

module_stat1_lecture7

Fri, 1/21 12:23PM 14:51

SUMMARY KEYWORDS

pie chart, crimes, corresponds, violent crimes, data, nova scotia, pie, newfoundland, variable, labrador, provinces, occurs, type, compare, property crimes, excel, pine trees, traffic violations, different types, bar graph

SPEAKERS

Sumon Majumdar



Sumon Majumdar 00:04

So so far, we've been talking about ways of visually presenting data for a qualitative variable. So let me remind you again, what a qualitative or categorical variable is. So it's a variable which comes in a particular category. For example, it could be the type of trip, it could be the vaccination status of someone who has tested positive for COVID, it could be the, the major of university students, or religion of some, right. So these are all examples of variables, which occur in pre specified categories. And once we have the data, what we look at is the number of each type, the number of pine trees, the number of spruce trees, the number of hemlock trees, and so on. And we have talked about essentially two ways of presenting that data visually. So one is using what's called a bar graph or a column chart. So either you use horizontal bars, or vertical bars to represent the frequency of each time. So for example, here, so this column here tells you how many pine trees are there in your lot of 30 trees, how many cedar trees are there in your lot of 30? Trees? Now, if you're interested in like, awful lot of 30 trees, right, what's the composition, what's the percentages, which has the which occurs the lake which has a higher relative fraction, compared to the others, that's when you use what's called a pie chart. And the idea of the pie chart is that you have an entire pie, and the size of the slice shows you which occurs relatively more compared to others. So in this particular pie chart, the light blue one, this is has the biggest slice of the pie, right, so the light blue corresponds to pie. So you see that pine trees occur relatively most frequently in this forest, then comes the size of the yellow one, which is spruce. So that means spruce is the is the next common type of tree here. And then, then comes the orange one, which corresponds to fir and then comes the gray and the dark blue, which corresponds to Hemlock and cedar. So what the pie chart tells you is relative composition of each time. So what I want to do next is talk about, I want to talk a little bit more about the pie chart. And now I want to talk about what if we had so so far, we have just one variable. But what if we had two or more qualitative variables? How would we represent that? So that's what we're going to cover in this clip, and the following, and following. So I've now gone back to the Excel spreadsheet that I had. And remember earlier, we, we dealt with an example of crime data 2020 crime rates in the different Canadian provinces, and there we covered only the violent crimes 400,000 residents, and we plotted them using this bar graph. So now what I'm going to add to that is other types of crimes. Okay, so let me go to the sheet which has the total crime rates. So in this case, what I've added is apart from the violent

crimes, I've also added property crimes, things like thefts, break and enter things like that, but then I have other crimes with that, which are things like disturbing the peace, pornography and things like that. And the last one is traffic violations. So now what we have is more than one qualitative variable right, we have this is one, this is another, this is another and this is a fourth and we have this by provinces. So now suppose let's focus on one province. Let's Newfoundland and Labrador, right. And what we want to know is among the crimes in Newfoundland and Labrador, right, what's so which, which is the relatively most common crime in Newfoundland, Labrador, and so on, which is the next common and so on. So for that, so the ideal tool to use is a pie chart. So let's, let's do that. So what we're going to do is insert, and let's go to the pie chart. So let's do a 3d pie. So this is the 3d pie of the different types of crimes in Newfoundland. But so what it immediately shows you that the most common type of crime, or the biggest slice of the pie is the orange one, which corresponds to property crimes. Then the next common are the gray and the blue, which are other crimes and violent crimes, and traffic violations. This corresponds to the the least amount of crimes, least common type of crimes. And so that's what your pie chart gives you an immediate visual idea of. Now, Excel is great at giving you many different options for this pie chart, right? So if you go up there, and you can choose a different style, right, so here is a style, which actually gives you also the percentages of each. You can have the word without the percentage, you can have it with percentage, you can have the person, you can have it by the names of these different types of crimes. Without percentages, it really depends on your aesthetic and what you're trying to portray. So let's, let's choose this. So again, the the pie chart, right, so what it does is again, the whole pie is 100%. And each slice shows you how big portion of the pie is constituted by that particular type of crime here, or if we're looking at the forest, that particular type of tree. Now the second thing is that see, here we have data on the various provinces as well. So what if we wanted to compare, let's say, between Newfoundland and Labrador, and another Atlantic province, let's say Nova Scotia so now, so let's try to draw a pie chart for Nova Scotia. So I've selected the data here, again, I'm going to go to insert, and I am going to insert 3d pie for Nova Scotia. So one problem is that this new pie chart that has been constructed by Excel right on top of my pie chart for Newfoundland and Labrador, so I can't see the two of them. So what I'm going to do is just move this to the side helps. So let me just move this to the side and make this a little bit smaller, so that I can see both of them at the same. Now you can see, you can get a rough idea that if you compare Nova Scotia with Newfoundland and Labrador, roughly, again, both of them the orange one which corresponds to property crimes, that constitutes the biggest type of crime, right. And again, in Nova Scotia, like Newfoundland and Labrador, the blue and the gray, they're almost equal and, and yellow is the is the least. And again, if you want to get the percentages as well, right, what you can do is, is choose a different style. So one problem with the Nova Scotia one is that if you look at if you look at the Newfoundland and Labrador, it exactly tells you what the different types of the colors correspond to. In the case of Nova Scotia, it doesn't it just says 123 and four and the reason is that see what when we did Newfoundland and Labrador we we selected, we selected this row and this row, right? So from there, Excel immediately understood that violent crimes, property crimes, other crimes, traffic violations, those were the headers. But when we chose for Nova Scotia, which was this, right, it did, I didn't understand that this corresponds to this. This corresponds to property crimes, other crimes, traffic violations, and so on. Right. So what we have to go in and tell, tell this pie chart here that these corresponds to what's depicted in row three. So how do we do that? So we go to the sorry, we try to figure out where the data is for this. So here, in the last one, it says The Select Data. So when you go to select data, so see here, it says, the horizontal axis labels are 1, 2, 3, and 4, right? You can edit it, you can edit it, and you can tell it that this actually occurs in in B, in B3 to E3. So you can tell it, that the locker from B3 up to go up to E3. And now if I hit OK, right, now, it realizes that the categories are these, right? And it depicts those here. So this is how you can, you can ensure that Excel knows what these things actually correspond to what

variables are we talking about. So this is how you create a pie charts in Excel. And as I said, there are lots of different styles of pie charts. And you can play around with that. And choose one that you like the best. There's also another cool trick that you can do so so suppose you want to highlight the violent crimes, right? Suppose you want to lay focus on violent crimes. So what do you can do is that if you go on to the pie, right, and double click it, right, so this comes up with this, you see, there are three small blue dots which come up here. And now, suppose you, you increase what's called the point explosion, you see what's happening is that this part of the pie is coming out. So sometimes you can use this that if you want to highlight the fact that you want to focus on violent crimes, right. So you can use this trick to hide out highlight one particular slice of the pie. So this is a useful trick that you sometimes see used in pie charts as well. So hope you have got now an understanding of how to debate data, by pie charts in Excel. The problem with pie charts is although they're very nice at showing you how the relative composition of different types of variables are there in the data. Right? The problem with pie charts is that you find it hard to use them to compare between, let's say for comparing between two provinces. Yes, there may be little differences in the sizes of the pie, but they may be hard to make out from the pie chart. So what we'll be doing next is only want to show you that if you want to really compare between two situations. We'll use another version of bar charts and column charts, which helps highlight those differences. So let's stop this here. And then we'll come back and do what's called grouped, grouped and stacked bar and column charts to highlight differences here between different provinces or between different situations in general