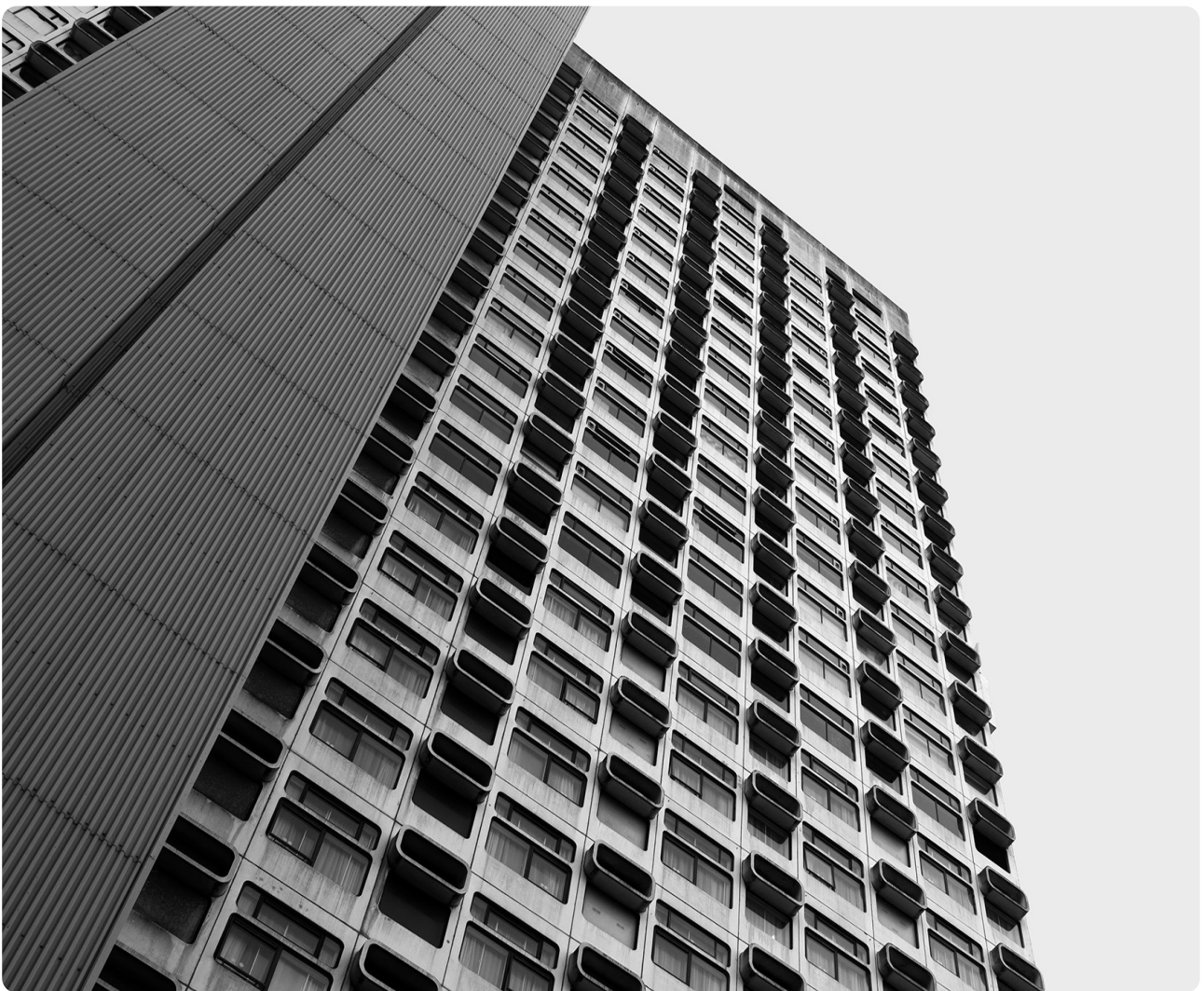


---

# Teaching with data

Findings from the UK Data Service Higher Education  
Teaching Consultation 2021

Jennifer Buckley, Alle Bloom and Nigel De Noronha



---

Summary and recommendations for the UK Data Service .....	3
Introduction.....	10
Methods .....	11
Data for teaching .....	15
Teaching Resources .....	27
Software.....	36
Dissertation students .....	42
Computational social science and reproducibility.....	47

## Acknowledgments

The authors would like to express their gratitude to all the participants in the consultation. We would also like to acknowledge the input of Vanessa Higgins, Deborah Price and Patty Doran.

## Summary and recommendations for the UK Data Service

The UK Data Service ran a consultation with higher education (HE) teaching staff in the Summer of 2021. The aim was to gather information to help the UK Data Service understand how to support the use of quantitative data from the UK Data Service within teaching and learning in HE.

### Methods

The consultation consisted of a survey of teaching staff and interviews.

Through non-probability sampling, the survey received 109 valid responses. Participants came from varied disciplines and career stages, but most were lecturers or above and sociology and politics were the most common disciplines. Nearly half the sample indicate they have substantial experience of quantitative data and methods and the teaching of quantitative methods.

Three interviewers carried out 16 interviews online via Zoom. The interviews combined open questions and participant evaluation of existing teaching resources produced by the UK Data Service.

### What data is using in teaching and why?

We find that a wide variety of quantitative data is being used in social science teaching in HE including data from the UK Data Service and other sources. Just over 60 percent of those surveyed report using data accessed from the UK Data Service, with 50 preferred datasets named. The most mentioned datasets are the British Social Attitudes survey, Understanding Society, Crime Survey for England and Wales and the Census.

Factors that are commonly influencing the choice of data for teaching include:

- A preference for real world data to answer real research questions (including some interest in replication)
- Students' interests
- Size and complexity of a dataset
- Data access and licensing
- Suitability
- Lecturer familiarity

Students' interests can relate to general topical issues or topics covered within their degree. They are often seen as varied, especially since many lecturers teach classes with students from mixed disciplinary backgrounds. The BSA is useful in these circumstances due to the variety of topics. Students prefer working with recent data.

Many lecturers use small and simple datasets, especially for introductory courses. Many also vary the data they use, often to add more complexity or because of the suitability of different data sources to teach specific methods.

Some lecturers do not use data from the UK Data Service due to issues around registration and access.

The limited number of continuous variables that are good for teaching correlation and regression in many datasets is a perceived problem.

Using new data for teaching can be challenging and time consuming, as a result, lecturers tend to find things that work and stick with them.

Some lecturers seek to develop students as data users with some explicit integration of the UK Data Service within teaching.

Most lectures are adapting data for use in teaching. The most common adaptation is to reduce the number of variables to something around 20-30 for teaching or 30-100 for classes where students select variables or use the dataset for assignments.

Based on the discussions, we identify five main principles for good teaching datasets.

1. Should use real data with manageable issues
2. Can be very small and simple (for example up to 20-30 variables) or something larger and more complex (up to 100 variables). Due to individual preferences, tend towards including some additional variables to offer choice.
3. Be intuitive and understandable (through easy-to-understand variables and accessible documentation that provides well-formatted information about variables including level of measurement)
4. Include multiple continuous variables that can be dependent and independent variables, even if they need to be derived using additive scores or factor analysis
5. Be current and connect to student interests (topical or linked to core social science themes)

## Teaching resources

The consultation considered whether additional teaching resources and materials could help support the use of data in teaching.

While there is interest in all types of resources, materials for lecturers such as information about datasets along with ideas and examples seem to be one of the most useful. We also found evidence of potential benefits for resources that lecturers can signpost students to for self-learning. However, lecturers may opt to signpost to only a few resources within a course. Importantly, many lecturers explained that they do not know whether students ultimately use the resources.

Interest in resources to use with students was more limited. Many lectures prefer to create their own resources and can be constrained in the resources they use due to preferred

formats or the use of house-styles. Most think any ready-made teaching materials would need to be customisable.

There is variation in the types of resources used. For example, some use step-by-step worksheets while others prefer video or code-based resources. In some cases, preference about format was linked to software choice.

Various topics for resources were suggested ranging from resources on basic techniques, illustrations of key issues such as Simpson's Paradox and resource relating to published work. Some expressed the view that the UK Data Service should focus mainly on issues relating to data and documentation.

Discussions in the interview also suggest a need for better promotion and labelling of resources.

## Software

Varied software is used in HE teaching but SPSS continues to be used most widely (73 percent use SPSS at least sometimes and 54 percent usually) and many indicate they expect to use SPSS in the future. Stata and R are used less. For instance, more than half never use Stata and R, but they are both used more by those who teach advanced courses.

There are strong advocates for R (and the use of open software) but those using SPSS tend to consider it to be the most student friendly option, with some arguing that programming could be especially off-putting to students already nervous about maths and statistics.

Suggestions in relation to software and resources include developing multiple versions of resources or software neutral resources. There was also interest in R based teaching resources (that could help those new to teaching in R).

## Dissertation students

Many indicate dissertation students could benefit from resources on varied topics including improved versions of the current UK Data Service resources. However, one of the biggest issues for lecturers is perceived barriers to students doing any secondary analysis of quantitative data. These barriers relate to student interest and the skills and time needed to find and work with large datasets. As a result, some of the main requests for resources include resources that either quickly convey the possibility of secondary analysis or help students identify useful data sources.

## Computational social science and reproducibility

Less than half of those surveyed are teaching about reproducibility and transparency of research including good practice in writing and storing code files and tools available to help. However, most agreed these are important topics to teach. A majority indicate that



demonstrations and examples of how to teach these topics would be useful, especially those already teaching these topics.

Only a minority (19/98) indicate they currently teach about computational methods such as using programming to collect, combine, synthesise or process data. A larger number indicate these are important topics to teach and also that resources on these topics would be useful. However, interest in teaching resources is greatest among those teaching these topics already.

Comments in the survey indicate interest in training for lecturers on these topics, with some indicating they are 'behind on this'. However, many perceive the inclusion of programming, as well as advanced statistical topics, may come at the expense of more general statistical literacy and potentially alienate some students.

## Recommendations for the UK Data Service

### Data for teaching

The UK Data Service can recognise that many lecturers are using data from the UK Data Service in teaching. Moreover, many seek to introduce their students to the UK Data Service and raise awareness of the wealth of data that is available. Specific barriers to using data from the UK Data Service in teaching relate to access and topicality. Many report finding the registration requirements too difficult to implement for their classes and opt to use data from other sources. Many are concerned that they cannot rely on students to be registered. The current focus on creating open access teaching datasets may help in some cases and, reassuringly fits with the finding that many are using similarly simple datasets for teaching. The popularity of British Social Attitudes (BSA) data for teaching is also encouraging as we have recently negotiated approval and created two recent open access teaching datasets based on this survey. To address issues, around topicality, we can seek to highlight data relevant to current issues, highlight potential uses of new datasets and keep teaching datasets up to date.

1. Develop a plan for creating teaching datasets in line with the [principles for a good teaching dataset](#) (chapter 4) and include open access data where possible. Specific tasks could be to:
  - a. Plan to create regular (every 2 years) teaching datasets using the BSA. Consider multiple versions, one small open access version and a larger EUL version.
  - b. Keep up to date teaching datasets for other popular surveys (including the National Survey of Sexual Attitudes and Lifestyles (NATSAL) when the new data is available).
  - c. Consider a dataset suited for teaching correlation and regression i.e. one with multiple continuous variables with clear independent and dependent variables.

- d. Ensure teaching datasets are effectively labelled and promoted, so lecturers can easily identify their potential value
  - e. Consider whether and how to evaluate the impact of teaching datasets – for instance, regularly monitor use, look for case studies. Data owners will also be interested in this information as they need to report to their funders.
2. Support lecturers promoting the UK Data Service to their students with appropriate student-centred content such as simple information about how to register (and why they do). Make sure lecturers know the easiest way for students to complete registration.
3. Review information for lecturers about using data in teaching
4. (When staffing resources allow) highlight data for exploring topical issues - for example, ethnic inequalities or political trust. It might be possible to link to specific media discussions, which could be especially good if also using our data.
5. Develop a plan for supporting the use of Census data in teaching. Will there be teaching datasets? Will there be open access microdata? How can we convey the possibilities for teaching with census data?

## Teaching resources

When it comes to resources for teaching, we can usefully separate resources aimed at lecturers and resources for students. The consultation indicates that lecturers can benefit most from ideas and case studies that help them develop their own teaching content. For instance, resources that highlight uses of datasets or good data for teaching certain topics or methods. Lecturers are less likely to use ready-made teaching materials, especially, if they are not customisable. However, examples of teaching materials available can help promote the use of data and support the sharing of pedagogy. There is also variation among lecturers, and some may benefit from more detailed examples of how data could be used. Many lecturers signpost students to additional resources, including resources from the UK Data Service but often only 1 or 2 links to materials on specific topics. This use of signposting to specific resources on specific topics could be important to consider in relation to the scope and labelling of resources; potentially, smaller more specific resources would be more useful in this context.

6. Extend resources offering ideas for using UK Data Service data in teaching (from univariate methods up to regression), and optionally, provide supporting editable files with example code or instructions. This work could take different forms, for instance,
  - a. Add to existing Teaching Ideas
  - b. Case studies discussing general approaches
  - c. Develop guides or instructor notes for other datasets
  - d. Create a resource on teaching with open data
  - e. Highlight data with continuous variables good for teaching correlation and regression. For example, write an article such as 'Five datasets for teaching regression'.

7. Work on student centred resources on key issues that lecturers can easily signpost students towards. Some already exist and some can be repurposed from existing materials. Consider that a more extensive resource (such as a whole Data Skills Module) can be harder to integrate into teaching and harder for students to work through. Suggested ideas for topics and formats can be found in Chapter 5.

## Software

The UK Data Service should acknowledge that a wide variety of software is being used in teaching with SPSS the most common software used. There are strong arguments for the use of varied software and we should cater to both by providing software neutral resources or resources in different formats where applicable. There is also interest in resources specifically related to R.

8. Be software neutral by developing resources that are software neutral or resources for multiple software
9. Support those wanting to develop teaching in R with examples on Github (or even an R package) (some of these can come from existing resources and those developed through recommendations). Could also run an event/training about teaching in R.
10. Review the best options for Excel Users

## Dissertation students

11. Edit and adapt the Using Survey Data Guide into a series of short guides or factsheets (see Chapter 7 for further content suggestions).
12. Develop more dissertation case studies ensuring a variety of topics, datasets, and universities
13. Review Dissertation Award for fairness and accessibility across social science disciplines
14. Promote secondary analysis and the value of specific datasets to students. This work may include
  - a. a short promo video about secondary analysis using data from the UK Data Service (ideally, which includes varied students)
  - b. Student centred introductions to the main surveys

## Computational social science and reproducibility

Relatively few are teaching about computational social science and reproducibility. We also find some strong views about integrating programming into social science degrees due to concerns about alienating students and pressures on teaching time. Nevertheless, a wider number see these as important topics and are interested in resources in these areas. Given that the UK Data Service has a substantial programme of activity in this area, there is potentially an opportunity to develop some innovative resources. It should be noted that



interest in resources was greater among those already teaching these topics, which suggests there are opportunities for knowledge sharing among a community of practice.

15. Support teaching in relation to reproducibility. For example, appropriate materials for dissertation students at UG and PGT level and activities relating to coding and replication (connects to recommendation 11).
16. Consider scope to support teaching in relation to computational social science for instance
  - a. Explore opportunities for sharing ideas and knowledge through networks, events, and case studies
  - b. Create teaching resources for CSS methods (such as Machine learning, mapping) using data from the UK Data Service

## Introduction

The UK Data Service conducted a consultation with lectures and other higher education (HE) teaching staff in the Summer of 2021. The aim of the consultation was to gather information to help the UK Data Service understand how to support the use of quantitative data from the UK Data Service within teaching and learning in HE.

## Background

The UK Data Service has always sought to support the use of data in HE teaching by providing access to data through the Access Agreement for Teaching, developing teaching datasets including some that are unrestricted open access, and creating teaching resources and events for lecturers.

The UK Data Service previously ran consultations with lecturers and other HE teaching staff to identify needs in 2014, with follow-up feedback sessions in 2015 and 2016. These led to new developments including the UK Data Service Dissertation Award and Student webpages. However, the landscape of HE teaching has changed since these consultations. For instance, the Nuffield funded Q-step programme, designed to address the shortage of quantitatively skilled social science graduates, has now been running for 8 years.<sup>1</sup> We are also seeing the increasing use of programming in the social sciences and interest in issues around reproducibility. Therefore, it is timely to run a new consultation to understand current issues in using quantitative data in teaching.

## Focus of the consultation

The topics for the consultation fall under five broad areas.

1. Use of data in teaching including how data is selected and adapted in teaching and views about teaching datasets
2. Teaching materials and other resources that can support the use of data
3. Use and views of different software
4. Needs of dissertation students
5. View on teaching reproducible research and computational social science

Each area is discussed in a separate chapter.

---

<sup>1</sup> For further details, see information from the Nuffield Foundation:  
<https://www.nuffieldfoundation.org/students-teachers/q-step> [accessed November 21].

## Methods

The consultation combined three elements. We first used a survey with a short questionnaire covering key topics. The survey questionnaire consisted primarily of closed questions with additional free-text questions to allow participants to clarify, elaborate and comment beyond the main survey topics. The survey topics were:

- use of data in teaching, including finding and selecting datasets
- types of teaching materials
- statistics software
- needs of dissertation students
- views on reproducibility and computational social science

The survey also collected details of those willing to be interviewed. The interviews comprised two elements:

- open questions about participants experiences and views
- interviewees evaluating teaching related materials produced by the UK Data Service.

## Survey

With no suitable sampling frame, the survey used non-probability sampling. First, we emailed HE teaching staff known to be involved in teaching research methods (using Qualtrics). This stage included identifying names and contact information from university websites. We sent an initial request and (to comply with GDPR) one follow-up email. We also promoted the survey via relevant mailing lists, a UK Data Service news item, and on twitter via the UK Data Service twitter account and personal accounts of UK Data Service staff.

Since an aim was to hear from all those involved in teaching within quantitative data in social sciences, we also approached discipline specific organisations and networks such as the Political Studies Association teaching group and asked them to pass on the details through their network.

The survey received 109 valid responses from those teaching social sciences at a higher education institution and who use quantitative data in teaching or teaching research methods. Table 1 shows that most were lecturers or senior lecturers and above (over 80 percent). Postgraduate students and ECR Researchers accounted for 16 percent. Turning to disciplines, Sociology (29 percent) and Politics (21 percent) were most common disciplines. The remaining respondents came from varied disciplines including Biostatistics, Business, Communication and Media, Education, Social Statistics, Epidemiology, and public health.

*Table 1 Which of the following describes you most closely? Survey participants by position and discipline*

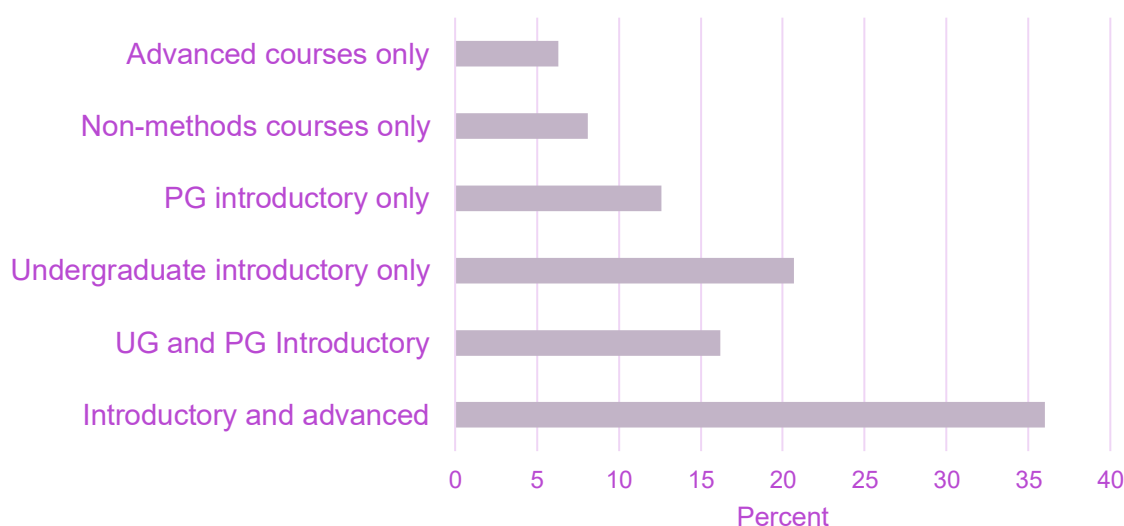
		Frequency	Percent
<b>Position</b>	Postgraduate	12	11
	ECR Researcher	6	6
	Lecturer	25	23
	Senior lecturer or above	65	60
	Other	1	1
<b>Discipline</b>	Politics	22	20
	Sociology	29	27
	Criminology	5	5
	Economics	11	10
	Psychology	6	6
	Geography	10	9
	Other	25	23

Figure 1 gives a breakdown of respondents by the type of teaching they do. About half are teaching only introductory research methods (at either undergraduate or postgraduate level or both), with some also using data on non-methods courses. Just over a third (36 percent) are teaching both introductory and advanced courses. Then, we have small numbers of those teaching only advanced topics (6 percent) or non-methods courses (8 percent) only. When rating their own expertise in quantitative methods and teaching, three-quarters indicated they had wide-ranging knowledge and experience of quantitative research methods including some more advanced techniques (Table 2). A majority in this group, also indicated they had lots of teaching experience. Most of the remaining third indicated they had good knowledge and experience of the basics, and a few 'little knowledge and experience of quantitative research methods'.

Based on these responses, we infer that the sample consists of HE teaching staff with substantial experience of quantitative data and methods and the teaching of quantitative methods. The information is therefore coming from many years of combined experience. We cannot be sure how representative the sample is of all those involved in teaching with quantitative data. Those with less knowledge and experience may have been less likely to have been asked to complete this survey as well as less inclined to notice or respond to invitation emails. We can therefore use the information to understand experiences and views

of varied HE teaching staff involved in teaching with quantitative data, but not necessarily the views and experiences of all those involved.

Figure 1 Breakdown of survey respondents (percent) by teaching type



*n*=109

Table 2 Background of survey participants (number of cases)

Knowledge and experience of quantitative methods	All	Among those with lots of experience having taught for many years and in many different contexts	Among those with some experience teaching or are new to teaching
Wide-ranging knowledge and experience of quantitative research methods including some more advanced techniques	75	51	24
Good knowledge and experience of the basics from either teaching or research	31	9	21
Little knowledge and experience of quantitative research methods	3	—	—
Total	109	60	48



## Interviews

Following the survey, we contacted those agreeing to be interviewed to arrange an interview. To try to capture diversity of experience, the initial set of interview invitations were sent to potential interviewees with diverse characteristics. We then sent further invitations to all those who agreed to be interviewed.

Three interviewers carried out 16 interviews in total. Interviews were conducted online using Zoom. The interview schedule included questions about

- the interviewees teaching background
- how they select and prepare data for teaching
- their teaching resources and materials
- dissertation students
- plans for changing their teaching in the future.

Participants also reviewed and evaluated specific teaching resources

- two teaching datasets and supporting instructor notes
- a worksheet for students on linear regression
- a document outlining an idea for teaching with data
- a guide for dissertation students
- case studies of dissertation projects

The schedule was designed to take 50 minutes, with the actual length of interviews varying from 30min to 2 hours. In one case, a shorter interview was agreed in advance and a shortened schedule was planned. In other interviews, interviews were slightly adapted due to technical issues.

## Data for teaching

An aim of the consultation was to better understand what data is being used in teaching and why. We wanted to learn about what makes a dataset good for teaching and what lecturers look for when selecting datasets. Equally, in relation to finding and accessing data for teaching, what are the key challenges and possible solutions.

### What data is used?

We found that a wide variety of quantitative data is being used in social science teaching in HE including data from the UK Data Service and other sources.

Many of those taking part in the consultation use data from the UK Data Service. Just over 60 percent of those surveyed report using data accessed from the UK Data Service. Responses indicate the range of data used in teaching is wide but that some data sources are being used widely. For instance, among those who have used data from the UK Data Service in teaching, respondents named 50 different preferred datasets. However, as we can see in Table 4, some datasets were named more often, these were the British Social Attitudes survey (mentioned 21 times) and Understanding Society (20), followed by the Crime Survey for England Wales (16). Census data was also mentioned by 10 respondents. Specific UK Data Service teaching datasets were named 8 times among the preferred datasets (see below for more discussion of teaching datasets).

From comments in the survey and interviews, we can identify a range of data from other sources that is being used in teaching. Survey data accessible from other sources was mentioned most frequently, with the European Social Survey (ESS), British Election Study (BES) and the US General Social Survey (GSS) mentioned most often. However, the interviews also highlight the use of varied other data sources with examples including data on the Olympics, Police data, data from a RCT – which they had to “beg and borrow” from a colleague – and data on pupil preferences for Justin Bieber, which may have been fake data but offered a “relaxed” way to practice multilevel modelling. Several of the interviewees had used synthesised or fake datasets.

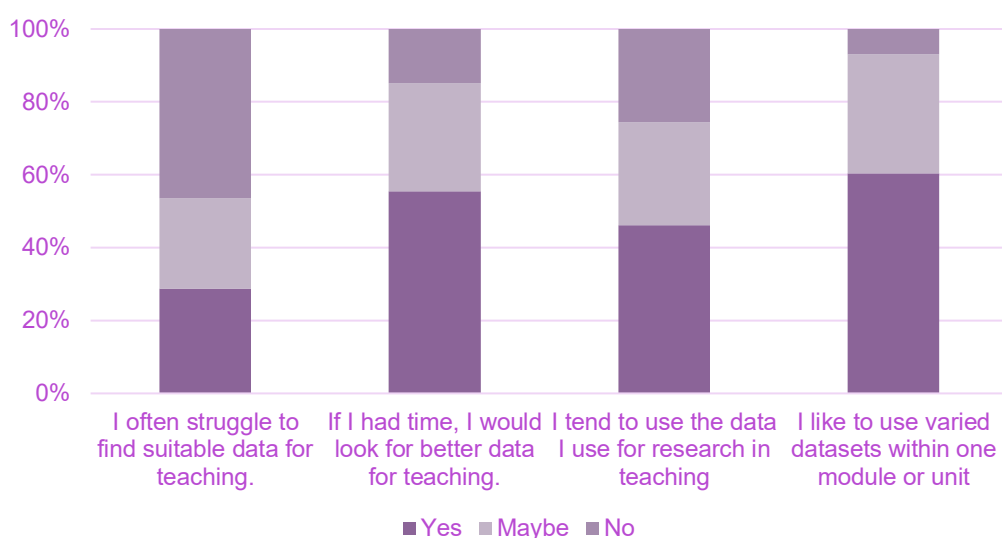
*Table 3 Top data from the UK Data Service used for teaching (from the survey)*

Type of data	Dataset mentioned	Times mentioned
Top five surveys	British Social Attitudes	21
	Understanding society	20
	Crime Survey for England and Wales/BCS	16
	Health Survey for England	9
	British Election Study*	7
Census	All combined	10
	Census (general)	2
	Census Aggregate	2
	Census Boundary	2
	Census Microdata	4
Teaching datasets mentioned	Crime Survey for England and Wales (teaching)	2
	ELSA (teaching)	2
	Labour Force Survey (teaching)	1
	National Child development study (teaching)	1
	Teaching datasets (in general)	1
	Understanding Society: Ethnicity and Health Teaching Dataset Wave 1, 2009-2010 (SN 8465)	1

Notes: \* British Election Study data can be accessed directly from the British Election Study website

The survey results in Figure 2 indicate a common preference towards using varied datasets within a module or unit; with more than half indicating that ‘yes’ they prefer to use varied datasets and less than 10 percent indicating ‘no’. Also in Figure 2, we find a mixture of some lecturers tending to use the same data as they do in research (45 percent) while the rest either may do (28 percent) or don’t (12 percent). The survey responses suggest that finding data for teaching is not necessarily difficult, but lecturers often perceive some room for improvement. For instance, while just over half indicate they do not struggle to find suitable data for teaching, more than half agreed that “If I had time, I would look for better data for teaching.” We also have nearly 30 percent reporting that they often struggle to find data and only 15 percent saying they would not look for better data.

Figure 2 Do you agree with the following statements about finding and using data for teaching?



## Common factors influencing the choice of data

From comments in the survey and the interviews, we identify six factors that are commonly influencing the choice of data for teaching and help explain why some datasets are especially popular.

### Real world data and research

Many lecturers emphasise the value of working with real world data. In many cases, this connects to a focus on research-led teaching, with students answering research questions using real data and appropriate methods. For instance,

**the idea is that the students get real life data collected by places such as the UK Data Service and they use research-based methods for a given scientific research question. (Interview 3)**

As a result, for many 'Real world and topic come first as priorities' when selecting data.

An extension to the research-led approach was an interest in replication. With some lecturers seeking to use data where students can replicate previous studies.

### Students' interests

Student interests are another factor influencing the selection of data, with many interviewees indicating that this is the primary factor in their decision making. As the selection of quotes from the interviews below shows, students' interests can relate to general topical issues or topics covered within their degree (so often topics on modules outside the research methods classroom). Students' interests are also seen as varied, especially since many teach classes with students from mixed disciplinary backgrounds. Several interviewees made the connection

between broad interests and the use of data from surveys such as British Social Attitudes Survey and Understanding Society due to the variety of topics covered. For instance,

**“How do I select a dataset? something relevant to the interests of our students, which is pretty wide to be fair” (Interview 4)**

**British Social attitudes is a very good dataset to use for students with interdisciplinary backgrounds....I can cover stuff on labour markets, health and political attitudes about immigration” (Interview 16)**

**Likes BSA, useful, matches to modules (Interview 12)**

Student’s interests can also come from some prior familiarity with the data, with the example of criminology students having some prior awareness of the Crime survey for England and Wales (CSEW). Some lecturers select data to include elements of fun in data exploration by using data students seem to like such as police data. Others also consider student comfort with a topic, for instance, when asked to evaluate the suitability of a teaching dataset based in the National Survey for Sexual Attitudes and Lifestyles (Natsal), some wondered if the topic of the data might be an issue for some students and would therefore hesitate on whether to use it.

In relation to students’ interests, many lecturers perceive the need for data to be *somewhat* current. While there is an acceptance that survey data is often a few years old due to the timescales involved, using older data is seen as problematic in terms of student interests. For some lecturers, an issue is that older data undermines the message that these datasets and methods are relevant because they give access to up-to-date evidence.

**I am very keen to use recent data. If I am telling students this is great as it’s update to date evidence on these topics, it’s no use using data from 2015 (Interview 5)**

In relation to the UK Data Service, issues around data being current were noted in relation to teaching datasets and that students can be interested in very topical data and using the UK Data Service it can be ‘hard to find current data’.

## Size and complexity

The size and complexity of data is a further factor, though we found a variety of different considerations made by lecturers. On the one hand, we found many lecturers opting to use very small and clean datasets for teaching, especially when teaching large groups introductory materials. Many perceive small datasets as being less likely to overwhelm students or helping engage students with the data by making it easier to understand and navigate. Typically, the size of data relates to variables (columns) rather than cases (rows). For instance, one interview summarised how they want



---

**Something that is small enough, so that it's easily understood. Size relates to number of variables not observations. Datasets cut down to 10-20. Flat files are enough for this introductory stage. Linking data across tables – is a bit too advanced for students at this stage. Being able to get it and share it. (Interview 4)**

Teaching time is also an issue as there is not always sufficient teaching time to cover data cleaning and in early stages these topics can be distracting to students. In a survey comment, the complexity of data is given as the reason for not using UKDS teaching datasets,

**I don't find the teaching datasets useful - they're still vastly overcomplicated relative to where my students are at, so I find myself stripping loads of stuff out of the full versions of datasets and sharing those with my students (Survey comment)**

In contrast to the small and simple datasets, we found many examples where lecturers look to use datasets that show some of the reality of data analysis and offer students some choice about variables they use. In these cases, they look for larger datasets with more variables and ones that are not artificially too clean. In this case, we find the contradictory view that pre-made teaching datasets are often too limited and do not offer enough variety. However, even in these cases there are many still adapting data for their purposes and look for data that is 'understandable' or straightforward, with the British Social Attitudes Survey specifically mentioned by some as good for teaching for this reason. For example, one described how they are "consistently trying to use data sets that people find manageable and intuitive". They want data that is accessible to as wider group of people as possible, which requires a little bit of data manipulation but not too much. Not as easy as "pressing a button and getting the answer" but also not prohibitively difficult.

Another explained how they

**need enough variety for me and them, to make it interesting for students. There aren't quite enough different outcomes for a class like I have. With the British Social Attitudes cut to about 100 I can cover stuff on labour markets, health and political attitudes about immigration (Interview 5)**

These different positions come together in relation to a common pattern of simple data being used initially and then moving towards bigger more complicated datasets. The progression from simple to more complex datasets can occur across years of a degree programme, with often a difference between data used in compulsory versus elective courses; for example,

**For first years, data is uber clean, it does exactly what they need it to do, we won't scare them with a massive dataset...second year is somewhere in between – longer list, clean it a bit. Third year – this is how horrible it can be to access data, this is how you would go about cleaning (Interview 2)**

Alternatively, progressions from simple to complex come through weeks of a course. For example, one lecturer talks about how they get students used to dealing with variables and add in as a layered approach over the weeks, for example, adding in weighting variables when needed (Interview 13). We also found another approach is to use more simple data for initial teaching and then a larger dataset when it comes to student projects and assessment as at these stages students need a choice of variables. For instance, “once you want students to decide on a project themselves within parameters, then you want the whole data” (Interview 12). In these progressive approaches, there is often the view that students are taught the ‘how to’ and then ‘set loose’ on real data.

## Data access and licensing

Another consideration when it comes to the choice of data is accessibility for both lecturers and students. Most critically for the UK Data Service, many commented in both the interviews and the survey that the current process for using data from the UK Data Service available under the End Users License (EUL) is difficult.<sup>2</sup> As a result, some opt to use data from other sources with the European Social Survey (ESS) and British Election Study (BES) most noted. The comment below is an example of this view, but it should be noted that similar views were expressed by many more.

**TBH when I've taught with UKDS resources before (it's been a while so this might have changed) the killer is the stuff around getting loads of signatures from students to confirm they're going to work with the data appropriately, which is just wildly impractical for the kinds of classes that 99% of people working in higher education are teaching, so (if you haven't done this already) more options to avoid this would be great. TBH this is the main reason I use BES direct from the BES website (Survey comment)**

---

<sup>2</sup> Most data from the UK Data Service are available under the End Users License (EUL), which requires users to register themselves and the use of data with the UK Data Service. Currently, the Access Agreement for Teaching enables lectures to share EUL data with students if they collect the names of all those accessing the data through a paper form or electronically via a virtual learning platform.

Interviewees also noted the relative ease of accessing survey data from the US such as the General Social Survey or the Survey of Health and Retirement. We also found requests for modified data for teaching that can be accessed without ‘jumping over hurdles’ or an online system for student signatures. For example,

**I did look into using the data in the past. However, I seem to recall that it was somewhat involved with not only me needing to register but my students also. I appreciate the need to proper record keeping and to ensure the data is used only for legitimate purposes. However, surely teaching data could be modified in such a way that the need for jumping over the hurdles could be removed (Survey comment)**

**Registration in the elephant in the room. The minute I have to rely on students having to register I have big problems. It’s almost impossible to meet that requirement to get students to register. There is no way that all students are going to be registered and accessing data independently. (Interview 5)**

A further issue for the UK Data Service to consider is the clarity of information around using data in teaching as we found both some misunderstandings around what data can be used and some comments that lectures felt the process was not clear. For instance, one lecturer thought that they could only use teaching datasets in teaching. A further comment also highlighted the need to clarify the process for those accessing data for teaching but not sharing the data with students.

**There is an issue when registering a data usage. When you select teaching as the purpose of data use it assumes you will give the data to the students. I don’t ever do this I get the students to download themselves so they know how. But I still download data for teaching to prepare materials etc. It would be useful if the registration for teaching use didn’t assume you will share data and make you agree to various things in relation to this.**

## Suitability

The suitability of the data to teaching a topic is also a factor in influencing the choice of data, most commonly it was the presence of continuous variables. Many noted the issue that social science datasets, in particular survey microdata which many students are interested in, contain few continuous variables, which makes teaching topics like correlation and regression difficult. Even when some variables are present, there is often a need for other variables for assessment or the analyses produce weak or null results. For example, one interview summarised the problem

**[It is] difficult to teach correlation and regression with social science data because of the lack of continuous variables. Especially when setting assignments and don't want to do same one as before. Ideally, we need age and two others at least to mix and match over years. (Interview 5)** .

In some cases, getting suitable data for teaching regression is one reason for switching dataset within a module or unit. At this point, some switch from using microdata to aggregate data (such as data from the World Bank) or a dataset used specifically for teaching these methods.

**What I do with my third years, is we have a cut down, very cutdown bit of the Health Survey for England and Wales, and just to explore using scale data in a linear regression and its literally height weight and blood pressure scales. And we make a joke about it not being interesting data per se, for social science students but it's really useful to get the elements of it. So we start with that one and we get the idea that if we increase your weight by this much your blood pressure might go up by that amount and it's nice to work with, then they get their flying wings if you like. And they can get a dataset they've chosen...We just picked that one because everyone understands getting fatter and the risks of that. And I picked that one, when I was desperately trying to find some nice data. (Interview 2)**

**For student projects, I switch to the World Bank data, where countries are cases and I have hundreds of options for variables. Really good as you get nice strong correlations (Interview 5)**

Other solutions to this problem include using non-continuous variables (for example, ordinal variables from attitude questions) or creating new continuous variables from additive scores or techniques such as factor analysis.

Other challenges around the suitability of data for teaching a method were not emphasized by interviewees as a problem. In some cases, lecturers accept some problems with the data. For instance, they might flag problems such as a skewed distribution or non-constant residuals, but in most instances, these are topics that fall beyond what is being taught. A view expressed in some instances was that if you find something that works sufficiently than a few problems were ok – and can even aid learning. However, on more advanced courses, critical students may question the inappropriate use of methods.

## Familiarity

The final factor is lecturer familiarity with data. A certain degree of familiarity with data is needed to use it in teaching; as a result, many lecturers use data they are already familiar with, especially if they are under time-pressure. In some cases, lecturers opt to use data they know from research. This brings many benefits in terms of understanding of any debates around measurement. Additionally, lecturers with many years of experience spoke of the knowledge they have of good datasets for teaching different topics.

As summarised in the comment below, a common theme was that once they find something that works, they often stick with it because there is considerable work in using a new dataset for teaching.

**I should probably find another one, but it works ...it takes a while to plough through a dataset. So, when you find one that works, you kind of stick to it a little bit (Interview 2)**

Though the desire to use recent data can prompt change after a few years. For instance, one interviewee explained how they have a system of creating a new teaching dataset of BSA data every two years as they want it to be current but cannot change it every year.

## Students and the UK Data Service

In addition to simply using data in teaching, several lecturers in the survey and interviews refer to developing students as data users with some explicit integration of the UK Data Service within teaching. For instance, as part of their teaching some seek to emphasise the availability of data and the idea of a data service. For instance,

**I always stress to my students how accessible the data is and how much effort is made by the data service to make that data usable (survey comment)**

**I have been proselytizing for UKDS for more than a decade and am concerned that it is not routinely included in EVERY social studies course. (survey comment)**

Relatedly, several interviewees talked about getting students to register with the UK Data Service as part of their class. And in some cases, students are encouraged to find data that interests them and explore it independently. We also find lecturers focusing on developing students as data users in several ways. For example, some emphasised the need for students to be able to understand documentation and metadata.

**Very important to teach students about the documentation and metadata, as this is something they miss and then analysis goes wrong! (Interviewee 12)**



**Skills in searching for data are important too, don't want to make it available too easily (Interview 16).**

## Adapting data for teaching

A substantial number of lecturers, though not all, are using some form of adapted data for teaching, which often requires extensive work to adapt and prepare. For example, Table 3 shows that from those taking part in the survey 89 percent usually or sometimes prepare data to make datasets more suitable by, for example, reducing the number of variables, setting missing values and recoding variables. In contrast, only 16 percent of survey respondents indicated that they usually 'make no prior changes to the data, students see the full data'. Teaching datasets from the UK Data Service are used by about half of those surveyed but only usually used by 20 percent.

*Table 4 When using data for teaching, how often have the following applied?*

	Never	Sometimes	Usually
I make no prior changes to the data, students see the full data	28%	56%	16%
I prepare data to make datasets more suitable (e.g. such as reducing the number of variables, setting missing values and recoding variables.	12%	43%	46%
I use teaching datasets from the UK Data Service	47%	34%	20%

*n=101*

As discussed above the extensive use of adapted data for teaching reflects the need for simple dataset that students can intuitively understand and handle in limited teaching time. As one expressed, they used teaching datasets because they are simpler, which leads to 'fewer challenges or confusion' and students are 'less overwhelmed by sheer volume of data'. It can also constrain the number of mistakes even if does constrain the type of questions.

What kind of adaptations are lecturers making? As the issue of simplicity is important, one of the most common adaptations mentioned by lecturers is reducing the number of variables. Those interviewed gave varied suggestions around optimal sizes of teaching datasets, with the number of variables ranging from around 20 to 100. The variation corresponds with the use of different datasets depending on the teaching context. For instance, there is an interest in small and simple datasets (up to 20-30 variables) that can be used for introductory methods teaching. Conversely, many opt to use larger teaching datasets ranging from 40 to 100 variables when they want to offer students choice in selecting variables for analyses. When removing variables some opt to remove details of weights and other sample design variables. Other adaptations include recoding variables into a simpler format and changing variable names and labels "to something sensible". Some re-document datasets, often providing a simple variable list with level of measurement included. In addition, to lecturers adapting

datasets, one lecturer described creating a prepared syntax file for students to use to simplify the dataset.

In addition to creating teaching datasets for lecturers to use, we received a request for advice on creating teaching datasets.

**There are some good teaching datasets but it would be good to be able to create my own, so some guidance on how to do this would be great (Survey comment)**

## Evaluations of the National Survey of Sexual Attitudes and Lifestyle Survey teaching datasets

To get further insight into good and bad features of teaching data, we asked interviewees to evaluate two teaching datasets prepared by the UK Data Service based on data from the National Survey of Sexual Attitudes and Lifestyles (Natsal). Starting with the general topic, many identified it as a good choice for students with comments such as “Spot on” and “Can’t think of a more perfect dataset.” However, others had some concern about student comfort with the topic and the potential for ‘sniggers’. Linked to a preference for current data, the age of the Natsal data was raised as problematic in one instance.

Comments about the selection of variables varied. Several interviewees commented positively about the variables in both the EUL and open datasets. In particular, the presence of continuous variables, including a purposely derived variable measuring attitudes (from factor scores). With some specifically commenting that including these additional variables was useful. In contrast, some suggested the datasets did not contain the variables they would choose. This reaction highlights how individual preferences make it hard to create general purpose teaching datasets. More specific comments noted that fewer attitude variables and more behavioural variables would be preferable.

The two teaching datasets differed in terms of size and access conditions, with one small open license dataset and one larger dataset available under the End User License (EUL). For some the open access teaching dataset matched what they would look to use in an introductory course, with some suggesting that if it had a further five variables, they could use it for most teaching. In some discussions, it was suggested that a range of these datasets covering key topics such as crime, health, and work, would be ideal. These reflections on the size of the datasets suggest many lecturers would be able (and happy) using small open access teaching datasets if available on a range of topics and reasonably current.

Similarly, some commented that the EUL dataset was larger than what they would typically use. However, many of those usually preferring smaller datasets suggested the Natsal teaching dataset of around 100 variables would be workable. Associated with this evaluation were comments about the formatting of the variable list, with the grouping of variables by topic

making the overall size more manageable. These comments suggest that preferences on the size of a dataset depend on the presentation of the data.

The interviews included many positive evaluations of the user guides for the Natsal teaching datasets, including the overall appearance. A specific feature well commented upon were the links contained in the User Guide to the original study including the questionnaire. For many it was important to be able to trace variables back to the survey questions. A further link to a summary of findings from the survey published in the Lancelot was also noted as useful.

Comments from interviewees conflict about the short descriptions of the survey design and missing data. Some specifically noted that the having short descriptions is useful as they are more accessible for students. However, others see these topics as too advanced and off-putting and questioned whether they would show students the documentation at all. Alternatively, it was suggested this kind of information should be presented after information about the topic and variables. Suggestions for additional information include adding information about level of measurement in the User Guide. It was also noted that we could highlight any variables where there are substantial missing data due to questionnaire routing as this is an issue that can confuse students. A further suggestion was that “If it's dedicated to being a teaching data set it might be nice to have a glossary in it to explain some of the survey and design related terms, e.g. response rate definitions”.

## Suggested principles for creating teaching datasets

Based on the discussions around data used for teaching and the Natsal teaching dataset, we can identify the following principles for a good teaching dataset:

1. Should use real data with manageable issues
2. Can be very small and simple (for example up to 20 (or 30 variables) or something larger and more complex (up to 100 variables). Due to variations in lecturer preferences, tend towards including some additional variables to offer choice.
3. Be intuitive and understandable (through easy-to-understand variables and accessible documentation that provides well-formatted information about variables including level of measurement)
4. Include multiple continuous variables that can be dependent and independent variables, even if they need to be derived using additive scores or factor analysis
5. Be current and connect to student interests (topical or linked to core social science themes)

Other factors to consider

- Some lecturers prefer to change variable names to be more intuitive. However, such as approach would make it harder to trace variables back to the original data and the questionnaire so may not be desirable to everyone.
- Where possible link to reports or articles summarising findings from the data.

## Teaching Resources

The consultation considered whether additional teaching resources and materials could help support the use of data in teaching. This chapter discusses lecturer perceptions of the usefulness of different types of resources, distinguishing between resources

- for lecturers such as information about datasets along with ideas and examples
- for use in teaching (such as data exercises and worksheets etc.)
- for signposting student to for self-learning

## Views on the usefulness of different resource types

In the survey, we asked lecturers how useful different types of resources would be for teaching. The responses in Table 5 indicate many would consider resources to be useful but that certain types of resources were more commonly indicated as useful than others.

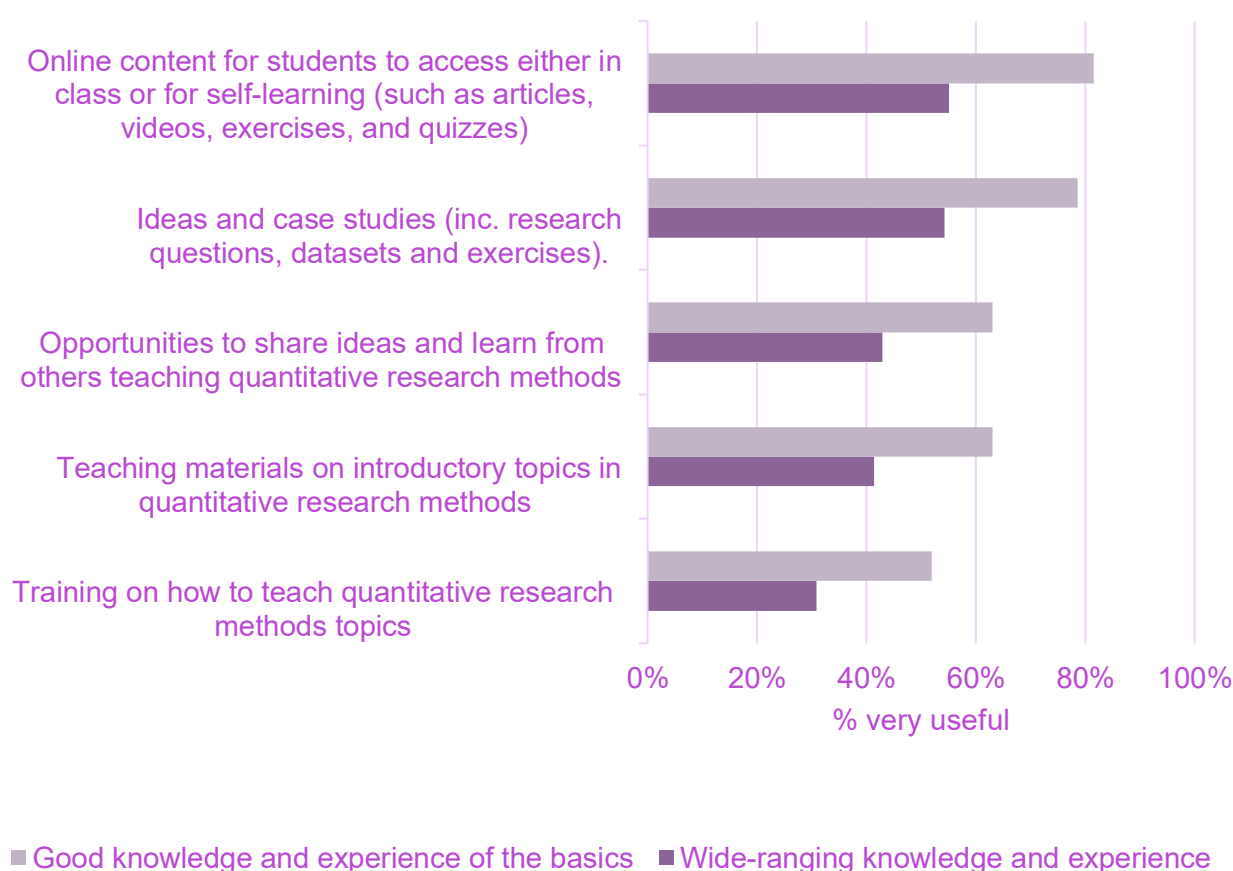
- Most popular were ideas and case studies (including research questions, datasets and exercises) were with 62 percent indicating these would be very useful and a further 34 percent moderately useful.
- Also popular was online content for students to access either in class or for self-learning (such as articles, videos, exercises, and quizzes) with just over 60 percent responding that these would be very useful and a further 33-35 percent moderately useful.
- Fewer indicated that teaching materials on introductory topics in quantitative research methods would be very useful, however just under half still indicated they could be very useful and a further 40 percent moderately useful.
- There is some interest in training in using data in teaching, with 38 percent (37/97) percent indicating this would be very useful. More popular than training are opportunities to share ideas and learn from others teaching quantitative research methods, with half indicating this would be very useful.

*Table 5 How useful do you think the following would be for your teaching (%)?*

	Very useful	Moderately useful	Not at all useful	N
Teaching materials on introductory topics in quantitative research methods	49	40	11	99
Ideas and case studies (inc. research questions, datasets and exercises).	62	34	4	100
Opportunities to share ideas and learn from others teaching quantitative research methods	50	43	7	100
Training on how to teach quantitative research methods topics	50	43	7	100
Online content for students to access either in class or for self-learning (such as articles, videos, exercises, and quizzes)	63	34	3	100

The survey results suggest that some lecturers might find resources more useful than others with some differences by expertise and experience. Figure 3 shows differences by background in quantitative methods. Here, those indicating they have ‘good knowledge of the basics’ more often report resources would be ‘very useful’ than those with ‘wide ranging knowledge and experience’ of quantitative methods. For instance, 54 percent of those with wide ranging experience report that ideas and case studies would be very useful compared with 79 percent of those with good knowledge of the basics. However, the relative popularity of different resources is approximately the same by level of experience with online content for students and ideas and case studies for lecturers most popular.

*Figure 3 Usefulness of varied resources by knowledge and experience of quantitative methods*





---

## Resources for lecturers

As part of the interviews, interviewees evaluated the potential usefulness of two resources developed by the UK Data Service designed to help lecturers develop teaching. We find that in general, the discussions about these resources echoed the view that resources for lecturers (such as ideas and case studies) could be very useful (and more useful than teaching materials to use in class).

### Instructor notes

The first resource was a document containing instructor notes that accompany a teaching dataset. The notes discuss the contents of the dataset and how it might be used in teaching. Many interviewees indicated the document would be useful with comments such as:

**Fantastic shortcut for finding a dataset with interesting topics. Super useful...This is awesome actually – ideas for teaching quantitative methods, I really like that it's very helpful, short. Very quickly gives some ideas of what we can do with it. I wish I had this last year. (Interview 4)**

**like this document as a starting point either if you're not confident or haven't used the data set before it breaks down barriers of having used it for the first time. Also good for doing own teaching materials - removes labour. (Interview 13)**

**The questions are very interesting, useful, the interval and continuous variables. There's some examples of some analysis that you could do for regression, so yeah, I mean, that's really useful". (Interview 16)**

Specific features of this resource noted by lectures as useful were a list of topics that could be examined using the data (to assess interest to students), clear information about the continuous variables in the dataset and ideas for how the data could be used to teach specific methods. The conciseness of the 2-page document was explicitly emphasized by several, suggesting resources aimed at lecturers need to be concise and easy to navigate. One interviewee did question whether lecturers need to do this work themselves as familiarising yourself with a dataset is necessary before teaching with it. However, such a guide could potentially be a 'useful starter'. Further suggestions for this type of resource were to ensure there is sufficient coverage of basic techniques such as univariate and bivariate analysis as "One of the things that's hard about teaching quants is finding real-world examples of very basic techniques". Other suggestions included having syntax/code for the suggested activities and having versions for other popular datasets.

## Teaching Ideas

The second resource used in the interviews was a document outlining ideas for teaching based on a research topic and specific dataset. The aim of the resource is to provide information about relevant variables, possible analyses, and the results obtained. Again, many lecturers identified such a resource as a useful shortcut with comments such as

**very timely, I think this is very good and It allows you to quickly suss if it can be used. (Interview 3)**

**I would say the teaching idea, 2-pager, is most useful. What I am usually doing is creating my own resources. I wouldn't take a worksheet and use it as it is. But ideas for datasets and how to use them is very useful. (Interview 4)**

**A couple of pages like this is quite useful. Good to have concise resources that highlight something that might work for one week.... This type of resource could be very useful for theme-based courses rather than methods courses (Interview 5)**

Several interviewees mistook this resource as something to use with students, a perception that seemed to connect to the aesthetics of the document. In these cases, some liked the idea of having something that looked good and “quite off-the peg”; however, others thought they would not use it they would want to tailor it to their own purposes. In several cases, the resource was perceived as being closer to an assessment rather than something for lab classes.

Beyond the issues of the document purpose, specific features of this resource noted by lecturers as useful include the choice of topic and dataset, the connecting of a research question, dataset, and relevant variables, and it being short. Suggestions for improvements were to make the document shorter focusing on just the research topic, dataset, and information about relevant variables. It was also suggested to have more using varied datasets and topics. A further observation from the interviews was that the resources for lecturers all presume knowledge of the suggested methods. None of the interviewees suggested this was a problem for themselves but some wondered whether those with less expertise would need more information on how to do the suggested analyses.

## Teaching materials for use with students

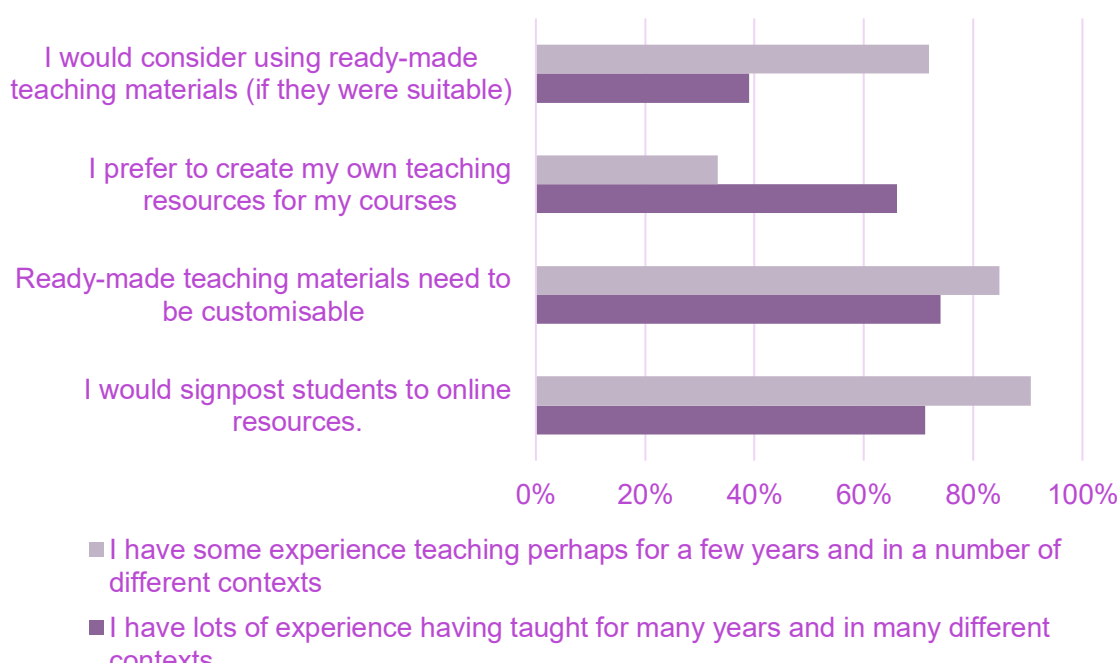
The survey and interviews offer several insights into whether and how lecturers might use ‘ready-made’ teaching materials in their own teaching. From the survey, we identify a preference towards creating your own resources, for instance in Table 6 around half the survey respondents indicated this to be their preference. However, a similar number indicated they would use ready-made teaching resources. A majority thought teaching materials needed

to be customisable. We find some difference in responses by teaching experience. In Figure 4, those indicating they have ‘some experience teaching’, 72 percent said probably yes to the question about using ready-made teaching materials, compared to 39 percent of those with lots of experience. Also, over 60 percent of those with ‘lots of experience’ indicated they would prefer to create their own teaching resources, compared to 33 percent of those with some experience.

Table 6 Views about teaching materials (%)

	Probably yes	Might or might not	Probably not	n
<b>I prefer to create my own teaching resources for my courses, rather than using ready-made ones</b>	52	39	9	99
<b>I would consider using ready-made teaching materials (if they were suitable).</b>	53	44	3	99
<b>Ready-made teaching materials need to be customisable.</b>	76	19	5	98
<b>I would signpost students to online resources.</b>	78	20	2	98

Figure 4 Preferences around teaching materials (percent ‘probably yes’)



## Feedback on UKDS student SPSS worksheet

We asked interviewees to evaluate the usefulness of a worksheet with step-by-step instructions and screenshot images for doing regression in SPSS. Overall, feedback on the worksheet was more varied than the resources for lecturers, with fewer indicating they would use it in teaching.

Many indicated that they did use similar resources, suggesting students need these step-by-step instructions. Some of these further indicated they would be interested in having this type of resource available, especially on key univariate and bivariate analyses. Echoing the survey results about resources needing to be customisable, several asked if the worksheet would be in a customisable format. Others said they would probably not use the worksheet as they have these types of materials already or they prefer to create their own. However, some did suggest that it might be useful for others 'starting out'. Several viewed the worksheet as a potential resource for student self-learning rather than something to be integrated into their teaching.

We also heard the view that worksheets were not a resource they would use in general. Some described worksheets as outdated, often saying that videos work better. There was also the view that worksheets are less relevant for those teaching with Stata and R, where teaching materials are more likely to be code files. In these cases, the issue was with the format of the resource rather than the content of the exercise.

## Barriers to using ready-made resources

Both the interviews and some comments in the survey indicate some barriers around the use of ready-made teaching materials, which the below points summarise:

- Limited time to change content of teaching generally, for example, "There are huge pressures to just use what worked ok previously".
- Finding resources is time-consuming, especially as there is not one clear place to look.
- Resources are never quite right
- Hard to integrate resources into existing courses, for example "I have a framework for how I am doing it, might use it for a new course"
- Adapting resources takes time – for instance, "it's simply that adapting your templates is not much quicker than drafting a new exercise based on my own existing materials".
- Resources need to be branded, fit house style

In addition to identifying barriers, comments from lecturers include some suggestions for ways to reduce or remove them. For instance, in relation to finding resources one commented that "coming up with an easy to find and navigate interface to this kind of resource would be top of my list". To help with the time needed to adapt resources, resources can be provided in an easy to customise format (for example, not in PDFs).

## Other benefits of resources: promotion of data and pedagogy

In addition to being used in teaching directly, the interviews pointed to other ways that teaching materials can support the use of data. For instance, in one interview, a lecture explained that they chose to use a dataset for teaching because they had seen existing teaching resources based on that dataset. They developed their own resources but seeing the online resources had indicated to them that the dataset was available and that it included relevant (continuous) variables for teaching correlation and regression. We heard a similar story from someone in relation to data not from the UK Data Service, that they had come across the data and resources and chose to use the data because the “exercises weren’t from scratch” there was code included for R and other software. The code also produced nice, well-formatted output that was more student friendly than some R output. Therefore, though resources were not used directly, they did inform choices around using data in teaching. Equally, another interviewee explicitly made the point that they would “be inclined to choose a dataset because of all available resources”.

A more general pedagogical benefit of making resources available online was suggested by another interviewee. They suggested that “good resources can get people to think about how they teach. Why have you done this in this order?”.

It therefore seems that resources using data, even if not used as they are, can support lecturers by providing ideas and examples for datasets or activities or for pedagogical approaches.

## Signposting students to resources

Both the survey and interviews indicate that many lecturers would signpost students to online resources. Table 6 shows around three quarters indicate they would sign-post students to online resources. We found most interviewees talked about signposting students to resources, including UK Data Service resources. We also found that when reviewing UK Data Service resources many looked at them as something they would signpost students to rather than a resource they would use in teaching.

The interviews also gave insights into how signposting works in practice. One such insight is that though many lecturers signpost, several interviewees placed limits on how much they would signpost. As one explained

**I think there is a danger we might just end up redirecting students to other places. But I think it’s important to tailor it to students. I will flag up one or two resources. I will say there is a really nice guide here that covers the same material but in a slightly different way (Interview 5)**

A further issue around signposting is that lecturers signposting students does not necessarily equate to use by students. Framed in relation to dissertation events and resources, one lecturer explained how they do signpost students to the UK Data Service resources but as lectures they have no way to monitor take up. As a result, even those lecturers signpost

students to resources, the actual impact on students may be less as in practice students may not access additional or supplementary resources

## Format and topics

The interviews included various insight into preferences about the format and topics of teaching resources.

## Worksheets, videos and guides

Several indicated they used worksheets like the UK Data Service example or had developed workbooks and guides. Conversely, others suggested workbooks and guides were outdated.

Video was mentioned in several interviews with some indicating it was a better method than step-by-step instructions or that increasingly it was the preferred format for students. As part of these discussions, several talked about increasingly using videos in teaching as part of ‘flipped learning’ even before the move to online during the pandemic. However, notably many of the interviewees contrasted this preference for video with their own preference for downloadable resources including PDFs.

When it comes to teaching resources, many talked about using code files as a teaching resource, especially those teaching with R or Stata.

## Focus on basic techniques

One of the most common suggestion was the need for resources on basic techniques. These suggestions included exploratory univariate and bivariate descriptive statistics as well as some specific mention of handling data in Excel. A related suggestion was for “Dashboards and visualisation tools that allow students to explore the data independently”.

## Case studies and examples

A suggestion echoed across several comments concerns case studies and examples on varied issues, for instance

- examples of dangers e.g. Simpson’s paradox
- interpretation and rationale for a method
- resources linked to key sociological themes/issues covered on the degree programme

A research-led approach appeared in many comments, with some suggesting direct links to published research.

- Resources that connect research questions to a dataset and then to means of analysing the dataset, in that order.
- published papers reproduced so students could see the whole process from data prep to analysis
- Stories of real, published research, written or told by original authors



- Tasks that enable students to 'match' published findings would be useful

## Sampling design and missing data

Other suggestions related to issues of supporting students in areas that might not be covered by introductory modules but are important to working independently with data such as sampling design, weights and missing data. One, described a gap in relation to resources for dealing with missing data from attrition in longitudinal studies. While unsure whether it should be from UK Data Service, they suggested some teaching ideas on this topic would be useful.

## Focus on data and documentation

We also had some comments indicating the view that the UK Data Service should focus on issues relating to data rather than analysis. For instance, some emphasized that “suitable simplified 'real life' data sets are the most valuable resources for me” or that what they can do with data is their “creative domain and our focus is data and documentation”. Areas mentioned in these comments include how to get data (via our website), understanding documentation, issues relating to samples, and research data management.

## Promotion and labelling

Several discussions in the interviews linked to issues around the promotion and labelling of resources. Most notably, though some interviewees were aware of the resources shown in the interviews, we heard lecturers express the view that ‘they wished they had seen this’ when planning their teaching. It may therefore be useful to consider the organisation of resources online and their promotion. Some indicated potential interest in a training session such as a webinar running through what’s available.

A further issue to consider is appropriate labelling of resources. We found that on some occasions, lecturers misunderstood the intended purpose of resources. Appropriate titles, descriptions and listing online will help lectures to quickly identify what might be useful to them. This issue was also raised directly by an interviewee

**This worksheet assumes that students know what linear regression is and what it's for. Useful to make explicit who resources are for example if students are like this, then you might want to do this etc. (Interview 6)**

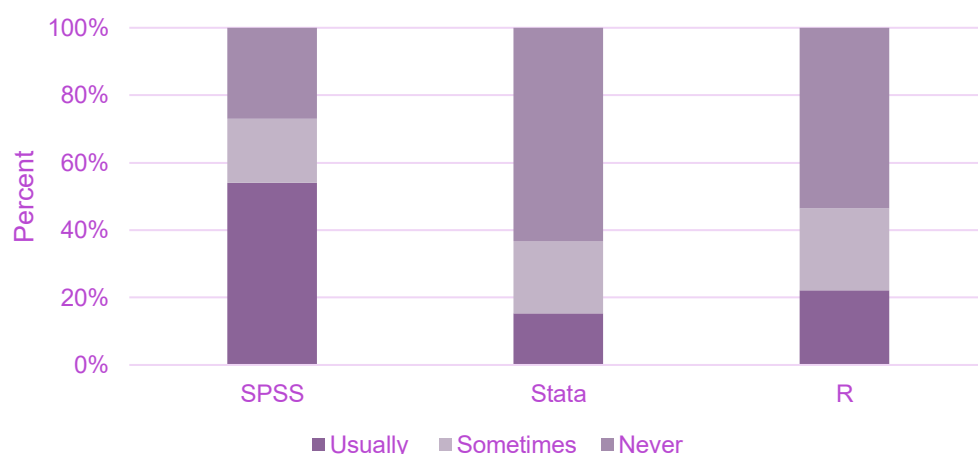
## Software

Use of software to analyse data is a key part of working with quantitative data, and often an important element of teaching. There are various software options with Stata, SPSS, and R the most common in the social sciences. This section examines the use of different software when teaching with data and the factors influencing choices about software.

### Use of different software

In the survey, we asked how often they use SPSS, R and Stata for teaching. Responses (Figure 5) show the continued widespread use of SPSS. Nearly, three-quarters (73 percent) of our respondents use SPSS at least sometimes and over half (54 percent) ‘usually’ use SPSS for teaching. Stata is used by a smaller proportion, with just under 40 percent using it at least sometimes and 15 percent indicating they usually use it. In our sample, R is slightly more commonly used for teaching than Stata, 22 percent usually use R and 24 percent sometimes use R. More than half never use Stata and R while only 27 percent said they ‘never’ used SPSS. The survey results and the discussions in the interviews indicate many use a mix of software. For example, in the interviews we had some using R alongside Stata and SPSS and then some using R, SPSS and Excel.

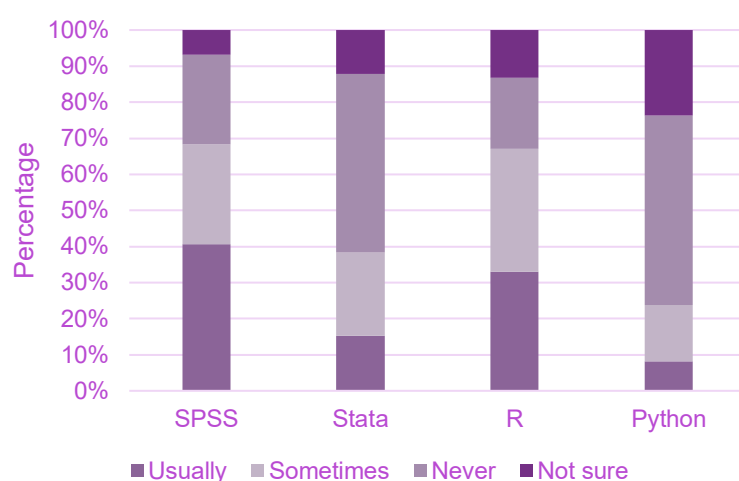
Figure 5 how much do you use the following software for teaching?



The survey also asked about how those teaching expect to use different software for teaching in the next few years. While the dominance of SPSS remains evident, responses indicate some potential for change. Around 40% of respondents said they expected to be ‘usually’ using SPSS over the next few years (Figure 6), which is a decrease from the 54 percent who use it in their teaching currently (Figure 5). This change suggests that some expect their use of SPSS to decrease in the future. In contrast, a third of respondents (33 percent) expect they will ‘usually’ use R in their teaching in the next few years compared to 22 percent who use it in teaching currently (Figure 5). The proportion expecting to use R sometimes in the future is also greater than the proportion who sometimes use it currently. The numbers expecting to use Stata are like the numbers using it now (around 15%). While SPSS remains dominant,

these figures point towards a potential increase in the use of R in teaching over the coming years – or at least a view among those teaching that the use of R will increase. The question about future use included the option of Python. Just under a quarter indicate they expect to use Python in the future at least sometimes, half expect they will never use Python and the remaining quarter are unsure.

*Figure 6 In the next few years, how much will you use the following software for teaching?*



In the survey results, we find some evidence that software use varies by the type of teaching, with SPSS used more by those teaching only introductory courses. For instance, in Table 7, we see that among those teaching only introductory courses (either UG or PG), 60 percent usually use SPSS and 19 percent never use SPSS. Among this group, half indicate they will still usually use SPSS in the future with the proportion never using SPSS not changing substantially. For those who are teaching some advanced courses, current use of SPSS is slightly less common, with 55 usually using SPSS and 30 percent never using SPSS. We also see a greater shift in the proportion indicating they will use SPSS in the future among those teaching some advanced topics.

Conversely, use of Stata and R are more common among those teaching advanced level courses, more than a third (36 percent) of those teaching advanced courses usually use R and nearly 20 percent use Stata, compared to only 11 percent of those teaching introductory courses only. Among those teaching introductory only, 72 percent and 65 percent never use Stata or R respectively compared to 51 percent and 40 percent among those who also teach advanced topics. Expected use of R in the future is also greater for those teaching advanced topics, with just under 40 percent expecting to use R 'usually' and over 60 percent expecting to use it 'sometimes' or 'usually'. While R is used more than Stata, at least half of those teaching some advanced topics expect to use Stata 'sometimes' or 'usually'. Finally, coming to Python, 60 percent of those teaching only introductory courses expect they will never use Python in teaching in the next few years. Those remaining 40 percent are evenly split

between 'not sure' and expecting to use it (either 'sometimes' or 'usually'). For those teaching advanced topics as well, the proportion indicating they would never use python is less at 41 percent. Again, the remaining are split between not sure and expecting they will use it to some degree.

*Table 7 Use and expected use of software by level of teaching*

Introductory only					Advanced or Introductory and advanced			
Current Use	SPSS	Stata	R		SPSS	Stata	R	
Never	19%	72%	65%		30%	51%	40%	
Sometimes	21%	17%	24%		16%	30%	24%	
Usually	60%	11%	11%		55%	19%	36%	
Expected use	SPSS	Stata	R	Python	SPSS	Stata	R	Python
Never	21%	63%	23%	60%	30%	37%	14%	41%
Sometimes	21%	19%	33%	11%	32%	28%	27%	24%
Usually	52%	8%	23%	11%	36%	23%	39%	7%
Not Sure	6%	10%	21%	19%	2%	12%	21%	29%

Lecturers also highlighted the use of additional software in a free-text question. Most commonly, Excel was mentioned as an obvious omission from the survey questions by a few respondents with comments highlighting how use of Excel is expected on work placements and in employment post-graduation. A few comments highlighted it's use as a basic pre-cursor prior to teaching more advanced software, or as useful add-on for visualisation.

Below is a list of all additional software mentioned.

- Excel (including Plugins)
- Jamovi
- QGIS/ARCGIS
- Nesstar
- Minitab
- MLWin
- JASP
- Machine Learning packages: Knime, RapidMiner, WEKA
- JULIA

## What influences software choice?

Comments in the survey and interviews highlight conflicting views around software and programming in teaching the social sciences.

### SPSS

The widespread use of SPSS comes mainly as some think SPSS is the most intuitive software for students. Those coming to this conclusion often do so while identifying negatives of SPSS but still that on balance it is the better option, especially for introductory courses. For instance,

**“While recognising the benefits of moving to R, for introductory courses for students with minimal non-quants experience i still feel SPSS has huge advantages as the most intuitive software to use...”**

**"SPSS is most user-friendly software university offers....I wish there was a package that was better than SPSS 'cause I hate so much about it. It's horrible, it's bloated, it's nasty looking, but it's I've looked at Stata. I've looked at R studio, it's just it's not going to be any easier for the students".**

The main features of SPSS that are seen as beneficial is the ability to easily view the data spreadsheet and to see the codebook. For instance, “R isn’t as good as you don’t get to see the data in the same way”. SPSS can also be used for a good range of methods. In one case, employer expectations were given as a reason for continued use of SPSS. In this case, they explained that after moving their undergraduate courses to R, comments from employers of their advisory board prompted a move to reintroduce SPSS. The view was that some employers use SPSS and like students who are familiar with it. However, use of SPSS is also controversial with comments describing it as *‘totally incomprehensible’* and *‘hideous’*. The main argument against SPSS centres on it being propriety software with some arguing that teaching ‘closed source’ software is irresponsible and should be discouraged. Some also propose that the user interface in SPSS, where students point and click, means that students do not fully understand what they are doing.

### R (and python)

The arguments for using R tend to focus on it being open software, code based, and (more generally) that it is ‘on trend’ or the ‘future’. The benefits of open software were sometimes described in relation to cost (e.g. saving money by not buying licenses), the more general

view that open software should be preferred or that open software was more cutting edge. For instance,

**We feel like R is more on trend in terms of what is out there, partly because it's free and also the things developed in Stata are good and reliable to use, but the cutting edge stuff is found in R before it's found in Stata. It will also save the department money, not having to buy a bunch of Stata licenses"**

Some discussions focused on issues of employability as with open software students can more easily take their skills to an employer. Coding in general is also seen as a better approach, with some suggestion that it is better for student understanding than pointing and clicking. We also had some using additional features of R in teaching such as RMarkdown (for combining code with text and results) as well as cloud-based services.

However, comments about R show some perceive its growing use as controversial. Some emphasise the large learning curve and that the lack of students' technical ability mean it may cause issues and further 'put-off' those who are already nervous about maths and statistics, especially first year students. For example,

**"I have never met a UG who is nervous about statistics and maths, who would not be further intimidated by having to learn programming before being able to do any stats"**

**"It would be a nightmare for students because they're worrying about whether they had a backslash or forward slash..."**

The argument that programming is too much for students and there is not enough time to teach is discussed further in chapter 8 in relation to Computation Social Science (CSS). In terms of pedagogy, like arguments that pointing and clicking is not good for students understanding, some argue that with code-based approaches you can end up with students copying and pasting, which is not necessarily a good use of time either.

We also find some that are interested in using R but face barriers in terms of department choices and their own experience. Issues arise with lecturers themselves not being confident using R or not having enough time to learn or teach it. For instance,

**"Moving to Stata/R requires planning and coordination. My R skills are only ok, not a fluent user"**

**"We know that R and python most useful for students, but not enough time to learn..."**



## Mixed use of software

A further view indicated in some survey comments was that students can increasingly choose the software that work best for them, or use whichever tool is best for what they are doing:

**“Different software suit different needs and different students”.**

**“It does not matter particularly what tools are used, students can increasingly use whatever tool they feel most comfortable with”**

## Software and resources

In the survey and interviews, several participants gave suggestions in relation to issues around software in resources. Though considering the issues discussed above, these suggestions are quite varied. One suggestion was to create resources in multiple versions for users of different software. Alternatively, it was suggested that resources should avoid making assumptions about what data package students are using. A few comments made specific suggestions for r-based teaching resources. These included

- an R-based teaching software package based on UK Data Service data.
- a GitHub repository with all resources and code and webpages.

It was also suggested that training to help people who are “left behind” upskill would be useful and that if this is a common challenge for teachers have a “train the trainers” session.

## Dissertation students

In the survey and interviews, we asked about the needs of dissertation students and the potential usefulness of resources on different topics. The interviews included evaluation of an existing UK Data Service guide for dissertation students, Using Survey Data, and a dissertation case study.

## Experiences of dissertations

Most survey respondents and interviewees were supervising dissertation students or had done in the past. When discussed in the interviews, interviewees expressed varied experiences around the popularity of quantitative dissertations using secondary analysis. Several interviewees had students regularly using data from the UK Data Service in their dissertations, including access to special licence data. One explained how they define dissertation projects including suggested data to use. We also find lecturers integrating UK Data Service and our resources into their dissertation support.

However, many identified issues around students doing secondary analysis of quantitative data in their dissertations. For some, the issue was student interest or concern that they would get lower marks for secondary analysis. Others identified secondary analysis as a big challenge to dissertation students, who can struggle to find relevant data and once they do, they can have a large amount of work to prepare the data for analysis.

## Usefulness of resources

Table 8 shows many lecturers perceive dissertation students could benefit from additional resources on the suggested topics to some extent. The topics most seen as beneficial are statistical methods, finding and accessing data and data management. Fewer indicated that resources on reproducible research and complex samples and weights would benefit students; however, still about half indicated they would benefit students ‘a lot’ and only 7-8 percent ‘not at all’.

*Table 8 Useful for dissertations students*

	A lot	A little	Not at all	n
<b>Statistical methods</b>	74%	24%	2%	95
<b>Finding and accessing data</b>	70%	27%	3%	96
<b>Data management (merging data, handling large files)</b>	68%	30%	2%	95
<b>Reproducible research (writing, storing and sharing code)</b>	53%	39%	8%	95
<b>Complex samples and weights</b>	51%	42%	7%	94

## Feedback on the Using Survey Data Guide

Many viewed the Using Survey Data guide as a useful resource and some were signposting students towards it. It was also suggested the usefulness of the guide extends beyond dissertations as the content is relevant to undergraduate courses. Many commented on the usefulness of a supporting worksheet shown in the session, suggesting it was a good approach and something that students could do before meeting a supervisor. Not everyone who thought the content was useful indicated they would use it as they prefer to integrate contents into their teaching.

A repeated critical evaluation was that there is too much content in the guide. Related to this evaluation, some identified some content as unnecessary (e.g. levels of measurement) as students know this content already – or “if they don’t, they won’t learn from a guide like this”. There was also the suggestion that the content was too advanced for undergraduates. However, views on the level of content contrasted for example one interviewee indicated the guide goes too quickly into weighting and another that weighting is buried.

A common suggestion was that the most important topics centred around how to find and access data for a dissertation from the UK Data Service. Many of these issues are reflected in the comment here, which also include the suggestion to “chunk it”.

**" too much information, cut it down to bare bones, chunk it. How to access and register. Later parts redundant as covered on course. (Interview)**

Dividing content into small chunks could address most of the issues above as lecturers can more easily signpost students to relevant parts, which could also be more easily integrated into lecturers’ other materials.

Suggestions for further topics were

- transparency and reproducibility include code files
- finding comparative data.
- what data looks like from the UK Data Service
- missing data
- assessing variables – how to deal with data when it’s not as expected e.g. skewed and content on manipulating data e.g. recoding data.
- data management

## Dissertation case studies

In the survey, several comments mentioned examples and case studies as potentially useful. These comments included the suggestion for resources covering the whole research process such “from research question to selected dataset” and “examples of dissertations that are data driven.”

In the interview, we sought feedback on a case study of undergraduate dissertation students. Many interviewees were positive about the dissertation case studies, with positive evaluations of the structure and content. Case studies in general were seen as a useful tool for encouraging students and some produce similar resources for their students or share real examples of dissertations. In these, some suggested they would not need to use UK Data Service case studies and suggested they would be more useful for lecturers and departments where few students do secondary quantitative analysis. Some suggested creating more and covering more topics and disciplines.

More critically, some suggested including more information would increase their value, especially in relation to the methods used and the findings. For instance, rather than just indicating they used regression, it would be good to describe how many models were included. This additional detail can help students understand the scope of a dissertation using secondary analysis of quantitative data.

The representativeness of the dissertation case studies was also raised as an issue. For instance, some noted how the one shown in the interviews was from a previous UK Data Service award winner at the University of Cambridge. While the full range of case studies is more diverse, several emphasised the need to “showcase the normal ones” and “Be kind of achievable, otherwise they’re counterproductive.”

## Dissertation award

A few interviewees commented on the UK Data Service dissertation award (though the topic was not part of the interview schedule). In these instances, the award was viewed as positive in general but with some problems. For one, the issue is around self-nomination, as they feel students doing good work are not always very aware of how good their work is, so it is hard to get them to enter. More generally some suggested the award acknowledge students working at a higher level than their own students. It was also suggested that it is not an even playing field as some institutions/individuals (such as those on Q-Step courses) have an advantage.

## Further suggestions for resources

From the interviews and responses to a survey question, we can identify several further suggestions about what would be most useful for dissertation students.

### Promotion of secondary analysis

One of the most common suggestions was for resources to help address the need to “overcome the barrier of students worrying it will not be as good to use existing data or that they will be marked down for not collecting data when the opposite is probably true”. Suggestions in relation to this need often focused on the idea resources that simply and quickly convey the possibility of secondary analysis. For instance, one described

**“the most important thing for dissertation students is to know they can use the data for their dissertations and a quick way of saying that is probably the most important thing. A big encouragement”.**

Another interviewee suggested that a short video could be effective way of promoting this message.

## Quick start guides to the major datasets

Another repeated suggestion was for resources that promote and introduce some of the main datasets. Several thought there could usefully be short orientation guides to the major surveys likely to be used in dissertations, with one commented that such a resource could be useful for other data users as the getting familiar with surveys is time-consuming. There could also be resources making connections between topics and datasets. These resources could help promote the value of datasets and secondary analysis.

## Finding and accessing data from the UK Data Service

In line with discussions in the interviews, many survey respondents perceive a need for resources that help students with finding and accessing data from the UK Data Service. Some of these comments referred to improving the navigability of the UK Data Service website and catalogue while others made suggestions about ways to inform students about how to find and evaluate data. Specific suggestions included:

- introductory lectures giving examples of potential suitable data
- a guide specifically from the student perspective
- how to evaluate datasets
- help to make sense of the many versions of datasets that are available
- case studies – from research question to selected dataset to results

Several comments also highlighted information about other publicly available data including international data.

## Weighting and complex samples

We have some comments referencing resources on weighting and complex samples; in one case, this was connected to information for specific surveys

**Practical stuff on complex samples - we do our best with this but every survey is different and presents its own challenges**

## Data skills development

Among the survey comments, several indicate a need for data analysis skills development. Many focused on basic data handling and presentation skills as well as statistical methods. For example,

**The level of sociology / criminology students who use quants in dissertations is relatively low. They need the basic stuff on how to present data, how to manipulate data in SPSS, but also excel. How to label tables/charts and what's needed to tell an appropriate story... much more than some of the more complex stuff”.**

**more resources/examples of how to run statistical tests such as chi-square / t-test / OLS / logistic regression using UK datasets available from the Data Service**

**in depth tutorials on Nesstar and downloading & pre-processing data / variables**

In some cases, suggestions related programming skills for data collection, management or analysis. For instance,

- examples of linking packages (e.g. Knime) with raw languages (e.g. Python) - and passing data in both directions
- more on webscraping - I do A LITTLE on this but there is a lot of demand
- more resources on how to collect their own data, especially in relation to text analysis (both manual annotation and machine learning models).

## Example code files

There were also suggestions around example code files for doing data handling and analysis with UK Data Service data.

- standard syntax for file merging BSA / USoc / LCFS etc would be v useful
- example of codes and scripts with datasets to illustrate methods



## Computational social science and reproducibility

Computational social science and reproducibility and the transparency of research are two growing themes relating to the use of data in the social sciences. As part of the consultation, we examined whether topics in these areas were being taught and lecturers' views about teaching such topics and any related need for resources.

### Reproducibility and transparency of research

The survey asked questions about topics relating to the reproducibility and transparency of quantitative research including good practice in writing and storing code files and tools available to help. Overall, under half of those surveyed reported teaching these topics (43 percent). The proportion was lower among those teaching introductory topics at 37 percent and higher among those who taught advanced topics at 55 percent.

*Table 9 Currently teaching reproducibility and transparency of research*

All		Type of teaching		
		Advanced or (introductory/advanced)	Introductory only	Non-methods courses only
<b>Yes</b>	43%	55%	37%	17%
<b>No</b>	57%	46%	63%	83%
<b>Total</b>	99	44	49	6

Though not all were teaching social science students about reproducibility and the transparency, we find that most of those surveyed agreed these are important topics to teach. Table 10 shows that 67 percent responded 'definitely yes' and a further 25 percent 'probably yes'. Even among those not currently teaching these topics, 86 percent responded either definitely or probably yes. A majority also indicated that demonstrations and examples of how to teach these topics would be useful (88 percent either 'probably yes' or 'definitely yes'). Those currently teaching these topics were more likely to report that such resources would be useful; however, even among those not currently teaching these topics, 85 percent indicated yes.

*Table 10 Is it important to teach about reproducibility and transparency of research and would demonstrations and examples be useful, by whether currently teaching these topics*

	Is it important to teach these topics to social science students?			Demonstrations and examples of how to teach these topics would be useful.		
	All	Currently teaching these topics	Not currently teaching these topics	All	Currently teaching these topics	Not currently teaching these topics
<b>Definitely yes</b>	67%	95%	45%	54%	74%	39%
<b>Probably yes</b>	25%	5%	41%	34%	19%	46%
<b>Probably not</b>	7%		13%	11%	7%	14%
<b>Definitely not</b>	1%		2%	1%		2%
<b>Total</b>	99			100		

Table 11 shows counts by discipline of those teaching these topics, those indicating yes to whether these topics are important to teach and those indicating examples and demonstration would be useful; the latter two categories include both 'probably yes' and 'definitely yes'. We cannot make inferences about disciplines in general as we do not have a representative sample and numbers from some disciplines are small. However, we can see that those teaching these topics come from a variety of disciplines and teaching staff across disciplines view these as important topics to teach and indicate that examples and demonstrations on how to teach these would be useful.

*Table 11 Reproducible research, numbers currently teaching the topics and numbers who indicate the topics are important to teach and that examples and demonstrations would be useful for teaching*

	Currently teach these topics	Important to teach	Examples and demonstrations would be useful	n
<b>Politics</b>	7	17	16	19
<b>Sociology</b>	15	26	25	28
<b>Criminology</b>	1	5	5	5
<b>Economics</b>	4	9	9	10
<b>Psychology</b>	1	5	5	5
<b>Geography</b>	5	7	7	9
<b>Other</b>	9	21	20	23
<b>Total</b>	42	90	87	99

## Computational methods

The survey asked about computational methods such as using programming to collect, combine, synthesise or process data, citing the examples of web-scraping, text-mining, and machine learning. In Table 12, we see just under 20 percent (19/98) currently teach these topics, rising to 30 percent among those teaching more advanced topics.

*Table 12 Currently teaching reproducibility and transparency of research*

All		By type of teaching		
		Advanced or introductory and advanced	Introductory only	Non-methods courses only
<b>Yes</b>	19%	30%	13%	
<b>No</b>	81%	71%	88%	100%
<b>Total</b>	98	44	48	6

A larger proportion indicate that these topics are important to teach. For instance, in Table 13, we see 43 percent respond ‘definitely yes’ and 41 percent ‘probably yes’. In a similar way to responses about reproducible research, we find that a majority indicate that demonstrations and examples of how to teach these topics would be useful but there is more interest among those already teaching the topic. For instance, when it comes to whether examples and demonstrations would be useful 90 percent of those teaching the topic indicate ‘definitely yes’ compared to 37 percent among those not teaching the topic.

Table 14 shows counts by discipline of those teaching these topics and of those indicating yes to whether these topics are important to teach and to whether examples and demonstration would be useful; the latter two categories include both ‘probably yes’ and ‘definitely yes’. We find that those teaching these topics come from a variety of disciplines. As we do not have a representative sample, we cannot make inferences about what staff in each discipline. But like with reproducible research, many indicate that computational social science topics are important to teach social science students and see demonstrations and examples of how teach as being potentially useful. This pattern applies even in politics, where none of those surveyed currently teach these topics.

*Table 13 Is it important to teach about computation social science and would demonstrations and examples be useful, by whether currently teaching these topics*

Is it important to teach these topics to social science students?				Demonstrations and examples of how to teach these topics would be useful.		
	All	Do you currently teach any of these topics?		All	Do you currently teach any of these topics?	
		Yes	No		Yes	No
<b>Definitely yes</b>	43%	95%	30%	47	90%	37%
<b>Probably yes</b>	41%	5%	49%	40	11%	47%
<b>Probably not</b>	14%		18%	10		13%
<b>Definitely not</b>	2%		3%	2		3%
<b>n=</b>	98	19	79	97	19	78

*Table 14 Computation social science by discipline, numbers teachers and consider demonstration and examples useful*

	Currently teaches these topics	Important to teach these topics	How to teach demonstrations and examples would be useful	n
<b>Politics</b>	0	15	17	18
<b>Sociology</b>	6	26	26	28
<b>Criminology</b>	1	4	5	5
<b>Economics</b>	1	8	8	10
<b>Psychology</b>	1	5	5	5
<b>Geography</b>	2	7	7	9
<b>Other</b>	8	16	16	23
<b>Total</b>	19	81	84	98

## Views in relation these topics

While the survey did not include the option to directly comment on these topics, several respondents used the general open questions in the survey to express views relating to these topics.

### Training for teaching staff

One theme from the comments concerns the need for training for teaching staff relating to reproducibility and computational techniques. Several participants directly mentioned the need for training for themselves:

- Much more on reproducibility, also use of GitHub etc – I am very behind on this
- Resources on reproducibility
- Teaching reproducibility
- Computational techniques

More generally, other respondents referring to the need for introductory training in relation to R and how to teach it.

### Programming versus statistical literacy

A further theme within the comments is that the inclusion of programming, as well as more advanced statistical topics, may come at the expense of statistical literacy. For example, one respondent commented:

**I am concerned about the direction in which the teaching of statistical analysis has been headed in the last few years. There seems to have been a fetishization of programming languages and a focus on advanced techniques at the expense of basic statistical literacy.**

They then proposed that a greater emphasis on programming and advanced methods may 'put off' some students from any statistical analysis and increase the knowledge gap among students.

**UG courses increasingly ask students to learn R or Python as a prerequisite to conducting any statistical analysis. This will doubtless put many students off continuing with any statistical analysis and could deter students from applying to some of the 'with statistical methods' courses that have resulted from the Q-Step initiatives. Knowledge of these programming languages will only be a transferable skill for a**

**small minority of graduates but the 'overhead' in terms of teaching time devoted to learning programming is considerable.**

An alternative proposed, by this respondent and others is a broad focus on statistical literacy rather than programming and advanced techniques within the 'finite amount of teaching time available'. Similar themes are found in other comments that emphasise the challenge of getting students interested in quantitative methods.

**The biggest problem in teaching quantitative methods is getting social science students interested in the first place and making them accessible. Quantitative methods are generally seen as hard and mathematical and the fetishisation of techniques and software in this field is off-putting for many students. To be clear, I speak as someone who is quantitatively minded both in research and teaching. I am not anti-quants. But the focus needs to be placed far more on getting students interested initially, not entirely at high level skills and packages.**

**Ultimately, I think the most important thing is to engage students with analysing data. As such, excessive emphasis on method or technique, at least in the first instance, can lead to students acquiring technical skills in a rather abstract way, but lacking adequate experience with actually engaging with data. This in turn can lead to technically competent analyses which are more problematic in terms of the way in which data/findings are handled/interpreted.**



[www.ukdataservice.ac.uk](http://www.ukdataservice.ac.uk)

[help@ukdataservice.ac.uk](mailto:help@ukdataservice.ac.uk)

+44 (0) 1206 872143

We are supported by the University of Essex, the University of Manchester, UKRI through the Economic and Social Research Council, and Jisc