# An Introduction to the UK Data Service: Resources for Social Research



Simon Parker and Valerija Kolbas 22nd November 2018



# Can you hear us?



# Can you hear us?

#### If Not:

 Check your volume, and that your speaker/headset is plugged in

 Your invitation also included a phone number; you can call that to listen in





## Overview

- Who are we?
- What kinds of data do we hold?
- Online resources
- Further help
- Questions?



**UK Data Service** 

Introduction to the UK Data Service

**22 November 2018** 



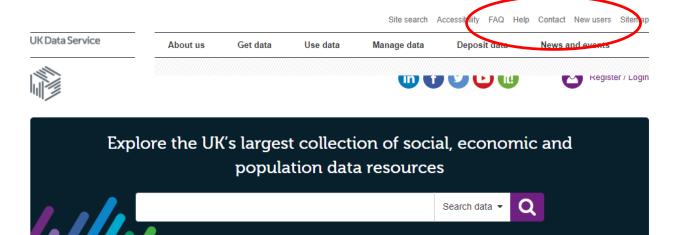
## What is the UK Data Service?

- A comprehensive resource funded by the ESRC
- The UK Data Service provides a single point of access to a wide range of secondary social science data
- The UK Data Archive is the lead organisation
- Support, training and guidance





## **UK Data Service**



#### About the UK Data Service



#### Guides and resources

Topic guides	
Methods and software guides	
Guides to exploring online	Video tutorials
See more >	See our range of training videos

See data from all over the world

Browse our data map



## Who is it for?

- Academic researchers and students
- Government analysts
- Charities and foundations
- Business consultants & data analysists
- Independent research centres & think tanks





## Sources of data



- National statistical authorities
- UK government departments
- Intergovernmental organisations
- Research institutes
- Individual researchers















## Types of data

#### **Survey Microdata**

Major UK surveys

**Cross-national** 

Longitudinal

#### Census

Aggregate data for 1971-2011

Microdata for 1991-2011

#### **International Macrodata**

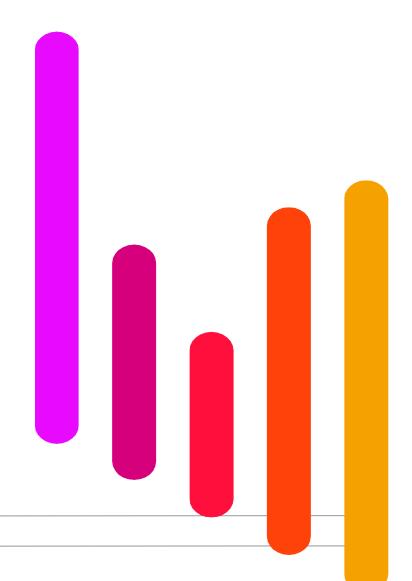
Multi-nation aggregate databanks and survey data

#### **Qualitative/Mixed methods**

Range of multimedia
qualitative data sources
Mix of qualitative and
quantitative data



# Survey Microdata



**UK Data Service** 



## Survey microdata

- Individual or household level data which have been anonymised to prevent disclosure
- Needs to be analysed using an analysis package like SPSS or Stata
- Very flexible as you can produce your own tables, populations and attributes



# Survey microdata: working

(D) Age 16-65+ in ten year age bands \* (D) BMI grouped with Obese categories I, II, III Crosstabulation

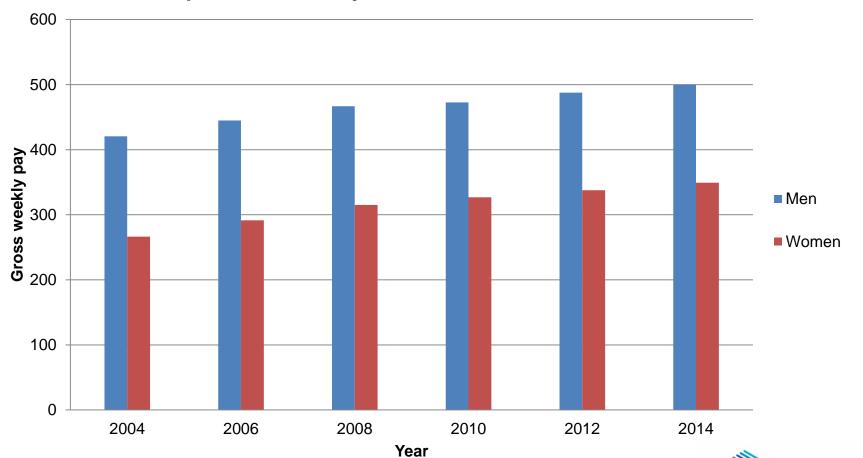
#### Count

			(D) BMI grouped with Obese categories I, II, III					
		Underweight: less than 18.5	Normal: 18.5 to less than 25	Overweight: 25 to less than 30	Obese I: 30 to less than 35	Obese II: 35 to less than 40	Obese III: 40 or more	Total
(D) Age 16-65+ in ten	16-24	25	235	84	29	14	9	396
year age bands	25-34	9	306	180	83	37	11	626
	35-44	5	302	324	134	43	20	828
	45-54	5	301	341	190	68	33	938
	55-64	4	210	322	184	58	25	803
	65+	9	328	591	274	71	17	1290
Total		57	1682	1842	894	291	115	4881



# Survey microdata outputs

Gross Weekly Pay (£) in Main Job by Sex, 2004-2014 Annual Population Survey



## **UK Surveys**

- Data about individuals or households
- Large sample sizes
- Nationally-representative
- Cross-sectional surveys
- Many surveys are repeated annually
- Cover many topics:
  - health
  - work
  - crime
  - social attitudes
  - family expenditure
  - living costs
  - housing





## Key data: UK surveys

- Labour Force Survey (1975 onwards)
- Annual Population Survey (2004 onwards)
- Family Expenditure Survey (1961-2001)
- Households Below Average Income (1994/95 2015/16)
- Health Surveys (England/Wales/Scotland) (1991 onwards)
- Crime Survey for England and Wales (1982 onwards)
- Family Resources Survey (1993 onwards)
- Citizenship Survey (2001-2011)
- British Social Attitudes Survey (1983 onwards)



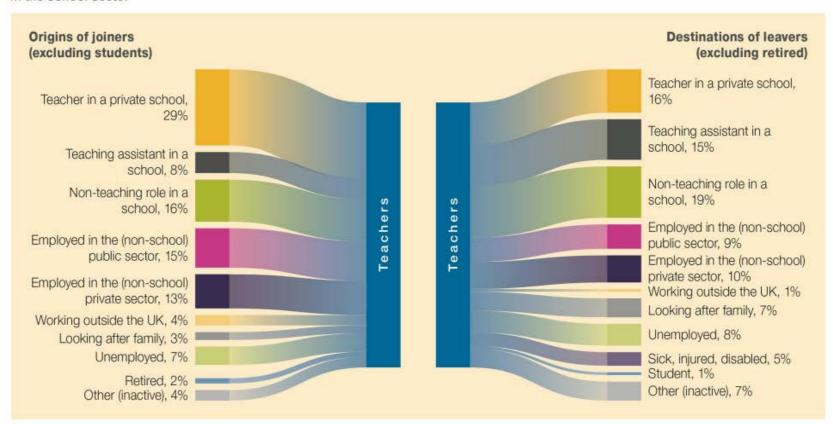
## Labour Force Survey

- Started as an annual survey in 1975
- Quarterly from 1991
- Uses international definitions of employment, unemployment & economic inactivity
- Used to inform social, economic & employment policies
- Topics include:
  - Occupation
  - Training
  - Hours of work
  - Personal characteristics of household members aged 16 & over



# Should I Stay or Should I Go? NFER Analysis of Teachers Joining and Leaving the Profession

Figure F More than half of non-student joiners come from jobs in the school sector; more than half of non-retiring leavers go to jobs in the school sector





# Longitudinal data

#### Similar to 'UK Surveys'

- Individual level data
- Large samples, nationallyrepresentative
- Surveys are repeated, often annually

#### BUT

- They follow the same individuals over time
- New respondents are added regularly to keep numbers up
- Allow researchers to analyse change at the individual level





## Key data: Longitudinal / Panel studies

- National Child Development Study (NCDS)
- 1970 British Cohort Study (BCS70)
- Millennium Cohort Study (MCS)
- British Household Panel Survey (BHPS)
- Understanding Society (UKHLS)
- English Longitudinal Study of Ageing (ELSA)
- Growing Up in Scotland (GUS)
- Next Steps



## British Household Panel Survey

- Collected and deposited by the Institute for Social and Economic Research at the University of Essex
- Follows the members of 5500 households first sampled in 1991 every year to 2008
- Multi-topic survey:
  - income, labour market behaviour, social and political values, health, education, housing and household organisation



# **Understanding Society (UKHLS)**

- Began in 2009, and builds on the long running British Household Panel Survey (BHPS) established in 1991
- Expanded to 40,000 households and 100,000 individuals including the original BHPS sample
- Tracks change at individual and household level by examining:
  - what people are experiencing (circumstances and key events in their lives as they happen)

Understanding Society

- attitudes, identity and subjective well-being (expressions)
- how individuals and households respond to policies and key events (behaviours)
- changes in societal conditions (outcomes)

## **Cohort Studies**

- 1958 National Child Development Study (NCDS)
- 1970 British Cohort Study (BCS70)
- Millennium Cohort Study (MCS)
- Study of Twins in the NCDS and the 1970 BCS, 2008
- Different decades generational change
- Health/medical focus
- Social and economic circumstances



# Data collection sweeps

1965 - age 7

1969 - age 11

1974 - age 16

1981 - age 23

1991 - age 33

1999-2000 – age 41-42

2004-05 – age 46-47

2008-09 – age 50

2013-14 – age 55

NEXT: 2018 – age 60

1958 National Child Development Study	1970 British Cohort Study	Millennium Cohort Study
1958 - birth	1970 - birth	2000-02 – birth
1000 bildi	1010 bildi	2000 02 BIITI

2001-03 – 9 months

2006 – age 5

2008 - age 7

2012 - age 11

2015 - age 14

NEXT: 2018 – age 17

1975 - age 5

1980 - age 10

1986 - age 16

1996 - age 26

2012 - age 42

1999-2000 – age 29-30

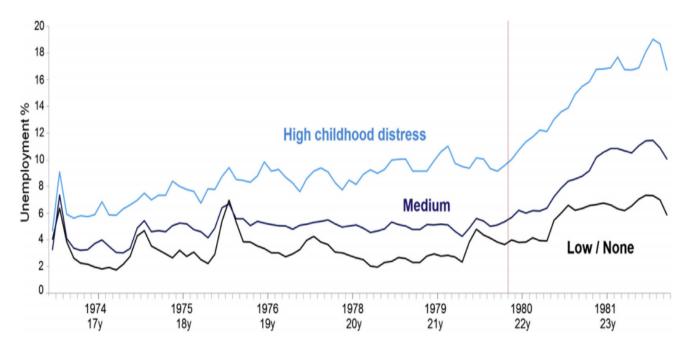
2004-05 - age 34-35

2008-09 - age 38-39

NEXT: 2016 – age 46

FUTURE: 2020 – age 50

# Childhood psychological distress and youth unemployment: Evidence from two British cohort studies Mark Egan, Michael Daly, Liam Delaney



**Fig. 2.** Descriptive statistics describing unemployment in the NCDS from August 1974—November 1981 by levels of childhood distress. The year above the cohort members' age refers to March of that year, the month when the cohort members were born. The vertical line denotes the onset of the U.K. recession in January 1980. Five months at the beginning and end of the data-range are omitted due to small sample sizes.



# **Cross-national Surveys**

- Individual or household level data
- Allow comparisons across different countries
- Topics covered include:
  - education
  - income
  - working conditions
  - poverty
- Examples:
- Eurobarometer Surveys
- European Quality of Life Survey
- European Social Survey
- European & World Values Studies
- International Social Survey Programme





## International Macrodata



**UK Data Service** 



# Aggregate data

- Aggregated: statistics or counts
- Tabular format
- Units are geographical areas
- Can be used
  - as they are
  - for further analysis trends, relationships
  - as contextual information



#### International Macrodata

- Time series data aggregated to country/region
- Regularly updated
- International governmental organisations (IMF, OECD, IEA, World Bank)
- Only available to staff and students from UK institutions of higher and further education (World Bank data are open access)





#### International Macrodata themes

#### Databanks cover:

- Economic performance and development
- Trade, industry and markets
- Employment
- Demography, migration and health
- Governance
- Human development
- Social expenditure
- Education
- Science and technology
- Land use and the environment



## International Macrodata



Login | Contact Us | Guides | Getting Starts

_					
	UK	DS	.8	ta	t

Disseminating international time-series data for the UK Data Service

20 October: (%) We celebrate World Statistics Day

15 July: We have launched a new version of UKDS.Stat! Please take a look at some of the new features and enhancements.

Login to view the full data portfolio (UK HE/FE only), or browse World Bank, OECD and IMF Open Data.

To get started, select data in the left-hand menus or select a dataset from the table below.

UKDS. Stat user guides and video tutorials provide extensive help on accessing the data. Please do get in touch if you have any guestions for us.

UKD S. Stat platform	Other platforms and data formats		
World Bank datasets: ADI   IDS   WDI	IMF eLibrary Text Collection Currently unavailable		
IMF datasets: BOPS   DOTS   GFS   IFS   WEO			
OECD delegate	OECD International Trade by Commodity Statistics (2010) in CSV format		
IEA datasets	United Nations Comtrade Database (premium access)		
Unido industrial statistics: IDSB   INDSTAT2   INDSTAT4	Eurostat New Cronos Database		

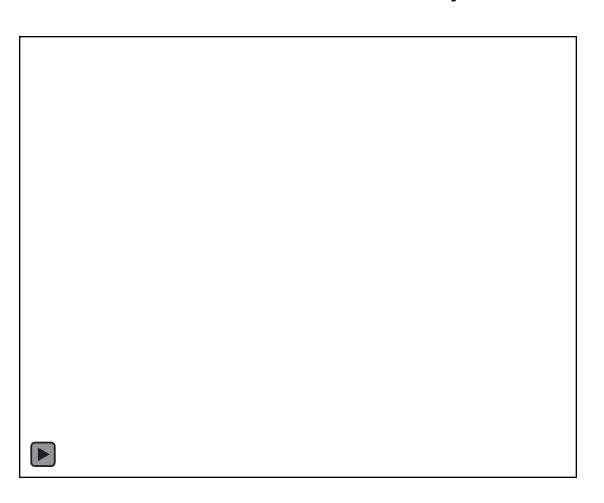
Download the series you want from the appropriate online tool: UKDS.Stat or IMF eLibrary.



#### Explore our data on Sustainable Development Goals No poverty: Poverty headcount Infant mortality rate Percentage living on less than \$1.25 a day Zero hunger: Prevalence of undernourishment Depth of food deficit Good health and wellbeing: Maternal mortality Rrevalence of HIV Measles immunisations Quality education: Youth literacy rates Gender equality: Adolescent fertility rate Realization Parliamentary seats held by women Clean water and sanitation: Access to improved sanitation facilities Industry, Innovation and Infrastructure: Mobile cellular subscriptions R&D expenditure and technology exports Life on land: ᇽ Land use in African countries 🙀 Protected terrestrial and marine areas Mammal and bird species, threatened Peace, justice and strong institutions: Bribery incidence % Intentional homicide

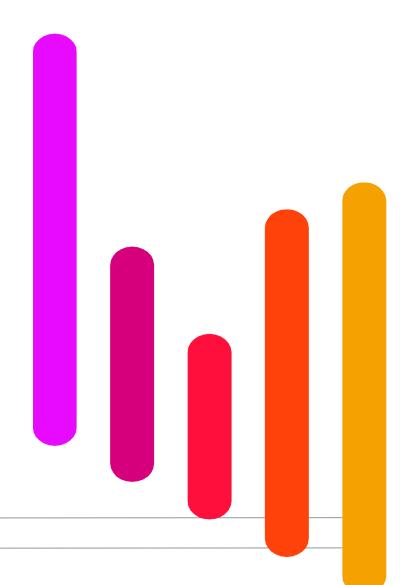


## International Macrodata Outputs





## Census Data



UK Data Service

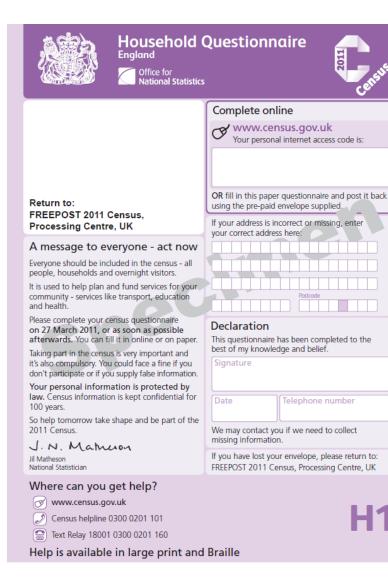


## UK census data

- 1971-2011 census data
- Gold standard
- Used as baseline for other statistics
- Detailed combinations of characteristics
- Small geographies
- Census outputs
  - Aggregate data
  - Boundary data
  - Flow data
  - Microdata



## **Topics**



- age
- sex
- health
- religion
- ethnicity
- caring
- travel to work
- household composition
- housing
- car ownership
- employment
- education
- national identity
- second homes
- intention to stay



# Census aggregate data

- Counts of people, households, etc. with particular combinations of characteristics for an area
- Areas vary from large to very small

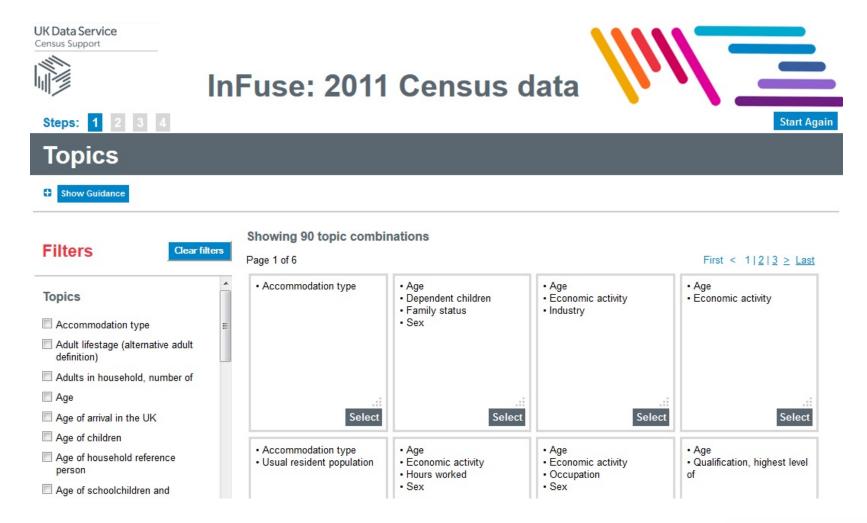


• Example:

Numbers of people who are aged over 50 and are unemployed (a combination of characteristics) within each of the wards in the district of Greater Manchester (a set of geographical zones).



## Census Tools - InFuse





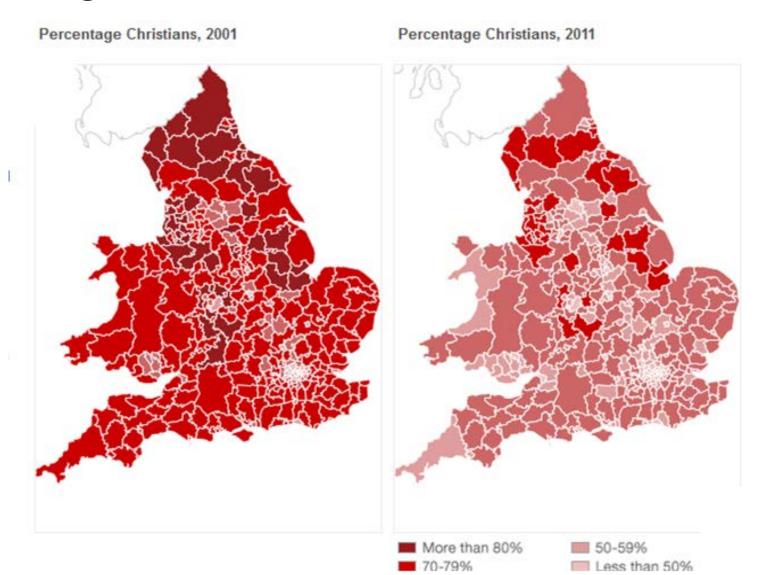
# Using InFuse

### Example data for 2011. This table was generated using InFuse.

CDU_ID	GEO_CODE	GEO_LABEL	GEO_TYPE	GEO_TYP2	F997	F998	F999	F1000	F1001
					Household is	Household is	Household is	Household is	Household is
					not deprived in	deprived in 1	deprived in 2	deprived in 3	deprived in 4
					any dimension	dimension	dimensions	dimensions	dimensions
8	E12000001	North East	Regions	RGN	448560	356794	248597	71195	4789
9	E12000002	North West	Regions	RGN	1218768	953693	626702	192951	17435
		Yorkshire and The							
10	E12000003	Humber	Regions	RGN	909078	720188	457389	125936	11468
11	E12000004	East Midlands	Regions	RGN	811134	614226	372252	90406	7586
12	E12000005	West Midlands	Regions	RGN	916431	745093	489986	131364	12035
13	E12000006	East of England	Regions	RGN	1084585	799157	434169	95983	9141
14	E12000007	London	Regions	RGN	1287458	1117962	626841	204345	29567
15	E12000008	South East	Regions	RGN	1695912	1145825	569744	129939	14043
16	E12000009	South West	Regions	RGN	1013722	751243	398302	91503	9871



# Religion



Census Boundary Data



# Census Tools – Boundary Data

#### ACCESS BOUNDARY DATA

These tools are restricted to staff and students from UK further/higher education institutions.



#### EasyDownload

The most regularly requested census boundaries available as ready-to-use national datasets in popular formats



#### Boundary Data Selector

This facility lets you select the boundaries you want, for the area you want, in the format you want



#### Postcode Directory Download

This facility allows you to download complete versions of current and historical postcode directories (sometimes referred to as look-up tables)



#### Postcode Data Selector

This facility allows you to download the set of postcodes that you want from postcode directories released between 2001 and the present day



#### GeoConvert

A geography matching and conversion tool. Obtain and manipulate complex geographical and postcode data in a straightforward way



#### Thematic Mapper

Create Choropleth thematic maps from your own uploaded CSV file and boundary data from Census Support



#### Metadata Viewer

This facility allows you to search for and view metadata records associated with the census geographic datasets



#### WICID (Web-based Interface to Census Interaction Data)

This facility allows users to select and download migration and journey-to-work flow data collected by the Census of Population.



#### Casweb

This facility contains boundary data bundled with census aggregate data for the

http://census.ukdataservice.ac.uk/get-data/boundary-data.aspx

http://census.ukdataservice.ac.uk/us e-data/quides/boundary-data.aspx



### **Census Flow Data**

- Special type of aggregate data, where there is an origin and destination
- Two main types
  - commuting
  - migration



http://census.ukdataservice.ac.uk/use-data/guides/flow-data.aspx

UK Data Service

## **WICID**

Origin geography: UK Government Office Regions (1999-) Destination geography: UK Government Offi

Layer: Total migrants [Table MG101]

	Destinations										
Origins	North East	North West	Yorkshire and The Humber	East Midlands	West Midlands	East of England	London				
North East	222328	5380	8203	3238	2261	2858	4279				
North West	5225	585973	17830	9527	10645	6490	11596				
Yorkshire and The Humber	7977	14930	460361	15232	6177	6337	9582				
East Midlands	2634	8742	14993	353045	12530	11273	9943				
West Midlands	2187	10709	7480	14682	421606	6648	10592				
East of England	2764	6649	8549	17004	7812	423178	28470				
London	3370	9678	8267	11015	9513	53477	641404				
South East	3919	11556	11125	18175	14081	25314	51487				
South West	2128	6691	5674	6974	10733	8577	15404				



### Census Microdata

- Individual-level data which has been anonymised
- Sample of Anonymised Records (SARs)



- Looks like data you would get from doing a survey yourself
- Analysed using an analysis package like SPSS or Stata
- More flexible as you can produce your own tables, populations and attributes
- 1991, 2001 and 2011 SARs data available



## **Qualitative Data**



UK Data Service



## **Qualitative Data**

### Qualitative data is non-numeric information for example:

- interview transcripts
- visual data
- focus groups
- Essays
- Diaries
- observation notes
- Documents
- audio data
- open-ended survey questions
- case notes
- newspaper clippings, etc.



**Mixed methods** approaches combine qualitative data with quantitative (numeric) data



# Examples of qualitative data collections

- Family Life and Work Experience before 1918, Middle and Upper Class Families in the Early 20th Century, 1870-1977 (SN 5404)
- Gender Difference, Anxiety and the Fear of Crime, 1995 (SN 4581)
- Mothers Alone: Poverty and the Fatherless Family, 1955-1966 (SN 5072)
- Affluent Worker in the Class Structure, 1961-1962 (SN 6512)
- Qualitative Election Study of Britain, 2015 (SN 6861)
- Retail Competition and Consumer Choice, 2002-2004 (<u>SN</u> 5049)
- Religion, martyrdom and global uncertainties, 2013-2015 (<u>SN</u> 852553 & <u>SN 852554</u>)



# Registering with the UK Data Service



**UK Data Service** 



# Registering with the UK Data Service

- You will need to be registered with the UK Data Service
- If you are at a UK university, you can use your own institutional username and password to log in and register with the UK Data Service
- If you are not at a UK University you will have to apply for a UK Data Archive username and password
- https://www.ukdataservice.ac.uk/get-data/how-toaccess/registration





# Useful resources and Further help



**UK Data Service** 



# Useful tips - Advice for new users

- We have webpages for new users with information on:
  - How to find data with our search application – Discover
  - How to register and access data
  - What kinds of data we hold
  - How you can get in touch?

### Advice for new users

"Learn about our service and how to use our resources"



The UK Data Service provides access to, and support for, the use of an extensive range of key economic and social data, spanning many disciplines and themes. Find out <u>about us</u>.

Our search and browse application, <u>Discover</u>, provides access to data and related resources, such as case studies, support guides and publications. Discover contains over 6,000 digital data collections for research and teaching purposes.

#### Using our data

Anyone can register for our service but access to our resources may differ depending on your affiliation.

Academic	Public Sector	Voluntary	Commercial	Personal
		Sector		



### Interactive Events

- Webinars
  - Introduction to the UKDS
  - Introduction to key types of data
  - Specialised webinars for topic area, methods
  - E.g. An Introduction to survey data on health
  - Can be collaborative

#### **UK Data Service**

### Introductory Webinars 2018



To help you get the most from the UK Data Service we run a series of webinars which introduce different aspects of the Service and explain our key datasets. The webinars run from 15.00-16.00 and they are free to attend.

You can find other webinars on more specialised topics on our events pages.

To book your place click on the dates below or go to <u>ukdataservice.ac.uk/news-and-events/events</u>.

#### Introduction to the UK Data Service

The UK Data Service holds the UK's largest collection of research data. This introductory webinar is for anyone with an interest in social research who wants to hear about the vast array of resources we offer. Participants will learn about the different types of quantitative and qualitative data we hold, how to find those data collections using our online resource discovery tools, and how to register to access datasets in a number of different formats. We will also demonstrate how some of our data collections can be viewed and analysed online, as well as providing an overview of our support and training resources for students, researchers and teachers.

26 April 22 November

### Finding and accessing data in the UK Data Service

This introductory webinar is intended for anyone who wants to hear about finding data from the UK Data Service. Participants will be given a practical overview of the tools that the service 29 November provides, particularly our Discover tool to help you to make the most of your data searches. We will discuss best approaches to searching data and how to use the search tool to narrow down search results. We will also explore other service resources provided to help you get the data you want and related information.

3 May

#### Key issues in reusing data

Participants will hear about the key issues in secondary analysis as a method. The introductory session will briefly cover the pros and cons of reusing data and the importance of learning about the origins of your data. Quantitative and qualitative secondary analysis will be discussed with examples and issues of context, sampling and ethics will be raised.

22 May 4 December



# Want to learn more about finding data and resources?

Then join us next week for that next webinar in this series -

Finding and Accessing Data in the UK Data Service 29<sup>th</sup> November 2018 3-4pm

https://www.ukdataservice.ac.uk/news-and-events/eventsitem/?id=5220



# Resources for using data

**UK Data Service** 



Advice and training

Guides

Video tutorials

Data in use

Citing data

Secure Lab

Teaching with data

Student resources

Data Skills Modules

Secondary analysis



Home > Use data

#### Use data

"Information, training and support for our users"



SHARE <

Behind everything we do is the desire to see data used to its maximum potential. To help you to make best use of our collection we offer resources and support for users.

This section contains information on:

- · advice and support, including information about our helpdesk and training
- · a number of guides, video tutorials and pointers to other useful information
- · examples of how the data have been used by others
- · information and resources designed for teachers and learners



# Surveys

Unit 1: What are survey data?

This unit will explore survey data.

You will look at what the data are, some key concepts relating to survey data and how to find datasets that you can use.

You can work through this unit in order using the next button, or use the menu to view the list of sections and jump to the part you are most interested in.





#### Demo: Opening a data set

This video illustrates how to open a data set in PSPP.

SPoint	SHARD	PopBand	PORT	Weactor	CHUNG	PORTO	ABLVIII	Country	Ksex	Kage	Househil	PZSes	PZAge	PZKels	PZRetZ	PZKet	PfSet
101	101.	2	12	1.02	1.65	-11	2	2	1	31	- 4	1	35		- 1	- 1	1
101	191	- 1	12	.52	.55	11	į.	1	- Pr	47		0	-1		8		0
101	101	2	12	.40	.55	11	2	- 2	1	Fem	ale 1	- 3	-4			- 0	0
101	101	2	U	.40	.55	п	3	1	- 2	57	- 1	0	:-1		. 1	. 0	0
101	101	2	12	.54	1.10	11	i	2	1	12	2	2	13	- 1	- 1	1	
101	101	2	12	.99	1.10	11	2	- 2	2		2	i	- 44	- 2	- 2	- 2	0
101	101.	2	12	.86	1.10	н	, i	2	2	62	2	J.	65	- 1	.1	1	0
101	101	- 2	12	LIJ	1.65	11	2	2	1	76	1	ž	72	1	- 1	1	Ł.
101	101	2	12	1.37	1.65	11	3	- 2	2	52	5	2	22	- 2	- 2	- 1	1
101	101	5	12	.55	.55	111	j,	2	2	. 16	1	0	:4	.0	- 1	. 0	- 6
101	191	2	12	1.61	2.26	11	- 2	2	- 1	52	. 4	1	51	i	1	- 1	i,
191	101	2	u	1.26	1.10	11	. 3	2	1	46	. 2	2	:29	.1	1	1	6
101	191	- 2	12	.46	55	- 11	1	1		42	- 2	, i	- 1	2	2	- 2	0
191	101.	. 2	12	-61	55	11	- 1		4	45	- 1	9	-1		0		0
101	101	2	12	.72	1.10	11	2	2	- 1	56	- 2	2	21	- 2	- 2	- 2	0
101	191	2	12	1.09	1.10	11	3	2	1	36	- 4	2	35	- 1	1	- 1	1
101	191	2	12	.47	55	11	1	- 2		51	- 1		-1	- 1		- 9	0
102	101	2	:12	1.15	1.10	- 11	1	2	- 1	31	- 4	2	34	1	-1	1	1
102	191	2	12	.40	.55	.11	2	2	1	74	- 1	9	-1				0
102	191	2	12	.76	1.10	11		2	3	58	2		50	- 1	- 1		0
102	101	ž	12	.87	1.10	н	)	1	7	20	5	i	40	- 1	- 3	1	i.
102	101.	2	12	.50	1.10	11	2	2	1	72	2	2	69	- 1	- 1	- 1	
102	.101	2	12	.40	55	11	- 2	2		72	i.		- 4		- 1		0
102	101	2	12	.40	55	.11	3	2	- 1	14	1	9	-4	. 4	- 1		6
102	191	2	12	.90	1.10	11	3	2	1	- 66	2	2	67	1	- 1	1	0
103	192	3	12	1.02	.55	- 11		2	1	25	- 1	9	-1	- 4	. 1	. 0	0.
193	102	3	12	.56	55	11	1	1	1	71	i	0	-4				0



# Always cite your data!

- Acknowledges the author's sources
- Makes identifying data easier
- Promotes the replication of research results
- Allows the impact of data to be tracked
- Provides a structure which recognises and can reward data creators





# User Support Helpdesk

- Data-related queries
  - Finding suitable data
  - Identifying specific variables
  - Understanding the data coding frames, etc.
  - Data problems/anomalies



### Questions

ukdataservice.ac.uk/help/

Follow us at:
ukdataservice@jiscmail.ac.uk
twitter.com/ukdataservice
facebook.com/ukdataservice

