

Block 2: Analysing one variable

2.3 Data transformations

2.3.1.2b2 Select, rename and recode specimen answer [Draft only: 7 June 2013]

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

Exemplar: British Social Attitudes 1989

- Task 1:
- a) Retrieve file [newbsa89.sav](#)
 - b) Select out demographic variables **v1408 v1411** and **v1412** and rename them.
 - c) Save the file as **bsa89rename1.sav** in folder **e:\weebly downloads**

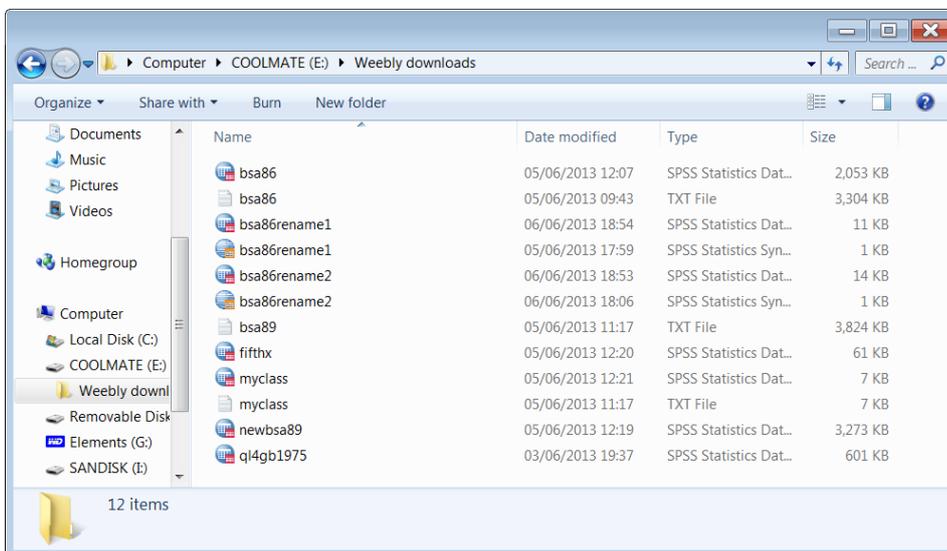
Task 2:

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

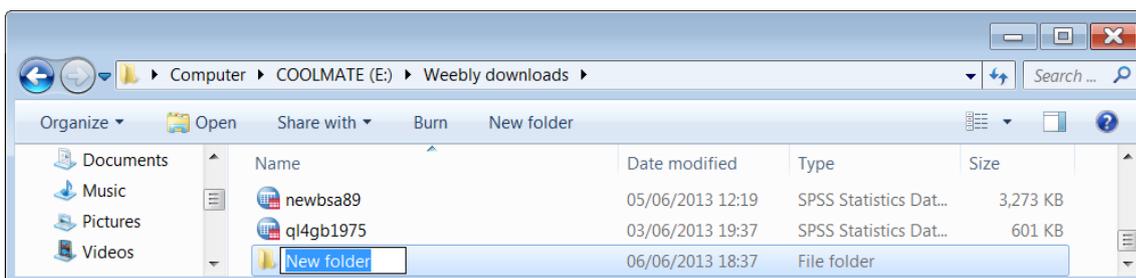
- b) Save the file as **bsa89rename2.sav** in folder **e:\weebly downloads**

Preliminary step: More housekeeping

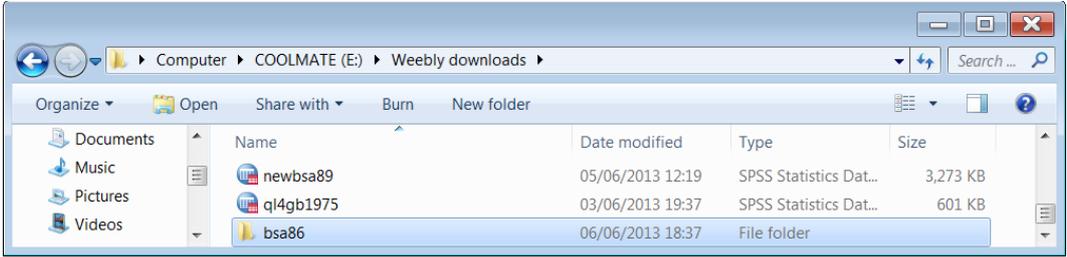
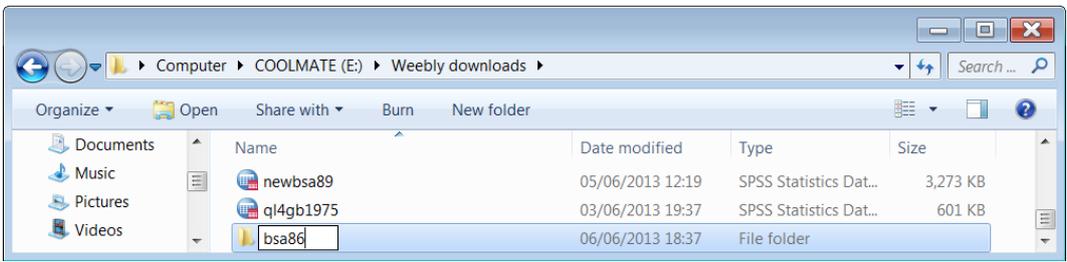
Navigate to folder **e:\weebly downloads**



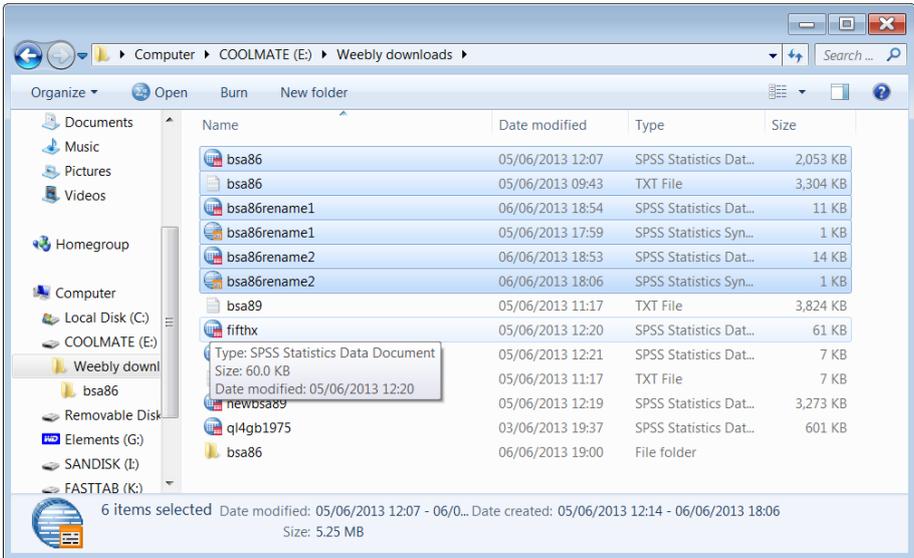
This is beginning to look a bit cluttered so we need to create a new folder **bsa86** to hold all the **bsa86***** files: Click on **New folder**:



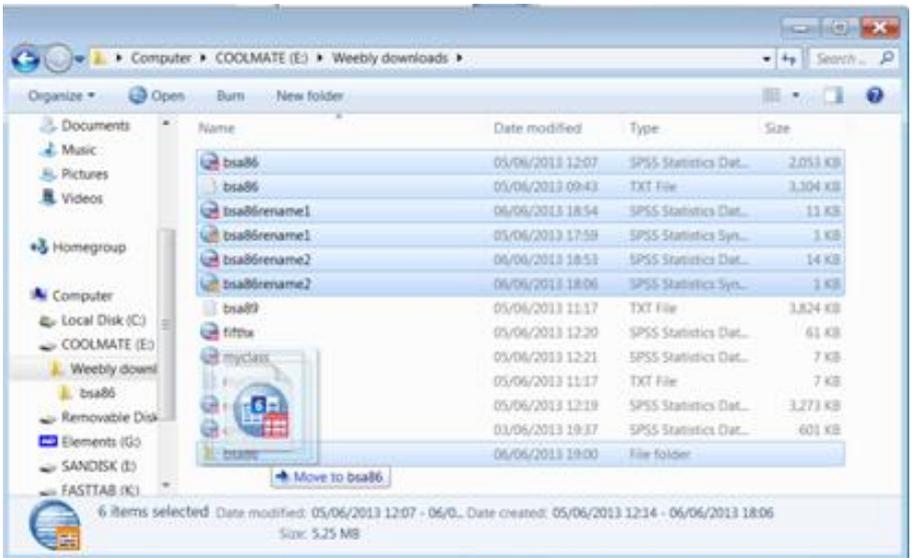
.. and call it **bsa86**:

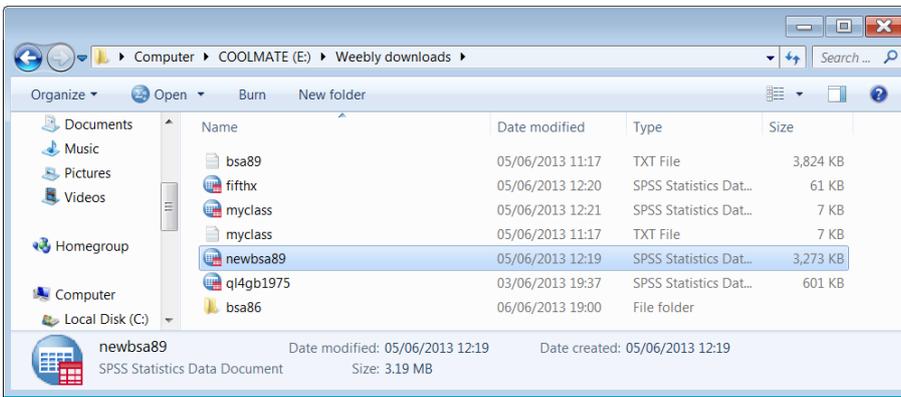


Highlight the **bsa86****** files



and drag them into folder **bsa86**





Task 1: Rename demographic variables

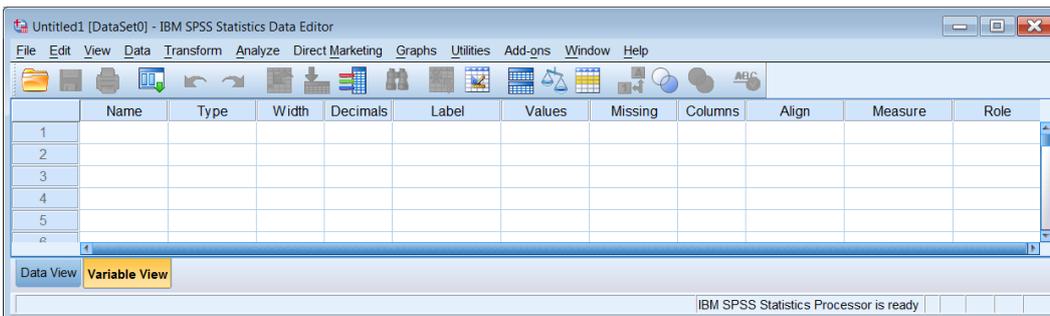
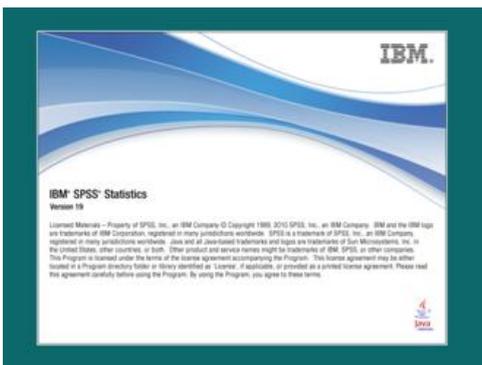
Old name	New name
v1408	marital
v1411	sex
v1412	age

If you navigate to drive **e:** **e:\weebly downloads** and open file **newbsa89.sav** by double-clicking you will get all 631 variables, but we only want three variables, **v1510**, **v1511** and **v1512**. To limit the number of variables in the active file, use instead the SPSS command:

GET¹ command with **/KEEP ~ ~ ~**.

Click on  to open SPSS:

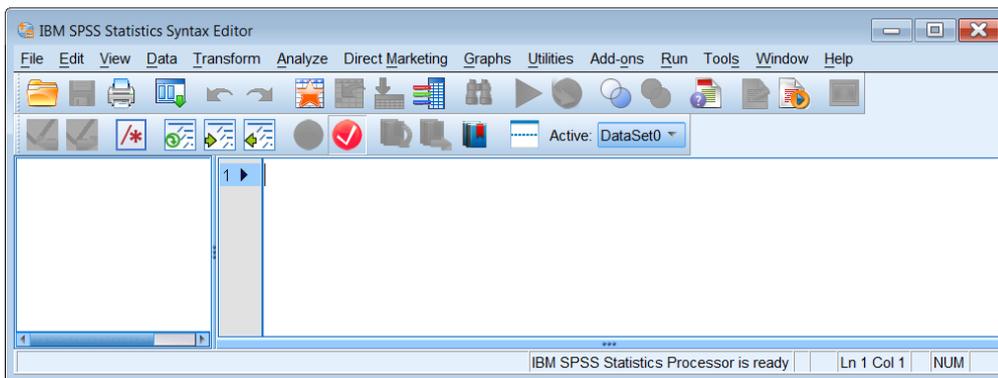
... then click on **Cancel**:



[Untitled1 Dataset0]

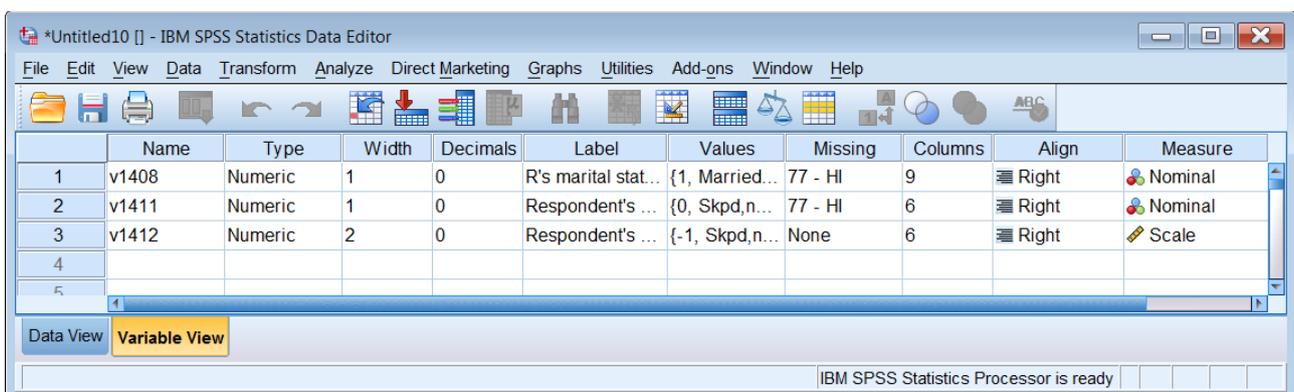
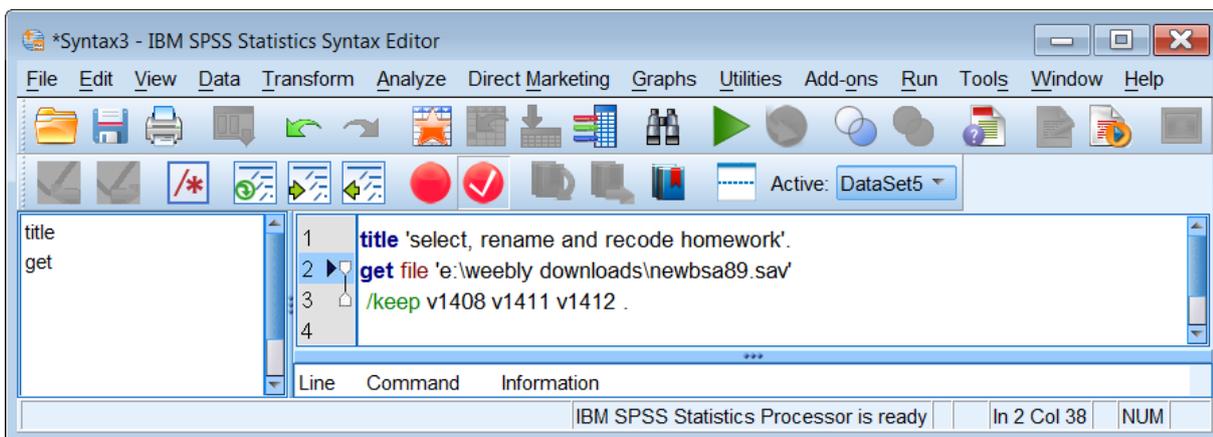
¹ format: **GET** '<filename>'. **/KEEP** <varlist>.

If your SPSS settings don't open a new Syntax Editor on start-up, click on **File** > **New** > **Syntax** to open a new Syntax Editor:



Type in:

```
title 'select, rename and recode homework'.  
get file 'e:\weebly downloads\newbsa89.sav'  
/keep v1408 v1411 v1412 .
```

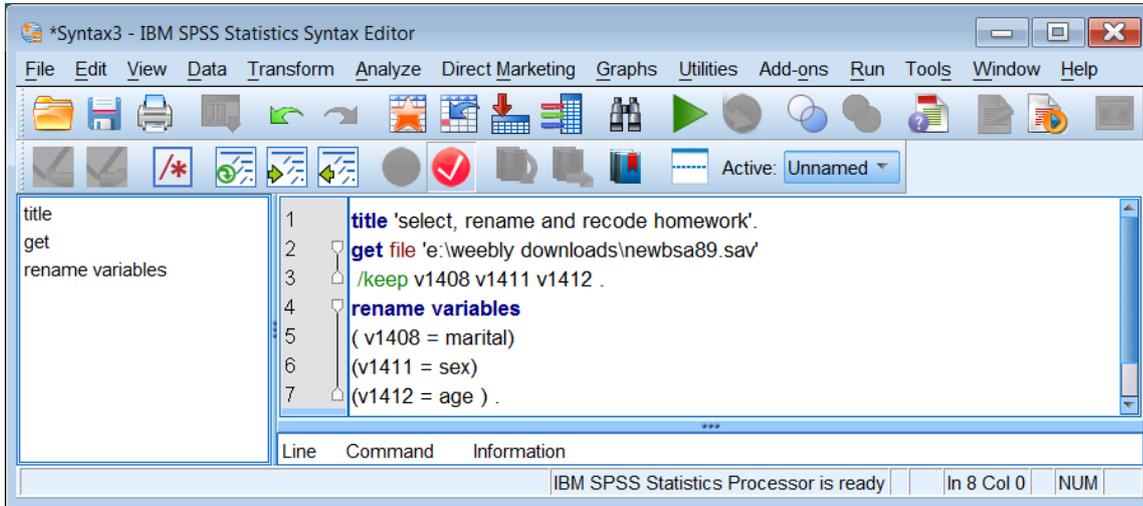


Despite my predilection for **positional** variable names, certain demographic variables used repeatedly in data analysis merit **mnemonic** names. Thus **marital**, **sex** and **age** are much more convenient to remember than **v1411**, **v1408** and **v1412**.

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**.

In the **Syntax Editor** type in:

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

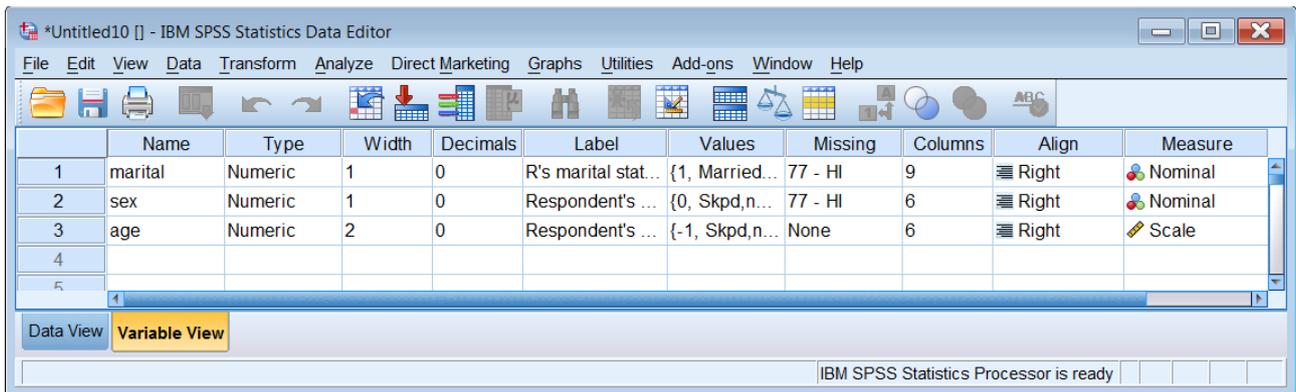


[NB: This command can also be written:

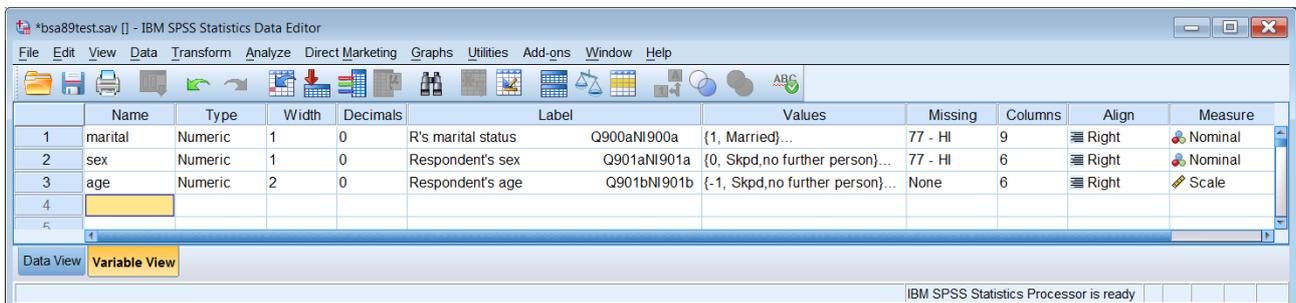
```
rename variables
  ( v1408 v1411 v1412 = marital sex age) .
```

... but make sure there are the same number of (implied) variables each side of the = sign.]

Run the rename command to get:



Drag the column separators to see the labels more clearly:



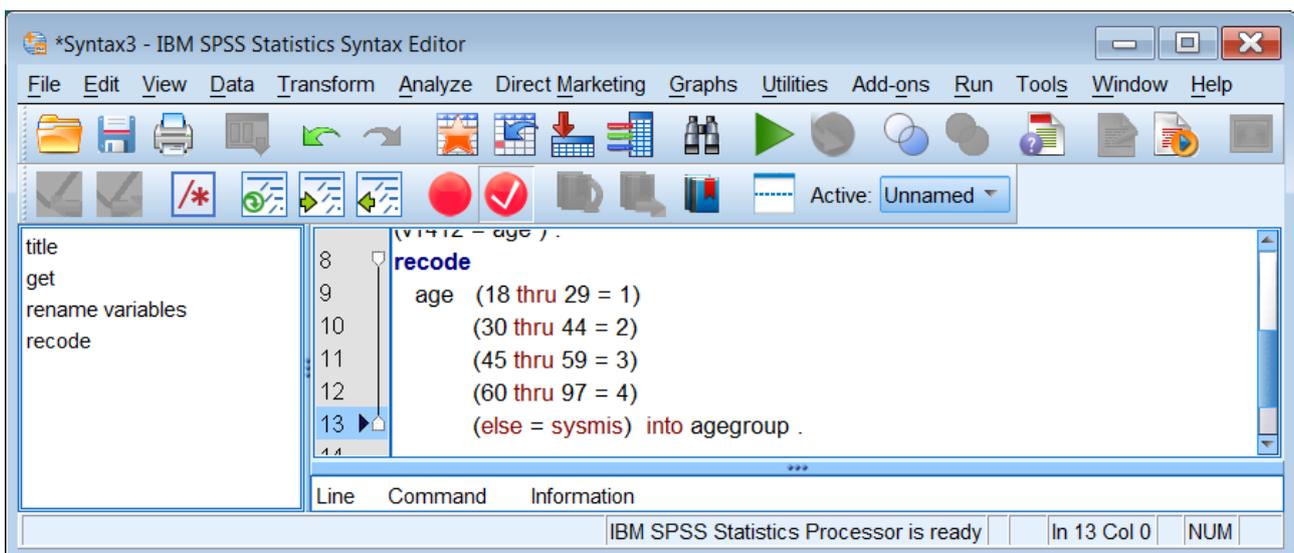
The existing variable labels are a bit untidy, but you can sort that out later.

	Label
R's marital status	Q900aNI900a
Respondent's sex	Q901aNI901a
Respondent's age	Q901bNI901b

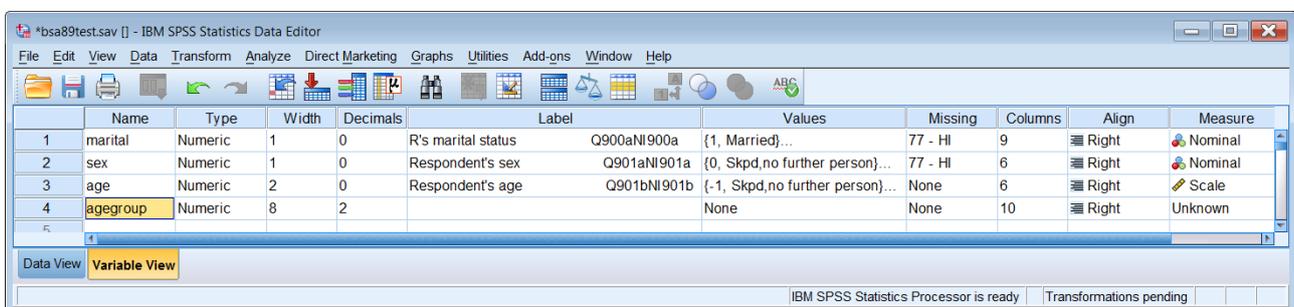
Task 2: Create a new variable **agegroup** by recoding **age** into four groups, 18 -29, 30 -44, 45 -59, 60 and over.

recode

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

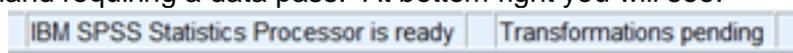


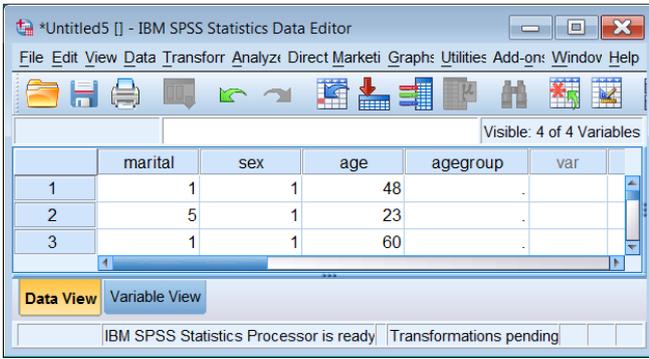
Run the job to get:



Note that for **agegroup** Measure is displayed as **Unknown** and Decimals as **2**.

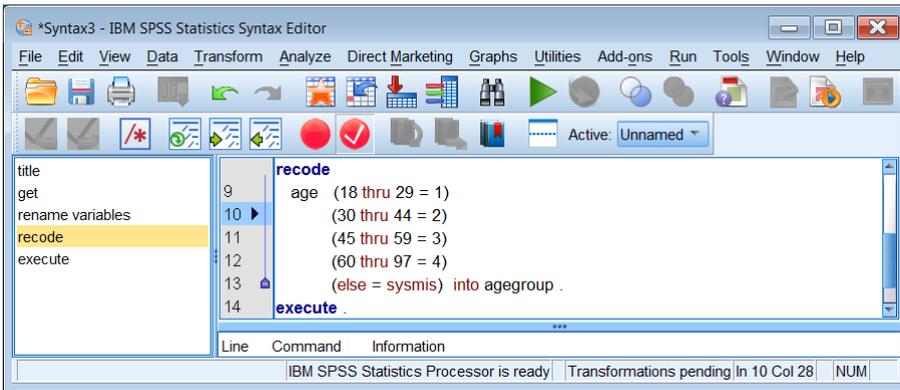
If you switch to **Data View**, you will see that there are no values entered for **agegroup**. This is because SPSS is waiting for a command requiring a data pass. At bottom right you will see:



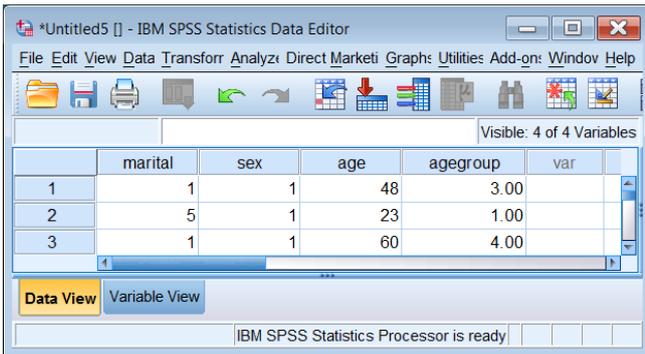


Go back to the Syntax Editor and type in :

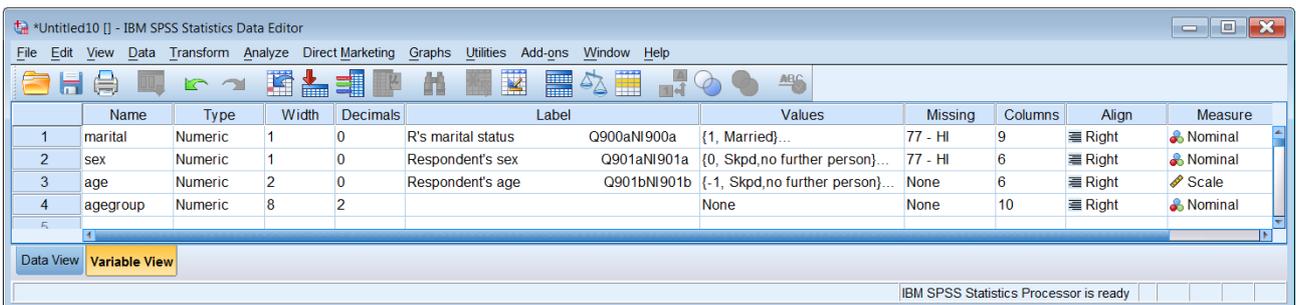
execute .

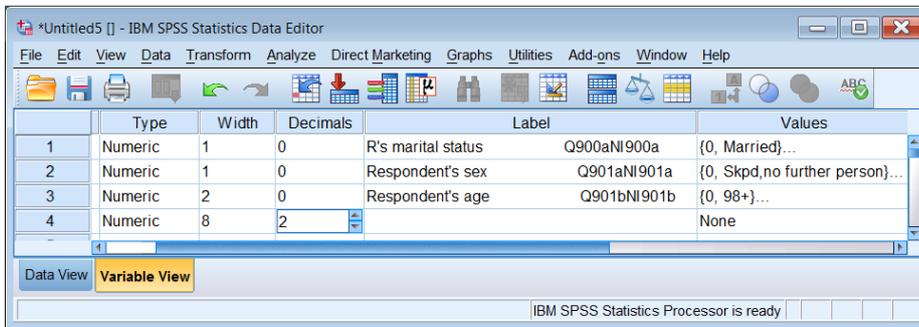


. . then press the green triangle ► to get:

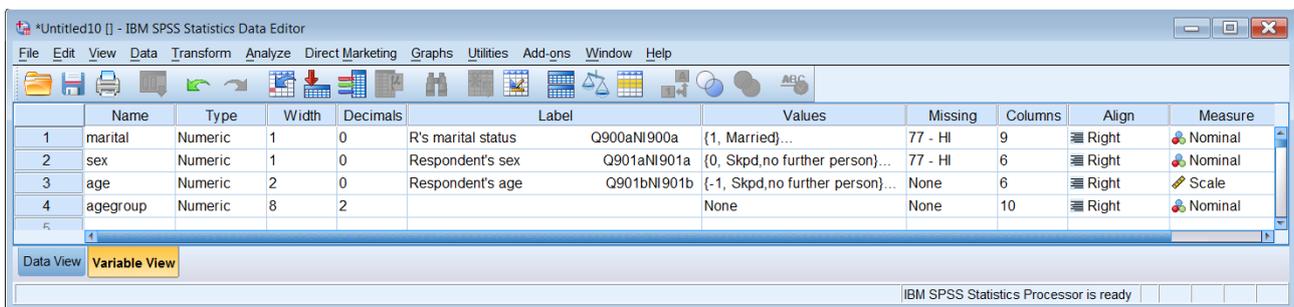
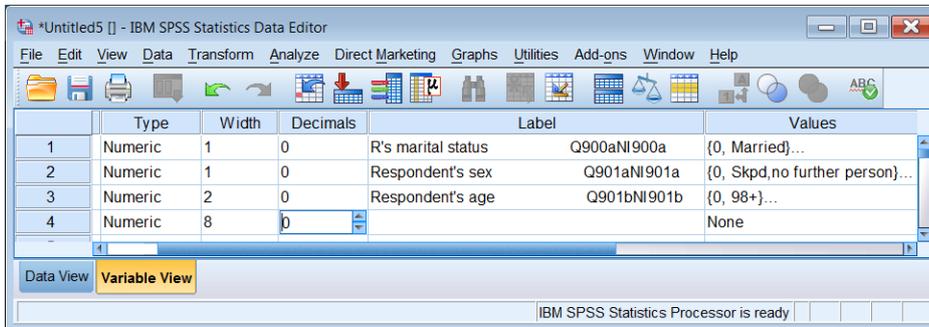


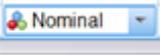
Values have now been entered for **agegroup**, but there are two superfluous decimal places. You can get rid of these manually in the **Data Editor** by clicking on the cell under **Decimals**

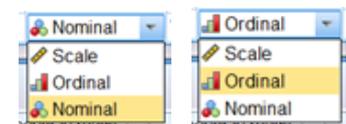




.. and changing the 2 to 0



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**:

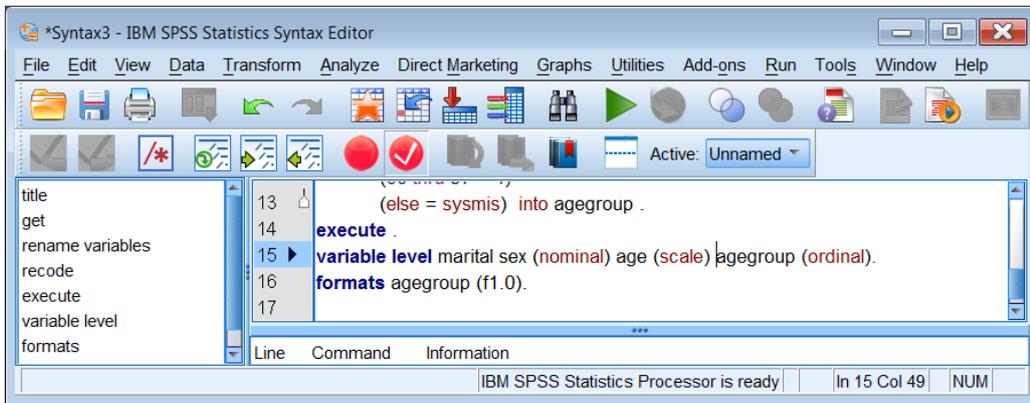


This is easy for just one variable, but cumbersome for many and doesn't save any syntax for possible later use. It's better practice to use syntax to set measurement levels and formats:

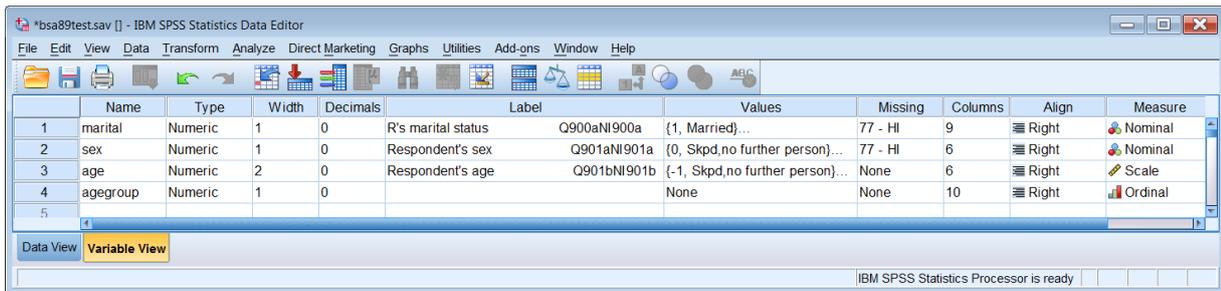
Use the **VARIABLE LEVEL** and **FORMATS** commands

Go back to your **Syntax Editor** and type in:

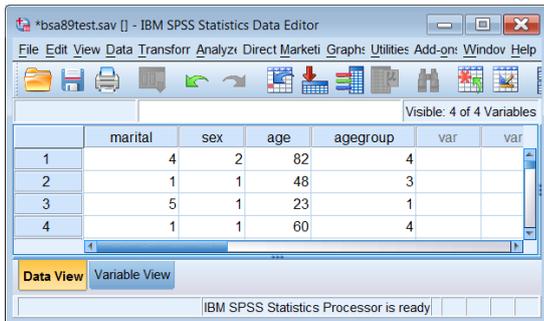
**variable level marital sex (nominal) age (scale) agegroup (ordinal).
formats agegroup (f1.0).**



Run to get:



If you switch to **Data View**, you will see that the values for **agegroup** are now all integer (no decimal places).



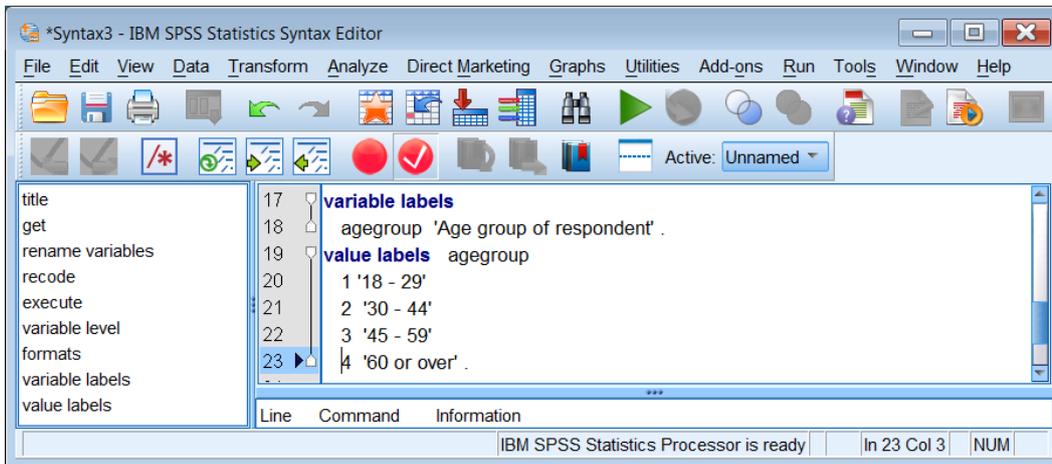
Add data dictionary information:

variable labels

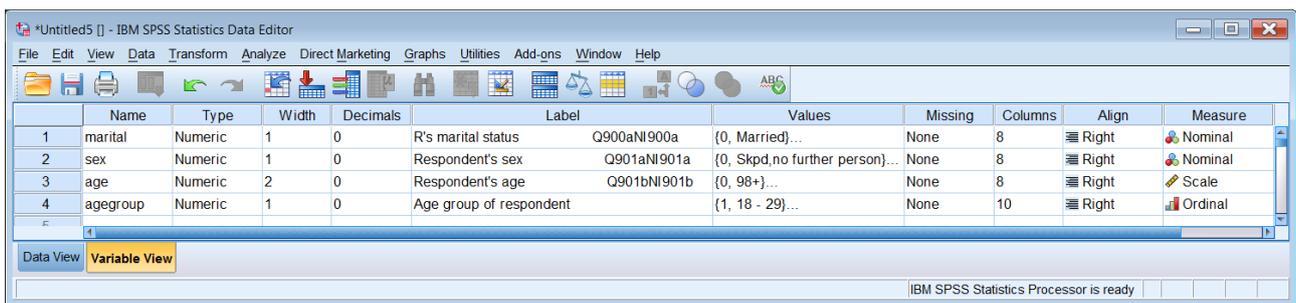
agegroup 'Age group of respondent' .

value labels

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

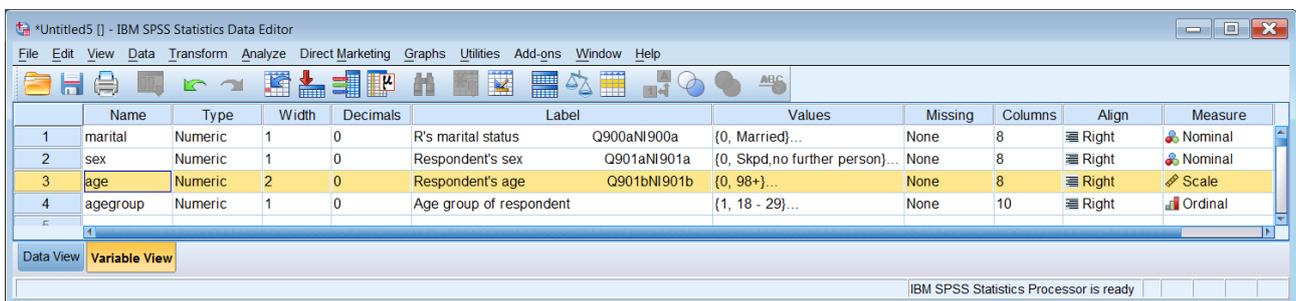


Run both new commands: variable and value labels have now been added to **agegroup**:

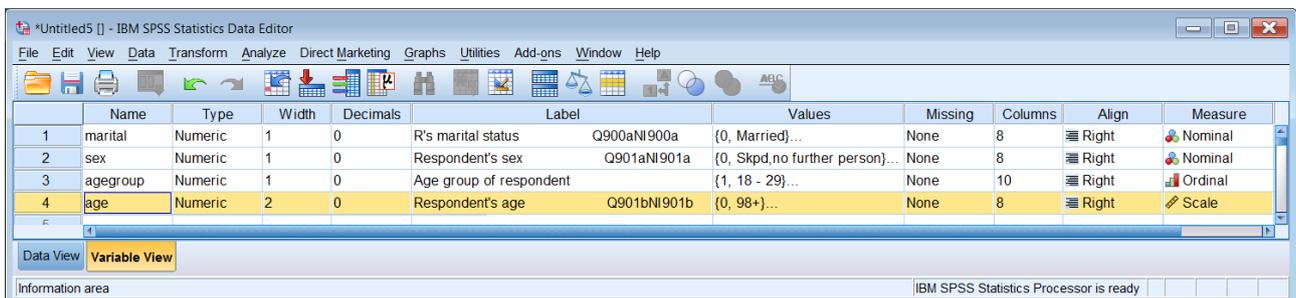


The file now needs tidying up a bit. Since analyses with demographic variables are often specified using **<varname> to <varname>** as in **marital to agegroup** we need to make sure that variables with many values are outside this block. Variable **age** needs to be moved to the end of the file, well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

Left click on row number 3:

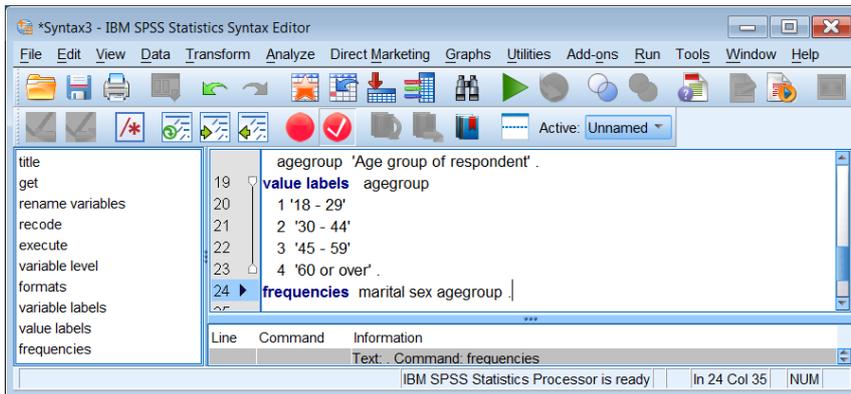


Left click on 3 again, but **hold the button down** and drag **age** to the bottom of the file:



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 .



Run the command to get:

Statistics

		R's marital status Q900aNI900a	Respondent's sex Q901aNI901a	Age group of respondent
N	Valid	3024	3025	3015
	Missing	1	0	10

There are 3025 cases in the file. There is one missing case for **marital**, no missing cases for **sex**, and 10 cases where **agegroup** is missing.

Frequency Tables

R's marital status Q900aNI900a

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Married	1948	64.4	64.4	64.4
	Livng as married	114	3.8	3.8	68.2
	Separtd/divorced	183	6.0	6.1	74.2
	Widowed	276	9.1	9.1	83.4
	Not married	503	16.6	16.6	100.0
Total		3024	100.0	100.0	
Missing	99	1	.0		
Total		3025	100.0		

Respondent's sex Q901aNI901a

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	1393	46.0	46.0	46.0
	Female	1632	54.0	54.0	100.0
Total		3025	100.0	100.0	

Age group of respondent

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

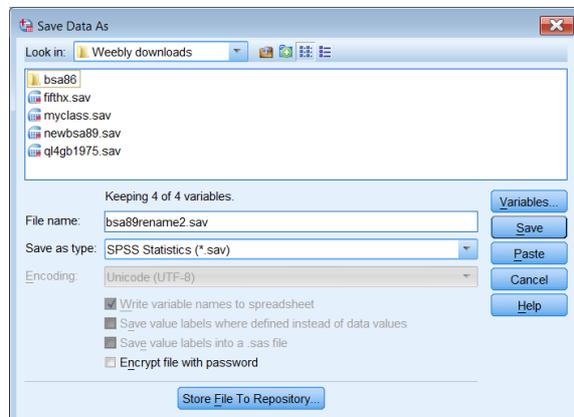
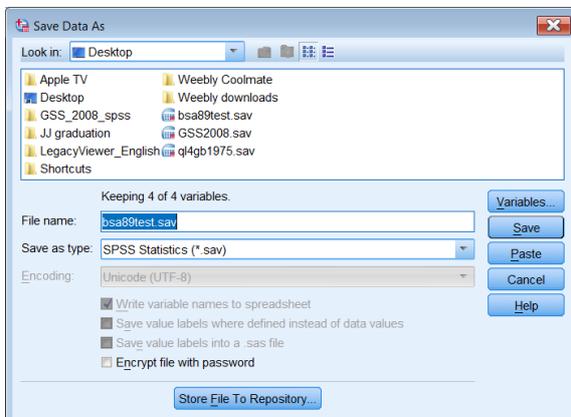
If we had left **age** where it was and used:

frequencies marital to agegroup .

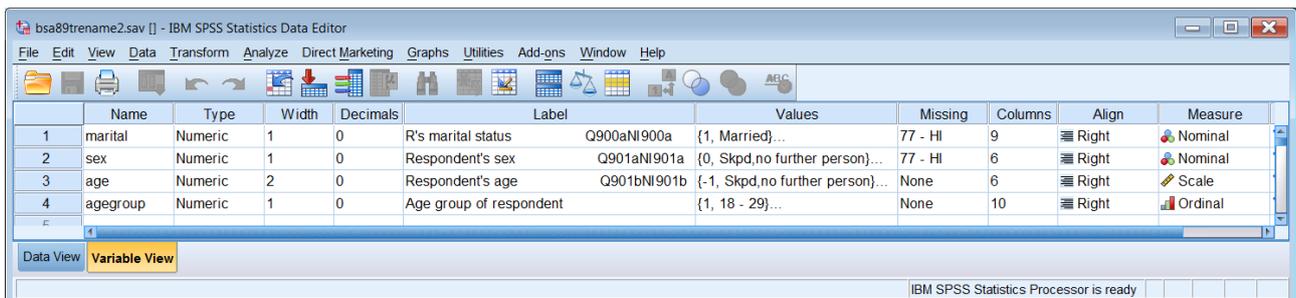
.. there would have been a very large frequency table for age as well !!

Save the new version of the file as **bsa89rename.sav**.

File > Save as:

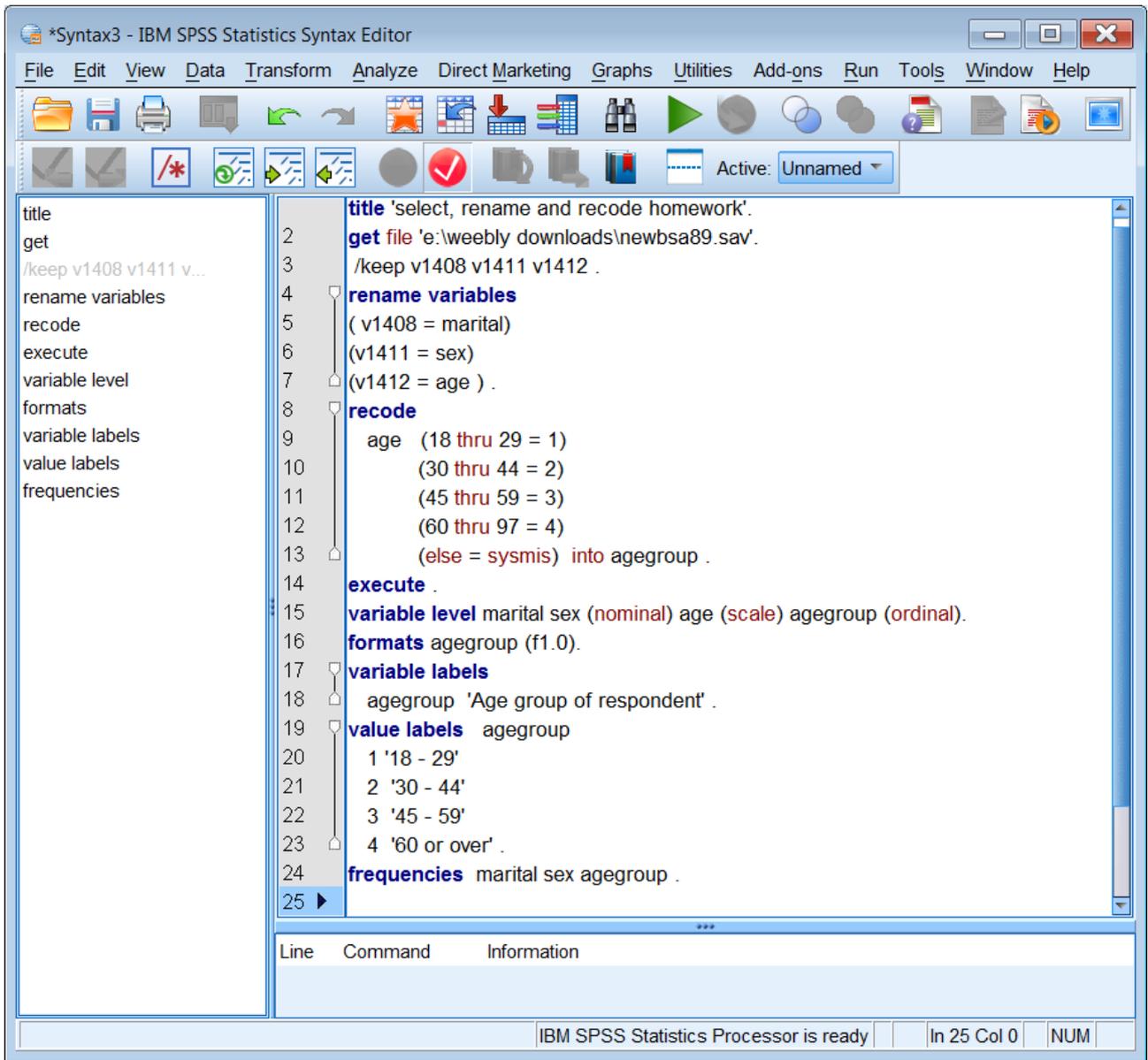


.. and click on **Save** :

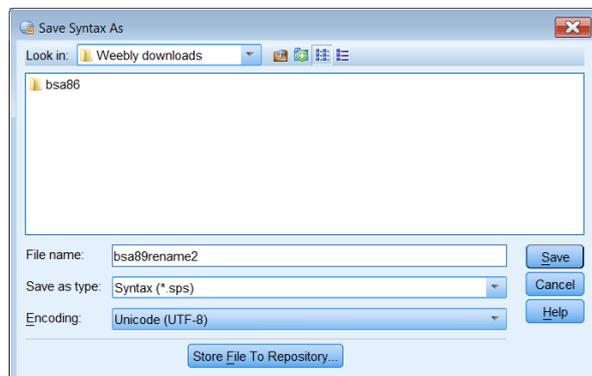
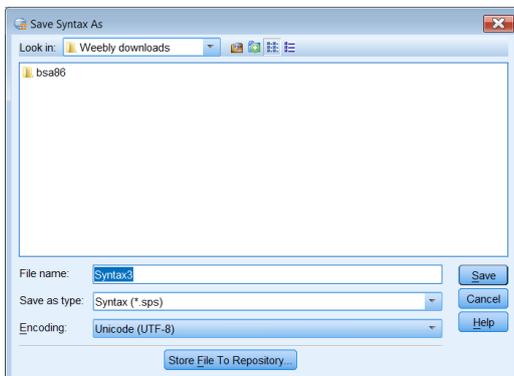


bsa89rename2.sav

You should also save the syntax file as **bsa89rename.sps**:



You can do this bit by yourself! . . . but if you can't . . .

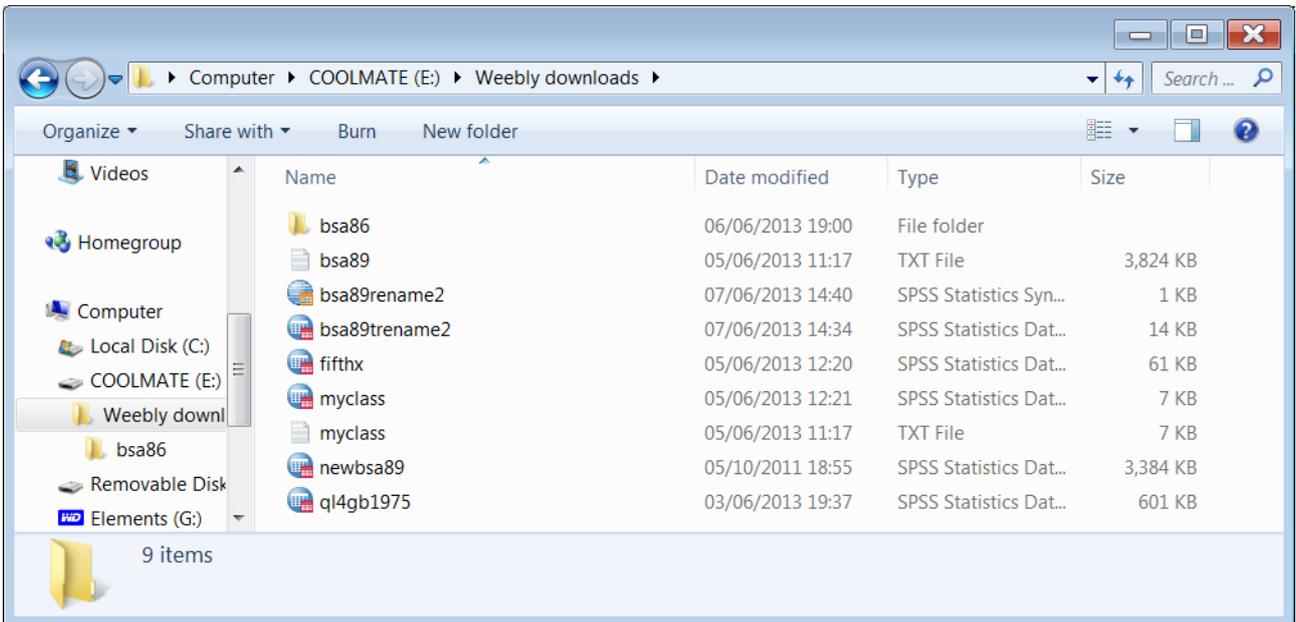
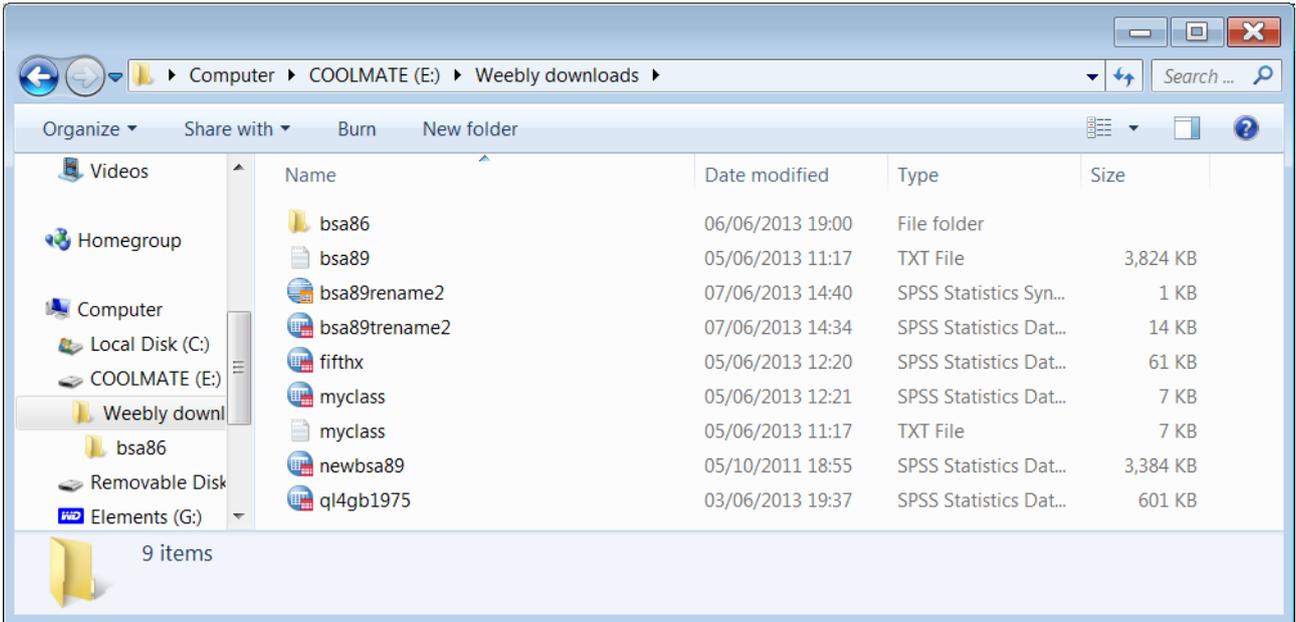


Change Syntax ** to **bsa89rename** and press **Save**.

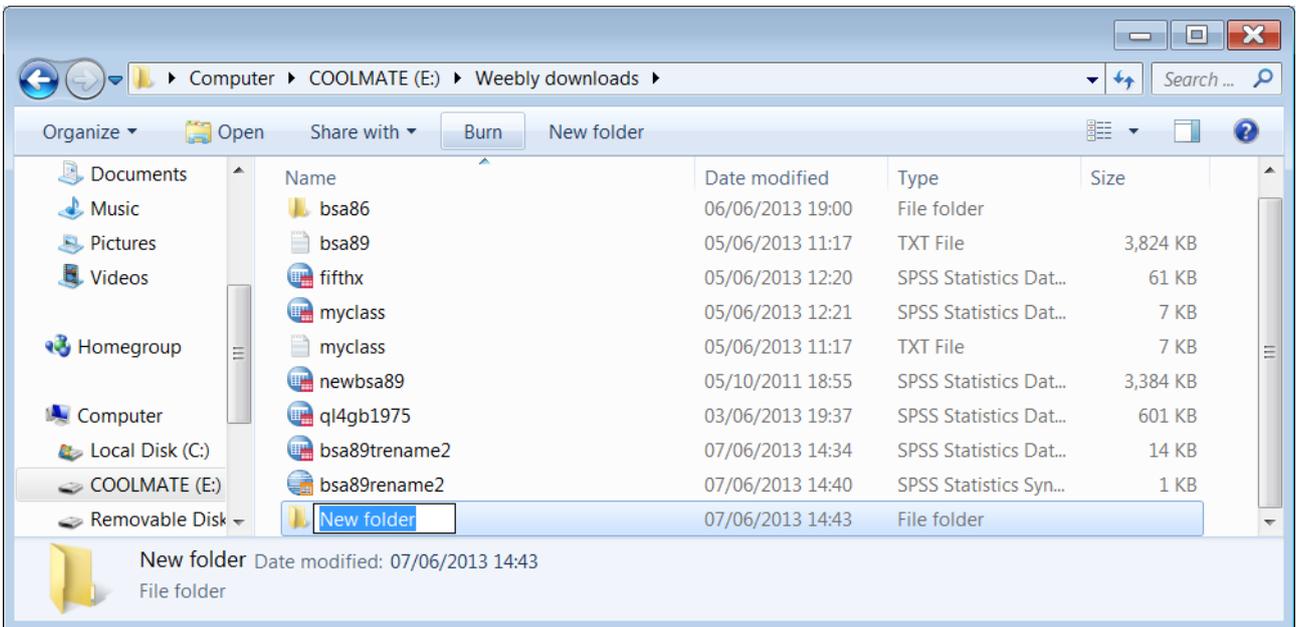
Alternatively you can use the SPSS **SAVE** command:

save out 'e:\weebly downloads\bsa89rename.sav'.

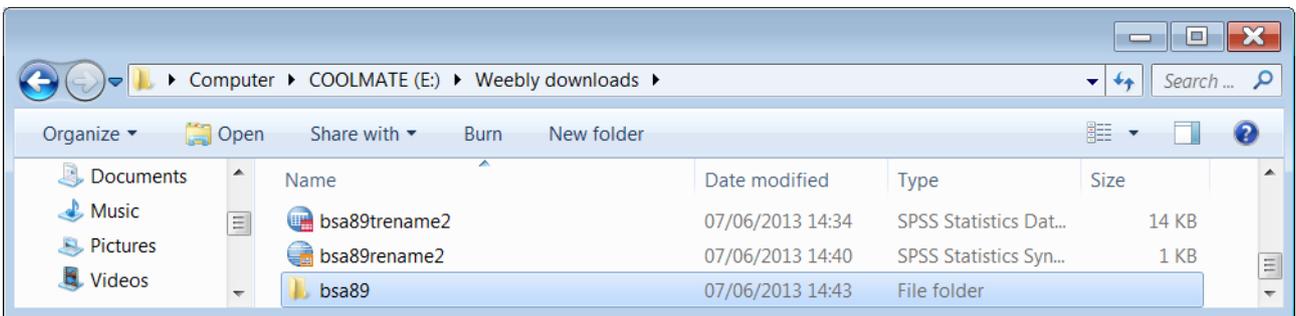
Your folder **e:\weebly downloads** should now look like this.



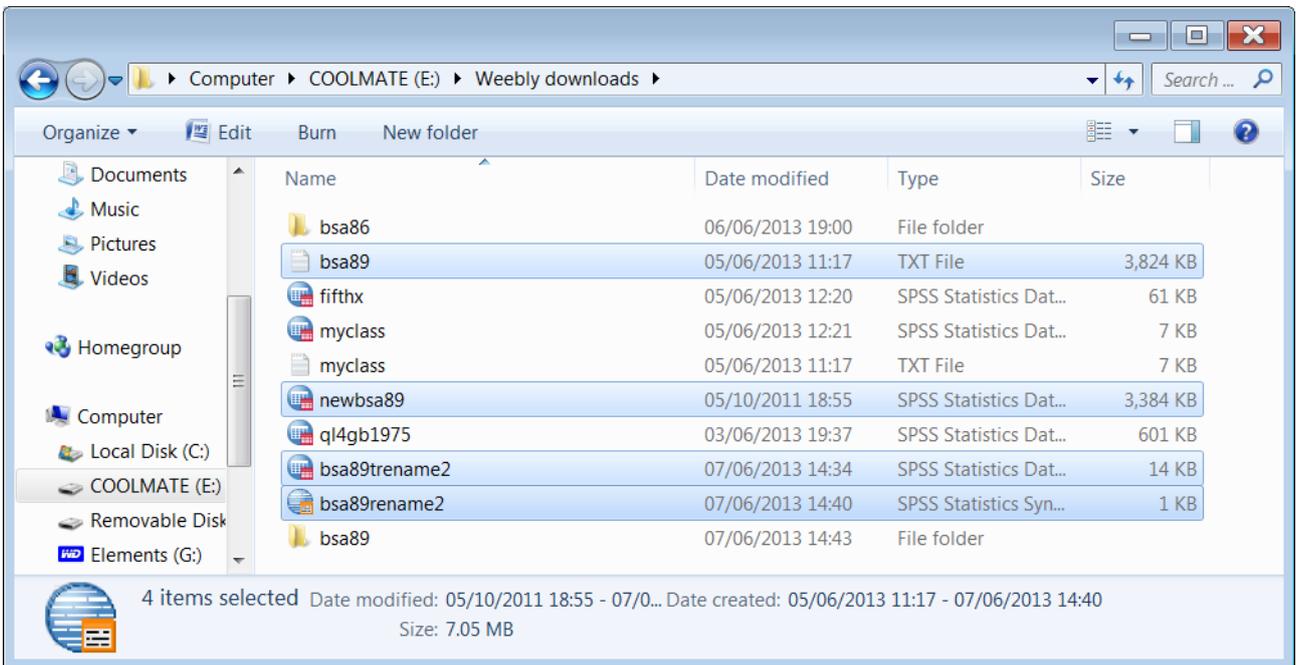
This folder could also now do with a bit of house-keeping to create a sub-folder for the **bsa89***** files. Click on **New folder** :

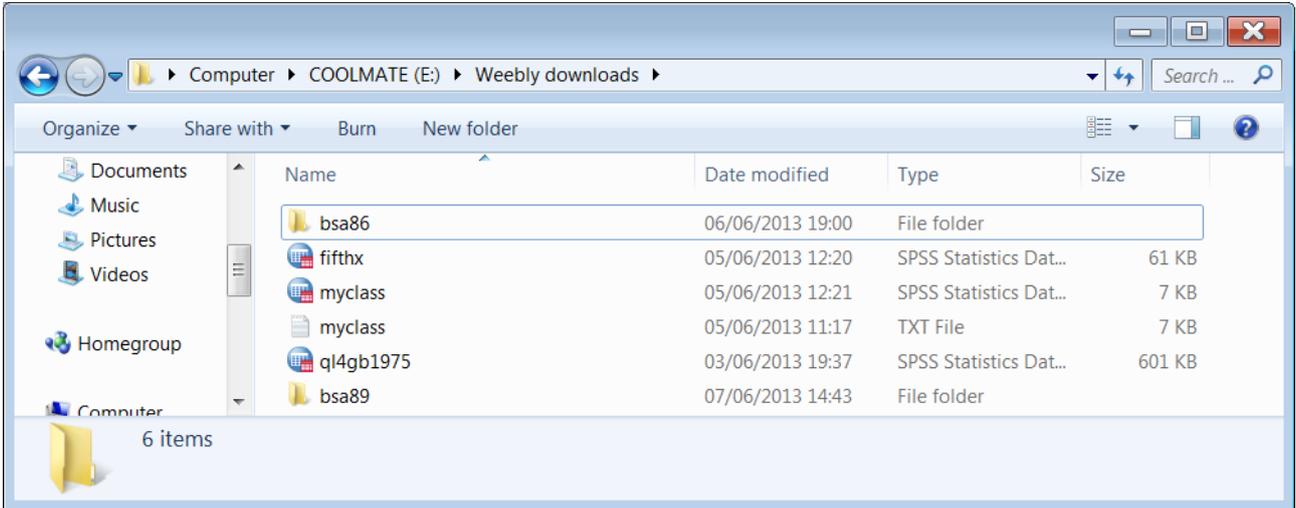


.. and call it **bsa89**.

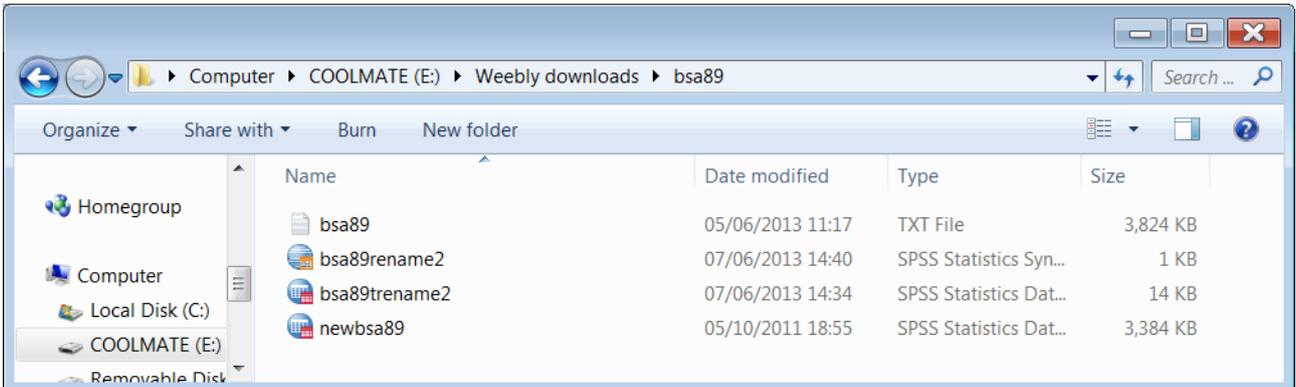


Highlight all the **bsa89****** files and drag them to the new folder:





Folder **e:\weebly downloads\bsa89** should now look like this:



End of 2.3.1.2b Rename and recode homework

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 page [2.3 Data transformations](#)

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

Forward to page: [Block 3: Analysing two variables \(and sometimes three\)](#)