

module_stat1_lecture8

Fri, 1/21 12:23PM 16:37

SUMMARY KEYWORDS

crimes, provinces, compare, violent crimes, data, stacked bar chart, newfoundland, type, grouped, property crimes, chart, bars, called, prince edward island, saskatchewan, traffic violations, stacked, clustered column, index, labrador

SPEAKERS

Sumon Majumdar



Sumon Majumdar 00:05

So the last clip, I talked about this data that we had for the different Canadian provinces of different types of crimes. And for a particular province, we could present the data using a pie chart and see which type of crime constitutes the biggest fraction of crimes in that particular province. Now, well, the pie chart is good for that. Now, if we were to compare across the different provinces, it becomes very difficult because that means you'll have, let's say, you want to compare across four provinces, what you will have to do is generate four pies, and then compare which segment is bigger, in which pie and so on. So what I want to show you in this clip is, is for that sort of comparison, right, there is another type of visual representation, another type of column chart, which is really helpful for doing. So that's what I want to show you. So, so here's my data, right? And so and I use the pie chart to compare between, let's say, Newfoundland, and Nova Scotia. But as I said, this is it makes it very difficult if you are to compare across more than two provinces. So what if we wanted to now compare across, let's say, all our four Atlantic provinces Newfoundland, Prince, Edward Island, Nova Scotia, New Brunswick. So what I'm going to do is first let me select the data for these four provinces. And the type of graph that I'm going to use to compare is it's going to be a type of a bar chart, right? So I've gone to insert and looked at bar chart, but what I'm going to do is, is this type what's called a clustered column, or grouped bar chart, grouped columns. So let's hit this right. So see what it does, right. So what it has done is it has created bars for the different types of crimes. Right. So this is for example, violent crimes, and it has one vertical bar for Newfoundland and Labrador. The orange colored is for Prince Edward Island, the green for Nova Scotia, the yellow is for New Brunswick. And what it is done essentially is it is created four bars for the four different provinces one for violent crimes, then it has done the same for property crimes, other crimes have traffic violations. And what this does great is that enables you to compare like suppose you are interested in violent crimes. And, and easily you can see, okay, Newfoundland and Labrador has the most Prince Edward Island has the least and the Nova Scotia, New Brunswick, the green and the yellow, they are almost equal. Right now if you look at property crimes again Newfoundland and Labrador has the highest then comes Prince Edward Island, then Nova Scotia, New Brunswick. So, this type of chart is called a grouped bar chart. So this is called a group bar chart. So, so this is called the grouped bar chart. And this is especially useful if you want to compare right so here we have data on one province, you

have your data on another province data, third province and so on. But if you're to compare across this different this data from this different provinces, this is a great tool to do that. So I'd like to point out one of the crucial aspects of this grouped bar chart is this idea that we have grouped here by the type of crimes. So see we have looked at violent crime And then compared across the four provinces, then we've looked at property crimes and then compare it across the four provinces and so on. So in terms of the data, what we've done is we've looked at violent crimes here, and then compared across the four provinces in this category, then same thing with property crime, compare across the four provinces, and so on. So this is called the grouping variable. So the grouping variable here is the type of crime. So, so far, we have compared across only the four Atlantic provinces, what if we wanted to do now compare across all the provinces, including the three territories. So we can do the same thing. So let's select all the data, go to my Insert, go to my bar charts, and then do this the clustered column. Now, look at what has happened here, right. So this is the chart that it has generated. But what it has done now is that here, it has gone by provinces. So it has looked at Newfoundland and Labrador, and, and done the four categories of crimes, violent property, other crimes and traffic violations, then it has gone to Prince Edward Island and done the same. Now, compared to what we had before, this is not as useful because what we cannot do, like this is what the this has generated is that for each province, we can see the different types of crimes, but we cannot look at just violent crimes and see which province has the most right, or look at property crimes and see which provinces the least. And the reasoning. The reason here is that it is grouped here by provinces, and not by the type of crimes. And this is a little problem with with Excel that if you don't tell it the grouping variable, what it does is that it chooses the one which has the fewest. So the one which, for which it has to create the fewest number of bars, that's what it chooses. But the good thing about Excel is that you can you can change this. So how do you change this? So if I go to the chart design, so what do I want to change is, instead of going by province, I wanted to go by the type of crime, right? So what I'm going to do is just go up here and ask it to switch between row and column. So if I hit switch to row and column, and now see what it has done, it has grouped by the type of crimes, violent crimes and looked across all the provinces, property crimes, look across all the provinces, and what we exactly what we wanted to do before. So this, so let me blow it up a little bit. So I'm just going to move this here and blow it here. So one problem is, again, Excel has done this chart exactly on top of the previous chart. So what you can do is that you can move this below so that you can now see both of these charts. So let me go down here. And now see, this gives you a good idea of comparing for each crime across all the provinces. So for example, if we look at violent crimes, the three bars which really stand out are this dark blue, which is Yukon, the green, which is Northwest Territories, and the third one, Nunavut, which is the light blue. So these have the largest number of violent crimes. So remember, these are not the absolute number of crimes, because these are really small provinces. But what these are crimes per 100,000 residents, right? So this is divided by population. So now the next if you look at violent crimes, the next two are these two here. That's Manitoba. And this one is Saskatchewan, right. Now, similarly, if you look at property crimes, and again, the three territories just stand out, then comes. This one here is Manitoba. The next one here is Saskatchewan. And then comes Alberta and British Columbia. So this again, this is a grouped chart right across the provinces. And this can help you in comparing each type of crime across all the provinces. Now, remember, I said that you could either do this as columns, or you could do this as bars, right? So what happens if we change this instead of column, suppose we change the into bars, so you can try it out. So if so if you try it out, okay. Now see, this doesn't look as good as the, the vertical columns, right. So in this case, the, this horizontal bars don't look as well. So I'm going to go back and undo this. And, and so this is a good chart for comparing across the different provinces. Sorry, comparing crime, for each type of crime comparing across the different provinces. Now, if you go to the Excel sheet, you'll see that I have another, the next sheet here. And this is an exercise for you.

So what I have done is given you three components of what's called the Human Development Index, so this is an index, which is used to compare quality of life across countries. And this index consists of three things. One is an index on education. So if a country has a higher value of this index, that means it has a better quality of education. Next is an income index. And the third is a life expectancy. And what I've done is for each country, I have collected the data for 2018. So what I would like you to do is use this idea of grouped bar charts to compare comparing each of these indices across all the countries. So this will give you help you understand how to actually construct such charts in Excel, because as I've said over and over again, for any type of software, you can only learn how to use it if you actually go in and at use it. Right. So I'm hoping that you will take this opportunity to do this exercise. Now, let me end this module by showing you another type of chart which is sometimes used. And this is called the stacked bar chart. And again, let me go back to my crime data. So this is my crime data here, right across the different Canadian provinces. So let me give select this. And what I would like to do is insert what's called this is a stacked bar, okay. So, this is a stacked bar. Let me first do that. And now tell you what this thing does right. So, what it does is present this data by stacking one on top of the other. So what it does is that so see, if you look at none of it, this first blue portion is violent crimes. The orange portion is property crimes, the gray portion is on the crimes and the yellow portion is traffic violations. And what it has done is it has stacked these bars on top of each other. And it has done this for all of the provinces. This is so this is called a stacked bar chart. So let me write this. So stacked. So this sort of stacked bar chart also sometimes helps in comparison. So for example, it's easy to compare that in terms of violent crimes, right? You look at the size of the blue in the different provinces and you see that these three immediately stand out. Then comes Saskatchewan and Manitoba. And then if to look a little bit more closely then comes British Columbia now. Similarly, if You want to compare across, let's say the gray, the other crimes, right? That's not as easy. But see, even then you can see that these three broad territories, these stand out in terms of the size of their gray bars, right, which is the other crimes, then comes Saskatchewan. And then maybe these three, Manitoba, Alberta, British. So what I've done again, as the name says, this is called a stacked bar chart, and it stacks these bars one on top of and this works fine when you have few things to stack on one another. Right? If you had 10 categories of crimes, then this wouldn't have looked very pretty. But what I wanted to present to you here in these clips is to show you how you can present data visually, this for a categorical variable. And not just if you had only one set of data. But if you had set data from different sets right here we had different provinces, if we have data from different countries, or maybe data across by gender, right, or data across different points in time, how you could use these different either the stacked chart or the group chart to compare across them.