

Article

Occupied address-level ethnicity measures for multivariate statistics, England and Wales: 2020

Approaches used to identify occupied address-level ethnicity for feasibility research on producing multivariate statistics for England and Wales from administrative data.

Contact:
Joanna Harkrader and Teresa
Tinklin
admin.based.
characteristics@ons.gov.uk
+44 1329 444974

Release date:
16 February 2023

Next release:
To be announced

Table of contents

1. [Main Points](#)
2. [About our transformation research](#)
3. [Data used to produce the admin-based housing by ethnicity dataset for 2020](#)
4. [Ethnic group of address reference person](#)
5. [Ethnic group summary variable for all individuals at an address](#)
6. [Combining the ethnic group of ARP2 and the ethnic group summary variable](#)
7. [Comparing the three approaches: accommodation type by address-level ethnicity](#)
8. [Occupied address-level ethnicity measures for multivariate statistics, England and Wales: 2020 data](#)
9. [Glossary](#)
10. [Data sources and quality](#)
11. [Future developments](#)
12. [Related links](#)
13. [Cite this article](#)

1 . Main Points

- This article summarises research undertaken to derive an occupied address-level ethnicity measure from administrative data, as part of feasibility research on producing housing by ethnicity statistics from administrative data.
- We explored three potential approaches to deriving address-level ethnicity: selecting an address reference person and using their ethnicity, deriving an ethnic group summary variable for all individuals at an address, and creating a variable that combines these two approaches.
- Using the five-category ethnic group of the address reference person showed more potential to publish statistics at lower levels of geography than the other two approaches because it has fewer categories, which reduces the need for suppression of small numbers at lower geographical levels.
- Using the ethnic group of the address reference person provides an incomplete summary of the ethnicity of an occupied address, particularly where there are two or more ethnic groups living at the address.
- We seek feedback on the potential usefulness of these statistics and to further develop these approaches in accordance with user needs.

These are not official statistics and should not be used for policy or decision-making. They are published as feasibility research into a new method for producing subnational multivariate statistics on housing by ethnic group using administrative data.

2 . About our transformation research

At the Office for National Statistics (ONS) we are exploring the feasibility of producing statistics on a range of topics using administrative data sources. This might remove the need for us to collect data through a census or surveys.

This research forms part of [our population and social statistics transformation programme](#), which aims to provide the best insights on population, migration and society using a range of data sources. The findings will form part of the evidence base for the National Statistician's Recommendation in 2023 on the future of population, migration and social statistics in England and Wales.

This article presents three approaches to measuring occupied address-level ethnicity as part of our feasibility research on the potential to produce [subnational multivariate housing by ethnicity statistics using administrative data](#).

The census identifies a household reference person (HRP) to represent the characteristics of a household using economic activity, age, and order on the census form.

Our first approach emulates this concept and creates an admin-based address reference person (ARP). Our second approach summarises the ethnic groups of all persons at each occupied address. Our third approach combines these two approaches to capture the maximum information on the ethnic group makeup of an occupied address. These approaches are described in detail in this article, and an initial assessment is made of the strengths and weaknesses of using each to explore accommodation type by ethnic group. Further statistics on housing by ethnic group at regional and local authority levels for England and Wales, are published in our [Developing subnational multivariate housing by ethnicity statistics from administrative data article](#).

3 . Data used to produce the admin-based housing by ethnicity dataset for 2020

To create the admin-based housing by ethnicity dataset (ABHED), we linked the [admin-based ethnicity dataset version 3.0](#) (ABED V3.0) for 2020, the admin-based household estimates version 3.0 (ABHE V3.0) for 2020, and the [admin-based housing stock version 1.0](#) (ABHS V1.0) dataset for 2020. The data sources are linked together using pseudonymised identifiers. In producing statistics using linked administrative data, particularly for small populations, we apply the same rigour in data security and privacy as with all official statistics. For further information about the security of these linked data, see our [Population and social statistics transformation: 2019 progress update](#).

The ABHE V3.0 and ABED V3.0 are derived from multiple administrative data sources, and both use the [Statistical Population Dataset version 3.0](#) (SPD V3.0) as the population base. The SPD V3.0 is a record-level dataset, which includes individuals that meet one, or more, "activity-based" rules, meaning they were considered part of the usually resident population in England and Wales. Because they are both derived from the SPD V3.0, the ABHE V3.0 was linked to the ABED V3.0 using a unique identifier. This assigned a unique property reference number (UPRN) to as many individuals as possible in ABED V3.0. This then allowed individuals in the ABED V3.0 to be grouped into addresses using UPRN. To obtain property characteristics for each UPRN, the ABHS V1.0 was joined to the linked datasets using UPRN. The ABHS V1.0 provides information about residential addresses, so communal establishments and special population groups are removed from our dataset. This step removed 1.0% of records from the ABHED.

In this article, all the individuals at an address are considered to form one household; however, it should be noted that there may be more than one household at each UPRN. Because this is a different definition of household compared with the one used by the [Government Statistical Service \(GSS\) or 2011 Census and social surveys](#), we refer to "occupied addresses", rather than households, in this report. We have used the ethnicity variable from ABED V3.0, which makes use of 2011 Census data, in addition to multiple administrative data sources, to derive an individual's ethnic group. For more information on this method, see our [Changes to data and methods article](#) on this topic.

The ABHED for 2020 provides a dataset with multiple people per address, some with a stated ethnicity and some with no stated ethnicity. Figures quoted in this article refer only to occupied addresses and to individuals and properties that are linked across all three datasets.

4 . Ethnic group of address reference person

The 2011 Census and Census 2021 define a household as one person living alone or a group of people (not necessarily related) living at the same address who share cooking facilities, and share a living room, sitting room or dining area. It is currently not possible to clearly identify multiple households at an address from administrative data alone. In our [admin-based housing by ethnicity dataset](#) (ABHED), a single unique property reference number (UPRN) may contain more than one household. For this reason, we call our admin-based version of the census household reference person (HRP) concept, the address reference person (ARP).

Selecting an address reference person with a stated ethnicity

We produced two versions of the admin-based ARP variable basing it on age because our linked dataset did not contain information on economic activity, which is part of the approach used to define an HRP in the census. The first, ARP1, takes the oldest person of working age (aged 16 to 67 years) as the ARP, regardless of whether they have a stated ethnicity or not. If there is no one of working age, it takes the oldest person of retirement age (aged 68 years and over), and in a small number of cases, the oldest child (aged under 16 years old). Where there were two people of the same age eligible to be ARP1, the first record appearing in our dataset was selected. Using this approach, 83.7% of occupied addresses in England and 87.2% in Wales had an ARP with a stated ethnicity (Table 1).

Our second approach, ARP2, sought to maximise the number of occupied addresses with an ARP with a stated ethnicity. It did this by sorting the individuals at each property into working age first, then retirement age, then children, further sorting each group in descending order of age. We then took the first person in this ordered group that had a stated ethnicity as the ARP. If there were two people of the same age eligible to be ARP2 and only one had a stated ethnicity, that person was selected as ARP2. If there were two people of the same age and both had a stated ethnicity the first person appearing in our dataset was selected. This increased the percentage of occupied addresses having an ARP with a stated ethnicity by 10.4 percentage points in England and 7.4 percentage points in Wales (Table 1).

Table 1: ARP2 approach increases the proportion of addresses with an address reference person (ARP) with a stated ethnicity

Proportion of addresses with an ARP with a stated ethnicity, England and Wales, 2020

	England	Wales
ARP1	83.7	87.2
ARP2	94.1	94.6

Source: Office for National Statistics

Notes

1. Proportions may not sum to 100.0% because of rounding.

Address reference person by age and sex

A comparison of the age and sex distributions of the admin-based ARP1 and ARP2 with HRP estimates from the 2011 Census (Figure 1) was done. It showed that both ARP approaches produced a higher proportion of ARPs aged 50 to 64 years and a lower proportion of ARPs aged 35 to 49 years when compared with HRPs. This shows the potential impact of not having information in the ABHED to identify economically active individuals, before ordering on age. The ARP2 approach tended to match the 2011 Census HRP distribution more closely for those aged 16 to 34 years than for the older age groups, except for males in Wales. Both ARP approaches included a small percentage of children, which HRP did not.

Figure 1: The ARP approaches produce a higher proportion of ARPs aged 50 to 64 years compared with the 2011 Census HRP estimates

Proportion of admin-based 2020 ARP1, admin-based 2020 ARP2 and 2011 Census HRP by age and sex, England and Wales

Notes:

1. Proportions may not sum to 100.0% because of rounding.

Download the data

[.xlsx](#)

Address reference person by ethnic group

Neither admin-based ARP approach in the ABHED 2020 provided an exact agreement with the 2011 Census HRP in terms of ethnic group. Table 2 shows that in England and Wales, both ARP approaches produced smaller proportions of ARPs in the White British ethnic group compared with the 2011 Census HRP proportions. Census 2021 showed [a decline in the proportion of the overall population in the White British ethnic group in England and Wales between 2011 and 2021](#). The differences between the admin-based ARP and the 2011 Census HRP distributions were generally smaller in Wales than in England. The different proportions of ethnic groups for ARPs in ABHED and HRPs in the 2011 Census are because of definitional differences as well as the different reference period of the data. We plan to produce the ABHED for 2021 and compare it with Census 2021 data, when available, to better understand these differences. For the purpose of this initial feasibility research, we use ARP2 in the rest of this article because it provides a stated ethnicity for a higher proportion of occupied addresses than ARP1, which helps us produce more data at smaller geographical areas.

Table 2: Proportion of address reference persons (ARPs) in the admin-based housing by ethnicity dataset (ABHED) 2020 and household reference persons (HRPs) in the 2011 Census by ethnic group, England and Wales

Ethnic group	England			Wales		
	ARP1	ARP2	HRP	ARP1	ARP2	HRP
Asian Bangladeshi	0.5	0.5	0.5	0.2	0.3	0.2
Asian Chinese	0.6	0.6	0.6	0.3	0.3	0.3
Asian Indian	2.0	2.2	2.0	0.4	0.4	0.5
Asian Pakistani	1.4	1.5	1.2	0.3	0.3	0.2
Asian Other	1.5	1.6	1.2	0.4	0.5	0.4
Black African	1.6	1.8	1.6	0.4	0.4	0.3
Black Caribbean	1.1	1.1	1.3	0.1	0.1	0.2
Black Other	0.7	0.7	0.4	0.1	0.1	0.1
Mixed White and Asian	0.3	0.3	0.3	0.2	0.2	0.2
Mixed White and Black African	0.2	0.2	0.2	0.1	0.1	0.1
Mixed White and Black Caribbean	0.3	0.4	0.5	0.3	0.3	0.3
Mixed Other	0.6	0.7	0.3	0.2	0.3	0.2
White British	81.6	79.8	83.3	93.3	92.3	94.3
White Gypsy, Roma or Irish Traveller	[low]	[low]	0.1	0.1	0.1	0.1
White Not specified	0.5	0.5	[z]	1.2	1.3	[z]
White Irish	0.6	0.6	1.3	0.4	0.4	0.6
White Other	5.7	6.4	4.3	1.8	2.2	1.7
Other Arab	0.1	0.1	0.3	0.1	0.1	0.3
Other Any other ethnic group	0.9	1.0	0.5	0.2	0.3	0.2

Notes

1. Proportions may not sum to 100.0% because of rounding.
2. Proportions are based on those with a stated ethnicity.
3. [low] denotes that the proportion is less than 0.05.
4. [z] denotes not applicable, as the category was not used in the data collection. The Higher Education Statistics Agency (HESA) data for England and Wales only have categories for White and Gypsy or Traveller within the higher-level White ethnic group. Those with a sub-category ethnicity of White in HESA were recoded as White not specified.

5 . Ethnic group summary variable for all individuals at an address

Our second approach, the ethnic group summary variable, summarised the ethnic groups of all persons present within an occupied address. This initial approach is based on five-category ethnic group. This approach is similar to [the one used for Census 2021](#), with the addition of categories to describe occupied addresses that have both a single ethnic group and individuals with no stated ethnicity. Our admin-based housing by ethnicity dataset (ABHED) classified 6.6% of occupied addresses in England and 2.8% in Wales as having two or more ethnic groups (Table 3). However, Census 2021 identified 10.4% of households in England and 5.3% of households in Wales as multiple-ethnic group households. This difference is accounted for by the missing ethnicities in the ABHED.

Table 3: Proportion of occupied addresses in the admin-based housing by ethnicity dataset (ABHED) by ethnic group summary variable, England and Wales, 2020

Ethnic group summary variable	England Wales	
	England	Wales
Asian only	3.0	0.8
Asian and no stated ethnicity	1.8	0.4
Black only	1.7	0.2
Black and no stated ethnicity	0.8	0.1
Mixed only	0.4	0.3
Mixed and no stated ethnicity	0.3	0.2
White only	63.2	76.2
White and no stated ethnicity	15.9	13.3
Other only	0.2	0.1
Other and no stated ethnicity	0.2	0.1
Two or more ethnic groups	6.6	2.8
No stated ethnicities	5.9	5.4

Source: Office for National Statistics

Notes

1. Proportions may not sum to 100.0% because of rounding.

Comparison between ethnic group of ARP2 and ethnic group summary variable

Our second admin-based address reference person (ARP2) approach is more comparable with the approach taken by the census and surveys. It enables comparisons with figures on housing by ethnicity from the 2011 Census. Publishing statistics on housing by ethnicity using administrative data involves applying disclosure rules aligned with the requirements of administrative data suppliers, which suppress both small counts and percentages. The ARP approach provides greater potential to publish statistics at lower levels of geography than the other two approaches because of its fewer categories. However, for addresses with two or more ethnic groups, information is lost about the other ethnic groups at the address when using this approach (Table 4). In both England and Wales, occupied addresses where the ARP is of Mixed or Other ethnic group are more likely to be multiple-ethnic group addresses.

Table 4: Proportion of ARP2s by five-category ethnic group and number of ethnic groups in the occupied address, England and Wales, 2020

Ethnic group	Living in an occupied address with a single ethnic group		Living in an occupied address with a single ethnic group and individuals with no stated ethnicity		Living in an occupied address with two or more ethnic groups	
	England	Wales	England	Wales	England	Wales
Asian	49.1	48.4	29.1	24.6	21.7	27.0
Black	49.7	39.4	23.1	20.9	27.1	39.7
Mixed	30.7	32.2	18.1	18.4	51.2	49.4
White	76.9	83.7	19.4	14.6	3.7	1.7
Other	22.1	26.3	15.6	31.6	62.3	42.0

Source: Office for National Statistics

Notes

1. Proportions may not sum to 100.0% because of rounding.

6 . Combining the ethnic group of ARP2 and the ethnic group summary variable

Our third method combined the ethnic group of the second admin-based address reference person (ARP2) approach with the ethnic group summary variable. This maximised the granularity of information (the level of detail) about the ethnic group composition of an occupied address. Table 5 shows that this third approach provides a better summary of address-level ethnicity but may reduce our ability to publish statistics at lower levels of geography. This is because the higher number of categories increases the need for suppression of small numbers at lower geographical levels following the application of statistical disclosure control measures.

Table 5: Proportion of occupied addresses in the admin-based housing by ethnicity dataset (ABHED) by combined ethnic group variable, England and Wales, 2020

Combined ethnic group variable	England	Wales
Asian ARP	4.8	1.2
Asian ARP and one or more other ethnicities	1.3	0.4
Black ARP	2.4	0.4
Black ARP and one or more other ethnicities	0.9	0.2
Mixed ARP	0.7	0.5
Mixed ARP and one or more other ethnicities	0.7	0.4
White ARP	79.1	89.6
White ARP and one or more other ethnicities	3.0	1.5
Other ARP	0.4	0.2
Other ARP and one or more other ethnicities	0.6	0.2
No stated ethnicities	5.9	5.4

Source: Office for National Statistics

Notes

1. Proportions may not sum to 100.0% because of rounding.

7 . Comparing the three approaches: accommodation type by address-level ethnicity

Analysis of accommodation type by address-level ethnicity was carried out to explore how accommodation type varied for each ethnic group category when using the different approaches. Because of the differences in categories, comparisons of the different approaches with 2011 Census data are only possible for the second admin-based address reference person (ARP2) approach. These results and comparisons can be found in our [Developing subnational multivariate housing by ethnicity statistics from administrative data article](#). This section presents the findings for the other two approaches.

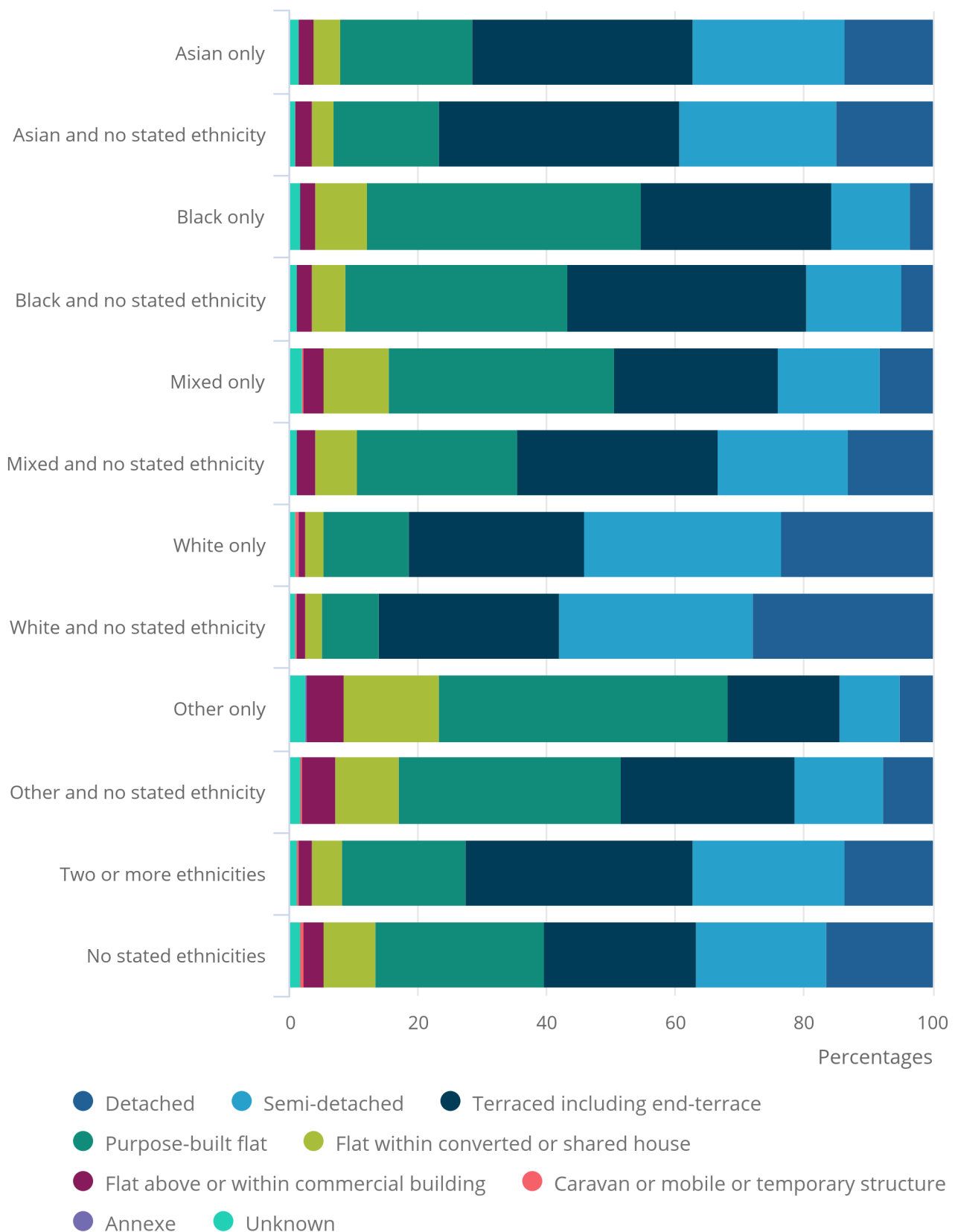
Figures 2 to 5 present initial analysis of accommodation type by ethnic group using the ethnic group summary and combined variables. Figures 2 and 3 show that Mixed and Other ethnic groups have more variation in accommodation type between occupied addresses with and without missing ethnicity data. Figures 4 and 5 suggest that there is variation in accommodation type by whether an address reference person (ARP) lives in a multiple-ethnic group occupied address.

Figure 2: The Mixed and Other ethnic groups show more variation in accommodation type between occupied addresses with and without missing ethnicity data

Proportion of occupied addresses by accommodation type and ethnic group summary variable, England, 2020

Figure 2: The Mixed and Other ethnic groups show more variation in accommodation type between occupied addresses with and without missing ethnicity data

Proportion of occupied addresses by accommodation type and ethnic group summary variable, England, 2020



Source: Office for National Statistics

Notes:

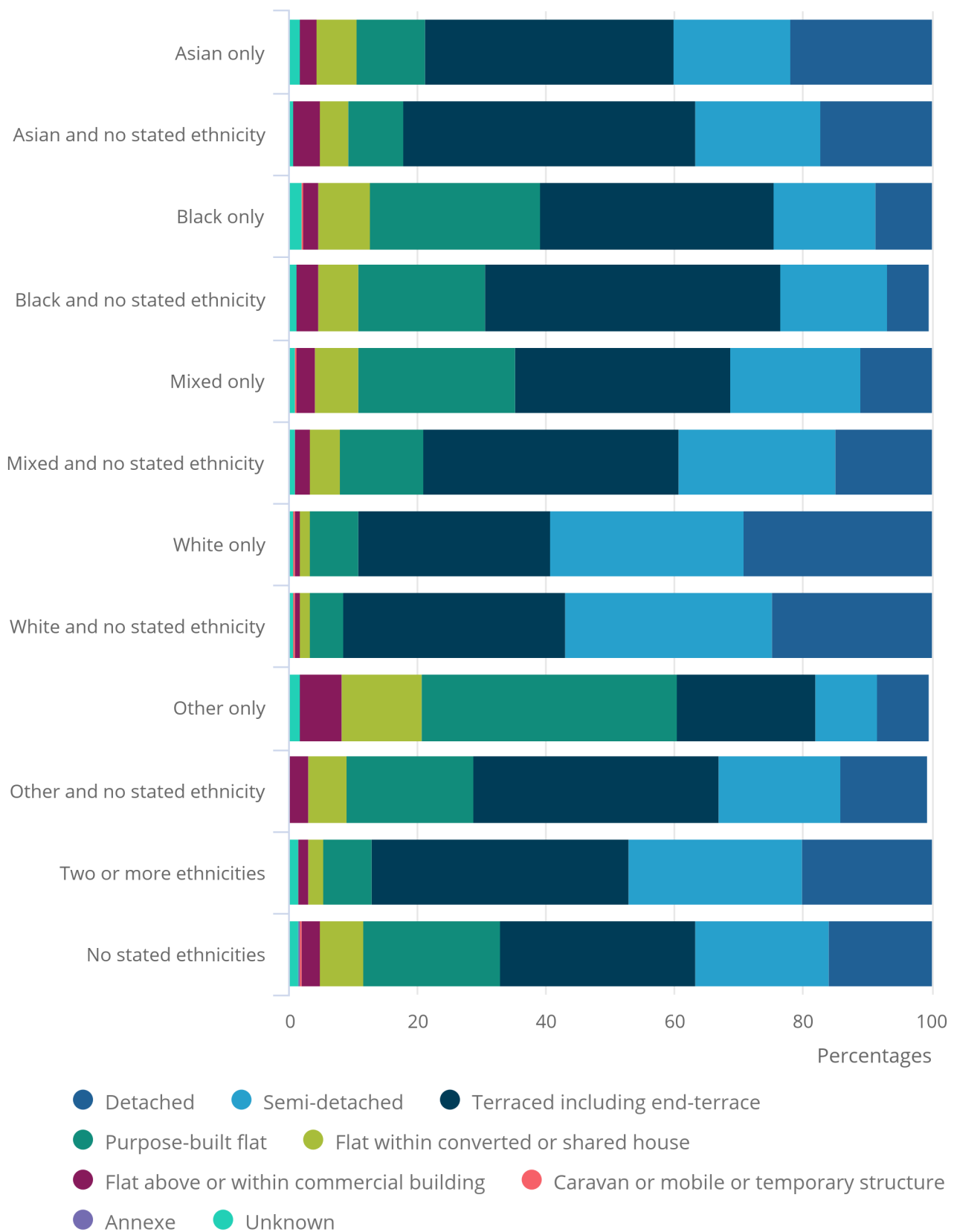
Proportions may not sum to 100.0% because of rounding.

Figure 3: The Mixed and Other ethnic groups show more variation in accommodation type between occupied addresses with and without missing ethnicity data

Proportion of occupied addresses by accommodation type and ethnic group summary variable, Wales, 2020

Figure 3: The Mixed and Other ethnic groups show more variation in accommodation type between occupied addresses with and without missing ethnicity data

Proportion of occupied addresses by accommodation type and ethnic group summary variable, Wales, 2020



Source: Office for National Statistics

Notes:

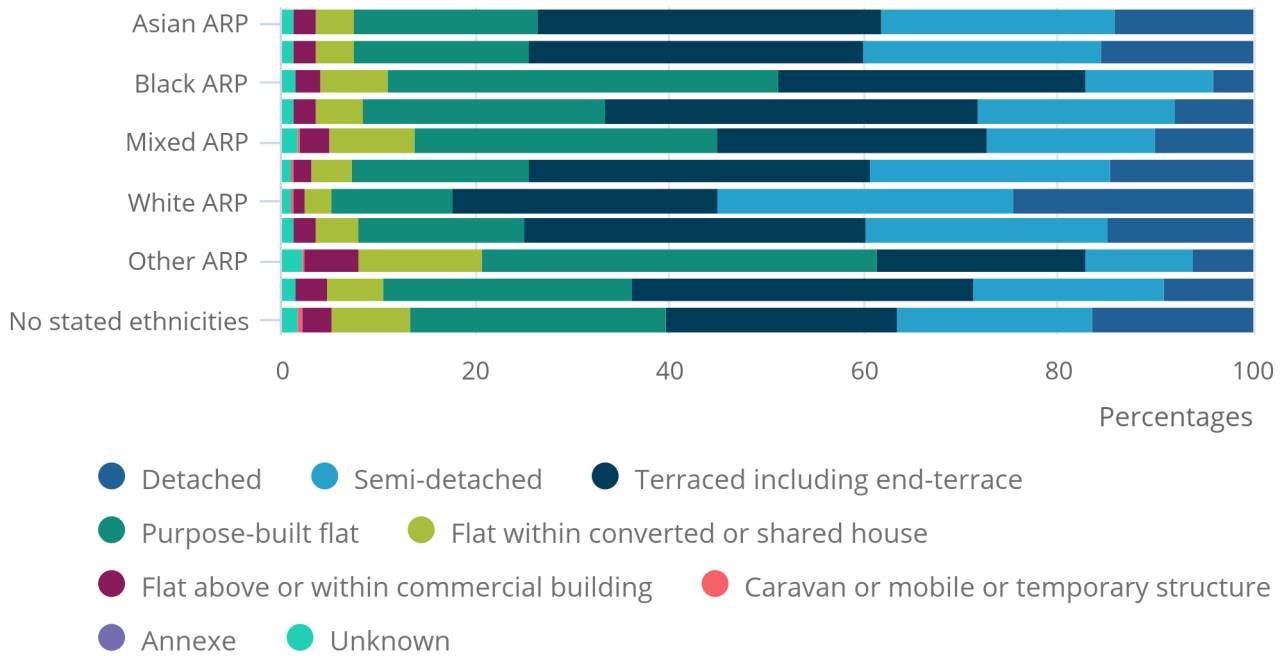
Proportions may not sum to 100.0% because of rounding.

Figure 4: Accommodation type varies by whether an address reference person (ARP) lives at an address with at least one other ethnic group or not

Proportion of occupied addresses by accommodation type and combined ethnic group variable, England, 2020

Figure 4: Accommodation type varies by whether an address reference person (ARP) lives at an address with at least one other ethnic group or not

Proportion of occupied addresses by accommodation type and combined ethnic group variable, England, 2020



Source: Office for National Statistics

Notes:

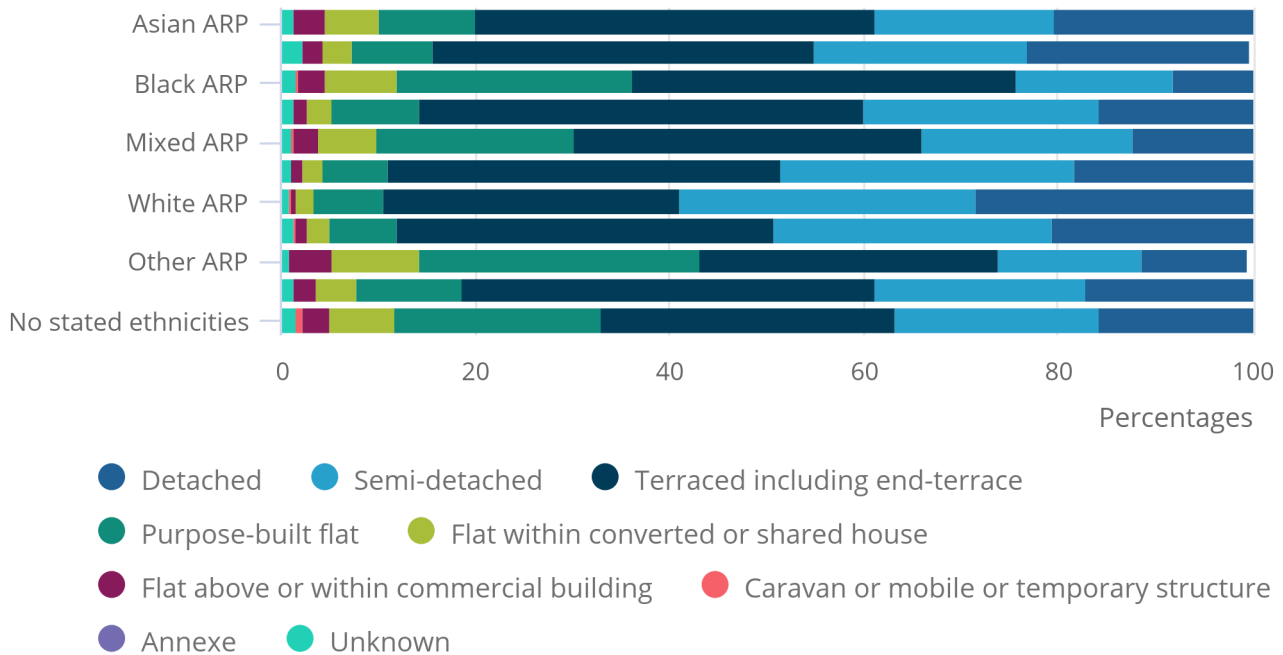
Proportions may not sum to 100.0% because of rounding.

Figure 5: Accommodation type varies by whether an address reference person (ARP) lives at an address with at least one other ethnic group or not

Proportion of occupied addresses by accommodation type and combined ethnic group variable, Wales, 2020

Figure 5: Accommodation type varies by whether an address reference person (ARP) lives at an address with at least one other ethnic group or not

Proportion of occupied addresses by accommodation type and combined ethnic group variable, Wales, 2020



Source: Office for National Statistics

Notes:

Proportions may not sum to 100.0% because of rounding.

Applying statistical disclosure control

Publishing statistics on housing by ethnicity using administrative data involves applying disclosure rules aligned with the requirements of administrative data suppliers, which suppress both small counts and percentages. An assessment of the percentage of cells suppressed in tables when using each approach to understand accommodation type by ethnicity at local authority (LA) level (Table 6) was undertaken. It showed that the ethnic group of ARP2 offers more potential to publish statistics at lower levels of geography. This is based on the current approach to statistical disclosure control set out in agreements with data suppliers.

Table 6: Percentage of suppressed cells when using each approach to analyse accommodation type by address-level ethnicity at local authority level, England and Wales, 2020

Address-level ethnicity variable	England	Wales
ARP2	29.0	34.8
Ethnic group summary variable	42.7	52.9
Combined ethnic group variable	39.5	49.9

Source: Office for National Statistics

Notes

1. Annexes and properties with unknown property characteristics were included.
2. Properties with no stated ethnicities were included.
3. The percentage of suppressed cells was based on the number of cells suppressed when calculating the percentage of each level of an occupied address-level ethnicity measure living in each property type.

In summary, ARP2 was more comparable with the current household reference person (HRP) approach used in surveys and the census than the other two approaches. It also offered more potential to publish statistics at lower levels of geography. It was less effective, however, in capturing the ethnic group composition of occupied addresses where the ARP was of Other or Mixed ethnic group, as these ARPs were more likely to live in properties with two or more ethnic groups. This is an area for further exploration in future research. Further statistics on housing by ethnicity, using the five-category ethnic group of ARP2, at regional and LA levels for England and Wales, are published in our [Developing subnational multivariate housing by ethnicity statistics from administrative data article](#).

8 . Occupied address-level ethnicity measures for multivariate statistics, England and Wales: 2020 data

[Occupied address-level ethnicity measures for multivariate statistics, England and Wales: 2020](#)

Dataset | Released 16 February 2023

Data for feasibility research on producing occupied address-level ethnicity measures for multivariate statistics for England and Wales from administrative data.

9 . Glossary

Accommodation type

The accommodation type variable is derived from the Valuation Office Agency (VOA) property type and VOA dwelling code variables to resemble the census accommodation type as closely as possible. It also adds an additional eighth category for annexes. "Annexe" is not a category in the 2011 Census accommodation type variable, but it is a new category we propose for the VOA property type of "annexe." The VOA describe an annexe as a building, or part of a building, which has been constructed or adapted for use as separate living accommodation.

Full information on the category names and mapping method can be found in our [Admin-based accommodation type statistics for England and Wales, feasibility research: 2011 methodology](#).

Address reference person (ARP)

An address reference person (ARP) is a variable derived from administrative data designed to serve the same purpose as the household reference person (HRP) variable from the census. An individual within an occupied address acts as a reference point for producing further derived statistics. They also characterise a whole occupied address according to the characteristics of the chosen reference person. We have created this variable as an admin data-based equivalent of household reference person to identify an individual whose stated ethnicity can represent each occupied address. This was used for the purpose of this feasibility research.

Communal establishments

A communal establishment is an establishment providing managed residential accommodation. "Managed" in this context means full-time or part-time supervision of the accommodation. Communal establishments include sheltered accommodation units (including homeless temporary shelter), hotels, guest houses, bed and breakfasts (B&Bs) and inns and pubs, and all accommodation provided solely for students (during term-time). More information is available in the [2011 Census glossary](#).

Ethnic group

The self-reported ethnic group of the individual, according to their own perceived ethnic group and cultural background. Five categories are presented in this article. The ethnic groups included in each category are:

- Asian ethnic group: Bangladeshi, Chinese, Indian, Pakistani, Asian Other
- Black ethnic group: African, Caribbean, Black Other
- Mixed ethnic group: White and Asian, White and Black African, White and Black Caribbean, Mixed Other
- White ethnic group: British, Gypsy, Roma or Irish Traveller [note 1], Irish, White Other, White not specified [note 2]
- Other ethnic group: Arab, Any other ethnic group

Household reference person (HRP)

In the census, household reference persons provide an individual person within a household to act as a reference point for producing further derived statistics. It also characterises a whole household according to characteristics of the chosen reference person. The full definition used can be found in the [2011 Census glossary](#). In this analysis, we have derived an address reference person (ARP) variable from administrative data designed to serve the same purpose as the HRP variable from the census, as described in the ARP glossary entry.

No stated ethnicity

No stated ethnicity refers to the ethnicity being recorded as refused or unknown, in line with the [methods used to derive an individual's ethnic group in the admin-based ethnicity dataset version 3.0](#) (ABED V3.0). No stated ethnicity also includes individuals who are in the Statistical Population Dataset version 3.0 (SPD V3.0) but have not been linked to any sources of ethnicity data.

Occupied address

For this research, an occupied address is a unique property reference number (UPRN) on the Address Frame which has been successfully linked to at least one individual in the [Statistical Population Dataset version 3.0](#) (SPD V3.0). It is different to the concept of a household, which uses a definition based on shared facilities. More information on the differences between a traditional "household" and an "occupied address" is available in our [Occupied address \(household\) estimates from Administrative Data: 2011 and 2015 release](#).

Special population groups

Special population groups include armed forces personnel and dependants stationed in the UK, foreign armed forces based in the UK (mainly US Air Force personnel and dependants) and the prison population.

Stated ethnicity

Stated ethnicity refers to the ethnicity being recorded as a specific ethnic group and not refused or unknown on their most recent administrative data record in 2020. This is in line with the [methods used to derive an individual's ethnic group in our admin-based ethnicity dataset version 3.0 \(ABED V3.0\)](#).

Unique property reference number (UPRN)

A unique property reference number (UPRN) is a unique identifier for every address in Great Britain. It is allocated by local government and Ordnance Survey (OS).

Notes for: Glossary

1. The Gypsy, Roma and Irish Traveller ethnic groups have been aggregated because of differences in response options across data sources meaning that it is not possible to separate them. Hospital Episode Statistics (HES) and Improving Access to Psychological Therapies (IAPT) do not include any Gypsy, Roma or Irish Traveller response options.
2. The Higher Education Statistics Agency (HESA) data for England and Wales only have categories for White and Gypsy or Traveller within the higher-level White ethnic group. Those with a sub-category ethnic group of White in HESA were recoded as White not specified.

10 . Data sources and quality

Population base

The [Statistical Population Dataset version 3.0](#) (SPD V3.0) was used as the population base for the [admin-based ethnicity dataset version 3.0](#) (ABED V3.0) for 2020 and the admin-based household estimates version 3.0 (ABHE V3.0) dataset for 2020. The SPD V3.0 is a record-level dataset, which includes individuals that meet one or more "activity-based" rules, meaning they were deemed to be part of the usually resident population in England and Wales as of 30 June 2020. [The quality of the population base](#) will have an impact on the quality of the ABED V3.0 and ABHE V3.0. More information about the coverage of the population base can be found in our [Population and migration statistics system transformation – recent updates: evaluating coverage and quality in the admin-based population estimates article](#).

Admin-based household estimate (ABHE)

The ABHEs are derived from the Statistical Population Datasets (SPDs). The ABHEs are created by taking all usual residents from the SPD that can be assigned a unique property reference number (UPRN) and grouping them into addresses to estimate the size and composition of occupied addresses. To create the ABHE V3.0, the SPD V3.0 successfully assigns a UPRN to 98.3% of usual residents directly from the Personal Demographic Service (PDS) data.

Our ability to accurately identify occupied addresses depends on the quality and coverage of the SPD V3.0 as well as the quality of the UPRN assignment of our address index matching service to address strings in the PDS data.

Admin-based housing stock dataset

Our [admin-based housing stock version 1.0 \(ABHS V1.0\) dataset](#) for 2020 brings together data from several administrative sources. This is with the aim of developing a new method for producing more regular census-like statistics for occupied residential addresses (down to small geographies) across England and Wales. The ABHS V1.0 2020 was produced by linking a residential Address Frame from June 2020 to Valuation Office Agency (VOA) data from June 2020 and the ABHE versions 2.0 and 3.0 (ABHE V2.0 and V3.0). To align more closely with the 2011 Census definition of a household, communal establishments were removed from the Address Frame. A more detailed description of how we developed the ABHS V1.0 and assessed its quality can be found in our [Developing admin-based housing stock statistics for England and Wales: 2020 article](#).

Admin-based ethnicity dataset

The ABED V3.0 was produced using the following administrative data sources:

- [English School Census \(ESC\)](#), 2011 to 2020
- [Hospital Episode Statistics \(HES\)](#), 2009 to 2020
- [Emergency Care Data Set \(ECDS\)](#), 2020
- [Improving Access to Psychological Therapies \(IAPT\)](#), 2012 to 2018
- [Higher Education Statistics Agency \(HESA\)](#), 2010 to 2020
- [Birth Notifications](#), 2006 to 2020
- [Individualised Learner Record \(ILR\)](#), 2008 to 2020
- [Welsh School Census \(WSC\)](#), 2011 to 2020

For more information about these data sources, see our [Developing admin-based ethnicity statistics for England and Wales: 2020 article](#).

Ethnicity records from these data sources were linked to the 2020 SPD V3.0 using unique identifiers. A method to select a final ethnicity per person was then implemented, as described in our [Producing admin-based ethnicity statistics for England: changes to data and methods article](#).

11 . Future developments

We will continue to explore how we can develop our subnational multivariate admin-based housing and ethnicity dataset (ABHED) alongside the wider priorities of population and social statistics transformation. This will be in line with feedback received from users. We will explore the potential for including admin-based income estimates and admin-based labour market statistics in the development of admin-based address reference person methods. This will be with a view to aligning more closely to census household reference person definitions.

Feedback

We welcome feedback on this research and our planned future developments, in particular on how occupied address-level ethnicity measures are used. We are also interested in knowing what level of detail in terms of ethnic group breakdowns, and multiple-ethnic group occupied addresses, would be required to meet user needs. Please email your feedback to admin.based.characteristics@ons.gov.uk, including "Housing by ethnicity" in the subject line.

12 . Related links

[Developing admin-based housing stock statistics for England and Wales: 2020](#)

Article | Released 5 December 2022

An overview of methods, data sources and data quality for the feasibility research on producing census-like housing stock statistics using administrative data.

[Admin-based housing stock profile for England and Wales: 2020](#)

Article | Released 5 December 2022

Summary statistics demonstrating the feasibility of producing census-like housing stock statistics using the admin-based housing stock 2020 v1.0 dataset.

[Developing subnational multivariate income by ethnicity statistics from administrative data, England: tax year ending 2016](#)

Article | Released 31 August 2022

Feasibility research on an initial case study producing income by ethnic group statistics for England from administrative data.

[Developing admin-based ethnicity statistics for England and Wales: 2020](#)

Article | Released 7 February 2023

Update on feasibility research on producing statistics on the population by ethnic group for England and Wales from administrative data, including comparisons with Census 2021 estimates.

[Change over time in admin-based ethnicity statistics, England: 2016 to 2020](#)

Article | Released 23 May 2022

Feasibility research on producing statistics on the population by ethnic group for England for 2016 to 2020 from administrative data, exploring change over time. This builds on work to improve the 2016 admin-based ethnicity statistics. These research outputs are not official statistics.

[Admin-based population and migration estimates: research update](#)

Article | 26 November 2021

Provides an overview of all the latest research into the transformation of population and migration statistics to an administrative data first methodology.

13 . Cite this article

Office for National Statistics (ONS), released 16 February 2023, ONS website, article, [Occupied address-level ethnicity measures for multivariate statistics from administrative data, England and Wales: 2020](#)