

UK Census 2021 Challenges and data access

data on the LGBTQ+ population and
privacy

Oliver Duke-Williams
Vassilis Routsis

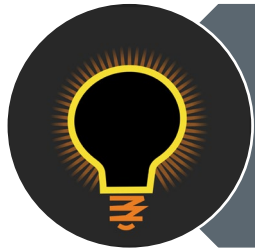
Department of Information Studies, UCL, UK



The UK Data Service and census data



UKDS provides access to the latest as well as past census data from 1971 onwards



It brings expertise, knowledge, and vast experience on handling census data. Some of our colleagues have been working on census data for decades



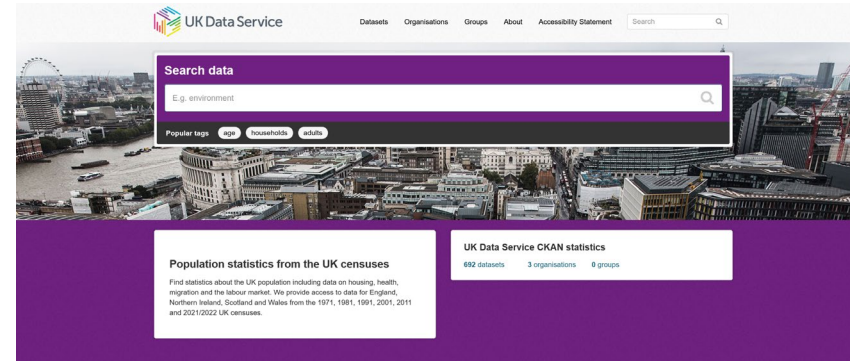
Separate groups within UK Data Service work with different types of census data



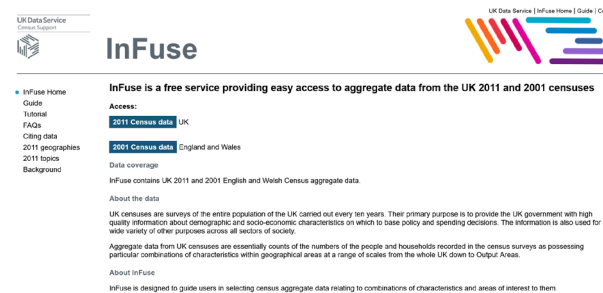
Aggregate data

Aggregate data are the most widely used of census data: they are the aggregate observations for areas running from national level to small local units, and are what would most commonly be recognised as 'the census data'.

UK Data Service provides online access to aggregate data from the 1971 census onwards.



<https://statistics.ukdataservice.ac.uk/>



<https://infuse.ukdataservice.ac.uk/>

Flow data

Flow data enumerate movements of people from one place to another. They include migration and journeys to work, and travel to second residences. These data are characterised by their complexity, volume, and sparsity.

Data are available from censuses from 1981 onwards. Data from 2011 onwards have different levels of access. 2021 census data have not been released yet.



The screenshot shows the 'Flow Data' support page on the UK Data Service website. The page is titled 'Census Support: Flow Data' and provides information about the data, how to access it, and where to find more resources. The page includes a navigation menu on the left, a main content area with sections for 'Getting the data', 'New data for un-registered users', and 'About the data', and a footer with a URL.

UK Data Service
Census Support

Flow Data

Flow data links
Flow data home
Information and Publications
Available Data
Downloads
Contact us
News
FAQs
Credits

Support
Accessibility statement
Feedback
Help

Census Support: Flow Data
Flow Data is part of UK Data Service Census Support, providing data and user support for anyone interested in using flow data from the Census.

Census flow data (also referred to as interaction data) relate to the movement of people between places. These can be regular short distance moves such as the daily commute to work, or less frequent - and sometimes longer distance - moves such as migration to a new home or an extended stay in a second address.

Flow data differ from other Census data in that they link two locations - an origin and a destination. With two geographies, flow datasets can be much larger than other Census data products.

Getting the data

There are three principal ways of getting to the data. In order to use most of the data sets, it is necessary to be registered.

- WICID - the flexible query builder
- WICID Downloads - flow data tables ready to be downloaded
- WICID Flow Summaries - a quick summary of the most significant flows associated with a chosen area

New data for un-registered users

Un-registered users and interested registered users can now access a time series of UK inter-district migration estimates, which have been produced by CIDER. These data are available directly from the Office for National Statistics - for more information about the dataset and how to gain access, click [here](#).

About the data

Currently migration and commuting data from the 1981, 1991, 2001 and 2011 Censuses are available through WICID. Additional non-census datasets are also available from the system.

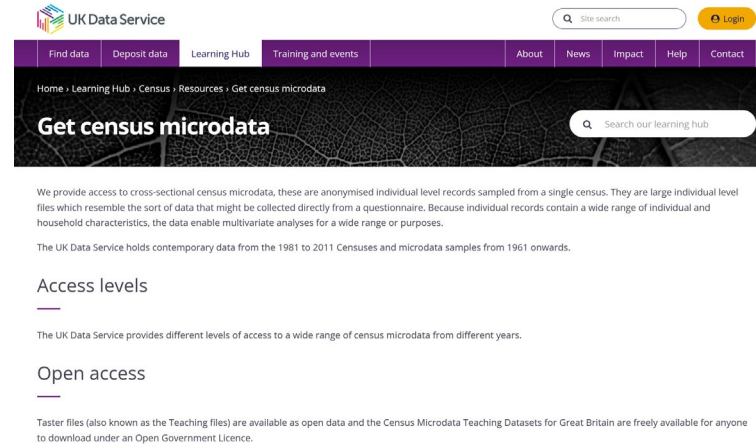
[Detailed information about our data sets](#)
[Tutorial information](#) is available describing how to use the interface

<https://wicid.ukdataservice.ac.uk/>

Microdata

Microdata are anonymised records at individual level sampled from a single census. They were introduced in 1991, but since then samples have been released from censuses from 1961 onwards.

They also have different levels of access.



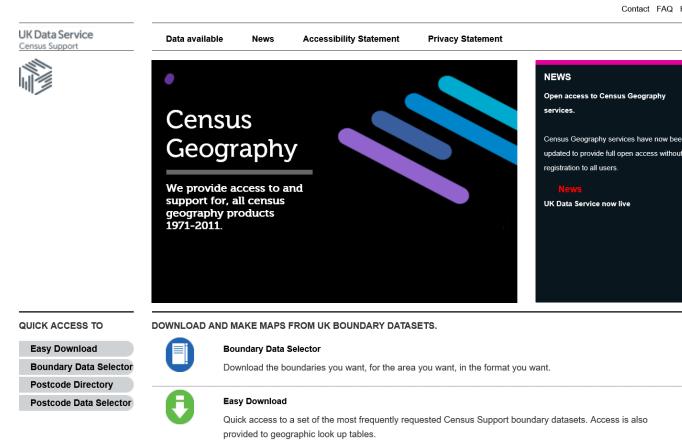
The screenshot shows the UK Data Service website. The header includes the UK Data Service logo, a site search bar, and a login button. The navigation menu contains links for Find data, Deposit data, Learning Hub, Training and events, About, News, Impact, Help, and Contact. The main content area is titled 'Get census microdata' and includes a search bar for the learning hub. The text below explains that the service provides access to cross-sectional census microdata, which are anonymised individual level records sampled from a single census. It also mentions that the UK Data Service holds contemporary data from the 1981 to 2011 Censuses and microdata samples from 1961 onwards. The page is divided into sections for 'Access levels' and 'Open access', with the latter noting that taster files are available as open data and teaching datasets are freely available under an Open Government Licence.

<https://ukdataservice.ac.uk/learning-hub/census/resources/get-census-microdata/>

Boundary data

Digital boundaries are used for mapping and spatial analysis, not just for census data but also for many other datasets. This data is provided by EDINA in Edinburgh, Scotland.

Boundary data are available from 1971 onwards in a number of different formats.



The screenshot shows the UK Data Service Census Support website. The main header includes 'UK Data Service Census Support' and navigation links for 'Data available', 'News', 'Accessibility Statement', and 'Privacy Statement'. The central content area features a large banner for 'Census Geography' with the text: 'We provide access to and support for, all census geography products 1971-2011.' To the right of the banner is a 'NEWS' section with the headline 'Open access to Census Geography services' and a sub-headline 'Census Geography services have now been updated to provide full open access without registration to all users.' Below the banner, there are two sections: 'QUICK ACCESS TO' with links for 'Easy Download', 'Boundary Data Selector', 'Postcode Directory', and 'Postcode Data Selector'; and 'DOWNLOAD AND MAKE MAPS FROM UK BOUNDARY DATASETS' with a link for 'Easy Download' and a description: 'Quick access to a set of the most frequently requested Census Support boundary datasets. Access is also provided to geographic look up tables.'

<https://borders.ukdataservice.ac.uk>

New systems for accessing UK census data



New systems for census data

CKAN for aggregate data

DCAT for improved data discovery

API-driven tools for flow data to supersede WICID



Web Interface to Census Interaction Data (WICID)

WICID was originally developed in the 1990s and went online in 2000 to provide a web-based interface to flow data. It was redesigned in early 2000s to support 2001 census outputs and was significantly upgraded in early 2010s to accommodate 2011 census data. WICID remains today the only available web tool in the UK dedicated to flow data with sub-setting capabilities.



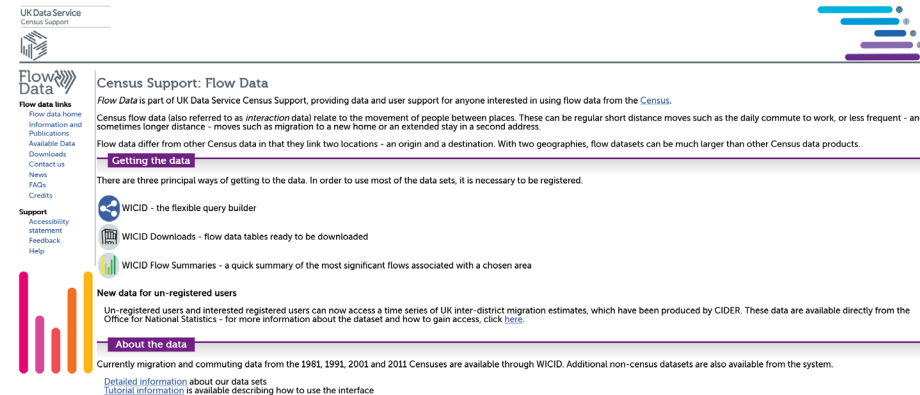
The screenshot shows the WICID web interface. At the top, there is a navigation menu with 'WICID' in a large font, followed by 'Query', 'Data', and 'Select by table (Select table)'. Below this, it says 'Logged in as: guest' and 'Customise Logout now'. There are also links for 'Help', 'Census links', 'WICID project', 'WICID query', 'Current query', and 'Save / restore queries'. A status bar indicates '0 data items currently selected (Unselect all) (Edit list)'. Below this is a list of 11 tables, each with a title and a link to the table.

Table 1	All migrants, age (5 broad age groups) by sex
Table 2	Wholly Moving Households and residents in Wholly Moving Households: counts
Table 3	All migrants, age (5 year groups) by sex
Table 4	All migrants, marital status by sex
Table 5	All migrants, ethnic group
Table 6	All migrants, whether resident in households by whether suffering from limiting long term illness
Table 7	All migrants aged 16+, economic position
Table 8	Wholly Moving Households: tenure
Table 8S	Wholly Moving Households: tenure
Table 9	Wholly Moving Households: sex and economic position of head
Table 10	Residents in Wholly Moving Households: sex and economic position of head
Table 11S	All migrants: Gaelic speakers
Table 11W	All migrants: Welsh speakers

What now?

You must select some data before you can produce any output!

- Select one of the above tables
- OR go back to the [dataset selection](#) list to choose a different data set
- OR go back to the general [data selection](#) page
- OR go back to the general [query](#) interface



The screenshot shows the UK Data Service Census Support Flow Data page. The page title is 'Census Support: Flow Data'. Below the title, there is a description of flow data and its uses. There are three main sections: 'Getting the data', 'New data for un-registered users', and 'About the data'. The 'Getting the data' section lists three ways to get the data: WICID (the flexible query builder), WICID Downloads (flow data tables ready to be downloaded), and WICID Flow Summaries (a quick summary of the most significant flows associated with a chosen area). The 'New data for un-registered users' section mentions that un-registered users can now access a time series of UK inter-district migration estimates. The 'About the data' section provides information about the data sets and how to use the interface.

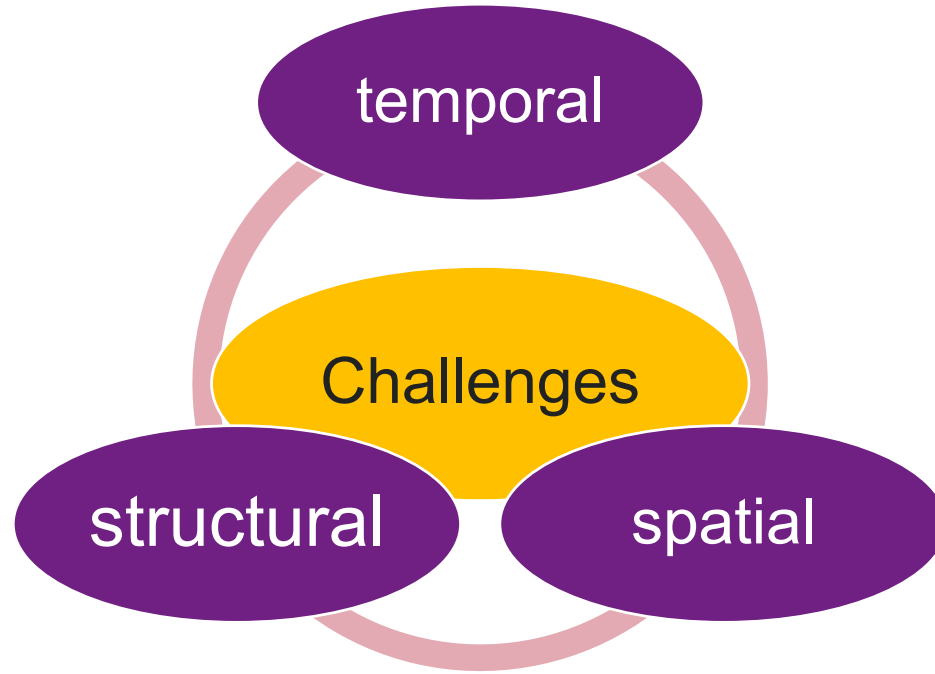
Challenges and handling of sensitive variables in UK censuses



Peculiarities of the 2021/22 census

In England, Wales, and Northern Ireland the census took place in March 2021 at the height of the Covid-19 pandemic.

Scotland decided to move its census to March 2022 instead.



Sensitive questions in UK census

Religion	Religion	Sexual orientation	Gender identity
Northern Ireland	England and Wales, Northern Ireland, Scotland	England and Wales, Northern Ireland, Scotland	England and Wales, Scotland
Pre 2001	2001 onwards	2021 onwards	2021 onwards

Sensitive questions in census 2021/22

England & Wales

26 Which of the following best describes your sexual orientation?

This question is **voluntary**

- Straight/Heterosexual
- Gay or Lesbian
- Bisexual
- Other sexual orientation, write in

27 Is the gender you identify with the same as your sex registered at birth?

This question is **voluntary**

- Yes
- No, write in gender identity

16 What is your religion?

This question is **voluntary**

- No religion
- Christian (including Church of England, Catholic, Protestant and all other Christian denominations)
- Buddhist
- Hindu
- Jewish
- Muslim
- Sikh
- Any other religion, write in

Scotland

8 Which of the following best describes your sexual orientation?

This question is **voluntary**

- Answer only if you are aged 16 or over
- Tick **one** box only

- Straight / Heterosexual
- Gay or Lesbian
- Bisexual
- Other sexual orientation, please write in:

4 Do you consider yourself to be trans, or have a trans history?

This question is **voluntary**

- Answer only if you are aged 16 or over
- Trans is a term used to describe people whose gender is not the same as the sex they were registered at birth
- Tick **one** box only

- No
- Yes, please describe your trans status (for example, non-binary, trans man, trans woman):

21 What religion, religious denomination or body do you belong to?

This question is **voluntary**

- None
- Church of Scotland
- Roman Catholic
- Other Christian, please write in below:
- Muslim, write in denomination or school below:
- Hindu
- Buddhist
- Sikh
- Jewish
- Pagan
- Another religion or body, please write in:

Northern Ireland

24 Which of the following best describes your sexual orientation?

- Straight/Heterosexual
- Gay or Lesbian
- Bisexual
- Other sexual orientation, write in

Prefer not to say

No question on gender identity

13 What religion, religious denomination or body do you belong to?

- Roman Catholic
- Presbyterian Church in Ireland
- Church of Ireland
- Methodist Church in Ireland
- Other, write in

None

14 What religion, religious denomination or body were you brought up in?

- Roman Catholic
- Presbyterian Church in Ireland
- Church of Ireland
- Methodist Church in Ireland
- Other, write in

None

Privacy and confidentiality in flow data

Dual geography identifiers

- Large matrices with low expected cell values

Flows often cluster towards short distance interactions

- The longer the distance between pairs the more likely to contain small or zero cell values



Protection of privacy

Disclosure controls

Access controls

Legal implications



Disclosure controls

Targeted record swapping

- Swap people and households that stand out in the data with a similar record in the local area

Cell key perturbation

- Make slight changes to cell counts, which adds some noise to the data

Publishing rules

- Rules on publishing data with very low counts (usually less than 3)



Access controls

Public

• Available to any user via the Open Government Licence

Safeguarded

• Available via UKDS to members of academia, local and central government, NHS, and UK parliaments and assemblies via End User Licence (EUL) so need to register and login

Secured

• Available to approved researchers via the Approved Researcher Scheme using the ONS Secure Research Service (SRS)



Flow data can be disclosive



Up to 37% of people in the GB/UK migration data, have unique combinations of origin and destination at the most detailed level of geography

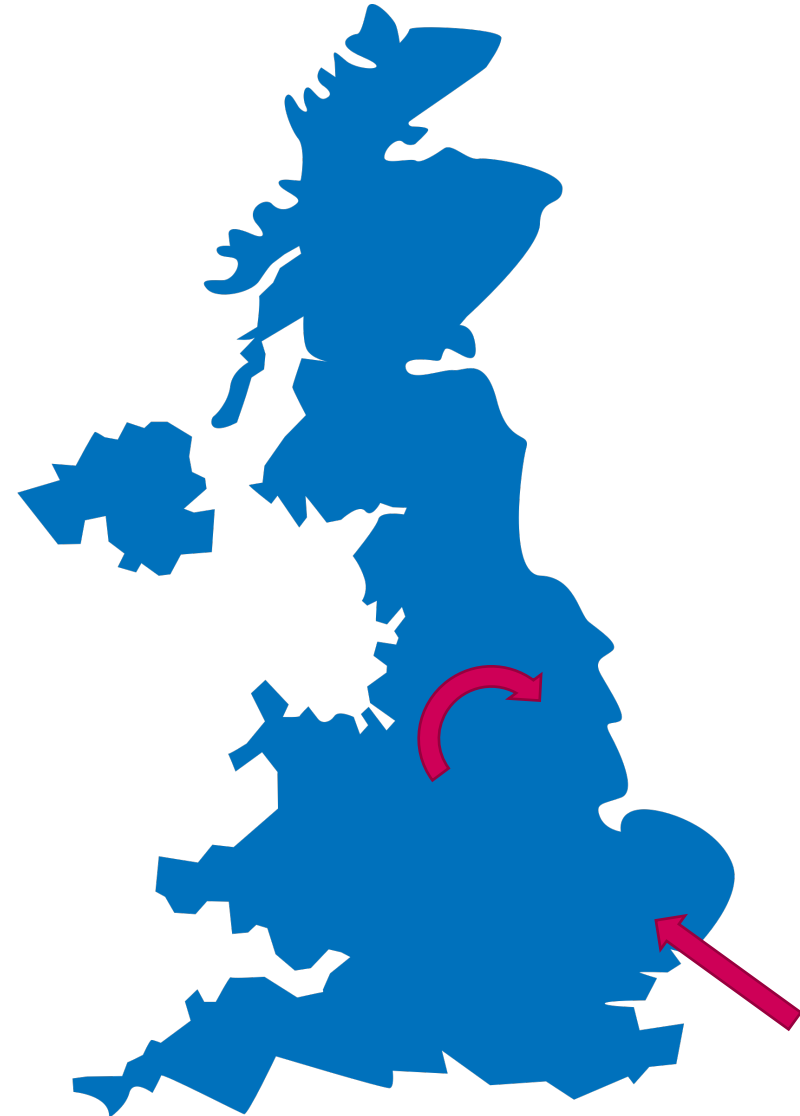
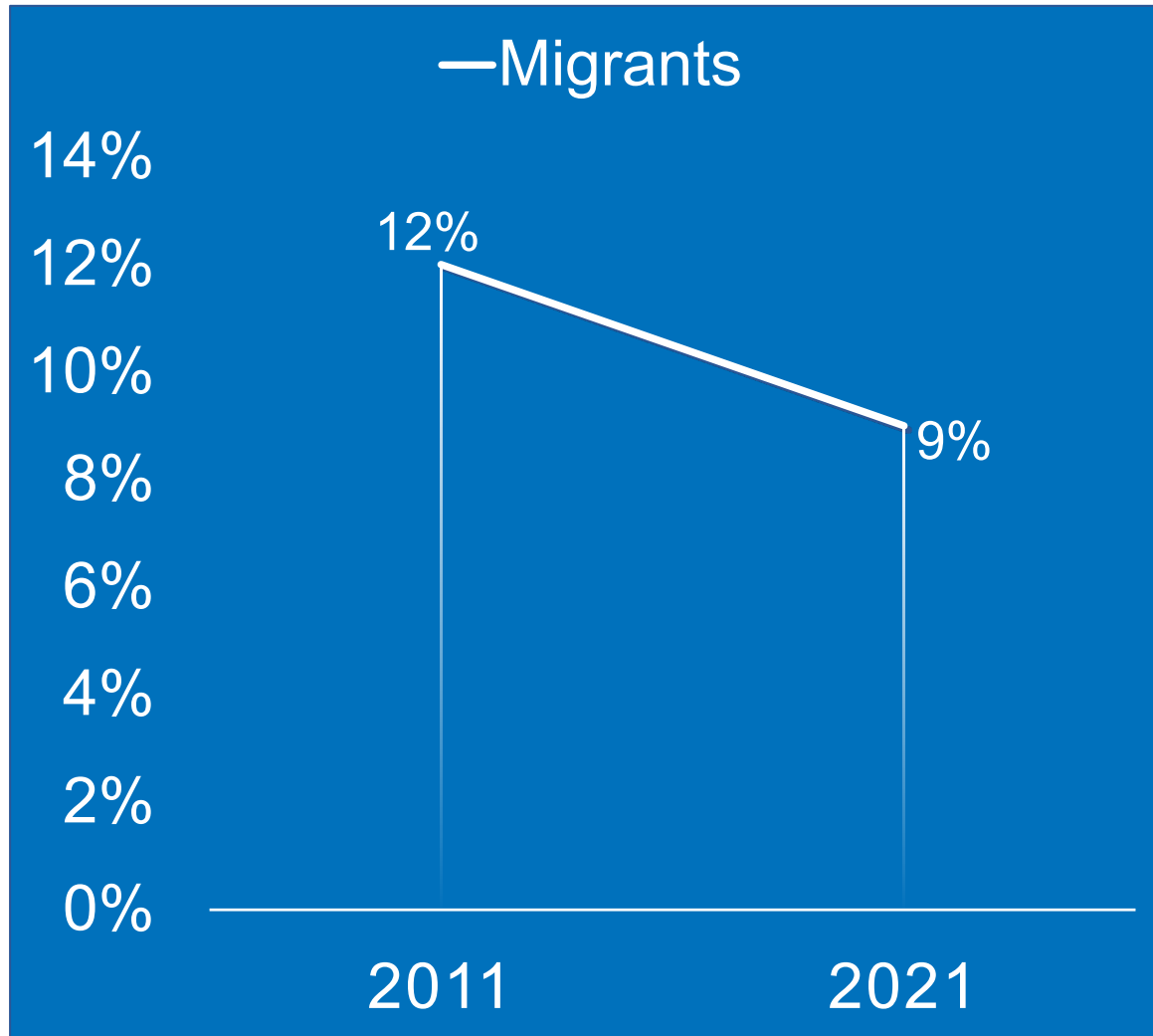


Up to 58% of people in the GB/UK journey to work data, have unique combinations of origin and destination at the most detailed level of geography

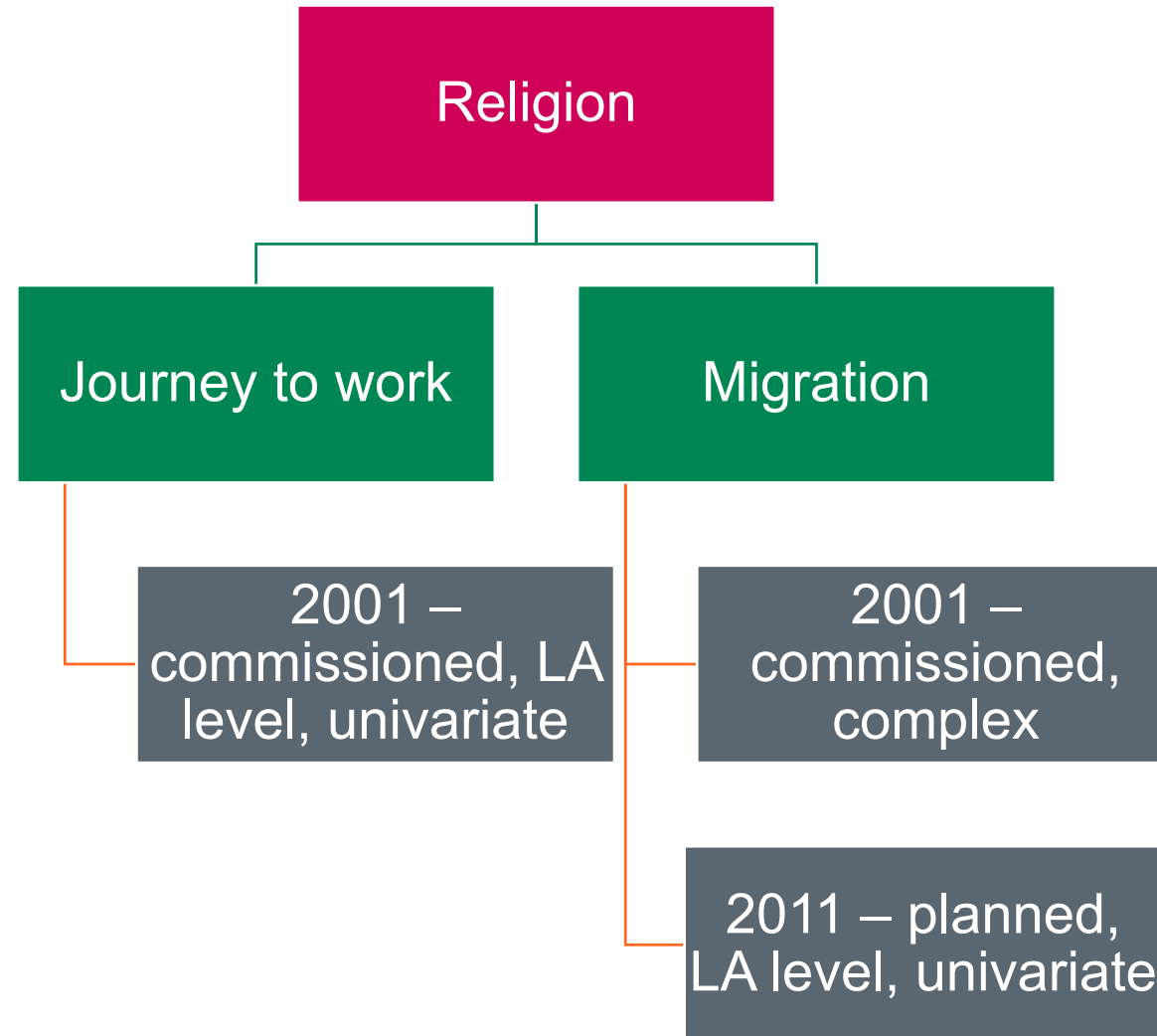
Duke-Williams, O. Location Tracing and Potential Risks in Interaction Data Sets. *Appl. Spatial Analysis* **12**, 377–394 (2019).
<https://doi.org/10.1007/s12061-017-9247-1>



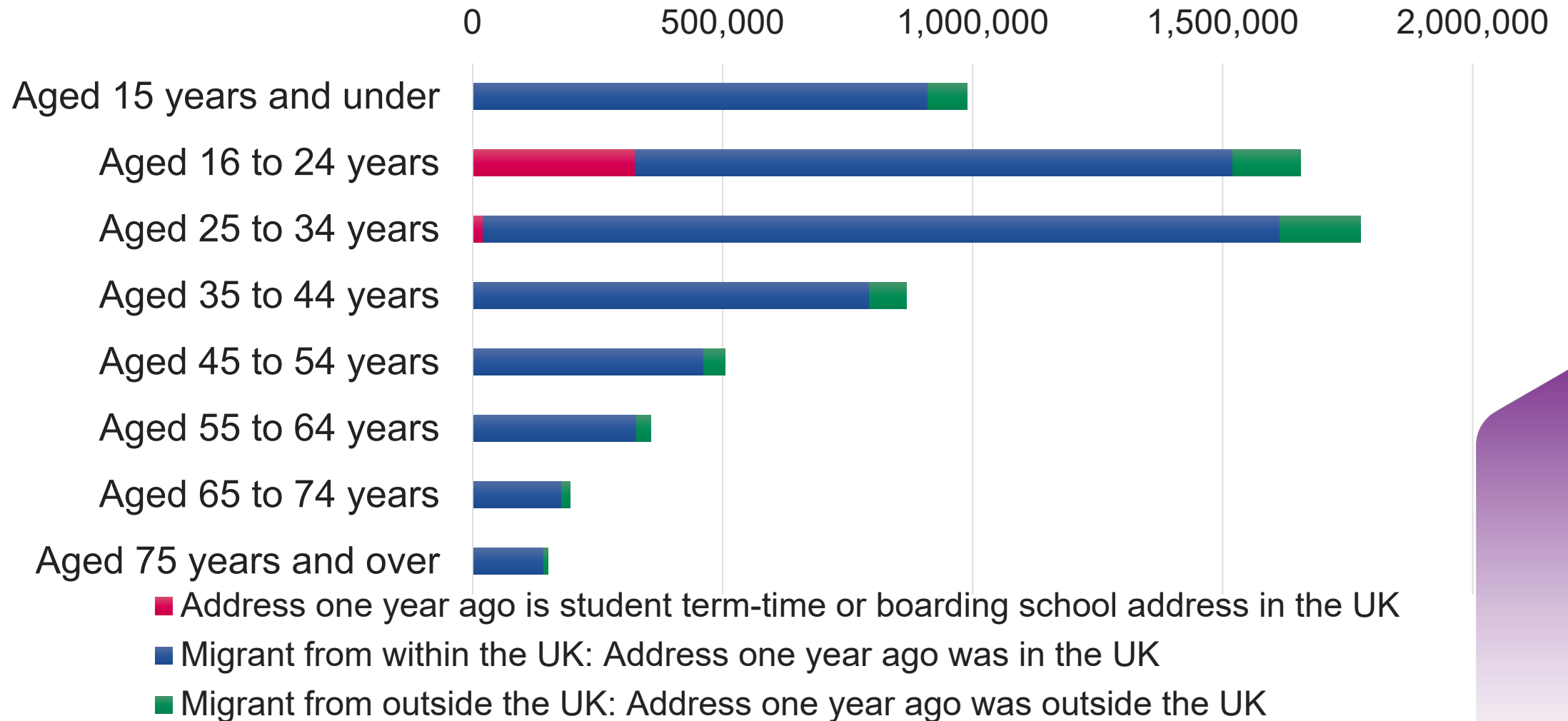
Significant fall in migration between 2011 and 2021 in England and Wales



Handling of sensitive variables for flow data in previous censuses



One-year migrants in England and Wales, 2021



Source: 2021 Census, England and Wales

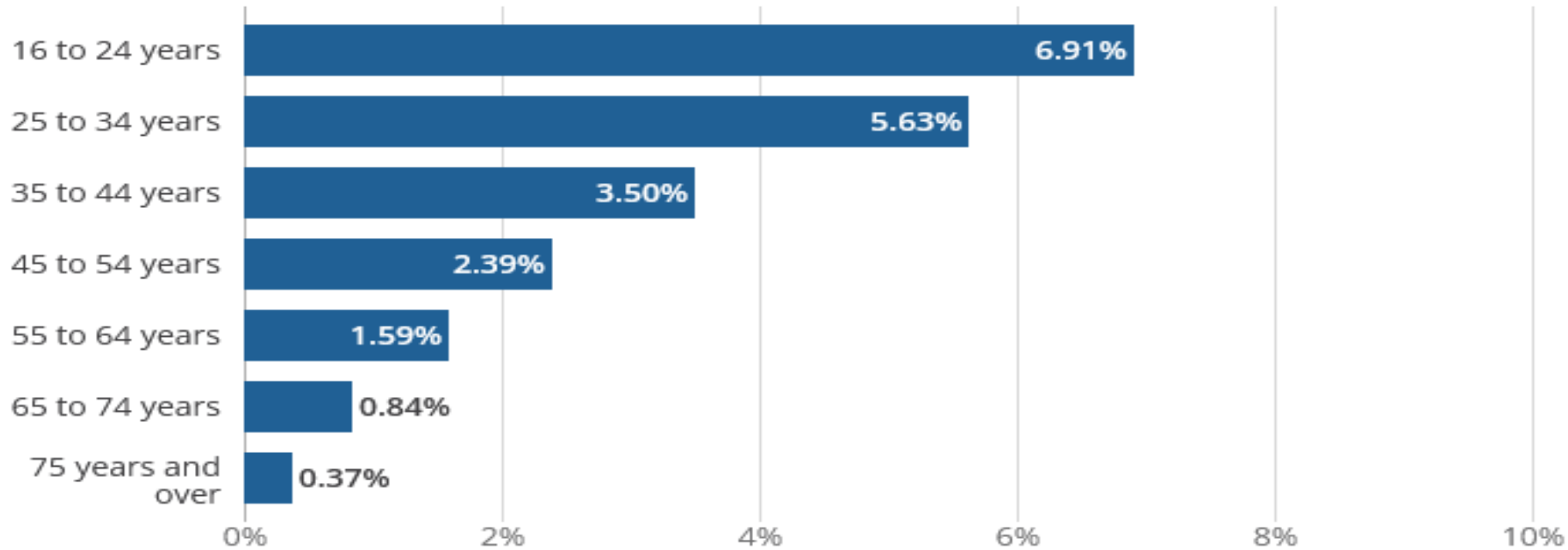
<https://www.ons.gov.uk/datasets/create/filter-outputs/44d65627-fb10-444f-a6de-9e7d39a4be8a>

LGB+ population by age in England and Wales

Figure 1: Young people were the most likely to have identified as lesbian, gay, bisexual or another minority sexual orientation (LGB+)

Percentage of usual residents aged 16 years and over who identified as LGB+ by age, England and Wales, 2021

92.5% answered
3.2% identified as LGB+



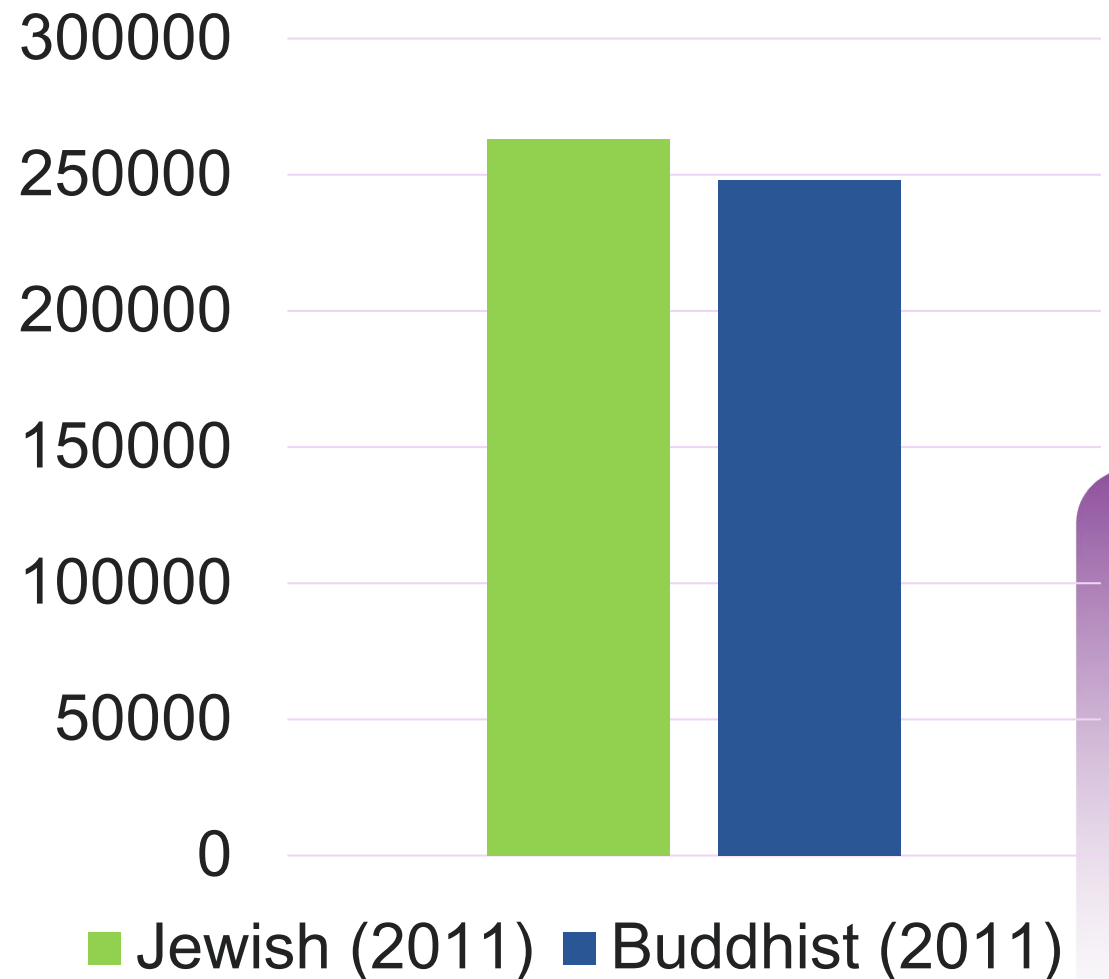
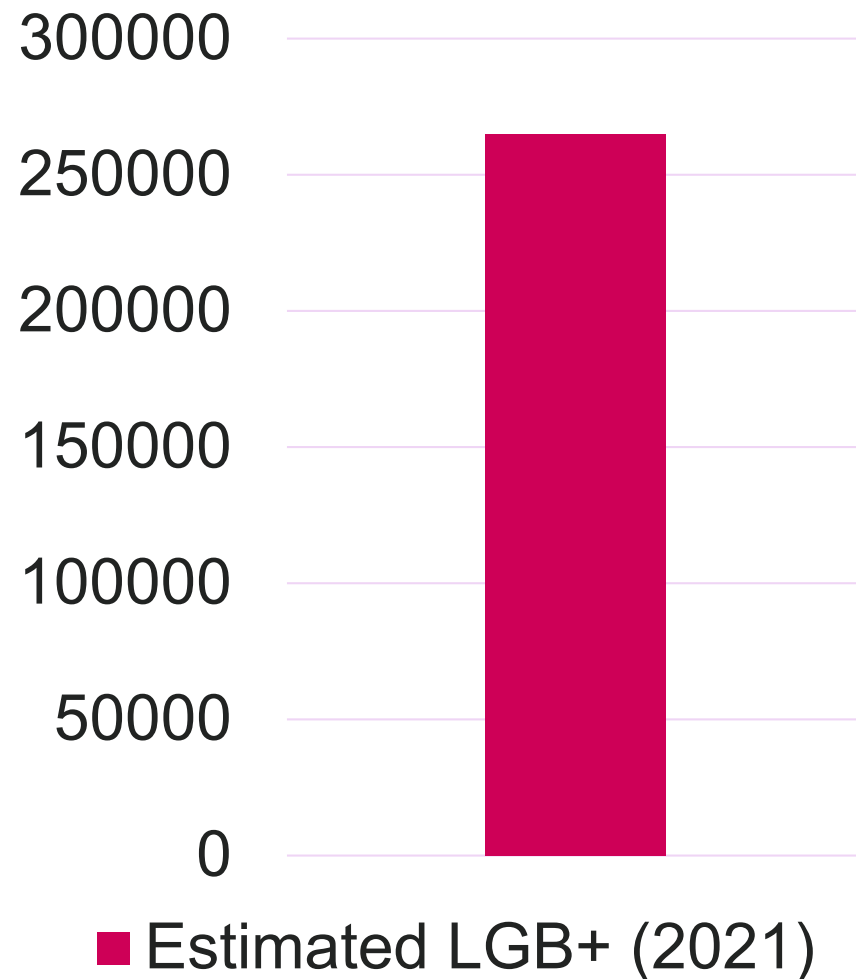
Estimated numbers of LGB+ migrants in England and Wales

Applying % LGB+ identification by age to internal migrants by age, we get an estimate of c. **265K LGB+ migrants** in England and Wales, 2020-21.

However, this assumes that LGB+ and non-LGB people have the same propensity to migrate.

We need data on migrants by age by sexual orientation to confirm or reject this.

Population sizes are comparable with past sensitive data

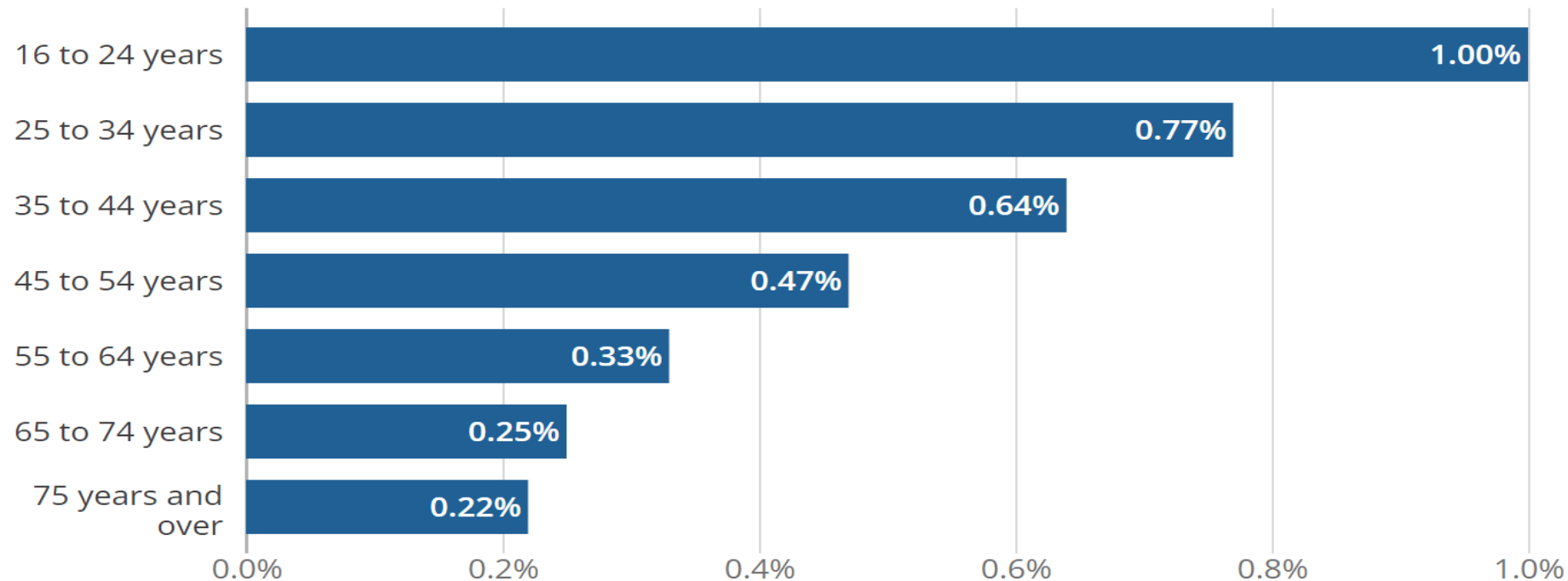


Trans population by age in census 2021

Figure 1: 1 in 100 people aged 16 to 24 years identified as trans

Percentage of usual residents aged 16 years and over who identified as trans by age, England and Wales, 2021

94% answered
0.54% identified as trans



Source: Office for National Statistics – Census 2021

Gender identity

The starting estimate (0.54% of migrants) would give around 35,000 trans migrants. Applying age specific migration and trans identification rates would give a revised estimate of about **40,000 trans migrants**.

Again, this relies on an untested assumption that propensity to migrate is the same in trans and cis populations.

Gender identity would likely need stronger access controls.

Journey to work data

Far more people travel to work than migrate. However, the disclosure risks posed by location tracing are far more pertinent in journey to work data, and fine scale geography is problematic.

Applying similar calculations we reach estimates of:

- **568,000** people travelling to work in the UK (not WFH) and LGB+
- **92,000** people travelling to work in the UK (not WFH) and trans

Again we assume LGB+/trans people have similar likelihood of travelling to work as the whole population.

What bespoke tables might be feasible with 2021 data?

	Sexual orientation	Gender identity
Migration	No public tables	No public tables
	Safeguarded at LA level with grouped responses	No safeguarded tables
	Secured at LA/wards level with detailed responses	Secured at regions/LA level with detailed responses
Journey to work	No public tables	No public tables
	Safeguarded at LA level with grouped responses	No safeguarded tables
	Secured at LA/MSOA level with detailed responses	Secured at regions/LA level with grouped or detailed responses



Thank you.

Oliver Duke-Williams, o.duke-Williams@ucl.ac.uk

Vassilis Routsis, v.routsis@ucl.ac.uk

