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

Workshop 2: Data skills - what are survey data?

Patty Doran
UK Data Service, University of Manchester
patty.doran@manchester.ac.uk
@Patty_NZ

Date: 2pm, Thursday July 9th, 2020

UK Data Service

What we are covering:

- Introduction
- Presentation – social surveys and secondary analysis
- Quiz
- Demonstration – exploring with Nesstar
- 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.

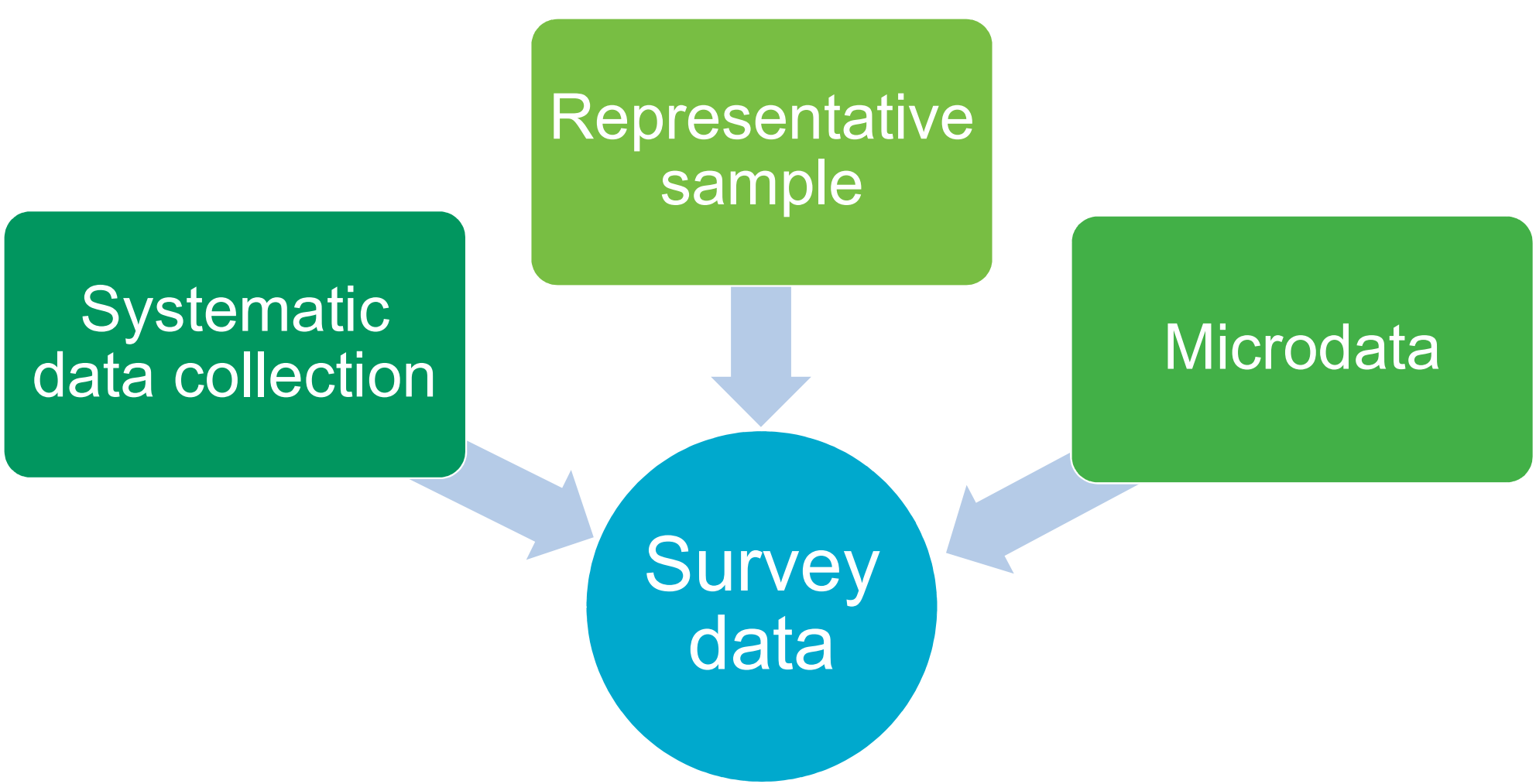
Social Surveys and the Third Sector

- 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

Social Surveys


Representative sample
Systematic data collection
Microdata
Survey data

Where do survey data come from?

1.

Overall, how satisfied are you with your mental well-being?

Where 0 is 'not at all satisfied' and 10 is 'completely satisfied'

(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
○	○	○	○	○	○	○	○	○	○	○

2. Sex of Respondent

○ 1 Male

○ 2 Female

1 Responses to survey

1.

Overall, how satisfied are you with your mental well-being?

Where 0 is 'not at all satisfied' and 10 is 'completely satisfied'

(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
○	○	○	○	○	●	●	○	●	●	●

2. Sex of Respondent

- ● 1 Male
- ● 2 Female

2 Creates individual level data →

Case	Wellbeing	Sex
1	8	2
2	10	1
3	8	1
4	6	2
5	9	2
6	5	2
7	9	2

Surveys

Example:

British Social Attitudes Survey

BRITISH SOCIAL ATTITUDES

NatCen Social Research

British Social Attitudes Survey 2018

Economic and Social Research Council, Government departments, research foundations etc.

Randomly selected households in Britain.
One adult per household,
3879 interviews

Personal interviews
Follow-up self-completion form

http://doi.org/10.5255/UKDA-SN-8606-1

Copy study DOI

British Social Attitudes Survey, 2018

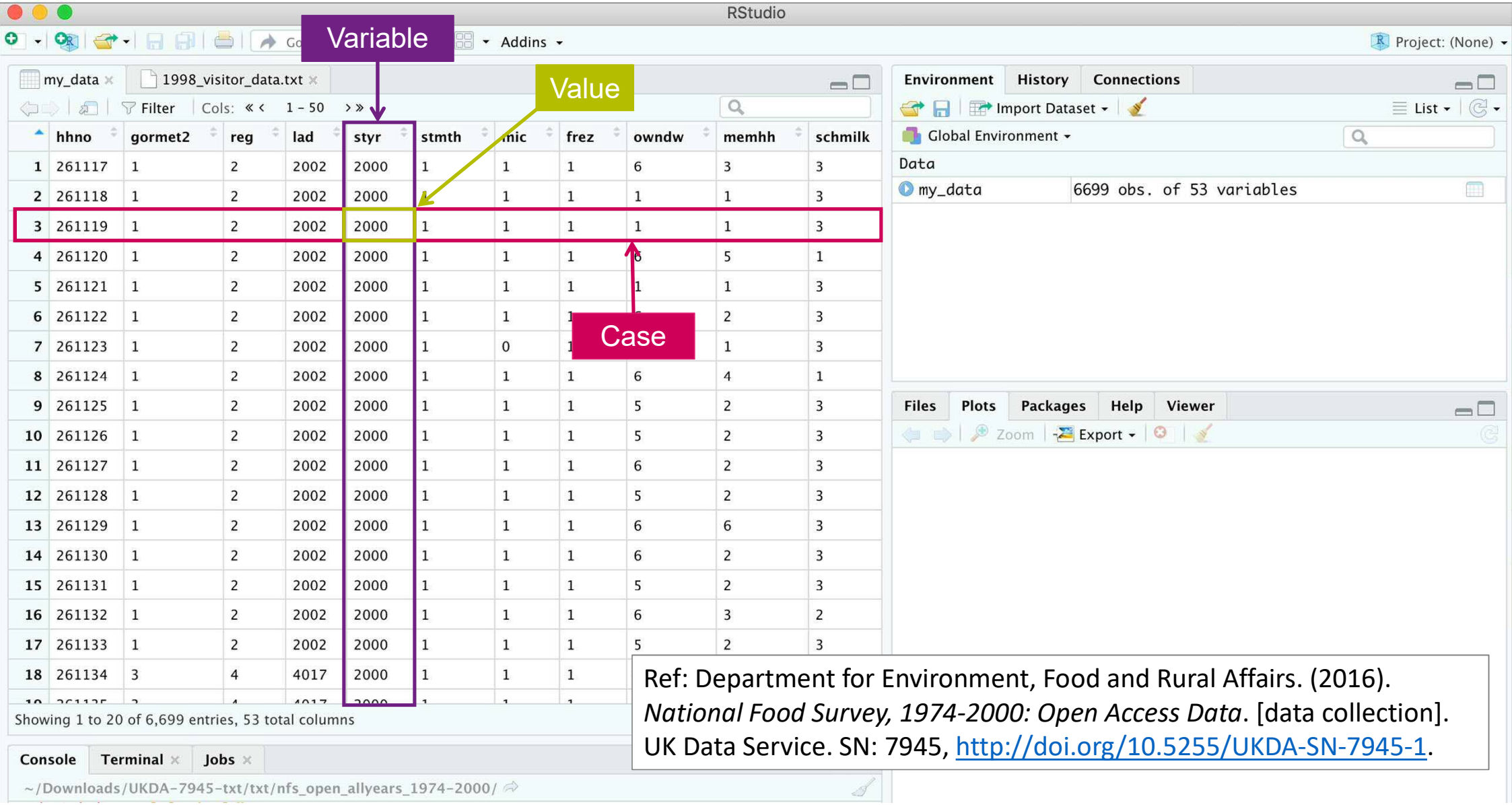
Details | Documentation | Resources | Access data

Details

Title:	British Social Attitudes Survey, 2018
Alternative title:	BSA
Study number (SN):	8606
Access:	These data are safeguarded
Persistent identifier (DOI):	10.5255/UKDA-SN-8606-1
Series:	British Social Attitudes Survey
Principal investigator(s):	NatCen Social Research

Surveys

What do they look like?

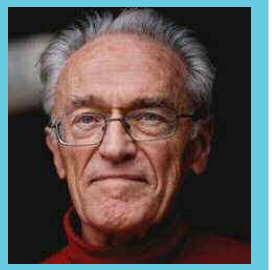


	hhno	gormet2	reg	lad	styr	stmth	mic	frez	owndw	memhh	schmilk
1	261117	1	2	2002	2000	1	1	1	6	3	3
2	261118	1	2	2002	2000	1	1	1	1	1	3
3	261119	1	2	2002	2000	1	1	1	1	1	3
4	261120	1	2	2002	2000	1	1	1	6	5	1
5	261121	1	2	2002	2000	1	1	1	1	1	3
6	261122	1	2	2002	2000	1	1	1		2	3
7	261123	1	2	2002	2000	1	0	1		1	3
8	261124	1	2	2002	2000	1	1	1	6	4	1
9	261125	1	2	2002	2000	1	1	1	5	2	3
10	261126	1	2	2002	2000	1	1	1	5	2	3
11	261127	1	2	2002	2000	1	1	1	6	2	3
12	261128	1	2	2002	2000	1	1	1	5	2	3
13	261129	1	2	2002	2000	1	1	1	6	6	3
14	261130	1	2	2002	2000	1	1	1	6	2	3
15	261131	1	2	2002	2000	1	1	1	5	2	3
16	261132	1	2	2002	2000	1	1	1	6	3	2
17	261133	1	2	2002	2000	1	1	1	5	2	3
18	261134	3	4	4017	2000	1	1	1			

Ref: Department for Environment, Food and Rural Affairs. (2016). *National Food Survey, 1974-2000: Open Access Data*. [data collection]. UK Data Service. SN: 7945, http://doi.org/10.5255/UKDA-SN-7945-1.

Secondary Analysis

Reusing data

Data collector		Secondary Analyst
Office for National Statistics; Understanding Society; NatCen Social Research; 	Data Sharing →	
Primary research		Secondary research
Data collection & analysis for planned purpose		Re-analysis (can be for a different purpose)

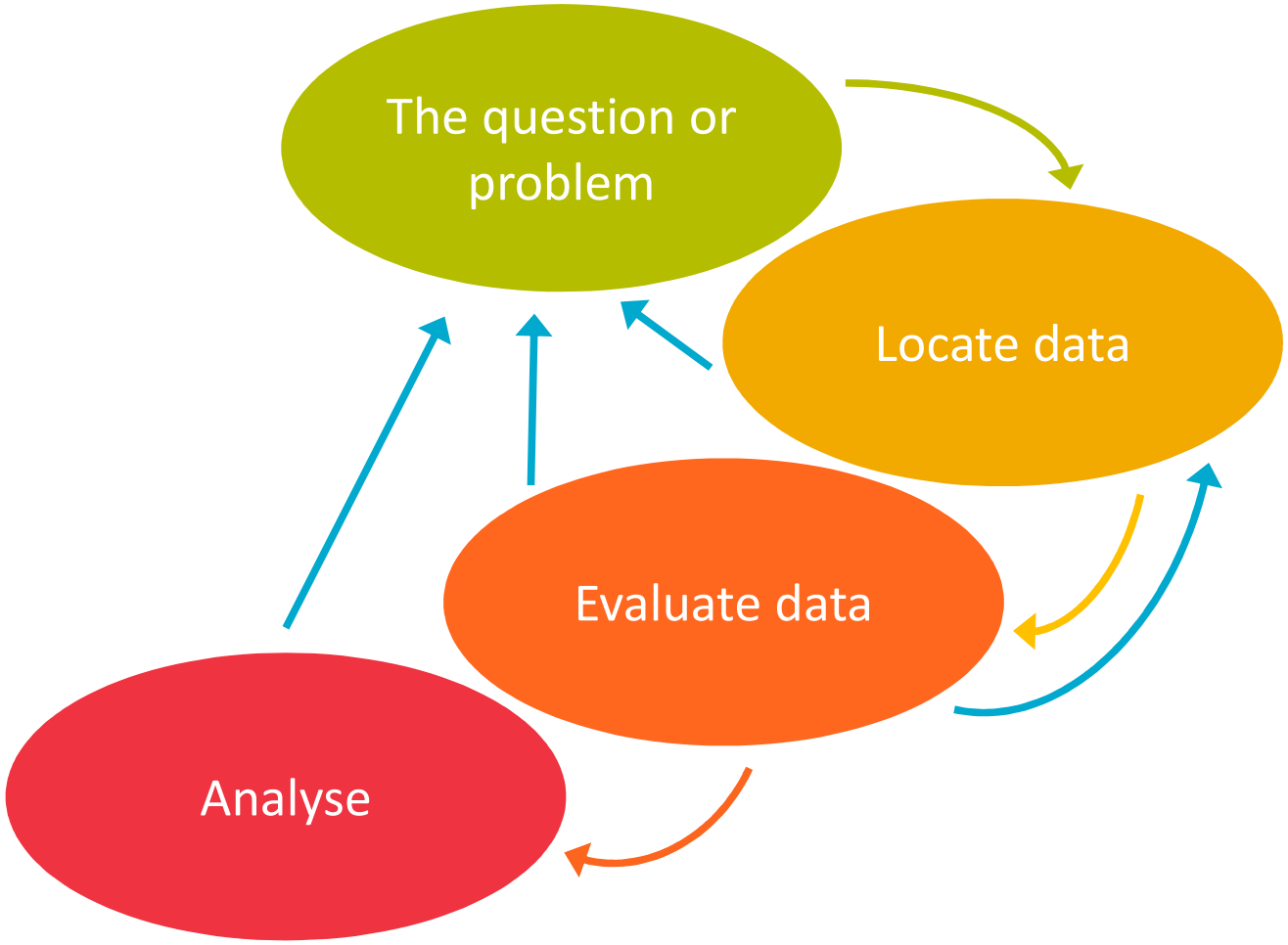
Pros and cons of reusing data

Pros	Cons
datasets impossible to create	do not have insider understanding of data and data collection
cost effective	effort to get to know the data
ethical issues about data collection dealt with	ethical issues about data use still apply (limit access to sensitive data)
do not need to re-contact data subjects	data may not match research question
reuse data used by others to make claims	cannot extend studies

Researchers reusing data need to:

1. Make effort to understand the data
2. Be pragmatic about whether data are good enough for your purpose

Research plan



Making sense of your data

To understand data and it's suitability for your research you must understand…

- what information was collected
- from whom
- when and where
- any changes to the raw data before being archived

Use documentation

- user guides
- questionnaires
- interview schedules

Simple Statistics

Types of variables

1. Interval variables are measured on a continuous scale, with each step along the scale being equal.
2. Categorical variables are divided into discreet categories. There are two types of categorical variables:
    - nominal variables (no natural order)
    - ordinal variables (naturally ordered)

Question 1	Question 2
What type of school does your oldest child attend? 1.Academy 2.City Technology College 3.Free School 4.Private / Independent School 5.Other (including home schooling)	How satisfied are you with the standard of teaching at the school attended by your oldest child? Completely satisfied 2.Mainly satisfied 3.Neither satisfied nor dissatisfied 4.Mainly dissatisfied 5.Completely dissatisfied
Question 3 How much did your family pay in school fees in previous year? _____ Record in whole GBP (£) (if unknown, give best estimate)	

Descriptive statistics

Continuous variables include:

- measures of central tendency (mean, mode, median)
- measures of variation (standard deviation)
- histograms (to visualise the distribution)

Categorical variables include:

- counts
- percentages
- frequency tables

Two-way tables

	What time do you get out of bed?		
Age	Before 9am	After 9am	
Under 18	20 (29%)	40 (80%)	60 (50%)
Over 18	50 (71%)	10 (20%)	60 (50%)
	70 (100%)	50 (100%)	120 (100%)

Quiz

Nesstar Demonstration

Activity

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

Workshop 2. Data skills: What are survey data?

Activity

Task:

1) **Find data to answer a question using Nesstar.**
2) **Create a two-way table to answer a question.**

About Nesstar

Nesstar is an online tool provided by the UK Data Service for exploring and accessing to a wide variety of datasets.

Website: http://nesstar.ukdataservice.ac.uk/webview/

Activity: Creating descriptive statistics

You are interested in the risk of adverse effects of COVID-19 and how this relates to income inequality:

> *Are people with low income more likely than people with higher income of being in the high risk group for COVID-19?*

To explore the question, you will look at data from the Health Survey for England. The key variables to explore are:

1. Having a long-term health condition

Tab
Q
W
E
R
T
Y
U
Caps Lock
A
S
D
F
G
H
Shift
Z
X
C
V
B
N
Ctrl
Alt

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

First, did you complete the activity?

1. How many respondents from the Health Survey for England reported having a physical or mental health condition or illness lasting or expected to last 12 months or more?
2. How many people reported that the condition affected stamina, breathing or fatigue?
3. Of those respondents with conditions affecting stamina, breathing or fatigue, what percentage are in the lowest income quintile?
4. Compared to all long-term health conditions, are conditions affecting stamina, breathing or fatigue more prevalent or less prevalent amongst people in the lowest income quintile?

Dataset: Health Survey for England, 2016

Health Survey for England, 1991-

Variable ILL12m: Whether have any physical or mental health conditions or illnesses lasting or expected to last 12 months or more

LITERAL QUESTION

Do you have any physical or mental health conditions or illnesses lasting or expected to last 12 months or more?

Values	Categories	N	
1	Yes	3914	38.9%
2	No	6151	61.1%
-9	Refused	1	
-8	Don't know	1	
-1	Not applicable	0	

SUMMARY STATISTICS

Valid cases	10065
Missing cases	2
Minimum	1.0
Maximum	2.0
This variable is numeric	

Dataset: Health Survey for England, 2016

Health Survey for England, 1991-

(D) Equivalised Income Quintiles: Categories | Whether condition…athing or fatigue: Categories

Type: Column percentage

Whether conditions or illnesses affect: Stamina, breathing or fatigue	Not mentioned	Mentioned	Total
(D) Equivalised Income Quintiles			
Lowest Quintile (<=£14,733)	22.0	30.7	24.9
Second lowest Quintile (>£14,733<=£21,985)	19.8	24.5	21.3
Middle Quintile (>£21,985 <=£31,798)	21.5	21.0	21.4
Second highest Quintile (>£31,798 <=£49,400)	19.9	13.6	17.9
Highest Quintile (>£49,400)	16.7	10.3	14.6
Total	100.0	100.0	100.0
N=	2,084	1,014	3,098

Dataset: Health Survey for England, 2016

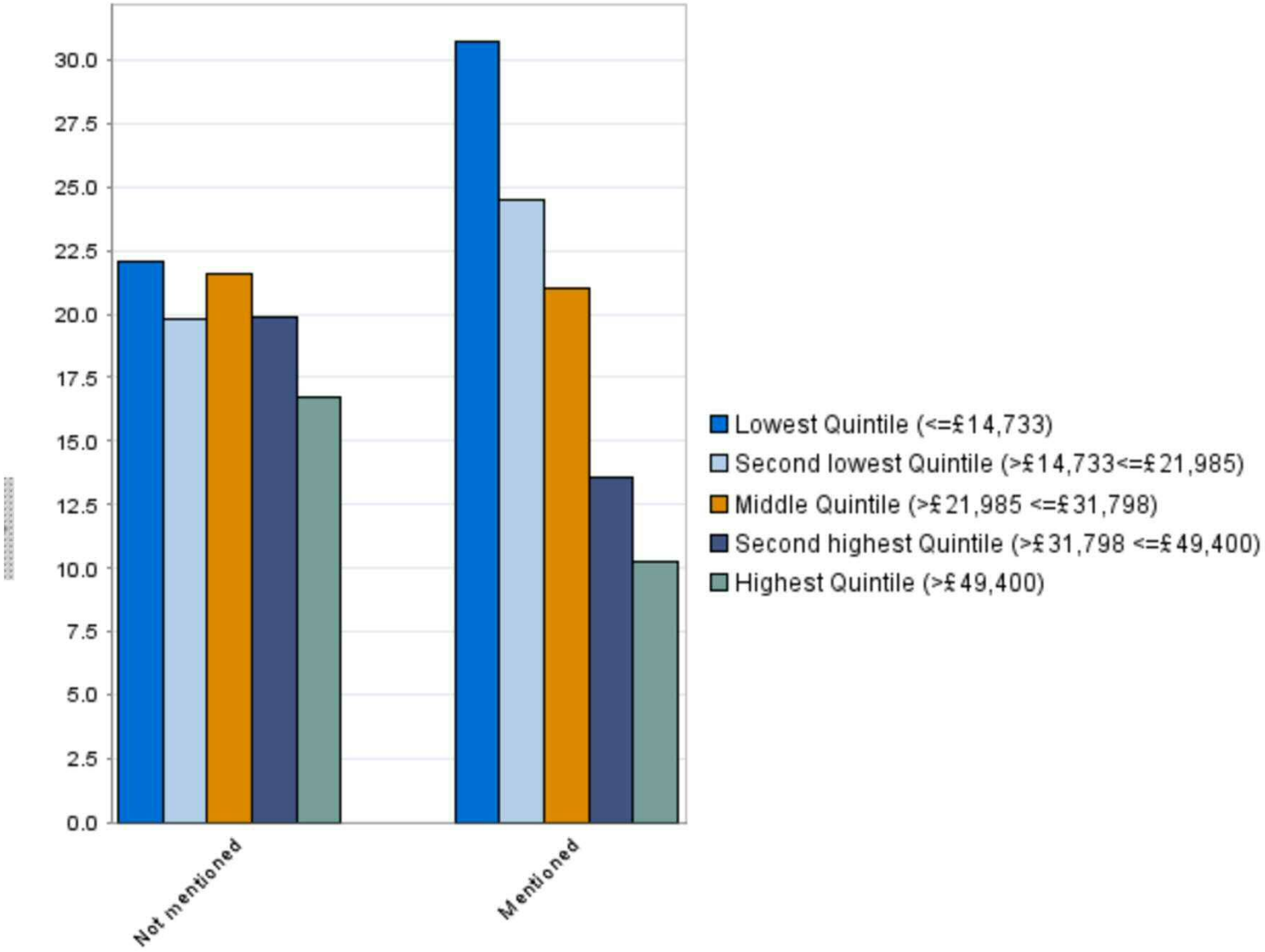
Health Survey for England, 1991-

Whether condition...athing or fatigue: Categories | (D) Equivalised Income Quintiles: Categories

Type: Column percentage

(D) Equivalised Income Quintiles Whether conditions or illnesses affect: Stamina, breathing or fatigue	Lowest Quintile (<=£14,733)	Second lowest Quintile (>£14,733<= £21,985)	Middle Quintile (>£21,985 <=£31,798)	Second highest Quintile (>£31,798 <=£49,400)	Highest Quintile (>£49,400)	Total
Not mentioned	59.6	62.5	67.8	75.0	77.0	67.3
Mentioned	40.4	37.5	32.2	25.0	23.0	32.7
Total	100.0	100.0	100.0	100.0	100.0	100.0
N=	770	661	662	553	452	3,098

Crown copyright material is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland



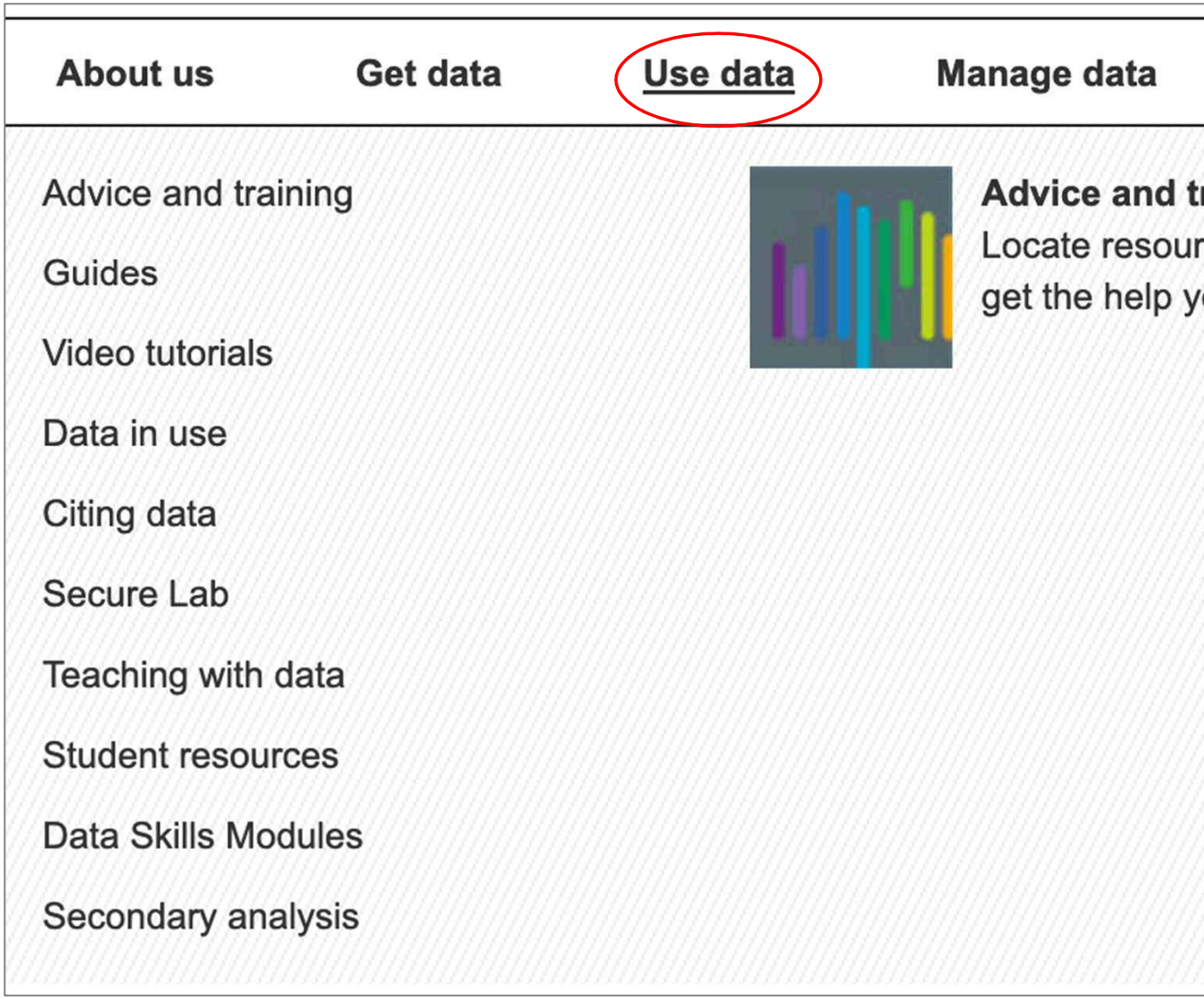
/hether conditions or illnesses affect: Stamina, breathing or fatigue

Discussion

Resources

Online resources

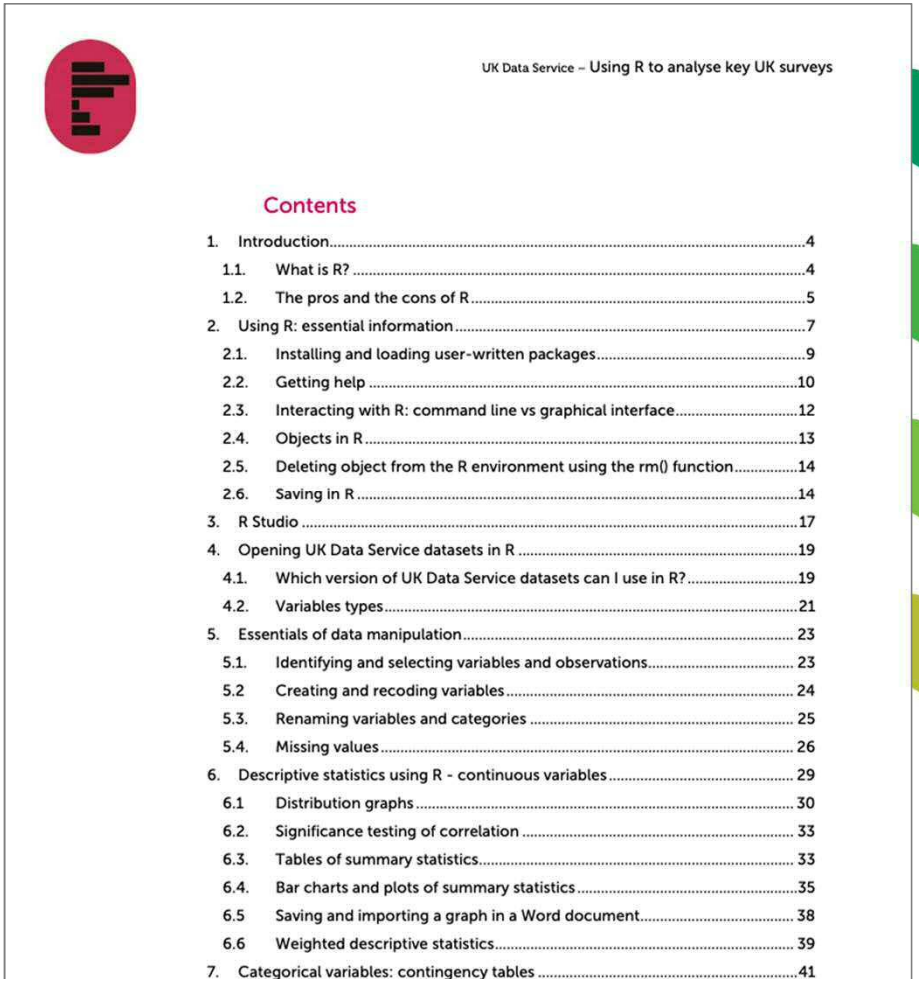
- Guides
- Videos
- Webinars
- Data Skills Modules



Methods and Software Guides

For example:

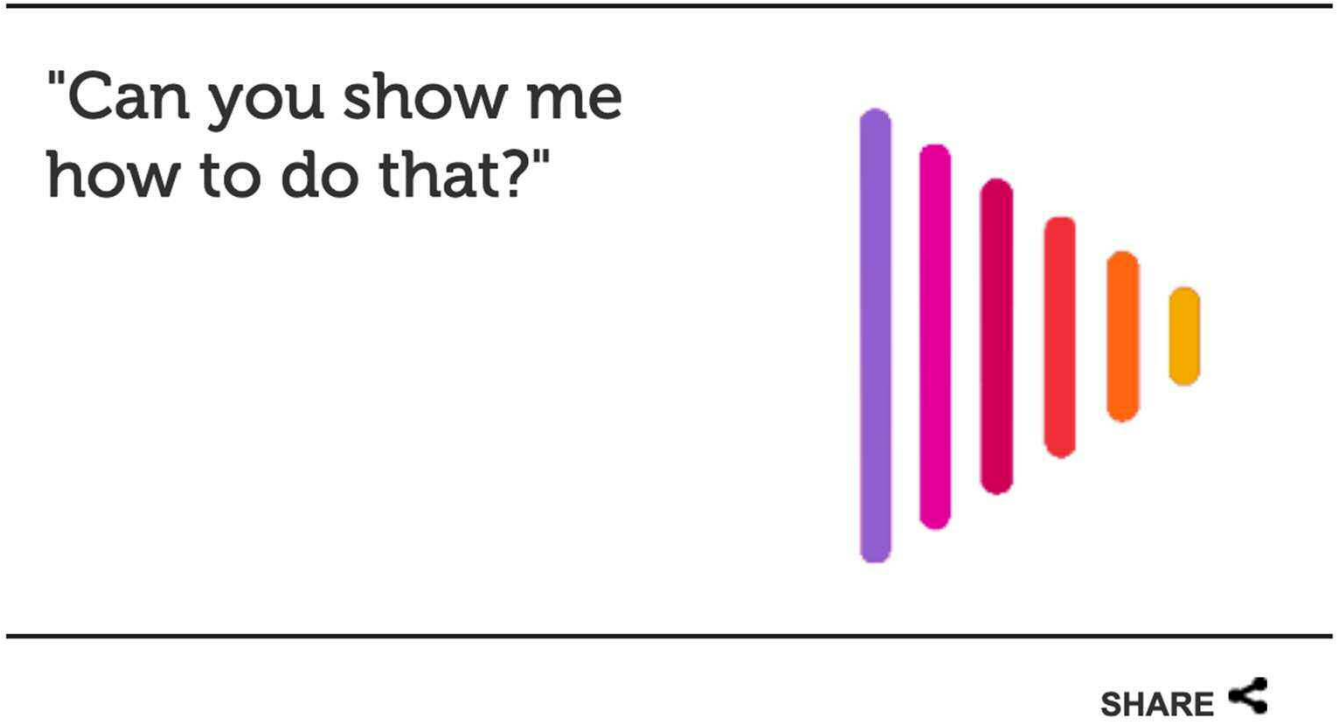
- Software guides
  - What is Stata?
  - What is SPSS 20 for Windows?
  - Using R to analyse key UK surveys
- Survey data
  - What is weighting?
  - What is complex sample design?
  - What are hierarchical files?



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

https://ukdataservice.ac.uk/news-and-events/events

Showing 1 to 10 of 18 entries 1 2

Online workshops: Social data and the third sector: evidence, skills and tools
02 - 16 Jul 2020
Online

Family finance surveys user conference 2020
08 Jul 2020
Online, 09.45 - 15.15

Webinar: Energy data for research
13 Jul 2020
Online, 15.00 - 16.00 BST

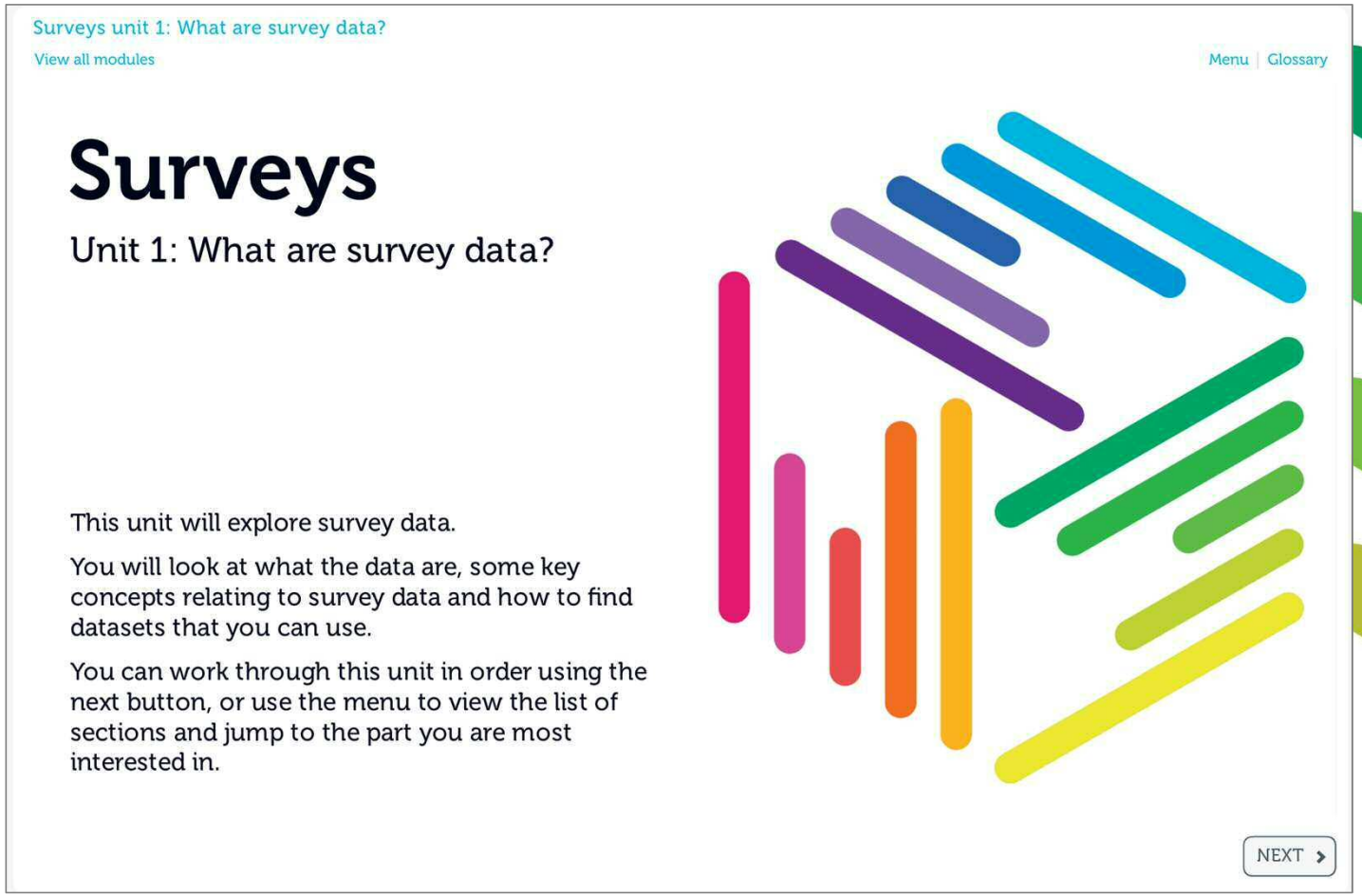
Introduction to Big Data for the Social Sciences - Presented by UK Data Service
20 - 24 Jul 2020
Online

Webinar: Key data: UK and cross-national surveys
08 Oct 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



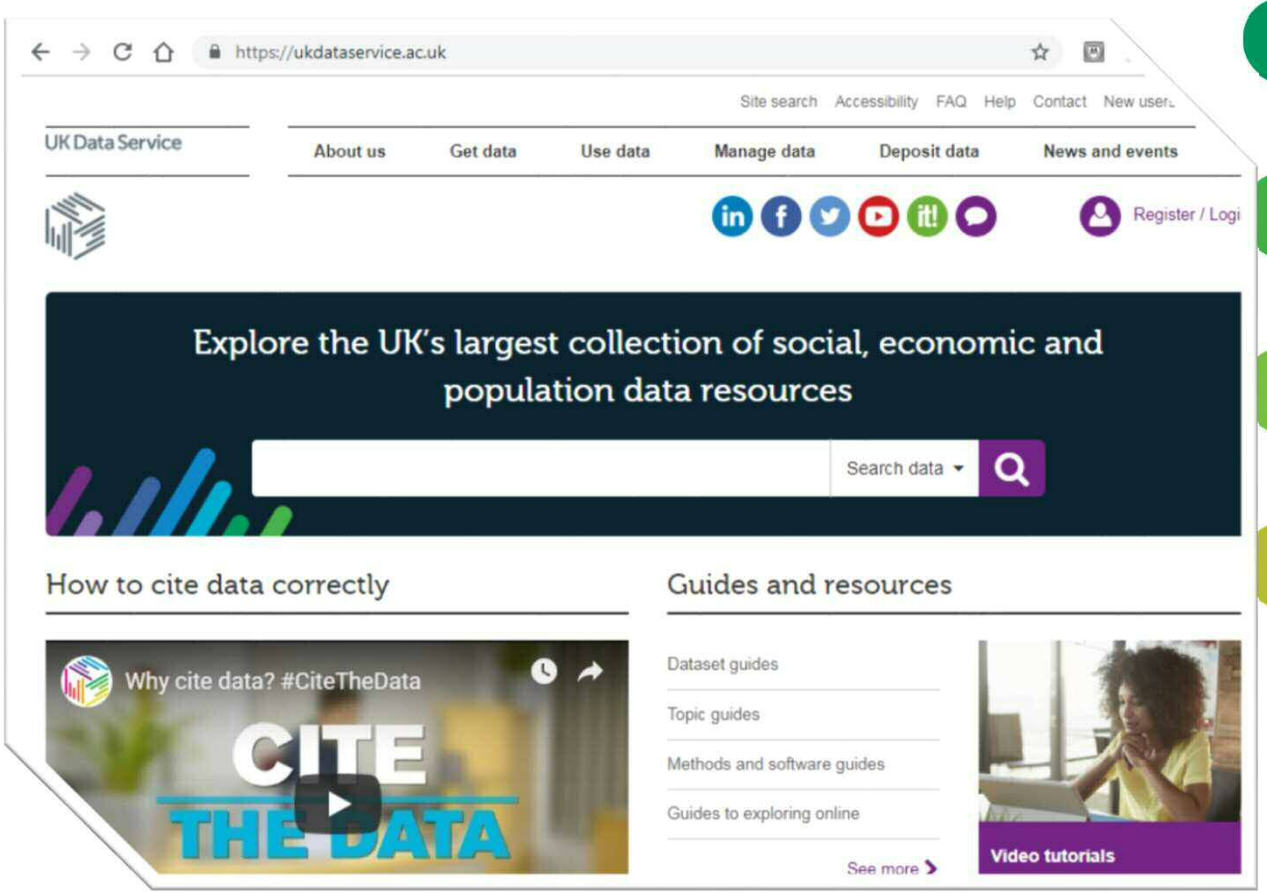
Thank you

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