A. Download the data spreadsheet, open it, and select the tab labeled Murder. This has the FBI Uniform Crime Statistics reports of “Murder and non-negligent manslaughter” in the United States for the years 1985-2004. Column A (Year) is the number of years since 1985 and column B is the number of murders reported that year. Select the data (block A2..B21) and graph it with an XY (Scatter) plot. Looking at the data, we can see several turning points, which is the usual sign that a polynomial fit might be in order. Lines, power curves, and exponentials have no turning points. Quadratics have one turning point and must be symmetric about that turning point.

B. Click on one of the data points marked in the plot, then right-click to get a pop-up menu of options. Select “Add Trendline …” This will give you a dialog box with a variety of different types of models to choose from. Select the Polynomial model and set Order to 3. Then click on the Options tab and check the “Display equation on chart” box. Also check the Custom option for Trendline name and give the name “3rd order.” Click OK and a polynomial curve will be drawn that is the “best fit” 3rd degree polynomial curve for the data provided. The equation will also appear on the chart. Note that you can click on the equation and move it around on the chart so that it isn’t obscured by the data points and the trendline. Your graph should now look something like this:

\[
y = 9.5516x^3 - 321.34x^2 + 2591.5x + 17529
\]

C. This trendline doesn’t seem to fit the data very well. We might do better if we used a higher order polynomial, which could have more turning points. Click one of the data points and go through the same procedures as in step 2 again, except that this time set the Order to 6 (and set the name to 6th order of course). Once you have the trendline on the graph, you can right-click on it to get a pop-up menu of options. Select Format Trendline… and you will get the dialog box back. But this time you will have a Patterns tab. If you select the Patterns tab, you will get
options to change the color and weight (width) of the trendline, so it is easier to distinguish the two trendlines. Your graph should now look something like this:

\[
y = 9.5516x^3 - 321.34x^2 + 2591.5x + 17529 \\
y = -0.0228x^6 + 1.1035x^5 - 17.552x^4 + 93.83x^3 - 65.645x^2 + 319.71x + 19363
\]

D. This trendline seems to do a much better job. But looks can sometimes be deceiving. A trendline is supposed to show the overall trend, and not necessarily follow all the little bumps of the data. Indeed, if your trendline follows the bumps too much, you end up “modeling the noise” and actually getting a less appropriate model. To get a sense of some of the difficulty, suppose we use both these models to predict how the murder rate will look in 2005 and 2006. Right-click on a trendline to bring up the pop-up menu and select Format Trendline. Click on the Options tab and then set the Forecast to Forward 2 units. Repeat this for the other trendline. Your graph should now look like

So if the 6^{th} degree polynomial fit is accurate, in 2006 we should have solved the problem of murder in the United States. Unfortunately, we haven’t. While final numbers aren’t yet available, the FBI’s preliminary report says murders went up from 16,137 in 2004 to
16,910 in 2005, which is not all that far off from the 3rd degree model’s prediction (which was 17,236). It is typical of polynomial models that in making extrapolations you can have trouble when the model starts rapidly heading off to either positive or negative infinity. You can also have trouble, particularly if the data is not evenly spaced, with the model making big curves between data points. All these problems are magnified as you use higher degree approximations (which is one reason why Excel limits you to at most an order 6 approximation). For example, using the data in problem 24 in section 4.2 for births to females under 15 in the United States, we could use a 9th degree polynomial which allows for enough flexibility to passes exactly through every data point. But this wouldn’t produce a useful model, as the graph below illustrates. The model predicts wild gyrations in the birth rate peaking at over 500,000 in 1968 and falling to -315,000 in 1999, which are obviously absurd values. By the way, the teenage birth rate did fall sharply in the late 90’s with a strong turning point just at the end of this data set.

There are advantages to using unevenly spaced data in terms of giving extra weight to more recent data in building your model. We won’t go into weighing these advantages vs. the disadvantages in this class. For now, I just want you to get the following points. Understanding these points is why we study how the graphs of polynomials look.

- Polynomial models can be useful when the data you are modeling has turning points.

- It is important to choose as low a degree polynomial as reasonable to capture just the trends and avoid “modeling the noise.”

- Polynomials will usually only forecast accurately for a limited range at the end of your data before they take off to plus or minus infinity.
Assignment:

1. Click on the GTA tab. This sheet has the Motor vehicle theft data for the last 20 years. Graph the data and pick an appropriate model. Use the model to predict motor vehicle thefts for 2005. List the model you used, a brief explanation for why you picked it, and the prediction for 2005 car thefts below.

2. Do problem 26 in section 4.2. List your answers below.
3. Do problem 31 in section 4.2. List your answers below.

4. In problem 31 part d, you compared the model values to get the ratio of the juvenile crime rate in 1999 to the juvenile crime rate in 1990. What is the ratio of the observed data for 1999 and 1990? Write a short paragraph about the advantages and disadvantages of comparing model values vs. observed values.