UK Data Service

Census microdata

Nigel de Noronha

Economic and Social Research Council

Overview

- what is on offer
- types of analysis
- preparing the data
- simple analysis
- practical exercise
- question and answer

UK Data Service

What is on offer

Census microdata

Datasets

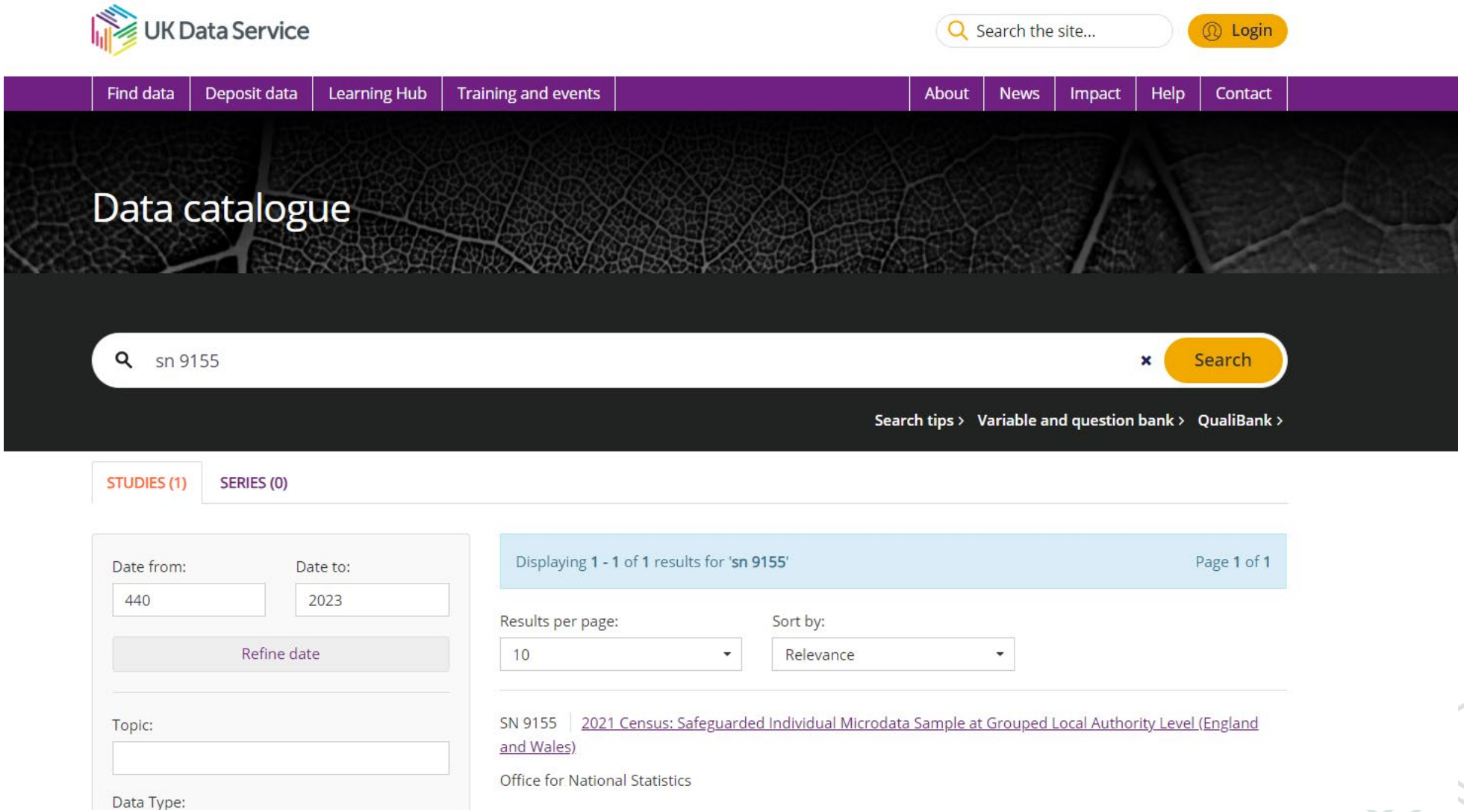
Individual

- 1% open access at regional level (18 variables)
- 1% safeguarded access IPUMS (internationally deposited)
- 5% safeguarded access combined local authority (85 variables)
- 5% safeguarded access region (86 variables)
- 10% secure access local authority

Household

- 1% safeguarded access household data at regional level (55 variables)
- 10% secure access local authority

Safeguarded data (1)



Safeguarded data (2)

Home > Data catalogue > Studies > Study

Studies

Series

https://doi.org/10.5255/UKDA-SN-9155-1

Copy study DOI

2021 Census: Safeguarded Individual Microdata Sample at Grouped Local Authority Level (England and Wales)

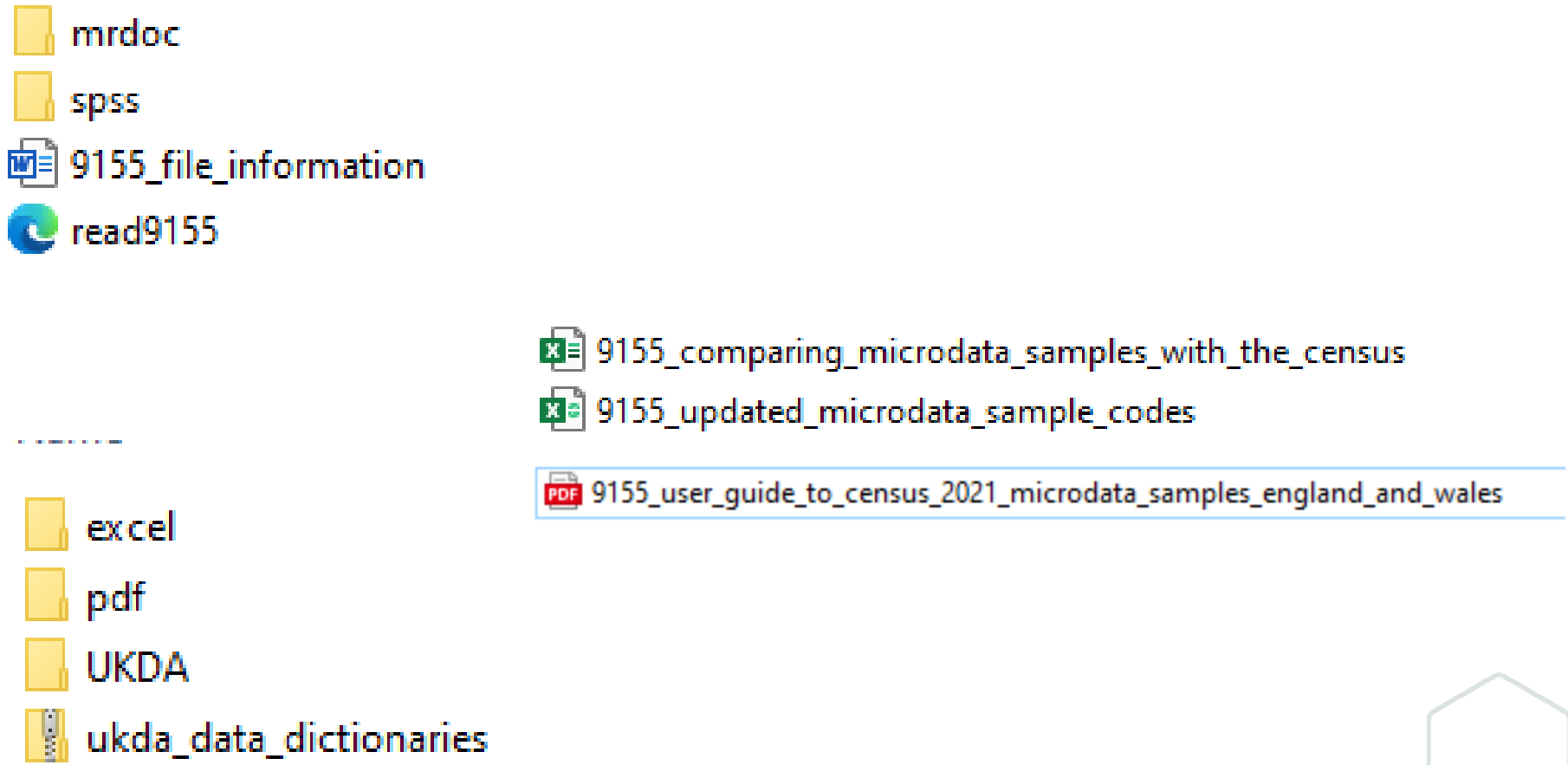
Details | Documentation | Resources | Access data

Details

Title:	2021 Census: Safeguarded Individual Microdata Sample at Grouped Local Authority Level (England and Wales)
Study number (SN):	9155
Access:	These data are safeguarded
Persistent identifier (DOI):	10.5255/UKDA-SN-9155-1
Series:	UK Census
Data creator(s):	Office for National Statistics

Sponsors and contributors

Supporting materials



UK Data Service

Types of analysis

Census microdata

Beyond aggregate data

Multiple variables enable

- more complex tables and charts
- exploratory analysis of different factors
- regression (combining variables in a single analysis to identify associations with a variable of interest) e.g.

    housing deprivation ~ age, sex, disability, ethnicity, migration history

Results depend on type of variable – housing deprivation is binary, so results are the odds of being housing deprived by each of the variable characteristics selected

UK Data Service

Preparing the data

Processing suggestions

Select

- Variables and records you want to explore

Prepare

- Recode categories and descriptions for variables
- Derive variables e.g. revised age bands

Aggregate

- To save space combine into groups with count of number of the same observations

Selecting variables and records

Access in a format that you can use

(SPSS, Stata, TAB)

Filter out records you do not want e.g. create a London dataset

Load the file into a format that you can manipulate, preferably programmatically and drop the variables you don’t need

Prepare data

Recode variables

Category	Description	Count
1	Aged 4 years and under	162078
2	Aged 5 to 9 years	178345
3	Aged 10 to 15 years	217524
4	Aged 16 to 18 years	106126
5	Aged 19 to 24 years	241037
6	Aged 25 to 29 years	196507
7	Aged 30 to 34 years	208137
8	Aged 35 to 39 years	199771
9	Aged 40 to 44 years	187913
10	Aged 45 to 49 years	189559
11	Aged 50 to 54 years	206318
12	Aged 55 to 59 years	201803
13	Aged 60 to 64 years	173119
14	Aged 65 to 69 years	147355
15	Aged 70 to 74 years	148941
16	Aged 75 to 79 years	108654
17	Aged 80 to 84 years	75877
18	Aged 85 years and over	72547

Category	Description	Count
1.00	under 16	557947
2.00	16-24	347163
3.00	25-34	404644
4.00	35-44	387684
5.00	45-54	395877
6.00	55-64	374922
7.00	65-74	296296
8.00	75 or over	257078

Derive variables

Some variables overlap categories so may need to be considered differently: e.g. simplified economic activity to full-time, part-time, unemployed, retired or other economically inactive **AND** added a flag for those who were students

Some desirable variables are not there

e.g. household composition

You can identify some useful categories e.g. single person, couple by relationship to HRP perhaps, dependent children.

Aggregated data

Combine items with the same combination of variables and a count of the number of observations:

1	age	sex	disable	ethnic	yar	region	tenure	hd	count
2	1.00	1.00	-8.00	-8.00	-8.00	0.00	-8.00	-8.00	36
3	1.00	1.00	-8.00	-8.00	-8.00	0.00	0.00	0.00	204
4	1.00	1.00	-8.00	-8.00	-8.00	0.00	0.00	1.00	7
5	1.00	1.00	-8.00	-8.00	-8.00	0.00	1.00	0.00	985
6	1.00	1.00	-8.00	-8.00	-8.00	0.00	1.00	1.00	19
7	1.00	1.00	-8.00	-8.00	-8.00	0.00	2.00	0.00	226
8	1.00	1.00	-8.00	-8.00	-8.00	0.00	2.00	1.00	22
9	1.00	1.00	-8.00	-8.00	-8.00	0.00	3.00	0.00	716
10	1.00	1.00	-8.00	-8.00	-8.00	0.00	3.00	1.00	22
11	1.00	1.00	-8.00	-8.00	-8.00	0.00	4.00	0.00	4
12	1.00	1.00	-8.00	-8.00	-8.00	1.00	-8.00	-8.00	8

Age	**Banded age**	
Code	*Description*	*Count*
1	under 16	557947
2	16-24	347163
3	25-34	404644
4	35-44	387684
5	45-54	395877
6	55-64	374922
7	65-74	296296
8	75 or over	257078
Disable	**Disabled (as Equality Standard)**	
Code	*Description*	*Count*
1	Disabled, activities limited a lot	223107

UK Data Service

Some considerations

When looking further into the data

Issues to be aware of

- To access the safeguarded microdata, you need to be registered with the UK Data Service
- Use the documentation to ensure you understand the variables available
- Consider grouping categories
- Some categories automatically exclude responses e.g. using economic activity removes young people from the analysis
- Geography is limited to combined local authority
- As well as the individual you have attributes of the household reference person

Using outside statistical software

Demonstration SPSS script that can be used to create Excel worksheets

You can translate this to other programmatic languages using ChatGPT – but please check for accuracy

This can then be explored or loaded into other tools for analysis

You will need to create metadata (variable descriptions and category definitions) from the source files as these do not transfer directly to Excel

For regression models

- Limited number of dependent variables –
  - no scale variables unless you generate them
  - Binary variables covering aspects of household deprivation (education, employment, health and housing)
- Some attributes are held in more than one place – make sure you resolve, or the overlap will mess up any regression

UK Data Service

Demonstration of multivariate analysis

Census microdata

Exploring household deprivation

Using combination of individual and household characteristics:

- Age (15 year bands)
- Sex
- Disability
- Occupational social class of household reference person (three groups)
- Ethnicity
- Year of arrival in the UK (decades since 1971)
- Tenure

For England and Wales and London

Outcomes

Using logistic regression, I have identified the odds of someone with a particular characteristic being housing deprived compared to a reference group within each variable. I also show the statistical significance of the result

*	Less than 5%
**	Less than 1%
***	Less than 0.1%

All the variables were considered together so I can claim that the results reflect the odds ratio of a particular characteristic after controlling for the other variables in the model.

Results

Odds ratio of experiencing housing deprivation in England and Wales and in London				
Variable	**E&W**	**Significance**	**London**	**Significance**
Age				
under 16 (reference group)				
16-29	1.08	***	-1.17	***
30-44	-1.57	***	-1.73	***
45-59	-1.85	***	-1.97	***
60-74	-3.37	***	-3.09	***
75 or over	-4.07	***	-3.61	***
Sex				
Male	1.03	***	1.03	
Occupational social class of household reference person				
Professional or managerial (reference category)				
Intermediate	1.72	***	1.86	***
Routine	2.01	***	2.17	***
Never worked or long-term unemployed	1.96	***	1.76	***
Student	2.10	***	1.65	***

Results

Odds ratio of experiencing housing deprivation in England and Wales and in London (2)				
Variable	**E&W**	**Significance**	**London**	**Significance**
Ethnic group				
White British (reference group)				
White Irish	1.15	***	-1.09	*
White Gypsy or Traveller	3.25	***	1.87	***
Roma	2.90	***	1.84	***
White other	1.86	***	1.50	***
Indian	2.58	***	1.85	***
Pakistani	5.16	***	3.11	***
Bangladeshi	5.75	***	4.01	***
Chinese	1.45	***	1.05	
Asian other	2.90	***	1.99	***
Black African	3.73	***	2.79	***
Black Caribbean	2.27	***	1.59	***
Black other	2.95	***	2.21	***
Mixed white and black African	1.92	***	1.79	***
Mixed white and black Caribbean	1.61	***	1.47	***
Mixed white and Asian	1.41	***	1.14	***
Mixed other	1.84	***	1.50	***
Arab	2.75	***	2.09	***
Other ethnic group	2.79	***	1.91	***

Results

Odds ratio of experiencing housing deprivation in England and Wales and in London (3)				
Variable	**E&W**	**Significance**	**London**	**Significance**
Disability				
Not disabled (reference group)				
Activities limited a little	-1.02	**	-1.06	***
Activities limited a lot	-1.05	***	-1.19	***
Tenure				
Owned outright (reference group)				
Owned with a mortgage or shared ownership	-1.07	***	1.10	***
Social rented	2.90	***	3.56	***
Private rented	1.98	***	2.61	***
Rent free	2.65	***	4.24	***
Year of arrival in the UK				
Born in UK (reference group)				
before 1971	-1.17	***	-1.49	***
1971 - 1980	1.08	***	-1.20	***
1981 - 1990	1.24	***	1.02	
1991 - 2000	1.44	***	1.32	***
2001 - 2010	1.27	***	1.40	***
2011 - 2021	1.39	***	1.53	***

UK Data Service

Questions and discussion

nigel.denoronha@manchester.ac.uk

https://ukdataservice.ac.uk