#1. Eruptions of the “Old Faithful” Geyser
This gives you practice in looking (really looking) at graphs.

A geyser is a hot spring in which water intermittently boils, sending a plume of water and steam high into the air. Such fun to watch! Importantly, eruptions tend to follow a reasonably stable pattern.

Suppose you are a tourist at the Yellowstone National Park (or, in my case, Iceland), keen to witness first hand an eruption. How long will you have to stand there, twiddling your thumbs, waiting for the next eruption?

The following are two graphical summaries of the distribution of 222 waiting times, a histogram and a stem and leaf diagram.

For Discussion: For each summary: (a) explain the graphical technique; and (b) state the “facts” of the summary of the 222 waiting times.
#2. Eruptions of the “Old Faithful” Geyser – continued

This is a reminder. **Beware** - Not all exploratory techniques are effective!

Below are two graphical summaries of the distribution of the 222 waiting times.

Left graph -
This is a **box plot** of all 222 waiting times.

Right graph
This is a **side-by-side box plot**. Here the 222 waiting times have been partitioned into 2 groups. The groups are defined by the duration of the *preceding* geyser eruption: short (< 3 minutes) versus long (> 3 minutes). A separate box plot is shown for each group, side-by-side.

**For Discussion:** (a) explain the graphical technique; (b) why is the left hand picture *not* an effective graph; and (c) what do you conclude from the right hand picture?
#3. Calls to the NYC Auto Club

More practice in looking at graphs, this time as an entrée to regression ....

The Automobile Club of New York city provides all kinds of services to its members. One of them is emergency road service, including tow truck assistance. As you might imagine, the demand for tow truck assistance might vary depending on the time of day or the weather. Here we consider variations in the demand for tow truck assistance in relationship to the weather. Bad weather! Cold weather, in particular. The Automobile Club of New York would love to be able to anticipate greater or lesser demands for tow truck assistance, so that it can be sure to have enough staff and trucks on hand.

The question is – Are the number of calls placed to the Automobile Club of NYC in any given day predictable from the previous days’ weather forecast?

To address this question, the Automobile Club of NYC has collected some data. Here we are looking at data from two periods of time: the 2nd half of January 1993 and the 2nd half of January 1994. The following are two graphical summaries, a time series plot and a scatter plot. In the scatter plot, the data plotted are X-Y pairs defined as follows: X = LOW = lowest temperature for the previous night; and Y = CALLS = # emergency road service calls received the next day.

For Discussion: For each summary: (a) explain the graphical technique, (b) state the “facts” of the summary, and (c) what conclusions do you draw?
#4. Calls to the NYC Auto Club - continued

All regressions should be preceded by: 1) an understanding of the study design, 2) a translation of the research question(s) into specific aims and/or statistical hypotheses, 3) understanding of the data, 4) descriptive statistics and, depending, 5) "model"-free explorations of the associations of interest.

The setting is the same. For concreteness, suppose we have a data set containing 28 paired observations (X=low, Y=calls), with X and Y as previously defined. Recall the research question: are the number of calls placed to the Automobile Club of NYC in any given day predictable from the previous days' weather forecast? Below is a larger scatter plot (this time produced in Stata).

For Discussion: (a) what type of question is of interest in this regression: prediction, estimation or testing, (b) what sampling plan was used to obtain the study data: retrospective or prospective, observational or intervention, (c) what type of variable is X=low? Y=calls? , and (d) suggest a “to do” list for producing preliminary descriptive statistics (both numerical and graphical).
#5. Calls to the NYC Auto Club - continued

More practice in looking at graphs, again as an entrée to regression.

Below is an overlay plot showing: the scatter of XY (X=low, Y=calls), the fitted line and the associated 95% confidence bands.

For Discussion: (a) state what is shown, (b) explain why the width of the confidence band changes, (c) relating this graph back to the original question of interest, what advice do you have for the Automobile Club of NYC?