

Visualizing data in Stata – Section 2.1

Najib Mozahem

As mentioned in the last section, this section will introduce some graphical tools that are used to tell us some information about certain variables. In general, there are two types of variables. The first type, which will be the focus of this section, can take on many values. Take the variable *gpa* for instance. One student might have a GPA of 87.65%, while another student can have a GPA of 74.21%. Other students can have many other values.

The second type of variables are not used to measure something. Instead, they are used to categorize individuals. Take the variable *college* for example. In our dataset, students were enrolled either in the school of engineering or the school of business. The variable *college* does not really measure anything. It just tells us to which group does an individual belong. What is interesting about these types of variables is that in a dataset, they are coded using numbers. In our dataset, the value of the variable *college* is either a 0 or a 1. In order to make things readable for us, we create a value label and we attach it to this variable. This value label specifies that a 0 indicates that the student is enrolled in the business school while a 1 indicates that the student is enrolled in the engineering school.

It is very important to know the distinction between both types of variables because what we are actually interested in are the values of the first type. For example, I care about the average GPA, or the standard deviation of GPA. This is why this section starts with these types of variables. This does not mean that the second type, categorical variables, are not important. However, first we will study single continuous variables (variables that can take on a large number of values).