

module_stat1_lecture5

Fri, 1/21 12:23PM 14:06

SUMMARY KEYWORDS

column chart, data, vaccinated, provinces, inference, chart, create, violent crime rate, cases, ontario, frequency distribution, unvaccinated, change, column, statistics, bar chart, choose, vaccination status, fully, crime rate

SPEAKERS

Sumon Majumdar



Sumon Majumdar 00:05

In the last clip I presented to you data about new COVID cases in Ontario with vaccination status. And I asked you to, you produced column chart. And I asked you to think about what inference can you draw from that? Now, if you look at the column chart, like, right off the bat, it seems that the number of cases among unvaccinated and fully vaccinated is almost equal. And from there, very naive inference could be that vaccination doesn't matter, because there are equal number of cases among non vaccinated and fully vaccinated. So what's the problem with that inference? The problem is that remember, at this point in Ontario, there are a lot lot more people who are fully vaccinated than who are unvaccinated. So about 75 to 80% of the population are fully vaccinated in Ontario, and only 20 to 25% are unvaccinated. So this number of cases that you're seeing is among a smaller group of people. And this number of cases that you're seeing on the fully vaccinated is among a much larger group of people. So that's why there's always the warning that Be careful with statistics big, like data doesn't lie, but you have to look at the right variable. So if you want to draw an inference like that, so it would, is not enough just to look at the raw number of cases in Ontario by vaccination status. So what do you really need to know is that among the population who are unvaccinated, right, divide this by the number of people who are unvaccinated? What's the rate of infection? Similarly, if you look at among the vaccinated people, what's the rate of infection among them? So again, I went back to the website. And there they also give the frequency distribution of cases per 100,000 per 100,000 unvaccinated people, how many cases are there. So that's 14.35. So among the partially vaccinated 100,000, partially vaccinated people, what's the rate of being infected with new COVID cases that 6.17 And among the fully vaccinated is 3.91. So they don't have data on how many people are of unknown status. So that eliminated, but what they do have data on is this rate, or cases per 100,000, among unvaccinated partially vaccinated and fully vaccinated. And now, if you look at this data, right, and we can do a column chart using this. And here is what it created. Now you see, in here, it's very clear that the rate of infection among the unvaccinated is the highest rate, and it's quite a far distance off from that among the fully vaccinated. So then comes the partial vaccinated, and the rate of infection among the fully vaccinated, although it's not completely eliminated, but it is much, much less than those who are unvaccinated. So that's why there's this warning. And this is an example of that. Be careful with statistics, be careful about making inferences, you need to

look at the right data variable, you need to make the right inferences. So now, as I promised, in the last clip, I'm going to show you how to create this column charts or bar graphs using Excel. And I'm going to do three examples there. So in two of the examples, the frequency distribution is already given. Right? So for example, here, I've already told you this is the frequencies have the four categories, right? But sometimes you may get raw data. And you may have to tabulate those categories by yourself. So in another example, I'm going to show you how you can create a frequency distribution from data and then how you can use that to actually create a column chart. And I'm also going to show you another related tool which is very useful, especially with categorical or qualitative data, that's called the pie chart, which shows the composition as percentages. So now let me move on to my Excel sheet. And I'll show you how to create these charts. So here is my Excel sheet with my data. So here I have the data on the total number of cases and new cases in Ontario on December 7, right among non vaccinated, partial vaccinated, fully vaccinated and unknown status. So now suppose I want to create a column chart using these right, so first, I select this data. Then I go to insert, like going to the Insert tab at the top. And here, you will see a lot of different types of charts. So I'm going to choose this first one, which says Insert bar or column chart. And see, it gives you lots of different choices. So I'm going to pick the first one, which is going to be a column chart. So here is the column chart and see it has created this. Now, you can move this around, right, so I don't didn't want to obstruct this data here. So I moved it a little bit here. Now, I want to give it a chart title. So again, if I click on that, and that'll give me an option to edit my title, so I'll put in new COVID cases in Ontario on 12 720 21. By vaccination status. So that gives you the chart title. And now see, there are lots of different options on this on this chart. So see, you can change things like here, you can also you can change the colors if you want, you can change the chart style. So if you keep going choose something that you like. So there are lots of different styles. So this is the one I chose. But if you don't like the dark shades of that, you can choose a lighter shade. That's aesthetics. And that's up to you. You can also change the color if you don't like the blue color, choose something else you can choose an orange color or a green color, right choose your favorite color. Now, you can also change this. So so this is how you create a column chart. And instead of this, if you want to change this instead into a, into, let me go here, I want to change this into a bar graph. And so if I chose this, it already shows you how your bar graph is going to look. So let me go back to the column chart, right. So this is how your column chart. Now you can try. There is also this data on cases per 100,000. Right, so you can try this out on your own that try and create a column chart using this data. And that will give you some practice on how to create these things in Excel. Let me do one other example. So let me go to my next sheet. So this is on. This is at the bottom. So if I click here, this is my next sheet. So in this sheet, I have 2020 crime rates in Canadian provinces. So this is Data from Statistics Canada, and what it lists on the left are the provinces. So our provinces and here is the data on violent crimes per 100,000 residents. So this is essentially the crime rate the violent crime rate. So this would be like things like murders firearm, this church sexual assault, so any sort of violent crimes would be under this country. Okay, so we have data across all the provinces, this is 2020. And now suppose I want to create a column or a bar chart using this. So here, right, so if I create a column chart, now you see, and here's the problem that I was talking to, trying, telling you about before, that if you have a lot of categories, right, so here, you have all the all our 10 provinces, right. And all of the names are pretty long. So this doesn't look that one. Right. So in a case like this, what may be better is to use a bar chart, because then these names would be along the Y axis, and they would all be horizontal. So let me change that. So let me go back to insert. And again, I'm going to choose this, right? And instead, let me choose my bar chart. Now you see, it looks much neater, the names of the provinces are horizontal, on down the Y axis. And I can still use this data, this the length of these bars to see, okay, which province has the highest violent crime rate? Okay, that's Nunavut, then comes Northwest Territories, then the Yukon. And these three are head and shoulders above all of the other

provinces means it here, it's not such a nice thing, because it's the violent crime rate, right. But what you can see is that the crime violent crime rate in these three way above that in any of the other provinces. So if we look at the other provinces, then Saskatchewan and Manitoba is high, right. And then comes this group, British Columbia, Alberta, New Brunswick, Nova Scotia, Newfoundland, Labrador, and Ontario, Prince Edward Island art. So this is how you would create a bar chart in Excel. And again, I can put in the title. So So let me put it violent provinces just been. And you can, you can make this look as you wish, you can expand this, you can change the color. And so basically, if you click on the graph, all of these options comes on. And if you want to change the color, you can change it to a different color. Take your pick. But this is how you create column charts and bar charts in Excel. And, again, these depict, are an easy depiction and easy visual depiction of the data and gets you a first look at the data in a pretty nice, convenient and informative way. So that's why bar graphs and column charts these are some of the most common pictures that that you've seen statistics, and it would be useful if you can generate them on your



own