

Visualizing data in Stata – Section 1.1

Najib Mozahem

Hello and welcome to the course Visualizing Data in Stata. The purpose of this course is to teach the student to use the proper visual tools that come with Stata in order to make it easier to know what the data is telling us. Usually when there is a book or a course that teaches how to plot graphs, several dataset would be used in order to illustrate what type of graph is used with what type of data. Although useful, I think that this technique is lacking. The best way to show data visualization is not to tell the user that to visualize a certain type of variable you have to use a certain type of graph, and then we move on to the next graph with a different data set. The best way to go about explaining visualization is to use a complete dataset and try to extract the information using the graphical tools that are available for us. This is why in this course I will be using just a single dataset. The data contains different types of variables. This way we can use different types of graphs on different types of variables in the same dataset. By doing this, we can go over the material in a more intuitive way. If you have taken my previous course An Introduction to Stata, you will know that I believe that the purpose of data analytics is to extract the story that is hidden in the data. This is the same approach that I will use in this course. We will be using the same dataset, and our job will be to uncover the story that is hidden within it using the different graphical methods of Stata.

You do not need to be an expert in Stata. In fact, this is an introductory course to the Stata graphical tools. The course does not assume any previous knowledge in plotting graphs in Stata, but it does assume that you know the basics of data management in Stata. You need to know how to load datasets into Stata, how to execute commands, and how to navigate the user interface. You do not need to know any statistics or any programming.

Since we will be using the same dataset, it would be instructive to get to know the dataset and the different variables that are included in it. This is why the purpose of the next section is to introduce the user to the dataset that we will be using.