

Block 2: Analysing one variable

2.3 Data transformations

2.3.1.2 Exercise to rename variables

[Draft only: 5 October 2011]

Previous session: [2.3.1.1 Data transformations](#)

Exemplar: British Social Attitudes 1989

File: **mybsa89\_3.sav**

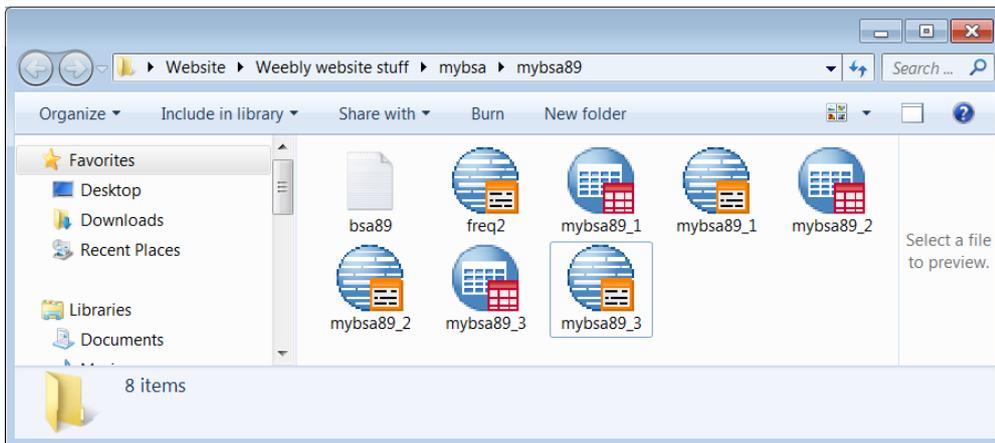
SPSS commands: **RENAME VARIABLES**  
**RECODE**

Task 1: Rename demographic variables

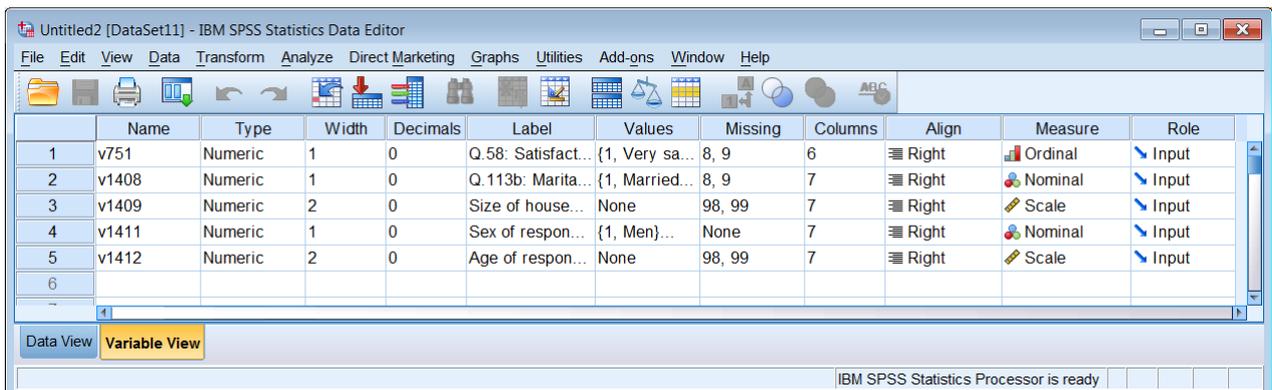
Old name	New name
v1408	marital
v1411	sex
v1412	age

Task 2: Create a new variable by recoding **age** into **agegroup** with four categories: 18 -29, 30 -44, 45 -59, 60 and over.

Go to folder **mybsa89**



Open file **mybsa89\_3.sav**



**mybsa89\_3.sav** (imported as Untitled)

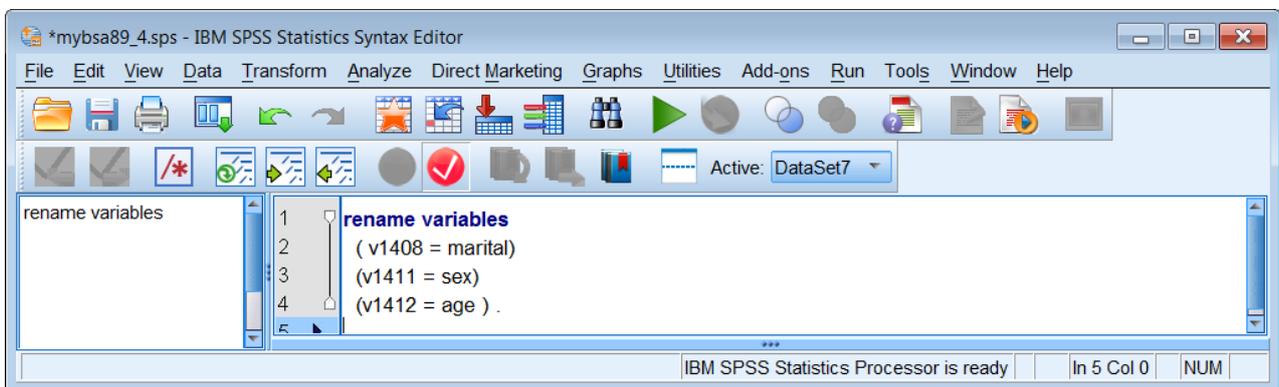
Despite my predilection for positional variable names, certain demographic variables used repeatedly in data analysis merit mnemonic names. Thus **sex**, **marital** and **age** are much more convenient to remember than **v1411**, **v1408** and **v1412**. Also variables with many values, such as **age**, need to be grouped into fewer categories for tabulation purposes. Standard practice is to keep demographic variables together in a block, often at or near the end of the file, so that they can be used with the **TO** keyword, eg **sex to agegroup**.

**Task 1:** Rename demographic variables

```
rename variables
(v1408 = marital)
(v1411 = sex)
(v1412 = age) .
```

This command can also be written:

```
rename variables
(v1408 v1411 v1412 = marital sex age) .
```



Run the job to get:

The screenshot shows the IBM SPSS Statistics Data Editor window. The title bar reads '\*mybsa89\_3.sav [DataSet7] - IBM SPSS Statistics Data Editor'. The menu bar includes File, Edit, View, Data, Transform, Analyze, Direct Marketing, Graphs, Utilities, Add-ons, Window, and Help. The toolbar contains various icons for data manipulation. The main area displays a table of variables in Variable View:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	v751	Numeric	1	0	Q.58: Satisfact... {1, Very sa...	8, 9		6	Right	Ordinal	Input
2	marital	Numeric	1	0	Q.113b: Marita... {1, Married...	8, 9		7	Right	Nominal	Input
3	v1409	Numeric	2	0	Size of house...	None 98, 99		7	Right	Scale	Input
4	sex	Numeric	1	0	Sex of respon... {1, Men}...	None		7	Right	Nominal	Input
5	age	Numeric	2	0	Age of respon... None	98, 99		7	Right	Scale	Input
6											

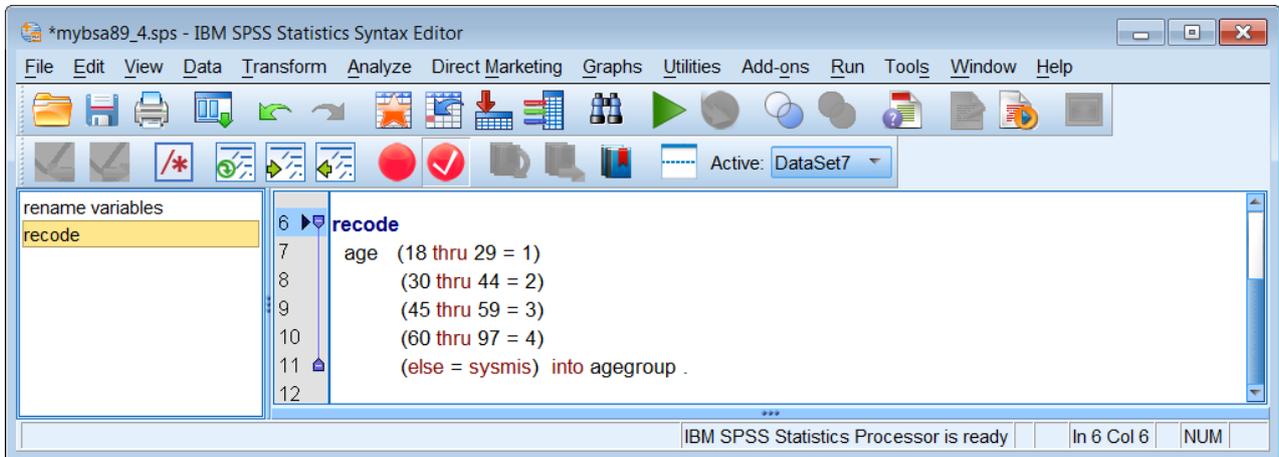
The status bar at the bottom indicates 'IBM SPSS Statistics Processor is ready'.

\*mybsa89\_3.sav after renaming variables

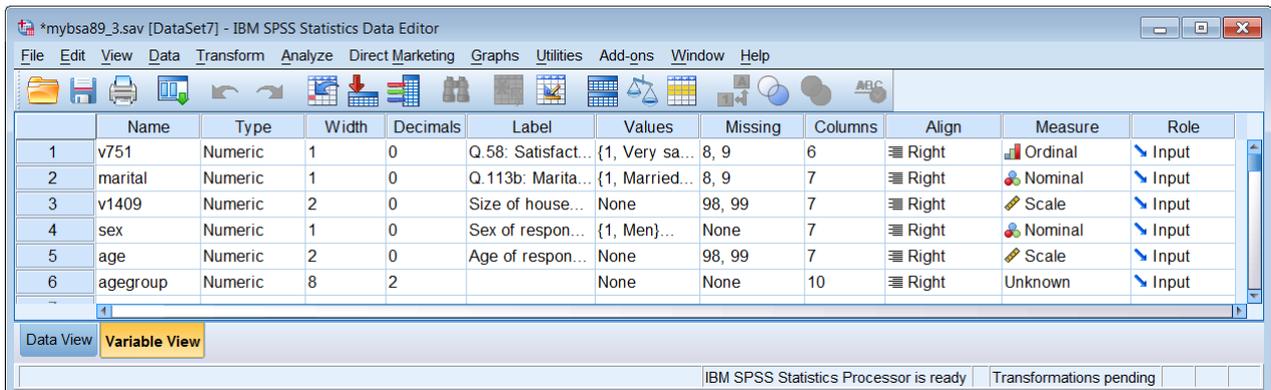
**Task 2:** Create a new variable **agegroup** by recoding **age** into four groups, 18 -29, 30 -44, 45 -59, 60 and over.

**recode**

**age** (18 thru 29 = 1)  
 (30 thru 44 = 2)  
 (45 thru 59 = 3)  
 (60 thru 97 = 4)  
 (else = sysmis) into agegroup .



Run the job to get:



\*mybsa89\_3.sav after recoding age into agegroup

[NB: new variable **agegroup** has no labels or missing values yet]

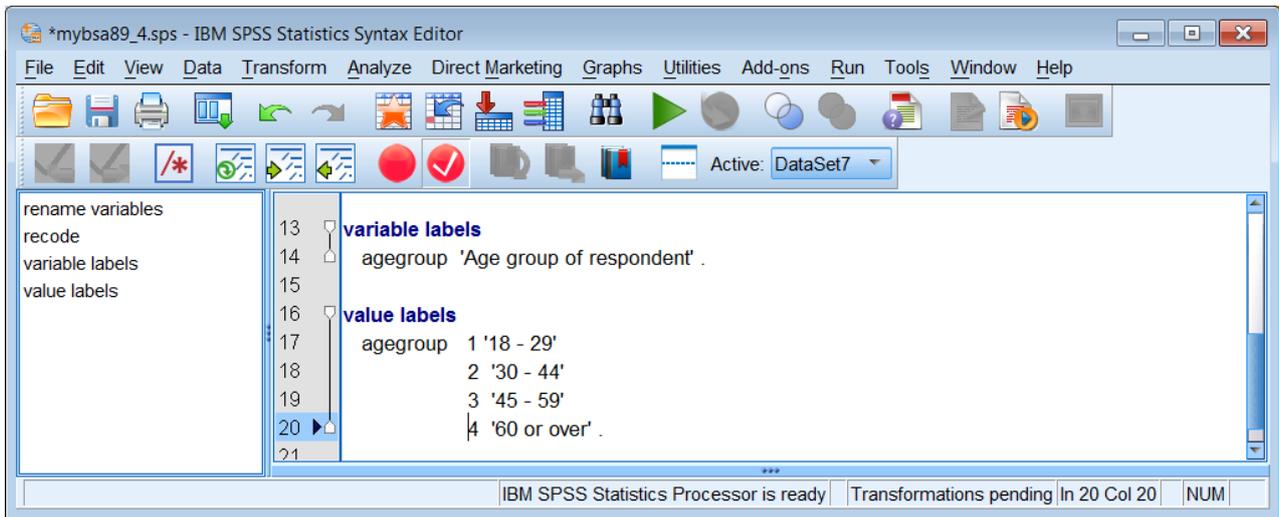
Add data dictionary information:

**variable labels**

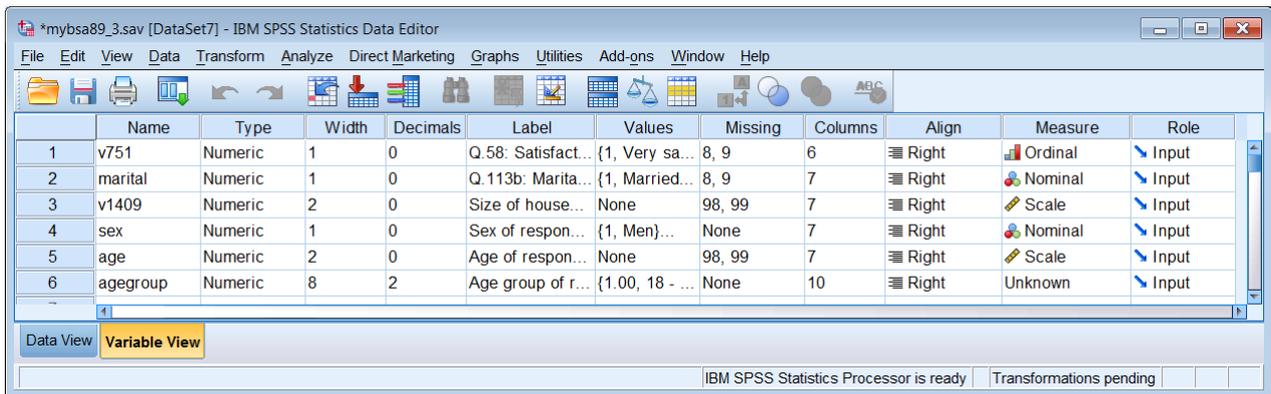
**agegroup** 'Age group of respondent' .

**value labels**

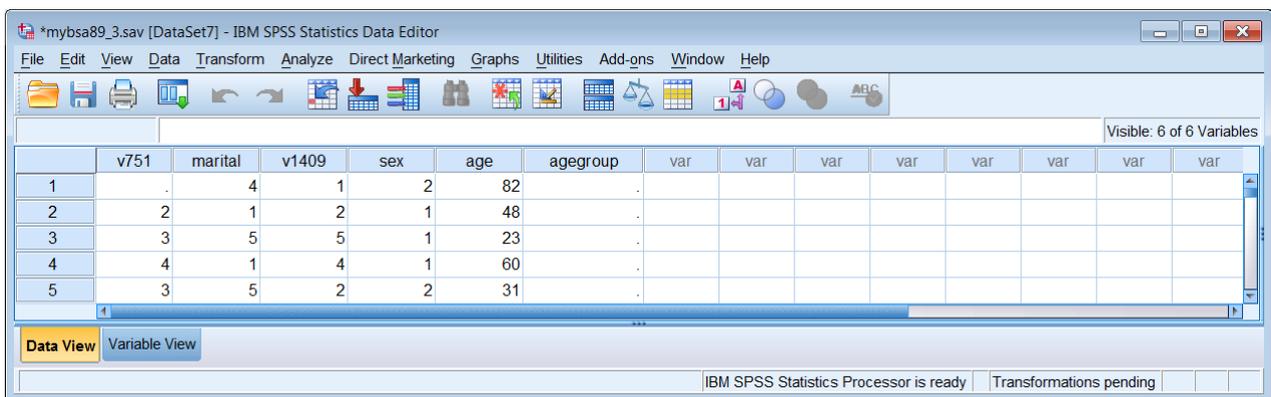
**agegroup** 1 '18 - 29'  
 2 '30 - 44'  
 3 '45 - 59'  
 4 '60 or over' .



Variable and value labels have now been added to **agegroup**:



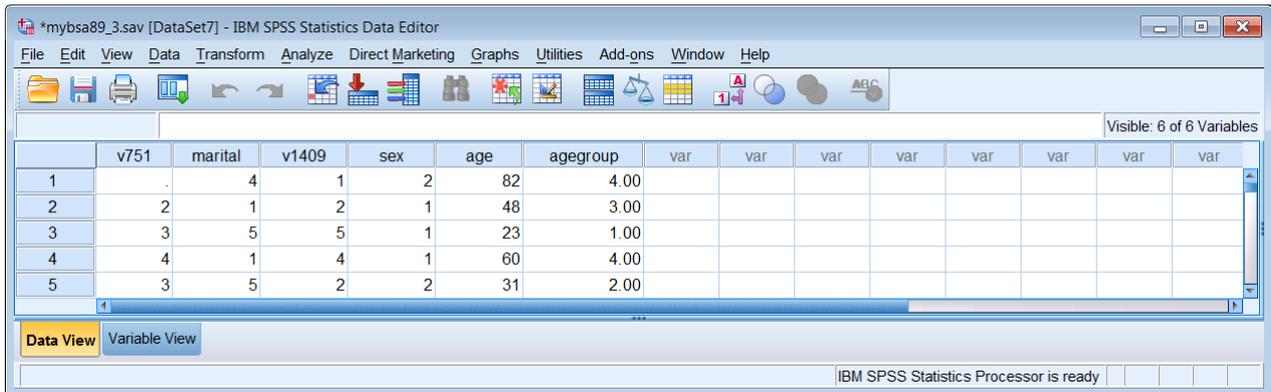
However, if you switch to Data View, you will see that there are no values entered for agegroup.



This is because SPSS is waiting for a command requiring a data pass. Type in **execute** .



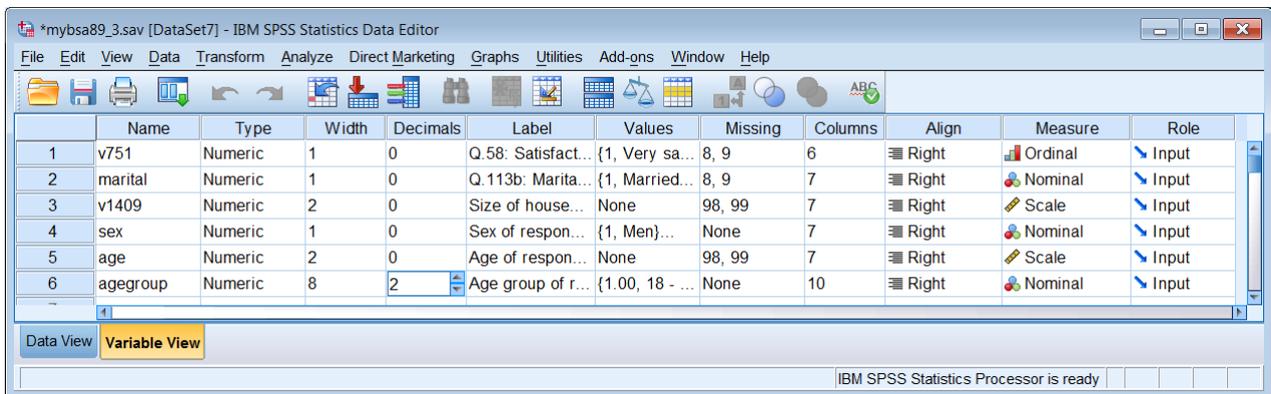
Run the command to get:



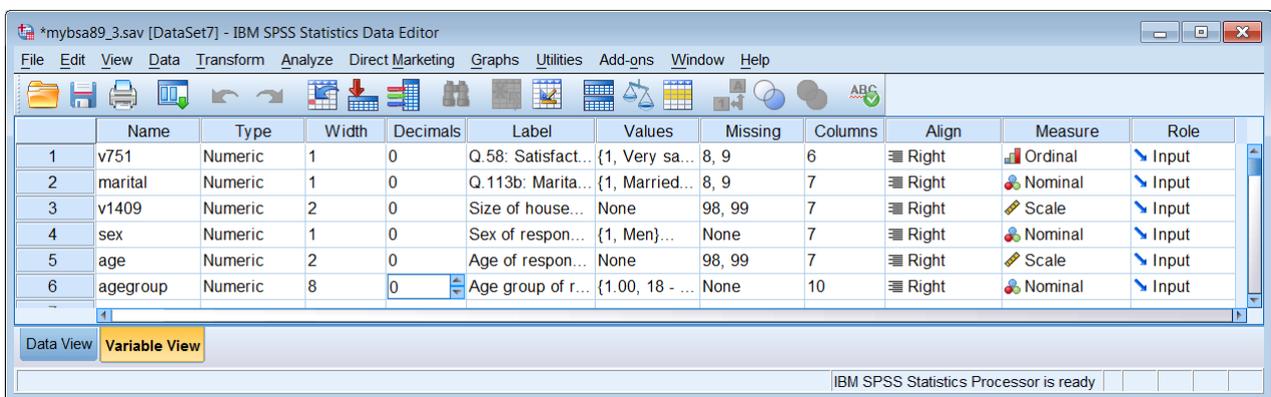
The values for agegroup have now been entered, but they have two redundant decimal places (the default format in SPSS for numeric variables). You can change this in syntax by:

**formats agegroup (f1.0) .**

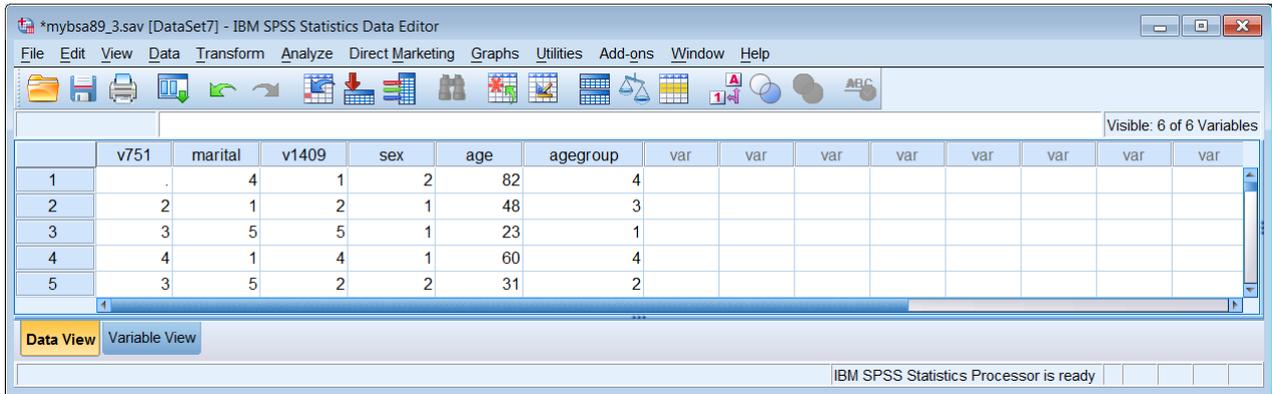
... or by switching back to Variable View, clicking on the decimals cell for agegroup:



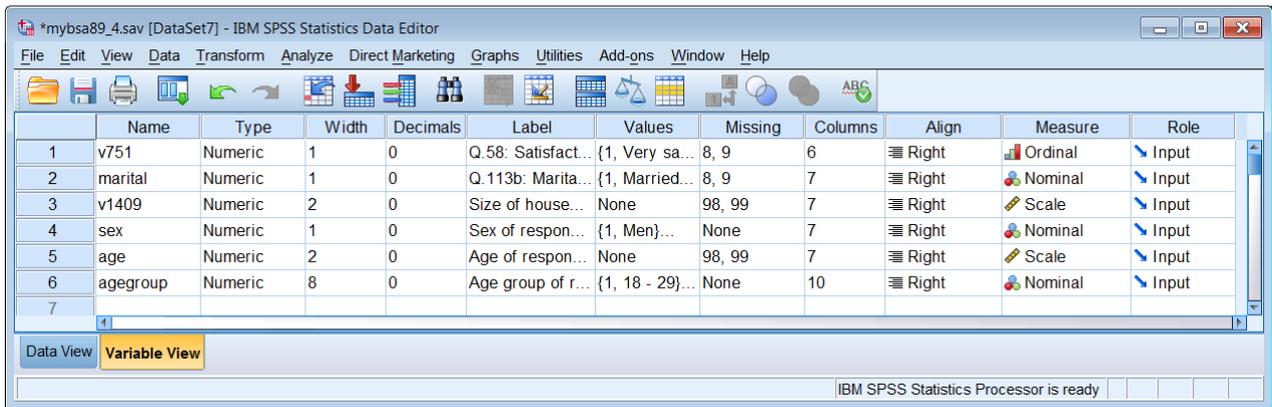
... and changing the 2 to 0.



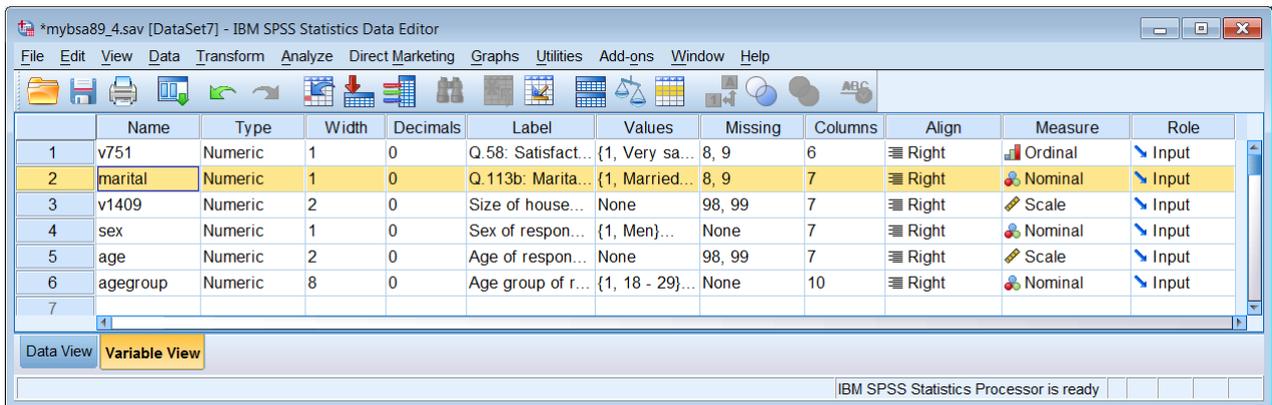
Switch to Data View and you will see that the decimal places have been removed and that the values are now all integer.



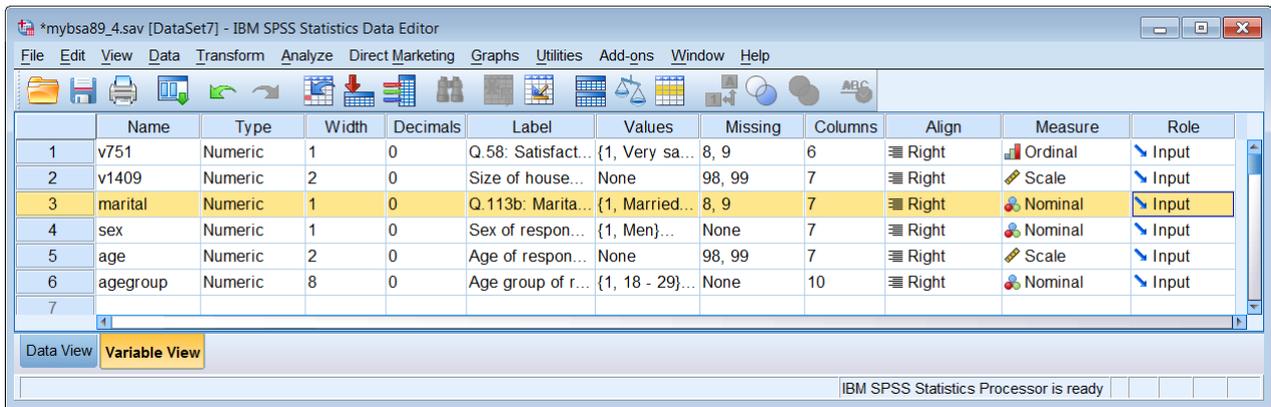
Switch back to Variable View:



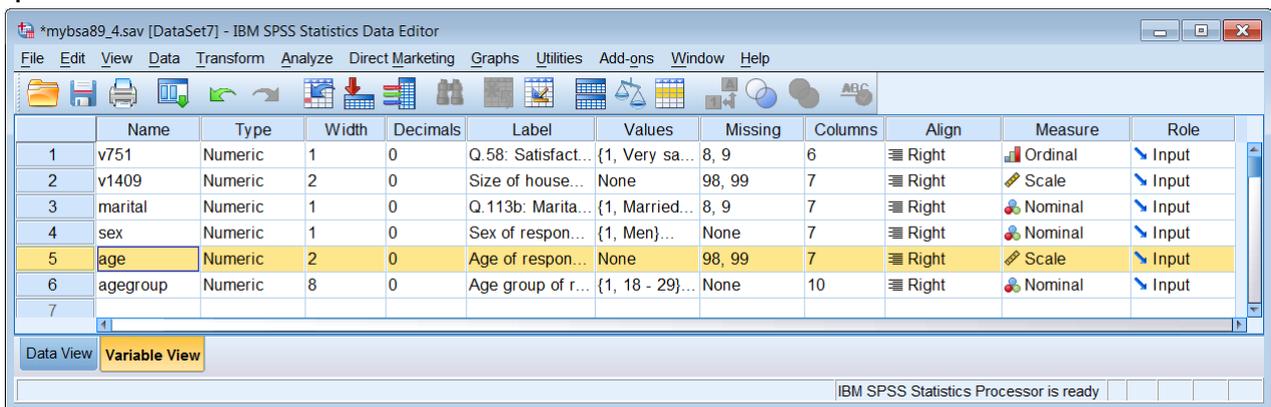
The file now needs tidying up a bit. Mindful of the earlier advice about keeping demographic variables together at the end of the file, we need to move **marital** down a bit. We can do this in the Data Editor by highlighting the row (click on row number 2):



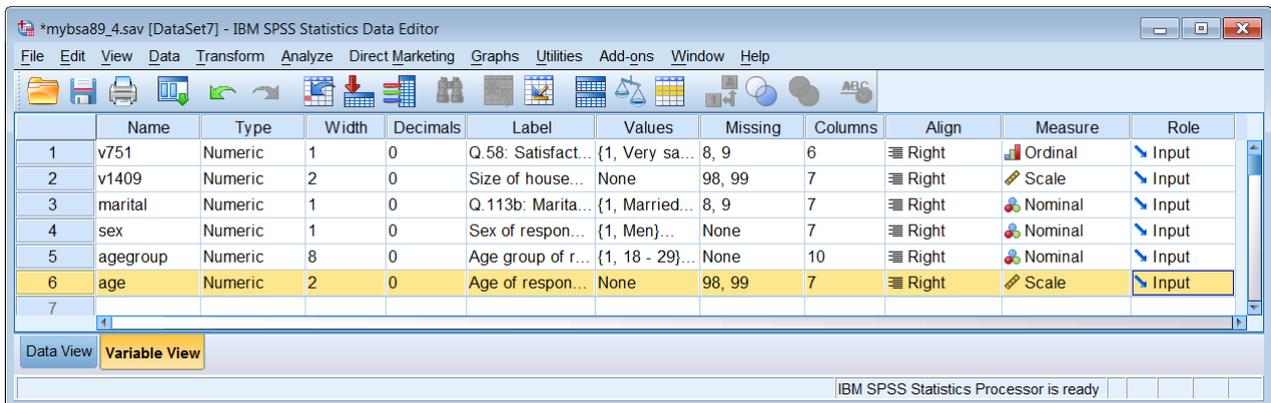
. . . keeping left button depressed and dragging the row down to below v1409 (a thin red horizontal line will display where the row will be moved):



Since analyses with demographic variables are often specified using `<varname> to <varname>` as in `marital to agegroup` we need to make sure that variables with many values are outside this block. Variable `age` needs to be moved to the end of the file, well out of harm's way.



Left click on row number 5, hold the button down and drag `age` to the bottom of the file:



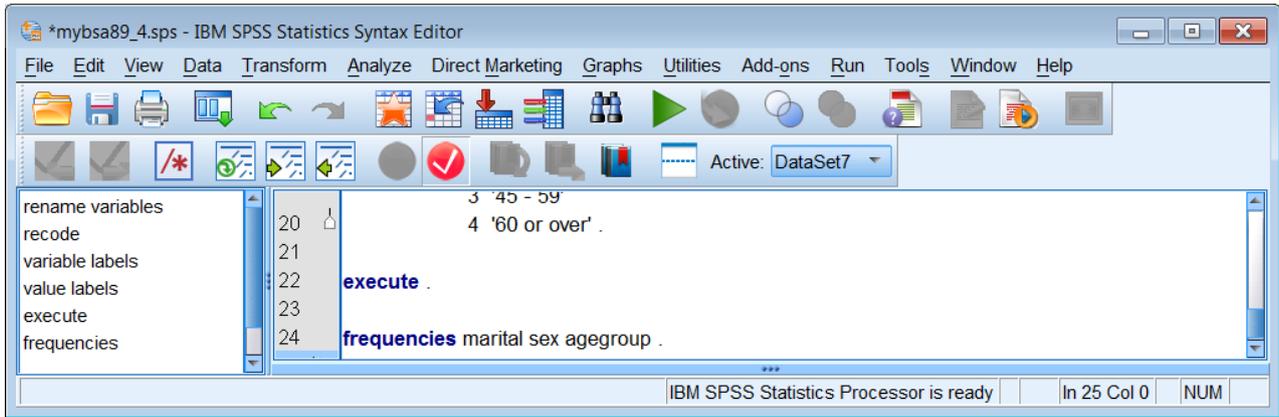
Each time you rename variables or recode your data, you should always check that the transformations have been correctly executed. In this case we need to check the renamed variables `marital` and `sex` and the new variable `agegroup`:

**frequencies marital sex agegroup .**

If we had left `age` where it was and used:

**frequencies marital to agegroup .**

.. there would have been a large frequency table for age as well.



**frequencies marital sex agegroup .**

. . . will yield:

**Statistics**

		Q.113b: Marital status of respondent	Sex of respondent	Age group of respondent
N	Valid	3024	3025	3015
	Missing	1	0	10

**marital Q.113b: Marital status of respondent**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Married	1948	64.4	64.4	64.4
	2 Living as married	114	3.8	3.8	68.2
	3 Separated or divorced	183	6.0	6.1	74.2
	4 Widowed	276	9.1	9.1	83.4
	5 Not married	503	16.6	16.6	100.0
	Total	3024	100.0	100.0	
Missing	9	1	.0		
Total		3025	100.0		

**sex Sex of respondent**

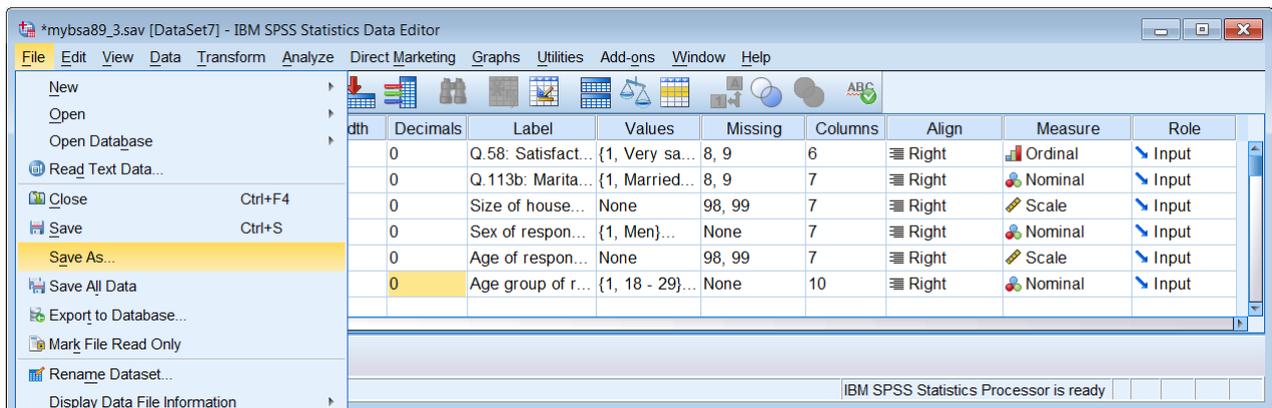
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Men	1393	46.0	46.0	46.0
	2 Women	1632	54.0	54.0	100.0
Total		3025	100.0	100.0	

agegroup Age group of respondent

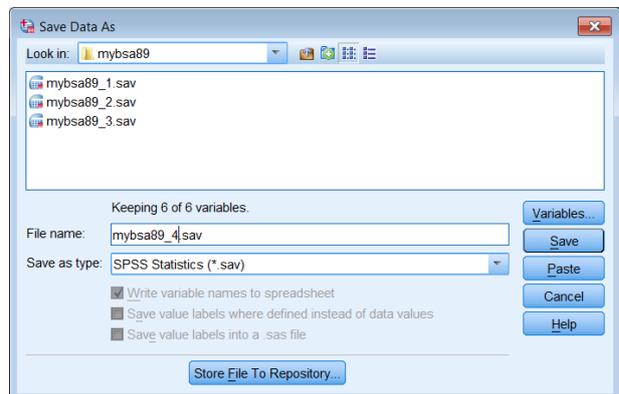
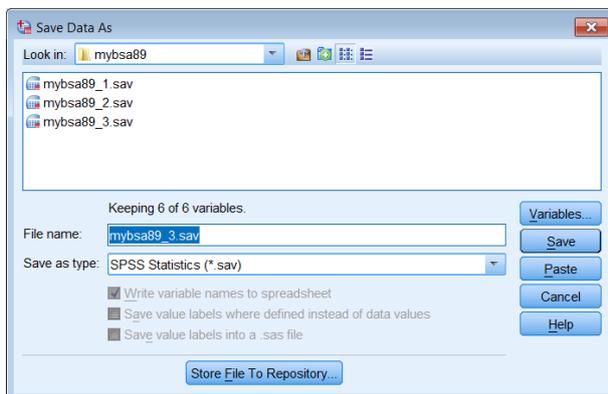
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 18 - 29	668	22.1	22.2	22.2
	2 30 - 44	839	27.7	27.8	50.0
	3 45 - 59	725	24.0	24.0	74.0
	4 60 or over	783	25.9	26.0	100.0
	Total	3015	99.7	100.0	
Missing	System	10	.3		
Total		3025	100.0		

Save the new version of the file as **mybsa89\_4.sav**.

File > Save as:



Change **3** to **4** in the filename. .



.. and click on **Save** :

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	v751	Numeric	1	0	Q.58: Satisfact...	{1, Very sa...	8, 9	6	Right	Ordinal	Input
2	marital	Numeric	1	0	Q.113b: Marita...	{1, Married...	8, 9	7	Right	Nominal	Input
3	v1409	Numeric	2	0	Size of house...	None	98, 99	7	Right	Scale	Input
4	sex	Numeric	1	0	Sex of respon...	{1, Men}...	None	7	Right	Nominal	Input
5	age	Numeric	2	0	Age of respon...	None	98, 99	7	Right	Scale	Input
6	agegroup	Numeric	8	0	Age group of r...	{1, 18 - 29}...	None	10	Right	Nominal	Input
7											

mybsa89\_4.sav

You should also save the syntax file as **mybsa89\_4.sps**:

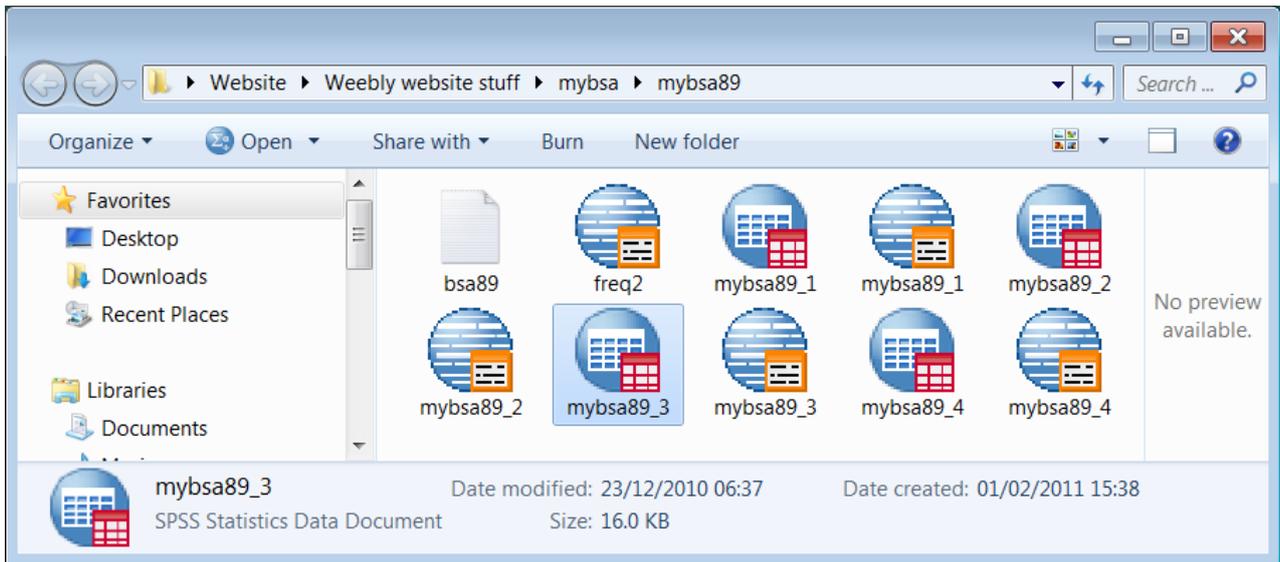
```

1  rename variables
2  ( v1408 = marital)
3  (v1411 = sex)
4  (v1412 = age ) .
5  recode
6  age (18 thru 29 = 1)
7  (30 thru 44 = 2)
8  (45 thru 59 = 3)
9  (60 thru 97 = 4)
10 (else = sysmis) into agegroup .
11 variable labels
12 agegroup 'Age group of respondent' .
13 value labels
14 agegroup 1 '18 - 29'
15           2 '30 - 44'
16           3 '45 - 59'
17           4 '60 or over' .
18 formats agegroup (f1.0) .
19 execute .
20 frequencies marital sex agegroup .
21

```

You can do this bit by yourself!

Your folder **mybsa89** should now look like this.



### End of exercise 2.3.1.2

**Next session:** [to be decided as remaining tutorials in this section are being revised to add more interesting and apposite variables, and to provide incremental steps via conditional frequencies towards some serious data analysis of two or more variables using **CROSSTABS** and **MEANS**.]

[\[Back to Block 2 menu\]](#)