British Social Attitudes 1983 to 2014: Constructing cumulative files

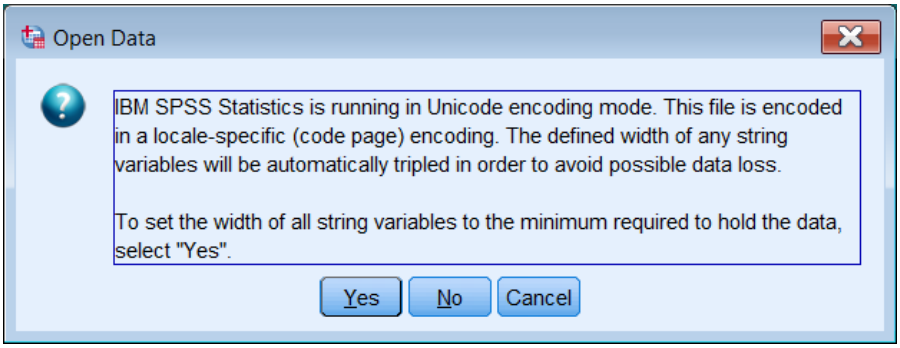
 John F Hall [Draft only: 5 April 2016]

These notes refer to downloads from the UK Data Service (UKDS) of the SPSS files for the British Social Attitudes surveys. Files for all waves 1983-2014, plus two cumulative sets and one panel data set, were downloaded *dans l'état* from UKDS.

Index to UKDS downloads for British Social Attitudes 1983 - 2014 is an Excel file detailing, for each wave, **year** of survey, **link** to UKDS, download **filename**, **size** of file, number of **cases**, number of **variables**, number of variables with **non-numeric** formats and the **new filename** assigned to amended files. The amended SPSS *.sav files will be sent to Natcen for approval and possible deposit with UKDS. . Not all files include year, month or date of interview.

1983 – 1992 Variables are all Numeric
1993 – 2010 Some variables are Strings, some in different formats for the same variable.
2011 – 2014 Variables are all Numeric

For 1993 and some other files, the following message appears when opening a file:



Clicking on Yes sets all Strings to AMIN (minimum width needed to accommodate a particular string) but the same variable names may have different widths in different surveys. When using **ADD FILES** to merge waves, any such Inconsistency will generate a warning message and the command will not be executed. The width specification will need to be changed to a single common value when merging files. This has already been done for 1983-1994 and 2011-2014: see Non-numeric variables in British Social Attitudes

Some strings seem to be (? BLAISE ?) route-ing/coding instructions that have crept in as labels.

The purpose of the following exercise is to check the metadata for all variables, in particular specifications of **missing values** and **measurement levels**.

Missing values are almost always specified as **Lo thru -1** but values ending in **7**, **8** and **9** which should be treated as missing are still there. This leads to distorted tables and errors in statistical calculations. See: Exploring British Social Attitudes 2014 and also Cautionary tale for the unwary which specifically demonstrates the importance of checking data before generating attitude scores.

Many measurement levels are incorrectly specified, possibly because of automatic archiving software, but I need to check with Natcen to see what they actually sent to UKDS.

Here's what I have done so far. Some screenshots include clips of the full text of the SPSS **Syntax Editor** and therefore SPSS colour coding of commands, subcommands and operators. However I work almost exclusively syntax, using **lower case** text and **abbreviated** commands and operators: in cases where I have written syntax in the text, I have colour coded these abbreviations to match.

Step by step:

1983 – 1991 waves

```
add files
    file  'C:\Users\John\Desktop\SPSS files\bsa1983.sav'
   /file 'C:\Users\John\Desktop\SPSS files\bsa1984.sav'
  /file 'C:\Users\John\Desktop\SPSS files\bsa1985.sav'
  /file 'C:\Users\John\Desktop\SPSS files\bsa1986.sav'
  /file 'C:\Users\John\Desktop\SPSS files\bsa1987.sav'
  /file 'C:\Users\John\Desktop\SPSS files\bsa1989.sav'
  /file 'C:\Users\John\Desktop\SPSS files\bsa1990.sav'
  /file 'C:\Users\John\Desktop\SPSS files\bsa1991.sav' .
```

As a quick check:

```
freq rsex.
```

		Frequency	Percent
Valid	1 MALE	9124	45.8
	2 FEMALE	10807	54.2
	Total	19931	100.0

Years 1992 – 1994 are almost compatible with 1983 -1991.

Combined in file bsa1983-1994.sav

Years 1995 – 2008 contain numerous inconsistencies which will take some time to sort out.

Years 2009 and 2010 are, with one small change to one variable, compatible with 2011-2014 and have been combined.

Years 2011 to 2014

No conflicts
No String variables
Years 2009 to 2014 are combined in file **bsa2009-2014.sav**

Cumulative data 1983 – 1989

Not available in SPSS format

Panel survey

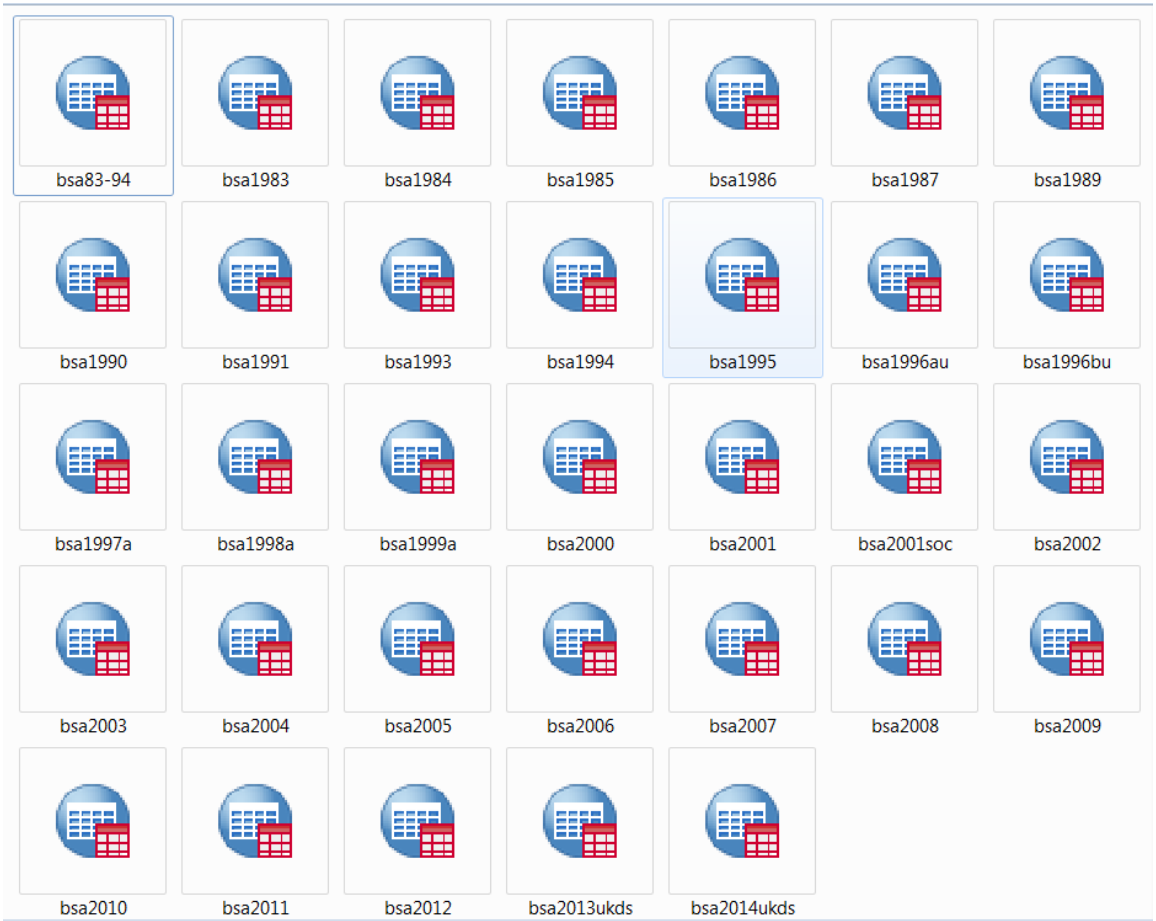
All vars NUMERIC
No missing values
Labels all in UPPER CASE
Variable **[Date]** is N6

Progress to date:

Starting on 17 March 2016, I have worked systematically through the data for each wave looking for Strings and changing them to a75. Didn't find them all as there were still some incompatible formats being thrown up.

Computed a new variable **[year]** as **f4.0** for each file (or changed format to **f4.0** if already present in another format).

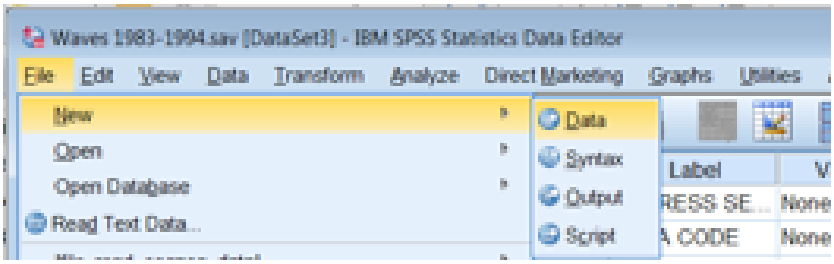
1983 to 1994 waves now in combined file **bsa83-94.sav** in folder **UKDS BSA files** which contains the following files exactly as downloaded from UKDS, but with new variable **[year]** added:



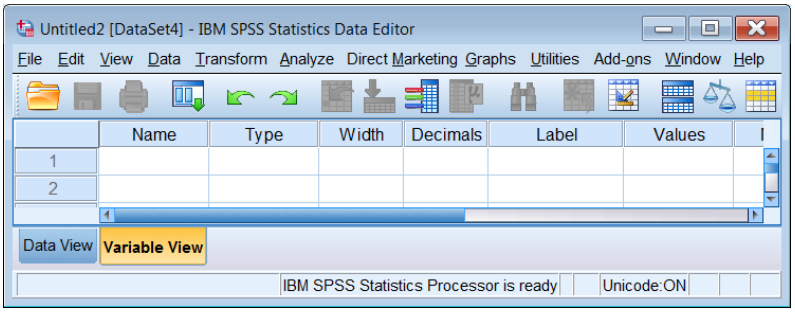
Still need to check with Natcen if **1996au** and **1986bu** should be combined, also **bsa2001** and **bsa2001soc**.

Combining files for 1983-1994

File >> New >> Data



. . to open new **Data Editor** Untitled



Run **ADD FILES** to combine data

```
* Encoding: UTF-8.
*1983-1994.
add files
 file  'C:\Users\John\Desktop\UKDS BSA files\bsa1983.sav'
 /file 'C:\Users\John\Desktop\UKDS BSA files\bsa1984.sav'
/file 'C:\Users\John\Desktop\UKDS BSA files\bsa1985.sav'
/file 'C:\Users\John\Desktop\UKDS BSA files\bsa1986.sav'
/file 'C:\Users\John\Desktop\UKDS BSA files\bsa1987.sav'
/file 'C:\Users\John\Desktop\UKDS BSA files\bsa1989.sav'
/file 'C:\Users\John\Desktop\UKDS BSA files\bsa1990.sav'
/file 'C:\Users\John\Desktop\UKDS BSA files\bsa1991.sav'
/file 'C:\Users\John\Desktop\UKDS BSA files\bsa1993.sav'
/file 'C:\Users\John\Desktop\UKDS BSA files\bsa1994.sav'.

freq year.
```

Check: **cro** year by rsex /cel cou.

year Year of Interview * rsex Q91A RESPONDENTS SEX Crosstabulation

Count

		rsex Q91A RESPONDENTS SEX		
		1 MALE	2 FEMALE	Total
year Year of Interview	1983	807	954	1761
	1984	780	895	1675
	1985	822	982	1804
	1986	1445	1655	3100
	1987	1326	1521	2847
	1989	1396	1633	3029
	1990	1256	1541	2797
	1991	1292	1626	2918
	1993	1227	1718	2945
	1994	1507	1962	3469
Total		11858	14487	26345

File>> Save As .. (write in filename) **bsa83-94.sav**

If SPSS is set up to include syntax in output, the **Viewer** file will contain:

```
SAVE OUTFILE='C:\Users\John\Desktop\UKDS BSA files\bsa83-94.sav'
/COMPRESSED.
```

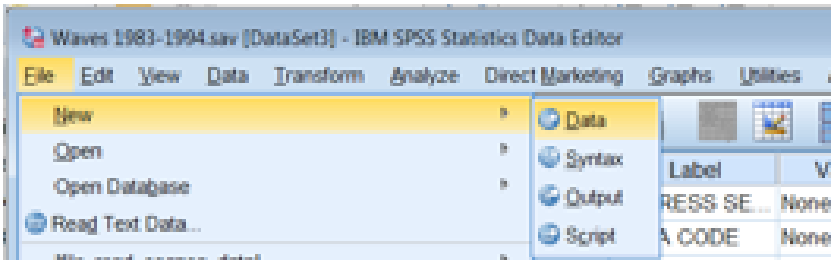
[Note that the pathname is enclosed in primes.]

Years 2011 to 2014

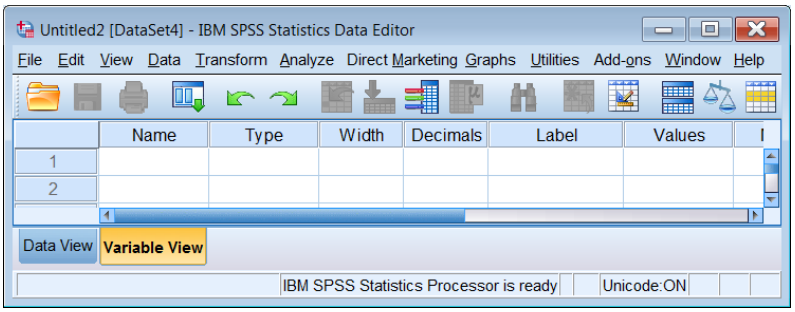
No conflicts
No String variables

Open new **Data Editor** Untitled

File >> New >> Data



. . to open new **Data Editor** Untitled



Run **ADD FILES** on 2011 – 2014 data

```
*2011-2014.
add files
    file *
        /file 'C:\Users\John\Desktop\UKDS BSA files\bsa2011.sav'
        /file 'C:\Users\John\Desktop\UKDS BSA files\bsa2012.sav'
        /file 'C:\Users\John\Desktop\UKDS BSA files\bsa2013ukds.sav'
        /file 'C:\Users\John\Desktop\UKDS BSA files\bsa2014ukds.sav'.
```

cro year by rsex /cel cou..

year Year of survey * Rsex Person 1 SEX Crosstabulation

Count

		Rsex Person 1 SEX		
		1 Male	2 Female	Total
year Year of survey	2011	1450	1861	3311
	2012	1444	1804	3248
	2013	1452	1792	3244
	2014	1255	1623	2878
Total		5601	7080	12681

File>> Save As .. (write in filename) **bsa2011-14.sav**

```
SAVE OUTFILE='C:\Users\John\Desktop\UKDS BSA files\ 'bsa2011-14.sav'
/COMPRESSED.
```

A useful check on file contents is **DISPLAY DICTIONARY**

disp dic / var year libauth leftrigh welfare2.

Variable Information

Variable	Position	Label	Level	Width	Format	Missing Values
year	1	<none>	Scale	10	F4	
libauth	788	Libertarian-authoritarian scale (TradVals to censor) dv	Scale	9	F6.4	-9.0000 through -1.0000
leftrigh	787	Left-right scale(redistrb to indust4) dv	Scale	10	F6.4	-9.0000 through -1.0000
welfare2	2	Welfarism scale (welfhelp to proudwlf) dv	Scale	10	F6.4	-9.0000 through -1.0000

(*Table modified to remove superfluous information*)

(**NB:** From the GUI you can use File > Display Data File Information > **Working File** or **External File** but this produces a table for the entire file, which contains 2588 variables!)

[welfare2] appears in position 2, but **[libauth]** and **[leftrigh**] are in positions 788 and 787: missing values for the scales are declared as -9.0000 through -1.0000. This is because, when merging files, SPSS takes metadata for variables from the first file read (2011) but it is quite easy to re-order the variables and put **[welfare2]** in another place. If the files are combined in reverse year order, this problem disappears, but the positions and missing values are different:

Variable	Position	Label	Level	Width	Format	Missing Values
year	650	<none>	Scale	10	F4	
libauth	637	Libertarian-authoritarian scale (TradVals to censor) dv	Scale	9	F6.4	Lowest through -1.0000
leftrigh	636	Left-right scale(redistrb to indust4) dv	Scale	10	F6.4	Lowest through -1.0000
welfare2	638	Welfarism scale (welfhelp to proudwlf) dv	Scale	10	F6.4	Lowest through -1.0000

With small changes to one variable, **[sttim]** *(Start time HH:MM:SS :Q341)* the data for the 2009 and 2010 waves can be combined with data for the 2011 to 2014 waves. The format was a50 in 2009 and time10 in 2010: both have been changed to time8. Data for 2009 and 2010 were then added to the cumulative file using **ADD FILES** with the years in reverse order from 2014 to 2009 and the file saved as **bsa2009-14.sav**

Check: **freq** year.

year

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2009	3421	17.6	17.6	17.6
	2010	3297	17.0	17.0	34.6
	2011	3311	17.1	17.1	51.7
	2012	3248	16.7	16.7	68.4
	2013	3244	16.7	16.7	85.2
	2014	2878	14.8	14.8	100.0
	Total	19399	100.0	100.0	

File>> Save As .. (write in filename) **bsa2011-14.sav**

SPSS **Viewer** file will contain:

```
SAVE OUTFILE='C:\Users\John\Desktop\UKDS BSA files\ 'bsa2011-14.sav'
/COMPRESSED.
```

However, a serious problem remains concerning missing values (see: Cautionary tale for the unwary).and my note: *British Social Attitudes 2009 to 2014: Missing values*

From this point there will be a number of complex and detailed changes: it is best to **make a copy** of the file and work on that.

I usually append my intials **jfh** to the new filename and at each stage, to make it easier to track changes, I append an edition number to the filename (eg **bsa09-14jfh_1.sav, bsa09-14jfh_2.sav** etc. and give associated *.sps setup files the same root.)

File>> Save As .. (write in filename) **bsa09-14jfh_1.sav**

SPSS **Viewer** file will contain:

```
SAVE OUTFILE='C:\Users\John\Desktop\UKDS BSA files\ 'bsa09-14jfh_1.sav'
/COMPRESSED.
```

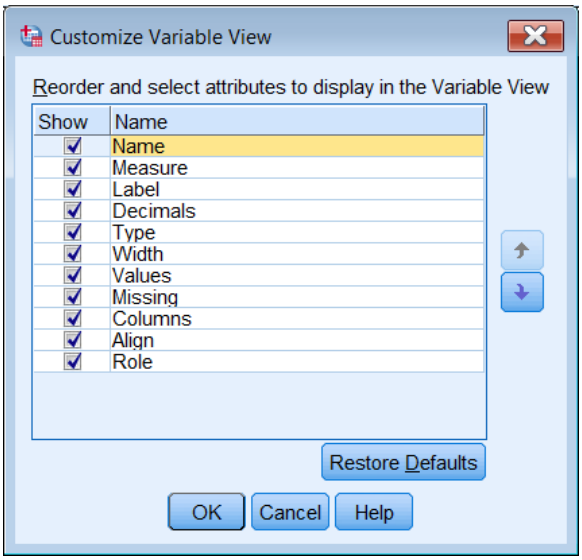
File **bsa09-14jfh_1.sav** is now my working copy: it will be saved many times (and accrue many edition numbers) before it is in a fit state to release as a "mother" file.

To make life easier working in Data View, I always change the priority order of variable properties to my preferred sequence:

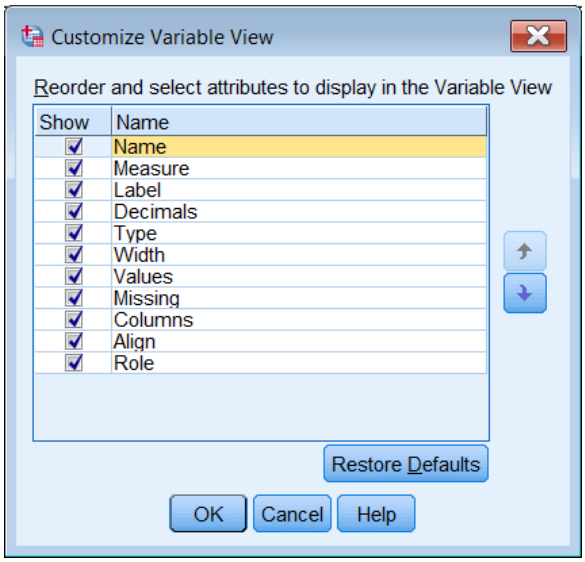
Name	Measure	Label	Values	Missing	Decimals

This can be done by clicking once on a column header to highlight the whole column, then click again and, holding the left mouse down, dragging it to the required position. It can also be done from the GUI with View > Customize Variable View then using the blue arrows to move (or uncheck) attributes.

Before



After



I have also moved variable **[year]** to position 2 in the file. (Note also that you can drag the column edges out to see more of the labels.)

	Name	Measure	Label	Values	Missing
1	SSerial	Scale	External Serial Number	None	LO - -1
2	year	Scale	Year of survey	None	None

There are separate notes on **Measurement Levels**, **Missing Values**, **Multiple Response** etc