Social data and the third sector: skills, tools and evidence

Workshop 1. Evidence and impact using census data

Patty Doran, UK Data Service, University of Manchester
Dave Rawnsley, UK Data Service, Jisc

Date: 2pm, Thursday July 2nd, 2020

UK Data Service

What we are covering:

- Introduction
- Presentation – what are aggregate data?
- Quiz
- Demonstration – exploring with InFuse, GeoConvert, DKAN
- Activity – in own time, 10 mins
- Feedback from activity
- Questions and discussion

Introduction

What is the UK Data Service?

- A comprehensive resource funded by the Economic and Social Research Council (ESRC)
- A single point of access to a wide range of secondary social science data
- Support, training and guidance

UKRI Economic and Social Research Council

Workshop objectives

- Promote UKDS data and resources to third sector users.
- Increase understanding of how data can be used within a variety of organisations and settings.
- Support you to access relevant data.
- Enable you to produce evidence by enhancing data knowledge and skills.

Social Surveys and the Third Sector

- Services: Improving outcomes for marginalised groups, reducing inequalities, providing a range of support to those most in need.
- Funding: relies on demonstrating the need and the potential impact of an intervention.
- UKDS: Social survey and census data can help provide context and demonstrate where services are most needed.
- Many third sector organisations are small, often there is no dedicated research team. Often it may fall to the people delivering the services to also provide the evidence of need and impact.

Our main data types

- **Aggregate Data**
  - International macrodata – country level time series data
  - Census tabular data
- **Microdata**
  - UK surveys (e.g. Labour Force Survey)
  - Census microdata
  - International comparative surveys (e.g. European Social Survey)
  - Business data
- **Other data**
  - Mixed methods data
  - Qualitative data

Aggregate data

What are aggregate data?

David Rawnsley
UKDS Senior Data Technical Officer

Evidence and impact using census data

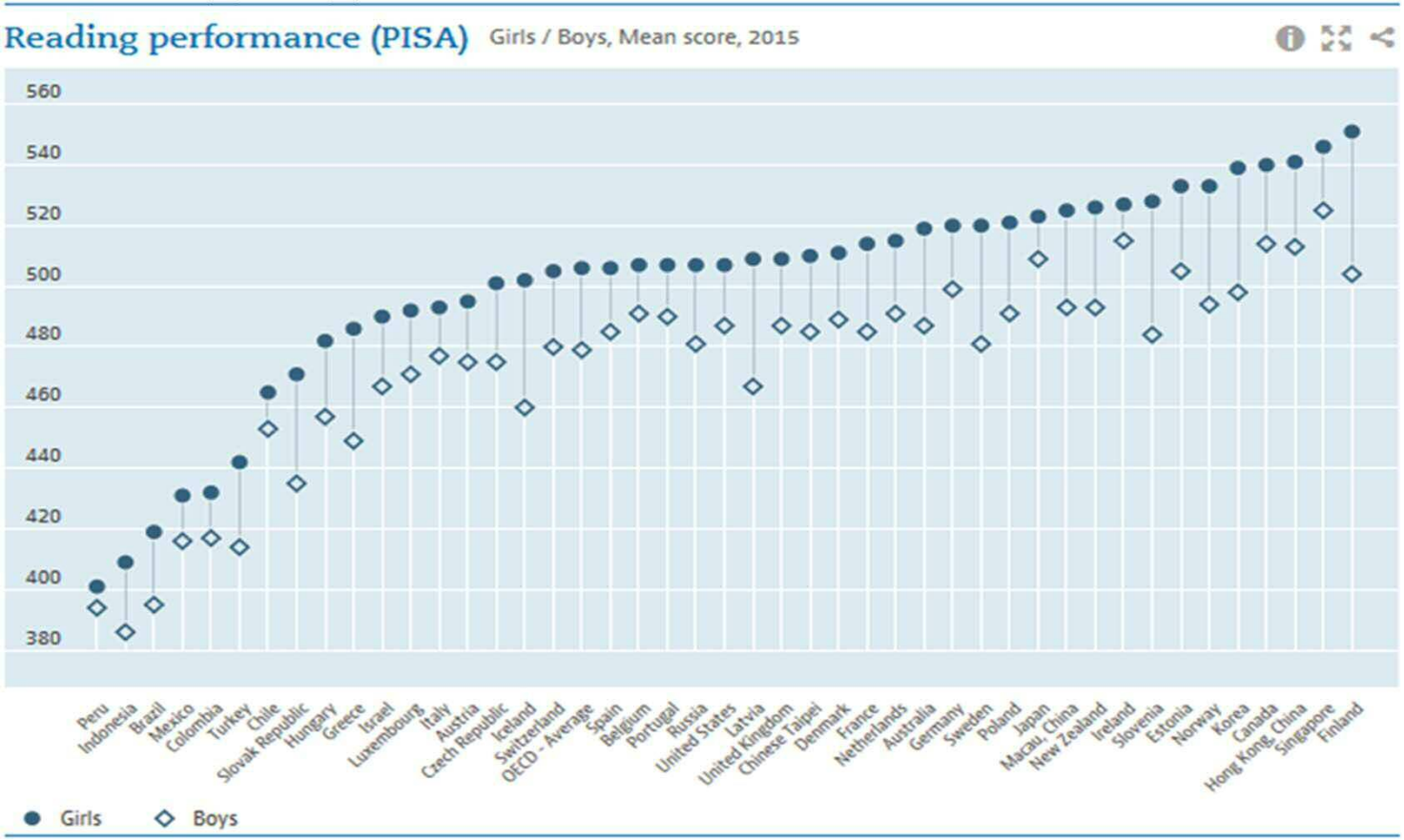
July 2nd 2020

Copyright © [year] UK Data Service. Created by [Organisation], [Institution]

CC BY-NC-SA

Aggregate data are data about
populations, groups, regions or countries

Averaging individual level data



OECD Chart: Reading performance (PISA), Girls / Boys, Mean score, Annual, 2015

59
35
17
22

Totalling headcounts

Age: 16-24 years old – religion (England and Wales): Muslim – Unit: people

GEO_LABEL	GEO_TYPE	F103600	F103603	F103606	F103609	F103612	F103615	F103618	F103621	F103624	F103627
		Age : Age 16 to 24 - Religion [E][S][W] : Sex : Total\ Sex - Unit : Persons									
Allerdale	Local Authorities	9230	6461	10	3	2	24	3	17	2143	567
Barrow-in-Furness	Local Authorities	7570	4636	10	1	1	13	0	19	2419	471
Carlisle	Local Authorities	11925	6903	36	37	2	45	8	40	4061	793
Copeland	Local Authorities	7170	5307	16	5	3	22	1	6	1392	418
Eden	Local Authorities	4758	2870	16	4	4	17	0	15	1486	346
South Lakeland	Local Authorities	8822	5048	27	3	7	19	3	26	2986	703

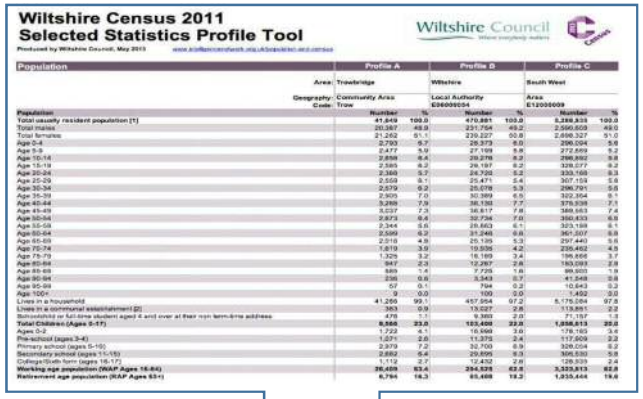
Office for National Statistics ; National Records of Scotland ; Northern Ireland Statistics and Research Agency (2017): 2011 Census aggregate data. UK Data Service (Edition: February 2017). DOI: http://dx.doi.org/10.5257/census/aggregate-2011-2

What Can Census Data Tell Us

- The most complete source of information about demographic and socio-economic characteristics of the UK population.

  - Population
  - Employment, occupation, socio-economic class, qualifications
  - Ethnicity, religion, national identity, language
  - Housing, tenure, residence type
  - Household and family composition, occupancy
  - Carers and caring
  - Long term health problems or disability
  - Travel to work, migration

How are census aggregate data produced

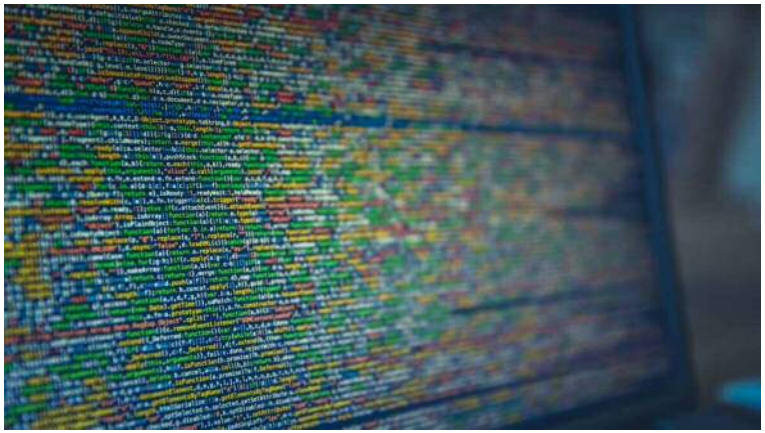


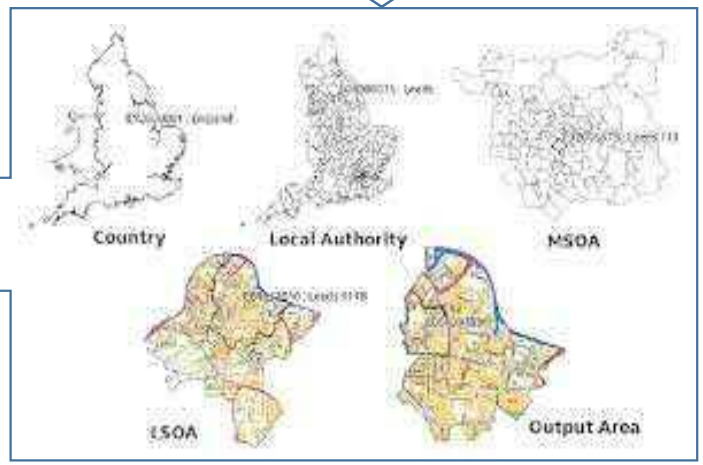
Wiltshire Census 2011
Selected Statistics Profile Tool

Wiltshire Council

Produced by Wiltshire Council, May 2013

Population	Profile A		Profile B		Profile C	
Area:	Trowbridge		Wiltshire		South West	
Geography:	Community Area		Local Authority		Area	
Code:	Trow		E06000054		E12000009	
Population	Number	%	Number	%	Number	%
Total usually resident population [1]	41,649	100.0	470,881	100.0	5,288,838	100.0
Total males	20,387	48.9	231,754	49.2	2,590,509	49.0
Total females	21,262	51.1	239,227	50.8	2,698,327	51.0
Age 0-4	2,793	6.7	26,373	6.0	296,094	5.6
Age 5-9	2,477	5.9	27,199	5.8	272,669	5.2
Age 10-14	2,858	6.4	29,278	6.2	298,892	5.6
Age 15-19	2,585	6.2	29,197	6.2	328,077	6.2
Age 20-24	2,369	5.7	24,720	5.2	333,168	6.3
Age 25-29	2,559	6.1	25,471	5.4	307,159	5.8
Age 30-34	2,579	6.2	25,078	5.3	296,791	5.6
Age 35-39	2,905	7.0	30,389	6.5	322,354	6.1
Age 40-44	3,298	7.9	36,136	7.7	375,938	7.1
Age 45-49	3,037	7.3	36,817	7.8	389,663	7.4
Age 50-54	2,873	6.4	32,734	7.0	350,633	6.6
Age 55-59	2,344	5.6	28,663	6.1	323,188	6.1
Age 60-64	2,599	6.2	31,246	6.6	361,307	6.8
Age 65-69	2,016	4.8	25,135	5.3	297,440	5.6
Age 70-74	1,819	3.9	19,936	4.2	235,462	4.5
Age 75-79	1,329	3.2	16,169	3.4	196,666	3.7
Age 80-84	947	2.3	12,267	2.6	153,093	2.9
Age 85-89	589	1.4	7,725	1.6	99,503	1.9
Age 90-94	236	0.6	3,343	0.7	41,548	0.8
Age 95-99	57	0.1	794	0.2	10,643	0.2
Age 100+	9	0.0	100	0.0	1,492	0.0
Lives in a household	41,266	99.1	457,954	97.2	5,175,084	97.8
Lives in a communal establishment [2]	383	0.9	13,027	2.8	113,861	2.2
Schoolchild or full-time student aged 4 and over at their non-term-time address	476	1.1	9,383	2.0	71,157	1.3
Total Children (Ages 0-17)	**8,566**	**23.0**	**103,400**	**22.0**	**1,058,613**	**20.0**
Ages 0-2	1,722	4.1	16,998	3.6	178,165	3.4
Pre-school (ages 3-4)	1,071	2.6	11,373	2.4	117,609	2.2
Primary school (ages 5-10)	2,979	7.2	32,700	6.9	326,054	6.2
Secondary school (ages 11-15)	2,882	6.4	29,895	6.3	306,530	5.8
College/Sixth form (ages 16-17)	1,112	2.7	12,432	2.6	128,939	2.4
Working age population (WAP Ages 16-64)	**26,409**	**63.4**	**294,525**	**62.6**	**3,323,813**	**62.8**
Retirement age population (RAP Ages 65+)	**6,794**	**16.3**	**65,408**	**18.2**	**1,035,444**	**19.6**

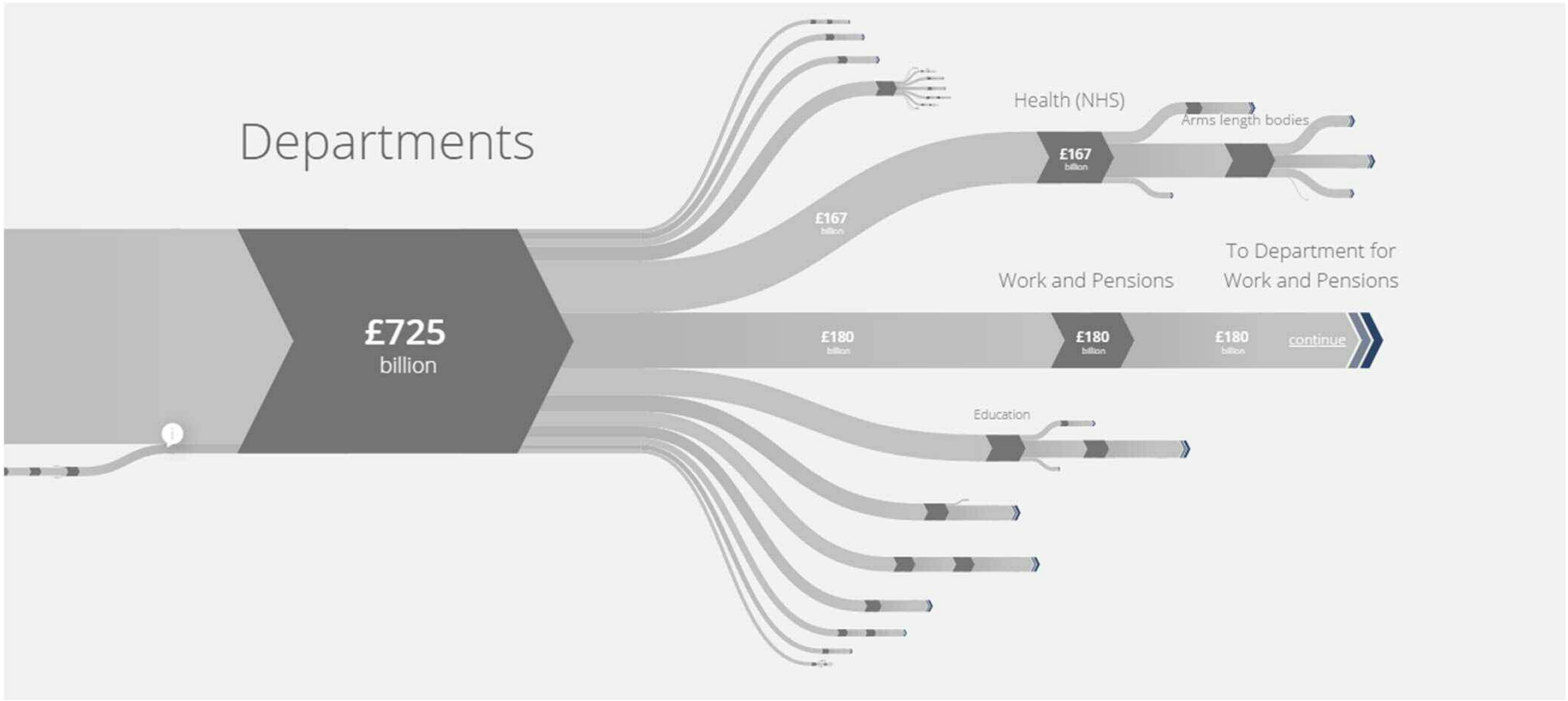




The Tricky Issue of Census Geography

- Building block is the Output Area
- These are used to create Super Output Areas (SOA)
  - Two types – Lower layer SOA (LSOA) and Middle layer SOA (MSOA)
  - In Scotland – Data Zones (DZ) and Intermediate Geographies (IG)
  - In Northern Ireland – just LSOA
- None of the above relate to anything ‘real’.
- Regions, Counties, Local Authorities, Wards and Electoral Divisions
- http://geoconvert.digitalresources.jisc.ac.uk/

Aggregate data: shaping policy



https://uk.wikibudgets.org/w/united-kingdom-budget-2015

Activity

UKDS Online Workshops: Social data and the third sector: skills, tools and evidence

Workshop 1. Evidence and impact using census data

Worksheet

Task:

1) **Find data to answer a question using InFuse.**
2) **Create and download a subset of data from InFuse**

About InFuse

InFuse is a free service providing easy access to aggregate data from the UK 2011 and 2001 censuses. InFuse is designed to guide users in selecting census aggregate data relating to combinations of characteristics and areas of interest to them.

Website: http://infuse.ukdataservice.ac.uk/

Activity: Downloading data

You are interested in the impact of COVID-19 on older people living alone in Birmingham. In particular, you want to answer the following question:

> ***What is the percentage of people aged 65 or over and living alone across all wards in Birmingham?***

Instructions

1. Select 2011 Census Data

2. Select preferred route to your data
   - Choose *Geography*

Work through the handout

- Follow the handout instructions
- You can post questions in the chat box in Go To Webinar
- Work through the exercise in your own time
- Answer the questions at the end
- We will **come back together at 3pm** to discuss the activity and any other questions

Activity questions

1. Did you download the data?
2. How many variables (different pieces of information) did you download?
3. How many one-person households, where the householder is 65 or over, are there in Bartley Green?
4. What percentage of households in Bartley Green is this?
5. What is the percentage of people aged 65 or over and living alone across all wards in Birmingham?

Steps: 1 2 3 4

Download

Show Guidance

Selected Category Combinations

- 1808 - Household composition : Total: Household composition - Unit : Households
- 1810 - Household composition : One person household - Unit : Households
- 1812 - Household composition : Aged 65 and over - Unit : Households

Selected Areas

- All Wards and Electoral Divisions (40 areas) in Local Authorities (324 areas) - Birmingham

File reference HHDCOM_UNIT

Download Data

Previous Get the Data

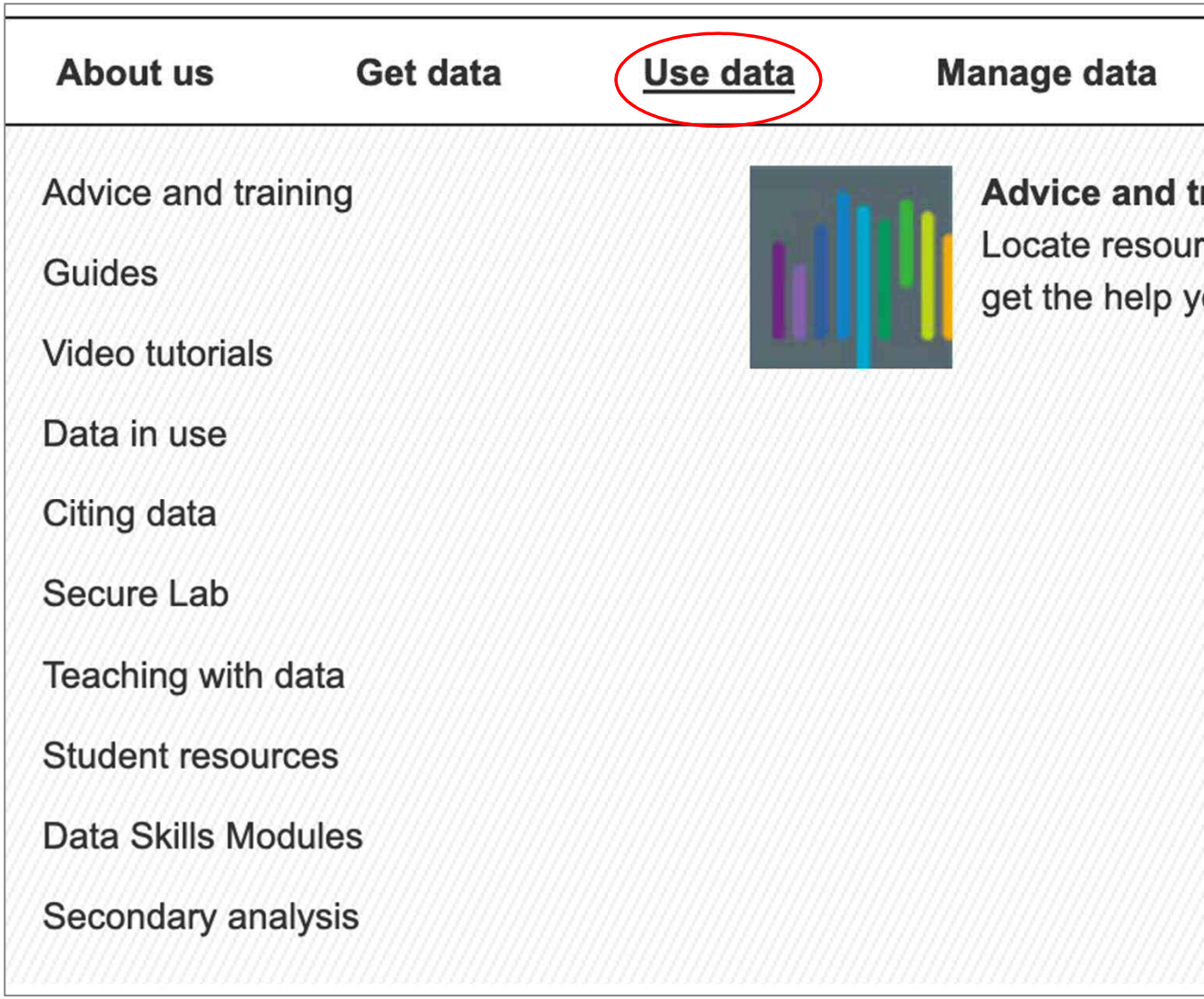
CDU_ID	GEO_CODE	GEO_LABEL	GEO_TYPE	GEO_TYP2	F1808	F1810	F1812	
					Total Households	One-person Households	One-person, 65+	Percentage
2205	E05001193	Ladywood	Wards and Electoral Divisions	WED	15661	7770	698	4.5%
2228	E05001216	Washwood Heath	Wards and Electoral Divisions	WED	9004	1960	667	7.4%
2207	E05001195	Lozells and East Handsworth	Wards and Electoral Divisions	WED	9532	2694	707	7.4%
2217	E05001205	Soho	Wards and Electoral Divisions	WED	10300	2947	770	7.5%
2219	E05001207	Sparkbrook	Wards and Electoral Divisions	WED	9406	2495	723	7.7%
2194	E05001182	Bordesley Green	Wards and Electoral Divisions	WED	9228	1851	719	7.8%
2220	E05001208	Springfield	Wards and Electoral Divisions	WED	9309	2178	759	8.2%
2191	E05001179	Aston	Wards and Electoral Divisions	WED	9742	2929	805	8.3%
2209	E05001197	Nechells	Wards and Electoral Divisions	WED	12045	4578	1058	8.8%
2214	E05001202	Selly Oak	Wards and Electoral Divisions	WED	8194	2022	746	9.1%
2200	E05001188	Handsworth Wood	Wards and Electoral Divisions	WED	9296	2565	923	9.9%
2196	E05001184	Brandwood	Wards and Electoral Divisions	WED	10912	3653	1431	13.1%
2192	E05001180	Bartley Green	Wards and Electoral Divisions	WED	10728	3712	1409	13.1%
2193	E05001181	Billesley	Wards and Electoral Divisions	WED	10749	3336	1418	13.2%
2204	E05001192	Kingstanding	Wards and Electoral Divisions	WED	10133	3249	1352	13.3%
2221	E05001209	Stechford and Yardley North	Wards and Electoral Divisions	WED	9871	3095	1326	13.4%
2195	E05001183	Bournville	Wards and Electoral Divisions	WED	11270	3889	1528	13.6%
2224	E05001212	Sutton New Hall	Wards and Electoral Divisions	WED	9433	2612	1290	13.7%
2225	E05001213	Sutton Trinity	Wards and Electoral Divisions	WED	10663	3146	1480	13.9%
2226	E05001214	Sutton Vesey	Wards and Electoral Divisions	WED	9635	2745	1344	13.9%
2211	E05001199	Oscott	Wards and Electoral Divisions	WED	10244	3023	1434	14.0%
2215	E05001203	Shard End	Wards and Electoral Divisions	WED	11603	4101	1635	14.1%
2198	E05001186	Erdington	Wards and Electoral Divisions	WED	10236	3872	1475	14.4%
2210	E05001198	Northfield	Wards and Electoral Divisions	WED	11370	3883	1643	14.5%
2216	E05001204	Sheldon	Wards and Electoral Divisions	WED	9203	2916	1413	15.4%
2223	E05001211	Sutton Four Oaks	Wards and Electoral Divisions	WED	10156	2890	1618	15.9%
				Totals:	**410736**	**131200**	**47586**	**11.6%**

Discussion

Resources

Online resources

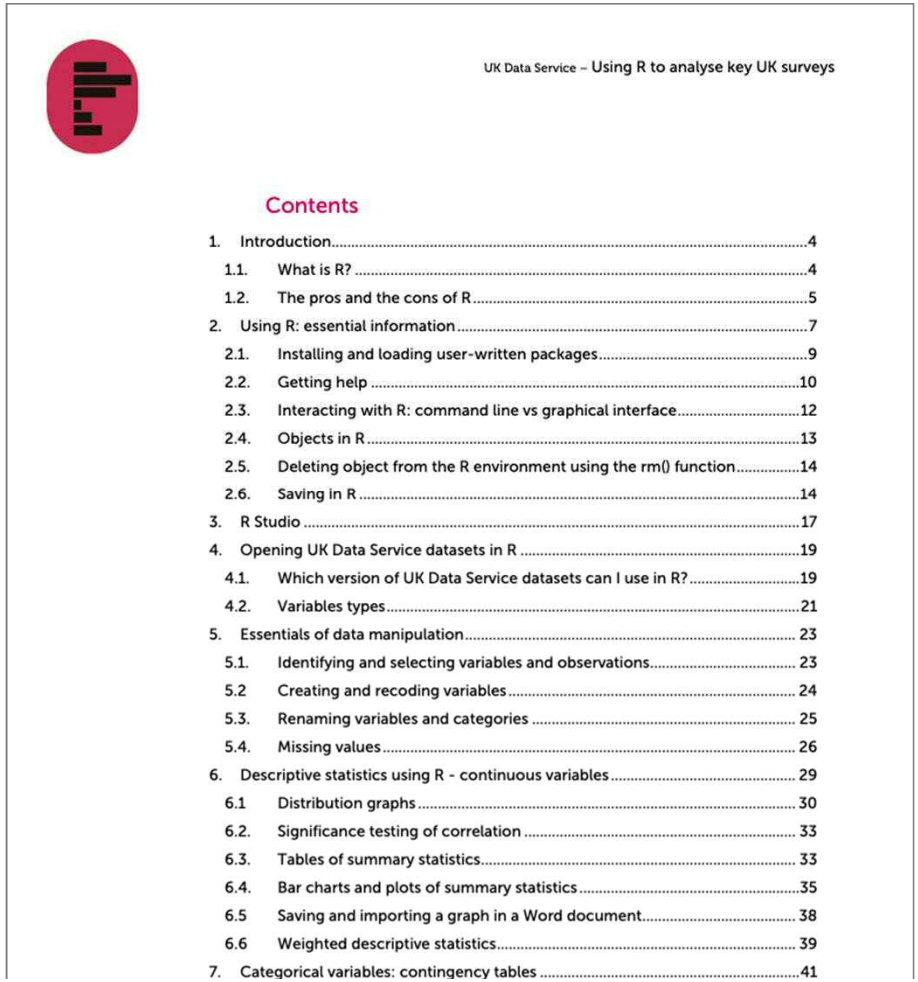
- Guides
- Videos
- Webinars
- Data Skills Modules



Methods and Software Guides

For example:

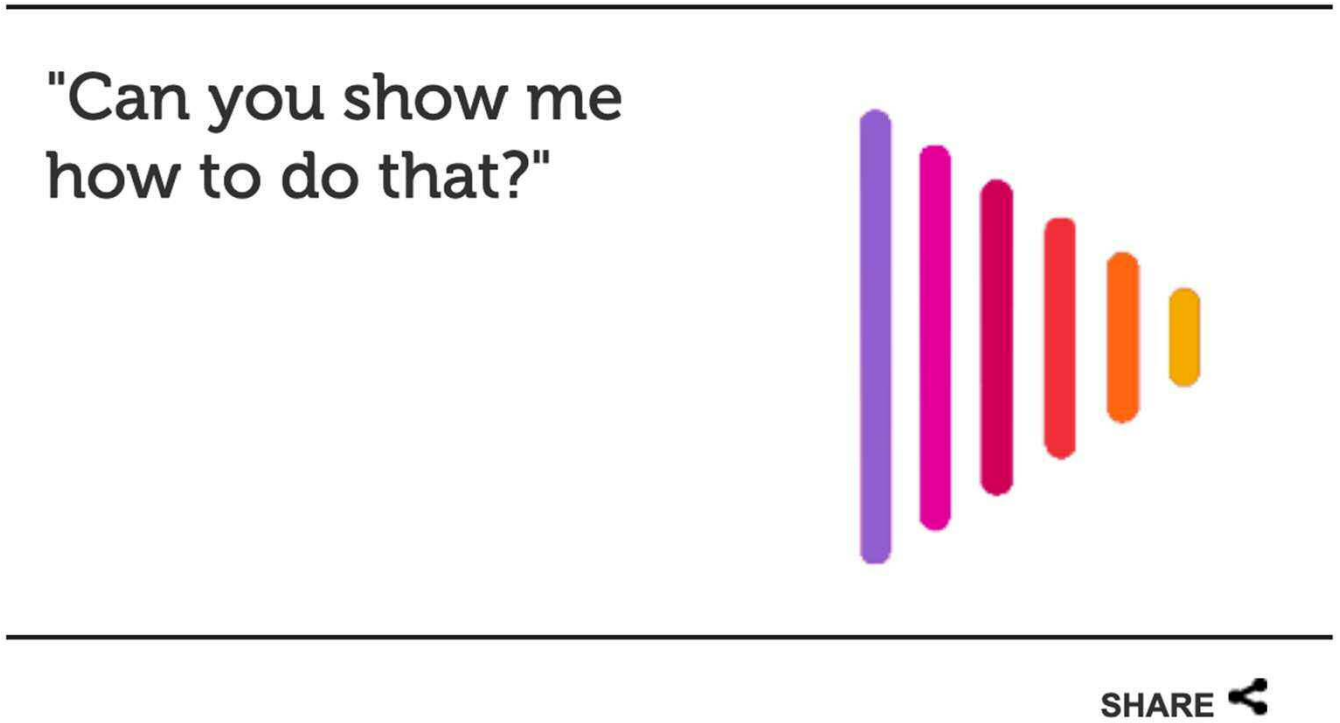
- Software guides
  - What is Stata?
  - What is SPSS 20 for Windows?
  - Using R to analyse key UK surveys
- Survey data
  - What is weighting?
  - What is complex sample design?
  - What are hierarchical files?



Contents

1. Introduction....4
1.1. What is R?....4
1.2. The pros and the cons of R....5
2. Using R: essential information....7
2.1. Installing and loading user-written packages....9
2.2. Getting help....10
2.3. Interacting with R: command line vs graphical interface....12
2.4. Objects in R....13
2.5. Deleting object from the R environment using the rm() function....14
2.6. Saving in R....14
3. R Studio....17
4. Opening UK Data Service datasets in R....19
4.1. Which version of UK Data Service datasets can I use in R?....19
4.2. Variables types....21
5. Essentials of data manipulation....23
5.1. Identifying and selecting variables and observations....23
5.2 Creating and recoding variables....24
5.3. Renaming variables and categories....25
5.4. Missing values....26
6. Descriptive statistics using R - continuous variables....29
6.1 Distribution graphs....30
6.2. Significance testing of correlation....33
6.3. Tables of summary statistics....33
6.4. Bar charts and plots of summary statistics....35
6.5 Saving and importing a graph in a Word document....38
6.6 Weighted descriptive statistics....39
7. Categorical variables: contingency tables....41

Videos



YouTube: https://www.youtube.com/user/UKDATASERVICE

Webinars

Showing 1 to 10 of 20 entries

2021 Census: What is planned and how will you use it? A UK Data Service users update
06 Mar 2020
Online, 11.00 - 12.00 GMT

Webinar: Investigating religion in the UK - what data can I use?
18 Mar 2020
Online, 15.00 - 16.00 GMT

Webinar: Investigating language spoken in the UK - what data can I use?
25 Mar 2020
Online, 15.00 - 16.00 GMT

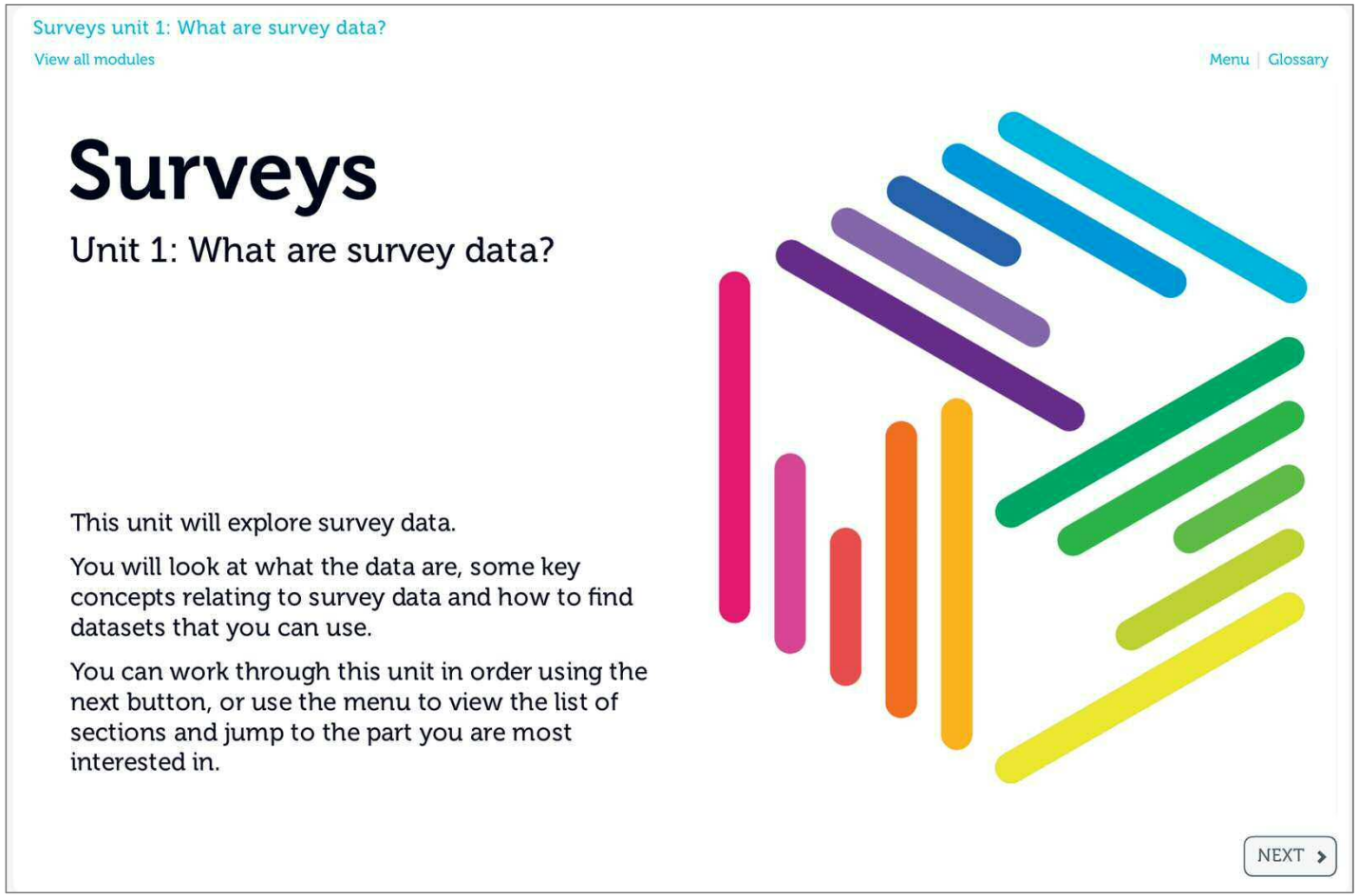
Webinar: Investigating political behaviour in the UK - what data can I use?
01 Apr 2020
Online, 15.00 - 16.00 BST

Webinar: Introduction to the UK Data Service
07 May 2020
Online, 15.00 - 16.00 BST

Data Skills Modules

- Survey data
- Aggregate data
- Longitudinal data
- Glossary

https://www.ukdataservice.ac.uk/use-data/data-skills-modules.aspx



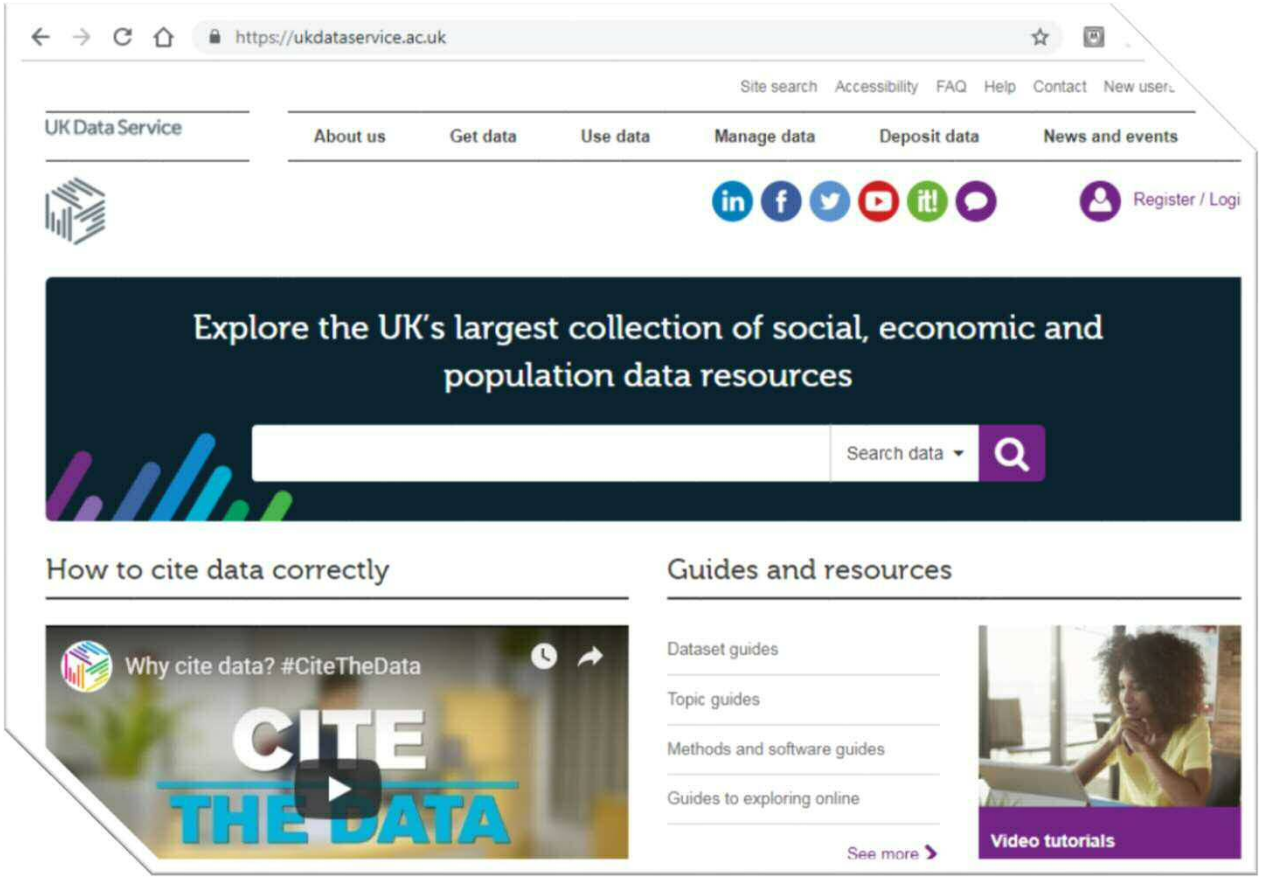
Thank you

web: https://ukdataservice.ac.uk/ YouTube: https://www.youtube.com/user/UKDATASERVICE

email: help@ukdataservice.ac.uk

twitter: @UKDataService

Visit the
UK Data Service



UK Data Service