Learning Analytics Tools

Professor Ramkumar Rajendran

Education Technology

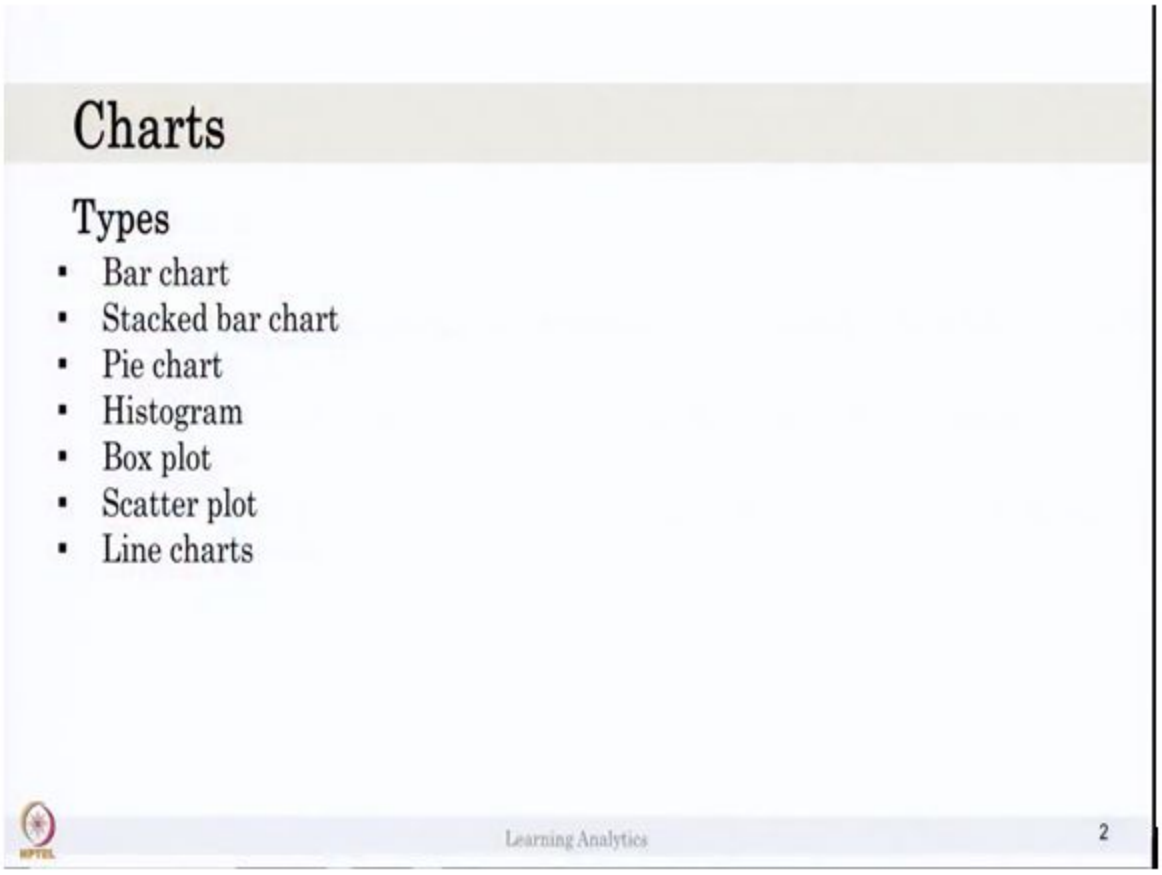
Indian Institute of Technology, Bombay

Lecture no. 4.5

Charts – III

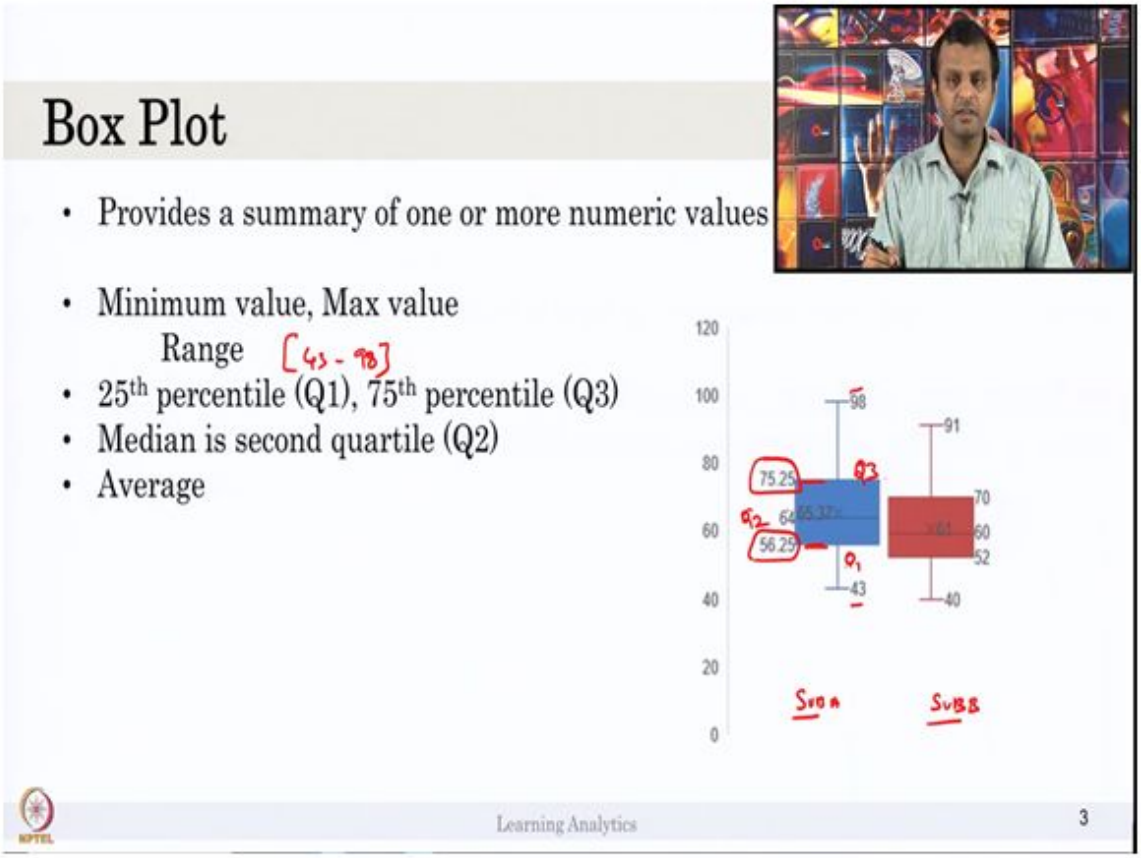
In this video, we will continue discussing Chart types.

(Refer Slide Time: 00:22)



So, we saw the Bar chart, Stack bar charts, Pie chart and Histogram. Let us look at what is Box plot, Scatter plot and Line charts in this video.

(Refer Slide Time: 00:29)



Box plot is used to provide a summary of one or more numeric values. For example, in the last video, we saw 60 students marks percentage as a Histogram, we can use the same student's marks and you can represent using the Box plot. So, but Box gives more information and it gives the feel of the data. So, here is the sample Box plot using the data we discussed in the previous slide. So, here there is a minimum value, this colour indicates marks in the subject A and this indicates marks in a subject B

There is a minimum value of 43 and some maximum value 98. So, the range is 43 to 98 that is the range of the marks from the student's performance. And you see there is 25th percentile that is Q 1 this is the 25th percentile. And this is the 75th percentile and the median is the 50th percentile. So, the 25th percentile is this value(lower were the edge of the box) and the upper edge of this rectangle is the 75th percentile.

The percentile indicates that 25 per cent of value in this performance will be below this percentile value. That is we have 60 students, 15 students, overall marks will be less than 56.25 and there are 15 students who would have got mark more than 75.25 that is a good thing. And this length indicates that there is more deviation here.

And this middle line indicates the median value median (6)4, for 60 students median will be the value of 31st student's mark. If you arrange the number in ascending order the middle number the median number. So, that is the median value (the 50th percentile). That is it when we say we want to see the median score and select the students above the median score which, means you are selecting the 50 per cent of students in your class.

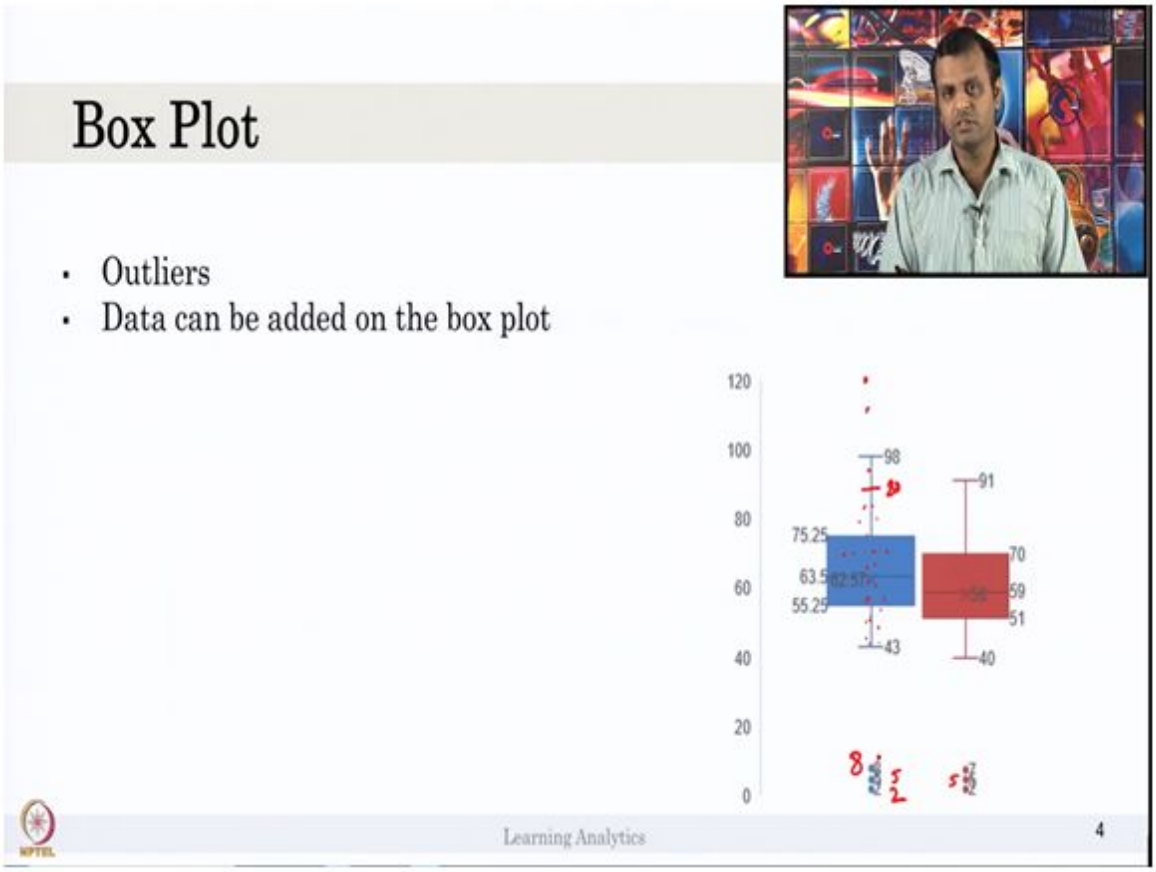
And the median is the Q2 so this is Q1 which is below this. This is Q1, this is Q2, this is Q3. So, what is the average? The average is this cross mark you can see. This cross mark is the average so that is the 65.37 in this particular course. And this subject average is 61 marks

So, this Box plot provides you with the minimum value maximum value how this marks is distributed across all the students. For example with average median value and deviation is more here. A lot of students (15 students) in the range of 56-43 but there is less deviation there. Also, we can use the same Box plot to compare the marks in a second subject.

For the same set of students, the marks in a second subject can be viewed there see. There the marks are almost similar. 40 is the minimum and 91 is the maximum. So, the performance in subject A and subject B is equal or you can use the Box plot to compare the performance of two classes say if you are teaching the same subject to two classes class A and class B and you conduct the test. And that marks can be plotted in a Box plot and you can compare why one particular class is doing better or not, you can know from this figure.

So, by Box plot, you can get the distribution(not like a Histogram) but you can get the distribution of data and you get the detail like range, deviation in each quartile everything can be seen from the Box plot.

(Refer Slide Time: 05:06)



In a Box plot we can also have outliers, for example, in this particular example I add 3 marks I arrange 3 marks (2,5,7) in this and (2, 5, 8). So, these 3 marks are the outlier because they lie below this mark that is 43 and 40. So, if you have outliers or the Box plot will indicate this. So, if you remove that as an outlier it does not change much so they remove these three marks as the outlier.

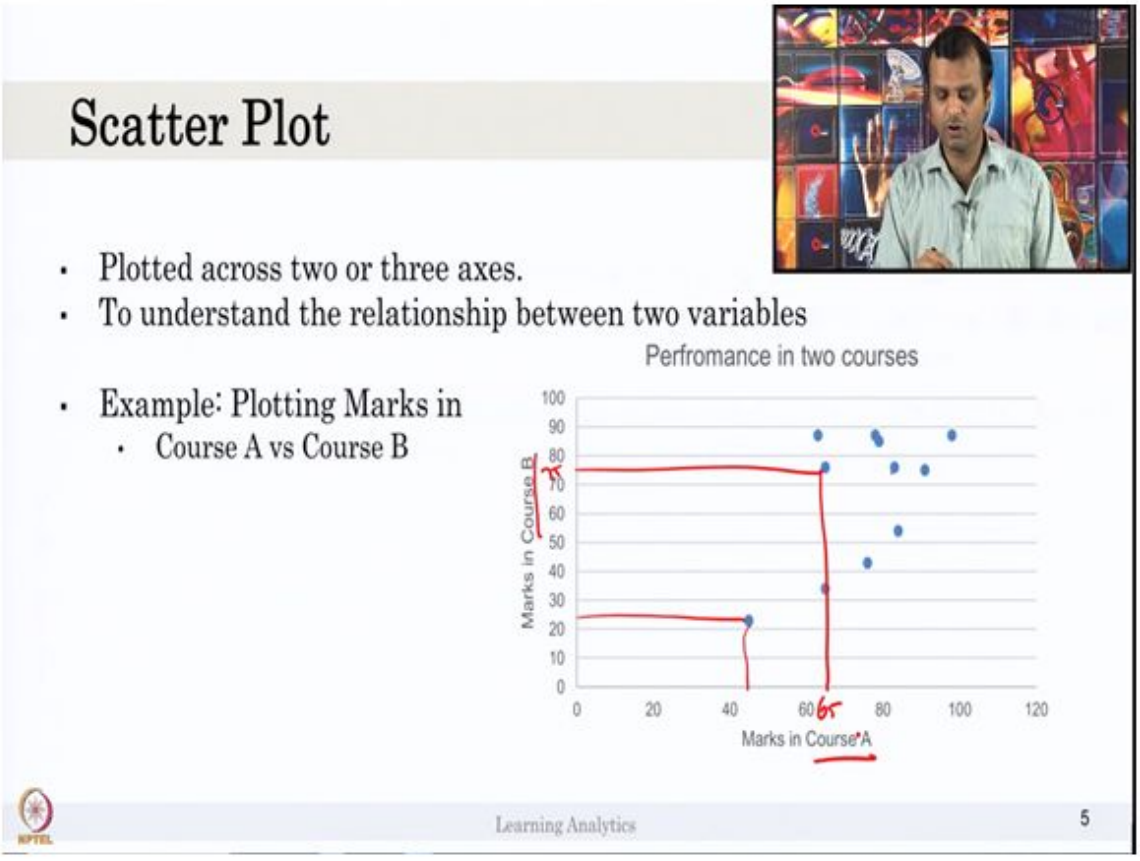
And, outlier can be seen here. Similarly, if the maximum mark is 98 only one student above that so suppose consider that if we have a chart here about 80 only two or three students got around 91 and 92 they will be like an outlier. When I was talking about processing the data I mention that you have to be careful about an outlier.

So you should be very careful about the outlier, so in some research, you might want to remove them or you want to consider why these students are outlier why there were extra marks for the students how are they not able to perform well. There might be these students who are not able to do good in the exam because they were not feeling well during that day. Because if we know

these students did not perform well not because of they did not attend the class or did not understand instruct they are not feeling well so they are not able to sit in the class or something like that. So, you should be careful about the outlier.

So, a box plot is one plot which tells you an outlier. There are other charts which can you tell outlier but the box plot is the easiest one. Also, the data can be added on the box plot. Also, you can have a data around this, so each data plotter will be looks like this so the data around this will look like that. The distribution of data everything can be plotted in the same chart to get a more feeling or more sense of the data. It is really good but it is just you can add the data on the box plot.

(Refer Slide Time: 07:47)



Let us look at the Scatter plot, Scatter plot is plotted across two or three axes usually two axes as it is easy to compare. It is to understand the relationship between two variables it is not just a

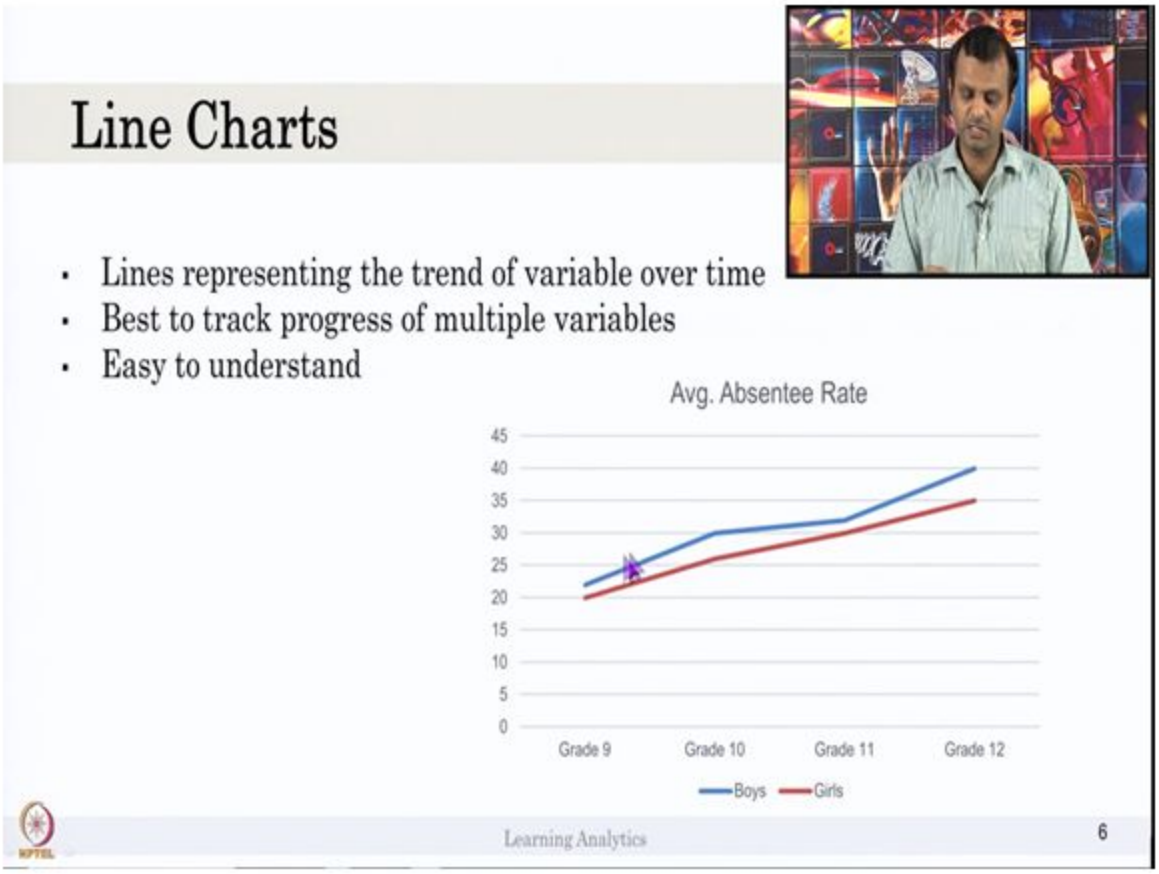
distribution. Example, plotting the marks in Course A and Course B we saw in the last slide course A and course B marks has been shown in the Box plot. And plotting the marks of course A and course B that is subject A and subject B in a Scatter plot.

Here I am not using all the 60 students data obviously, so because that will not lead to a good figure so I just removed most of the data. I kept only 15 data points, so the Scatter plot shows here that the student who got so marks 40 something says 45 or something got around 25 marks in a subject B. Similarly, the student who got say 65 or something got around 35 to 80 marks.

So, the student who got 65 in a subject A got over 75 marks in course B. So, why we are plotting this Scatter plot I want to understand the relationship between the marks in course A verses course B. Whether the student who can do well in course A also can do well in course B or there are some students who can really do good in this course say 45. But he is not able to do a well in the other course. Some students who are doing okay in this course but it can get up to 90 marks in the other course.

So, we might want to know why the students are not able to do well in this course or why these students are not able to do well in this course. So, to understand the relationship between these two variables i.e. marks in course A and course B we can use the Scatter plot. Or, you can do a Scatter plot for attendance basis performance or Scatter plot of students engagement in the performance lot of other variables.

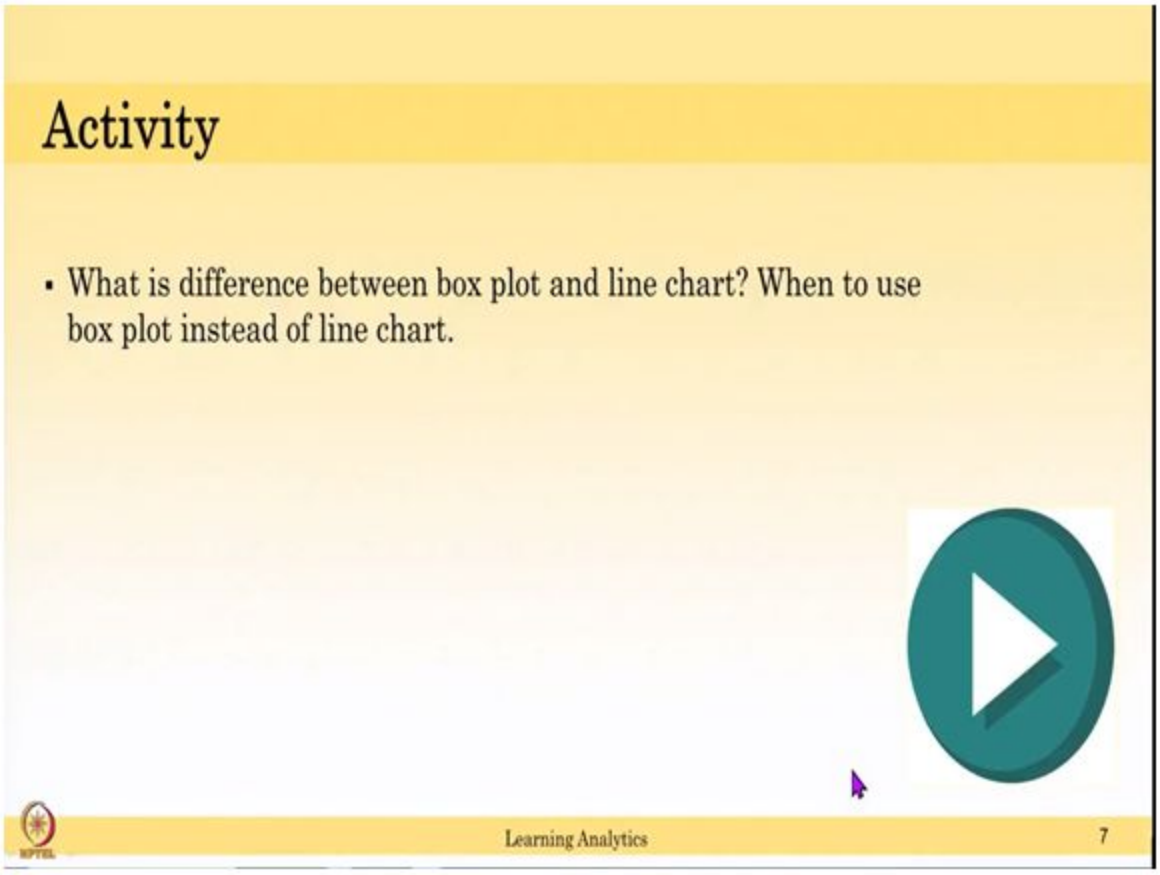
(Refer Slide Time: 10:06)



So, let us move on to the next chart line chart. Line chart represents a trend of the variable over time. So, the Line chart you might have seen it everywhere. The line chart is very common and it is very useful to track the multiple variables over time. And it is very easy to understand that is why it is used commonly. Let us look at the line chart of the average absentee rate from grade 9 to grade 12 across boys and girls. It is not over time it is over different grades. A line chart is not always plotted against the time, it can be like to check the percentage variance over different grades.

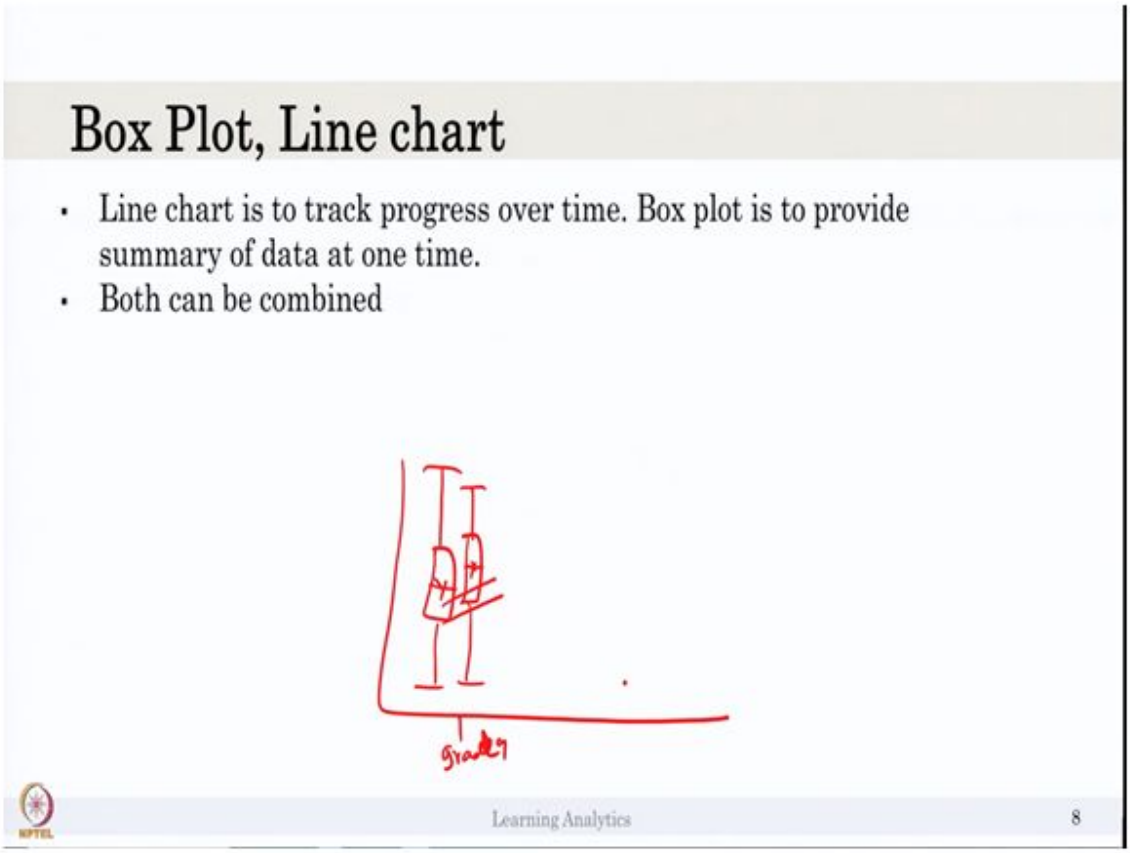
We saw this similar chart in a Stack bar chart also in the bar chart example, so the same value can be plotted as the line chart. In this chart, it is easy to understand that girls average absentees rate is always below the boys average absentee rate. And boys average absentee rate is increasing for every grade.

(Refer Slide Time: 11:18)



So, you saw a Box plot, Line chart also Scatter plot, so what is the difference between Box plot and a Line chart. Also, can he list down when to use Box plot instead of a Line chart?

(Refer Slide Time: 11:32)

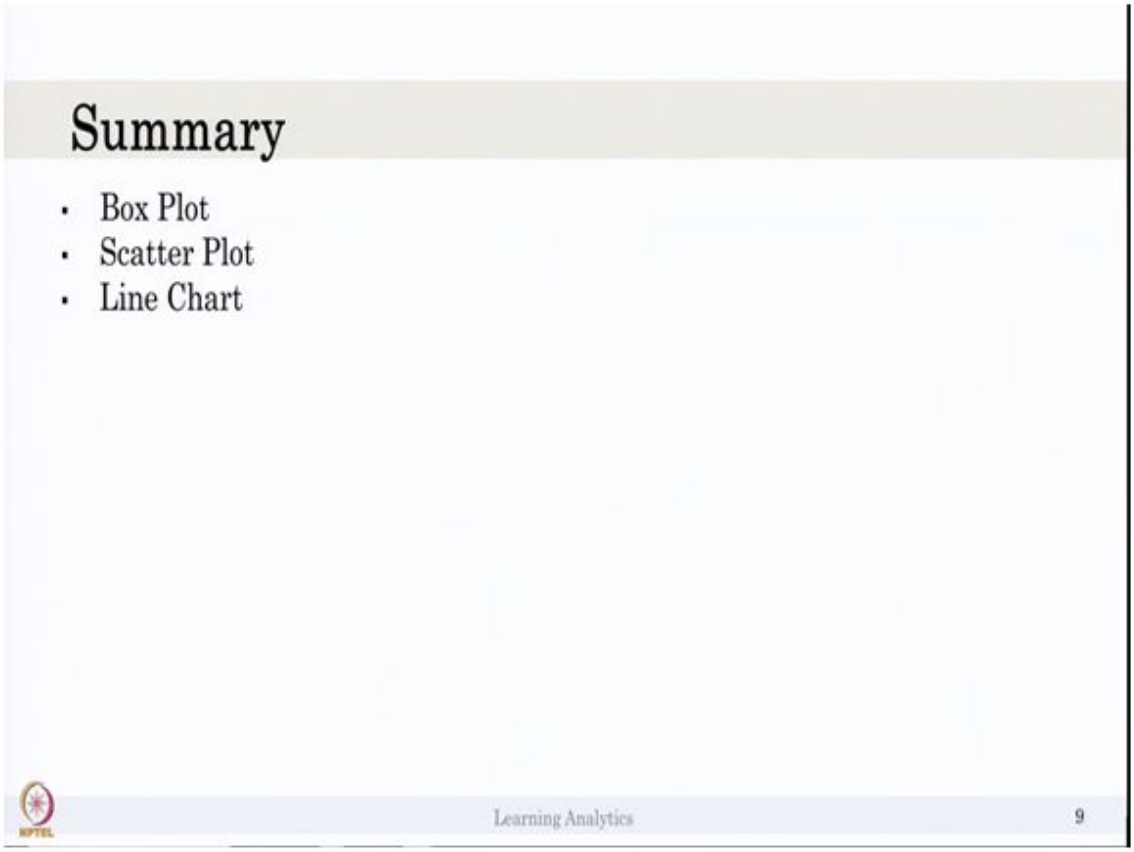


The line chart is used to track progress over time and Box plot is to provide a summary of data at one particular time say class A students marks or grade 9 boys students attendance rate or something like that. And, Box plot will give more sense of data like what is the meaning of range or some maximum value also the distribution of data in the particular grade or particular class. But if we can combine both there will be a chart like we can combine both Line chart verses and Box plot. You can have the Box plot here you can have a Bar plot here something like that.

So, for grade 9 there are two so you can have a lot a Box plot also with the trend change(shown using the Line chart). That will be more helpful to understand there is trend between each grade and boy and the girls absentee rates observed. Also, you can get the sense of data in each class.

So, you can combine this kind of basic charts to create a new chart, so it is not that you have to use only one chart for representation you can combine this chart and make a new chart which makes more easy to understand and more sense of data and more inference from the data as possible.

(Refer Slide Time: 13:06)



So, in this video, we talked about Box plot, Scatter plot and Line chart. Thank you.