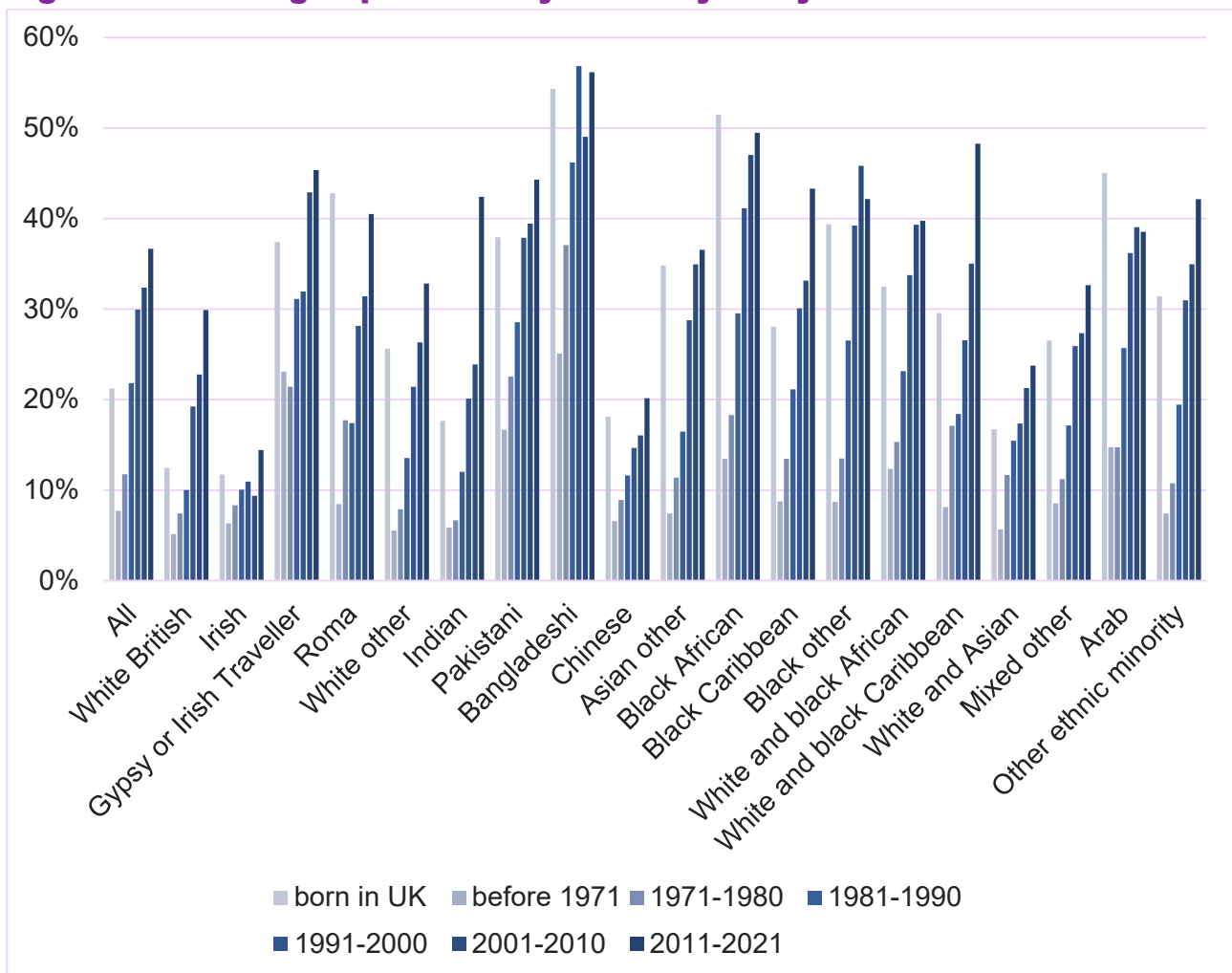
Continue

For tables with counts of households use the first option, for individuals use either *Usual residents in households* if you want to include household characteristics such as tenure, housing composition, deprivation measures or multi-ethnic households or *usual residents* if

you only want individual characteristics such as age, ethnicity, social class or sex. The other two categories may also be useful for specific investigations.

In the worked example we will use *Usual residents in households* to generate an analysis of housing deprivation by ethnicity and year of arrival in the UK for local authorities in London as shown in Figure 1.

Figure 1 – housing deprivation by ethnicity and year of arrival for London



Source: ONS custom table builder

Selecting the Usual residents in household and clicking the Continue button displays the following screen

Custom dataset

All 331 areas available

Variables

Population type	Usual residents in households	
Area type	Lower tier local authorities	Change
Coverage	England and Wales	Change

> **Variable and dataset information**

Add a variable

Geographical scope

We can leave the area type the same for London. Selecting <change> the area type would allow you to select England and Wales, countries, regions, NHS regions, Integrated care boards, parliamentary constituencies, wards or middle layer super output areas.

We will select all lower tier local authorities within a larger area. Select Regions, type London in the area name or code. Click <search>, then <add>, then <Continue> changes the scope to London.

Coverage

Select the geography you would like your dataset to cover

- All lower tier local authorities within England and Wales
- Search for lower tier local authorities by name
- Select all lower tier local authorities within a larger area

Larger area type
For example, a region or local authority

Regions
▼

Enter an area name or code

London

🔍 Search

Results

London

Add

Continue

Selecting variables

Clicking <Add a variable> allows you to search for or browse available variables. Use the browse option and identify the variables you want to use

- Ethnic group
- Household deprived in the housing dimension
- Year of arrival in the UK

Scrolling to the bottom of the screen and selecting <Continue> tells you how many of the selected areas are available for the breakdown of categories for each variable. Next to each variable is a <Change> link which allows you to select more or less detailed breakdown for ethnic group and year of arrival in the UK. The default of 20 ethnic groups and 11 periods covering year of arrival in the UK are fine in this case so we can move on the <Get the data>. If the screen does not have an XLSX option then refresh it and it will be displayed (the advantage is that this includes the metadata for the variables you have selected).

Once you have downloaded and opened the data, you will need to prepare it for analysis.

Preparing the data

Missing values

The first task is to check the missing values (-8) in the code for the data. For some variables these may have observations so it is worth checking to see if we want to include them in the analysis.

In Excel select <Data><Filter> and -8 under the ethnic code. We can check if there are any observations by checking the filter on the column observations. In this case they are all 0 so we can delete the rows selected. We now need to select all ethnic groups.

We can repeat this for housing deprivation and year of arrival. In both cases the number of observations is 0.

Category changes

We can now make any changes to the categories to match our analysis. In this case I want to combine the years of arrival since 2011 into a single value. We again use the filter and select codes 7,8,9 and 10. We then change this to 7 and the description to *Arrived 2011 to 2021*.

We can then copy this to the selected records. You could also simplify the ethnic descriptions.

Category order

I have presented the ethnic groups as shown below. The existing category code is shown in brackets:

- white (British (13), Irish (14), Gypsy or Irish Traveller (15), Roma (16) and other (17))
- Asian (Indian (3), Pakistani (4), Bangladeshi (1), Chinese (2) and other (5))
- Black (African (6), Caribbean (7) and other (8))
- Mixed (white and black African (10), white and black Caribbean (11), white and Asian (9), and, other (12))
- Other (Arab (18), Other ethnic minority (19))

To change these, you can again use the filter to change the existing categories. I used an alphabetic code (a-s) so that they appear in the required order when I carry out my analysis.

**Please save your work into your
own file space now**

Analysing the data

The quickest way to complete this analysis is to generate a pivot table. I use the code rather than the name so that the order of presentation matches my requirements.

- <Insert> <Pivot table> then place
- Local authority name in the Filter space
- Ethnic code in the Row variable
- Year of arrival and housing deprivation in the Column variable
- Observations into the Values variable

This will show a table like this

Sum of Observation	Column Labels				Column Labels				Column Labels				Column Labels			
Row Labels	0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3
a	2589554	368809	2958363	33150	1796	34946	14476	1162	15638	15299	1701	17				
b	57156	7586	64742	26335	1779	28114	7890	717	8607	12732	1421	14				
c	2945	1760	4705	40	12	52	44	12	56	104	47					
d	1828	1368	3196	130	12	142	130	28	158	209	44					
e	153901	53043	206944	25553	1497	27050	18504	1583	20087	25482	3990	29				
f	212264	45525	257789	41925	2618	44543	57574	4110	61684	25133	3437	28				
g	93399	57088	150487	9605	1921	11526	9048	2634	11682	7390	2952	10				
h	77784	92526	170310	3245	1087	4332	7834	4614	12448	15239	13084	28				
i	30861	6822	37683	4374	308	4682	9219	902	10121	6363	836	7				
j	77999	41683	119682	6525	524	7049	14005	1798	15803	17402	3430	20				
k	141480	150164	291644	5656	880	6536	8658	1938	10596	32056	13437	45				
l	160327	62516	222843	45354	4355	49709	6430	1000	7430	5461	1464	6				
m	72056	46756	118812	4074	389	4463	1173	183	1356	2152	777	2				
n	33860	16284	50144	419	59	478	575	104	679	1255	378	1				
o	81191	34082	115273	1445	128	1573	339	70	409	372	84					
p	81747	16438	98185	2167	130	2297	1097	145	1242	1236	226	1				
q	84316	30429	114745	2792	261	3053	1727	218	1945	2266	469	2				
r	24788	20304	45092	856	148	1004	3151	544	3695	5883	2035	7				
s	93995	43031	137026	12583	1010	13593	14877	1793	16670	17348	4189	21				
Grand Total	4071451	1096214	5167665	226228	18914	245142	176751	23565	200306	193382	54001	247				

You can clear the totals for each year of arrival by right ticking in the column and selecting the Subtotal box which has a tick against it. My approach was to create new tables in case I made any mistakes in the formula. The table shows the observations for those who did not experience housing deprivation (column 0 for each year of arrival). I generated a similar one for those experiencing housing deprivation.

	A	B	C	D	E	F	G	H
1	Not deprived							
2	Ethnic group	born in UK	before 197	1971-1980	1981-1990	1991-2000	2001-2010	2011-2021
3	White British	2589554	33150	14476	15299	30194	38969	59183
4	Irish	57156	26335	7890	12732	6711	8882	18389
5	Gypsy or Irish Traveller	2945	40	44	104	164	197	718
6	Roma	1828	130	130	209	942	2693	16852
7	White other	153901	25553	18504	25482	83542	210009	405484
8	Indian	212264	41925	57574	25133	25464	64599	81302
9	Pakistani	93399	9605	9048	7390	12544	25221	23914
10	Bangladeshi	77784	3245	7834	15239	8968	16372	22459
11	Chinese	30861	4374	9219	6363	9673	18938	36549
12	Asian other	77999	6525	14005	17402	35686	60922	57053
13	Black African	141480	5656	8658	32056	58110	75135	48557
14	Black Caribbean	160327	45354	6430	5461	14511	12926	7988
15	Black other	72056	4074	1173	2152	3436	3663	2665
16	White and Black African	33860	419	575	1255	3129	4468	6876
17	White and Black Caribbean	81191	1445	339	372	924	1386	5101
18	White and Black other	81747	2167	1097	1236	2171	3945	9863
19	Mixed other	84316	2792	1727	2266	5228	8955	18272
20	Arab	24788	856	3151	5883	10937	13246	25709
21	Other ethnic group	93995	12583	14877	17348	35717	44018	61173

The final stage is to calculate the percentage experience housing deprivation for each cell.

This is the number deprived / (number not deprived and number deprived).

23	Percentage deprived							
24	Ethnic group	born in UK	before 197	1971-1980	1981-1990	1991-2000	2001-2010	2011-2021
25	White British	12%	5%	7%	10%	19%	23%	30%
26	Irish	12%	6%	8%	10%	11%	9%	14%
27	Gypsy or Irish Traveller	37%	23%	21%	31%	32%	43%	45%
28	Roma	43%	8%	18%	17%	28%	31%	40%
29	White other	26%	6%	8%	14%	21%	26%	33%
30	Indian	18%	6%	7%	12%	20%	24%	42%
31	Pakistani	38%	17%	23%	29%	38%	39%	44%
32	Bangladeshi	54%	25%	37%	46%	57%	49%	56%
33	Chinese	18%	7%	9%	12%	15%	16%	20%
34	Asian other	35%	7%	11%	16%	29%	35%	37%
35	Black African	51%	13%	18%	30%	41%	47%	49%
36	Black Caribbean	28%	9%	13%	21%	30%	33%	43%
37	Black other	39%	9%	13%	27%	39%	46%	42%
38	White and Black African	32%	12%	15%	23%	34%	39%	40%
39	White and Black Caribbean	30%	8%	17%	18%	27%	35%	48%
40	White and Black other	17%	6%	12%	15%	17%	21%	24%
41	Mixed other	27%	9%	11%	17%	26%	27%	33%
42	Arab	45%	15%	15%	26%	36%	39%	39%
43	Other ethnic group	31%	7%	11%	19%	31%	35%	42%