

## Module 3 - The Big Picture

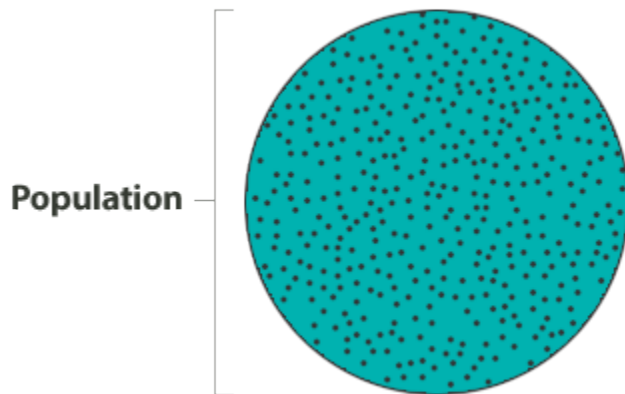
We organized this course around the Big Picture of Statistics. As we learn new material, we will always look at how these new ideas relate to the Big Picture. In this way the Big Picture is a diagram that will help us organize and understand the material we will learn throughout the course.

The Big Picture summarizes the steps in a statistical investigation.

We begin a statistical investigation with a research question. The research question is frequently something we want to know about a *population*. The population can be people or other things, such as animals or objects. For example, we might want to know the answer to questions such as:

- What percentage of U.S. adults supports the death penalty? (Population: U.S. adults)
- Do cell phones affect bees? (Population: bees)
- Do cars get better gas mileage with a new gasoline additive? (Population: cars)

The population is the entire group that we want to know something about:



In most cases, the population is a large group. Often, the population is so large that we cannot collect information from every individual in the population. So we select a *sample* from the population. Then we collect data from this sample. This is the first step in the statistical investigation. We call this step *producing data*.



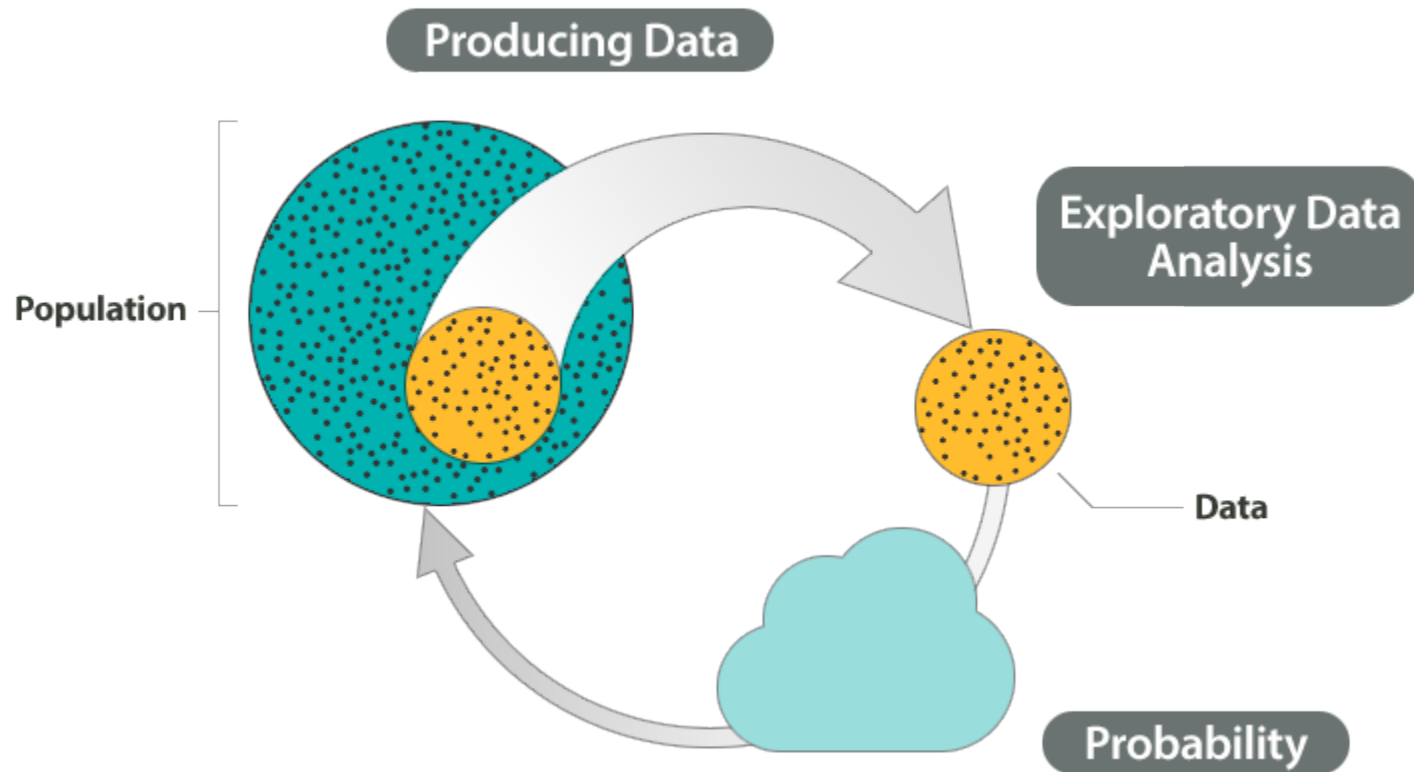
Of course, we need a sample that represents the population well. This involves careful planning but also involves chance. For example, if our goal is to determine the percentage of U.S. adults who favor the death penalty, we do not want our sample to contain only Democrats or only Republicans. So we can give everyone the same opportunity to be in the sample, but we will let chance select the sample.

At this step of the investigation we also carefully define what kind of information we plan to gather. Then we collect the data.

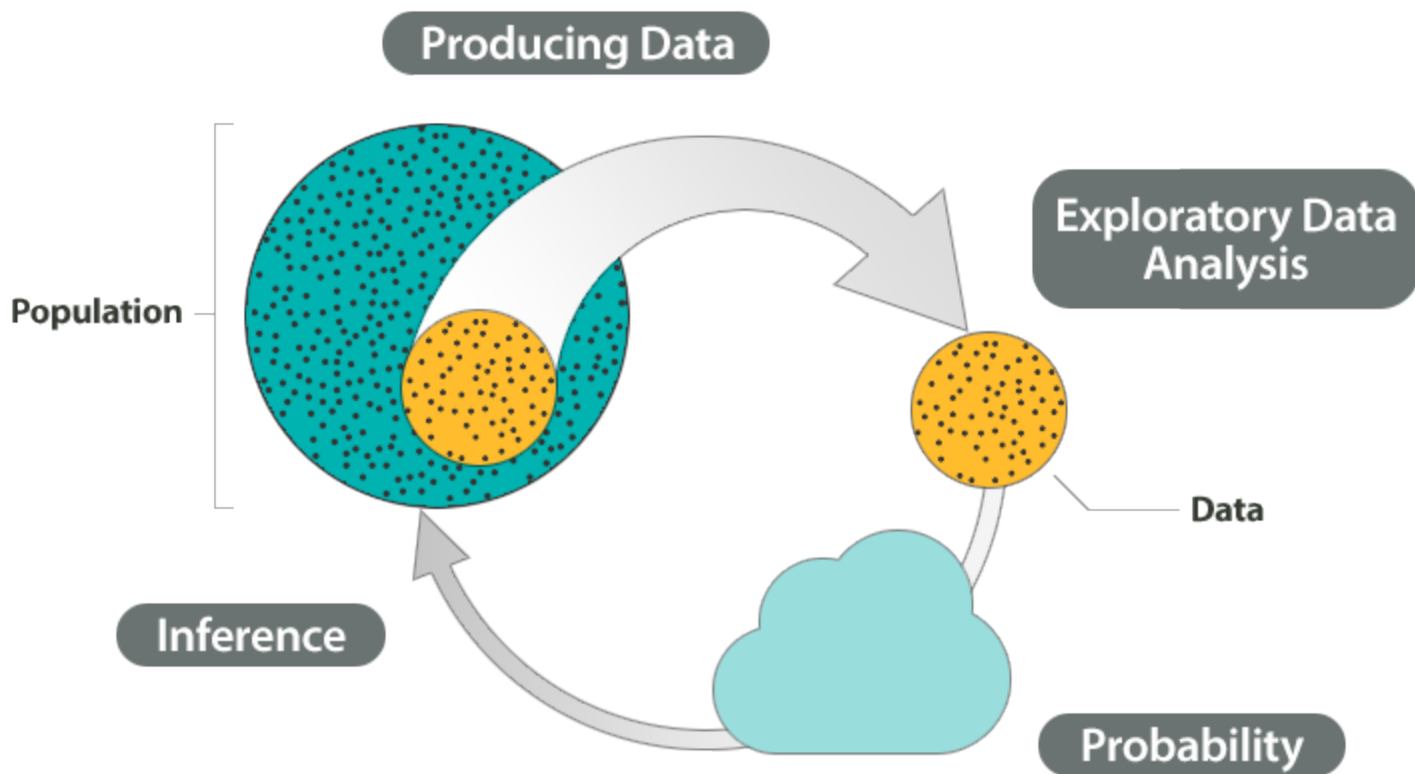
Data is often a long list of information. To make sense of the data, we explore it and summarize it using graphs and different numerical measures, such as percentages or averages. We call this step *exploratory data analysis*.



Remember, our goal is to answer a question about a population based on a sample. Of course, samples will vary due to chance, and we will need to answer our question in spite of this variability. So we need to understand how sample results will vary and how sample results relate to the population as a whole when chance is involved. This is where *probability* comes in.



Probability is the “machinery” behind the last step in the process called *inference*. We infer something about a population based on a sample. This inference is the conclusion we reach from our sample data that answers our original question about the population.



## EXAMPLE

### The Big Picture of Statistics

At the end of April 2005, ABC News and the *Washington Post* conducted a poll to determine the percentage of U.S. adults who support the death penalty.

*Research question:* What percentage of U.S. adults support the death penalty?

Steps in the statistical investigation:

1. *Produce Data:* Determine what to measure, then collect the data.

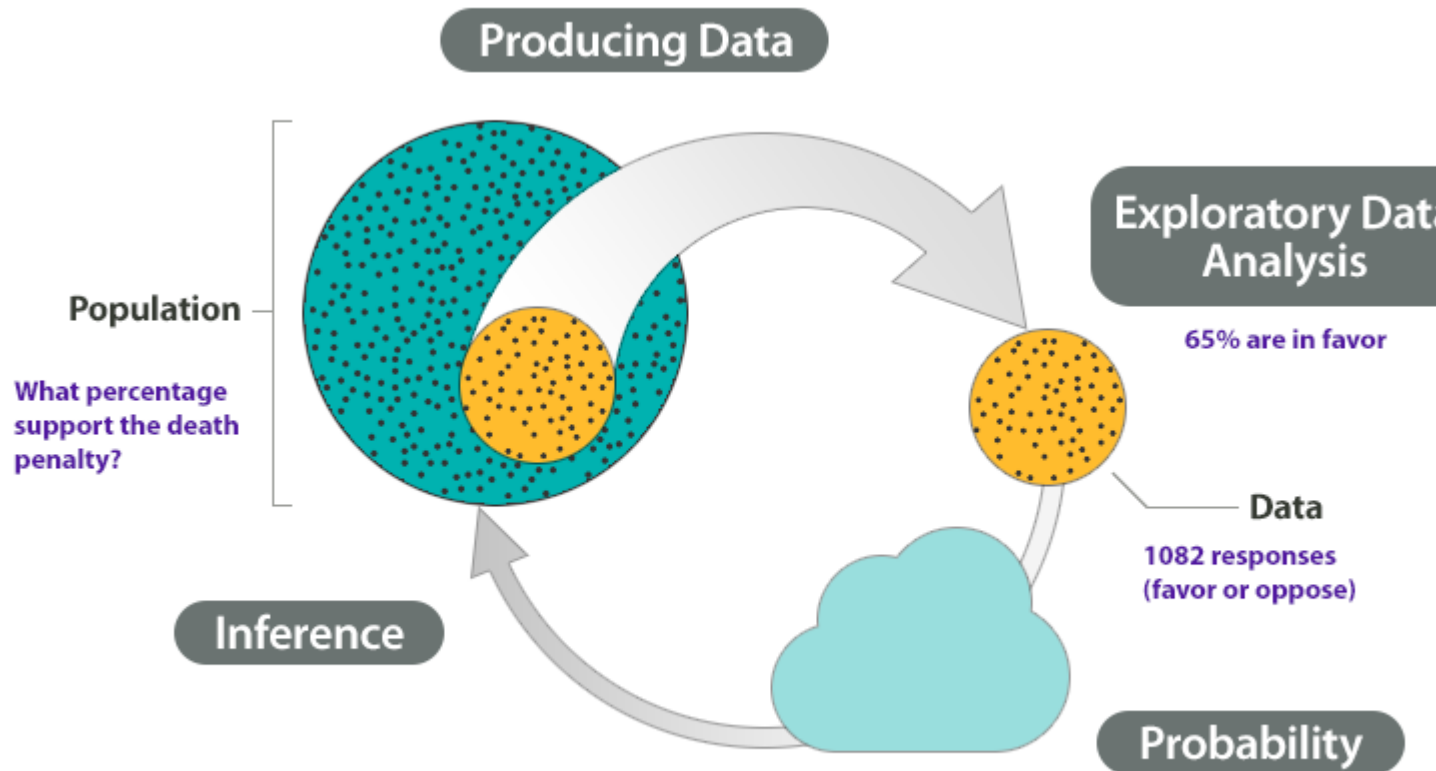
The poll selected 1,082 U.S. adults at random. Each adult answered this question: “Do you favor or oppose the death penalty for a person convicted of murder?”

2. *Explore the Data:* Analyze and summarize the data.

In the sample, 65% favored the death penalty.

3. *Draw a Conclusion:* Use the data, probability, and statistical inference to draw a conclusion about the population.

Our goal is to determine the percentage of the U.S. adult population that supports the death penalty. We know that different samples will give different results. What are the chances that our sample reflects the opinions of the population within 3%? Probability describes the likelihood that our sample is this accurate. So we can say with 95% confidence that between 62% and 68% of the population favor the death penalty.



**Conclusion: We can be 95% sure that the population percentage is between 62% and 68%.**

## Let's summarize:

A statistical investigation begins with a research question. Then the investigation proceeds with the following steps:

- Produce Data: Determine what to measure then collect the data
- Explore the Data: Analyze and summarize the data (also called exploratory data analysis),
- Draw a Conclusion: Use the data, probability, and statistical inference to draw a conclusion about the population.

In this course, we discuss these steps in this order. In reality, a statistical investigation may not be this orderly. For example, an initial exploration of the data may reveal a more interesting question that we decide to investigate. This change in the research question may require a

The following material comes from *Concepts in Statistics* written by the Open Learning Initiative (OLI) in conjunction with Carnegie Mellon and Stanford Universities (<http://oli.cmu.edu>)

different plan for collecting data. So we restart the investigation before we draw a conclusion. Or, after exploring the data, we might realize that the data does not meet the requirements for a statistical inference procedure we want to use. So we go back and collect a new set of data.

Content by the [Open Learning Initiative](#) and licensed under [CC BY](#).