Industrial Liaison: Dr. Genetha Gray,   Intel Corporation,

Consider the data at: http://ag.arizona.edu/azmet

Use the information related to solar radiation in order to calculate some exceedance probabilities that would demonstrate to financial backers how to consider the risks that arise from the uncertainty of weather when deciding whether or not to build new renewable (in this case, solar) power stations.

Here are the issues that arise:
- There are pieces of "bad" data, i.e. times when the equipment failed to take a reading- how do we deal with those?
- There are both hourly and daily readings- how do these differ and how can we take this into account?
- How does this method compare to NRELs TMY (typical meteorological year)?

There is also a lot of other information in the data set besides just the solar radiation so some work could be done with the other information.
A student could try to understand using what has happened previously to predict the future. He/she could see how these predictions changed as additional information was added. For example, given years 1990-1999, try to predict what will happen in 2000.   Use different subsets of the data and observe how the predictions change. Compare your prediction to what actually happened in 2000. What other information might help improve the prediction? Etc, etc.

Starting Point for Faculty

I could certainly help set up an abbreviated version of what I worked on.

There is a lot of work to be done with the data and the students can be quite clever about how to do it. They also might learn some R to complete this task or another related language. Matlab may also help.

In terms of the math and stats, some things are very basic like just calculating and using means and standard deviations or quartiles and even just deciding what to use. The more advanced stats ideas would be things like exceedance probabilities and forecasting techniques (like ARIMA models perhaps).   There might also be some building of a math model from data (and what kind) or just generally considering a response surface built from the data and how to use it to make decisions or predictions.

Finally, the students could also try to learn some of the basic data mining techniques that are not taught in class. I think that everyone learns about regression but other things like clustering are not taught in an intro class.