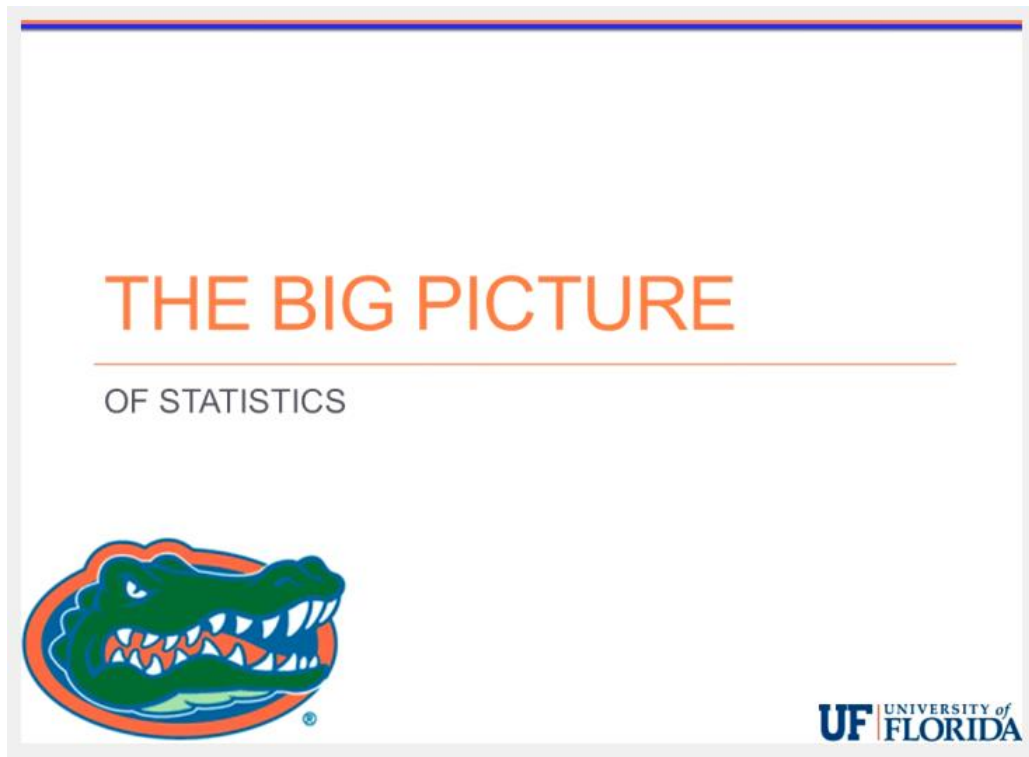


## Transcript: The Big Picture



How does statistics work?

In particular, we are interested in understanding the process of statistical inference where we use the information from one sample to draw conclusions about an entire population.

To begin to address this question, and to set the stage for this course, we introduce the "Big Picture of Statistics" which illustrates the theoretical process of statistical inference.

## The Big Picture



Population: the entire group that is the target of our interest  
Usually too large to study all subjects in the population

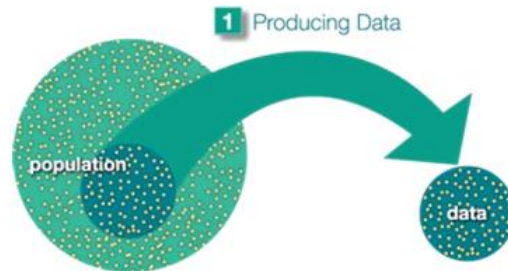
We always begin with the population, which is the entire group that is the target of our interest.

In the health sciences, often our entire population will be something like "All U.S. Adults" or "All U.S. Adults with a particular condition."

However, we are not restricted to people, we could have animals, plants, objects, and so on.

Usually, the population is too large to study all subjects.

## The Big Picture



Producing Data: choosing a sample and collecting data from it

**Choices made here can greatly impact all remaining steps**

The solution in statistics is our first component of the big picture - to obtain a representative sample from that population - which, in the big picture, we label Producing Data.

As we mentioned in the steps of a research project, the choices we make in planning, including how we choose our sample, can greatly impact what we will be able to do with our data later on.

## The Big Picture

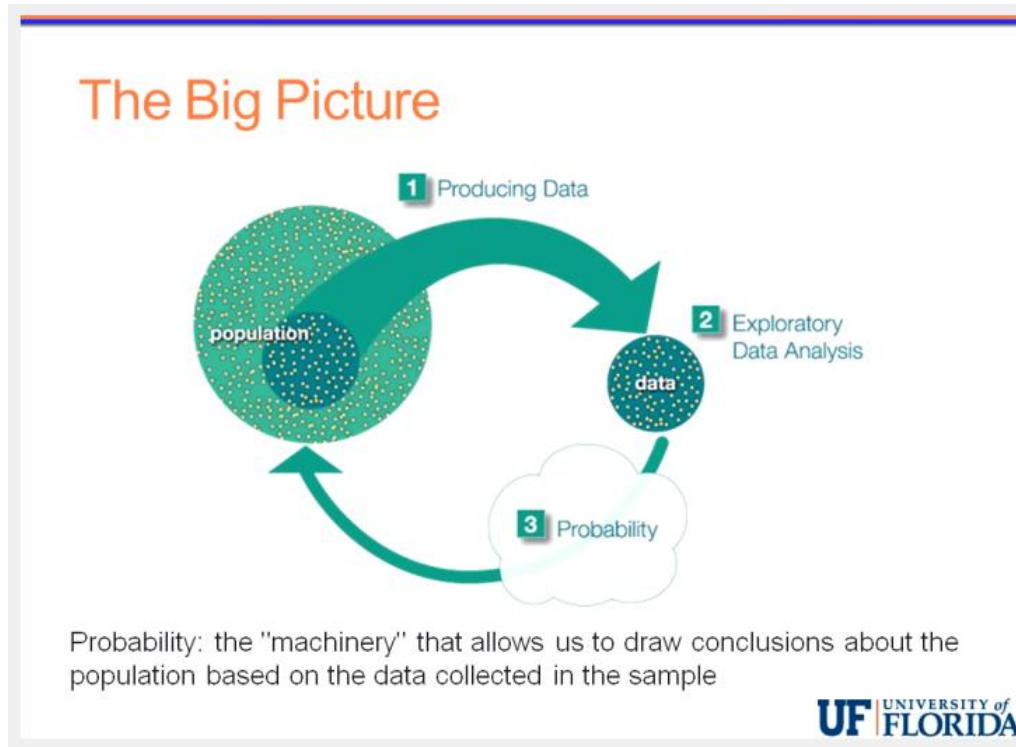


Exploratory Data Analysis: summarizing the collected data numerically/visually

Once we have our sample and have gathered our data, the second component of the big picture is Exploratory Data Analysis, which we often abbreviate, EDA.

Exploratory Data Analysis is also often called Descriptive Statistics.

It consists of summarizing our data with numeric measures and visual or graphical displays.



Now we have summarized the results from our sample, but we are not done!  
Our goal is to study the population.

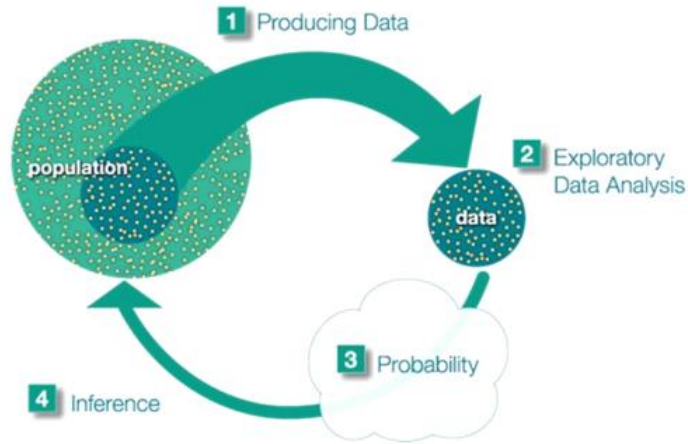
In order to be able to draw conclusions about the population, we need to look at how the sample we are using may differ from the population.

This third component is Probability, which is the "machinery" that allows us to draw conclusions about the population based upon the data in only our single sample.

The "cloud" is a good visual representation of how many student feel about probability!

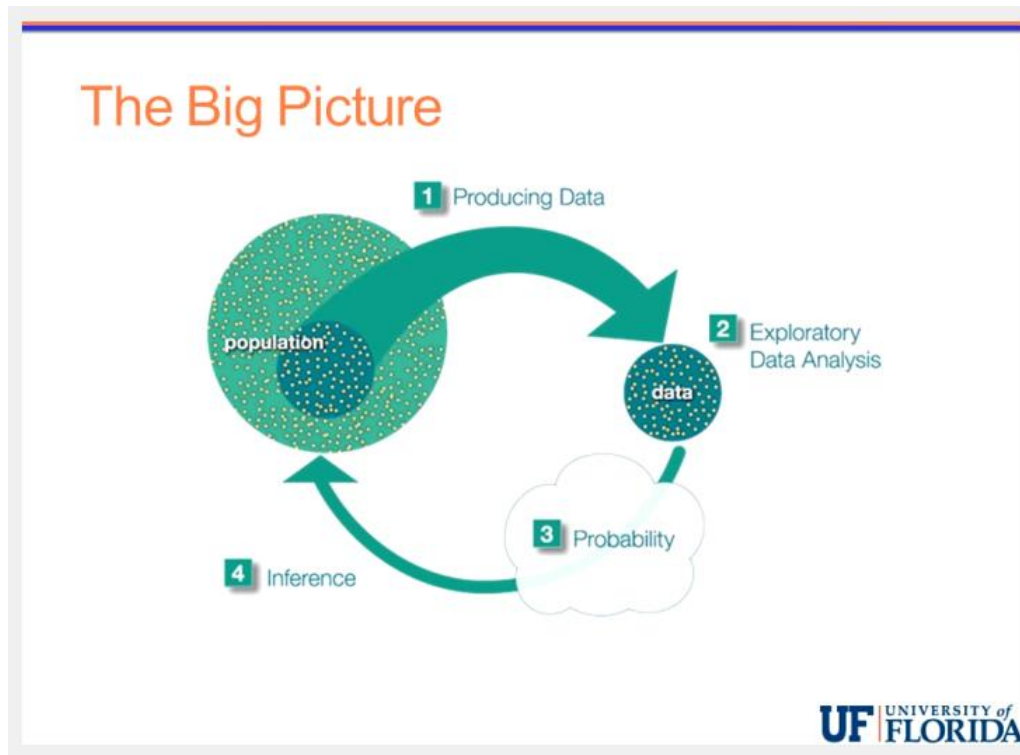
You don't YET need to understand how this requires probability, but one of our goals is that, by the end of the course, you have a solid understanding of the basics of this theoretical part of the process of the "big picture of statistics."

# The Big Picture



Inference: use what we've discovered about our sample to draw conclusions about our population

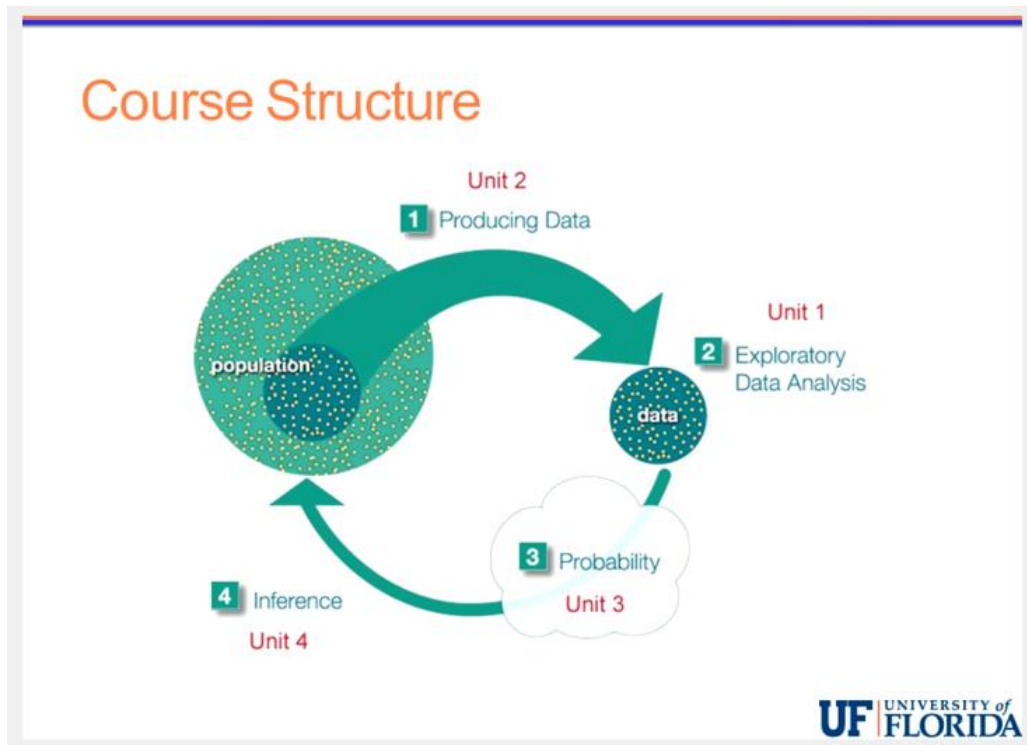
The final component is Inference, where we combine the results from Exploratory Data Analysis and Probability to draw conclusions about our population based upon only the information contained in our one single sample.



So that is the big picture of statistics, with four components: Producing Data (which includes the concepts of the population and a sample), Exploratory Data Analysis, Probability, and Inference.

How much you understand of these ideas now greatly depends on your previous experience with statistics.

It will take a good amount of the semester before we have completely outlined this "big picture" so you should keep it in mind and know where we are in this process as you proceed in the course.



Let's talk briefly about the course structure as it relates to the big picture.

Although in the big picture, Producing Data is first and Exploratory Data Analysis second, we will cover Exploratory Data Analysis first. Some reasons are:

We need to understand something of what we will do with data before we can discuss the issues related to how best to produce data in practice.

We can also cover the concepts of exploratory data analysis early and have time to hone the significant amount of required software skills related to exploratory data analysis while at the same time covering the concepts of producing data and probability which require little or no software skills.

After we have covered exploratory data analysis and then producing data, we will discuss the required theoretical background from probability which is needed to proceed, with a good foundation, into the concepts related to statistical inference, which is the practical endpoint we are aiming for!

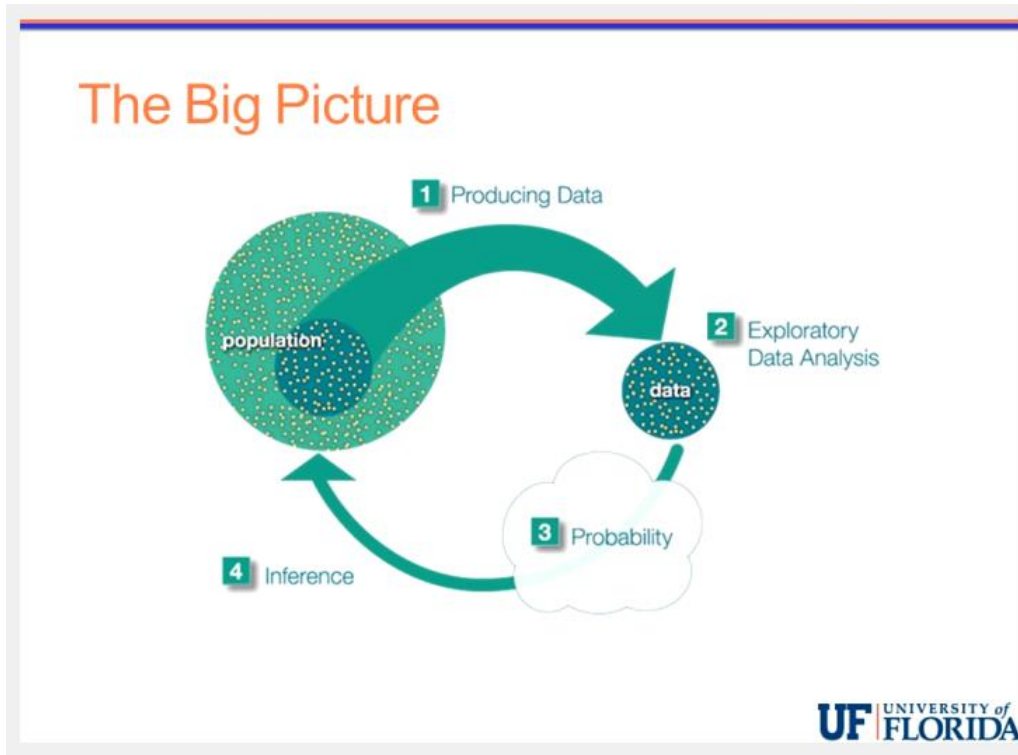


I would like to point out at the beginning that many students find probability to be frustrating. There are numerous reasons for this and they vary depending on the student.

In this course, we will be interested in covering some basic probability concepts and introducing you to the theoretical considerations which are important in statistical inference.

We are NOT however, interested, in learning everything there is to know about probability!

This sometimes leaves students with many unanswered questions which we will not have time to address in this course. Just try to keep this in mind when we get there!



So, that is the big picture of statistics and an overview of how it ties into the course structure.

Many of you already have some desire to be able to carry out a research project from beginning to end and to inform real-world decisions based upon your work. In which case, a full understanding of this big picture is a necessity.