MATH 124 – Drawing boxplots
bolstad_math124@bmbolstad.com
http://math124sfsu.bmbolstad.com

The purpose of this document is to guide you through the steps needed to draw a boxplot in Excel, including multiple boxplots depending on the value of a categorical variable. For this document we will use the data from exercise 2.11 in the textbook. However, you should be able to follow the same methods on the body data for the assignment. For your reference the data file in both tab delimited text and Excel 2000 formats is on the webpage. Note that Excel does not let you draw a boxplot with outliers indicated.

Making a simple boxplot

1. First we need to make a small table containing the LQ, UQ, Median, Minimum and Maximum values of the variable we want to boxplot. In this document we will boxplot the Mass variable. First go to an empty part of the spreadsheet and type in the words "LQ", "Median", "Minimum", "Maximum" and "UQ"each in a separate cell so that they form a column. You might like to put a heading on the adjacent column. So your spreadsheet should look something like this:
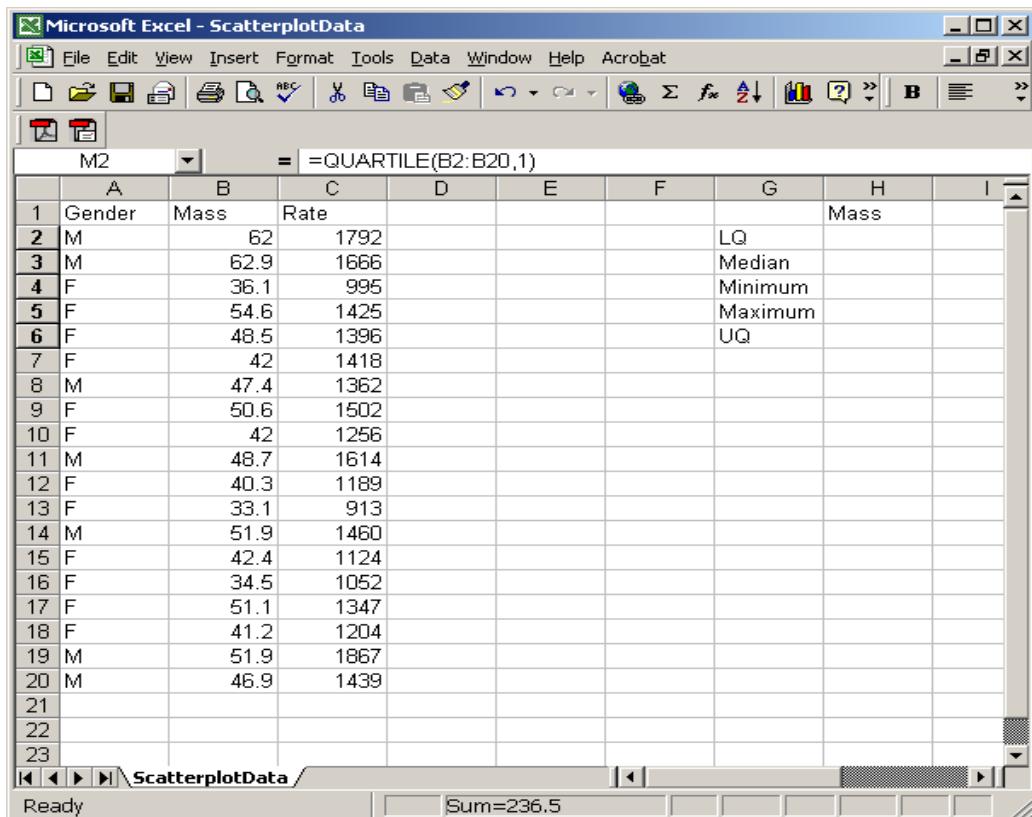
2. Next we need to compute each of the different quantities. So go to the cell adjacent to the word LQ and type "=quartile(B2:B20,1)" to compute the lower quartile of the mass data. Similarly, go next the cell adjacent to UQ and type "=quartile(B2:B20,3)". Now go to the cell next to Median and type "=median(B2:B20)". For the minimum type "=min(B2:B20)" and the maximum type "=max(B2:B20)". Your spreadsheet should now look something like this

	A	B	C	D	E	F	G	H	I
1	Gender	Mass	Rate					Mass	
2	M	62	1792				LQ	41.6	
3	M	62.9	1666				Median	47.4	
4	F	36.1	995				Minimum	33.1	
5	F	54.6	1425				Maximum	62.9	
6	F	48.5	1396				UQ	51.5	
7	F	42	1418						
8	M	47.4	1362						
9	F	50.6	1502						
10	F	42	1256						
11	M	48.7	1614						
12	F	40.3	1189						
13	F	33.1	913						
14	M	51.9	1460						
15	F	42.4	1124						
16	F	34.5	1052						
17	F	51.1	1347						
18	F	41.2	1204						
19	M	51.9	1867						
20	M	46.9	1439						
21									
22									
23									

ScatterplotData

Ready

3. It is now time to start drawing the boxplot. The first thing you need to do is select the whole table that you just created.

	A	B	C	D	E	F	G	H	I
1	Gender	Mass	Rate					Mass	
2	M	62	1792				LQ	41.6	
3	M	62.9	1666				Median	47.4	
4	F	36.1	995				Minimum	33.1	
5	F	54.6	1425				Maximum	62.9	
6	F	48.5	1396				UQ	51.5	
7	F	42	1418						
8	M	47.4	1362						
9	F	50.6	1502						
10	F	42	1256						
11	M	48.7	1614						
12	F	40.3	1189						
13	F	33.1	913						
14	M	51.9	1460						
15	F	42.4	1124						
16	F	34.5	1052						
17	F	51.1	1347						
18	F	41.2	1204						
19	M	51.9	1867						
20	M	46.9	1439						
21									
22									
23									

ScatterplotData

Ready   Sum=236.5
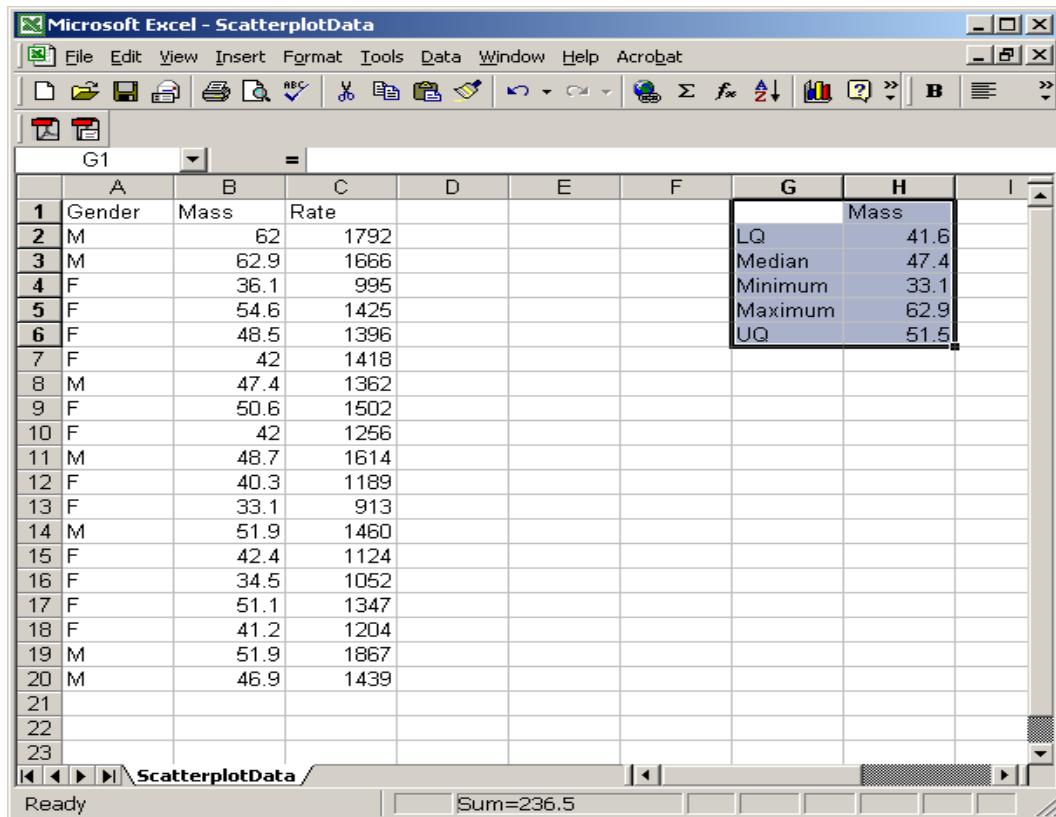
4. Then start the chart wizard and select the line type chart. I recommend that you also select the line chart with marker points. This will make it a little bit easier later on. Move to the next step by clicking next.
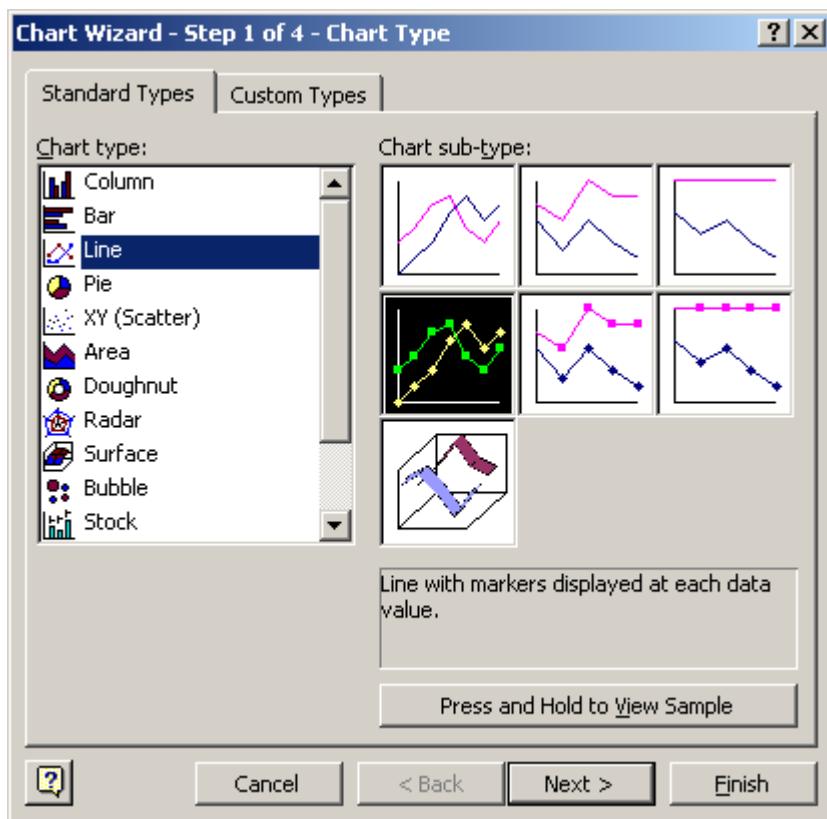
5. On step 2 of the chart wizard make sure that you click "Series in rows". You could now click finish or you could click next move to the next step where you could set the title of the plot and then click finish.
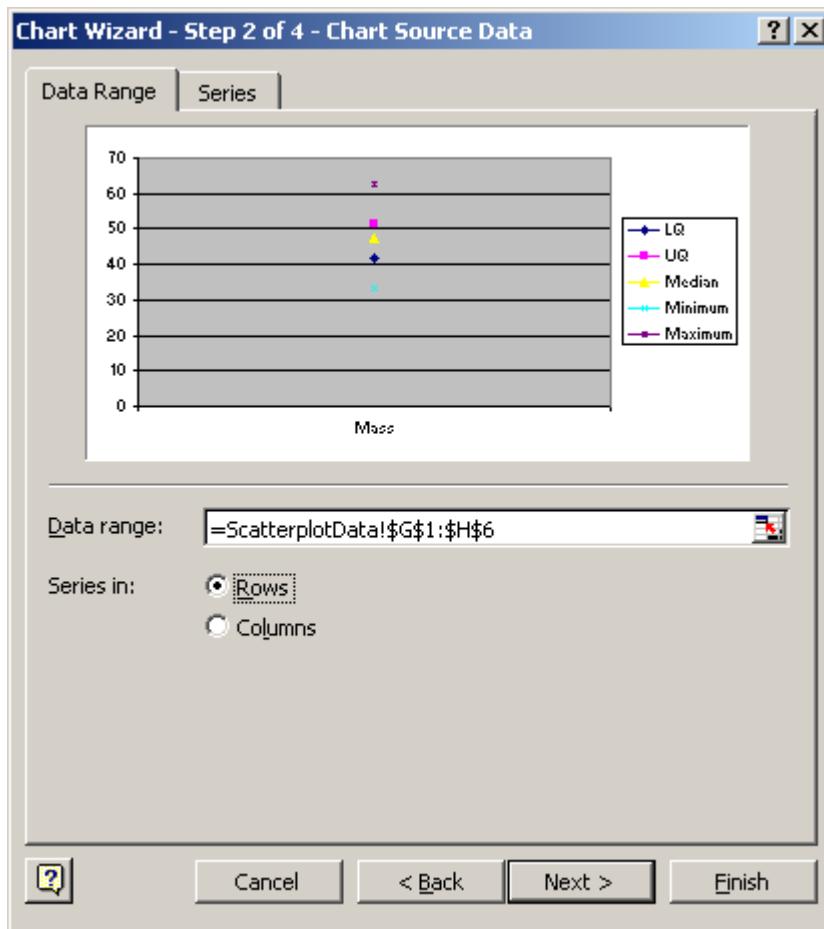

6. Now you should have a plot that looks something like the following
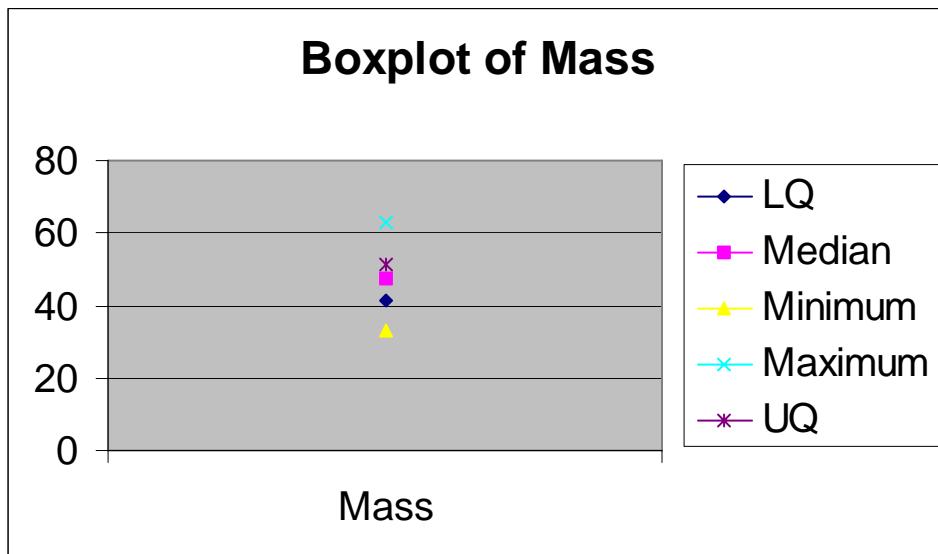
7. The next step is a little tricky. You need to double click on one of the points. I recommend trying for the median. Once you double click on the data point the *Format Data Series* dialog opens. Click on the *options* tab. Then make sure that *High-low lines* and *Up-down bars* get checked. Finally click ok.
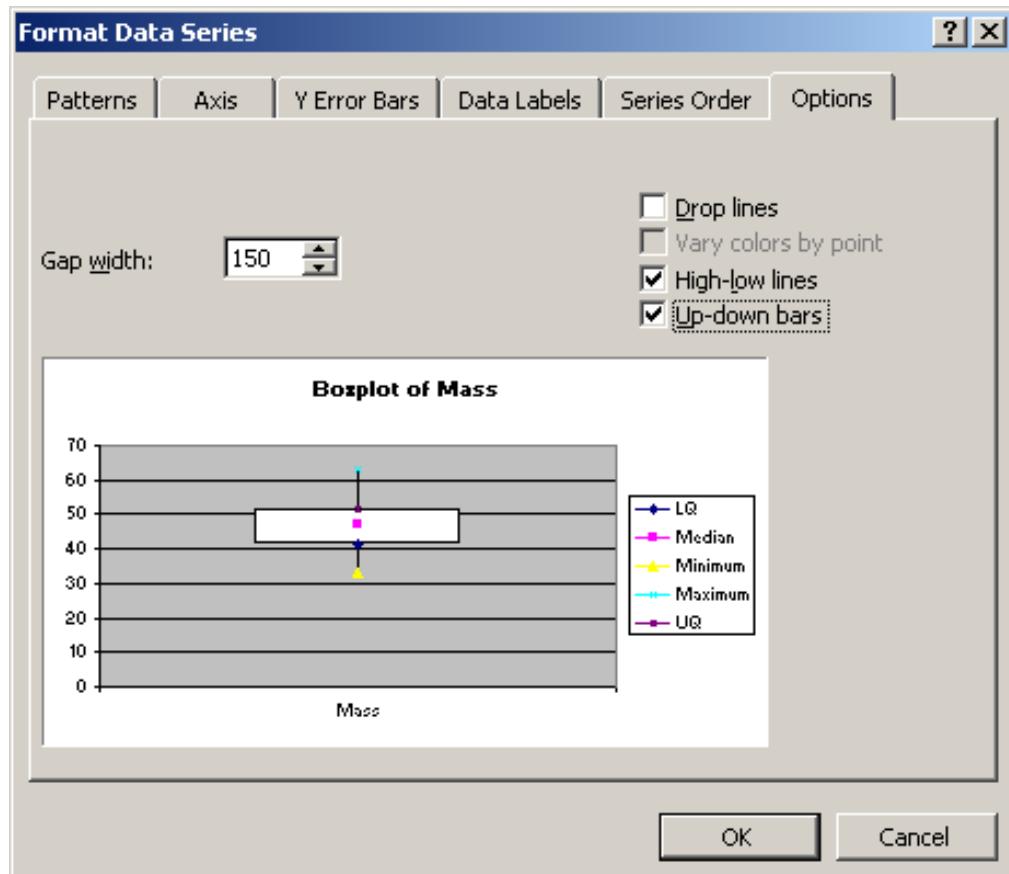
8. Your boxplot is now done. It should look something like the following.

Boxplot of Mass
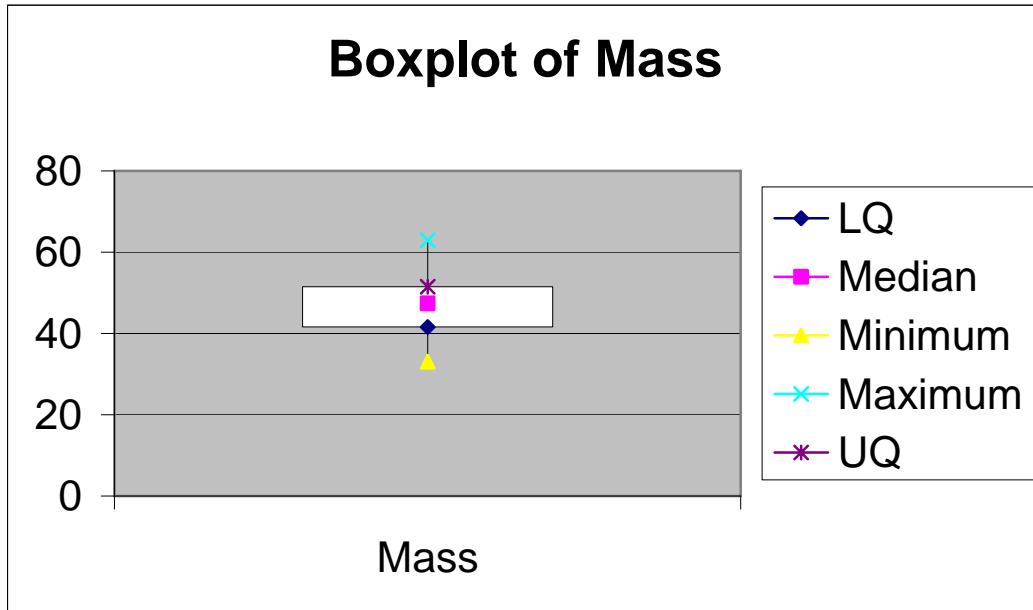
Making boxplots of a variable by levels of a categorical variable

1. In class we talked about using different boxplots based on different levels of a categorical variable. Excel allows us to do this, but it is a little bit harder than the simple plot we did before.

2. The first step is to sort the data so that all the female measurements are together and the male measurements are together. I recommend that you copy and paste the data columns into another location on the spreadsheet, so that the original data is not affected. To do this select the three data columns and select copy from the edit menu.
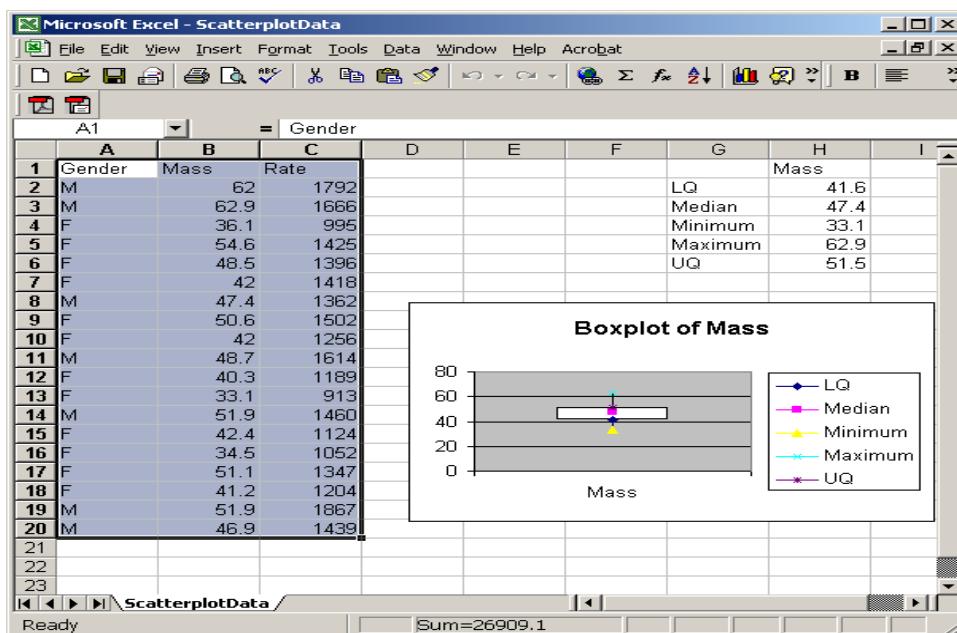
3. Now move over to an empty part of the spreadsheet (say column K). Click on cell K1 and then choose paste from the edit menu. Your spreadsheet should now look something like this
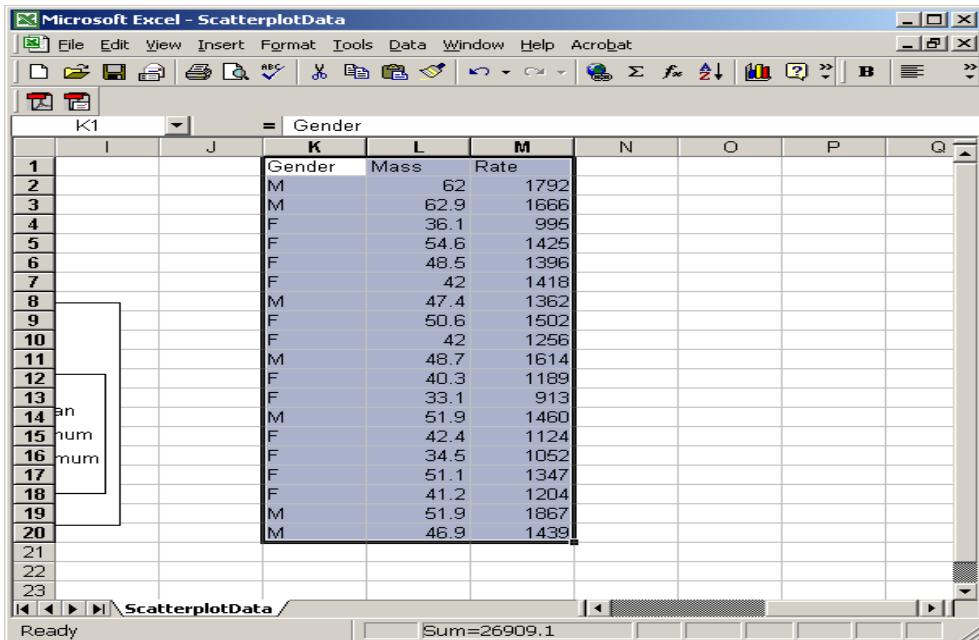
4. Now we will sort the data. Select the data in the new columns (ie K, L, M) and then select Sort from the Data menu. In this sort dialog box make sure that *Gender* is in the Sort by drop down menu. Also make sure that the *My list has header row* check box is checked. To finish this dialog box click ok.

5. The data should now be sorted by gender. Your spreadsheet window should look like this:

	Gender	Mass	Rate
1	Gender	Mass	Rate
2	F	36.1	995
3	F	54.6	1425
4	F	48.5	1396
5	F	42	1418
6	F	50.6	1502
7	F	42	1256
8	F	40.3	1189
9	F	33.1	913
10	F	42.4	1124
11	F	34.5	1052
12	F	51.1	1347
13	F	41.2	1204
14	M	62	1792
15	M	62.9	1666
16	M	47.4	1362
17	M	48.7	1614
18	M	51.9	1460
19	M	51.9	1867
20	M	46.9	1439

6. The next stage is to create yourself a table of LQ, Median, Minimum, Maximum and UQ values broken down by Gender. Set yourself up a table that looks something like this:

	Gender	Mass	Rate			Women	Men
1	Gender	Mass	Rate			Women	Men
2	F	36.1	995		LQ		
3	F	54.6	1425		Median		
4	F	48.5	1396		Minimum		
5	F	42	1418		Maximum		
6	F	50.6	1502		UQ		
7	F	42	1256				
8	F	40.3	1189				
9	F	33.1	913				
10	F	42.4	1124				
11	F	34.5	1052				
12	F	51.1	1347				
13	F	41.2	1204				
14	M	62	1792				
15	M	62.9	1666				
16	M	47.4	1362				
17	M	48.7	1614				
18	M	51.9	1460				
19	M	51.9	1867				
20	M	46.9	1439				

7. And of course you need to enter the formulas for computing all these quantities:

	Formula for Women	Formula for Men
LQ	=QUARTILE(L2:L13,1)	=QUARTILE(L14:L20,1)
Median	=MEDIAN(L2:L13)	=MEDIAN(L14:L20)
Minimum	=MIN(L2:L13)	=MIN(L14:L20)
Maximum	=MAX(L2:L13)	=MAX(L14:L20)
UQ	=QUARTILE(L2:L13,3)	=QUARTILE(L14:L20,3)

8. Next select the entire table like this:
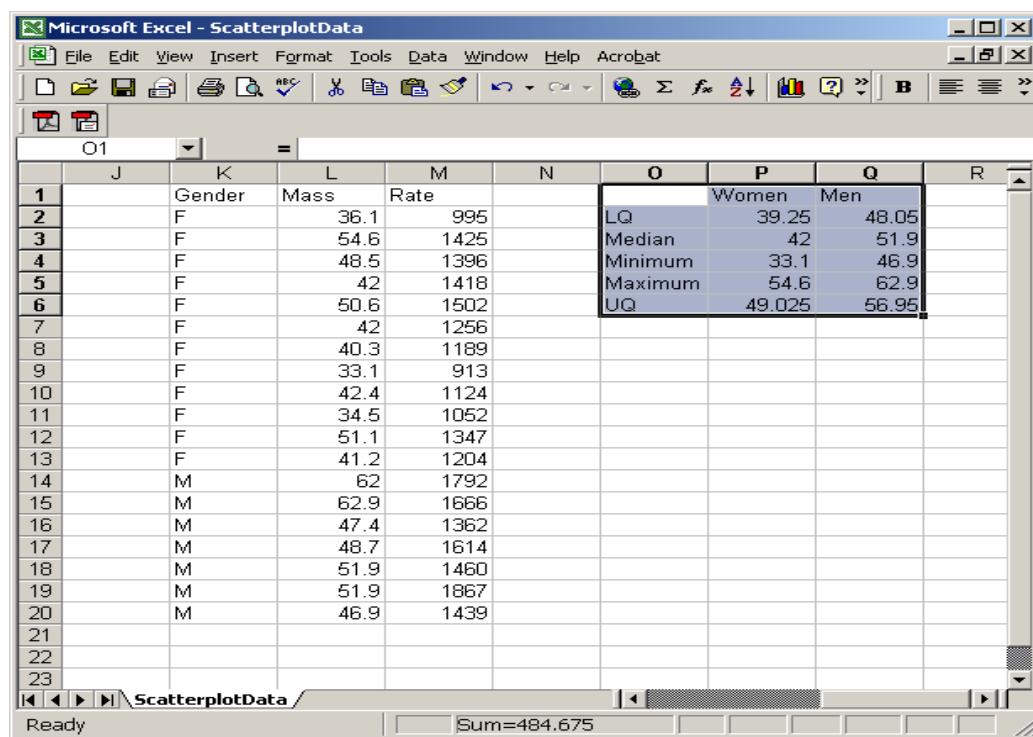
9. Start the chart wizard. Select the line chart type as before then click next
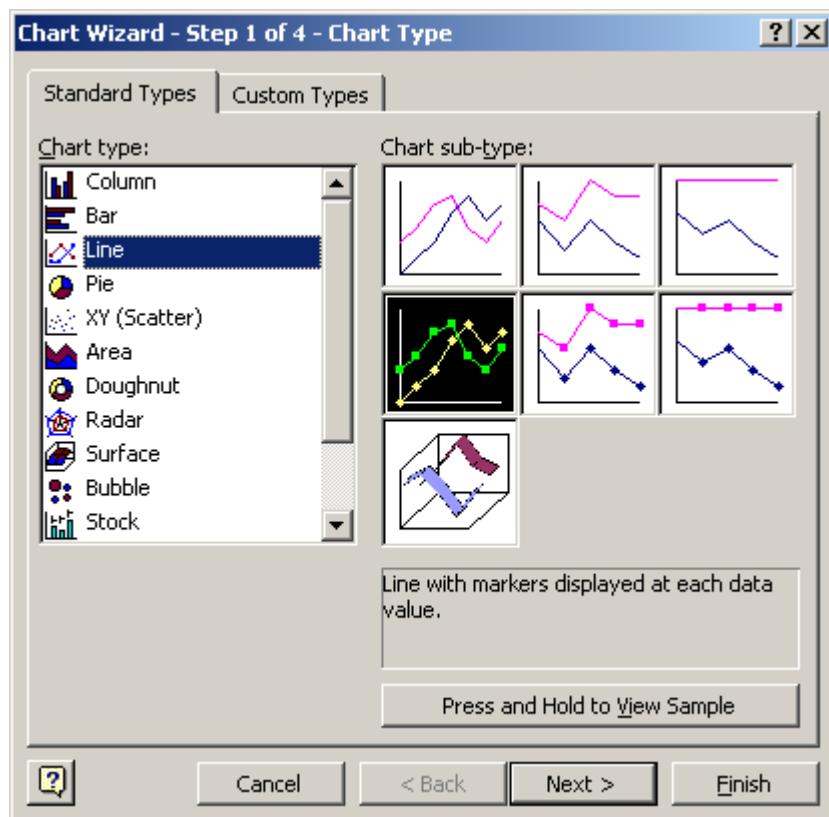
10. As before make sure that select Series in rows. Click finish to draw the chart.
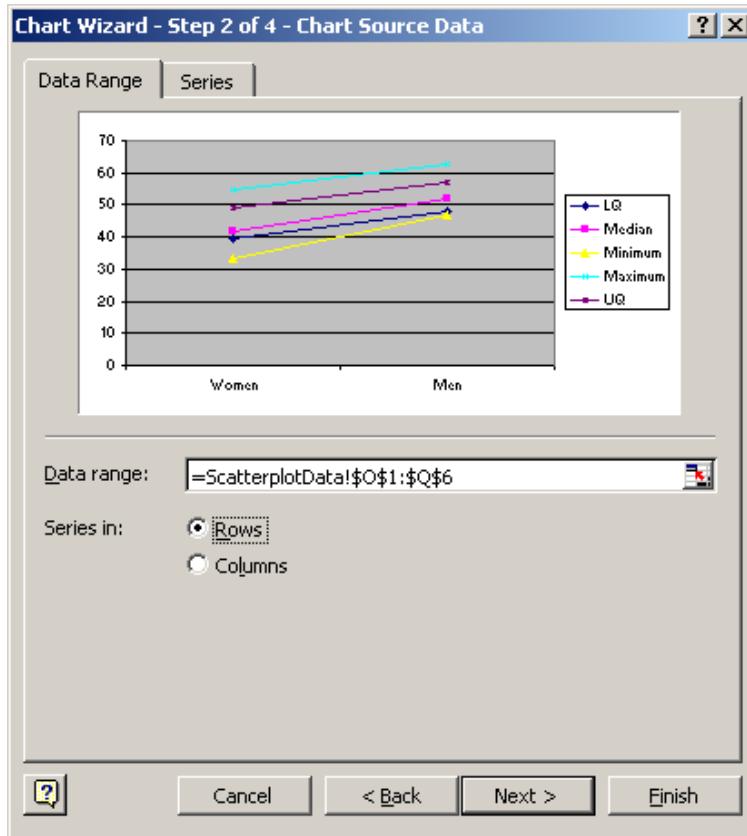


11. Your chart should look something like this:
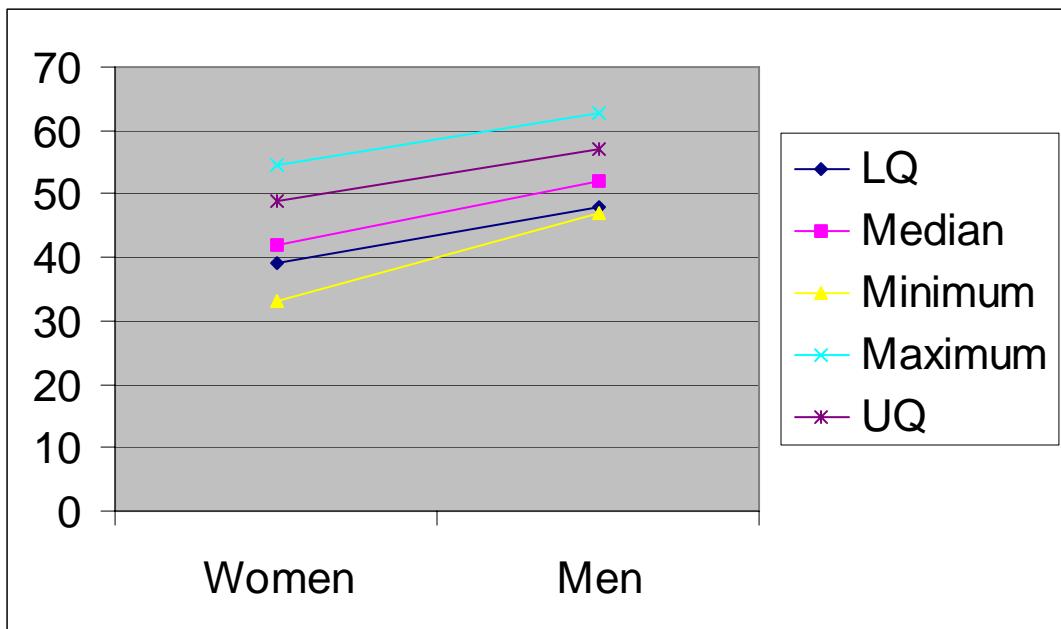
12. The next stage is to remove the connecting lines. To do this you need to double click on each line individually bringing up the Format Data Series dialog. Make sure that you click None in the Line radiobox group. The click ok. The line should have disappeared. You should repeat this process another four times to get rid of the remaining lines.
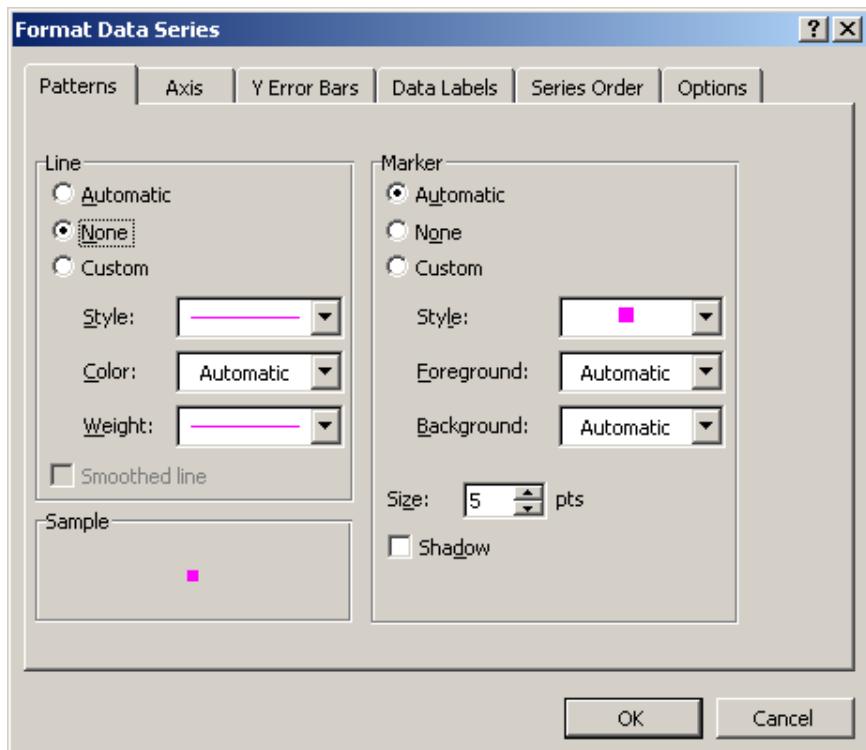
13. Once you've finished removing all the lines your plot should look like this:
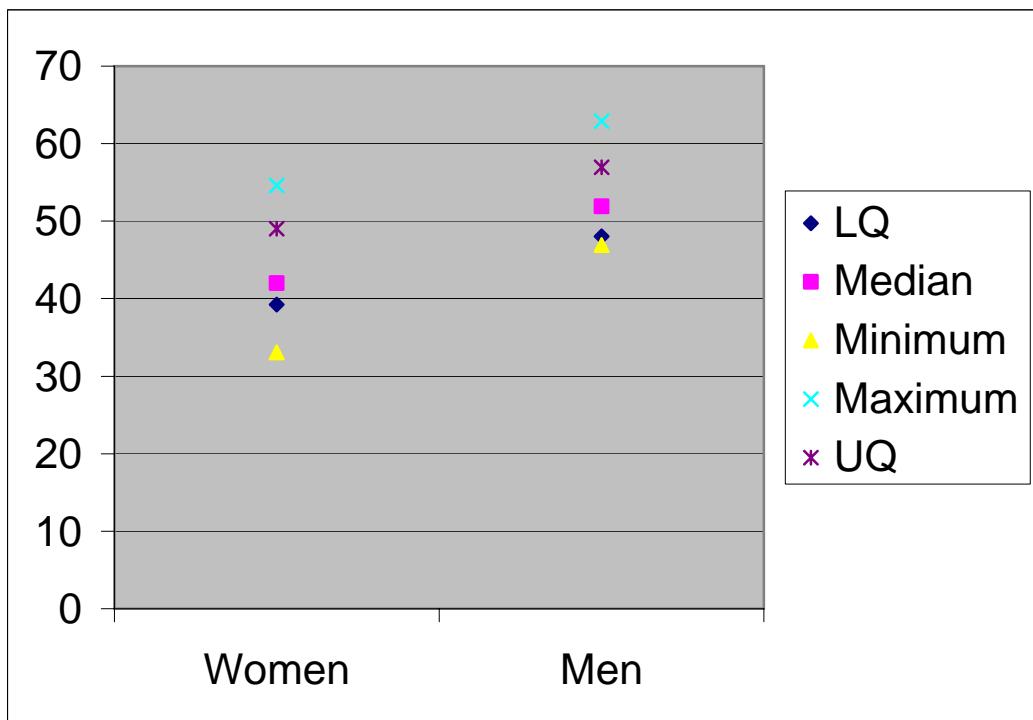
14. Now double click on one of the points (I recommend the median) to again bring up the format data series dialog box. This time go to the Options tab and then make sure that *High-low lines* and *Up-down bars* get checked. Finally click ok.
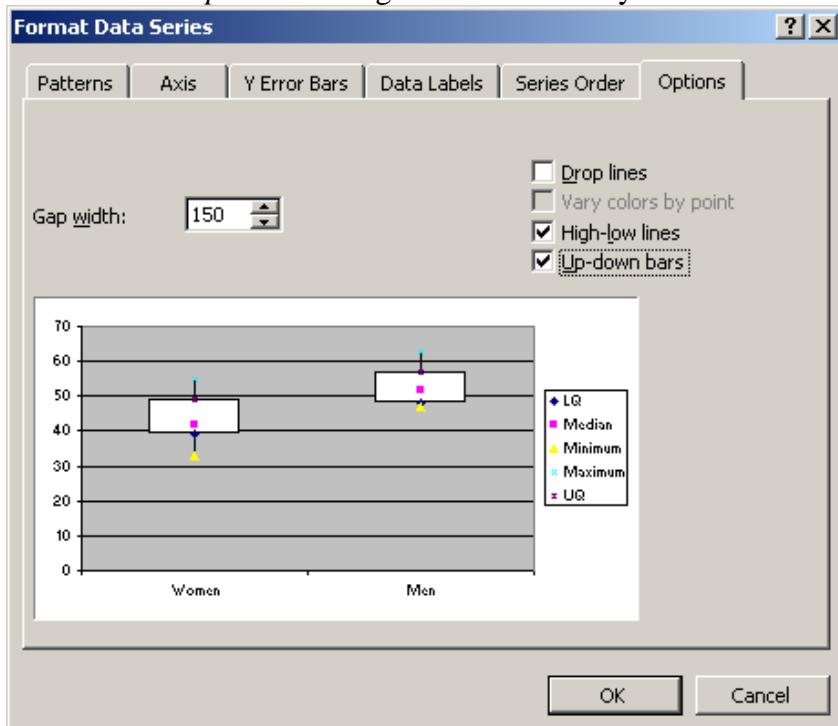
15. Congratulations you are now done. Your boxplots should look something like this: