# Survey Analysis Workshop                    Copyright © 2013 John F Hall

## Block 2: Analysing one variable          2.3  Data transformations

### 2.3.1.2a1  Select and rename variables exercise          [Draft only: 5 June 2013]

**Previous sessions:**  2.3.0  Preliminary file downloads
2.3.1.1  Data transformations

**Exemplar:**    British Social Attitudes 1986

**File:**    bsa86.sav

**New commands:**    **GET FILE** . . . /KEEP
**RENAME VARIABLES**
Utilities > Variables
Data > Define Variable Properties

**Task:**    a) Retrieve file bsa86.sav
b) Select out demographic variables **v1510 v1511** and **v1512** and rename them.

| Old name | New name |
|----------|----------|
| **v1510** | **marital** |
| **v1511** | **sex** |
| **v1512** | **age** |

c) Correct the measurement levels
d) Add variable and value labels.

_____

### Step 1:  Retrieve the SPSS saved file.

If you have already downloaded file bsa86.sav from this site and saved it to a memory stick as
e:\weebly downloads\bsa86.sav skip to **Step 2** on page 4.  Otherwise download it now.

Go to page 2.3.0  Preliminary file downloads, right-click on file **bsa86.sav** and follow instructions to:
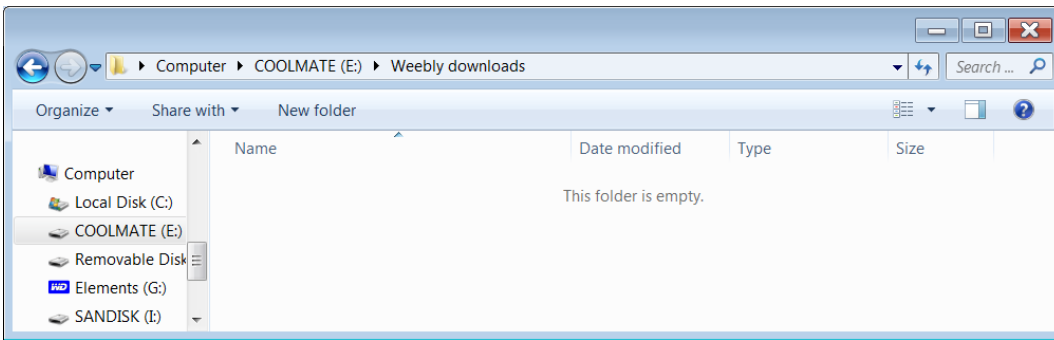
Open in a New Window  .

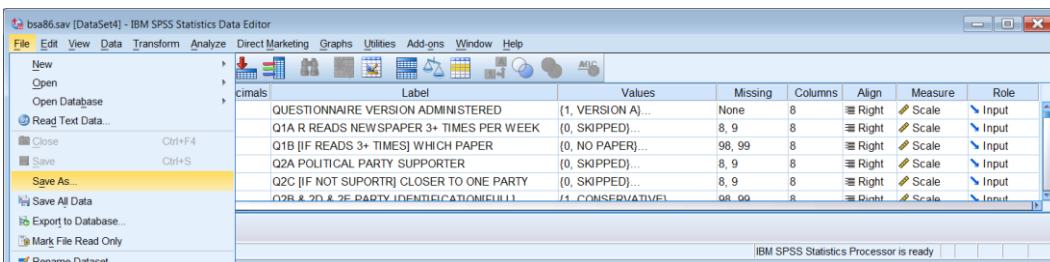It will probably download  in  Data View  so change to   Variable View .
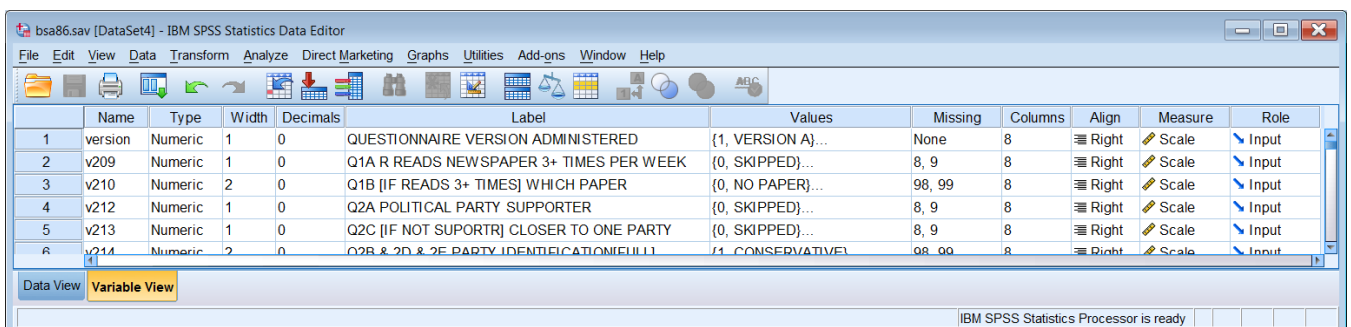


**bsa86.sav** as downloaded

Explore the file if you want to, then:

1

File **bsa86.sav** is 2.05 mb.  Insert a stick with sufficient capacity in drive **e:** and create a new folder **weebly downloads**.  Check that destination folder **e:weebly downloads** is empty:
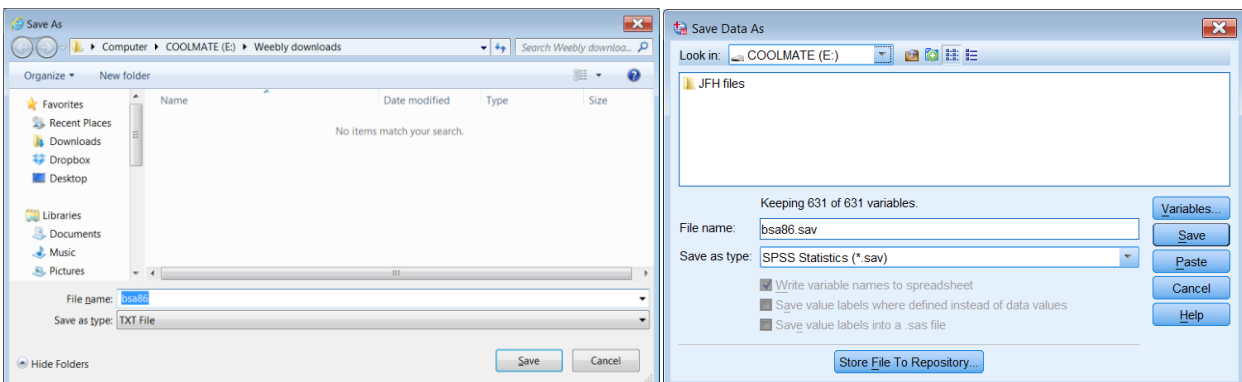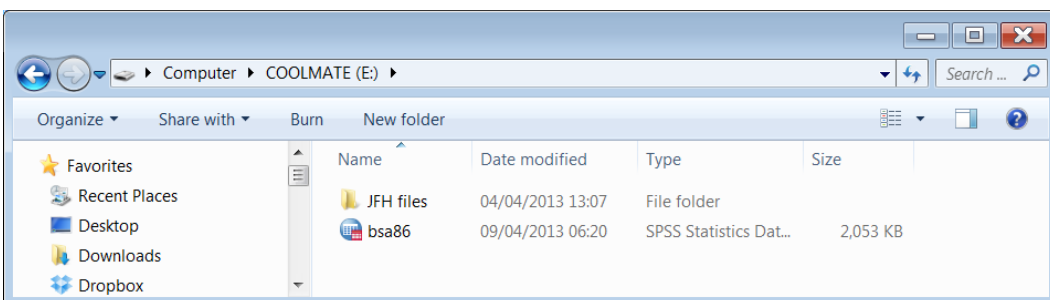


Go back to the SPSS Data Editor:



File > Save as:

and navigate to **e:weebly downloads**:



Check that the file name is bsa86.sav , then click on Save:



Close file **bsa86.sav** and exit from SPSS.

**End of Step 1.**

**[Note]**

**Variable names and labels**

I have renamed the variables in **bsa86.sav** (the saved file for the 1986 wave in 2.3.0 Preliminary file downloads) from the original **mnemonic** names to my preferred **positional** variable names as this makes it easier to work direct from the original questionnaire (see:1.3.1 Conventions for naming variables in SPSS). However, the file still has the original UPPER CASE variable labels and value labels written in 1986 by John Curtice and Ann Mair using a much earlier version of SPSS. Given the limit in those days of 40 characters for variable labels and 20 for value labels (only 16 for column headers) writing them then was quite tortuous (as is reading them now).

In later versions of SPSS the limits on number of characters in labels have been almost totally relaxed, so it's best to build up your own SPSS syntax and saved file(s) with Mixed Case labels, as in the tutorials, exercises and specimen answers. You should try the exercises yourself before looking at my specimen *.sps files. They are there as a crib (or as backup) not for you to copy and run, otherwise there's no point in you taking the course. See entries 2.2.1.2 to 2.1.2.5

**Downloaded *.sav and *.txt files**

The screenshots in the following steps assume that all files were downloaded to **weebly downloads** on a USB stick in drive e: as per instructions in 2.3.0 Preliminary file downloads

If all files have been correctly downloaded, folder e:weebly downloads (sorted by Type) should look something like like this:
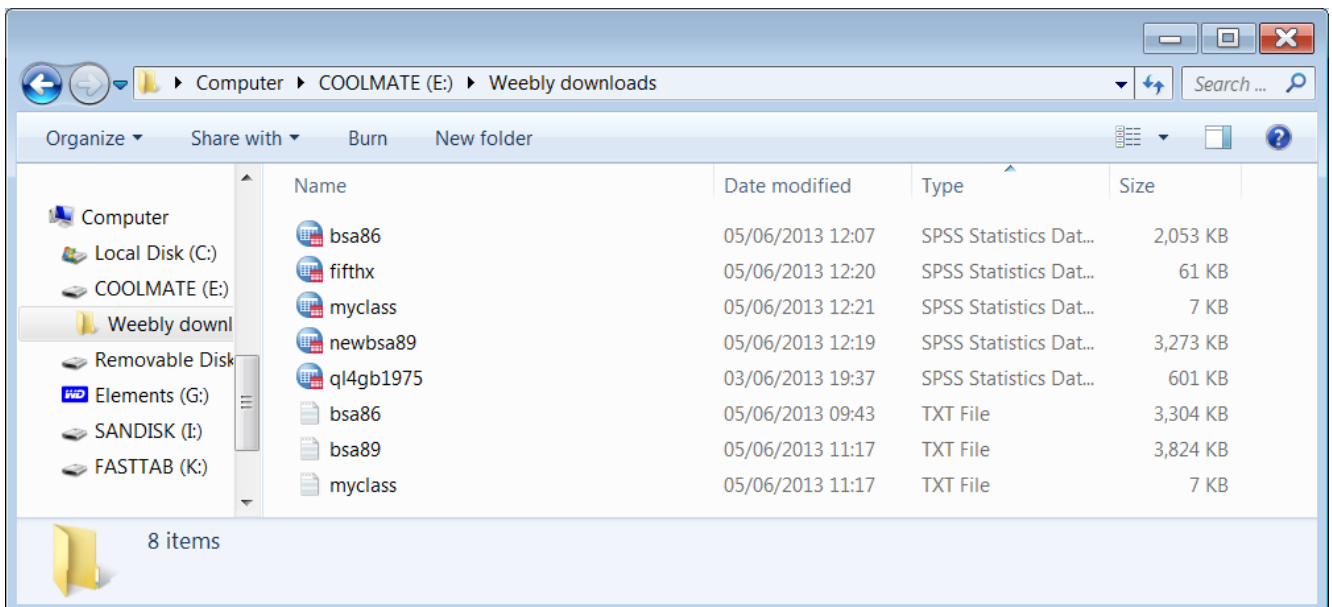


More files will be added from subsequent tutorials and exercises, but things can be kept manageable by creating sub-folders for each study and/or for each topic.

**Step 2:  Selecting variables for analysis**

The following assumes you are starting SPSS from scratch.

If you navigate to drive **e:weebly downloads** and open file **bsa86.sav** by double-clicking you will get all 631 variables, but we only want three, **v1510 v1511** and **v1512**.  To limit the number of variables in the active file, we use instead the SPSS **GET**[1] command:
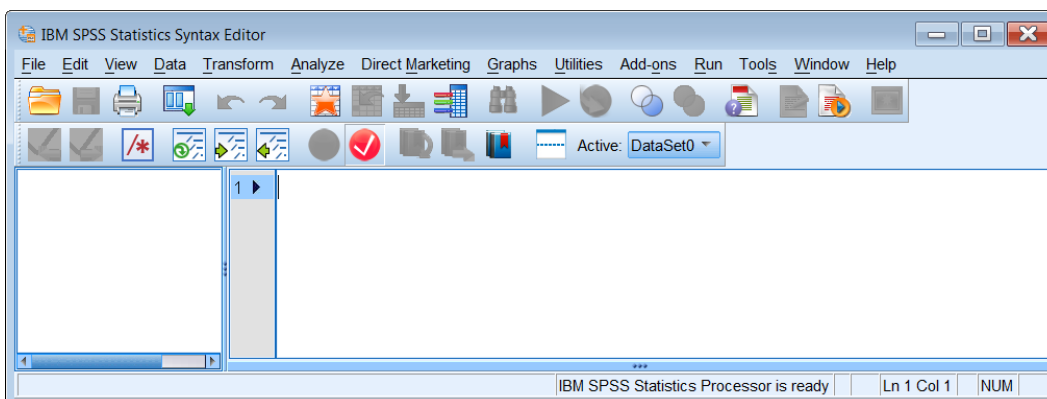
Click on  to open SPSS:                                          . . . and click on Cancel:
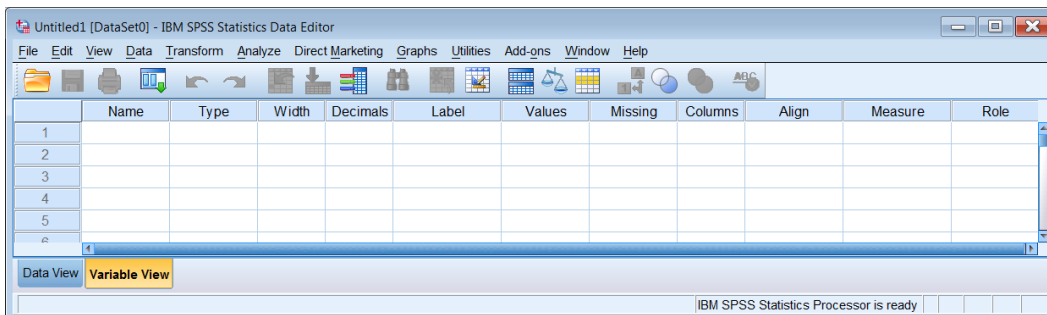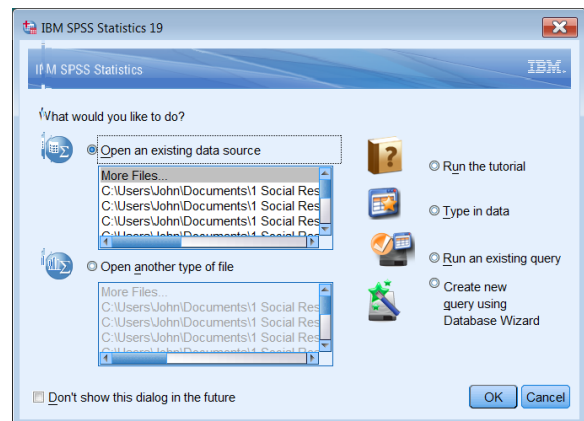
  



[Untitled1 Dataset0]

If your SPSS start-up settings don't open a new Syntax Editor on start-up, click on:

File > New > Syntax



---
[1]  **format:  GET '**<filename>**'.**and the subcommand /KEEP <varlist>.

  . . . eg:  **get file 'e:bsa86.sav' /keep v1510 v1511 v1512.**

4

. . . and then type in:                                    <span style="color:magenta">[Don't forget the primes and full stops!]</span>

> **title  'Select and rename variables exercise'.**
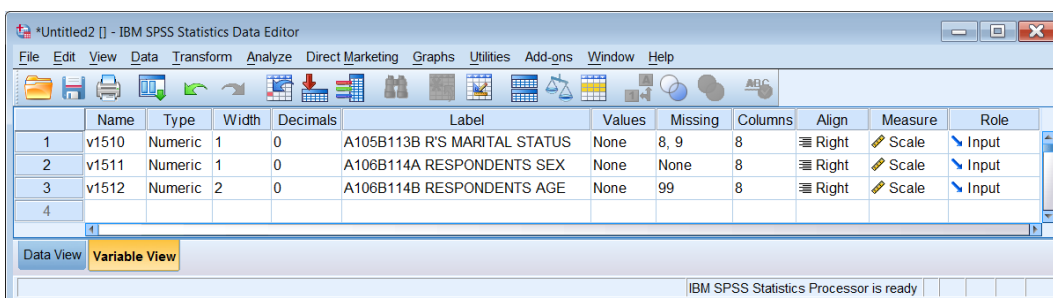> **get file 'e:bsa86.sav' /keep v1510 v1511 v1512.**
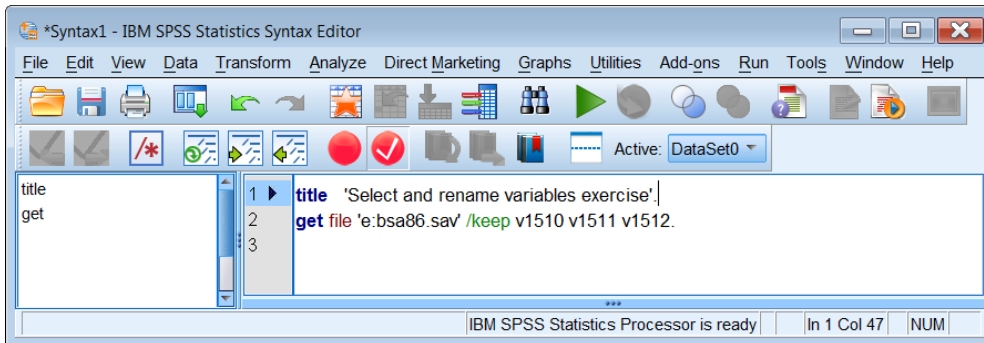


[File name changes to *Syntax1]

Click on Run > All:
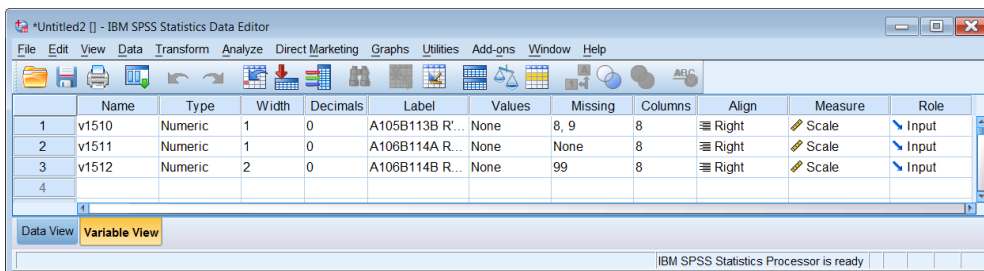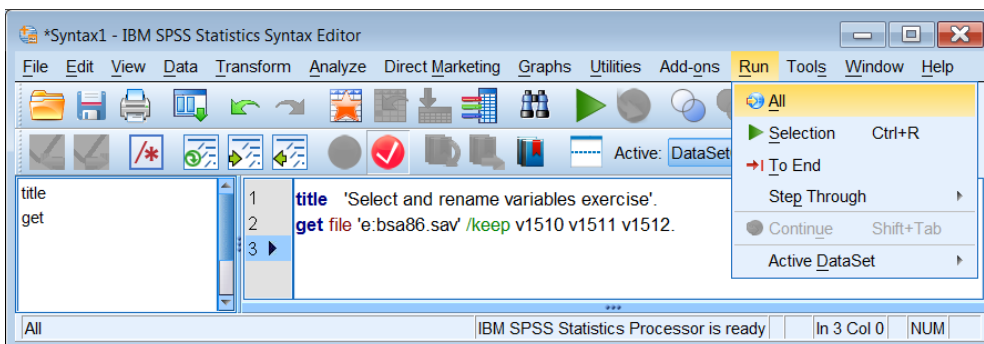




*Untitled2

Adjust the column separators to see the labels more clearly:



As you can see the variable labels are in UPPER CASE: they are also confusing because the question number was different in versions A and B.  The first part of the label combines the question numbers from each version.  There are no value labels, so these need to be added.

5

Don't blame me!  I didn't write the original SPSS setup files: they came from John Curtice and his team at Strathclyde University in 1973.  I've only changed the variable names to make the file easier to use with the questionnaire.

We also need to modify the measurement levels of **v1510** and **v1511**.  When reading the raw data, SPSS has automatically assigned SCALE to these variables because of the range of values found (1 -  99).  This can be done with syntax[2], but it's quicker in this case to do it manually in the Data Editor.

For each variable **v1510** and **v1511**, in the <span style="color:red">Measure</span> column, click on <mark>Scale</mark>, then click on <mark>Nominal</mark>
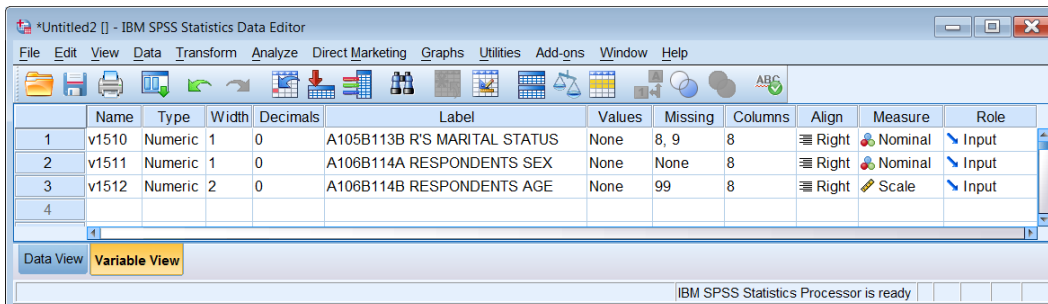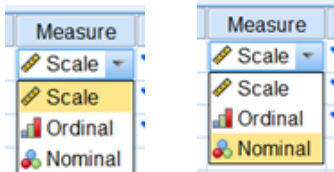




Despite my predilection for **positional** variable names, certain variables used repeatedly in data analysis merit **mnemonic** names.  For demographic variables, **marital, sex** and **age** are much more convenient to remember than **v1510 v1511** and **v1512**.  Also, for tabulation purposes, variables with many values, such as **age**, need to be grouped into fewer categories.  Standard practice is to keep frequently used 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**.

---

[2]    **VARIABLE LEVEL** <varname(s)> (<LEVEL>)
           eg   **VARIABLE LEVEL** v1510 v1511 (nominal).

**Step 3:          Rename demographic variables**

**RENAME VARIABLES**[3] changes the names of variables permanently.

Type in:

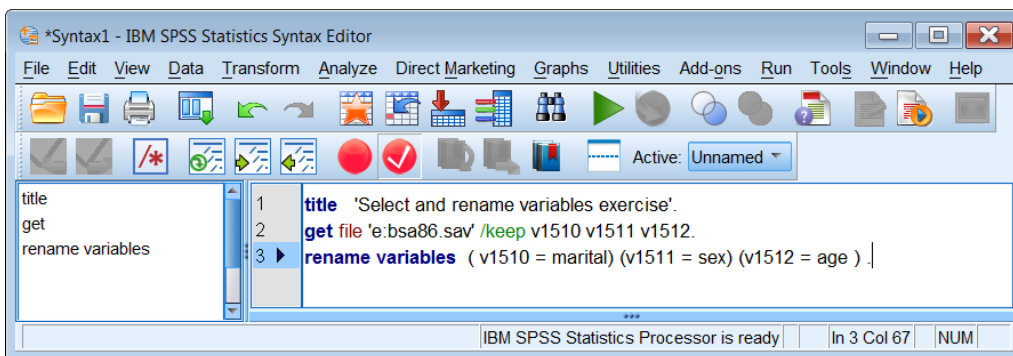> **rename variables  ( v1510 = marital) (v1511 = sex) (v1512 = age ) .**

This command can also be written:

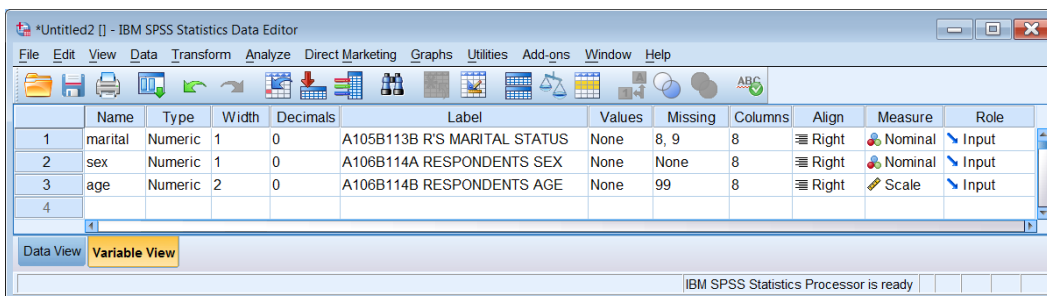> **rename variables  ( v1508 v1511 v1512 = marital sex age) .**

. . or, because the variables are contiguous in the file:

> **rename variables ( v1508 TO v1512 = marital sex age) .**

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



Run the command to get:



*Untitled2 **after renaming variables**

in the Data Editor, you can change the variable labels and add value labels manually, but this won't yield any syntax for future reference or use, so. . .

---

3    **RENAME VARIABLES** (<old name(s)> **=** <new name(s)>)

## Step 4:  Change variable labels

Type in:    **variable labels**
              **marital 'Marital status of respondent'**
              **/sex 'Sex of respondent'**
              **/age 'Age of respondent'.**



Run to produce:



## Step 5:  Add value labels

Type in:    **value labels**
              **marital 1 'Married ' 2 'Living as married ' 3 'Separated or divorced '**
                      **4 'Widowed ' 5 'Not married '**
              **/ sex    1 'Men ' 2 'Women '.**        [Don't forget the forward slash **/** ]



. . and run again to get:

**Step 6: Check your data**

Switch to Data View: (note that the variable names have also changed in the top row):
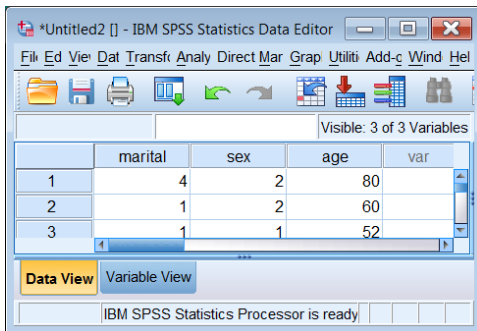


Another way to check is:

Utilities > Variables
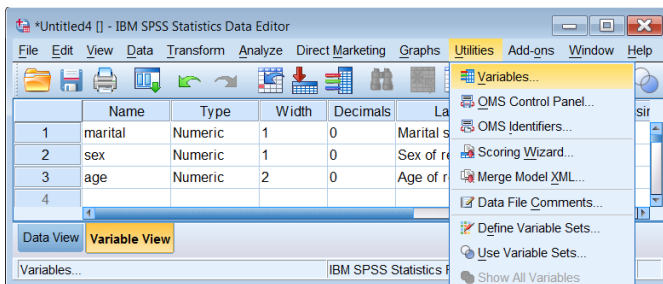


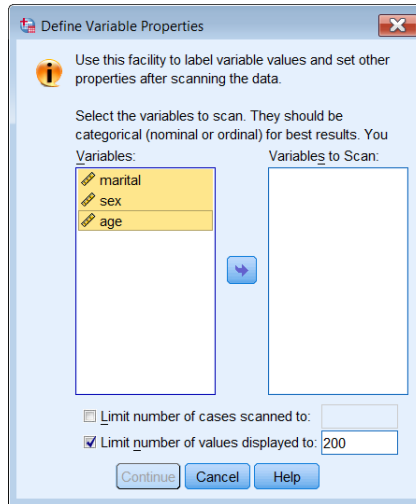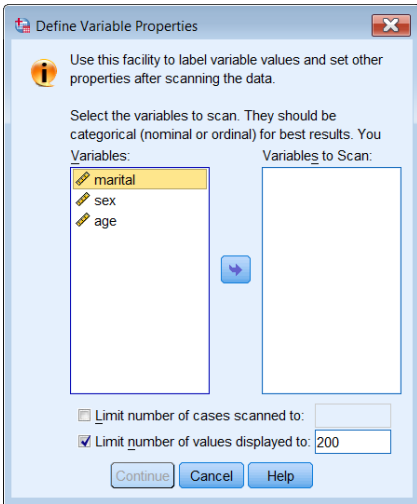Highlight a variable to check its properties:
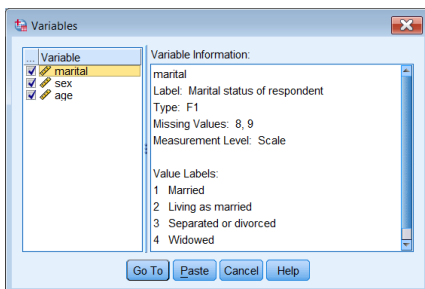


**marital**
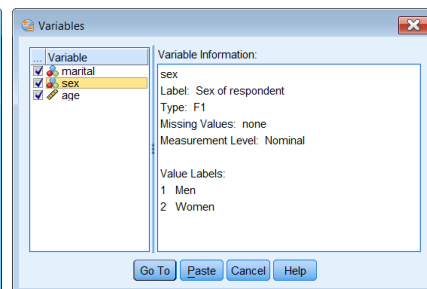


**sex**
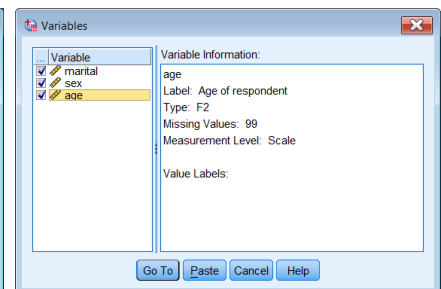


**age**

Now try Data > Define Variable Properties
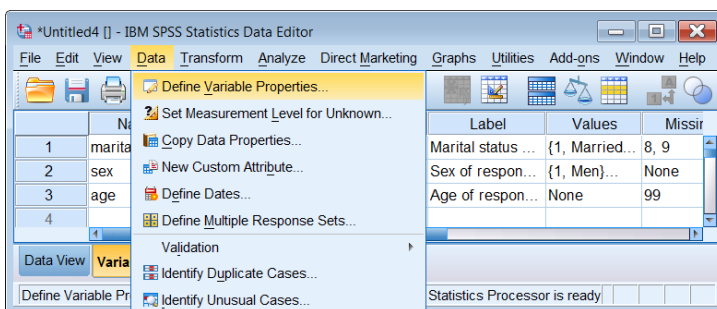


This is a **really useful facility** as it not only allows you to edit variable **properties**, such as **variable labels**, **value labels** and **missing values**, but also gives you a **count** for each **value** found.
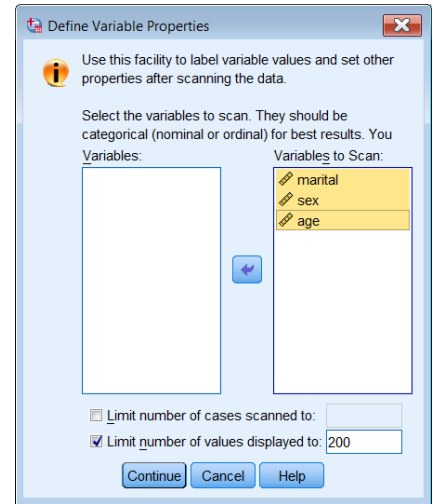
Highlight all three variables,        transfer them to the right pane        and click on Continue

Again highlight each variable to check its properties.  If you didn't correct the measurement levels manually in the Data Editor, you can do so here.  You can also edit the variable and value labels.
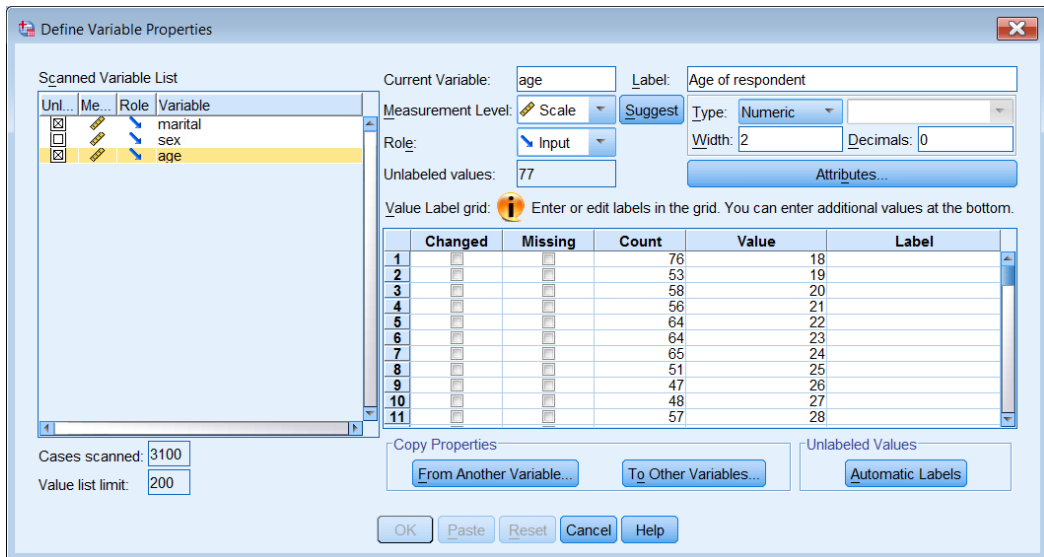
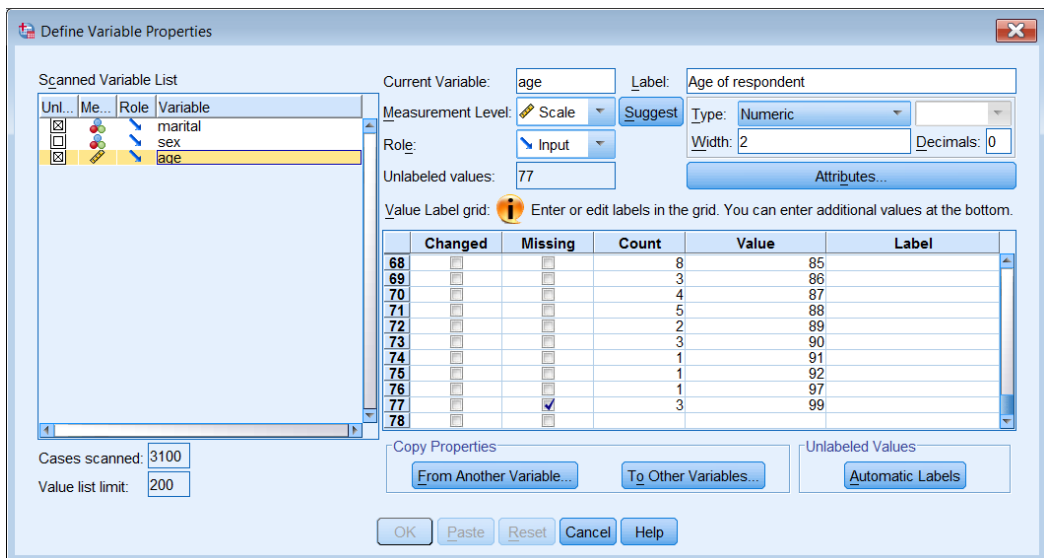Screenshots from Data > Define Variable Properties
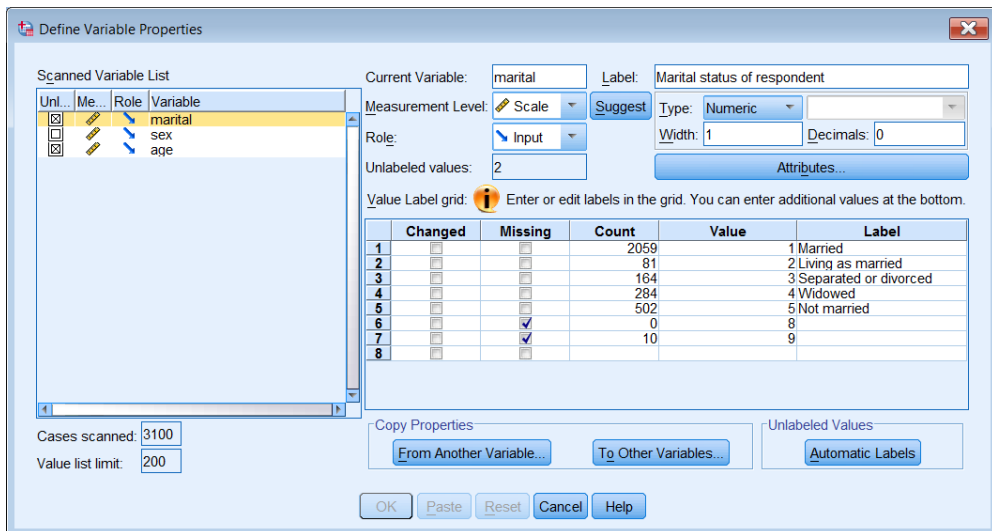


**marital**
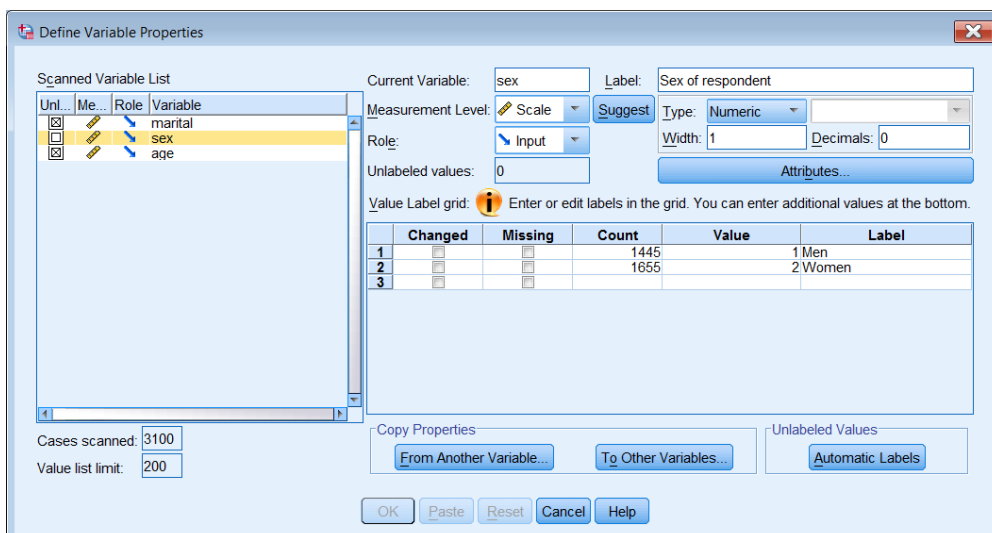


**sex**

Values 8 and 9 have been declared as missing for **marital**, but are not labelled: there are no missing values for **sex**



**age**

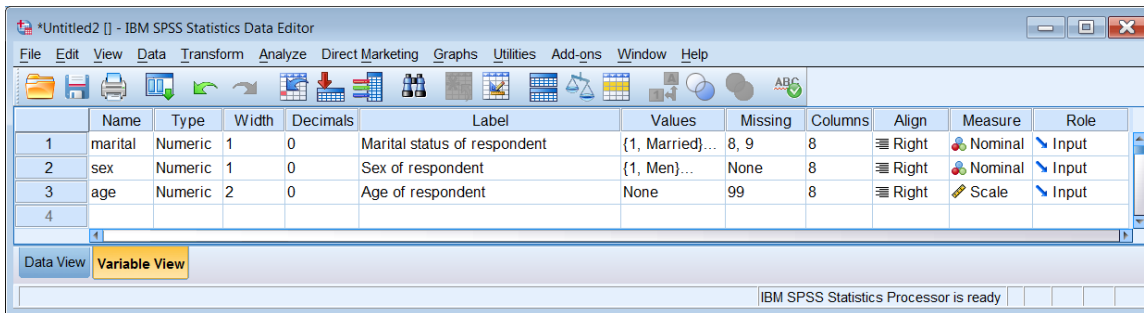Scroll to the bottom to see top range of values for **age**, including 99 declared as missing
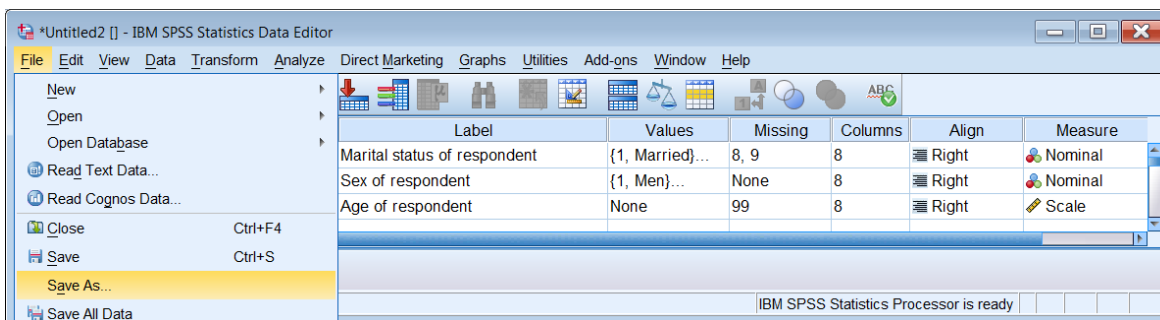


**age**

**Step 7:  Save your active file**

Once you are happy with everything, you need to **save** your active file and your syntax.
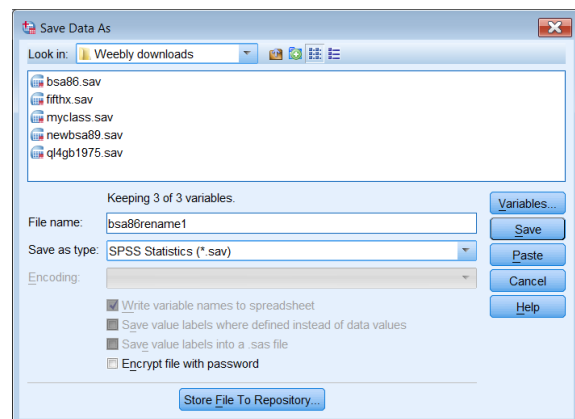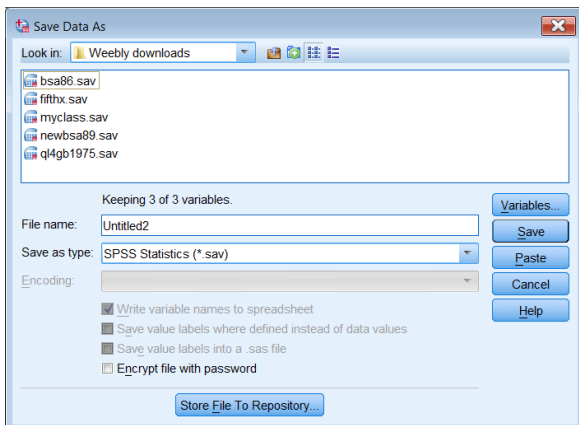
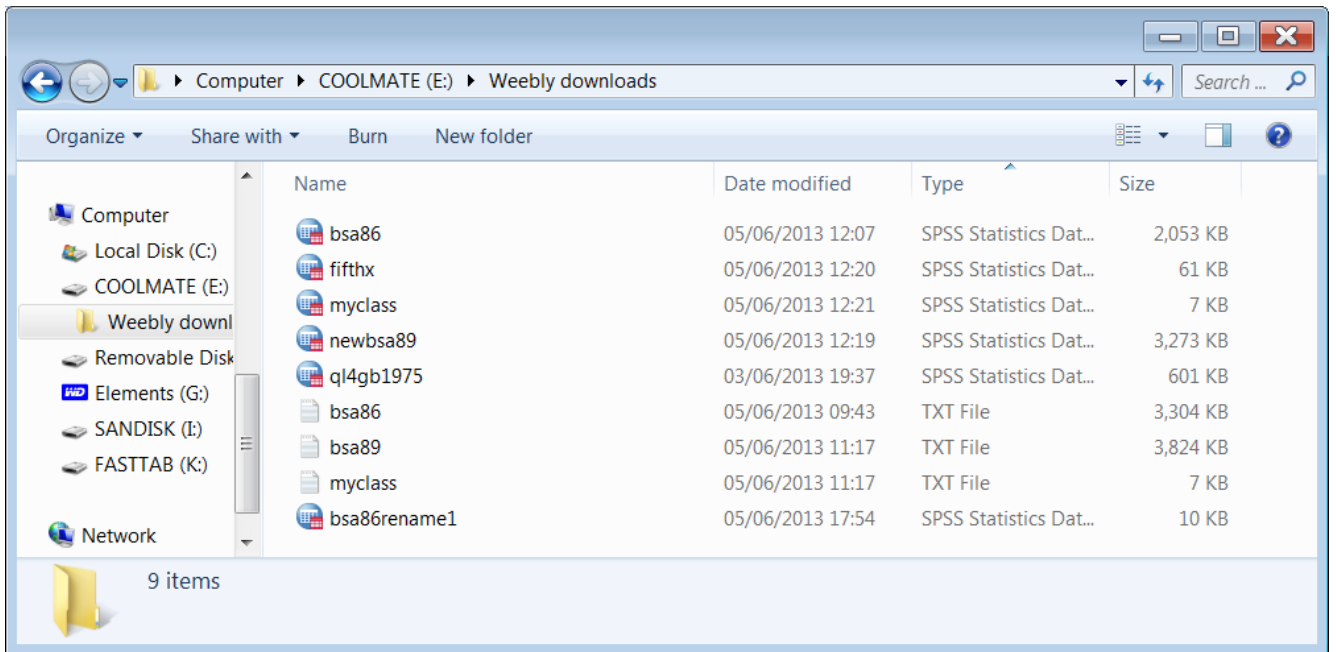Go back to your **Data Editor**:



. . and click on File > Save As



Navigate to **e:weebly downloads** and change the file name from Untitled2 to bsa86rename1.sav
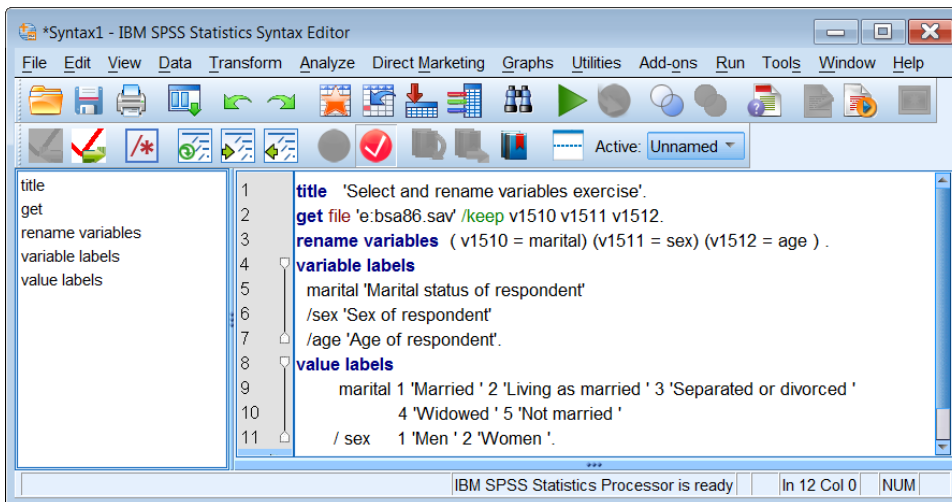


and press Save:

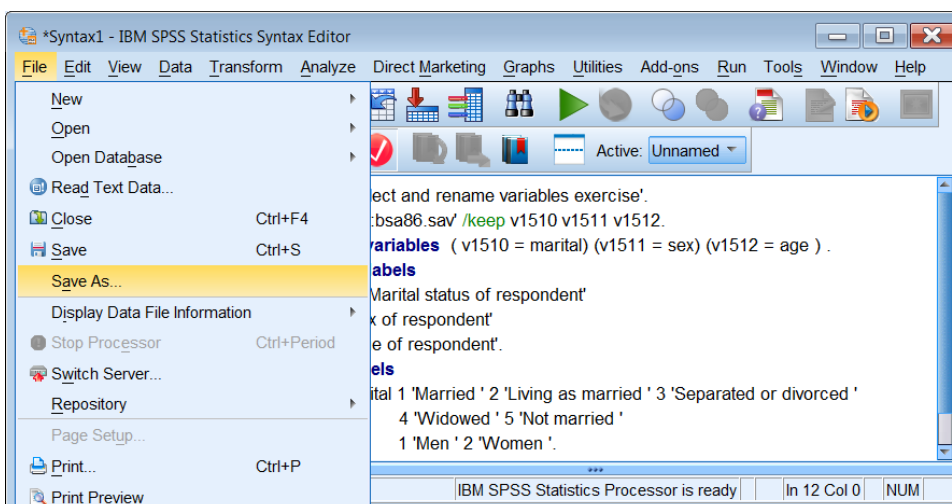The file will be saved as **bsa86rename1.sav** on drive **e:**

## Step 8:  Save your syntax
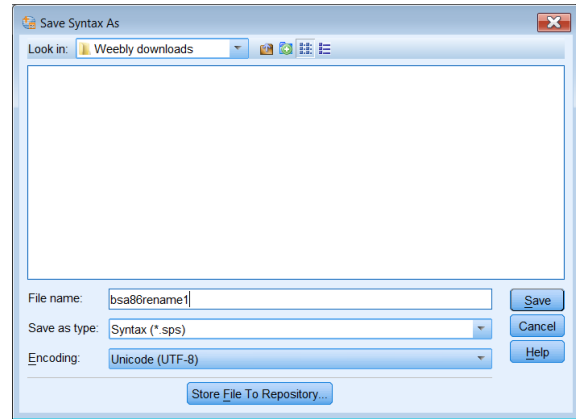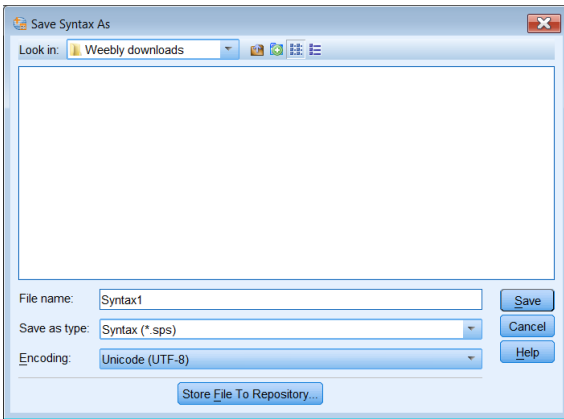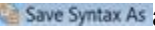
Go to your back to your **Syntax Editor**:



File > Save As

Navigate to **e:weebly downloads**



No, you haven't lost your *.sps file!  It's not shown because SPSS is checking for the presence of other *.sps files in folder **e:weebly downloads** and there aren't any yet.  It's waiting for further instructions.  Save Syntax As appears at top left.  Change the name from Syntax1 to bsa86rename1 and click on Save:
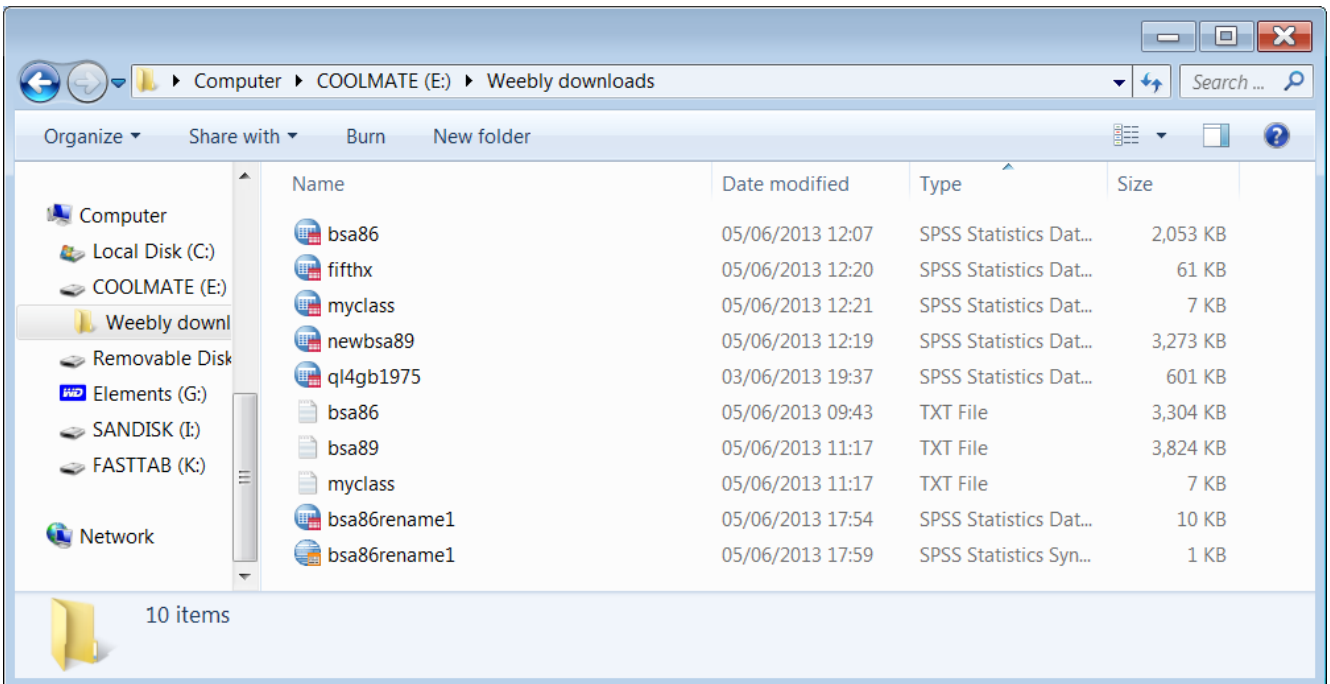


Your syntax file has now been saved in folder **e:weebly downloads**


**End of exercise:     2.3.1.2a1  Select and rename variables**

You can close the saved file and exit SPSS now or leave it open and go straight to the next exercise

**Next exercise:**        2.3.1.2a2 Recode into new variable
**Back to page:**        2.3 Data transformations

[Back to Block 2 menu]