

Block 1 - From questionnaire to data file

1.2.3 First look at real data from a major survey

[Updated 20 September 2010]

Exemplar: British Social Attitudes survey¹ 1986

Below is a facsimile of part of a page of the 1986 questionnaire, dealing with household information.

- 43B -
SECTION NINE

			Col./Code	Skip to																																																							
113.a)	<p>ASK ALL</p> <p>Finally, a few questions about you and your household. Including yourself, how many people live here regularly as members of this household? <u>INTERVIEWER:</u> CHECK INTERVIEWER MANUAL FOR DEFINITION OF HOUSEHOLD IF NECESSARY.</p> <p style="text-align: right;">WRITE IN: <input style="width: 20px; height: 15px;" type="text"/> <input style="width: 20px; height: 15px;" type="text"/></p>		CARD 15	(1505-7)																																																							
b)	<p>And can I just check your own marital status? At present are you ... <u>READ OUT</u> ...</p> <p style="text-align: right;">... married, ... married, living as married, separated or divorced, widowed, or - not married?</p>		(1510)	1 2 3 4 5																																																							
114.	<p>Now I'd like to ask for a few details about each person in your household. Starting with yourself, what was your age last birthday? WORK DOWN COLUMNS OF GRID FOR EACH HOUSEHOLD MEMBER.</p> <table border="1" style="width: 100%; border-collapse: collapse; margin-top: 10px;"> <thead> <tr> <th style="width: 10%;"></th> <th style="width: 10%;">Resp-ondent</th> <th style="width: 10%;">2</th> <th style="width: 10%;">3</th> <th style="width: 10%;">4</th> <th style="width: 10%;">5</th> <th style="width: 10%;">6</th> <th style="width: 10%;">7</th> <th style="width: 10%;">8</th> <th style="width: 10%;">9</th> <th style="width: 10%;">10</th> </tr> </thead> <tbody> <tr> <td></td> <td>1511</td> <td>1515</td> <td>1520</td> <td>1525</td> <td>1530</td> <td>1535</td> <td>1540</td> <td>1545</td> <td>1550</td> <td>1555</td> </tr> <tr> <td>a) <u>Sex:</u></td> <td>Male</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> </tr> <tr> <td></td> <td>Female</td> <td>2</td> <td>2</td> <td>2</td> <td>2</td> <td>2</td> <td>2</td> <td>2</td> <td>2</td> <td>2</td> </tr> <tr> <td>b) <u>Age last birthday:</u></td> <td>1512-3</td> <td>1516-7</td> <td>1521-2</td> <td>1526-7</td> <td>1531-2</td> <td>1536-7</td> <td>1541-2</td> <td>1546-7</td> <td>1551-2</td> <td>1556-7</td> </tr> </tbody> </table>		Resp-ondent	2	3	4	5	6	7	8	9	10		1511	1515	1520	1525	1530	1535	1540	1545	1550	1555	a) <u>Sex:</u>	Male	1	1	1	1	1	1	1	1	1		Female	2	2	2	2	2	2	2	2	2	b) <u>Age last birthday:</u>	1512-3	1516-7	1521-2	1526-7	1531-2	1536-7	1541-2	1546-7	1551-2	1556-7			
	Resp-ondent	2	3	4	5	6	7	8	9	10																																																	
	1511	1515	1520	1525	1530	1535	1540	1545	1550	1555																																																	
a) <u>Sex:</u>	Male	1	1	1	1	1	1	1	1	1																																																	
	Female	2	2	2	2	2	2	2	2	2																																																	
b) <u>Age last birthday:</u>	1512-3	1516-7	1521-2	1526-7	1531-2	1536-7	1541-2	1546-7	1551-2	1556-7																																																	

For question 113a (Number of people in the household) . . .

- 43B -
SECTION NINE

			Col./Code
113.a)	<p>ASK ALL</p> <p>Finally, a few questions about you and your household. Including yourself, how many people live here regularly as members of this household? <u>INTERVIEWER:</u> CHECK INTERVIEWER MANUAL FOR DEFINITION OF HOUSEHOLD IF NECESSARY.</p> <p style="text-align: right;">WRITE IN: <input style="width: 20px; height: 15px;" type="text"/> <input style="width: 20px; height: 15px;" type="text"/></p>		CARD 15
			(1508-9)

. . . the interviewer will write the number (right justified or with leading zero) in the pair of boxes numbered (1508 - 9).

	WRITE IN: <input style="width: 20px; height: 15px;" type="text"/> <input style="width: 20px; height: 15px;" type="text"/>		(1508-9)
--	---	--	----------

¹ [British Social Attitudes](http://www.data-archive.ac.uk/findingData/snDescription.asp?sn=2315) has its own site. Data files and user documentation obtainable via the UK Data Archive at Essex University as SN 2315. See <http://www.data-archive.ac.uk/findingData/snDescription.asp?sn=2315> for details and conditions for access. The raw data file [bsa86.txt](#) for this exercise can also be downloaded from my site.

Question 113b (Marital status of respondent) is **precoded**

b)	And can I just check your own marital status? At present are you ... <u>READ OUT</u> married,	(1510)	1
		living as married,		2
		separated or divorced,		3
		widowed,		4
		or - not married?		5

The interviewer will circle the appropriate code (1 to 5) in the box numbered **(1510)** in the right hand column.

... married,	(1510)	1
living as married,		2
separated or divorced,		3
widowed,		4
or - not married?		5

Question 114 is a typical **grid** for collecting information about members of the respondent's household. The layout is slightly more complex, but quite straightforward.

114.	Now I'd like to ask for a few details about each person in your household. Starting with yourself, what was your age last birthday? WORK DOWN COLUMNS OF GRID FOR EACH HOUSEHOLD MEMBER.											
		Resp- ondent	2	3	4	5	6	7	8	9	10	
a)	<u>Sex:</u>	Male	1511	1515	1520	1525	1530	1535	1540	1545	1550	1555
		Female	2	2	2	2	2	2	2	2	2	2
b)	<u>Age last birthday:</u>	1512-3	1516-7	1521-2	1526-7	1531-2	1536-7	1541-2	1546-7	1551-2	1556-7	

Question 106a (Sex of respondent) is precoded and again the interviewer will circle the appropriate code (**1** for Male, **2** for Female) underneath **1511**.

	Resp- ondent
	1511
Male	1
Female	2

Finally, for question 106b (Age of respondent last birthday) the interviewer will write in the age of the respondent in the first pair of boxes on the line (numbered **1512 - 3**).

b)	<u>Age last birthday:</u>	1512-3	

The grid continues (not shown) with questions on the relationship to the respondent of each additional person in the household and whether that person has any legal responsibility for the accommodation, but for now this doesn't concern us.

If there are other people in the household, the interviewer will, for each additional person, circle the code for sex and write in the age in the appropriate pair(s) of boxes, starting with the second person and coding sex in the box numbered **1515** (at the top) and age in the boxes numbered **1516 - 7**.

2
1515
1
2
1516-7

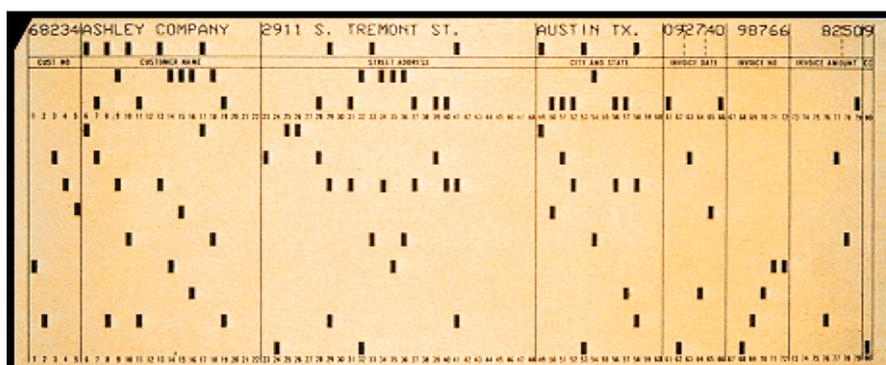
From questionnaire to raw data file

How do the responses to these questions get into the computer?

The data from this survey were originally keyed in as fixed format 80-column lines (card images), using the data layout template printed in the right hand margins of the questionnaire. This format derives from the days before visual display units (VDUs) and keyboards, when all data (and also programs) had to be punched on 80-column Hollerith cards and fed into the computer via a card reader. Each questionnaire takes up 23 lines of data and each line is numbered from 01 to 23.

80-column Hollerith card

From Computer Desktop Encyclopedia
© 2000 The Computer Language Co. Inc.



The 1986 British Social Attitudes survey consisted of two versions, A and B, with separate questionnaires and a separate self-completion questionnaire for each version. Each version carried a core of common questions plus a different set of questions on key topics which were split between versions. Altogether there were 3,100 respondents, of whom 1,542 were administered version A and 1,548 version B.

On the computer, when you open the raw data file, each record is displayed as a line on the screen. There are 23 lines of data for each respondent, but, apart from the serial number and record number, the lines for the data from versions A or B² will be blank depending on which version the respondent completed. In each of the 23 records for each of the 3,100 cases (i.e. every record in the entire data set) columns 1 to 5 contain a unique serial number and columns 6 and 7 the record number (from 01 to 23).

Have another look at the questionnaire extract.

In the top right hand corner you can see two boxes, one with **CARD 15** and the other with **1506 - 7**.

Col./ Code	Skip to
CARD 15	1506-7

This means that the data from questions on this page are to be punched on card 15 and that, for each respondent, the card number **15** will be punched as two digits in columns 6 and 7.

The data for Q.113a (number of people in the household) will be punched (right-justified) in record 15 columns 8 - 9 as indicated by **(1508-9)** in the margin.

WRITE IN:			(1508-9)
-----------	--	--	----------

² The code for which version was administered (A = 1, B = 2) is in column 8 of record 2 (viz, **208**)

For question 113b (marital status), you can see the **(1510)** in smaller type in the margin above the response codes 1 - 5.

	(1510)
... married,	1
living as married,	2
separated or divorced,	3
widowed,	4
or - not married?	5

For question 114a (sex of respondent) there you can see **1511** above the codes 1 and 2 in the margin. . .

	Resp- ondent
	1511
Male	1
Female	2

...and further down, for question 114b (age of the respondent last birthday) **1512 - 3** above a pair of boxes where the age will be written in by the interviewer.

		1512-3
b)	Age last birthday:	

This means the code for **marital status** is to be punched in record 15 column 10, the code for **sex** in record 15 column 11 and **age** of the respondent in record 15 columns 12 - 13.

The 3- and 4-digit numbers in the margin are a **template** for where the data are to be punched. The first pair of digits indicates the **record** (card) number and the second pair the **field** of one or more columns numbered from 01 to 80 where the data each of the above questions will be punched. In this survey there are 23 data cards (records) for each respondent. Each card has 80 columns.

It is because the data layout template is printed on the questionnaire that I prefer to use **positional**³ rather than **mnemonic** variable names for most variables, as the questionnaire can then serve as user documentation. This survey has a full user-manual⁴, but for many surveys the questionnaire is the only documentation available.

The raw data for the 1986 British Social Attitudes survey were initially supplied by the UK Data Archive in a *.dat file which, on my previous computer, Windows interpreted as a Wordperfect file. On this machine Windows thinks it's a movie!

Wordperfect displayed the data in **Times New Roman proportional** font which is not only horrible to look at, but also, because the columns are not vertically aligned, is impossible to interpret visually.

³ See 1.3.1 [Conventions for Naming Variables in SPSS](#)

⁴ <http://www.esds.ac.uk/doc/2315/mrdoc/pdf/2315userguide1.pdf>

Here are the actual data⁵ for the first respondent.

```
102030101011072187061 020311 12
102030222 1 018111188821213 18111112112211213453311 10
1020303
1020304
1020305203 2 443 090414112322222332 31811123 233220401111111
102030603 8813112311 1332111100601 003120 3222222 21212 5
1020307 22222888121776 2112112122451 23636666343111131180107030411
10203081 1 38 8 1 2 2 1 6 4 2
1020309
1020310
1020311
1020312
1020313
1020314 01 2 2
10203150142801 2 032
1020316 999021117998999 99999 933
1020317 03031 5 03 11 11290601870300486231
1020318
1020319
102032044414114132244441352222215333333311111111331232522121212222211211212111
10203211133334341422324424418881111812222323111111322212225542245415511
1020322151122421124422113
1020323 01.0000 15 7 1 02 2 1 01 1 08 9 9 9 9 99 3 7 1 1 3 1
```

SPSS will read the data in this **proportional** font if the file name is enclosed in primes (apostrophes) i.e. **'bsa86.dat'**, but it is always useful to be able to inspect the original data visually.

For tutorial purposes the data need to be displayed in a **fixed-width** font to make it easier to find our way around inside the file. We shall be using these data in later exercises and there's a little test at the end to see if you've understood how the data relate to the questionnaire.

The original data file was supplied as a WordPerfect file **GB8601.DAT**, but I have converted it from **Times New Roman proportional** to **Courier New fixed-width** font in ***.txt** format in file **bsa86.txt**⁶.

⁵ You can download the original raw data file [bsa86.txt](#) (3.4 mb) from this site.

⁶ WordPerfect is not installed on my current machine and Windows thinks **bsa.dat** is a movie!!

Here are the reformatted data⁷ for the first respondent.

```
102030101011072187061      020311      12
102030222  1 018111188821213  18111112112211213453311  10
1020303
1020304
1020305203      2      443      090414112322222332  31811123      233220401111111
102030603      8813112311  1332111100601      003120  3222222      21212      5
1020307      222222888121776  2112112122451      236366666343111131180107030411
10203081  1 38      8      1 2      2      1      6      4 2
1020309
1020310
1020311
1020312
1020313
1020314      01  2      2
10203150142801      2  032
1020316      999021117998999  99999  933
1020317      03031  5      03  11 11290601870300486231
1020318
1020319
102032044414114132244441352222215333333311111111331232522121212222211211212111
10203211133334341422324424418881111812222323111111322212225542245415511
1020322151122421124422113
1020323  01.0000  15 7 1 02 2 1 01 1 08      9  9  9  9  99      3 7 1 1 3 1
```

This is a conversion of the original data from **Times New Roman variable-width** font to plain text in **Courier New fixed-width** font. The data are displayed in 80-column lines in which all 80 columns are vertically aligned. The 1986 British Social Attitudes survey has 3,100 cases and 23 lines (records) per case (that's a lot of lines!).

Question: What are the sex and age of this respondent?

Try to answer this yourself before reading the next page.

⁷ You can download the original raw data file [bsa86.txt](#) (3.4 mb) from this site.

The record number is in columns 6 – 7, highlighted in pink. Count down to find record 15 (highlighted in pale blue) then count along to find 2 in column 11 (sex) and 80 in columns 12-13 (age).

```

102030101011072187061          020311    12
102030222  1 018111188821213  18111112112211213453311  10
1020303
1020304
1020305203          2    443    090414112322222332  31811123    233220401111111
102030603    8813112311  1332111100601    003120  3222222    21212    5
1020307          22222888121776  2112112122451    23636666343111131180107030411
10203081 1 38          8          1 2          2          1    6          4 2
1020309
1020310
1020311
1020312
1020313
1020314          01  2          2
10203150142801          2  032
1020316          999021117998999  99999  933
1020317          03031    5          03  11  11290601870300486231
1020318
1020319
102032044414114132244441352222215333333311111111331232522121212222211211212111
10203211133334341422324424418881111812222323111111322212225542245415511
1020322151122421124422113
1020323  01.0000  15 7 1 02 2 1 01 1 08    9    9    9    9    99    3 7 1 1 3 1

```

This respondent is therefore a woman aged 80. What is her marital status? How many people live in the accommodation? Which version of the questionnaire did she complete? (See the footnote on page 3.) Find the answers yourself!!

If you want to make sure you understand the relationship between the questionnaire and the raw data, there's a repeat session on the 1989 survey which also runs through the conversion procedure from Times New Roman proportional font to Courier New fixed-width font.

End of session

Now go to the repeat session using data for the 1989 survey:

1.2.4 Second look at data from a major survey

Next section: [1.3 Reading raw data into SPSS](#)

Next step: [1.3.3.1 Preparing the ground](#)

[\[Back to Block 1 menu\]](#)