Okay we're back and we're going to continue with the Lesson A notes. And starting out here it's going to be a pretty definition-heavy set of lessons. But that's just to get the ball rolling here. So let's just go ahead and continue.

The first thing we want to do is try and identify the different pieces of gathering data. And so one of the terms, and there are various terms for the different facets of data collection, but for now what we're going to do is we're going to stick to just a certain set of definitions.

So first thing we want to know is that when we are looking at individuals that are being studied, so when we care about, that we call those experimental units. And experimental units can really be anything. It's just what we're actually studying. And individuals sometimes have implication of people. But when we say experimental units it's just what are we actually testing on. Now when we have people that is not uncommon to hear them referred to as subjects. So, it's a relatively common term when we're dealing with people, but experimental units or individuals are essentially just kind of a general term.

Now there are two kinds of variables. And so two sets of characteristics that we want to measure. The first one is what we call a response variable. And a response variable is really what we want to measure on each experimental units. So it's what we're interested in when we're performing the study. Now when we talk about the other variable we call that an explanatory variable.

Now there are actually various terms to describe the explanatory variable. Sometimes they are called predictor variables or response explanatory. You might see independent and dependent. Well I guess it would be dependent and an independent respectively. But the explanatory variable is what we use to explain changes in response. So when I see different responses from different experimental units, what can I use to explain that? What sort of features can I use to explain that difference. Sometimes these are also called treatments. So I guess I should say explanatory variable or variables. We'll get to multiple explanatory variables towards the end of the [AP] course. So it's going to be a while till we see that. But we could have more than one explanatory variable. We could have more than one response variable. That's getting a little bit far into the material, so we won't go quite that far to multiple response variables. But that is something you can continue at the end of this. But again, were way, way, way off in the future there.

Okay, let's take a look at an example and try and pick out these different pieces. So let's start here with Example two. And let's suppose that scientists want to determine which of three types of seed will result in the largest wheat yield. So here's the study that we will follow through. And we're going to prepare three identically sized plots of land. And we're going to make sure soil types are similar. I will talk about that here in a moment why we want that. And we're going to do- each plot will have its own seed and we'll choose plots at random. And we'll talk about that in a minute. We want to water and fertilize the plots in the same way. And then we're going to harvest the wheat and measure the grown amounts for each plot. And at the end, if one seed type produces substantially more or less than the others, then we can start identifying which one is better or which one's the best.

So in this case, what are our experimental units? What are we actually performing the test on? Well, the experimental units in this case are actually the plots of land. So-so the plots of land in which we actually plant the seeds-those are experiments units because those are the individuals that we're studying. And you might have initially think that the seed type is the experimental units, but the explanatory variable is actually the type of seed. And our response variable is going to be the amount grown. So, it feels a little strange that I'm identifying them in this way. Lets kind of keep in mind, what are we actually want to measure? We want to measure the amount grown. So this basically tells me that it's my response variable. And then we want to say okay I see different amounts grown between these different plots and what is explaining that? And so we're going to hope that the type of seed is what will explain that.

Alright, so this is actually one type of study that is called an experiment. So. there are actually two different kinds of studies. Experiment is just one in which the person who's doing the investigation assigns the levels of explanatory variable to the experimental external units at random. So, as we come back up here and take a quick look. [with all these in view] We are noting that we are going to take [sorry about that] We're going to choose the plots at random. So, we're assigning the seed to the experimental units randomly. And from that we measure the amount grown-our response variable. So the experiment is is what is being done here, or an experiment I should say is what's been done here, mostly because we are choosing which plot gets which seed type. And when I say choosing, I mean we're assigning it to the various plots.

Now we'll contrast that with an observational study. And this is a case where we don't actually assign the different levels of explanatory variable to the groups. It's just what we observe. So an example of this, or that we could think of as an example, is how does the color of a car effect it's the drivers tendency to stop at a four-way stop. So, so that's actually a question we can ask. So, so how does the car color influence the driver's tendency [Sorry about this] to stop we'll say at a four-way stop. So, the issue here is that the experimental units, in this case, are the individuals {that] so the drivers and my response variable is going to be whether they stop or not. Stop or don't stop. [It's hard to not break into "Don't Stop Believing" {song} there.] And so my explanatory variable- variable in this case is the car color.

Now why is this not an experiment? Well because I can't assign people's car color to them. What I would do is just maybe sit at a four-way stop and note how many actually come to a complete stop, at the end of, you know, approaching the- approaching the stop sign. So because I'm not actually assigning the color of the car to the people being studied this would be considered an observational study. By contrast, an experiment is a case in which I would [I'm sorry] a study in which I would assign the different levels, the different treatments of, the explanatory variable, or treatment levels of the explanatory variable, to the experimental units. And I would want to do so randomly, and that's something we'll hit in the next video just to kind of close that idea. So we will pick up in the last video for Lesson A.