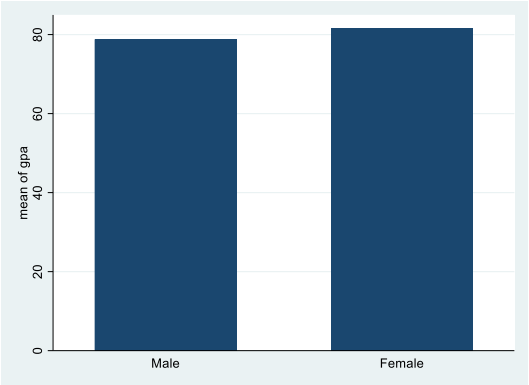
Visualizing data in Stata – Section 5.2
Najib Mozahem

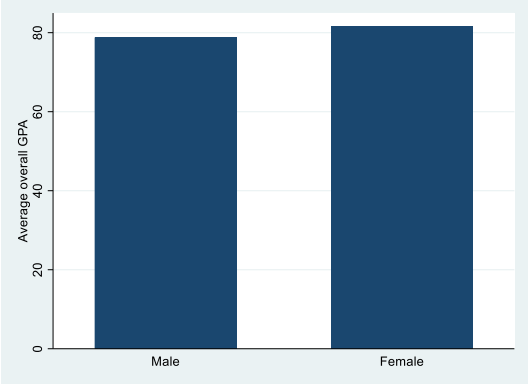
We already know that to plot the variable *gpa* for each gender we use the following command:

graph bar gpa, over(gender)



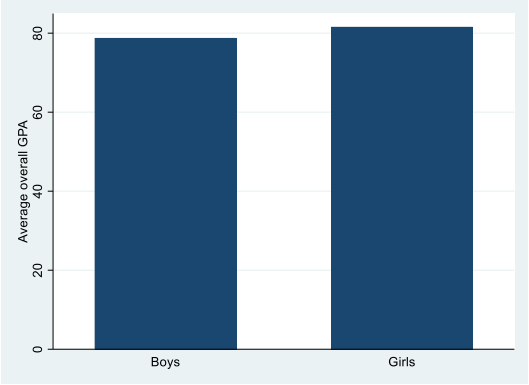
One issue with the above graph is that the label of the vertical access is not what we wish it to be. I would rather the vertical axis be labelled as "Average overall GPA". To do that we can use the **ytitle()** option:

graph bar gpa, over(gender) ytitle("Average overall GPA")



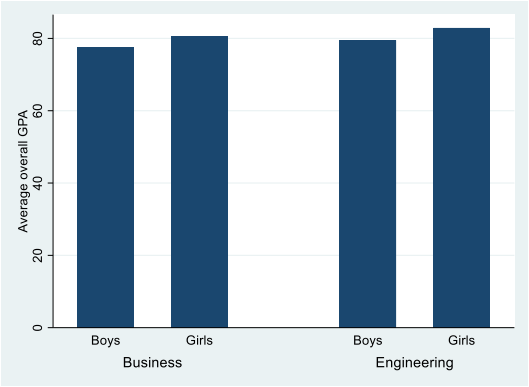
It's also possible to control the labels used for the categories. Instead of "Male" and "Female", I might wish to use the labels "Boys" and "Girls". To do that, I need to use the **relabel()** option, which is a sub option of the **over()** option:

graph bar gpa, over(gender, relabel(1 "Boys" 2 "Girls")) ytitle("Average overall GPA")



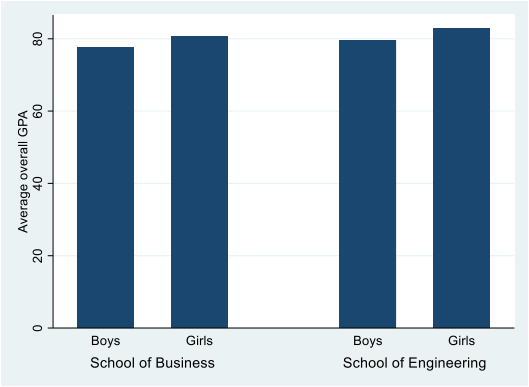
Here, we are telling Stata that what we want is to label the first category as “Boys” and the second as “Girls”. This would also work if we had specified the **over()** option more than once:

graph bar gpa, over(gender, relabel(1 "Boys" 2 "Girls")) over(college) ytitle("Average overall GPA")



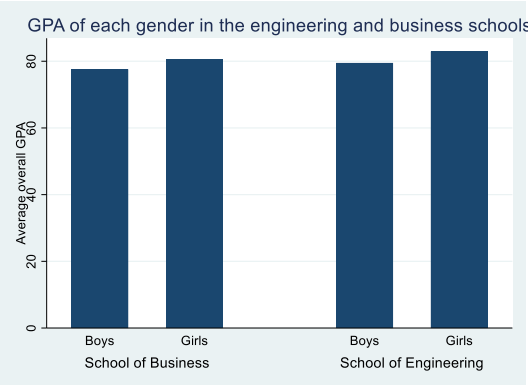
We can also do the same for the variable college:

graph bar gpa, over(gender, relabel(1 "Boys" 2 "Girls")) over(college, relabel(1 "School of Business" 2 "School of Engineering")) ytitle("Average overall GPA")



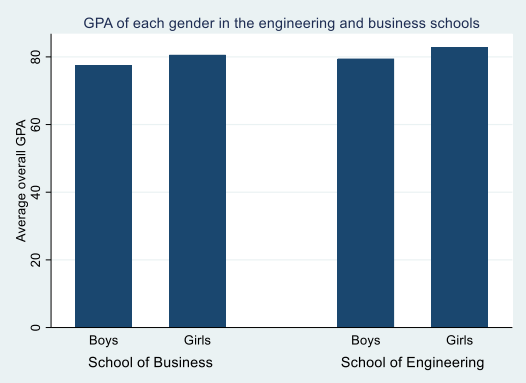
It is also possible to give the graph a title. This is done by specifying the **title()** option:

graph bar gpa, over(gender, relabel(1 "Boys" 2 "Girls")) over(college, relabel(1 "School of Business" 2 "School of Engineering")) ytitle("Average overall GPA") title("GPA of each gender in the engineering and business schools")



However, something is wrong with the title. We notice that it is too big. Here we have two options. We can either decrease the size of the font of the title, or we can tell Stata to write the title on two lines. To decrease the size of the title we can do the following:

graph bar gpa, over(gender, relabel(1 "Boys" 2 "Girls")) over(college, relabel(1 "School of Business" 2 "School of Engineering")) ytitle("Average overall GPA") title("GPA of each gender in the engineering and business schools", size(medium))



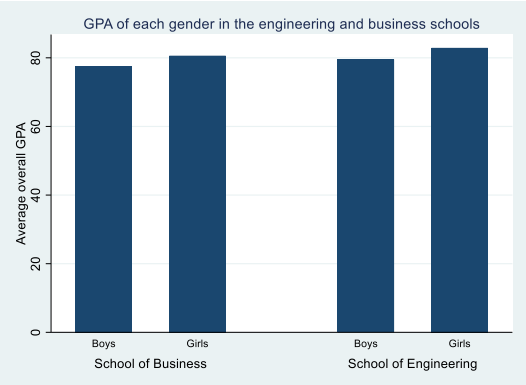
We used the option **size()** as a sub option of the option **title()**. In that option, we specified that we want the size of the font used for the title to be "medium". To see a list of font sizes, you can execute the following command:

help textsizestyle

As you can see, there are many sizes to choose from. We might also wish to change the size of the fonts used in the different categories. This is accomplished by using the **label(labsize())** option inside the **over()** option:

graph bar gpa, over(gender, relabel(1 "Boys" 2 "Girls") label(labsize(small))) over(college, relabel(1 "School of Business" 2 "School of Engineering")

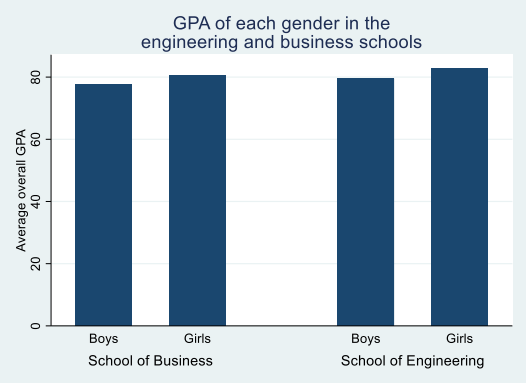
label(labsize(medsmall))) ytitle("Average overall GPA") title("GPA of each gender in the engineering and business schools", size(medium))



In this command, we specified that we wanted the size of the labels for tge gender categories to be **small**, while we wanted the size of the labels in the college category to be **medsmall**.

The other thing that we can do is to tell Stata to keep the default font size but instead to split the title by using more than one line. This is accomplished as such:

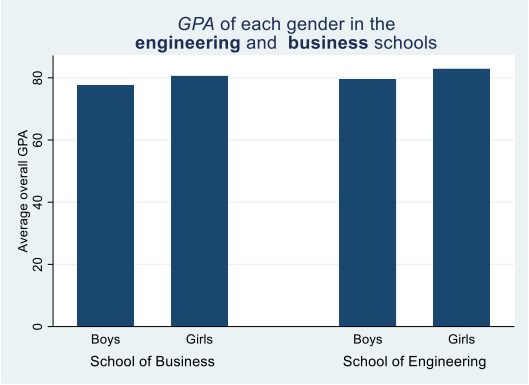
graph bar gpa, over(gender, relabel(1 "Boys" 2 "Girls")) over(college, relabel(1 "School of Business" 2 "School of Engineering")) ytitle("Average overall GPA") title("GPA of each gender in the" "engineering and business schools")



By including different parts of the title in different quotations, we are telling Stata to write on the first line “GPA of each gender in the” and to write on the second line “engineering and business schools”.

We can also use the italics and bold options as such:

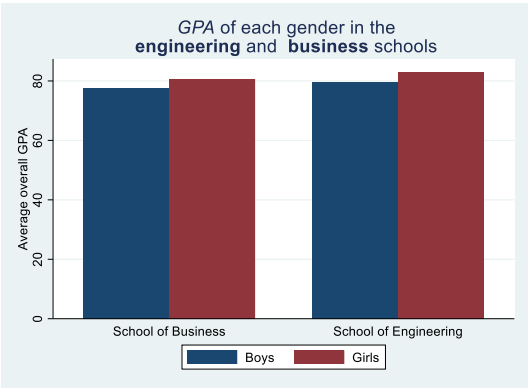
graph bar gpa, over(gender, relabel(1 "Boys" 2 "Girls")) over(college, relabel(1 "School of Business" 2 "School of Engineering")) ytitle("Average overall GPA") title("{it: GPA} of each gender in the" "{bf: engineering} and {bf: business} schools")



In the above command, we italicized the word "GPA" and we used the bold font with the words "engineering" and "business".

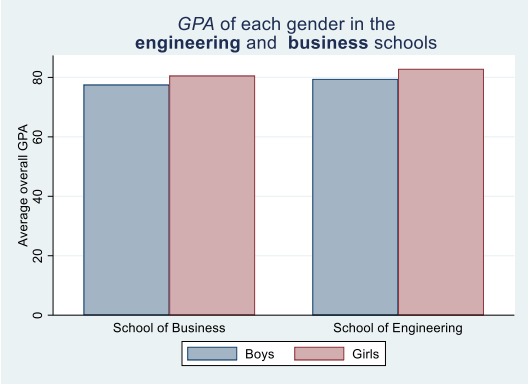
All of the above works the exact same way if we used the **asyvar** option in the **graph bar** command:

graph bar gpa, asyvar over(gender, relabel(1 "Boys" 2 "Girls")) over(college, relabel(1 "School of Business" 2 "School of Engineering")) ytitle("Average overall GPA") title("{it: GPA} of each gender in the" "{bf: engineering} and {bf: business} schools")



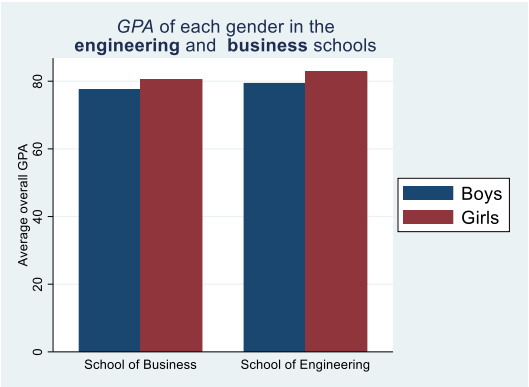
Notice that the only things that change are the colors and the fact that there is a legend now. Perhaps we decide that the colors are too bright for us. We can use the **intensity()** option in order modify the intensity of the colors with which the bars are drawn:

graph bar gpa, asyvar over(gender, relabel(1 "Boys" 2 "Girls")) over(college, relabel(1 "School of Business" 2 "School of Engineering")) ytitle("Average overall GPA") title("{it: GPA} of each gender in the" "{bf: engineering} and {bf: business} schools") intensity(40)



We can also change the position, size, and shape of the legend. All three can be accomplished by including the **legend()** option. Inside this option, we can specify whatever sub option that we need. To change the position, we can use the **position()** sub option. To change the size, we use the **size()** sub option. Finally, to change the shape, we can use the **cols()** sub option to tell Stata the number of columns included in the legend:

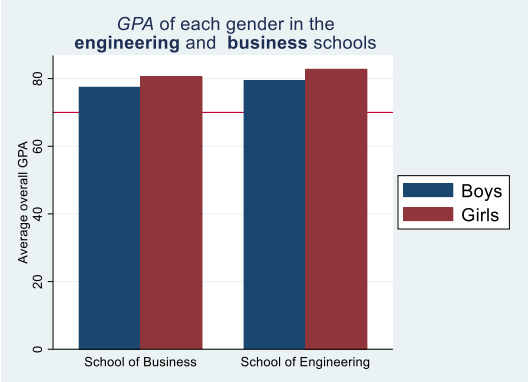
graph bar gpa, asyvar over(gender, relabel(1 "Boys" 2 "Girls")) over(college, relabel(1 "School of Business" 2 "School of Engineering")) ytitle("Average overall GPA") title("{it: GPA} of each gender in the" "{bf: engineering} and {bf: business} schools") legend(position(3) cols(1) size(large))



The **position(3)** option tells Stata to place the legend in the three o'clock position. The **size(large)** option tells Stata to make the size of the legend large. Finally, the **cols(1)** option tells Stata to list the items inside the legend in one column.

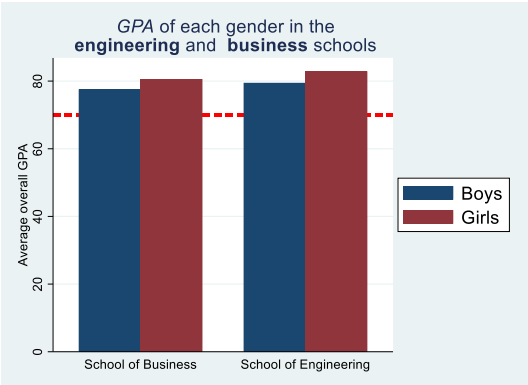
We can also choose to draw a line to illustrate a point. For example, the passing GPA in some universities in 70%. Therefore, I might choose to draw a line at the value 70 for the variable *gpa*. This is accomplished using the **yline()** option:

graph bar gpa, asyvar over(gender, relabel(1 "Boys" 2 "Girls")) over(college, relabel(1 "School of Business" 2 "School of Engineering")) ytitle("Average overall GPA") title("{it: GPA} of each gender in the" "{bf: engineering} and {bf: business} schools") legend(position(3) cols(1) size(large)) yline(70)



Here, we are telling Stata to draw a horizontal line that represents *gpa* = 70. We can also control how the line looks:

graph bar gpa, asyvar over(gender, relabel(1 "Boys" 2 "Girls")) over(college, relabel(1 "School of Business" 2 "School of Engineering")) ytitle("Average overall GPA") title("{it: GPA} of each gender in the" "{bf: engineering} and {bf: business} schools") legend(position(3) cols(1) size(large)) yline(70, lwidth(thick) lcolor(red) lpattern(dash))



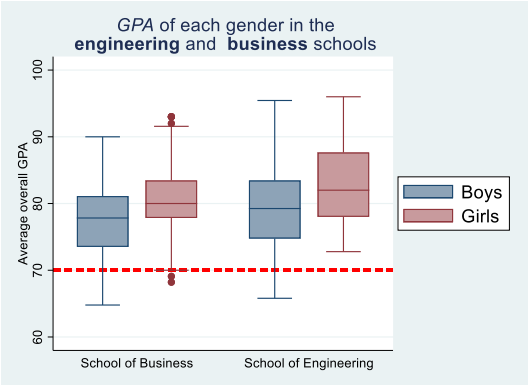
In the above command, we used several sub options of the **yline()** option to change the thickness of the line, the color of the line, and the pattern used to draw the line. As you can see, there are a lot of options. Thankfully, Stata makes it easy to access these options. For example, execute the following command:

help graph bar

We now see the help file for the graph bar command. Assume that I wish to see what types of patterns I can use to draw the horizontal line. We used the **yline()** option to draw the line, and the **lpattern()** option is a subset of that option, so we go to the section that gives us information about the **yline()** option. If we click on the blue link for the **yline()** option, we are taken to the help file of that option. Looking ta the sub options section of this option, we see that the list includes the **lpattern()** option. If we click on the **linepatternstyle** link, we are taken to the help file that lists all the patterns that we can use. All you have to do is to pick one of these and include it in the command.

Every single option used so far works the exact same way for the dot plots and for the box plots:

graph box gpa, asyvar over(gender, relabel(1 "Boys" 2 "Girls")) over(college, relabel(1 "School of Business" 2 "School of Engineering")) ytitle("Average overall GPA") title("{it: GPA} of each gender in the" "{bf: engineering} and {bf: business} schools") legend(position(3) cols(1) size(large)) yline(70, lwidth(thick) lcolor(red) lpattern(dash))



graph dot gpa, asyvar over(gender, relabel(1 "Boys" 2 "Girls")) over(college, relabel(1 "School of Business" 2 "School of Engineering")) ytitle("Average overall GPA") title("{it: GPA} of each gender in the" "{bf: engineering} and {bf: business} schools") legend(position(3) cols(1) size(large)) yline(70, lwidth(thick) lcolor(red) lpattern(dash))