

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

[New page: 22 January 2018, updated 16 June 2018]

John MacInnes

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

(Sage, Dec. 2017)

Video 12: Boxplots ¹ (3'48")

Chapter 4, paragraph 23: Creating a Boxplot with SPSS (p 101)

Unlike earlier tutorials, JM does not use syntax in the boxplot video: the boxplot is produced entirely from the GUI and associated menus. At no point does he PASTE any syntax, but this may not be possible with some chart editing procedures. It would be incredibly complex anyway and is much easier in the GUI. I had difficulty in exactly replicating the boxplot chart. SPSS 24 is different from JM's version and my computer is a PC whereas his seems to be a Mac. OK, here we go.

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](#)

[4.1.7 Guide to video tutorials 10 and 11](#)

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](#))

¹ For further information about boxplots see [Appendix: What is a boxplot?](#)

The **Boxplot** video assumes that you have already created the **filter** restricting the data to a subset of people aged between 25 and 54 and that you have also created a new variable **[country1]** in which the first group contains only cases from selected countries (France, Germany, UK, Netherlands, Portugal, Russian Federation and Sweden) and the second group contains all other countries.

If you haven't already done this, the safest method is to copy the following syntax and run it on the original SPSS saved file **ESS6_Practice.sav**. Instead of **FILTER_\$** it recodes **[agea]** into a new variable **[agegrp3]** with three age groups (16-24, 25-54, 55+) and temporarily selects those aged 25-54.

autorecode variables = cntry /into country .

compute country1 = country .

recode country1 (7 12 13 20 23 24 25 = 1)(ELSE = 0)
/agea (16 thru 24 = 1)(25 thru 54 = 2)(missing = -1)(else = 3) into agegrp3 .

formats country1 agegrp3 (f2.0) .

variable level

cntry (nominal)

/agegrp3 (ordinal) .

missing values agegrp3 (-1) .

variable labels

country1 'selection of 7 countries'

/agegrp3 'Age recoded into three groups' .

value labels

agegrp3 1 '16-14' 2 '25-54' 3 '55+'

/country1 0 'Not selected' 1 'Selected' .

***check1** .

frequencies agegrp3 country1 .

Age recoded into three groups

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	16-14	6579	12.0	12.1	12.1
	25-54	26202	47.9	48.0	60.1
	55+	21759	39.8	39.9	100.0
	Total	54540	99.8	100.0	
Missing	-1	133	0.2		
Total		54673	100.0		

selection of 7 countries

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not selected	39134	71.6	71.6	71.6
	Selected	15539	28.4	28.4	100.0
	Total	54673	100.0	100.0	

***check2** .

temporary .

select if (agegrp3 = 2) and (country1 = 1).

crosstabs country by gndr .

Country * Gender Crosstabulation

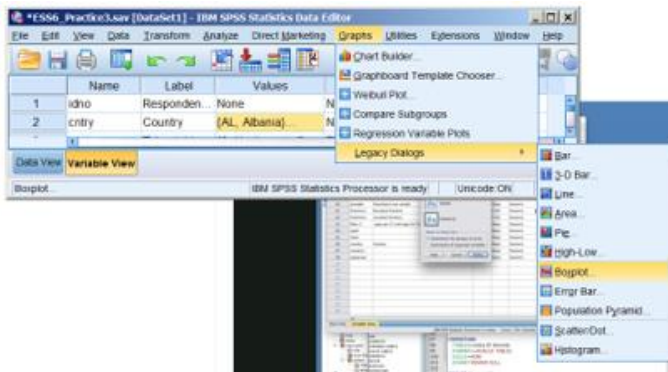
Count

		Gender		Total
		Male	Female	
Country	Germany	699	696	1395
	France	428	487	915
	United Kingdom	420	595	1015
	Netherlands	412	479	891
	Portugal	394	534	928
	Russian Federation	529	751	1280
	Sweden	433	415	848
Total		3315	3957	7272

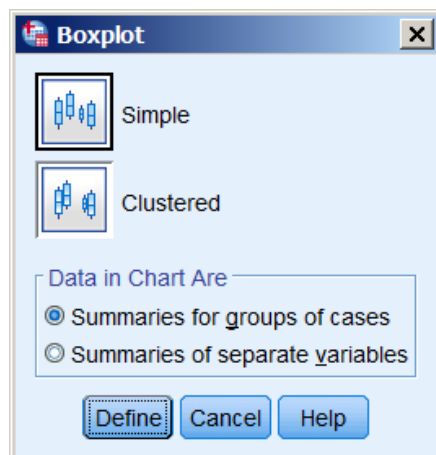
Producing the boxplot

To produce the boxplot from the **Data Editor**, JM uses the GUI.

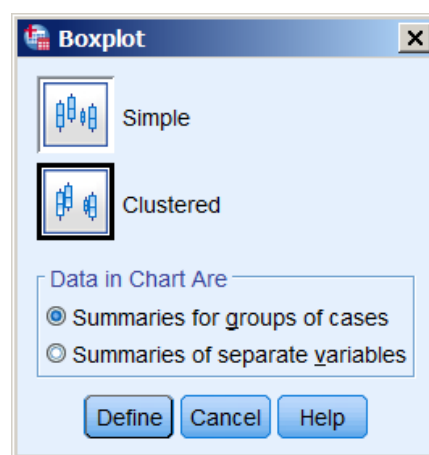
Graphs >> Legacy Dialogs >> Boxplot :



Opening **Boxplot** dialog

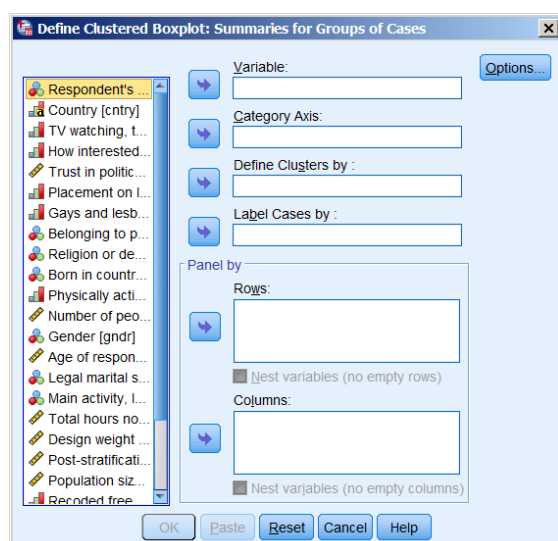


Checks **Clustered**

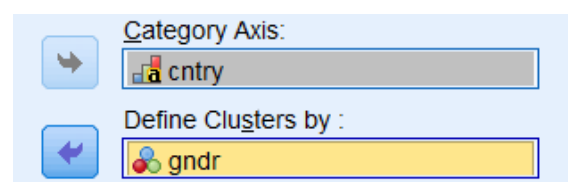


(1'12") Clicks **Define** to get:

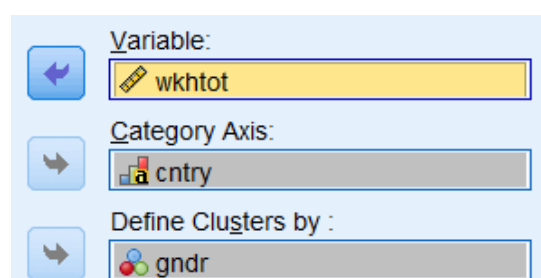
Define Clustered Boxplot opening dialog



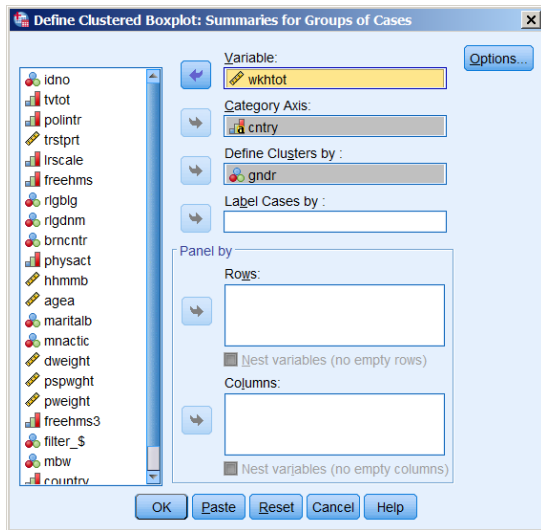
Two of the boxes have already been filled in:



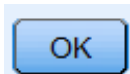
JM transfers **[wkhtot]** to the Variable box:



He explains that he has defined the clusters by **[gndr]** as he wants separate boxplots "for each country for men and for women."




(1'43") . . . JM clicks on



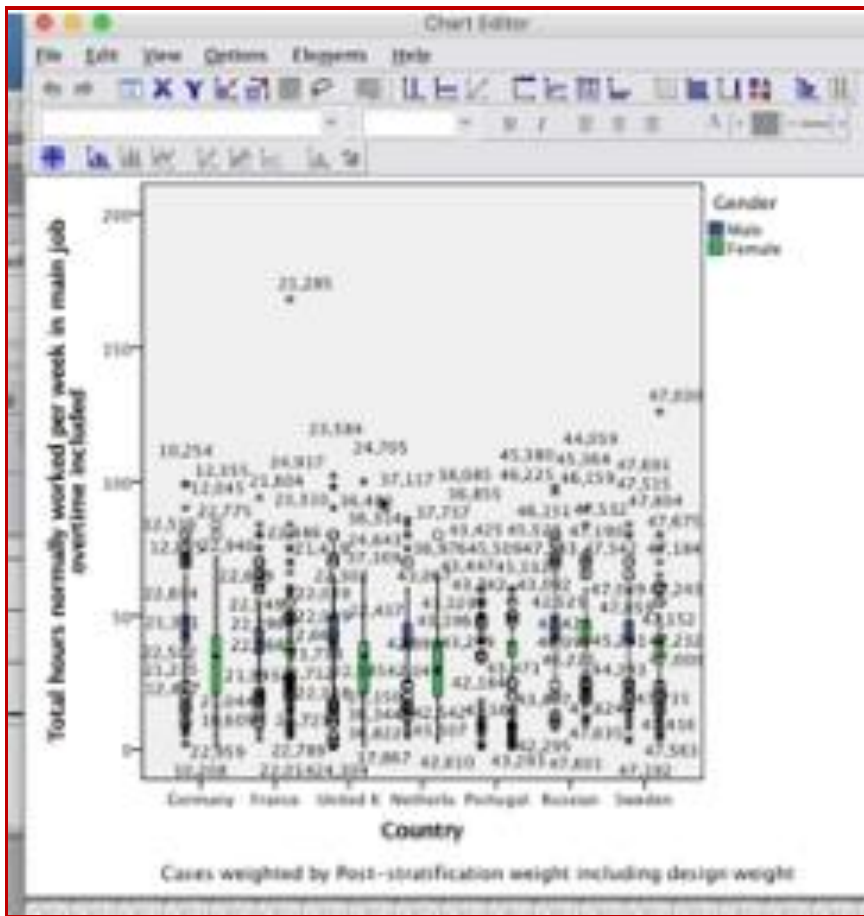
²

to get this summary table:

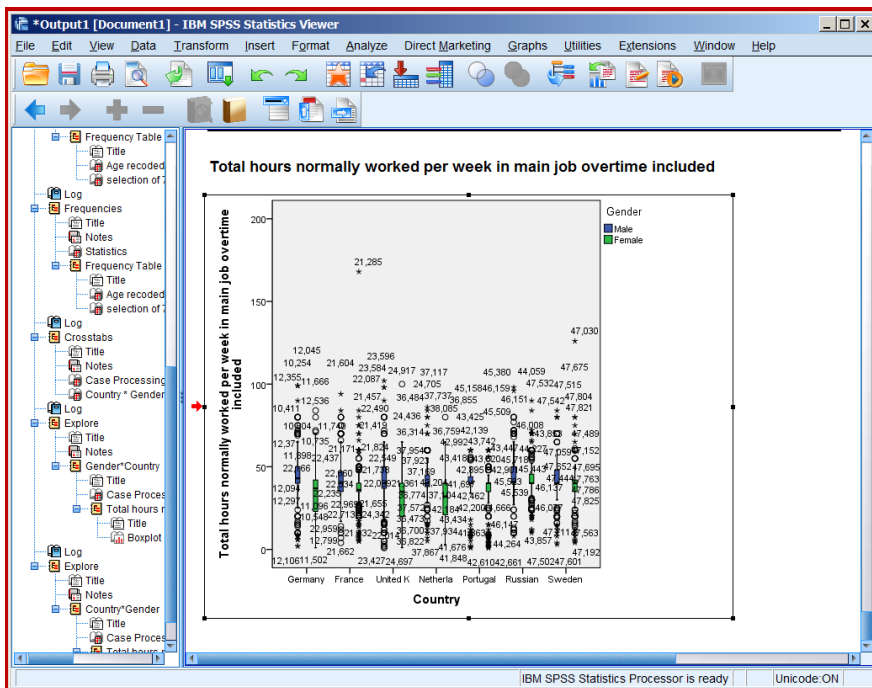
Case Processing Summary								
	Gender Country		Cases					
			Valid		Missing		Total	
			N	Percent	N	Percent	N	Percent
Total hours normally worked per week in main job overtime included	Male	Germany	687	98.3%	12	1.7%	699	100.0%
		France	423	98.8%	5	1.2%	428	100.0%
		United Kingdom	400	95.2%	20	4.8%	420	100.0%
		Netherlands	401	97.3%	11	2.7%	412	100.0%
		Portugal	300	76.1%	94	23.9%	394	100.0%
		Russian Federation	380	71.8%	149	28.2%	529	100.0%
		Sweden	428	98.8%	5	1.2%	433	100.0%
	Female	Germany	669	96.1%	27	3.9%	696	100.0%
		France	465	95.5%	22	4.5%	487	100.0%
		United Kingdom	544	91.4%	51	8.6%	595	100.0%
		Netherlands	453	94.6%	26	5.4%	479	100.0%
		Portugal	384	71.9%	150	28.1%	534	100.0%
		Russian Federation	543	72.3%	208	27.7%	751	100.0%
		Sweden	401	96.6%	14	3.4%	415	100.0%

² Using the GUI to get the boxplot chart is complex and frustrating: I found it much easier and quicker to use the syntax from 

.. and, as he says, "this incredibly messy looking graph."



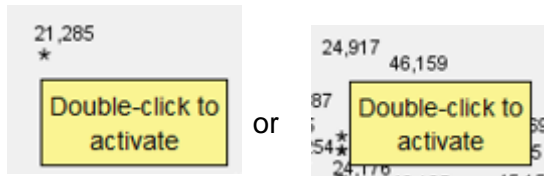
[Screenshot from JM's video]



[Screenshot from SPSS 24 viewer]

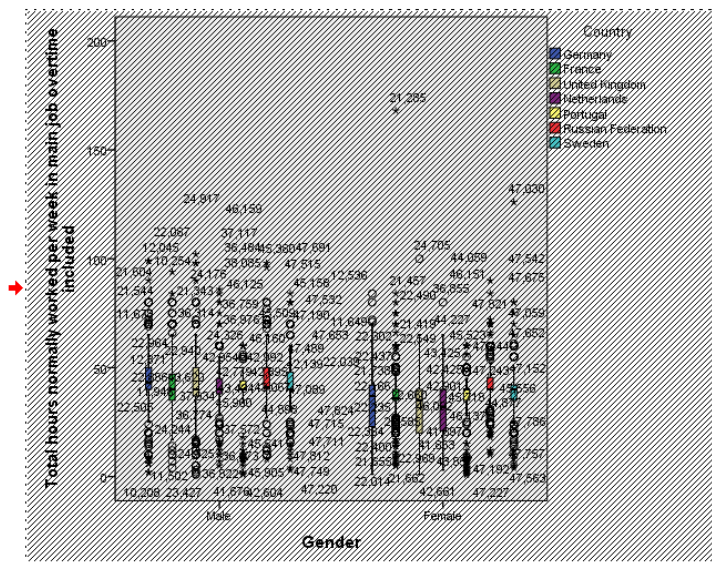
Editing the boxplot

If you hover the cursor over the chart you will see:

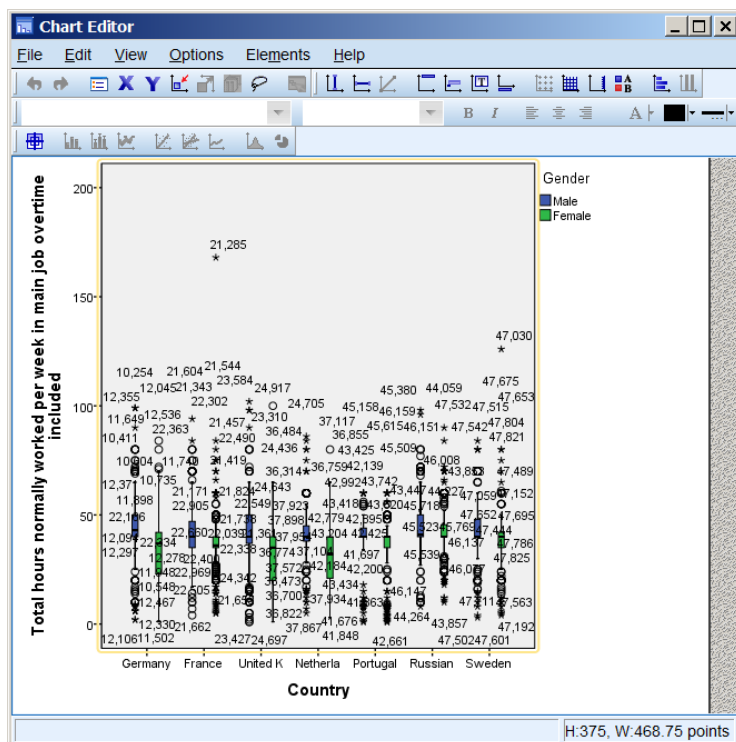


SPSS has automatically labelled the data points with case numbers, but JM gets rid of these by double clicking (anywhere inside) the chart:



Total hours normally worked per week in main job overtime included



.. to activate it and get the **Chart Editor** dialog:




To replicate JM's example in SPSS 24:

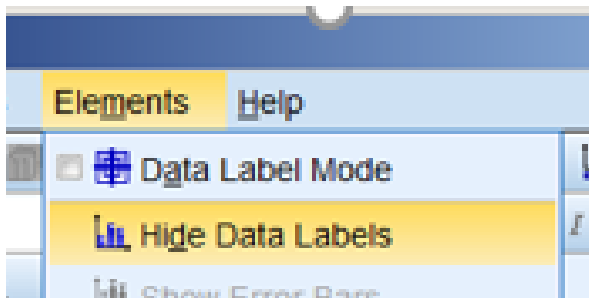
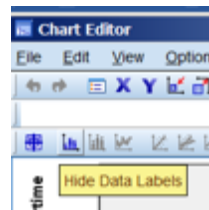
Hover over the SPSS icon  in the taskbar: a new icon  **Chart Editor** has appeared:



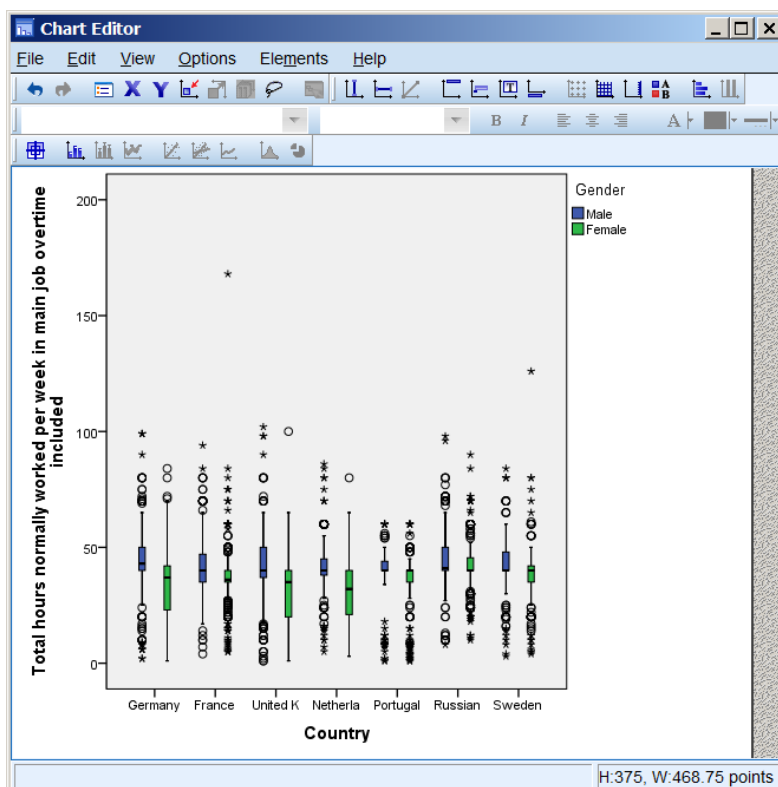
In the **Chart Editor**

click on **Elements** or on  at top left

then on **Hide Data Labels**

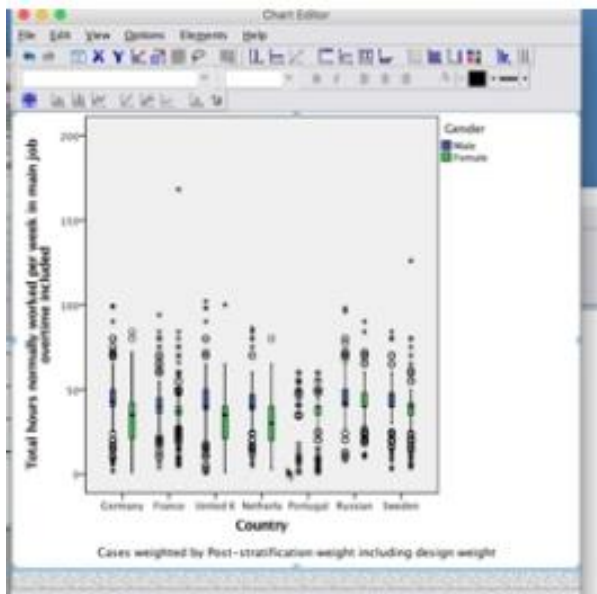


... to get:

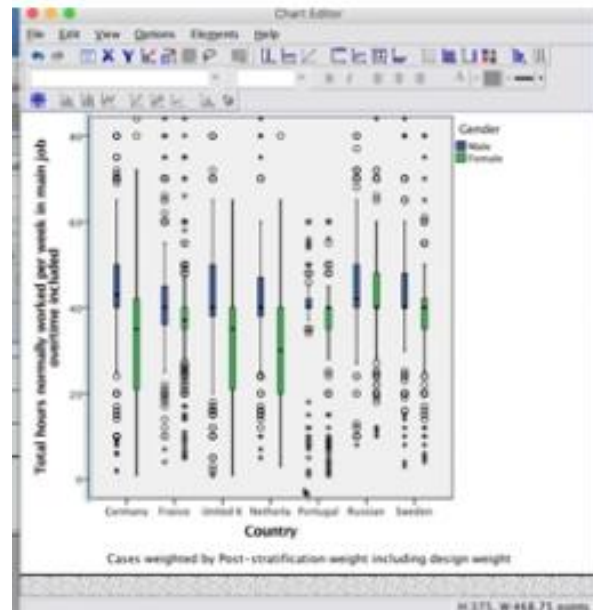


(2'40") JM now changes the scale on the vertical axis from

max **200**

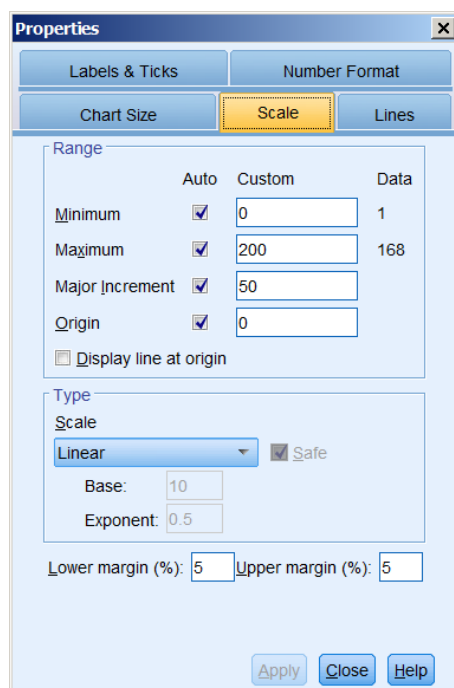
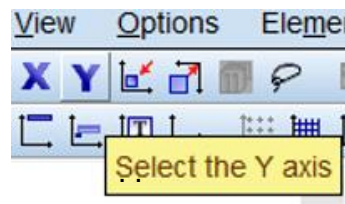


to max **80**

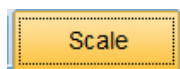


To do this in SPSS 24

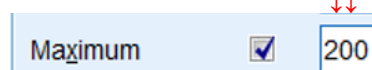
In the **Chart Editor** Click  in the toolbar



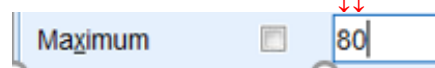
Click on



.. change **200**



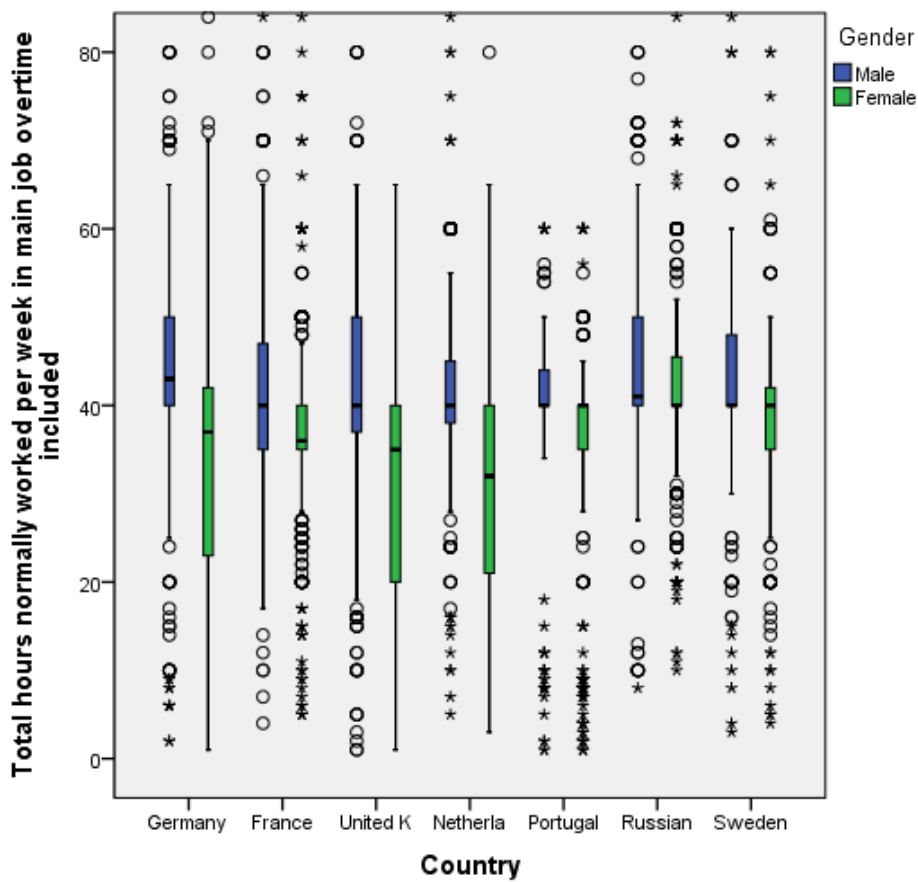
.. to **80**



.. and click

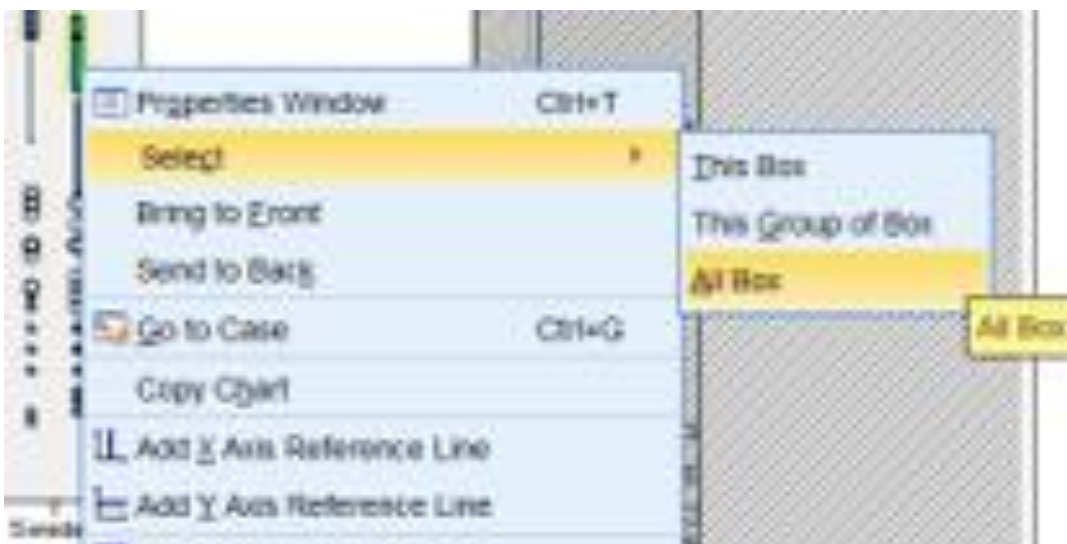


.. to get:



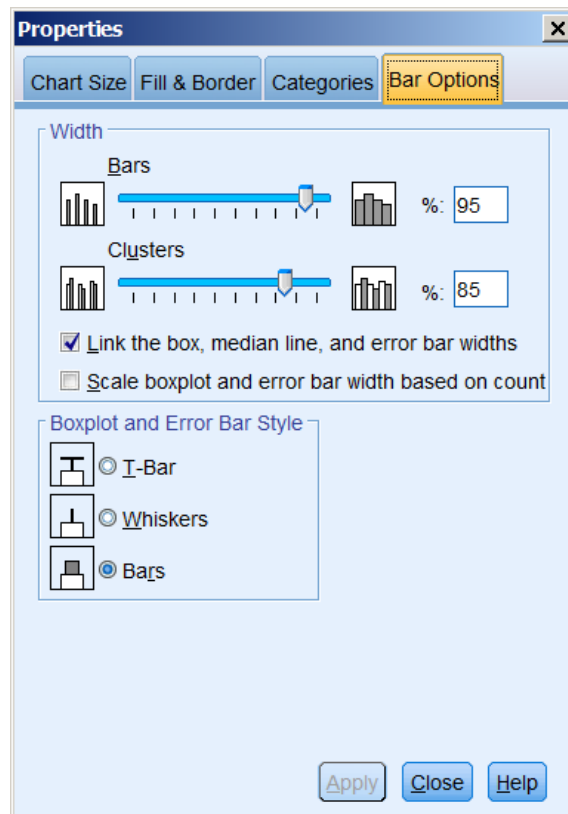
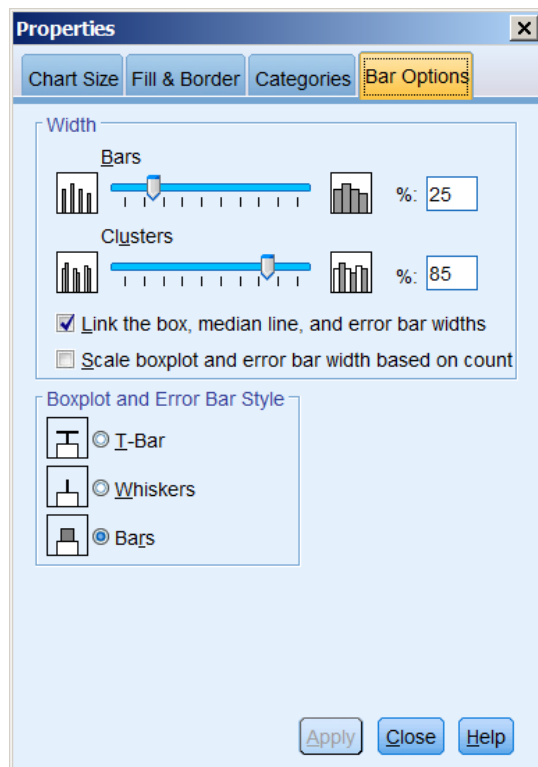
JM now widens the bars to make the chart clearer:

Right click on any box:

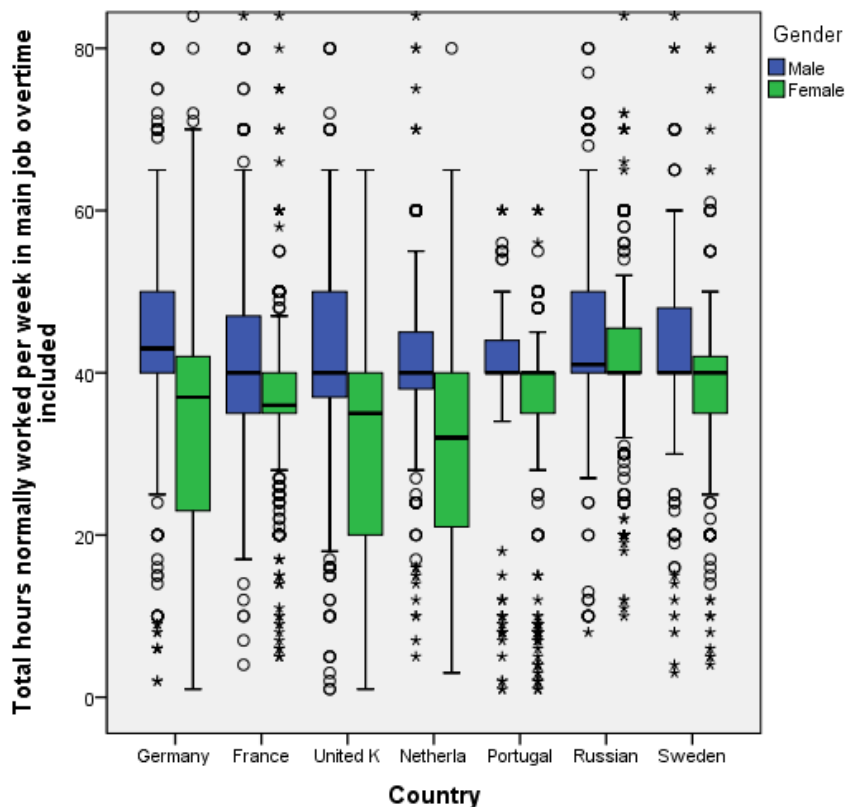


Select >> All Box >> Bar Options

JM drags Bars slider to 95%

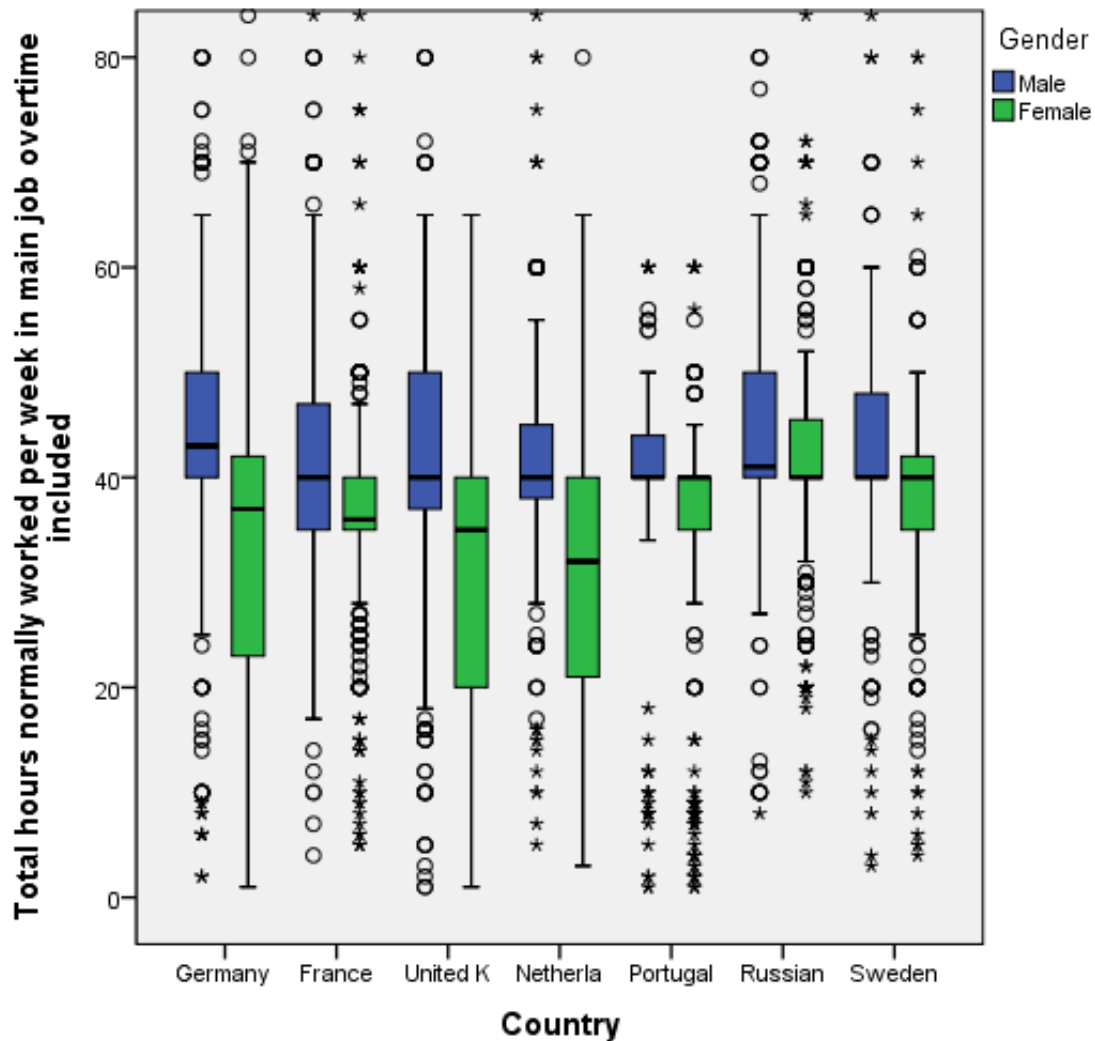


.. and gets:



(See Fig. 4.26 on page 102)

. . but increasing to 70% yields a clearer chart.



[NB: JM keeps calling it a **barchart**, but it's a **boxplot**.]

End of: 4.1.8: Guide to video tutorial 12 for Chapter 4

Back to: [4.1.7: Guide to video tutorials 10 and 11 for Chapter 4](#)
 Back to: [MacInnes \(2017\)](#)

Appendix: What is a boxplot?

A boxplot is a diagram which illustrates the range of values for a variable, from lowest to highest, and displays a box, the edges of which are determined by the lower and upper quartile values, and which also has a line depicting the median value. It also has whiskers illustrating the values between the quartiles and the extremes.

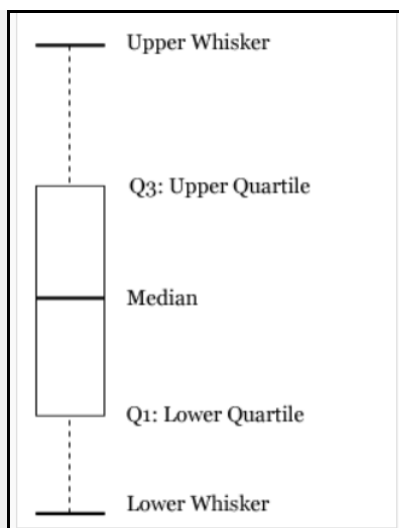
[Boxplots](#) is an explanation within SPSS: [Boxplot examples](#) explains how to get different types,

There are several sites and blogs which explain what box-plots are and how they are created and used. I can recommend:

[Box Plot: Display of Distribution](#) (College of Saint Benedict & Saint John's University)

[How to Read and Use a Box-and-Whisker Plot](#) (Flowing Data)
(Did you know the box and whisker plot was invented by Tukey?)

[Box Plot](#) (from [Information and Visualisation](#))

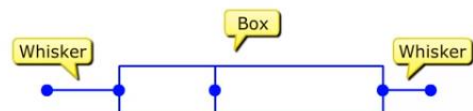


© Copyright Information & Visualization

Uses data from the [World Freedom Atlas](#)

[Box plots](#) (from [BrainingCamp](#))

A box plot, also known as a box-and-whisker plot, is a good way to summarize large amounts of data.



© Copyright  **Brainingcamp**

Uses snowfall in ski resorts as an example, but explains the principles of ordering the data from lowest to highest. A very basic tutorial is [Variability, Range, Interquartile, Mean Absolute Deviation \(Measures of Variability\)](#) Other basic maths videos are available on [BrainingCamp](#)

A quick search on Google will throw up other useful links such as:

[Box and whisker plots](#) (Khan Academy)

[Box plot - Wikipedia](#)

[What a Boxplot Can Tell You about a Statistical Data Set](#) (Dummies.com)

[Statistics: boxplot \(English\) - YouTube](#) (and links to others)

[Interpret the key results for Boxplot - Minitab Express](#)