# THE HISTORIAN AND THE COMPUTER: THE EXAMPLE OF STYAL

Simon Goodwin

**Some historians might be surprised by the description of a computer as an historical source. In fact the use of this term is entirely justified and there is a growing body of work to prove this. This is not more evident than in the field of local history where legions of amateur and professional historians are transcribing documentary sources into computerised form. In the North West alone projects are being undertaken at Manchester Polytechnic (the Didsbury census) and Lancaster and Liverpool Universities, in addition to the work of local groups. When such records are held on computer they can be processed in such large quantities and so quickly that analysis that was previously too time consuming and costly becomes easy work. The use of computers threatens to revolutionise the practise of local history.**

It would be foolish to pretend that computers are easy to understand for the uninitiated. In many cases the mere thought of using a computer is enough to send shivers down the spine. This is simply fear of the unknown, though admittedly people educated before the so-called 'computer revolution' often find it difficult to adapt to the new ideas. Usually all that is required is a little patience and some 'hands-on' experience. Once the fear factor is overcome, what follows is usually something akin to a religious conversion. Almost every University or Polytechnic History department will contain a member of staff who will delight in describing the power of his new machine. Of course the computer zealot will label anyone who has not undergone this conversion as a dyed-in-the-wool reactionary. Nevertheless I have yet to meet an academic who, having tried the new technology, refused to use it.

Having absorbed all the hype surrounding computers I chose to make a study of the value of computers for historical research, using the factory community at Styal in Cheshire as a case study.[1] Within a few weeks I had produced pages and pages of statistical tables and graphs, building up a picture of the demographic make up of the community. Using traditional methods the work would have taken months. Quite possibly the volume of material could have discouraged detailed work. In fact the two main existing works on the subject by Mary Rose and Frances Collier only used census data to trace individual families.[2] The computer widened the horizons of my study. The whole project was far more rewarding than a conventional study would have been. Of course the journey to this computer-inspired nirvana was strewn with pitfalls (both practical and theoretical), but more of that later.

Using a computer it was relatively easy to compile general demographic statistics for Styal. For example the population graph (see figure 1) was produced by splitting the population by sex and by age groups (such as 0-9, 10-19 etc). This took a matter of minutes. The high number of persons under twenty denoted a high birth rate, while the correspondingly low number of people over sixty was evidence of a high death rate. Such conditions were common in the nineteenth century and were replicated in the 1861 and 1871 censuses. Similarly the sizes of families and households were measured. The computer was instructed to count the number of individuals in each household and draw a frequency graph from the data (see figure 2). The average household size was 5.4, slightly above the national average.

Such calculations were simple to make, but the computer facilitated a further depth of analysis that was previously impractical. This could be taken to ridiculous extremes. For example, when at a loose end one day it occurred to me it would be interesting to know how many mothers over the age of 30 years, with two or more children worked in the mill. Moments later I had the answer. There were endless variants of this type of question. The breakdown of the population into occupational groups, stratified by age and marital status, demonstrated how the local mill owner Samuel Greg dominated the community. 58% of the village worked for Greg or on his land. However more prominent members of the community had more independence from the mill, preferring to take up a trade or run a farm (see figure 3). By contrast women were more dependent. Greg

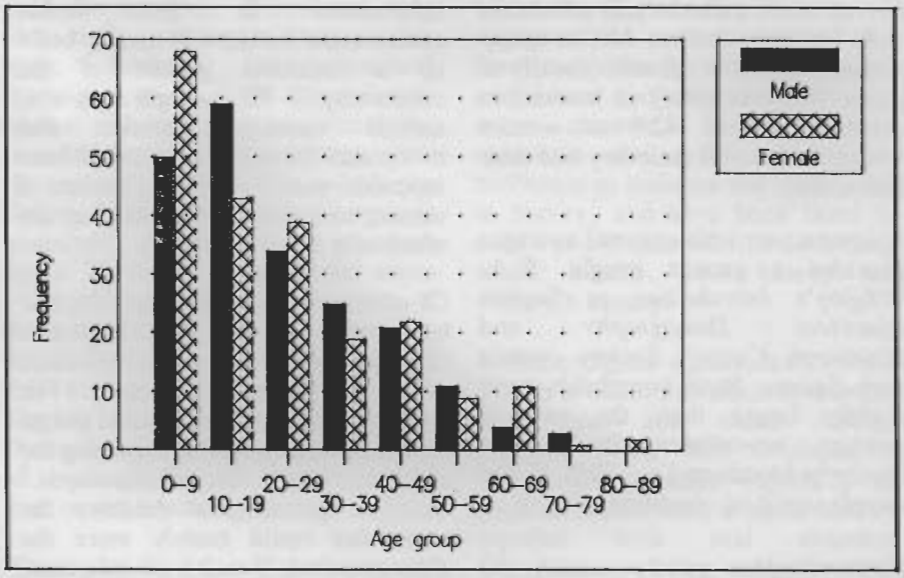


**Figure 1.** *Styal 1851, Population by age and sex.*

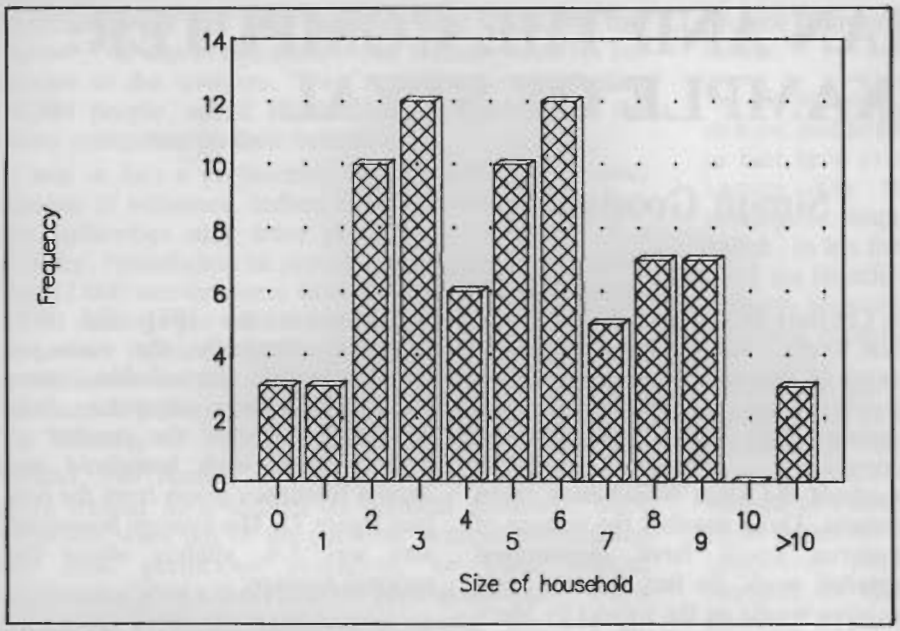


**Figure 2.** *Styal 1851, Household size.*

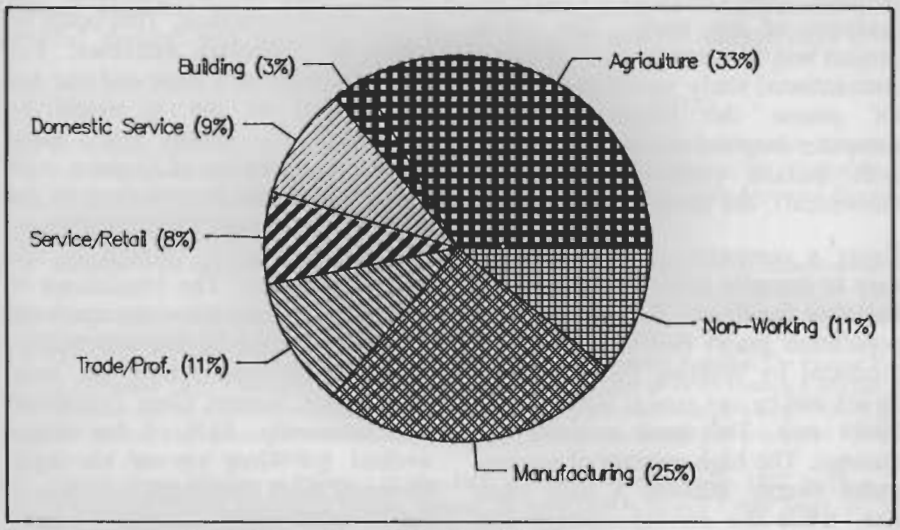


**Figure 3.** *Styal 1851, Occupation of household heads.*

preferred female workers as they proved more reliable and of course were cheaper to hire. Just as many women worked in the mill (usually as power loom weavers) as worked in domestic service. 42% of women worked in the mill until they had their first child.

A comparison with national averages provided a greater insight. E.A. Wrigley's *Introduction to English Historical Demography* and *Nineteenth Century Society* contain such figures. Styal households were slightly larger than the national average, nevertheless the nuclear family predominated — 70% of the sample could be classified as such.

These samples merely scratch the surface of what was possible using a computer. Every single piece of information in the census enumerators' books was used to build up a statistical picture of the community — for example age, sex, marital status, occupation and birthplace. Once all the data had been inputted, most of the questions I wanted to ask were answered at the touch of a key.

Of course a census database can also be used to browse for long lost relatives, or just to provide individual colour in a larger project. For genealogists a computer would put an end to fruitless hours of searching for one name in a list of thousands.[3] Using a genealogical database the researcher could merely enter the name required. If such a record exists, its full details, and that of the entire household it belongs to would be displayed. For example, using the 1851, 1861 and 1871 censuses of Styal three families, the Platts, the Henshalls and the Venables were traced over 30 years.

The results of the Styal project stand as a new historical source, because they contain information that other sources do not. The best example of this would be the addition of household size to the census data.

Furthermore old records are presented in a new format, so as to make them more accessible, and thus unlock some of the insights contained therein. A computer database is just as useful to the amateur genealogist or curious layman as to the professional historian.

However, such advantages could only be realised through using a strict methodology which guaranteed the accuracy of the statistics. This meant careful choice of hardware and software, careful sampling and a strict regimen for transcription. In this case these procedures were applied to the transcription of census returns, but they apply equally to most other documentary sources used in local history, such as parish registers of births, deaths and marriages.

A description of such methodology usually entails using computer jargon which may seem meaningless to the layman. There is no need to feel daunted by jargon. Once you begin using computers, terms like *ROM* and *RAM* will be flowing from the tongue as freely as any term beloved of historians, from *ancien regime* to *cold war*. I am assured that computer scientists are baffled by nineteenth century history, but there is no reason why they can't learn!

The choice of computer will have a great bearing on the success of any project. The machines on sale in the high street are of variable quality and power. A comparison of two popular machines demonstrates this. The IBM PC can store large amounts of data, which it can process quickly, while Amstrad PCW working with the same memory operates at less than half the speed, making complex operations impractical. Today micro-computers can be divided into two groups — those that are IBM compatible and those that are not. IBM compatibility offers the possibility of data being transferred between virtually any

machine, and is commonly found in schools and colleges. The addition of a hard disc drive makes the storage of data more reliable, and a printer is essential to make a permanent record of work.

Important though it is, the choice of computer is secondary to the choice of software. The same software packages can usually be made to work on any IBM machine with similar results. However the quality of individual packages varies greatly, so it is vital to make a considered choice.

An examination of the latest computer magazines will reveal the latest packages.[4]

There are many applications computers can be put to, and a correspondingly wide range of software packages. The analysis of census returns involves the storage and retrieval of textual data. For this a database is used, although most spreadsheets perform a similar function. Both store data in the form of a table. Each row of the table represents one *record* (or in the case of census returns the entry for one individual). Each column of the table is known as a *field*. For example, the name, age, sex, marital status and so on are the fields that make up a record. Databases are more adept at processing textual data, while spreadsheets are more suitable for numeric data — for example calculating the age difference between spouses or a mean age. Among the most popular and best software packages are the databases DBASE III and PARADOX, the spreadsheet LOTUS 123 and the word-processor WORD PERFECT. Hybrid packages, incorporating the best of both have been developed by some academic institutions.[5]

Most computer projects in history require the use of sampling techniques because the amount of data to be processed is so great. This is roughly the same technique used by opinion pollsters to get a representative survey of the population. Sampling should not impair the accuracy of the work being undertaken. Various techniques can be used but the most common is systematic random sampling which involves taking every fifth or tenth household from the census (for a 20% or 10% sample respectively). The size of sample taken should depend on the size of the area under study. For example a 10% sample of a town of Preston's size would be more than adequate.

Sampling takes place during the transcription of documentary sources to transcription sheets, which can be taken to the computer to be entered into a database. Census returns are found on micro film or xerox copy from Public Record offices and libraries. Unlike parish records, census returns are easy to find (local history societies produce indexes to this end).

Ideally documentary sources should be transcribed as they stand, warts and all, and without abbreviation if possible. Since in effect the historian is creating a new historical source the original should be reproduced exactly where possible. However, the alteration of data during transcription is a more complex question than it first appears. Errors are common in census enumerators' books and parish registers but since they are not common enough to be statistically significant, they should be transcribed. Censuses of the Celtic fringes often misspelled names, thus making comparison with sources written in local dialect more difficult.

The conscience and literacy of the author of a document affects non-computer analysis too, so spending time correcting 'mistakes' is a red herring. Nevertheless it is possible to make the transcription more accurate if a strict set of guidelines is followed. This should only apply to obvious mistakes, for example, where the forename clashes with the sex in a census book or a marriage register.

Once the data has been entered into the computer, the software program is used for analysis. Procedures may vary from program to program but most databases use standard techniques and jargon. For example an historian may want to research the occupations of married women over the age of 30. For this search the fields marital status, sex and age are examined. All the records in which age > 30, sex = f and marital status = m will be displayed. The results of the search can be placed in a new file and subjected to further analysis. In this way an historian's question in English is translated into a form which the computer can understand. All that is required is logical thought on the part of the historian.

There are limitations to the use of computers. Despite the recent fall in prices, computers are still expensive. As late as January 1991 the bottom-of-the-range machine, the Amstrad PCW could be obtained from around £300 which includes a printer. For a large project an IBM compatible machine is needed. Prices for these machines vary from four hundred pounds to several thousand, usually excluding a printer. However these machines can be used for other tasks too, such as word processing. The investment is definitely worthwhile.

Inevitably coming to terms with new technology causes problems for many people, particularly those involved in the arts and humanities. Admittedly computer scientists and manufacturers do not make the transition any easier. Any manual which claims that its program is user friendly is being economical with the truth. Computer packages are written by people who have long forgotten what it is like to learn to use a computer for the first time and have no sympathy for the beginner. In fact computer scientists have as great a propensity for meaningless jargon as some noted American sociologists. Such people have left academics way behind in their search for more sophisticated software.

In addition to these practical problems (which should not prove to be a major obstacle to most people), sceptical historians have raised objections to computers in principle. The theoretical issues involved are those which apply equally to any demographic research, using statistics in addition to documentary sources. The issues are raised largely by those for whom the document is king. Using a computer does not divest the historian of his responsibility to engage all the available sources. Far from it, it actually widens his scope, and if used correctly makes his work more accurate. Most of these criticisms are symptomatic of a resistance to embrace any new trends in history, and have been faced by each wave of innovation, most notably social history, oral history and women's history.

In fact computer aided history (if such a term is valid) has much in common with oral and social history. Computers are best employed in demographic studies especially the population data used in social history. together with oral evidence. Computers are being used to research the history of the People, not just their

rulers, the details of their lives, not just the special events which are recorded on paper. These details are the substance of local history. Meanwhile the use of computers in History classrooms and higher education institutions has almost become the norm. The HiDES projects, developed at Southampton University is a fine example of this.[6] HiDES is a sophisticated package used for interrogating historical documents in schools and higher education. There are countless other examples of software programs in use.

Using computers the potential exists to pool the vast array of computer records which are being created into a central database. This had already been achieved in France and Germany using the software program known as CLIO, and at Hull University where the Domesday Book has been computerised. A group of social historians at Cambridge University (CAMPOP) has obtained a large grant to begin work on a major scheme to pool computerised parish records from all parts of the U.K. Academic institutions and individuals could access this database via the telephone network. The potential is virtually unlimited. Some years ago the French historian LeRoy Ladurie claimed "tomorrow's historian will have to program in order to survive". This may be an exaggeration, but the benefits are such as to make the acquisition of a computer increasingly common among local historians.

**NOTES**

1. S.T.Goodwin, 'A study of the value of computers for historical research with a case study of the social history of Styal, 1851-1871.' (unpublished BA. dissertation, Manchester Polytechnic, 1990).
2. M.B.Rose, *The Gregs of Quarry Bank Mill*. F. Collier, *The Family Economy of the Working Class*.
3. I.Macalpine, 'The Use of Computers to Analyse Parish Register Entries', *Manchester Genealogist*, (4), 1987, pp. 94-95. J.Bunting, 'Family Life', *8000 Plus*, (13), October 1987, pp29-30.
4. The best general computer magazine is Personal Computer World.
5. M.Thaller, 'Methods and Techniques of Historical Computation' in P.Denley and D.Hopkin (eds), *History and Computing* (Manchester, 1987).
6. For details of HiDES contact: The Hides Project, the History Department, Southampton University, Southampton, Hampshire SO9 5NH.