Now that we know how to create a dataset, we can use Stata in order to make it readable. The first step in making data readable is to use variable names that make sense. It is good practice to use variable names that start with a small letter. If you want, you can name a variable by joining two words with an underscore. I generally do not like this because it has been my experience that the shorter the variable name, the easier it is for me to perform the analysis. The rule that I follow is that I pick a name that is as short as possible but still makes sense.

After naming the variables, we should use labels. Labels are a description for variables. A label is a statement that describes what the variable is. Labels are really helpful when we generate graphs using Stata because Stata is smart enough to use these labels in the graphs in order to make them more readable.

Another thing that we do in order to make the data readable is that we can label certain values in a dataset. This section will hopefully show you how you can give meaning not only to the variables, but also to specific values that a value can take. For example, I once received a dataset in which a certain variable took on one of two values, either a one or a zero. When I inquired about this I was told that a zero meant that the respondent was female while a one meant that the respondent was male. In this section, we will see how we can tell Stata what each value represents.

In order to work through the section, we will be using a version of the grades dataset that we used in section one. Remember the dataset that was called dataset1 and contained grade information about students? I have attached to this lecture a similar dataset that contains the same information but without the variables names, labels, or even value labels. Our job is to make the dataset as readable as the one we used in the first section.