

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

[**Draft only**. Last updated: 14 June 2018]

John MacInnes

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

(Sage, Dec. 2017)

Chapter 4: Getting Started with SPSS

4.1.5: Guide to video tutorials 7 and 8

Previous guides:

[Aide-mémoire for easier navigation of companion website](#)

[4.1.1 Overview of video tutorials 1 to 6](#)

[4.1.2 Downloading the European Social Survey Practice File](#)

[4.1.3 Downloading the SPSS syntax](#)

[4.1.4 Checking the SPSS files](#)

Page Contents

2	Video 7: Means procedure
2	4.14: Creating a barchart in SPSS (p87)
7	4.15: Using the MEANS procedure (p88)
9	Video 8: Recoding vars
10	4.16: Recoding variables and creating new variables (p 89)
19	4.17 Using Syntax in SPSS

The following notes assume you are familiar with copy/paste, highlighting and dragging with left mouse down and that you have access to a licenced copy of SPSS. They are based on communications with John MacInnes and Sage when I first accessed the companion website and are offered, not as criticism, but as supplementary comments intended to help guide users through Chapter 4.

The direct link to the video tutorials for Chapter 4 is: [Chapter 4 video tutorials](#)

Warning!

All the video tutorials are on the same webpage and can be accidentally triggered by stray mouse pointers. If you are not careful you can have two or more simultaneous commentaries playing and can't always tell which commentary relates to which video. It would be far better to split all these videos across separate pages.

All the syntax for all the analyses is in the same **Syntax Editor**, which can get very complicated and possibly confusing. The syntax needs to be split into separate ***.sps** files for each topic, preferably with file names corresponding to the relevant paragraph in the text:

The rationale for the order of video topics is not particularly evident: in fact, the chapters can be read in almost any order without losing pedagogic efficacy. Some topics could do with much more preliminary explanation of what is being done and why. However, at over 300 pages the book is already quite big and additional material could well make it unwieldy.

Video 7: Means procedure

Data set used: **ESS6_Practice.sav**
 [modified for this guide to **ESS_Practise_jfh.sav** as per
[4.1.2 Downloading the European Social Survey Practice File](#)]

Method: **GUI**
 Variable used: **[Irscale]**
 Procedure: **FREQUENCIES**

[Extract from core questionnaire]


B19 CARD 8 In politics people sometimes talk of “left” and “right”.
 Using this card, where would you place yourself on this scale,
 where 0 means the left and 10 means the right?

Left											Right	(Don't Know)
	00	01	02	03	04	05	06	07	08	09	10	88

Placement on left right scale

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Left	2184	3.8	4.6	4.6
	1	1033	1.8	2.2	6.8
	2	2573	4.5	5.4	12.2
	3	4065	7.2	8.6	20.8
	4	4222	7.4	8.9	29.8
	5	17672	31.1	37.4	67.1
	6	4293	7.6	9.1	76.2
	7	4280	7.5	9.0	85.2
	8	3430	6.0	7.3	92.5
	9	1313	2.3	2.8	95.3
	Right	2241	3.9	4.7	100.0
Missing	Total	47305	83.2	100.0	
	Refusal	910	1.6		
	Don't know	8569	15.1		
	No answer	51	0.1		
	Total	9529	16.8		
Total		56835	100.0		

JM produces a summary table and bar-chart, but doesn't explain, except by inference, the reason for using a bar-chart (rather than a histogram) when he later uses it as **scale** rather than **ordinal**. Strictly speaking **[Irscale]** has no fixed interval between the points and is therefore **ordinal**.

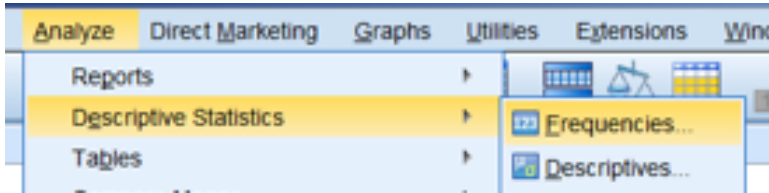
6	Irscale	 Ordinal	Placement on left right scale
---	---------	---	-------------------------------


The points are not contiguous so a barchart is used because it has spaces between the points.

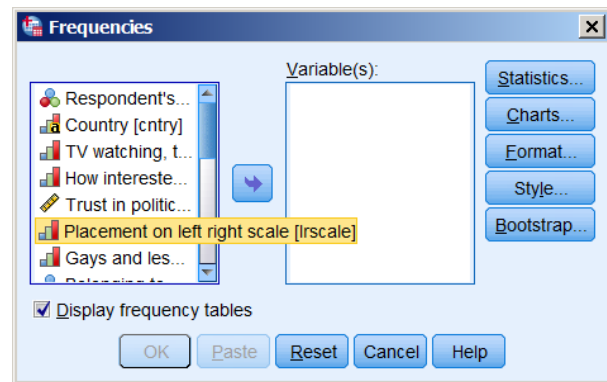
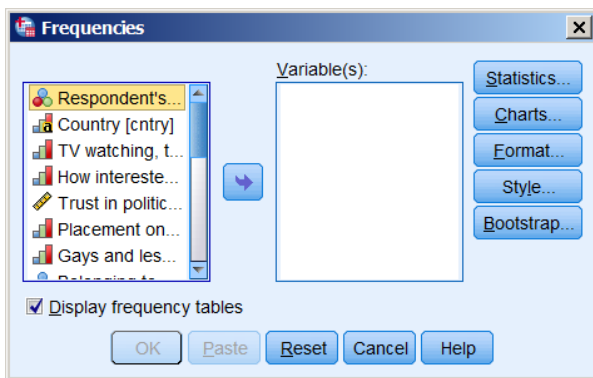
4.14: Creating a barchart in SPSS (p87)

JM doesn't demonstrate this in detail, but suggests:

Analyze >> Descriptive Statistics >> Frequencies



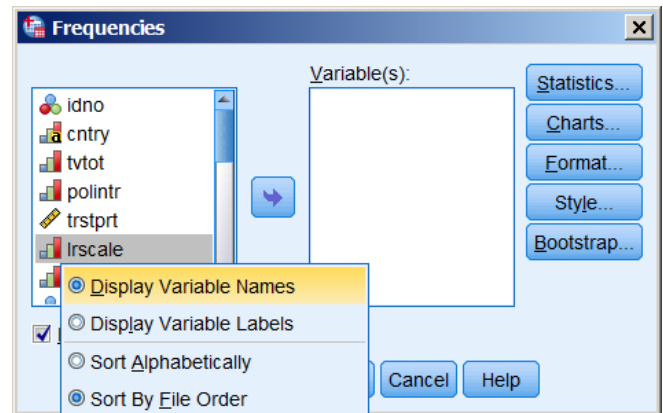
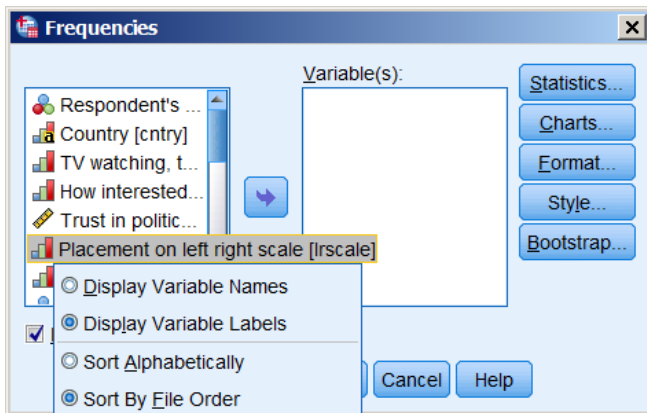
Highlight **Placement on left right scale** and click on  to move it to the Variable(s) box:



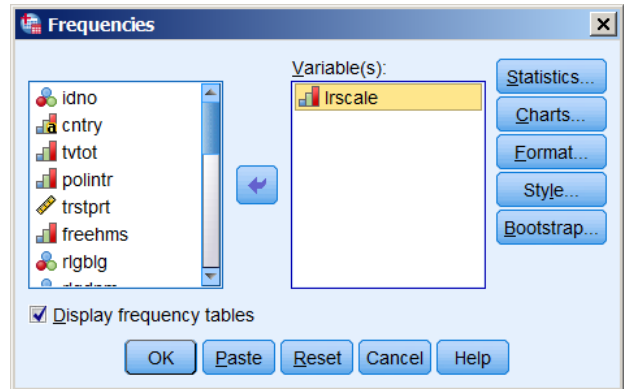
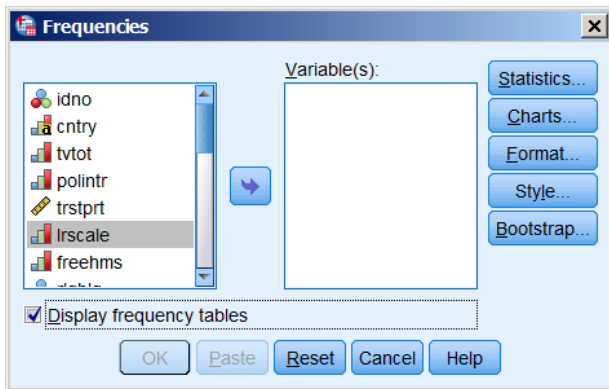
The dialog box displays variable **labels**, but navigation is much easier if you use variable **names**.

Right click on the label:

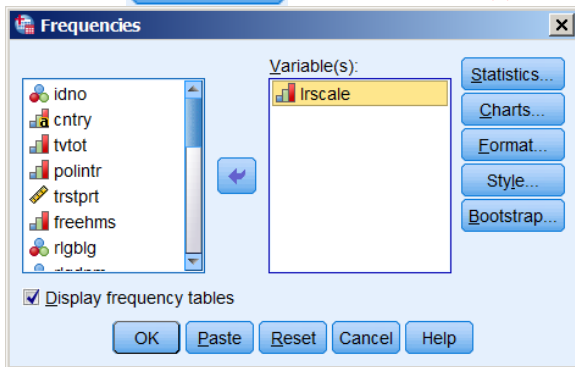
and check **Display Variable Names**



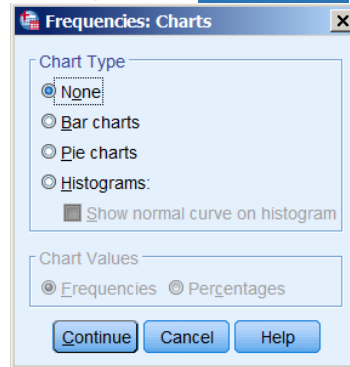
Click on  to move **Irscale** to the **Variable(s)** box:



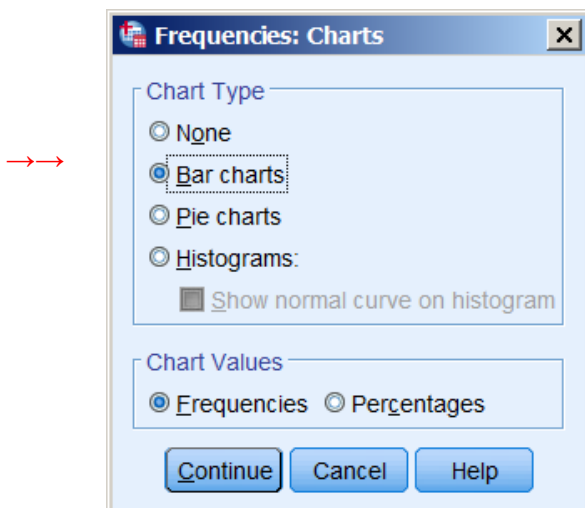
Click on **Charts...**



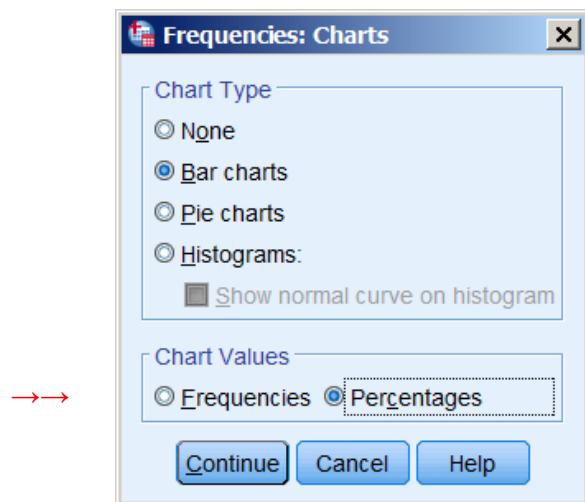
... to open the **Chart Type** dialog box:



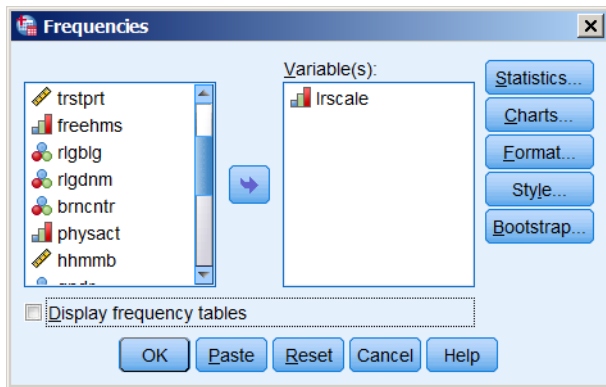
Check **Bar charts**



Check **Percentages**



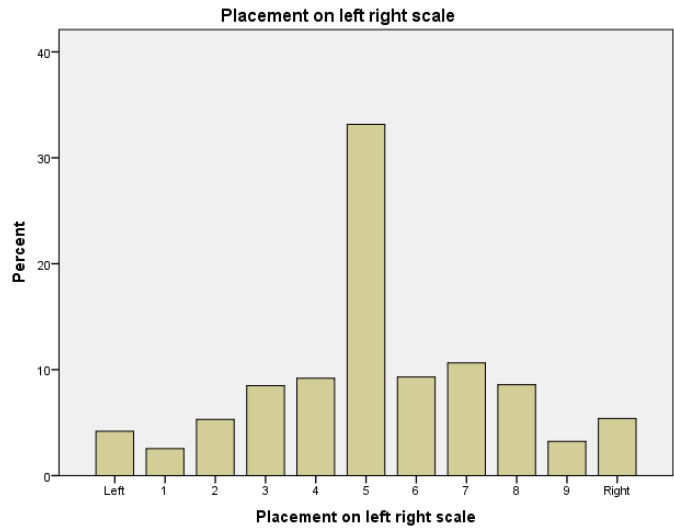
Click on **Continue**




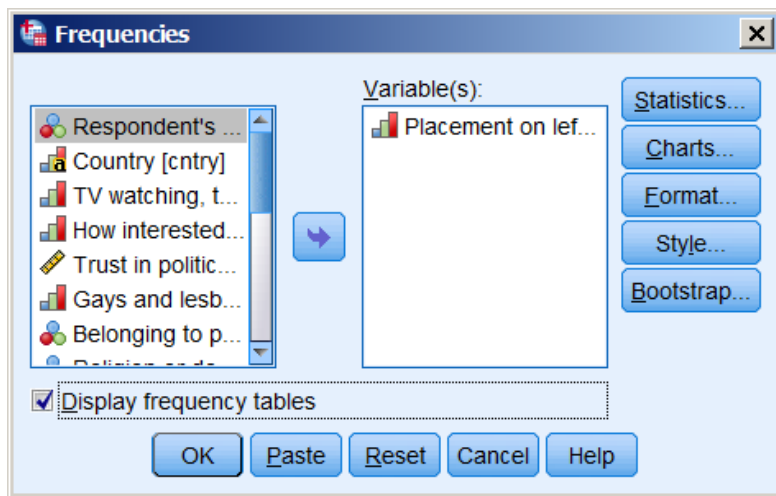
Click on  to get → →

Statistics
Placement on left right scale

N	Valid	47305
	Missing	9529



If you want to see the frequency table as well check  **Display frequency tables**



Click on 

Statistics
Placement on left right scale

N	Valid	47305
	Missing	9529

Placement on left right scale

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Left	2184	3.8	4.6	4.6
	1	1033	1.8	2.2	6.8
	2	2573	4.5	5.4	12.2
	3	4065	7.2	8.6	20.8
	4	4222	7.4	8.9	29.8
	5	17672	31.1	37.4	67.1
	6	4293	7.6	9.1	76.2
	7	4280	7.5	9.0	85.2
	8	3430	6.0	7.3	92.5
	9	1313	2.3	2.8	95.3
	Right	2241	3.9	4.7	100.0
Total		47305	83.2	100.0	
Missing	Refusal	910	1.6		
	Don't know	8569	15.1		
	No answer	51	0.1		
	Total	9529	16.8		
Total		56835	100.0		

The syntax generated by  from the GUI is:

```
FREQUENCIES VARIABLES=Irscale
  /BARCHART PERCENT
  /ORDER=ANALYSIS.
```

If all you want is a bar-chart, there's no need to use the GUI at all: it's much quicker and easier to use syntax:

```
FREQUENCIES =Irscale
  /FORMAT NOTABLE
  /BARCHART PERCENT
```

. . and even quicker if it's abbreviated to:

```
freq Irscale /for not /bar per.
```

. . which produces exactly the same chart, but with far fewer key depressions.

4.15: Using the MEANS procedure (p88)

Variables: [freehms] [lrscale] [agea]

Procedure: MEANS

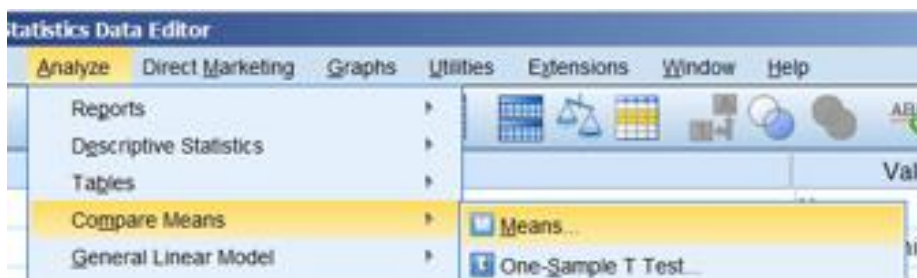
Method: GUI

Task: Compare mean [lrscale] and [agea] for each level of [freehms]

This analysis does not make sociological sense, making [lrscale] and [age] dependent variables and calculating means for each category of [freehms]. Variable [age] is unlikely to be caused by [freehms] as any causal link will be in the other direction.

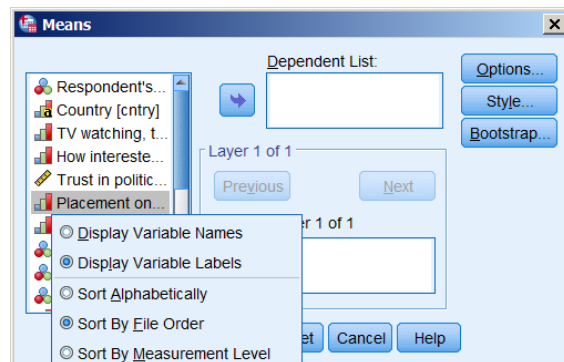
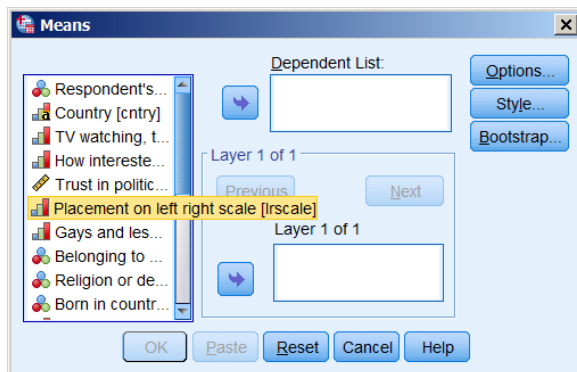
Refers to barchart in video 6

Analyze >> Compare Means >> Means




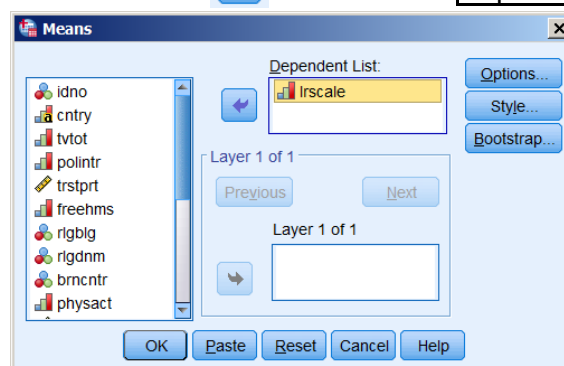
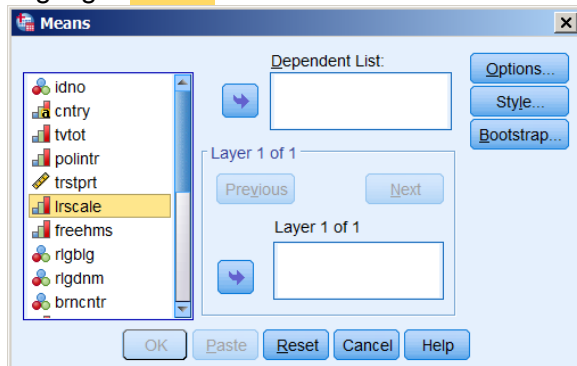
Difficult to navigate with labels, so right click:
on Placement on left right scale [lrscale]

Click on ☐ Display Variable Names

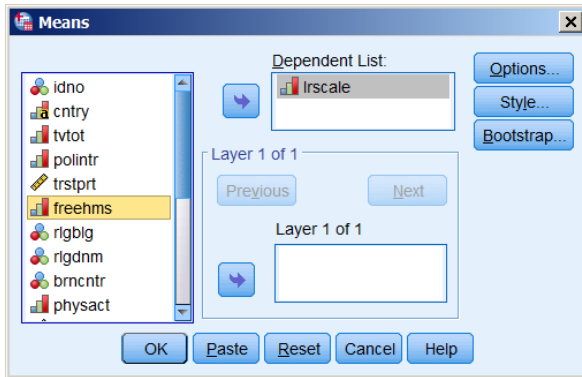



Highlight lrscale

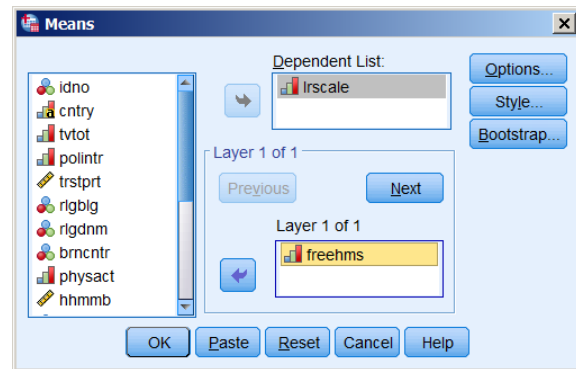
.. and click on  to move it to **Dependent list**



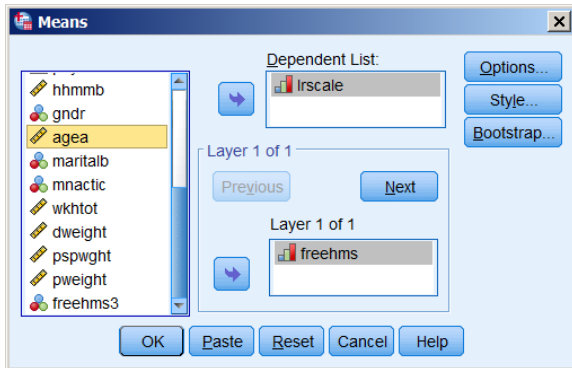
Highlight **freehms**




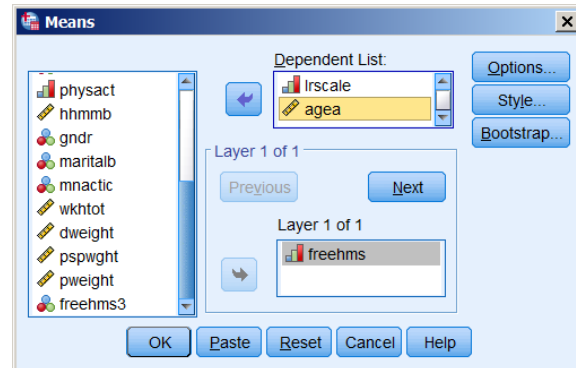
.. and click on  to move it to **Layer 1 of 1**



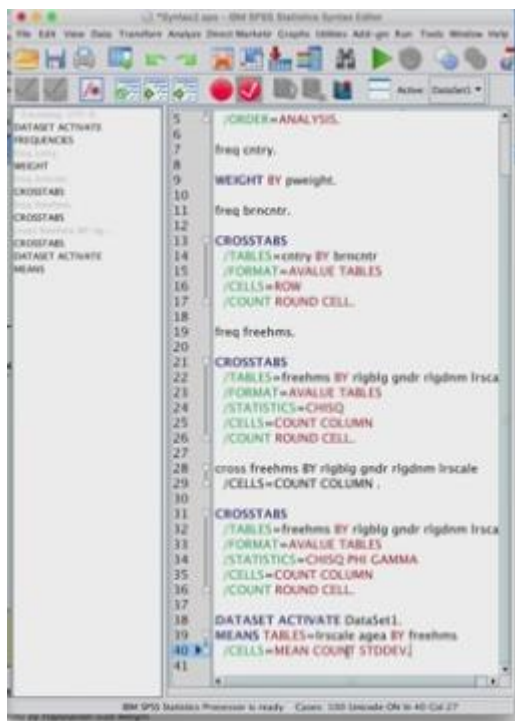
Highlight **agea**



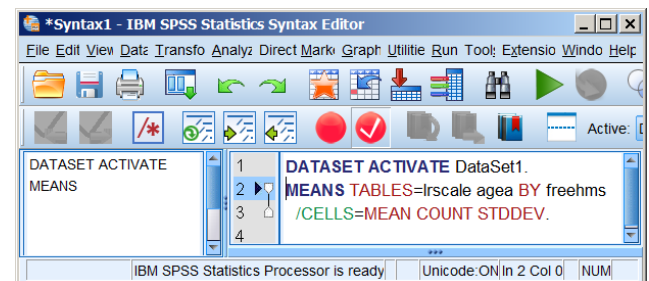
.. and click on  to move it to **Dependent List**



JM then clicks on **Paste**



The syntax generated is appended to the syntax editor, but it's difficult to see on the screen as it's very small and right at the bottom.



Running the syntax produces the following table, which doesn't tell you very much, except that people who disagree strongly (that gays and lesbians should be free to live life as they wish) describe themselves as slightly more to the political right and are slightly older.

Report

Gays and lesbians free to live life as they wish		Placement on left right scale	Age of respondent, calculated
Agree strongly	Mean	4.53	44.47
	N	14099	15475
	Std. Deviation	2.342	17.691
Agree	Mean	5.19	48.75
	N	15487	17585
	Std. Deviation	2.152	18.946
Neither agree nor disagree	Mean	5.39	47.12
	N	6447	7886
	Std. Deviation	2.073	19.050
Disagree	Mean	5.54	48.43
	N	4607	5753
	Std. Deviation	2.286	19.647
Disagree strongly	Mean	5.52	49.26
	N	4718	6432
	Std. Deviation	2.346	19.381
Total	Mean	5.08	47.29
	N	45358	53131
	Std. Deviation	2.270	18.833

To be honest I can't see the point of this example. **[freehms]** is unlikely to affect **[agea]** (except by DUP¹-style apoplexy?) and is much more likely to be a component of the multi-dimensional construct underlying **[lrsc]**. It would be better to treat subjective measures as dependent variables and classification variables (country, sex, grouped age) as independent.

There is no need for **MEANS** at this stage. It loses sight of the **shape** of the distribution. It's much easier for beginners to compare **percentages** of **[freehms]** within age groups using **CROSSTABS** or charts. However, the practice data set only has actual age in years, so it would first need to be recoded into (not too many) age groups.

¹ Democratic Unionist Party in Northern Ireland.

Video 8: Recoding vars

Data set used: ESS6_Practice.sav
Method: GUI
Variable used: [freehms]
Procedures: RECODE
CROSSTABS

4.16: Recoding variables and creating new variables (p 89)

It would also help to see the original question and show-card used together with some discussion of respondents' use of the 0-10 scale.

[Extract from core questionnaire]

CARD 11 Using this card, please say to what extent you agree or disagree with each of the following statements. **READ OUT EACH STATEMENT AND CODE IN GRID²**

		Agree strongly	Agree	Neither agree nor disagree	Disagree	Disagree strongly	(Don't know)
B27	Gay men and lesbians should be free to live their own life as they wish ¹⁶ .	1	2	3	4	5	8

JM stresses the importance of collecting data at *disaggregated* level and explains that disaggregated data sometimes need to be grouped for tabulation, but doesn't mention the **golden rule** (at least among survey professionals) that people who collect age in groups, and not actual age last birthday, should be shot!

Variable [freehms] has 5 categories:

```

1 = "Agree strongly"
2 = "Agree"
3 = "Neither agree nor disagree"
4 = "Disagree"
5 = "Disagree strongly"

```

JM wishes to reduce them to 3:

```

Agree strongly + Agree
Neither agree nor disagree
Disagree + Disagree strongly

```

At 1' 14" in the video JM uses the following SPSS syntax "because it's quicker" [Note the abbreviation]

```

FREQUENCIES VARIABLES = freehms
  /BARCHART FREQ
  /ORDER = ANALYSIS.

```

. .to produce the following output:

² ¹⁶ Freedom of lifestyle is meant, 'free/entitled to live as gays and lesbians'.

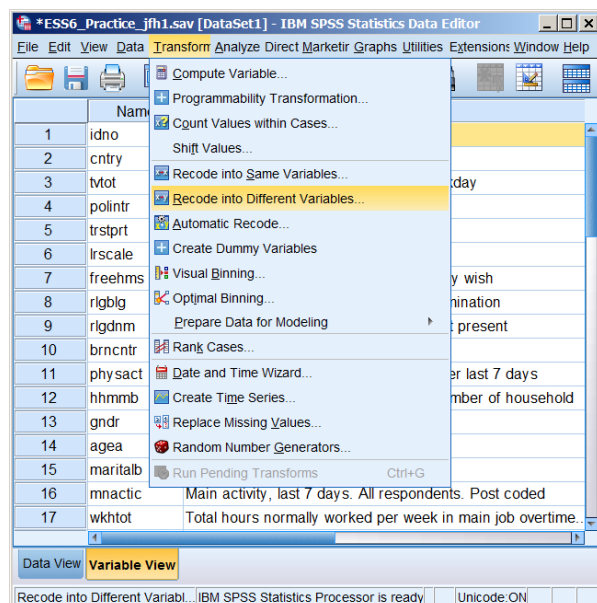
Gays and lesbians free to live life as they wish

		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	0.3		
	Don't know	2900	5.3		
	No answer	60	0.1		
	Total	3111	5.7		
Total		54673	100.0		




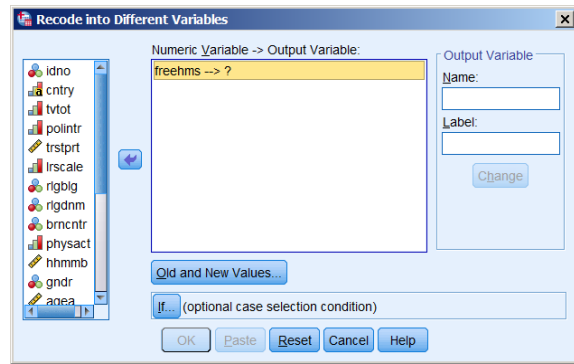
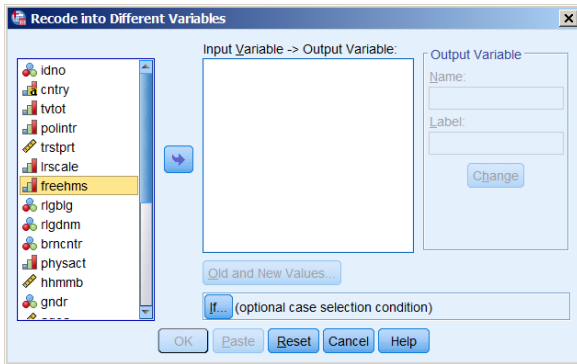
He then uses the GUI to generate a new variable with recoded values, but admits that using the dialog boxes “can be a little bit clunky”.

Transform >> Recode into Different Variables

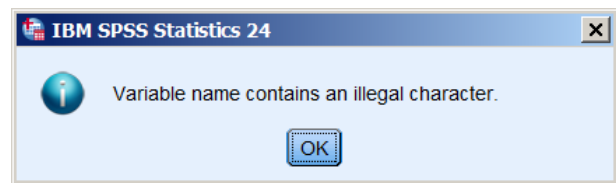
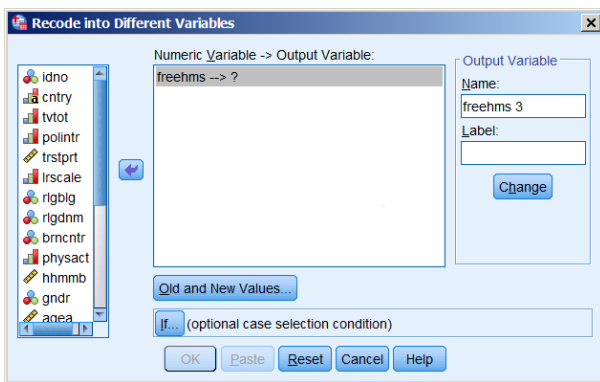


JM highlights **freehms**

. . and clicks on  to move it to the **Output variable** box:



In the **Name** box he types **freehms 3** which in my version of SPSS 24 creates a **warning message**



This is because he has typed a **space** in the new variable name **freehms_3**. If you look at the video in full-screen mode and freeze it, you can just make it out. He should have typed **freehms3**.

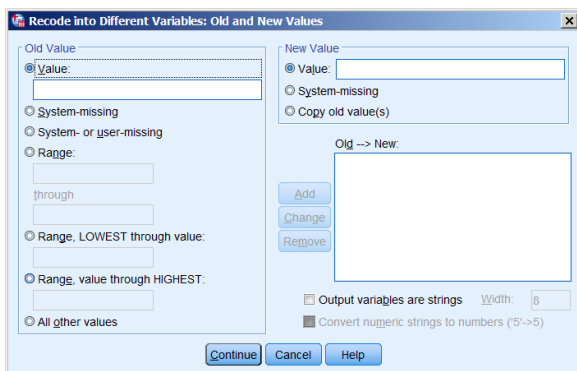


This error is copied into the syntax file on the companion site: if you try to run the syntax it causes an error.

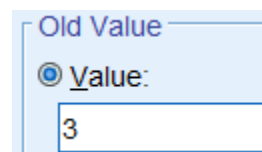
The correct GUI sequence (in full clunky mode) should look like this:

Transform >> Recode into Different Variables

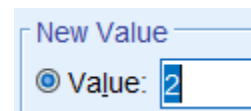
Opening dialog



In the **Old Value** box, type 3



In the **New Value** box type 2



Click on **Add** to display

Old --> New:
3 --> 2

Now click on **Range:**

Enter **1** in the upper box
and **2** in the lower box

Enter **1** in the **New Value** box

New Value

☒ Value:



.. and click on

Add

Recode into Different Variables: Old and New Values

Old Value

☒ Value:

☐ System-missing

☐ System- or user-missing

☒ Range:

through

☐ Range, LOWEST through value:

☐ Range, value through HIGHEST:

☐ All other values

New Value

☒ Value:

☐ System-missing

☐ Copy old value(s)

Old --> New:

3 --> 2

Add **Change** **Remove**

☐ Output variables are strings Width:

☐ Convert numeric strings to numbers ('5'-->5)

Continue **Cancel** **Help**



Recode into Different Variables: Old and New Values

Old Value

☐ Value:

☐ System-missing

☒ System- or user-missing

☒ Range:

through

☐ Range, LOWEST through value:

☐ Range, value through HIGHEST:

☐ All other values

New Value

☒ Value:

☐ System-missing

☐ Copy old value(s)

Old --> New:

3 --> 2
1 thru 2 --> 1

Add **Change** **Remove**

☐ Output variables are strings Width:

☐ Convert numeric strings to numbers ('5'-->5)

Continue **Cancel** **Help**

Repeat this process to recode values 4 and 5 to 3 until:

Recode into Different Variables: Old and New Values

Old Value

☐ Value:

☐ System-missing

☐ System- or user-missing

☒ Range:

through

☐ Range, LOWEST through value:

☐ Range, value through HIGHEST:

☐ All other values

New Value

☒ Value:

☐ System-missing

☐ Copy old value(s)

Old --> New:

3 --> 2
1 thru 2 --> 1
4 thru 5 --> 3

Add **Change** **Remove**

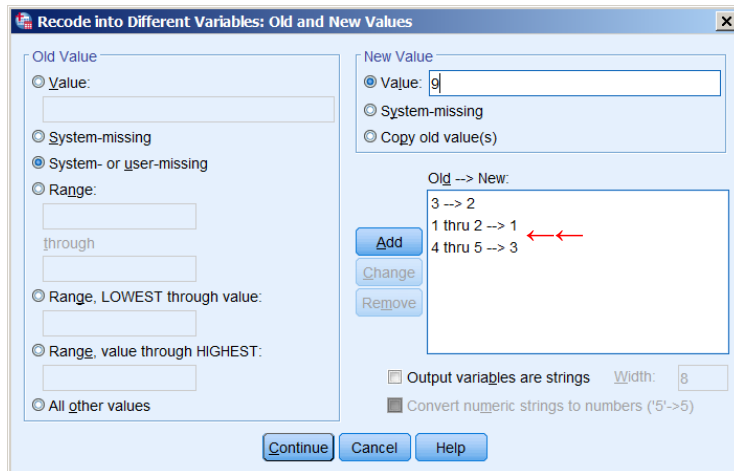
☐ Output variables are strings Width:

☐ Convert numeric strings to numbers ('5'-->5)

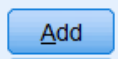
Continue **Cancel** **Help**

Variable **freehms** also has 3 user-missing values declared and JM uses the GUI to recode them to all to 9 noting the convention of using 8 and 9 or 88 and 99 as missing values.

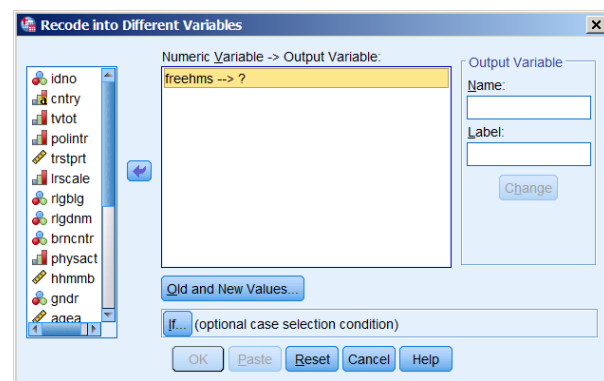
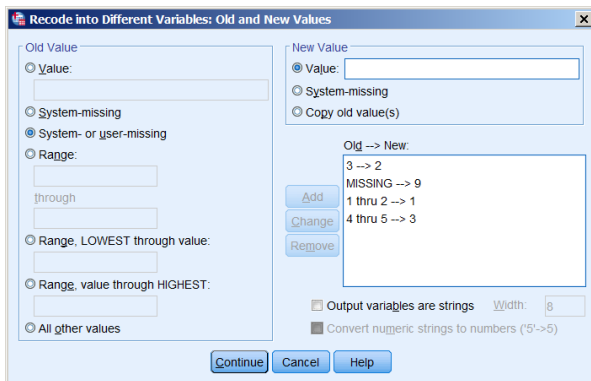
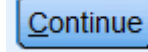
Click on ☒ **System- or user-missing** enter **9** in the **Value** box



then click on

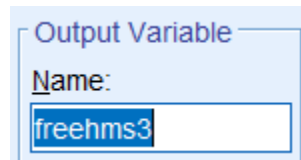


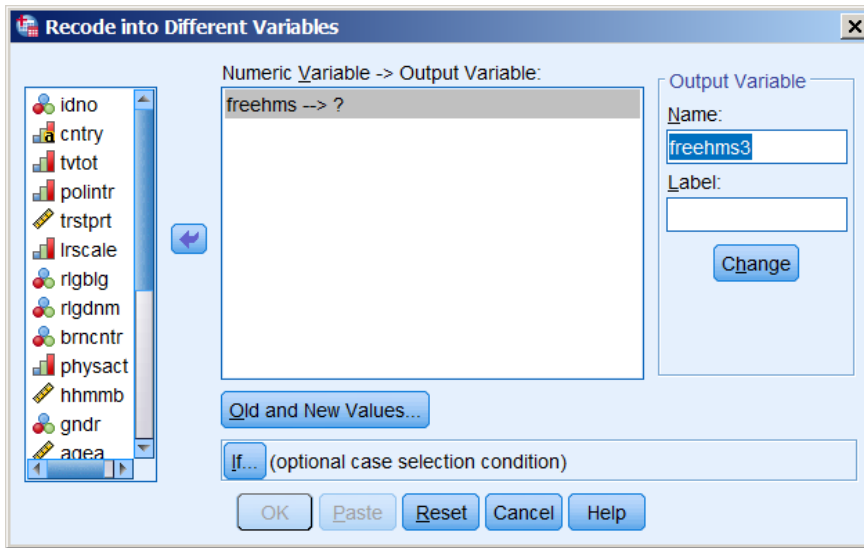
Click on



SPSS is waiting for you to give a name to the new variable:

Write **freehms3** in the **Output Variable** box





Click on **Change** then **OK**

The new variable **[freehms3]** is appended to the **Data Editor**:

20	pweight	Populat...	None	None	Numeric	Scale
21	freehms3		None	None	Numeric	Nominal
22						

>>

The syntax created from the GUI is displayed in the output viewer:

```
RECODE freehms (3=2) (MISSING=9) (1 thru 2=1) (4 thru 5=3) INTO freehms3.
EXECUTE.
```

. but can be copied into your syntax file with **Paste**

EXECUTE.

[NB: The recode specifications generated are not in the order entered. Direct syntax is neater]

RECODE freehms (3=2) (**MISSING** = 9) (2 =1) (4, 5=3) **into** freehms3 .

In fact, there is no need to recode the missing values at all as they won't affect any subsequent analysis.

NB: SPSS has assigned the new variable **[freehms3]** a measurement level of **Nominal**, but it should be **Ordinal**.

21	freehms3		None	None	Numeric	Nominal
----	----------	--	------	------	---------	---------

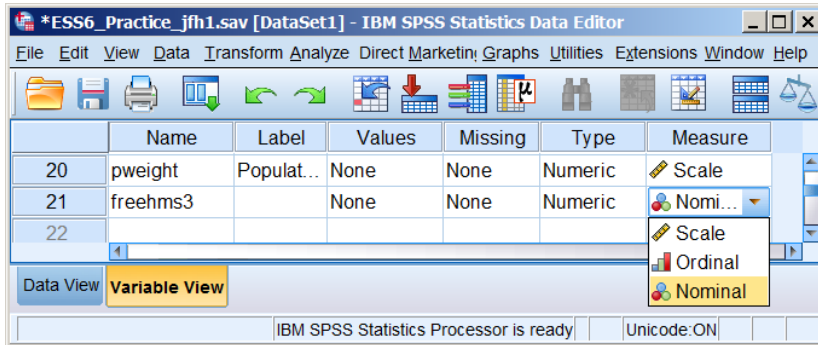
↓↓

This can be corrected with syntax:

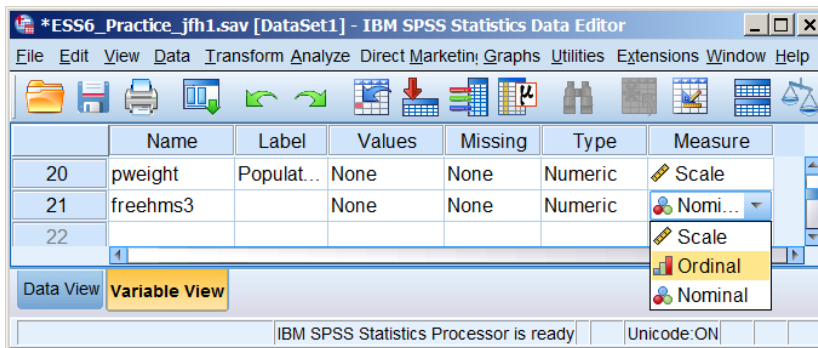
VARIABLE LEVEL freehms3 (**Ordinal**) . [var lev freehms3 (**ord**) .]

.. or manually in the **Data Editor**.

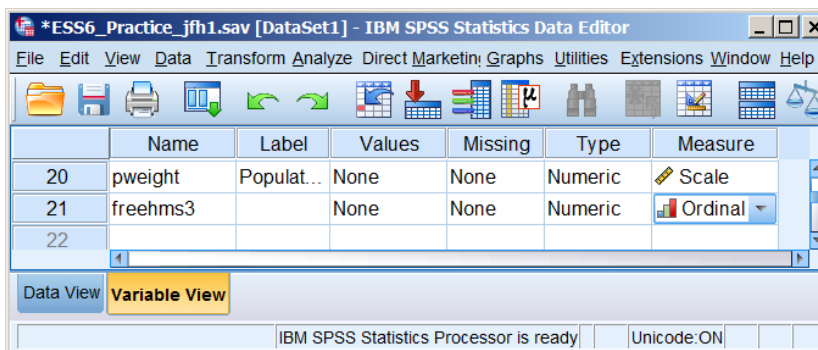
Click on  **Nominal**



Click on  **Ordinal**



In the **Measure** column [freehms3] has now changed to **Ordinal**:



JM also uses the incredibly laborious and tortuous GUI route to write value labels for the new variable. Whilst the GUI can be a useful learning tool and avoids syntax errors, this is a classic example of how direct syntax is so much quicker and easier:

There is actually a much quicker way to achieve a (temporary) grouping of **[freehms]**.

TEMPORARY .
RECODE freehms (1=2) (5=4) .
FREQUENCIES freehms .

. . which leaves value 3 alone and saves writing new labels.

Gays and lesbians free to live life as they wish

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	33321	60.9	64.6	64.6
	Neither agree nor disagree	7691	14.1	14.9	79.5
	Disagree	10550	19.3	20.5	100.0
	Total	51562	94.3	100.0	
Missing	Refusal	151	0.3		
	Don't know	2900	5.3		
	No answer	60	0.1		
	Total	3111	5.7		
Total		54673	100.0		

Note the SPSS colour coding as JM types and the use of **FREQ** as a check.

In the text, but not in the video, JM creates another target variable **[freehms2]** with only two categories (1 = Agree 2 = Disagree) but commits the cardinal sin of recoding other values to **SYSMIS**. (See Fig 4.16, p 91) Admittedly it's a convenient way for JM to get values other than 1, 2, 4 and 5 out of the way whilst experimenting with the data, but it is a very dangerous practice to demonstrate in front of learners as it risks the original values being permanently lost.

He then uses **CROSSTABS** to check the recoded values of the target variable against the values of the source variable, but in the resulting table the values of **[freehms2]** have 2 superfluous decimal places. Again, this is understandable for speed, but for learners it is sloppy practice. He should have used:

FORMATS freehms2 (f2.0) .

4.17 Using Syntax in SPSS

JM rightly encourages the use of syntax rather than the GUI, but repeats the dangerous practice of recoding to **sysmis**:

RECODE³ freehms (1 2 =1)(4 5=2)(**else = sysmis**) into freehms3 .

It is far better practice to recode to an existing (positive or negative) user-missing value, to a new user-missing value (to be declared) or to a value which is included in an existing range of user-missing values declared for other variables.

It's better practice, and much safer, to use something like:

RECODE freehms (2 =1)(4 5=2) (**else=copy**) into freehms2 .
FORMATS freehms2 (f2.0) .
MISSING VALUES freehms2 (3 7 thru 9) .
VARIABLE LABELS freehms2 'Recoded freehms'.
VALUE LABELS freehms2
 1 'Agree' 2 'Disagree' 3 'Neither agree nor disagree'
 7 'Refusal' 8 "Don't know" 9 'No answer' .
FREQUENCIES freehms2.

		freehms2			
		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	Neither agree nor disagree	7691	14.1		
	Refusal	151	0.3		
	Don't know	2900	5.3		
	No answer	60	0.1		
	Total	10802	19.8		
Total		54673	100.0		

End of: **MacInnes 4.1.5 Guide to videos 7 and 8 for Chapter 4**

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

Back to: [MacInnes 1 Aide-mémoire for easier navigation of companion website](#)

Back to: [MacInnes 2 Downloading the European Social Survey Practice File](#)

Back to: [MacInnes 3 Downloading the SPSS syntax for Chapter 4](#)

Back to: [4.1.1 Overview of video tutorials 1 to 6](#)

Back to: [4.1.2 Downloading the European Social Survey Practice File](#)

Back to: [4.1.3 Downloading the SPSS syntax](#)

Back to: [4.1.4 Checking the SPSS files](#)

Forward to: [4.1.6 Guide to video tutorial 9](#)

³ Note that this over-writes the existing **[freehms3]** (which has three groups) and does not, as he possibly intends, create **[freehms2]** (which has only two).