[Commentary by **John F Hall**] [Last updated: 10 August 2017]

John MacInnes
An Introduction to Secondary Data Analysis with IBM SPSS Statistics
(Sage, Dec. 2017)

Chapter 4: Getting Started with SPSS

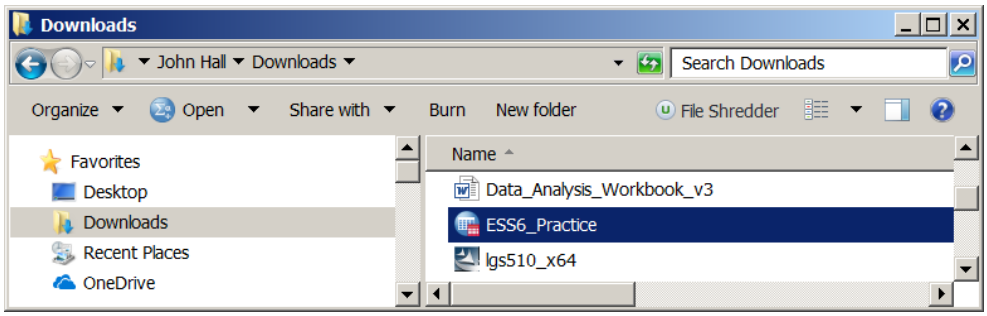
4.1.4: Checking the SPSS files

The following notes assume you have access to SPSS and are familiar with highlighting, dragging and copy/paste. They are based on an earlier communication with John MacInnes and Sage when I first accessed the companion website. They are offered not as criticism, but as constructive supplementary comments intended to help guide users through Chapter 4.

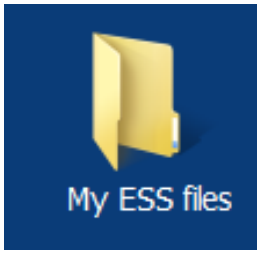
Getting Started with SPSS (Chapter 4) uses two files from the companion website: an SPSS saved file **ESS_Practice.sav** and **Syntax_Ch4.sav**, which is **not** an SPSS *.sav saved file, but a Notepad *.txt plain text file containing lines of SPSS "syntax": it really should be called **ESS_Practice.txt**.

For use in (hypothetical) teaching these are the steps I followed and would also recommend to users of the book, especially if they are new to survey data and/or to SPSS.

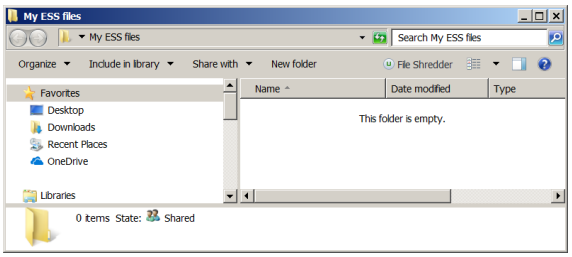
1: Download **ESS_Practice.sav** from the companion website. (See **4.1.2 Downloading the European Social Survey Practice File**)



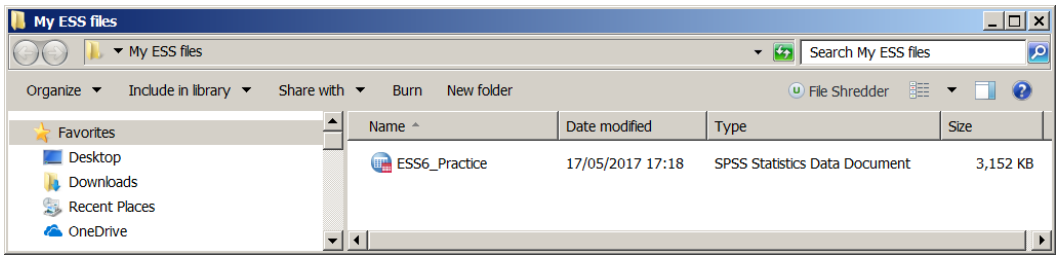
2: If you haven't already done so, create a new folder **My ESS files** in your Desktop

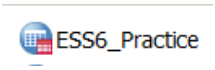


and open it



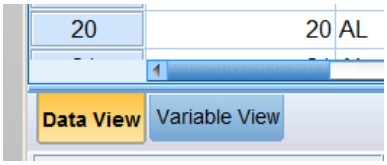
3: Copy **ESS_Practice.sav** from your download folder to folder **My ESS files**



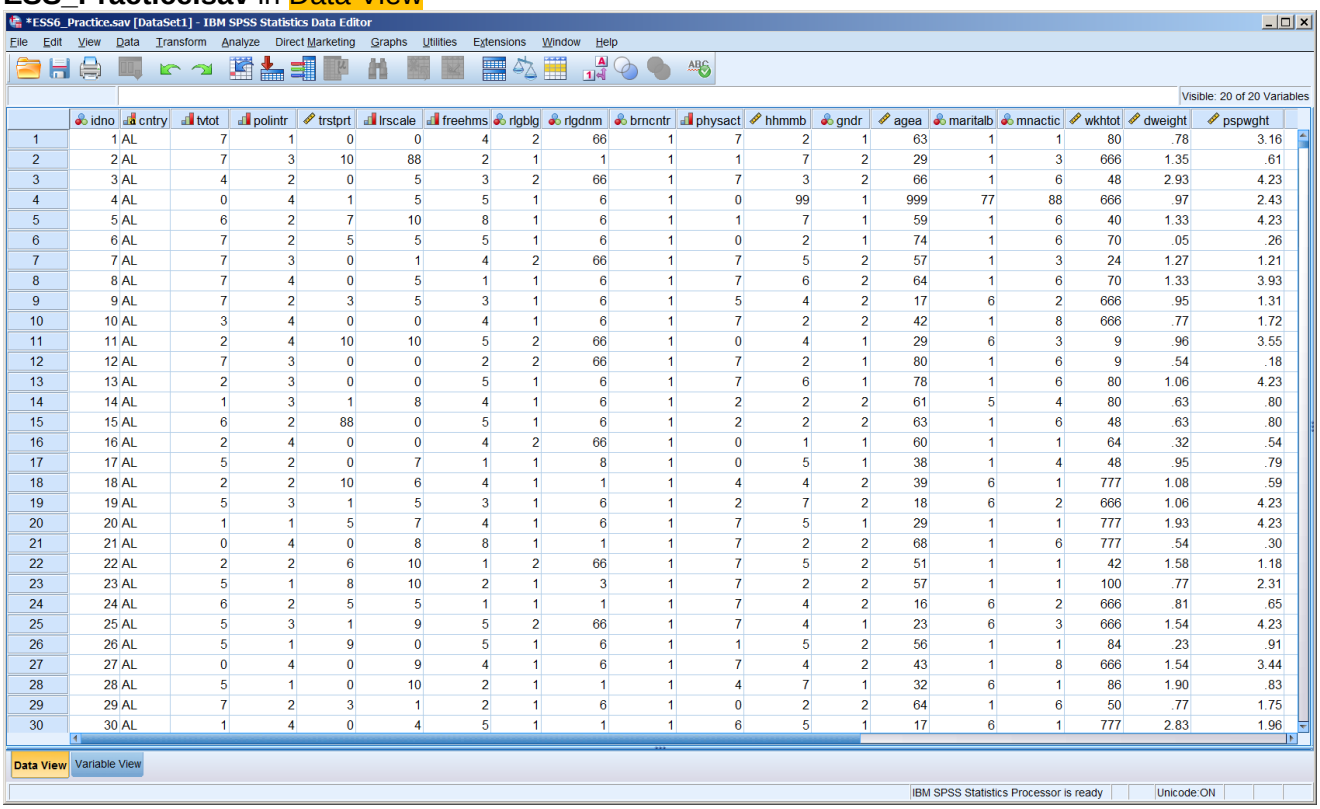
4: Double-click on 

to open the file

ESS_Practice.sav opens in Data View (see bottom left corner)



ESS_Practice.sav in Data View



	idno	cntry	tvtot	polintr	trstprt	lrscale	freehms	rlgblg	rlgdnm	brncntr	physact	hhmmb	gndr	agea	maritalb	mnactic	wkhtot	dweight	pspwght
1	1	AL	7	1	0	0	4	2	66	1	7	2	1	63	1	1	80	.78	3.16
2	2	AL	7	3	10	88	2	1	1	1	1	7	2	29	1	3	666	1.35	.61
3	3	AL	4	2	0	5	3	2	66	1	7	3	2	66	1	6	48	2.93	4.23
4	4	AL	0	4	1	5	5	1	6	1	0	99	1	999	77	88	666	.97	2.43
5	5	AL	6	2	7	10	8	1	6	1	1	7	1	59	1	6	40	1.33	4.23
6	6	AL	7	2	5	5	5	1	6	1	0	2	1	74	1	6	70	.05	.26
7	7	AL	7	3	0	1	4	2	66	1	7	5	2	57	1	3	24	1.27	1.21
8	8	AL	7	4	0	5	1	1	6	1	7	6	2	64	1	6	70	1.33	3.93
9	9	AL	7	2	3	5	3	1	6	1	5	4	2	17	6	2	666	.95	1.31
10	10	AL	3	4	0	0	4	1	6	1	7	2	2	42	1	8	666	.77	1.72
11	11	AL	2	4	10	10	5	2	66	1	0	4	1	29	6	3	9	.96	3.55
12	12	AL	7	3	0	0	2	2	66	1	7	2	1	80	1	6	9	.54	.18
13	13	AL	2	3	0	0	5	1	6	1	7	6	1	78	1	6	80	1.06	4.23
14	14	AL	1	3	1	8	4	1	6	1	2	2	2	61	5	4	80	.63	.80
15	15	AL	6	2	88	0	5	1	6	1	2	2	2	63	1	6	48	.63	.80
16	16	AL	2	4	0	0	4	2	66	1	0	1	1	60	1	1	64	.32	.54
17	17	AL	5	2	0	7	1	1	8	1	0	5	1	38	1	4	48	.95	.79
18	18	AL	2	2	10	6	4	1	1	1	4	4	2	39	6	1	777	1.08	.59
19	19	AL	5	3	1	5	3	1	6	1	2	7	2	18	6	2	666	1.06	4.23
20	20	AL	1	1	5	7	4	1	6	1	7	5	1	29	1	1	777	1.93	4.23
21	21	AL	0	4	0	8	8	1	1	1	7	2	2	68	1	6	777	.54	.30
22	22	AL	2	2	6	10	1	2	66	1	7	5	2	51	1	1	42	1.58	1.18
23	23	AL	5	1	8	10	2	1	3	1	7	2	2	57	1	1	100	.77	2.31
24	24	AL	6	2	5	5	1	1	1	1	7	4	2	16	6	2	666	.81	.65
25	25	AL	5	3	1	9	5	2	66	1	7	4	1	23	6	3	666	1.54	4.23
26	26	AL	5	1	9	0	5	1	6	1	1	5	2	56	1	1	84	.23	.91
27	27	AL	0	4	0	9	4	1	6	1	7	4	2	43	1	8	666	1.54	3.44
28	28	AL	5	1	0	10	2	1	1	1	4	7	1	32	6	1	86	1.90	.83
29	29	AL	7	2	3	1	2	1	6	1	0	2	2	64	1	6	50	.77	1.75
30	30	AL	1	4	0	4	5	1	1	1	6	5	1	17	6	1	777	2.83	1.96

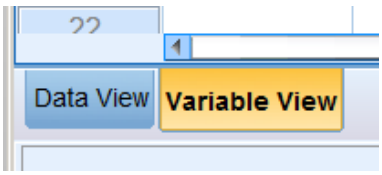
[**NB:** I have narrowed the columns to include all variables and still keep the icons on a single row]

Scroll up and down the file or use CTRL+End and you will see there are 54,673 **cases**:

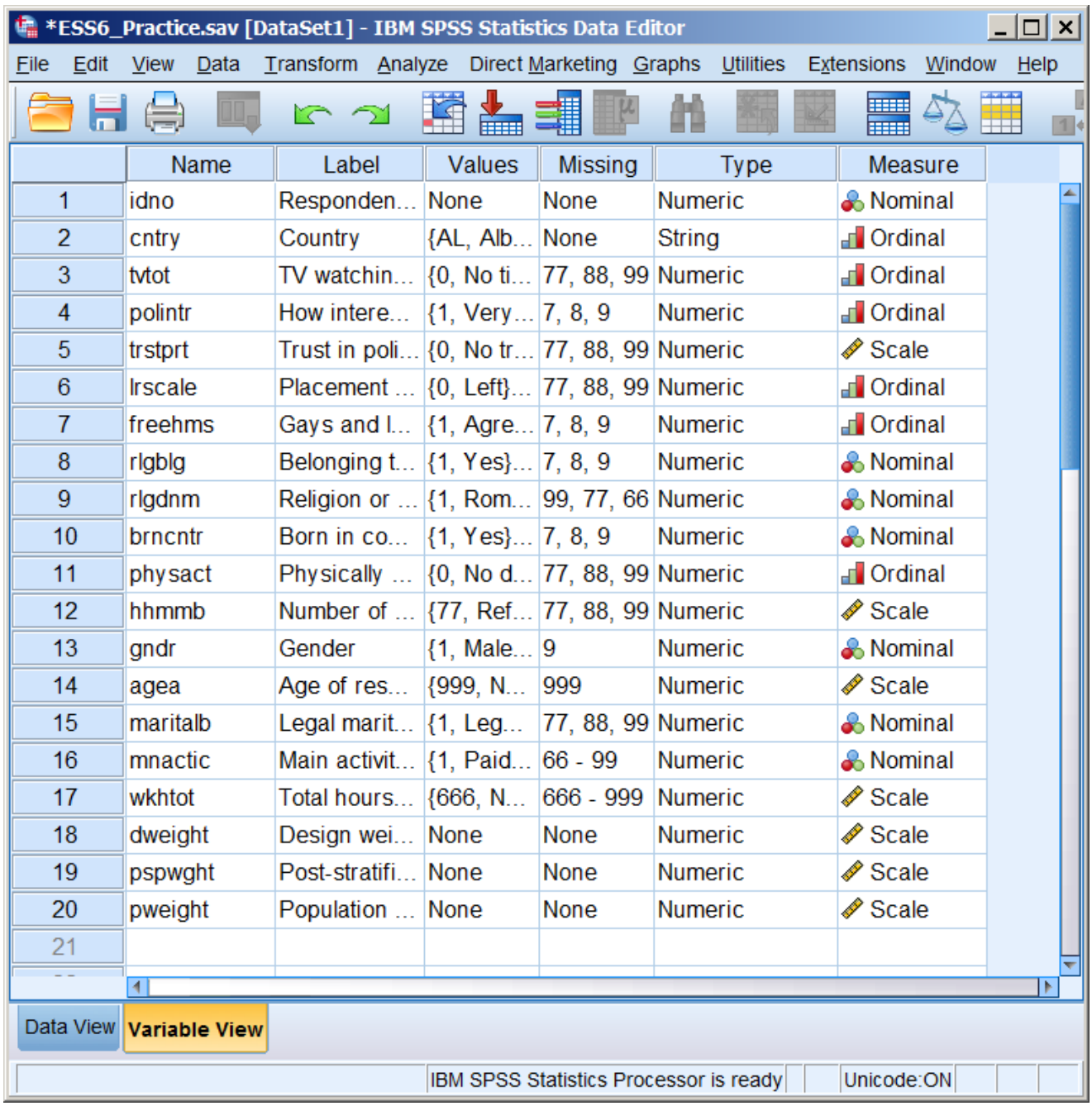
54671	1	2	52	6	6	888	.03	.04	.00
54672	1	1	69	4	6	888	.03	.04	.00
54673	5	1	29	1	4	888	.17	.23	.02
54674									

This doesn't really tell you anything so:

5: Click on Variable View



ESS_Practice.sav in Variable View

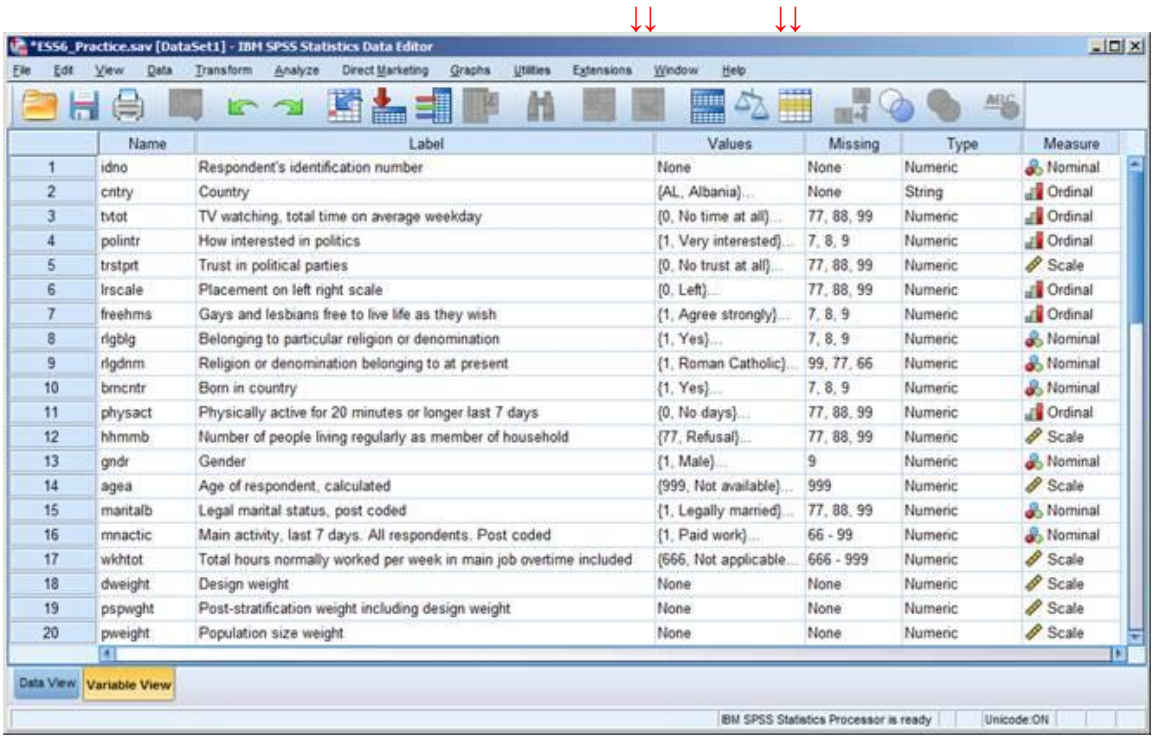


	Name	Label	Values	Missing	Type	Measure
1	idno	Responden...	None	None	Numeric	Nominal
2	cntry	Country	{AL, Alb...	None	String	Ordinal
3	tvtot	TV watchin...	{0, No ti...	77, 88, 99	Numeric	Ordinal
4	polintr	How intere...	{1, Very...	7, 8, 9	Numeric	Ordinal
5	trstprt	Trust in poli...	{0, No tr...	77, 88, 99	Numeric	Scale
6	lrscale	Placement ...	{0, Left}...	77, 88, 99	Numeric	Ordinal
7	freehms	Gays and l...	{1, Agre...	7, 8, 9	Numeric	Ordinal
8	rlgblg	Belonging t...	{1, Yes}...	7, 8, 9	Numeric	Nominal
9	rlgdnm	Religion or ...	{1, Rom...	99, 77, 66	Numeric	Nominal
10	brncntr	Born in co...	{1, Yes}...	7, 8, 9	Numeric	Nominal
11	physact	Physically ...	{0, No d...	77, 88, 99	Numeric	Ordinal
12	hhmmb	Number of ...	{77, Ref...	77, 88, 99	Numeric	Scale
13	gndr	Gender	{1, Male...	9	Numeric	Nominal
14	agea	Age of res...	{999, N...	999	Numeric	Scale
15	maritalb	Legal marit...	{1, Leg...	77, 88, 99	Numeric	Nominal
16	mnactic	Main activit...	{1, Paid...	66 - 99	Numeric	Nominal
17	wkhtot	Total hours...	{666, N...	666 - 999	Numeric	Scale
18	dweight	Design wei...	None	None	Numeric	Scale
19	pspwght	Post-stratifi...	None	None	Numeric	Scale
20	pweight	Population ...	None	None	Numeric	Scale
21						

. . from which you can see there 20 **variables**.

[NB: I always open in Variable View as it is more informative (and intuitive for sociology students).]

6: Drag the column separators to the right to widen the **Label** and **Values** columns:

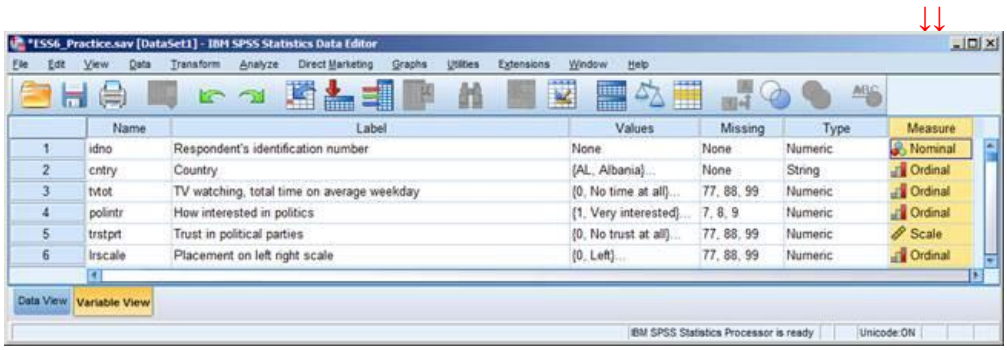


	Name	Label	Values	Missing	Type	Measure
1	idno	Respondent's identification number	None	None	Numeric	Nominal
2	cntry	Country	{AL, Albania}...	None	String	Ordinal
3	tvtot	TV watching, total time on average weekday	{0, No time at all}...	77, 88, 99	Numeric	Ordinal
4	polintr	How interested in politics	{1, Very interested}...	7, 8, 9	Numeric	Ordinal
5	trstprt	Trust in political parties	{0, No trust at all}...	77, 88, 99	Numeric	Scale
6	lrscale	Placement on left right scale	{0, Left}...	77, 88, 99	Numeric	Ordinal
7	freehms	Gays and lesbians free to live life as they wish	{1, Agree strongly}...	7, 8, 9	Numeric	Ordinal
8	rlgblg	Belonging to particular religion or denomination	{1, Yes}...	7, 8, 9	Numeric	Nominal
9	rlgdnm	Religion or denomination belonging to at present	{1, Roman Catholic}...	99, 77, 66	Numeric	Nominal
10	brncntr	Born in country	{1, Yes}...	7, 8, 9	Numeric	Nominal
11	physact	Physically active for 20 minutes or longer last 7 days	{0, No days}...	77, 88, 99	Numeric	Ordinal
12	hhmmb	Number of people living regularly as member of household	{77, Refusal}...	77, 88, 99	Numeric	Scale
13	gndr	Gender	{1, Male}...	9	Numeric	Nominal
14	agea	Age of respondent, calculated	{999, Not available}...	999	Numeric	Scale
15	maritalb	Legal marital status, post coded	{1, Legally married}...	77, 88, 99	Numeric	Nominal
16	mnactic	Main activity, last 7 days. All respondents. Post coded	{1, Paid work}...	66 - 99	Numeric	Nominal
17	wkhtot	Total hours normally worked per week in main job overtime included	{666, Not applicable...	666 - 999	Numeric	Scale
18	dweight	Design weight	None	None	Numeric	Scale
19	pspwght	Post-stratification weight including design weight	None	None	Numeric	Scale
20	pweight	Population size weight	None	None	Numeric	Scale

[**NB:** I would normally keep **Decimals**, but in a small teaching set this is not an over-riding need]

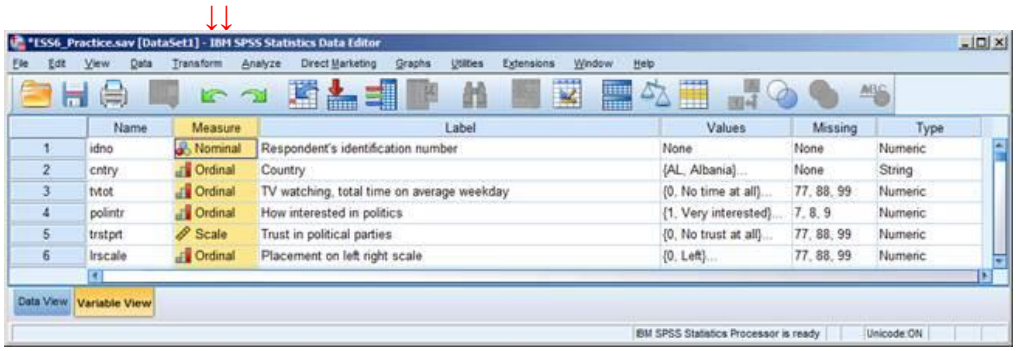
I much prefer working with the variable attributes in a different order, so:

7: Highlight the **Measure** column:



	Name	Label	Values	Missing	Type	Measure
1	idno	Respondent's identification number	None	None	Numeric	Nominal
2	cntry	Country	{AL, Albania}...	None	String	Ordinal
3	tvtot	TV watching, total time on average weekday	{0, No time at all}...	77, 88, 99	Numeric	Ordinal
4	polintr	How interested in politics	{1, Very interested}...	7, 8, 9	Numeric	Ordinal
5	trstprt	Trust in political parties	{0, No trust at all}...	77, 88, 99	Numeric	Scale
6	lrscale	Placement on left right scale	{0, Left}...	77, 88, 99	Numeric	Ordinal

. . then hold left mouse down and drag the column across next to the **Name** column:



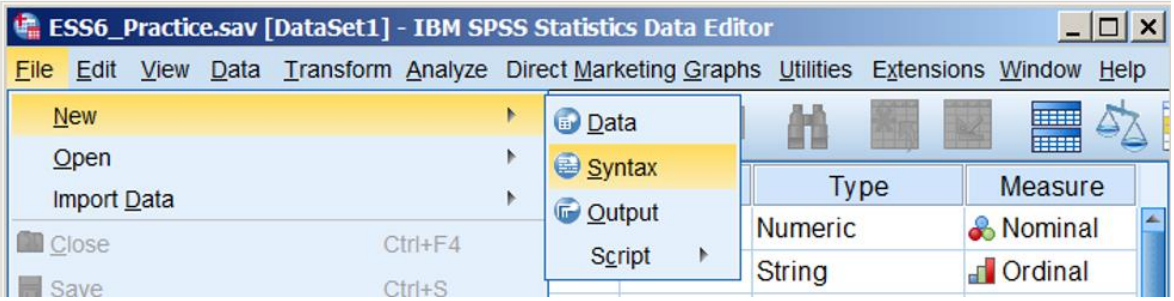
	Name	Measure	Label	Values	Missing	Type
1	idno	Nominal	Respondent's identification number	None	None	Numeric
2	cntry	Ordinal	Country	{AL, Albania}...	None	String
3	tvtot	Ordinal	TV watching, total time on average weekday	{0, No time at all}...	77, 88, 99	Numeric
4	polintr	Ordinal	How interested in politics	{1, Very interested}...	7, 8, 9	Numeric
5	trstprt	Scale	Trust in political parties	{0, No trust at all}...	77, 88, 99	Numeric
6	lrscale	Ordinal	Placement on left right scale	{0, Left}...	77, 88, 99	Numeric

[**NB:** This can also be done with View >> Customize Variable View]

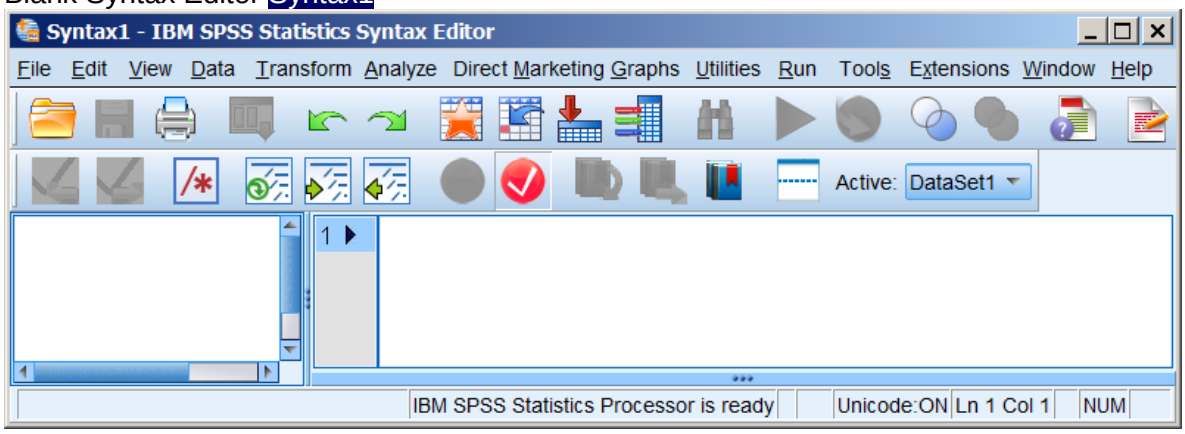
SPSS will retain the column order for the next start-up, but not the column widths.

8: Open a new Syntax Editor

File >> New >> Syntax



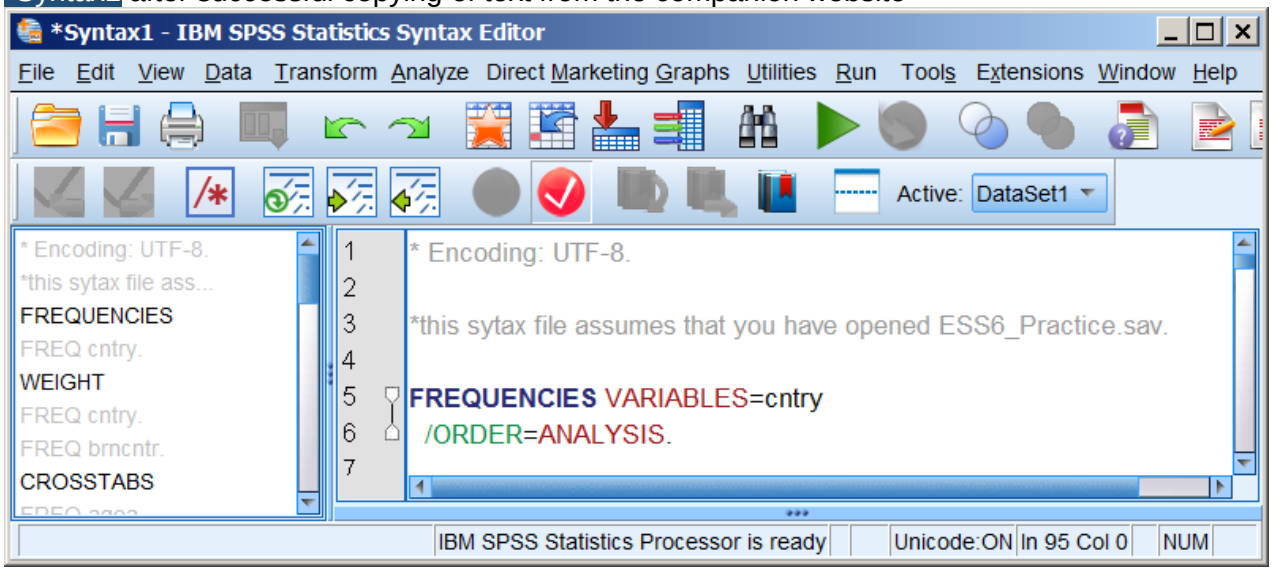
Blank Syntax Editor Syntax1



9: Copy the contents of **Syntax_Ch4.sav** into Syntax1.

This is more complicated than it sounds: for details of how to do it see **4.1.3 Downloading the SPSS syntax**. If anything is written or copied to Syntax1 the name will change to *Syntax1 as it becomes a temporary working file until it is saved.

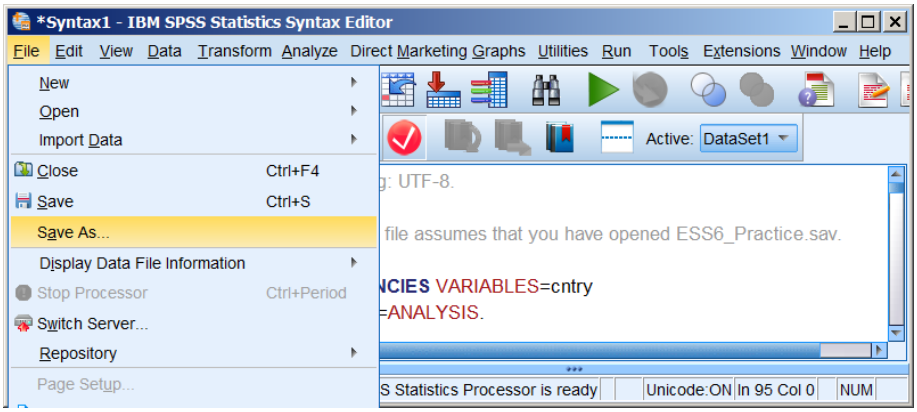
*Syntax1 after successful copying of text from the companion website



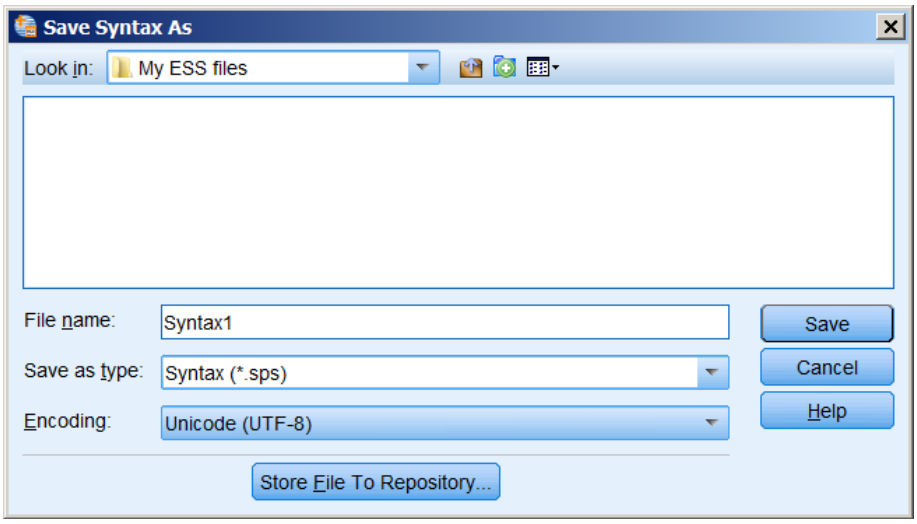
Scroll up and down the file to see what's in it: note the colour coding of the syntax (except where abbreviations are used). The colour of the comments can be changed to make them more readable (see **4.1.3: Downloading the SPSS syntax** page 8)

10: Save the Syntax Editor

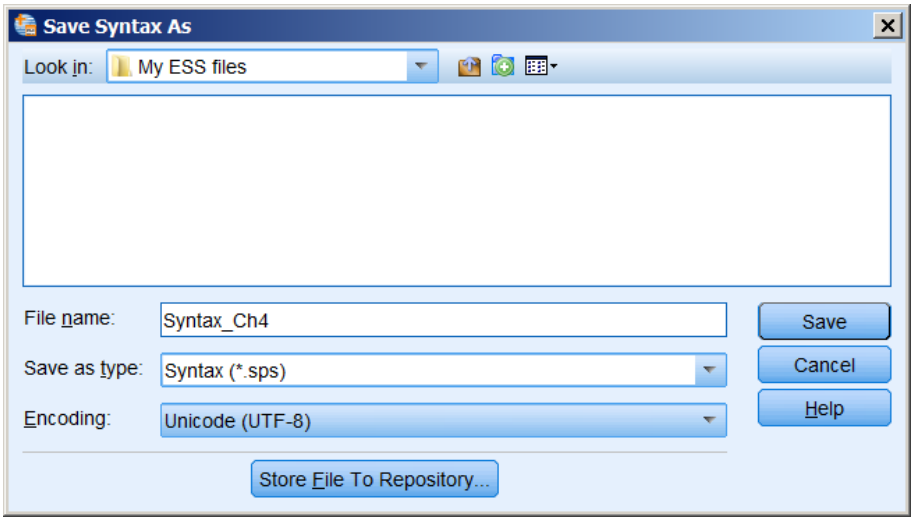
File >> Save As



Navigate to folder **My ESS files** (It seems empty, but it's because there are no ***.sp**s files in it)



Change Syntax1 to Syntax_Ch4



and click on Save

The file name changes from *Syntax1 to Syntax_Ch4.sps

Folder **My ESS files** now contains three files:

Name	Date modified	Type	Size
ESS6_Practice	17/05/2017 17:18	SPSS Statistics Data Document	3,152 KB
Syntax_Ch4.sav	09/06/2017 18:29	Text Document	2 KB
Syntax_Ch4	22/07/2017 17:58	SPSS Statistics Syntax File	2 KB

11: I usually use Edit >> Options to open a new syntax file on start-up and also change the output settings to **Names and Labels** for variables and to **Values and Labels** for values so that I can see what I'm working with. I can always change them back again if I need tables for publication.

12: With a class of newbies I would, like MacInnes, run some sample **FREQUENCIES** and **CROSSTABS** commands, but with different variables, simpler syntax and sticking to default output.

Students in sociology and related areas are unlikely to be very interested in whether respondents in 27 different countries were actually born in those countries. Much more interesting to them are respondents' attitudes and opinions on social and political issues such as trust in political parties or whether they have authoritarian or liberal positions on selected issues.

13: Documentation

I would also encourage them to investigate the contents of the data file with:

```
DISPLAY labels .
```

and

```
DISPLAY dictionary .
```

. . neither of which is available from the Graphic User Interface (GUI): the output can then be printed up and used as documentation.

Except for user-supplied variable and value labels, SPSS syntax is case insensitive: it can also interpret abbreviated commands. Once students have got the hang of opening a syntax editor, writing syntax and then running it, I tend to use lower case only and abbreviated syntax:

. . so it's nice to see MacInnes occasionally using them as well.

DISPLAY labels .

Variable Labels

Variable	Position	Label
idno	1	Respondent's identification number
cntry	2	Country
tvtot	3	TV watching, total time on average weekday
polintr	4	How interested in politics
trstprt	5	Trust in political parties
lrscale	6	Placement on left right scale
freehms	7	Gays and lesbians free to live life as they wish
rlgblg	8	Belonging to particular religion or denomination
rlgdnm	9	Religion or denomination belonging to at present
brncntr	10	Born in country
physact	11	Physically active for 20 minutes or longer last 7 days
hhmmb	12	Number of people living regularly as member of household
gndr	13	Gender
agea	14	Age of respondent, calculated
maritalb	15	Legal marital status, post coded
mnactic	16	Main activity, last 7 days. All respondents. Post coded
wkhtot	17	Total hours normally worked per week in main job overtime included
dweight	18	Design weight
pspwght	19	Post-stratification weight including design weight
pweight	20	Population size weight

Variables in the working file

DISPLAY dictionary .

Variable Information

Variable	Position	Label	Level	Role	Width	Alignment	Print	Write	Missing Values
idno	1	Respondent's identification number	Nominal	Input	5	Right	F9	F9	
cntry	2	Country	Ordinal	Input	5	Left	A2	A2	
tvtot	3	TV watching, total time on average weekday	Ordinal	Input	5	Right	F2	F2	77, 88, 99
polintr	4	How interested in politics	Ordinal	Input	7	Right	F1	F1	7, 8, 9
trstprt	5	Trust in political parties	Scale	Input	6	Right	F2	F2	77, 88, 99
lrscale	6	Placement on left right scale	Ordinal	Input	7	Right	F2	F2	77, 88, 99
freehms	7	Gays and lesbians free to live life as they wish	Ordinal	Input	8	Right	F1	F1	7, 8, 9
rlgblg	8	Belonging to particular religion or denomination	Nominal	Input	6	Right	F1	F1	7, 8, 9
rlgdnm	9	Religion or denomination belonging to at present	Nominal	Input	7	Right	F2	F2	99, 77, 66
brncntr	10	Born in country	Nominal	Input	6	Right	F1	F1	7, 8, 9
physact	11	Physically active for 20 minutes or longer last 7 days	Ordinal	Input	7	Right	F2	F2	77, 88, 99
hhmmb	12	Number of people living regularly as member of household	Scale	Input	7	Right	F2	F2	77, 88, 99
gndr	13	Gender	Nominal	Input	5	Right	F1	F1	9
agea	14	Age of respondent, calculated	Scale	Input	5	Right	F3	F3	999
maritalb	15	Legal marital status, post coded	Nominal	Input	8	Right	F2	F2	77, 88, 99
mnactic	16	Main activity, last 7 days. All respondents. Post coded	Nominal	Input	7	Right	F2	F2	66 through 99
wkhtot	17	Total hours normally worked per week in main job overtime included	Scale	Input	6	Right	F3	F3	666 through 999
dweight	18	Design weight	Scale	Input	7	Right	F4.2	F4.2	
pspwght	19	Post-stratification weight including design weight	Scale	Input	8	Right	F4.2	F4.2	
pweight	20	Population size weight	Scale	Input	7	Right	F8.2	F8.2	

Variables in the working file

[NB: To squeeze in the above table, I had to narrow the column widths and reduce the font to 8 points]

14: SPSS syntax

I always taught, and still teach, SPSS as a **language**, using SPSS **commands** direct, accepting default outputs, and only gradually adding options, statistics and abbreviated syntax as and when I feel students are ready. See: 1.3.3.2 Introduction to SPSS syntax

However, MacInnes has frequently used the **PASTE** facility to generate SPSS syntax (in my view unnecessarily complex and detailed) from the GUI. If I had my way, I would hide the GUI menu completely from beginners. They invariably start clicking on everything and quickly get completely lost. I do actually use the GUI myself, but only when absolutely necessary, when it is easier and quicker or when it can do something I can't do in syntax.

15: Syntax errors on the companion website:

In file **Syntax_Ch4.sps** you will find a small error in lines 22 and 23:

```
22  FREQ agea.
23    /histogram.
```

. . which I didn't spot until I had created a frequency table of all respondents' ages.

There's a stop after **agea** so /histogram should have caused an error, but I only clicked on the first line. If I highlight both lines and run them an error message appears;

```
  /histogram.

Error # 1.  Command name: /histogram
The first word in the line is not recognized as an SPSS Statistics command.
Execution of this command stops.
```

If you delete the stop after **agea**, the syntax produces a histogram as well as the frequency table. I would have used a different, more interesting variable or suppressed the frequency count with:

```
frequencies agea
        /format notable
        /histogram.
```

There's another error in lines 42 and 43:

```
40  RECODE freehms (1 2= 1) (4 5 =2) (ELSE = SYSMIS) into freehms3.
41  VARIABLE LABELS freehms3 "recoded freehms".
42  VALUE LABELS freehm3 1 "Agree" 2 "Disagree".
43  FREQ freehm3.
```

[**NB:** Recoding to SYSMIS is not good practice: best to keep user-defined missing values]

I was about say that, after **RECODE** <varname1> into <varname 2> I would normally use:
FORMATS <varname2> (f2.0) .

. . but when I ran the syntax:

```
RECODE freehms (1 2= 1) (4 5 =2) (ELSE = SYSMIS) into freehms3.
VARIABLE LABELS freehms3 "recoded freehms".
VALUE LABELS freehm3 1 "Agree" 2 "Disagree".
```

I got:

```
Warning # 4474.  Command name: VALUE LABELS
The (ADD) VALUE LABELS command specifies an unknown variable name.  The name
will be ignored.
The error is associated with 'freehm3'
FREQ freehm3.
```

Warnings

Error in FREQUENCIES command.
Execution of this command stops.
The FREQUENCIES command requires a variable list. The most common cause of this error is a variable list that contains one or more undefined variables, and the variable list is not preceded by the optional VARIABLES keyword.

It took me quite a while to spot it, but an "**s**" has been left out (a sure sign of haste without checking).

FREQ freehm3 .

. . should be

FREQ freehm**s**3.

freehms3 recoded freehms

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	33321	60.9	76.0	76.0
	2.00	10550	19.3	24.0	100.0
	Total	43871	80.2	100.0	
Missing	System	10802	19.8		
Total		54673	100.0		

Like I said, to get rid of the superfluous decimals in the value labels, it needs:

FORMATS freehms3 (f2.0) .

freehms3 recoded freehms

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	33321	60.9	76.0	76.0
	2	10550	19.3	24.0	100.0
	Total	43871	80.2	100.0	
Missing	System	10802	19.8		
Total		54673	100.0		

Actually there are no value labels either because an "**s**" got left out there as well.

VALUE LABELS freehm3 1 "Agree" 2 "Disagree".

. . should be:

VALUE LABELS freehm**s**3 1 "Agree" 2 "Disagree".

. . which does the trick:

freehms3 recoded freehms

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Agree	33321	60.9	76.0	76.0
	2 Disagree	10550	19.3	24.0	100.0
	Total	43871	80.2	100.0	
Missing	System	10802	19.8		
Total		54673	100.0		

For publication, the output settings need changing to **Labels** and **freehms3** needs a more informative variable label, eg:

VARIABLE LABELS freehms3 "Gays and lesbians free to live life as they wish (recoded)".

The full set of syntax should then read:

```
RECODE freehms (1 2= 1) (4 5 =2) (ELSE = SYSMIS) into freehms3.
FORMATS freehms3 (F2.0) .
VARIABLE LABELS freehms3 "Gays and lesbians free to live life as they wish (recoded)".
VALUE LABELS freehms3 1 "Agree" 2 "Disagree".
FREQUENCIES        freehms3 .
```

to yield:

Gays and lesbians free to live life as they wish (recoded)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	33321	60.9	76.0	76.0
	Disagree	10550	19.3	24.0	100.0
	Total	43871	80.2	100.0	
Missing	System	10802	19.8		
Total		54673	100.0		

These errors need to be corrected on the website before I add yet more slides to SPSS usage in major surveys ! I'll work my way through the rest of the site and keep JM posted with anything I find.

16: General points

Recoding to SYSMIS is not recommended professional practice. Experienced SPSS users always keep to user-defined missing values, in this case **7** "Refused" **8** "Don't know" and **9** "No answer".

In the table above 7,691 respondents with value **3** "Neither agree nor disagree" have been completely omitted because value **3** has been recoded to SYSMIS.

It would perhaps have been better to start with:

FREQUENCIES freehms.

Gays and lesbians free to live life as they wish

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree strongly	15880	29.0	30.8	30.8
	Agree	17441	31.9	33.8	64.6
	Neither agree nor disagree	7691	14.1	14.9	79.5
	Disagree	5075	9.3	9.8	89.4
	Disagree strongly	5475	10.0	10.6	100.0
	Total	51562	94.3	100.0	
Missing	Refusal	151	.3		
	Don't know	2900	5.3		
	No answer	60	.1		
	Total	3111	5.7		
Total		54673	100.0		

. . before deciding on an arbitrary recoding for which no rationale is given and which omits 14.9% of the sample who gave a non-missing answer.

We can then proceed to recode responses with something like:

```
RECODE freehms (1 2= 1) (4 5 =2) (ELSE = COPY) into freehms3.
FORMATS freehms3 (F2.0) .
VARIABLE LABELS freehms3 "Gays and lesbians free to live life as they wish (recoded)".
VALUE LABELS freehms3
  1 "Agree" 2 "Neither agree nor disagree" 3 "Disagree"
  7 "Refused" 8 "Don't know" 9 "No answer".
MISSING VALUES freehms3 (7 8 9).
FREQUENCIES freehms3 .
```

Gays and lesbians free to live life as they wish (recoded)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	33321	60.9	64.6	64.6
	Neither agree nor disagree	10550	19.3	20.5	85.1
	Disagree	7691	14.1	14.9	100.0
	Total	51562	94.3	100.0	
Missing	Refused	151	.3		
	Don't know	2900	5.3		
	No answer	60	.1		
	Total	3111	5.7		
Total		54673	100.0		

I know I'm a complete pedant: I would get out more, but I just ♥ SPSS and want others not only to love it too, but also enjoy pursuing research ideas using real data from real surveys of real people.

End of: **4.1.4: Checking the SPSS files**

Back to: MacInnes (2017)