

Block 2: Analysing one variable

[Draft only: 17 July 2013]

2.3.1.6.2 Specimen answer for conditional frequencies homework [Tasks 3 and 4]

Exemplar: British Social Attitudes 1989

File: e:weebly downloads\bsa89\2.3.1.6.1.sav [or whatever you called it]

Previous sessions:

[2.3.1.5 Conditional frequencies homework exercises](#)

[2.3.1.6.1 Specimen answer for conditional frequencies homework \[Tasks 1 and 2\]](#)

SPSS commands: **TEMPORARY  
SELECT IF  
FREQUENCIES  
CROSSTABS**

**Research question 1:**

What is the distribution of respondents' personal gross income from paid work (Q.918b)? What shape does the distribution have? What is the distribution for women only? What is the distribution for men only? Are there any differences?

**Research question 2:**

What other variables might affect income regardless of gender? What effect do they have by themselves?

**Research question 3:**

What effect does gender have when these other variables are taken into account?

Last time we performed two tasks related to research question 1:

**Task 1:** a) Read in raw data for Q.901a (Sex of respondent) and Q.918b (Gross income of respondent from paid work, if working).

b) Produce initial frequency counts as a check on the data.

**Task 2:** a) Specify missing values, variable labels and value labels.

b) Produce a frequency count for sex and a separate one for income (with barchart).

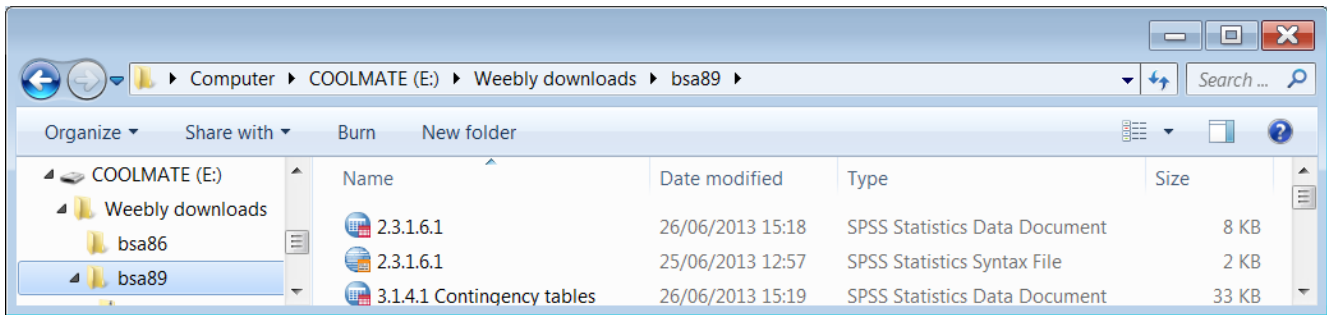
Tasks 1 and 2 were done in the previous exercise:

[2.3.1.6.1 Specimen answer for conditional frequencies homework \[Tasks 1 and 2\]](#)

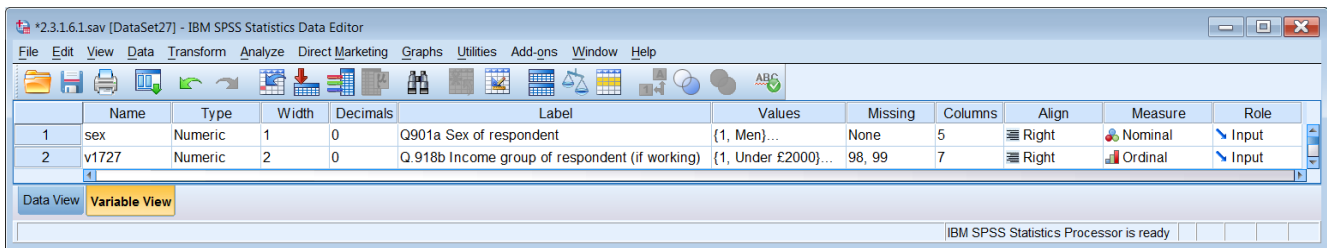
**Task 3:** Produce conditional frequency counts (with bar-charts) of income group, first for men only, then for women only.

**Task 4:** Produce a contingency table of income group by sex (with correct percentages for comparing the incomes of men and women)

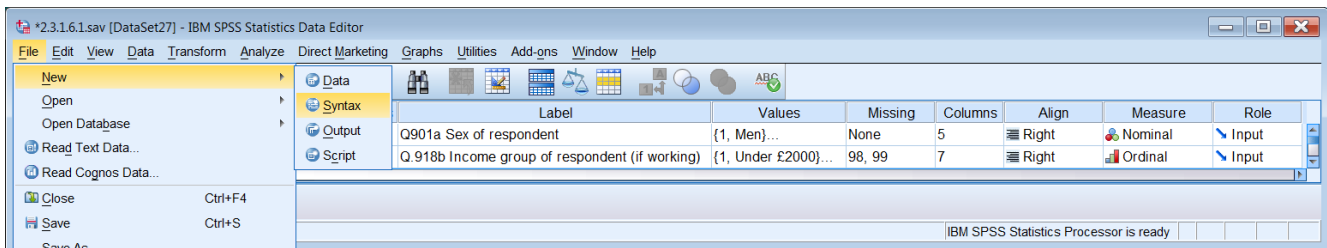
We ended the last session by producing a frequency count for sex and a frequency count and bar chart for income. We then saved our syntax and data editors using either the default SPSS names, or in my case **2.3.1.6.1.sps** and **2.3.1.6.1.sav** in folder **e:weebly downloads\bsa89**



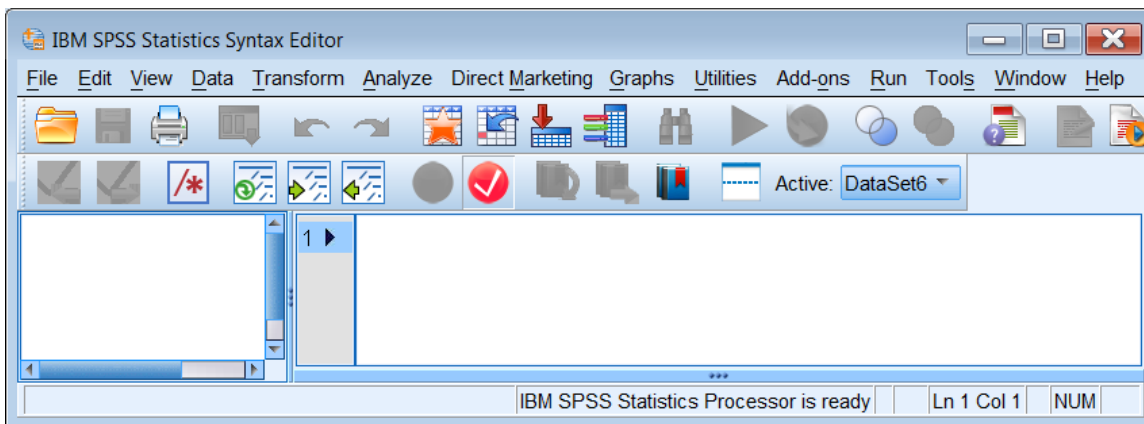
Open **2.3.1.6.1.sav** by double-clicking on it: if you don't have the file download [2.3.1.6.1.sav](#) from this site and open it when asked:



**File > New > Syntax**



to open a new syntax editor:

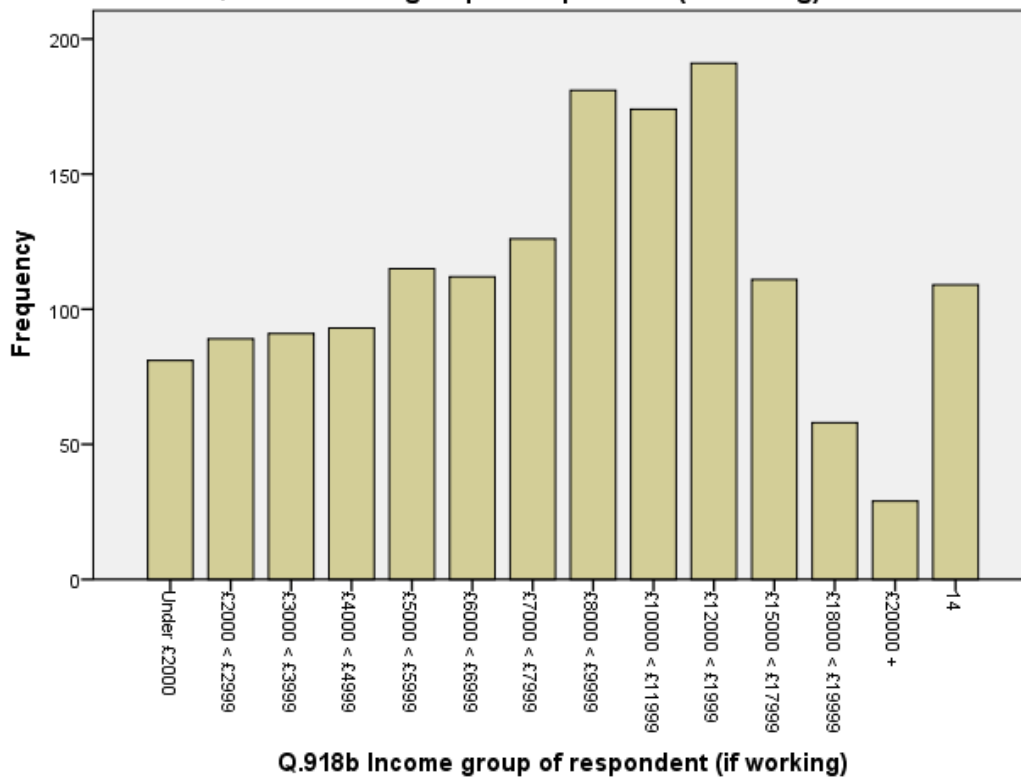


The distribution of income for the whole sample looked like this:

**v1727 Q.918b Income group of respondent (if working)**

	Frequency	Percent	Valid Percent	Cumulative Percent
1 Under £2000	81	2.7	5.2	5.2
2 £2000 < £2999	89	2.9	5.7	10.9
3 £3000 < £3999	91	3.0	5.8	16.7
4 £4000 < £4999	93	3.1	6.0	22.7
5 £5000 < £5999	115	3.8	7.4	30.1
6 £6000 < £6999	112	3.7	7.2	37.2
7 £7000 < £7999	126	4.2	8.1	45.3
Valid 8 £8000 < £9999	181	6.0	11.6	56.9
9 £10000 < £11999	174	5.8	11.2	68.1
10 £12000 < £1999	191	6.3	12.2	80.3
11 £15000 < £17999	111	3.7	7.1	87.4
12 £18000 < £19999	58	1.9	3.7	91.2
13 £20000 +	29	1.0	1.9	93.0
14	109	3.6	7.0	100.0
Total	1560	51.6	100.0	
98 Don't know	17	.6		
99 Not answered	108	3.6		
Missing System	1340	44.3		
Total	1465	48.4		
Total	3025	100.0		

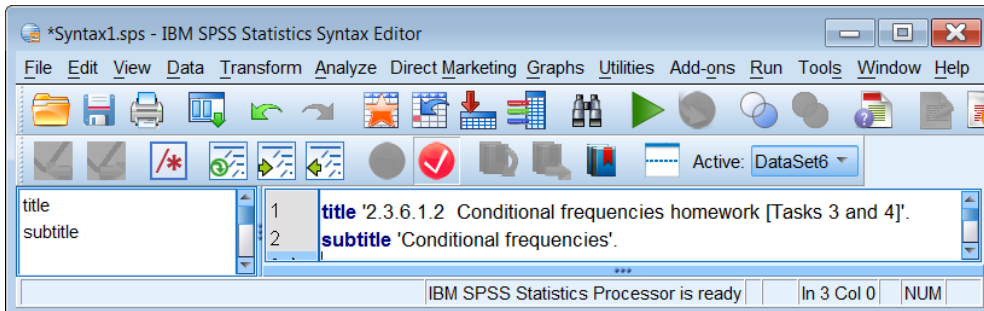
**Q.918b Income group of respondent (if working)**



**Task 3:** Produce conditional frequency counts (with barcharts) of income group, first for men only, then for women only.

In the syntax editor, type:

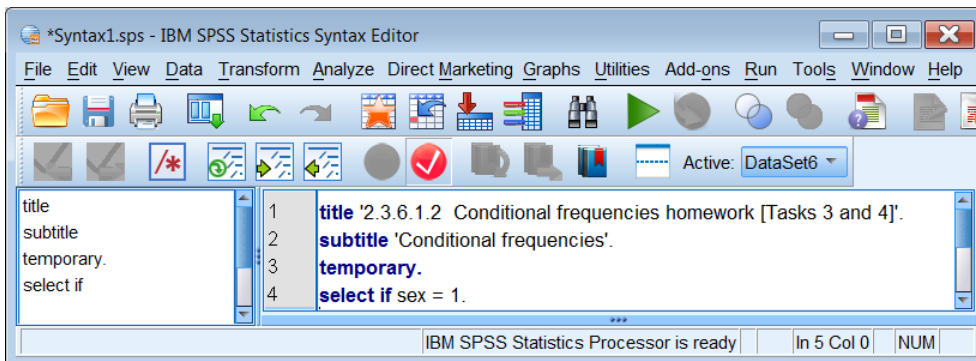
**title '2.3.6.1.2 Conditional frequencies homework [Tasks 3 and 4]'.  
subtitle 'Conditional frequencies'.**



We now need to run two separate frequency counts, one for men only and one for women only. One way of doing this is to use the **SELECT IF** command. However if we do this we need to use the **TEMPORARY** command, otherwise the unselected cases will be permanently written out of the file.

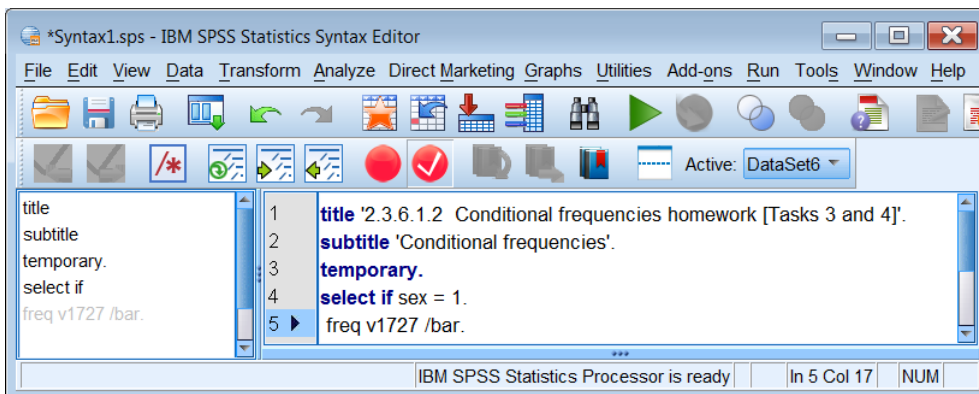
To select only men, type in:

**temporary.  
select if sex = 1.**



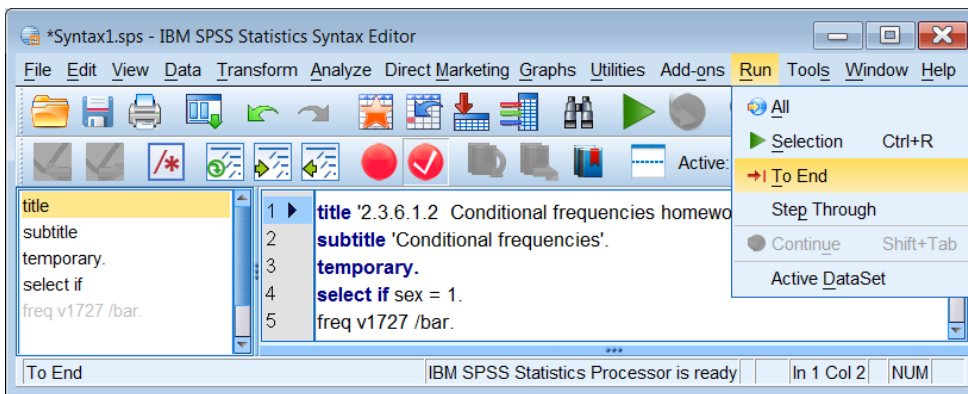
To produce a frequency count with a bar-chart, type in:

**freq v1727 /bar.**



[NB: freq v1727 /bar. is grayed out in the left pane because I've used abbreviated syntax.]

Place the cursor in the title line and press **Run** > **→| To End**:



## Men only

### Statistics

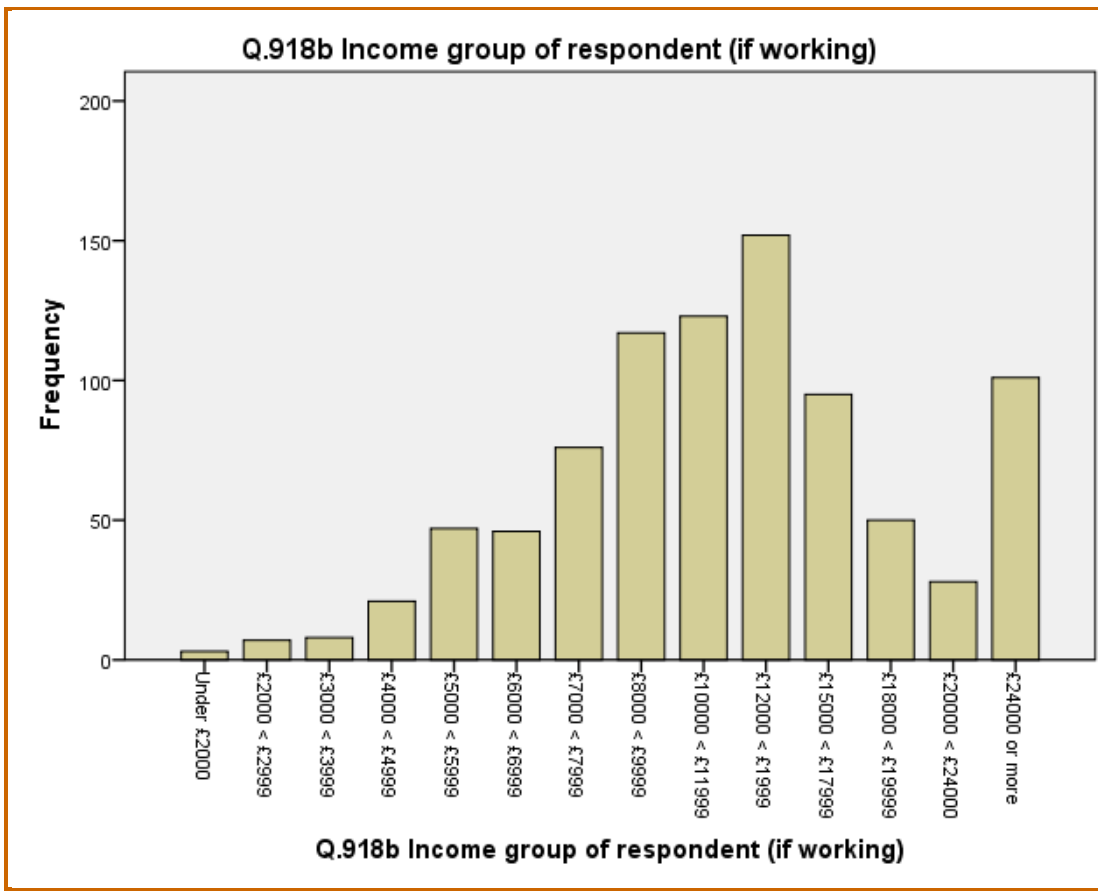
v1727 Q.918b Income group of respondent  
(if working)

N	Valid	874
	Missing	519

v1727 Q.918b Income group of respondent (if working)

	Frequency	Percent	Valid Percent	Cumulative Percent
1 Under £2000	3	.2	.3	.3
2 £2000 - £2999	7	.5	.8	1.1
3 £3000 - £3999	8	.6	.9	2.1
4 £4000 - £4999	21	1.5	2.4	4.5
5 £5000 - £5999	47	3.4	5.4	9.8
6 £6000 - £6999	46	3.3	5.3	15.1
7 £7000 - £7999	76	5.5	8.7	23.8
Valid 8 £8000 - £9999	117	8.4	13.4	37.2
9 £10000 - £11999	123	8.8	14.1	51.3
10 £12000 - £14999	152	10.9	17.4	68.6
11 £15000 - £17999	95	6.8	10.9	79.5
12 £18000 - £19999	50	3.6	5.7	85.2
13 £20000 - £24000	28	2.0	3.2	88.4
14 £24000 or more	101	7.3	11.6	100.0
Total	874	62.7	100.0	
98 Don't know	7	.5		
Missing 99 Not answered	62	4.5		
System	450	32.3		
Total	519	37.3		
Total	1393	100.0		

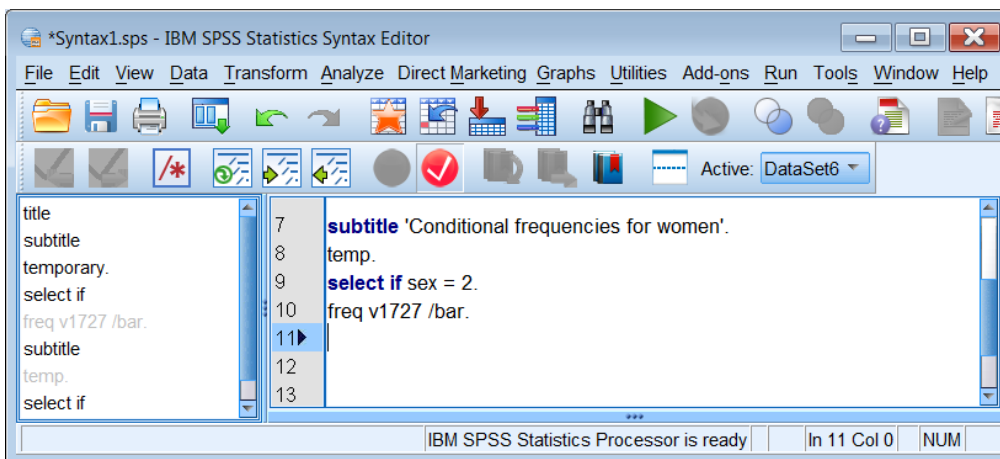
## Men only



## 2: For Women

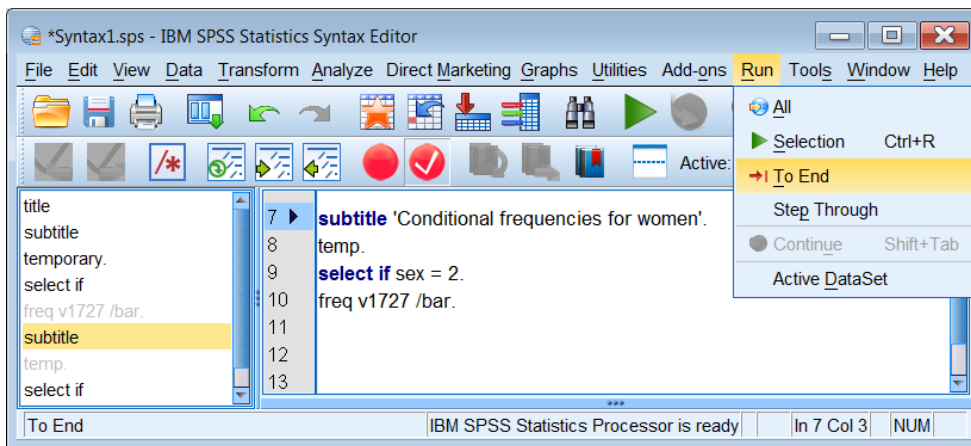
Go back to the syntax editor, add a subtitle and select only women by typing in:

```
subtitle 'Conditional frequencies for women'.  
temp.  
select if sex = 2.  
freq v1727 /bar.
```



[NB: temp. and freq v1727 /bar. are both grayed out in the left pane because I've used abbreviated syntax **temp.** and **freq** instead of **temporary** and **frequencies**.]

Place the cursor in the second **subtitle** line and press **Run > →| To End**:



## Women only

### Statistics

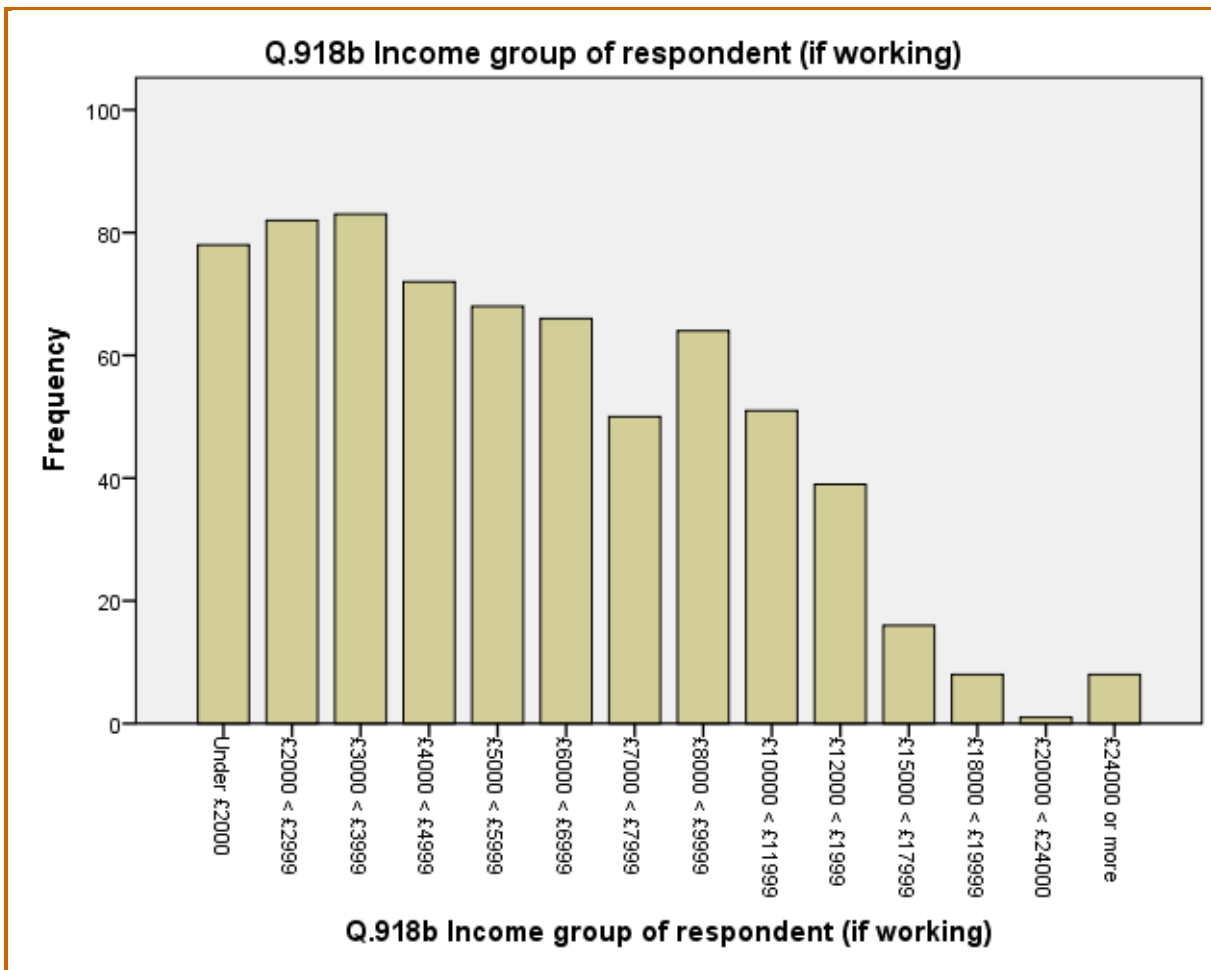
v1727 Q.918b Income group of respondent  
(if working)

N	Valid	686
	Missing	946

**v1727 Q.918b Income group of respondent (if working)**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid				
1 Under £2000	78	4.8	11.4	11.4
2 £2000 - £2999	82	5.0	12.0	23.3
3 £3000 - £3999	83	5.1	12.1	35.4
4 £4000 - £4999	72	4.4	10.5	45.9
5 £5000 - £5999	68	4.2	9.9	55.8
6 £6000 - £6999	66	4.0	9.6	65.5
7 £7000 - £7999	50	3.1	7.3	72.7
8 £8000 - £9999	64	3.9	9.3	82.1
9 £10000 - £11999	51	3.1	7.4	89.5
10 £12000 - £14999	39	2.4	5.7	95.2
11 £15000 - £17999	16	1.0	2.3	97.5
12 £18000 - £19999	8	.5	1.2	98.7
13 £20000 - £24000	1	.1	.1	98.8
14 £24000 or more	8	.5	1.2	100.0
Total	686	42.0	100.0	
Missing				
98 Don't know	10	.6		
99 Not answered	46	2.8		
System	890	54.5		
Total	946	58.0		
Total	1632	100.0		

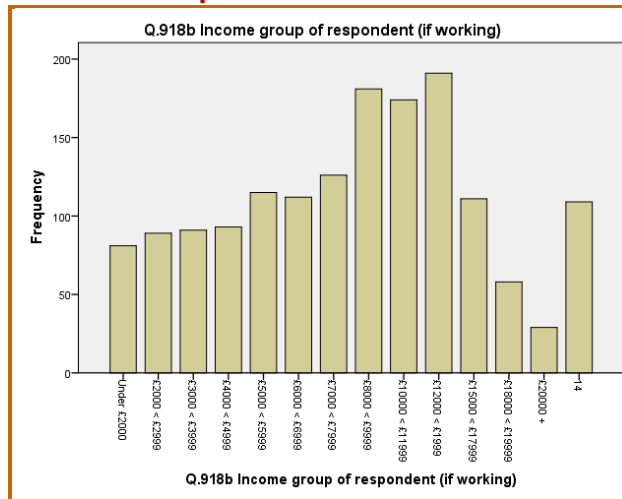
Women only





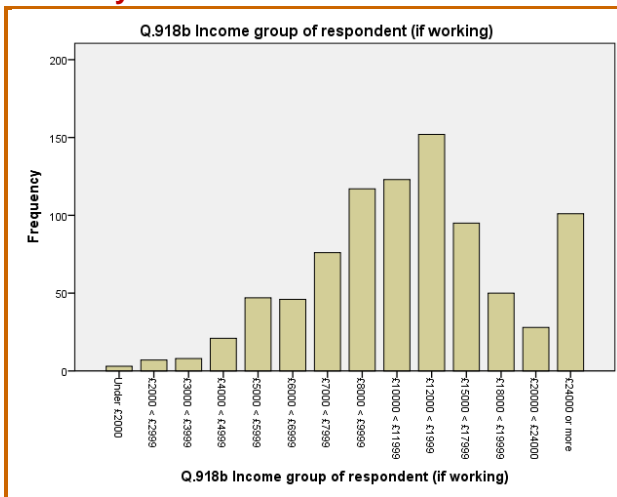
When the distribution for the whole sample:

**Whole sample**

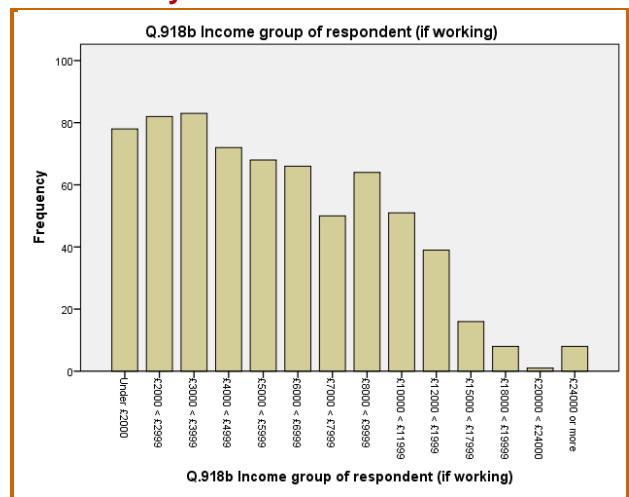


. . is partitioned into separate **conditional distributions** for men and women, there is an obvious difference between the distribution of the earnings of men and women, which you can see much more clearly when the bar-charts are placed side by side.

**Men only**



**Women only**



[NB: These results are based on un-weighted data, but life is complicated enough at this stage so there's no need to start worrying just yet about weighting corrections]

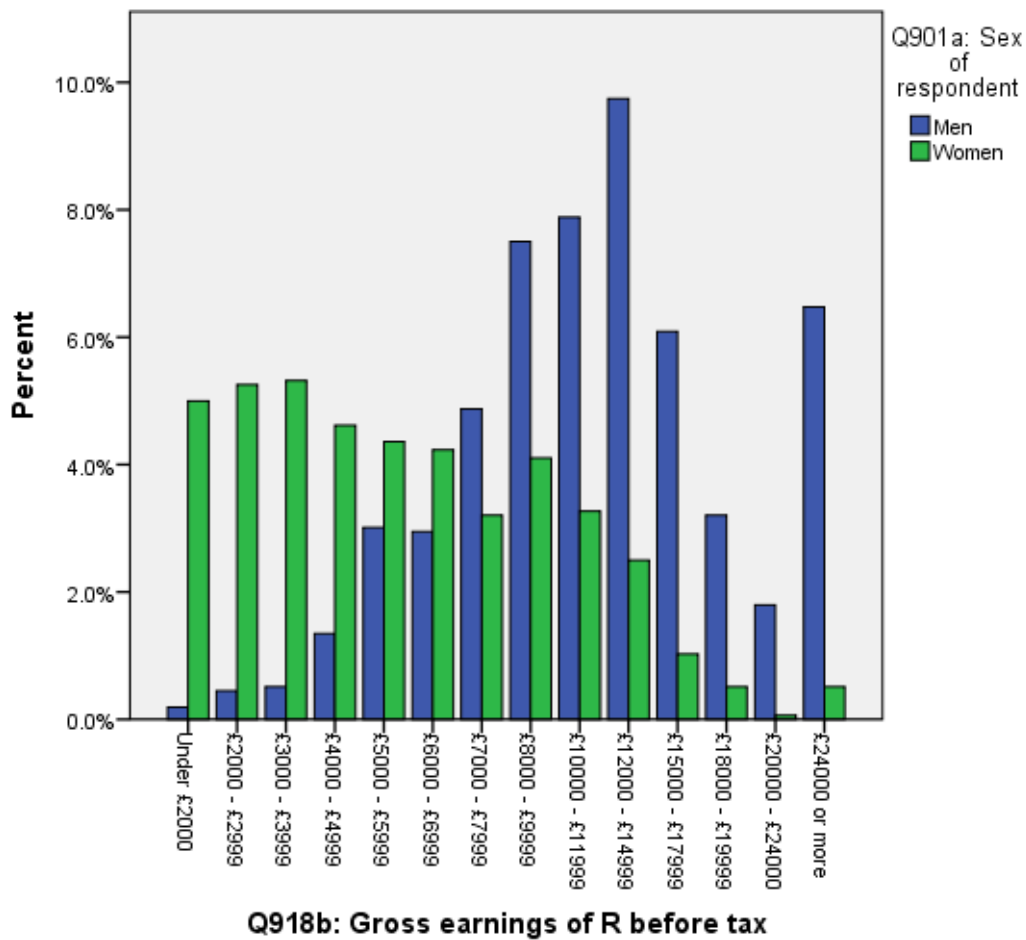
SPSS has another way of producing the above tables and charts using the **SPLIT FILE** command:

**split file by sex.  
freq v1727 /bar.**

This produces exactly the same conditional frequencies as above, but remember to **turn it off** if you want to revert to tables for the whole sample.

**split file off.**

There is also a way of displaying both conditional distributions on the same chart (See page 17)



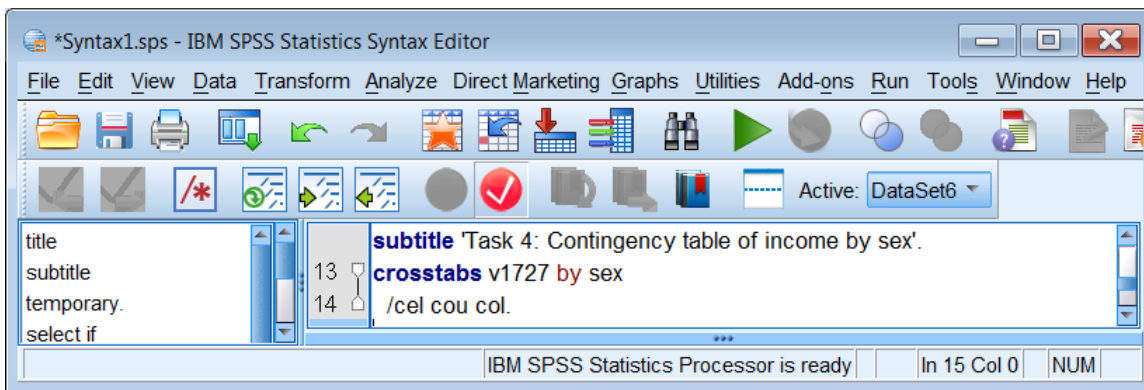
**Task 4:**

Produce a contingency table of earnings by sex (with correct percentages for comparing the earnings of men and women).

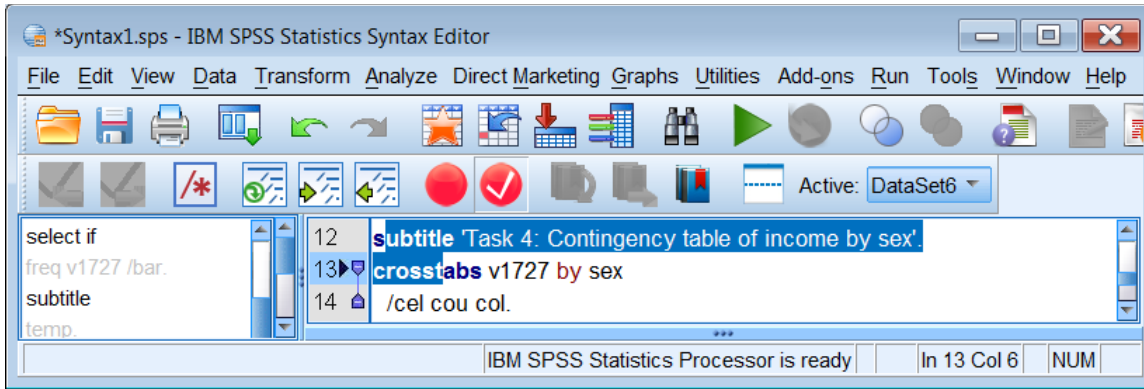
Go back to the **Syntax Editor** and type in:

```

subtitle 'Task 4: Contingency table of income by sex'.
crosstabs v1727 by sex
/cel cou col.
    
```



Highlight both commands:



Click on the green ► to get:

**Case Processing Summary**

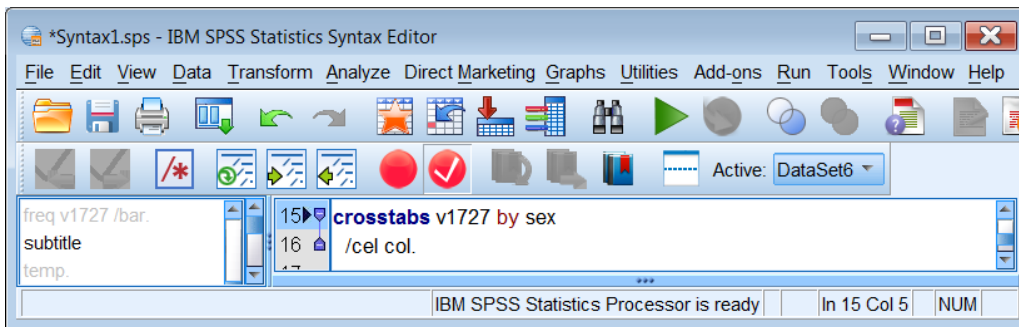
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
v1727 Q.918b Income group of respondent (if working) * sex Q901a Sex of respondent	1560	51.6%	1465	48.4%	3025	100.0%

**v1727 Q.918b Income group of respondent (if working) \* sex Q901a Sex of respondent Crosstabulation**

			sex Q901a Sex of respondent		Total
			1 Men	2 Women	
v1727 Q.918b Income group of respondent (if working)	1 Under £2000	Count	3	78	81
		% within sex Q901a Sex of respondent	0.3%	11.4%	5.2%
	2 £2000 - £2999	Count	7	82	89
		% within sex Q901a Sex of respondent	0.8%	12.0%	5.7%
	3 £3000 - £3999	Count	8	83	91
		% within sex Q901a Sex of respondent	0.9%	12.1%	5.8%
	4 £4000 - £4999	Count	21	72	93
		% within sex Q901a Sex of respondent	2.4%	10.5%	6.0%
	5 £5000 - £5999	Count	47	68	115
		% within sex Q901a Sex of respondent	5.4%	9.9%	7.4%
	6 £6000 - £6999	Count	46	66	112
		% within sex Q901a Sex of respondent	5.3%	9.6%	7.2%
	7 £7000 - £7999	Count	76	50	126
		% within sex Q901a Sex of respondent	8.7%	7.3%	8.1%
8 £8000 - £9999	Count	117	64	181	
	% within sex Q901a Sex of respondent	13.4%	9.3%	11.6%	
9 £10000 - £11999	Count	123	51	174	
	% within sex Q901a Sex of respondent	14.1%	7.4%	11.2%	
10 £12000 - £14999	Count	152	39	191	
	% within sex Q901a Sex of respondent	17.4%	5.7%	12.2%	
11 £15000 - £17999	Count	95	16	111	
	% within sex Q901a Sex of respondent	10.9%	2.3%	7.1%	
12 £18000 - £19999	Count	50	8	58	
	% within sex Q901a Sex of respondent	5.7%	1.2%	3.7%	
13 £20000 - £24000	Count	28	1	29	
	% within sex Q901a Sex of respondent	3.2%	0.1%	1.9%	
14 £24000 or more	Count	101	8	109	
	% within sex Q901a Sex of respondent	11.6%	1.2%	7.0%	
Total	Count	874	686	1560	
	% within sex Q901a Sex of respondent	100.0%	100.0%	100.0%	

This table is incredibly cluttered and it's virtually impossible to interpret with all the raw counts included. However, if we ask for a table with only column percent:

**crosstabs v1727 by sex  
/cel col.**



Click on the green ► to get:

**v1727 Q.918b Income group of respondent (if working) \* sex Q901a Sex of respondent Crosstabulation**  
% within sex Q901a Sex of respondent

		sex Q901a Sex of respondent		Total
		1 Men	2 Women	
v1727 Q.918b Income group of respondent (if working)	1 Under £2000	0.3%	11.4%	5.2%
	2 £2000 - £2999	0.8%	12.0%	5.7%
	3 £3000 - £3999	0.9%	12.1%	5.8%
	4 £4000 - £4999	2.4%	10.5%	6.0%
	5 £5000 - £5999	5.4%	9.9%	7.4%
	6 £6000 - £6999	5.3%	9.6%	7.2%
	7 £7000 - £7999	8.7%	7.3%	8.1%
	8 £8000 - £9999	13.4%	9.3%	11.6%
	9 £10000 - £11999	14.1%	7.4%	11.2%
	10 £12000 - £14999	17.4%	5.7%	12.2%
	11 £15000 - £17999	10.9%	2.3%	7.1%
	12 £18000 - £19999	5.7%	1.2%	3.7%
	13 £20000 - £24000	3.2%	0.1%	1.9%
	14 £24000 or more	11.6%	1.2%	7.0%
Total	100.0%	100.0%	100.0%	

We have now lost the bases for calculating the percents. The table can be edited manually to get rid of the % signs in the table body, insert % in the column headers instead, then replace the **Total** row **100%** with the **base** for percentaging:

**v1727 Q.918b Income group of respondent (if working) \* sex Q901a Sex of respondent Crosstabulation**  
% within sex Q901a Sex of respondent

		sex Q901a Sex of respondent		Total %
		1 Men %	2 Women %	
v1727 Q.918b Income group of respondent (if working)	1 Under £2000	0.3	11.4	5.2
	2 £2000 - £2999	0.8	12.0	5.7
	3 £3000 - £3999	0.9	12.1	5.8
	4 £4000 - £4999	2.4	10.5	6.0
	5 £5000 - £5999	5.4	9.9	7.4
	6 £6000 - £6999	5.3	9.6	7.2
	7 £7000 - £7999	8.7	7.3	8.1
	8 £8000 - £9999	13.4	9.3	11.6
	9 £10000 - £11999	14.1	7.4	11.2
	10 £12000 - £14999	17.4	5.7	12.2
	11 £15000 - £17999	10.9	2.3	7.1
	12 £18000 - £19999	5.7	1.2	3.7
	13 £20000 - £24000	3.2	0.1	1.9
	14 £24000 or more	11.6	1.2	7.0
	<b>(n=100%)</b>	<b>874</b>	<b>686</b>	<b>1560</b>

.. but the table is still difficult to interpret (except for accountants and statisticians who like things to add up to 100% in columns). I find it easier to compare %% in rows rather than columns, but if you do that,

**crosstabs sex by v1727 /cel row.**

.. the **Viewer Window** displays a very wide table in which the columns have to be tediously edited to fit on to a Word page:

**sex Q901a: Sex of respondent \* v1727 Q918b: Gross earnings of R before tax Crosstabulation**

% within sex Q901a: Sex of respondent

		v1727 Q918b: Gross earnings of R before tax														Total
		1 Under £2000	2 £2000 - £2999	3 £3000 - £3999	4 £4000 - £4999	5 £5000 - £5999	6 £6000 - £6999	7 £7000 - £7999	8 £8000 - £9999	9 £10000 - £11999	10 £12000 - £1999	11 £15000 - £17999	12 £18000 - £19999	13 £20000 - £24000	14 £24000 or more	Total
sex Q901a:	1 Men	0.3%	0.8%	0.9%	2.4%	5.4%	5.3%	8.7%	13.4%	14.1%	17.4%	10.9%	5.7%	3.2%	11.6%	100.0%
Sex of respondent	2 Women	11.4%	12.0%	12.1%	10.5%	9.9%	9.6%	7.3%	9.3%	7.4%	5.7%	2.3%	1.2%	0.1%	1.2%	100.0%
	Total	5.2%	5.7%	5.8%	6.0%	7.4%	7.2%	8.1%	11.6%	11.2%	12.2%	7.1%	3.7%	1.9%	7.0%	100.0%

.. then de-cluttered as above and amended with **n = 100%**:

**sex Q901a Sex of respondent \* v1727 Q.918b Income group of respondent (if working) Crosstabulation**

% within sex Q901a Sex of respondent

		v1727 Q.918b Income group of respondent (if working)														Total
		Under £2000	£2000 - £2999	£3000 - £3999	£4000 - £4999	£5000 - £5999	£6000 - £6999	£7000 - £7999	£8000 - £9999	£10000 - £11999	£12000 - £1999	£15000 - £17999	£18000 - £19999	£20000 - £24000	£24000 or more	Total
		%	%	%	%	%	%	%	%	%	%	%	%	%	%	n = 100%
sex Q901a	1 Men	0.3	0.8	0.9	2.4	5.4	5.3	8.7	13.4	14.1	17.4	10.9	5.7	3.2	11.6	874
Sex of respondent	2 Women	11.4	12.0	12.1	10.5	9.9	9.6	7.3	9.3	7.4	5.7	2.3	1.2	0.1	1.2	686
	Total	5.2%	5.7	5.8	6.0	7.4	7.2	8.1	11.6	11.2	12.2	7.1	3.7	1.9	7.0	1560

This table has been modified below to reorganise the rows so that the frequency distribution of income for the whole sample now appears in the top row. The next two rows partition this distribution into **conditional frequencies** for men and women. This is a much more logical presentation which cannot be produced by SPSS without special programming.

**sex Q901a Sex of respondent \* v1727 Q.918b Income group of respondent (if working) Crosstabulation**

% within sex Q901a Sex of respondent

		Q.918b Income group of respondent (if working)														Total
		Under £2000	£2000 - £2999	£3000 - £3999	£4000 - £4999	£5000 - £5999	£6000 - £6999	£7000 - £7999	£8000 - £9999	£10000 - £11999	£12000 - £1999	£15000 - £17999	£18000 - £19999	£20000 - £24000	£24000 or more	Total
		%	%	%	%	%	%	%	%	%	%	%	%	%	%	n = 100%
<b>Total</b>		<b>5.2</b>	<b>5.7</b>	<b>5.8</b>	<b>6.0</b>	<b>7.4</b>	<b>7.2</b>	<b>8.1</b>	<b>11.6</b>	<b>11.2</b>	<b>12.2</b>	<b>7.1</b>	<b>3.7</b>	<b>1.9</b>	<b>7.0</b>	<b>1560</b>
<b>Men</b>		0.3	0.8	0.9	2.4	5.4	5.3	8.7	13.4	14.1	17.4	10.9	5.7	3.2	11.6	874
<b>Women</b>		11.4	12.0	12.1	10.5	9.9	9.6	7.3	9.3	7.4	5.7	2.3	1.2	0.1	1.2	686
<b>Epsilon</b>		<b>-11.1</b>	<b>-11.2</b>	<b>-11.2</b>	<b>-8.1</b>	<b>-4.5</b>	<b>-4.3</b>	<b>+1.4</b>	<b>+4.1</b>	<b>+6.7</b>	<b>+11.7</b>	<b>+8.6</b>	<b>+4.5</b>	<b>+3.1</b>	<b>+10.4</b>	

One analysis we can perform on this table is to calculate the percentage point difference (**epsilon**) between men and women for each category, which I've inserted as an additional row.

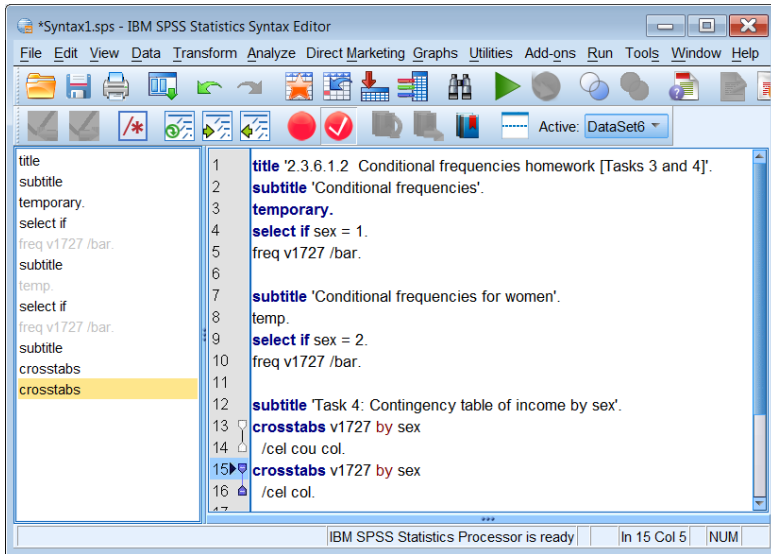
This table gives us a summary of the difference between men and women at each income level with a clear shift from (high) **negative** to (high-ish) **positive** epsilons as the income bands increase. The next stage of analysis is to see what happens to these differences when controlling for one or more test variables. This process is known as **elaboration**.

When we start introducing test variables into contingency tables, the number of cells in the tables can rapidly get very large and the counts within the cells rapidly very small, too small in fact to act as a base for percentages. We therefore need to think about grouping variables into far fewer categories.

Before progressing to Block 3, you might try deciding from the above table where the optimal cutting points should be to reduce income groups from fourteen categories to three.

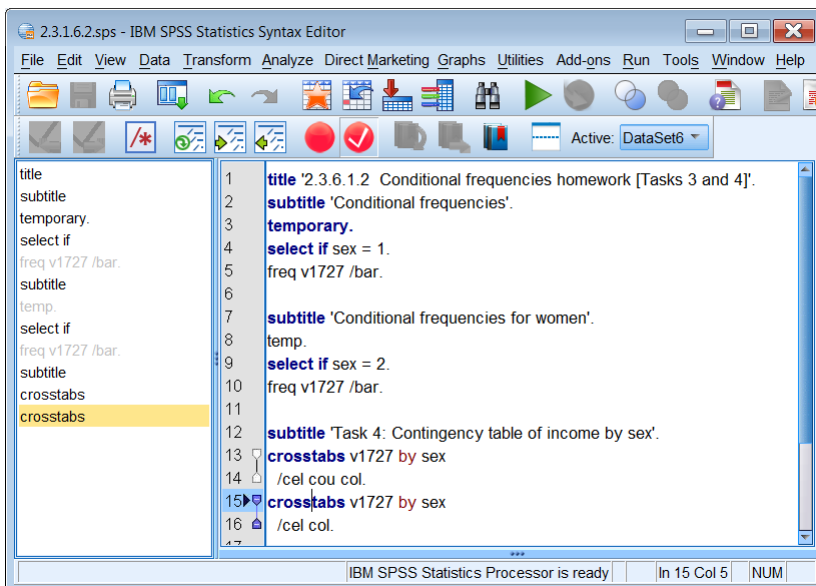
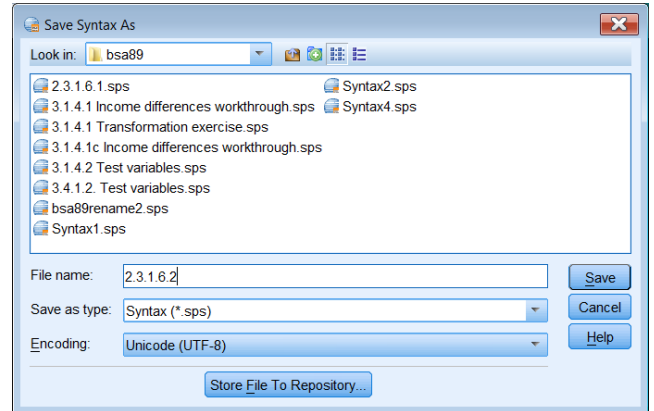
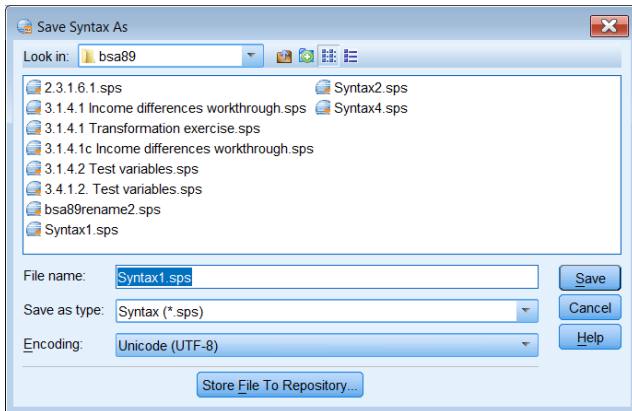
You can also be thinking about other variables that might affect income.

There's no need to save the **Data Editor** as you haven't made any changes, but it's a good idea to save the **Syntax Editor**. You should be able to do this yourself by now, but mine is on the site as **2.3.1.6.2.sps** (to tally with the title of this tutorial).



Click on **Syntax1.sps**

and change it to **2.3.1.6.2**



**End of tutorial:** 2.1.3.6.2 Specimen answer for conditional frequencies homework

**Forward to page:** [Block 3: Analysing two variables \( and sometimes three\)](#)

**Back to:** [2.3.1.5 Conditional frequencies homework exercises](#)

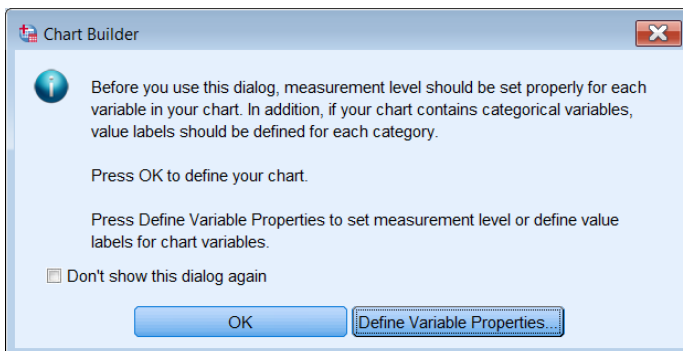
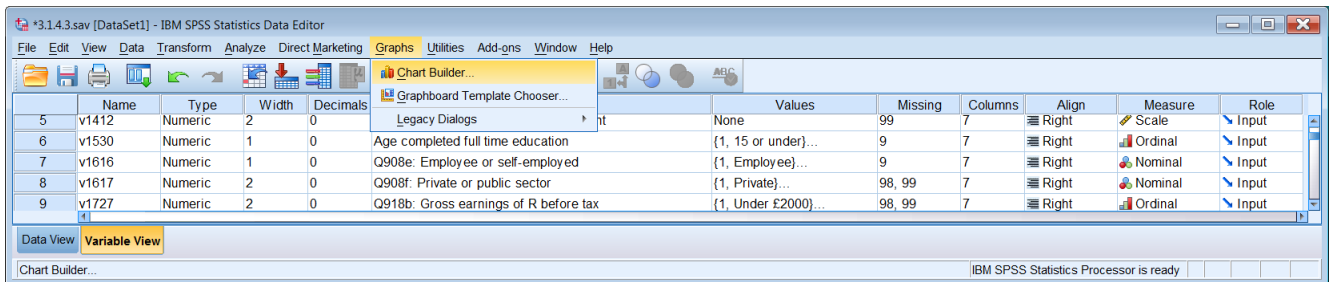
**Back to:** [2.3.1.6.1 Specimen answer for conditional frequencies homework \[Tasks 1 and 2\]](#)

**Back to page:** [Block 2: Analysing one variable](#)

**Back to page:** [Summary guide to SPSS tutorials](#)

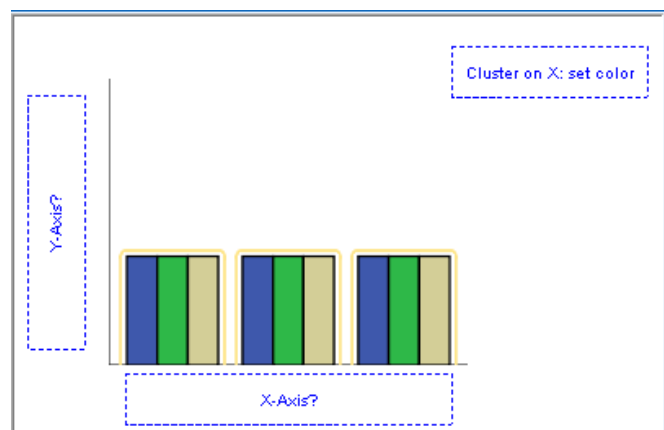
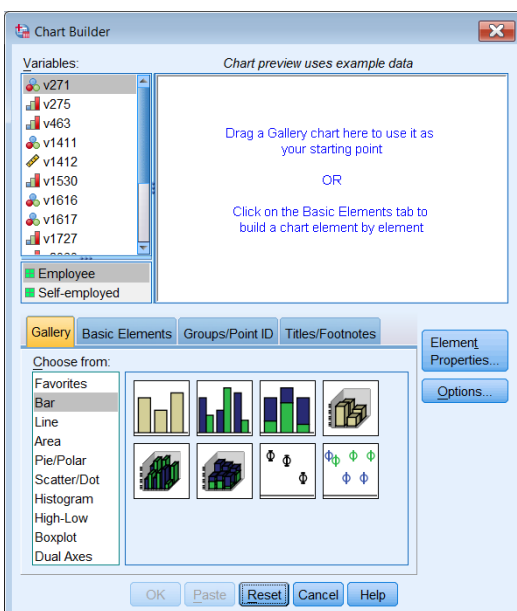
The chart on page 10 above was produced with:

**Graph > Chart builder**

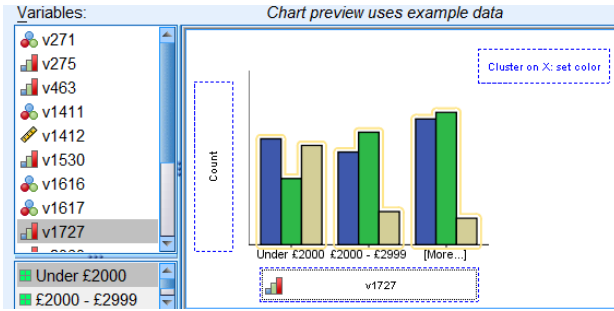


Both variables are Nominal or Ordinal so click on **OK**

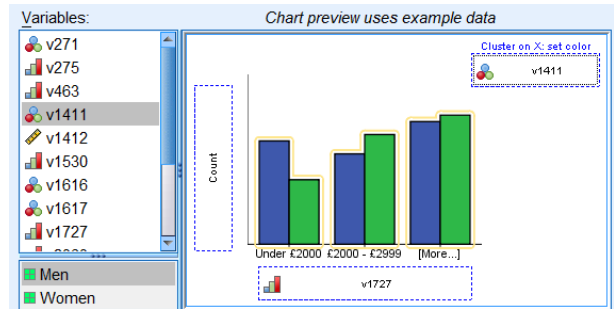
Click on **Bar** and drag cluster gallery chart to the white pane:



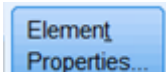
Drag **v1727** to the X-axis



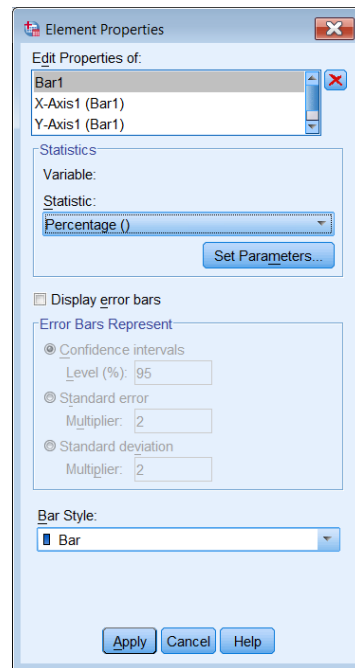
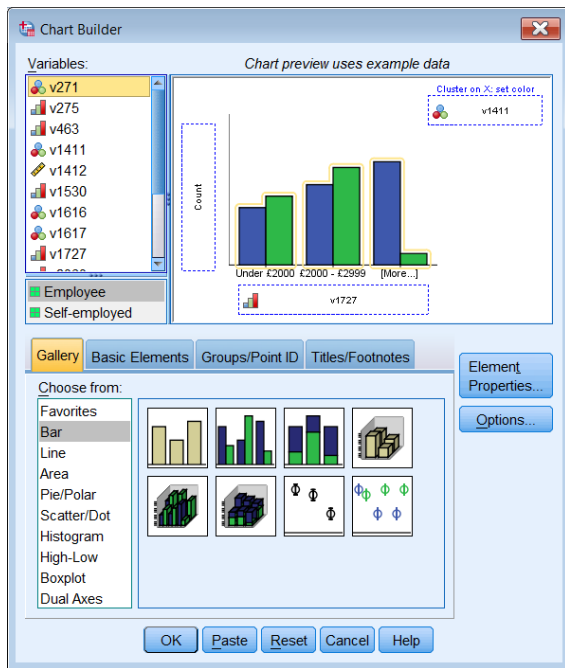
and **v1411** to the Cluster on X: box



Click on



and set **Statistic** to **Percentage ()**



This produces a horrendous mass of syntax, so stick to the GUI!

```

1 * Chart Builder.
2 **GRAPH
3 /GRAPHDATASET NAME="graphdataset" VARIABLES=v1727 COUNT(1)[name="COUNT"] v1411 MISSING=LISTWISE REPORTMISSING=NO
4 /GRAPHSPEC SOURCE=INLINE.
5 **BEGIN GPL
6 SOURCE: s=userSource(id("graphdataset"))
7 DATA: v1727=col(source(s), name("v1727"), unit.category())
8 DATA: COUNT=col(source(s), name("COUNT"))
9 DATA: v1411=col(source(s), name("v1411"), unit.category())
10 COORD: rect(dim(1,2), cluster(3,0))
11 GUIDE: axis(dim(3), label("Q918b: Gross earnings of R before tax"))
12 GUIDE: axis(dim(2), label("Percent"))
13 GUIDE: legend(aesthetic(aesthetic.color.interior), label("Q901a: Sex of respondent"))
14 SCALE: cat(dim(3), include("1", "2", "3", "4", "5", "6", "7", "8", "9", "10", "11", "12", "13", "14"))
15 SCALE: linear(dim(2), include(0))
16 SCALE: cat(aesthetic(aesthetic.color.interior), include("1", "2"))
17 SCALE: cat(dim(1), include("1", "2"))
18 ELEMENT: interval(position(summary.percent(v1411*COUNT*v1727, base.all(acrossPanels()))), color.interior(v1411), shape.interior(shape.square))
19 **END GPL.

```