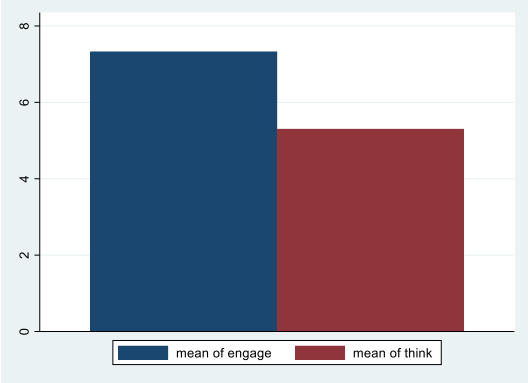
Visualizing data in Stata – Section 7.2
Najib Mozahem

In a previous section of this course, we discussed the use of box plots, bar graphs, and dot plots in studying single variables. We also used these tools to study how a single variable varies between groups. However, this is not the end of the story. These exact same tools can be used to study two or more variables at the same time. All we have to do is to include the extra variables in our command.

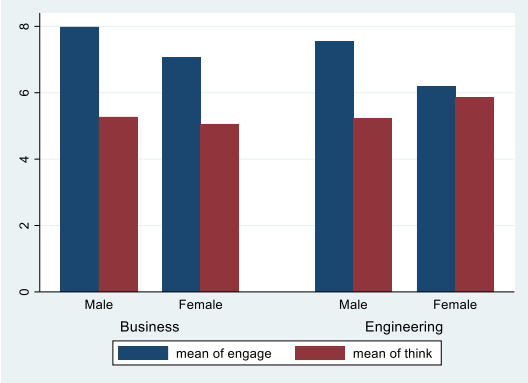
We have previously commented on the fact that the value of *engage* is higher than the value of *think*. We can produce a bar chart to compare the means:

graph bar engage think



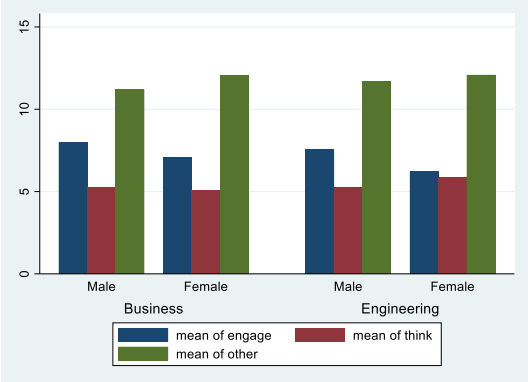
Notice that in this example, all we did was to add a new variable to the command. That is all it takes. Even when we want to study group differences, nothing changes:

graph bar engage think, over(gender) over(college)



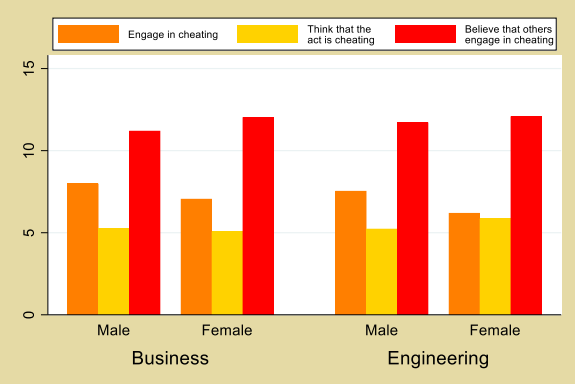
We note that the value of *engage* is lower than that of *think* in every single group. We can even include three variables:

graph bar engage think other, over(gender) over(college)



What we see is that in every single group, students always believe that other students are cheating more than they report that they themselves cheat. We can use any of the options that we previously used in order to change the look of the graph:

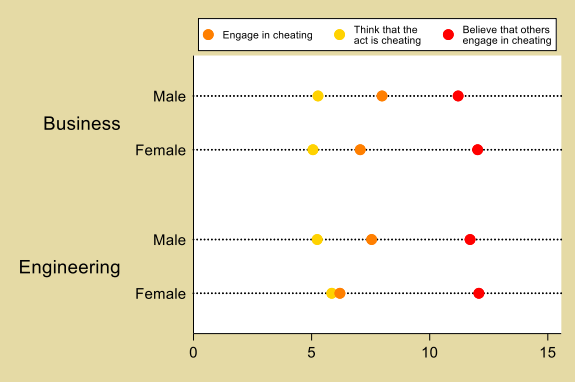
graph bar engage think other, over(gender) over(college) scheme(vg_brite) legend(position(12) cols(3) size(small) label(1 "Engage in cheating") label(2 "Think that the" "act is cheating") label(3 "Believe that others" "engage in cheating"))



Here we specified a certain scheme to change the look of the graph. We also used the **legend()** option to customize the legend. We renamed the texts in the legend to make them clearer for example. Note that we split the text between two sets of quotation marks in order to have the text of some labels displayed on two lines. We also set the location of the legend to the 12 o'clock location.

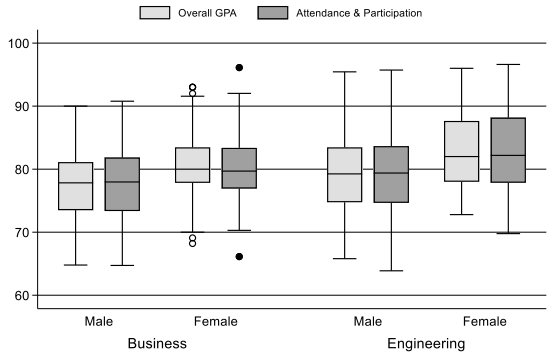
The exact same command can be executed to produce a dot plot:

graph dot engage think other, over(gender) over(college) scheme(vg_brite) legend(position(12) cols(3) size(small) label(1 "Engage in cheating") label(2 "Think that the" "act is cheating") label(3 "Believe that others" "engage in cheating"))



We can also look at more than one variable in the case of box plots:

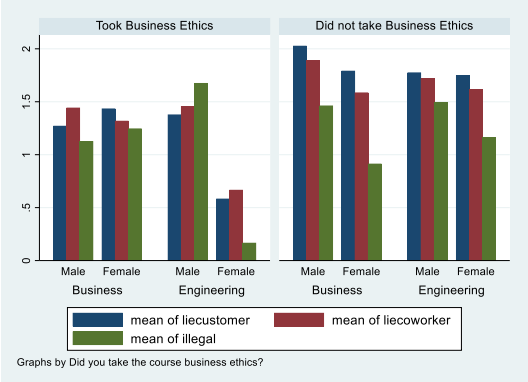
graph box gpa attendance , over(gender) over(college) scheme(lean2) legend(position(12) cols(3) size(small) label(1 "Overall GPA") label(2 "Attendance & Participation"))



As you can see, these command follow the exact same logic that we used before. By including more than one variable, we can see how the values of all variables vary. For example, in this graph, we see that females have a higher GPA and a higher attendance score than males in both colleges. We also see that female engineers rank on top with regards to both of these variables. The graph also shows that the within each group, the variables *gpa* and *attendance* have the same values. In all groups we see that the lines representing the medians are on the same level. The same can be said about the lines that represent the 0.25 and the 0.75 quantiles.

Another example can also illustrate the power of including more than one variable. Imagine that we now want to look at the variables *liecustomer*, *liecoworker*, and *illegal*. These variables measure the respondents belief about whether it is acceptable to lie to customers, coworkers, or even the government respectively. I would like to see which of these three actions are the respondents most willing to partake in, and whether there are group differences, so I use the following command:

graph bar liecustomer liecoworker illegal , over(gender) over(college) by(course)



In the command, we used the **by()** option to produce two graphs, one for those who took the course business ethics, and one for those who didn't take the course yet. We also group the results by the variables *gender* and *college*. What we see is that students who took the course business ethics are less likely to think that it is acceptable to lie to customers and colleagues, or to do something that is illegal. The only exception to this are females in the business school when it comes to the question about doing something that is illegal. We also see that In general, respondents are least willing to do something that is illegal. One interesting finding is that students who hadn't taken the course are most willing to lie to customers, while students who did take the course seem to be most willing to lie to colleagues. This means that the course seems to have an affect on students but mostly with regards to how they treat customers. Therefore, we might suspect here that the course covers material that deals with ethical situations regarding the customer but less so when it comes to the relationship between colleagues.