⚠ Lagunita is retiring and will shut down at 12 noon Pacific Time on March 31, 2020. A few courses may be open for self-enrollment for a limited time. We will continue to offer courses on other online learning platforms; visit http://online.stanford.edu.

Course > Introduction > The Big Picture > The Big Picture

🔖 Bookmark this page

The Big Picture

In a nutshell, what statistics is all about is **converting data into useful information**. Statistics is therefore a process in which we

- Collect data,
- Summarize data, and
- Interpret data.

To really understand how this process works, we need to put it in a context. We will do that by introducing one of the central ideas of this course—the **Big Picture of Statistics**. We will introduce the Big Picture by building it gradually and explaining each step. At the end of the introductory explanation, once you have the full Big Picture in front of you, we will show it again using a concrete example.

The process of statistics starts when we identify what group we want to study or learn something about. We call this group the **population**. Note that the word *population* here (and in the entire course) does not refer only to people; it is used in the broader statistical sense to refer not only to people, but also to animals, objects, and so on. For example, we might be interested in

- The opinions of the population of U.S. adults about the death penalty
- How the population of mice react to a certain chemical
- The average price of the population of all one-bedroom apartments in a certain city

Population, then, is the entire group that is the target of our interest:
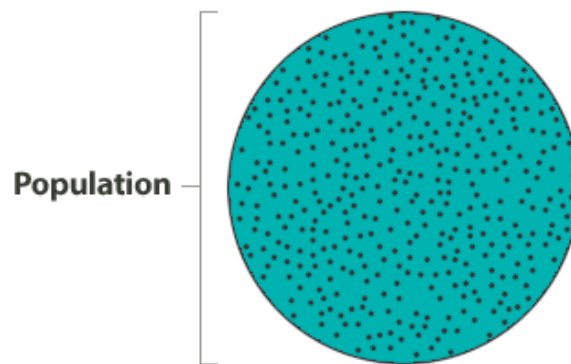
In most cases, the population is so large that, as much as we want to, there is absolutely no way we can study all of it (imagine trying to get the opinions of *all* U.S. adults about the death penalty). A more practical approach would be to examine and collect data only from a subgroup of the population, which we call a *sample*. We call this first step, which involves choosing a sample and collecting data from it, **producing data**.


It should be noted that since, for practical reasons, we need to compromise and examine only a sub-group of the population rather than the whole population, we should make an effort to choose a sample in such a way that it will represent the population well. For example, if we choose a sample from the population of U.S. adults, and ask their opinions about the death penalty, we do not want our sample to consist of only Republicans or only Democrats.

Once the data have been collected, what we have is a long list of answers to questions, or numbers, and in order to explore and make sense of the data, we need to summarize that list in a meaningful way. This second step, which consists of summarizing the collected data, is called **exploratory data analysis**.

Now we've obtained the sample results and summarized them, but we are not done. Remember that our goal is to study the population, so what we want is to be able to draw conclusions about the population based on the sample results. Before we can do so, we need to look at how the sample we're using may differ from the population as a whole, so that we can factor that into our analysis. To examine this difference, we use **probability**.

In essence, probability is the "machinery" that allows us to draw conclusions about the population based on the data collected about the sample.
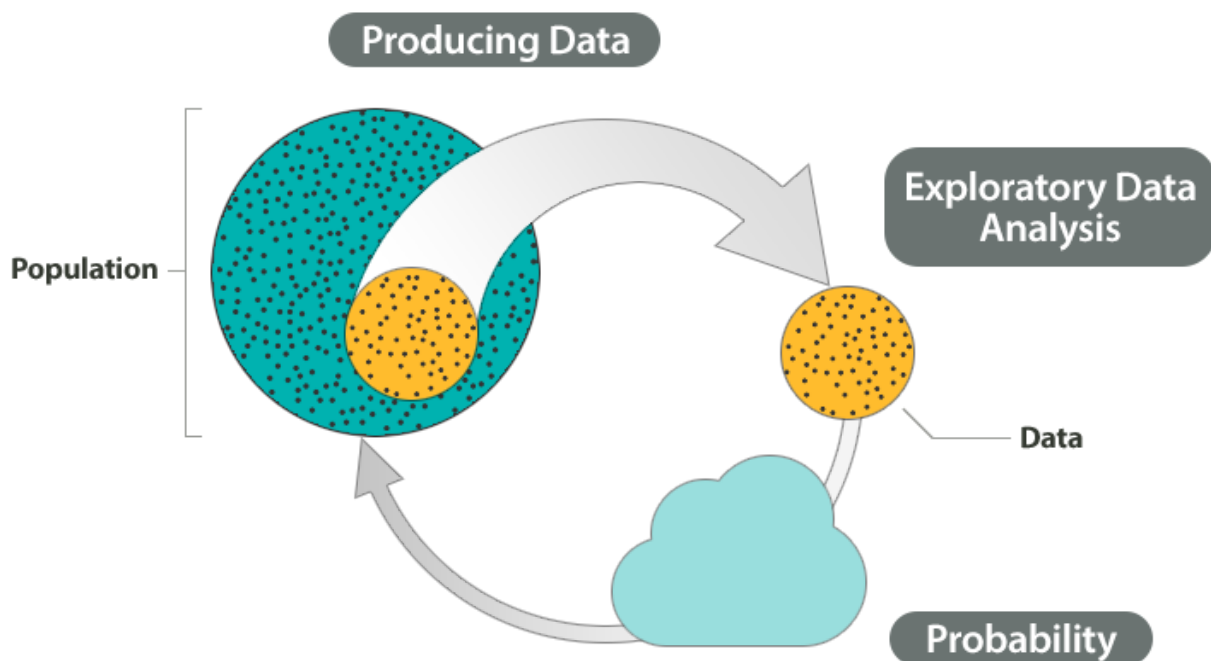
Finally, we can use what we've discovered about our sample to draw conclusions about our population. We call this final step in the process **inference**.
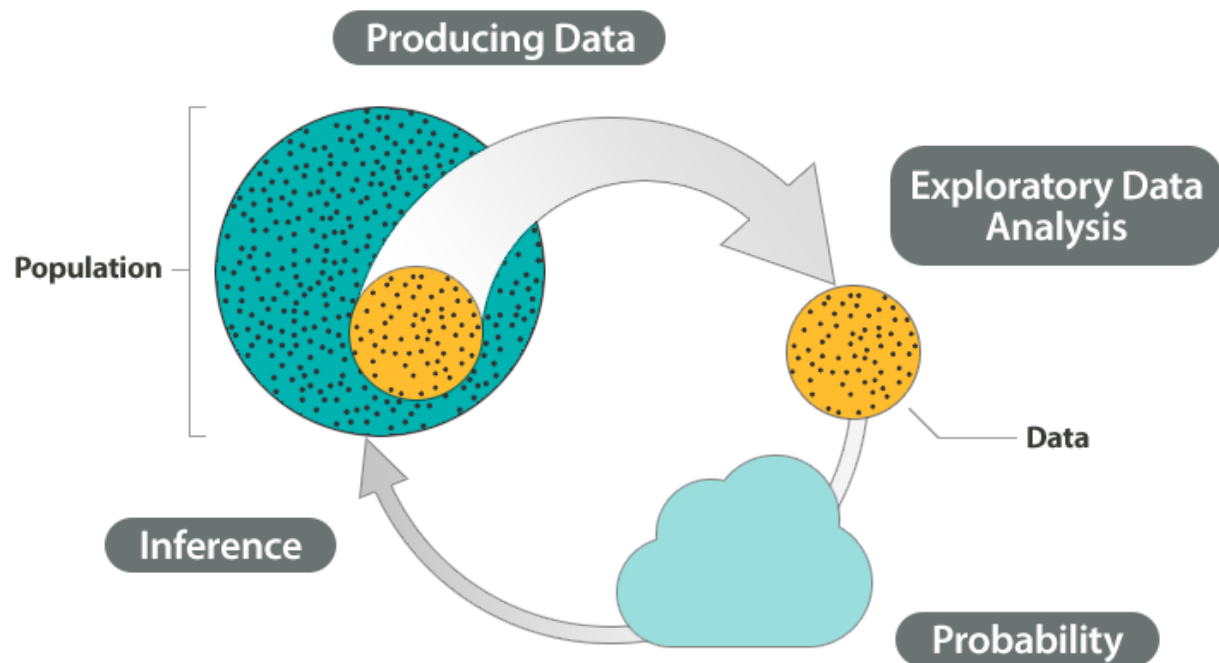
This is the **Big Picture of statistics**.

> ### Example

At the end of April 2005, a poll was conducted (by ABC News and the *Washington Post*) for the purpose of learning the opinions of U.S. adults about the death penalty.

1. Producing Data: A (representative) sample of 1,082 U.S. adults was chosen, and each adult was asked whether he or she favored or opposed the death penalty.

2. Exploratory Data Analysis (EDA): The collected data were summarized, and it was found that 65% of the sampled adults favor the death penalty for persons convicted of murder.

3 and 4. Probability and Inference: Based on the sample result (of 65% favoring the death penalty) and our knowledge of probability, it was concluded (with 95% confidence) that the percentage of those who favor the death penalty in the population is within 3% of what was obtained in the sample (i.e., between 62% and 68%). The following figure summarizes the example:
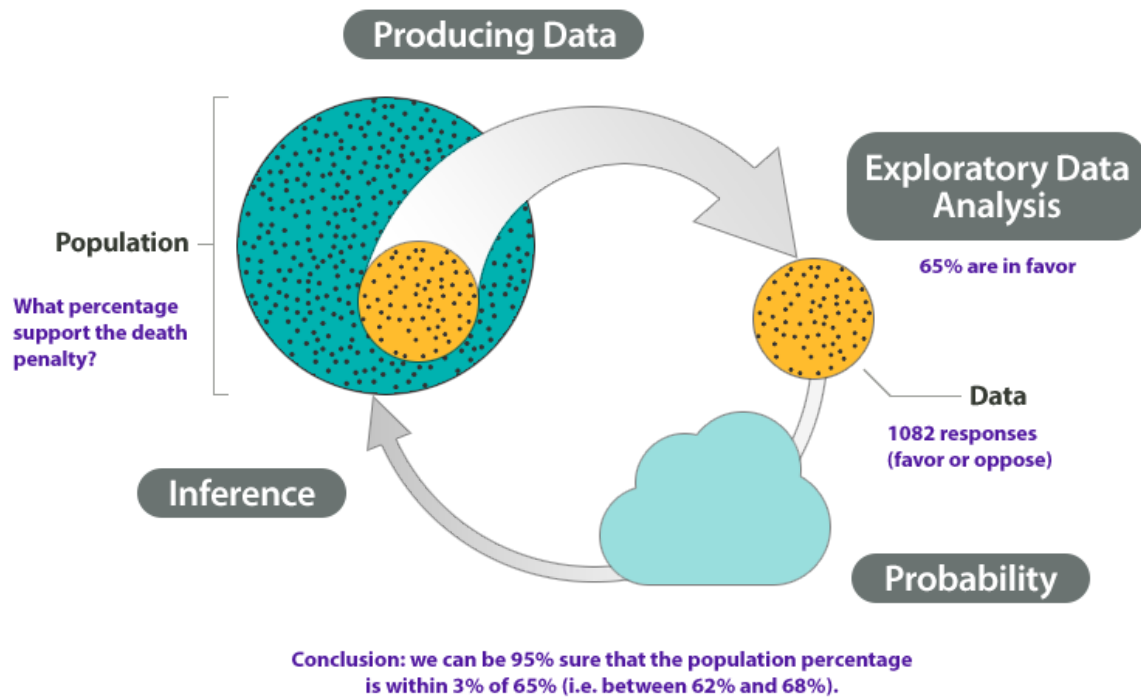
Course Structure

The structure of this entire course is based on the Big Picture. The course has one section, divided into subsections, for each of the steps in the Big Picture. As the figure below shows, we start this course with EDA (even though it is second in the process of statistics), continue to discuss producing data, then proceed to probability, so that at the end we'll be able to discuss inference. The following figure summarizes the structure of the course.
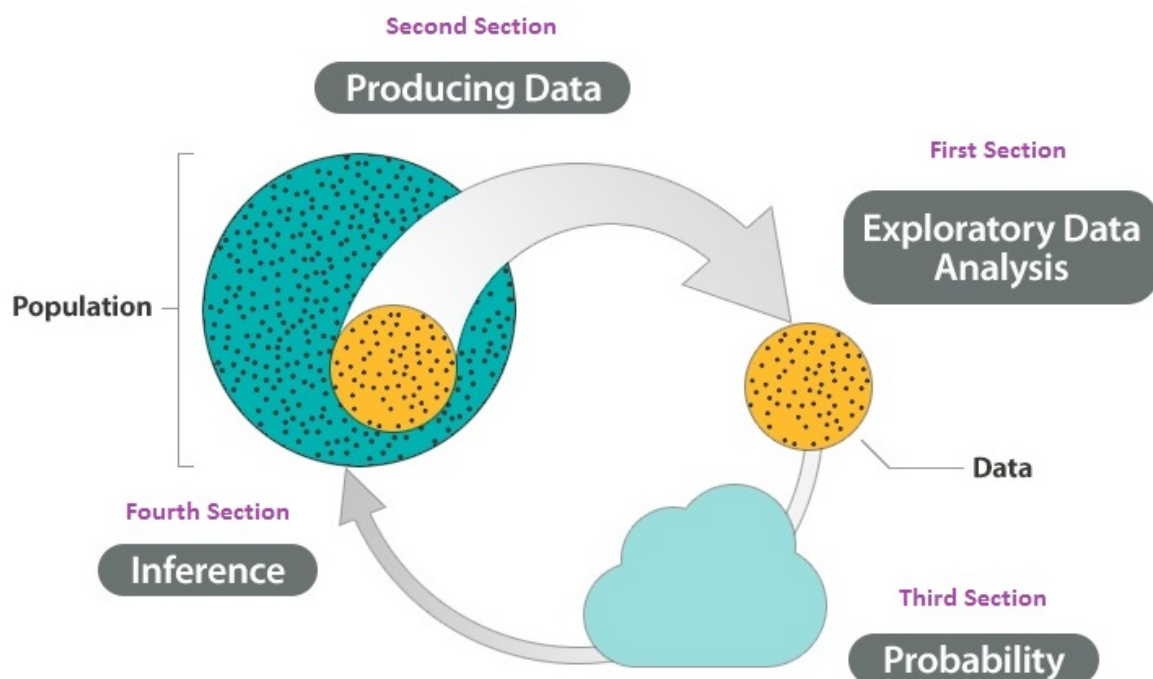
As you'll see, the Big Picture is the basis upon which the entire course is built, both conceptually and structurally. We will refer to it often, and having it in mind will help you as you go through the course.

Open Learning Initiative 🔗