In this segment, we want to reinforce the message that how you choose to sample can give different results, and the choice of sampling is important.

Suppose you're interested in measuring the average family size in some population.

Suppose that there are families that are small and have just one person in them, and there's also a family that has many people in it.

What does it mean to measure the average family size?

One possibility is to pick at random a family, each family being chosen with equal probability, and talk about the expected value that you get, or the average value if you sample that way.

In this particular example with probability 1/4, you get a 1, with probability 1/4, you get a 1 with probability 1/4, you get a 1.

So the answer would be with probability 3/4 you get a 1, and with probability 1/4, you get a 6.

But suppose that instead, you pick a person at random and you ask for that person, how big is your family?

What's the expected value of the answer you're going to get?

Here we have nine people.

Out of those nine people, 3 of them will give you an answer my family has size one, and six of those people will give you an answer, my family has a size of six.

And this number is going to be larger than the previous number.

You're going to get different answers.

So it is possible to have a situation where you can make a statement such as the following.

The average family size is three, but the average person lives in a family of size four.

There is no contradiction between these two statements because we're measuring different things.

Another example of the same flavor.

You're interested in the average bus occupancy.

You're interested in whether buses are crowded or not in your city.

One way of carrying out this calculation is to pick buses at random, each bus is equally likely to be picked, and see how many people are riding this bus.

Another possibility is to take a typical passenger, a random passenger, and ask them, how crowded was your bus?

Take an extreme case.

Suppose that half of the buses have 0 people in them, and half of the buses have 50 people in them.

If you look at random buses, then the average occupancy would be 25.

But if you ask passengers, all of the passengers would report 50, and it would be, again, a different answer.

A similar situation is if you're talking about average class sizes.

One method is to look at all the classes, see how many students there are in each class, and take the average.

Another method would be to ask a typical student, how large is your class?

Because more students are in large classes, when you pick a student at random, you are likely to get a higher answer, as opposed to when you look at a random class.

The moral from all these examples is that it is very important to be careful about what you choose to sample.

When you pick at random, what exactly are you picking at random?

And you need to be aware that different sampling methods measure different things, and will generally give you different results.