

module_stat1_lecture9

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

SUMMARY KEYWORDS

gdp, plot, data, variable, trends, capita, interested, year, line, cases, chart, type, number, horizontal axis, ontario, visual representation, graph, tool, tree, change

SPEAKERS

Sumon Majumdar

Sumon Majumdar 00:04

So the type of variable we have been focusing on so far for visual representation is what's called qualitative or categorical. So an example was the type of tree. So the data on that is where that particular tree is referred to your pine tree, your hemlock tree, or so on. And what we have been showing in either in terms of the bar charts, column charts, or pie chart, is how many fir trees? How many pine trees, how many Hemlock trees and so on. So now we'll be coming to the other type of variable, which is called quantitative or a numeric variable. So examples of those are what somebody's income, what's the country's GDP per capita? What's the height of a tree? What's the amount of rainfall? So the data here is already numeric. So whether it rained 10.5 millimeters yesterday, or 5.2 millimeters, that whether the country's GDP per capita is \$1,500? Or is it \$10,500? So what we'll now be focusing on is visual representation of data like that. So so the particular visual tool that we'll be using will depend on what's our objective, right? What do we want to represent? What do we want to see? So let's start with an example. So this is something which we had done before. So remember, I talked earlier about data on the number of new COVID cases in Ontario per day. And we are data from December 7 2021. And this was the number of COVID cases on over the past seven days. So for data like this, what what are we interested in seeing? We're not only interested in seeing how many COVID cases are there, we're probably primarily interested in seeing whether that number is increasing or decreasing over time. So what's the trend over time? So for that sort of data, what do we really want to see is the plot over time that whether the number of cases are going up, whether they're coming down or are they stabilizing. So this, so what we're looking for is some kind of graph like this, where so if you look at this, right, so what I have on this side is the dates, right, so this is the time. And on the Y axis, I have the number of cases. And what I've plotted here is the number of cases on on the first of December, the number of cases on the second of December, the third of December, and so on. So this is this is a plot or a time plot of the number of COVID cases over the last seven days. And what it shows us is that and this was in December 7 right. So what it shows us is that this has been going up, it peaked around here, and then it has been coming down. This is the type of trend over the last seven days. Now, this sort of this sort of track of time plot, right, or what we are really doing is establishing the relationship between one variable here, which is the time, right, and the other variable here, which is the number of COVID cases. So what this is, is again, it's a plot between two variables, one side is time. The other variable here is X, which is the number of cases. Now this

sort of time trends are a very common type of data representation that one is often interested in looking at. So what are lots of examples of this. So for example, you could be interested in seeing how GDP of a particular country has gone up or gone down over time. What about population or immigration? Right? And typically, you see this trends over year. by you, right? How has population changed from last year to this year? And how is it going to change tomorrow, there could be other things where you may be interested in looking at over days, months or quarters. So typically things like sales, or profits, or costs, right. So these is what businesses are interested in seeing how these are changing over days, or over months, or over quarters. And all of these are our time trends. So this is an important type of data that you often encounter in the social sciences. And this type of plots that I just showed you, which is called a line plot, you can easily do it using Excel. So in Excel, you can easily do a plot like this with time on one dimension, and whatever is the other variable of interest, whether it's COVID cases or sales or profits on the other axis, you can usually do that by using the command offer line. So let me show that to you. I've now come again to my Excel spreadsheet. And here I have data on new COVID cases over the past 7 days. I have it for Ontario, as well as for Quebec. So what I'm going to first focus on is Ontario. So I'm going to select this data, so date, Ontario. So I have it over the seven days, written the dates here, and the number of cases here. And then again, I'm going to go to Insert. And now see, I'm going to hit this button here, which says Insert line or area chart. So if I go there, I'm going to pick the first one, which is called the line plot or the and this shows trends over time, or categories. So let me just hit that. And this gives you the plot for exactly what I showed you in my slide earlier. And if I move it here and change the style of the graph, that is exactly what I presented to you. So it's as easy as that on the x axis, or the date on the y axis, or the number of cases. Now, I suppose you want to do the same thing, but for both Ontario and Quebec, and maybe you want to compare across them. So I've selected the data, again, I'm going to do exactly the same go here. Choose the line plot. And this gives you the plots for for Ontario in blue, Quebec, in orange. And again, if I move this here, and maybe if you want to change the style, there are lots of different styles. There's a style, which actually also incorporates the numbers, there is one with which shows you these lines, right, you can get one without those lines, you can take your pick, depending on whichever app you use to your aesthetic. So these are easy enough to do. And these, this is a very useful tool, especially for time plot, because here on one axis is has the time and the other axis, it has the numbers. So let's do yet another example. So this is Canadian GDP per capita, measured in 2021. US dollars. And what I have is data on GDP per capita. So again, GDP per capita means GDP per head. So it's the total GDP divided by the population. And I have this for Canada from years 1960 up to up to the present. So let me just, me through these two, because that way, you'll have the headers in your labels of the variables in your chart. So I have the data to 2020. And again, let me go Insert, go to the line chart and plot it. And let me move it to this area here. Now, see what it has done. Its it has done a line plot, all right. But see, on the x axis instead of the particular years, it has just done 1234, and so on. And plotted first the year against that that's a flat line. And then it has plotted the GDP per capita. The reason is that sometimes Excel has difficulty in realizing that this is a year that this is a time variable here, it doesn't know that. So that's why I thought that it needs to plot this separately, and it needs to plot this separate, what you can do is you can go in into the chart and actually change that. So if you go to the third thing here, it's a distance chart filters. And if I go to Select Data, see it has plotted year, and it has plotted GDP per capita. So I don't want the years. So let me take that out. Secondly, what I want is the horizontal axis labels, right? So that's done 1234. Like that, I want to change them, and I want to change it to the years. So what I'm going to do is first let me move this here. Let's do edit. And I wanted to incorporate these years, so I'm going to hit this. So this is L four, two, and I want to go down. And this is my last point. So I want to hit and I do okay. And see now the horizontal axis labels are 1960 1961, and so on. So if I hit OK, and if I go back to my graph here, so see, it has the years on the horizontal axis as I

wanted, and the GDP per capita, here, right, and, and this is something which economists would be interested in. So see, if you look at Canadian GDP per capita, in the 1960s, it was growing, but growing rather slowly. And here, it starts taking off here, it really shoots up, right. And at this point, it drops. And then again, it climbs. And again, it drops. And here, it's growing, but not very much. So this is what economists, a lot of other people would be interested in, right. And you can get this sort of a plot a time plot of GDP per capita, you could do the same thing with population, you could do the same thing with immigration or any other variable which changes over time, you could generate a youth using this line plot. So let me stop this clip here. So I hope you've got the hang of how to generate such a line plot in Excel because this is a very important tool to have in your arsenal. And as I said, you encounter it in a lot of places. For example, if you're running a business, you may be interested in seeing how your sales are changing month to month or week to week or year to year. If you're an epidemiologist, you may be interested in seeing how infections are moving from day to day or week to week, and so on. So let's stop here. And I'll show you other tools for visually representing quantitative variables.