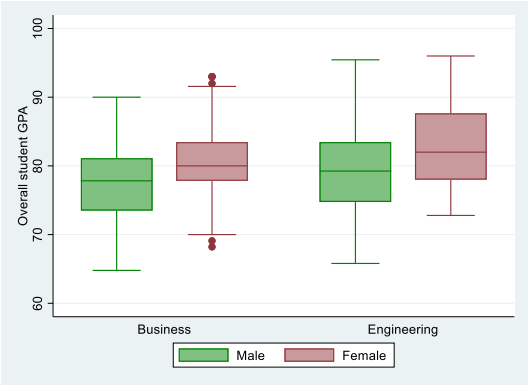
Visualizing data in Stata – Section 5.4
Najib Mozahem

The previous two lectures covered some of the most used options that are common between box plot, bar graphs, and dot plots. In the following three lectures, we will go over some specific options for each of the three types of graphs, starting with box plots. To change the color of the boxes, we can use the **box()** option together with many of the sub options that come with it:

graph box gpa, asyvar over(gender) over(college) box(1, bcolor(green))

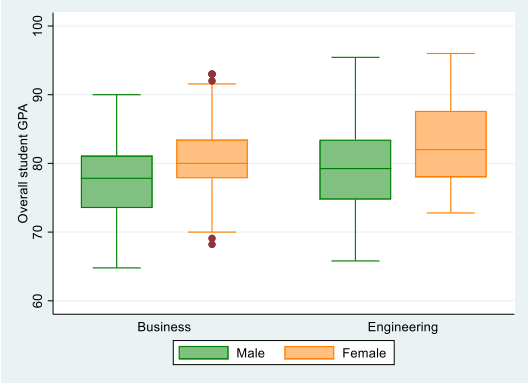


The option **bcolor()** specifies the color for the box. To see a list of the colors that you can specify, execute the following command:

graph query colorstyle

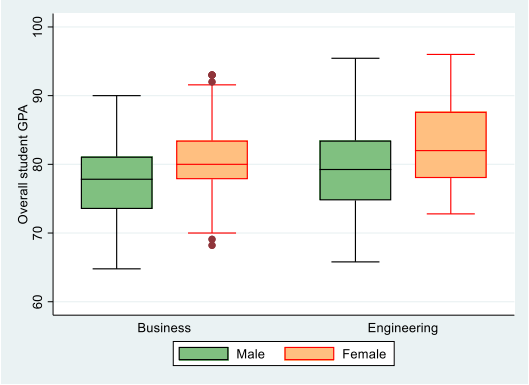
If we want to change the color of the female boxes, we would execute:

graph box gpa, asyvar over(gender) over(college) box(1, bcolor(green)) box(2, bcolor(orange))



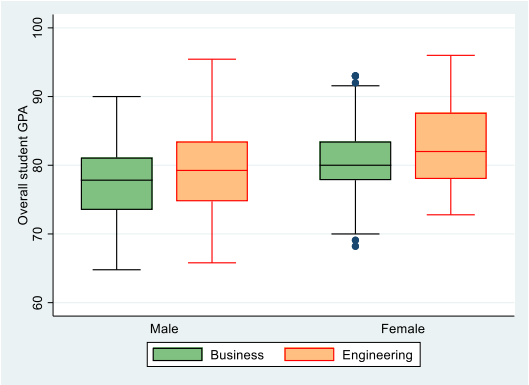
We can also change the color of the border using the **blcolor()** option:

graph box gpa, asyvar over(gender) over(college) box(1, bcolor(green) blcolor(black)) box(2, bcolor(orange) blcolor(red))



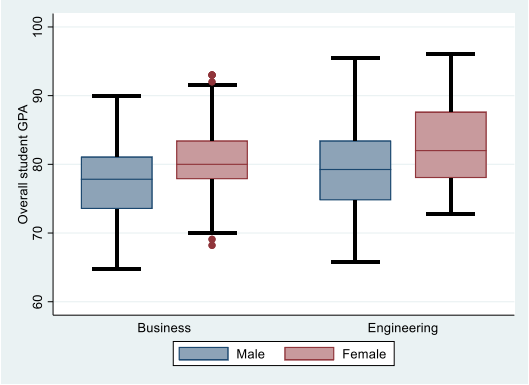
Here, we told Stata to use black as the borderline color of the first box and red as the borderline color of the second box. Notice that the numbers used to identify the boxes depend on the order in which we specify the over() options. In the above commands, the over(gender) option came before the over(college) option. Therefore, box number 1 refers to the first gender category, which is male in our case, and box number 2 refers to the second gender category. If we specify over(college) before over(gender), box number 1 will refer to the first college category, which is business:

graph box gpa, asyvar over(college) over(gender) box(1, bcolor(green) blcolor(black)) box(2, bcolor(orange) blcolor(red))



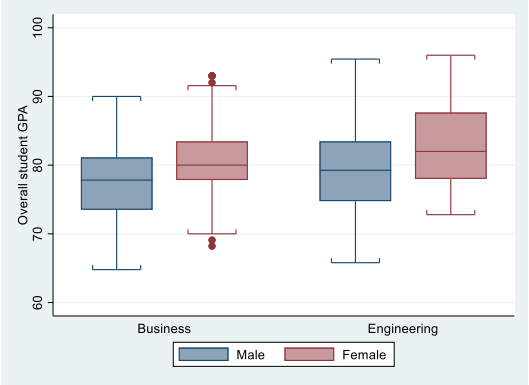
The above options change the look of the boxes. It is also possible to change the look of the whiskers:

graph box gpa, asyvar over(gender) over(college) cwhiskers lines(lcolor(black) lwidth(thick))



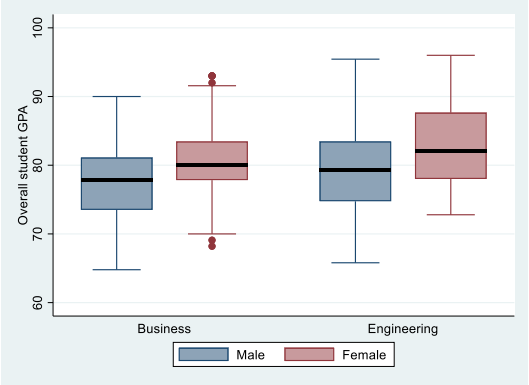
This option is slightly different from previous ones because it involves specifying two options, the first of which is **cwhiskers** followed by the option **lines()**. In the above command, we specify that the whiskers should be black and thick. We can also modify the look of the horizontal lines that represent the 0.25 and the 0.75 quantiles by using the **capsize()** option:

graph box gpa, asyvar over(gender) over(college) capsize(4)



We can also change the look of the median line, which is represented by the horizontal line in the middle of the boxes using the **medtype()** option:

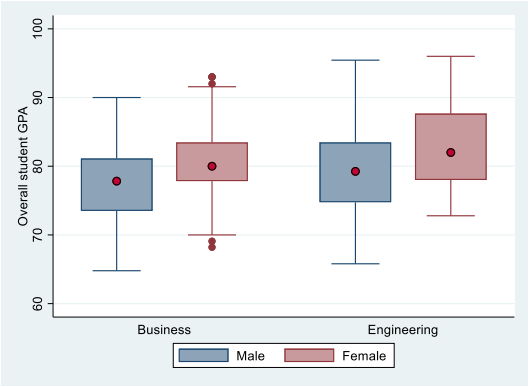
graph box gpa, asyvar over(gender) over(college) medtype(cline) medline(lcolor(black) lwidth(thick))



Here, we told Stata that the median type is a **cline**, which means that it will be a colored line. Since we specified **cline** as the **medtype**, we use the **medline()** option to customize the look of the line.

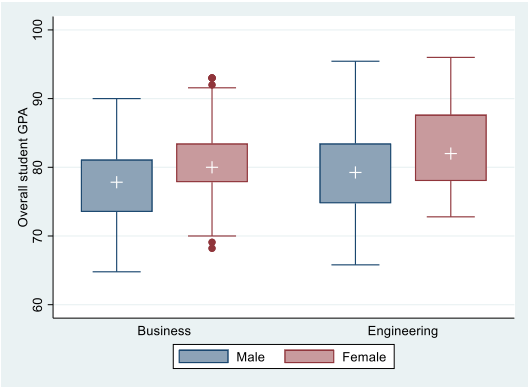
We can also use a marker instead of a line to draw the median. This is accomplished by specifying **marker** as the **medtype()**:

graph box gpa, asyvar over(gender) over(college) medtype(marker)



The default shape used is a circle, but we can customize the look of the marker just as we did when the market was a line, but this time we do it using the **medmarker()** option:

graph box gpa, asyvar over(gender) over(college) medtype(marker)
medmarker(msymbol(+) mcolor(white) msize(vlarge))

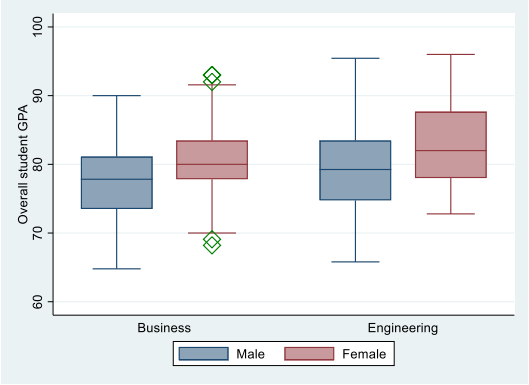


In this command we told Stata that the symbol to be used is a "+" sign, and that its color should be white and that the size should be very large. To see a list of the symbols that you can use just execute the command:

help symbolstyle

Finally, we can also change the shape of the markers that specify the outside values by using the **marker()** option:

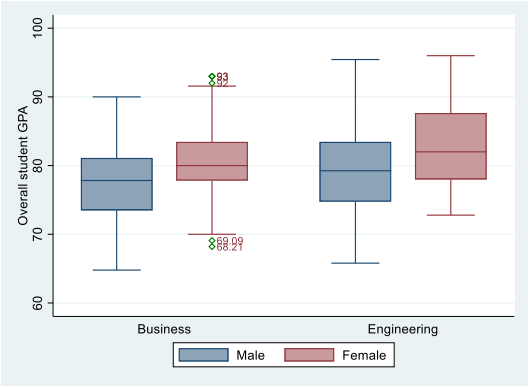
graph box gpa, asyvar over(gender) over(college) marker(2, msymbol(Dh)
mcolor(green) msize(vlarge))



To use this option, we need to specify which groups of markers are we modifying. This is because it is possible to assign different signs to different groups. In our case, there are outlier values only for females, which is the second group, so in the **marker()** option, we need to specify the number 2 first. We then specify that the markets should be shaped like hollow diamonds and that they should be green and very large.

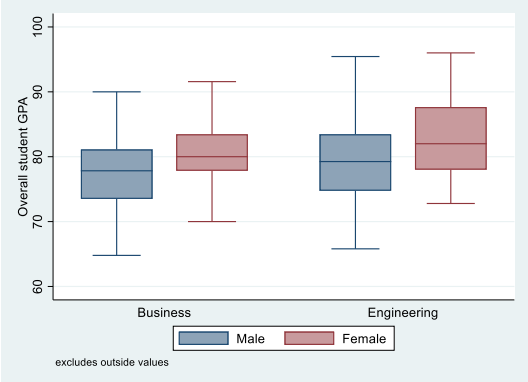
We can also use the **marker()** option to label the outliers:

graph box gpa, asyvar over(gender) over(college) marker(2, msymbol(Dh) mcolor(green) msize(small) mlabel(gpa))



Here we told Stata to label the outliers by the value of the variable *gpa*. In case we don't want the outliers to show on the graph, we can use the **nooutsides** option:

graph box gpa, asyvar over(gender) over(college) nooutsides


Overall student GPA
100
90
80
70
60
Business
Engineering
Male
Female
excludes outside values