

Block 2: Analysing one variable

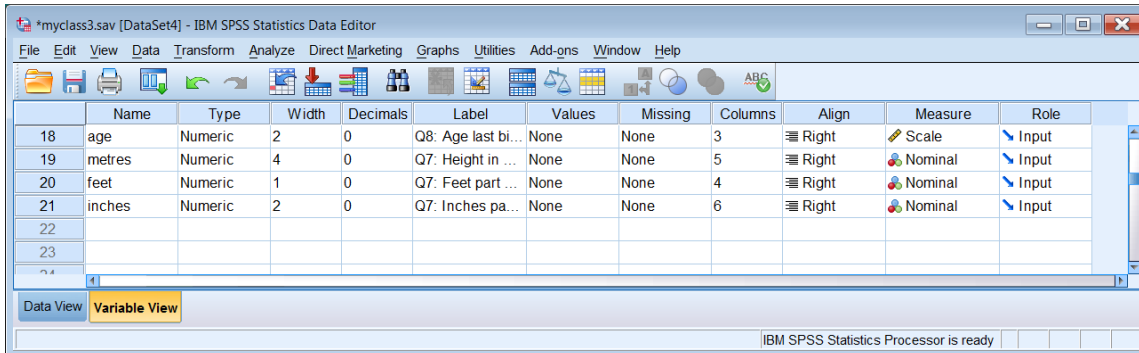
2.3 Data transformations

Screen messages during data transformations

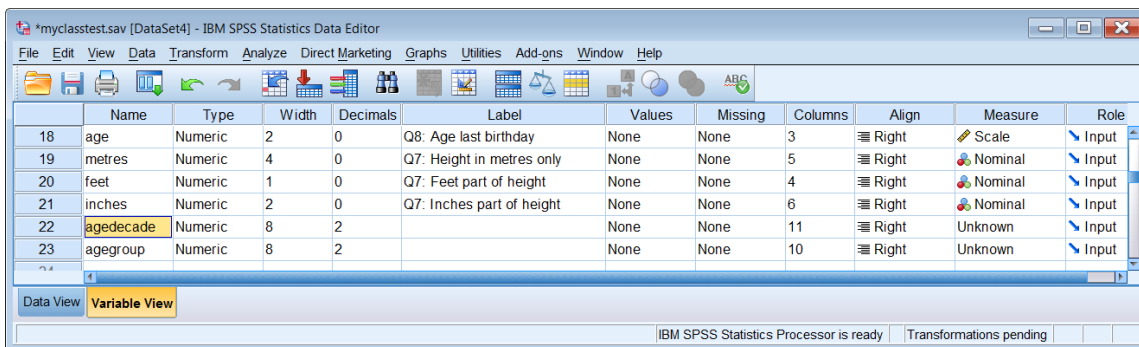
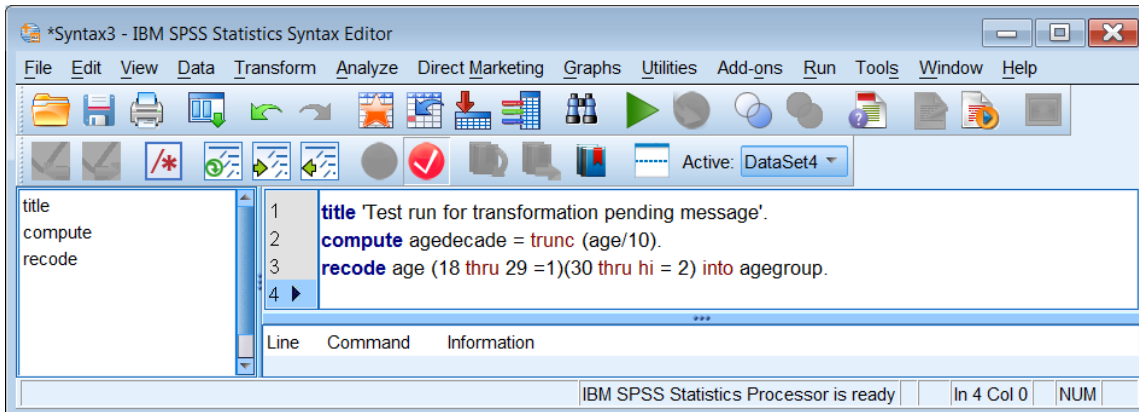
[New: 15 April 2013]

Exemplar: Pre-course questionnaire
 File: [myclass3.sav](#)

Task: Create two new variables: **agedecade** and **agegroup** from **age**
agedecade (integer from **age** divided by 10) and **agegroup** (18-29, 30+)



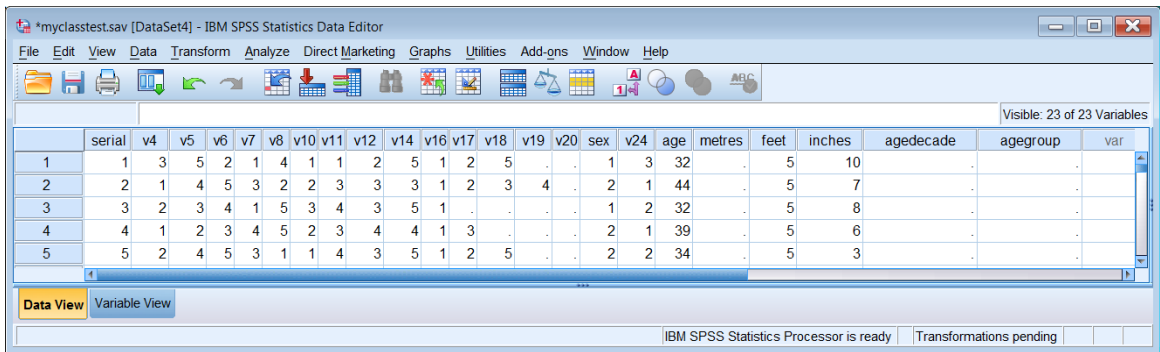
compute agedecade = trunc (age/10).
 recode age (18 thru 29 =1)(30 thru hi = 2) into agegroup.



Both new variables are shown in Decimals as 2 and in Measure as Unknown

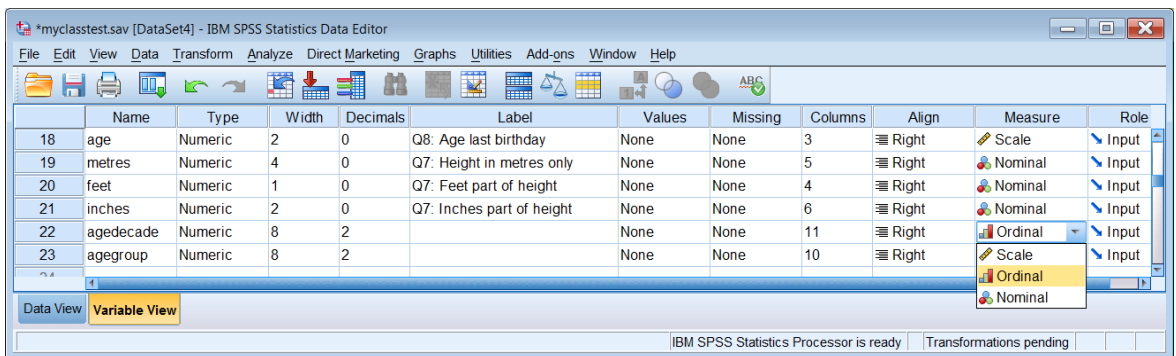
If you look at the lower right, you will see Transformations pending

In **Data View** there are no values displayed for **agedecade** or **agegroup**:

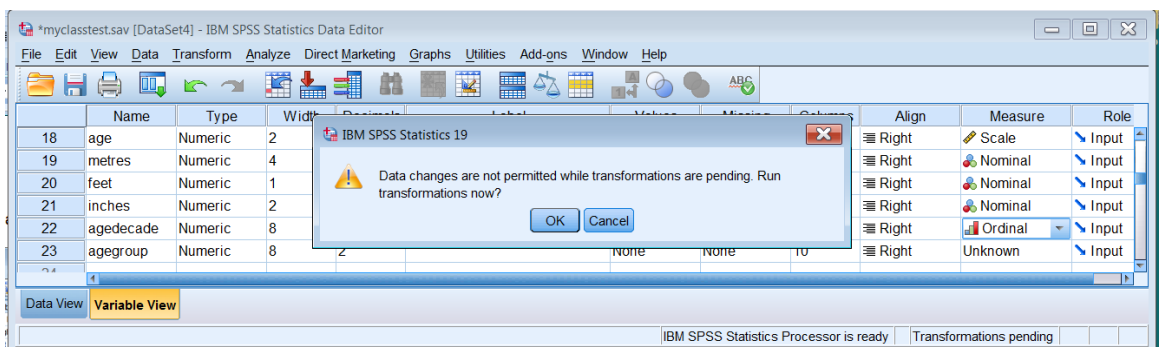


This is because SPSS has not yet made a data pass and is waiting for a command which requires one.

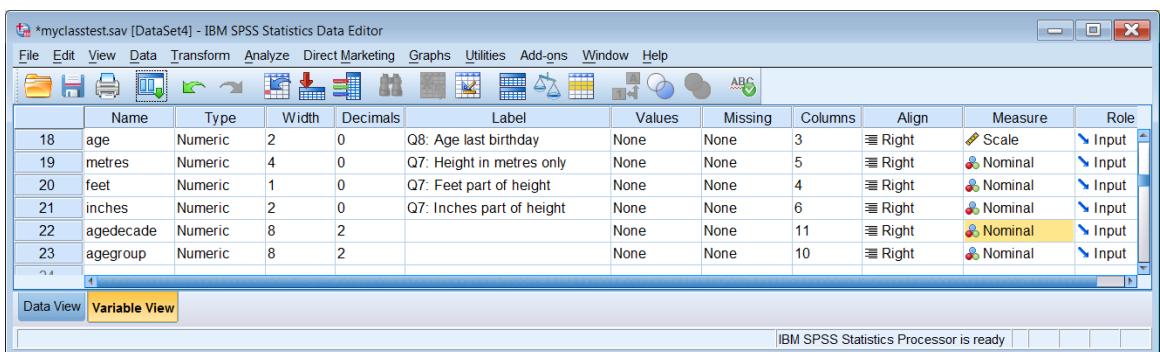
If you go back to **Variable View** and try to change **Unknown** to **Ordinal** for **agedecade**:



you get an error message:



If you press **OK**, both new variables are reset to **Nominal**.



and the values fill up for both variables:

	serial	v4	v5	v6	v7	v8	v10	v11	v12	v14	v16	v17	v18	v19	v20	sex	v24	age	metres	feet	inches	agedecade	agegroup	var
1	1	3	5	2	1	4	1	1	2	5	1	2	5	.	.	1	3	32	.	5	10	3.00	2.00	
2	2	1	4	5	3	2	2	3	3	3	1	2	3	4	.	2	1	44	5	7	4.00	2.00		
3	3	2	3	4	1	5	3	4	3	5	1	1	2	32	.	5	8	3.00	2.00	
4	4	1	2	3	4	5	2	3	4	4	1	3	.	.	.	2	1	39	.	5	6	3.00	2.00	
5	5	2	4	5	3	1	1	4	3	5	1	2	5	.	.	2	2	34	.	5	3	3.00	2.00	

Likewise, If you try to change 2 to 0:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
18	age	Numeric	2	0	Q8: Age last birthday	None	None	3	Right	Scale	Input
19	metres	Numeric	4	0	Q7: Height in metres only	None	None	5	Right	Nominal	Input
20	feet	Numeric	1	0	Q7: Feet part of height	None	None	4	Right	Nominal	Input
21	inches	Numeric	2	0	Q7: Inches part of height	None	None	6	Right	Nominal	Input
22	agedecade	Numeric	8	0		None	None	11	Right	Unknown	Input
23	agegroup	Numeric	8	2		None	None	10	Right	Unknown	Input

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
18	age	Numeric	2	0	Q8: Age last birthday	None	None	3	Right	Scale	Input
19	metres	Numeric	4	0	Q7: Height in metres only	None	None	5	Right	Nominal	Input
20	feet	Numeric	1	0	Q7: Feet part of height	None	None	4	Right	Nominal	Input
21	inches	Numeric	2	0	Q7: Inches part of height	None	None	6	Right	Nominal	Input
22	agedecade	Numeric	8	0		None	None	11	Right	Unknown	Input
23	agegroup	Numeric	8	2		None	None	10	Right	Unknown	Input

IBM SPSS Statistics 19

Data changes are not permitted while transformations are pending. Run transformations now?

OK Cancel

Clicking on **OK** executes all the pending transformations. Decimals is still 2 for both variables.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
18	age	Numeric	2	0	Q8: Age last birthday	None	None	3	Right	Scale	Input
19	metres	Numeric	4	0	Q7: Height in metres only	None	None	5	Right	Nominal	Input
20	feet	Numeric	1	0	Q7: Feet part of height	None	None	4	Right	Nominal	Input
21	inches	Numeric	2	0	Q7: Inches part of height	None	None	6	Right	Nominal	Input
22	agedecade	Numeric	8	2		None	None	11	Right	Nominal	Input
23	agegroup	Numeric	8	2		None	None	10	Right	Nominal	Input

and the values fill up for both variables:

Visible: 23 of 23 Variables

	serial	v4	v5	v6	v7	v8	v10	v11	v12	v14	v16	v17	v18	v19	v20	sex	v24	age	metres	feet	inches	agedecile	agegroup
1	1	3	5	2	1	4	1	1	2	5	1	2	5	.	.	1	3	32	.	5	10	3.00	2.00
2	2	1	4	5	3	2	2	3	3	3	1	2	3	4	.	2	1	44	.	5	7	4.00	2.00
3	3	2	3	4	1	5	3	4	3	5	1	1	2	32	.	5	8	3.00	2.00
4	4	1	2	3	4	5	2	3	4	4	1	3	.	.	.	2	1	39	.	5	6	3.00	2.00
5	5	2	4	5	3	1	1	4	3	5	1	2	5	.	.	2	2	34	.	5	3	3.00	2.00

Changing 2 to 0 in the **Data Editor** now works:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
18	age	Numeric	2	0	Q8: Age last birthday	None	None	3	Right	Scale	Input
19	metres	Numeric	4	0	Q7: Height in metres only	None	None	5	Right	Nominal	Input
20	feet	Numeric	1	0	Q7: Feet part of height	None	None	4	Right	Nominal	Input
21	inches	Numeric	2	0	Q7: Inches part of height	None	None	6	Right	Nominal	Input
22	agedecade	Numeric	8	2		None	None	11	Right	Unknown	Input
23	agegroup	Numeric	8	2		None	None	10	Right	Unknown	Input

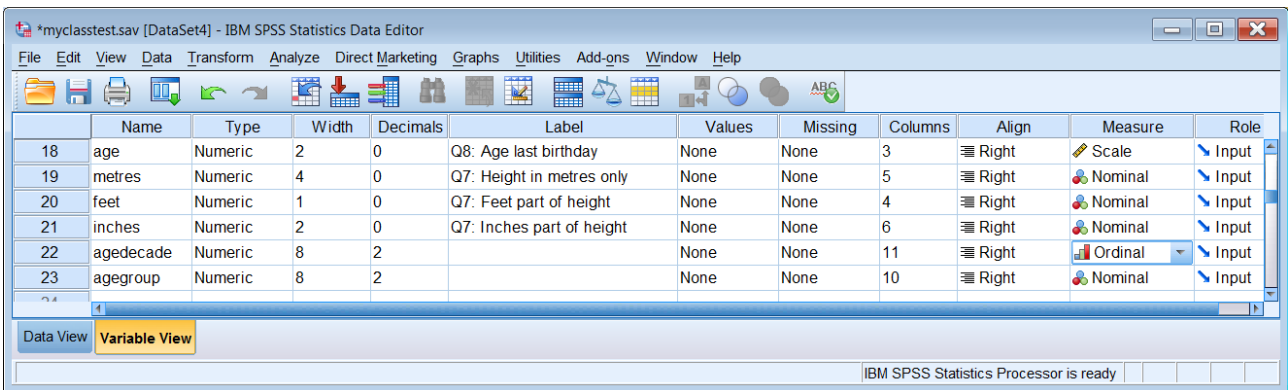
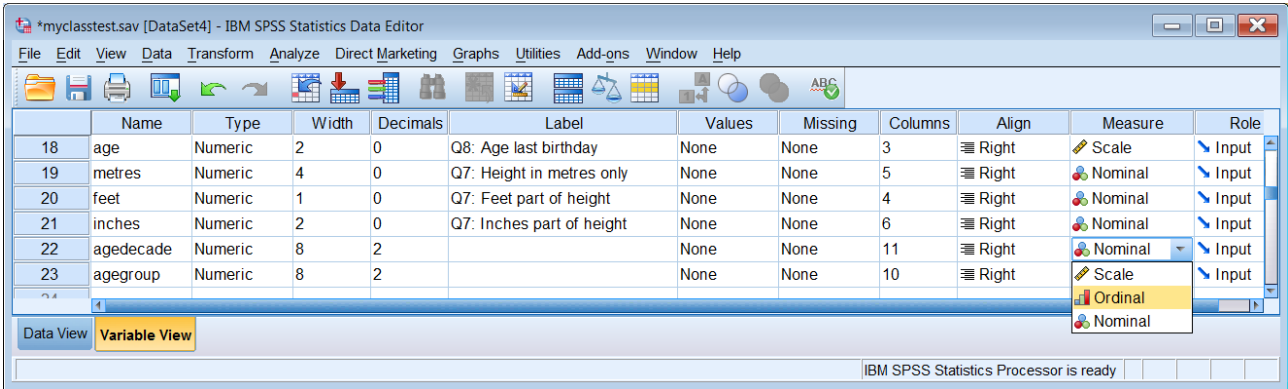
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
18	age	Numeric	2	0	Q8: Age last bi...	None	None	3	Right	Scale	Input
19	metres	Numeric	4	0	Q7: Height in ...	None	None	5	Right	Nominal	Input
20	feet	Numeric	1	0	Q7: Feet part ...	None	None	4	Right	Nominal	Input
21	inches	Numeric	2	0	Q7: Inches pa...	None	None	6	Right	Nominal	Input
22	agedecile	Numeric	8	0		None	None	11	Right	Nominal	Input
23	agegroup	Numeric	8	2		None	None	10	Right	Nominal	Input

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
18	age	Numeric	2	0	Q8: Age last bi...	None	None	3	Right	Scale	Input
19	metres	Numeric	4	0	Q7: Height in ...	None	None	5	Right	Nominal	Input
20	feet	Numeric	1	0	Q7: Feet part ...	None	None	4	Right	Nominal	Input
21	inches	Numeric	2	0	Q7: Inches pa...	None	None	6	Right	Nominal	Input
22	agedecile	Numeric	8	0		None	None	11	Right	Ordinal	Input
23	agegroup	Numeric	8	2		None	None	10	Right	Nominal	Input

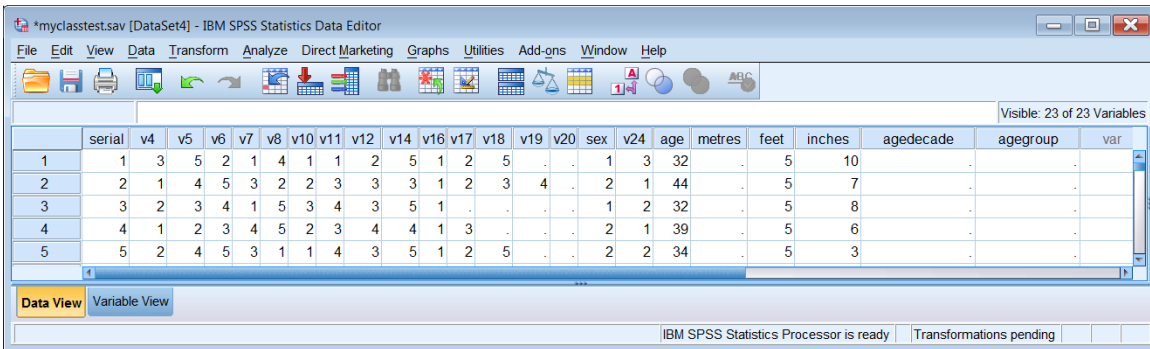
Visible: 23 of 23 Variables

	serial	v4	v5	v6	v7	v8	v10	v11	v12	v14	v16	v17	v18	v19	v20	sex	v24	age	metres	feet	inches	agedecile	agegroup
1	1	3	5	2	1	4	1	1	2	5	1	2	5	.	.	1	3	32	.	5	10	3	2
2	2	1	4	5	3	2	2	3	3	3	1	2	3	4	.	2	1	44	.	5	7	4	2
3	3	2	3	4	1	5	3	4	3	5	1	1	2	32	.	5	8	3	2
4	4	1	2	3	4	5	2	3	4	4	1	3	.	.	.	2	1	39	.	5	6	3	2
5	5	2	4	5	3	1	1	4	3	5	1	2	5	.	.	2	2	34	.	5	3	3	2

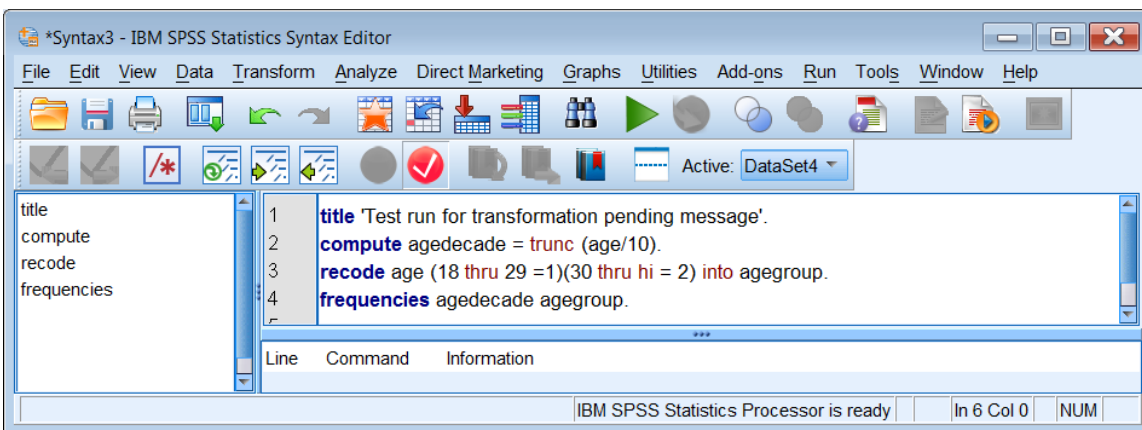
. . as does changing **Nominal** to **Ordinal** for **agedecade**.



At an earlier stage In **Data View**, when there are still no values for **agedecade** or **agegroup**:



frequencies agedecade agegroup.



produces the following output:

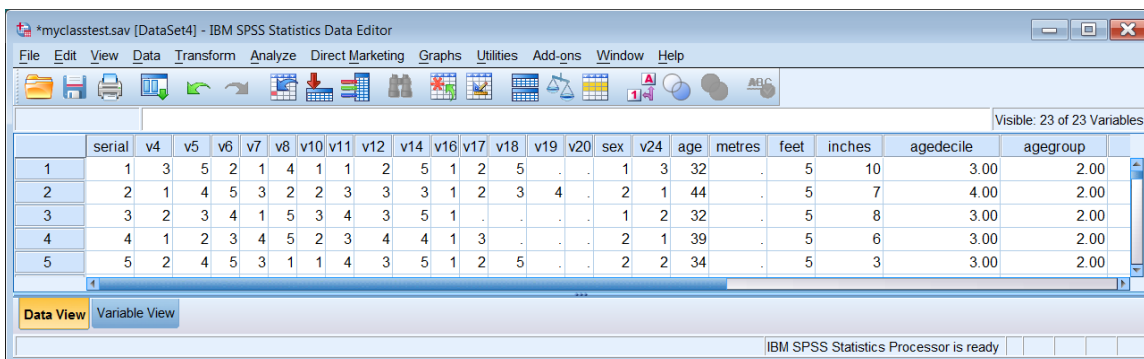
Statistics			
		agedecade	agegroup
N	Valid	166	166
	Missing	3	3

agedecade					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	7	4.1	4.2	4.2
	2.00	77	45.6	46.4	50.6
	3.00	58	34.3	34.9	85.5
	4.00	19	11.2	11.4	97.0
	5.00	4	2.4	2.4	99.4
	6.00	1	.6	.6	100.0
	Total		166	98.2	100.0
Missing	System	3	1.8		
Total		169	100.0		

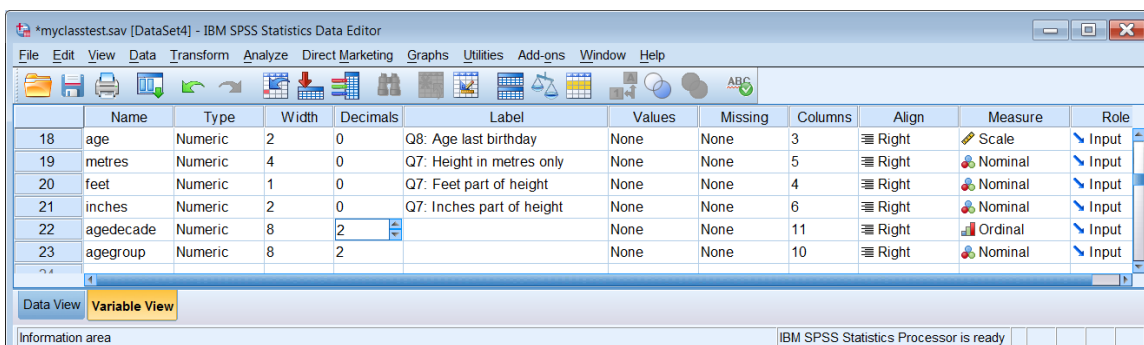
agegroup					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	84	49.7	50.6	50.6
	2.00	82	48.5	49.4	100.0
	Total	166	98.2	100.0	
Missing	System	3	1.8		
Total		169	100.0		

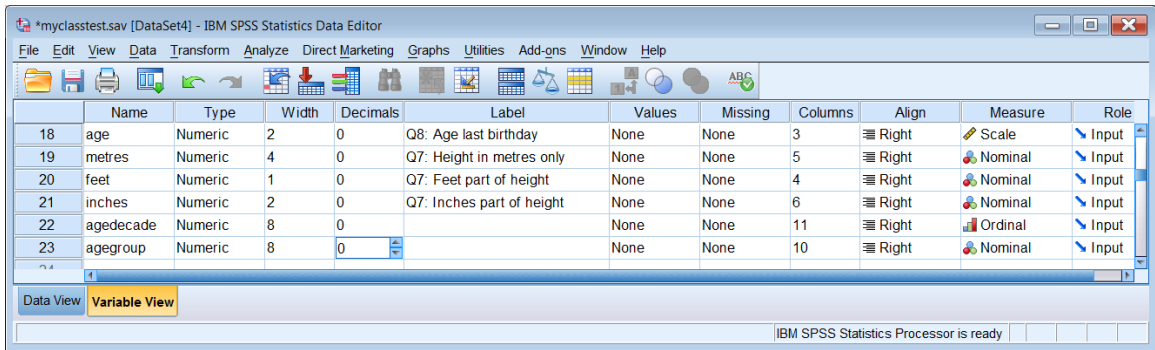
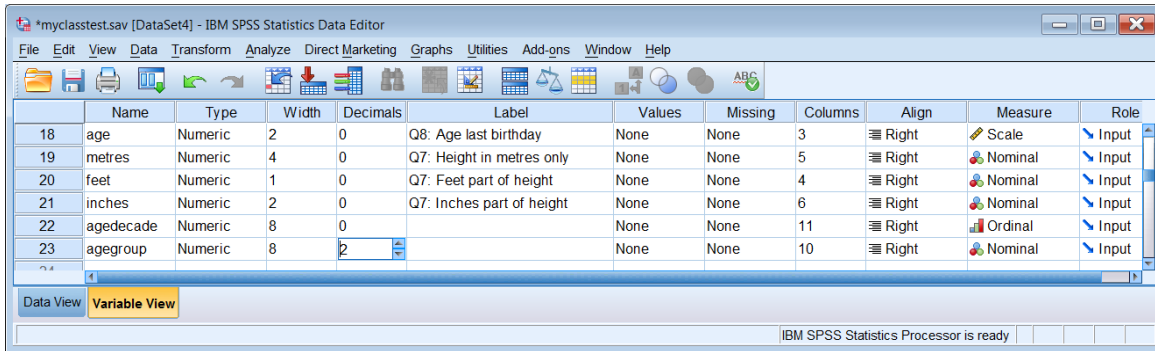
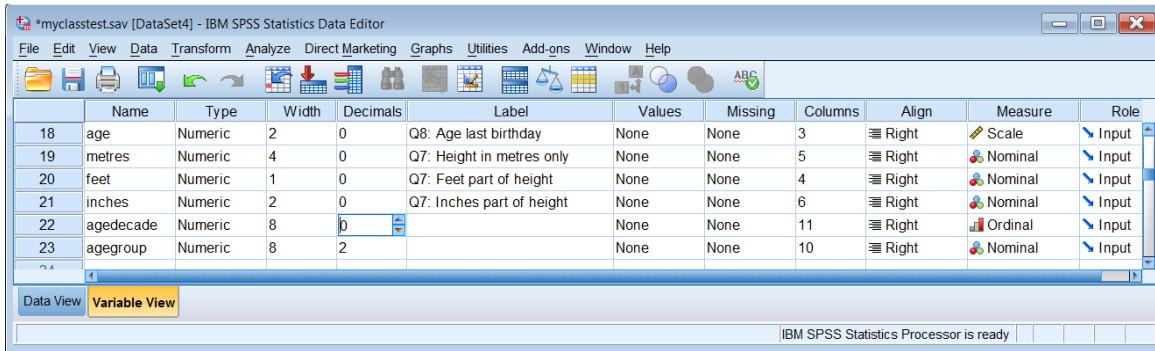
, but both variables have values with superfluous decimal places.

In **Data Editor** the values have now filled in for both variables:

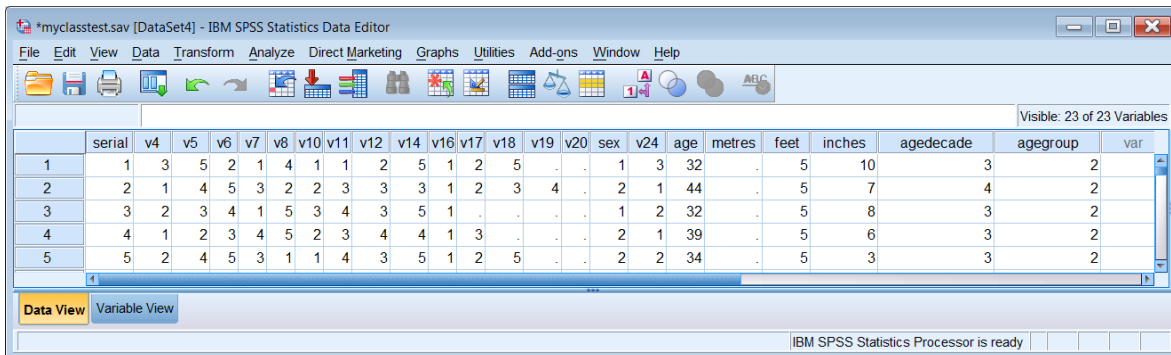


You can manually change the 2 to 0 in the data editor **Decimals** column:





Data Editor now looks like this:



but it's far better to specify the variable properties in syntax in the first place and save a lot of messing about.

Just after the transformations:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
18	age	Numeric	2	0	Q8: Age last birthday	None	None	3	Right	Scale	Input
19	metres	Numeric	4	0	Q7: Height in metres only	None	None	5	Right	Nominal	Input
20	feet	Numeric	1	0	Q7: Feet part of height	None	None	4	Right	Nominal	Input
21	inches	Numeric	2	0	Q7: Inches part of height	None	None	6	Right	Nominal	Input
22	agedecade	Numeric	8	2		None	None	11	Right	Unknown	Input
23	agegroup	Numeric	8	2		None	None	10	Right	Unknown	Input

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

```

title
compute
recode
frequencies
variable level
formats

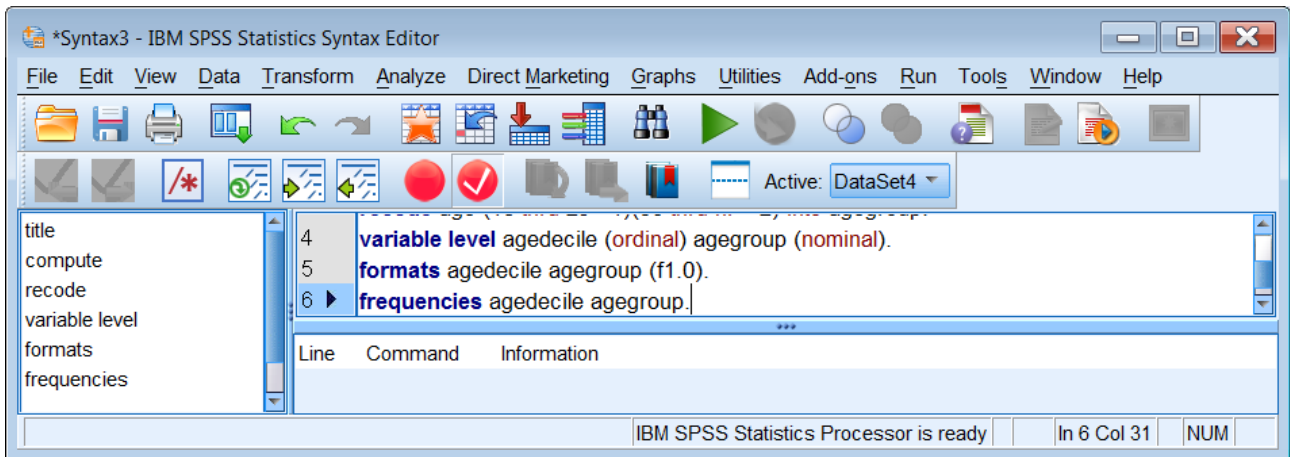
recode age ( 18 thru 29 = 1)(30 thru 44 = 2) into agegroup.
4 frequencies agedecade agegroup.
5 variable level agedecade (ordinal) agegroup (nominal).
6 formats agedecade agegroup (f1.0).
7

```

Measurement and Decimals now correct, but note Width has also changed from 8 to 1

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
18	age	Numeric	2	0	Q8: Age last birthday	None	None	3	Right	Scale	Input
19	metres	Numeric	4	0	Q7: Height in metres only	None	None	5	Right	Nominal	Input
20	feet	Numeric	1	0	Q7: Feet part of height	None	None	4	Right	Nominal	Input
21	inches	Numeric	2	0	Q7: Inches part of height	None	None	6	Right	Nominal	Input
22	agedecade	Numeric	1	0		None	None	11	Right	Ordinal	Input
23	agegroup	Numeric	1	0		None	None	10	Right	Nominal	Input

frequencies agedecade agegroup.



produces:

agedecade

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	7	4.1	4.2	4.2
	2	77	45.6	46.4	50.6
	3	58	34.3	34.9	85.5
	4	19	11.2	11.4	97.0
	5	4	2.4	2.4	99.4
	6	1	.6	.6	100.0
	Total	166	98.2	100.0	
Missing	System	3	1.8		
Total		169	100.0		

agegroup

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	84	49.7	50.6	50.6
	2	82	48.5	49.4	100.0
	Total	166	98.2	100.0	
Missing	System	3	1.8		
Total		169	100.0		

Job done! (apart from adding labels).

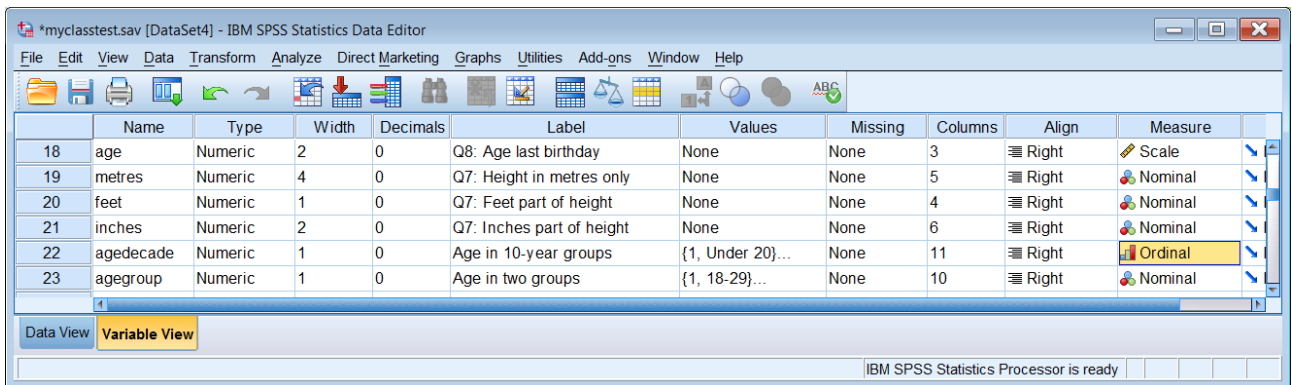
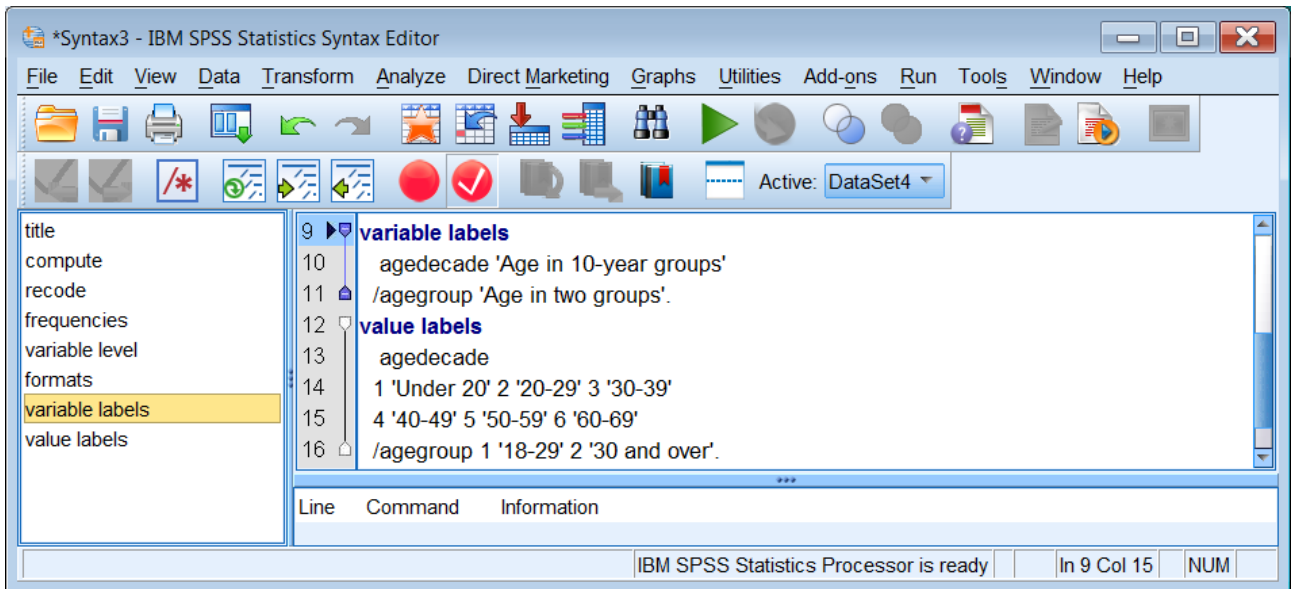
Now let's do a proper job, and add labels:

variable labels

agedecade 'Age in 10-year groups'
/agegroup 'Age in two groups'.

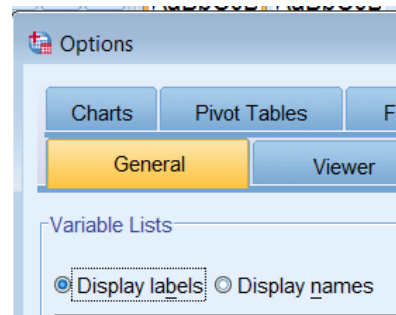
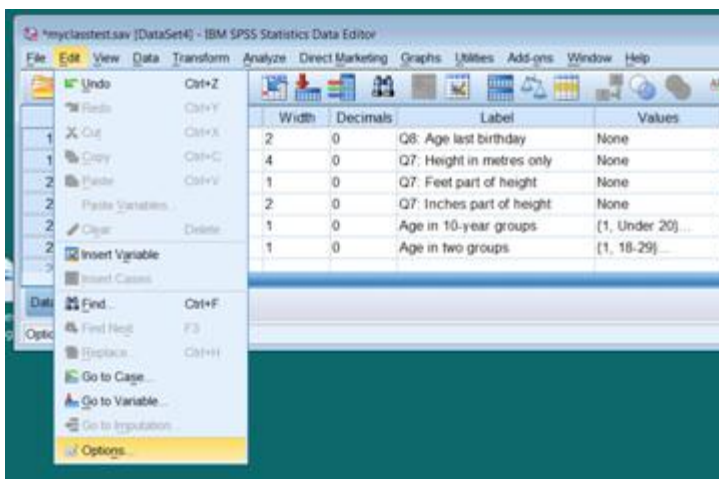
value labels

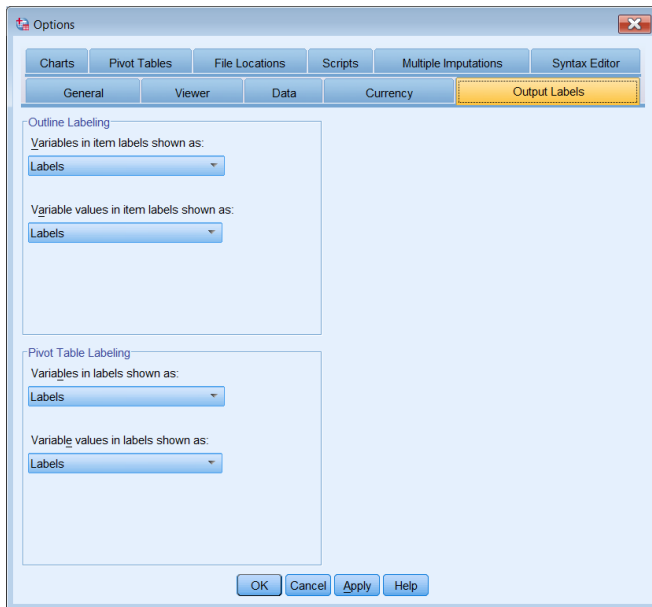
agedecade
1 'Under 20' 2 '20-29' 3 '30-39'
4 '40-49' 5 '50-59' 6 '60-69'
/agegroup 1 '18-29' 2 '30 and over'.



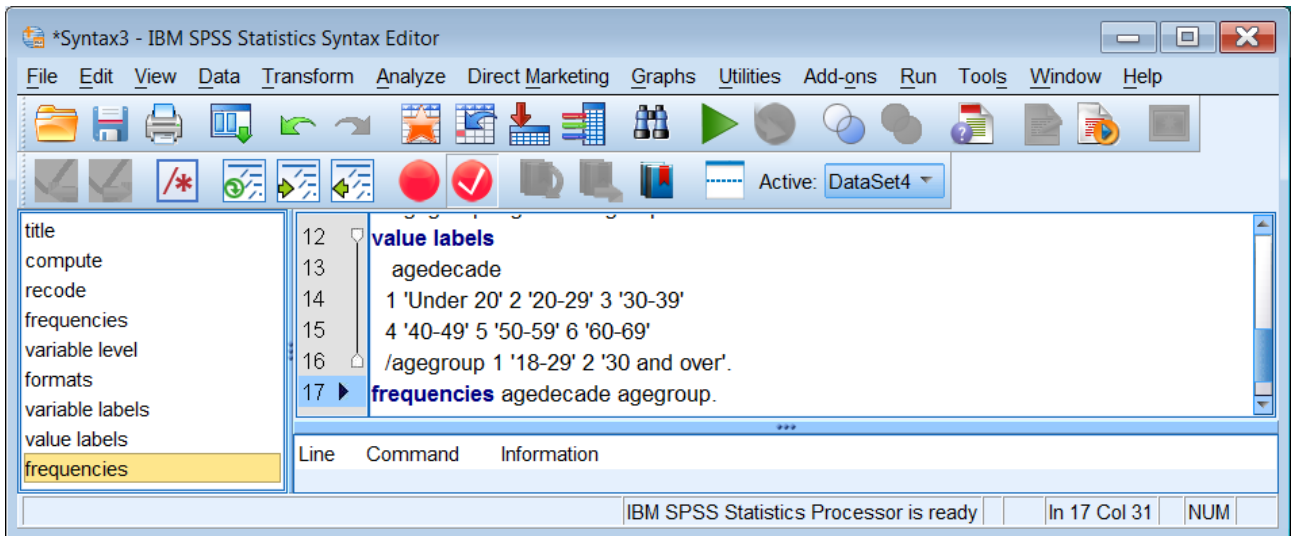
Until I'm happy with a file I usually have table settings to display both names and labels for variables and both values and labels for values. However, for this next run I've changed both settings to labels only, as the tables look neater.

Edit > Options





frequencies agedecade agegroup.



Age in 10-year groups

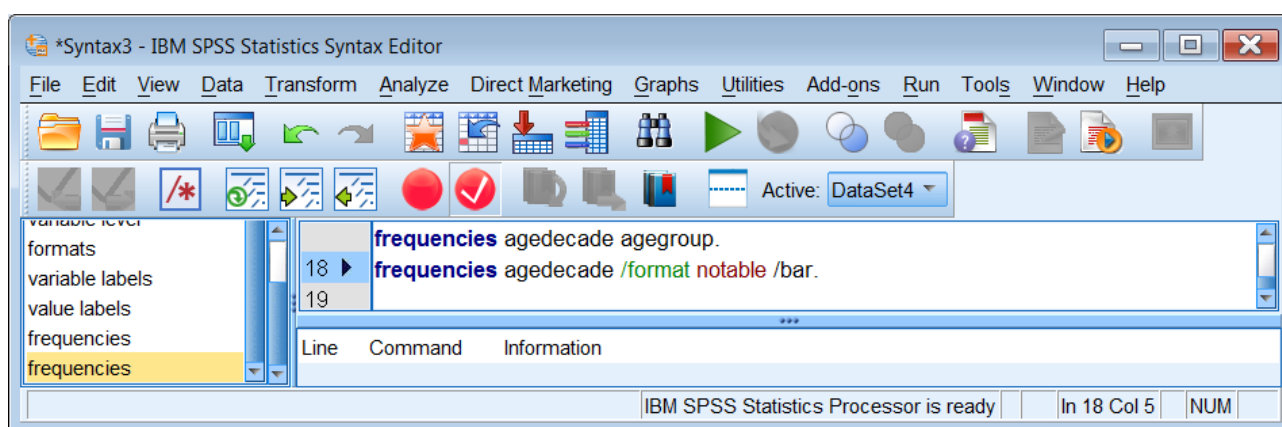
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid				
Under 20	7	4.1	4.2	4.2
20-29	77	45.6	46.4	50.6
30-39	58	34.3	34.9	85.5
40-49	19	11.2	11.4	97.0
50-59	4	2.4	2.4	99.4
60-69	1	.6	.6	100.0
Total	166	98.2	100.0	
Missing				
System	3	1.8		
Total	169	100.0		

Age in two groups

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	18-29	84	49.7	50.6	50.6
	30 and over	82	48.5	49.4	100.0
	Total	166	98.2	100.0	
Missing	System	3	1.8		
Total		169	100.0		

Because a picture is worth a thousand words. . .

frequencies agedecade /format notable /bar.



Age in 10-year groups

