

We will first start by creating a dataset within Stata using our keyboard. To do that, click on the “Data Editor (Edit)” button. Let’s say that I want to create a dataset that contains someone age, status (married or not), number of children, and country of birth. To do that I enter the data for each observation in the same row. Notice that when I enter the information, Stata labels each column. Notice also that some entries are in red while others are in black. With regards to the column names, since each column is a variable, Stata has taken it upon itself to name each variable. The names that Stata gives are not nice. In order to make our dataset more readable we would rename the variables. Click on the entry in the first column. You will notice that the table in the right-hand side of the screen is populated with information about the variable. The first field is “Name”. This is the name of the variable. Change it to *age*. You should make it a habit to use small letters when naming variables in Stata. We now notice that the name of the column has changed. Let us do the same for the other variables. Change their names to *status*, *children*, and *country*.

With regards to the colors used in each cell, Stata colors variables of type string in red and variable which are numeric in black. If you click in a cell under the *children* column, you will notice that the type is *float*. A float is a numerical value that allows for decimal places. This doesn’t make sense in the case of this variable because the number of children must be a whole number. We can change the variable type by choosing *int* from the drop down menu. Now we can enter a few more entries.

If you try to enter text into a cell that has a numeric value, Stata will not allow it. Now try deleting a numeric value. You will notice that Stata represents the missing value by a “.” This is how Stata treats missing numeric values. This is an important point because researchers usually want to know how many missing values of a certain variable are there.

Now close the data editor. You will see that Stata has been executing commands without us knowing about it. Whenever you use the user interface to modify the data, Stata executes a command in order to execute the requested changes. I have been using Stata for eight years and I have never entered data into it using the command prompt. If you look at the output you will know why. Entering the entire dataset using the command prompt is tedious. This is why you should only use the command prompt when you are performing operations on a dataset that is already created.

Now that we have entered our data, let’s save the dataset. To do that we can use the **save** command like this:

```
save mydata1
```

The output tells us that the data has been saved. In order to make sure that this is the case, we can ask Stata to list the contents of our working folder by using the **ls** command. The output shows that the dataset is now saved in the working folder under the name *mydata1.dta*. We have just saved the data, but it is still loaded in memory. If we want to remove the data from the memory, we can execute the following command:

```
clear all
```

This tells Stata to clear its memory. To see that this is the case, you can run the **describe** command.

As you can see, creating a dataset in Stata is easy, thanks to the friendly user interface. However, in most cases, I find myself importing the data into Stata. The reason is that not everyone has Stata. This is why I find myself in many cases importing the data from a spreadsheet program such as Excel. The next section will show you how to do this.