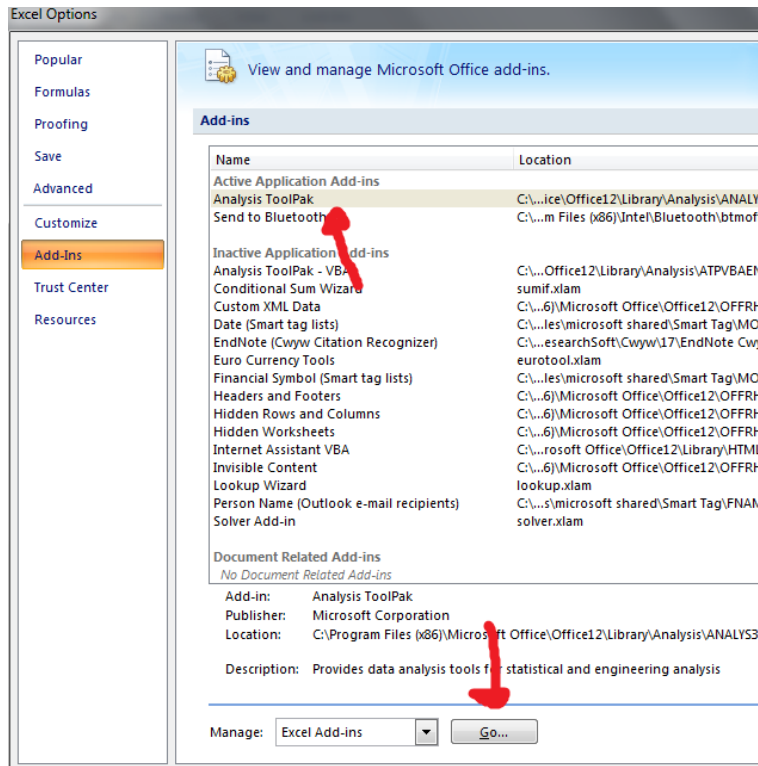ANALYZING DATA IN EXCEL:

Excel functions pretty similarly to Microsoft Word with how the font and such are formatted (and you can make tables in Excel, too), but Excel has analysis capabilities that Word doesn't.
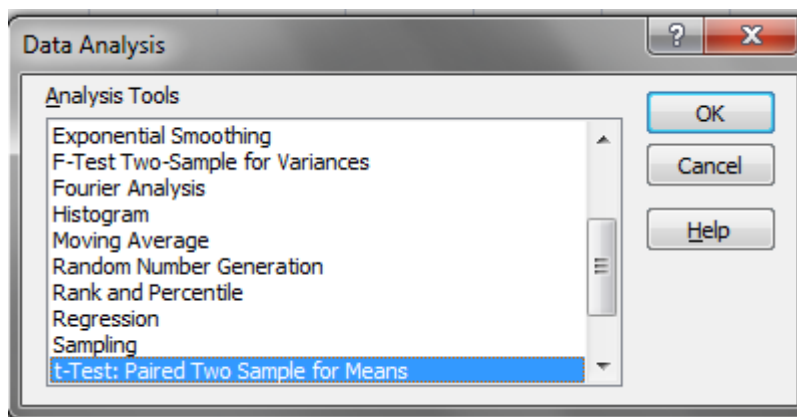
Preparing to analyze data:

One of the best ways to analyze data in Excel is to use the Analysis Toolpak. You can access this by clicking the Microsoft icon in the top left, then selecting Excel options. From that list, go to "Add-Ins." Have Analysis ToolPak selected and then click "Go" next to Manage Excel Add-ins:

Have "Analysis ToolPak" checked in this new menu and then hit okay. Now, under Data, there should be a new "Analysis" tab that had a "Data Analysis" function.
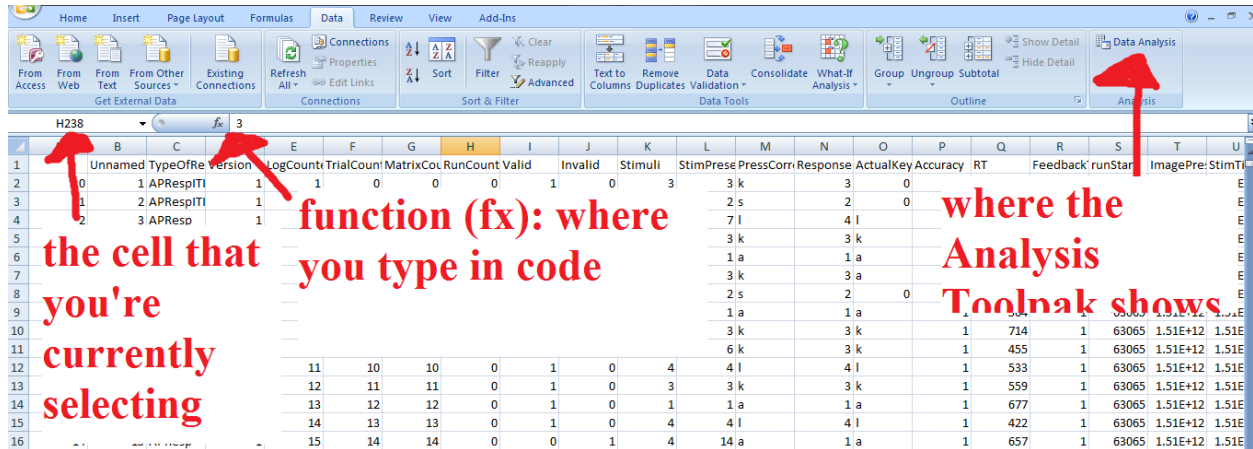
When you click this, you'll be presented with any number of options:

I typically have used Excel only for t-test comparisons, when I haven't felt like coding the data in another experiment. However, know that you can use Excel for other analyses.

Special codes in Excel:

Excel accepts **code** just like any other analysis program.

If you want to take the mean/average of a certain amount of cells, for instance, you'd type in = average(*insert cell numbers here*) and hit enter. It will execute the code and output the result in the cell that you were clicked inside before you typed in the code. Instead of typing the cell numbers to average, you can also just select a group of cells that you want to average, because of the way that Excel functions.

Basic code: * = multiplication; / = division; ^ = exponential; ** = create this many zeroes; etc.; all basic mathematical operations, like the fact that parenthesis are evaluated first, are also coded into any Excel code. So, if you want to add numbers up before you take their average, make sure you have = average(sum(cells)). Advanced code: if statements, etc.

Some useful functions are: average, stdev, sqrt, median. You can look up some others on Google or in Excel (Formulas -> Insert Function). Stdev is standard deviation. You'll learn more about this and what all the analysis options mean when you've taken a stats course.
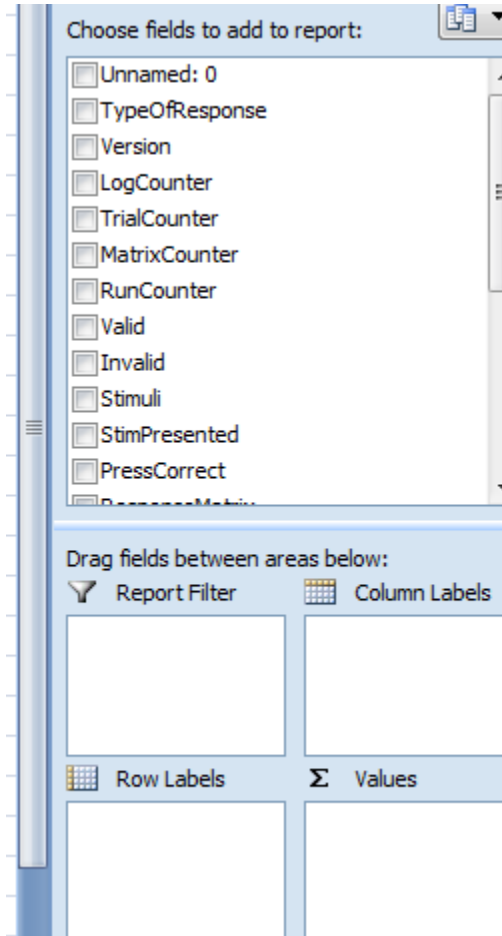
Preprocessing/Filtering data:

You can filter data with two different methods: using a filter or pivot table.

Filters:

Under data -> filter, you can create filters for the data that you have selected. Filters can be dangerous, though, especially when you use your only copy of the data. If using a filter, I'd recommend copying the data into the subsheet. Then, you'd just uncheck the numbers or labels that matched what you wanted to exclude. Filters are more permanent versions of pivot tables, so you'd have to make sure to use these correctly.
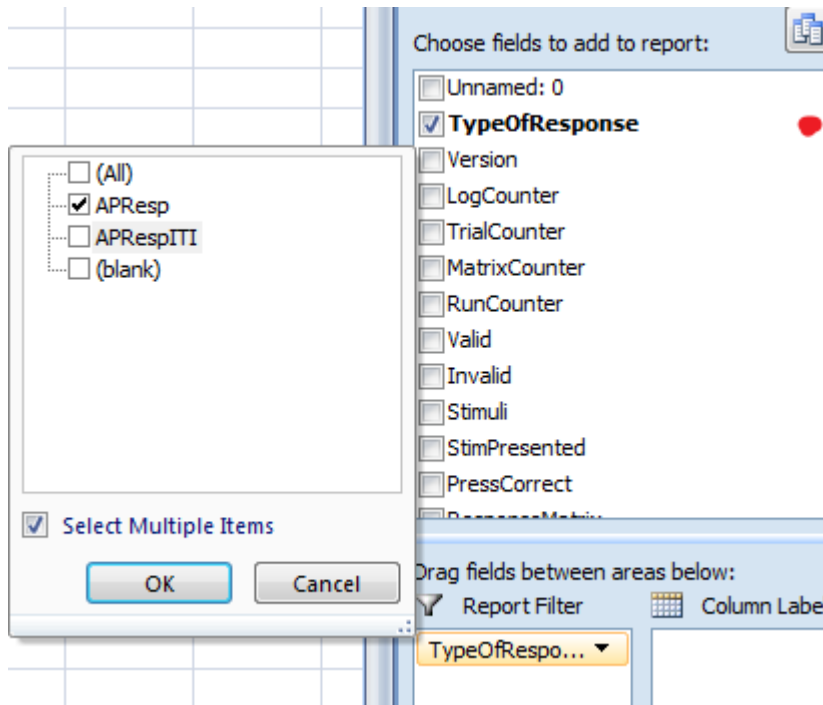
Pivot tables:

To create a pivot table, you'd select all the data that you want to be covered in the pivot table and then go insert -> pivot table (and then put this in a new subsheet or off to the side of your data in the same sheet).
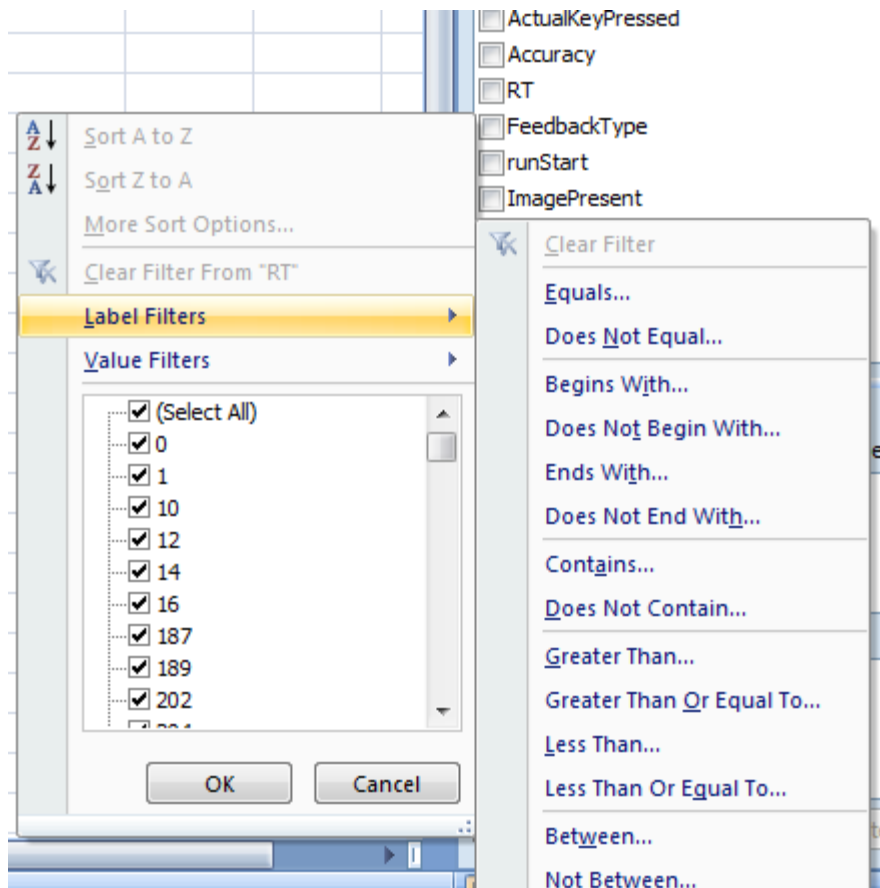

The pivot table will then take any column headers and assume that these are variables. "Report filter" is filtering the data; column and row labels are just how you want the data organized; values are the actual variable you'd be interested in. So, to give an example, when we're analyzing RTs, I'd put "Type of Response", "Accuracy", and "StimRep" until Filter, "subject" as my Row label, "Stim Presented" and "Trial Type" as my column label, and "RT" as my values.

For any variable that you include in the pivot table, you can also choose exactly which values from that variable you want to include. This helpful for the Filtering. So, I'd click on the down arrow (marked here where the red dot is) and then I'd unselect "ITI" or no responses, hit okay, and that would make only "Resp" responses be included in the table.
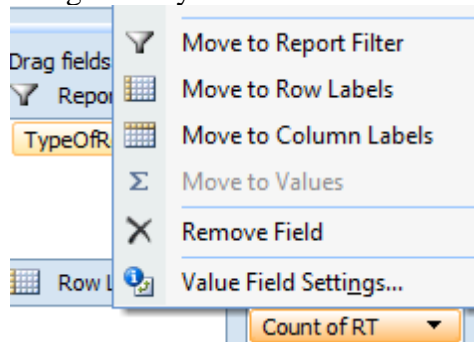
However, a variable can only be in one section in the pivot table (i.e., under filter, label, or value). So, there's also the option of creating a "value" or "label" filter for a variable that you end up putting as your value.

So for RT, we also typically exclude RTs outside a particular range. So, here, we'd select "between" and type in the lower and upper limits of our RT range (e.g., usually we don't take RTs less than 200 ms, since it's been shown that most cognitive processing does not occur until after the 200ms since stimulus presentation. Upper limit is usually how long the stimulus is shown for). The other code in the experiment should catch the "upper" RTs in that they should be marked as ITI/no responses, but sometimes JavaScript can be finnicky, so this is just an additional step.

Also note that when you put any variable in the "values" section, it defaults to showing the "count" of the variable. Usually we deal with means (or "averages" in Excel speak), so we'd change that by:

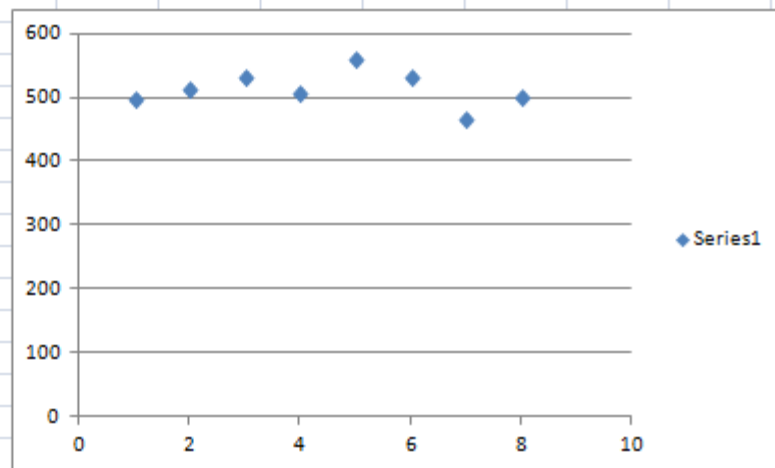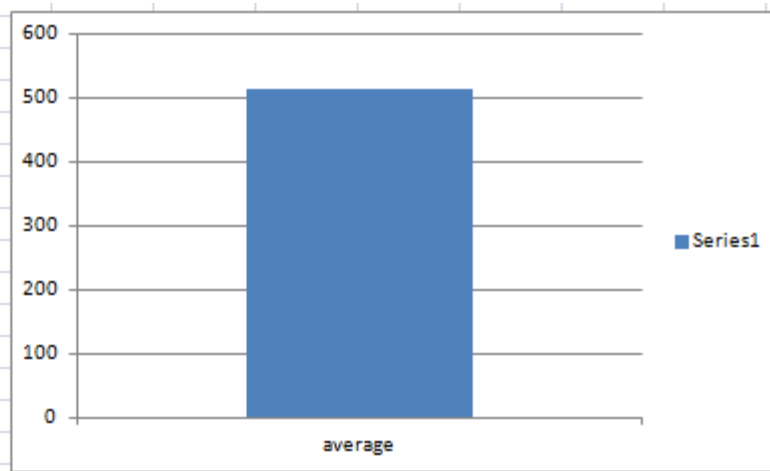selecting value field settings (after clicking the down arrow) and then selecting average instead of count.

Sorting data:

If you want to sort your data, this is also permanent like filtering your data, so I'd save another copy of the data or put the sorted data in another subsheet. You have to make sure to highlight *all* the data you want to sort. If you only sort a couple of columns, then any other columns you had associated with the data are no longer matched correctly. To sort, select all the columns of your data, go to data -> sort, and make sure you have "my data has headers" checked. You can choose any of the column headers as the variable by which you want the data sorted. Typically the data is already sorted alphabetically or numerically for the subject (just by the way I've accumulated the data). You can sort by any other variable you'd like.

Making graphs:

Excel allows you to make graphs, and many researchers use Excel to make their graphs for grant applications and papers. For this, you typically graph the "averages". So if you take some data and calculate the average, you'd select that average value, go to Insert -> and here you can choose any type of graph; we typically use bar graphs (which are called column graphs in Excel) and scatter plots. Bar graphs graph the averages; scatter plots, you'd select all your subject's individual means/averages instead of averaging them all together. In that way, a dot/scatter point represents one individual subject.

subject	RT
1	498.916
2	512.57
4	532.238
5	507.936
6	560.987
7	534.219
8	466.109
9	502.286
average	514.408
stdev	28.3979
sterror	10.0402

Each of these graphs can be further customized. You'd need labels for each axis. In this particular case, since there is only one variable "RT", you wouldn't need a legend (i.e., where it says Series1). You can format the axis to be within a particular range of values. You can change the font. You can add error bars - this is particularly important for bar graphs. You add error bars representing the "standard error" (you'll learn more about this in a stats course). To do that, you'd select the bar for which you'd want to add the error bars, go to layout, and select more error bar options. (For newer versions of excel, it actually just has a plus sign when you click on the graph, and you'd select error bars but also "more options" because the other ways excel calculates error bars are incorrect). From there you'd select "custom" and select the cell that has your standard error value calculation.

For scatter plots, you'd often be asked to provide a "linear regression line." To do this, you'd select the data points, right click and hit "add trendline." Then if you'd like, you can further customize how things look (e.g., do you want the equation to show? do you want to see the r-squared value on the graph?). Again, more of this will make sense when you've taken a stats course.

Special Notes:

Each statistical analysis program treats blank cells differently. Excel marks them as if nothing was there, and they're a blank cell. It can function just fine if you don't exclude the blank cells. Not every program is like this.

If you use a formula in excel, like average, you have to watch out when you copy and paste this number to other areas in the worksheet. Because the formula is what's in the cell, excel will copy the formula rather than the number. So, if you just want to copy the number, you'd have to make sure to hit "paste special" and then select "values" (pasting the value). Sometimes copying formulas can be useful, if you want to calculate multiple averages, but you have to make sure that the way you've structured the data is similar (e.g., if you're taking the average of data in a column that is one column away from where the average is being outputted, you have to do the same for your next average).

To quickly repeat a number, when you have a cell highlighted, there is a black box at the bottom corners. You'd select the black box and drag down. It will repeat whatever you want done. So if you had a formula in the cell, it'd repeat calculating the formula for its associated cell.

=AVERAGE(F5:F6)

D	E	F	G	H
	subject	RT		
	1	498.916		505.743
	2	512.57	◇	522.404
	4	532.238		520.087
	5	507.936		534.462
	6	560.987		547.603
	7	534.219		500.164
	8	466.109		
	9	502.286		

Here, the averages are calculated in pairs. So the first is F4:F5, then when the black box was dragged down, it calculated the average for F5:F6 to account for the fact that it's in one cell below the original, etc.