MITOCW | Investigation 3, Part 8

PROFESSOR: We've noticed a couple of things. Let me ask you a question-- in each one of these graphs, what is the total number of people represented? What was that, Steve?

AUDIENCE: 29?

PROFESSOR: 29. If I look here, I had 16 plus 13, that's 29-- or no? 16 plus 13, yeah, that's 29. If I look at all of these, I have 2 plus 2 plus 1 plus 1 plus 1 plus 1. You're all representing the same number of people, you're just representing it differently.

Here, it's hard to tell. Which of these graphs can you easily tell what's the average of the number of calories eaten?

AUDIENCE: 400.

AUDIENCE: 400.

PROFESSOR: Bin width the 400. Why do you say that, Chris?

AUDIENCE: Because there's a big peak right there.

PROFESSOR: OK. If we look at 400, it kind of goes down and up and then back down. If we look at 1,600 bin width, it's kind of like it drops off a little bit. But if we look at bin width 100, it's like up, down, down, down, up, down, up, down, up, down a little bit, so it's harder to tell.

Where did most of the people have their calories? Because up here where most of the people had their calories, it's like three or four bins, so you're splitting the people out. Whereas here in the middle, you've got a wider bin. It's easy to compare nine to one. It's kind of hard to compare two to one, or two to zero.

So in some cases you want lots of people in each bin. In some cases you want essentially lots of resolution for telling the difference between what people consumed. And in some cases you want things in the middle so you can see some patterns a little bit better.

Picking the bin size for a histogram is kind of an art form. It depends on what you're looking for, because there's different ways to represent the exact same data, so it depends on what you're looking for. So we're actually going to think now about how could we represent light, or how could we represent measurements of light, with histograms.