

In the previous lecture, we talked about generating value labels and attaching these labels to certain values. This, as we said, allows Stata to use deal with numbers while at the same time allowing us to understand what these numbers mean. But what if the data was not entered as zeros and ones, or ones and twos? What if the person who entered the data actually entered “Female” and “Male”. To see this, download and open the file *datasetother* which is attached to this lecture. Either double click on the file in windows to open it or save the file in a folder, navigate to the folder using the **cd** command, and then load the dataset using the **use** command. Once you have opened the dataset, open the data editor window. You will notice that instead of the variable *gender* we have a variable names *gender2*. The name is not the issue however. The issue is that the values in the cells that correspond to this variable are written in red. As you recall, data written in red is of type string. In the previous lecture, we saw that the labelled data was written in blue. We conclude therefore that in the current dataset (*datasetother*) the data has not been coded. Instead, it has been entered as text. As I have said previously, using numeric values and value labels together is better than using string data types for this sort of data. Luckily for us, Stata has a very simple command that will do the work for us, and this command is **encode**. With the dataset loaded into the memory, execute the following command:

```
encode gender2, gen(gender)
```

If you now look at the right-hand side of the screen, you will notice that there is a new variable which is named *gender*. This is because the command that we just executed told Stata to encode the variable *gender2* and to store the values in a new variable that it should name *gender*.

Open the data editor now. What you will see is that the variables *gender2* and *gender* contain the exact same values. But that the color of the text differs. Looking at the variable *gender*, we see that the text is blue, which means that what we are actually seeing are value labels and not the actual values themselves. Go back to the command prompt and run the command **describe**. You will notice that the variable *gender* is of type long, which is a numeric type, and that it has a value label attached to it, and that this value label is called *gender*. Let us ask Stata to list the labels included in this value label:

```
label list gender
```

We see that the values are actually ones and twos and that the text that we see is just labels that have been attached to these values. As you can see, using just one command, Stata created a new variable named *gender*, populated the variable with ones and twos, created a value label that gave meaning to the ones and twos, and applied the value label to the newly generated variable.