

An enquiry into the use of numeric data in learning & teaching

September, 2001

Robin Rice (Project Manager)
Edinburgh University Data Library

Peter Burnhill (Project Director)
Edinburgh University Data Library, EDINA

Melanie Wright (Project Team)
The UK Data Archive

Sean Townsend (Project Team)
British Library of Political and Economic Science

Part 1: Report and recommendations

Contents

Executive summary	3
Introduction and participants	4
Aims and objectives of the project	5
Nature of the enquiry (or why it matters)	5
Methods used in the enquiry	6
Discussion of results	7
Recommendations and conclusions	17
References	20
Appendix A: Related developments	21
Appendix B: Case studies of current practice	23

Part 2: Teachers' survey results 41

Executive summary

Within UK higher education the renewed attention to *learning and teaching* is an impetus for change. Advances in information technology create new space for learning beyond the traditional classroom lecture format. New initiatives are creating networked teaching materials for shared use across institutions. But little is known about the readiness of teachers and students to take advantage of these resources for teaching and study. Are universities providing the support needed for using these networked resources in classrooms, computer labs, and independent study?

An academic Task Force on the use of numeric data in learning and teaching has issued a report on the barriers faced by teachers and students to using national data services across a number of disciplines, including but not limited to the social sciences. The enquiry focused on numeric data, which involves a higher number of skills to use than many other types of information resources. Results were analysed from a national survey of teaching departments in universities, and seven case studies of real-life teaching scenarios in both post- and undergraduate classes in several disciplines. The Task Force contributed views from their own significant experience of teaching in academia as well.

The project is part of a national development programme on learning and teaching funded by the JISC (Joint Information Systems Committee). Its unique focus within the set of projects is on the value of introducing statistical data such as area census statistics, sample survey datasets, and economic trend data to the educational experience of students, particularly when students actively take part in analysing the data, and practice drawing conclusions from empirical evidence.

The enquiry found that despite established use of quantitative secondary analysis of national datasets in research, a number of issues make its use in teaching and students' independent study difficult, and therefore rare. Whilst print tables and graphs are often used by lecturers in teaching empirical subjects, statistical files requiring 'hands-on' computer analysis are not commonly built into the teaching design, except in methods courses. Yet these are transferable skills needed by today's graduates to enter the professions or advanced study.

Only one-quarter of survey respondents who said they used data in the classroom had considered using the nationally funded academic data services provided by the Data Archive (at Essex), MIMAS (at Manchester), or EDINA (at Edinburgh) as a source of the data used in their teaching. The survey uncovered a number of barriers experienced by teachers in the use of these services, namely a lack of awareness of relevant materials, lack of sufficient time for preparation, complex registration procedures, and problems with the delivery and format of the datasets available. These problems were elaborated in open-ended comments by respondents and in the case studies of current teaching practice.

A compounding problem is the lack of local support for teachers who would like to incorporate data analysis into substantive courses. A majority of the survey respondents said that the level of support for data use in their own institutions was ad-hoc. Peer support was more common than support from librarians and computing service staff, and over one-third received no support whatever. The top three forms of local support needed were data discovery/locating sources, helping students use data, and expert consultation for statistics and methods (for staff).

The Task Force analysed the results of the survey and the experiences expressed in the case studies and issued recommendations for UK higher education, summarised below:

1. A broad initiative is recommended to promote subject-based statistical literacy for students, coupled with tangible support for academic teaching staff who wish to incorporate empirical data into substantive courses.
2. The development of high-quality teaching materials for major UK datasets must be funded adequately, in order to provide salience to subject matter and demonstrate relevant methods for coursework.
3. The national data services need to improve the usability of their datasets for learning and teaching.
4. A more concerted and co-ordinated promotion of the national data services could then follow, which is responsive to user demand.
5. Universities should develop IT strategies that include data services and support for staff and students, and integration of empirical datasets into learning technologies.

A limited number of free copies of this report are available by request; contact datalib@ed.ac.uk. In addition, the complete survey findings, case studies, and other project information are on the Web, at <http://datalib.ed.ac.uk/projects/databeach.html>.

Introduction

An enquiry into the use of numeric datasets in learning and teaching within UK higher education, sponsored by the JISC (Joint Information Systems Committee), has been looking into barriers faced by post- and undergraduate teachers who wish to introduce students to the use of empirical datasets in the classroom.

UK higher education is rich in numeric datasets. In the socioeconomic field, for example, there are large-scale government surveys (e.g. the General Household Surveys), current and historical population censuses, international studies, academic studies, economic time series, and geographic data. These numeric data are disseminated in ways that support the work of many academic researchers, including some postgraduate students.

However, these data resources are under-used in the learning and teaching environment. Despite potential gains in numeracy, critical use of evidence, and empirically-grounded knowledge by students conducting data analysis at both the postgraduate and undergraduate levels, obstacles exist that make integration of numeric data resources into coursework difficult. As expectations about use of information technology in learning and teaching rise, the barriers which inhibit the use of this wealth of data in the classroom and in student projects need to be lowered.

Making effective use of numeric data in teaching requires specialised skills and more time for preparation than the use of printed materials or bibliographic databases, and both students and teachers require a high level of support. Whilst some problems can be solved by data providers or by a national approach to supporting learning and teaching, others need to be tackled at the local level.

Project participants

This project is funded by the Joint Information Systems Committee (JISC), as part of the DNER Development Programme for Learning and Teaching—see <http://www.jisc.ac.uk/dner/programmes/projects> for further information. (DNER stands for Distributed National Electronic Resource, “a managed environment for accessing quality assured information resources on the Internet which are available from many sources,” as defined on the JISC website).

Project Partners

Project Director:

Mr. Peter Burnhill, Director Edinburgh University Data Library / EDINA

Project Manager:

Ms. Robin Rice, Data Librarian Edinburgh University Data Library

Ms. Melanie Wright, Director Depositor Services The Data Archive, University of Essex

Mr. Sean Townsend, (formerly) Data Librarian London School of Economics

Mr. Keith Cole, Manager MIMAS, University of Manchester

Task Force

Chair:

Professor Peter Elias Institute of Employment Research, University of Warwick, Economic and Social Research Council (ESRC) Research Resources Board

Dr. Nina Bullen MIMAS, Manchester Computing, University of Manchester

Dr. Mark Brown Cathie Marsh Centre for Census and Survey Research, University of Manchester

Professor John Curtice Dept. of Government and Director, Social Science Research Laboratory, University of Strathclyde

Professor Ian Diamond Department of Social Statistics, University of Southampton, Chair, ESRC Research Resources Board

Ms. Kristina Drew Information Systems Unit, Institute of Child Health, London

Dr. Jane Fielding Department of Sociology, University of Surrey

Mr. Peter Lee	Centre for Urban and Regional Studies, University of Birmingham
Dr. Graeme Morton	Dept. of Economic & Social History, University of Edinburgh
Professor Clive Payne	Social Studies Faculty Centre, University of Oxford
Mr. Eric Tanenbaum	Dept. of Government and Director, Summer School in Social Science Data Analysis and Collection, University of Essex
Dr. Richard Wiggins	Department of Sociology, City University, London

Aims and objectives of the project

This project was designed to generate knowledge about good practice and pitfalls faced by teachers and learners who use national data resources and those who support them. The project partners are staff members of two academic data libraries (at the University of Edinburgh and the London School of Economics) and three national data centres—the Data Archive, EDINA (Edinburgh Data and Information Access), and MIMAS (Manchester Information & Associated Services).

The aims of the project were:

- to ascertain how to lower barriers to the use of numeric data in learning and teaching, and;
- to determine ways to foster and promote effective local data support services able to work independently and in effective collaboration with the national data services.

To achieve these aims, an enquiry was launched in February 2000. The following objectives were pursued:

- Set up a diverse Task Force made up of academic teachers and user supporters.
- Open a public electronic forum for solving problems in the use of data in learning and teaching. Use the email list, the project website, academic conferences and newsletters to disseminate project findings.
- Undertake a national survey of teaching departments to discover the current state of the use of numeric data in learning and teaching.
- Develop at least six exemplary case studies on the use of numeric data in the classroom in different disciplines and levels of education.
- Analyse the results from the studies, and in consultation with the Task Force, deliver a report of recommendations to JISC and other relevant bodies.

This document is the report mentioned in the final objective. Additional information can be found on the project website: <http://datalib.ed.ac.uk/projects/data teach.html>.

Nature of the enquiry (or why it matters)

Underlying this project are two assumptions that perhaps demand clearer articulation. The first is *the importance of teaching students to understand numeric data*. To many this may seem so obvious as to be trivial—the careers students may be entering could require quantitative skills, and there is a significant quantitative element to many degree programmes as a matter of course. But this project starts from the viewpoint that teaching hands-on quantitative skills is important not just as career preparation for those entering quantitative disciplines and posts, but is also an important part of creating critically numerate citizens, irrespective of their career goals.

‘Evidence-based practice’ and ‘evidence-based policy’ are currently steering significant portions of the national agenda—in medicine and health policy, in government and social policy, and in research across many disciplines. Strengthening the evidence base for decisions is seen as an intrinsic positive value, whether in clinical medical practice, or in formulation of social services policy and programmes. Accompanying this is an emphasis on engaging more fully with the end users of this evidence—public policymakers, voluntary organisations, and the press—not just with other researchers.

The research councils themselves have adopted this agenda. For example, in February 2001, the ESRC issued revised postgraduate training guidelines in an effort to address a perceived “deficiency in the research skills in some of the UK’s social science disciplines”. According to the ESRC Chief Executive, “British universities and colleges are not producing quantitatively competent social scientists in sufficient numbers.”¹ A number of social

science disciplines were named as having a “shortage of appropriately skilled analytical researchers,” and it was argued that the UK was slipping behind other countries competitively, as a result.

Of course, evidence-based policy and practice are only as good as the quality of the underpinning evidence, and the ability of those formulating the policy or practice to understand that evidence and evaluate and interpret it. To achieve this objective requires not just the training of a new generation of researchers who will generate quality evidence, but also a knowledgeable and critical body of consumers of research, in all domains. Since much of this evidence is quantitative in nature, teaching students to handle quantitative material in a hands-on fashion is an important part of that training.

The skills involved in statistical literacy are different than those involved in producing or ‘doing’ statistics, Milo Schield has found.² “Statistical literacy focuses on making decisions using statistics as evidence just as reading literacy focuses on using words as evidence. Statistical literacy is a competency just like reading, writing, or speaking.” Perhaps the goal should be statistical *literacy* for undergraduates, and statistical *competence* for postgraduates, at least in certain fields.

But further than simply advocating statistical literacy, this study is underpinned by a further assumption, that *the use of secondary datasets in teaching is important*. The reasons for this are manifold, not least of which is effective use of resources. Each year the government (both directly and via the research councils) expends millions of pounds collecting numerical data of one sort or another. Usually this data collection is quite narrowly focused on a particular research question. But data are frequently multi-purpose, and can be put to new and novel analyses. Original data collection, particularly when large samples are needed, is expensive; secondary analysis is comparatively cheap.

Secondary analysis is valuable in other ways beyond getting the most out of an investment. Knowledge is cumulative. The more secondary datasets that exist and are accessible, the more new research can be framed to build upon it. One of the most exciting methodological areas in development is that of data harmonisation and data merging (meta-analyses). When new research is constructed that is methodologically consistent with previous studies, disparate research can be brought together, and the greater their value becomes. Similarly, quantitative datasets can be used as benchmarks for other studies, including those with qualitative designs. But this value added to research can only occur if the user knows both about the availability of a dataset, and how to use it.

Likewise, the practice of secondary analysis may actually improve the quality of original data collection. The awareness that a collected dataset will be preserved and made available to other researchers to use, validate, or criticise, may well lead to increased care in the collection, construction, and documentation of the primary research.

But use of secondary data provides other attractions in the classroom. Using real data to add empirical evidence to a particular subject area can be particularly engaging to students. Quoting numbers in a vacuum may be far less interesting to students than discovering for themselves where these numbers came from, as well as the pitfalls of their creation and interpretation. Likewise, in pure methods teaching, the use of ‘real world’ data to teach the techniques used by real analysts can add a frisson of relevance to otherwise intrinsically rather dry material. This may be particularly true for ‘service courses’ taught outside of the students’ chosen subject area—a common practice in the teaching of Statistics.³

Fundamentally then, this study assumes that the use of numerical data in learning and teaching is an intrinsically positive value. The survey was designed to find out the current state of play in the use of numeric data in teaching and supervising students, users’ experience of the national data services, barriers to using data in teaching, and the extent of support available within their institutions. The case studies are intended to provide some examples of good practice which can illustrate effective use as well as uncover barriers to more effective use.

Methods used in the enquiry

A major objective of the project was to generate knowledge on issues such as the extent of use and the practicalities of using data in teaching, and the experiences teachers have of data support from both national data services and support staff in local institutions. Since user surveys tend to target those already registered for national services, there is no ready evidence about the larger population of UK university teaching staff on these issues. Therefore, a nationally representative sample survey was needed to discover the current “state of play” before recommendations about how to lower barriers could be made.

Teachers' survey

A sample postal survey was conducted of UK university teaching departments within the social sciences, plus other selected disciplines "outside" the social sciences, such as public health sciences. Two hundred and sixty-seven department heads were randomly selected from a universe of 1590 (1 in 6 sampling fraction). The sampling frame was purchased from the marketing company Mardev, extracted from the *Worldwide Academic & Library File*. Department heads were asked to complete the four-page questionnaire themselves and to pass copies to relevant teaching colleagues to garner their participation. (A Web version was also made available for on-line input.) There were 206 responses collected from 110 departments. Fifteen records were removed as ineligible (e.g. non-teaching department). Following telephone, e-mail, and postal follow-up requests to sample members, the final response rate (110 / 252) was 44 percent of departments sampled.

Heads of departments and lecturers were asked about their use of numeric data in teaching and supervising students, their experience of national data services, perceived barriers to using data in teaching, and the extent of support available within their institutions for using datasets in teaching. The complete findings, survey instrument, and other appendices are in the related document, *Teachers' Survey Results*, included as Part 2 of this report, and available on the Internet at <http://datalib.ed.ac.uk/projects/datateach/findings>.

Case studies

As part of the project design, case studies were conducted during Spring 2001, providing a handful of realistic 'snap-shots' of current practice in using numeric data in learning and teaching within higher education. Members of the Task Force authored the case studies themselves, thereby enhancing their own personal knowledge on the topic with an objective study of another teacher's experience. 'Subjects' for the case studies were volunteers selected by the Task Force, some of whom were survey respondents. In all cases an interview between author and subject took place, and where possible a site visit to the institution and an observation of the class in action.

There were two aims in conducting the project case studies:

1. To complement the empirical survey results by providing a richer, qualitative picture of data use in classroom activities in a range of situations.
2. To provide exemplars for the academic community, available through the project website, for teachers to explore new ideas for integrating the use of data in a variety of disciplines and course levels.

The written case studies are appended to this report, and are also available on the Internet, at <http://datalib.ed.ac.uk/projects/datateach/casestudies.html>.

Discussion of results

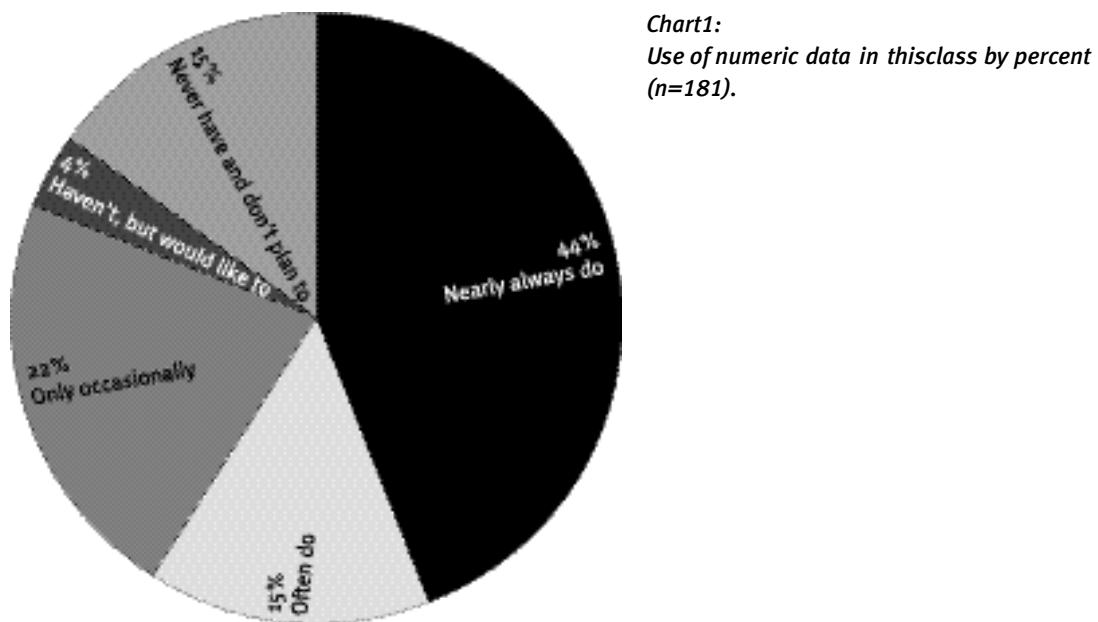
The *Teachers' Survey Results* document and the seven case studies provide a full picture of what was discovered by the Task Force's enquiry. In this section we draw on those materials as evidence to inform our recommendations. This discussion is organised under the following eight topics:

- 1) Use of data in teaching
- 2) Use of data in learning
- 3) Data sources and use of national data services
- 4) Pedagogical purposes
- 5) Perceived burden of data preparation
- 6) Sources and levels of data support
- 7) Forms of local support needed
- 8) Perceived barriers to use of national data services.

N.B. To avoid confusion, the charts and tables provided below retain the same numbering sequence as in the *Teachers' Survey Results* document (part 2 of this report). Case studies are referred to by the author's surname and are included as an appendix to this report.

1) Use of data in teaching

Due to the survey design and instructions to department heads, there was likely a skew toward data users among those in the sample who participated, as a result of self-selection. (Non-data users tended not to respond to the survey, as it was not felt to be relevant to them.) Seventy-nine percent of those survey respondents who taught or convened courses used data either “nearly always,” “often,” or “occasionally” (see Chart 1).



The sample also seemed to over-represent senior staff (perhaps because the request was sent to department heads), those particularly committed to quantitative analysis, and teachers of methods courses. (Appendix 3 shows course names given by respondents and how they were categorised into methods versus subject-based courses.)

As Table 9 shows, methods-based research courses across disciplines made an even higher use of numeric data (95% at least occasionally, and 72% ‘nearly always do.’) Subject-based courses showed fairly high usage, but a more casual approach (76% at least occasionally, but only 34% ‘nearly always’). Again, these results are likely to be higher than the actual population of teachers.

Table 9: Use of numeric data in class by course type.

	Col %	Methods	Subject	All
Nearly always do	72	34	44	
Often do	17	15	16	
Only occasionally	6	27	21	
Haven't, but would like to	0	6	4	
Never have and don't plan to	4	18	15	
n =	47	132	179	

Among those who used numeric data in teaching in some form, about two-thirds expected students to work with data on a computer, ‘hands-on’. Not surprisingly, a higher proportion of methods courses were hands-on than subject courses (85% vs. 54%). However, undergraduates were given hands-on coursework about as often as postgraduates (see Table 12), and hands-on work was also spread equally among class sizes (although the instrument failed to distinguish among class sizes larger than 40).

Table 12: Whether course is hands-on, by course level.

	Undergrad	Postgrad	All
Hands-on	63	64	63
Not hands-on	37	36	37
n=	100	47	147

The Task Force has taken the ‘hands-on’ filter to be a likely indicator of more intensive learning of data analysis skills by students. It is also more intensive of resources in terms of facilities, staff time for preparation, and direct support to students, especially when large classes are broken up into several lab sessions.

2) Use of data in learning

Although the survey was directed towards staff, not students, there was an attempt to understand the level of data use by students in their independent learning. Ninety-two percent of those who were supervisors recommended the use of numeric data for student’s research at least occasionally (depending on the nature of the research project). Table 22 shows only a slightly higher tendency to recommend use of datasets to postgraduate students than undergraduates.

Table 22: Recommend use of numeric data by level of students supervised.

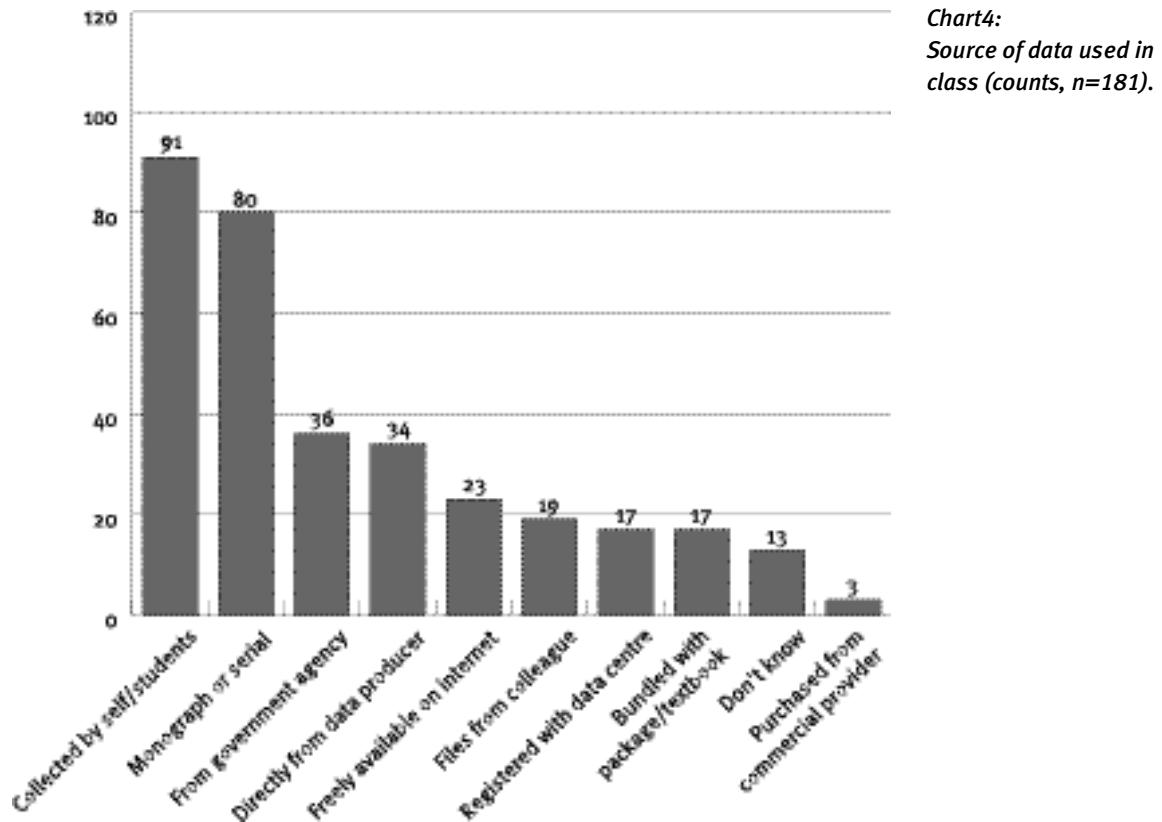
Col %	U'grad only	P'grad only	Both	All
Nearly always do	36	41	32	36
Often do	36	31	34	34
Only occasionally	12	16	28	22
Never have and don't plan to	6	10	5	7
Haven't yet but would like to	9	2	0	2
n =	33	51	96	180

The open-ended comments shed more light on teachers’ perception of this area, such as, “Many students are more inclined to qualitative research.” (Appendix 5 groups all the comments to this question.) But the survey instrument does not provide a clear picture of how students are guided in the search for data sources for their individual chosen topics. Recent research shows that students approach their tutors for help with sources more than anyone else, including librarians.⁴ There is room for concern about missed opportunities here, if tutors are not well-versed in ready sources of empirical data themselves.

3) Data sources and use of national data services

The survey showed quite clearly that, although the use of numeric data in higher education teaching and learning is widespread, the use of national data services that provide on or off-line access to secondary datasets is not. *Only one-quarter* of the respondents had “used or considered using” the national academic data services (namely the Data Archive, EDINA, and MIMAS) for teaching purposes. Somewhat discouragingly, the number of years experience in teaching the course did not affect this. Higher levels of use were found for heavy data users and methods course instructors, whilst casual data users and those teaching subject courses were less inclined to have considered accessing national data services.

So what are the sources of numeric data used in higher education classes? There was a wide variety, including government agencies, directly from the data producer or a colleague, freely available on the Internet, or bundled with a textbook, but each of these sources were used by less than 20% of teachers who use data (see Chart 4). Nearly half, 44%, used print data sources too, as in a monograph or serial. This reflects the relative ease with which print sources can be located and obtained by teachers, although as noted, these are less likely to involve hands-on work. This finding also seems to undermine the popular notion that anything needed can be found for free on the Internet.



Half the teachers either had the students collect their own data, or taught with data collected by themselves (see Chart 4). This reflects two important realities of current teaching practice, as supported in the open-ended comments and the case studies. One is the belief in the educational value of having students go through a research project from start to finish, or 'learn by doing.' Students are frequently asked to collect numerical data, perhaps in the form of surveys, and also qualitative data, through focus groups, in-depth interviews, or case studies, and—particularly in Psychology courses and the physical sciences—experimental data as well. The Task Force agreed this was an important way for students to learn about the research process. However, they also felt strongly that it was not always a sufficient way to learn how to analyse data, since data collected by students would tend not to be robust and would be prone to error. The value of using large-scale survey data, population census data, and economic time series provided by the national data services and elsewhere is that they can provide solid, reliable, and representative benchmarks for practising statistical analysis techniques and making evidence-based statements about society.

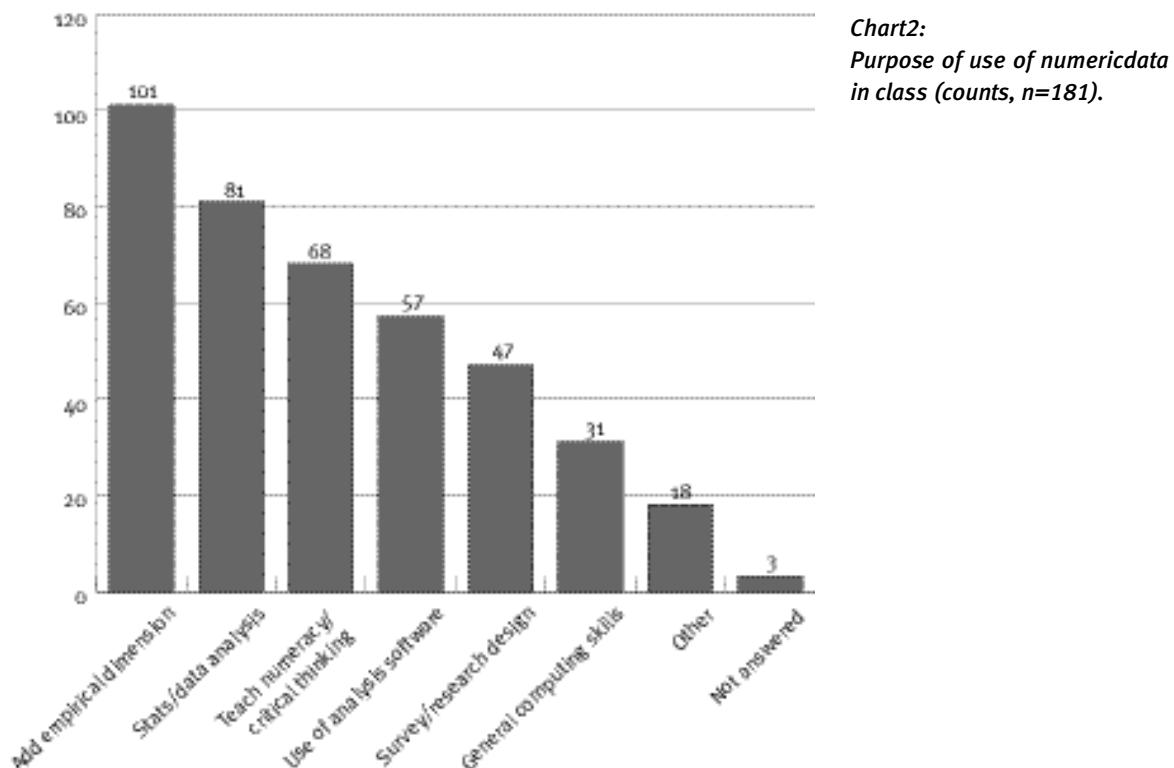
The second reality about current teaching practice which these results show is the continuing traditional belief in the educational value of lecturers using their own research material as the object of student learning. In a sense there is a perceived economy of effort achieved by academic staff in their dual roles of researcher and teacher. Additionally, the in-depth knowledge that the researcher gains through intensive familiarisation with an object of study, and the experience of grappling with real-world research problems, is transferred to students, as in a mentoring situation.

Although this tradition remains embedded in university teaching, a number of new realities are bringing the sense of this time-honoured paradigm into question. First, the increase in numbers of students enrolling in higher education means that, in addition to large class sizes and a more heterogeneous student population, a smaller proportion of students continue in an academic career, but need skills for other professions, including other research-related careers outside of academia. Second, the renewed attention on university teaching in its relation to research, and efforts to reward dedicated teachers as well as successful researchers in universities may lead to a greater division of labour between teachers and researchers. Third, a learner-centred philosophy of education is gaining prevalence, which gives the teaching a backseat to the process of learning itself, and may emphasise facilitating students' own educational goals and choices through a variety of learning activities other than the traditional lecture. And fourth, the rising expectations for learning technologies to enhance the educational experience of students, both in distance learning and in traditional university settings may have a large affect on the content of what is taught.

Academics need to come to grips with these changing realities and values in higher education as they adapt and evolve their teaching practices.⁵ The Institute for Learning Technology (ILT) and the new Learning and Teaching Subject Network (LTSN) are providing space for these sorts of debates within and between disciplines in the academic community.

4) Pedagogical purposes

It is important to understand the purposes of use in order to provide relevant support to teachers as they pursue their own pedagogical aims. Most teachers reported more than one purpose from the list provided in the questionnaire. As Chart 2 illustrates, the top three purposes arising were 'To add an empirical dimension to the subject'(56%) 'To teach statistics or data analysis methods'(45%) and 'To teach numeracy or critical thinking skills'(38%).



Again, the two types of courses differed in terms of the primary purpose for which data was used into the classroom (see Table 11). Not surprisingly, teaching statistics or data analysis methods was primary for the methods classes (71%), but adding an empirical dimension to the subject was a primary purpose for 41% of the subject-based courses. Again, course type proved to be more of a distinction than level of course (post or undergraduate) or department type (inside/outside Social Sciences). Developing the numeracy or critical thinking skills of students was a secondary purpose of both in over a third of responses. These various purposes have important implications for the development of teaching resources and support services.

Table 11: Primary reason for use of numeric data by course type.

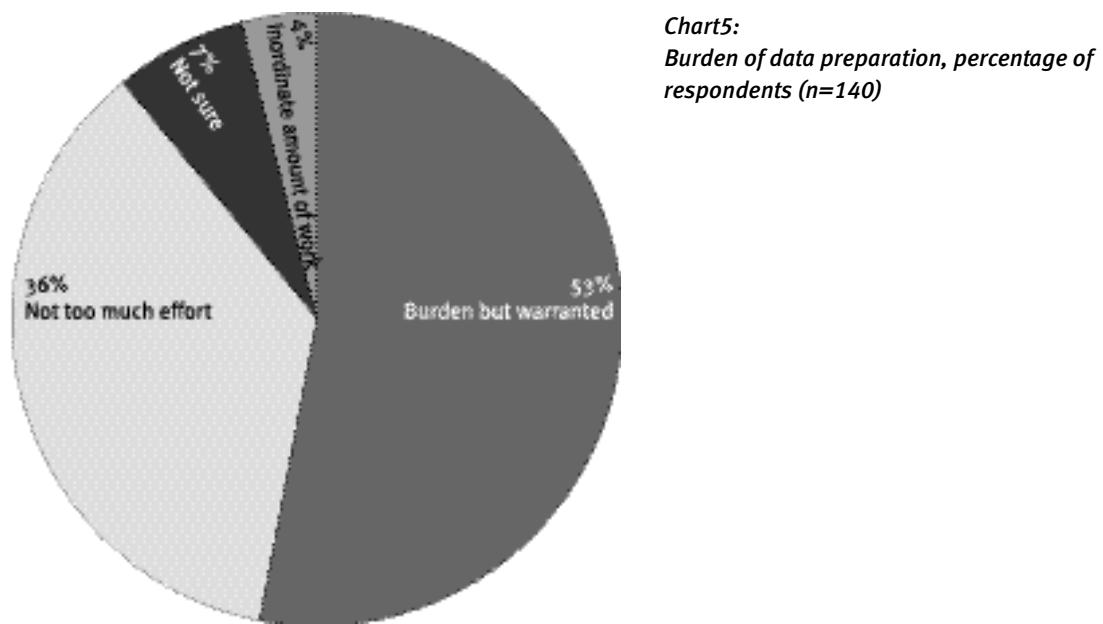
	Col %	Methods	Subject	All
Statistics or data analysis methods	71	23	38	
Survey or research design	10	5	7	
Use of statistical analysis software	5	0	2	
Numeracy or critical thinking skills	0	13	9	
General computing skills	0	3	2	
Add empirical dimension to subject	12	41	32	
Other	2	14	11	
	n=	41	91	132

The case studies show a more contextual view of the teachers' purposes in using data, including the choice of particular sources and how the observed component fits into the course of study as a whole. They also reinforce the idea that there is usually more than one pedagogical purpose, even though one is primary. The subject areas of different disciplines and the creativity of individual teachers ensures that no two courses' use of data will ever be alike, even if they use the same dataset. This diversity has ramifications for the production of 'teaching datasets.'

The use of datasets in independent learning introduces another layer of diverse purposes, as students look for empirical data that answers their personal questions about the world, whilst fulfilling course requirements. A range of educational research supports learner-centred approaches in higher education⁶, which increases students' motivation to learn by making the content relevant to their interests. Allowing students to select their own datasets for empirical analysis has major implications for learning support, staffing resource, and assessment.

5) Perceived burden of data preparation

The survey explicitly dealt with the question of whether the preparation of datasets for teaching use was a burden for teachers. A majority (53%) did feel that the amount of work was a burden, but warranted (see Chart 5). A related finding showed that while 86% agreed that datasets needed to be updated or refreshed for use in successive classes, 29% admitted they had insufficient time to do so.



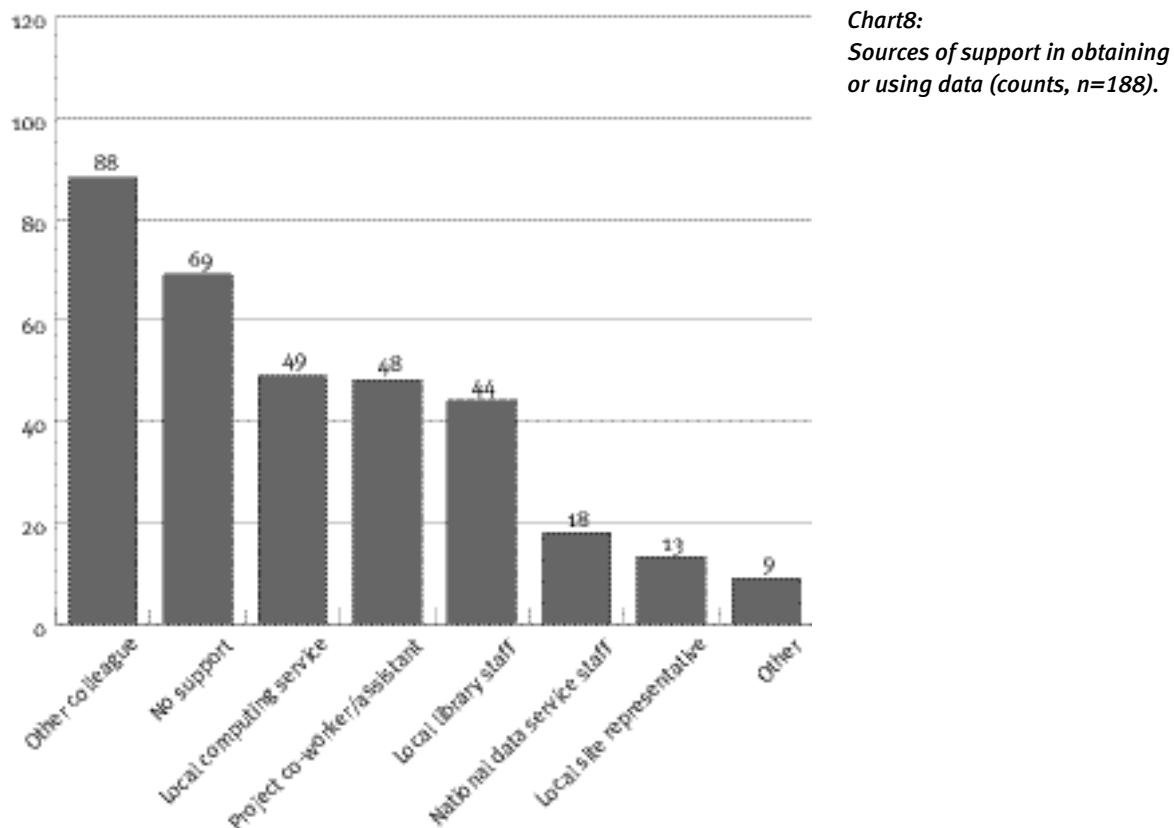
Most of the case studies addressed the level of difficulty in the preparation of data for the class as well. Several of the lecturers had written their own workbook exercises for the practical coursework, which represented very significant amounts of time and effort. Fielding's subject complained that no textbook addressed the methods she wished to teach, so had to – in effect – write her own, through a substantial series of handouts. Drew's subject also wrote workbook material from scratch but had not found the time to incorporate large empirical datasets into the exercises. Therefore his actual preparation time for the small, hand-entered data files was small. However, he found that the time needed for the development of a course website was a burden. Similarly, Morton's subject did not find the dataset preparation difficult for those which were sourced from the Internet and TLTP (Teaching and Learning Technology Programme) products, but found the time needed to set up the course website was very extensive, and regretted not having a break from other duties to allow time for this work.

Both Brown's and Bullen's subjects wrote extensive workbook material to teach students how to use data online at a national data centre. While one had significant help from the data centre staff, the other had help from no one. But both found the entire preparation effort easier due to the fact that they had used the selected datasets extensively in their own research and were well-versed with their content and structure. Townsend's subject had developed a 105-page workbook for the exercises as well. But he was the only one to have extensive local help, in the form of student teachers and a local data library. Interestingly, this was also the only case study that allowed students a choice of datasets to analyse.

6) Sources and levels of data support

Prior to the survey, only anecdotal evidence was available for knowing how teachers obtained support for classroom use of datasets. Members of the Task Force were familiar with the common reality of peer support for data use in both research and teaching via word-of-mouth. One member was aware he was considered to be ‘the data guy’ in the department, to whom others came for support. Although two data librarians were involved in this project, it is known that dedicated data libraries and librarians are not common in UK universities. (They are more common in major research universities in North America.) Site representatives for the national data services can be based in the library, computing service, or elsewhere in the institution, but it was not known how much support they actually provide to users. Finally, the national services themselves staff email and telephone-based helpdesks to answer user enquiries, which are regularly monitored on behalf of the JISC, but only an unknown portion of potential users contact them.

To get a more complete picture of data support in academia, the teachers’ survey asked respondents from whom they had received support in the use of data, either for teaching or research (see Chart 8). Although respondents could ‘tick all that apply’, more than a third (37%) had received no support at all. The related open-ended answers demonstrate that a significant minority were content with this situation, and that support was not needed. However, since the majority had also indicated that they used self or student-collected data, it is possible that more support would be sought if secondary datasets were used. Indeed, national data service staff were only called on for help by 10% of respondents, and local site representatives by only 7%.



The local computing service and library were each a source of support for about a quarter of respondents. This seems rather low, considering the question was asked about both teaching and research—the full range of academic activity. As anticipated, peer support was the most common form (26% from a project co-worker/assistant, and 47% from another colleague). Whilst this may be a completely acceptable and appropriate situation for departments and local institutions, it reinforces the majority perception that the data support provided by local institutions is ad-hoc (62%), as illustrated in Chart 10. A further 24% said their institutions do not provide co-ordinated support at all. (More heads of department characterised their institutions as not providing co-ordinated support than other respondents: 29% vs. 22%). Only 14% of respondents thought their institutions provided ‘very good support across the board’. (And only 9% of the hands-on course lecturers agreed with this.)

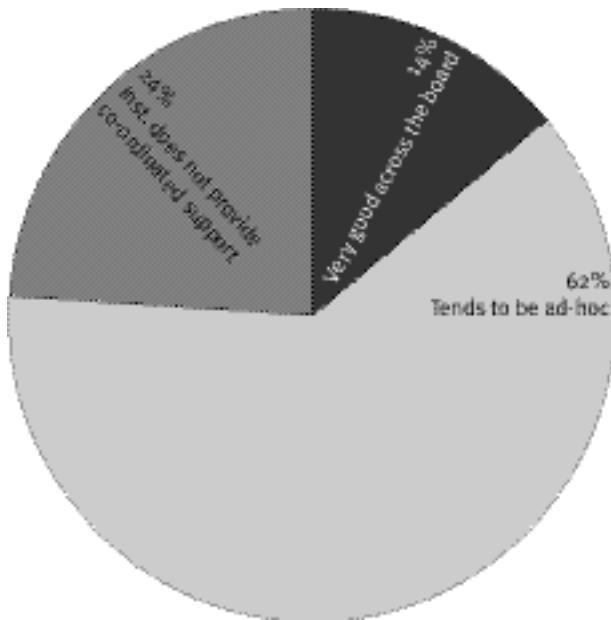


Chart 10:
Level of local support provided, percentage of respondents (n = 176).

These results are not a picture of success in the provision of local support for numerical data use. The Task Force perceived a gap between the library and the computing service of many institutions in providing adequate data-related support. Whilst librarians are trained to help people locate and use published sources, both in print and digital form, they may feel that numerical datasets are out of their range of expertise (even when they are provided by the same data centres that provide bibliographic services). On the other hand, whilst computing support staff are prepared to help users with statistical packages and other analysis software, and in some cases statistical methods, they may be unfamiliar with secondary data sources and their documentation, i.e. the content of datasets. Additionally, those with “site representative” responsibilities to the data centres may not be aware of each other, or the other national services. As one respondent lamented (in answer to Question 26):

Suggest better collaboration between Site reps, computing staff and library staff, converging in a more centralised system within the institution. There is no need for three points of contact in an electronic era. I had to contact 7 individuals before being granted access to library journals, datasets, software and national datasets. Two more applications are still pending...

A recent report by the Association of University Teachers, the union which represents both academic and academic-related staff across UK higher education, claims that “The delivery of higher education is changing. Academic-related staff—principally administrators, computer staff and librarians—are increasingly involved alongside academic staff in making a direct contribution to the provision of higher education,”⁷ in the form of teaching, preparation of learning materials, support for ICT-based study, and integrating learning technology into academic provision. Attention needs to be given at the institutional level as to how support for use of numeric data is integrated with these other contributions, resulting in better quality education for students and reduced burden on academic teaching staff.

7) Forms of local support needed

For those who would wish for more support in using numerical data, what forms of support are needed? The questionnaire anticipated a number of local support activities and asked respondents to tick all ‘forms of locally provided support needed by academic data users,’(see Chart 9). Respondents ticked three forms of support on average. All of the suggested forms of support received between one-third and two-third ‘votes’from those who answered the question (79%). They are:

- Data discovery / locating sources 66%
- Help students use data for learning and research 58%
- Current awareness service via email 49%
- Assist in the preparation of data for teaching 49%
- Teach use of new data interfaces 46%
- Provide expert consultation for statistics & methods 44%

- Help with registration procedures 35%
- Reformat, subset, manipulate datasets for the user 35%
- Acquire datasets for a local collection 33%

A related question asked about the primary form of support needed. The top two remained the same, but ‘expert consultation’ became the third most important form of support needed.

Departments, libraries, and computing services within universities may wish to consider if the support functions described above are provided at all, and if so by whom. Before libraries dismiss too many of the above functions regarding numerical dataset support as overly in-depth compared with the familiar bibliographic support, it is worth considering the words of John Akeroyd of South Bank University, on the future of academic libraries:

If there is one aspect of the future which seems to meet universal agreement it is that libraries will need to become even more service centred than before and bring higher levels of learning support and research support to a user body which is becoming more distributed. ...Libraries will need to become much closer to their users than now.⁸

If the local data support provided to staff and students for learning, teaching, and research is fragmented or ad-hoc, can strategies be identified to make a more seamless fabric of support for individual users? Are there existing institutional fora for discussion of cross-departmental support needs, in which an agenda for data support can usefully be inserted? Institutions are rapidly developing new modes of support for electronic learning technologies such as VLEs (Virtual Learning Environments) and course websites; these changes create opportunities for consideration of how support for using numerical datasets can be provided as part of traditional and new learning environments as well, so that it does not continue to slip through the cracks of institutional support functions.

A number of thoughtful replies were given in a related open-ended question which asked respondents for suggestions to improve data support in the local environment (see Appendix 6). Answers were grouped around four themes. The first group felt that either they were not in a position to make a positive suggestion, or that no additional support was needed, e.g. “My needs are specialist. Support not necessary or feasible.” There was a concern expressed for balancing of priorities with available resources (“Rising journal costs means prioritising resources.”)

Second, there were a number of suggestions revolving around the funding or structure of support, e.g. “Using numeric data and computers in class in learning and teaching is relatively labour intensive. A premium to allow additional support would be very helpful,” and “Central web-based facility plus faculty assistance.” There was not consensus on the need for centralised support though: “Data usage tends to be School-specific, so across the board support is difficult to envisage.” Some were critical of the quality or quantity of existing support structures: “Library staff being pro-active rather than re-active to the needs of academic community,” and “More computer support.” The third group related to staffing or specific services, e.g. “Appoint someone in the department to collect/gather the relevant data and set it up in a suitable form for analyses – teaching,” and, “Raising awareness and providing consultation on using data resources.” The fourth group of responses made suggestions on training needs. Issues raised included study leave to attend refresher courses, similarly, “helpful scheduling” of local teaching workshops, useful examples and case studies, and learner-centred training for staff.

8) Perceived barriers to use of national data services

A critical component of the enquiry was the exploration of barriers to use of the national data services in learning and teaching. Those who had some experience of the national data services were asked to rank a list of potential barriers to their use in learning and teaching (see Table 30). The barriers identified are listed below in declining order, along with comments from the accompanying open-ended question which asked for suggestions for improvement.

- **Lack of awareness of relevant materials.** Sample suggestions are: “Regularly be in contact with people like me,” and “The initiative needs to come from the National Services but better publicity would be a start.”
- **Lack of sufficient time for preparation.** “Time is my main problem.” Another comment focused on the elapsed time between ordering (off-line) data and receiving it.
- **Registration procedures.** “Faster registration - Can it be done online?” Also, “Make registration procedures simple and abolish restrictions on use (e.g. all students signing disclaimers).”
- **Interfaces for data downloading or analysing.** “Able to get data without learning special software.” Similarly, “Easy access and compatibility with different software.”
- **Format of datasets provided.** “Getting data in right format for the software that will be used is a problem.”

- **Documentation & supporting materials.** “Use of numeric data should include a revision of the methods for collecting this data. Num. data should offer complementary qualitative info.” Also, “Sample data with supportive documentation for teaching purposes.”
- **Lack of relevant teaching subsets.** “Easy to access datasets that would be interesting for students to use.” “Rapid access to key summary economic data in form tailored for teaching.” Some specific datasets were named: “e.g. BHPS [British Household Panel Study] from the Data Archive, documentation is great for research, easy to extract. A few series for teaching but dauntingly large for students ‘independent research’, subsets would be useful.”
- **Other.** The few “other” barriers supplied tended to tie in to the ones above in some way. However, two open-ended comments touched on an interesting approach: “One source of data,” and “Overall shift to a consistent and open system.”

These barriers are not that mysterious to experienced users and the data providers themselves, but solutions are slippery. With respect to raising awareness of available resources, whose role is it to do this? Answers must include JISC, national data services, site representatives and support staff within institutions, as well as users themselves, by availing themselves of current awareness email lists on JISCmail, etc. Also, how much should the providers ‘push’ information at the academic community in the absence of specific requests or voluntarily signing up to receive information through national academic email lists? There is a growing clutter of logos, acronyms, and URLs in the JISC’s DNER (Distributed National Electronic Resource) which cannot be expected to fill the heads of each and every academic. Elegant and tailored solutions must be found to the problem of raising awareness. Timing is also key, as information tends only to be taken in at the point of need.

Registration for social science datasets tends to be more complex than other research databases because of the need for confidentiality protection for respondents. Despite techniques to strip identifying information from individual records, producers are concerned about attempts to mis-use datasets for exploitative purposes. Another concern is non-academic use of expensive downloadable data. Therefore, users of numeric datasets are generally asked to complete registration forms, stating their purpose for use and contact details, and accepting lengthy legal terms and conditions with a signature that must be posted, not faxed, to the data provider before access is given. Often the site representative needs to first counter-sign the registration form. All of this processing takes substantial time and energy away from the work of class preparation and studying. Another problem is the issuance of multiple usernames and passwords for different services.

These procedures need to be re-conceptualised from a user-centred perspective, with particular streamlined access for teachers and their students, whenever opportunities for changes arise. The Office for National Statistics is setting a positive example with the commitment to free and accessible information and data on the Web, including a planned click-use license for public access to Census data. Academic service providers are concerned that such moves will prevent them from tracking and monitoring usage, to build the case for further funding based on identifiable needs within the academic community.⁹ However, such substantial lowering of barriers for end-users must be welcomed. New technologies should continue to be explored to ease user management for the providers. (Perhaps Sparta—the next generation Athens authentication system, or digital signatures, will provide some solutions.)

Regarding teaching datasets, some lessons have been learnt. The popularity of the teaching subsets of General Household Survey data created by the Sociology Department of the University of Surrey was confirmed within the open-ended comments and among Task Force members. The value-added work done on the dataset made it interesting for students, relevant for the teaching goals, and easy for teachers to incorporate into coursework. Unfortunately, these were not kept current due to lack of resource. The TLTP (Teaching and Learning Technology Programme) funded courseware products are another pre-cursor. Morton’s case study in History still used a TLTP-created CD-ROM, but was less enthusiastic about the format of the data and the proprietary software than the content. (The TLTP pre-dated the WWW.) A consultation recently done by the CHCC project, which is developing teaching materials for current and historical census data, found that teachers appreciate different formats for different class types and levels. Also teaching materials are best delivered in small chunks (units), rather than complete modules, so that they can “pick and mix” according to their teaching goals.¹⁰

Recommendations and conclusions

The following recommendations and conclusions emerge from synthesis of the results of our enquiry and discussion, and are meant to be considered alongside the other components of this report. We hope they will provide the JISC and others with a framework for moving forward. No single organisation can bring these recommendations to fruition; rather, strategic co-operation to achieve objectives is a fundamental message of this report.

1. A broad initiative is recommended to promote subject-based statistical literacy for students, coupled with tangible support for academic teaching staff who wish to incorporate empirical data into substantive courses.

- **Key skills** – Responsibility for building students’ ‘transferable skills’ which include statistical literacy, numeracy, critical thinking, data analysis, and computing skills needs to be addressed right across higher education. Such quantitative-based skills should be integrated with, and not overlooked, in the push for information literacy, IT (information technology) skills, and other skills deemed necessary as educational outcomes, along with discipline-based knowledge.
- **Teaching rewards** – More rewards for innovative teaching are needed, combined with adequate facilities, preparation time, and personal support for teachers who wish to integrate hands-on use of data by students into coursework. This is true for both methods teachers and subject teachers, but the latter may need extra help in making a start.
- **Training** – Many teachers need to build or rebuild confidence in their own quantitative skills for incorporating students’ use of data into coursework. ‘Refresher’ courses should be made available locally, which are convenient for staff with busy teaching schedules. Bursaries are needed for teachers to attend specialised short courses and summer courses.
- **Student-centred** – Both undergraduate and postgraduate students should be given adequate support for the use and analysis of secondary data sources as part of their independent research. It is unrealistic for support for students to fall solely on the tutor or supervisor, because there are usually several learning curves that need to be mastered by the student in order to get the empirical result desired. (A majority of survey respondents desired help at the local level with both data discovery/locating sources and helping students to use data for learning and research.) This would help lead to a more student-centred education as well as reduce the burden of teachers.

2. The development of high-quality teaching materials for major UK datasets must be funded adequately, in order to provide salience to subject matter and demonstrate relevant methods for coursework.

- **Fully-documented datasets** – Data-related teaching materials need to include subsets of large complex datasets, along with clear documentation about the original and subsetted dataset, practical exercises for students, and teachers’ notes. If teachers are not confident about what the dataset can demonstrate to the class, they will not use it. They also need to be in a usable format for the local environment.
- **Reinforce taught subjects** – Subsets need to be tailored to a range of subject disciplines, able to illustrate concepts that are actually taught. The differences in needs and purposes between methods-courses and subject-based courses found in the survey need to be taken into account when designing teaching materials, and the greater reluctance of subject-based teachers to incorporate hands-on work with data.
- **Supply interesting evidence** – Quantitative study has a reputation for being dry, irrelevant, or even dishonest among many students (‘Statistics lie’). A related problem is that at present, easy-to-use sample data available in standard statistical packages are either US-based or outdated. This can be combated through provision of current, interesting data, based in the UK or other geographic area of interest to the students. If enough user-friendly subsets of major studies are developed for learning and teaching, students can be encouraged to use more empirical data in their own research without added burden to teachers, leading to a more learner-centred education.
- **Define responsibility** – It is unclear at present who should lead the effort to create these new data-related teaching materials, but national data services have expertise regarding datasets they service, while LTSN (Learning and Teaching Subject Network) subject centres have knowledge of teachers’ needs in particular disciplines. Teachers themselves may have much to offer, given sufficient resource for development and an environment which encourages sharing. Learning and teaching materials should be free at the point of use to encourage uptake, but funding needs to be earmarked within existing structures for materials to be developed.

- **Develop partnerships** – One possible model for the creation of new teaching materials is currently being used by some of the other projects funded in this programme (see Appendix A). Partnerships between data centres and subject centres have been established to commission teachers or learning technologists in universities to develop and contribute on-line, locally customisable course materials for shared use. Other members of the academic community will then be encouraged to share their customised learning units, and newly authored units, with the rest of the community.

3. The national data services need to improve the usability of their datasets for learning and teaching.

- **On-line tools** – Intuitive search and extraction interfaces to downloadable data need to be developed that comply with current expectations of accessibility. At present, web-based access meets that norm. The Data Archive, MIMAS, and EDINA could also look for partnerships that pilot students' use of data within new technologies such as VLEs—Virtual Learning Environments. (This could provide exciting examples for teachers and learning technologists working together to modernise courses.)
- **Customise** – Datasets need to be provided in a variety of useful formats, conforming with software that is supported in local environments and appropriate for the course of study.
- **Reduce time-lag** – Delivery time for user access needs to be improved to be realistic for the time pressures of students and teaching staff.
- **Streamline student registration** – Registration procedures should be more user-friendly (currently they are depositor-friendly). Terms and conditions of use should be simplified for learning and teaching purposes, eliminating individual student usernames and signatures when publication of work is not an outcome (i.e. letting the teacher sign on behalf of a class, taking responsibility for usage).
- **Profile without paper** – Technology should be favoured for monitoring usage over giving users more paperwork to fill out. Registration must be recognised as a barrier to use in learning and teaching, stripped down and eliminated whenever possible. The current model of providing access free at the point of use however, should be kept.
- **Train the trainers** – Site representatives or other relevant support staff should be offered training in use of the national data services that is commensurate with their own professional development goals and their department's policies (e.g. travel, days away, and course fees are potential hurdles.) These 'champions' can then offer group or individual support to academic staff and students at the time of need.
- **Inter-operability** – Cross-references and web links should be used between data providers and data discovery 'portals' to increase awareness of relevant resources. New technologies and standards should be exploited that can generate links between numeric data sources and to and from text-based sources.

4. A more concerted and co-ordinated promotion of the national data services could then follow, which is responsive to user demand.

- **Effective marketing** – "Marketing is not just finding a better way to tell people about your service. The Chartered Institute of Marketing defines it as 'a management process which identifies, anticipates and satisfies customer requirements profitably' (or efficiently).¹¹ This involves understanding users and their motivation, segmenting them into groups with common needs, and developing strategies with clear objectives.
- **Roadshows** – Regional roadshows can provide stronger links with local support staff and accessible opportunities for over-burdened teaching staff. By providing a hands-on component to the promotional event, the national services can offer users a low-investment learning opportunity along with their chance to 'stump' to an audience.
- **Local templates** – National data resources need to be 'localised' to improve their accessibility within the education environment. For example, catalogue records of important datasets could be provided to university libraries for inclusion — whether or not there is a local dataset collection or 'data library'. Promotional materials should be customisable at the local level to enhance targeting of different user groups. The emphasis should be on adding value to local awareness-raising efforts rather than national or organisational branding.
- **Dialogue** – Consensus needs to be sought among national and local 'gatekeepers' and stakeholders, about appropriate publicity and promotion by different organisations within the DNER (Distributed National Electronic Resource) and beyond, to avoid confusion as promotional efforts are stepped up. Dialogue with research councils, national and local government agencies, scholarly and professional societies is also needed.

5. Universities should develop IT strategies that include data services and support for staff and students, and integration of empirical datasets into learning technologies.

- **Identify locus of support** – Universities need to identify appropriate staff for data and statistical support for both research and L&T. Data-related support need not be centralised in an academic data library, but it needs to be easily identifiable by users. Sufficient resource needs to be allocated for data-related support functions whether they are located in libraries, computing services, or specialist departments. Support staff in both libraries and computing services should be encouraged to ‘keep up’ with national developments in order to communicate relevant services to users at the point of need, as a component of professional development policies.
- **Balance resources** – Increased resource is being found for development of managed learning environments (MLEs) and Virtual Learning Environments (VLEs) in institutions to keep apace with student and staff expectations in learning technologies. Along with other concerns, the ease with which numeric data and other empirical evidence can be used in these electronic environments needs to be considered. Innovators should be encouraged to present examples of successful modes of data presentation and analysis in these new learning environments.
- **Support for students** – Whilst data-related support for academic staff may be slipping through the cracks, in many institutions support for students is non-existent, particularly undergraduates. For teaching departments to construct ambitious learner-centred curriculae, they need to know their students will have access to computing resources and be able to get personal support beyond their own office hours. Dialogue needs to take place across university academic and support departments to determine if support levels are adequate for all and to identify solutions for new learning goals and environments.

Particular thanks is given to the members of the Task Force, who committed valuable time to this small project and whose focused contributions greatly enhanced our understanding of the problem and our search for solutions. We see this project as the first step of what must be an ongoing effort to improve and enhance the provision of numeric data to teachers and students, and to promote its effective use in UK higher education.

References

- 1 Marshall, G. (January 2001) "Addressing a problem of research capacity." *Social Sciences: News from the ESRC* (47): 2.
- 2 Schield, M. (1999) "Statistical literacy: thinking critically about statistics." *Of Significance* (Journal of the Association of Public Data Users): 1. Available: <http://www.augsburg.edu/ppages/~schield/MiloPapers/984StatisticalLiteracy6.pdf>
- 3 Ahmet, K. (July 2001) "Inspiring students (or teaching students whose interests lie elsewhere)." Paper presentation made to Institute for Learning and Teaching Annual Conference: York. Abstract available: <http://www.ilt.ac.uk/iltac2001>.
- 4 "In fact only about 27% of the students noted that a tutor or lecturer had pointed them to a resource while 19% obtained help from a peer. Worryingly, the library and information service only provided help (either directly or via a course) to a further 16%," (p. 37). Lonsdale, R. and Urquhart, C. (May/June 2001) "Electronic information resources for higher education: something amassed or amiss?" *UKOLUG Newsletter* 12 (3): 35-37.
- 5 As a summary of current issues affecting higher education this only scratches the surface. For one recent overview see: D'Andrea, V. and Gosling, D. (July 2001) "Joining the dots: reconceptualising educational development." *Active Learning in Higher Education* 2 (1): 64-80.
- 6 For example, see: Tait, J. and Knight, P. (eds) (1996) *The Management of Independent Learning*. London: Staff and Educational Development Association and Kogan Page.
- 7 Court, S. (January 2001) *Building the Academic Team: a Report on the Contribution of Academic-related Staff to the Delivery of Higher Education*. London: Association of University Teachers. Available by request: hq@aut.org.uk.
- 8 Akeroyd, J. (March 2001) "The future of academic libraries." *Aslib Proceedings* 53 (3): 79-84.
- 9 Rees, P. (May 2001) "To partner or not to partner, to license or not to license, to register or not to register, those are the questions." Paper presented to the ESRC/JISC 2001 *Census Development Programme Fourth Workshop*, 15-16 May, 2001: University of Leeds. Available by request: c.macdonald@geog.leeds.ac.uk.
- 10 Russell, L. (May 2001) *CHCC Workshop Evaluation Report*. Available: <http://www.chcc.ac.uk/workshop/EvalReport.html>.
- 11 Bedwell, K. (July 2001) "Thinking as a marketeer." *ASSIGNation: ASLIB Social Sciences Information Group and Network Journal* 18 (4): 2-5.

Appendix A: Related developments

The project team is aware that positive developments have been taking place since the goals of the project were first formulated. The funding councils have been making more money available for national services in the learning and teaching arena, and the DNER team of the JISC has developed a coherent programme in this area since the first call for proposals was issued. A second round of funding called into existence a wide range of new projects, some of which have been clustered with this one under the rubric, "Enhancing JISC Services for Learning and Teaching." These include the following important service-led innovations to make numeric data more accessible in terms of interface, discovery, and development of relevant on-line teaching materials:

- **Biz/ed Virtual Learning Arcade** (1 October 2000 to 31 March 2002)

Led by the Institute for Learning and Research Technology at the University of Bristol, this project will develop a Virtual Learning Arcade containing a series of online business and economic models and simulations to support learning and teaching. The Arcade will run these models and simulations live online, developing interactive materials to support business and economics curricula. The learning and teaching materials will include worksheets and glossaries to facilitate understanding and guidelines on how they can be used and how lecturers can integrate them into their teaching.

- **CHCC: Collection of Historical and Contemporary Census data** (1 October 2000 to 30 September 2003)

Led by MIMAS with several partners, the central aim of the project is to develop the Collection of Historical and Contemporary Census data and related materials (CHCC) into a major DNER learning and teaching resource.

- **e-MapScholar: New Tools for Learning and Teaching using Digimap** (January 2001 to December 2002)

Led by EDINA, e-MapScholar aims to develop tools and learning and teaching materials to enhance and support the use of geo-spatial data currently available within tertiary education in learning and teaching, including digital map data available from the EDINA Digimap service.

- **Timeweb: MIMAS/ILRT Time Series Databanks** (1 March 2000 to 31 August 2001)

The UK academic community has access to a number of key national and international macro-economic time series databanks which are under utilised in learning and teaching. This proposal was funded to improve access to this data through a web-based interface and development of a range of supporting Internet-based interactive learning materials.

These efforts are laudable, and their uptake in learning and teaching needs to be monitored to make sure they are providing the kinds of enhancements needed in coursework. Likewise, recent developments at the Data Archive, such as new registration and ordering procedures, file delivery over the Internet via FTP, and the integration of new, major datasets into NESSTAR—an online data browser and extraction tool—need to be evaluated in terms of its affect on the use of datasets in learning and teaching.

The Economic and Social Research Council, which founded and partially funds the Data Archive, is addressing some of these issues as part of its strategic look at its data archiving policy.¹ Possible developments include a "virtual data centre" or data portal, that would allow data discovery without users having to know about various websites and organisations, centres of excellence around specific datasets, and increased resource for major datasets (at the expense of smaller, lesser-known datasets). Co-operation with JISC is considered to be essential for several of the aims.

In addition, the new LTSN (Learning and Teaching Subject Network), which aims to "become the primary information and advice resource on learning and teaching matters for all academic and related staff in HE"² has begun to make its presence felt in the academic community through its 24 subject centres and the 'generic' centre. It is crucial that adequate knowledge of, and support for numeric datasets and related materials are provided by the LTSN within their subject-specific domains, and integrated with other teaching materials and technologies.

1 Boddy, M. (February 2001) "Data policy and data archiving: report on consultation for the ESRC Research Resources Board." *Economic and Social Research Council*. Available:
<http://www.esrc.ac.uk/GreenPaperFindings.htm>.

2 _____ (Undated) "About the Learning and Teaching Subject Network."
<http://www.ltsn.ac.uk/about/default.asp>.

Appendix B: Seven case studies

Title: Using Census Data in a Computer Workshop

Subject: Peter Lee, BSc, MSc. Lecturer, Centre for Urban and Regional Studies, University of Birmingham

Author: Dr Mark Brown, Centre for Census and Survey Research, University of Manchester

This case study is based on observation of a computer workshop using 1991 UK population census data (Small Area Statistics), conducted by Mr. Lee for a group of 20 second year undergraduates. It is supplemented by interviews with Mr. Lee and students participating in the workshop and by course documentation.

Course: Urban Poverty and Planning Issues

Level: Second Year (Geography/Planning)

Teaching methods: Weekly lectures and workshops

The workshop – learning objectives

The workshop, one of a series based around the access and manipulation of numeric data using computers, is designed as an integral part of the learning process on this course. Essentially lectures are used to impart subject knowledge and identify policy issues while workshops provide hands-on experience of the practical tasks involved in exploring these with real data. A key objective is to demonstrate that the process of selecting and analysing data is a political as well as technical one.

At one level individual workshops are designed to address a particular skill in working with data - such as using SPSS or accessing census data. However, a key purpose of the workshops is to support students in their completion of an ongoing class assignment. This assignment (on which students are formally assessed) requires the access, manipulation and analysis of data in addressing a policy issue for a specific area chosen by the student. The skills learned in workshops thus acquire direct relevance, and contribute directly and cumulatively to the students progress on the assignment. This arrangement ensures continuity between workshops and also provides an incentive to attend (you need the skills to complete the assignment).

The observed class

The observed workshop was a session on accessing data from the Small Area Statistics using the web-based application Casweb ("Census Area Stats on the Web"), hosted at MIMAS, a national JISC-funded data centre.

Teaching environment

The workshop was run in a well-appointed computer lab. A manageable class size of just under 20 was achieved by dividing the class in two and running the workshop twice in back to back sessions. The changeover was not rigidly enforced – an arrangement that appeared to contribute to a relaxed environment and allowed students to work at their own pace – some students left a bit early, others stayed over the hour.

Documentation

The workshop included a minimal introduction by Mr. Lee, and was driven by a self-teach workbook, with Mr. Lee providing one-to-one assistance where requested. The workbook provided full instructions for each task including screen shots. This, in conjunction with web-access to the data, ensured students could pursue the tasks (and collect additional data) in their own time.

The workshop

Students were required to:

1. Access Casweb, a web-based interface to the small area statistics from the 1991 census
2. Select the specific area for which data is required by the student
3. Select the variables required to investigate the research question
4. Extract and download the selected data from Casweb into an Excel file
5. Transfer data into SPSS format, ready for analysis

The exercise incorporated a number of learning tasks of varying levels of difficulty:

- I. what 1991 census data actually looks like
- II. where and how to access small area statistics from the 1991 census
- III. the cognitive processes required to generate relevant data to inform on a policy question
- IV. the political dimension to data used to inform policy questions (implications of choices regarding selection of area and variables)
- V. empirical illustration of key concepts such as ecological fallacy and the modifiable area unit problem
- VI. skills in data management.

'Successful' participation could thus be measured according to different criteria and was therefore appropriate for students of varying ability and commitment – tasks i ii and vi could be achieved by simply working mechanically through the workbook. Tasks iii , iv and v were much more demanding.

The requirement for students to select and extract their own data from a national dataset (rather than working with a teaching sub-set) had some clear advantages. First, it made students think about the data they were handling rather than the technique itself. Second, it ensured tasks were set in a data environment more akin to that faced by researchers in the field.

Student reactions

The class was well attended, relaxed and productive – all students appeared to progress through the practical. Students seemed keen to help each other and this extended beyond simple instruction on how to make Casweb work (there appeared relatively little difficulty in using Casweb) with small groups frequently observed discussing appropriate choice of variables among themselves.

Two students interviewed in more detail had contrasting experiences. One appeared to have only a rudimentary grasp of the issues covered, and had missed the previous lecture. However, despite stating a lack of confidence with computers, she had been able to complete the exercise by mechanically working through the workbook. The other student, although again having little previous experience with the software (Excel and SPSS), was highly enthusiastic about this and previous workshops. She expressed strong support for the heavy emphasis on hands-on learning and use of real data which had made learning more interesting and helped her understanding of the issues covered in lectures. Both students valued the fact that the course was providing useful computer-based skills for future use.

Development of teaching materials

The format and content of the course (including the intensive use of census data) drew heavily on Mr. Lee's own research experience. Potential constraints of time and expertise in preparation of data were overcome by his familiarity with the census and ability to draw directly on datasets prepared for research. For example, earlier workshops used a data file on the West Midlands, containing census data and additional derived variables and data from other sources used previously in research. Developing such a dataset from scratch, and by someone with less experience, would require a much greater investment of time.

Although Mr. Lee had used census data from the beginning, he had modified course content over the three years he'd been delivering the course. This notably included allowing more time and a less rigid structure where learning was based on practical work with real data. To accommodate this he had dropped coverage of factor analysis from the programme this year.

Support for using data in teaching

Use of some census data (e.g. via Casweb) in teaching does require obtaining separate licences for students – each student must sign a separate agreement form. This process is conducted through a local MIMAS site representative. Mr. Lee questioned the need for this extra layer in the registration process, though he acknowledged the possible value of the local site rep. where teachers were unfamiliar with procedures and conditions. As long as teachers are aware in advance of the need for licence agreements, registration need not be a serious constraint to using data in teaching.

Apart from the need to go through the MIMAS site rep. for data registration, Mr. Lee was conducting this data-intensive course with minimal support from within his institution or a national provider. Datasets were prepared and deposited by him on the local server for student access. Other data was directly accessed by students (as in the Casweb workshop).

Mr. Lee suggested the considerable improvements in hardware (well -equipped teaching labs) and software (such as SPSS for Windows and Casweb) had made it much easier to utilise census data in mainstream teaching. This seems an important point. A workshop of this type would have been unforeseeable using tools previously available for the extraction and analysis of census data.

Further information

Data resource URL: <http://census.ac.uk/casweb/>

More information available from the Census Dissemination Unit, Manchester: <http://census.ac.uk/cdu/>

A sample exercise from the workbook is available in PDF format, for examination and reuse (Copyright, Peter Lee): *Downloading Census Data: Introduction to CASWEB*. Available from project website: http://datalib.ed.ac.uk/projects/databeach/workshop_casweb.pdf.

Title: Using the National Child Development Study in a 'Sampling and Surveys' Workshop

Subject: Dilys Bayes, Senior Lecturer, Computing and Mathematical Sciences, University of Brighton

Author: Nina Bullen, MIMAS Support Officer, Manchester Computing, University of Manchester

Introduction

The broad project aim is to examine the barriers which currently limit the use of numeric data in classroom teaching and student projects. The case studies provide illustrative examples of how empirical data has been introduced into learning environments. Using interviews with "innovative adapters" and collecting student feedback on the course, it was possible to identify some key issues which need to be addressed when introducing numeric data into teaching.

This case study is based on a telephone interview, supplemented by personal correspondence with the lecturer and additional material such as the course workshop handouts and practical assessment papers and student feedback collected from the final workshop. Unfortunately, it was not possible to observe a workshop at first hand. However, the lecturer offered to collect some student feedback specifically on the elements of the course which involved using the NCDS data at MIMAS.

Supplementary material

Numeric data resource: National Child Development Study (NCDS)

Further information on the NCDS data: <http://www.cls.ioe.ac.uk/Ncds/nhome.htm>

Further information on using the NCDS data at MIMAS: <http://www.mimas.ac.uk/surveys/ncds/>

Course: Sampling and Surveys

Level: Second year undergraduate

Teaching methods

The course is delivered by team teaching. Most of the theory (on the use of sampling to produce population estimates) is taught in weekly lectures, by a colleague. The practical workshops (run by Dilys Bayes) are used to give the students a realistic idea of what it is like to handle 'real'datasets.

Workshop aims

The aims of the workshops were:

- 1) to reinforce the teaching of statistics and survey design by applying methods and techniques to 'real'data;
- 2) to teach the use of statistical analysis software (SAS); and
- 3) to enhance the student's general computing skills.

From the lecturer's point of view, the most valuable aspect of the workshop sessions was to give the students some practical experience with using a real dataset. She felt that it was very important to get some applied experience, rather than just learning the statistical theory.

The lecturer emphasised to the students that the practical problems were typical of the sort of thing they would

eventually have to do when they worked in the real world. The practical experience would also be useful preparation for the students because they have to undertake a dissertation project during their final year.

"The intention is that this experience will improve their chances of obtaining work placement during their third year and might also help them during their final year when they are doing their dissertations - often the students won't have collected any data of their own and if they decide to use a MIMAS dataset, it comes as a bit of a culture shock when they are suddenly exposed to SIR etc. So this should give them some preparation for using the data for their own research projects."

Since some of the students go on to work for the ONS or in statistical consultancy, she felt that having this experience on their CV would give them an edge when applying for jobs. Indeed, she felt that the practical workshops were "*one of the most useful experiences that the students can have*".

Workshop delivery and assessment

The module consists of a series of eight workshops which are designed to give the students hands-on experience with using real data. The workshops take place in a computing laboratory, with around 15 students. They work individually, but tend to chat with their neighbours about the tasks and if they need help, they will ask the class lecturer.

The students are given weekly handouts which contain worked examples and exercises for them to complete. There are two assessed practicals at the end of the module. The assessments are marked separately (so that the results are not dependent on passing each module), and an average mark awarded.

Learning objectives

"We get the students to extract their data from MIMAS so that it is ready for SAS programming and we were particularly fortunate in that SAS version 8 (which we have just started to use) contains procedures for survey processing, tying in nicely with the theory that they are taught."

The students are required to:

- use the NCDS data dictionary (a stand-alone DOS application) to interrogate the NCDS metadata
- to access the survey data itself, which is held in a relational database package (SIR) and to write retrieval programs to extract selected records;
- to use a statistical analysis package (SAS) to apply specific procedures (covered in the class lectures) to the NCDS data.

The students are expected to work with the NCDS data using these packages, both on a remote computer (SIR is run on Irwell, the Unix server at MIMAS) and a local PC network (for SAS version 8). They also learn how to connect to a remote system (using Telnet to work on Irwell), some basic Unix commands for handling and editing files on Irwell, and how to transfer files to their local PC network using FTP.

Workshop summary

The first workshop begins with a recap on survey design issues, with specific reference to the NCDS and the Family Expenditure Survey (FES). It gives an overview of the procedure for extracting NCDS data from MIMAS, in preparation for subsequent workshops.

In Workshop 2, the students login to Irwell at MIMAS and learn some basic Unix commands for file handling and how to use the Joe editor to write a text file. The editor is used to write a SIR retrieval (database query), which is provided for them at this stage.

Workshop 3 gives an introduction to information retrieval. The structure of a large and complex dataset is explained and illustrated by the NCDS. The students use the NCDS Data Dictionary to search for specific variables (using various forms of keyword search) and to identify key items of metadata (eg. valid codes on variables). They also look at the construction of a SIR retrieval and an exercise asks them to edit an existing retrieval file to make specific changes to the data extraction program.

Workshop 4 explains how to run the SIR retrieval (saved from the previous session) on Irwell. The retrieval is run non-interactively (at the Unix prompt), rather than from within SIR. This is how researchers tend to work once they are familiar with the SIR command syntax. The various output files are explained and the students are shown how to check that a job has completed successfully.

Workshop 5 extends the SIR retrieval to export the selected cases/variables to SAS. It begins with an overview of SAS (which is taught in a separate module). The elements of the SAS programming language are identified,

including data steps and procedures. The workshop handout explains how to modify the SAS program (produced by the SIR export procedure) to enable it to be run with SAS on the local system at Brighton. The handout illustrates a SIR retrieval and the resulting SAS program file and data file. The students have to set up an FTP connection to Irwell at MIMAS. They then create and run a SIR retrieval to extract a subset of data from NCDS and save it as a SAS program and data file. The files are transferred by FTP to Brighton. The SAS program file is modified and run to create a local SAS dataset and perform various SAS procedures on the data.

Workshop 6 is a revision session. The students are asked to write a SIR program from scratch to extract a set of specified variables from the NCDS database at MIMAS and to save the data in SAS format. The files are then transferred to Brighton using FTP. Working with SAS on the local network, the SAS program file is then modified as necessary to read the accompanying ASCII data file into a SAS dataset and produce a range of descriptive statistics and tabulations.

Workshop 7 was missed due to illness of the instructor. Workshop 8 focuses on using SAS to estimate the Standard Error (SE) of the mean. SAS version 8 has procedures for survey data processing, including SURVEYSELECT to select a sample from a population dataset under various sampling designs and SURVEymeans to estimate means on a sample of data (accounting for the sample design). For this practical, a file containing a subset of NCDS data (2000 cases) was created for use as the source of population data source.

The first assessed practical focused on the tasks required to interrogate a database, extract a subset of data and export it to a statistical package (SAS). The tasks include using the NCDS Data Dictionary (a DOS application) to identify key items of information about some variables which they will later extract from the NCDS database at MIMAS. They then write a SIR retrieval to extract a specified subset of NCDS data and save it in a format suitable for processing by SAS. They individually log into a remote machine (irwell.mimac.ac.uk) and use a Unix editor (joe) to write a file containing the SIR database query. They then run this query and export the data to a SAS dataset.

The second assessed practical focuses on the analysis of data within SAS, using another SAS data subset which has been prepared by the class lecturer. They students are required to use SAS procedures to select a sample (the students have been taught simple random sampling and stratified sampling) and to calculate survey means.

Teacher's assessment

The lecturer noted that the students had no previous experience of computer programming at all and their initial reactions were "*horrified*". However, this was as expected by the lecturer, who has been teaching SAS programming for about eight years. The students went through a very steep learning curve (particularly with the NCDS SIR database) and after about four weeks they began to get to grips with it and produced some very good work. The lecturer felt that although it had been a painful process, the students had learnt a lot and this showed through in the final assessed workshops.

"The students don't find it easy – most of them have absolutely no programming experience so to be introduced to SIR and SAS (which they are taught in another module) is hard work. However they have come through it ok and obtained good marks in their first assessed practical."

Student feedback

From the student's point of view, there was almost unanimous agreement that the practical workshops were useful in helping their understanding of the subject (this was the feedback given on the course monitoring and evaluation form).

The students also gave some feedback which was of interest to the TaskForce enquiry:

1) Did the module increase your confidence in handling large datasets?

They all agreed that the workshops had increased their confidence. Some noted that this was because they had no prior knowledge and had struggled at the start of the course. This suggests that the use of numeric data within teaching can be used to break down perceived barriers in the minds of the student's on their own abilities and competences. This is an important step since confidence is the single most important determinant of student success (John Cowan).

2) How has the module increased your computing skills?

They all felt that the module had increased their level of computing skills. They had been introduced to new software (SIR, SAS), learnt new techniques and grasped new ideas. In particular, the course had increased their understanding of structured programming and emphasised the importance of accuracy. One student noted '*a large improvement, as I had very poor knowledge of programming at the start, now I feel fairly confident.*'

3) What is your reaction to the experience of working with the NCDS data at MIMAS?

These responses ranged substantially, from

"A BIG SHOCK!! It was hard to understand at the start as it was all new for me"

to "was fairly easy working with NCDS"; "NCDS is straightforward, easy to use. However, MIMAS's commands are not easy to remember" and "NCDS is easy to use, but SIR has proved difficult as you need to constantly refer to notes that have been made"; "It is not very user friendly, but given time and patience, it can become more 'natural'". One student even wrote: "*I enjoyed it!*"

4) What was the most difficult part of the practicals?

In terms of the difficulties they had experienced, most students identified SIR as being difficult to learn. Their comments related mainly to the command-based interface (ie. syntax, structure). A few mentioned SAS - specifically writing SAS programs to read the SIR data files and making the link between the statistical theory on weighting and writing programs in SAS.

5) What was the most valuable aspect of the practicals?

Several students highlighted the experience of working with real data and writing programs in SIR and SAS to extract and analyse the data as being very valuable. Others felt they had developed skills which would be useful for the future, when they entered the workplace: *'Beginning to become skilled in something that will be of benefit in many jobs I may hope to be employed in after my time at University'*.

6) Having done the module, would you consider using a large dataset for your dissertation. If not, why not?

A positive outcome was that no-one answered "no" to this question! The "maybe's" said that they did not yet have enough confidence or needed more practice. The 'yes' group said that this was because they felt more comfortable working with data now, or had found it *'fairly easy to learn and quite enjoyed it'*. Others hedged their bets with "probably" – they saw that the skills learned on the course had opened up access to a good range of data sources, but another student noted that their use of numeric data would *'depend on the dataset. [I] would choose a dataset which interests me, irrelevant of size'*.

Development of teaching materials: key issues

1. Time is required to develop teaching materials which use numeric data

This is the first year that this module has run and the time taken to prepare the course from scratch was around four months (from July to October 2000). The lecturer had previously used the NCDS for her own research, so was already familiar with the data. She had intended to use another survey (the Family Expenditure Survey) for the course, partly because some of the third year student placements are with ONS. But she found that the amount of time required to familiarise herself with the structure of the FES database and time lost through personal illness meant that she had to fall back on a survey she already knew well.

2. Both the data and the database system are complex

The main barrier to using the NCDS was the format of the data. The SIR database system has a very steep learning curve and the data itself has a complex structure. The documentation and supporting materials were good, although they took quite a while to work through.

3. Need for teaching subsets?

The lecturer did not feel that it would have been of any personal benefit to have teaching subsets, as the purpose of the course was to give the students experience with extracting data from a large database. However, the availability of teaching subsets might be of use to other colleagues (for example, some project supervisors have been interested in using census data).

4. Registration

The registration procedures were not liked, but there was help available through the local MIMAS representative. This lecturer did not regard registration as a major barrier to using the data for teaching.

However, she commented on the experience of dissertation students who had used national datasets for their own research. They had found that the registration for access to data took quite a long time and some got frustrated and went elsewhere for data (eg. downloading for free from the Internet).

5. Support for using numeric data in teaching

The software required (Exceed to access Irwell at MIMAS via Telnet, FTP and SAS 8) was installed on the PC's in the computer laboratory by local computing staff.

The lecturer obtained support from MIMAS support staff in terms of help and advice on accessing the survey data using SIR and interpreting the documentation (the FES User Guide). There was not felt to be a need for local support in reformatting, subsetting and manipulating datasets, as the support provided by MIMAS covered this area.

Help was also obtained from SAS Technical Support with regard to specific features of the SAS 8 software. There was no need for more local expert support in statistics and methods in this case, because the Statistics Department ran its own Statistics Consultancy Unit which provided free support and advice for University staff and postgraduates (but not undergraduates).

There was not felt to be any need for more current awareness services, over and above the current mailing list services (via Jiscmail). She pointed out that too many untargeted messages were annoying and just got deleted. She mentioned some of the sources which are used within her own department (eg. RSS News), so it may be that resource providers would be better advised to use these existing channels.

Finally, she felt that there might be a need amongst colleagues for help with data discovery/locating sources and for training on the use of new data interfaces.

Title: Research Methods in Health Care – Statistics component

Subject: Dr Peter Griffiths, Lecturer, Research in Nursing Studies Section, King's College London, School of Nursing and Midwifery

Author: Ms Kristina Drew, Assistant Computer Manager, Information Systems Unit, Institute of Child Health, University College London

This case study is based on an interview that took place between Dr Griffiths and Ms Drew at King's College London on Monday 19 March 2001.

About the course

Research Methods in Health Care is a compulsory module for all undergraduate nursing BSc students at the Florence Nightingale School of Nursing and Midwifery at King's College London. There are about 70 students in total attending. They are participants of one of the following degree programmes:

- 3 or 4 year nursing degree (the programme was formerly four years in duration and has recently been reduced to 3 years)
- Those holding a Diploma in Nursing and who are on a one year programme to upgrade to a degree (may be full time or part time)
- Those on a one year Community Nursing Degree programme (i.e. are converting to Community Nursing)

The course is run during the first term of 3rd year for the 3 year and 4 year programmes and during the first term of the one-year programme.

The aims of the Research Methods in Health Care module are:

- To extend skills in critically evaluating published nursing research
- To prepare students to conduct a final year research project (this may be non-numeric)

The overall purpose of the Research Methods course is to provide teaching about "doing research". It is not "hands-on" in the sense of seeing the whole research process through experience.

Course outline – Statistics component of Research Methods in Health Care module

Peter Griffiths has been responsible for the Statistics component for the past three years. The Statistics component is taught solely by him.

The Research Methods course forms part of a professional qualification and is therefore subject to ratification by the English National Board for Nursing and Midwifery. Consequently, there is (in effect), a fairly specific syllabus that

specifies, for example a range of statistical tests that should be covered. The lecturer is however, given some flexibility about the interpretation of the syllabus and how he/she chooses to deliver it.

The Statistics component takes place over a term and is a series of eight, one-hour lectures of basic statistical training followed by two, three hour hands-on computing workshops using the statistics computing package Minitab. For the workshops, the students are split into two groups of about 35 students. The students are divided according to whether they are on a one year or three/four year programme. Another lecturer from the school assists with these workshops.

Aims of the statistics component

The learning outcomes:

- Demonstrate an understanding of qualitative data analysis
- Utilise descriptive statistics where appropriate
- Describe statistical models and select appropriate inference tests
- Begin to utilise a computing system for interactive statistical analysis
- Demonstrate awareness and understanding of an interactive statistical computing package

Choice of statistics package

The decision to use Minitab is largely historical. Minitab and SPSS are made available as part of a centralised suite of programmes in the college's open access/training rooms. Given the basic entry level of some students' computer literacy, Minitab is deemed the most appropriate choice.

Computing skills of those attending the Minitab sessions

The computing skills of the students was deemed to be rather mixed. A small number of students are new to computing.

The only compulsory computing training that is provided as part of the degree courses is a basic half day introductory session during the first term of year one (all courses). Some informal remedial sessions are available. Students may also apply for ITcourses that are provided centrally by the college.

Data

The data used are a number of small datasets that have been constructed by Peter Griffiths for demonstrating particular statistical tests.

One dataset is generated by class participation during an Experimental Design lecture. Students are asked to partake in a decision-making exercise (2 variables, 70 cases) and the data is then input into Minitab by Peter Griffiths for class use.

For the other datasets, students enter the data into Minitab themselves.

Peter Griffiths commented that he had endeavoured to provide slightly bigger datasets that students would open rather than input. Due to time constraints, this had not been accomplished for the 2000/1 academic year.

Facilities

Three open access rooms (viewed by Kristina Drew), each housing 60 workstations, are available for computer training. These rooms were primarily designed as open access rather than teaching rooms. Teaching facilities were gradually being provided as add-ons and these were as follows:

Room 1: A white board. There is no data projector installed; instead one is hired.

Room 2: Data projection and sound installed; one projector screen and 2 plasma screens provided.

Room 3: Data projection to a white wall available. The trainer's screen can be viewed on the trainees' monitors.

Development of teaching materials/training notes

These include a handbook for using Minitab and a Statistics workbook that the student completes and hands in at the end of the Research Methods module. Both of these documents were written by Peter Griffiths. These documents along with lecture notes and datasets are available at the following Web page:
<http://www.kcl.ac.uk/ip/petergriffiths/UG/UG.html>

Course assessment

To assess the Statistics component, students complete a workbook. This is available as either an online form version (one that can be saved and typed on) or a version for printing out and writing on. The assessment workbook is designed to be completed using Minitab on a computer. However, it was acknowledged that some students found it difficult to get used to using a computer and completion of the workbook from manual working was also acceptable. The assessment workbook exercise makes a contribution of 10% of the final mark for the Research Methods module. The rest of the module is assessed by examination and a written assignment.

Student feedback

In the first year of running the course, Peter Griffiths set up a focus group of five students with whom he discussed issues related to the course.

Measuring the success of aims/objectives

Peter commented that an informal measure of the success of the course could be the reduction in the number of students returning to him for basic advice at the research project stage.

At the end of the course students are given an evaluation questionnaire. Broadly, this asks how much the student expected to understand and how much they felt they had actually understood.

Support for using data in teaching

In response to the "Using Numeric Data in Learning and Teaching" questionnaire, Peter Griffiths identified the following as forms of locally provided support needed by academic users:

- Most importantly, providing expert consultation for statistics and methods
- Data discovery/locating resources
- Teaching use of new interfaces
- Assisting in the preparation of data for teaching
- Helping students to use data for learning and research

Statistics support at King's College is based centrally in the Computer Centre and is provided by one member of staff in the form of a consultancy service. This person is considered by Peter Griffiths to be very over-stretched. In the past the Computer Centre had taught the statistics component of the Research Methods module to students in the School of Nursing and the School of Life Sciences.

Peter Griffiths felt that there would be some advantages in using datasets that were non-trivial. For example, it would be more interesting for students to look at the proportion of a population that had hypertension rather than calculating the mean blood pressure for a small group of subjects. He felt, that from his perspective, the profile for access to national datasets was low. However, since the merger of the Computing and Library services at King's College in the summer of 1999, he felt that the promotion of national services had improved immensely and he was optimistic that this would continue.

Title: Secondary Analysis and Statistical Modelling

Subject: Prof. Sara Arber and Mark Phillips, Department of Sociology, University of Surrey

Author: Dr. J. Fielding, Department of Sociology, University of Surrey

Level of Module: MSc (compulsory module)

Contact time: 20 hours

Number of students: 15-20

Aim of module

The aim of this course was for students to learn more about conducting secondary analysis of large-scale survey data. It was a practical course focussing on analysis of data from the Health Survey for England (HSE) for 1993. The survey contained a range of variables concerning health behaviours including smoking, diet, physical activity and drinking, as well as various socio-economic variables, family characteristics and measures of health. A thread running

throughout the course was the investigation of two research topics:

- i. Women and smoking
- ii. Factors influencing participation in health-promoting physical activity through choice.

The overall aim of the course was to enable students to analyse data in a conceptually informed way and to competently undertake, present and interpret data using logistic regression analysis.

Learning outcomes

At the end of the course students should be able to:

1. Understand how to order and set-up a data set for secondary analysis
2. Undertake exploratory data analysis
3. Derive new variables and construct scales, and be able to check them
4. Understand the statistical underpinnings of logistics regression
5. Undertake and interpret the results from logistics regression analysis
6. Write a sociological report on a secondary analysis using nested logistics regression models.

Course outline

Week	Lecture	Practical session
1	Develop research questions; Outline of aims of building multivariate models; Find out about datasets	Using the Question Bank (http://qb.soc.surrey.ac.uk/) and ordering data from the Data Archive (http://biron.essex.ac.uk/)
2.	Data management using SPSS. Identifying variables to answer research questions. Conceptual issues in selecting variables and cases. Initial exploration of the HSE dataset.	Creation of own “tailor-made” data subset
3	Investigating research questions. How a literature review informs the subsequent analysis. Drawing path diagrams to characterise conceptual models. Choice of variables to measure the concepts in conceptual model.	Creation of derived variables using COMPUTE in SPSS
4	Exploratory data analysis and graphical techniques.	Creating different kinds of graphical output and interfacing with Excel and Word
5	Introduction to factor analysis. Constructing scales	Creating more complex derived variables for own data subset
6	Review of multivariate analysis – multiple regression and dummy variables	Merging two years of the HSE data
7	Introduction to logistic regression	Running a simple logistic regression
8	Choice of variables for logistic regression. Handling missing data. Statistical significance and confidence intervals	Refining and recoding own variables for inclusion in logistic regression
9	Analysing nested logistic regression models. Choice of models.	Running nested logistic regression models and interpretation of output
10	Graphical presentation of data from nested logistic regression models.	Student analysis of data for their assessed exercise

Assessment

The course was assessed by two interlinked exercises both using secondary analysis of the Health Survey for England (HSE) for 1993 (and /or 1994).

The first exercise was to undertake exploratory data analysis to underpin the logistic regression to be conducted for the second exercise. In this first exercise students decided upon a conceptually informed research problem to be examined using the HSE data. They chose an appropriate dependent variable and described how they might dichotomise this variable for the subsequent logistic regression. In addition, between 4-8 independent variables, one of which should be derived, are selected and a conceptual model in the form of a path diagram is presented, along with preliminary exploratory analysis, in a report.

The second exercise was to undertake a logistic regression analysis to examine a research problem defined in the first exercise and to write it up in a sociologically informed way in the form of a journal article appropriate for *Sociology*, *British Journal of Sociology* or *Sociology of Health and Illness*.

Problems encountered by the course convenors

Access and Registration for students

This was no problem. Sara phoned the Data Archive who sent her a single permission form which was signed by all the students using the data and duly returned.

Preparation of dataset

This was also no problem since the dataset was already being used for research purposes and therefore Sara, and Mark Phillips, the Research Fellow on the HSE project and the tutor on the course, were very familiar with the data. In fact Sara insisted "You have to use a dataset you are familiar with." In addition, they had several worked-through examples that just needed simplifying for the students. Consequently, the time spent preparing the dataset was minimal. However, an enormous amount of time was spent creating student handouts to support the classes. Sara estimated that 25 days were spent creating these handouts. Of course, this 'set-up' time is saved the next time the course is run.

Supporting texts

One problem was that there were no good supporting texts, thus emphasising the importance of the handouts. Existing contemporary texts did not discuss the problems of secondary analysis such as creating and checking derived variables, and texts on logistic regression tended to be too complex.

Title: Death and Dying in early Modern England

Subject: Dr Jeremy Boulton, Senior Lecturer in History, University of Newcastle

Author: Dr Graeme Morton, Lecturer in Economic and Social History, University of Edinburgh

Site Visit: 13 March 2001.

The course, *Death and dying in early modern England* (His104) has been taught for two years. It is an undergraduate course with around 25 students, but some of the data is used for postgraduate teaching on other courses (Hist221) and for undergraduate, doctoral or masters' supervision. The course is based around a website – <http://www.ncl.ac.uk/history/his104index.htm>

Stated student objectives (from the web-site):

- I. To develop your ability to work in a team by preparing a group project.
- II. To enhance your communication skills by provision of electronic bulletin boards. You are expected to ask questions of myself and the rest of group, contribute comments and respond to the questions of others.
- III. To develop your ability to complete a substantial piece of written work: **the long essay**
- IV. To develop your ability to work with primary source material by writing **gobbet exercises**
[*N.B. According to the Oxford English Dictionary, a gobbet is an “extract from a text especially set for translation or comment in examination.”*]
- V. To develop your ability to understand and analyse a range of historical material. The evidence you will consider will include portraits, monuments and artefacts.
- VI. To develop your ability to see historical material in **contemporary context**. In particular contrasts with nineteenth and twentieth-century society are encouraged.

Course Examination

Group workshop tasks (best five)	20%
One long essay or ‘mini project’ of 3000 words	40%
Document report or similar exercise	5%
Document report or similar exercise	5%
Document report or similar exercise	5%
Group Web Project	25%
Total	100%

Using Numeric data

The use of numeric data is not a primary aim of the course; it is used, but only occasionally. Typical of Arts/Humanities History students there is no great desire to encounter numeric data *per se*, ensuring a preference for qualitative material. Yet once enrolled in the Hist 104 Course, students experience numeric data in a variety of settings. The material typically comprises parish data, various on-line resources and material produced and published by the Teaching, Learning and Technology Project (TLTP) History Courseware Consortium. The data is used to teach general computing skills, numeracy and critical thinking, some low level statistics and analytical methods, but primarily to add an empirical dimension to the subject.

The material is presented over the Web, often as raw data (which the course-organiser prefers) or by using spreadsheet or Internet software. Indeed, it is the Internet that is the major source of data used in the course rather than any of the national providers. Gathering, checking and refreshing the data from these sources is not regarded as a great burden, although it is acknowledged that the initial effort needed to establish the website for delivery of the course did take a great deal of time, and that was not helped by the lack of any reduction in other duties to allow for the task to be completed.

The course co-ordinator is someone who does recommend the use of numeric data –to add an empirical dimension to his subject – but he has not used any of the national academic services in his teaching. The main difficulties encountered were to do with registration, the format of the data sets, and the lack of historical teaching subsets that were relevant. The latter point has been a particular problem faced by the historical community.

Recommendations for change focused on an increase in the number of small-scale pre-1800 data sets and the use of raw – not coded or processed – data.

Satisfactory levels of support from local computing staff, local library staff and co-workers on the courses had been received, but the course co-ordinator had had no contact with national data services staff. Forms of local support that would be appreciated are discovering and locating sources, acquiring data sets for a local collection, then help reformatting and manipulating the data. The greatest difficulty remained the time taken to locate the data. The dominant wish for further institutional support was to embark on a rolling programme of hardware and software upgrading. Typically, only parts of the University would be able to upgrade to a level normally ‘expected’ by the national data providers and others.

Reflections on the numeric data and software

- The greater availability of raw material would cut down the need for further re-manipulation of data.
- Many of the software packages – esp. on CD-ROM – are too slow for teaching. This is a problem with some, but not all, of the TLTP History Courseware.
- The students have access to very good computing equipment in the labs, but the staff struggle to keep up with the current software and hardware.
- Greater data support would be welcomed, but often there is already good ad hoc support around the department and the University’s computing service.

Final reflections on the teaching experience

- The students can all point and click, but there is much variation in their skill levels
- The students rarely – if ever – come with any statistical knowledge.
- The students require a great deal of direction.
- Point and click software packages are found to be ‘boring’; interactive teaching with tasks to be achieved on a regular basis are much more successful – they also encourage class bonding.

- One or two in a class of c.25 find the level of numeracy required too much to handle and drop out of the course.
- The students come to do 'straight history' and are resistant to too much social science or number-counting.
- The web-based projects are useful, but students take a long time to focus on a topic.
- Few have any experience of writing web-pages.
- Undirected web-trawling is not successful.

But the students do appreciate what they have learned – on this and the 'Documents and Database' course. The students do regard it as a valuable transferable skill.

Title: Basic Data Analysis Using SPSS for Windows

Subject: Clive Payne, Director: Computing and Research Support Unit, Department of Politics and International Relations, University of Oxford

Author: Sean Townsend, Data Librarian: London School of Economics and Political Science Library

This case study is based on a day visit to Nuffield College, University of Oxford. The aim was to interview Clive Payne in as much detail as possible about his course on data analysis, gathering information on its content, objectives, environment and logistics. No interviews were conducted with students and no class was observed in action. Clive's honesty during the interviewing process does however shed important light on what it is like to run these sorts of courses in reality and as such makes this an interesting example case study.

Course: Basic Data Analysis Using SPSS for Windows.

Level: Second Year Politics, Philosophy and Economics (PPE) students.

Teaching Methods: All day lecture and practical classes with a report to be submitted by students at the end of the term.

Course description and objectives

The course is a one-day compulsory module for PPE students that requires a small data-based report to be submitted at the end of the term. It is a new course that got underway in the 2000/2001 academic year. The basic objective of the session is to introduce social science students to the various data types that exist, analysis software available (in this case SPSS), data output (such as scatterplots, line graphs and histograms) and also how to describe and summarise data via simple frequencies and cross-tabulations.

In terms of student numbers this is a large course. Some 300 students across the Oxford colleges would be included. This has necessitated the course to be run at five separate intervals, therefore splitting each individual session into approximately 60 students. This is done by having each day cover about seven colleges with the timetable published well in advance. Students therefore know where they need to be and when.

The course is largely designed to equip the students with the knowledge and skills they will need to complete the report successfully. Because the report is not optional, attendance at the course is an obvious pre-requisite. However, from another level the course also encourages students to think about the data and the possible interpretations of it. The ability to test hypotheses and explore current-day issues is the most important aspect of the course. Students should be able to appreciate how to investigate policy-relevant data and reach conclusions on their own initiative.

Teaching environment

Each individual class session comprises some 60 students. Although there was initially some anxiety about the class size, Clive suggested that it was not as significant a problem as previously thought. The day takes place within an equipped computing lab at the Oxford University Computing Service. The room houses approximately 40 terminals; both the analysis software and example datasets are pre-loaded. This allows the teachers to explain some of the key concepts of data and analysis alongside the practical use of SPSS to load and manipulate the data.

Documentation

The course is supported by a 105 page ring-bound guidebook. The guidebook is split into logical sections that follow the course sequence. The majority of the book is made up of describing how to use SPSS with accompanying

screen shots and menu sequences listed in bold. It is very much in the workbook model and includes opening data in SPSS, simple analyses, modification, subsetting, data management, data entry and import. Two appendices are given that describe the codebook for the 1995 General Household Survey and the 1992 and 1997 UK General Election Results. During the session help is on hand for those who have difficulties.

The end of term report

In some ways the course is geared towards giving the students some independence once they are familiar with data analysis concepts. As a principle this is certainly convincing. The report assignment is asked to be between five and ten pages in length with appropriate SPSS outputs included. Students will have already received training on how to do this. Nine datasets are made available to the students for the project. They are:

1. British Election Results 1992-1997
2. General Household Survey 1995
3. Economic Growth
4. Correlates of War
5. British Election Survey 1997
6. Eurobarometer Survey
7. US Election Survey
8. British Social Attitudes Survey 1997
9. British Household Panel Study

The data files (SPSS SAV files) and codebooks are downloadable by the students via the course web page. Suggestions for project themes are given ranging from the fairly straightforward to those that would require the students to think harder about the data. The suggested student projects are available as a Word document at <http://datalib.ed.ac.uk/projects/datateach/student-projects.doc>. These suggestions are designed to mirror current-day issues so that the students can relate closely to the sources and their relevance. There is also the option for the students to choose their own dataset, and data centre Web sites are listed if they want to explore this. Most choose the pre-set sources, which is understandable.

Full support is given to the students during the project weeks. This includes an email address that aims to reply within 24 hours in addition to instructors that are available from the Computing and Research Support Unit. Clive employs the use of graduate students for this support system who prove to be helpful and enthusiastic. One day per week is set aside to answer student queries on the project face-to-face. For those who find the report a struggle, one-to-one tuition is arranged to help them complete the assignment.

Student reactions

Despite the course being a departure from traditional Oxford teaching methods, Clive thought that the feedback from students had been positive. Some (about 10%) had reacted against the compulsory nature of the course and in extreme cases argued that they did not choose Oxford to study numbers. A very small minority therefore had real problems, both with the idea of the course, and in understanding the most basic principles of data analysis.

The vast majority (about 90%) were happy. This was clearly reflected in the quality of the final reports, some of which approached dissertation quality. Clive stressed that careful selection of data to make them relevant to the subject matter had been important, and that students understood the relevance of the data sources and the methods employed to analyse them as a result.

Support for using data in teaching

The experience of putting together and running a data-based course had thrown forward a number of important issues. There had been something of a mixed experience in terms of getting the datasets for students to use. Some of the data were freely available from the Internet. These had presented few problems and it was relatively straightforward to obtain these along with the associated documentation. Where barriers did occur was with the need to register users for Data Archive-supplied datasets. Clive strongly opposed the need to register individual students and that for a class of over 300, this was practically impossible. Clive also questioned the need for such information to be sent to the Data Archive when the resources are for teaching. Technically, the existing registration process makes such a class impractical.

Oxford is one of the few universities with local data support. This had proved convenient for accessing the datasets from the local data holdings. The Data Librarian had taken care of the administration and had also prepared the

datasets for use by students. Graduate students provided hands-on support both during the one-day course and as support during the report period. Their experience and expertise in using datasets had proved invaluable.

In terms of software and hardware things were a little complicated. The Oxford collegiate system was not conducive to this sort of course. Students need access to SPSS from their individual colleges and this (for a few cases) had proved a challenge to achieve. Other institutions, it was remarked, don't necessarily experience this sort of fragmentation. Clive suggested that a centralised solution would be the best way to proceed in the future. However once the software had been installed, students in general found it easy to get to grips with SPSS. The Windows version with its 'point-and-click' interface generally made data analysis painless and instant. The technology is at a stage where the old problems of complexity, capacity and interface are no longer an issue.

Conclusion

Clive was extremely positive about the course and hoped it would be expanded on in the future. The students themselves could see the relevance of such training, were largely motivated, and faced few problems in learning the technology. Clive stressed the importance of getting students to use real world datasets and not small overly specific studies. This enabled students to create valid results and see that the data they were using is the same that policy-researchers also use. The subsetting was therefore on variables, not cases. The major datasets also have a good literature to research. Clive admitted that care needs to be taken with regard to students that find the going tough. He suggested that the following year's course would contain an element that tackles the basics for the 10% or so who have particular difficulties.

There was some negativity towards registration procedures, which was entirely understandable. Clive argued for a class of datasets that should be in the public domain and freely accessible by users. He felt that the Data Archive's procedures, particularly for teachers, were counter-productive and wasn't convinced that this was necessarily always a demand from data depositors. If some data could be set-aside as free access, the problems of registration would go away.

Overall the course was a success, providing the technology to analyse large real-world datasets in a way that had been much more problematic in the past. The range of good quality datasets available also made tailoring to particular interests relatively straightforward. Students were confident and able to follow the training without many difficulties.

Websites

http://www.politics.ox.ac.uk/	Department of Politics
http://www.ssfc.ox.ac.uk/	Computing and Research Support Unit
http://www.ssfc.ox.ac.uk/datalib/index.htm	Data Library

Title: Training Courses in Multilevel Modelling

Author: R D Wiggins, PhD, C.Stat, City University, London

Subject: Dr Alastair Leyland, MRC Social and Public Health Sciences Unit, University of Glasgow (with Alice McCleod (Glasgow), Nigel Rice (York) and P Groenewegen (Utrecht))

Level of module: mainly post-doctoral, academic and government researchers in public health.

ContactTime: The courses held (during 1999-2000) vary from 1-day and 3-5 -day programmes. Altogether six 1-day courses were held, one 3-day course and two whole week courses were held during the year in a variety of locations (including Finland, Sweden and Holland).

Number of students: On average 20 per course. Typically, courses were free of charge. ESRC-ALCD underwrote four 1-day courses. Students are fortunate in that the MRC includes the dissemination of multilevel modelling as part of Dr Leyland's job description.

Aim of the course workshops

1-day:

Firstly, to introduce the practice of applying multilevel models in order to appreciate the use of simple variance component models, random regression coefficient modelling and the importance of contextual and compositional effects in population hierarchies. Secondly, to develop a critical appreciation of the application of multilevel modelling in the literature.

3-5 day:

To enhance the one day aims in order to achieve a deeper knowledge of the application of multilevel modelling to include non-linear modelling, reactive sessions to take forward student's own research problems ('fire-fighting') and hands-on applications (at least 8-hours training) using MLwiN and the local mortality dataset provided as part of TRAMSS (Teaching Resources and Materials for Social Scientists) – <http://tramss.data-archive.ac.uk>.

Course delivery and assessment

The 1-day courses are divided into 2 hours for formal lectures, a 1 hour discussion of a pre-assigned journal article and 3 hours of hands-on session in the use of MLwiN. Three to five day sessions are broadly balanced between formal and reactive sessions, group presentations and discussion and the critical evaluation of published material notably, articles issued in advance from the *Journal of Social Science and Medicine* and the *Journal of Epidemiology and Community Health*. One-day courses are not assessed. Wherever institutions require their research students to attend the longer courses for research training, certification has been developed. These students complete two assignments, one involving critical reading and the other a formulation of research problems as tasks involving the application of multilevel modelling techniques. The former involves issuing up to 15 abstracts at the beginning of the session and asking students to form groups of four or five, each group selecting a mutually exclusive abstract with the intention of obtaining the complete article and preparing a critical evaluation to present to the remaining groups. The latter requires individual students to systematically consider their own research questions in terms of the data they are planning to collect and the issues the analysis will raise for multilevel modelling. Volunteers are expected to present their research plans to the rest of the group. Tutorial support is targeted towards individuals during this endeavour and wherever appropriate sessions are scheduled to address questions of analysis, FAQ's about MLwiN, repeating aspects of the MLwiN practicals, consultation and group discussion. Of particular interest as an item of 'good practice', sessions in the longer courses involve a social theorist (Groenewegen) who helps students unpack the theoretical levels in their data and contrasts these with the ones operationalised for inclusion in any analysis.

TRAMSS is not used per se but sessions signpost its potential and mimic the analysis of worksheets using the local mortality data set. The courses are not presented as training sessions in the use of MLwiN but as general capacity building in multilevel modelling. Other software products for multilevel modelling (e.g. SAS PROC MIXED) and their parent websites are routinely issued.

Learning outcomes**1-day:**

At the end of the course students should be able to:

1. Read and understand a journal application involving the use of multilevel modelling.
2. Be able to communicate effectively with statisticians about the application of multilevel modelling.
3. Collaborate both orally and in writing about multilevel modelling with colleagues who are likely to have a range of statistical expertise.
4. Sustain the motivation to work independently on the application of multilevel modelling and with others more expert.

3-5 day:

All of the above at a deeper level of intellectual understanding as well as the practical know-how to use and apply multilevel modelling.

Reflections on the use of the TRAMSS resource

Clearly, opportunities exist around the resource to expand the training mode to include the modelling of recurrent events. However, of overwhelming importance is the need to lead applications by substantive examples that capture the research interest of participants.

Reflections on the teaching experience

Very positive. Highly task-orientated and motivated students help fuel a successful experience. Student evaluations demonstrate that the convenors actually improved as they delivered more workshops. If anything, 1-day courses were felt to be too short, 3-day courses too pressured (roughly, convenors attempt to cover the same amount of material in 3-days as in 5-days) and 5-day courses about right. During the 5-day courses students by their own violation use the evenings to prepare visual display material for the group sessions.

There are unlikely to be any more 1-day courses but two 5-day courses are planned this year.

Development of teaching materials

There is clearly work to be done to relate course material to the software product. In particular, on how to read data in to MLwiN and how to handle graphs and equation writers in MLwiN. There is an enormous amount of work that could be done to prepare more new data and exemplar analyses in medicine and public health. Given time these examples could be added to the TRAMSS site together with data source searching illustrations in BIRON. There was a clear recognition that such effort required time, money and expertise. The latter necessarily included the ability to write effectively, previous teaching and relevant software experience.

Support for teaching using numeric data

Dr. Leyland identified a clear need here. TRAMSS could be usefully augmented by the careful preparation of teaching datasets drawn from suitably health focussed sources, e.g. there might be example analyses on smoking cessation using BHPS [British Household Panel Study].

Sign-posting the future

With the growth in the application of multilevel modelling in social and health sciences there was a clear opportunity to enrich the TRAMSS model to include more data, suitably anchored in substantive areas that relates to individual researcher's interests.

An enquiry into the use of numeric data in learning & teaching

Part 2: Teachers' survey results

March, 2001
Revised June, 2001

by

Robin Rice (Project Manager)
Edinburgh University Data Library

and

Joan Fairgrieve (Statistical Analyst)
University of Edinburgh

with special acknowledgement to survey administrators:

Max Fox
Data Archive, University of Essex

and

Alistair Cairns
EDINA, Edinburgh University Data Library

Many thanks to our respondents for their participation and ideas.

Contents

Table of contents	41
Preface	43
Introduction	43
I. Contextual information about respondents	43
II. Current use of numeric data in teaching	44
III. Data for independent learning	55
IV. Support for using data in teaching—national services	56
V. Support for using data in teaching—local institutions	60
Conclusion	67
Appendix 1: Departmental categories	68
Appendix 2: Respondents' institutions & map	69
Appendix 3: Course names by type of course	70
Appendix 4: Names given for datasets by department type	72
Appendix 5: Recommend use of data to students — why or why not	75
Appendix 6: Changes in local data support (grouped by theme)	79
Appendix 7: Additional comments (grouped by theme)	82
Appendix 8: Covering letter and questionnaire	85

Preface

This document reports the results of a recently completed survey of teachers in UK higher education to the Task Force on the use of numeric data in learning and teaching. The Task Force is a diverse group of experienced academics and academic support staff from across the UK brought together as volunteer participants in the project of the same name. A preliminary analysis was presented in November 2000, and further specifications for analysis were given by the Task Force. After the survey was closed in January, 2001, another draft report was received by and discussed by the task force, with additional analyses requested. This is the final report of the Teacher's Survey results, and is one of the main deliverables of the project.

The Task Force will use the survey results and other information to deliver its recommendations to the JISC (Joint Information System Committee), the project's funder, in September 2001. This report, along with other information and outputs from the project, is available from the project home page, at <http://datalib.ed.ac.uk/projects/datateach.html>.

Introduction

Using a one in six sampling frame, 267 department heads were randomly selected from a universe of 1590 Heads of Department, from both within the Social Sciences and selected other departments. The list was acquired from the marketing company MARDEV, extracted from the *Worldwide Academic & Library File*. Department heads were asked to complete the survey themselves and pass it to relevant teaching colleagues to garner their participation. Two hundred-six responses were collected from 110 departments at 80 different institutions. Fifteen records were removed as ineligible (e.g. non-teaching department). With repeat telephone, email and postal follow-up requests to non-respondents, the final response rate (110 / 252) was 44% of departments sampled.

Heads of departments and lecturers were asked about their use of numeric data in teaching and supervising students, their experience of national data services, barriers to using data in teaching, and the extent of support available within their institutions for using datasets both for research and teaching. Complete results are given below, following the structure of the questionnaire.

I. Contextual information about respondents

Q1 Department name

Departments were categorised by 'inside social science' and 'outside social science' to aid parts of the analysis. Since any such categorisation would be somewhat arbitrary, the original classifications from the MARDEV *Worldwide Academic and Library File* used to construct the sample were retained. Some oddities—such as the inclusion of Environmental Sciences, but the exclusion of Geography—exist. See **Appendix 1** for department frequencies and categorisation.

Table 1: Numbers & percentage of departments in each category.

	%	Count
Inside social science	61.2	126
Outside social science	38.8	80
Total	100.0	206

Q2 Name of institution

The 206 respondents came from 80 different institutions. **Appendix 2** has the full list.

Q3 Speciality

Although the respondent was asked for his/her own discipline or subject speciality, this was not included in the analysis.

Q4 Job title

Thirty percent (61) were Lecturers, 25% (52) were Senior Lecturers, 29% (60) were Professors, 4% (8) were Readers, and 11% (23) did not answer or had another job title.

Q5 Teaching-related functions

Respondents were asked about their teaching roles in the department, ticking all that apply: Head of Department, Course Convenor, Class Lecturer, Supervisor of postgraduate students, Supervisor of undergraduate students, or none. Only one respondent had no teaching functions. About one-third of respondents were the head of their department, along with any teaching roles. (See Table 2.)

Table 2: Number of respondents who were heads of departments.

	%	Count
Head of Department	32	66
Not Head of Dept	68	140
Total	100	206

II. Current use of numeric data in teaching

This section was for course convenors or class lecturers only (185 respondents). They were asked to answer a set of questions about *one* of the courses they taught or convened.

Q6 Name of course

Course names given by respondents have been classified into ‘methods’ and ‘subject’ type-courses to aid analysis. Forty-eight (26%) were considered to be ‘methods’ courses and 132 (71%) were deemed to be ‘subject’ based courses. Five lecturers/convenors out of 185 (3%), did not answer this question. **Appendix 3** has the full list of course names grouped into categories.

Q7 Level of course

About two-thirds of the courses named were at the undergraduate level, and almost one-third were for postgraduates.

Table 3: Level of course.

	%	Count
Undergraduates	66	123
Postgraduates	31	58
Not answered or ‘other’	2	4
	100	185

Q8 Number of years teaching this course

Table 4 shows the spread of teaching experience of the survey respondents. The mean number of years was 6, median 4, longest time teaching was 30 years.

Table 4: Number of years teaching course.

	%	Count
Less than two	17	32
Two to five years	43	80
More than five years	36	66
Not answered	4	7
Total	100	185

Q9 Class size for this course

Table 5 shows the class sizes for the named courses. Unfortunately, 40+ was the largest category obtained from the survey instrument, although it is known that some courses have over 100 or 200 students.

Table 5: Class size for course.

	%	Count
Under 15 students (small)	18	33
15-40 students (medium)	39	72
Over 40 students (large)	41	76
Not answered	2	4
	100	185

Not surprisingly, more of the undergraduate courses were large, compared to the postgraduate courses.

Table 6: Class size by student level.

Row %	Class size			Count
	Small	Medium	Large	
Undergraduates	7	41	52	123
Postgraduates	41	38	21	58
Total (%)	18	40	42	181

Q10 Use of numeric data in this course

As Table 7 shows, 146 respondents had used data at some time in the named course—79% of all lecturers/class conveners in the survey. This outcome suggests that a self-selection process was occurring among those who chose to answer the survey, or were asked by their department heads to respond. Those whose courses were not quantitative or empirically-based tended not to respond to the survey. This means that while the survey succeeded in finding relevant targets, the results are undoubtedly skewed towards the responses of data users rather than non-users. This outcome was an inevitable result of the survey dissemination design, which looked for ‘relevant’ teaching staff in departments.

Table 7: Use of numeric data in course.

	%	Count
Nearly always do	43	79
Often do	15	28
Only occasionally	21	39
Haven't, but would like to	4	8
Never have and don't plan to	15	27
Not answered	2	4
Total	100	185

Chart 1 shows the proportion of responses without the four missing responses.

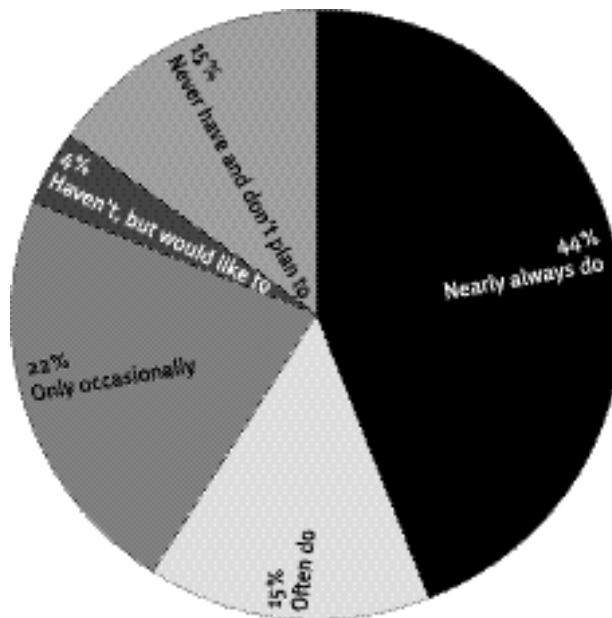


Chart1: Use of numeric data in this class by percent (n=181).

Somewhat surprisingly, no major differences were found in data use by post/under-graduate level courses, as shown in Table 8. Nor were there any meaningful differences found by class size, or in courses ‘inside’ vs. ‘outside’ the social sciences.

Table 8: Use of numeric data, by course level.

Col %	Undergrad	Postgrad	All
Nearly always do	45	42	44
Often do	15	18	16
Only occasionally	21	23	22
Haven't, but would like to	3	7	4
Never have and don't plan to	16	11	14
n =	123	57	180

Table 9 suggests that respondents teaching methods-type courses are more likely than those teaching subject-based courses to use numeric data in the course.

Table 9: Use of numeric data in class by course type.

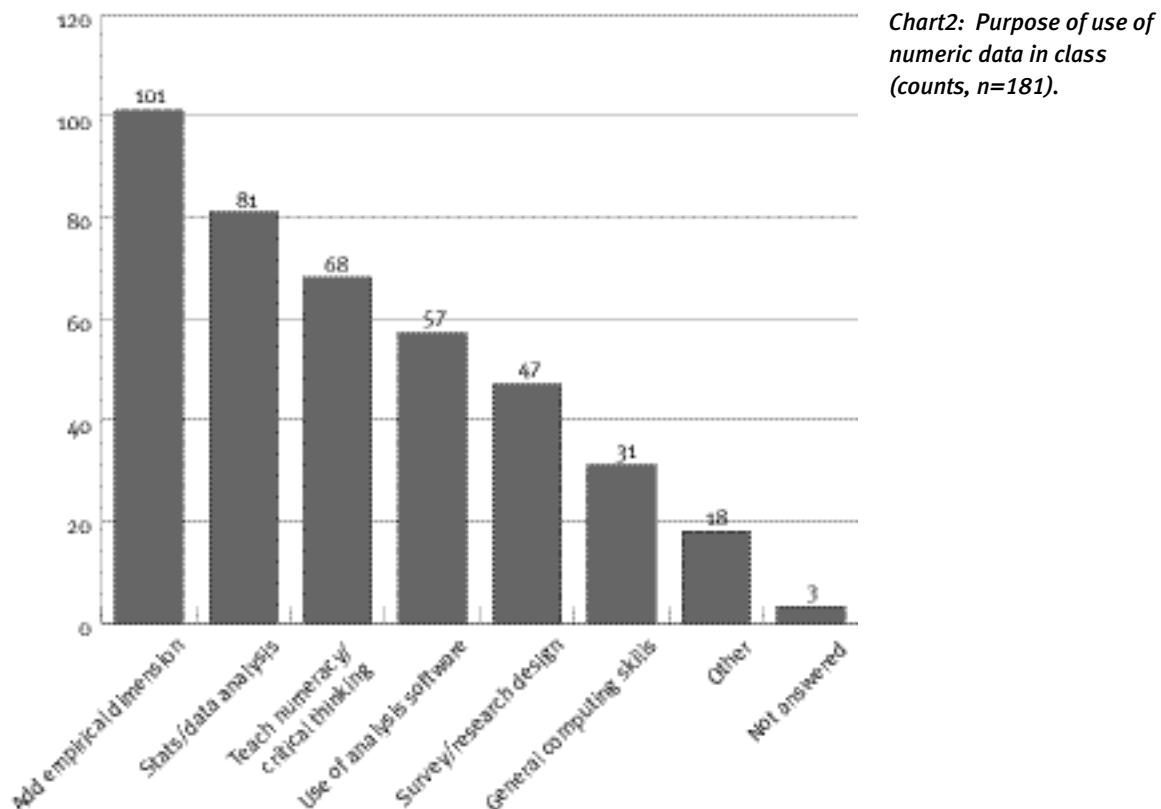
Col %	Methods	Subject	All
Nearly always do	72	34	44
Often do	17	15	16
Only occasionally	6	27	21
Haven't, but would like to	0	6	4
Never have and don't plan to	4	18	15
n =	47	132	179

Q11 Name of dataset

The survey deliberately avoided defining the terms data or dataset, so as to let the respondents answer according to their own definitions. However, some respondents objected to the question, “What are the data you use in this class called?” because they did not have names. Nevertheless, a wide variety of datasets were reported. **Appendix 4** shows the responses to this question, listed by department categorisation (inside/outside social sciences).

Q12 Purpose of use of numeric data in class

Teachers who used data in their courses were then asked to select one or more given responses that characterised their reasons for doing so (or to specify another), as shown in Chart 2. A majority, 56% of those teaching with data, ticked ‘To add an empirical dimension to the subject’. This was the sole purpose given by 23 people (16%). Nearly half, 45%, also used data “To teach statistics or data analysis methods.” More than a third, 38%, had another purpose: “To teach numeracy or critical thinking skills.” The average number of selected purposes was three.



Responses given by those who ticked ‘other’are listed below:

- All of the above
- Descriptive
- Empirical reflections on my theory
- Familiarise students with psychological data
- Illustrate Result
- Property investment performance analysis
- Show inadequacy of data coll. without qualitative approach
- Social History
- To highlight learning points in relation to particular lecture
- To inform about the economic environment of business
- To provide information in general
- To research the hierarchical structure of datasets
- To show that numbers are but one part of SA

- To teach GIS software
- To teach modelling
- To teach spreadsheets

Q13 Primary reason for using numeric data in class

When asked which was the primary purpose (see Table 10), the top two responses switched places, and ‘To teach statistics or data analysis methods’ became the highest selected response. This may reflect the large number proportion of methods-type courses that were represented in the survey.

Table 10: Primary reason for using numeric data in class.

	%	Count
Statistics or data analysis methods	34	50
Add empirical dimension to subject	27	39
Other	10	14
Numeracy or critical thinking skills	8	12
Survey or research design	6	9
General computing skills	2	3
Use of statistical analysis software	1	2
Total	100	134

‘Other’ primary reasons given were:

- Discipline Theory
- Primary reason is Critical Thinking
- a + d [statistics/data analysis + numeracy/critical thinking]
- To illustrate application of methods
- To quantify biological processes
- 1 and 3 [statistics/data analysis and teach use of statistical analysis software]

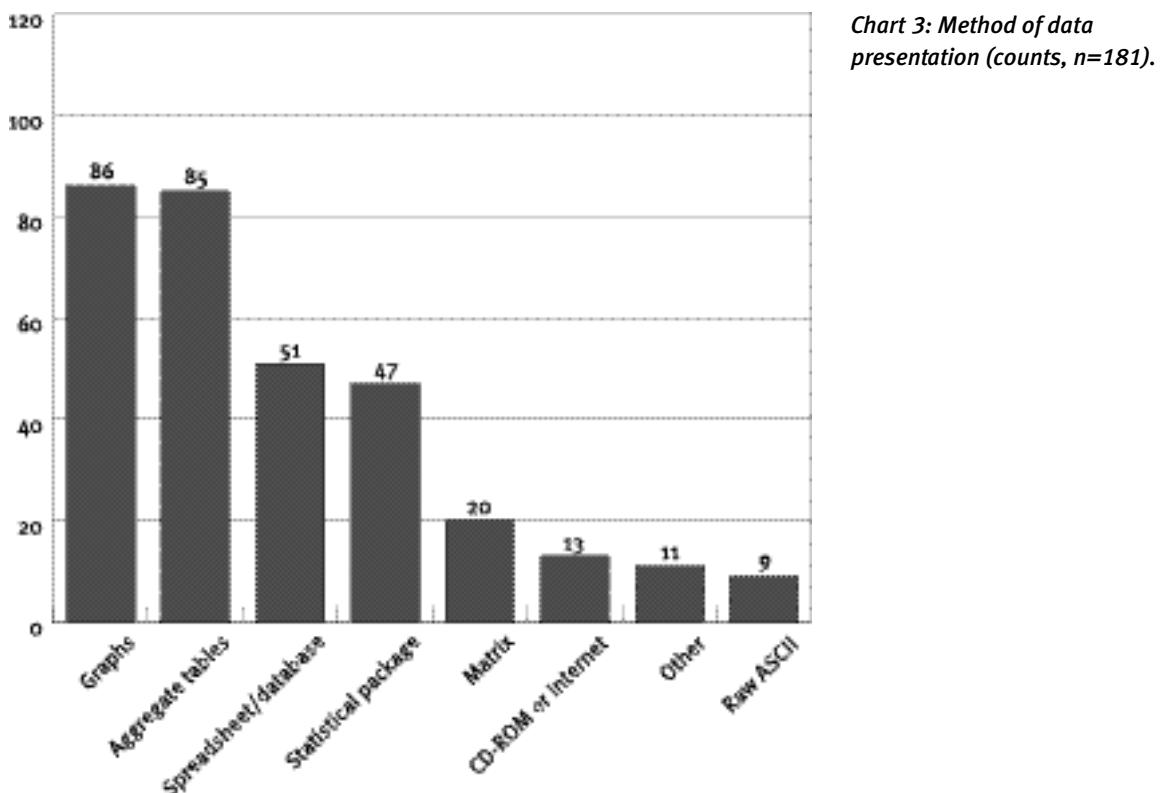
Table 11 suggests that the main reason given by those teaching methods courses was largely to teach statistics or data analysis, whereas the main reason for those teaching subject courses was more varied, but mainly to add an empirical dimension.

Table 11: Primary reason for use of numeric data by course type.

Col %	Methods	Subject	All
Statistics or data analysis methods	71	23	38
Survey or research design	10	5	7
Use of statistical analysis software	5	0	2
Numeracy or critical thinking skills	0	13	9
General computing skills	0	3	2
Add empirical dimension to subject	12	41	32
Other	2	14	11
n =	41	91	132

Q14 Method of presentation of data to class

The questionnaire also asked about the method of presentation, as a way of focusing on the level of difficulty introduced to students. This also helped to contextualise what was meant by ‘numeric data’ since the questionnaire avoided defining the concept in favour of respondents defining it for themselves. Again, more than one method could be selected. Nearly half of those teaching with data used graphs and aggregate tables as methods of presentation to the class (48% and 47% respectively). Twenty-eight percent used spreadsheet or database software. A statistical package was used by 26%. Data were presented as a case by variable ‘matrix’ by 11%, and as ‘raw’ ASCII file or number string by 5%. Only 7% used proprietary software on a CD-ROM or the Internet, and 6% used some other method. The average number of methods ticked was two, the maximum was eight.



‘Other’ methods of presenting data to the class that were given are listed below:

- As short lists of data
- Cardbox Plus
- during lectures
- Financial statements in a textbook
- Outside lectures refer to stats research
- Photocopied handouts
- Problems
- Published data from journals, etc.
- Raw data, indices etc
- Students create graphs etc. from this.
- Used as part of numerical demonstrations

Q15 Are students expected to work with the data on a computer ('hands-on') as part of their coursework?

Teachers who said they used data were asked if students were expected to work with the data on a computer ('hands-on') as part of their coursework. Slightly under two-thirds did, as Table 12 shows. This was virtually the same for both post and undergraduate classes.

Table 12: Whether course is hands-on, by course level.

	Undergrad	Postgrad	All
Hands-on	63	64	63
Not hands-on	37	36	37
n=	100	47	147

The proportion of hands-on courses was not affected by class size (presumably because large classes are broken down into smaller groups for computer practicals). However, methods courses had a much higher percentage of hands-on work than subject-based courses, as Table 13 shows.

Table 13: Whether course is hands-on, by course type.

Col %	Methods	Subject	All
Hands-on	85	54	64
Not hands-on	15	46	36
n =	46	100	146

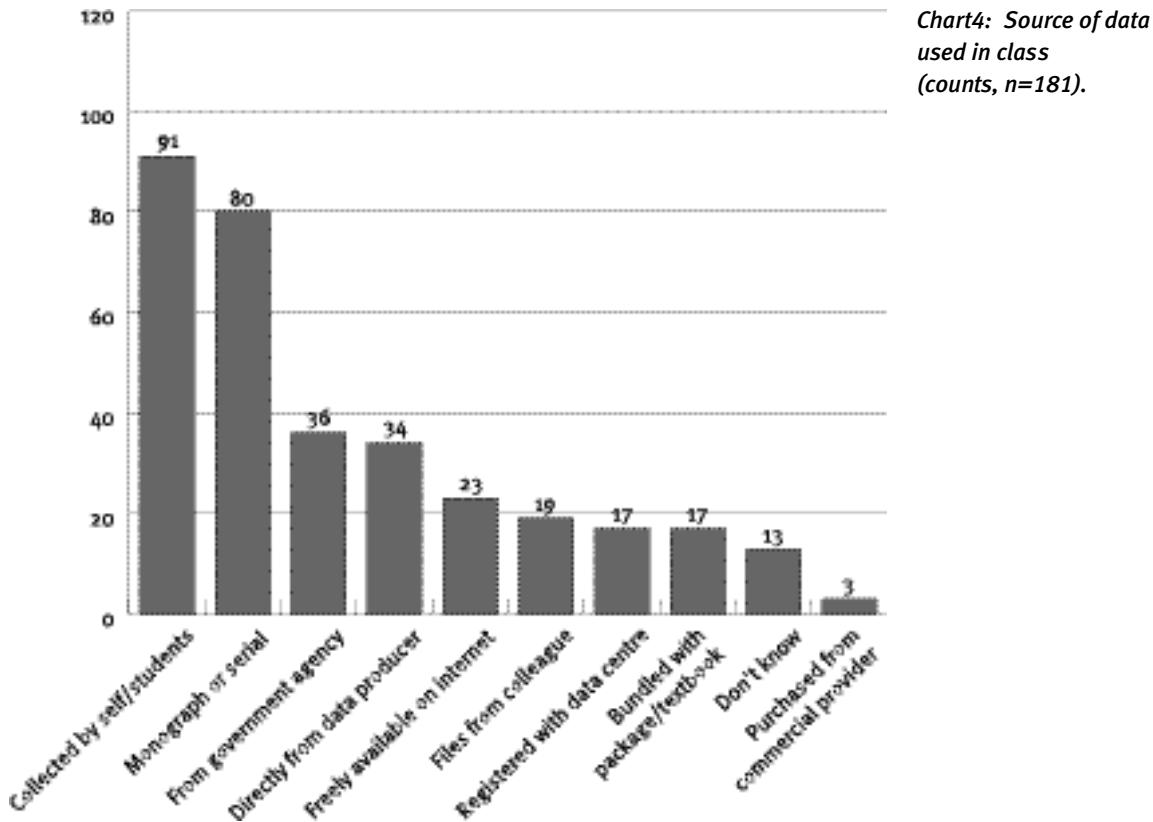
Table 14 combines the 'hands-on' responses with the teacher's methods of presenting data to the class. Graphs and tables, for example, can be easily shown to students on an overhead projector or in a handout. Asking the students to create graphs or tables in a computer program is a different educational experience.

Table 14: Hands-on ornot by method of presentation.

Row%	Hands-on	Not hands-on	Count
Raw ASCII	100	0	9
Case by variable matrix	95	5	20
Statistical package	89	11	47
Spreadsheet/database	86	14	51
Proprietary software on CD or Internet	62	38	13
Graphs	59	41	86
Aggregate tables	56	44	85
Other	36	64	11

Q16 Source of data used in class

Next, teachers were asked about the sources of data used for the course. The average number of sources used was two, the maximum was seven. Half of those who taught courses with numeric data used datasets collected by self or students (50%). Nearly half (44%) used data published in a monograph or serial. Each of the other sources were used by less than 20% of the teachers who use data. The proportions are showed in Chart 4. Twenty-five respondents used only data collected by self or students.



Specified sources are listed below:

Internet sources:

- CCDC communicable disease data
- Corporate Data
- Internet sources (cont'd):
- DETR, gov especially
- DETR/ONS/Environment Agency
- Eurostat
- FAO Database
- From anthropological sites and subject specific sites
- GIS files for environmental data
- <http://www.soc.surrey.ac.uk/uss>
- Human Genome Programme
- <http://cdiac.esd.ornl.gov/ftp/fossilfuel-co2-emissions/nation97.ems>
- NASA atmospheric data
- ONS
- open.gov
- QCA Key Skills
- Remote Sensing Imagery
- WHO NHS Cancer Registers etc
- www.ginasthma.com/

Purchased from data provider:

- Census on CD-ROM
- Met Office
- Spons [sic]

Bundled with textbook:

- Begg Economics
- Howell: Fundamental statistics for behavioural sciences
- CD ROM accompanying textbook “Statistics for Managers using Excel” (Prentice-Hall)
- Microfit
- SAGE test data
- SPSS: Analysis without anguish, Coakes and Stead 1999
- standard Microsoft package
- Starting SPSS
- Textbook Disk

Other:

- Generally simulated by me
- Included in SPSS package
- Ethnographic data/qualitative information processed according to the anthropological methodology
- Nat. Asthma Campaign data, Drug Refs – Al— & Harbing?
- School Research Work
- Self Created
- Simulated/made up
- SimulateP
- Synthetic data
- Unpublished manuscript sources

The data sources were cross-tabulated with whether the course was hands-on, to see which types of sources were more likely to be used in hands-on coursework.

Table 15: Source of data, by whether course hands-on.

	Col %	Hands-on	Not hands-on	All
Published monograph or serial	47	77	57	
Collected by self or students	68	65	67	
Obtained from data producer	28	18	24	
Files copied from colleague	19	4	14	
Local or national govt. agency	22	35	27	
Registered for access from data centre	17	4	12	
Freely available from Internet	14	20	16	
Purchased from data provider	1	4	2	
Bundled with textbook or software	17	6	13	
Don't know/other	9	12	10	
n =	90	51	141	

Percentages sum to more than 100 as more than one source could be given.

As Table 16 shows, there were no great differences in sources used by level of course.

Table 16: Source of data, by course level

Col %	Undergrad	Post-grad	All
Published monograph or serial	59	56	59
Collected by self or students	67	63	66
Obtained from data producer	18	35	24
Files copied from colleague	13	15	13
Local or national government agency	26	28	27
Registered for access from data centre	9	17	12
Freely available from internet	18	13	16
Purchased from data provider	3	0	2
Bundled with textbook or software	12	13	12
Don't know/other	12	7	10
n =	96	46	142

Percentages sum to more than 100 as more than one source could be given.

Q17 Time taken to prepare data

Three questions were asked to determine the extent to which teachers who used data felt that the preparation of datasets for teaching was a burden. The first of these asked “How much time did it take to prepare the data for course in total?” The open-ended responses were grouped as shown in Table 17.

Table 17: Time taken to prepare data for teaching.

	%	Count
Already prepared	8	12
Number of hours	23	33
Number of days or weeks	12	17
Other	8	13
Don't know	8	11
Not answered	41	60
All	100	146

Q18 How much of a burden was data preparation?

The next question asked respondents directly the extent to which they regarded data preparation a burden in teaching. Chart 5 shows the proportions of the responses (5 eligible respondents did not answer the question).

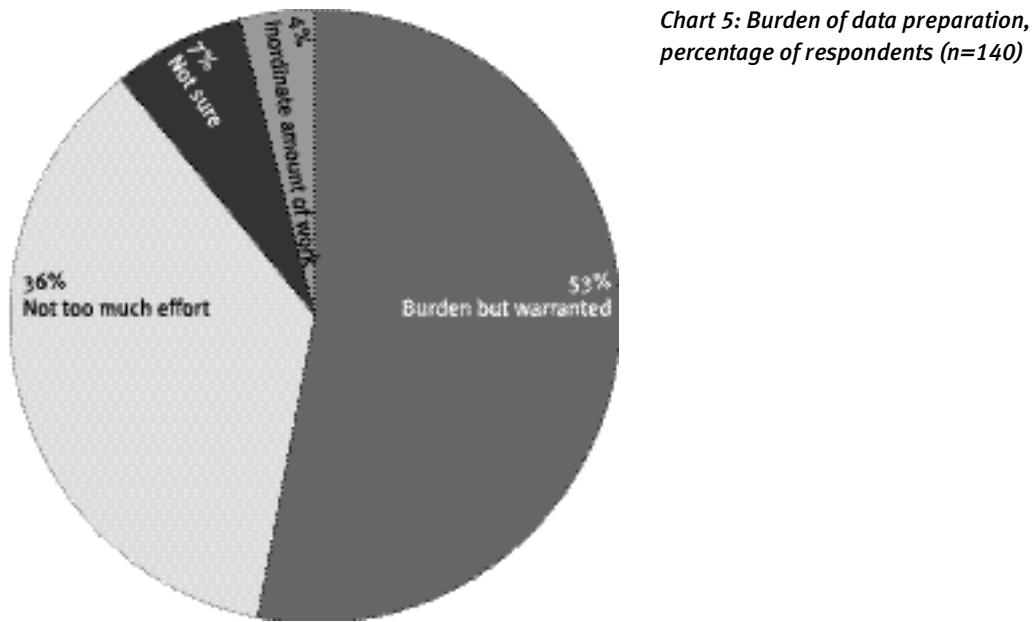


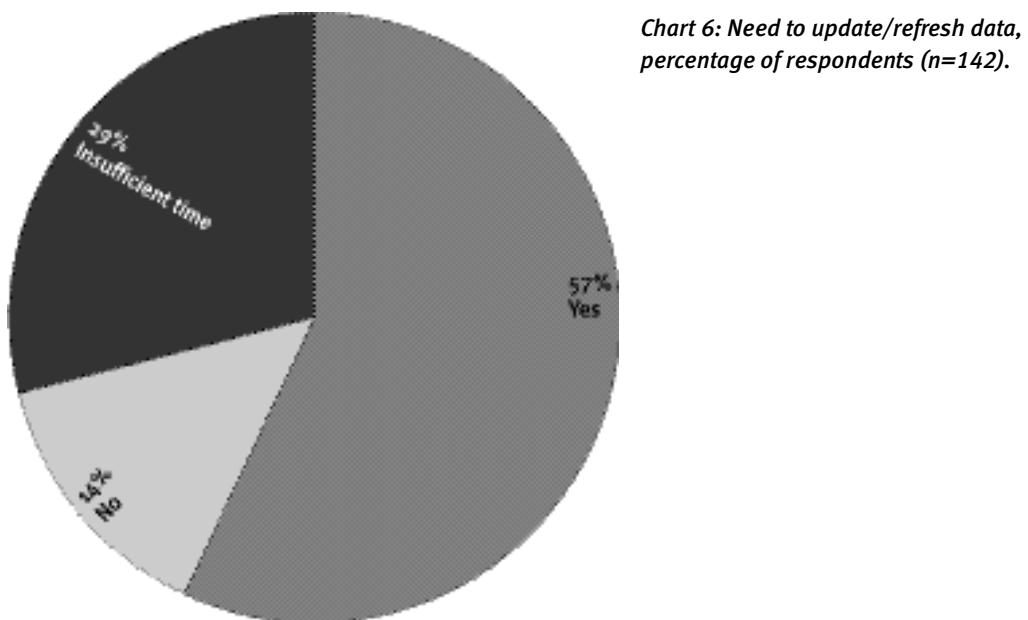
Table 18 shows a slightly greater burden felt by those who teach hands-on courses.

Table 18: Burden of data preparation, by whether course hands-on

	Col %	Hands-on	Not hands-on	All
An inordinate amount of work	5	0	4	
Not too much effort involved	32	41	35	
Yes, but warranted	59	43	54	
Not sure	3	16	8	
	n =	93	49	142

Q19 Need to update/refresh/revise the data used on a regular basis?

As the pie graph in Chart 6 shows, most respondents, 86%, felt the need to update / refresh / revise the data used on a regular basis. However, 29% felt there was insufficient time to do so. (Four out of 146 did not answer.)



No relevant differences were found between the hands-on courses and the others.

Table 19: Need to update data, by whether course hands-on.

Col %	Hands-on	Not hands-on	All
Yes	59	54	57
No	14	13	14
Yes, but insufficient time	27	33	29
n =	92	52	144

III. Data for independent learning

Q20 Respondents who supervise students undertaking independent research

This short section of the questionnaire dealt with staff in their role as supervisors of students' independent learning activities. Most (91%) of the respondents did supervise students in undertaking independent research.

Table 20: Level of students supervised.

	%	Count
None	6	13
Undergraduate only	16	33
Postgraduate only	26	53
Both	49	100
Not answered	3	7
Total	100	206

The remainder of this section was completed only by the 187 respondents who supervised students undertaking research.

Q21 Do you recommend the use of data to students for research?

Most supervisors (89%) did sometimes recommend use of data sources to students in their independent learning (Table 21).

Table 21: Recommend use of numeric data for students' independent research.

	%	Count
Nearly always do	35	65
Often do	33	61
Only occasionally	21	39
Never have and don't plan to	6	12
Haven't yet but would like to	2	4
Not answered	3	6
Total	100	187

There were not many differences in answers between undergraduate and postgraduate supervisors (Table 22).

Table 22: Recommend use of numeric data by level of students supervised.

Col %	U'grad only	P'grad only	Both	All
Nearly always do	36	41	32	36
Often do	36	31	34	34
Only occasionally	12	16	28	22
Never have and don't plan to	6	10	5	7
Haven't yet but would like to	9	2	0	2
n =	33	51	96	180

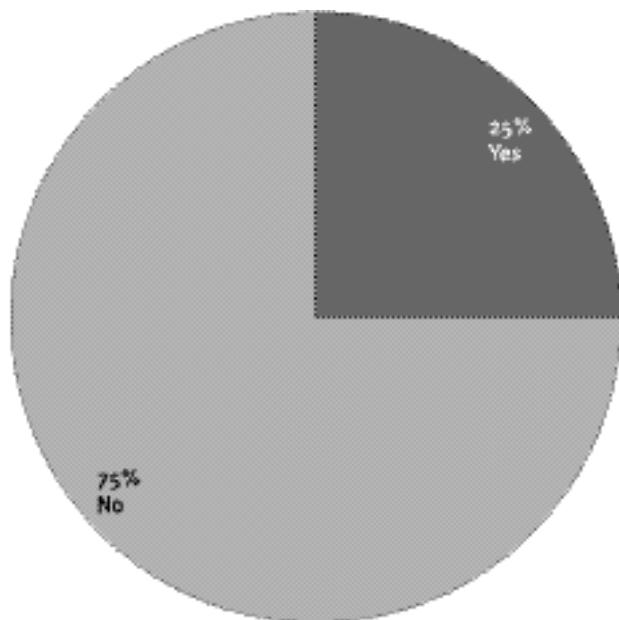
Q22 Why/why not recommend data

Appendix 5 has respondents' answers in full, grouped by their response to the previous question. Below are 'typical' responses for each category.

- *Nearly always do* (35%): "Statements made need to be backed up with evidence – often of an empirical nature."
- *Often do* (33%): "Depends on topic, but statistical sources can contextualise a topic."
- *Only occasionally* (21%): "Many students are more inclined to qualitative research."
- *Never have and don't plan to* (6%): "Not relevant to what I am teaching."
- *Haven't yet but would like to* (2%): "Not always appropriate & [I am] insufficiently briefed on numeric data available."

IV. Support for using data in teaching – national services**Q23 Ever used or considered using national data services to access numeric data for learning and teaching purposes?**

All respondents (n = 206) were asked if they have ever used or considered using national data services (i.e. The Data Archive, EDINA, MIMAS) to access numeric data for learning and teaching purposes. All but 10 answered the question. An overwhelming three quarters (147) had not; only 25% (49) said yes. Considering the skew towards data users in this sample, explained earlier, it is surprising that such a large number have never considered using the national data resources in teaching.

Chart7: Considered using national data centres, percentage of respondents (n = 196).

The number of years the respondent had been teaching a course for did not affect the likelihood of the use of national services, as Table 23 shows.

Table 23: Considered using data services by years teaching course.

(Column %)	Two years or less	Three to five years	More than five years
Used or considered using	26	26	28
Not considered using	74	74	72
Total	100	100	100

Table 24 shows a greater inclination to make use of national data services in courses that are hands-on (35%).

Table 24: Considered using data services by whether course hands-on.

Col %	Hands-on	Not hands-on
Considered using	35	23
Not considered using	65	77
Total	100	100

For those who “nearly always” use data in teaching, an even greater number—45% had used or considered using the national data services. For those who use data less often, the numbers dropped well below one-quarter, as shown in Table 25.

Table 25: Considered using data services by use of data in class.

Row %	Yes	No	Count
Nearly always do	45	55	80
Often do	18	82	28
Only occasionally	14	86	37
Haven’t, but would like to	14	86	7
Never have and don’t plan to	4	96	27
All	27	73	179

Surprisingly, there was a greater tendency for undergraduate supervisors to have considered national data services than postgraduate supervisors (Table 26). But the numbers outside of the “both” category are too small to draw any meaningful conclusions from this.

Table 26: Considered using data services by level of students supervised.

Row %	Yes	No	Count
None	8	92	12
Undergraduate only	34	66	32
Postgraduate only	27	73	51
Both	23	77	98
Total	25	75	193

The methods teachers were more likely to be national data service users (40%) than the subject teachers (21%), as Table 27 shows.

Table 27: Considered using data services by type of course.

Col %	Methods	Subject	All
Yes	40	21	26
No	60	79	74
n =	48	127	175

Those courses deemed to be ‘inside’ the Social Sciences were a bit more likely to have considered using national data sources (28%) than those ‘outside’ (21%).

Table 28: Considered using data services by department type.

Row %	Yes	No	Count
‘Inside’ soc sci	28	72	119
‘Outside’ soc sci	21	79	77
Total	25	75	196

Q24 Ranked barriers to using national data services

Respondents who were familiar with the national data services (n=46) were then asked to rank eight factors regarding the extent to which they act as barriers in using national data services for learning & teaching purposes. Table 29 shows the frequency of each ‘barrier’ ranked first, second, third, etc., and the ‘raw’ cumulative scores for each barrier.

Table 29: Ranking of barriers (number of respondents giving each ranking).

	1st	2nd	3rd	4th	5th	6th	7th	8th	Total score*
Lack of time for preparation	13	17	7	2	4	2	2	0	301
Lack of awareness of materials	21	11	1	3	4	4	2	0	298
Registration procedures	12	9	5	7	2	2	7	1	253
Format of datasets	4	1	9	8	8	4	5	0	187
Lack of teaching subsets	4	6	3	6	5	8	10	0	186
Interface	4	6	4	8	5	7	1	1	182
Documentation	1	3	9	7	9	6	3	0	178
Other	1	1	0	0	0	0	0	1	16

* if barrier was ranked 1st then a score of 8 was given, if 2nd then 7 and so on.

More usefully perhaps, Table 30 shows the mean scores for each barrier, in descending order, since many respondents did not rank all of the barriers. The barrier with the greatest mean ranking was “Lack of awareness of relevant materials.” The second greatest barrier was “Lack of sufficient time for preparation.” And the third greatest barrier was “Registration procedures” [of the national data services]. However, the other barriers received great enough scores to also be taken seriously.

Table 30: Average ranking of barriers (8=highest score, 1=lowest).

	Mean score	Median score
Lack of awareness of materials	6.5	7
Lack of time for preparation	6.4	7
Registration procedures	5.6	6
Interface	5.0	5
Format of datasets	4.8	5
Documentation	4.6	5
Lack of teaching subsets	4.4	5

'Other' barriers given by respondents are listed below:

- Don't know
- All are low priority
- Cost
- Have not considered using these sources
- Lack of expertise
- No barriers
- Not had info on how to do it easily available
- Ready availability of data from other sources
- These datasets are too detailed for an introductory course
- Time lapse before access to data is given

Q25 What changes are needed to national data services?

Users were then asked if they would like to see a change in services at the national level for using numeric data. If so, what are the most important improvements these services can make to support teachers and learners in the use of data? Thirty-six out of 46 eligible respondents provided answers, which have been grouped into six categories.

Interface/format

- Able to get data without learning special software
- Easier access and greater dissemination of data available
- Easy access and compatibility with different software
- Getting data in right format for the software that will be used is a problem.
- Reform data collection e.g. Published 1951 Census more useful than 10% 1991 version in some respects.
- [In answer to Q18] I would regard it as burdensome as I had to do the SIR retrievals from MIMAS myself!

Provide teaching datasets w/ supplementary documentation

- Cut out the registration hassle, make data sets customisable and downloadable directly, supply support teaching material, a 2000 equivalent of the Surrey EBS project? ["Exploring British Society" was a set of teaching datasets and materials developed at Surrey for the *General Household Surveys* series in the 1980s.]
- Ease of access to a teaching subset
- Easy to access datasets that would be interesting for students to use.
- No. I use LFS and GHS in research but would require time series data for teaching.
- Provision of teaching subsets (e.g. from BSA or BES) [British Social Attitudes Survey, British Election Studies]
- e.g. BHPS [British Household Panel Study] from the Data Archive, doc is great for research, easy to extract. A few series for teaching but dauntingly large for students 'independent research', subsets would be useful.
- Rapid access to key summary economic data in form tailored for teaching

- Sample data with supportive documentation for teaching purposes
- Use of numeric data should include a revision of the methods for collecting this data. Num. data should offer complementary qualitative info.

Simplify registration procedures

- Faster registration – Can it be done online?
- Improve speed of registration and downloading
- Easy access and registration facilities are required. Quick response would be greatly appreciated.
- Make registration procedures simple and abolish restrictions on use (e.g. all students signing disclaimers)

Raise awareness of national data services

- Initially, making users aware of products and support.
- Not aware of existing services at Nat. level
- Regularly be in contact with people like me.
- The initiative needs to come from the Nat. Services but better publicity would be a start.

None/Don't know

- Don't know
- n/a
- No
- No change seems fine to me
- Insufficient experience to answer this question.
- Not enough experience to comment
- Not enough information to form a judgement.
- Not sufficiently aware to make an informed comment
- [In answer to Q26] MIMAS and UKBORDERS staff are very helpful
- Reasonably good – much improved from 4-5 years ago!
- Time is my main problem

Other improvements

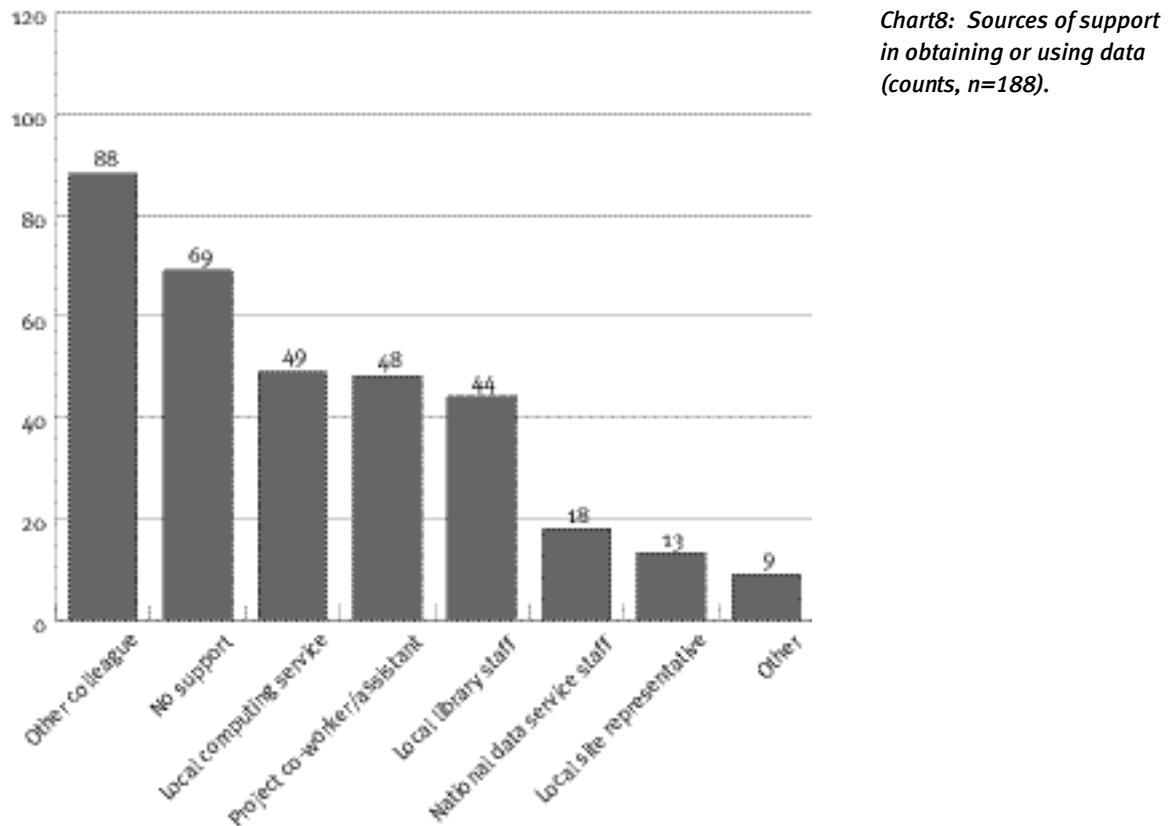
- In-service training; tutorial support - online or by distance learning.
- One source of data.
- Overall shift to a consistent and open system.
- Widen availability through libraries?
- Suggest better collaboration between Site reps, computing staff and library staff, converging in a more centralised system within the institution. There is no need for three points of contact in an electronic era. I had to contact 7 individuals before being granted access to library journals, datasets, software and national datasets. Two more applications are still pending... I would like to see one registration form only, involving all the data resources available for teaching. Lecturers and students should be able to tick the relevant boxes and obtain a single user-ID that follows them while they remain within registered institutions.

V Support for using data in teaching – local institutions

Q26 Sources of support for data use

This section was asked of the entire sample (206); 18 did not respond to this question, “From whom have respondents ever had support in obtaining or using data, whether for teaching or for research?” Of those who did, more than a third, (37%) had received no support at all. Again, more than one source could be ticked; the average number of sources of support received was two, the maximum was seven. Chart 8 illustrates the frequencies of response. Peer support was the most common form, either from a project co-worker/assistant or another colleague

(26% and 47% respectively). The local computing service (26%) was roughly matched with the local library service (23%) which had helped about a quarter of respondents each. National service staff provided help to 10% of respondents, and their local site representatives only helped 7% of them.



'Other' forms of support are listed below:

- Archivists
- [Department] Colleagues
- [Institution's] Student Statistics Team
- Industry
- International collaborators
- No
- Proprietary Data
- University
- W3

For those who had used/considered national data services, Table 31 shows a higher proportion of support from national data services staff and site representatives (27% each). Fewer of these users received no support (only 18%). All other sources of support were reported higher for these users, particularly local computing service staff, which was twice as high as for those who did not use the national services.

Table 31: Sources of support by whether used or considered using national data services.

Col %	Used/considered	Not used/considered	All
National data services staff	27	4	10
National data services local rep.	27	1	7
Co-worker/assistant on course or project	31	24	26
Another academic colleague	56	45	48
Local computing service staff	40	20	25
Local library staff	24	23	23
Other	7	4	4
No support received	18	43	37
n =	45	138	183

Percentages sum to more than 100 as more than one source could be given.

Those with hands-on courses also obtained higher levels of most forms of support.

Table 32: Sources of support, by whether course hands-on.

Col %	Hands-on	Not hands-on	All
National data services staff	15	6	12
National data services local rep.	11	4	9
Co-worker/assistant on course or project	35	18	29
Another academic colleague	53	51	52
Local computing service staff	33	16	27
Local library staff	28	25	27
Other	2	6	4
No support received	28	31	29
n =	89	49	138

Percentages sum to more than 100 as more than one source could be given.

According to Table 33, methods courses relied more on national data services staff and local representatives (not surprising since a higher proportion were users), whilst subject-based courses show a higher level of support from local computing service staff and librarians.

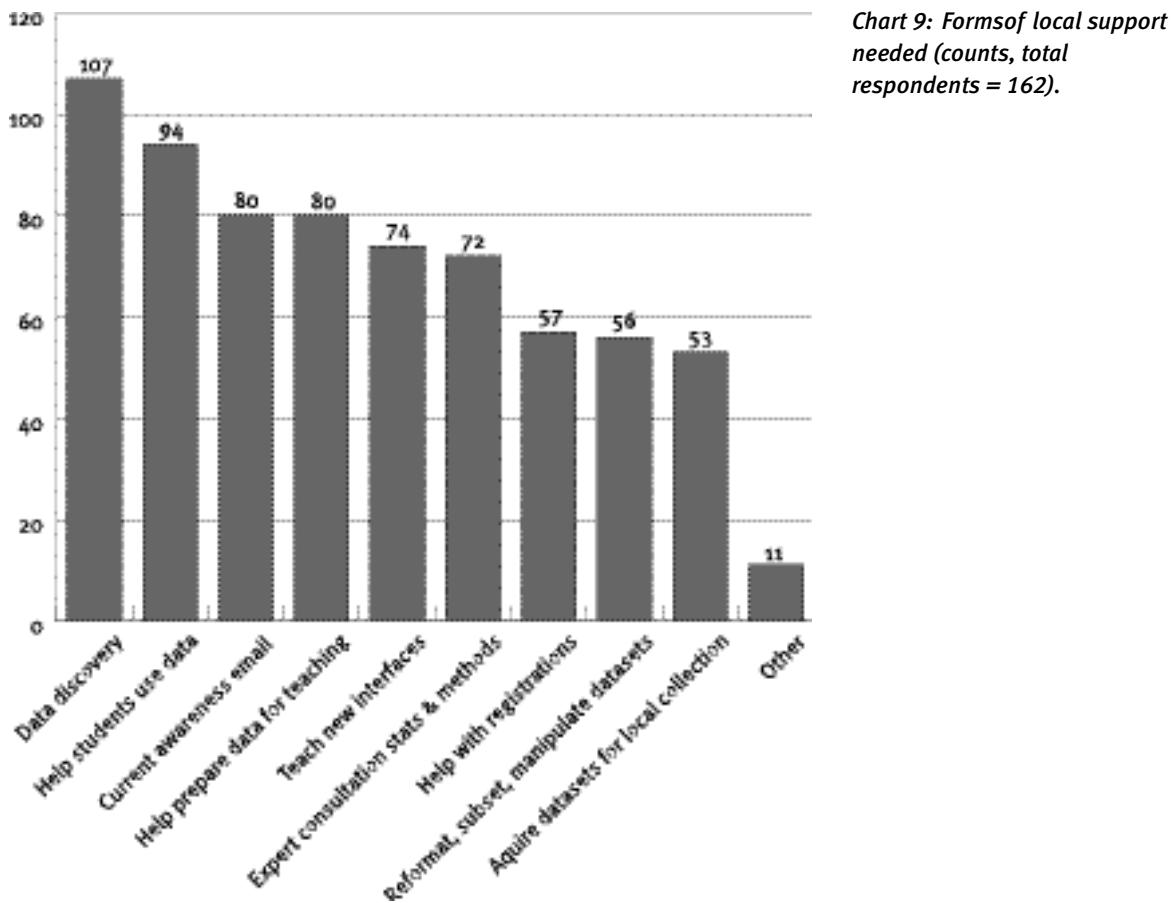
Table 33: Sources of support, by type of course.

	Col %	Methods	Subject	All
National data services staff	16	7	10	
National data services local rep.	16	4	7	
Co-worker/assistant on course or project	30	25	26	
Another academic colleague	44	46	46	
Local computing service staff	19	25	23	
Local library staff	14	26	23	
Other	2	4	4	
No support received	33	40	38	
	n =	43	122	165

Percentages sum to more than 100 as more than one source could be given.

Q27 What forms of local support are needed by academic data users?

Those who responded to this question (162 or 79% of total) reinforced the need for a number of forms of locally provided support, above all 'Data discovery / locating sources' (66%). All of the answers shown in Chart 9 received 'votes' from between one-third and two-thirds of those responding (except 'other'). The average number of needs ticked was three, the maximum was 10.



Q28 Which is the most important form of local data support needed?

Respondents were then asked to prioritise a single form of support above the others (see Table 34). One third of respondents (67) did not answer this question. Of those who did, the highest ranking form of support again was ‘Data discovery / locating sources,’ followed by ‘Help students use data for learning and research,’ and third, ‘Provide expert consultation for statistics & methods.’ Interestingly, although 35% said they would like ‘Help with registration procedures’ in the first question, only 1% deemed it to be the most important form of support needed.

Table 34: Most important form of local data support needed.

	%	Count
Data discovery/locating sources	18	37
Help students use data	11	22
Expert consultation statistics & methods	10	21
Help prepare data for teaching	7	14
Current awareness email	7	14
Teach new interfaces	4	9
Reformat, subset, manipulate datasets	4	9
Other	4	8
Acquire datasets for local collection	1	3
Help with registrations	1	2
Not answered	33	67
Total	100	206

‘Other’ answers are as follows:

- a) is most important but f) takes a lot of time.
[a = data discovery/locating sources, f = reformat, subset, manipulate datasets]
- All of them
- Most important local support is [local department name]
- All are important
- Depends on the individual.
- Hard to say
- Equally important
- First Hand Research
- From Library/Learning Centre Staff
- I don’t use this
- Library and Computing

There were only a couple of apparent differences based on course level (Table 35). A somewhat greater importance was placed on data discovery with undergraduate courses (31% vs. 20%), and on assistance in preparation of teaching datasets with postgraduate courses (16% vs. 9%).

Table 35: Most important local support needed, by level of course.

	Col %	Undergrad	Postgrad	All
Data discovery/locating sources	31	20	27	
Help with registration	0	4	2	
Teach use of new interfaces	6	9	7	
Acquire datasets for local collection	3	2	2	
Current awareness via email	11	2	8	
Reformat, subset or manipulate datasets	5	9	6	
Expert consultation for stats & methods	14	16	14	
Assist in preparation of data for teaching	9	16	11	
Help students use data	17	16	17	
Other	5	7	6	
n =	81	45	126	

There was a slightly greater call to “Provide expert consultation for statistics & methods” and to “Help students use data for learning and research” among those whose courses are not hands-on, as shown in Table 36.

Table 36: Most important local support needed, by whether course hands-on.

	Col %	Hands-on	Not hands-on	All
Data discovery/locating sources	28	29	28	
Help with registration	3	0	2	
Teach use of new interfaces	7	6	6	
Acquire datasets for local collection	1	6	3	
Current awareness via email	5	9	6	
Reformat, subset or manipulate datasets	11	0	7	
Expert consultation for stats & methods	13	17	15	
Assist in preparation of data for teaching	11	11	11	
Help students use data	16	20	17	
Other	5	3	5	
n =	75	35	110	

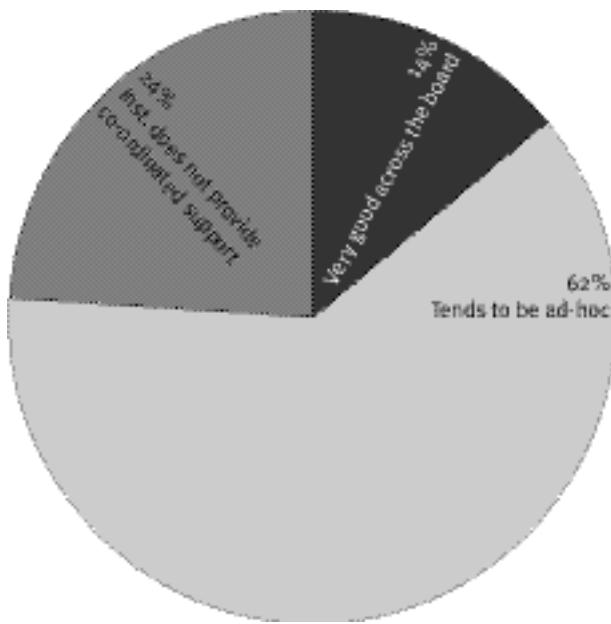
There were some differences between the Heads of Department who responded and the others, including a greater need for a current awareness service via email (see Table 37).

Table 37: Most important form of data support needed, by role in department.

Col %	HoD	Not HoD	Count
Data discovery/locating sources	27	26	37
Help students use data	22	13	22
Current awareness email	18	7	14
Expert consultation statistics & methods	10	17	21
Help prepare data for teaching	5	12	14
Reformat, subset, manipulate datasets	7	6	9
Acquire datasets for local collection	5	1	3
Other	5	6	8
Teach new interfaces	0	9	9
Help with registrations	0	2	2
Total	100	100	139

Q29 Level of support at institution for data use in teaching and research

Respondents were also asked about their perception of local data support currently provided. As illustrated in Chart 10, nearly one-quarter of those responding agreed with the statement, “In general, my institution does not provide co-ordinated local data support. The majority (62%) felt that “Local data support tends to be ad-hoc (e.g. from colleagues, general support services, etc.),” and only 14% felt that “Local data support here is very good across the board.”

**Chart 10: Level of local support provided, percentage of respondents (n = 176).**

Less of those who taught hands-on courses felt the level of data support was ‘very good across the board’ (9% vs. 19% not hands-on). Similarly, more of those who taught hands-on courses characterised data support in their institutions as ad-hoc.

Table 38: Level of local data support, by whether course hands-on.

	Col %	Hands-on	Not hands-on	All
Very good across the board	9	19	13	
Tends to be ad-hoc	69	58	65	
My institution does not provide coordinated support	22	23	22	
n =	88	48	136	

More Heads of Department characterised their institutions as not providing co-ordinated support than other respondents (29% vs. 22%).

Table 39: Level of local data support, by role of respondent.

	Col %	HoD	Not HoD	All
Very good across the board	12	14	14	
Tends to be ad-hoc	59	64	62	
My institution does not provide coordinated support	29	22	24	
n =	58	118	176	

Q30 Change in support given by local institution for using data in L&T

Respondents were then asked in an open-ended question if they would like to see any change in local support for the use data resources in learning and teaching. See **Appendix 6** for the responses, grouped by the following themes which were coded by the authors:

- No changes or no specific suggestions
- Resource or organisational-related issues
- Staffing or service-related issues
- Training-related issues

Q31 Case study volunteers

Respondents were asked if they were willing to volunteer to be subjects of case studies, or if they wished to recommend a colleague. The Task Force selected a handful of subjects — teachers in various situations — to interview and if possible, observe first-hand in the classroom, in order to add a qualitative aspect to the study and to further an in-depth understanding of the issues addressed in the survey. The case studies are available for viewing on the project website (URL given below).

Q32 Additional comments

A final open-ended question asked respondents if they had any additional comments to inform the survey. Thirty-six responses were given and are listed in **Appendix 7**. These are grouped by the following themes:

- Comments on the survey itself
- Questioning assumptions about numeric data
- Comments about one's self
- General suggestions on the survey topic

Conclusion

This report of findings forms one part of the entire report of the project on The Use Numeric Data in Learning and Teaching. Interested readers may obtain other documents including Task Force recommendations and case studies of teaching scenarios from the project website: <http://datalib.ed.ac.uk/projects/data teach.html>. The project team also welcomes comments; email datalib@ed.ac.uk.

Appendix 1: Departmental categories

In the interest of preserving confidentiality, specific department names of respondents are not given. Responding departments were categorised into two groups according to the original sampling frame provided by the MARDEV *Worldwide Academic & Library File* as follows:

Social Science departments (excluding Military Science)

Anthropology
Business & Management Studies
Communications
Economics
Education
Environmental Sciences
International Relations
Political Science
Psychology
Social Studies (including Social Work & Social Policy)
Sociology

Universe: 1223. Respondents: 126.

Selected non-Social Science departments

Agriculture
Archaeology
Epidemiology
Geography
History
Nursing & Ancillary Services
Public Health
Statistics
Town Planning

Universe: 367. Respondents: 80.

Appendix 2: Respondents' institutions & map

Bell College
 Birkbeck College
 Bournemouth University
 Bradford College
 Brunel University
 Cardiff University
 City University London
 College of Ripon & York St John
 De Montfort University
 European Business School
 Glasgow Caledonian University
 Goldsmiths College
 Heriot-Watt University
 Keele University
 King's College London
 Lancaster University
 Leeds Metropolitan University
 Liverpool John Moores University
 London Guildhall University
 London School of Hygiene & Tropical Medicine
 Luton University
 Manchester Metropolitan University
 Nottingham Trent University
 Open University
 Oxford Brookes University
 Paisley University
 Queen's University of Belfast
 Queen Mary and Westfield
 Royal Holloway, University of London
 Scottish Agricultural College
 Sheffield Hallam University
 South Bank University
 Southampton Institute
 St. Georges Hospital Medical School
 Strathclyde University
 The Robert Gordon University, Aberdeen
 University College Northampton
 University College Worcester
 University of Aberdeen
 University of Abertay, Dundee
 University of Bath
 University of Birmingham
 University of Bradford
 University of Brighton
 University of Central Lancashire
 University of Dundee
 University of Durham
 University of Edinburgh
 University of Essex
 University of Glasgow
 University of Greenwich
 University of Hertfordshire

University of Huddersfield	University of Sheffield
University of Hull	University of St. Andrews
University of Kent	University of Stirling
University of Leeds	University of Ulster
University of Leicester	University of Wales
University of Liverpool	University of Wales Lampeter
University of Manchester	University of Wales, Bangor
University of Newcastle	University of Wales, College of Medicine
University of North London	University of Wales, Glamorgan
University of Northumbria	University of Wales, Swansea
University of Paisley	University of Warwick
University of Plymouth	University of Westminster
University of Reading	University of Wolverhampton
University of Salford	University of York



Appendix 3: Course names by type of course

The authors grouped the course names provided by respondents into ‘methods’based courses and ‘subject’based courses, as follows. The question was asked of all course convenors/lecturers (n=185). Five eligible respondents did not answer this question. Each course name has a frequency of one, unless a number appears beside the title.

‘Methods’ courses (48):

Applied Demography	Quantitative Methods – applied econometrics component
Data Analysis (2)	Quantitative Research Methods (2)
Doctoral training (PhD. 1st Year) via MSc in Social Research	Quantitative Methods
Econometric Analysis of Panel Data	Research Methods (5)
Econometric methods	Research Methods and Issues
Economic Modelling	Research Methods I
Economic Statistics and Econometrics (2)	Research Methods in Criminology
Empirical Research Methods in the Social Sciences	Research Methods in health care
Gen'lized linear models & survival anal	Research Methods in Human Geography
Generalised line statistical modelling	Research Methods in Psychology
Generalised Linear Modelling	Research Methods in Psychology II
Initial data exploration and analysis in the social sciences	Research Methods/Analysis
Introduction to Research Methods (2)	Research Sources on Methods
Investigative Methods (2)	Spatial Analysis Statistics
Mathematical Statistics	Survey methods and data
Probability and Statistics II	Survival Analysis with Stata; Applications of Data Analysis
Quantitative Methods for Accountancy A/B	The Research Process
Quantitative Concepts & Methods for Geoxxxxxx	Time Series Models
Quantitative economics	Using Economic Data

‘Subject’ courses (132):

Accounting Information Systems	BSc Transport and Logistics Management
Administration and Politics of the EU	BSc(Hons) Midwifery Studies IRM
AFC 308 Financial Statement Analysis	BSc(Hons) Psychology
Agriculture and Conservation in Europe	Business Analysis
An Introduction to Criminal Justice	Business Forecasting
ANDS	Business Information Systems and Data Analysis
Asthma	Childhood through the ages
Atmospheric Pollution	Constructions of Childhood
B.Med.Sci (Hons) Nursing Studies	Consumption Cultures
BA Community & Criminal Justice	Contemporary Moral Issues
BA Community and Council function	Core Skills in Geography II
BA Hons in Sport Studies & Physical Education	Development Politics
BA(Hons) Tourism Studies	Diploma in social work
BEng(Hons) in Civil Engineering	Diploma in Social Work
Biological Science	Diploma of Higher Education ‘Mental Health Nursing’
BmedSci Nursing	Earth Sciences for the Environment 1
BSc (Hons) Nursary Studies	Economics
BSc Geographical Skills	Ed.D in Evidence based Policy and Practice
BSc Psychiatry	Environmental and community biology 1

'Subject' courses (cont'd):

Environmental Anthropology	PgDip/MA in Education for Health + Social Care Practice
Environmental Impact Assessment	Physical Science for the Environment 2
Europe's Countrysides	Physiological Psychology
Financial Management	Political Anthropology
Foundations of Finance + Advanced Corporate Finance	Political Thought: Hobbs to Mill
General Linguistics	Politics and the Environment
Geological occurrences of British Gold	Postgrad Diploma in Personnel Management
Global Finance	Power Modernity and Social Theory
GY271 GIS1	Practical Econometrics
Health and Behaviour	Pricing Policy for Underdeveloped Countries
Health Care Management	Principal Educational Studies
Health Policy	Property Investment and Performance Analysis
Health Psychology	PS377 Psychological Research Design and Analysis
Healthcare Management	Psychobiology
Historical Research	Psychology and the Law
History of Computing	Psychometrics
History of European Construction	Public opinion and policy making: Issues and Cases
Igneous Modelling	Registered nurse diploma in nursing
Industrial Revolution	Representation and Textual Analysis
Into to the EU	Research Appreciation
Intro to stochastic processes	Research Skills in Biology
Introductory Computer Applications	Scandinavia and the modern world
ITSkills	Scotland and India 1695–1813
Italian Fascism	Scottish Society 1750–1950 (2)
Linear Programming [xxxx] application	Several but one is research methods in human geography
Local societies and economics	Social Theory and Work
M.Med Sci.	Society, Belief and Culture in Early Modern Europe
Masters and Doctorate in Counselling Psychology	Sociological Debates
Masters in Health and Hospital Management	Sociology of Organizations
MBA Financial Management	Sources and Methods in Economic History
MbChB MED in Lifelong Learning	Special Subject: England at war 1290-1360
MEd Learning and Teaching	Specialist Community Nursing
Medical Demography	Strategic Management
Meteorology	Symmetry and Groups
MSc Collaborative Health Care	Teaching and Learning in HE
MSc in Environmental Protection	The Anthropology of Europe
MSc Palliative Nursing Care	The Darwinian Revolution
MSc. in Actuarial Science	The Physical Landscape (Physical Techniques)
Occupational Therapy	The UK Economy
Occupational Therapy BSc(Hons)	Tides and Modelling
Operations Management	Topics in Macroeconomics
Pain (module)	Urban Structure
PG Dip/MSc Leisure + Tourism Policy and Practice	Using Spreadsheets
PGCE IT and PGCE Maths	Visual Anthropology
PGCE Primary	Vocational Education
PGCE Secondary	Writing and Audiences

Appendix 4: Names given for datasets by department type

'Inside' Social Sciences:

- 1991 Census (Area Statistics) Vital reg. data (Birth/Deaths) Migration data (var.) Demographic Health Surveys (BHS)
- Basic economic, political, environmental statistics
- BHPS
- BHPS, GSOEP, JCHP
- Biological Statistical Data; Life-Tables, Quadrant Samples
- Cash flows, financial market prices and returns
- Census, Eurostats, Welsh Office data, SPSS, Nimbus
- Civil Engineering cost databases (eg Spons)
- CPU file name
- Criminal Statistics, British Crime Surveys, Judicial Statistics
- Data from experimental work in science
- Data generated by research activities - chemical analysis
- Data on Greater London drawn from a variety of statistical sources
- Data! Various datasets, inc. student practicals used within a research methods course
- Data/stats on environmental pollution or number of social groups supporting particular ideas/policies
- Digital Maps and Census Data
- Do not understand the question
- Does not understand Q.
- don't understand
- Don't understand the question
- Educational Statistics
- ethnic stats
- EXCEL spreadsheet exercises; SAGE integrated accounting software test data.
- FAO Price data series, production data
- Financial statements and the accounting information contained therein
- GHS95
- Government statistics Survey data generated from student body.
- Housing Needs Survey
- I don't understand this question – they are just datasets
- I provide many small datasets for class calculation. Not called anything.
- I'm sorry I don't understand this question. I don't usually christen my data.
- Indices, stock market data sets, risk analysis data sets.
- International Economic Statistics
- It is simulated data
- Macro-economic and financial data
- Macroeconomic data – e.g. GDP, inflation, unemployment etc.
- Many small datasets
- Mathematics and IT
- National UK CSO Statistics
- NHS resource utilisation data
- No names. I use a number of data sets which I have collected from my own research.
- No Specific Name
- Numbers!!

- Observations
- Occasional demographic statistics
- Official Statistics, Panel Data (e.g. BHPS), Eurostat etc.
- On education policies and practices
- ONS Economic Trends
- Personnel data of various kinds
- primarily drawn from OECD or IMF
- Project uses data each year... lectures use a mix of generated data and macroeconomic time series -
- Reaction time Data, Accuracy Survey/Questionnaire (Likert scales)
- Research Methods – Visual Data - Number Games
- Security returns, Option prices, Simulated auction, Simulated negotiation.
- Social Statistics, Official Statistics, Research Data
- SPSS Data Files
- Standardized Mortality Data
- Statistical Information usually for tabular form.
- Statistics
- Statistics
- Statistics
- Statistics
- Subjects' responses (e.g., reaction times, memory scores, signal detection parameters (d' and B), frequencies, most empirical psychological measures are involved at some time or other.
- Survey data
- Survey Data and Performance Data
- Survey/questionnaire data
- Survey: What determines exam success
- TEATS
- Texts, media materials – visual, audio, verbal
- They are mainly Eviews data files containing a variety of time series macroeconomic variables.
- time series data for illustrating the estimation of consumption functions; cross-section data for illustrating the use of logit/probit type model estimations.
- Time series of earnings and job titles for an internal labour market.
- UK Macroeconomic Data
- Usually government published statistical data e.g. numbers of children on child protection registers - annual returns etc.
- Variable data e.g. risk assessment results, opinion poll results, etc.
- Varied
- Various
- Various
- Various – primarily small ($n \leq 30$) example data set for worked example exercises.
- Various datasets covering various aspects of business eg sales figures, profit, cost, etc.
- Various major social and political UK surveys, UK census
- Various: World Bank, UNDP Country Stats. and Dev. Indicators, economic data, soc. survey data.
- World Development Indicators (2000) and (1995) and General Household Survey (1995)

'Outside' Social Sciences:

- Case studies
- Census
- Census Data
- Census Data, SARS
- Demographic, Employment Data
- Depends on Student led projects
- Descriptive Statistics
- Don't understand the question – I teach several classes using data sets ranging from national CSO sets to small packets of data from local companies connected with my consultancy interests.
- Economic Growth using the Industrial Revolution.
- Embodied Energy Data
- Epidemiology + Statistical data from various sources
- Epidemiology Research Compliance
- Experimental Data
- experimental results, physical constants, properties of substances, molecules and atoms
- Fiscal and General Economic Data
- Geochemical Analysis
- Geographic Data
- I call it 'Shoe Size'
- I don't understand the question, I use lots of data sets from a variety of sources.
- I use both demographic and epidemiological data
- Illustrative Examples
- Mainly government produced statistics from various publications and internet sites
- Meta-analysis
- No name so call them Global Carbon Balance
- Not at all sure what you're after here. I call my data "data" – what do you call yours? No single data returned.
- Numeracy Skills
- Primary data from staff research
- Quantitative data pertinent to midwifery practice/obstetrics
- Research data for use with SPSS
- Several data sets but one based on reg. soc-econ. data from Eurostat
- Social Class Surveys
- SPSS Package
- Statistical data sets
- Storm Discharge & TDS, Salt Marsh (various) – vegn., soil properties, distance & altitude
- There are many (mainly small) data sets
- They don't have names
- Use various sets of (usually) simulated data to illustrate techniques
- Usually quantitative findings from resources
- Variety of samples, mainly from GRO reports

Appendix 5: Recommend use of data to students – why or why not

The following open-ended answers to Q22, “Why or why not?” are grouped by their answers to the previous question, “Do you recommend that your students use numeric data in the course of their independent research?” Fifty-four of the 187 eligible respondents did not answer this question (29%).

Nearly always recommend:

- ?
- All of the research I supervise is experimental.
- Always a rule, even in qualitative research e.g. for pop. profile of field area.
- analyse and discuss
- Appropriate cognitive psychology research nearly always involves numerical data
- Balanced research in theoretical development and search of empirical evidence
- Because it is essential for the topic in hand.
- Because many important questions require a quantitative approach if they are to be securely addressed, to allow structured comparisons etc.
- Because the areas I am interested in - social work, social policy, services for children - are the areas that I supervise and they lend themselves to the use of such data.
- Because the topics (memory, fatigue, attention) dictate empirical data
- Chemical information reveal processes in genesis of rocks
- Data analysis is an essential part of any science course.
- Data collected is quantitative
- Empirical research in my areas of psychology almost always involves numeric data
- Essential element of research process - handling and interpretation of data
- Essential for Comprehension
- Essential for economic analysis
- Essential for economic research
- I advise students that (particularly in the area of Business and Economics) statements made need to be backed up with evidence – often of an empirical nature.
- I am interested in quantifying natural processes
- I believe all management reports need some numerical info
- I believe quantitative data leads to the most objective investigation.
- Interesting and do-able
- It ensures there is an independent objective measure for analysis and discussion
- it is an essential part of the study
- It is necessary in remote sensing and GIS research
- It is very much a function of the research topics they are dealing with
- It is what applied economics is about.
- Most students develop topics that require some data collection & analysis
- Motivation/Checking
- Nature of discipline – nat. sci. research is usually quantifiable at some stage
- Nature of discipline means num. data usually collected.
- Observations and measurements are often numeric; numbers are objective.
- Ours is ultimately a quantitative, empirical discipline.
- Part of the course requirements
- Psychology is an experimental discipline. Students collect a variety of types of data which enable them to quantify and analyse their results to determine their statistical significance, for hypothesis testing.
- Quantitative dimension key element of much research and students need to acquire evidence in all forms, quantitative and qualitative, to support argument, generalisation etc.

Nearly always recommend (cont'd):

- Quantitative methodologies
- Students are registered on an Environmental Science degree programme and so need to be able to obtain, manipulate and draw conclusions from data.
- The course is a quantitative one that requires familiarity with data
- The Nature of their work
- The sorts of topics our students deal with require some discussion as to the extent of particular crime issues. Looking at/for trends in issues. Numeric descriptions of populations etc.
- They need to demonstrate that they can analyse data.
- To motivate and illustration methods. To demonstrate understanding of formulae.
- To substantiate theory and argument
- Training students in use of data and packages is invaluable, for the postgraduate students I teach this follows a heavy theoretical econometrics course, learning is far more effective with hands on econometrics. For the undergraduate students I teach the course is taken prior to the core Econometric theory course. My course was introduced to enable students to understand, at an intuitive level, what econometrics is about – it is a “leg up” to the full course.
- Undergrad projects need to be constrained, clear and relatively easy to
- Using secondary data enables a quantitative form which is seldom possible with student collected data which tends to be more limited in size and tends to be a qualitative analysis.

Often recommend:

- Able to contrast it with qualitative data and to see patterns in large cohorts
- Appropriate to the subject
- As a tool for analysis/to provide descriptive statistics
- As appropriate depending on objectives of research
- As appropriate to problem in hand.
- By default all my students will be using numeric data since I only supervise students who are engaged in carrying out empirical work.
- Demonstrates ability to apply theory, so get better results.
- Dependent on what their research question is
- Depends (obviously!) on the nature of the research project!
- Depends on the nature of the dissertation research. Valid statistical analysis improves credibility of findings.
- Depends on the nature of the research. Support many who do quantitative research.
- Depends on the Student and the Project
- Depends on topic, but statistical sources can contextualise a topic.
- Enhances range of critical analysis and transferable skills.
- I suggest students use numerical data if that is what they want to do. The projects (both U/G and P/G) tend to be student led.
- If it is appropriate for the research question
- Illustrative examples
- Important to research
- Impossible really to do research and not use some type of numerical data.
- It depends on the topic, numeric data is not always appropriate
- It depends on what they are researching
- It is (they are!) fundamental to most aspects of business analysis
- It is essential
- It is relevant to the question they seek to answer!
- It is so important to health care/nursing

Often recommend (cont'd):

- Most prefer qualitative methods
- Needed for GIS based projects.
- Often a balance of quantitative and qualitative info is required in dissertation
- Often, numerical data is essential in the analysis of a question/problem. But, sometimes it isn't eg case study / qualitative study.
- Projects are decided by students – I supervise qualitative and quantitative
- Provides a complete coverage of the topic.
- Quantitative analysis provides more influential and publishable results
- Recommended only if relevant to research question(s) being addressed.
- Strong believer in triangulation of research through both quantitative and qualitative data.
- That's what science is all about.
- The 'basic science' of public health practice is quantitative, i.e. epidemiology. The undergraduate medical 'MBChB' curriculum here is truly problem-based and student-centred and so we do not 'lecture' to the medical students but facilitate their small group learning around paper-based clinical scenarios in each of their 2 week modules throughout the curriculum. There are triggers in their PBL scenarios for researching quantitative data and their assessments involve quantitative data handling.
- The epidemiological studies the dept. carries out, require it.
- The nature of the research often dictates evidence which should be measured.
- The use of numeric data depends on the specific needs of the student's research.
- To demonstrate their capacity for data use and manipulation and collection
- To establish empirical regularities and test hypotheses
- To indicate their powers of critical analysis / to illustrate their knowledge of relevant data sources.
- To reinforce and motivate statistical methodology and to interpret the data.
- To support arguments/claims
- We run an applied statistics degree
- Where it is relevant to the problem!
- Yes if data is relevant to research being undertaken

Only occasionally recommend:

- Anthropology emphasises the use of qualitative, non-numerical data
- Depends entirely on the Research Question
- Depends on the subject of research.
- I am basically a qualitative researcher
- I recommend it if the research approach requires it.
- I recommend it when it is relevant to the subject
- If appropriate and depending on the focus of the research.
- If it is appropriate for their research topic
- If it is relevant then I recommend numerical data
- if they were periodically updated – at present such data is only available
- It is not usually relevant to their research questions.
- Mainly qualitative work
- Many students are more inclined to qualitative research
- MBA dissertation uses numeric data if appropriate to question being pursued.
- Only occasionally recommend (cont'd):
- Moral analysis rarely depends upon this. Relevant for student dissertations?
- My emphasis (+most students I suppose) is on qualitative data.

Only occasionally recommend (cont'd):

- My own research background and skills are qualitative.
- My own research interests are primarily theoretical.
- Nature of the subject that I supervise
- Not often relevant to their research which relies on qualitative methods.
- Numeric data is often not appropriate or unavailable
- Question or area of enquiry does not often warrant a quantitative approach
- Recommended as appropriate to area of study
- scattered through numerous journals.
- Students nearly always engaged in small scale community based qualitative research
- The cost dataset is only relevant to a few design projects; more fundamental experimental research does not need the datasets. However, new datasets on say typical ranges of values of design parameters would be used.

Never recommend and don't plan to:

- Background of students + ocean science topic
- Has little relationship to the subjects I teach + interested in little data for
- History of ideas doesn't lend itself to a numerical approach.
- Irrelevant to the history of ideas.
- It's almost never relevant, and when it might be (e.g. in practical ethics topics) students tend to get mesmerized by statistics and quit the philosophy.
- My analytical focus is qualitative
- Not Applicable
- Not appropriate in the branch of mathematics
- Not my area of supervision or research competence
- Not relevant to what I am teaching
- Students that I supervise largely carry out qualitative research studies. My knowledge of numeric data is very limited and whilst I personally do not teach it, it is included in the course.

Haven't recommended yet but would like to:

- Lack of resources
- Normally they collect own numeric data, now I will suggest existing data sets.
- Not always appropriate & insufficiently briefed on numeric data available
- The material (mod. int. history) does not lend itself to numerical analysis.

Appendix 6: Changes in local data support (grouped by theme)

After the question about the level of support provided by the respondent's institution, the following open-ended question was asked: "Would you like to see a change in support given by your institution for using data resources? If so, what are the most important improvements **your institution** can make to support teachers and learners in the use of data?" Ninety-three respondents provided answers to this open-ended question. They have been grouped according to theme by the authors.

No changes or no specific suggestions:

- [Dept. name] is largely not a teaching unit as such. We have a masters in ODE which doesn't yet use large data sets. [Dept. name] does generate all the quantitative data this [institution] uses for student progress, retention and levels of satisfaction, so our staff provide support to units for QA and to a small extent for other data [xxx] as in teaching which is your main concern here.
- A more co-ordinated approach would be nice, but in practice responsibility for preparing numeric data for teaching comes down to the individual teachers who wish to utilise this material. No sign this will alter!
- Any more help would be useful
- Difficult to say as new to job
- Don't know
- Hard to be specific, to be honest I don't have a full sense of what you are asking about.
- Have not given this any thought.
- Help is available as and when required.
- I do not know enough about the data currently available
- I have all the support I need
- It has to be user driven
- It may be more effective than I think as I am only an occasional user of numerical 'data'. I would not therefore wish to comment.
- It would be rather impractical given the extent of its present use. [numeric data?]
- May not be aware of current provision locally. Biggest problem is tendency for Home Office and others to change the presentation of data from year to year which often precludes analysis of trends.
- My needs are specialist. Support not necessary or feasible
- no
- No
- No
- No
- No it is up to the lecturers to decide their own programme.
- No real problem at present
- No significant need in my area.
- No support – veys
- No, all the help one needs is available
- No. Resources are precious, could wish to know what was being traded for such support.
- None Needed
- Not a priority at present
- Not at present
- Not for my own purposes.
- Not necessarily feasible at uni. level, if dept. lacks resources/will
- Not really – I seem to manage very well.
- Not really. By the time I've interacted with yet another human being I could have accessed suitable data by myself.
- Not sure
- Not the sort of support I would expect at local level, other perhaps than funding for transcribing data.

No changes or no specific suggestions (cont'd):

- Our data use is small, notwithstanding prep. time, not clear that central support is appropriate
- Our students generate their own data, through practicals or project work, and we have little call for centrally supplied datasets.
- Rising journal costs mean prioritising resources.
- Seems fine to me, don't really need much support
- Where are the resources
- Yes

Resource or organisational-related issues:

- Awareness of availability, accessibility in MS Office compatible (readable) software.
- Better funding
- Central web-based facility plus faculty assistance.
- Centrally co-ordinated
- Co-ordination in support for staff and students.
- Data usage tends to be School-specific, so across the board support is difficult to envisage
- Develop on-line statistical information
- Facilities for collecting data more easily (computer marking) and for backing up data archives
- Generally sound – Maybe more standardised step by step ‘idiot’ guides
- I may not ask as much as I should. But an easily accessible register would be helpful.
- I would like to see support for use of relational databases rather than numeric. More useful for Historians.
- Library staff being pro-active rather than re-active to the needs of academic community
- More computer support
- More emphasis on qualitative methods of data collection: participant observation, interviews, fieldwork.
- More or dedicated computer suite to deliver data and applied methods rather than joint hardware or tutorial computing assistance, the future will be driven by content.
- Not sure, perhaps a resource centre.
- Qualitative gathering of data must be added to courses dealing with data resources in the field of humanities. Support grants for research for students.
- The provision of more resources.
- Using numeric data and computers in class in learning and teaching is relatively labour intensive. A premium to allow additional support would be very helpful.
- V. hard to find out who to ask and takes time needed elsewhere. Clearer gateways into resources would help a lot, esp. for people using/need data beyond their research fields.

Staffing or service-related issues:

- Appoint someone in the department to collect/gather the relevant data and set it up in a suitable form for analyses – teaching.
- Appointment of a specific ‘information’/Data Officer
- Assistance in data formatting and analysis to save time.
- Assistance in preparation of data for teaching, help for students using data for learning and research, datasets reformatted, subsetted and manipulated for the user; and help with registration procedures.
- Awareness of sources, their access and backup for users.
- Better support services from support staff primarily
- Centralised advice service on design and statistical methods
- Co-ordinated view of teaching qualitative methodologies – appointment of e.g. health statistician
- Data discovery, database maintenance and data preparation for teaching.
- Dataset prep. help possibly, but only in conjunction with the academic teacher, to provide necessary time to set up teaching sets.

Staffing or service-related issues (cont'd):

- Discuss requirements on a one-to-one basis and provide a custom made service; especially in relation to distance learning developments.
- Do 27 a,b,f [data discovery/locating sources; help with registration procedures; reformat, subset, manipulate datasets for the user]
- Greater co-ordination, clear policy on acquisition, newsletter or email list to relate new availability and/or suggestions.
- Raise awareness of how data can be used for teaching, learning and research.
- Raising awareness and providing consultation on using data resources.
- Raising awareness of types of accessible data
- Second me to develop a support facility.
- See 27e above [Current awareness service via email]
- Service is good but possibly understaffed. Greater staff availability would make service more accessible and possibly more widely publicised.
- Specific appointment of someone for data prep & skills in setting up teaching data sets on intranet. Teaching workshops are easily disturbed by tech. access/connex probs. Lect. lacks time or skills to deal with this.
- The number of people using or needing support is limited. Those people should be identified and helped on a mentor/mentee system.

Training-related issues:

- Awareness of and training in the use of data resources. However, this is merely one of a number of areas of staff development which are neglected because of a general lack of resources (funds).
- Become more proactive in tracking down sources; where course provision for teachers is appropriate, more helpful scheduling of these.
- Initially learner-centred short courses. I am v. under-confident & unskilled in using data resources, which, considering what I teach, needs urgent attention.
- Make aware what is available, provide support and training – need examples and case studies to get started.
- More preparation time. Study leave to attend refresher courses. Relevant staff-development courses.
- More staff development on use of numeric data for learning and teaching in social sciences
- Organise staff development courses, either in the faculty of education (too small?) or the main university
- Organise workshops and give support for IT skills
- Provide in-service for lecturers
- Training courses/workshops for lecturing staff
- Training days
- Training plus items from Q27

Appendix 7: Additional comments to inform survey (grouped by theme)

Thirty-six respondents (17%) chose to answer this optional open-ended question. Comments are grouped into four themes below.

Comments on the survey itself

- Failed to return originally as objected to unreasonable request to recruit and chase respondents.
- I notice from the covering letter (dated 31 July, and sent to a colleague) received by me today (8 Aug), that you would like to see all questionnaires returned within two weeks. This would seem to be optimistic at any time but, particularly, during 'vacation' periods.
- I think I misunderstood the meaning of course convenor in q5 (especially as it was placed next to class lecturer). I am course leader (convenor?) for a whole degree programme (course) as well as a module (class?) lecturer on several components of the programme. Therefore I have general and specific data requirements across a wide range of subjects. Your questionnaire seems to focus on a single class (module?) eventually – sorry!
- Perhaps I am missing something but I cannot really see the point of this survey
- Questions/answers in 27-29 not clear. Also I'm unclear as to what, "support for using data..." means? Perhaps I don't "use data in teaching" and hence don't understand. But there is no space to say this!
- This questionnaire is poorly constructed. Q27-29-30 not clear. 29+30 do you mean numeric data or data in general? Has a very positivist undertone throughout.

Questioning assumptions about numeric data

- As an infrequent user of the datasets you have primarily in mind my responses/comments may not be very relevant. However I believe there is a need for increased use of datasets within Civil Engineering, to allow benchmarking of ideas/designs with past successes and failures. Such databases would be useful for teaching and practice.
- Did it ever occur to you that there are some topics that can't be addressed numerically?
- I fear that I may have missed the mark somewhat. The course is an introductory stats course driven more by examples of exercises using small amounts of data rather than the large datasets in the sense you seem to mean them.
- I have a feeling, not surprisingly, this questionnaire is orientated/predicated on the idea that survey methods courses have to be based on the use of large datasets. I have taken such a course in the past, taught by an applied sociologist, in which the accent was firmly based on data retrieval, entry using SPSS etc, rather than on the teaching of STATS methods per se. My own feeling is that students learn such methods better when confronted by small data sets (often of interval data that can be aggregated as required) than say reference to huge apparently amorphous collections for which they feel no sense of ownership.
- I think you could have defined 'data' a bit better. (Almost) all scientists work with data. I presume you have in mind not a few numbers here and there, but a substantial body of data. In my field [of global environmental change] I find that the internet sources, mostly coming from the USA, have become very good. In other courses, I work with data that our students collect. This is obviously an important part of their learning experience.
- It is clear an extant service is behind this questionnaire, your classifications may not be appropriate for data in all contexts.
- Numeric data could be useful, but our concern in anthropology is with highlighting other forms of data.
- Numerical data is in my view of marginal relevance in philosophy. As I said earlier, students tend to have difficulty in separating empirical facts from philosophical arguments. Some do use data from various sources found by them, but they tend not to know how to make relevant use of it.
- On the whole I don't see these techniques as relevant to political science.
- Research tends to be qualitative
- The course I teach focuses on qualitative methods so numeric data is only used incidentally.
- There is no definition of 'numeric data'. From the context of many questions, it looks as if you mean data in spreadsheet format, but I will often use numbers as examples for demonstration.
- This questionnaire seems to target numeric data only. Depts in soc sci might find this q. does not apply to/reflect their teaching methods.

Questioning assumptions about numeric data (cont'd)

- We use small datasets and students tend to create their own for their own purposes/dissertations. No one master data set.
- You assume that computers are central to using numeric data. That depends on the nature, quality and size of the data set. One thing we teach students is when not to use a computer. GIGO!
- Your notion of numerical data is not defined, it is clear you have a definition but this is not shared with the respondent

Comments about one's self

- Couldn't answer all the questions as only teaching at institution since Sept.
- Have no competence in stat data so would rely on a co-worker etc.
- I don't do anything at the cutting edge.
- I have passed this on to colleagues directly involved in teaching statistics to our students.
- I suspect there may be more support available here and perhaps I have just not discovered it.
- In the past I have thought about using your data sources, but didn't know much about them. Now our MbChB course has changed, I can see there may be lots of opportunities for using this data, so could you send me info on what is available, and how to access it. Thanks.
- The most crucial thing is lack of time although my work is largely textual, certain kinds of data would be useful.

General suggestions on the survey topic

- I have found lack of continuity a considerable problem together with the variable way in which data is published. Particular problem dealing with Home Office and Department of Health. The use of internet for dissemination by central government is excellent but not all government figures are available that way and new TSO arrangements are poor when it comes to discovering publication of latest version of data set.
- It would be nice to have conferences/symposia on advanced data collection/analysis methods.
- Like many others (I think) I would like to include use of numeric data in my lectures. There are a number of obstacles including (if honest) a lack of confidence. Really we need support/training etc and good examples – templates perhaps, to make a start which can then be developed.
- New course, to bring qualitative and quantitative data analysis together. The primary source of data will be that collected by self or students.
- RAE has killed enthusiasm for the job. When one is dumped with (nearly) all the teaching, with no remuneration of teaching, why expect enthusiastic adoption of schemes to incorporate numeric datasets?
- Students collecting their own data for analysis is desirable, in my view, but likely to be seriously compromised by emerging focus on ethical considerations. So use of data archives will be forced upon us.

31 July, 2000

«Addressees»
 «Department»
 «Institute»
 «Address_1»
 «Address_2»
 «Address_3»
 «Address_4»

Dear Head of Department,

A project is underway to research and evaluate the use of numeric datasets in learning and teaching within Higher Education. This project is sponsored by the Joint Information Systems Committee (JISC) and is managed by the following institutional partners:

- British Library of Political and Economic Science, London School of Economics
- The Data Archive, University of Essex
- EDINA, Edinburgh University Data Library, University of Edinburgh
- MIMAS, University of Manchester.

Additionally, the project receives guidance from a Task Force, comprising a diverse group of academics who have interests in developing the use of numeric data for learning and teaching. The ultimate long-term objective of this project is to improve both local and national support for those individuals who either currently use, or would like to use empirical data resources in their teaching.

An important component of the project is a survey of academic departments in order to gather information on the current situation faced by teachers and supervisors. You will appreciate that information on the topic is in short supply; as a result we are placing a great deal of importance on both the level and quality of response. The survey results will form the basis of a comprehensive report back to the JISC, which is likely to have a positive and significant impact at policy level. I am therefore writing for your help with the distribution of this survey within your department.

Enclosed with this letter are five questionnaires. We ask you to complete one questionnaire yourself, and then to pass the remaining to undergraduate teachers or supervisors (two), and postgraduate teachers or supervisors (two), involved with subjects which have an empirical component. If that split is inappropriate, please feel free to use your own discretion. We also ask you to collect the completed questionnaires from your colleagues and then post them back to us using the reply-envelope provided. If individuals prefer, there is a web-based questionnaire located at the following address: <http://datalib.ed.ac.uk/projects/datateach/survey1.html>. All that is required to use the web version is for the respondent to enter the questionnaire number located in the top right corner of the front page of their paper form into the box provided on the screen.

We trust that two weeks will be sufficient time for all five respondents to give their attention to the survey and complete it, and request your help in ensuring the return of completed questionnaires whether paper or web in that timeframe.

On behalf of the project partners and task force, thank you very much in anticipation of the time taken to respond to this exercise. As soon as we have sufficient responses, results from this survey will be placed on the Project's website. <http://datalib.ed.ac.uk/projects/datateach.html>
 Sincerely,

Prof. Peter Elias
 Institute for Employment Research, University of Warwick
 Chair, Task Force on the Use of Numeric Data in Learning & Teaching



We appreciate that your time is valuable, but hope that you can help us gain an empirical understanding of the teacher's experience of incorporating numeric data into coursework. The survey is one component of a national project funded by JISC. Completed questionnaires will be treated as confidential with respect to both individuals and institutions, and, we will comply with the Data Protection Act.

*If you wish, you may fill out the Web version of this survey at
<http://datalib.ed.ac.uk/projects/datateach/survey.html>*

I Contextual information

1. What is the name of your department? _____
2. What is the name of your institution? _____
3. What is your own discipline or subject speciality? _____
4. What is your job title (e.g. Professor, Lecturer, etc.)? _____
5. Setting aside your research roles, what are your teaching-related functions in your institution?

(Tick all that apply.)

- | | |
|---|--|
| <input type="checkbox"/> Head of Department | <input type="checkbox"/> Supervisor—postgraduate students |
| <input type="checkbox"/> Course Convenor | <input type="checkbox"/> Supervisor—undergraduate students |
| <input type="checkbox"/> Class Lecturer | <input type="checkbox"/> None |

If your roles do not include Course Convenor or Class Lecturer, please turn to Question 20 on page 3.

II Your current use of numeric data in teaching

*Please answer the following questions for **one** of the courses you teach or convene.*

6. What is the name of the course? _____
7. This course is for: *(Tick one.)* Undergraduates Postgraduates Other: _____
8. How long have you been teaching/convening this course? _____
9. Class size is (approximately): *(Tick one.)* Under 15 students
 15 – 40 students
 Over 40 students
10. Which of the following most closely characterises your use of numeric data in this course? *(Tick one.)*
 a) I nearly always use numeric data. d) I haven't yet but would like to.
 b) I often do. e) I never have and don't plan to.
 c) I have, but only occasionally.

If you answered d) or e) then please turn to Question 20 on page 3.

11. What are the data you use in this class called? _____

12. How would you characterise the purpose of your use of numeric data for this class? (*Tick all that apply.*)
- | | |
|---|---|
| <input type="checkbox"/> a) To teach statistics or data analysis methods | <input type="checkbox"/> d) To teach numeracy or critical thinking skills |
| <input type="checkbox"/> b) To teach survey or research design | <input type="checkbox"/> e) To teach general computing skills |
| <input type="checkbox"/> c) To teach use of statistical analysis software | <input type="checkbox"/> f) To add an empirical dimension to the subject |
| <input type="checkbox"/> g) Other: _____ | |

13. And which would you say is the **primary reason**? _____

14. How do you present the data you teach with in this class? (*Tick all that apply.*)
- | | |
|---|---|
| <input type="checkbox"/> As aggregate tables | <input type="checkbox"/> Using spreadsheet or database software |
| <input type="checkbox"/> As graphs | <input type="checkbox"/> Using statistical package (e.g. SPSS, Minitab, Stata, SAS) |
| <input type="checkbox"/> As case by variable ‘matrix’ | <input type="checkbox"/> Using proprietary software on CD-ROM or Internet |
| <input type="checkbox"/> As “raw” ASCII file or number string | <input type="checkbox"/> Other: _____ |

15. Are students expected to work with the data on a computer (‘hands-on’) as part of their coursework?

- Yes No

16. What is the source of the data you use for this course? (*Tick all that apply.*)

- | |
|--|
| <input type="checkbox"/> Published (printed) monograph or serial |
| <input type="checkbox"/> Collected by self or students |
| <input type="checkbox"/> Obtained directly from data producer |
| <input type="checkbox"/> Copied files from a colleague who already had them |
| <input type="checkbox"/> Obtained from a local or national government agency |
| <input type="checkbox"/> Registered for access from a national data centre (i.e. Data Archive, MIMAS, EDINA) |
| <input type="checkbox"/> Freely available on the Internet (specify): _____ |
| <input type="checkbox"/> Data was purchased from a commercial data provider (specify): _____ |
| <input type="checkbox"/> Data bundled with software package or textbook (specify): _____ |
| <input type="checkbox"/> Don’t know, or Other: _____ |

17. How much time did it take to prepare the data for course use in total? _____

18. To what extent do you regard this preparation effort a burden? (*Tick one.*)

- | | |
|--|---|
| <input type="checkbox"/> Yes, an inordinate amount of work | <input type="checkbox"/> Yes, but warranted |
| <input type="checkbox"/> No, not too much effort involved | <input type="checkbox"/> Not sure |

19. Do you feel the need to update / refresh / revise the data used on a regular basis? (*Tick one.*)

- Yes No Yes, but there is insufficient time

III Data for independent learning

20. Do you supervise students when they undertake independent research? (*Tick all that apply.*)

- Yes, undergraduates Yes, postgraduates No

If no, then please go to Question 23.

21. Do you recommend that your students use numeric data in the course of their independent research?

(*Tick only one.*)

- | | |
|--|---|
| <input type="checkbox"/> I nearly always use numeric data. | <input type="checkbox"/> I have, but only occasionally. |
| <input type="checkbox"/> I often do. | <input type="checkbox"/> I never have and don't plan to. |
| | <input type="checkbox"/> I haven't yet but would like to. |

22. Why or why not?

IV Support for using data in teaching – *national services*

23. Have you ever used or considered using national data services (i.e., The Data Archive, EDINA, MIMAS) to access numeric data for learning and teaching purposes?

- Yes No

If no, then please go to Question 26.

24. Please rank the following factors regarding the extent to which they act as barriers in using national data services for learning and teaching purposes. (*Rank each item, 1 = largest barrier, 8 = smallest barrier.*)

- | | |
|---|---|
| ____ Registration procedures | ____ Format of datasets provided |
| ____ Lack of awareness of relevant materials | ____ Documentation & supporting materials |
| ____ Lack of sufficient time for preparation | ____ Lack of relevant teaching subsets |
| ____ Interfaces for data downloading or analysing | ____ Other (specify): _____ |

25. Would you like to see a change in services at the **national level** for using numeric data? If so, what are the most important improvements these services can make to support teachers and learners in the use of data?

V Support for using data in teaching – *local institutions*

26. **Whether for teaching or for research**, have you ever received support for obtaining or using numeric data from the following? (*Tick all that apply.*)

- | | |
|--|--|
| <input type="checkbox"/> National data services staff | <input type="checkbox"/> Local computing service staff |
| <input type="checkbox"/> National data services' local site representative | <input type="checkbox"/> Local library staff |
| <input type="checkbox"/> Co-worker or assistant on course/project | <input type="checkbox"/> Other (specify): _____ |
| <input type="checkbox"/> Another academic colleague | <input type="checkbox"/> Have received no support |

27.Which forms of locally provided support are needed by academic data users? (*Tick all that apply.*)

- | | |
|---|--|
| <input type="checkbox"/> a) Data discovery / locating sources | <input type="checkbox"/> f) Reformat, subset, manipulate datasets for the user |
| <input type="checkbox"/> b) Help with registration procedures | <input type="checkbox"/> g) Provide expert consultation for statistics & methods |
| <input type="checkbox"/> c) Teach use of new data interfaces | <input type="checkbox"/> h) Assist in the preparation of data for teaching |
| <input type="checkbox"/> d) Acquire datasets for a local collection | <input type="checkbox"/> i) Help students use data for learning and research |
| <input type="checkbox"/> e) Current awareness service via email | <input type="checkbox"/> j) Other (specify): _____ |

28.And which form of local support would you say is the **most** important? _____

29.How would you categorise the level of data support from your institution? (*Tick one.*)

- Local data support here is very good across the board
- Local data support tends to be ad-hoc (e.g. from colleagues, general support services etc.)
- In general, my institution does not provide co-ordinated local data support

30. Would you like to see a change in support given by **your institution** for using data resources? If so, what are the most important improvements your institution can make to support teachers and learners in the use of data?

VI Case studies and closing comments

31. We would very much like to follow-up this survey with some site visits for selected case studies to explore in more detail some of the issues covered by this survey. Could we have permission to contact you, or perhaps you could recommend a colleague whose work might make a useful case study? (*Tick one.*)

- No, please do not contact me.
 - Yes, you may contact me. Name: _____ Email: _____
 - Yes, I would like to recommend a colleague for a case study and suggest you contact him/her:
-

32.If you have any additional comments to inform this survey, please make them here or attach another page:

* * * **Many Thanks** * * *

We greatly appreciate your completing this survey. Please return the completed questionnaire(s) to:

Attn: M. Fox, The Data Archive, University of Essex, Colchester CO4 3SQ. Or, fax to: 01206 872 001.

N.B. Please feel free to visit our project website, <http://datalib.ed.ac.uk/projects/datateach.html>, or join our open discussion list, <http://jiscmail.ac.uk/lists/datateach.html>.