

Block 2: Analysing one variable

2.3 Data transformations

2.3.1.2a2 Recode into new variable

[Draft only: 6 June 2013]

Previous session: [2.3.1.2a1 Select and rename variables](#)

Exemplar: British Social Attitudes 1986

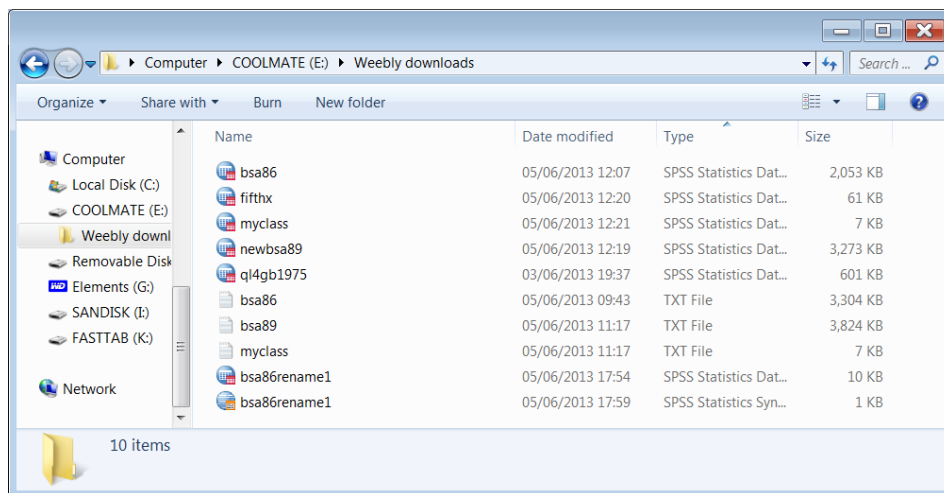
File: **bsa86rename1.sav**
[generated in previous session in folder **e:\weebly downloads**]

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

b) Add variable and value labels to **agegroup**.

New commands: **RECODE ~ ~ ~ INTO ~ ~ ~**
VARIABLE LEVEL
FORMATS

Step 1: Insert your memory stick from the previous session into drive **e:** and navigate to folder **e:\weebly downloads**:

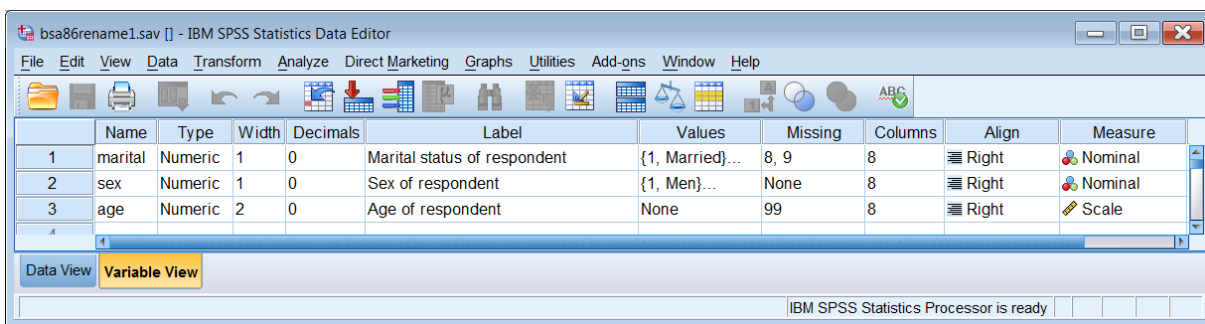
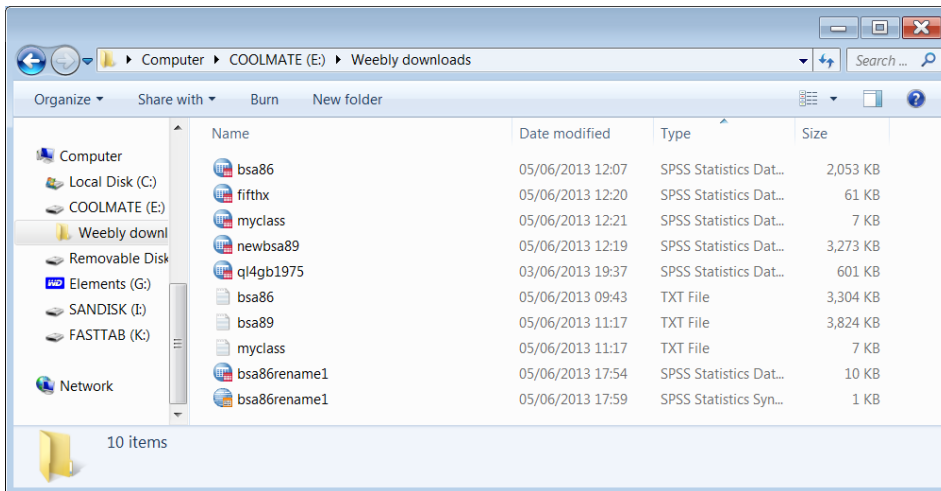


If you don't have file **bsa86rename1.sav** go back to exercise [2.3.1.2a1 Select and rename variables](#)

I love you all dearly, but I have neither time nor inclination to take you all the way back to what you should have done by now in previous exercises.

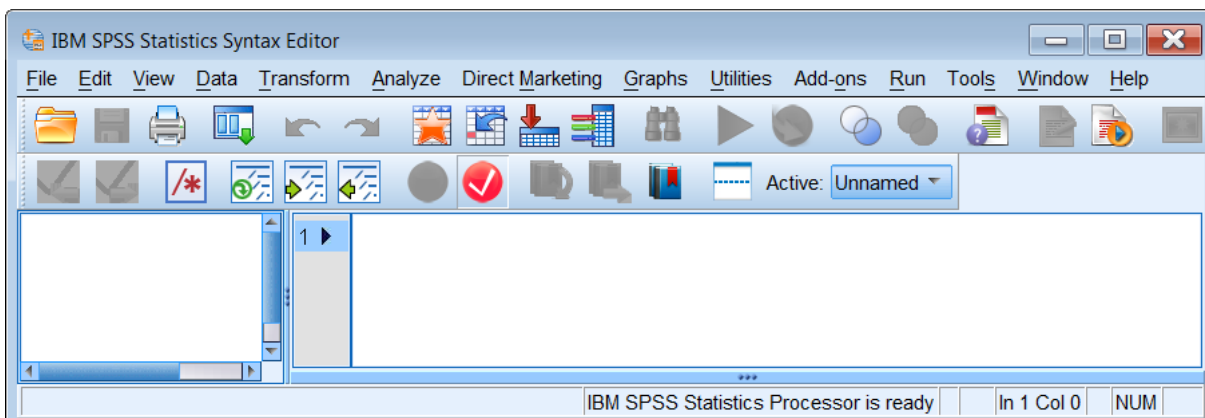
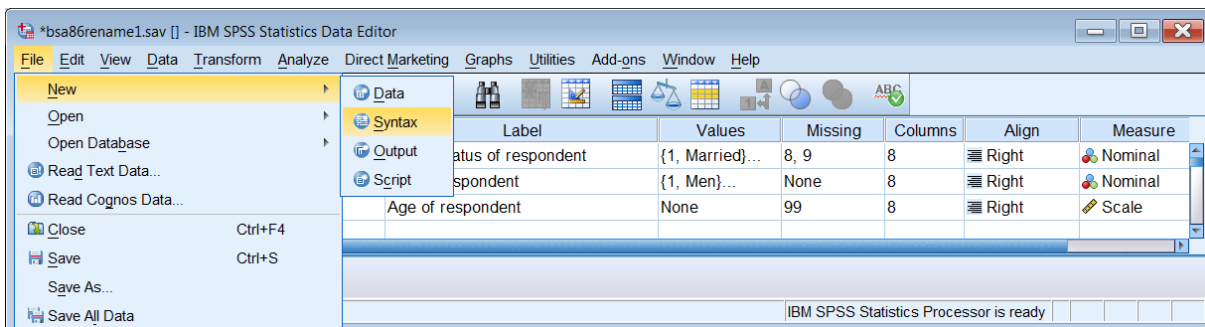
All subsequent steps start on separate pages.

Step 2: In drive **e:** double click on file **bsa86rename1.sav**



Step 3: Open a new Syntax Editor

File > New > Syntax



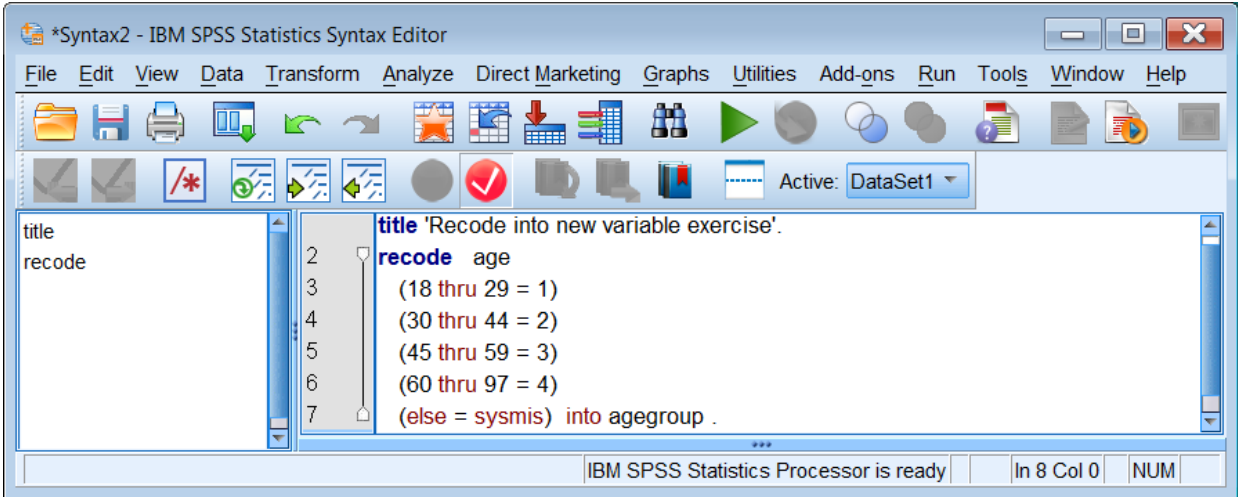
Step 4: Recode age into new variable agegroup

Type in:

```

title 'Recode into new variable exercise'. [Always give your job a title or insert comments]
recode age
    (18 thru 29 = 1)
    (30 thru 44 = 2)
    (45 thru 59 = 3)
    (60 thru 97 = 4)
    (else = sysmis) into agegroup .
    
```

[Don't forget the primes or the full stops!]



Run > All to get:

The screenshot shows the IBM SPSS Statistics Data Editor window in Variable View. The title bar reads '*bsa86rename.sav [DataSet1] - IBM SPSS Statistics Data Editor'. The menu bar includes File, Edit, View, Data, Transform, Analyze, Direct Marketing, Graphs, Utilities, Add-ons, and Window. The table below shows the properties of the variables in the dataset.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	marital	Numeric	1	0	Marital status of respondent	{1, Married}...	8, 9	8	Right	Nominal	Input
2	sex	Numeric	1	0	Sex of respondent	{1, Men}...	None	8	Right	Nominal	Input
3	age	Numeric	2	0	Age of respondent	None	99	8	Right	Scale	Input
4	agegroup	Numeric	8	2		None	None	10	Right	Unknown	Input

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

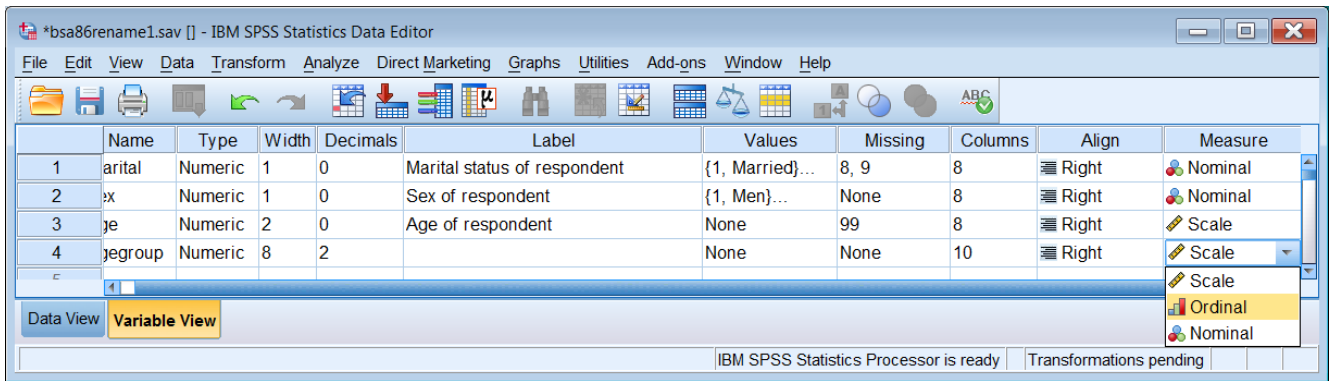
Note that the Measure of agegroup is displayed as Unknown and Decimals as 2.

The screenshot shows the IBM SPSS Statistics Data Editor window in Variable View. The title bar reads '*bsa86rename1.sav [] - IBM SPSS Statistics Data Editor'. The menu bar includes File, Edit, View, Data, Transform, Analyze, Direct Marketing, Graphs, Utilities, Add-ons, and Window. The table below shows the properties of the variables in the dataset.

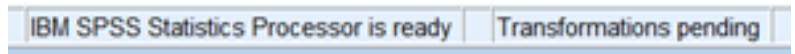
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	arital	Numeric	1	0	Marital status of respondent	{1, Married}...	8, 9	8	Right	Nominal
2	ex	Numeric	1	0	Sex of respondent	{1, Men}...	None	8	Right	Nominal
3	ge	Numeric	2	0	Age of respondent	None	99	8	Right	Scale
4	gegroup	Numeric	8	2		None	None	10	Right	Unknown

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

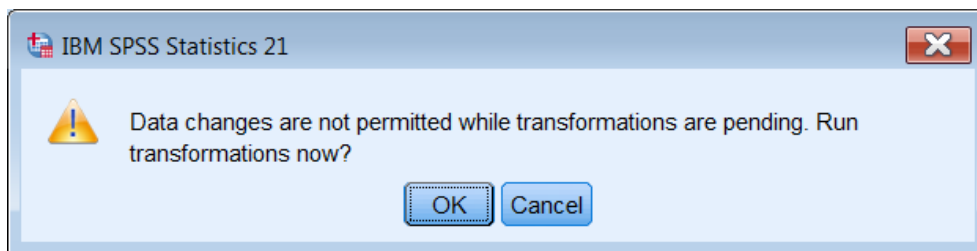
There are two quick ways to change **Decimals** to 0 and **Measure** to Ordinal



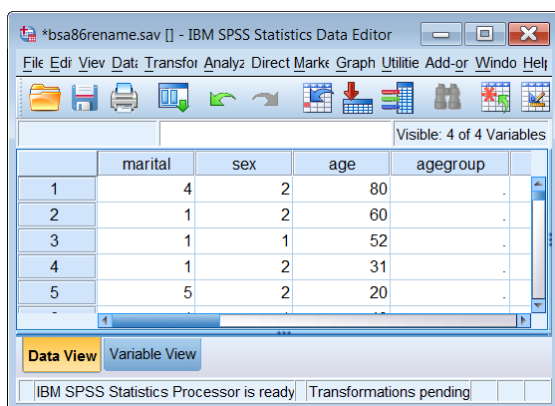
... but these are not allowed whilst transformations are pending (See message displayed right at the bottom of the Data Editor).



If you try to change them in the **Data Editor** whilst this is displayed. SPSS will display a message:

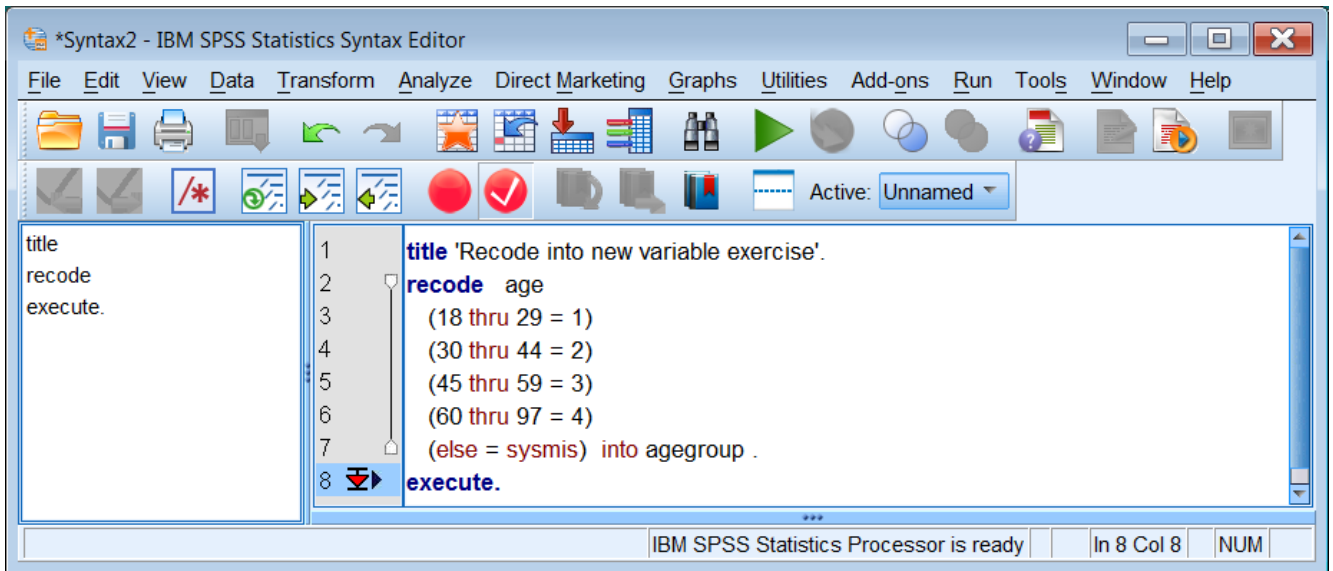


If you switch to **Data View**, you will see that there are still no values entered for **agegroup**. This is because SPSS is waiting for a command requiring a data pass.

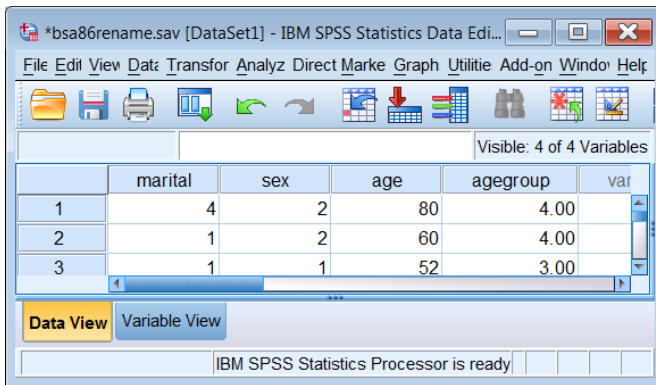


Go back to the Syntax Editor and type in :

execute .

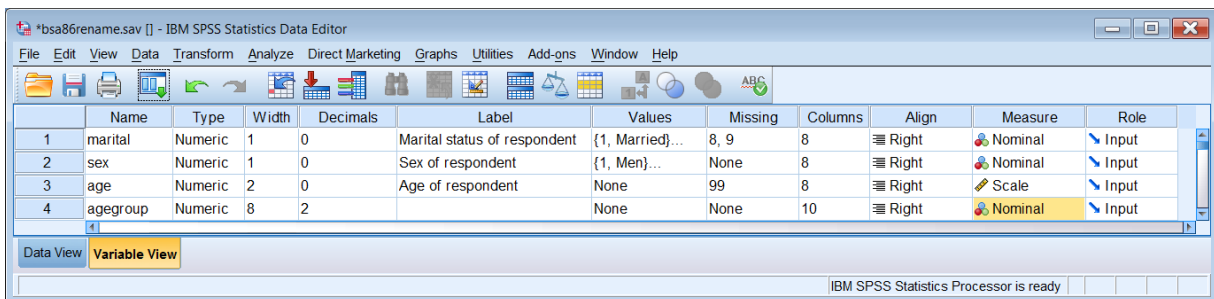


Run the job to get:

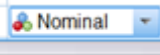


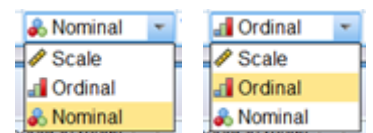
The values for **agegroup** have now been entered,

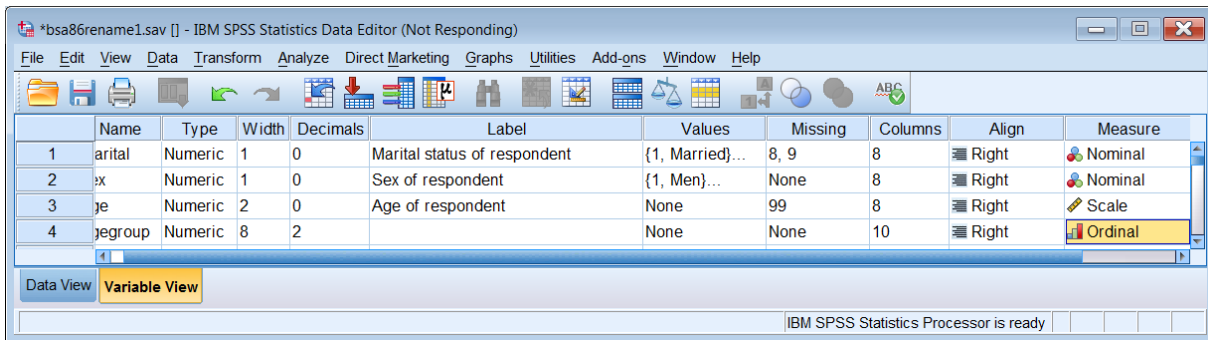
Go back to **Variable View**



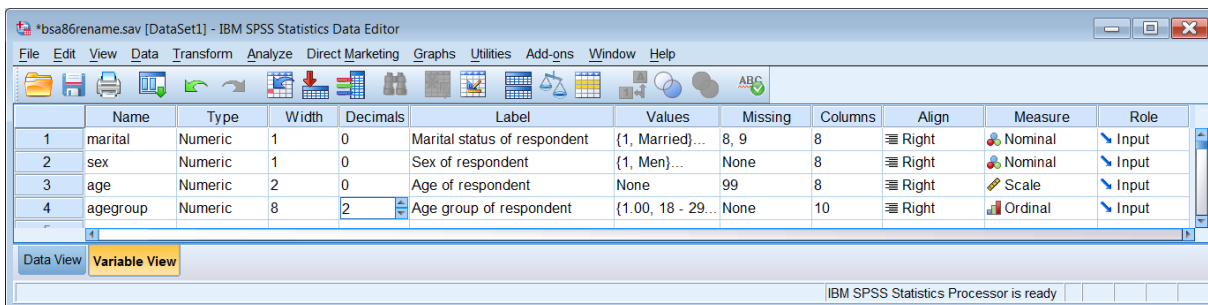
In the Measure column for **agegroup**, **Unknown** has changed to **Nominal**. This because SPSS detected only four values and set it to the default **Nominal**. You can change this to **Ordinal** in the Data Editor by clicking on

 and changing it to **Ordinal**:

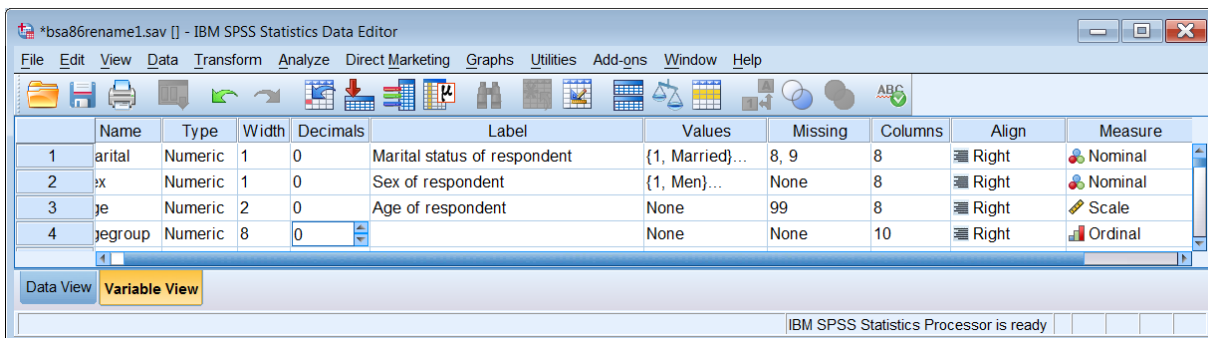




The new variable **agegroup** which should have four integer values, has two superfluous decimals. You can change this in the **Data Editor** by clicking on the cell under **Decimals**



... and changing the 2 to 0.



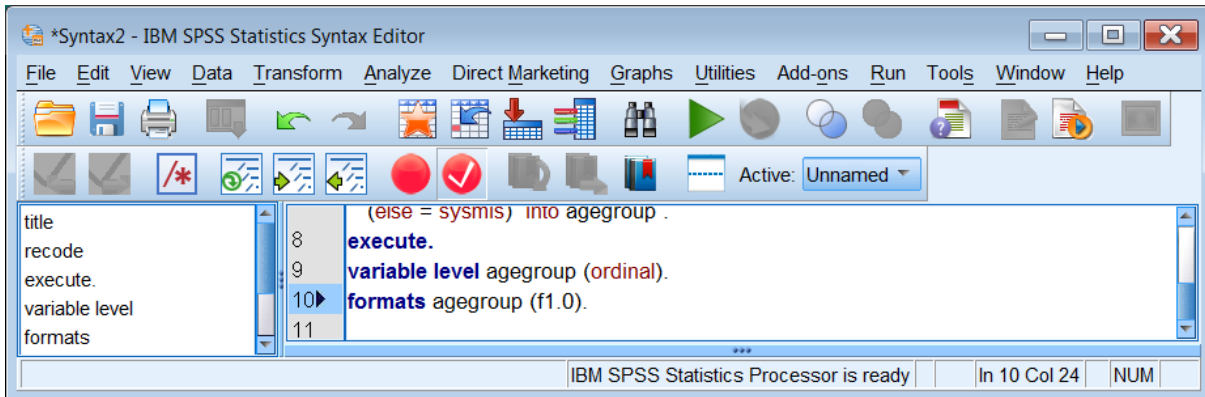
With only one or two variables these modifications are quicker to do manually in the **Data Editor**, but if you have dozens or even hundreds of variables it's much preferable to have some syntax for future use in case you (or someone else) later need to recreate the file.

Step 5: Change measurement level and format of **agegroup**

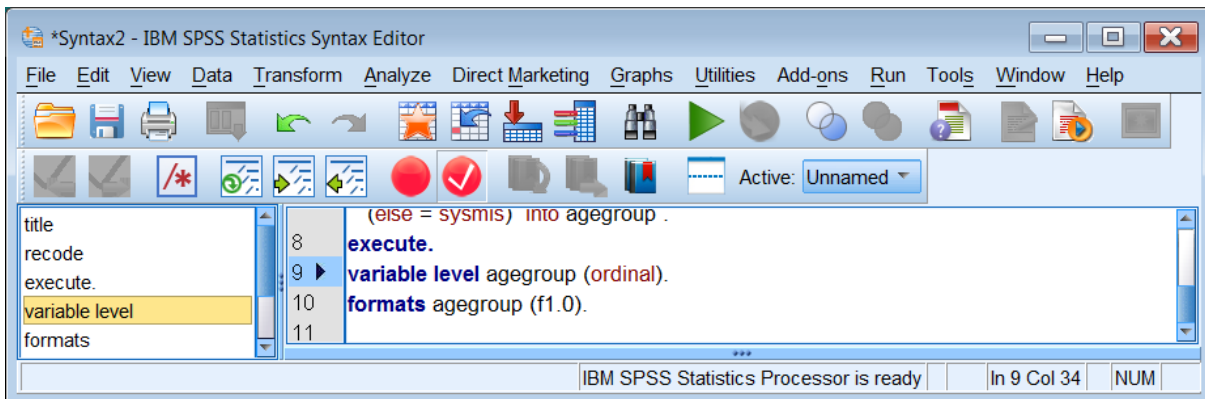
Use the **VARIABLE LEVEL** and **FORMATS** commands

Go back to your Syntax Editor and type in:

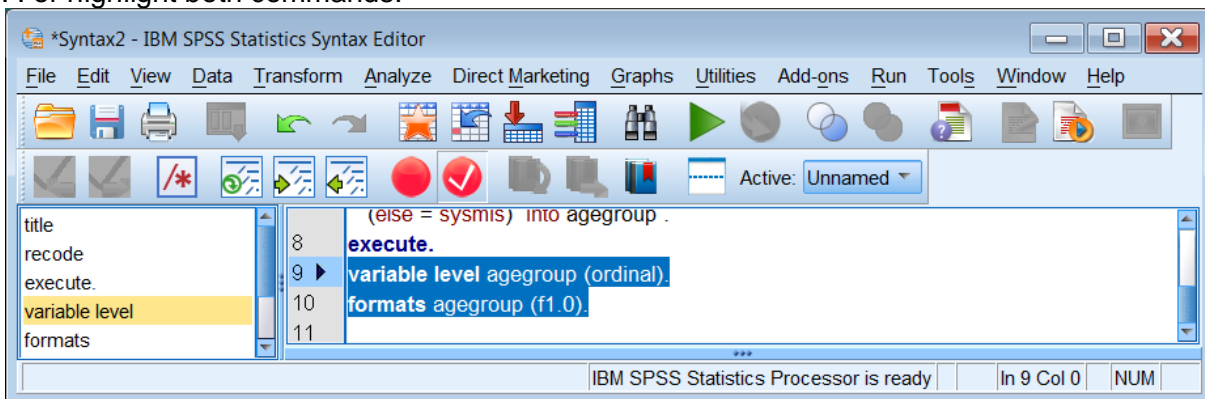
variable level agegroup (ordinal).
formats agegroup (f1.0).




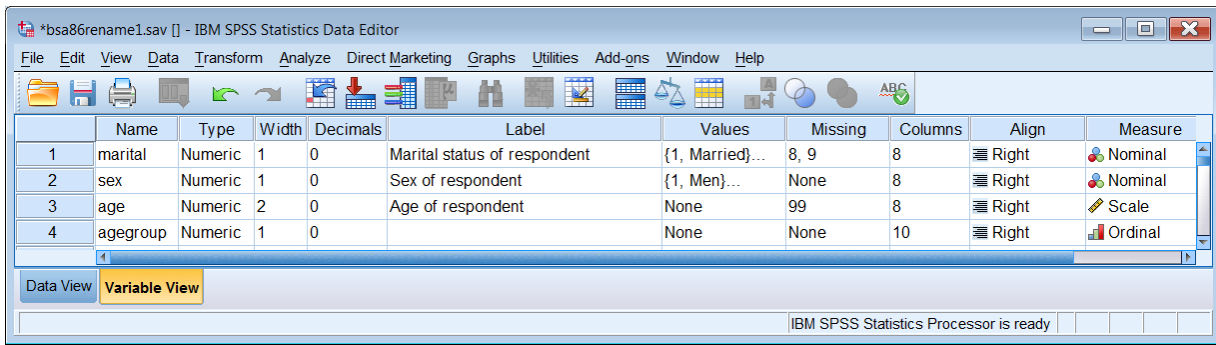
Highlight **variable level** and click on **Run > To End**



.. or highlight both commands:

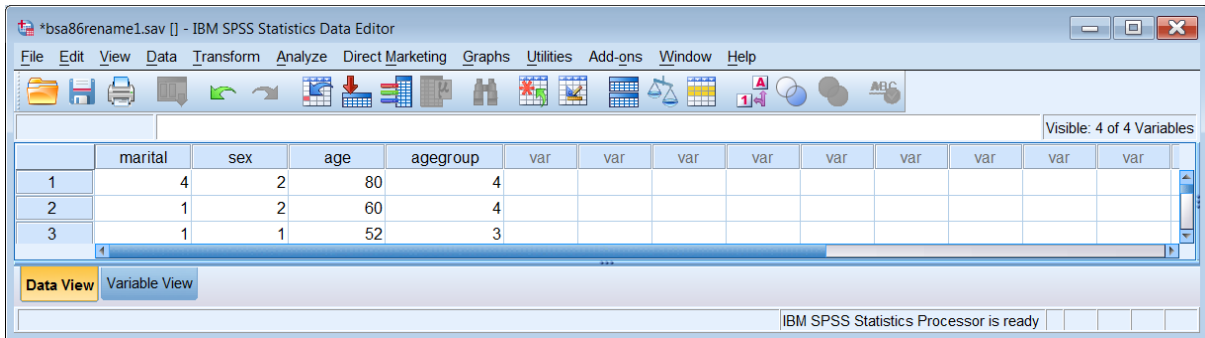


and press the green triangle  to get..



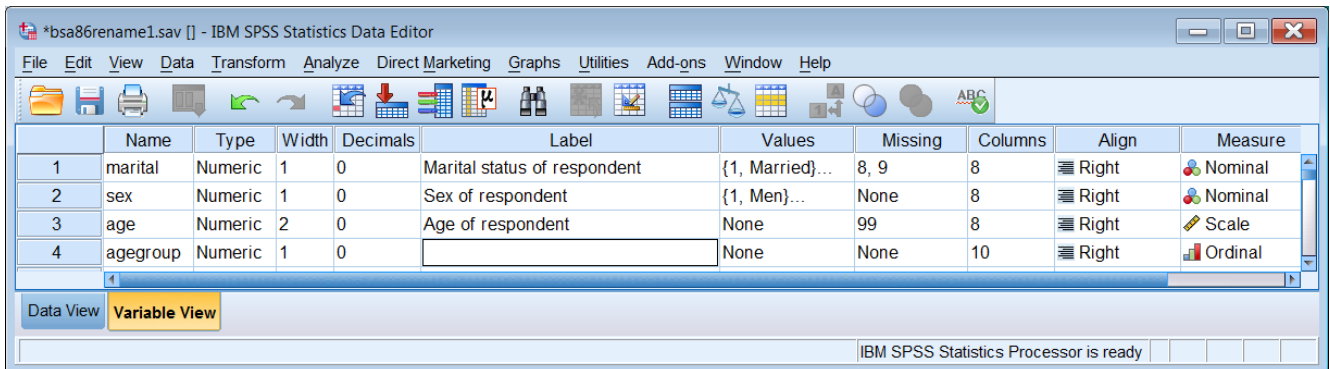
In the **Measure** column, **agegroup** is now displayed as **Ordinal**.

Switch to **Data View** and you will see that the decimal places have been removed and that the values for **agegroup** are now displayed as integers.



Step 6: Add a variable label for agegroup

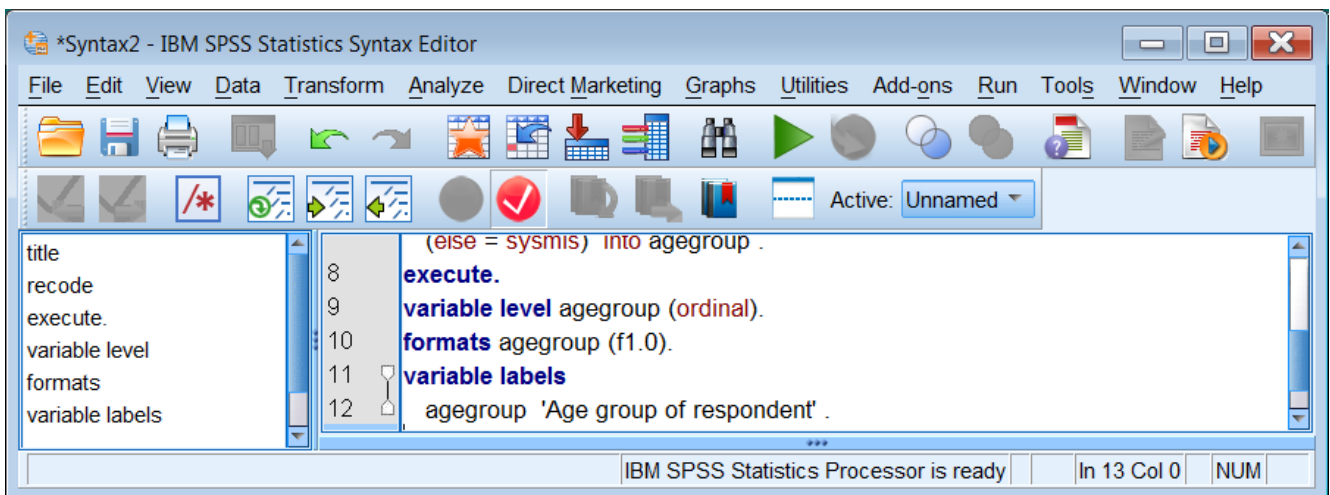
Go back to **Variable View**:



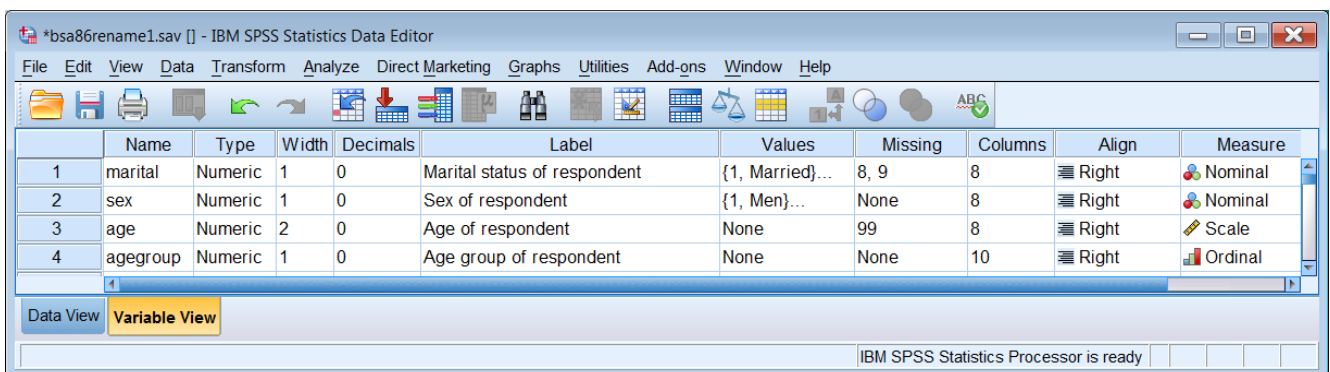
With only one variable you can do this manually by typing **Age group of respondent** into the **Data Editor**, but it is always good practice to generate and keep syntax for future use.

Go back to the Syntax Editor and type in the following:

variable labels
agegroup 'Age group of respondent' .



Run to get:

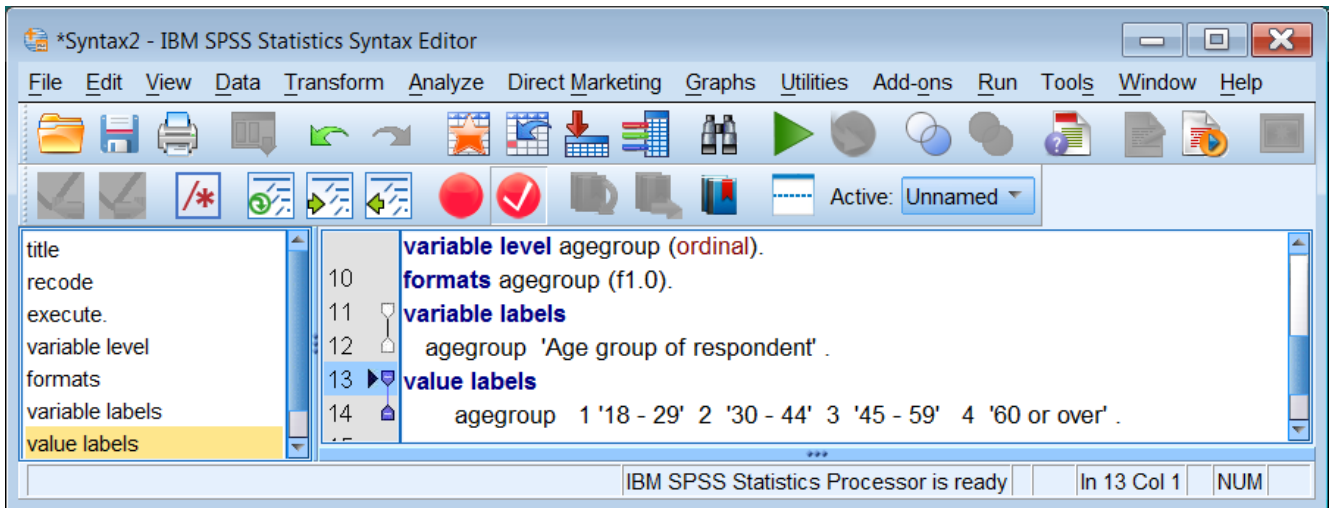


[NB: Sometimes this doesn't work, so the best thing to do is continue to Step 7. If that doesn't work, add an **execute . command and run that to add both variable and value labels.]**

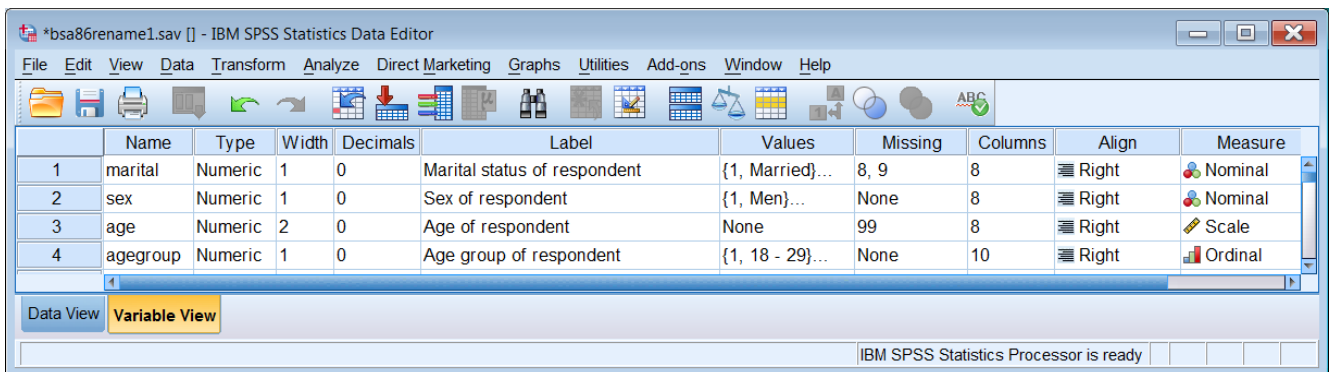
Step 7: Add the value labels for agegroup

Type in:

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



Run to get:

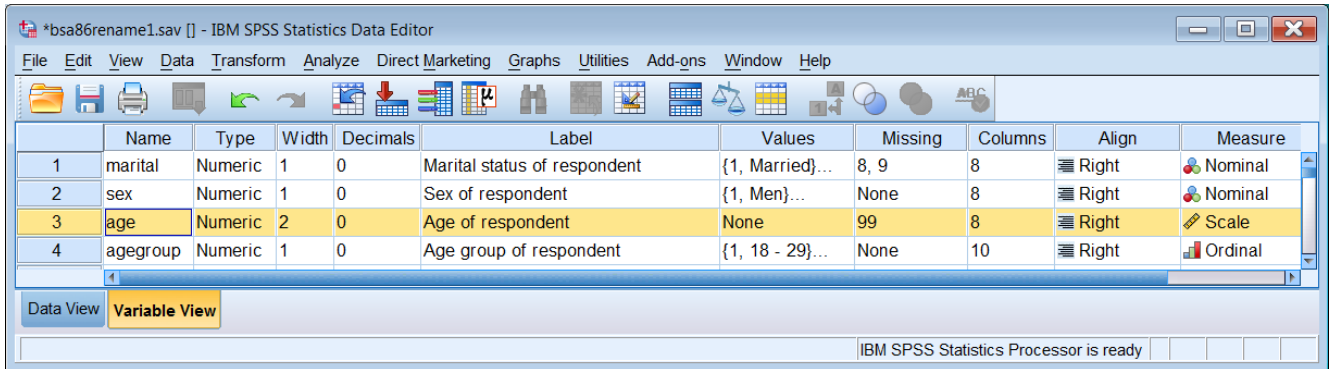


Variable and value labels have now been added to **agegroup**:

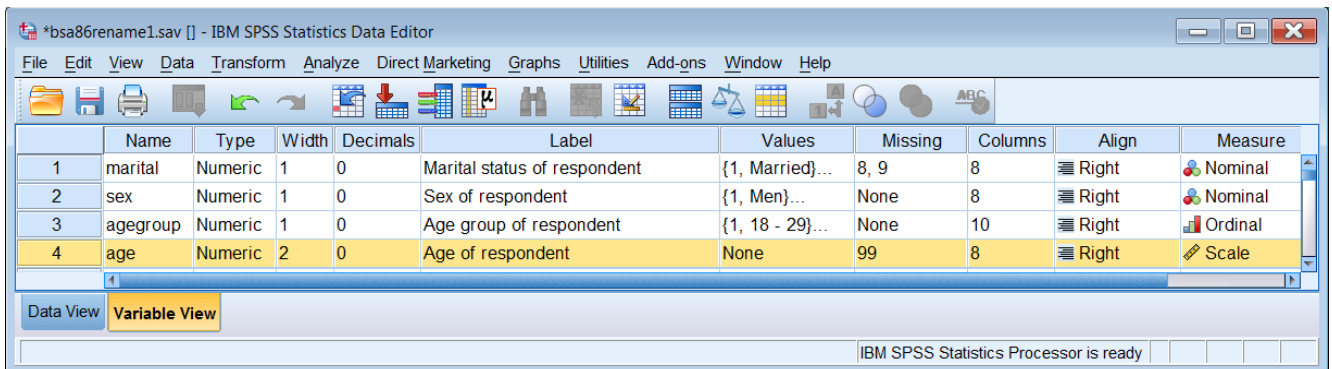
Step 8: Re-arrange the variable order

The file now needs tidying up a bit. 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** therefore needs to be moved to the end of the file, well out of harm's way.

We can do this in the **Data Editor**. Left click on row number **3**:



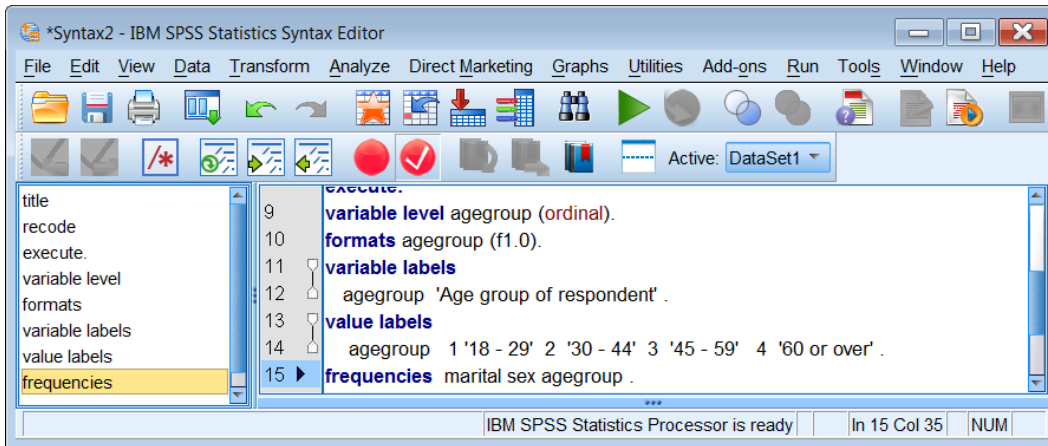
Hold the left button down on **3** and drag **age** to the bottom of the file, below **agegroup** (a thin red horizontal line will display where the row will be moved):



Step 9: Check your file

Each time you perform data transformations such as **RENAME** or **RECODE**, you should always check that they 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 .



Run the **FREQUENCIES** command to yield:

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

There are 3025 cases in the file. One person is missing for marital status and 10 missing for age

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		

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

frequencies marital to agegroup .

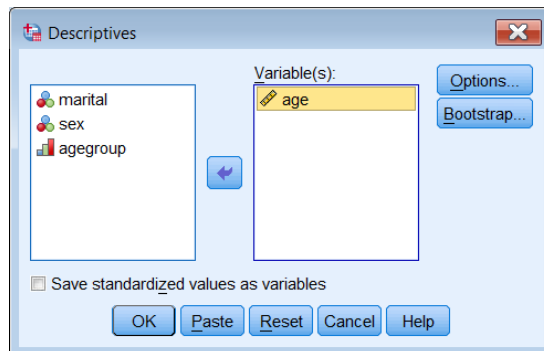
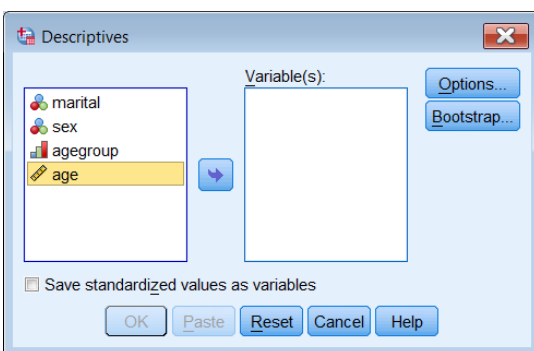
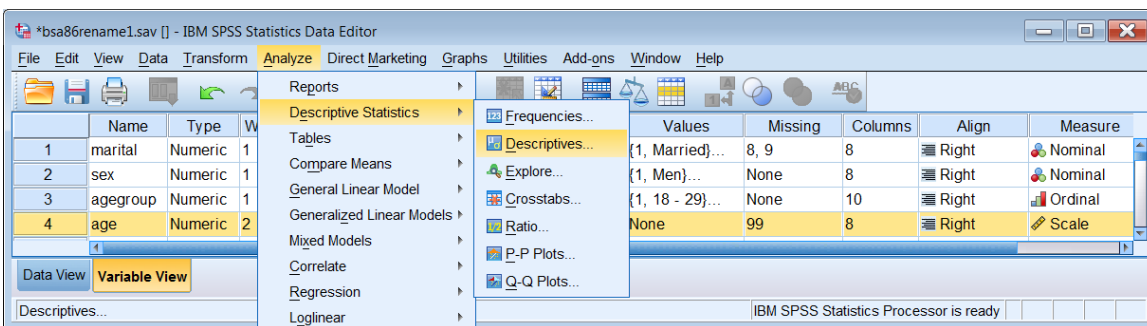
. . there would have been a very large frequency table for age as well!

Best check for **age** is:

descriptives age.

OR:

Analyze > Descriptive Statistics > Descriptives



Press **OK** to get:

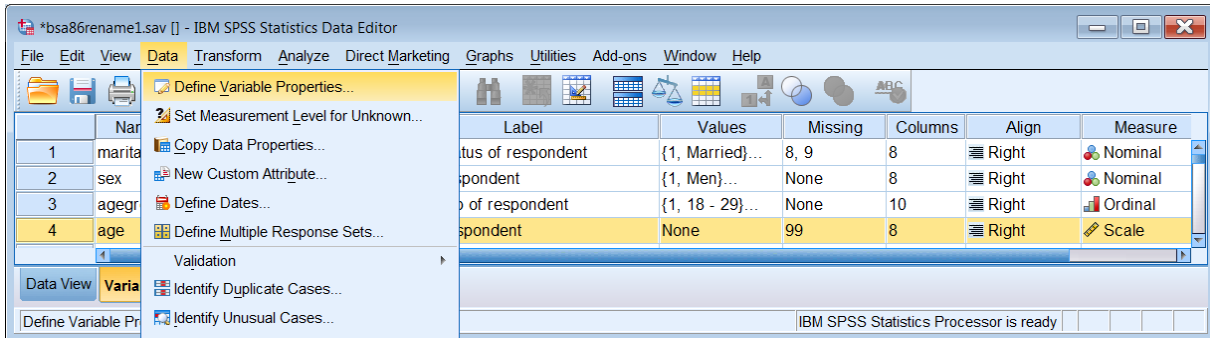
Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
age Age of respondent	3097	18	97	45.23	17.556
Valid N (listwise)	3097				

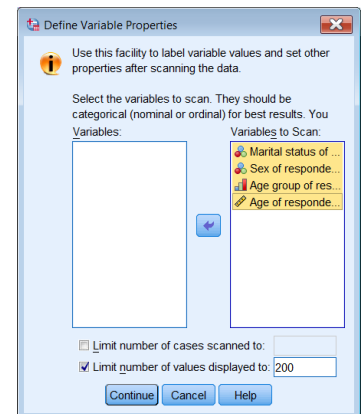
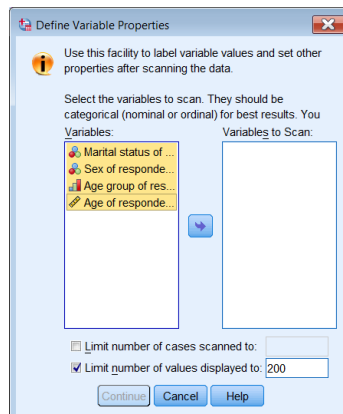
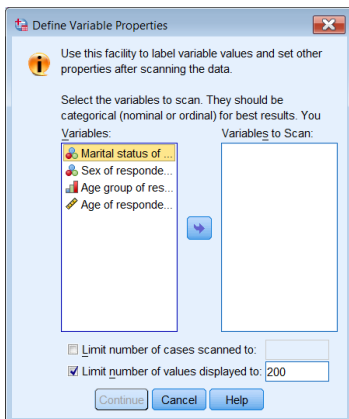
A particularly useful way of checking variable properties is:

Data > Define Variable Properties

[NB: It wasn't designed to this, but it's a lot quicker than asking for frequencies.]



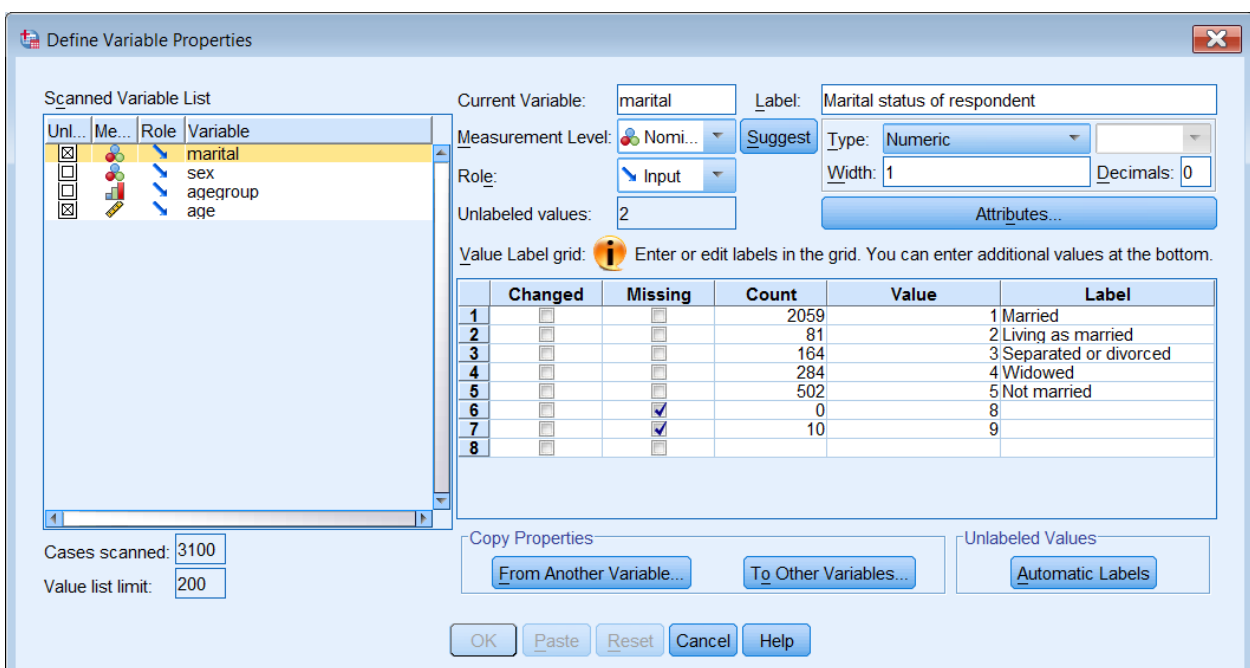
Move cursor to **ress** {Shift} click to highlight all four variables and]



[Shift] click on **Age** . .

click on blue arrow to transfer to right pane . .

click on **Continue**



Define Variable Properties

Scanned Variable List

Unl...	Me...	Role	Variable
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	marital
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	sex
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	agegroup
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	age

Current Variable: sex Label: Sex of respondent

Measurement Level: Nomi... Type: Numeric Width: 1 Decimals: 0

Role: Input

Unlabeled values: 0

Value Label grid: Enter or edit labels in the grid. You can enter additional values at the bottom.

	Changed	Missing	Count	Value	Label
1	<input type="checkbox"/>	<input type="checkbox"/>	1445	1	Men
2	<input type="checkbox"/>	<input type="checkbox"/>	1655	2	Women
3	<input type="checkbox"/>	<input type="checkbox"/>			

Cases scanned: 3100 Value list limit: 200

Copy Properties: Unlabeled Values:

Define Variable Properties

Scanned Variable List

Unl...	Me...	Role	Variable
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	marital
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	sex
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	agegroup
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	age

Current Variable: agegroup Label: Age group of respondent

Measurement Level: Ordinal Type: Numeric Width: 1 Decimals: 0

Role: Input

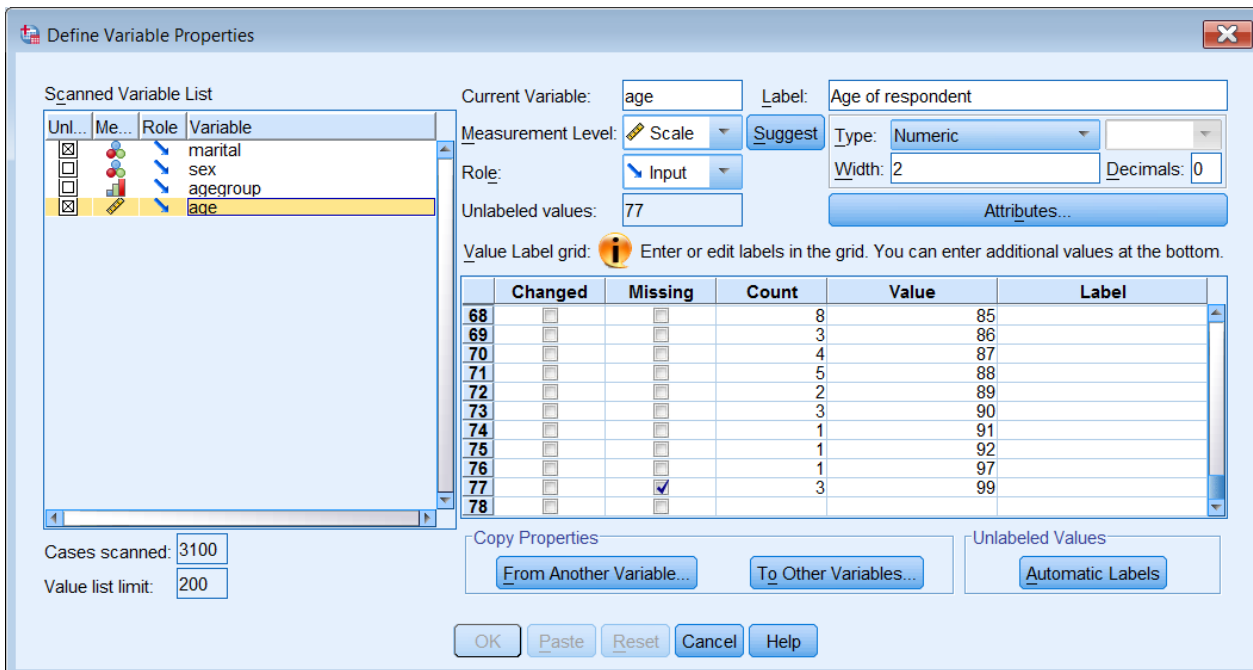
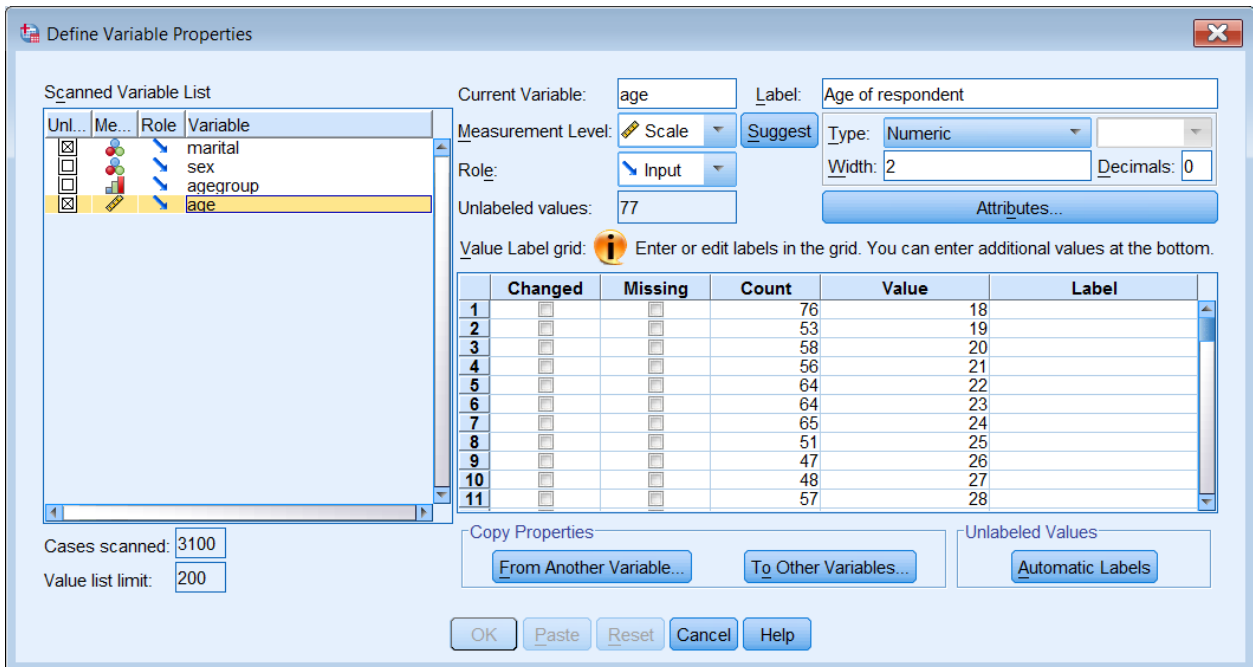
Unlabeled values: 0

Value Label grid: Enter or edit labels in the grid. You can enter additional values at the bottom.

	Changed	Missing	Count	Value	Label
1	<input type="checkbox"/>	<input type="checkbox"/>	701	1	18 - 29
2	<input type="checkbox"/>	<input type="checkbox"/>	917	2	30 - 44
3	<input type="checkbox"/>	<input type="checkbox"/>	748	3	45 - 59
4	<input type="checkbox"/>	<input type="checkbox"/>	731	4	60 or over
5	<input type="checkbox"/>	<input type="checkbox"/>			

Cases scanned: 3100 Value list limit: 200

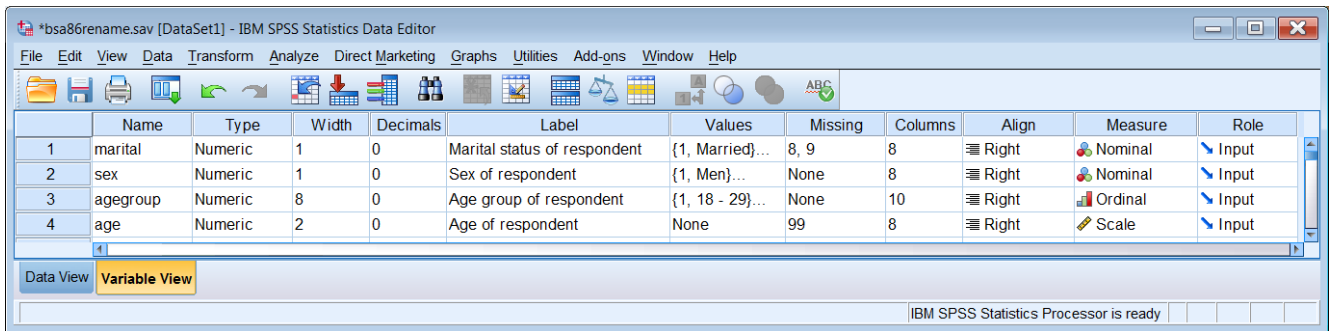
Copy Properties: Unlabeled Values:



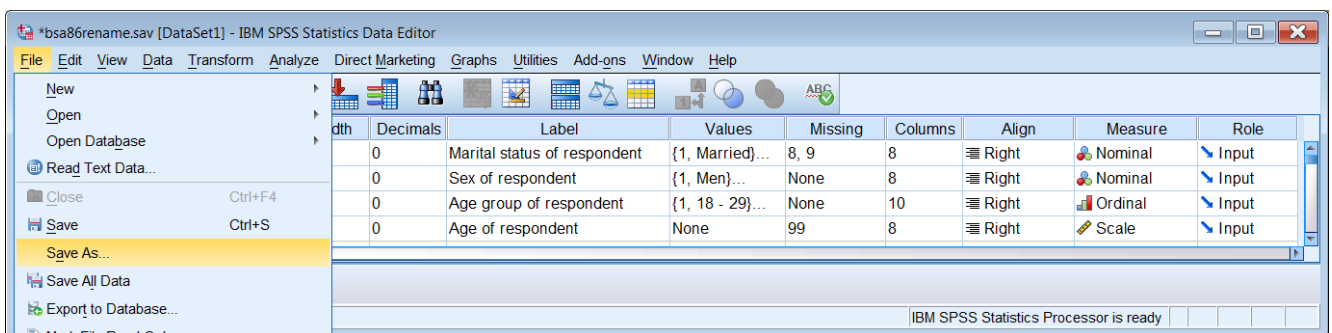
Data > Define Variable Properties is probably one of the most useful facilities in SPSS as it allows not only instant checking of all the main properties of a variable, but also a facility for changing them.

Step 10: Save the new version of the file as **bsa86rename2.sav** in folder **e:\weebly downloads** .

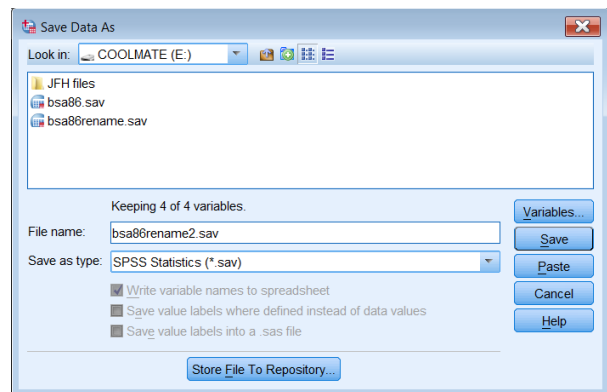
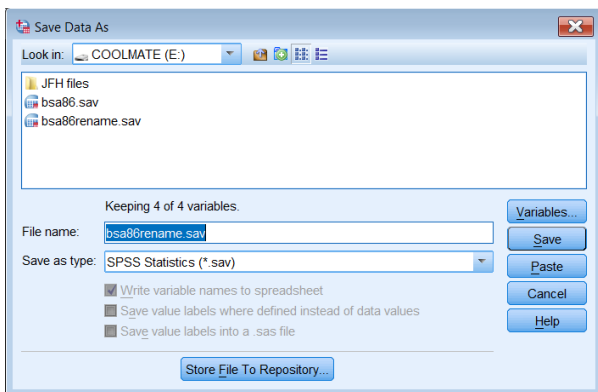
From the **Data Editor**:



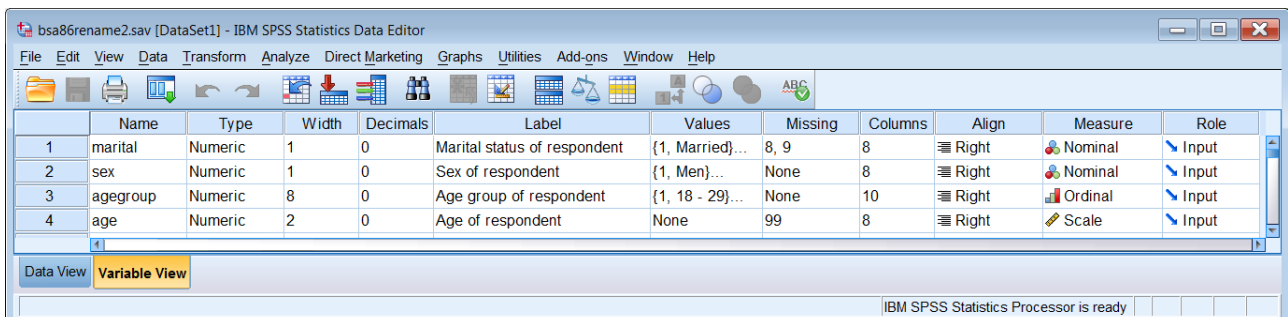
File > Save as:



Change **bsa86rename1.sav** to **bsa86rename2.sav**



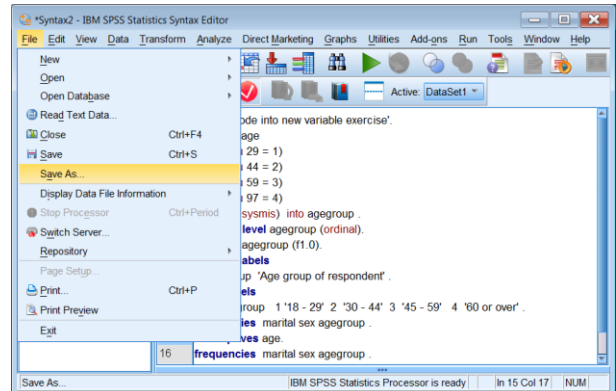
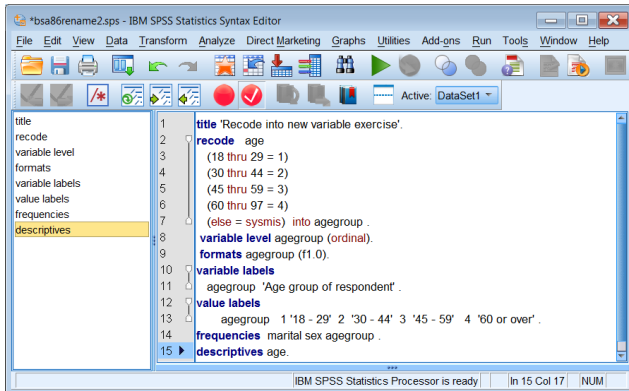
.. and click on **Save** :



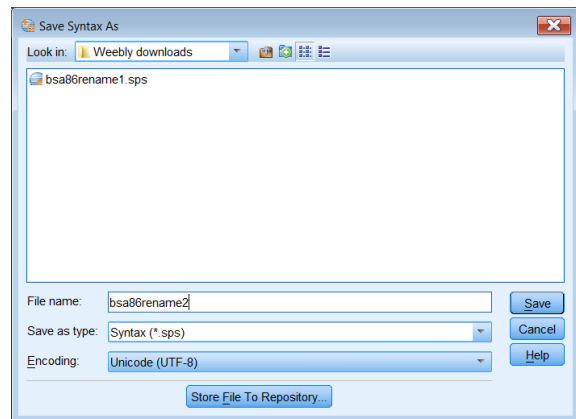
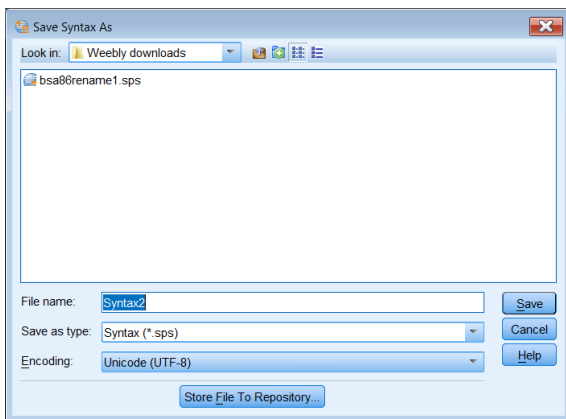
You also need to save the syntax file as **bsa86rename2.sps**. You should now be able to do this bit by yourself, but in case you can't (or won't):

From your syntax file:

File > Save as:

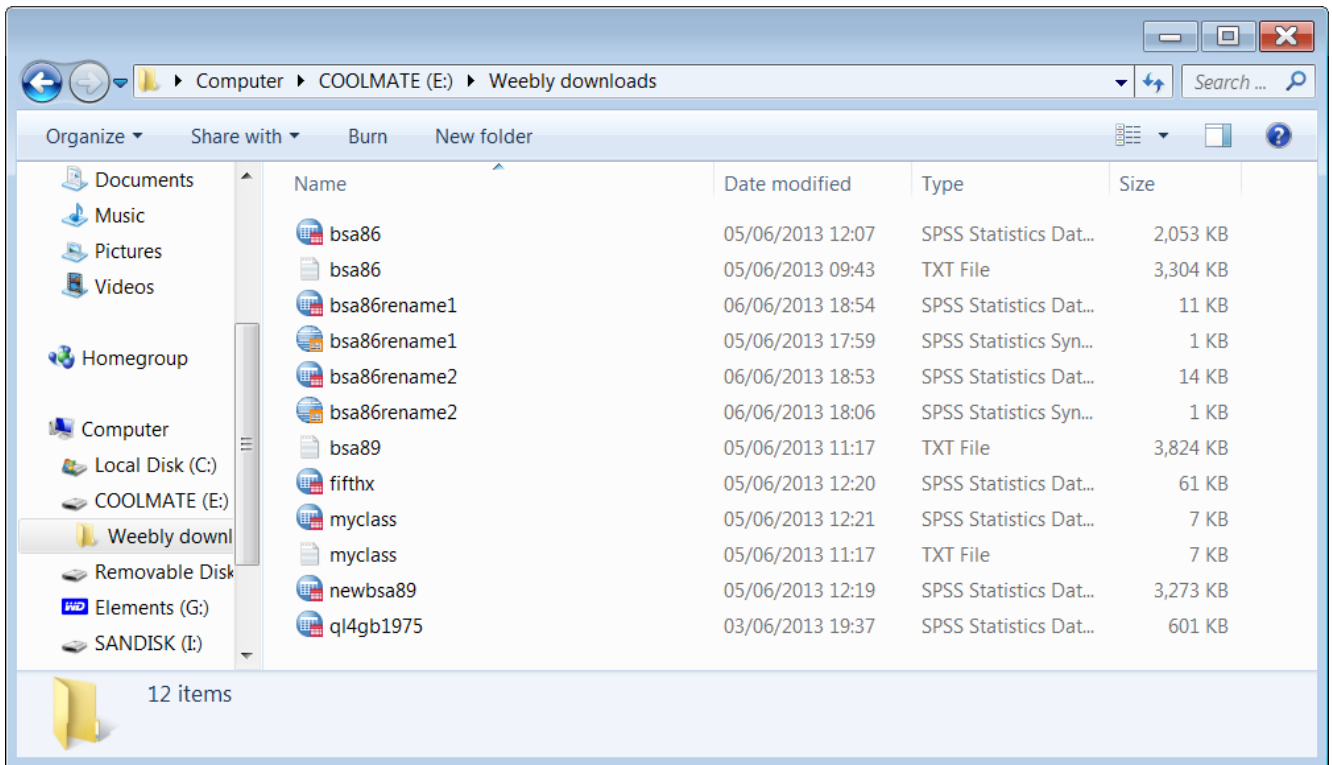


Change the name in the File name box to **bsa86rename2.sps**



.. and click on **Save** :

Your folder **e:\weebly downloads** should now look something like this:



End of exercise 2.3.1.2a2

Next session: [2.3.1.2b1 Select, rename and recode homework](#)

Back to page: [2.3 Data transformations](#)

Back to page: [2.3.1.1 Data transformations](#)

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

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**.

These will be found in [Block 3: Analysing two variables \(and sometimes three\)](#)