How to construct an Access Database - and why you should seriously consider it

Jeremy Boulton (Newcastle University). Email: Jeremy.Boulton@ncl.ac.uk

Database Course for the ESRC Newcastle-Durham DTC

This manual, related databases and other material are available to download at:

http://www.staff.ncl.ac.uk/j.p.boulton/esrcdtcdatabasetraining.htm

Contents:
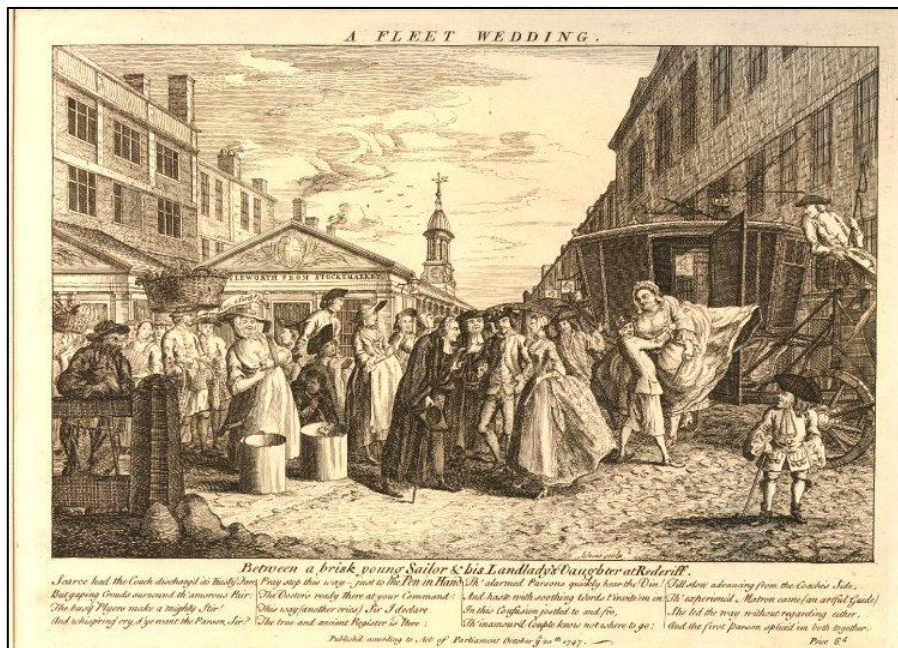
A Fleet wedding between a brisk young sailor & his landlady's daughter at Rederiff: Satirical view of London life, with a clergyman marrying the couple in the Fleet Market. 20 October 1747. British Museum.

Timing of first session

First session to take place in the FAYOL cluster, room 2.96 Armstrong Building, Newcastle University. Those taking the course should meet in the foyer of the School of Historical Studies, first floor, Armstrong Building, at 2.45 pm, Friday, December 2nd, 2011. The session should take no more than two hours.

Introduction: What am I doing here?

There are some things that you need to think about before constructing your very own database. You might also have some questions about it. This opening section is designed to answer some of those questions.

1) **Do I really need to do this?**
2) **What sort of historical material is suitable for a database?**
3) **I am frightened of statistics and I 'hate computers'?**
4) **How can I be sure that my material would make a suitable database?**

****

1) **Do I really need to do this?**

A great deal of time spent researching for PHD's is completely wasted. Students follow paths that do not lead anywhere, inadvertently duplicate other people's work, or come across material that undermines their methodology or that suggests new and better ways of approaching the same problem.

It is important, therefore, that your database will be a useful tool for analyzing data, and does not, on its own, become a new way of wasting precious research time.

A well constructed database *can save you immense amounts of research time*, enable you to answer research questions you might not otherwise be able to examine, and may well *shed unexpected light on material that were not apparent before the analysis*.

A poorly constructed database might not do any of the above, and might actually cost you time, drive you mad and make you curse the day you ever heard the words 'database' or 'research', or, for that matter, 'Jeremy Boulton'.

2) **What sort of historical material is suitable for a database?**

It is vitally important that you only construct a database if your primary research material is suitable. Nowadays almost *all* historical research material *is* suitable (since both qualitative and quantitative database software exist) but, that said, databases are especially highly recommended for:

Any sort of historical material which has a clear structure and pattern (such as those that take a regular format, are in lists or columns)

Any sort of historical material which contains regular information (such a date or occupation field) even though the text of the material may be highly varied (such as wills)

Any sort of historical material which is being used to gathered to shed light on people (e.g. the construction of biographical material on a group of individuals) or perhaps buildings (accounts of payments for construction, for example)

Virtually any data that exists on a very large scale.

3) **I am frightened of statistics and I 'hate computers'**

This is still a depressingly common reaction to the use of the word 'databases'. Here are my responses:

a. **Get over it**. As other sessions on this course will make clear, all serious historians now engage with databases on a daily basis. Databases containing bibliographic material are now the norm and a number of specialist databases can be delivered on disk. Databases containing a wealth of primary source material, delivered online, are also becoming increasingly common. Some of these may look like simple transcripts, but are in fact 'tagged' so that particular types of information can be extracted and analyzed by the user. These simply are simply a different type of database to the ones run normally on a PC, but they are still databases.

b. **Statistics have nothing to do with databases**. It is still a common misconception amongst some historians (and especially among many lecturers) that only those with statistical awareness or advanced IT skills can or should use a database. This is, in fact, stupid. Modern databases are ideal ways of sifting, sorting and manipulating historical material. They can also perform very simple counts and generate basic statistics of the sort that most people did at GCSE or equivalent. No serious historian can embark on historical research without engaging with databases in some form.

c. **You only hate what you fear.** Fear of computers almost always means lack of familiarity and practice. Running databases is not rocket science. If it was, Microsoft Access would not be the money spinner it is today. Everything about such software is designed to make it as simple to use as possible. This is so much the case that some of the 'step-by-step' procedures detailed below will seem over simple or obvious to many of you. I apologize if this is so, but for those unfamiliar with database software (or, actually) any software, following a few simple step-by-step procedures early on helps to make the essentials clear. It also means you can repeat this lesson at your own speed later. You will soon find that much of it is intuitive.

d. Remember that you cannot 'break' a computer by pressing the wrong buttons and that it is actually quite difficult to destroy a database by doing the same thing.

e. Lastly, it is worth remembering that having some knowledge of Microsoft Access is a *very marketable transferable skill*.

4) **How can I be sure that my material would make a suitable database?**

    a. This session contains an introduction on how to construct a database from scratch. It uses a pre-designed database as a teaching aid.

    b. Think about the structure and format of the data you hope to use for your doctoral work.

    c. The person running the session will be happy to discuss the suitability of any primary material you have brought with you.

The Fleet Marriage Database (© Jeremy Boulton and Jacob Field)

Introduction

THE SAILORS FLEET WEDDING ENTERTAINMENT.
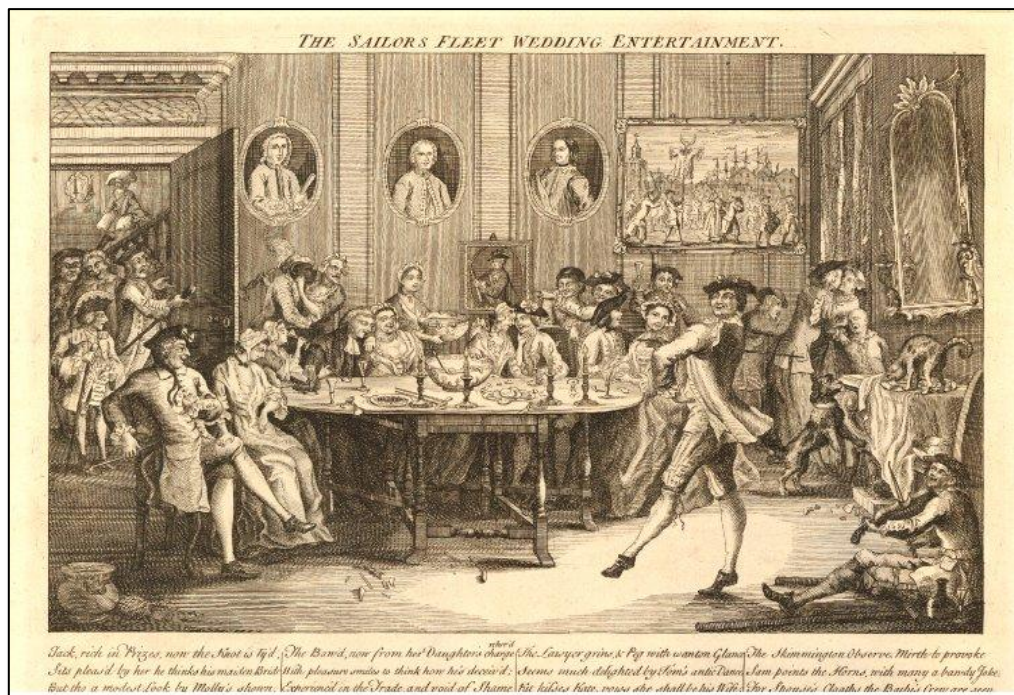
Satirical view of London life, with a riotous wedding party: 10 November 1747 Etching. British Museum.

One of the oddest features of London life in the first half of the eighteenth century must be the huge popularity of Fleet Marriages. The custom of marrying 'clandestinely' began in the previous century, when thousands of Londoners married in a few churches located outside ecclesiastical jurisdictions. With the closure of such sites, however, increasingly large numbers of London couples turned to the Fleet Prison and chapels located within its 'Rules' (an area around the prison exempt from criminal and civil jurisdiction). Although such Fleet Weddings were technically clandestine in law (not taking place in the couples' local church, lacking consent of parents and so on) they were nonetheless legally valid unions. They were also relatively cheap, often included an all-in package that included a wedding party in an adjoining alehouse or tavern (as above), and could be undertaken quickly and secretly. By the mid eighteenth century *thousands* of London couples were marrying in the Rules of the Fleet *every year*. Estimates vary, but between *one half and three quarters of all London weddings were taking place at the Fleet in the middle of the eighteenth century*. The custom ended abruptly in 1753, when Chancellor Hardwicke, tiring of the legal ambiguities such unions produced in court, got an Act onto the Statute books that removed their legal validity completely. Henceforth, with few exceptions, only couples marrying in a local parish church according to the rites of the Church of England, had a valid, legally-binding marriage.

Research questions

The Fleet Marriage Registers exist in very large numbers and are held at The National Archives under the RG 7 call mark. They represent both a challenge and an opportunity for those interested in London's economic and social history. Those historians interested in London's economy are beginning to use the registers to look at the occupational structure of the capital. Arguably, before the censuses of the nineteenth

century, only the Fleet marriage registers supply occupational data on a London-wide scale. This is important because there were about 150 parishes in the capital, each one of which might have had a distinctive occupational structure. For such exercises to produce valid results, however, we need to know whether the Fleet marriages were socially biased (towards the poor, who may have been keener to pay lower fees) or spatially biased (perhaps some parishes were more likely to have couples marrying at the Fleet than others). Moreover, the registers themselves are chaotic, with very large numbers of duplicate entries. Demographers interested in London's marriage patterns are similarly interested in the representativeness of the Fleet marriages. Demographers hoping to carry out 'Family Reconstitutions' in any London parish, for example, would need to recover 'missing' marriages in the Fleet registers – a very challenging undertaking given that there are around 237 *separate* registers containing, in all, over half a million marriages, many of them duplicates.

The Fleet Marriages pose such challenges to the [Pauper Lives Project](), which focuses on the Westminster parish of St Martin in the Fields. One of the aims of this project is to study infant mortality in the parish, based partly on a limited Family Reconstitution. For this reason both the representativeness, coverage and biases deriving from the phenomenon of the Fleet Marriages need to be confronted. The research questions therefore are:

> How common was it for couples in St Martin's to marry in the Fleet?
> What occupations are represented in such marriages?
> How many entries are duplicates?
> What characteristics (social and marital status) did Fleet couples possess?

To answer these questions a hapless Research Assistant was hired to extract all marriages from the 237 surviving Fleet Registers that stated that either the bride or groom or both were from 'St Martin in the Fields', for the period 1710-1753 inclusive. This produced the database we are using today. Each marriage is a single record in the database. The database contains 9,417 'records' i.e. marriages.

Select Bibliography

Brown, R. L. (1981). 'The Rise and Fall of the Fleet Marriages', in R. B. Outhwaite, ed., *Marriage and Society*, London: Europa, 117-136.

Boulton, J. (1993). 'Clandestine marriages in London: an examination of a neglected urban variable'. *Urban History Yearbook* 20(2): 191-210.

Herber, M. (1998). *Clandestine Marriages in the Chapel and Rules of the Fleet Prison 1680-1754*. London, Francis Boutel Publishers.

Outhwaite, R. B. (1995). *Clandestine marriage in England, 1500-1850.* London, Hambledon.

Probert, R. and Brown, L. D. A. (2008). 'The impact of the Clandestine Marriages Act : three case-studies in conformity', *Continuity and Change* 23(2): 309-330.
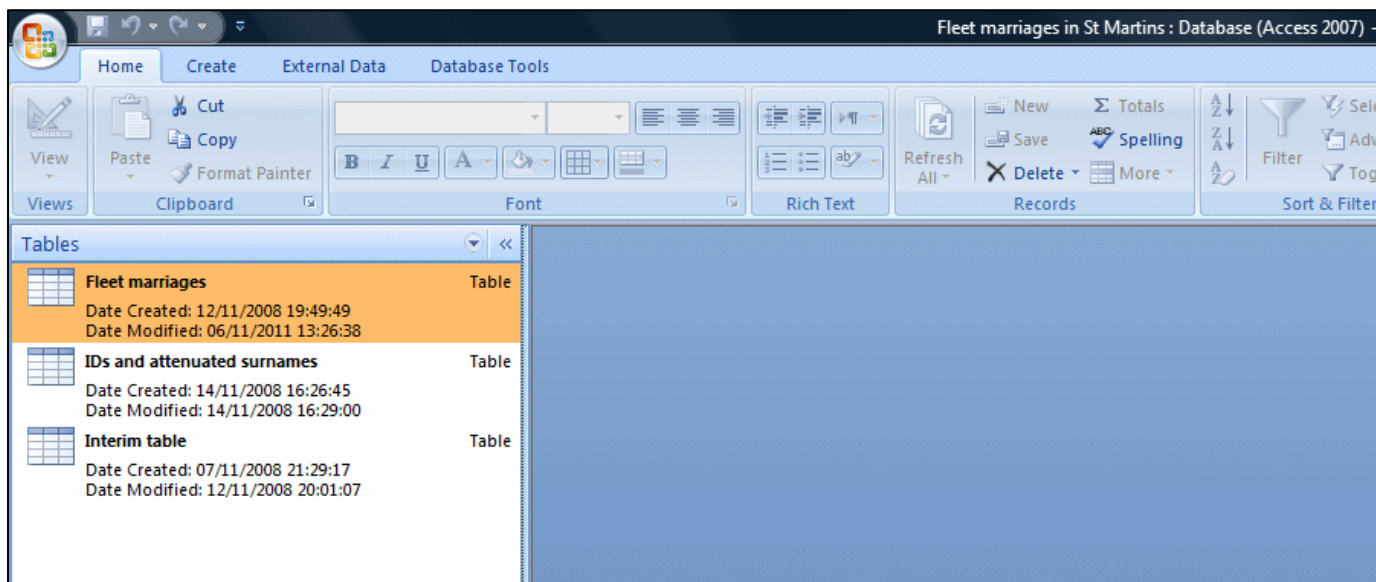
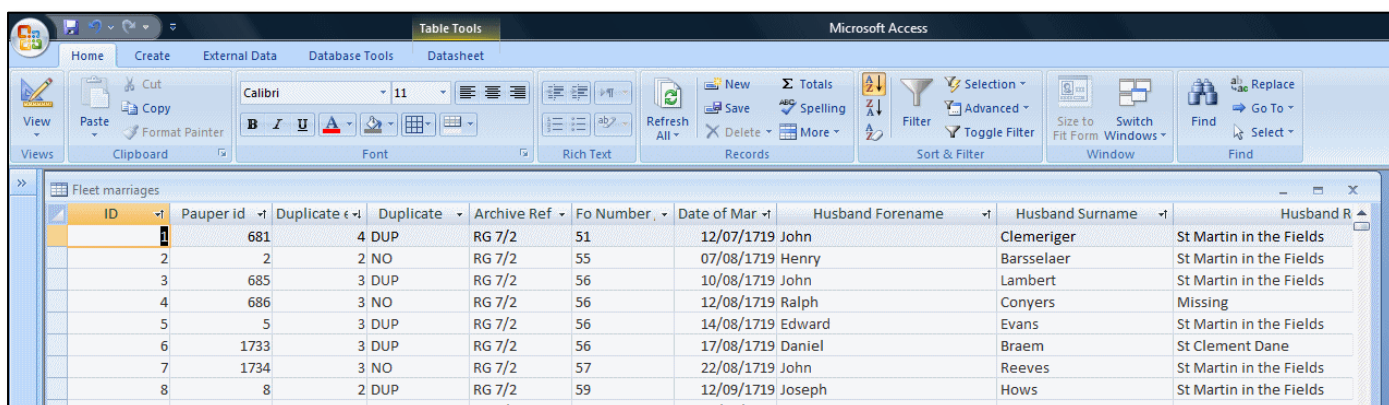**The structure of the database: a step by step look**

Open Access 2007 and load the database:

Fleet marriages in St Martins.accdb

1. You should see something like this:

2. Double click on the highlighted Fleet Marriages Table, and you should see something like this:

Use the mouse or tab or arrow keys to move along left to right to look at the whole thing. Each marriage is one line of data ('a record'). Each record ('entity') has a number of 'attributes', entered as separate fields of information.

Each marriage has its own unique ID number, there is a Pauper ID number to enable the compiler to identify the same individuals. The duplicate entries per ID field assigns a number to let the compiler know how many duplicates there are for each Pauper ID. The Duplicate field simply identifies whether or not a given record is duplicated elsewhere. Original records therefore are given a NO, subsequent duplicates a DUP. This field was entered only after all the data had been entered, and represents the result of a very extensive exercise in nominal linkage within the dataset.

The next two fields identify archive and folio numbers. The date field is the date of the marriage. The next fields detail information about bride and groom, giving their names, residence, marital status and

occupation. Notice there is no field for bride's occupation, since such information was never supplied. Other fields supply details (where given) of the presiding minister, any fee charged, details on marriage location and two general fields for notes recorded by the RA when entering the data.

There are some general points that derive from this brief examination:

Before constructing a database you must know the structure and content of your data. Fields can be added later, but it is better to begin with a structure that captures most of the data from the start.

Many fields may contain missing information and are blank.

Each record has its own unique ID. The ID number will enable the compiler *to reconstruct the order in which the data was entered*. You should set this field as an auto-number, so that Access will add it as you enter your data.

It is important to enter data like archive and page or folio numbers so that you can cite individual records in your publications.

Data collection takes place first. Only when all data is entered should the database be edited to add further information, coding, etc.

Never *edit* data as you *enter* it.

Entering data can be productive, since knowing your data will give you ideas about how to use it, and what problems there may be within it.

Entering data, of course, can also be mindless and boring. However, constructing a large database is useful not just for the compiler, but, if done properly, can also benefit future generations of scholars.

Future users of this database, for example, could augment the database by adding more information about the couples, or by adding more parishes to the data, or by asking different questions. They could, for example, look up these Fleet couples in the massive online archive represented by the London Lives Project (http://www.londonlives.org/). This huge online database of primary material actually includes other classes of record from the Pauper Lives Project, notably workhouse admission and discharge registers and settlement examinations.

Each field of information has its own characteristics. These should be set and decided upon before data is collected. This can only be done, therefore, if you know the structure of your data.

3. **In order to look at the structure of the Table**, click on the View button.
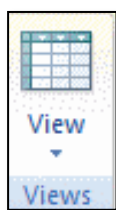
This should produce the following breakdown of the Table structure:
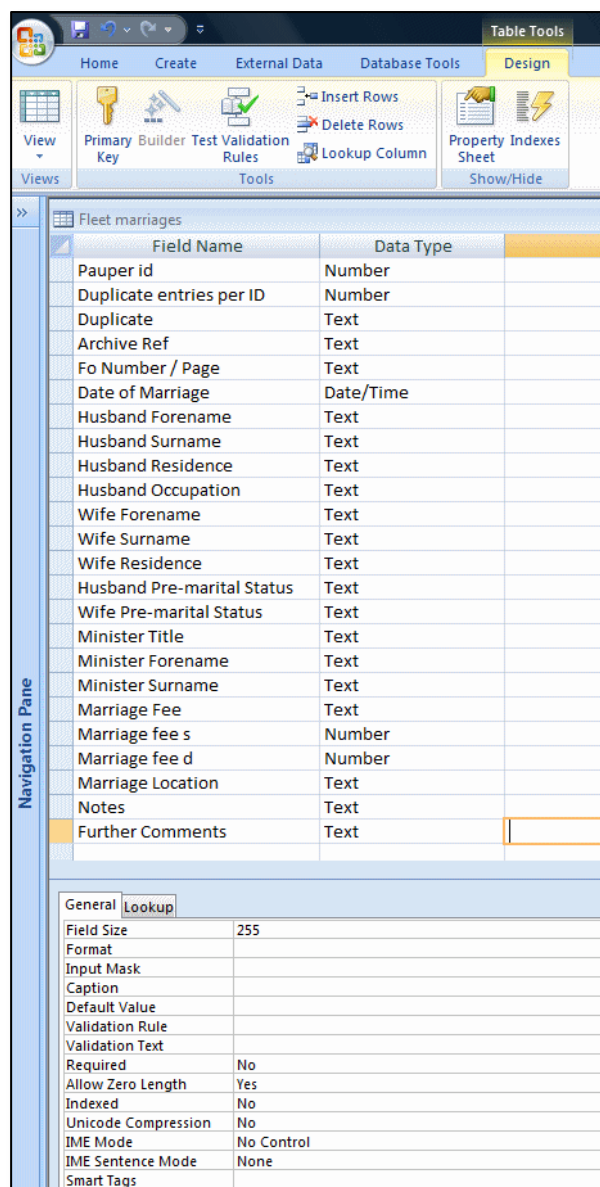
By clicking in the data type column, you will see the characteristics of each field displayed in the General tab at the bottom of the screen. Here, the 'Further comments' field is defined as 'text', with the maximum field size of 255 characters. Remember that the size of the database (the space it takes up on your computer) depends partly on the amount of space you allocate to each field. Even empty space takes up resources.

At this point, when constructing a database, you can force 'default values', require a field to be completed (required = yes), set formats and add validation rules (that limit the size or nature of data entered). It is possible at this point to add a 'lookup' table, from which to choose values.

As you can see, almost all fields here happen to be *text* fields, with the exception of the date field, marriage fees (s and d) and the ID fields. Text fields are just that, *text*. They *cannot* be counted or summed. Even numbers are treated as text in text fields. This is important because if you enter data which looks numeric, but includes a text character, it will be treated as text or it will generate an error. Thus the old English currency 2s 6d (two shillings and six pence in pre-decimal currency) has, here, been entered as a *text* field. To recover and count the fees, we need two number fields derived from the original text field to preserve the numeric fee. It is defensible to ENTER the fees as text as in the original, since you should not *alter* data when you enter it, if at all possible, or slow down the process of data entry by making numerical

9

transformations. *Always enter the data you see*. Editing and adding fields should be done only after data entry is completed. An alternative method to that used here would have been to enter the data in three numeric columns (*£, s, d*).

4. Manipulating data in the Table.

Fortunately an RA has already entered all the data, and the PI (Principal Investigator) has added the other necessary fields. It is now possible to use filters to have a look at the data in more detail.

Databases are particularly good at the manipulation of data. For example, you can use the Selection button to sort out a subset of the data. To do this, click on the Duplicate Field and highlight a 'No'. Then click on the Selection button and click on Equals 'NO':
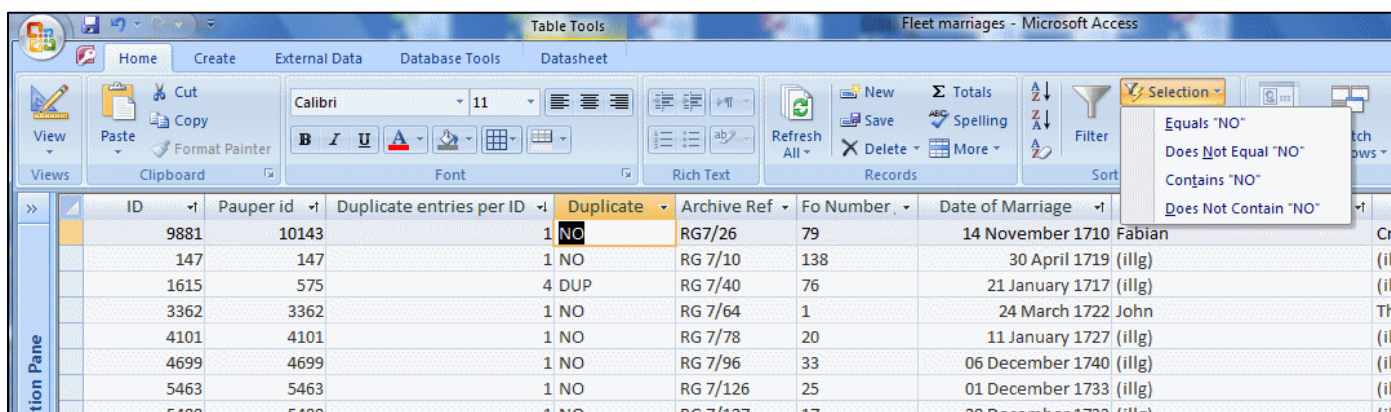
You should get a subset of the unique marriages, consisting of 6751 records, with all the NO fields. To revert to the original database, just click on 'Toggle Filter'. You can also filter by using the Advanced button.
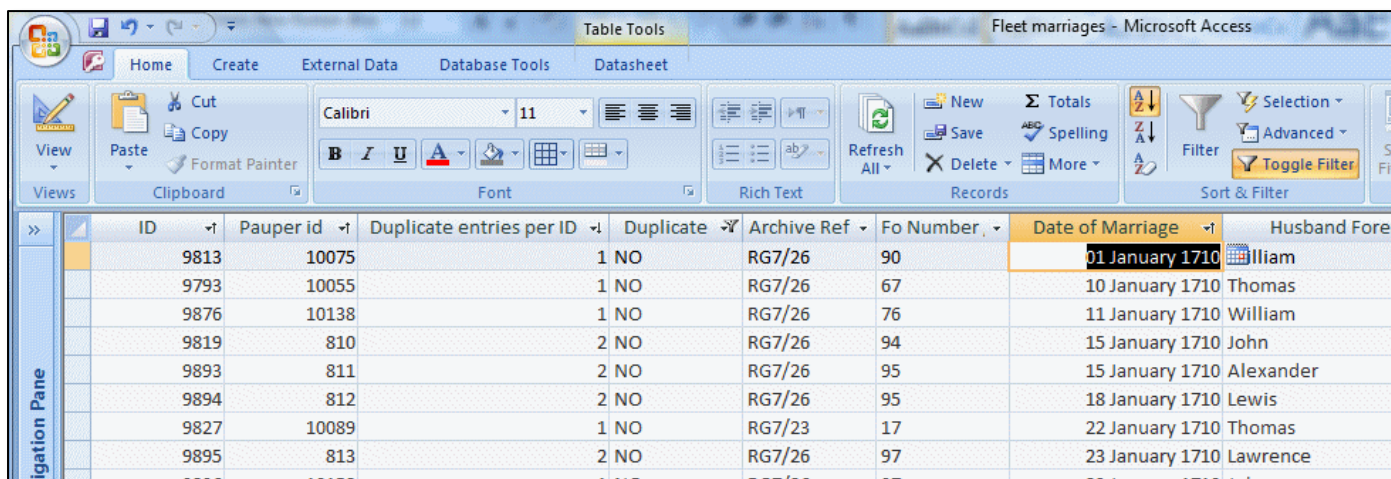
With the filter on, you can sort the subset, by any of the variables in the various fields. You could also filter the subset. Thus, for example, to find out how many unique records took place in 1710, just select 1710 and choose 'ends with 1710' via the Selection button. **There should be 155 unique entries ('NO' in duplicate field) which took place in 1710**. Such filters are useful ways of finding out about your data, before you run queries. They can also be used to EDIT subsets of data, using the Find and Replace buttons. When you have

finished hit the toggle button to get back to the original unfiltered data. Try playing around with filters to find out:

How many unique marriages took place involving a groom from 'Chelsea' (*Answer*: 54 if you select 'Contains Chelsea' in the husband's residence field and NO in Duplicate field).

How many unique entries involving a waterman? (*Answer*: 73 if you select 'contains Waterman' in husband's occupation field and 'equals NO' in Duplicate field).

How many unique entries involved a widower? (*Answer*: 1324 if you select 'equals widower' in husband pre-marital status and 'equals NO' in Duplicate field).

How many unique entries involved a widower marrying a widow? (*Answer*: 753 if you select 'equals widower' in husband pre-marital status field and 'equals widow' in wives pre-marital status field and 'equals NO' in Duplicate field).

Return to contents page

## Querying the Fleet Marriages Table

A simple count:

To query the table, you need to switch to the list of pre-prepared queries.
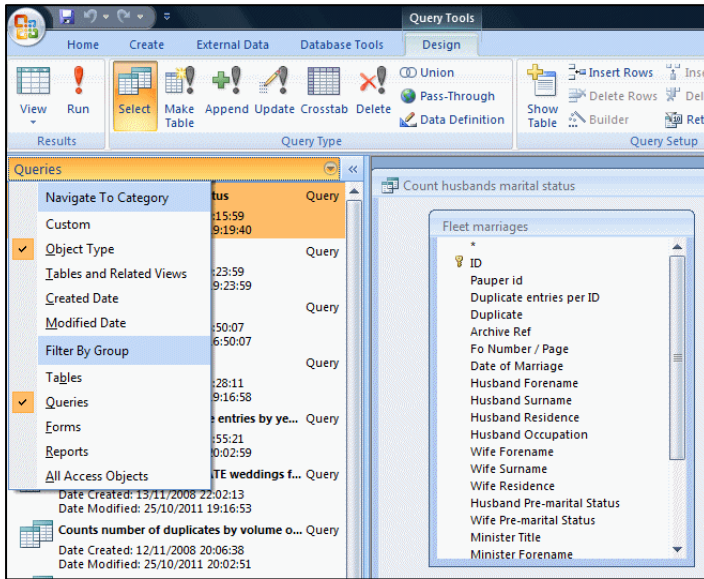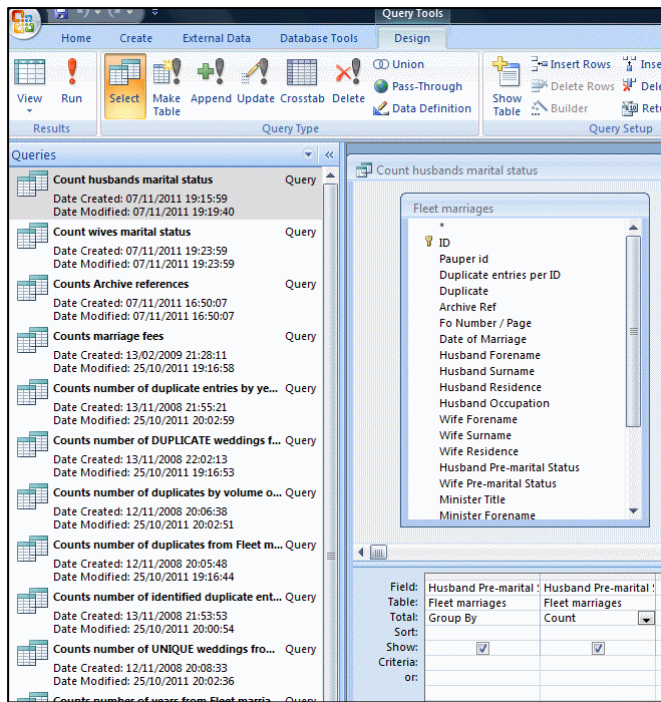

Click on the drop down menu to get the Object type, Queries view. If you double click on this, it will *automatically* run the query.

To see the query STRUCTURE rather than the results, RIGHT click on the Count husbands marital status query, as depicted below, and click on Design View:
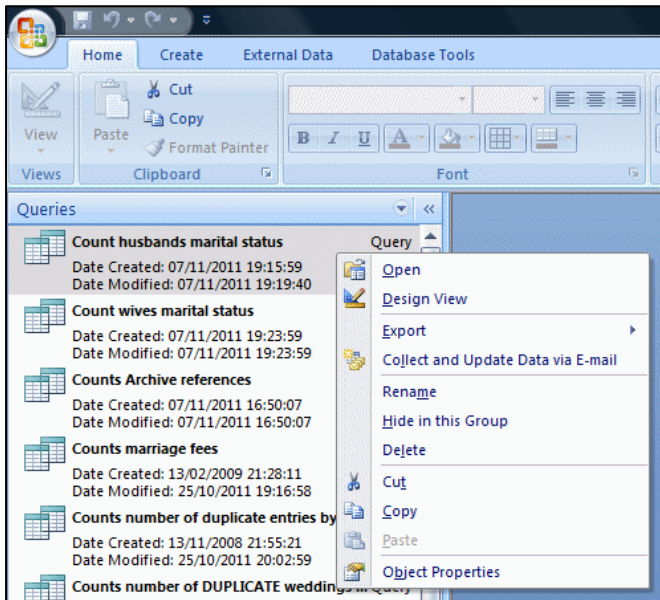
This is the structure of the Count husbands' marital status query.

Clicking on View will run the query.

And this should be the result:

This is a simple count of the contents of the husband's pre-marital status field. 69 were illegible, 6632 were stated to be bachelors, 856 records had missing information and 1860 were stated to be widowers. Note that this is a count of *all* records, including those which were duplicates.

A simple count using a query rather than a table: how to generate annual counts from DATE fields

In Access you can run queries against existing queries. Access processes the existing query and then runs your new query against its output. This is incredibly useful.

One particularly useful way of using this Access function is to turn DATE fields into Annual counts. If you counted a DATE field (that is a field formatted and defined as a DATE field in Access) you would only get a count grouped by each date. In order to generate an *annual* count from a date field,

you need firstly to create a query that generates a year from a date. Then you construct a count of that initial query.

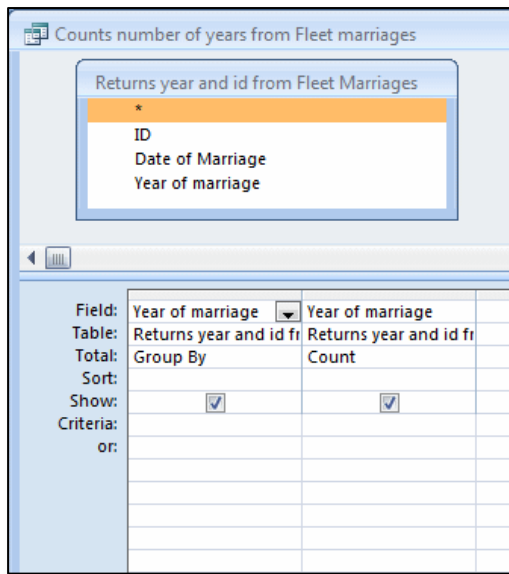Open the query **Returns year and id from Fleet Marriages**
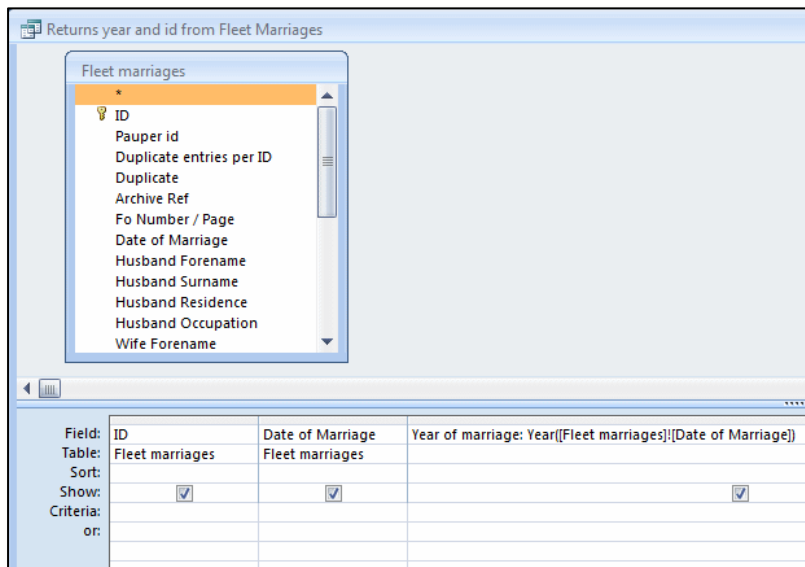
Note that the Field: Year of Marriage: Year ([Fleet marriages]![Date of Marriage]) has been created by right clicking in that field and using the Build function. YEAR is a simple command which returns years from DATE fields. There is a similar function called MONTH which returns months from DATE fields. If you run the query, you get:

This is a basic query that you should always construct if you have a DATE field in your database. Some users construct separate fields in their tables in which they enter year values, separately from date values. This is extra work. With the **Returns year and id from Fleet Marriages** query, you can generate a year value for each unique ID in any subsequent query.
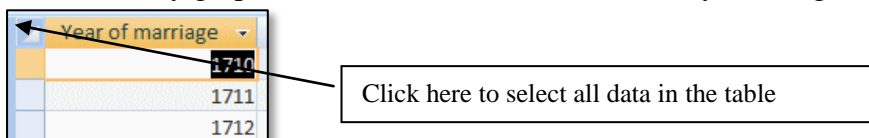
One thing you can do is to *count* the **Returns year and id from Fleet Marriages** query to generate an annual count. Look at the **Counts number of years from Fleet Marriages** query.

This returns the following result:

You can easily graph this in Excel. Select all the data by clicking on the top left hand cell:

Click here to select all data in the table

Copy the selection. Then Open Excel and paste in the data. Delete the Year of marriage column *heading* (this makes it easier to draw graphs from data in Excel), select all the data and *insert* a simple line graph, and you should get a graph that looks like the figure below:

CountOfYear of marriage
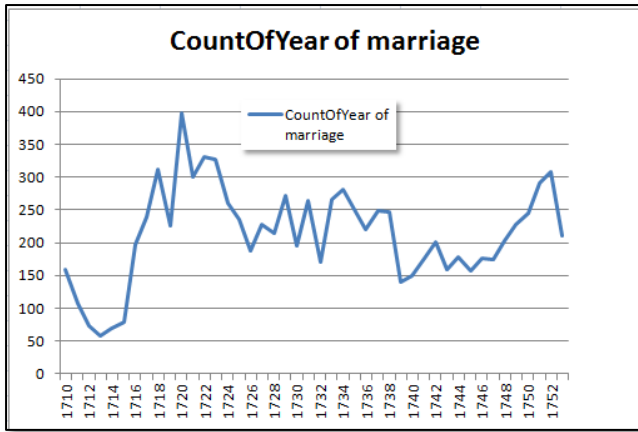
How to develop simple queries

The above procedure counts the annual number of all Fleet marriages, including the duplicates. You could drag down the Duplicate field and select only cases equalling NO to count the unique marriages. This then gives you a simple annual picture of the incidence of duplicates in the database:
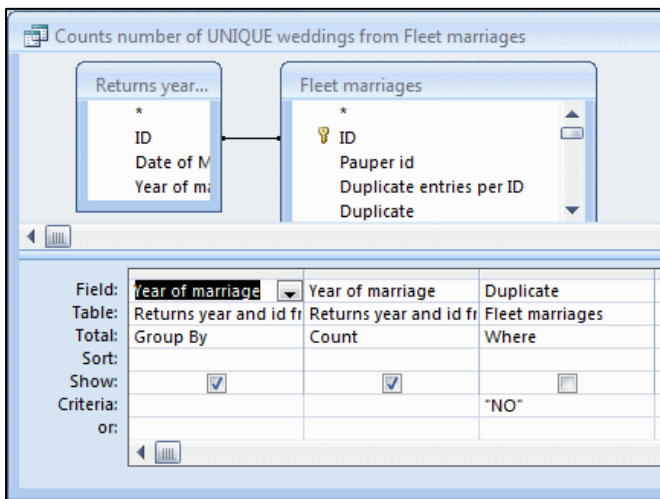
This should generate the following results:

Year of marriage	CountOfYear of marriage
1710	155
1711	95
1712	69
1713	50
1714	65
1715	70
1716	124
1717	120
1718	145
1719	138
1720	190
1721	184

Record: 1 of 44   No Filter   Search

16

You could then find out how many *individual* soldiers were getting married from the parish at the Fleet every year. Note that you need to have some idea about how such soldiers were described before you can design the query. The *Where* function in the *Total* line of the query is a very useful command:
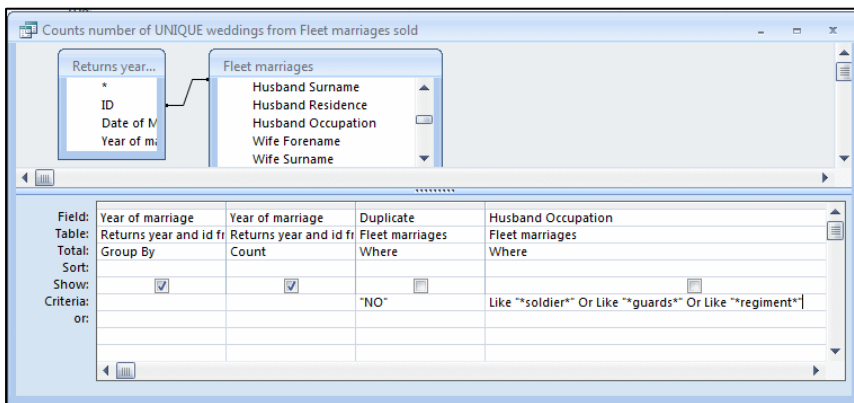
This is a count of all individuals whose occupation field *contained* the words 'soldier' *or* 'guards' *or* 'regiment'. This should generate the following results:
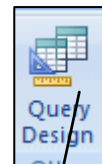

The obvious message here is that with queries *you should start from a simple query and build up to more complex ones*.

> Note that * is a wild character, which stands for *any other character(s)*. So selecting *where* in the Total field and the criteria *stuff* would return all records *containing* the word 'stuff'. If you just put stuff in the criteria *without* the wild character * on both sides, Access would return records that contained *only* the word stuff and nothing else.
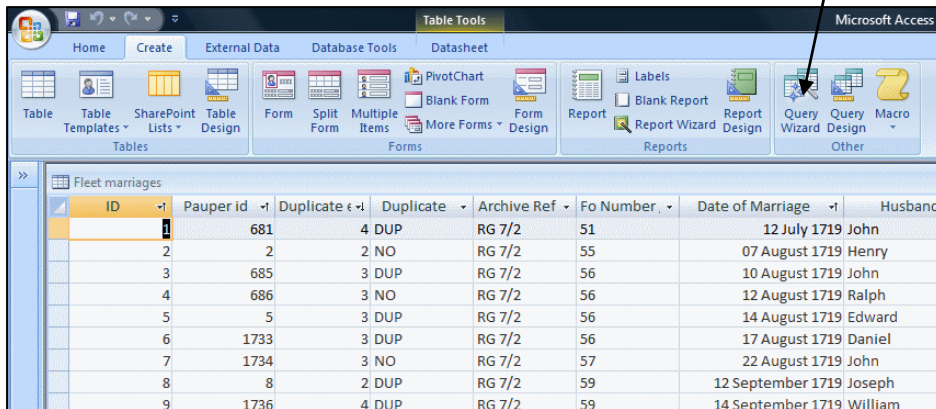
How to construct a query

With the database open, begin by clicking on Create/ Query   Design button :

This will produce the Show Table/Query dialogue Box. Select the Fleet Marriages Table.

To construct a query simply drag the field you want to count down TWICE.

In the Table row select 'Group by' in one column and 'Count' in the other, and then click on the Sum button:
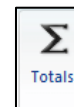
Try constructing a query to count the Husbands' forename field. You can sort the results by either smallest or largest (ascending or descending).
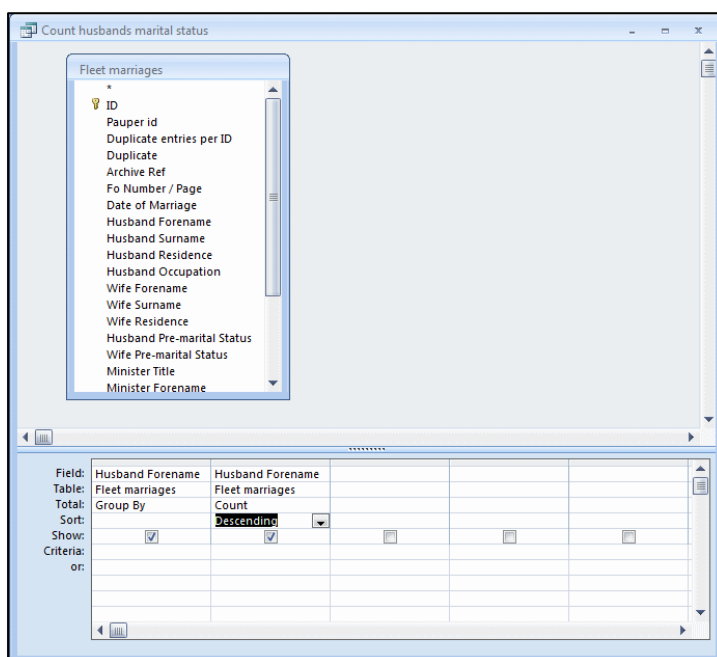
The result should look like exactly like this, sorted by descending forename…

Play around with some simple queries of your own.

How to construct a database from scratch: The Westminster Infirmary (founded 1720)

This can be regarded as homework.

There is no reason that you should not use your own material to construct a database in class.
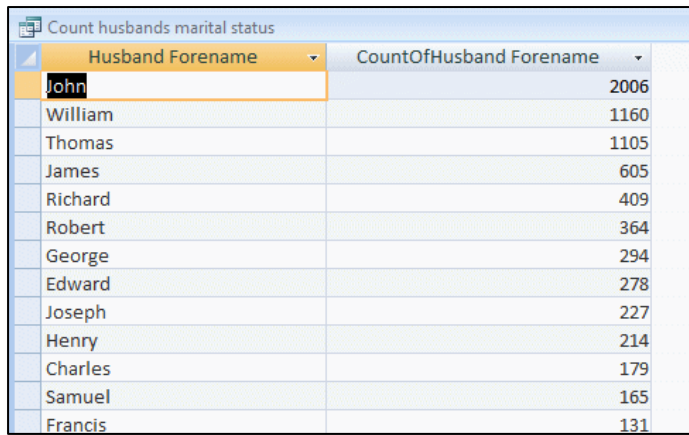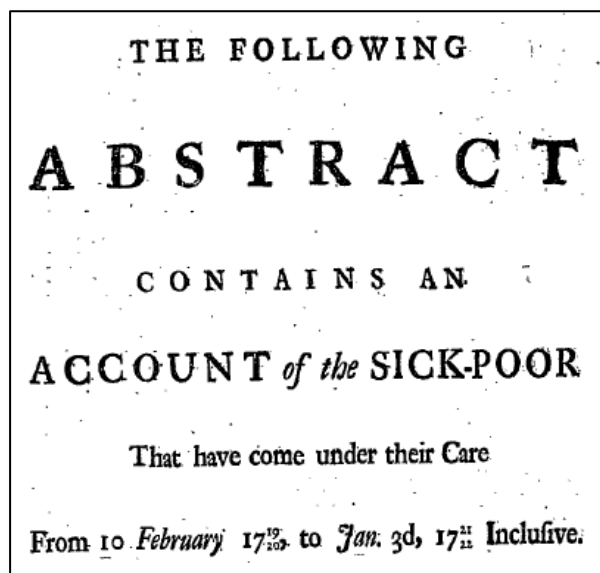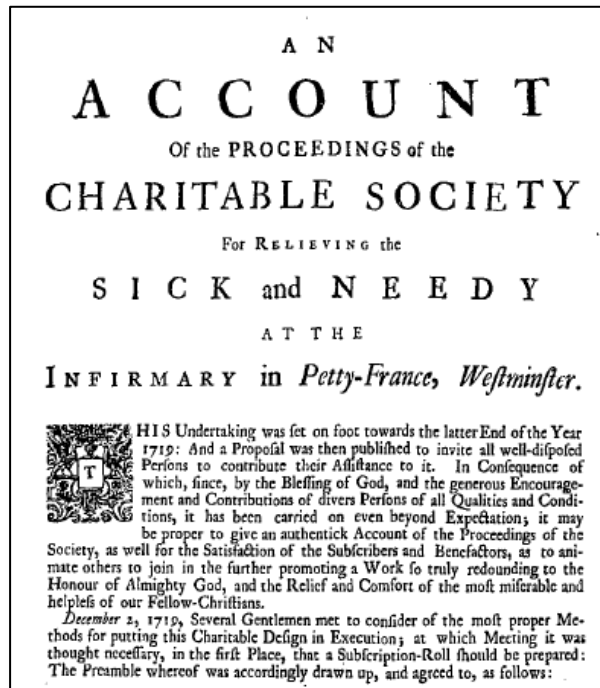
In case you haven't got one to hand, I have provided some suitable material. This is a list of patients admitted to the Westminster Infirmary, which opened in 1720.

A N

A C C O U N T

Of the PROCEEDINGS of the

CHARITABLE SOCIETY

For RELIEVING the

S I C K and N E E D Y

AT THE

INFIRMARY in *Petty-France, Weftminfter.*

HIS Undertaking was fet on foot towards the latter End of the Year 1719: And a Propofal was then publifhed to invite all well-difpofed Perfons to contribute their Affiftance to it. In Confequence of which, fince, by the Bleffing of God, and the generous Encouragement and Contributions of divers Perfons of all Qualities and Conditions, it has been carried on even beyond Expectation; it may be proper to give an authentick Account of the Proceedings of the Society, as well for the Satisfaction of the Subfcribers and Benefactors, as to animate others to join in the further promoting a Work fo truly redounding to the Honour of Almighty God, and the Relief and Comfort of the moft miferable and helplefs of our Fellow-Chriftians.

December 2, 1719, Several Gentlemen met to confider of the moft proper Methods for putting this Charitable Defign in Execution; at which Meeting it was thought neceffary, in the firft Place, that a Subfcription-Roll fhould be prepared: The Preamble whereof was accordingly drawn up, and agreed to, as follows:

.THE FOLLOWING

A B S T R A C T

C O N T A I N S A N.

A C C O U N T *of the* SICK-POOR

That have come under their Care

From 10 *February* 17⅑₀. to *Jan.* 3d, 17²¹₂₁ Inclufive.

This is how you begin to construct a database in Access 2007.

Open Access and you should get the opportunity to create a new database. This new file will hold *all* subsequent tables, queries, forms and so on. It is, essentially, a large electronic filing cabinet.
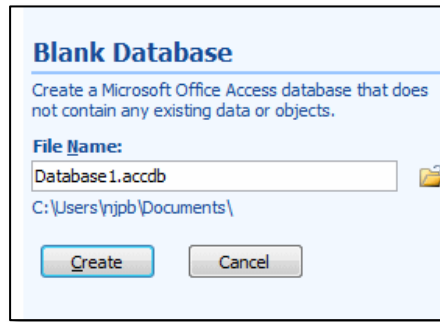
Create a database, naming it something sensible, such as *Infirmary*. This file will hold your tables and queries.

You should see something like this:
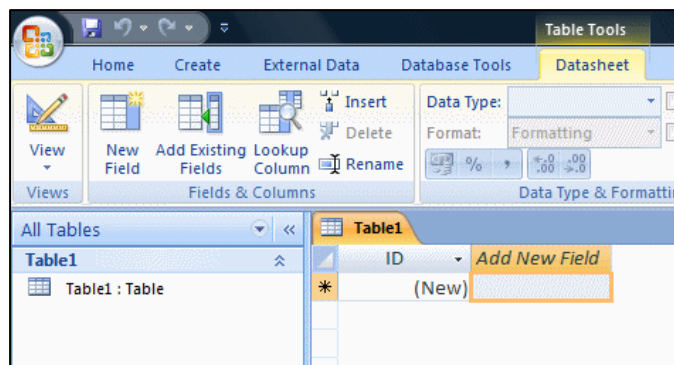
Now click on the View button and begin defining the fields in your table. Now you have to decide what types of fields you need, what sort of data they will contain, how many characters the text fields should contain and so on. As part of this course, it would be good if you enter all the data in the Infirmary records. We will be using the database you create in the next session. Don't forget to save your data to your file space. If you need to, copy it to a USB stick as well.

See you at the next session…

If you need any help or have any further queries, please feel free to contact me: Jeremy.Boulton@ncl.ac.uk