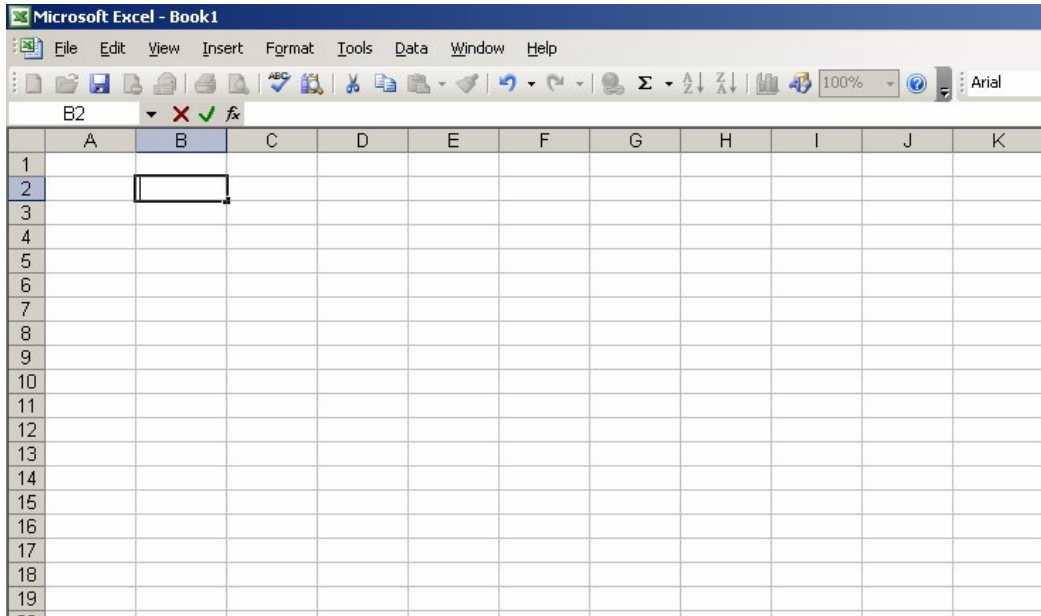
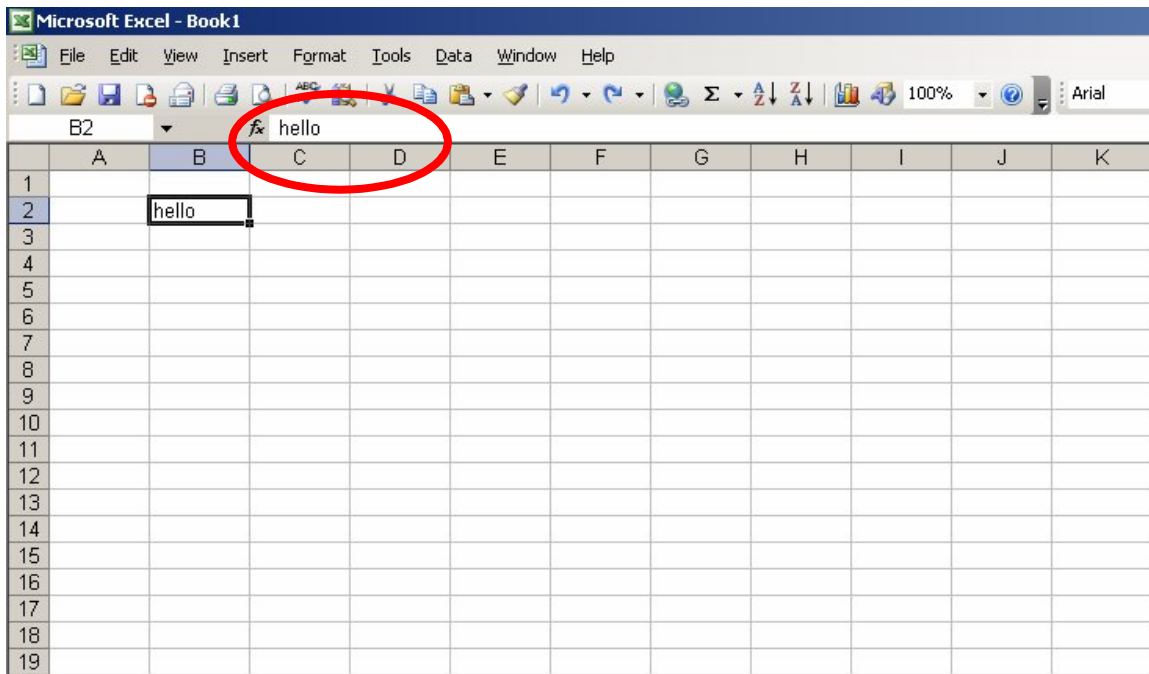


Homework 1—Excel Basics

Excel is a software program that is used to organize information, perform calculations, and create visual displays of the information. When you start up Excel, you will see the following grid, which is called a spreadsheet.



Each box is referred to as a cell. Each cell can hold text or numbers. Each cell is uniquely referenced by its column letter and row number. For example, cell B2 is located in column B and row 2. To enter information into a cell, click on a cell and type in the information.



Note that the text you type into a cell appears above the spreadsheet, as circled in the picture above.

To open a file, click on the File pull-down menu in the top left corner of your screen, and choose Open. Select the file on your computer and click Open.

For this exercise, open the data file called “ExcelBasicsDataSet.xls”. To get a copy of the data set, go to the course website <http://faculty.washington.edu/tamre/stp.shtml> and select the **Data Sets** link found on the upper left of the page. This will open a page that has a link to the data set you need. Click **ExcelBasicsDataSet.xls** to access the data set. You should get a dialog that allows you to either **Open** or **Save** the file. Make sure to save a copy of the Excel file clicking **Save** and entering a name for the file—**make sure you pay attention to where you save the file so that you can open it from Excel.** This file contains some sample data that you will use to practice various to complete the rest of this exercise. The first tab in the file contains the data. The second tab contains a description of the variables in the data set.

USING FORMULAS

SUM

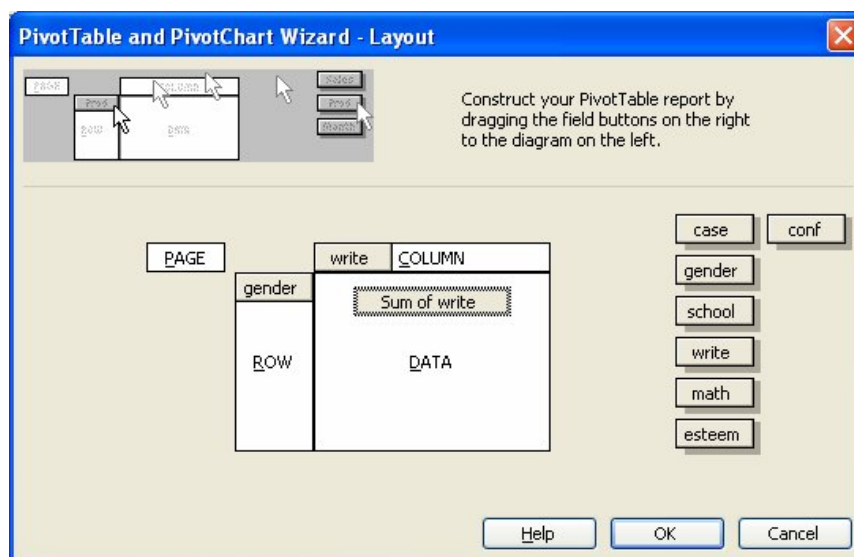
Next, sum the total writing score for the 30 students in the sample. To do this, click on a blank cell in your spreadsheet. To tell Excel to use a formula to perform a calculation, you must start with an equal sign. For example, to calculate a sum, enter the following into the blank cell:

=SUM(

Then use the mouse to select the group of cells you wish to sum and press **Enter**. Note that Excel enters the reference to the cells you select and this becomes part of your =SUM, so you end up with =SUM(D2:D31)—you must remember to enter the closing “)”. You can also enter the cell references manually with the keyboard, rather than selecting them with the mouse. If you entered the formula correctly, you should get a sum of 8411.

In the cell next to your calculated sum, type some sort of description of what that number means, such as "Sum of Writing Scores".

What if you were interested in the sum of writing scores, by gender? A useful tool for doing this is Pivot Tables. Select **Data > PivotTable and PivotChart Report** from the main Excel toolbar. Choose **Next >** in the dialog that opens. To enter the Range, use the mouse to select all of the data, including the variable labels, then choose **Next >**. You have the choice of having the Pivot Table written to a new worksheet or put in the existing worksheet. If you want the table to remain in the existing worksheet, choose an empty cell that will become the top, left-hand corner of the table (recommend cell A34). Now, choose **Layout**. Notice that the variable names show up on the right hand side of the layout dialog. Click on gender and drag it to the area labeled ROW; click on write and drag it to the area labeled COLUMN; click on write again, and drag it to the area labeled DATA. If you have done this correctly, your layout dialog should look like the screen shot below.



Now click **OK** to return to the previous dialog. Click **Finish** to create the table.

The Pivot Table function treats all the variables as categorical. Thus, it will create a cell for each unique writing score that is in the data set. Also, notice that it breaks the scores out by gender (remember 1 = female and 2 = male). Since you want the sum of writing scores by gender, go to the last column of the Pivot Table where the Grand Totals are given. If you did this correctly, you should get a sum of 4346 for females and 4065 for males. Notice that $4346 + 4065 = 8411$, the sum of all writing scores.

Next, repeat the process above for the sample math scores.

AVERAGE (MEAN)

To calculate the mean (average) height, click on another blank cell, and type the following:

=AVERAGE(

then select the group of cells to take the average of as you did before for the sum. Be sure to include a description in the cell next to this number. Do this for both the writing and math scores.

Next, calculate the average writing and math score for each gender separately. This can easily be done using the Pivot Table function. This time you do not need to repeat the entire Pivot Table set of dialogs. Rather, go to your Pivot Table output for the writing scores. Double click on the cell that says “Sum of write” (top left corner of the table). Now select **Average** from the Summarize by drop-down list. Press **OK**. The Pivot Table now displays averages, by gender, for the write variable. If done correctly, you should get the overall average as 280.37, with an average of 289.73 for females and 271.00 for males.

Repeat this process to find the averages for the math scores (overall and by gender).

MEDIAN

To calculate the median writing score, find another blank cell, and type the following:

=MEDIAN(

and select the group of cells as before. Be sure to include all students.

Which is bigger, the mean (average) or the median?

MODE

The formula for calculating the mode is:

=MODE(

Use this function to find the mode of the writing data. Don't forget to include a description!

STANDARD DEVIATION

The formula for calculating the standard deviation is:

=STDEV(

Calculate the standard deviation for the writing data, and include a description.

QUARTILE

The quartile formula can be used to calculate the minimum value, the lower quartile, the median, the upper quartile, and the maximum value. Enter the formula as below:

=QUARTILE(

Then select the group of cells (write), and press the comma, ",". Next, enter one of the following numbers, depending on which quartile you'd like:

- 0 Gives the minimum value
- 1 Gives the lower quartile (25th percentile)
- 2 Gives the median
- 3 Gives the upper quartile (75th percentile)
- 4 Gives the maximum value

In your spreadsheet, use the QUARTILE formula to calculate these 5 values. Be sure to label each one correctly.

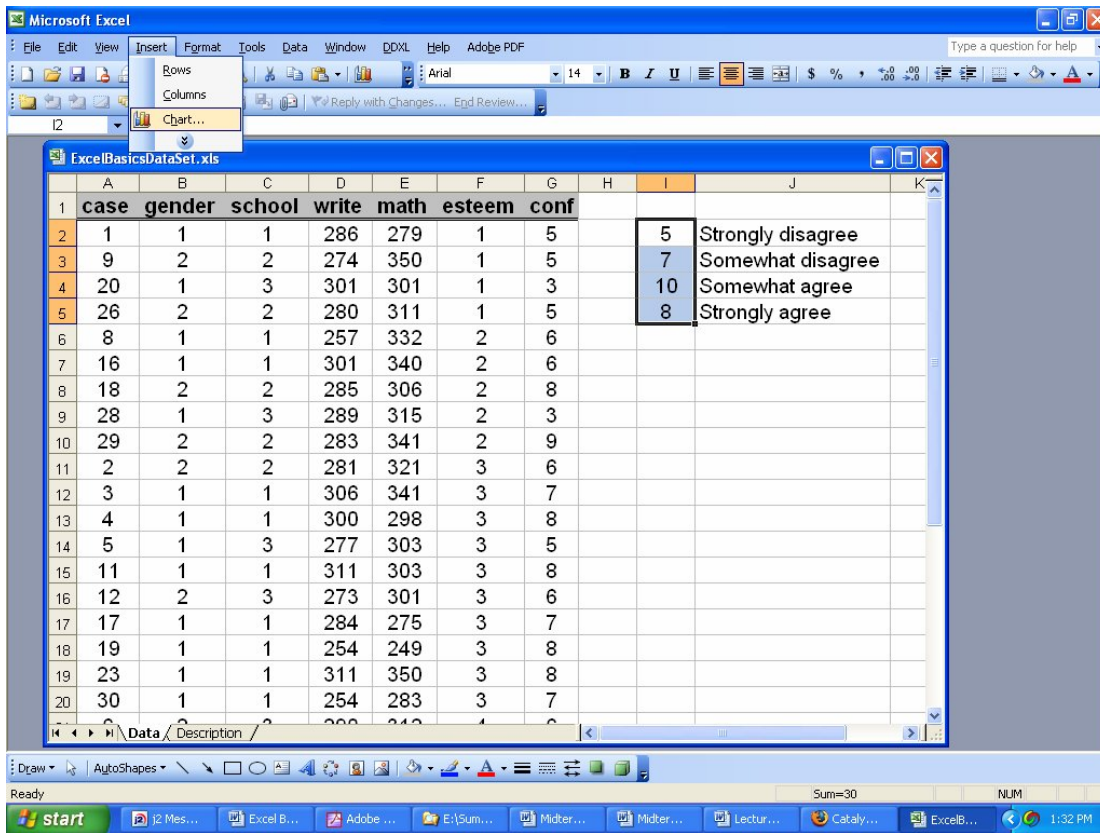
Using the functions described above, find the median, mode, standard deviation, and quartiles for the math variable. Just do this for all students. Compare the mean and the median.

MAKING GRAPHS

BARPLOT

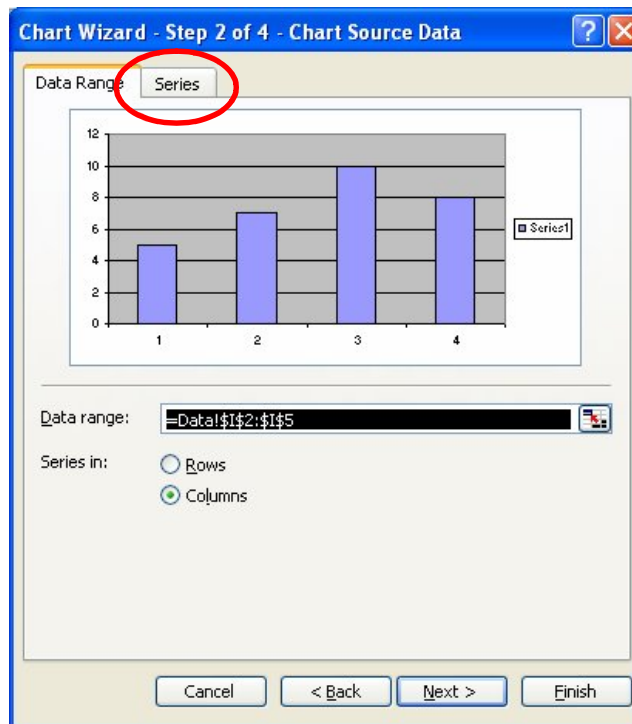
Excel needs to have counts of how many individuals fall into each group in order to make a bar plot. We will make a barplot of responses to the esteem question. To start, you need to count how many students came from each of the four possible esteem categories (strongly disagree, somewhat disagree, somewhat agree, and strongly agree). You can either count these up by hand (sorting or filtering the data can make this relatively easy for a small data set) or, you can use the Pivot Table dialogs to get the counts. Once you have the counts, enter them into cells I2 through I4. In cells J2 through J4, enter esteem response labels (Strongly disagree, Somewhat disagree, Somewhat agree, and Strongly agree).

To make a bar plot, use your mouse to highlight the column of counts. Then click on the Insert drop-down menu at the top of your screen and select **Chart...**, as shown in the screen shot below.

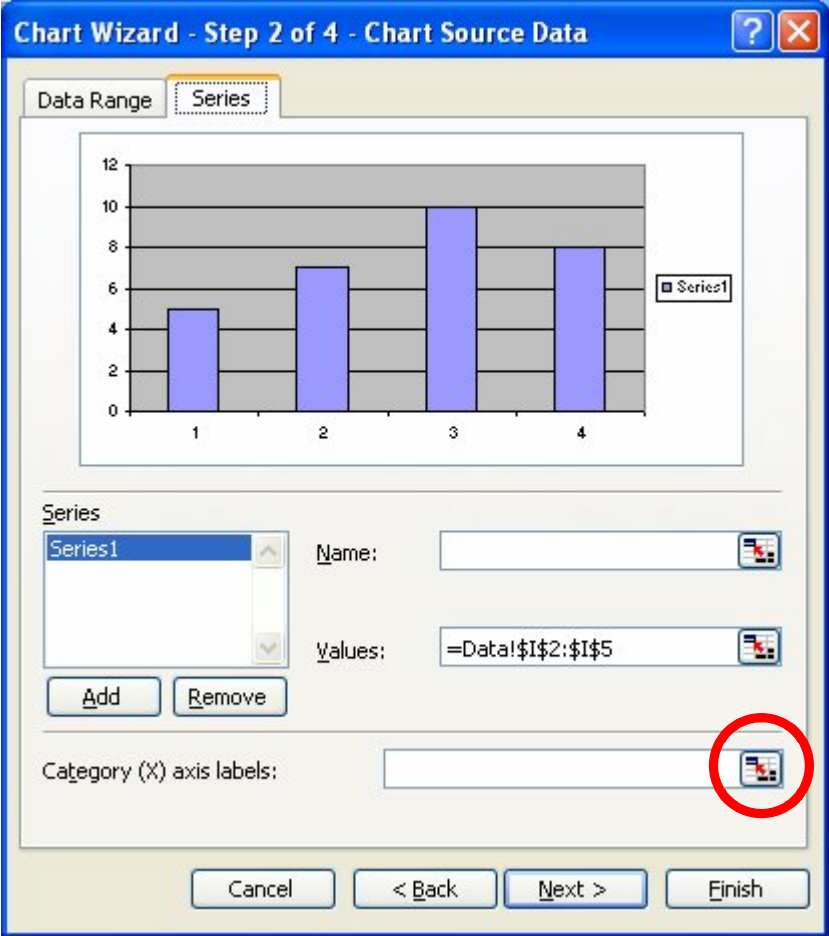


Select the "Column" chart type (the first one) and click Next.

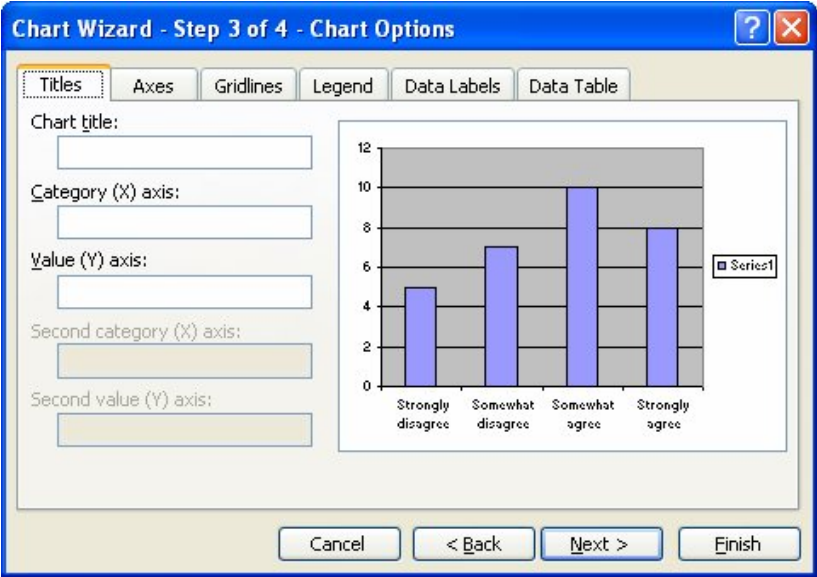
On the next screen, the Data Range tab should show the series in columns, as below—your chart should look a bit different, as the numbers were changed for this example.



Click on the Series tab circled at the top of the screen shot shown above . At the bottom, click on the Category (X) axis labels, circled in the screen shot below.



With your mouse, select the column of labels (You should select cells J2 through J4). Select Next >.



For step 3, enter an appropriate label for the Category (X) axis, such as "Feelings about Self-Esteem", and another for the Value (Y) axis, such as "Number of Students". Feel free to change the title as well if you choose. Select the **Legend** tab and unclick the box marked "Show Legend"—there is no reason to include a legend on a boxplot when only one variable is summarized (esteem in this case). Click **Next >**. For this example, it is not necessary to change anything on the other tabs in this screen, but you may experiment with the options if you would like. Then click **Next >**.

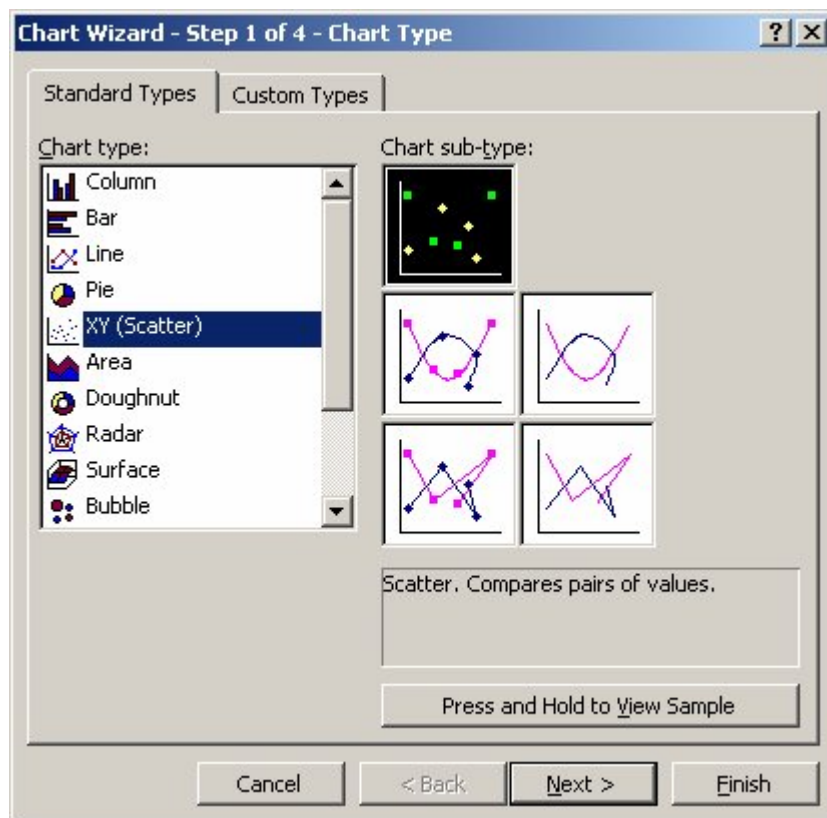
For step 4, choose to place your chart as an object in Data, and click **Finish**.

SCATTERPLOT

Scatterplots are useful look at the relationship between two quantitative variables.

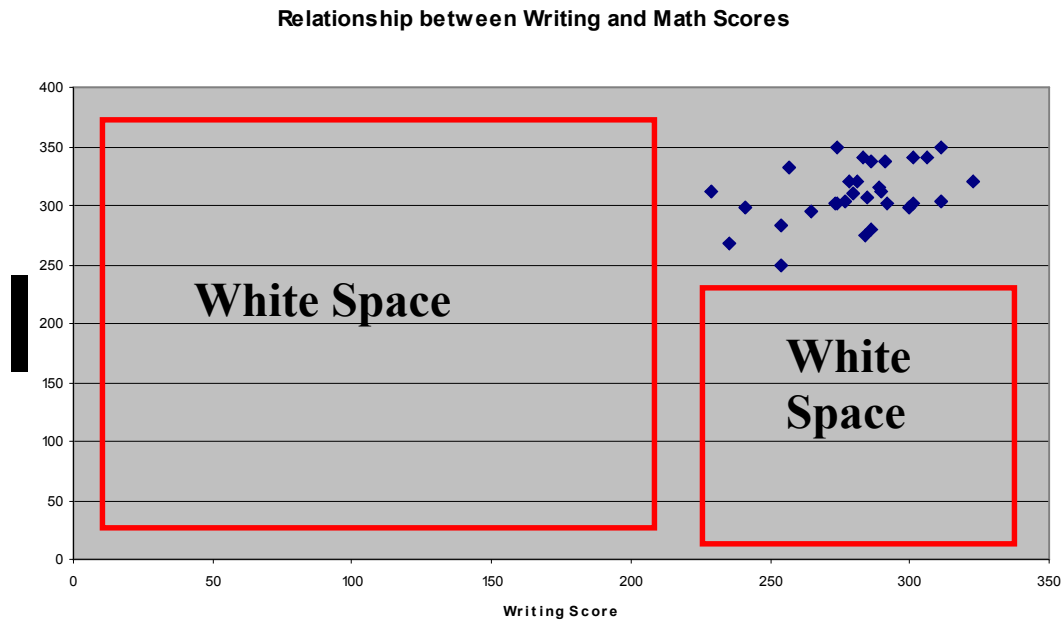
Select the two columns labeled “write” and “math” with your mouse—just select the data, not the column labels. Click on the Insert drop-down menu and choose Chart.

Choose XY (Scatter), and select the first option available to you, as shown below. Click Next.

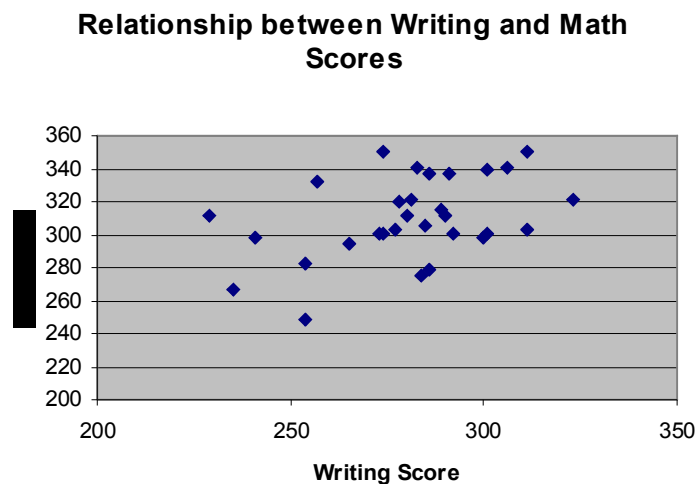


Click Next to move on to step 3. On the Titles tab, type an appropriate title (optional), and be sure to label the x -axis as the “Writing Score”, and the y -axis as the “Math Score”. Note: when you choose two columns at a time to create a scatterplot, Excel automatically uses the first column as the x variable and the second column as the y variable. You can control which variable is assigned to x and y by using the Series tab to enter the data. Under the Legend tab, unclick the box marked "Show Legend"—there is no reason to include a legend on a scatter plot when you only have one set of data. Click **Next >**.

Insert this chart as a new sheet. Your completed scatterplot should look something like the plot shown below. Notice the how Excel defaults the scaling of the x- and y-axes. There is a lot of wasted “white space,” which can often make it hard to see potential patterns in the data.



To rescale the axes, place the mouse pointer on one of the x -axis values, and right click. Select **Format Axis...**Select the **Scale** tab and then adjust the minimum value of the x -axis---set the minimum value to 200. Repeat this process for the y -axis. You can experiment with other chart options if you would like. For example, you can change the shape, size, and color of the plotting symbols by selecting one of the symbols with the mouse pointer and then right-clicking. After you right-click, select **Format Data Series...** Other options can be found by pointing to different parts of the plot and right-clicking. Your final chart should look something like:



Print only the chart; don't worry about printing the data for this graph. **On your printout (or in your word document), write a sentence that describes the observed pattern between the math and writing scores.**

To Turn In

(due Thursday July 31 in Statistics quiz section)

Turn in a computer print-out that includes the sum, average, median, mode, standard deviation, and quartiles that you calculated. Also be sure to include the sums and means by gender for the writing and math scores. Also include your barplot and scatterplot.

You can either turn in a nicely-formatted printout from Excel, or a Word document that you have copied all of the results into – the key is to make sure that whatever you turn in has all of the results and graphs you created in this exercise, in a clear and readable format—everything should be appropriately labeled. If you plan on printing from the Excel file, be sure to use the Print Preview dialog found under the File drop-down menu in order to make sure the material you print is appropriately formatted. Ask if you need help formatting the output.

Summary References

A3:A8 cells A3 through A8
A3, A8 cells A3 and A8, no others

Formulas

=SUM(cells)	returns the sum of the selected cells
=AVERAGE(cells)	returns the mean of the selected cells
=MEDIAN(cells)	returns the median of the selected cells
=MODE(cells)	returns the mode of the selected cells
=QUARTILE(cells, quartile)	returns the quartile (0,1,2,3 or 4) of the selected cells. 0 returns minimum, 1 returns 1st quartile (lower), 2 returns median, 3 returns 3rd quartile (upper) and 4 returns maximum.
=STDEV(cells)	returns the standard deviation of the selected cells