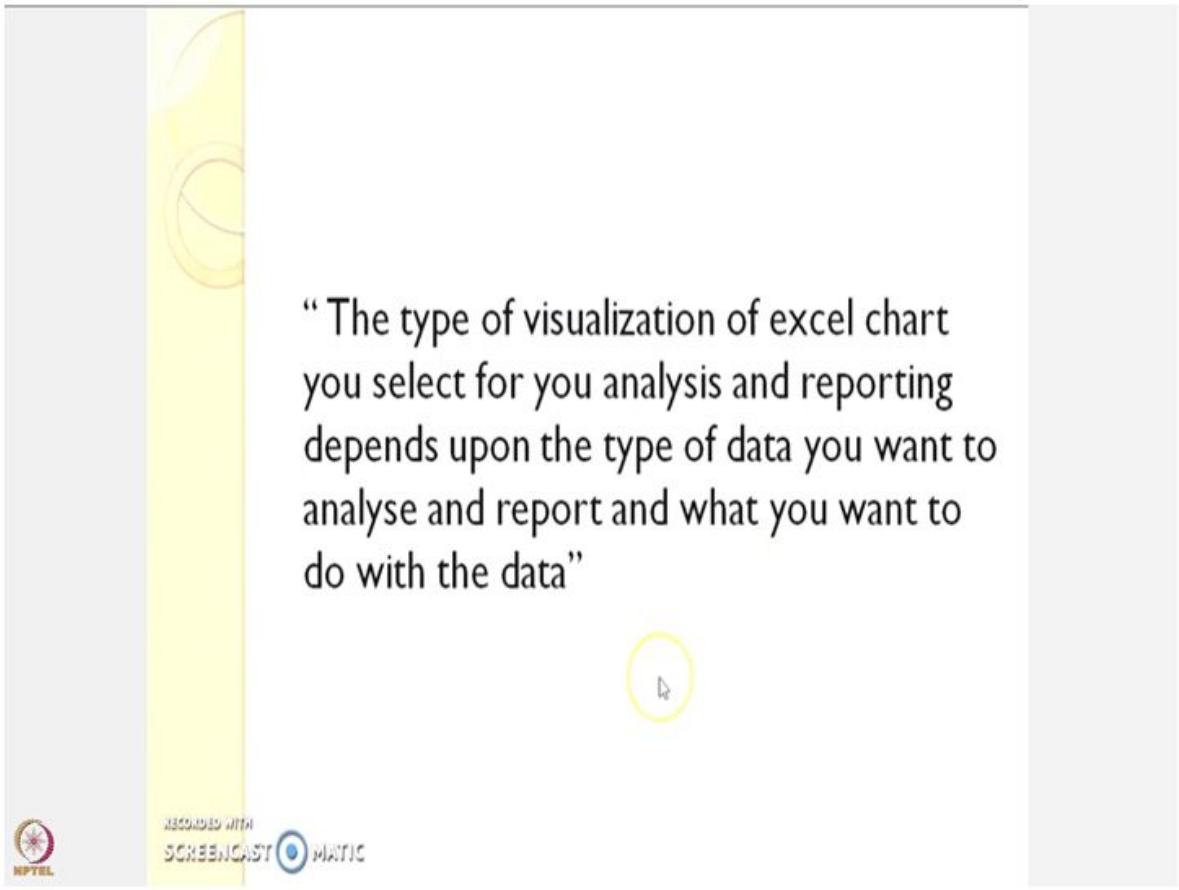
Learning Analytics Tools

Priya Lotlikar
Research Scholar
IDP-Educational Technology
Indian Institute of Technology, Bombay

Lecture 25

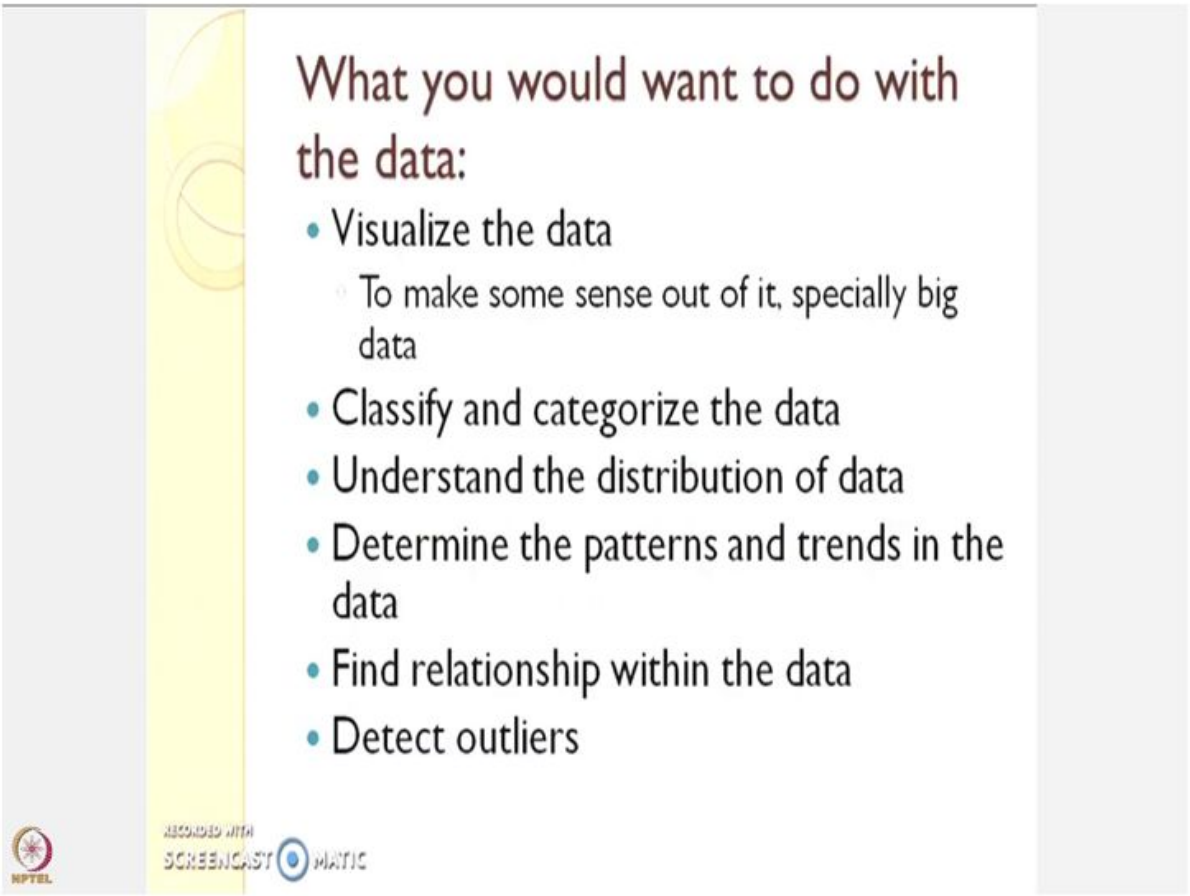
Data Visualization in Excel

(Refer Slide Time: 0:31)



Hi everyone, I am Priya Lotlikar, a first-year PhD research scholar from the Department of Education Technology, IIT Bombay. Today, we are going to look at data visualization in excel. The type of visualization of excel chart you select for your analysis and reporting depends upon the type of data you want to analyze and report and most importantly what you want to do with the data.

(Refer Slide Time: 0:45)

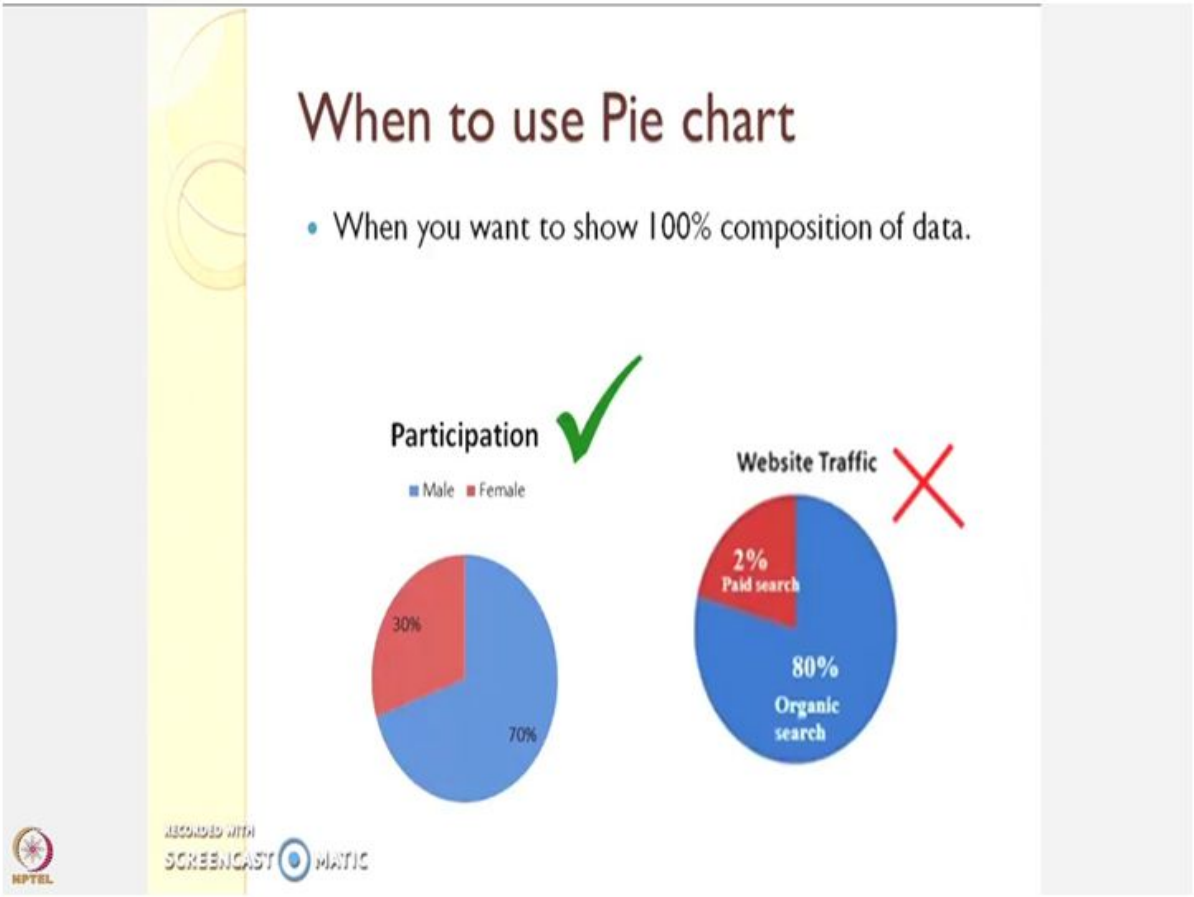


Some operations that you will want to do on your data would be, visualize the data to make some sense out of it if it is a big data, secondly classify and categorize the data or understand the distribution of data, determine the patterns and trends in the data or find relationship within the data or detect the outliers.

(Refer Slide Time: 1:09)

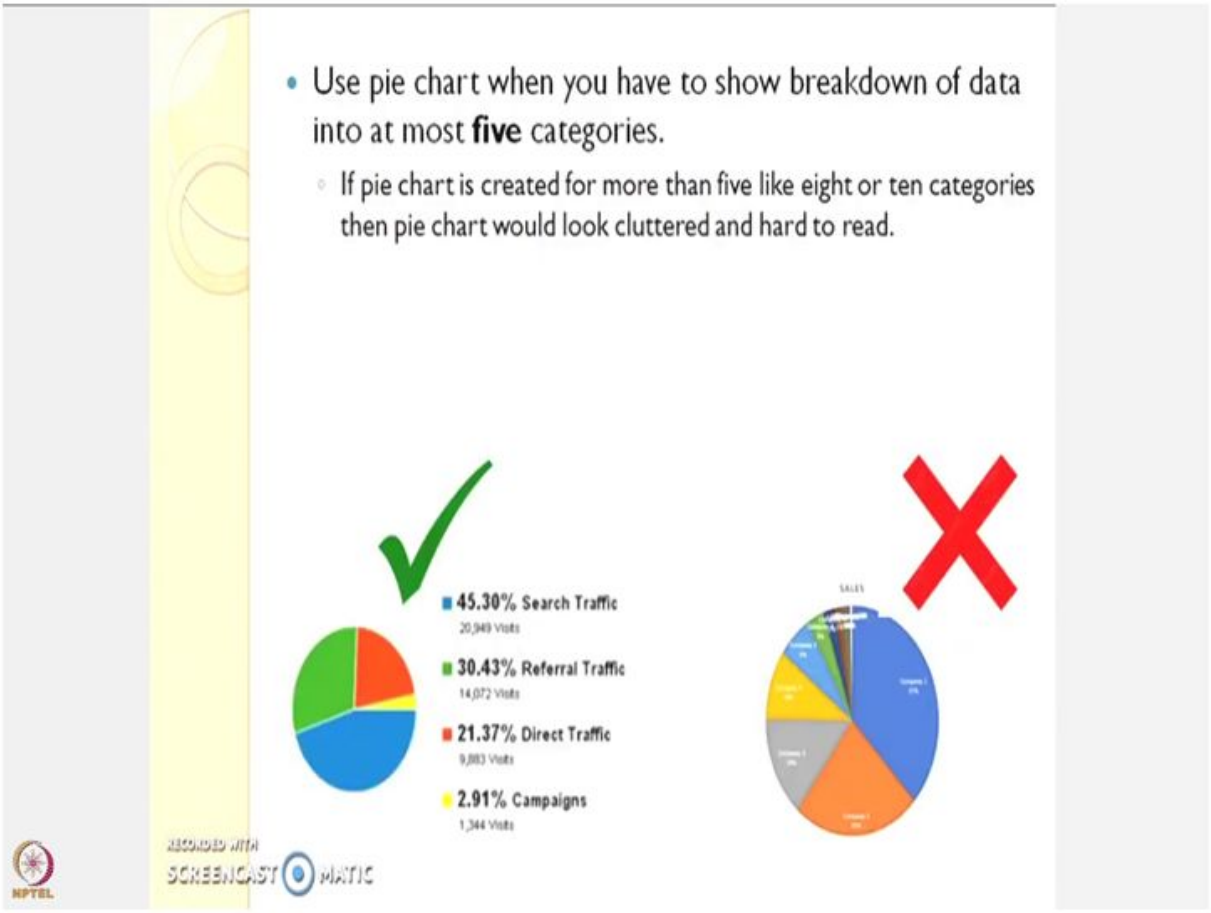
Here are some popular charts, which I will use for data visualizations in Excel - pie chart, bar and column chart, grouped bar chart, line chart, scatter plot, stacked bar chart, histogram and heat map. We will look into this in detail in the coming session.

(Refer Slide Time: 1:27)



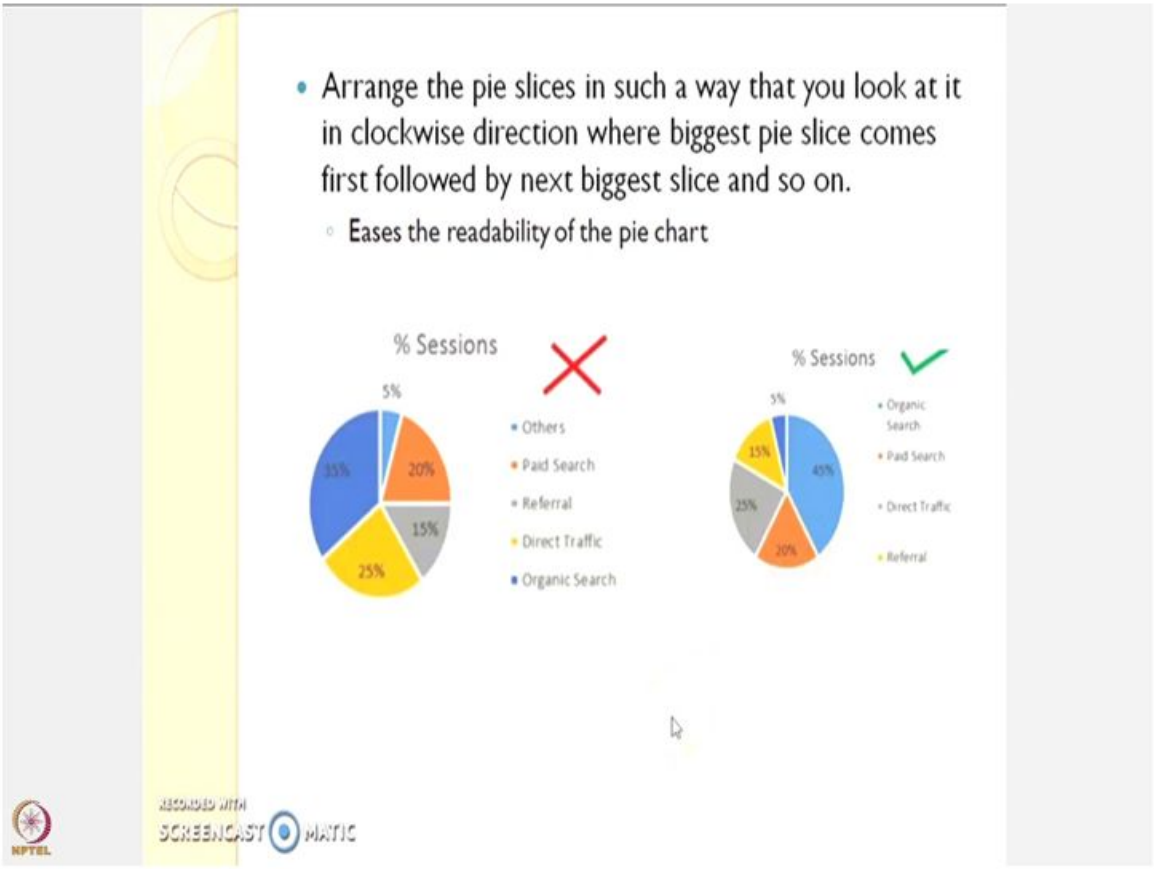
First, the pie chart. When to use a pie chart? You use a pie chart when you want to show 100 per cent composition of your data. For example, look at this image to the left side. Over here, the composition of data is 100 per cent whereas in this image the composition is 82 per cent, so this is a wrong pie chart.

(Refer Slide Time: 1:51)



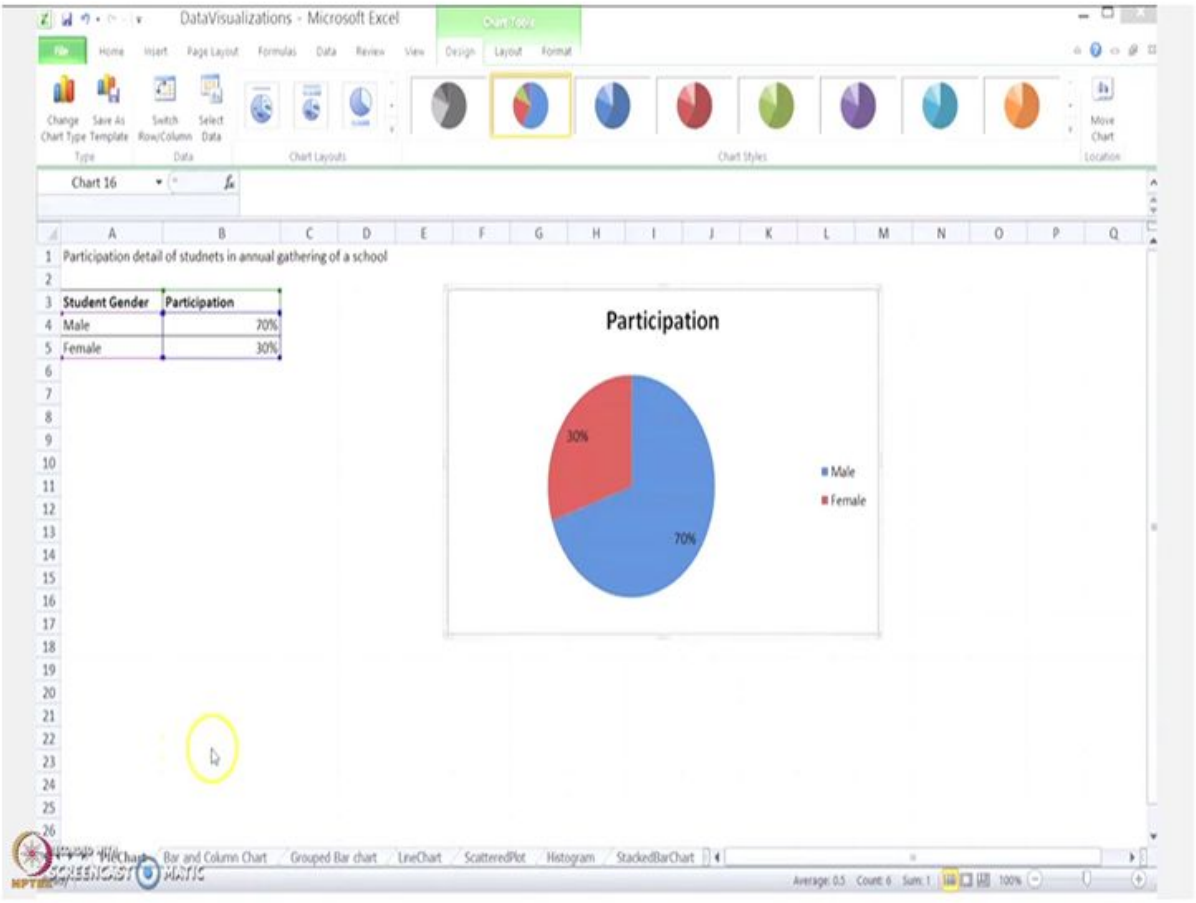
Use a pie chart when you have to show a breakdown of data into at most five categories. That is, if you create the pie chart for more than 5, like 8 or 10 categories, then the pie chart would look cluttered as you can see in this image.

(Refer Slide Time: 2:10)



Thirdly, arrange the pie slices in such a way that you look at it in the clockwise direction where the biggest pie comes first followed by the next bigger slice and so on. This facilitates the ease in the readability of the pie chart.

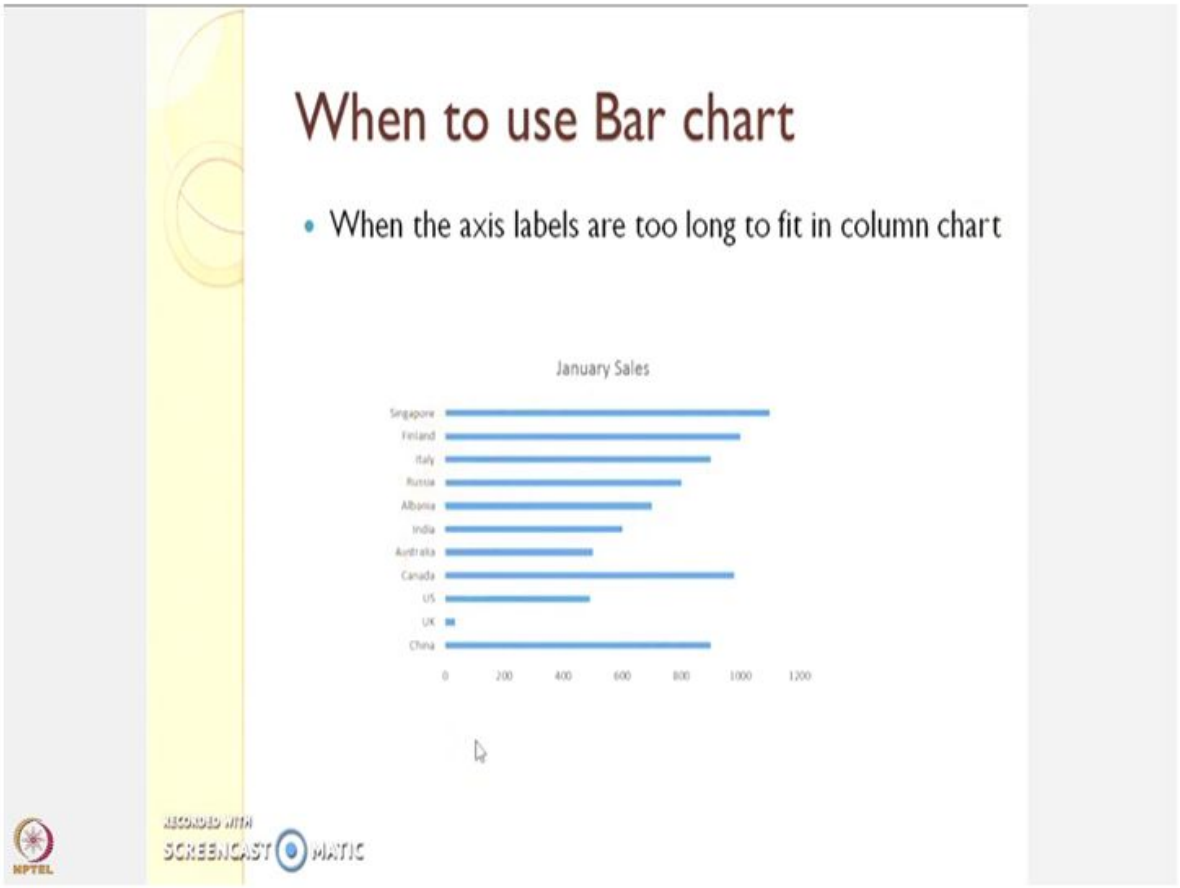
(Refer Slide Time: 2:26)



Let us see how to construct a pie chart. Here, I have the student data, which shows me the participation of students in the annual gathering of the school. I want to visualize this data using a pie chart. So, for that, first I need to select the entire table, go to insert, under the chart section, I need to choose the “pie” option. When I click on this, there are various representations provided to me, choose whichever you want. I will choose the first one.

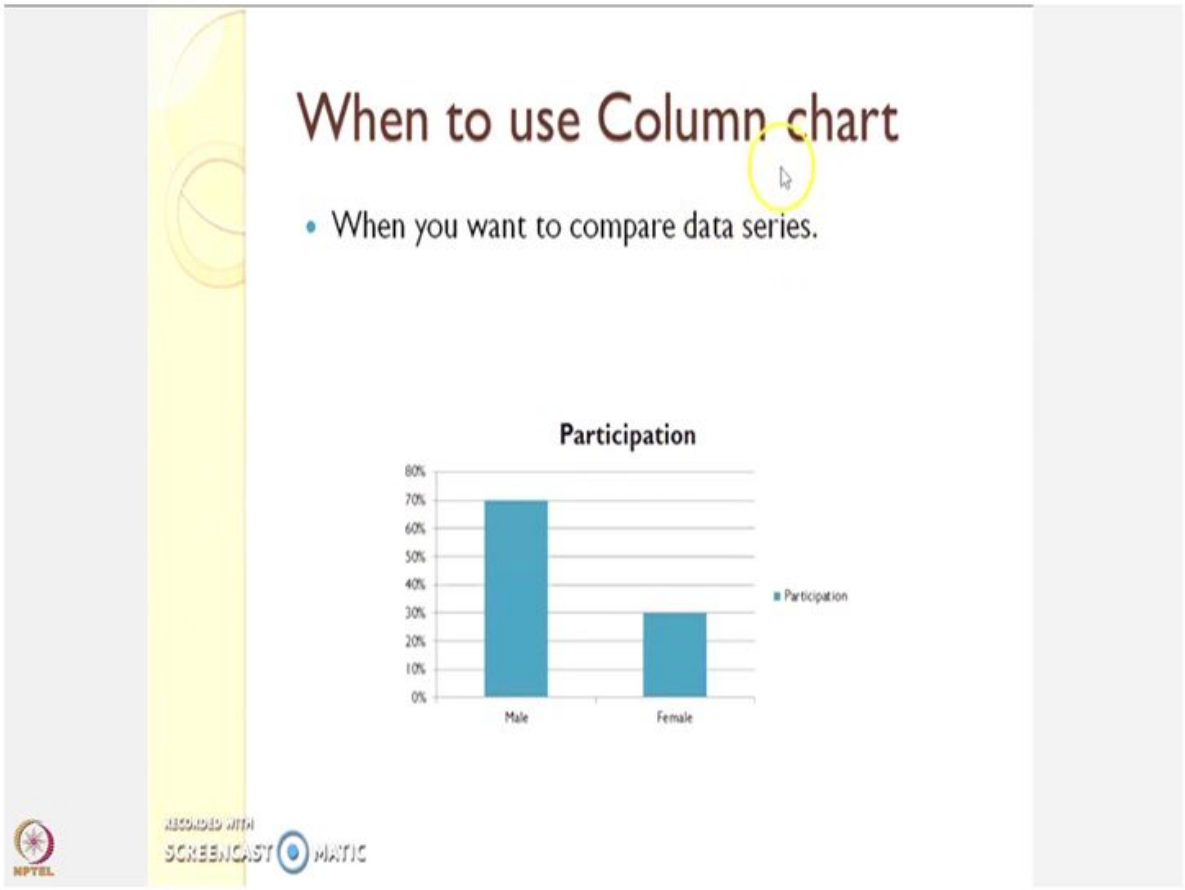
So, this is how the pie chart would look like, but currently, there are no data labels on the pie chart. So, to add the data labels, click on the pie chart, right-click and click on add data labels. So now, this is a complete pie chart for a given particular table. Here, every column of the table comprises one slice from the pie chart. As you can see, female participation is represented by the red colour and male participation is represented by blue colour. One thing to remember with the data is that rows should be sorted by numerical column in descending order, so the slices will be in the size order.

(Refer Slide Time: 3:40)



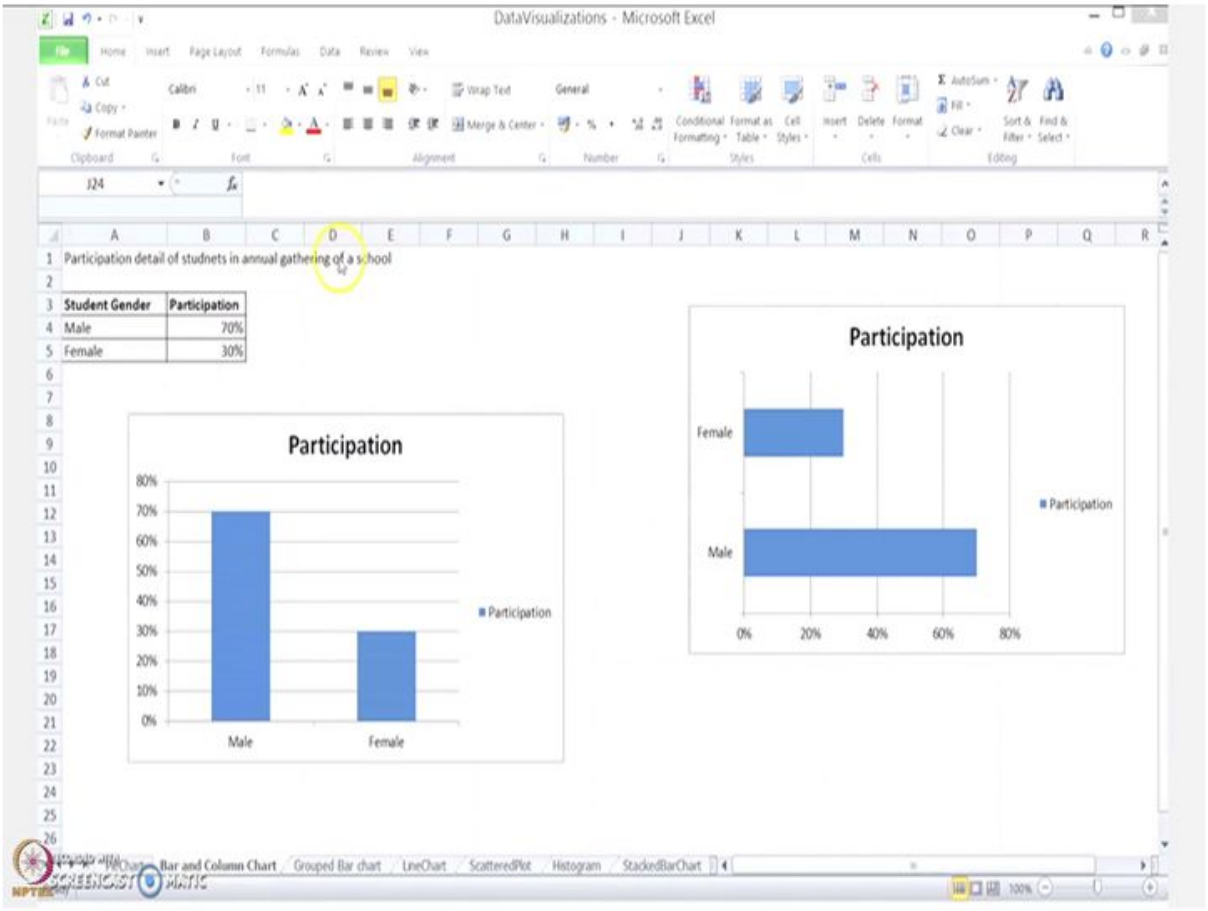
Next is the bar chart. Now, when to use a bar chart? To use a bar chart, when the axis labels are too long to fit in the column chart. So, a bar chart is basically arranged along the horizontal axis, that is the X-axis.

(Refer Slide Time: 3:57)



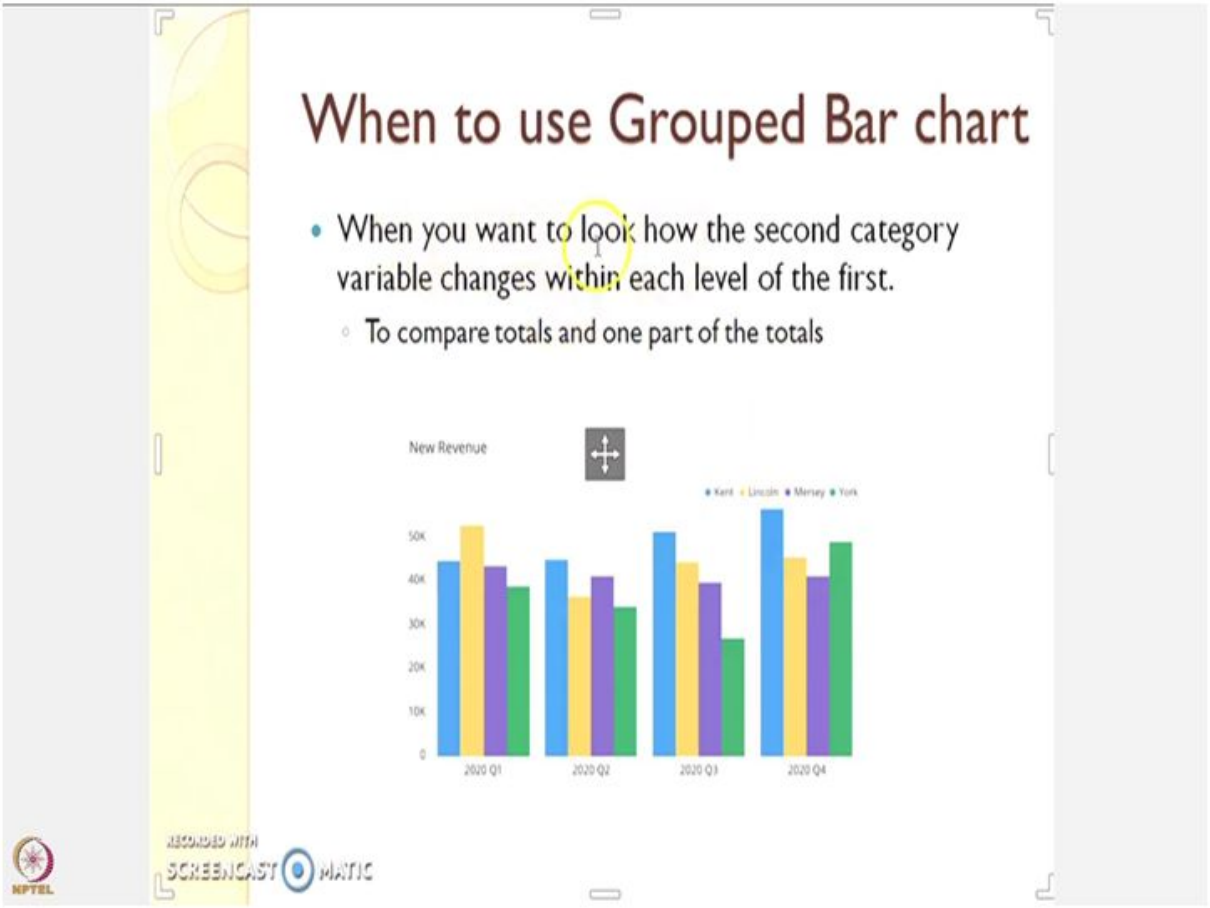
Next, you have the column chart. The column chart is basically used to compare the data series.

(Refer Slide Time: 4:05)



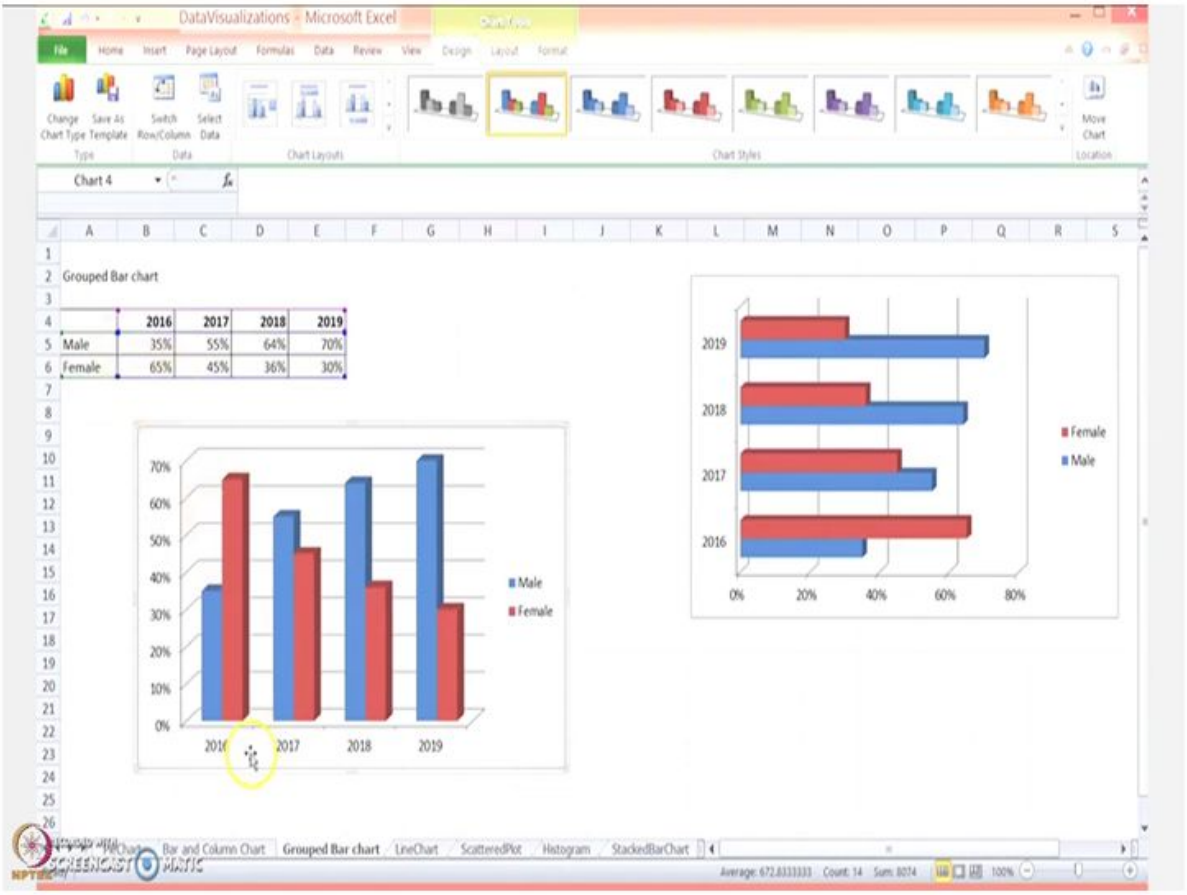
Let us see how to construct a bar chart or a column chart for the same data. So, choose the entire table, under insert in the chart section, select the column. In this you have to 2D representation as well as 3D, let me choose the 2D representation. So, this is how the column chart would look like. Now, if I want to construct a bar chart again, select the entire table, insert, under charts I have the bar option, choose the 2D representation, so this is how the bar chart for this table would look like. Here, each row will become a separate bar, so the rows should be sorted in a numerical column in descending order.

(Refer Slide Time: 4:56)



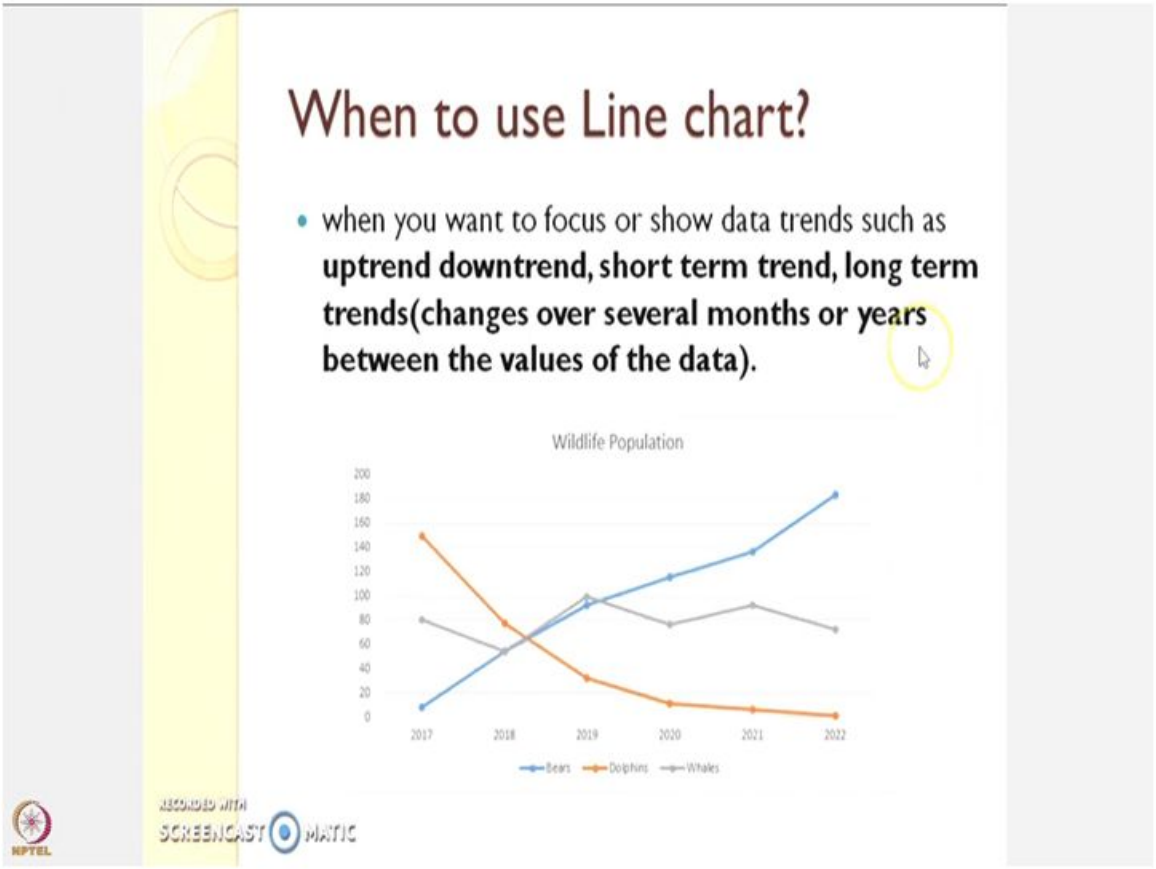
Next, we go to the grouped bar chart. When to use it? You use it when you want to look how the second category variable changes within each level of the first, that basically compares the "totals" and one part of the total.

(Refer Slide Time: 5:12)



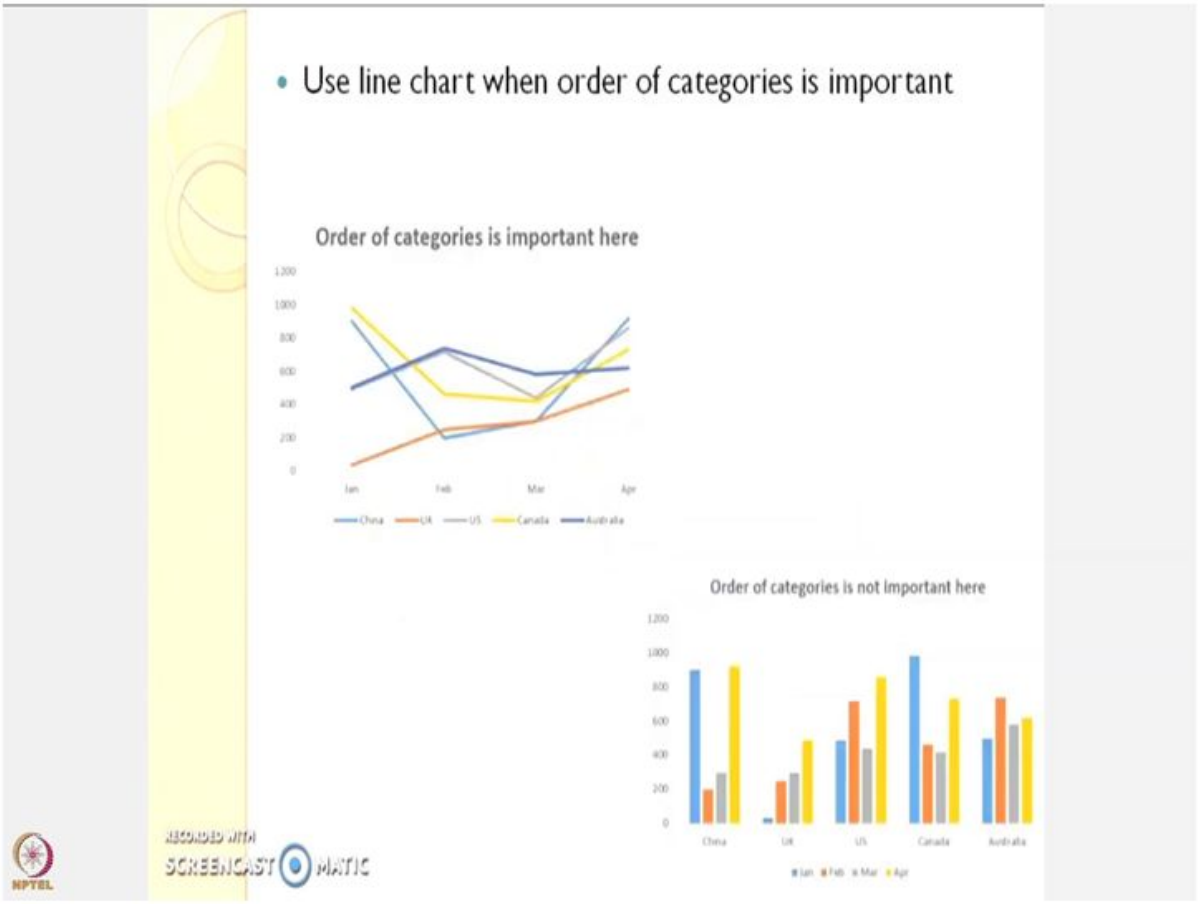
Let us see how to construct a grouped bar chart. I have this data on student participation over a period of four years. So, in order to view this in a grouped bar chart, select the entire table, go to insert. Under insert choose the bar option, in this bar you choose the 3D bar. So, this would give you the bar visualization. I can even choose the column representation, so choose the entire table, under insert, in chart section, column, choose the 3D column. So, this is how the 3D bar and column representation looks like. Here, the colour-coding is provided for every group. So, you can distinguish between these groups. As you can see the male and female participation over the years are being represented with two colour codings.

(Refer Slide Time: 6:12)

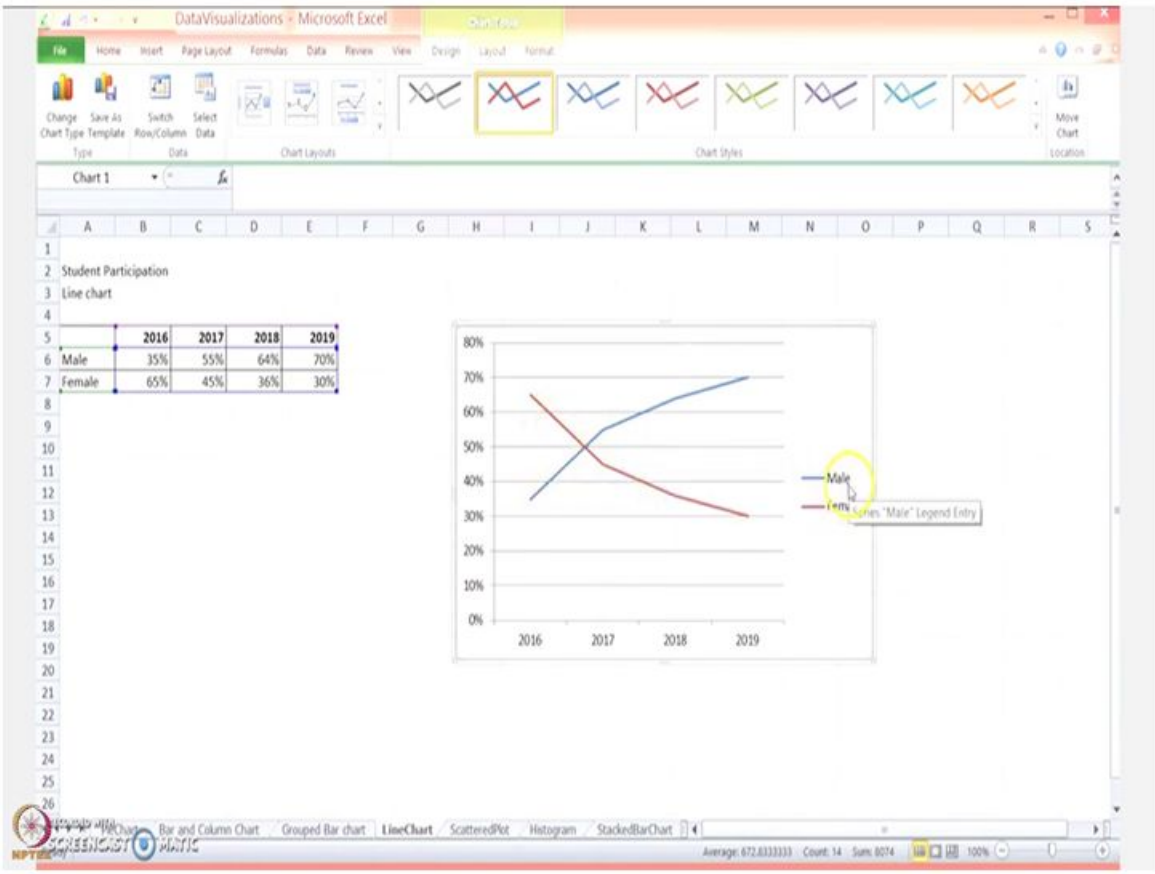


Next, we will look at the line chart. So, line chart is basically used when you want to focus or show data trends such as uptrend, downtrend, short term trend, long term trend, that is changes over several months or years between the values of the data as you can see in this image.

(Refer Slide Time: 6:33)

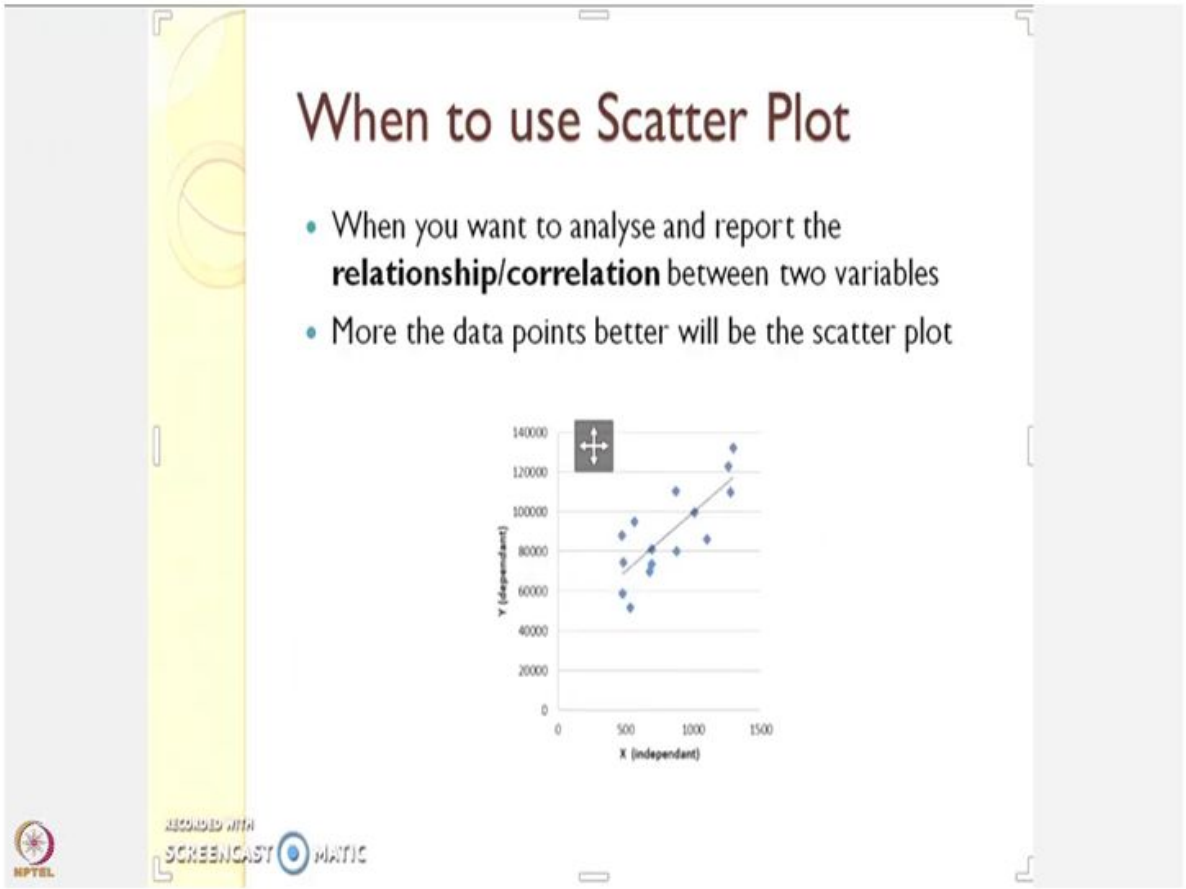


Also, use line chart when the order of categories is important. Like you can see in this the order of categories of respective product is important over here rather than in a bar chart.
(Refer Slide Time: 6:49)



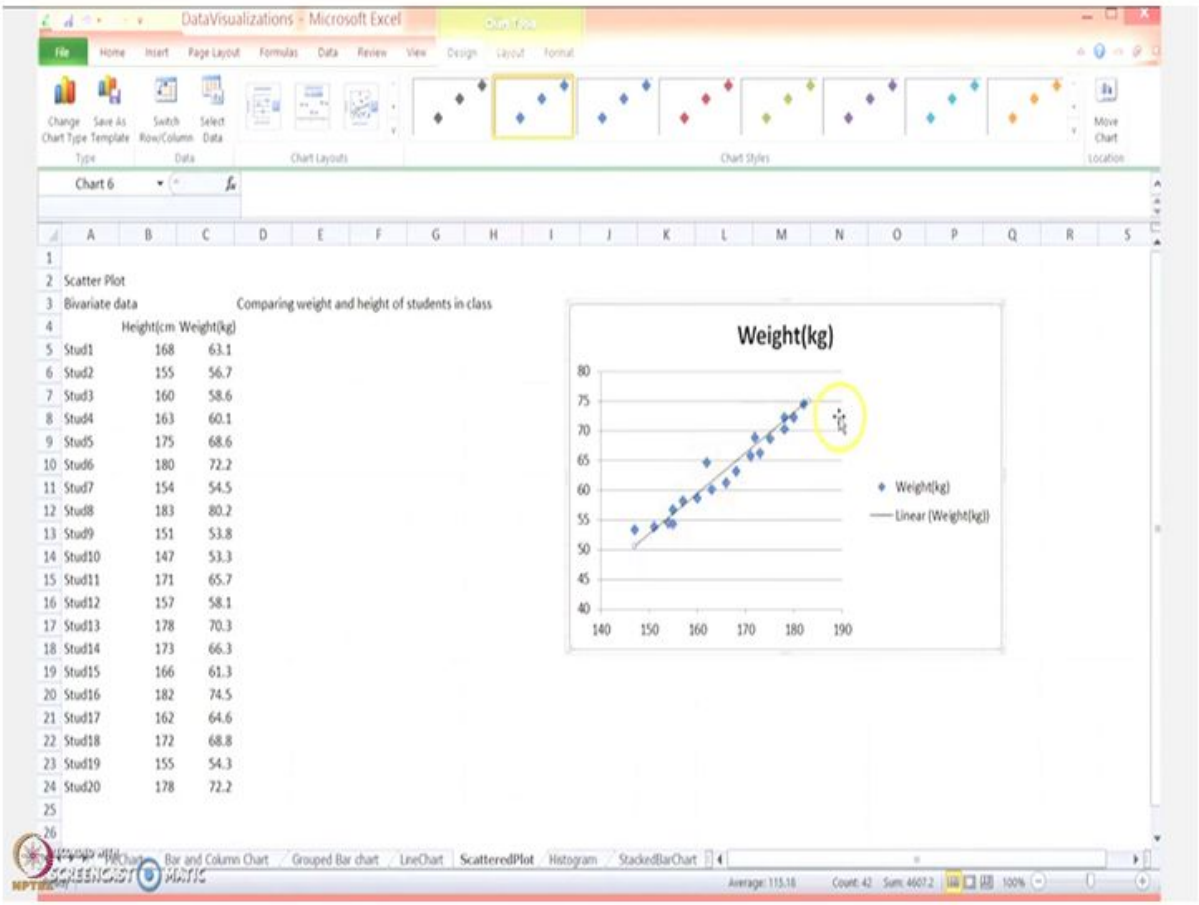
Let us see how to construct a line chart. This is the same data that we used to construct the group bar chart, so basically this is multiple group data, we can also use a line chart to represent this data, so for that, select the entire table, go to insert, under chart section choose the line option, you can choose any representation you want. Let me choose the first one. So, this is how the line representation for this particular data looks like. As you can see, the lines red and blue represent the trend in male and female participation.

(Refer Slide Time: 7:28)



Next, we will look at the scatter plot. So, the scatter plot is basically used to analyze and report the relationship or the correlation between two variables. So, more the data points, better will be the scatter plot. You can see the independent variable is on the X-axis and the dependent variable is on the Y-axis.

(Refer Slide Time: 7:50)



Let us see how to construct a scatter plot. Here, I have the data which captures the height and weight of students from a class. Now, since you have variables, this is a bivariate data. So, we use a scatter plot to basically compare or show the relationship between the two variables. So, let us develop a scatter plot to compare the weight and height of students in this class.

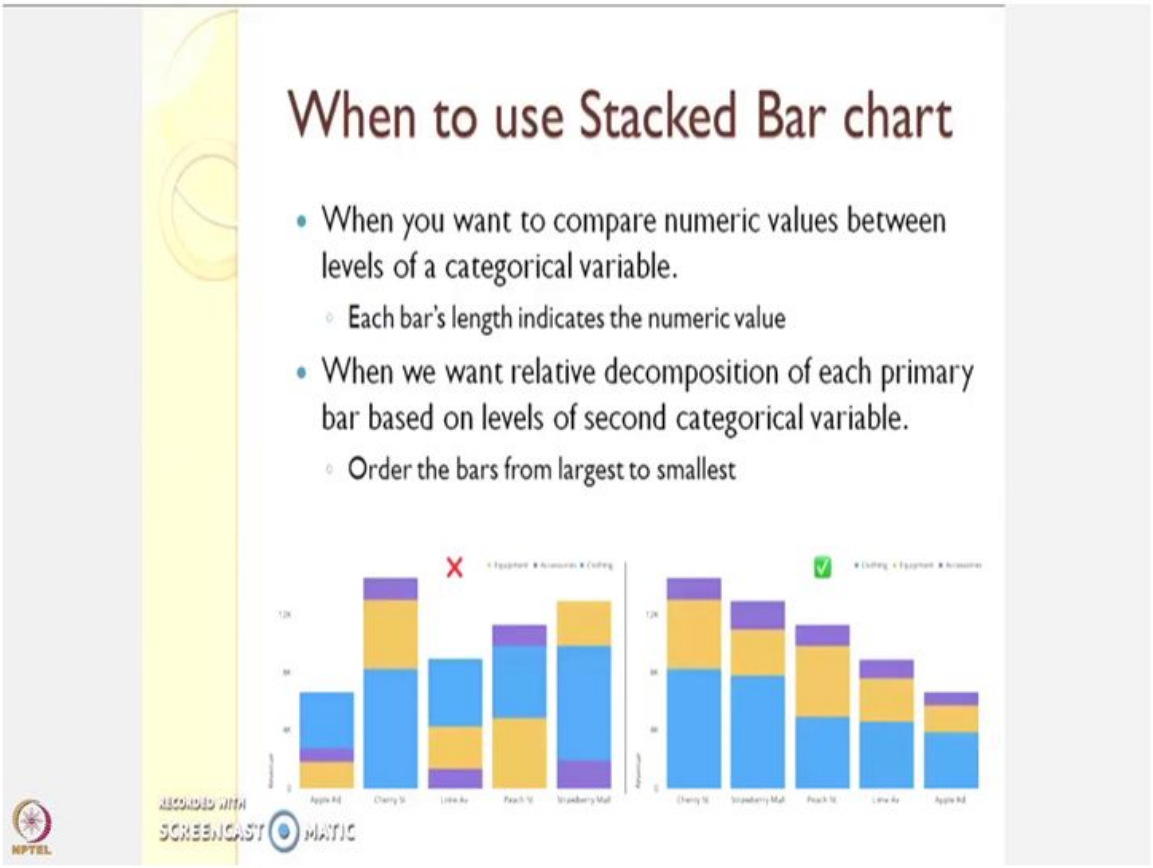
For that, I need to choose the columns of the two variables, so choose these two columns, go to insert, under chart section choose the scatter plot. You have various representations over here. Let me choose the first one. If you go to see, this does not look like a clear representation of the data, the minimum value would lie somewhere 130-140 and maximum value would lie somewhere from 190-200, so let us modify the X-axis to get a clear picture of this data.

For that, click on the X-axis, right-click, format axis. So, over here, in the axis option, set the minimum value to 140 and maximum value to 190 and say close. As you can see this gives a clear picture. The same thing, we can do with respect to Y-axis. So, click on the Y-axis, right-click, format axis, so I can set the axis minimum value to somewhere 40 as you can see in the image and maximum value to somewhere 80, close.

So, now this gives us a clear picture of the entire data. If you want, you can even add a trend line to give a more clear picture about this data, so for that, I can choose the data points, right-click and click on “add trend line”. So, this gives us a clear picture of a linear increase in

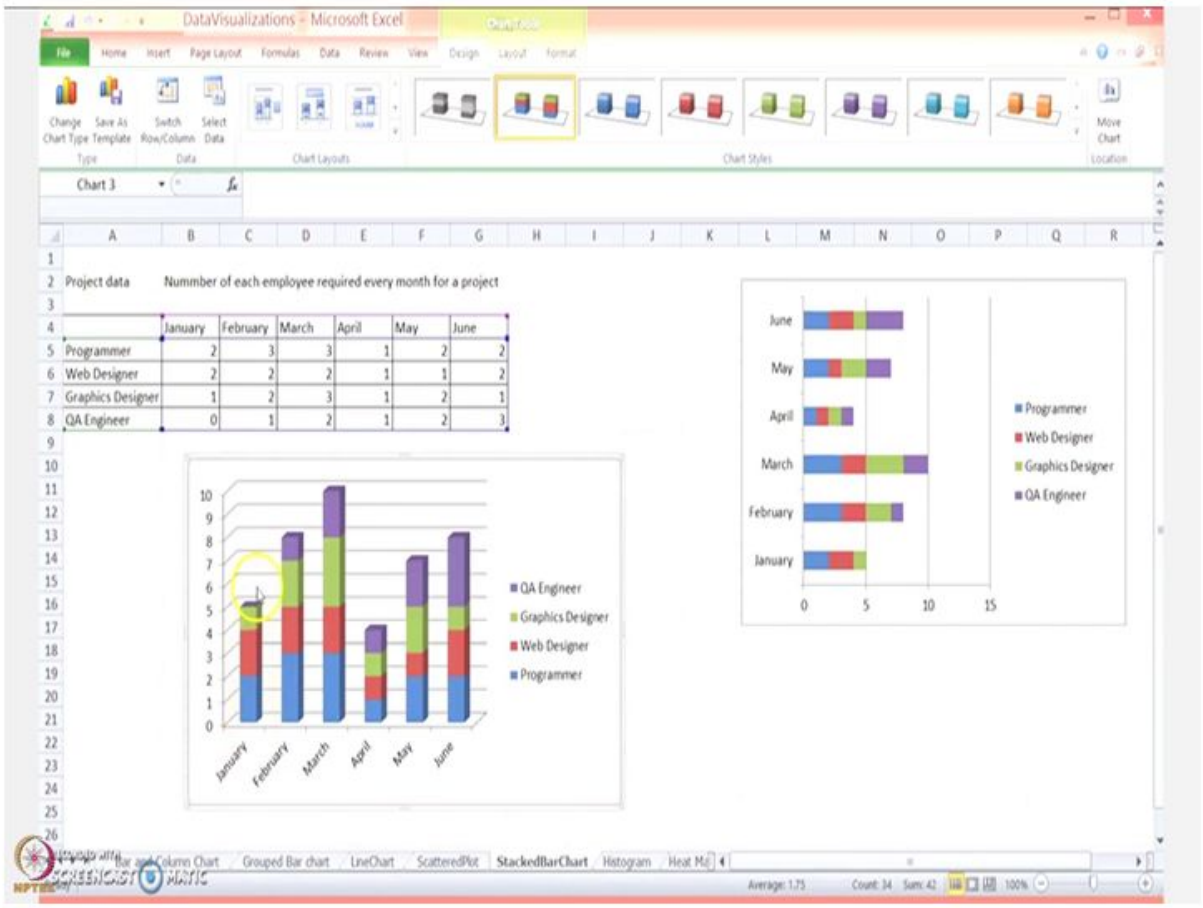
the trend of the height and weight, that is height and weight are linearly proportionate to each other.

(Refer Slide Time: 10:11)



Next, we will look at the stacked bar chart. So, you use a stacked bar chart when you want to compare the numeric values between the levels of the categorical variable. That is each bar indicates a numerical value. So, when you want the relative decomposition of each primary bar based on the levels of the second categorical variable, we use a stacked bar chart. That is, the order of the bar is from the largest to the smallest.

(Refer Slide Time: 10:43)

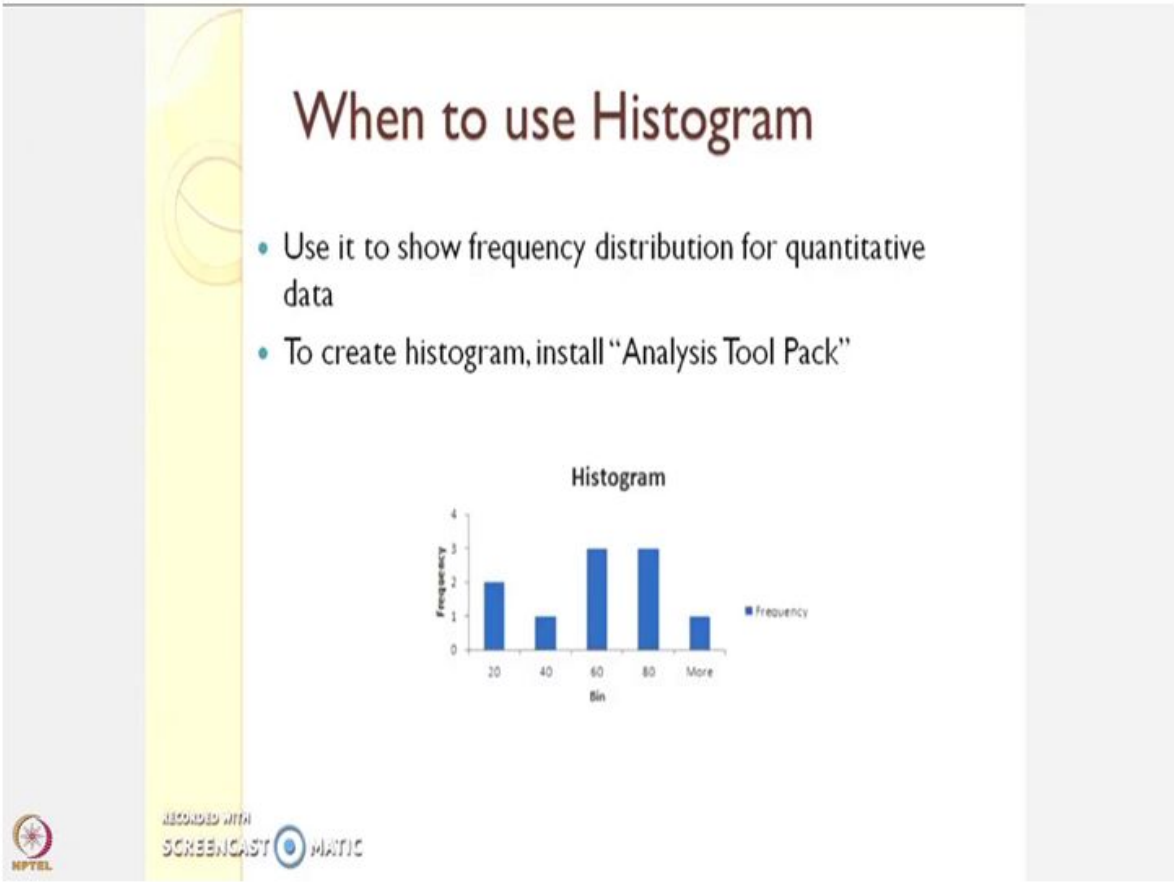


Let us see how to construct a stacked bar plot. I have this data over here, which gives me the number of employees required on a particular project for this respective month and the number of employees is being categorized into programmer, web designer, graphic designer and QA engineer.

So, in order to have this stacked look of these employees every month, I need to construct a stacked bar plot. So, for that, I need to choose the entire table. Go to insert under chart section, choose the bar plot, in bar plot I need to use the stacked bar. So, this is how the stacked bar plot would look like. If you want, you can even have a column representation of this. So, for that choose the entire table, insert, under charts go to columns, you can even choose the 3D visualization or the stack, then I click on this, so this is how the 3D visualization of the entire data looks like.

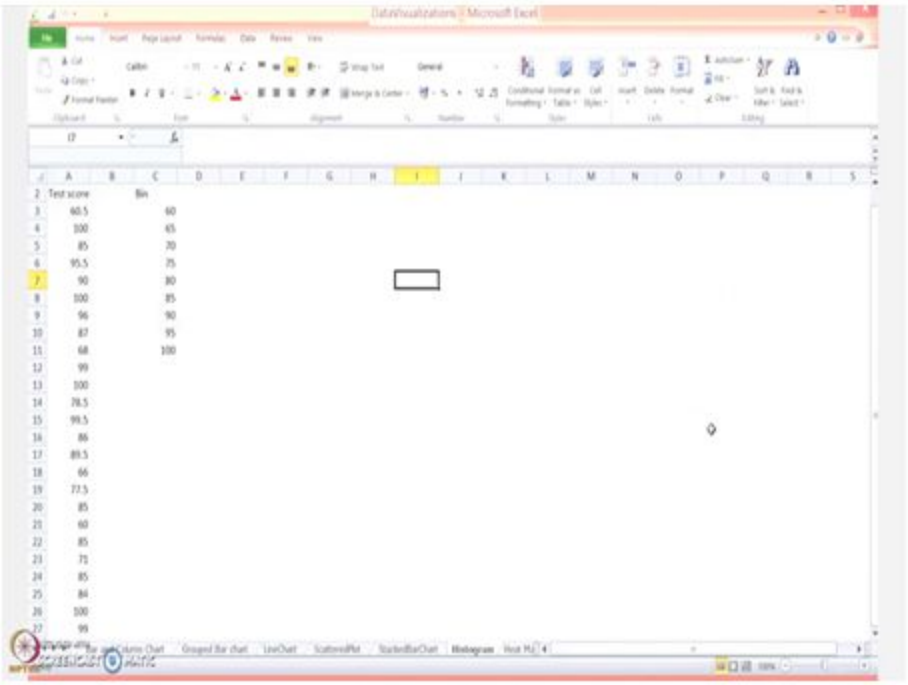
So, every category of the employees that is programmer, web designer and graphic designer and QA engineer are being colour coded so you can see in January the total number of employees required were 5 and so on.

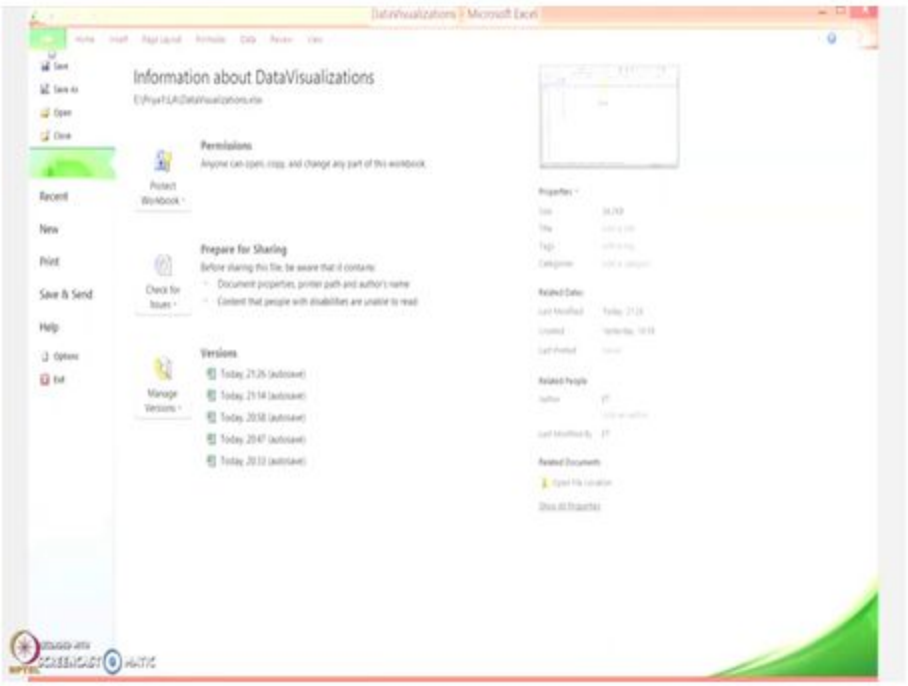
(Refer Slide Time: 12:13)

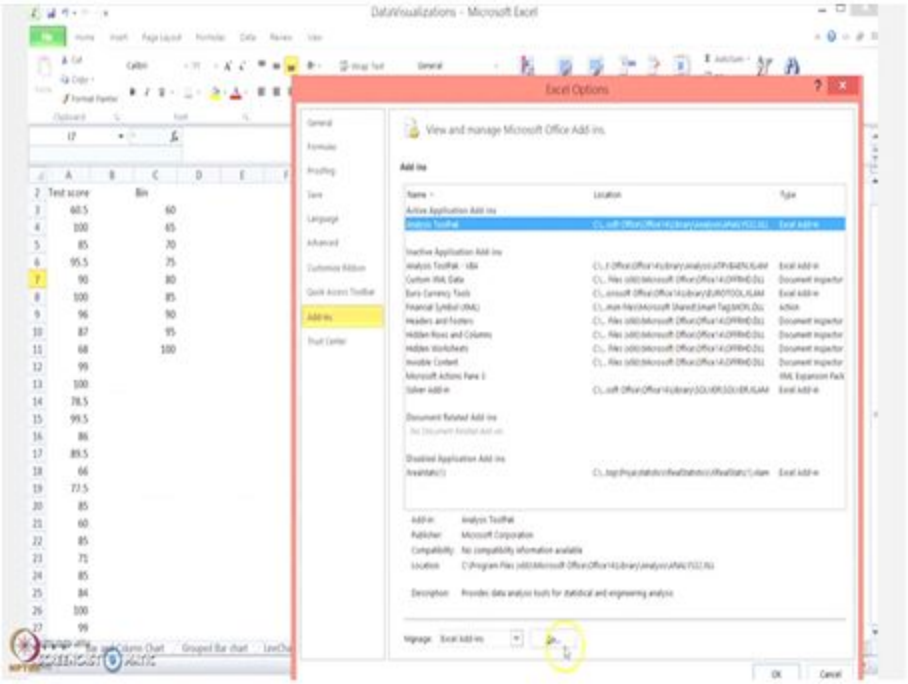


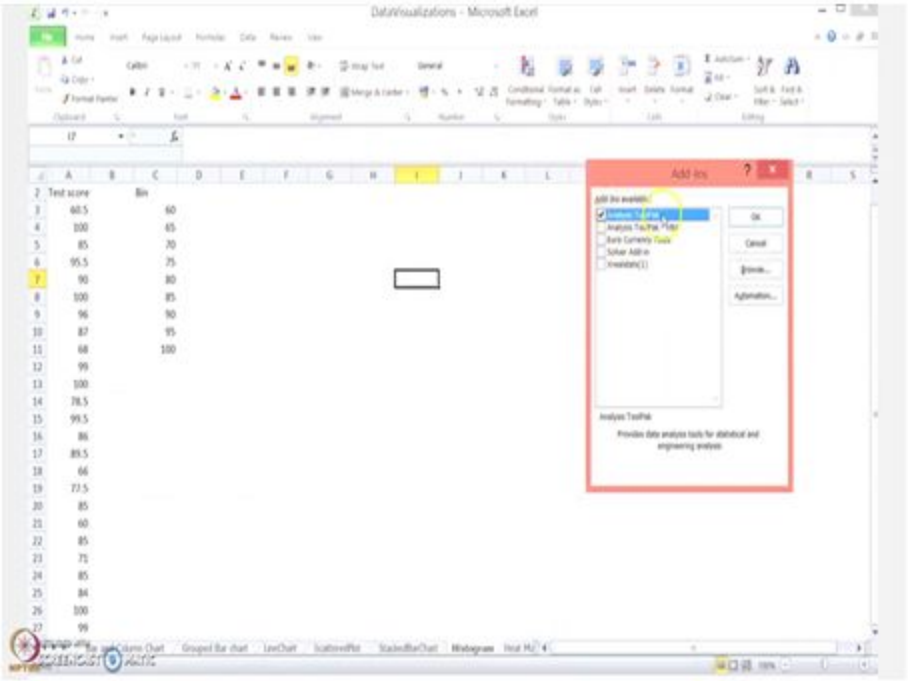
Next, we will look at the histogram. You use histogram when you want to show the frequency of distribution for quantitative data. In order to create a histogram, you need to install the analysis tool pack.

(Refer Slide Time: 12:28)


DataVisualizations - Microsoft Excel
Test score
Bin
60.5 60
100 65
85 70
95.5 75
90 80
100 85
96 90
87 95
68 100
99
100
78.5
99.5
86
89.5
66
77.5
85
60
85
71
85
84
100
95
Grouped Bar chart
LineChart
ScatterPlot
StackedBarChart
Histogram


DataVisualizations - Microsoft Excel
Information about DataVisualizations
E:\Priya\1.IA\DataVisualizations.xlsx
Permissions
Anyone can open, copy, and change any part of this workbook.
Protect Workbook
Prepare for Sharing
Before sharing this file, be aware that it contains:
Document properties, printer path and author's name
Content that people with disabilities are unable to read
Check for Issues
Versions
Today, 21:26 (autosave)
Today, 21:14 (autosave)
Today, 20:58 (autosave)
Today, 20:47 (autosave)
Today, 20:33 (autosave)
Manage Versions
Recent
New
Print
Save & Send
Help
Options
Exit

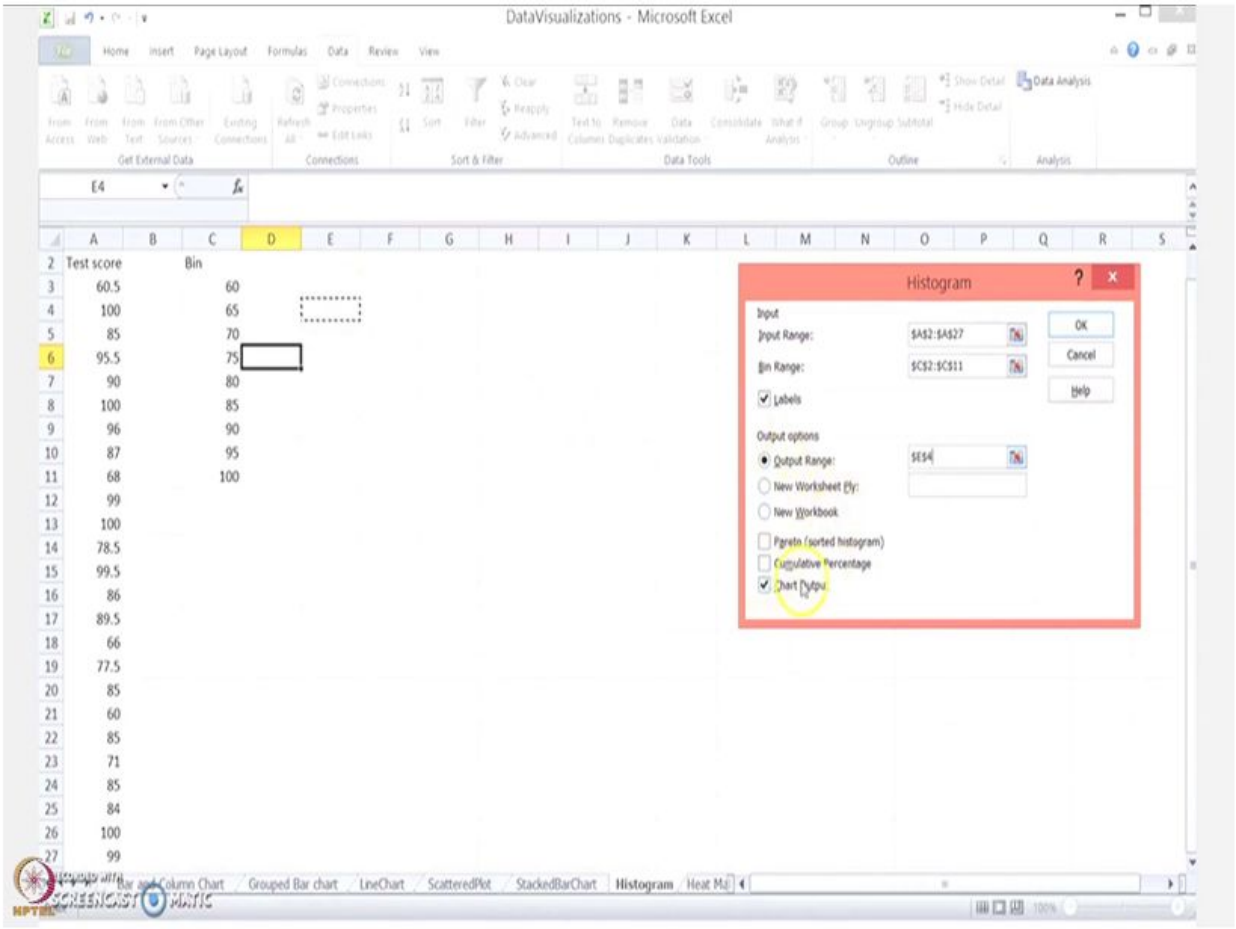




Let us see how to construct a histogram. You need to install the analysis tool pack. For that, go to files, under options, click on add-ins, choose the analysis tool pack, click on go, check box of analysis tool pack should be checked and say okay. Once it is done, go to data and you will see data analysis over here under the analysis category.

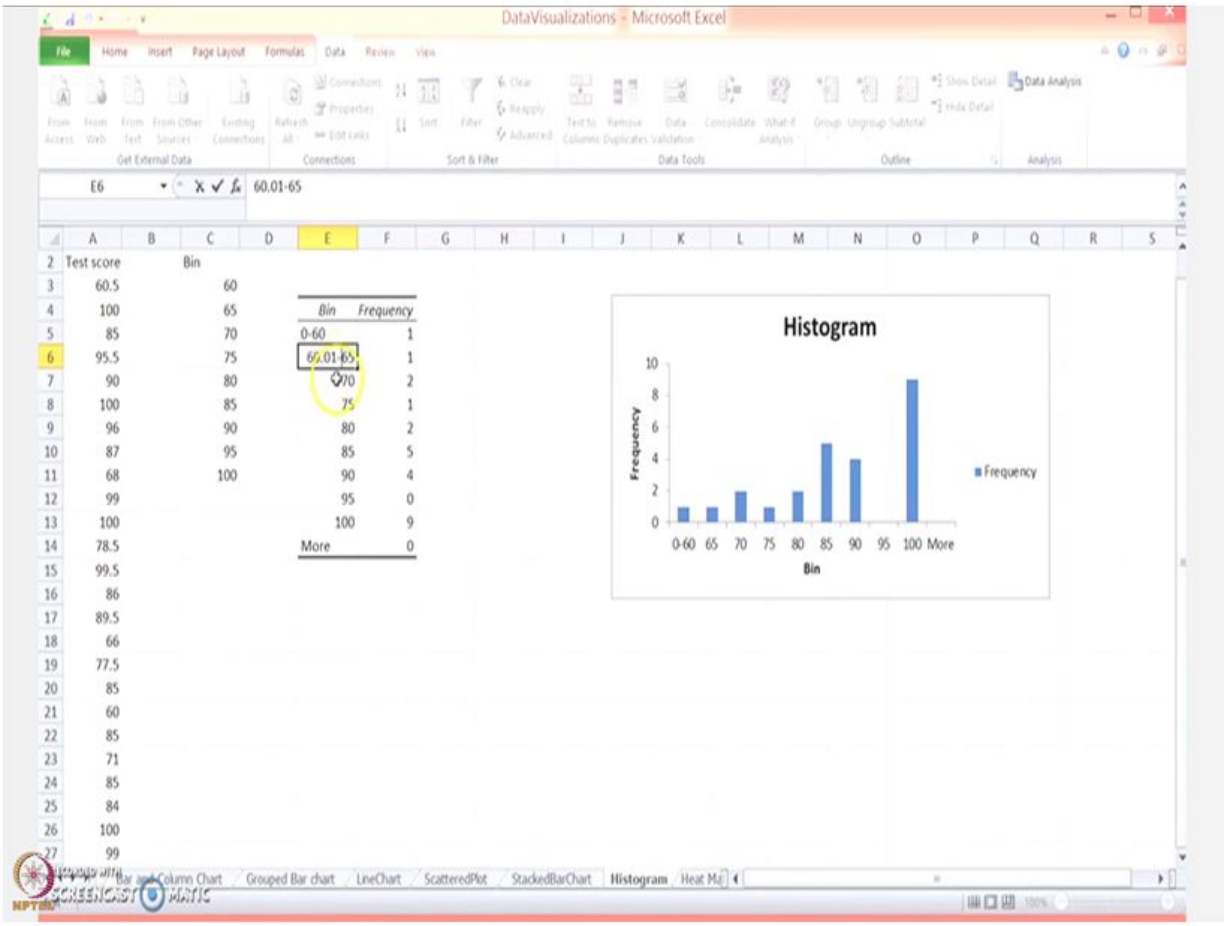
Now, histogram gives you the frequency distribution of the data within particular ranges that is called bin. I have this data over here that is the test score of students from a class and this is the bin range that I have set, that is 60-100, so this would be 0-60, 60-65, 65-70 and so on.

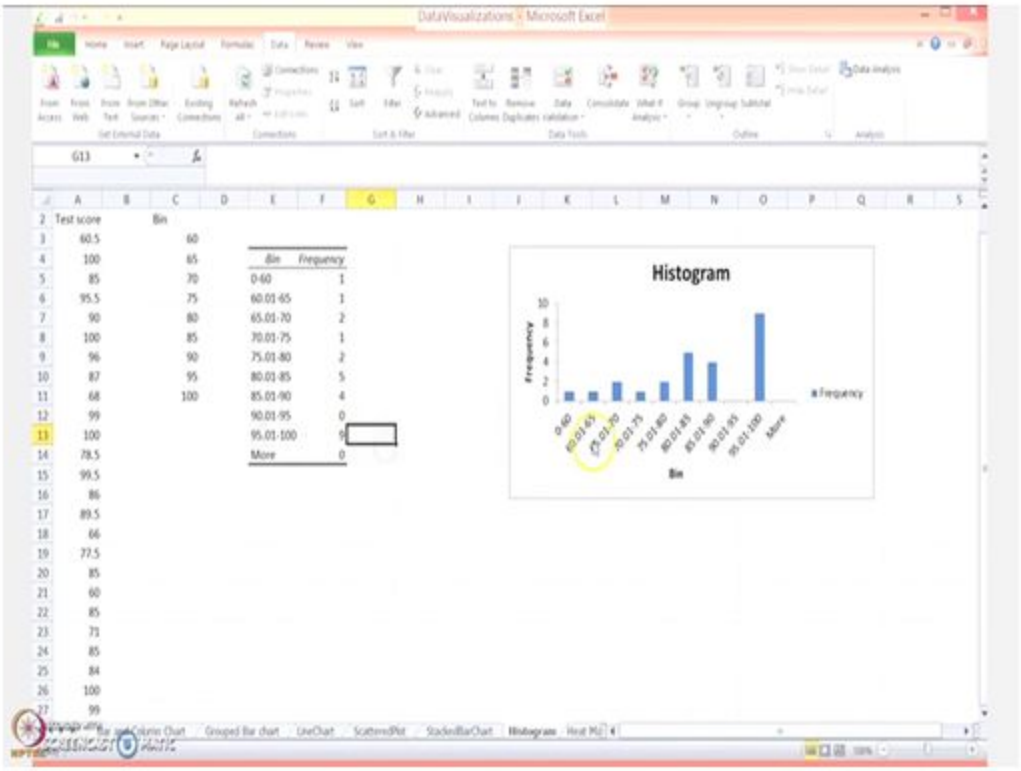
(Refer Slide Time: 13:27)

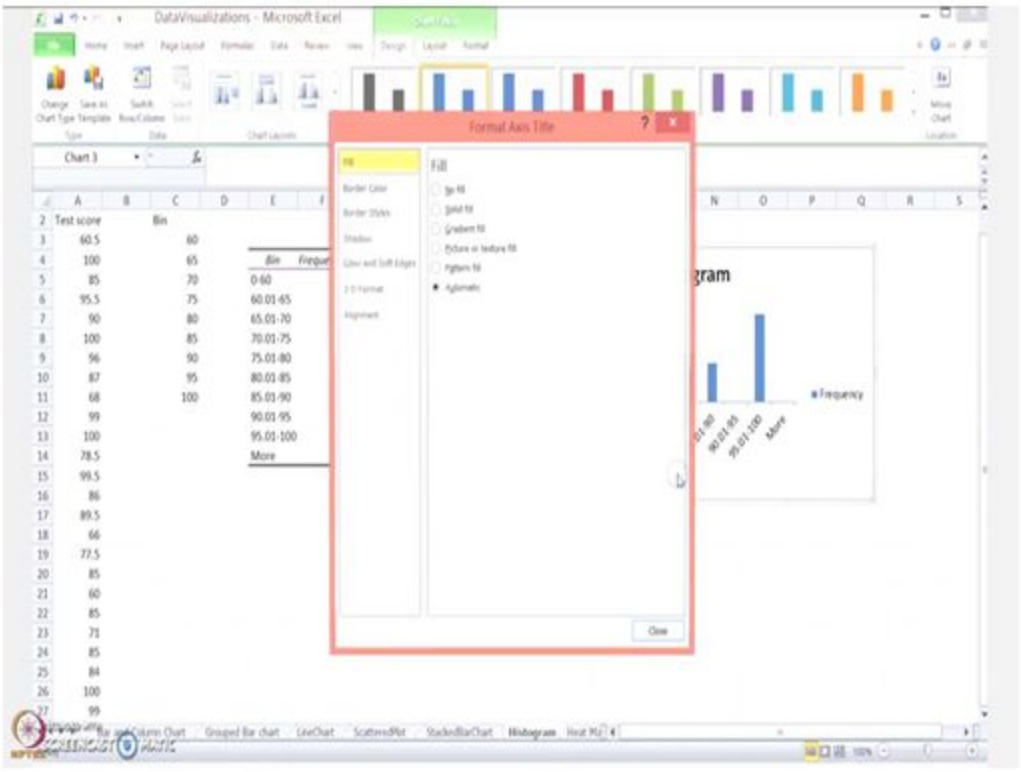


So, to construct a histogram on this data, click on data analysis in the data section, choose the histogram option and say okay. I need to choose the input range, the bin range for the histogram. Input range will be the test score column, so for that choose the entire column, in the bin range choose the bin column, do not forget to check the labels because they have also included the test score in the data. Now, since we want the output in the same excel, choose the output range and give the reference address. Do not forget to check the chart output since we also want the histogram chart.

(Refer Slide Time: 14:28)

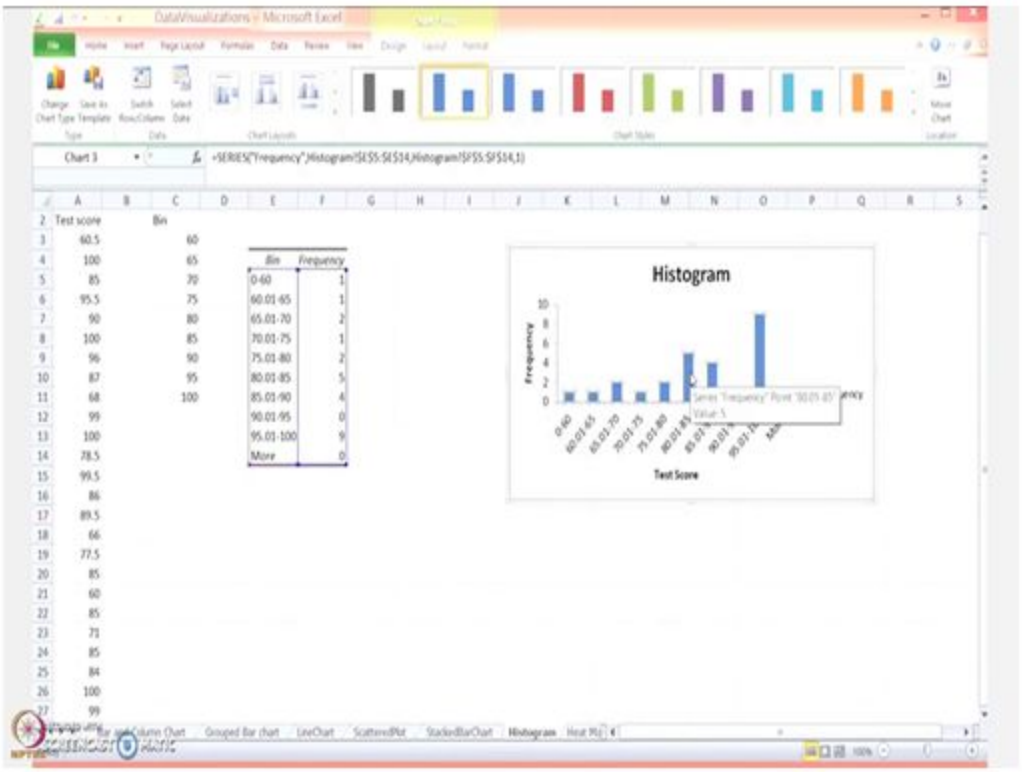

DataVisualizations - Microsoft Excel
E6
60.01-65
Test score
Bin
Bin Frequency
0-60 1
60.01-65 1
70 2
75 1
80 2
85 5
90 4
95 0
100 9
More 0
Histogram
Frequency
Bin
0-60 65 70 75 80 85 90 95 100 More
Frequency
Bar and Column Chart
Grouped Bar chart
LineChart
ScatteredPlot
StackedBarChart
Histogram

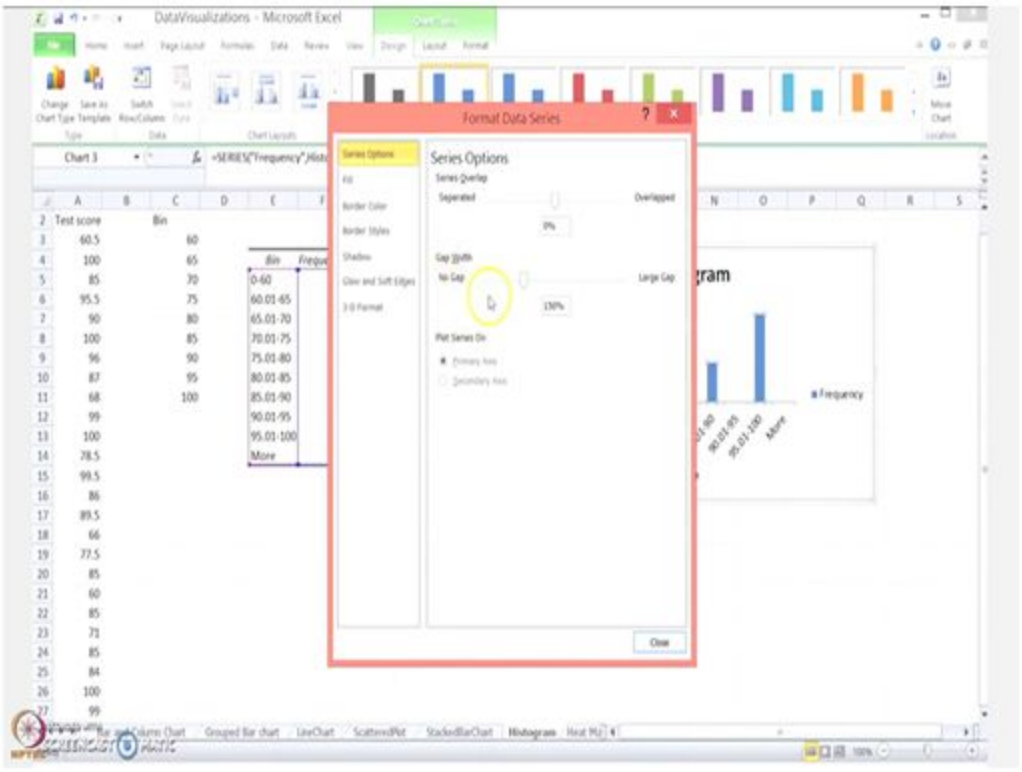

DataVisualizations - Microsoft Excel
Histogram
Frequency
Bin
Bar and Column Chart
Grouped Bar chart
LineChart
ScatterndPlot
StackedBarChart
Histogram

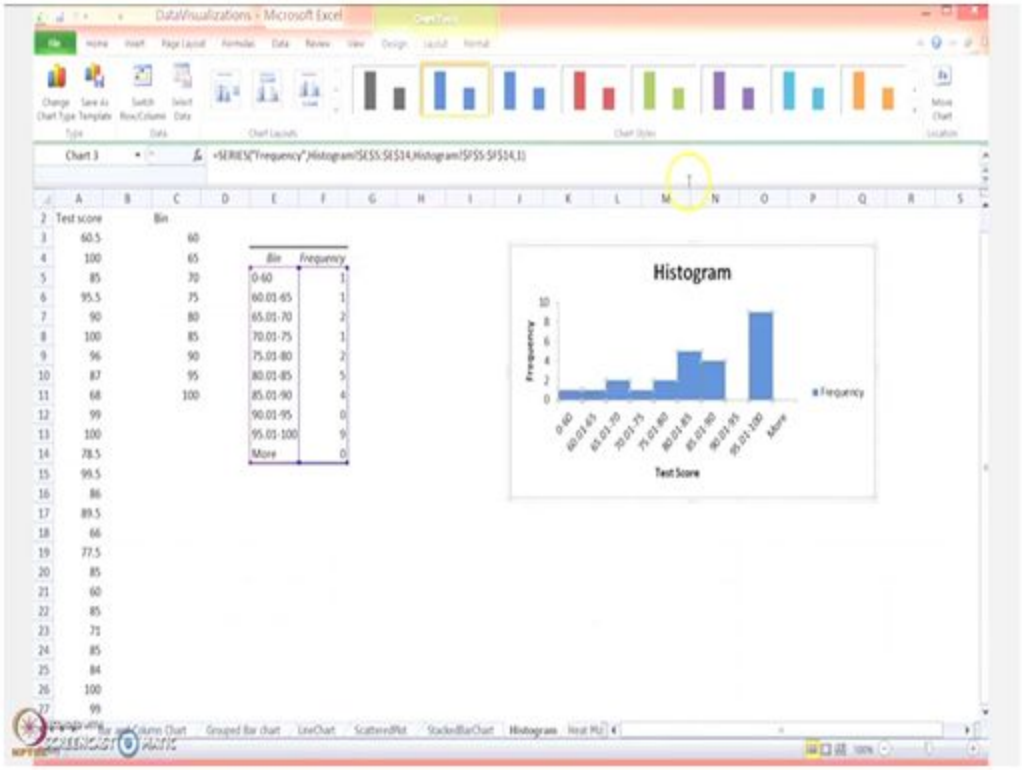

DataVisualizations - Microsoft Excel
Format Axis Title
Fill
Close
Frequency
Bar and Column Chart
Grouped Bar chart
LineChart
ScatterndPlot
StackedBarChart
Histogram

Now, this is how the histogram looks like, let me modify it a bit by expanding it, so the values on the X-axis are a bit misleading, so let me modify these values. 60 would be 0-60, this would be 60.01 to 65, then 65.01 to 70, and 95.01 to 100. There you go, this gives a better clear picture. I can even change the name of the X-axis variable.

(Refer Slide Time: 15:17)

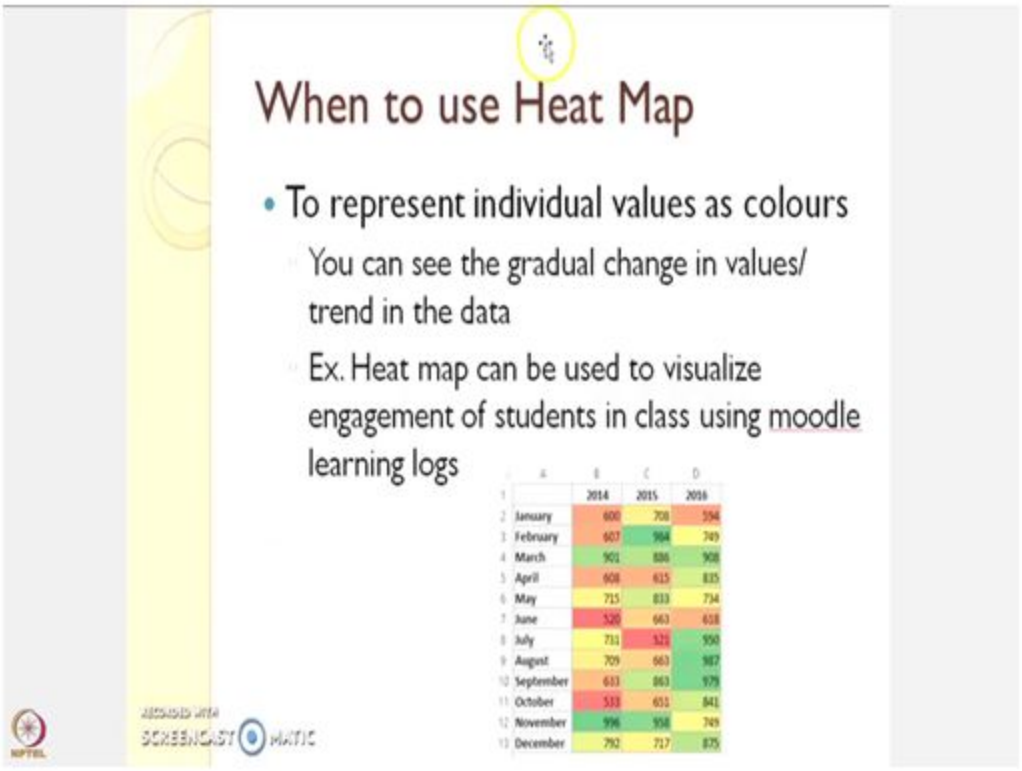

DataVisualizations - Microsoft Excel
Chart 3
=SERIES("Frequency",Histogram!$E$5:$E$14,Histogram!$F$5:$F$14,1)
Test score
Bin
Bin
Frequency
0-60
60.01-65
65.01-70
70.01-75
75.01-80
80.01-85
85.01-90
90.01-95
95.01-100
More
Histogram
Frequency
Test Score


DataVisualizations - Microsoft Excel
Format Data Series
Series Options
Series Overlap
Separated
Overlapped
Gap Width
No Gap
Large Gap
Plot Series On
Primary Axis
Secondary Axis
Close
Frequency

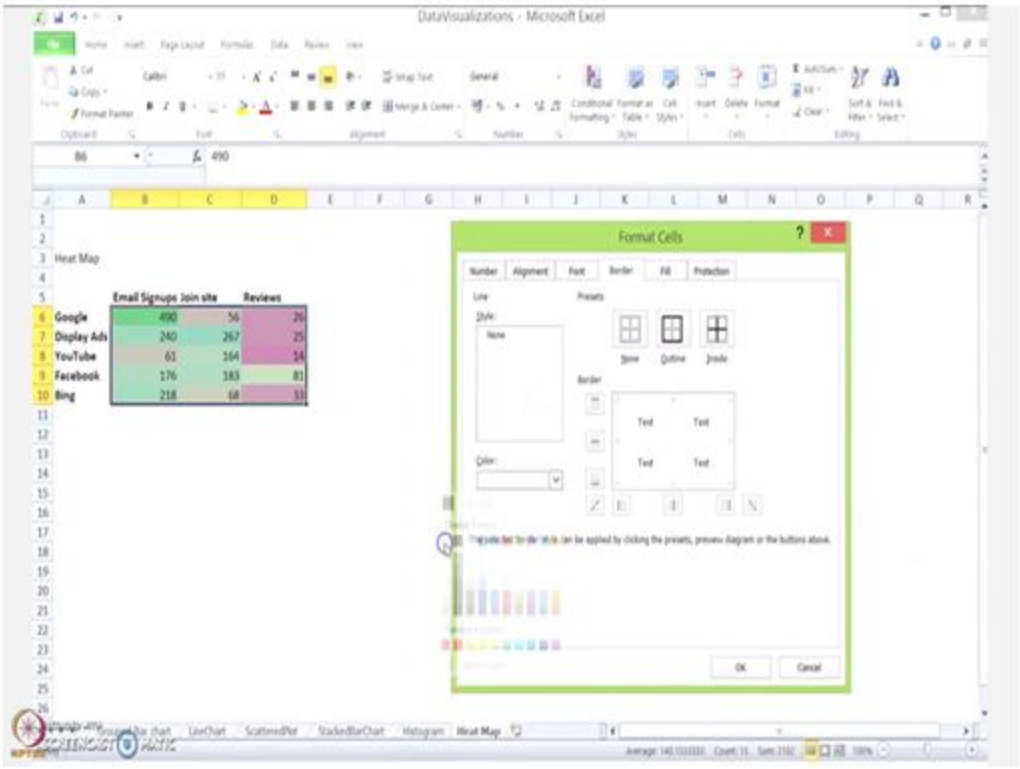


So, this, let me write it as a test score. If you go to see, histogram basically gives a continuous distribution of the data without any gaps. So now, since these gaps are over here, let me modify this chart a bit. Click on the bars, right-click, choose the format data series. Under the series option, you have the gap width option over here. Make the gap width to 0 per cent and click on close. This is an appropriate histogram for the data.

(Refer Slide Time: 15:55)

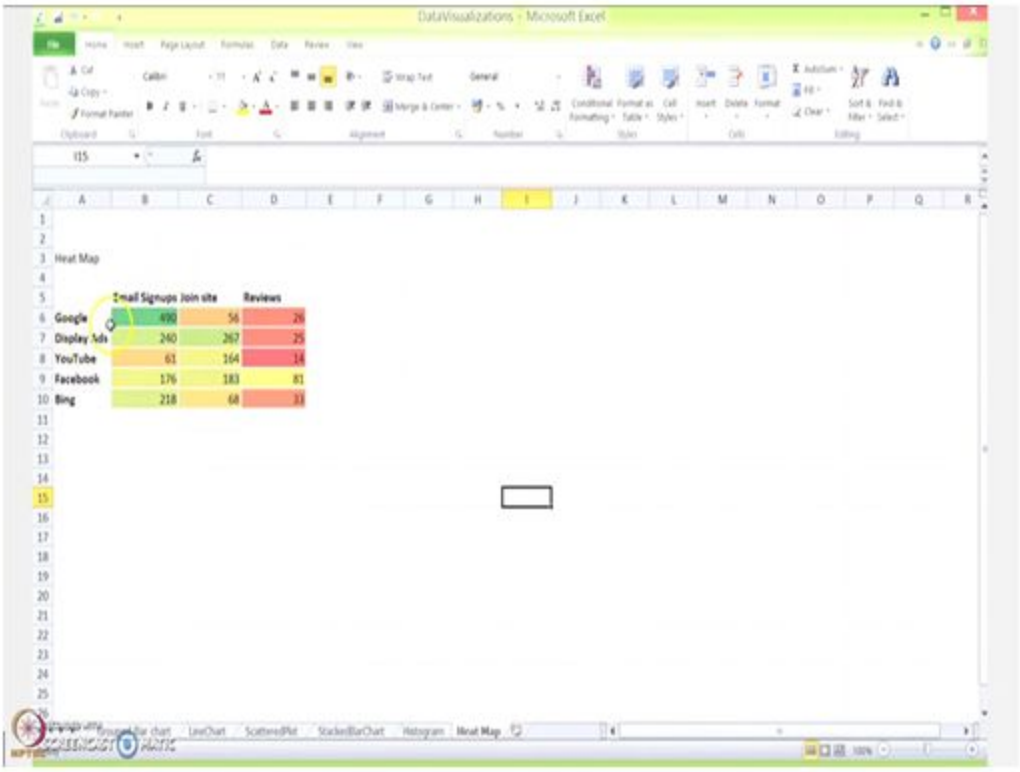


And the last one heat map. A heat map is used to represent the individual values as colours. You can see the gradual change in the trend of the value in the data. For example, a heat map can be used to visualize the engagement of students in class using the moodle learning logs.
(Refer Slide Time: 16:16)



Let us see how to construct a heat map. So, I have this data over here which has internet facilities on one side as one variable and the second variable would be the email signups, the join site, reviews for these respective facilities. Now, in order to construct the heat map for these, I need to select the numerical values only. So, let me select numerical values. Under home, go to condition formatting, click on this and choose the colour scales. In this, choose the first one. Let me format this table a bit, right-click, format cells, border, I choose this border for colour white and I will say okay.

(Refer Slide Time: 17:11)



So, this is how a heat map looks like. So, the heat map will give you relationship between two variables, one plotted on each axis, so you get to see the change in the colours through the axis, you get to see the pattern in the changing trend of the two variables.