

In this section, we saw how to create new datasets from within Stata, how to import datasets from both Excel files and CSV files. We have also seen how to create our own variable names and how to import these names from another spreadsheet program. We ended the section by loading an already saved dataset into Stata.

Although all these commands are important, they are not enough. It is not enough to just have the dataset in Stata. The dataset needs to be clean and it needs to be readable. Looking at the dataset that we worked with in this section, when someone sees a variable named country, he or she might think that this variable refers to nationality for example, or perhaps the current location of an individual. You might argue that since we created the dataset, we already know what each variable refers to. This argument however breaks down when a dataset contains a large number of variables with some variables containing similar information. In addition, you might find yourself having to return to a dataset on which you previously worked. There is a high probability that you will not be able to remember what every single variables refers to. This is why the next step after entering the data and loading it into Stata is to make sure that the dataset is as readable as possible. This will make our job easier when we look at the data, and it will also make it easier to read the output from Stata. As we will see in the next section, the readability of the output produced from Stata depends on the effort that we have invested in making the data as readable as possible.