

[Commentary by [John F Hall](#)]

[Last updated: 8 August 2017]

John MacInnes[An Introduction to Secondary Data Analysis with IBM SPSS Statistics](#)

(Sage, Dec. 2017)

5.1 [Chapter 5 video tutorials](#) (direct link to companion website)

[NB: All video tutorials for chapter 5 are on the same web page and cannot (yet) be disaggregated]

Video tutorial 5.1.2: Identifying the missing cases on **[cldgng]** (4'55")

The survey carried a number of psychological well-being questions, eight of which (questions D5 to D12) asked people whether, during the past week, they had experienced symptoms of "depression".

CARD 25 I will now read out a list of the ways you might have felt or behaved during the past week. Using this card, please tell me how much of the time during the past week...**READ OUT...**

		None or almost none of the time	Some of the time	Most of the time	All or almost all of the time	(Don't know)
D5	...you felt depressed?	1	2	3	4	8
D6	...you felt that everything you did was an effort?	1	2	3	4	8
D7	...your sleep was restless?	1	2	3	4	8
D8	...you were happy?	1	2	3	4	8
D9	...you felt lonely?	1	2	3	4	8
D10	...you enjoyed life?	1	2	3	4	8
D11	...you felt sad?	1	2	3	4	8
D12	...you could not get going ³¹ ?	1	2	3	4	8

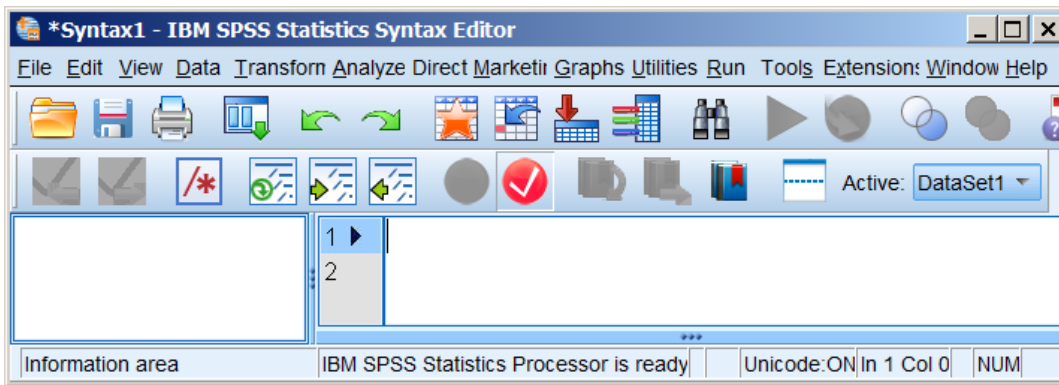
The eight associated variables are on rows 197 to 204 of the **Data Editor**:

	Name	Measure	Label
197	ftdpr	Ordinal	Felt depressed, how often past week
198	fiteeff	Ordinal	Felt everything did as effort, how often past week
199	slprl	Ordinal	Sleep was restless, how often past week
200	wrhpp	Ordinal	Were happy, how often past week
201	fltlnl	Ordinal	Felt lonely, how often past week
202	enjlf	Ordinal	Enjoyed life, how often past week
203	fltsd	Ordinal	Felt sad, how often past week
204	cldgng	Ordinal	Could not get going, how often past week

IBM SPSS Statistics Processor is ready | Unicode:ON

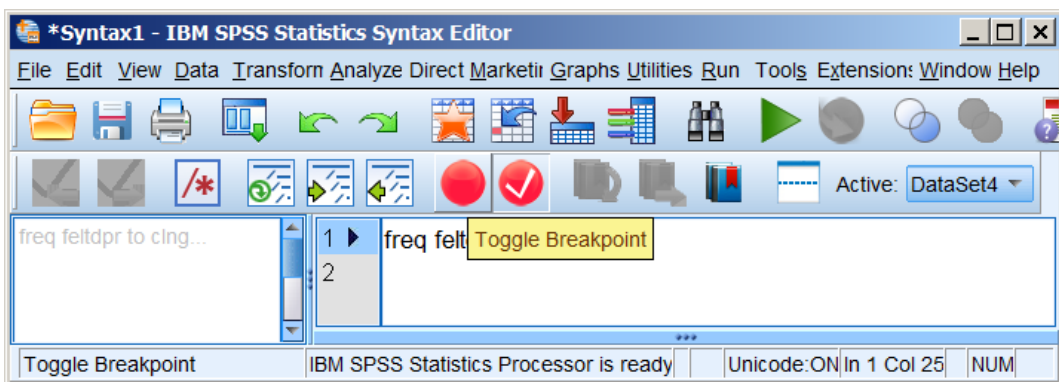
JM produces frequency tables for each of these by opening a new **Syntax Editor**:



File >> **New** >> **Syntax**



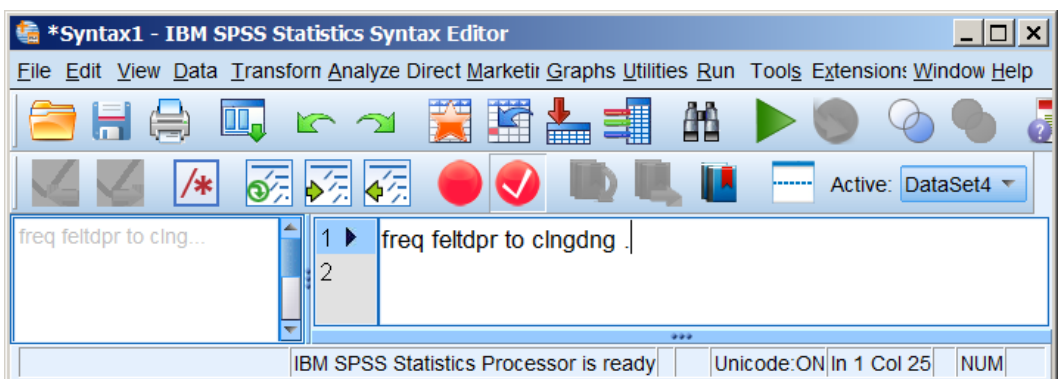
. . and typing in (abbreviated) syntax:

freq feltdpr to cldgng .



Note that, as soon as he starts typing, the **Toggle Breakpoint** icon changes from  to 

He does not explain the use of keyword **TO**¹ in the specification, nor (unless **[pweight]** is still in force from earlier in the aggregated syntax file) does he apply a weight.



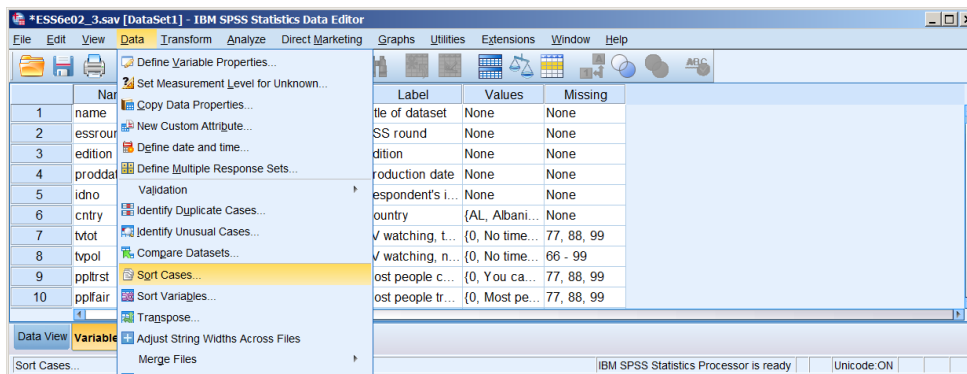
The frequencies command produces a separate frequency table for each of the eight symptoms.

JM then uses the tables to demonstrate that most of these variables have very few missing cases (around 1.0% to 1.5%) except for an unusually high number on the last variable **[cldgng]** which has 3.7%, including 2.2% have the system missing value **SYSMIS**.

He surmises that something has gone wrong in one of the countries. To track down the culprit(s) with the unusually high number of missing values for **[cldgng]** he uses:

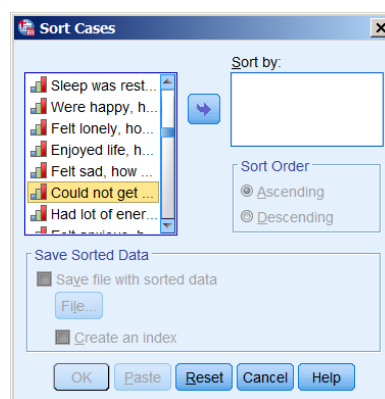
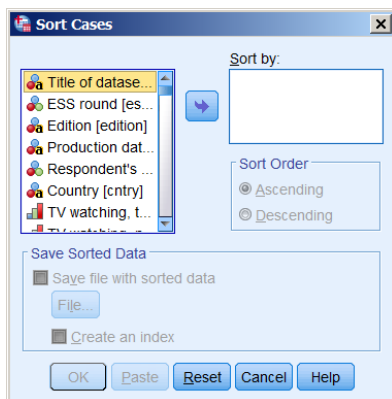
¹ Provided the variables are contiguous in the file **<varx> TO <vary>** saves typing all the variable names separately.


Data >> Sort Cases

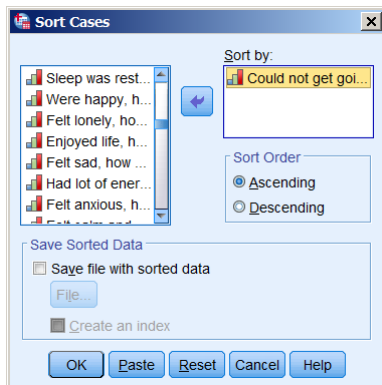


Scroll down to find **[cldngng]**

(Not easy when display set to **Labels** only!)



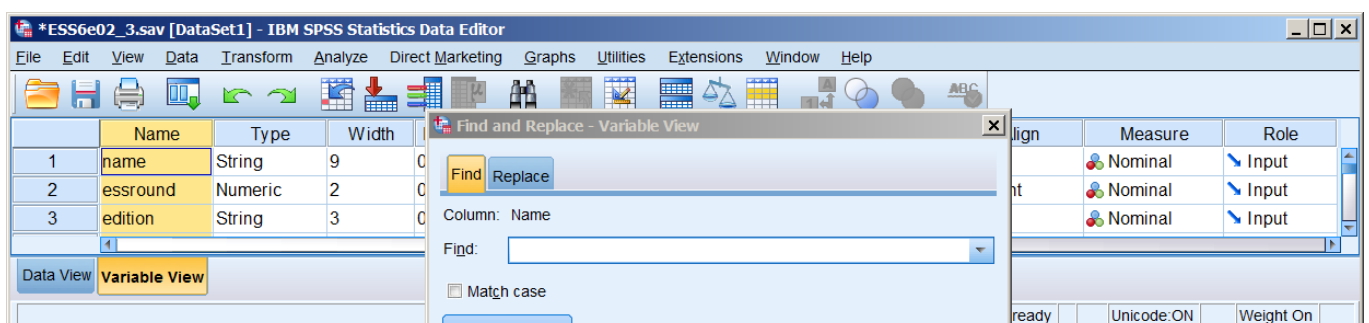
Click on  to transfer **[cldngng]** to right hand pane.




Syntax generated by **Paste**

```
SORT CASES BY cldngng (A).
```

Highlights the **Name** column:

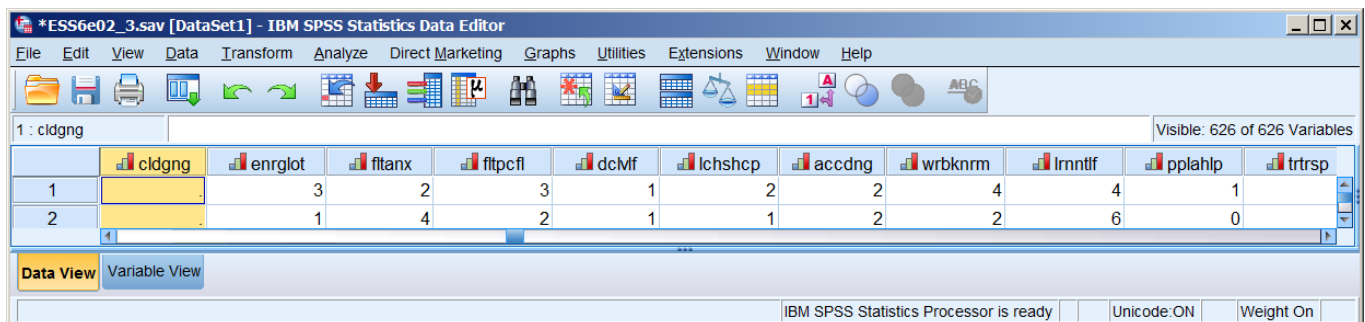


.. and uses  to search for **[cldgng]**

	Name
202	enjl
203	fltsd
204	cldgng
1	
Data View	Variable View

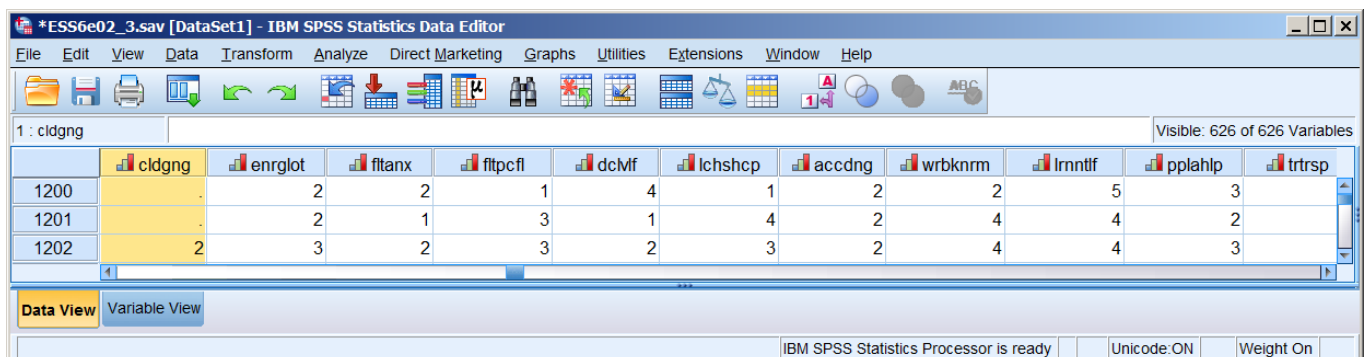
". . double click next to that variable" (actually on the **204** cell)

Data View



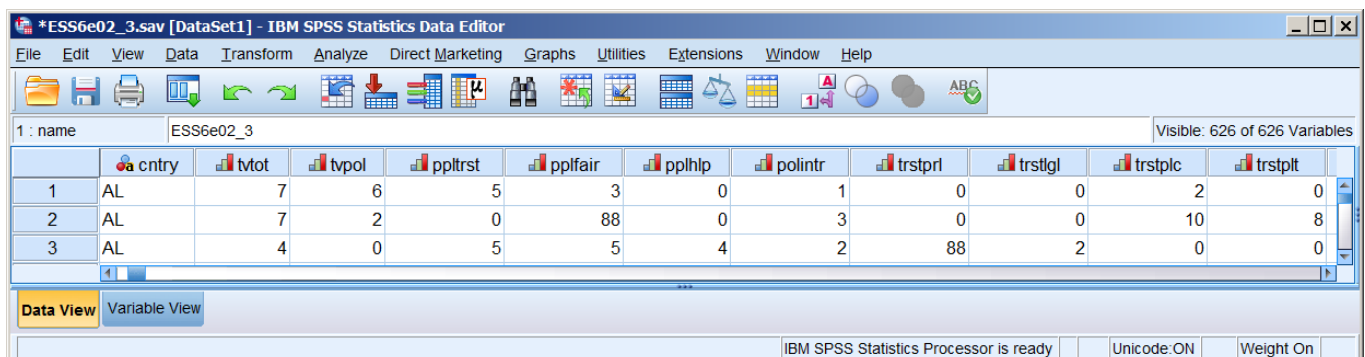
*ESS6e02_3.sav [DataSet1] - IBM SPSS Statistics Data Editor											
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Extensions Window Help											
1 : cldgng Visible: 626 of 626 Variables											
	cldgng	enrglot	fltanx	fltpcfl	dcMf	lchshcp	accng	wrbknrm	lrrnttf	pplahp	trtrsp
1	.	3	2	3	1	2	2	4	4	1	.
2	.	1	4	2	1	1	2	2	6	0	.

". . and there's all my system missing . .". Scroll down to row **1201**



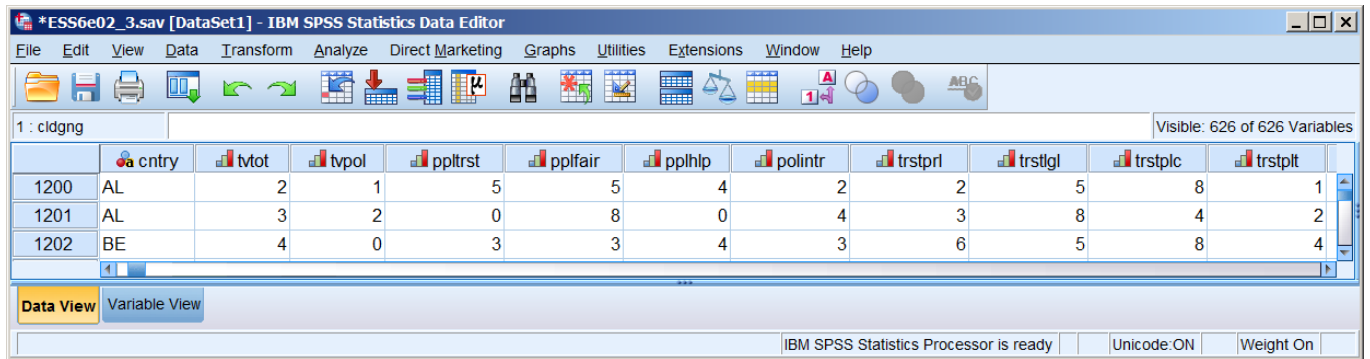
*ESS6e02_3.sav [DataSet1] - IBM SPSS Statistics Data Editor											
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Extensions Window Help											
1 : cldgng Visible: 626 of 626 Variables											
	cldgng	enrglot	fltanx	fltpcfl	dcMf	lchshcp	accng	wrbknrm	lrrnttf	pplahp	trtrsp
1200	.	2	2	1	4	1	2	2	5	3	.
1201	.	2	1	3	1	4	2	4	4	2	.
1202	2	3	2	3	2	3	2	4	4	3	.

If you scroll up and down, you will see that, in rows 1 to 1201, all cases of **[cldgng]** are system missing. JM then scrolls across to find the **[cntry]** column:



*ESS6e02_3.sav [DataSet1] - IBM SPSS Statistics Data Editor											
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Extensions Window Help											
1 : name ESS6e02_3 Visible: 626 of 626 Variables											
	cntry	tvot	tpol	ppltrst	pplfair	pplhlp	polintr	trstprl	trsttgl	trstplc	trstplt
1	AL	7	6	5	3	0	1	0	0	2	0
2	AL	7	2	0	88	0	3	0	0	10	8
3	AL	4	0	5	5	4	2	88	2	0	0

.. scrolls up and down, and finds that, in rows 1 to 1201, all cases of **[cntry]** have string value "AL":



*ESS6e02_3.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Extensions Window Help

1 : cldgng Visible: 626 of 626 Variables

	cuntry	tvot	tpol	ppltrst	pplfair	pplhlp	polintr	trstprl	trstlgl	trstplc	trstplt
1200	AL	2	1	5	5	4	2	2	5	8	1
1201	AL	3	2	0	8	0	4	3	8	4	2
1202	BE	4	0	3	3	4	3	6	5	8	4

Data View Variable View

IBM SPSS Statistics Processor is ready Unicode: ON Weight On

JM searches the documentation for country reports, and discovers that, in Albania, this question was not even asked (or answered). The problem was caused by a CAPI software error, so he proposes to drop Albania for the analysis of "depression" and will use **SELECT CASES**.

Comment:

As an exercise in data analysis (exploring the symptoms of "depression") the point of this video tutorial is not immediately clear: it's more of a technical demonstration in navigation of the **Data Editor**.

A less obvious way of tracking down the culprit is:

frequencies cntry.
crosstabs cntry **by** cldgng .

Total column from **Frequencies**

Valid	AL Albania	223
	BE Belgium	921
	BG Bulgaria	635
	CH Switzerland	676
	CY Cyprus	72
	CZ Czech Republic	896
	DE Germany	7101
	DK Denmark	459
	EE Estonia	113
	ES Spain	3917
	FI Finland	451
	FR France	5319
	GB United Kingdom	5217
	HU Hungary	852
	IE Ireland	359
	IL Israel	563
	IS Iceland	25
	IT Italy	5229
	LT Lithuania	256
	NL Netherlands	1383
	NO Norway	406
	PL Poland	3272
	PT Portugal	898
	RU Russian Federation	12139
	SE Sweden	790
	SI Slovenia	176
	SK Slovakia	457
	UA Ukraine	3892
	XK Kosovo	135
	Total	56835

Total column from **Crosstabs**

	BE Belgium	919
	BG Bulgaria	599
	CH Switzerland	674
	CY Cyprus	71
	CZ Czech Republic	873
	DE Germany	7086
	DK Denmark	457
	EE Estonia	112
	ES Spain	3908
	FI Finland	448
	FR France	5307
	GB United Kingdom	5192
	HU Hungary	846
	IE Ireland	357
	IL Israel	551
	IS Iceland	24
	IT Italy	5141
	LT Lithuania	250
	NL Netherlands	1383
	NO Norway	404
	PL Poland	3222
	PT Portugal	897
	RU Russian Federation	11532
	SE Sweden	789
	SI Slovenia	175
	SK Slovakia	449
	UA Ukraine	3610
	XK Kosovo	131
		55407

Albania appears in the frequencies table, but, because all cases of **[cldgng]** for Albania are coded as **SYSMIS**, they are not in the crosstabs table,

End of: **5.1.2 Identifying the missing cases on [cldgng]**

Forward to: **5.1.3 Creating a numeric version of country**

Back to: [MacInnes \(2017\)](#)

Back to: **5.1.1 Downloading the full ESS6 dataset & documentation**