2.1.2.6 Checking your file contents

[30 November 2010]

Previous session: 2.1.2.5 Extending your data dictionary

Exemplar: British Social Attitudes (1986 survey)

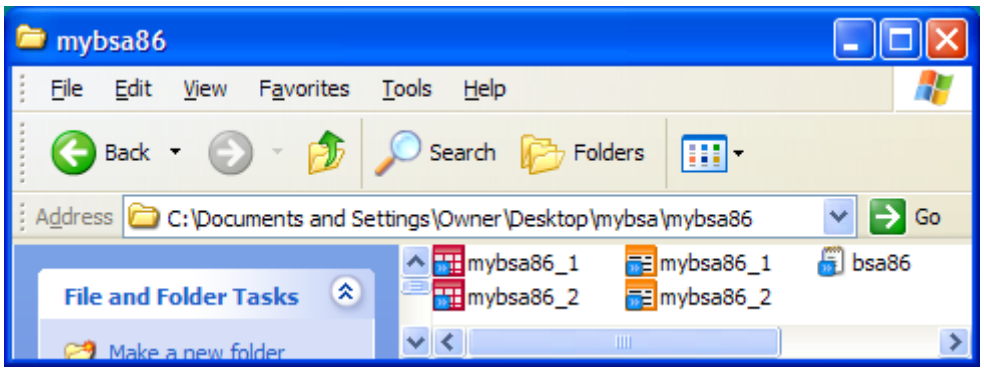
File: mybsa86_2.sav (British Social Attitudes 1986, 2nd saved file)

Task: Check contents of file for accuracy and completeness.

SPSS syntax used: DISPLAY[1]

SPSS drop-down menus: Utilities > Variables
Data > Define Variable Properties
Analyze > Reports > Codebook

Go to your folder **mybsa** and double-click on **mybsa86_2.sav** (the file you saved last session).

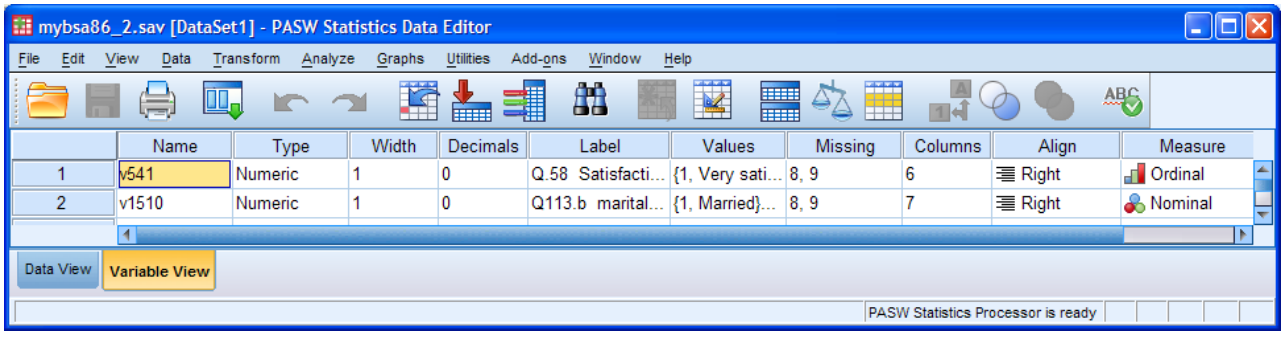


*[If you don't have folder **mybsa** or file **mybsa86_2.sav**, go back and do the exercises in 2.1.2.3, 2.1.2.4 and 2.1.2.5.]*

The output file will display the syntax automatically generated by SPSS:

```
GET
  FILE='C:\Documents and Settings\Owner\Desktop\mybsa\mybsa86_2.sav'.
DATASET NAME DataSet1 WINDOW=FRONT.
```

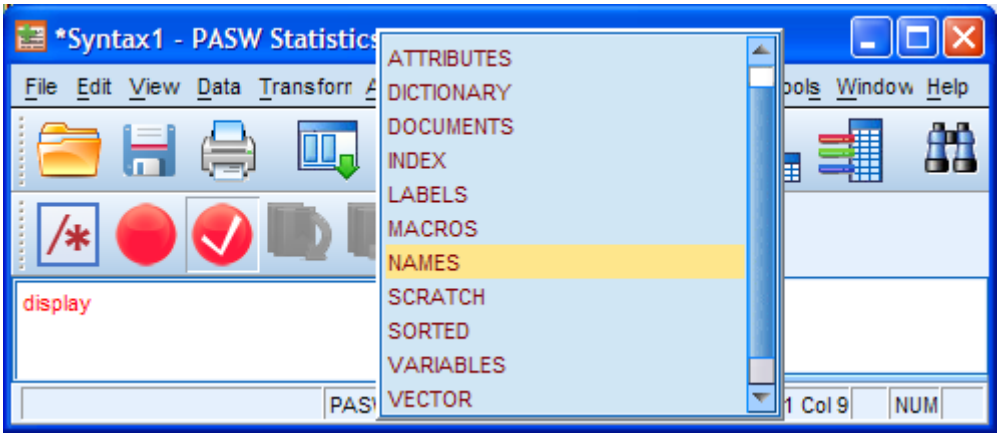
. . . the data editor will open:



[1] **DISPLAY** <options>

DISPLAY

As a check on file contents you can use the **DISPLAY** command *(not available in the drop-down menus)*. This command has several keywords, but for now we shall use only four of these:



The simplest displays just the variable **names** (default):

display . displays only the **names**

Variable Names

v541
v1510

Currently defined variables

display index . displays the **names** and **positions** in the file

Variable Indices

Variable	Index
v541	1
v1510	2

Variables in the working file

display labels . displays the **names**, file **positions** and **labels** for each variable in the file, but the output is a bit squashed:

Variable Labels

Variable	Position	Label
v541	1	Q.58 Satisfaction with running of NHS
v1510	2	Q113.b marital status of respondent

Variables in the working file

You can drag the right edge outwards after double-clicking the table in the output file and adjusting the pivot table, or after copying it to Word.

Either way it looks neater like this:

Variable Labels

Variable	Position	Label
v541	1	Q.58 Satisfaction with running of NHS
v1510	2	Q113.b marital status of respondent

Variables in the working file

display variables . displays variable labels plus technical information:

. . which gives another squashed up table with more information than you really need at this stage:

Variable Information

Variable	Position	Label	Measurement Level	Role	Print Format	Write Format	Missing Values
v541	1	Q.58 Satisfaction with running of NHS	Ordinal	Input	F1	F1	8, 9
v1510	2	Q113.b marital status of respondent	Nominal	Input	F1	F1	8, 9

Variables in the working file

Again you can adjust the columns manually (working from right to left) to get:

Variable Information

Variable	Position	Label	Measurement Level	Role	Print Format	Write Format	Missing Values
v541	1	Q.58 Satisfaction with running of NHS	Ordinal	Input	F1	F1	8, 9
v1510	2	Q113.b marital status of respondent	Nominal	Input	F1	F1	8, 9

Variables in the working file

display dictionary . gives you the works [already edited to squeeze it on the page]

Variable Information

Variable	Position	Label	Measurement Level	Role	Column Width	Alignment	Print Format	Write Format	Missing Values
v541	1	Q.58 Satisfaction with running of NHS	Ordinal	Input	6	Right	F1	F1	8, 9
v1510	2	Q113.b marital status of respondent	Nominal	Input	7	Right	F1	F1	8, 9

Variables in the working file

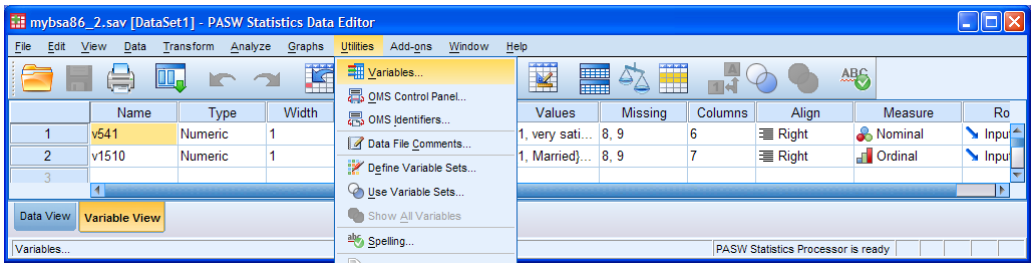
. . . including the value labels:

Variable Values

Value		Label
v541	1	Very satisfied
	2	Quite satisfied
	3	Neither satisfied nor dissatisfied
	4	Quite dissatisfied
	5	Very dissatisfied
v1510	1	Married
	2	Living as married
	3	Separated or divorced
	4	Widowed
	5	Not married

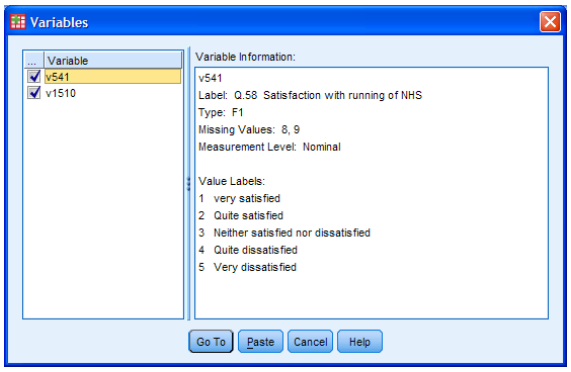
Checking file contents with the drop-down menus

You can also get immediate displays of file information from the drop-down menus in the GUI by clicking on Utilities > Variables

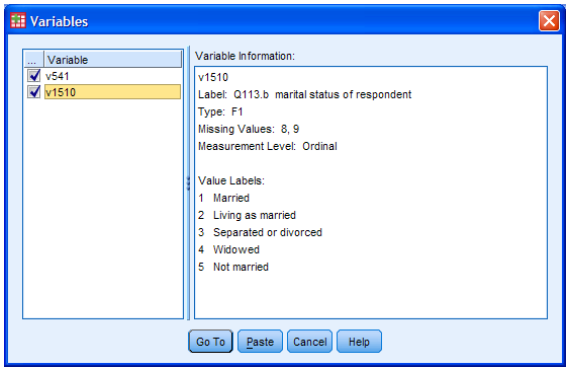


. . . but only one variable at a time, starting with the first variable in the file:

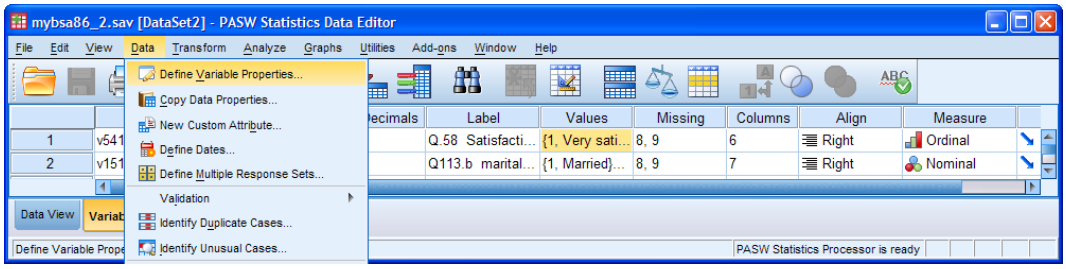
v541



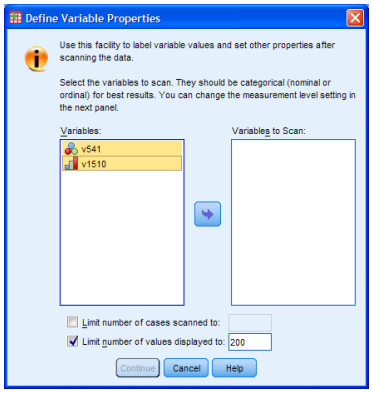
v1510

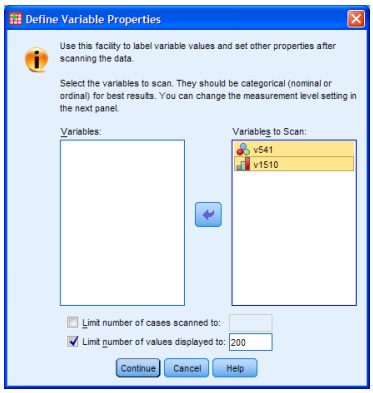


A particularly useful menu, intended not as a data check, but as a means of entering variable properties (measurement level, type, labels, formats etc.) is available from the GUI by clicking on Data > Define Variable Properties

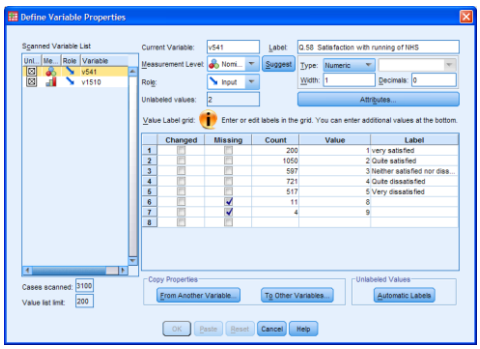


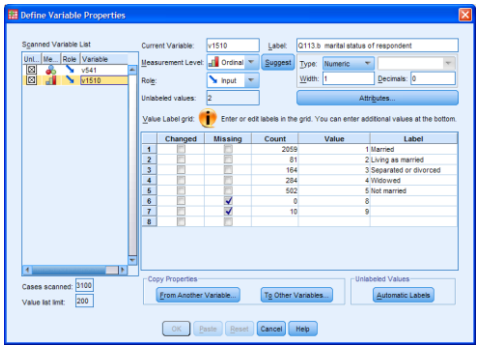
Highlight both variables . . and click on the  to drag them across:





Click on Continue . . . and click on **v1510** to get:

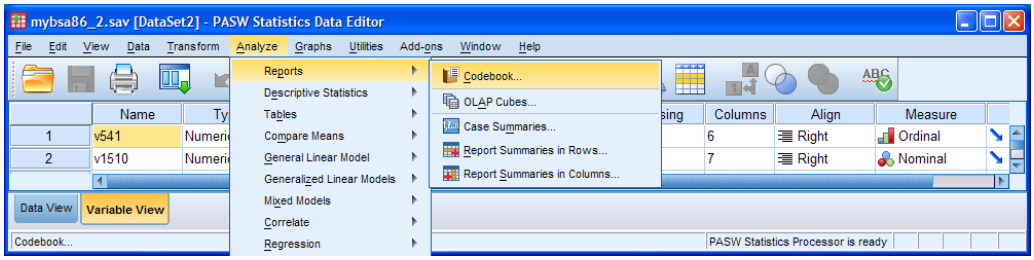




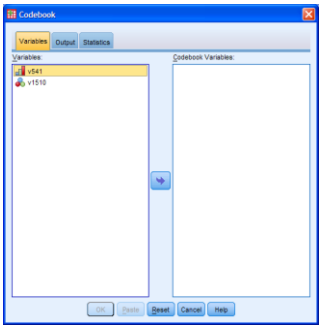
This gives all the information at a glance, including a frequency count for each value: possibly not quite what the designers intended, but very useful all the same!

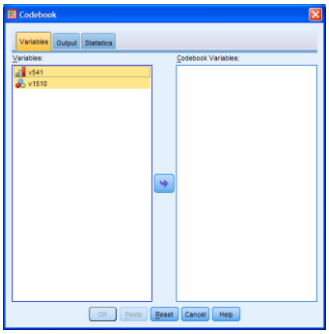
Finally, you can produce a full **CODEBOOK** for the file, but the syntax for this can be quite complex, as we shall see. For once it's simpler to use the drop-down menus.

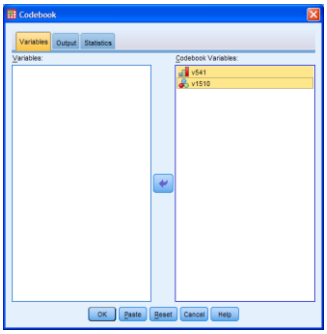
Click on Analyze > Reports > Codebook



Highlight both variables and click on the blue arrow to drag them across:







SPSS automatically generates the following syntax: now you know why I used the drop-down menu instead of writing out the syntax!

```
CODEBOOK  v541 [o] v1510 [n]
  /VARINFO POSITION LABEL TYPE FORMAT MEASURE ROLE VALUELABELS MISSING
ATTRIBUTES
  /OPTIONS VARORDER=VARLIST SORT=ASCENDING MAXCATS=200
  /STATISTICS COUNT PERCENT MEAN STDDEV QUARTILES.
```

[Try experimenting with the syntax above by changing the order, or reducing the number, of elements requested after the `/VARINFO` *subcommand.]*

This will produce full dictionary information for each variable, together with percentages and frequency counts. Again the tables will be a bit squashed, so in the tables below I've adjusted the columns, reduced the font size to 8-point and changed the text to single space

v541

		Value	Count	Percent
Standard Attributes	Position	1		
	Label	Q.58 Satisfaction with running of NHS		
	Type	Numeric		
	Format	F1		
	Measurement	Ordinal		
	Role	Input		
Valid Values	1	Very satisfied	200	6.5%
	2	Quite satisfied	1050	33.9%
	3	Neither satisfied nor dissatisfied	597	19.3%
	4	Quite dissatisfied	721	23.3%
	5	Very dissatisfied	517	16.7%
Missing Values	8		11	.4%
	9		4	.1%

v1510

		Value	Count	Percent
Standard Attributes	Position	2		
	Label	Q113.b marital status of respondent		
	Type	Numeric		
	Format	F1		
	Measurement	Nominal		
	Role	Input		
Valid Values	1	Married	2059	66.4%
	2	Living as married	81	2.6%
	3	Separated or divorced	164	5.3%
	4	Widowed	284	9.2%
	5	Not married	502	16.2%
Missing Values	9		10	.3%

These tables are the default for **CODEBOOK**, but by clicking on the Output tab you can select the elements to display by checking or unchecking the boxes in the left pane: in the same window you can also check boxes in the right pane to obtain information about the file. Play around with this menu and see what you get.

This session was to emphasise the important practice of checking your files for completeness and accuracy before attempting to do any serious analysis.

End of session:

Next session: 2.1.2.7 Frequencies for nominal and ordinal variables

[Back to Block 2 menu]