

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

[New page: 18 January 2018]

John MacInnes

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

(Sage, Dec. 2017)

### 4.1.7: Guide to video tutorials 10 and 11 for Chapter 4: Getting Started with SPSS

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](#)

[4.1.5 Guide to video tutorials 7 and 8](#)

[4.1.6 Guide to video tutorial 9](#)

Page	Video	Paragraph in book
2	<a href="#">Video 10: Mbw excel chart</a> (9'19")	4.20: Exporting output to other applications
4		4.21: Examining Gender and Employment
9	<a href="#">Video 11: Auto recode</a> (8'17")	4.22: Creating Numeric Versions of String Variables

The following notes assume you are familiar with copy/paste ( **Ctrl+C** / **Ctrl+V** ) 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 for Chapter 4 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.

Chapter 4: Getting Started with  
SPSS

Video tutorials

Syntax files and Datasets

Datasets

Syntax

Clicking on [Datasets](#) downloads SPSS saved file **ESS6\_Practice.sav**. Clicking on [Syntax](#) downloads file **Syntax\_Ch\_4.sav** which is **not** an **SPSS \*.sav** file, but a **Notepad \*.txt** file containing all the syntax commands for all the analyses (See: [4.1.3 Downloading the SPSS syntax](#))

The syntax text really needs to be split into separate **\*.sps** files for each topic, preferably with file names corresponding to the relevant paragraph in the text.

## Video 10: Mbw excel chart (9'18")

### Chapter 4, paragraph 20: Exporting output to other applications (p 96)

JM refers to the modified three-way table he created in 4.1.6 from which he proposes to produce the barchart in the book (Fig 4.24: p 98). Deletes missing values of **[mbw]** and reruns the syntax:

#### CROSSTABS

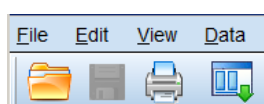
```
/TABLES=cntry BY mbw BY gndr
/FORMAT=AVALUE TABLES
/CELLS= ROW
/COUNT ROUND CELL.
```

Of the table he says, "As it stands that table isn't very easy to understand. Let's turn to the more interesting part . . . economic activity of women in different countries. Let's do that by exporting the table to Excel and then producing a barchart."

In the SPSS viewer he clicks on the table and a small red arrow appears halfway down on the left side:

Country * Economic activity * Gender Crosstabulation					
% within Country					
Gender			Economic activity		Total
			active	at home	
Male	Country	Albania	99.1%	0.9%	100.0%
		Belgium	96.9%	3.1%	100.0%
		Bulgaria	89.2%	10.8%	100.0%
		Switzerland	99.2%	0.8%	100.0%
		France	87.6%	12.4%	100.0%
		United Kingdom	80.3%	19.7%	100.0%
		Hungary	81.3%	18.7%	100.0%
		Ireland	58.8%	41.2%	100.0%
Total	Country	Albania	72.9%	27.1%	100.0%
		Belgium	87.2%	12.8%	100.0%
		Bulgaria	74.6%	25.4%	100.0%
		Switzerland	86.0%	14.0%	100.0%
		Kosovo	56.2%	43.8%	100.0%
		Total	85.7%	14.3%	100.0%

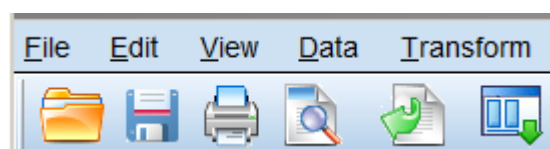
In the tool bar



a new icon



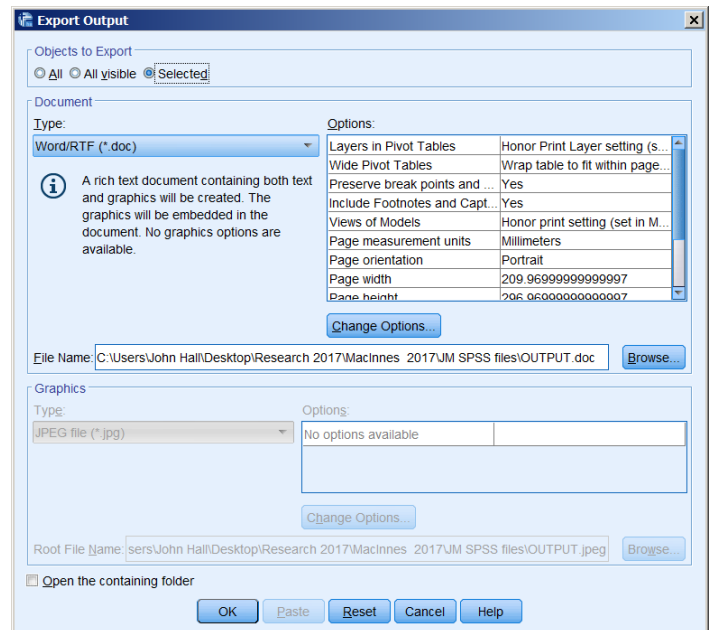
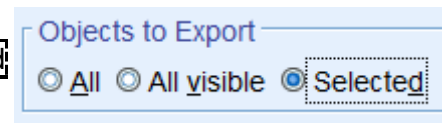
appears:



JM clicks on  to get **Export Output** dialog

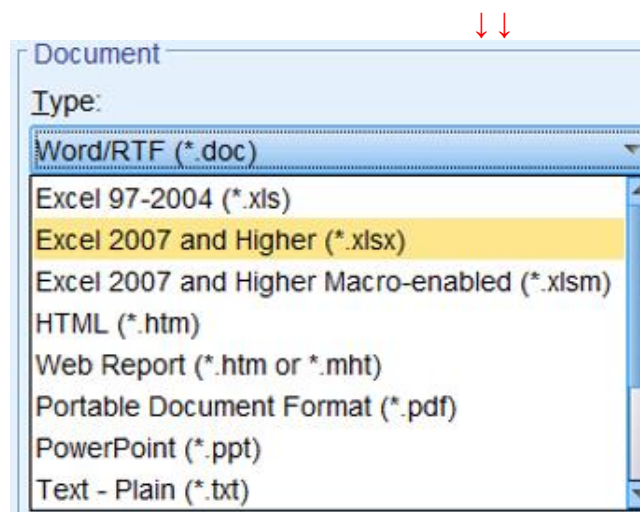


→ → Checks **Selected**



In JM's version of SPSS the default destination is **Excel**, but in mine it is **Word**

I click on the ▼ in **Word/RTF (\*.doc)** to get **Document Type**

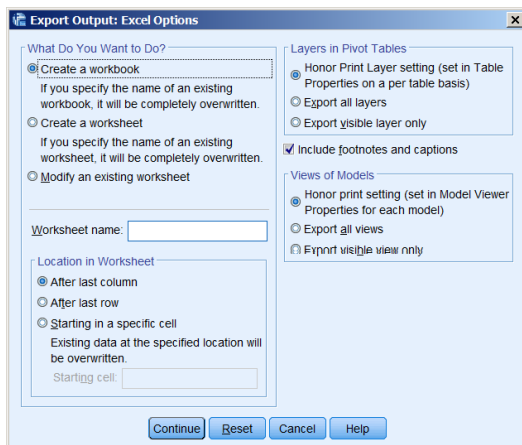


Click on **Excel 2007 and Higher (\*.xlsx)** . . then on

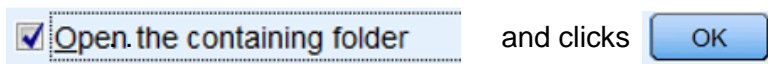
**Change Options...**

. . to get **Export Output: Excel Options**

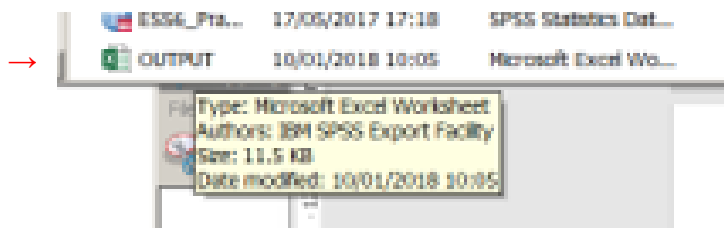
## Export Output: Excel Options opening dialog:



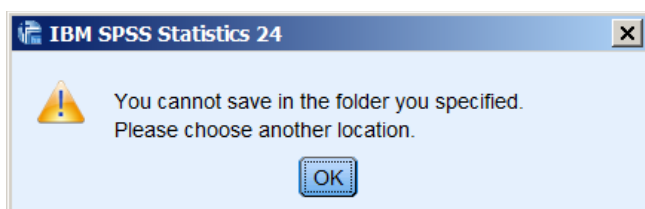
JM types **mbw** in Worksheet name: **mbw** then clicks **Continue** to get:



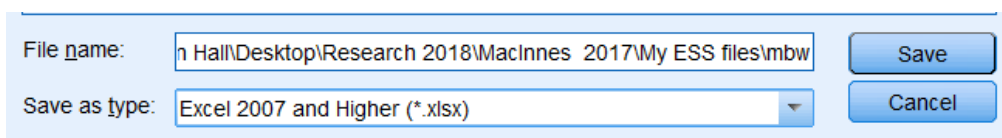
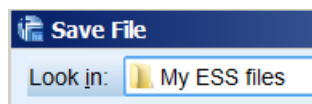
A new Excel file **OUTPUT** appears in his **containing** (destination) folder:



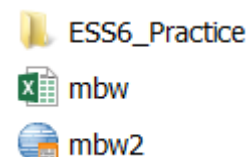
When I tried it I got:



Browse to folder:



Click on **Save** then on **OK** to save Excel file **mbw.xlsx** to folder **My ESS files**



## Chapter 4, paragraph 21: Examining Gender and Employment (p 99)

JM opens the Excel file: (video point 4'45")

	A	B	C	D	E	F
1	Country * Economic activity * Gender Crosstabulation					
2	% within					
3	Country					
4	Gender		Economic activity		Total	
5	Male	Country	Albania	99.1%	0.9%	100.0%
6			Belgium	96.9%	3.1%	100.0%
7			Bulgaria	89.2%	10.8%	100.0%
8			Switzerland	99.2%	0.8%	100.0%

90			Slovenia	87.7%	12.3%	100.0%
91			Slovakia	89.1%	10.9%	100.0%
92			Ukraine	78.1%	21.9%	100.0%
93			Kosovo	56.2%	43.8%	100.0%
94		Total		85.7%	14.3%	100.0%

"Now, I want to produce a barchart . . . , but before I do that I'm going to re-order the rows to make the interpretation easier."

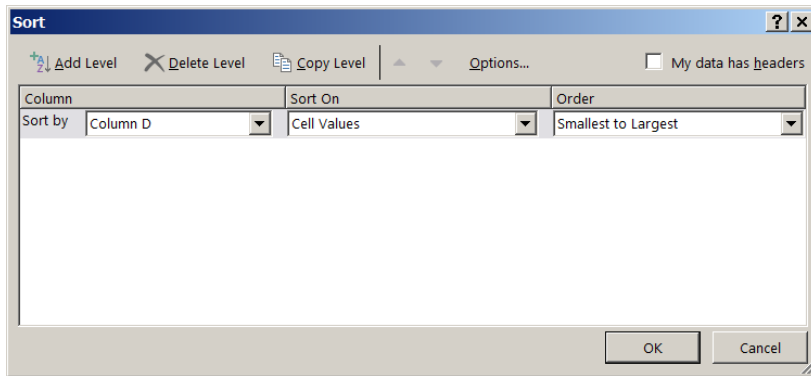
Highlights rows for women in the second section

35	Female	Country	Albania	50.7%	49.3%	100.0%
36			Belgium	77.8%	22.2%	100.0%
37			Bulgaria	64.7%	35.3%	100.0%
38			Switzerland	73.3%	26.7%	100.0%
39			Cyprus	68.6%	31.4%	100.0%
40			Czech Republic	86.7%	13.3%	100.0%
41			Germany	73.4%	26.6%	100.0%
42			Denmark	79.3%	20.7%	100.0%
43			Estonia	88.5%	11.5%	100.0%
44			Spain	70.1%	29.9%	100.0%
45			Finland	90.5%	9.5%	100.0%
46			France	87.6%	12.4%	100.0%
47			United Kingdom	80.3%	19.7%	100.0%
48			Hungary	81.3%	18.7%	100.0%
49			Ireland	58.8%	41.2%	100.0%
50			Israel	75.5%	24.5%	100.0%
51			Iceland	74.2%	25.8%	100.0%
52			Italy	77.3%	22.7%	100.0%
53			Lithuania	85.6%	14.4%	100.0%
54			Netherlands	71.8%	28.2%	100.0%
55			Norway	88.4%	11.6%	100.0%
56			Poland	79.4%	20.6%	100.0%
57			Portugal	74.3%	25.7%	100.0%
58			Russian Federation	87.7%	12.3%	100.0%
59			Sweden	94.9%	5.1%	100.0%
60			Slovenia	80.3%	19.7%	100.0%
61			Slovakia	83.9%	16.1%	100.0%
62			Ukraine	68.5%	31.5%	100.0%
63			Kosovo	31.2%	68.8%	100.0%

"What has it done?" he asks.

His first attempt has sorted column C (country) by alphabetical order, because it was the (default) first column encountered.

He goes back to the table, and sorts on column D ("active") <sup>1</sup>



. . to produce the following table in which countries are sorted in **ascending order** of the percentage of women who are economically active:

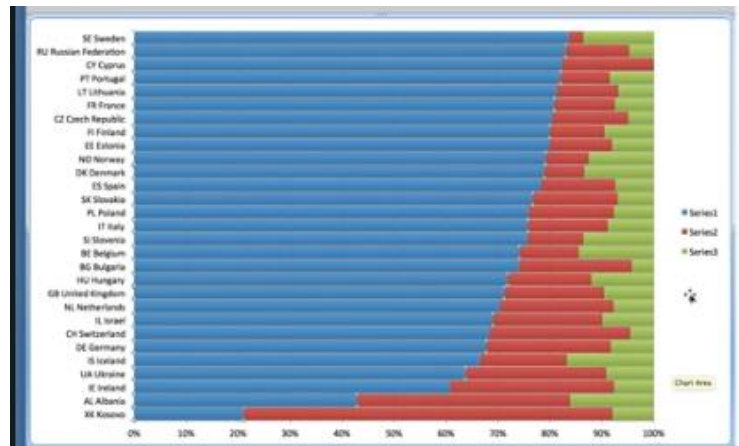
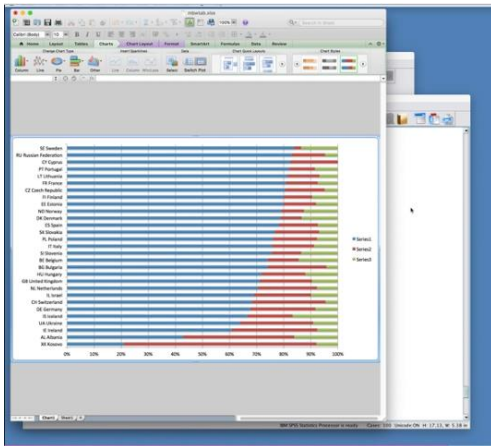
↓↓

Albania	31.2%	49.3%
Belgium	50.7%	22.2%
Bulgaria	58.8%	35.3%
Switzerland	64.7%	26.7%
Cyprus	68.5%	31.4%
Czech Republic	68.6%	13.3%
Germany	70.1%	26.6%
Denmark	71.8%	20.7%
Estonia	73.3%	11.5%
Spain	73.4%	29.9%
Finland	74.2%	9.5%
France	74.3%	12.4%
United Kingdom	75.5%	19.7%
Hungary	77.3%	18.7%
Ireland	77.8%	41.2%
Israel	79.3%	24.5%
Iceland	79.4%	25.8%
Italy	80.3%	22.7%
Lithuania	80.3%	14.4%
Netherlands	81.3%	28.2%
Norway	83.9%	11.6%
Poland	85.6%	20.6%
Portugal	86.7%	25.7%
Russian Federation	87.6%	12.3%
Sweden	87.7%	5.1%
Slovenia	88.4%	19.7%
Slovakia	88.5%	16.1%
Ukraine	90.5%	31.5%
Kosovo	94.9%	68.8%

<sup>1</sup> To sort the columns in my version of Excel I had to right click inside the highlighted area and select **Sort** >> **Custom Sort** to get the **Sort dialog**.

JM now proceeds to create a barchart (6'03")

Goes to **Charts** and gets a **clustered barchart** (6'45") but now has **three** categories rather than **two**.

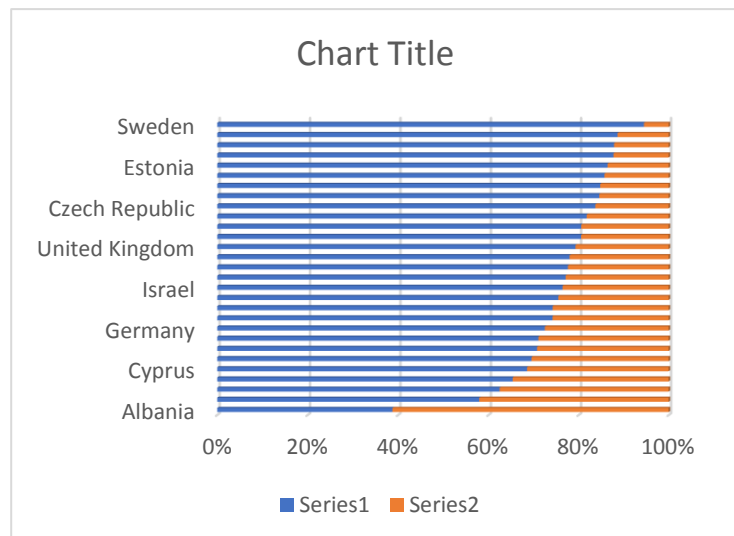


He narrows the gap between the bars and adds labels to each category: **blue** is active, **red** is at home, **green** is Other, but would the curve be smoother if "Other" were omitted?

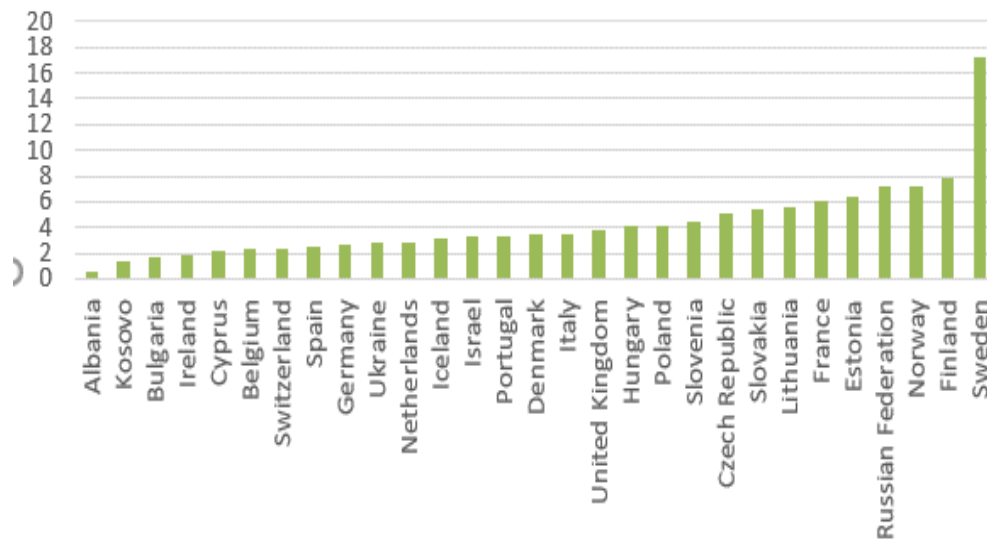
Create ratio [% active : % at home] for Albania  $\text{=J35/K35}$  and copy to all countries to get:

Albania	31.2%	49.3%	0.6
Kosovo	94.9%	68.8%	1.4
Bulgaria	58.8%	35.3%	1.7
Ireland	77.8%	41.2%	1.9
Cyprus	68.5%	31.4%	2.2
Belgium	50.7%	22.2%	2.3
Switzerland	64.7%	26.7%	2.4
Spain	73.4%	29.9%	2.5
Germany	70.1%	26.6%	2.6
Ukraine	90.5%	31.5%	2.9
Netherlands	81.3%	28.2%	2.9
Iceland	79.4%	25.8%	3.1
Israel	79.3%	24.5%	3.2
Portugal	86.7%	25.7%	3.4
Denmark	71.8%	20.7%	3.5
Italy	80.3%	22.7%	3.5
United	75.5%	19.7%	3.8
Hungary	77.3%	18.7%	4.1
Poland	85.6%	20.6%	4.1
Slovenia	88.4%	19.7%	4.5
Czech	88.6%	13.3%	5.1
Slovakia	88.5%	16.1%	5.5
Lithuania	80.3%	14.4%	5.6
France	74.3%	12.4%	6.0
Estonia	73.3%	11.5%	6.4
Russian	87.6%	12.3%	7.1
Norway	83.9%	11.6%	7.2
Finland	74.2%	9.5%	7.8
Sweden	87.7%	5.1%	17.2

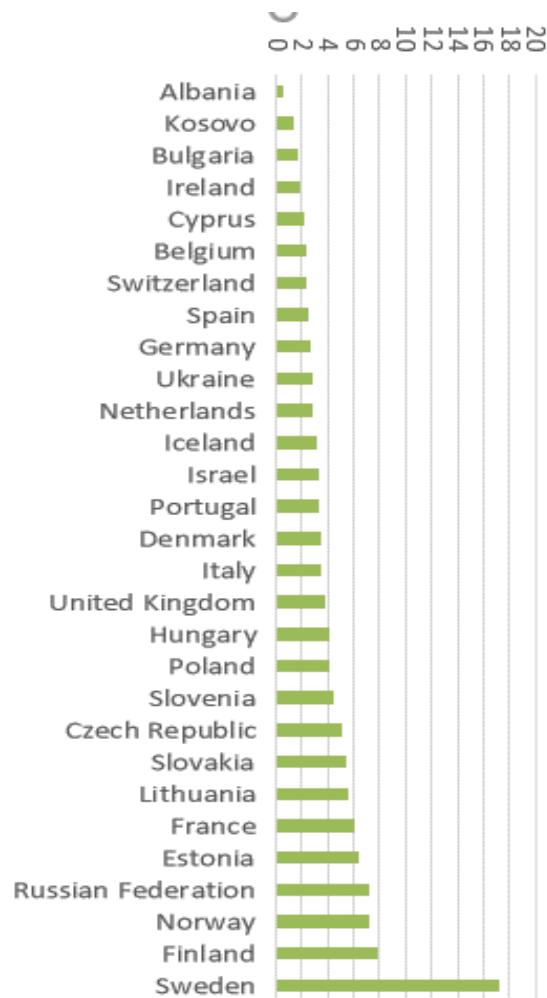
Not sure how I produced the chart below, but some countries got omitted. Nice chart though!



The above chart is missing some countries, but the one below includes all countries:



It's easier to read if you rotate it:





## Video 11: Auto recode (8'17")

### Chapter 4, paragraph 22: Creating Numeric versions of String Variables (p 100)

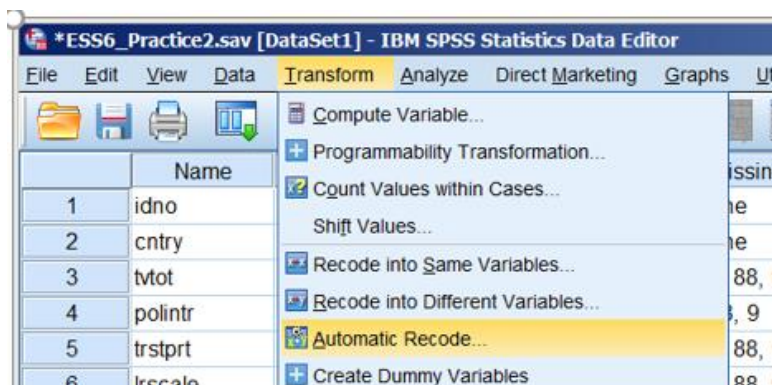
Computers are far better at handling numeric than alphabetic (string) data. Although they can handle strings, the procedures can take longer (and be more expensive). For many statistical analyses, string variables can be converted to numeric using:

**RECODE** <"string value"> **into** <numeric value> .


The SPSS procedure **AUTORECODE** automatically recodes strings to numeric.

2	cntry	Country	{AL, Albania}...	None	String	Ordinal
---	-------	---------	------------------	------	--------	---------

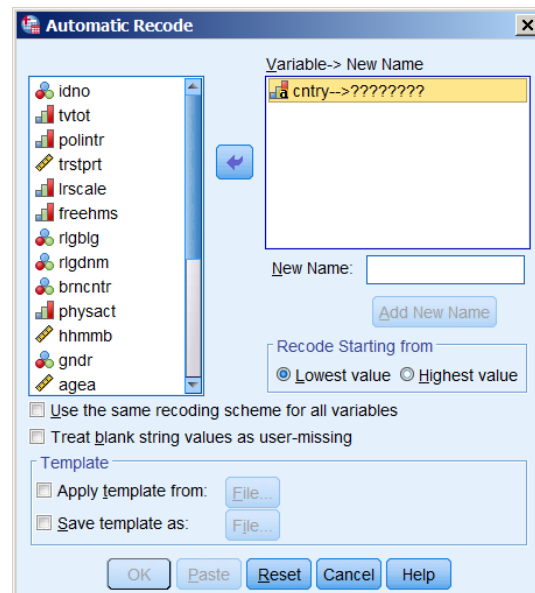
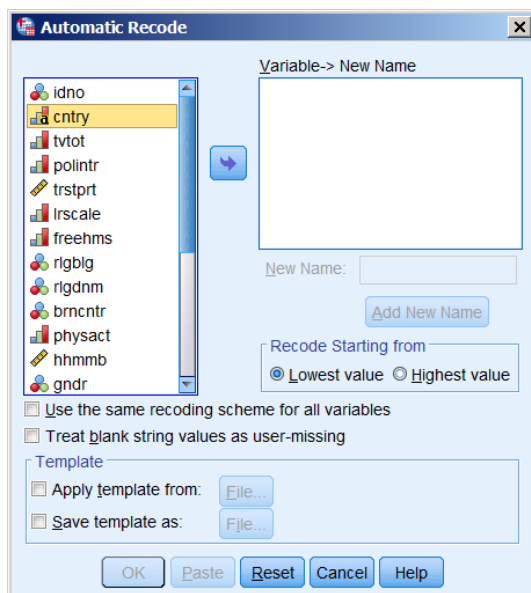
Transform >> Automatic Recode



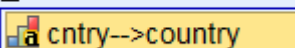
... to get the **Automatic Recode** dialog.


Clicks on **cntry** then on 

to move it to the Variables > New name box



Writes **country** in New Name box, **New Name:**

Clicks **Add New Name** to get **Variable-> New Name**  



Clicks on  to produce the following syntax:

```
DATASET ACTIVATE DataSet1.  
AUTORECODE VARIABLES=cntry  
/INTO country  
/PRINT.
```

(2'21") Runs the syntax to get:

Old Value	New Value	Value Label
AL	1	Albania
BE	2	Belgium
BG	3	Bulgaria
CH	4	Switzerland
CY	5	Cyprus
CZ	6	Czech Republic
DE	7	Germany
DK	8	Denmark
EE	9	Estonia
ES	10	Spain
FI	11	Finland
FR	12	France
GB	13	United Kingdom
HU	14	Hungary
IE	15	Ireland
IL	16	Israel
IS	17	Iceland
IT	18	Italy
LT	19	Lithuania
NL	20	Netherlands
NO	21	Norway
PL	22	Poland
PT	23	Portugal
RU	24	Russian Federation
SE	25	Sweden
SI	26	Slovenia
SK	27	Slovakia
UA	28	Ukraine
XK	29	Kosovo

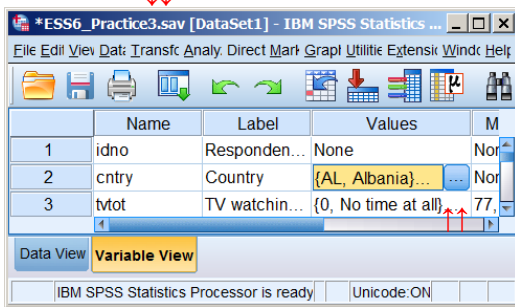
The variable **[country]** is appended to the **Data Editor**

24	country	Country	{1, Albania}...	None	Numeric	 Ordinal
----	---------	---------	-----------------	------	---------	---

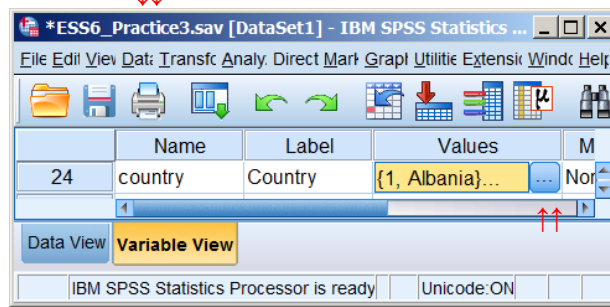
[NB: In the **Level** column SPSS has assigned **[country]** to **Ordinal**: it needs changing to **Nominal**]

If you click on the cells (one at a time) in the **Values** column for

**[cntry]**



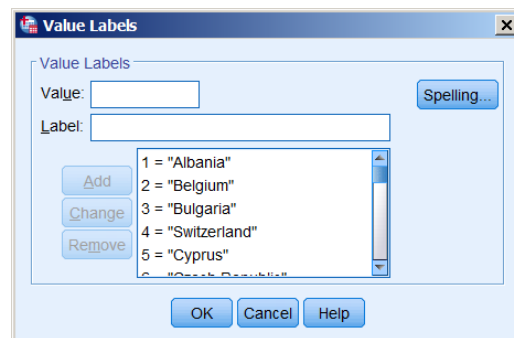
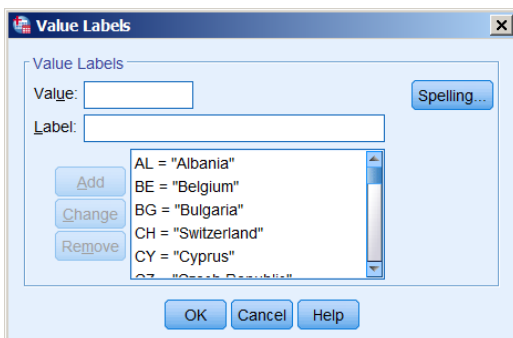
and **[country]**



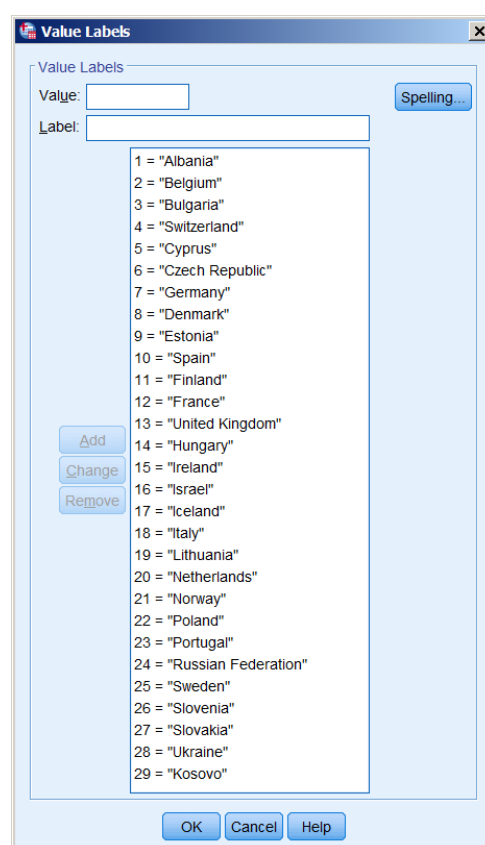
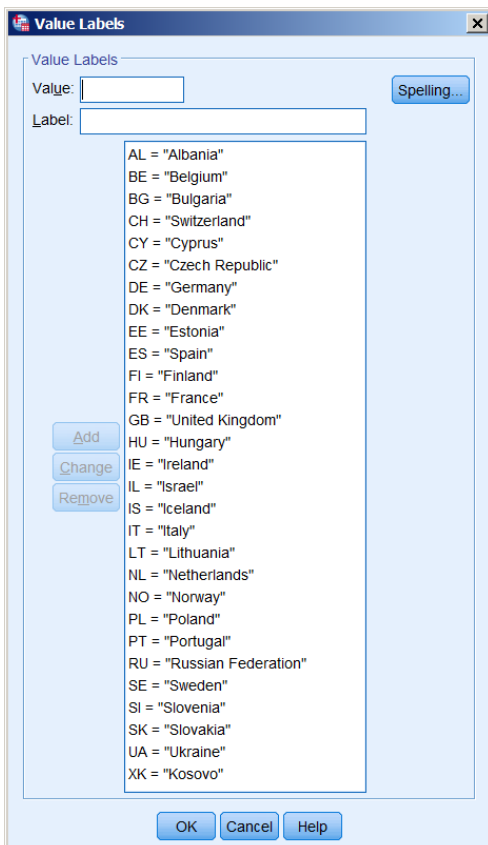
.. then click on the blue squares



you can see that the string values for **[cntry]** have been replaced with numeric values for **[country]**



[NB: You can drag the bottom edges of the dialog boxes down to see the whole list]



Jm now says, "Let's practice our syntax skills a little." and uses direct syntax to create a new variable **country1** with one group containing France, Germany, UK, Netherlands, Portugal, Russian Federation and Sweden and another group containing all other countries.

```
compute country1 = country .
```

```
recode country1 (12 7 13 20 23 24 25 = 1)(ELSE = 0) .
```

.. and runs a check:

**cross** country1 by country .

[NB: Note that JM over-rides the SPSS prompt which would display the **RECODE** command in UPPER CASE, but that he leaves the display for **ELSE**. He also uses abbreviated syntax **cross** in lower case which is **not colour-coded** by SPSS.]

The syntax produces a very wide table so he swaps the rows and columns.

country1 Country Crustalation																														
Count	Country																													
	Albania	Belgium	Bulgaria	Switzerland	Cyprus	Czech Republic	Germany	Denmark	Estonia	Spain	Finland	France	United Kingdom	Hungary	Ireland	Israel	Iceland	Italy	Lithuania	Netherlands	Norway	Poland	Portugal	Russian Federation	Sweden	Slovenia	Slovakia	Ukraine	Kosovo	
country1 selected countries	1201	1889	2262	1493	1116	2009	0	1650	2390	1899	2197	0	0	2014	0	2628	2088	752	860	2109	0	1624	1898	0	0	1257	1847	2178	1295	39334
country1 other countries	0	0	0	0	0	0	2459	0	0	0	0	1969	2298	0	0	0	0	0	0	1845	0	0	2151	2484	1847	0	0	0	0	10519
Total	1791	1889	2262	1493	1116	2009	2459	1650	1990	1899	2197	1969	2298	2014	2628	2088	752	860	2109	1624	1898	1624	1898	1257	1847	2178	1295	39334		

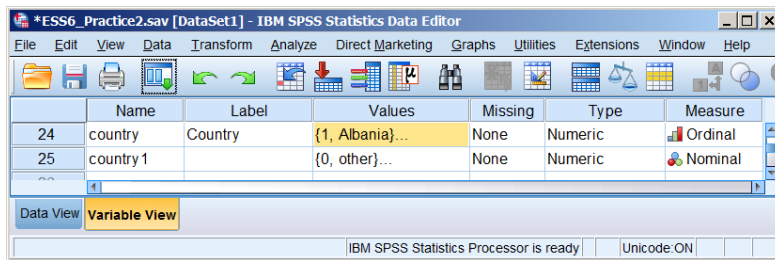
**cross** country by country1 .


Country * country1 Crosstabulation				
Count		country1		Total
		0.00	1.00	
Country	Albania	1201	0	1201
	Belgium	1869	0	1869
	Bulgaria	2260	0	2260
	Switzerland	1493	0	1493
	Cyprus	1116	0	1116
	Czech Republic	2009	0	2009
	Germany	0	2958	2958
	Denmark	1650	0	1650
	Estonia	2380	0	2380
	Spain	1889	0	1889
	Finland	2197	0	2197
	France	0	1968	1968
	United Kingdom	0	2286	2286
	Hungary	2014	0	2014
	Ireland	2628	0	2628
	Israel	2508	0	2508
	Iceland	752	0	752
	Italy	960	0	960
	Lithuania	2109	0	2109
	Netherlands	0	1845	1845
	Norway	1624	0	1624
	Poland	1898	0	1898
	Portugal	0	2151	2151
	Russian Federation	0	2484	2484
	Sweden	0	1847	1847
	Slovenia	1257	0	1257
	Slovakia	1847	0	1847
	Ukraine	2178	0	2178
	Kosovo	1295	0	1295
	Total		39134	15539

It's worked: gives it some labels. (6'23")

```
val labels = country1 1 'selected countries' 0 'other' .
```

[NB: He's still using abbreviated syntax and lower case, even in the labels in which the first letter would normally be capitalised (e.g. 'Selected countries' 'Other'). There is no variable label for **country1**. This is clearly a working file used for exploring the data, but standard practice would normally be to specify metadata (labels, missing values, measurement levels, formats etc. **before** running the analysis.)

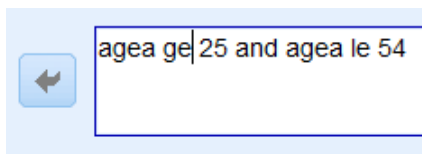


He still has his enormous **Syntax Editor** open. He selects cases by going to the **Select Cases** icon  and extending the logical expression from the filter he generated in Video 9 for:

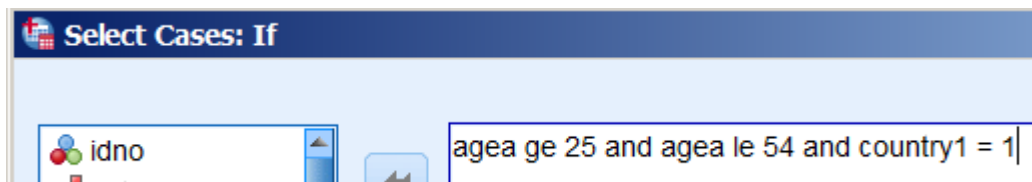
### Chapter 4, paragraph 18: Subsetting and Selecting Cases (p 93)

#### Task 1: Select only those cases aged between 25 and 54<sup>2</sup>

(see: [4.1.6 Guide to video tutorial 9](#), page 4)



Adds: " **and country1 = 1** "



In the video his table has no variable or value labels for **country1** just values 0.00 and 1.00.

Country \* country1 Crosstabulation

Count		country1		Total
		0.00	1.00	
Country	Albania	1201	0	1201
	Belgium	1869	0	1869
	Bulgaria	2260	0	2260

<sup>2</sup> On page 93 JM talks about selecting men and women between the ages of 25 and 49, but in the video selects people aged between 25 and 54.

However, because I saved the file modifications in new files **ESS6\_Practice2.sav** and **ESS6\_Practice3.sav** with **[filter\_\$]** (and also entered the value labels) I can simply write:

**filter** by filter\_\$ .  
cro country by country1 .

. . to get:

**Country \* country1 Crosstabulation**

Count

		country1		Total
		other	selected countries	
Country	Albania	566	0	566
	Belgium	912	0	912
	Bulgaria	965	0	965
	Switzerland	728	0	728
	Cyprus	573	0	573
	Czech Republic	1004	0	1004
	Germany	0	1395	1395
	Denmark	710	0	710
	Estonia	1059	0	1059
	Spain	1039	0	1039
	Finland	991	0	991
	France	0	915	915
	United Kingdom	0	1015	1015
	Hungary	1005	0	1005
	Ireland	1380	0	1380
	Israel	1187	0	1187
	Iceland	363	0	363
	Italy	476	0	476
	Lithuania	960	0	960
	Netherlands	0	891	891
	Norway	812	0	812
	Poland	924	0	924
	Portugal	0	928	928
	Russian Federation	0	1280	1280
	Sweden	0	848	848
	Slovenia	591	0	591
	Slovakia	950	0	950
	Ukraine	1019	0	1019
	Kosovo	716	0	716
Total		18930	7272	26202

Having selected a subset of people aged between 25 and 54, JM will proceed to examine working hours in the next tutorial.

**End of:** 4.1.7: Guide to video tutorials 10 and 11 for Chapter 4

**Forward to:** [4.1.8 Guide to video tutorial 12 for Chapter 4](#)

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

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