

Historical Insights

Focus on Research



QUANTITATIVE SKILLS FOR HISTORIANS

Mark Freeman





Historical Insights: Focus on Research
Quantitative Skills for Historians

ISBN 978-0-9564603-0-1
February 2010

Acknowledgements

I am grateful to Donald Spaeth, who has made me aware of many of the data sources and literature that are cited in this guide.

An electronic version of this Historical Insights guide, with additional resources, can be found at www.historysubjectcentre.ac.uk/quantitativeskills/

Published by History at the HEA
in conjunction with the Institute of Historical Research

University of Warwick
Coventry
CV4 7AL
t: 024 76 150 892
e: heahistorysubjectcentre@warwick.ac.uk
w: www.historysubjectcentre.ac.uk

Contents

Introduction	2
A note to student researchers	4
A note to teachers of history	5
The importance of numbers in history	6
Quantification and history: an exercise	8
Promoting 'dataracy': manipulating, presenting and interpreting numerical data in history	10
More advanced 'dataracy': graphing data effectively	10
Other examples of quantification in history	16
Databases and history: the Victorian censuses	16
Pollbooks and politics	18
Height, health and history	19
How to find data	21
Some thoughts on assessment	22
Bibliography	24
General introductions	24
Quantification and its place in history	25
The Victorian censuses	25
Pollbooks and prosopography	25
Height, health and history	26

Introduction

A lecturer in biological sciences has recently written: 'It is no secret that many undergraduates enter university lacking some basic skills in literacy and numeracy.' This lecturer, Harriet Jones, went on to describe some of the methods used in science departments to address the poor literacy and numeracy skills of their students, including diagnostic tests and remedial classes.¹ On the arts side, many universities now provide writing support to their students: at the University of Glasgow, for example, a basic skills test is required of all first-year students, and those whose performance is below what is expected are given additional help to improve their basic literacy. However, numeracy remains a problem in many arts-based disciplines, as the Higher Education Academy has recognised, for example in the creation of its Numeracy Bank, which will provide online numeracy support materials for teachers across the higher and further education sectors.² It is hoped that this will help to address a situation in which, according to a recent study, the status of UK social sciences is 'threatened by [a] dearth of quantitative teaching'.³ Indeed, Melanie Newman has identified a pervasive marginalisation of quantitative methods across the social sciences:

in many departments, the teaching of quantitative methods is dependent on just a few staff members, and they are heavily involved in research. On average, students receive about 12 hours of teaching in quantitative methods across a three-year degree ... Less than one in ten student projects has a significant quantitative element. Many academics draw a clear distinction between quantitative and non-quantitative approaches and ... [a]cademics with 'quants' knowledge are therefore defined by their methodology rather than their substantive areas of interest.⁴

It is becoming widely recognised that numeracy is an integral part of the corpus of transferable skills that are associated with attributes of 'graduateness'. As universities increasingly promote an agenda of employability and transferable skills, numeracy is likely to be highlighted as a key area in which improvement is possible. In 2006 it was claimed by the Confederation of British Industry that, among employers, 16 per cent 'had concerns about graduates' numeracy skills'.⁵ This does not mean that all students need to learn complicated mathematical procedures and concepts; rather, as Claire Morris has emphasised, it is important to promote a culture of 'dataracy', which she

1 *Times Higher Education*, 2–8 July 2009, pp. 37–9.

2 www.heacademy.ac.uk/projects/detail/oer/OER_IND_AngliaRuskin/ [accessed 12 Feb. 2010].

3 Melanie Newman, 'Want of "quants" bad news for UK', *Times Higher Education*, 3 Dec. 2009, www.timeshighereducation.co.uk/story.asp?sectioncode=26&storycode=409323 [accessed 12 Feb. 2010].

4 *Ibid.*

5 Cited in Claire Morris, 'Cosine of the times', *Times Higher Education*, 1 June 2007, www.timeshighereducation.co.uk/story.asp?storycode=209064 [accessed 12 Feb. 2010].

defines as ‘the ability to look at numerical information intelligently and to link the numbers to the practical situation to which they relate’.⁶ That this ability is highly valued by graduate employers is demonstrated in the UK civil service fast stream entrance process, which requires candidates to undertake both verbal and numerical reasoning tests. This kind of formal examination is not uncommon in graduate-level recruitment, and most employers will value an ability to understand, interpret and produce numerical data. It is therefore important that students are exposed to numerical reasoning in their university study. This is recognised in the Quality Assurance Agency History Benchmark Statement, although this document falls short of recommending the inclusion of ‘numeracy and quantitative methods’ in all courses.⁷

Among students of history, quantification, even of a comparatively basic kind, often arouses fear, especially among those who come to history with a background in arts subjects. Even those who approach history from a social-science standpoint – those who may also study sociology or politics, for example – are easily disheartened when confronted with historical numbers. The same may be true of their teachers, who doubt the extent to which students will be able to use quantitative data with advantage, and who are aware of the resource implications of introducing computer-lab sessions and the like into their courses.

Alternatively, ‘number-crunching’ is seen as the province of economic historians, something to be feared and avoided by ‘mainstream’ researchers. They can easily feel alienated when they encounter regression analyses, correlation coefficients or even simple tabular data in articles they read in conventional history courses; and they need to be encouraged to believe in their ability to use the literature to advantage, even where their understanding of the numerical techniques involved may be limited. Most numerical data presented by most historians is relatively simple, and should not be difficult to use profitably. Confidence is a key aspect of ‘dataracy’, and all historians need to be helped to acquire it. Handling data will help them to gain this confidence, and embedding data analysis within history courses will help them to understand the wider significance of the numerical data that they encounter.

This guide suggests ways in which historians and students of history can appreciate the importance and value – and limitations – of numbers in history, and overcome their fears of numbers through the integration of small amounts of quantitative work into history courses and independent research. The intention here is not to set out a curriculum for a course on quantification in history. Some universities offer this to researchers, and many postgraduate history students are required by their universities, or more usually by funding bodies, to undergo quantitative methods

6 *Ibid.*

7 QAA, History Benchmark Statement, section 13 and annex 2, www.qaa.ac.uk/academicinfrastructure/benchmark/honours/history.asp [accessed 12 Feb. 2010].

training. Some of the exercises in this guide may be of some use in such a course, but the focus is on exercises and ideas that could be adapted to existing history courses and basic methods which can be integrated into historical research. Useful textbooks for quantification in history are listed in the bibliography.

A note to researchers

It is important for all students of history to be aware of the potential for using quantitative data in their research and writing, and to learn how to present it effectively. Even where the focus of your work is on qualitative analysis, introducing some basic quantitative material can help to establish the background and wider context. Presenting a small amount of material in graphical form can establish quickly the main trends that you discuss in more depth in your text.

This guide suggests ways in which you should – and should not – present numerical data, and offers avenues for further exploration. There are some very basic ‘dos’ and ‘don’ts’ of which you should be aware, and examples of good and bad practice abound in the academic historical literature. Just as when introducing quoted material or other evidence, you need to tell your readers what the material is, why it matters, and where it comes from.

The exercises in this guide will introduce you to key aspects of historical numeracy, using real examples that are used within undergraduate teaching at the University of Glasgow. They introduce you to key software; in this case Microsoft Excel is used, rather than SPSS, which is often preferred in the social sciences. The aim of the guide is to introduce you to fairly basic quantitative methods: there is no ‘hardcore’ statistical analysis. For an introduction to social science statistics, see David S. Moore and George P. McCabe, *Introduction to the Practice of Statistics* (4th edn., New York, 2003). For a guide specifically aimed at historians, see Pat Hudson, *History by Numbers: an Introduction to Quantitative Approaches* (London, 2000).

A note to teachers of history

It is recognised that existing courses may not offer any obvious opportunities for the introduction of formal assessment based on numerical analysis or a computing project; however, it is still possible to encourage students, within normal learning and teaching environments, to think about numerical data and its importance. This can be done simply through the selection of appropriate reading material. If students are able to engage with historical arguments that rest on numerical data and its interpretation, they will be encouraged to think beyond the figures towards what they do – and do not – designate and explain. One way of doing this is to conduct a debate on the value and limitations of quantitative methods in history. Such a session could form part of a course on historiography, for example, and has also been used in courses

introducing students to history and computing, in which quantification obviously plays an important role. It could also naturally arise from courses on historical topics in which quantification has been important: for example, American slavery, the standard of living during the industrial revolution, British political history in the 18th century, and so on.

One possible way to assess students' engagement, at this fairly basic level, would be an article report: students can choose a journal article based on quantitative evidence, and write an evaluation, perhaps comparing the findings of the article with alternative sources. Beware, however, of giving students a free choice of articles: you may find yourself with the time-consuming task of reading not only a number of evaluations, but also all the articles themselves! A lecturer can select a few articles, from his/her areas of interest, to which students can be directed.

The importance of numbers in history

Numbers are encountered in all areas of history: they are not, and should not, be left to economic historians. In the course of a broad historical curriculum, students of history are likely to encounter demography, prosopography, psephology and many other areas of study in which quantification is fundamental. Even in histories which are not explicitly quantitative in approach, we find numbers: to take a random example from political history, John Vincent's classic *Formation of the Liberal Party* (London, 1966) contains tabular data relating to, among other things, the wealth of Liberal peers, the numbers of speeches made by Richard Cobden and John Bright in the House of Commons, the circulations of newspapers, the voting patterns of occupational groups in specific localities, and the appointment by different ministries of magistrates in the Duchy of Lancaster. It is a rare or perhaps non-existent historian who never counts anything. Moreover, even where numbers are not reported, many historical judgements rest on some kind of underlying, or 'hidden', quantification.

The emergence of the 'new economic history' in the 1960s represented a fundamental challenge to the existing practices of historians. By exploiting new sources – or at least, by identifying patterns revealed in sources which had hitherto remained hidden – the new economic historians used quantification to challenge some established historiographical positions, and to suggest new methods for achieving a better understanding of the past. In many cases this involved bringing new source material to bear on old debates – for example, the use of anthropometric data, which is considered in more detail below – and in others it involved extracting the quantifiable material from other sources with which historians were already familiar.

We should remember, then, that quantitative research can be simply a matter of using established sources in a different way from other historians. As Roderick Floud has commented, '[q]uantitative questions complement qualitative questions, and quantitative evidence complements qualitative evidence; neither can replace the other, and neither can pretend to comprehend the whole of historical study.'⁸ Nevertheless, the obvious strengths and limitations of quantitative methods – on the one hand, they can help us to discern tendencies that are not otherwise apparent and to generalise effectively about aspects of human experience in the past, while, on the other, they often rely on incomplete and questionable data sources, and are clearly incapable of illuminating some immeasurable aspects of the past – allow a debate to be established about the value of quantification in history.

8 Roderick Floud, *An Introduction to Quantitative Methods for Historians* (London, 1973), p. 3.

It would be possible to select texts from virtually any area of history which rely on quantification. However, Robert W. Fogel and Stanley L. Engerman's influential and important book *Time on the Cross: the Economics of American Negro Slavery* (2 vols., London, 1974) remains a good starting point. The book presented the results of detailed quantitative research into many aspects of the slave economy and life, and challenged much of the accepted wisdom on slavery. For example, Fogel and Engerman argued that slave labour was profitable and efficient; that the material deprivations of slaves had been exaggerated; that the antebellum southern economy was growing in the years before the Civil War. Unsurprisingly, the book was heavily criticised, both for its methods and its findings: one historian called it 'amoral'. Attention should be especially directed to the introduction, which sets out what Fogel and Engerman believed quantitative methods could achieve, and also to their afterword to the 1989 (and 1995) Norton edition, which responds to some of the criticisms of the book.

Fogel, together with the British constitutional historian Geoffrey Elton, contributed to a helpful book which discusses, in the form of a debate, the differences between 'traditional' historians and the 'new economic historians'. This book is *Which Road to the Past? Two Views of History* (New Haven, Conn., 1983). It should be read with caution, because it is now almost 30 years old and reflects a situation in which the mutual suspicions of different groups of historians were particularly acute. One might also note the different subject-matter studied by the two historians – Fogel was an economic and demographic historian, Elton concentrated on high politics and diplomacy – and perhaps different methodologies were felt appropriate to different fields. There is also little indication of the growing importance of social and cultural history, which complicates the picture much more. An alternative, and much shorter, source is Mark R. Horowitz, 'Which road to the past?', *History Today*, 34 (Jan. 1984), 5–10.

It is also useful, at this stage, to counsel against thoughtless reliance on data that is not comparable over time. The use of numerical data can help researchers and students to think about the nature of historical source material; about its limitations as well as its strengths. A text which emphasises this, is critical of the reliance of some historians on long series of data, covering very long periods, and encourages 'short views', is D. C. M. Platt, *Mickey Mouse Numbers in World History: the Short View* (Basingstoke, 1989).

Quantification and history: an exercise

One way of emphasising this is through an exercise outlined by Daniel I. Greenstein.⁹ Robert W. Fogel, the Nobel Prize-winning economic historian and co-author of *Time on the Cross*, suggested in 1975 that the sceptic should, in Greenstein's words, 'choose a historical text at random, open it to [sic] any page, and count up the quantitative statements found there'. Greenstein uses examples from economic, diplomatic and religious history.

This exercise is a useful one to introduce the importance of numbers in history. The passage below is from John Stevenson's *British Society 1914–1945*, in the Pelican Social History of Britain series.¹⁰ I chose this extract partly because it discusses leisure and recreation, traditionally perceived as 'soft' subject areas which are less 'difficult' than economic history. Of course, as the exercise shows, historians of these areas are reliant on official and other statistics, as well as contemporaneous impressionistic accounts, when they discern larger trends and patterns in their social histories. The passage runs as follows:

One of the most important developments in twentieth-century society has been the growth of leisure and recreation. Already by the Edwardian era, many pastimes and pursuits had been fashioned or transformed to meet the needs of a primarily urban and industrial society. In sport, entertainment and private recreations, one of the major driving forces was commercialization, drawing upon the increased spending power of a mass consumer market. Another was the increased leisure time available as a result of shorter working hours, paid holidays, longer life expectancy after retirement, smaller families and, for some, enforced idleness through unemployment. But the growth of leisure illustrates more than commercialism and more free time from work. With the growth of the media, it was part of the development of a more uniform and homogeneous society, partaking of an increasingly common culture. Notwithstanding regional and class differences, by 1945 only the remotest parts of Britain were insulated against the pervasive influences of the latest popular tune or major sporting event. In contrast, some aspects of leisure, particularly those centred around hobbies and domestic life, reflected an increasing home-centredness. The two themes of an increasingly common culture, balanced by the cult of domesticity and individual choice, dominated the development of leisure in this period.

This passage with the quantitative statements or judgements underlined would look like this (others might do it differently, or take issue with some of the underlinings):

9 Daniel I. Greenstein, *A Historian's Guide to Computing* (Oxford, 1994), pp. 6–15.

10 John Stevenson, *British Society 1914–45* (Harmondsworth, 1984), p. 381.

One of the most important developments in twentieth-century society has been the growth of leisure and recreation. Already by the Edwardian era, many pastimes and pursuits had been fashioned or transformed to meet the needs of a primarily urban and industrial society. In sport, entertainment and private recreations, one of the major driving forces was commercialization, drawing upon the increased spending power of a mass consumer market. Another was the increased leisure time available as a result of shorter working hours, paid holidays, longer life expectancy after retirement, smaller families and, for some, enforced idleness through unemployment. But the growth of leisure illustrates more than commercialism and more free time from work. With the growth of the media, it was part of the development of a more uniform and homogeneous society, partaking of an increasingly common culture. Notwithstanding regional and class differences, by 1945 only the remotest parts of Britain were insulated against the pervasive influences of the latest popular tune or major sporting event. In contrast, some aspects of leisure, particularly those centred around hobbies and domestic life, reflected an increasing home-centredness. The two themes of an increasingly common culture, balanced by the cult of domesticity and individual choice, dominated the development of leisure in this period.

Several points might be made about this text. Some of the reliance on quantification is obvious: for example, Stevenson's comments about longer life expectancy, shorter working hours and increased spending power. Some identifiable trends – for example, the 'growth' of the media – reflect changes that are quantifiable, at least to an extent. Others are more subtle. For example, Stevenson refers to the remotest parts of Britain: presumably remoteness is a characteristic of regions with low population density, another quantitative measure. The identification of a tune as 'popular' or a sporting event as 'major' can pose quantitative questions: how many spectators make an event 'major'? In these cases, the numerical dimension is not the only component – the cultural significance of a sporting event might transcend such considerations – but it is important.

Naturally, this is only an introductory exercise; its main purpose is one of emphasis, to overcome suspicion of what quantification might entail. The idea behind the exercise, however, is one that is almost universally applicable: specialists in any area of history can easily select a passage, almost at random, and subject it to the same sort of interrogation.

Promoting ‘dataracy’: manipulating, presenting and interpreting numerical data in history

The best software for historians to use for manipulating basic numerical data is MS Excel, the most widely used spreadsheet programme. This allows simple, and more complex, calculations to be performed on data, and graphs to be created and modified. A 10-page workbook, ‘Introduction to MS Excel for Historians’, using very basic historical population data, is available at www.historysubjectcentre.ac.uk/quantitativeskills/.

This will introduce you to a number of key features of the Excel programme. The workbook is adapted from an introductory document used for various courses at the University of Glasgow. The data used in the workbook bears no relation to the actual data used in any of the exercises (see below for an example): it is helpful for lecturers to be able to use the same template for introductory documents such as this, without having to vary it every time different data is used. Thus at Glasgow we also use this introductory document as a basis for exercises that deal with very different data sources.

One such exercise, originally designed for second-year students, considers the welfare state in inter-war Britain. It is designed (a) to introduce the Excel programme, (b) to encourage closer manipulation of numerical data, and (c) to apply numerical and graphing techniques to the analysis of a historical question. Face-to-face, this is taught in three sessions: a one-hour lecture that introduces debates on the inter-war welfare state, and two one-hour lab sessions, in which a version of the introductory Excel document is used. In the exercise, participants are required to complete a worksheet, which asks a range of questions, of increasing difficulty. Some of the questions test the ability to carry out basic calculations in Excel; others ask for graphs based on the data; and others, most importantly, encourage them to demonstrate the ability to deploy numerical evidence in textual answers to questions about the historical subject-matter at hand.

For this exercise, the data is presented in a user-friendly Excel workbook, with spaces provided for carrying out the necessary calculations. The data – which is taken from a recent textbook, Bernard Harris, *The Origins of the British Welfare State: Social Welfare in England and Wales* (Basingstoke, 2004) – can also be downloaded from www.historysubjectcentre.ac.uk/quantitativeskills/.

Creating an exercise of this kind requires the designer to spend some time producing the appropriate documentation, and designing the appropriate assignment(s). Our usual practice is to create a printed workbook, and then to provide a worksheet, which must be completed as the assessed component of the exercise. An example of the worksheet, with a reading list, is available online at www.historysubjectcentre.ac.uk/quantitativeskills/.

More advanced ‘dataracy’: graphing data effectively

Bernard Harris’s data is also used in a more advanced graphing exercise, which considers in more depth the presentation of numerical data. It is perhaps surprising how many basic errors and infelicities of presentation occur when writers attempt to present data to an audience. Newspapers are serial offenders in this regard: graphs are often not properly labelled; the axes are frequently unidentified; the source of the data is unclear; the units of measurement are not given; the list goes on. Even in academic publications, graphs can appear inelegant, or may be unnecessary. Among historians, who often lack training even in basic quantitative methods, such problems may be especially acute.

On the following pages, various ‘do’s’ and ‘don’ts’ of graphing are presented, all based on Bernard Harris’s data. It is also worth making the point that graphs which look very clear in colour – in Excel itself, or on a Powerpoint slide – may not look so good on a black-and-white printed page.

One common error is presenting a graph when the figures could easily, and just as helpfully, be given in the text. Although properly labelled, Chart 1 is, unnecessarily, presented in 3D; moreover, it would be just as easy (in fact, better, because the exact number would be given) to present this information as a sentence in the text: ‘In the UK in 1917, 10.5m men and 4.9m women were entitled to benefits under the national health insurance scheme.’

Chart 1: Number of men and women entitled to benefits under the national health insurance scheme, UK, 1917

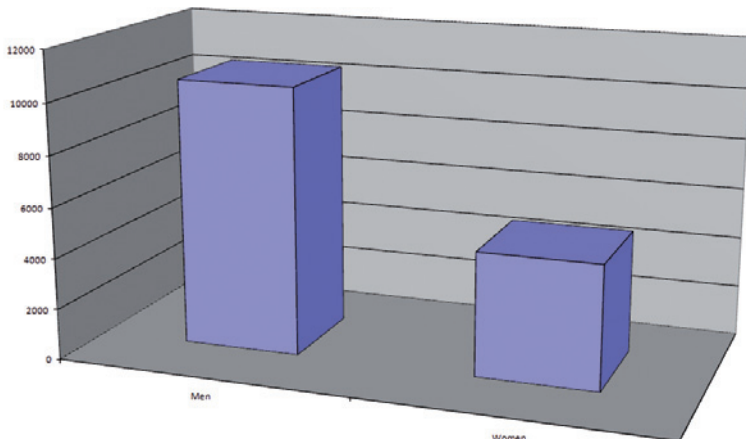
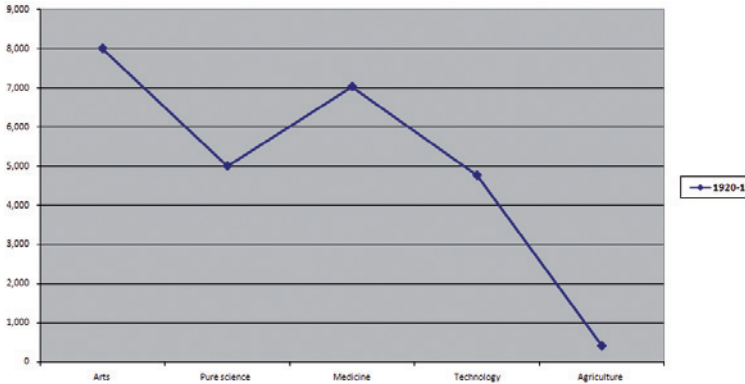


Chart 2 is an example where the wrong type of graph is used. A line chart should be used only to graph a series of data, and not for categorical variables. Moreover, there is a 'legend' to the right-hand side of this graph, which should *not* be used when there is only one data series.

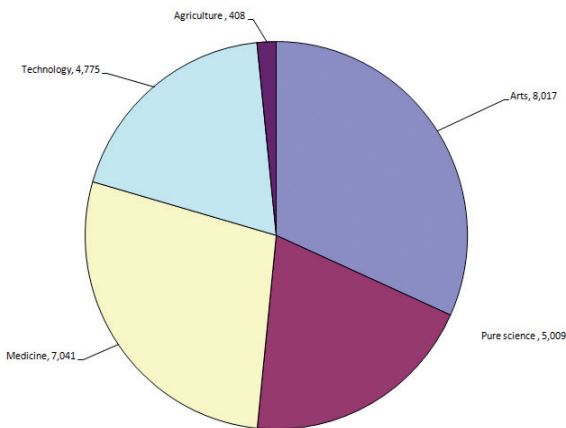
Chart 2: Number of full-time students studying in each subject area in universities in England and Wales, 1920–1



Source: Bernard Harris, *The Origins of the British Welfare State* (London, 2004), p. 279

This material should be presented, as shown in Chart 3, in the form of a pie chart, or alternatively as a column chart. There is no legend here: data labels have been used.

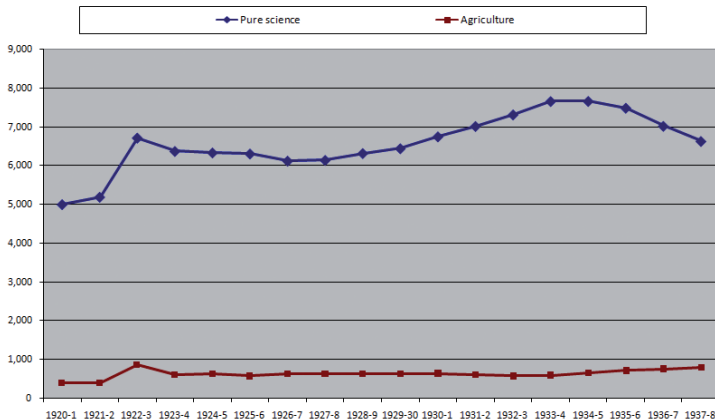
Chart 3: Number of students studying in each subject area in universities in England and Wales, 1920–1



Source: Bernard Harris, *The Origins of the British Welfare State* (London, 2004), p. 279

Chart 4 is an example of a chart that presents two data series which differ so much in magnitude that it becomes difficult to discern the trends in the smaller series. Here, a legend has been, correctly, used (and positioned at the top of the graph, which can look better than a legend placed at the side), the chart has been correctly labelled, and the source given. Moreover, a line chart is an appropriate way of representing this data. However, the difference in size between the two series hides a number of interesting features of the 'Agriculture' series. Although it is possible to discern the spike in the number of students taking agriculture degrees in 1922–3, it is not readily apparent to the casual eye that the number increased by 38 per cent between 1932–3 and 1937–8.

Chart 4: Number of full-time students studying pure science and agriculture at universities in England and Wales, 1920/1–1937/8

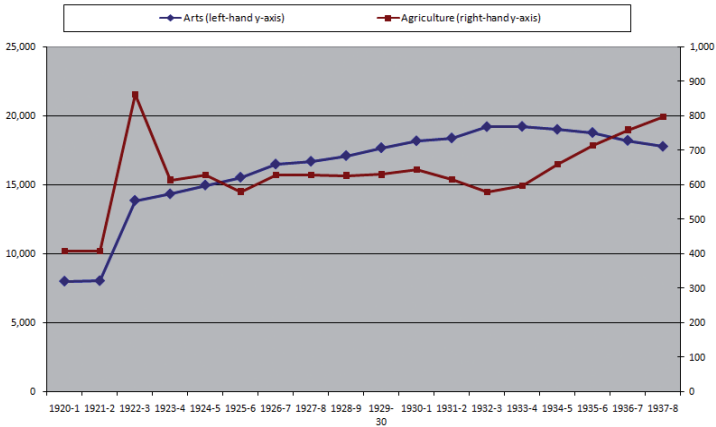


Source: Bernard Harris, *The Origins of the British Welfare State* (London, 2004), p. 279

In fact, as Chart 5 (on the next page) shows, the number of students taking agriculture exhibited greater fluctuation than was the case for 'Pure science'. Chart 5 shows the 'Agriculture' series against a 'secondary axis' – note that the legend has been carefully re-labelled to make this clear.

Chart 5 makes the trends easier to discern, but at the cost of some potential confusion regarding the different y-axes.

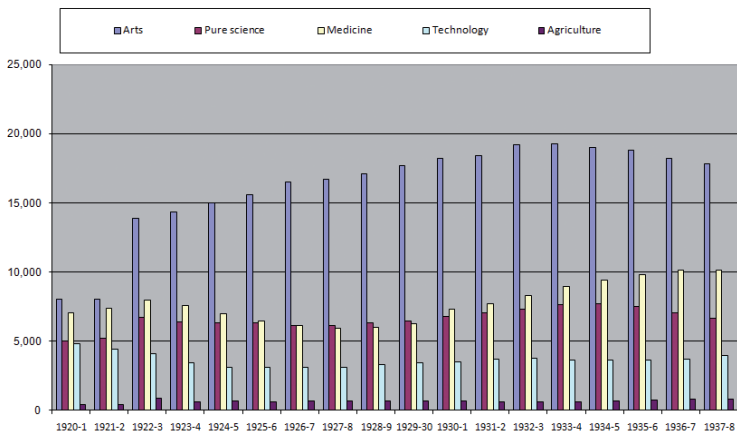
Chart 5: Number of full-time students studying pure science and agriculture at universities in England and Wales, 1920/1–1937/8



Source: Bernard Harris, *The Origins of the British Welfare State* (London, 2004), p. 279

In Chart 6, too much data is shown on one graph. In this case, a line graph would make the material more readable, but there remains the problem of the ‘Agriculture’ series, which would still be difficult to follow, even on a secondary axis.

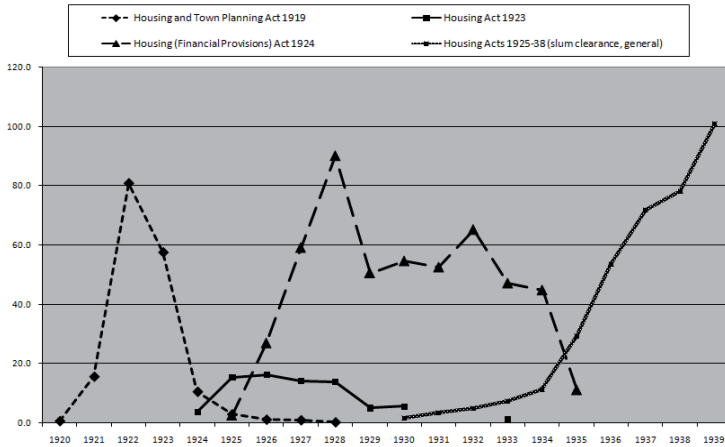
Chart 6: Number of students studying in each subject area in universities in England and Wales, 1920/1–1937/8



Source: Bernard Harris, *The Origins of the British Welfare State* (London, 2004), p. 279

You should also note that, when printing graphs in black-and-white, some of the distinctions between the data series may become blurred. Chart 7 is prepared entirely in black-and-white, to ensure clarity when printed.

Chart 7: Number of houses ('000s) built by local authorities under various Housing Acts, 1920–39

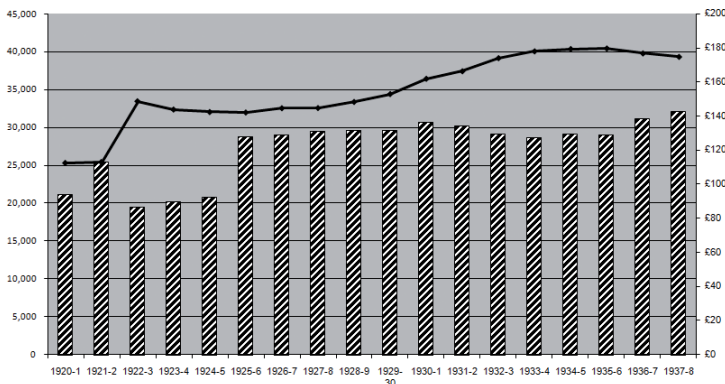


Source: Bernard Harris, *The Origins of the British Welfare State* (London, 2004), p. 261

A final example is given below. Here, two series have been charted in different ways. Notice that the right-hand y-axis makes clear the units – i.e., £ – and that the subtitle indicates which series is charted where.

Chart 8: Number of full-time students at universities in England and Wales, and universities' income per full-time student, 1920/1–1937/8

Number of students charted in the line, against the left hand y-axis; income per student charted in the bars, against the right hand y-axis



Source: Bernard Harris, *The Origins of the British Welfare State* (London, 2004), p. 279

Other examples of quantification in history

This section introduces three more areas of historical research and teaching as examples of the more advanced use of quantitative methods in history. The areas have been chosen mainly to give a flavour of the type of material that can be used. Most are taken from my own field of interest – modern British economic and social history – but there is also an example, ‘Pollbooks and politics’, from political history. Medieval and early modern historians, and historians of other countries, are also well served with quantitative data (see ‘How to find data’, on p. 21 below).

Databases and history: the Victorian censuses

The Victorian censuses offer perhaps the most accessible and easily usable source material for studying British economic and social history in its quantitative dimension. There are census-type sources for earlier periods of history, but the ‘mature’ Victorian censuses of the period 1841–91 offer the teacher of history unrivalled possibilities for devising ‘hands-on’ exercises. These can be made as easy or as complex as the lecturer wishes. Alternatively, there is a range of resources available on which the teacher can draw. The published reports of the censuses – which can be read at www.histpop.org.uk – can be mined for statistical data on individual places and themes in economic and social history. We have used the published data in Excel exercises, similar in scale to the welfare state exercise described above. Moreover, the surviving census enumerators’ books (CEBs), a staple of family history, carry individual-level data. Much of this is already available in electronic format, and many historians have used it in quantitative research into the history of Victorian communities. Some examples are given in the bibliography below.

For a number of years in my department we employed a small computerised section of the Victorian CEBs to introduce students to the use of databases in history, and to encourage them to undertake a very small amount of basic numerical calculation. For this purpose we use the MS Access database software. Most historical researchers are less likely to have encountered Access than Excel, and many find it more difficult to use. Therefore the supporting documentation may need to be more thorough than the equivalent Excel material. A more detailed case study of a CEB exercise can be found in Mark Freeman, ‘Population: the 19th-century census and the teaching of history’, in *Historical Research Methods for the Twenty-First Century*, ed. Ian Anderson and Derek Harding (online resource and print-on-demand book; London, forthcoming; see www.history.ac.uk/publications/historical-research-methods/19th-century-census/). This contains a commentary, together with electronic versions of the course materials.

The exercises involve no more than simple counts of occupations, ages and so on, and expressing these figures as percentages. Participants are asked to consider the implications of their findings for historians of the occupational and social structure of Victorian communities. This exercise also introduces the principles and practice of occupational coding, something which some researchers might want to explore further.

It has never been easier to obtain electronic extracts from the CEBs. Historians can use AHDS History (see below), which has the complete 1881 CEBs for the UK, and many extracts from other census years. Other websites contain free data. One example is the Victorian Census Project at Staffordshire University: www.staffs.ac.uk/schools/humanities_and_soc_sciences/census/vichome.htm.

One can also find useful datasets, as well as online teaching material, relating to the Victorian census and other sources from the Centre for Historical and Contemporary Census Collections (CHCC). There are eight tutorials, seven dealing with themes in Victorian economic and social history, and one with the database and spreadsheet skills required to manipulate and use the census data. These can be found at <http://chcc.arts.gla.ac.uk>. However, most are not kept regularly updated, and should therefore be used with caution. For teachers, they contain ideas as to how formal assessments might be constructed using census data. For researchers, there are also suggestions of how other information – quantifiable and unquantifiable – might modify or enhance the picture we get from the census.

The census is one of the easiest sources to use in this way, and possibly the least daunting for beginners, as well as being of interest in raising issues about the quality of the evidence that is being quantified.

Pollbooks and politics

This is another area in which the use of quantitative data can help students to think about wider historical questions. Before the introduction of secret ballots in British parliamentary elections (in 1872), pollbooks recorded who had voted for whom, and enable the historian to investigate electoral patterns and trends. As with many other such sources, the arrival of computer applications from the 1960s made exploitation much easier. Like the census enumerators' books, pollbooks have been used by family historians: they contain voters' places of residence, and in some cases an indication of their rank (e.g., 'Gentleman').

Eighteenth-century British political historians have made particularly widespread use of pollbooks to examine voting behaviour. The key text is W. A. Speck, *Tory and Whig: the Struggle for the Constituencies 1701–1915* (London, 1970). Speck, using quantitative analysis of pollbooks, found that party awareness among voters was high, and that

voters were inclined to switch frequently between candidates at elections. The existence of a large ‘floating vote’ suggested to Speck that British democracy in this period was ‘participatory’, in other words that voters were aware of issues and made political choices based on their own interest and initiative.

This challenged the dominant view, most famously expressed by Lewis Namier, that voters acted ‘deferentially’, following their social superiors; the evidence marshalled by Namier and others in support of this interpretation was predominantly correspondence which emphasised the provision of transport for voters, the role of bribery and the exertion of influence. Since Speck first worked on the pollbooks, a large literature has grown up around the source, and the availability of data allows lecturers to incorporate some ‘hands-on’ study of pollbooks into the curriculum.

A word of caution should be introduced here. Much of the literature on pollbooks is based on evidence from several elections, which shows the frequency with which individuals changed their electoral allegiance. This relies on the use of nominal record linkage, and may not be appropriate for beginners. Again, however, the data can be used in a reasonably simplified form and simple quantitative questions can be asked. Pollbooks from a number of areas are available from AHDS History (study numbers 3032–6, 3038, 3165, 3178, 3179),¹¹ including for 18th-century Essex, Herefordshire and Northampton. Nineteenth-century pollbooks are also available; and there are many studies of voting behaviour in the period.

Height, health and history

In the area of demography, the techniques available to the early practitioners of the ‘new economic history’ and their successors, as well as a range of source material, have enabled historians to use anthropometric data, in other words data on the physical characteristics of past populations, especially the height and weight of people in the past. Historians have used this data in a number of productive ways, contributing to the historiographies of France, American slavery, the ‘standard of living debate’ in Britain, and so on. Anthropometric history relates to the history of medicine and health, and the history of the human body, as well as to the more traditional concerns of economic and social historians. The methodological issues associated with the use of anthropometric data can also encourage students to think about the nature and imperfections of primary source material.

The study of height data from the past also has the advantage of being relatively easily understood in a contemporary context. As Roderick Floud and his collaborators, in the pioneering study in this area, pointed out, ‘we observe that some of our friends are much taller, some much shorter, than ourselves; in statistical terminology,

11 On AHDS History and its successors, see below.

there is variance in heights within the group'.¹² The meaning and importance of anthropometric data are certainly more apparent to the uninitiated than data on economic growth, wages and earnings, or even literacy and other social/demographic indicators of wellbeing.

The use of anthropometric data by historians has also provoked some surprisingly fierce exchanges in the pages of academic journals, which in my experience have interested students more than the majority of papers they are assigned to read. A lively debate in the *Economic History Review* between Peter Kirby and Jane Humphries, about the heights of children who worked in early Victorian coal mines, has proved particularly popular in seminars on the standard of living in Britain during the industrial revolution. Students have been surprised to read Humphries accusing Kirby of 'breathtaking ignorance'!¹³ More complex questions are opened up by recent studies of the impact of smallpox on height: this debate encourages students to consider the nature and implications of the historical evidence.

This feature of records of historical heights, as in the case of the census and other quantitative sources, provides a valuable insight into the processes by which historians use their source materials. Anthropometric data has been manipulated in various ways by historians, always with an eye to the limitations. The records of historical heights are quite sparse, and we have data only on some groups within the population. We have some data on schoolchildren; and, in the 19th century, measurements were regularly collected from military recruits and from convicts. Neither of these groups

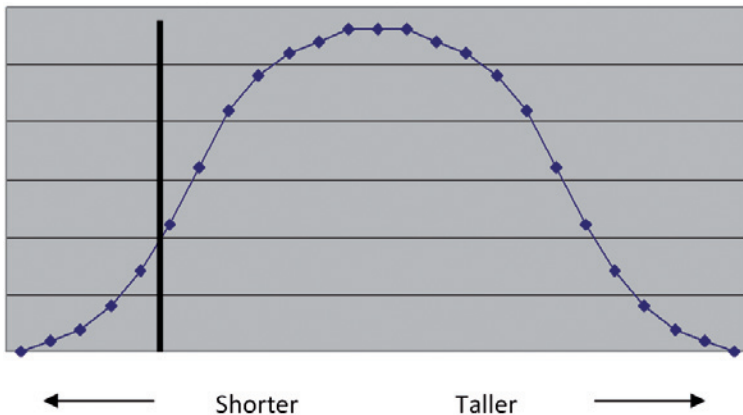


Chart 9: Normal height distribution in a modern population, showing a height limit

12 Roderick Floud, Kenneth W. Wachter and Annabel Gregory, *Height, Health and History: Nutritional Status in the United Kingdom* (Cambridge, 1990), pp. 5–6.

13 Jane Humphries, 'Short stature among coal-mining children: a comment', *Economic History Review*, 50 (1997), 537. See the bibliography for full references to Kirby, and to other articles on height, health and history.

were necessarily typical of the adult population. Convicts were likely to come from more deprived social backgrounds, where we would expect heights to be lower. Military recruiting was often subject to a height limit, below which potential recruits were rejected. It is, however, possible to allow for a ‘truncation bias’ in the sample (see Chart 9 below).

The chart above shows a normal height distribution. If a height limit is applied (the black line), all below a certain height are excluded. Therefore any calculations based on such data need to take account of the *truncation bias* of the sample. As military height limits have sometimes changed over time, the question of how heights change is further complicated.

Other problems include the clustering of measurements at convenient half-inch intervals, the possibility of changing practices over time (e.g., measuring with and without shoes) and the small numbers of observations that can make generalisations questionable. However, all this makes the data more, rather than less, interesting in historical enquiry. There is a large recent literature on historical heights (some of which has challenged the historical accuracy of the normal height distribution represented above), and a number of datasets are available for downloading.

How to find data

There are a number of ways in which data sources can be generated and made available. The data on the 20th-century welfare state, discussed above, was simply entered into a spreadsheet directly from Bernard Harris’s textbook. Similarly, census material can be selected from the original source and entered into a spreadsheet.¹⁴ Although entering data can be time-consuming, the relatively small amounts used in the exercises discussed above demonstrate that the time spent need not be excessive. It is important, however, to consider carefully the choice of data, and what you expect to do with it. A balance needs to be struck between, on the one hand, presenting material that will not be excessively cumbersome, and, on the other, offering enough data both to illuminate historical debates and to challenge readers in a meaningful way.

A wide range of data is available from the UK Data Archive (UKDA). Historical data was formerly held by the Arts and Humanities Data Service for History (AHDS History), but this service is no longer funded. Data can still be obtained, via the UKDA, from the AHDS website: www.ahds.ac.uk. Datasets created in historical research are stored here, and are publicly available, although registration is required. The range of material is wide: examples include, *inter alia*, data on ships in the English slave trade in the 1790s (study number 1385), early modern Scottish witchcraft (4667), late medieval Italian consumerism (4884) and 20th-century Irish agriculture

¹⁴ Digital facsimile versions of population reports, together with useful articles by Edward Higgs, Matthew Woollard and others, can be found at www.histpop.org.uk [accessed 12 Feb. 2010].

(3544), as well as census data from England and Wales, Scotland, Canada and elsewhere. It is difficult to imagine a researcher with interests that are not catered for, in some way, by data from the UKDA. The numerical data usually comes in text files, and some knowledge of the relevant software will be needed to convert this into Excel or Access format.

Some thoughts on teaching and assessment

It is difficult to offer more than tentative suggestions about how the study of historical numbers can be integrated into the curriculum in the early years of university history courses. The structure of existing courses, preferred methods of assessment, the availability of resources and the general institutional milieu can all militate against the introduction of any new curricular feature, and perhaps against quantification in particular. In current practice, anything especially difficult tends to be left to postgraduate research training courses, and in my experience this is the element of research training which postgraduate students dislike the most (It should, however, be added that some of those who approach quantitative research methods courses with the most trepidation actually perform extremely well.) On the whole, it is probably sensible to keep the more complex statistical techniques out of the mainstream undergraduate history curriculum.

However, as this brief guide has shown, there are many ways in which the study of quantitative data, at a fairly basic level, can be incorporated into mainstream history courses. The start-up costs need not be excessive, provided that computer laboratories are available; and the opportunities in terms of devising innovative assessed exercises are substantial. Assessment has been touched on briefly above, but some final remarks should be made.

Any assessment needs to bear in mind the various intended learning outcomes of the exercise, which are likely to include transferable skills as well as more academic outcomes. Where widely used software is employed, quantification exercises can incorporate assessment of a student's facility with the relevant package. It is, of course, important to test students' ability to perform at least the basic software functions and the simplest kinds of calculation, as these are necessary, although insufficient, conditions for presenting and interpreting the findings.

There are four stages in the process of using numbers in history. First, extracting the relevant data or carrying out the necessary calculations. Second, presenting the results in an appropriate form: is a table or a graph the best way to present a particular set of data? How should the data be labelled? What level of detail is required? All these questions need to be asked. Third, the findings must be interpreted: what does the data tell us about the occupational structure of the Victorian population, the results of 18th-century elections, the changing height of the population over time, and so on? Fourth, and most important, the interpretations must be fitted into wider historiographical contexts: does the occupational structure of the community reflect what we know about Victorian society more generally? Do

these pollbooks suggest a large ‘floating vote’? What does the anthropometric data tell us about the standard of living? The course designer needs to devise exercises which address all four of these stages in the use of evidence.

We have found that producing a worksheet for students to submit, featuring both numerical ‘right/wrong’ answers and more interpretative questions, allows the necessary balance to be struck. The simple questions, requiring the filling in of a table, for example, are followed up with questions requiring one- or two-paragraph answers, for which students are encouraged to read the secondary literature as widely as they would do in the preparation of a traditional history essay. Some examples of how these questions might progress through the four stages outlined above can be found in the CHCC online tutorials.

The kind of assessment used, of course, can vary as widely as in any other area of teaching. As suggested above, an article or book review, examining both the methodology and the wider context of a piece of historical writing that uses quantitative data, is one way of helping students to become familiar with historical quantification. Formal or informal skills tests can be used to assess the level of acquisition of the key technical and numerical vocabularies. Traditional essays can be set, requiring the use of pre-circulated data. More ambitiously, group exercises, where students interrogate datasets and present their results collaboratively, can be undertaken. These are only a few suggestions; experience will show which are the most appropriate in a given learning environment.

Bibliography

General introductions

The most recent general textbook on quantitative methods and approaches is Pat Hudson, *History by Numbers: an Introduction to Quantitative Approaches* (London, 2000).

See also the much older textbook: Roderick Floud, *An Introduction to Quantitative Methods for Historians* (London, 1973).

On databases, see Mark Freeman, 'Population: the nineteenth-century census and the teaching of history', in *Historical Research Methods for the Twenty-First Century*, ed. I. Anderson and D. Harding (online resource and print-on-demand book; London, forthcoming; see www.history.ac.uk/publications/historical-research-methods/19th-century-census/).

For introductions to history and computing, see:

Sonja Cameron and Sarah Richardson, *Using Computers in History* (Basingstoke, 2005).

M. J. Lewis and Roger Lloyd-Jones, *Using Computers in History: a Practical Guide* (London, 1996).

Charles Harvey and Jon Press, *Databases in Historical Research: Theory, Methods and Applications* (Basingstoke, 1996).

Daniel I. Greenstein, *A Historian's Guide to Computing* (Oxford, 1994).

Because the technology available changes so rapidly, history and computing textbooks date quickly. However, the general principles outlined in these textbooks remain applicable; only the technology available to the teacher and researcher of history has changed.

For a more advanced statistics textbook, see:

David S. Moore and George P. McCabe, *Introduction to the Practice of Statistics* (4th edn., New York, 2003).

Quantification and its place in history

The classic text of the 'new economic history' was Robert W. Fogel and Stanley L. Engerman, *Time on the Cross: the Economics of American Negro Slavery* (2 vols., London, 1974). This controversial book has gone through a number of editions: the Norton edition of 1995 has a new afterword by Fogel and Engerman.

Robert W. Fogel and G. R. Elton, *Which Road to the Past? Two Views of History* (New Haven, Conn., 1983).

Mark R. Horowitz, 'Which road to the past?', *History Today*, 34 (January 1984), 5–10.

The Victorian censuses

The most useful guide for undergraduate students is Edward Higgs, *Making Sense of the Census Revisited: Census Records for England and Wales 1801–1901* (London, 2005). It is also available under its previous title, *A Clearer Sense of the Census: Victorian Censuses and Historical Research* (London, 1996), and its original title, *Making Sense of the Census*.

Other examples of the uses of the census by historians:

Michael Anderson, *Family Structure in Nineteenth-Century Lancashire* (London, 1971).

Mark Freeman, 'Population: the nineteenth-century census and the teaching of history', in *Historical Research Methods for the Twenty-First Century*, ed. I. Anderson and D. Harding (online resource and print-on-demand book; London, forthcoming; see www.history.ac.uk/publications/historical-research-methods/19th-century-census/)

Eleanor Gordon and Gwyneth Nair, 'The economic role of middle-class women in Victorian Glasgow', *Women's History Review*, 9 (2000), 791–814.

Pollbooks and prosopography

W. A. Speck, *Tory and Whig: the Struggle for the Constituencies 1701–1715* (London, 1970).

W. A. Speck and W. A. Gray, 'Computer analysis of pollbooks: an initial report', *Bulletin of the Institute of Historical Research*, 43 (1970), 105–12.

W. A. Speck, W. A. Gray and R. Hopkinson, 'Computer analysis of pollbooks: a further report', *Bulletin of the Institute of Historical Research*, 48 (1975), 64–90.

The 'deference' argument was most famously made in Lewis Namier, *The Structure of Politics at the Accession of George III* (London, 1929; 2nd edn., 1957, with several reprints).

J. E. Bradley, *Religion, Revolution and English Radicalism: Nonconformity in Eighteenth-*

Century Politics and Society (Cambridge, 1990).

John A. Phillips, *Computing Parliamentary History: George III to Victoria* (Edinburgh, 1994; special issue of *Parliamentary History*).

Height, health and history

Roderick Floud, Kenneth W. Wachter and Annabel Gregory, *Height, Health and History: Nutritional Status in the United Kingdom* (Cambridge, 1990).

Bernard Harris, 'Health, height and history: an overview of recent developments in anthropometric history', *Social History of Medicine*, 7 (1994), 297–320.

Peter Kirby, 'Causes of short stature among coal-mining children 1823–1850', *Economic History Review*, 48 (1995), 687–99.

For a reply to this, see Jane Humphries, 'Short stature among coal-mining children: a comment', *Economic History Review*, 50 (1997), 531–7; and for Kirby's rejoinder, immediately afterwards, see Peter Kirby, 'Short stature among coal-mining children: a rejoinder', *Economic History Review*, 50 (1997), 538–42.

There has been a long debate in the *Economic History Review* concerning the impact of smallpox on height. For recent contributions, see:

Deborah Oxley, "'The seat of death and terror': urbanization, stunting, and smallpox', *Economic History Review*, 56 (2003), 623–56.

Deborah Oxley, "'Pitted but not pitied", or, does smallpox make you small?', *Economic History Review*, 59 (2006), 617–35.

History at the Higher Education Academy works towards the development of teaching and learning of history in Higher Education – reviewing current practices, discussing disciplinary research and innovations, and examining issues of strategic importance.

We offer a wide range of services: a programme of events and professional development seminars; funding for the enhancement of teaching and learning; advice and support for research and development in HE history education; resources including reports, tutor guides and case studies; support for academic networks in history; and support for early career historians including postgraduates and postdoctoral students.

History at the Higher Education Academy
University of Warwick
COVENTRY CV4 7AL

heahistorysubjectcentre@warwick.ac.uk

The Institute of Historical Research is one of 10 member Institutes of the School of Advanced Study, University of London. Founded in 1921, the Institute is home to three important research centres and a major open access library, hosts over 50 seminar series and offers doctoral supervision in a wide range of historical subjects. It has a substantial publishing programme, including the journal *Historical Research*, hosts a number of innovative digital research projects, administers Fellowships, runs specialist research training programmes and organises a variety of conferences and workshops each year. For further details see www.history.ac.uk.

