INDIAN INSTITUTE OF TECHNOLOGY, KANPUR

NPTEL

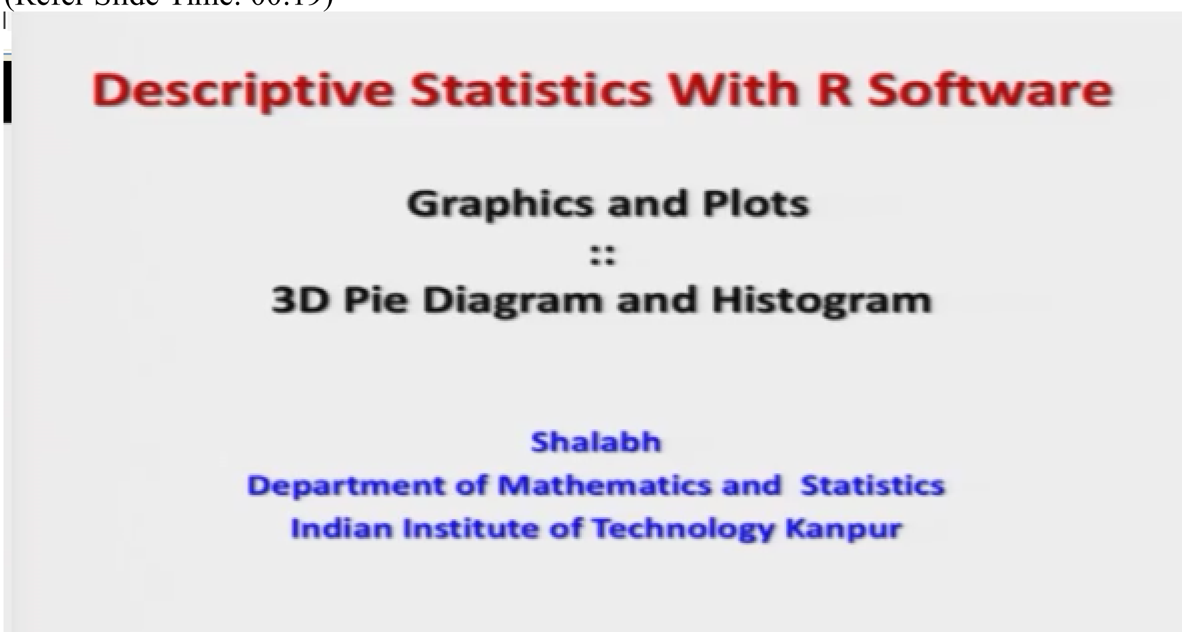
NATIONAL PROGRAMME ON TECHNOLOGY ENHANCED LEARNING

COURSE TITLE
DESCRIPTIVE STATISTICS WITH R SOFTWARE

LECTURE 12
GRAPHICS AND PLOTS – 3D PIE DIAGRAM AND HISTOGRAM

BY
PROF. SHALABH
DEPARTMENT OF MATHEMATICS AND STATISTICS
IIT KANPUR

Welcome to the next lecture on the course Descriptive Statistics with R Software.
(Refer Slide Time: 00:19)

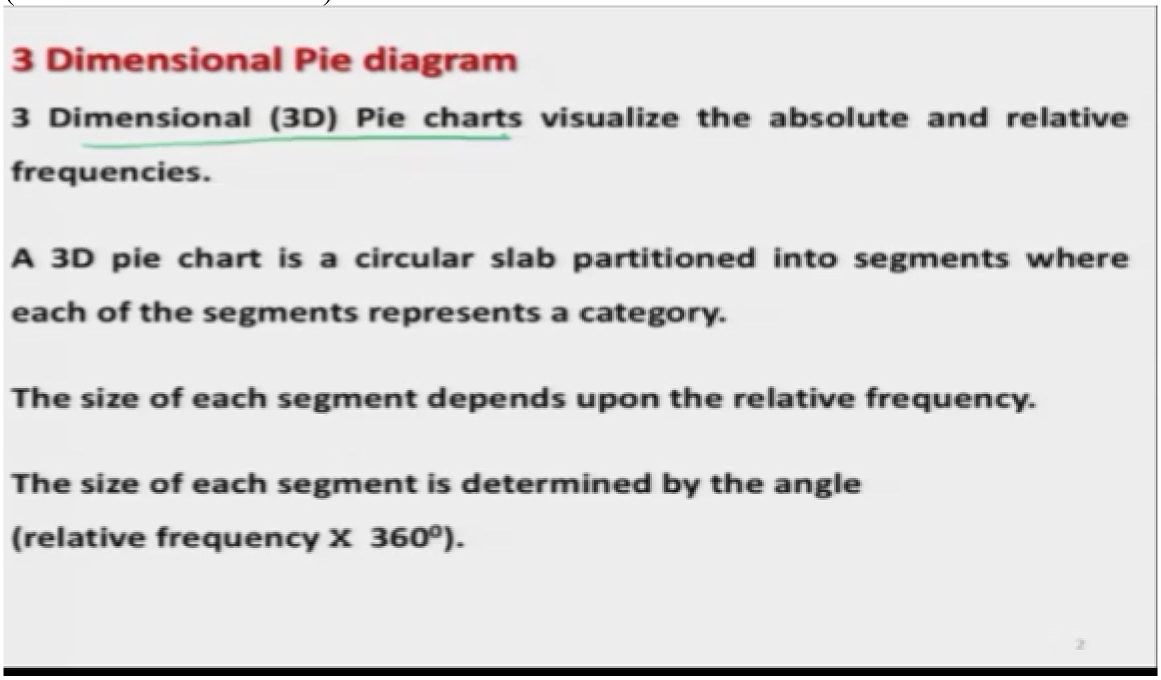



You may recall that we had a discussion on different types of graphics in the last lecture, and we had concluded our lecture whether discussion on pie diagrams.

So in this lecture I'm going to address two topics, two more types of graphics one is 3 dimensional pie diagram and another is histogram. This pie diagram and 3 dimensional pie diagram they are more or less similar, the only difference is in their look, the construction, the structure and the interpretation, they are the same as in the case of pie diagram.

So let us start our discussion first with the 3 dimensional pie diagram as in case of pie diagram there are different slices, and those slices represent the absolute or the relative frequencies.

(Refer Slide Time: 01:29)

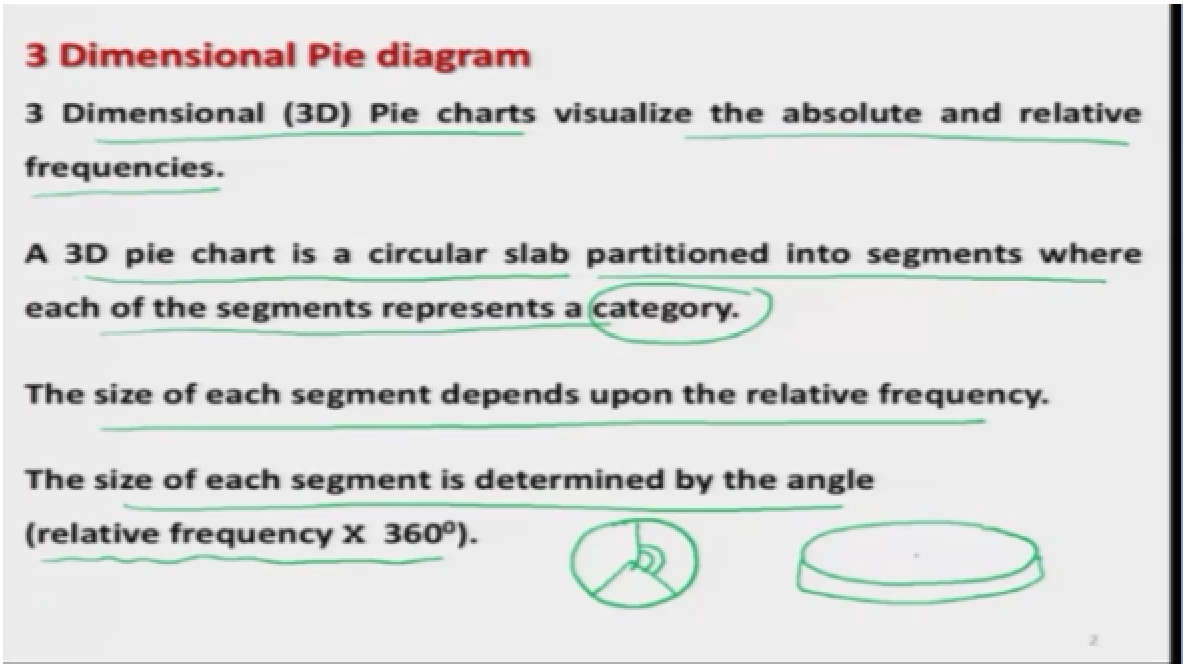



Similarly in case of 3 dimensional pie charts or 3 dimensional pie diagram they also represent the absolute and relative frequencies. The difference between a pie diagram and a 3 dimensional pie diagram is that in case of pie diagram there is a slice, but in a 3 dimensional pie diagram there is a circular slab and this slab is partitioned into different segments or slices, and every segment or every slice represents a category of the frequency distribution and the size of each segment this depends on the relative frequency, and this is the same case as it happens in the case of pie diagram also.

And here also the size of each segment is determined by the angle, and the angle is determined by the same formula as in the case of pie diagram that is relative frequency into 360 degree, so a pie diagram is a circular diagram which is partitioned into different segments and the size of the segment is determined by the angle.

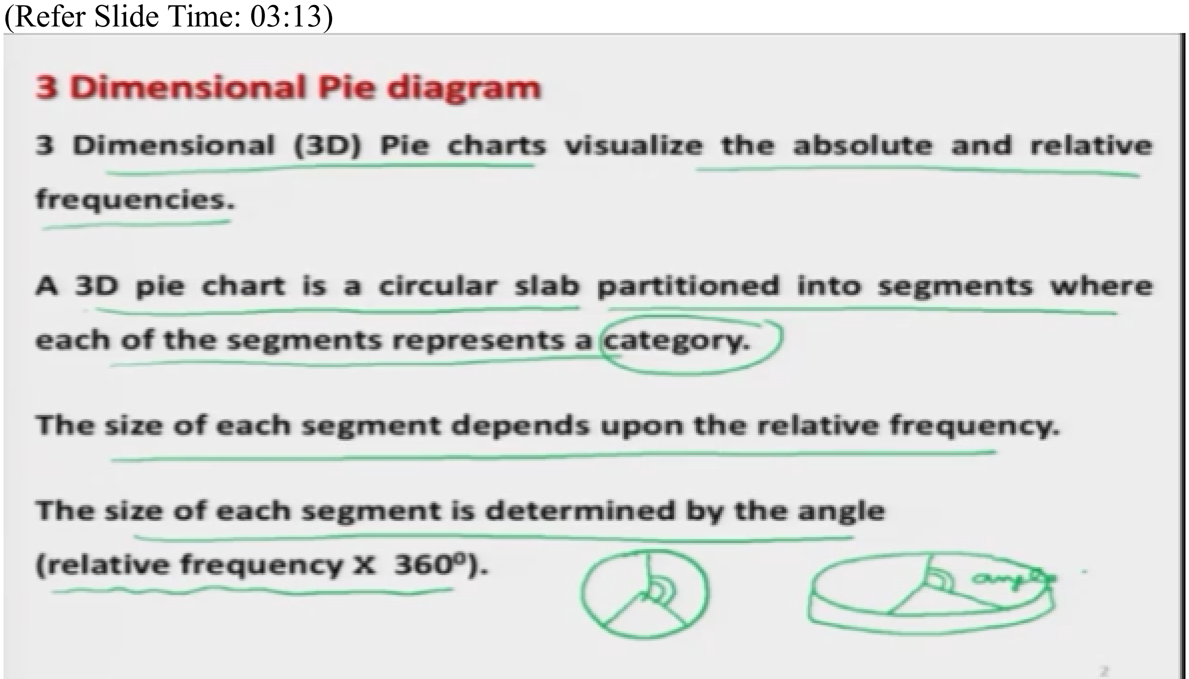
Similarly the 3 dimensional pie diagram this is a sort of circle having a third dimension as its height,

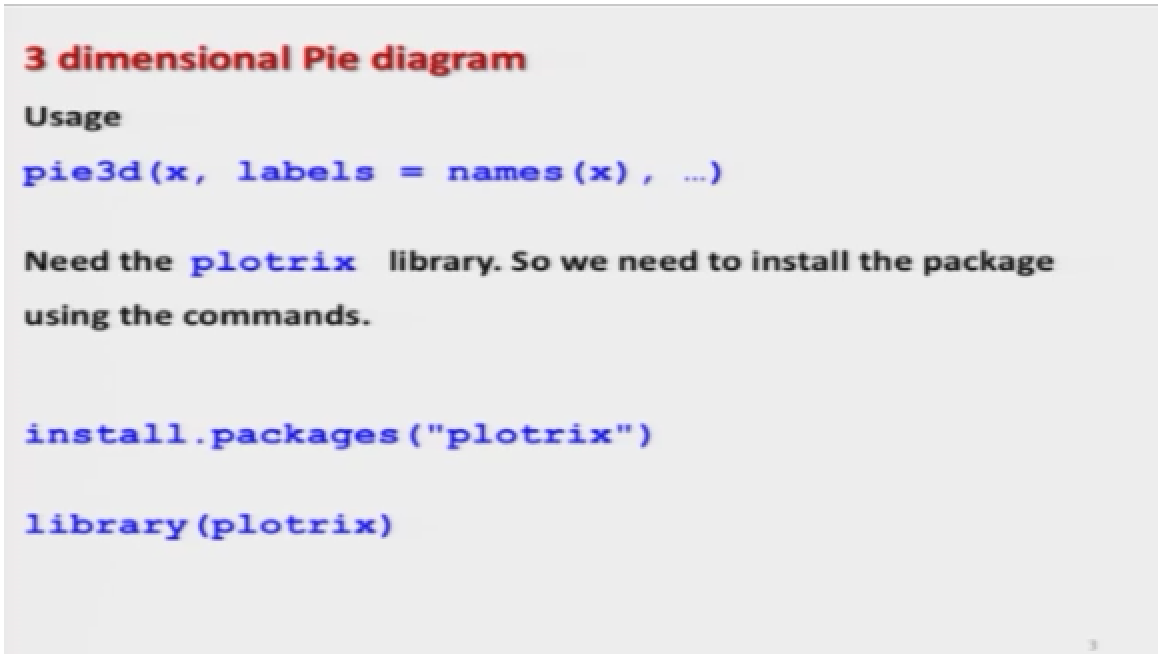
(Refer Slide Time: 03:00)




and same way as in the case of pie diagram we create the slices and the size of the slice is again determined by the angle.

(Refer Slide Time: 03:13)

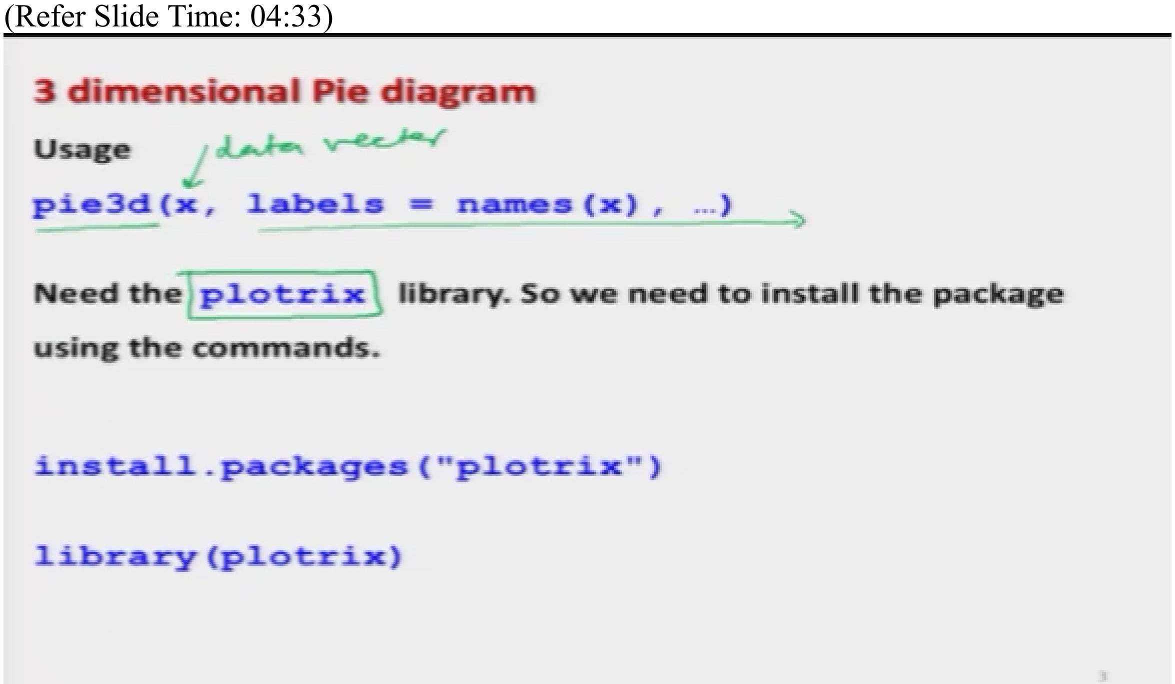





So we again now try to first understand how to create a 3 dimensional pie diagram in R software, so in order to construct the 3 dimensional pie diagram we have a command here, pie3d, pie that is small letters and 3 number and d, and then here is the data vector, exactly as in the case of pie diagram.

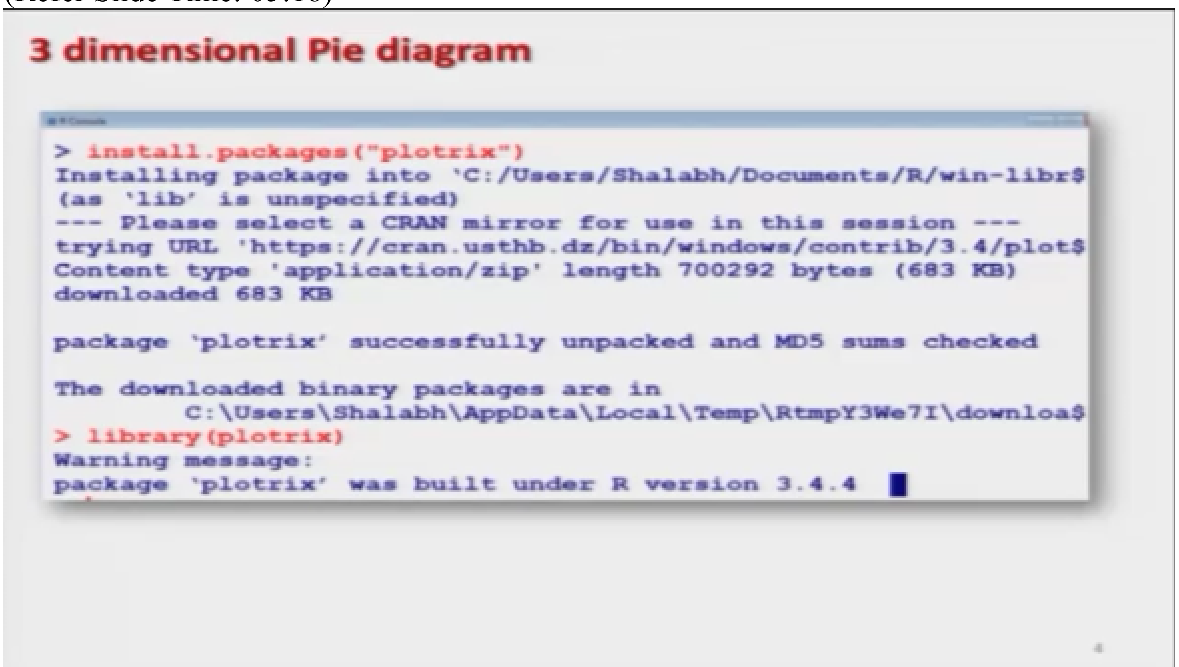
And then there are certain parameters which are given for different types of options like as labels and another things, the difference between pie diagram and 3 dimensional pie diagram is that construction of pie diagram is the part of the based package of R that is inbuilt in the base package, but in order to construct the 3 dimensional pie diagram we need to install a package or a library, so in order to do it we need here a library plotrix, PLOTRIX,

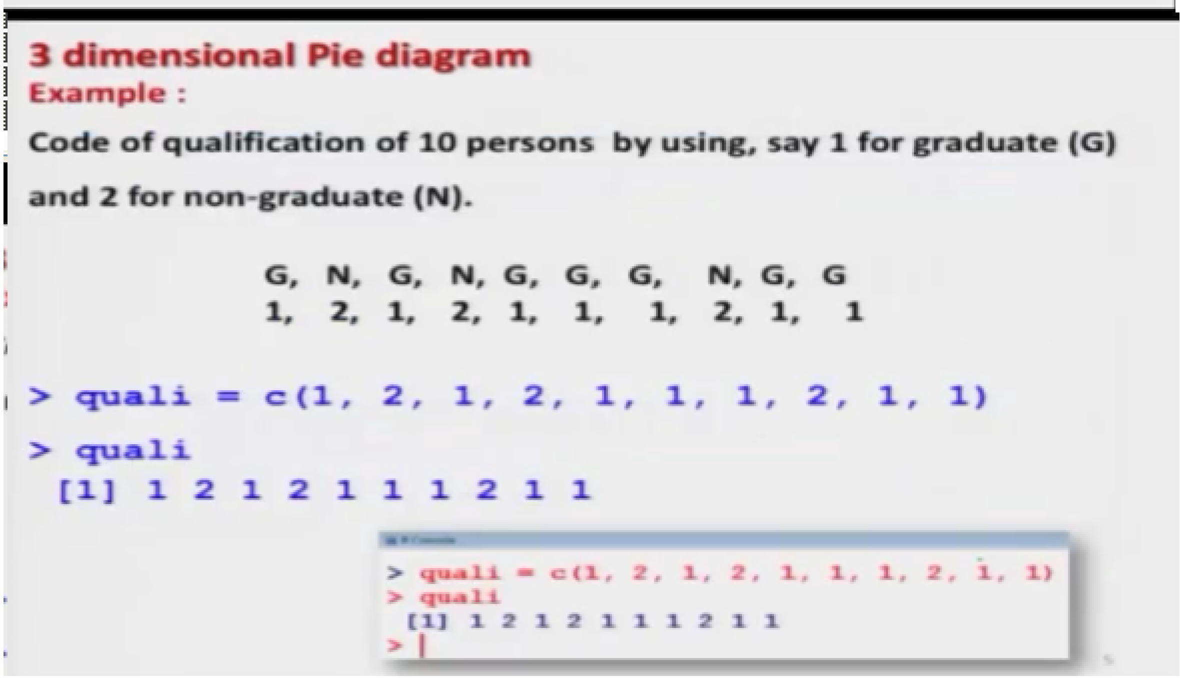
(Refer Slide Time: 04:33)



so first we need to install this library using the command install.packages and inside the arguments, inside double quotes we have to write plotrix, PLOTRIX, and once I do it then I have to use the command library plotrix.

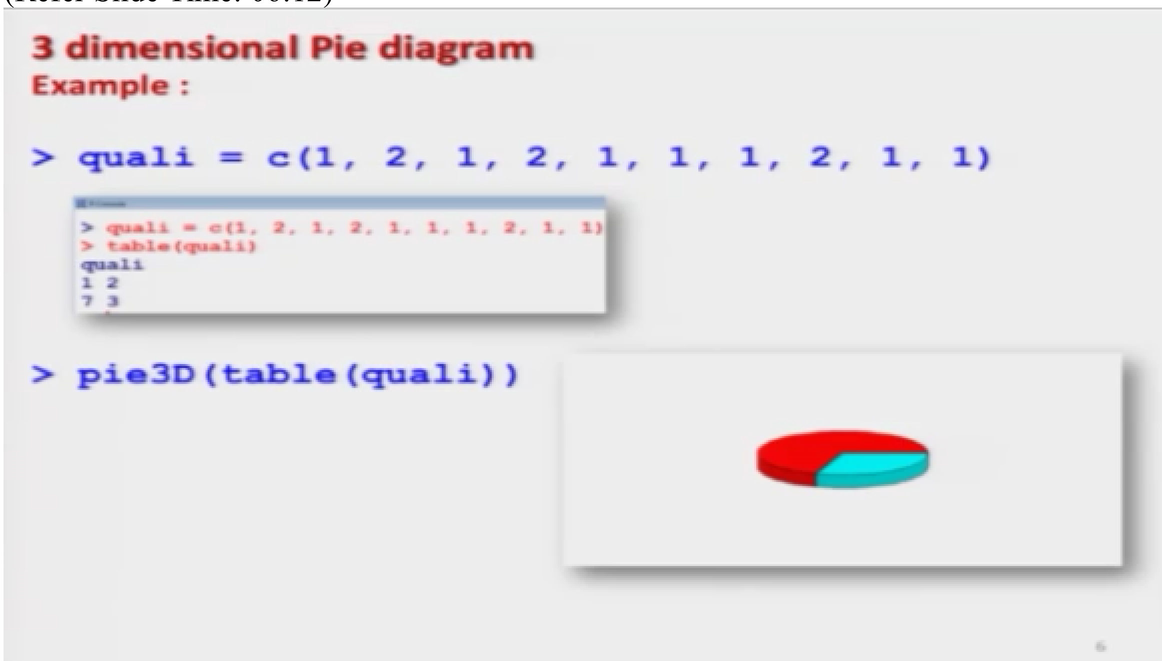
In case if you execute these two commands on the R console you can get the library plotrix on your computer and if you try to see I have installed this package on my computer and this is the screenshot and so on,
(Refer Slide Time: 05:18)



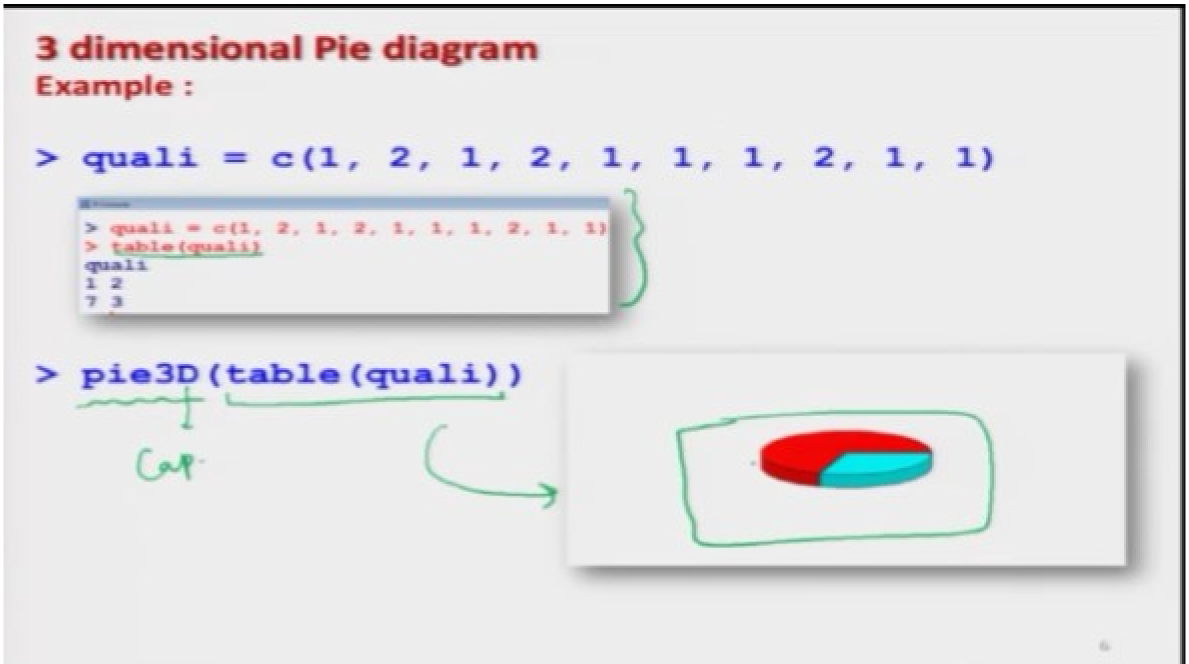

now I will try to take some examples to show you how to create the 3 dimensional pie diagram, so once again I'm continuing with the same example that I consider in the case of pie diagram that I have a data on 10 persons and we have recorded their educational qualification in 2 categories graduate and non-graduate and this data has been indicated by 1 for graduate and 2 for non-graduate and we have this data vector and we have stored the data in a variable name quali, and so now we have here this data vector quali consisting of two numbers 1 and 2 and we would like to create a 3 dimensional pie diagram for this data.

(Refer Slide Time: 06:12)



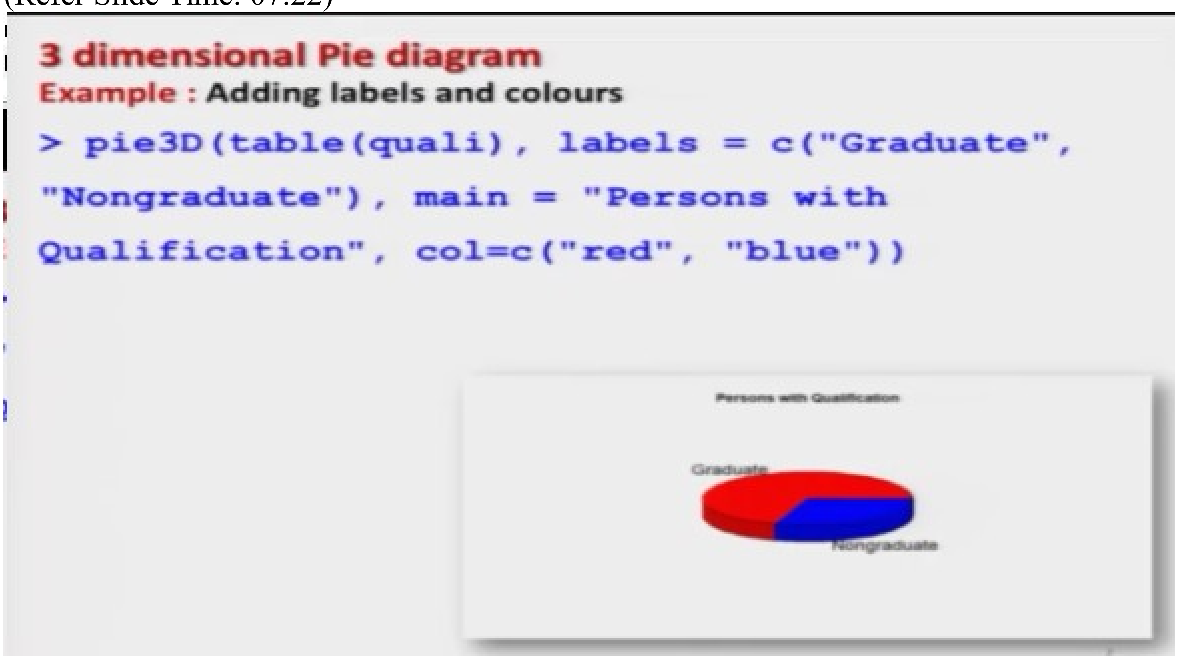

So obviously as we have discussed earlier that whenever you want to create pie diagram or say 3 dimensional pie diagram you need to input that data in the form of frequencies, so what I'll try to do? First I would try to create the frequency table using the data quali, using the command table quali and you can see here I already had done it in the last lecture, so I'm simply reproducing here a screenshot, and after this you simply have to use the command pie3D, remember one thing D here is in capital letter, and then you have to use the data that is obtained by table quali.

And once you do so you will get here an outcome like this one, so you can see here this is your 3 dimensional pie diagram,
(Refer Slide Time: 07:13)




the third dimension has been added by this height here, and in case if you want to make it here more informative

(Refer Slide Time: 07:22)



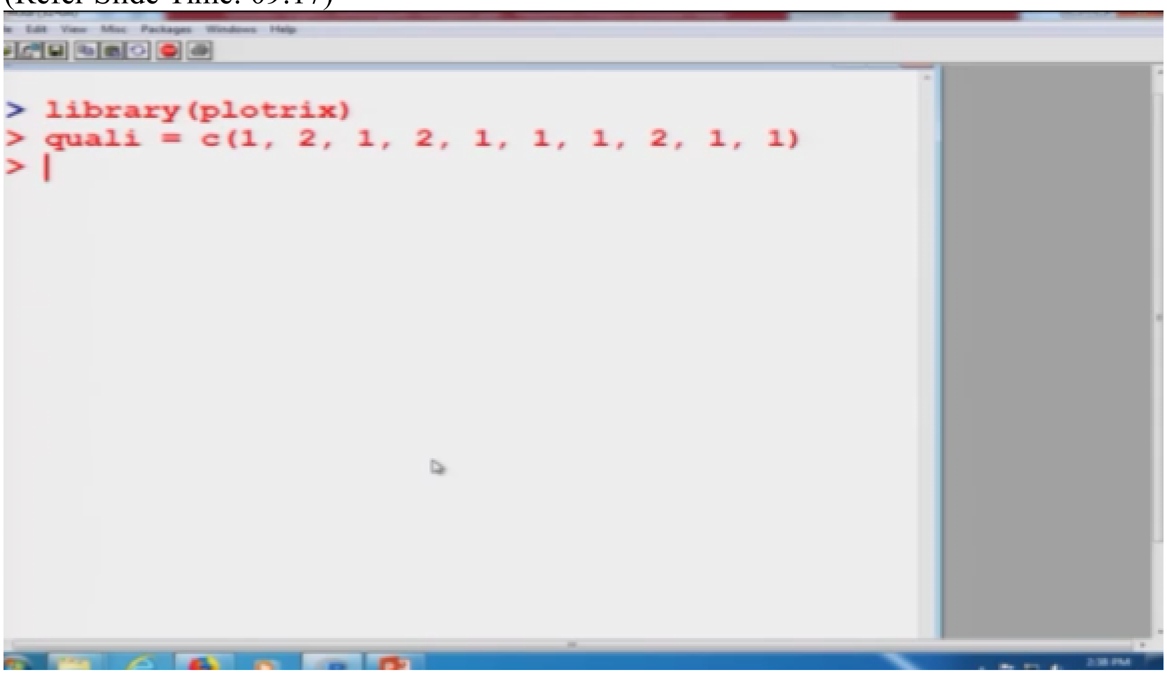

by adding the names to the slices like as non-graduate and graduate and if you want to add here title of the graphics like as persons with qualification, and if you want to change the colours you can use the similar commands what we used in the case of pie diagram. For example if you try to see pie3D table quali this is the same command that we used earlier.

And now in order to give here two categories graduate and non-graduate I'm using here are parameter labels, LABELS, labels equal to the graduate and undergraduate whatever we want to give the name inside the double quotes, and these two values are combined in a vector using the C command.

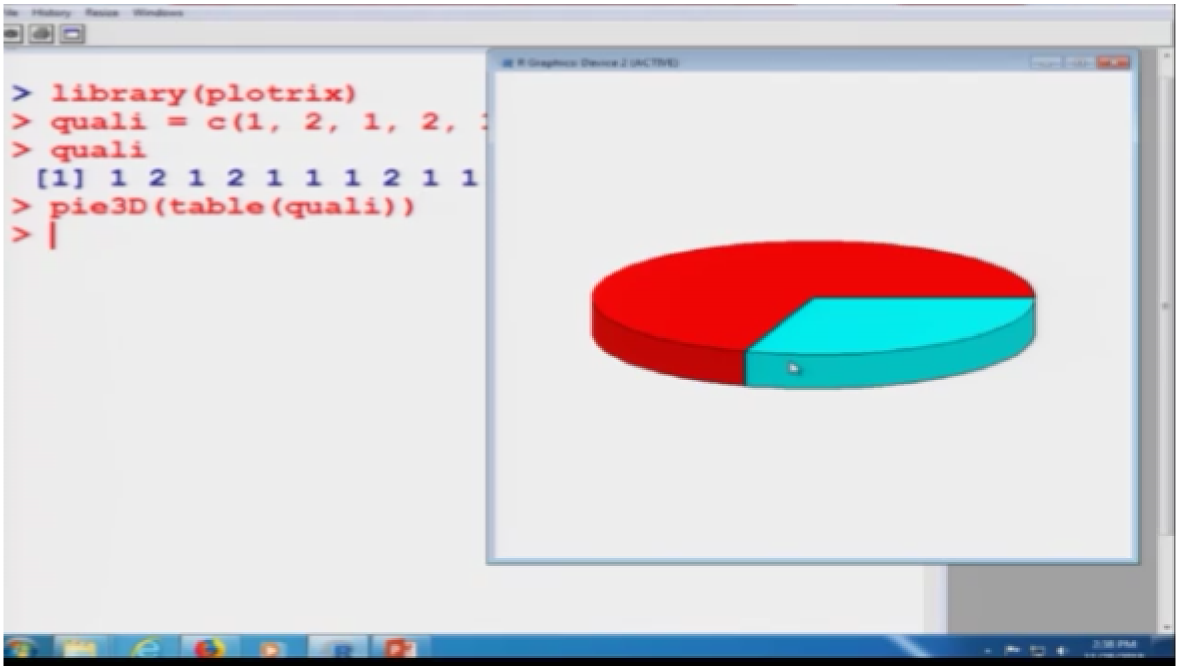
And similarly if you want to give here that title, this title is given by the parameter main, so I have to write main is equal to an inside the double quotes I have to write what word is that, title I would like to have.

And then you can see here one slice is in red colour, and another slice is in blue colour, so once again I will use the similar command here COL = red and blue that is the R command for the two colours inside the double quotes and separated by comma and they are combined with the C operator and I will mention here the colours of this one. And once you try to do it you will get a 3 dimensional pie diagram like this.

Now I would like to show it to you on the R console, so first I try to load here the library so you can see now there is no error, the library has been uploaded,
(Refer Slide Time: 09:17)

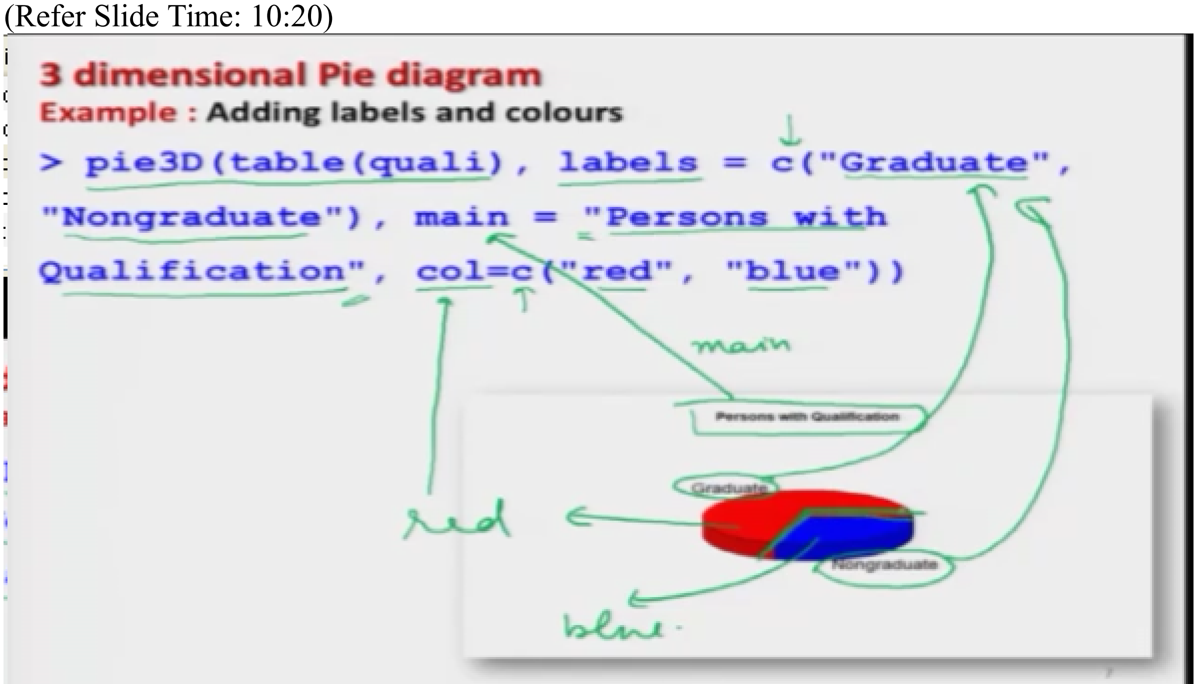



and now I defined here the data quali, and if I want to make it here the 3 dimensional pie diagram I have to first create the frequency table of the quali by, and if you try to do it here I get here a 3 dimensional pie diagram like this one.
(Refer Slide Time: 09:45)

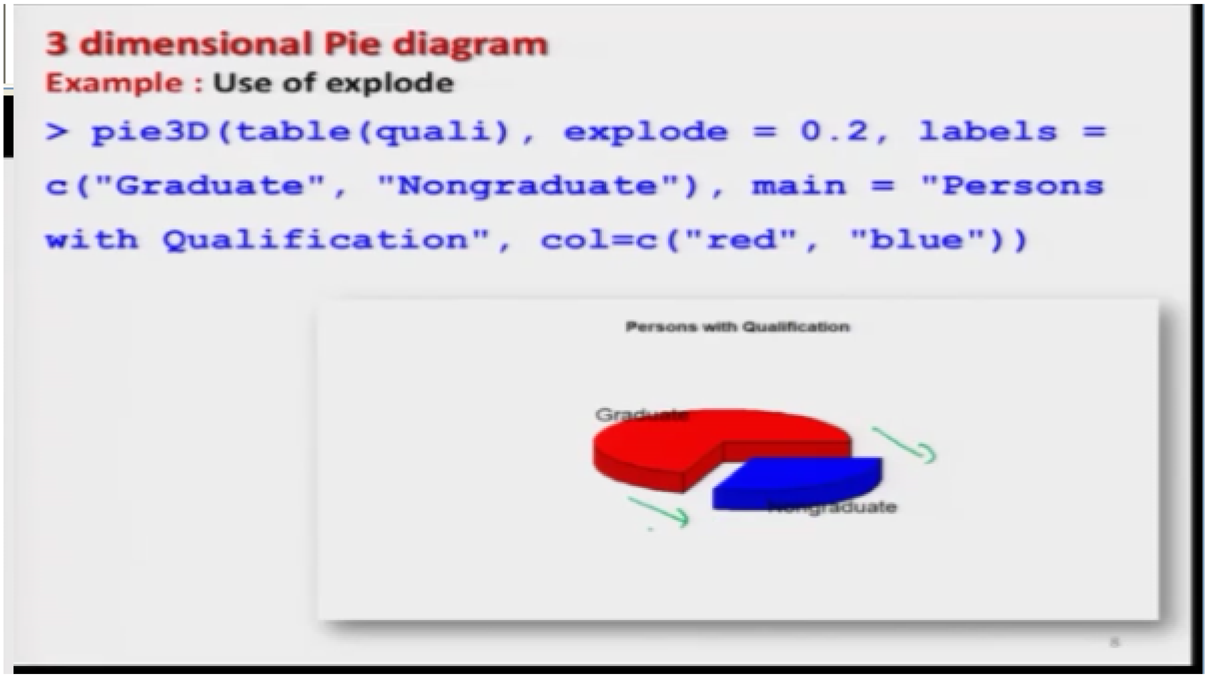



And similarly in case if you want to add here more information I can execute the same command over here, I will try to copy and paste the same command and you can see here that you are getting the same graph that you had obtained, right.

Now I will try to show you another feature in the same 3 dimensional plot, you can see here that here these two slices are joint,
(Refer Slide Time: 10:20)

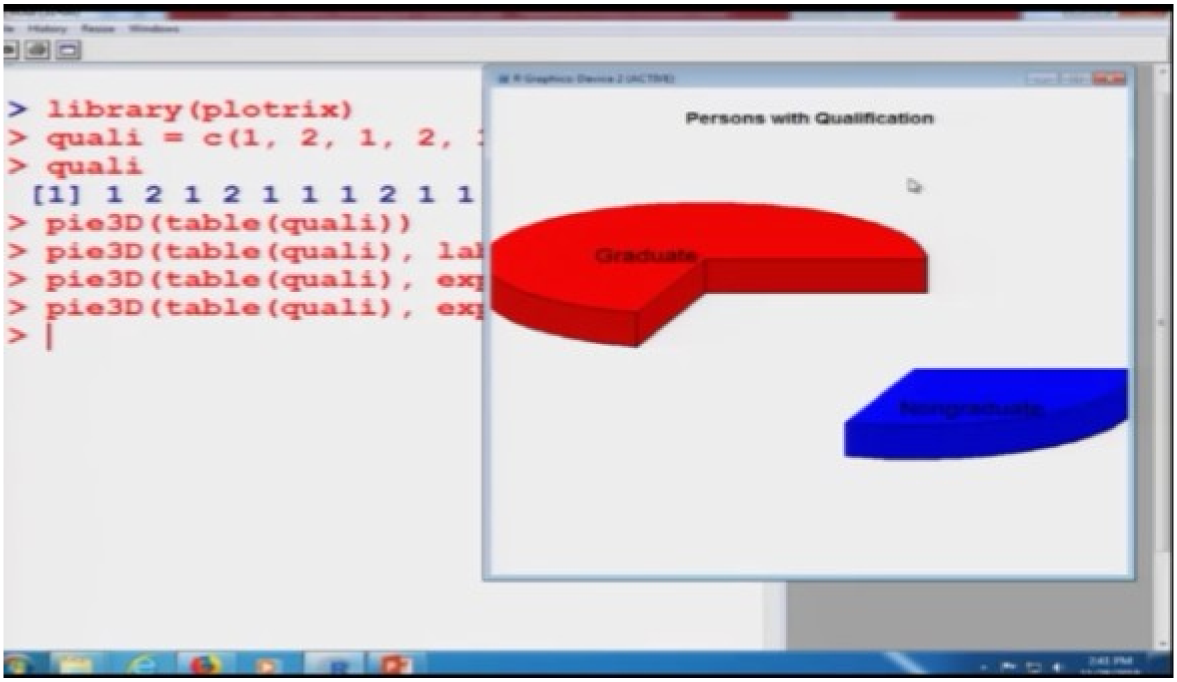



in order to make it more informative I can separate it, so that the graphic will look like this that you can see here that the red and blue parts are separated.
(Refer Slide Time: 10:35)




In order to make this type of graph I can use here one parameter that is explode, so you see here in this command which is the same as the earlier one, but now I'm adding here a new parameter explode, EXPLODE all in small letter is equal to 0.2, actually this 0.2 is the factor that is going to decide that how much separation do you want for example in this case, this is the space between the two slices or two slabs, so this 0.2 factor is going to determine this thing, so I'll try to show you on the R console so that you are more comfortable and then I'll try to take one more example and I'll try to show you all those things very quickly, so if you try to see here now I have used here the function see here explode, and suppose if you want to, well I'll try to show you that change here, suppose I try to change this explode value, so suppose if I try to make it here this explode is equal to suppose here, instead of here 0.2 suppose I make it here 0.8, you can see here what happens. You can see here now the separation becomes more,

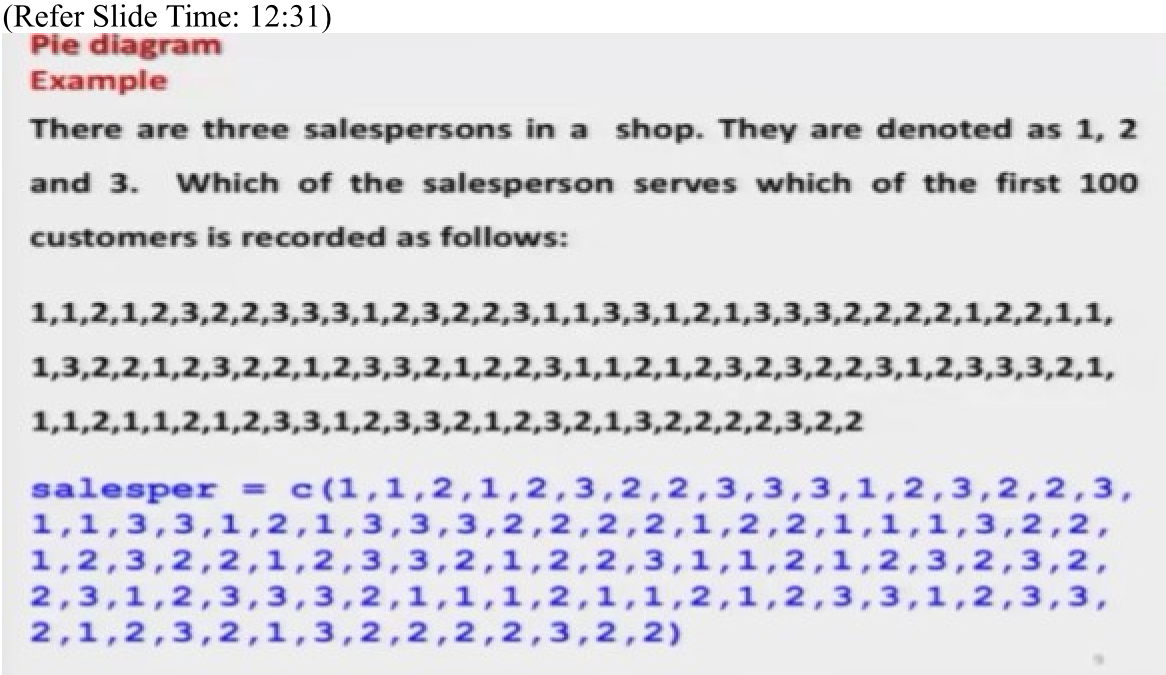
(Refer Slide Time: 12:12)




so by increasing the value of the parameter explode you can increase the separation between the, or I mean the slices.

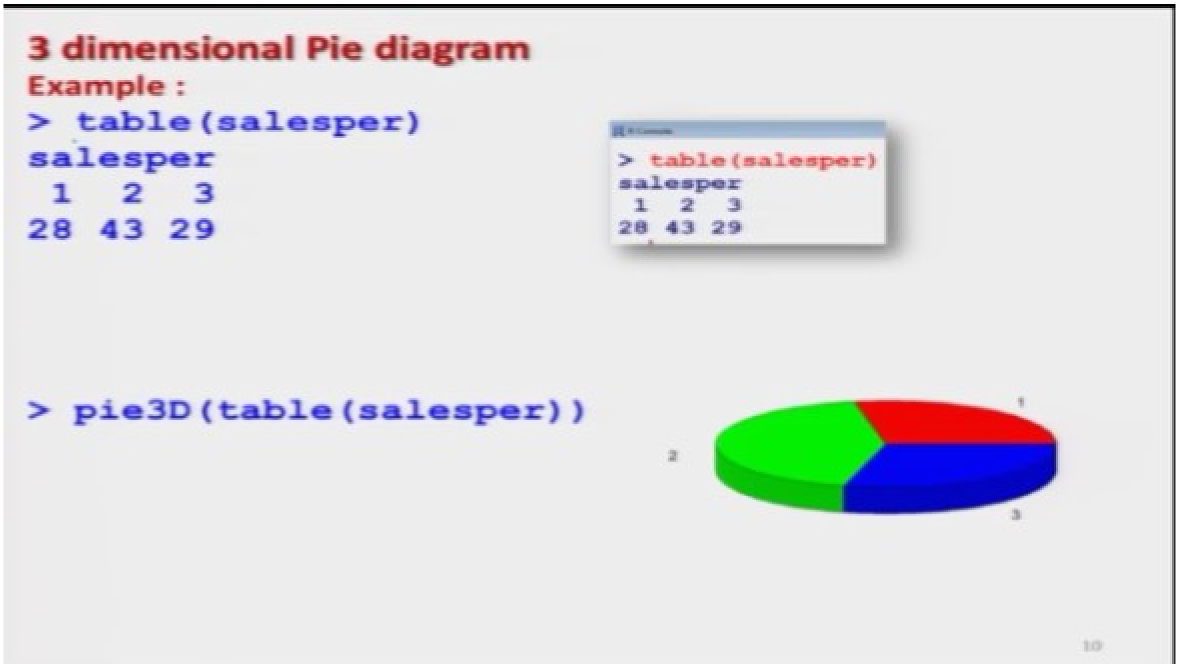
So now I would take one more example to make you comfortable, so once again I'll try to use the same data set which I had used earlier in the pie chart,

(Refer Slide Time: 12:31)



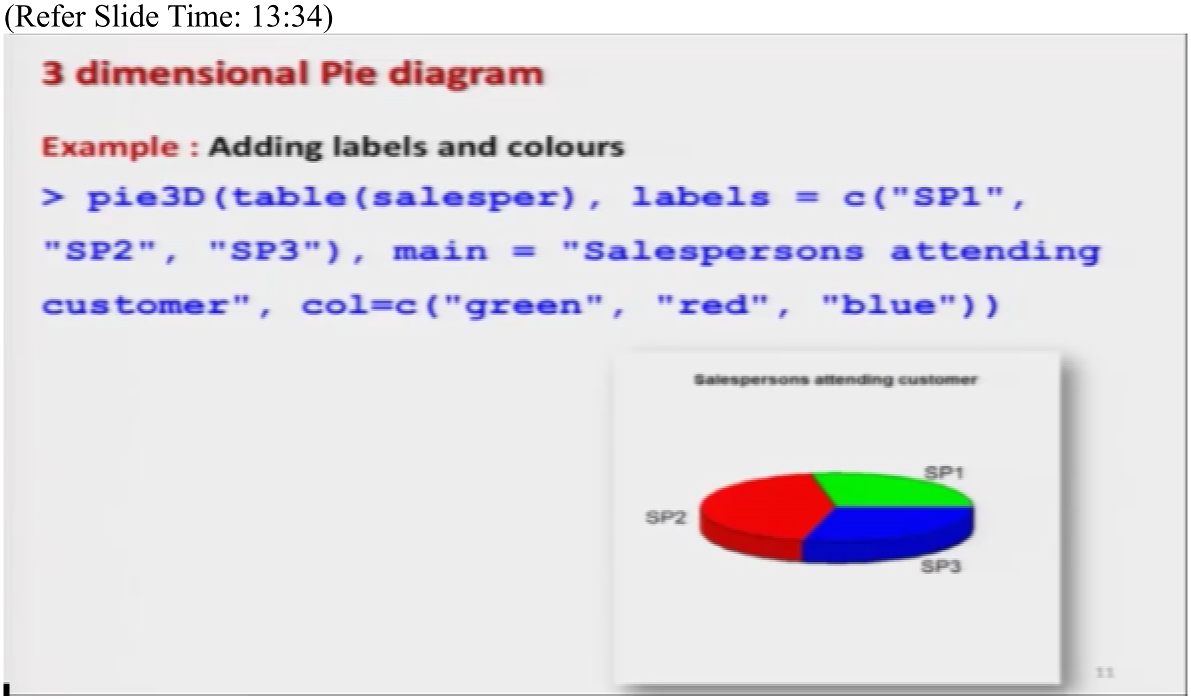

so that was about the 100 values or that was the data on the 100 customers, they were entertained by 3 sales person, 1, 2 and 3, and this is the data here that was stored in our variable salesper, and now I'm trying to create the frequency table using that table command,

(Refer Slide Time: 12:51)



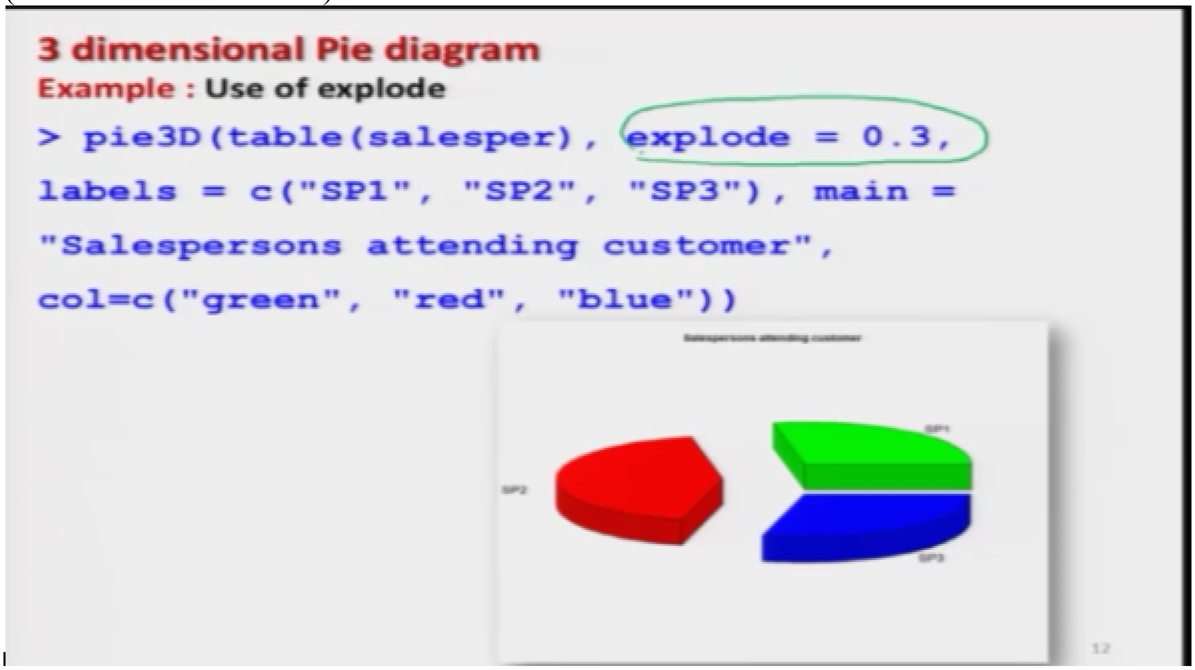

and now you can see here there are three categories 1, 2 and 3, and if I try to make the simple 3 dimensional pie diagram using the command pie3D I simply have to use the same command and I have to change the name of the variable which is now here the salesper, so you can see here this is the standard 3 dimensional pie diagram which is using the default values, so you can see here this 1, 2 and here 3 they are indicating the 3 classes that is the sales person 1, sales person 2 and sales person 3.

(Refer Slide Time: 13:34)



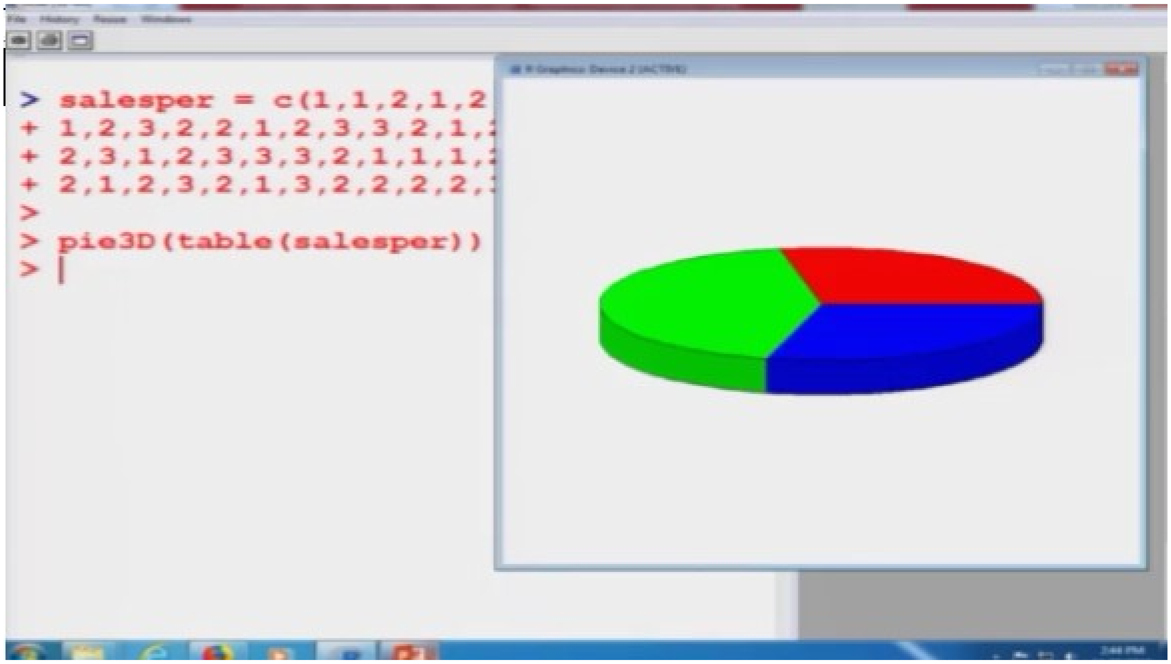

And similarly if you want to give here title or the names to this slabs, right, you can also do it here by using the same command labels, main, colour, but now I have here 3 categories, so now

I'm using here green, red and blue, 3 colours, exactly on the same lines and you will get here this type of plot.
(Refer Slide Time: 13:57)



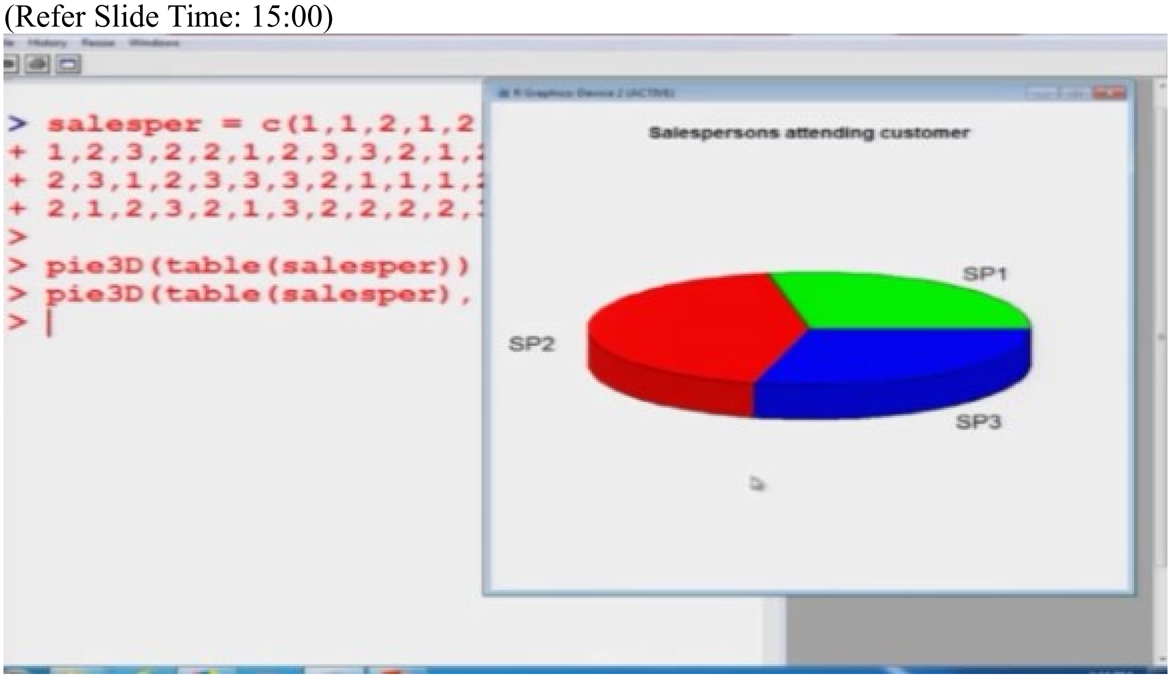

And in case if you want to use the parameter explode, so for example here I'm trying to use here explode = 0.3, then you will get here 3 separated slices, so you can see here these slices are now separated.

So I'd try to show you on the R console also so that you are more comfortable, so first I'll try to copy this data, and then I'll try to make it here pie3D on the sales person, but I need to give it in the form of a table, so table of this one, so right you can see here this is the same graphic that I have just shown you.
(Refer Slide Time: 14:43)

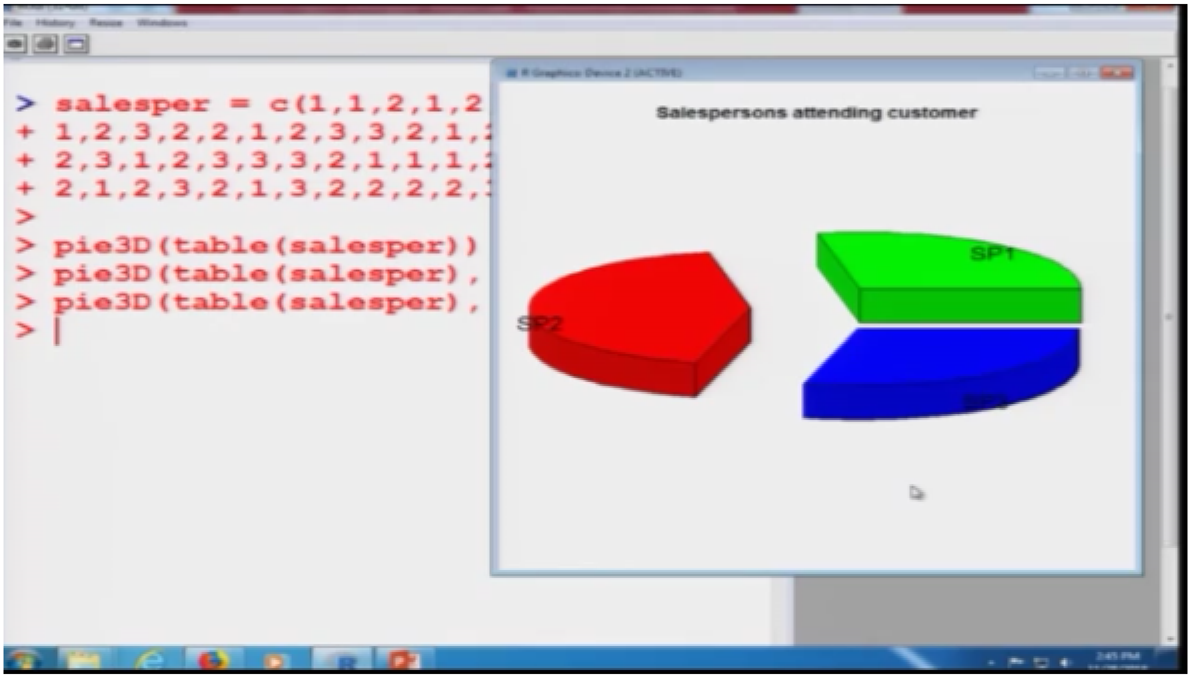



And similarly if you want to make it more clear I can by adding the titles and colours and legends you can use the same command here and I can show you the outcome is coming out to be like this, this is the same output that I just shown you.

(Refer Slide Time: 15:00)

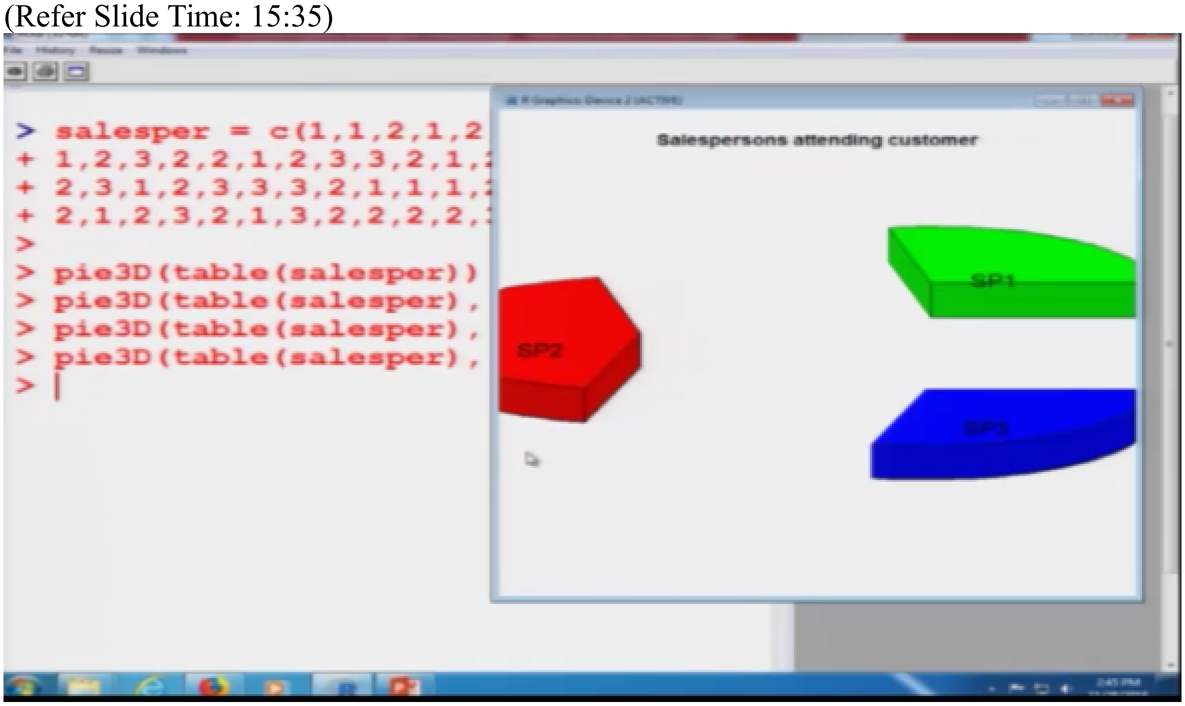



And in case if you want to use here the explode function just add this as one of the parameter inside the arguments and you will see here, now this is separated,
(Refer Slide Time: 15:17)




and once again in case if you want to make the separation bigger you simply have to increase the value of explode, suppose I make it here 0.8, so you can see here now the separation becomes more,

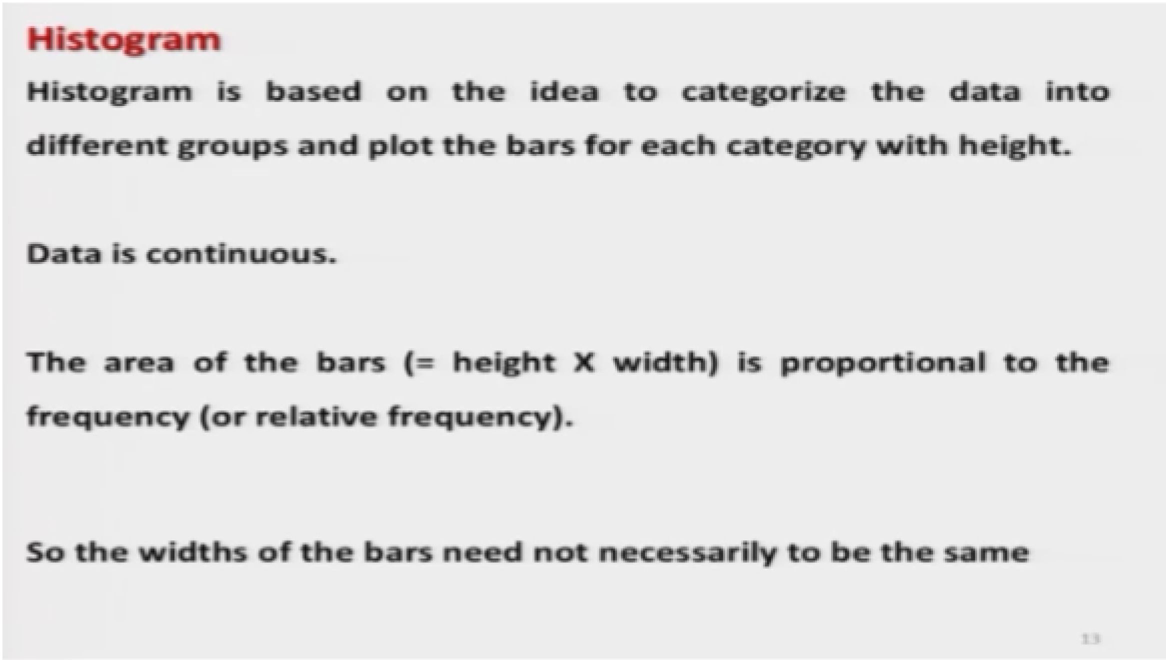
(Refer Slide Time: 15:35)




so now it essentially depends on the choice of the experiment that what exactly he or she wants.

Now after this pie diagram let me try to introduce here histogram,

(Refer Slide Time: 15:48)




so histogram is graphic but this is used for continuous data, you can recall, we had discussed the aspect of discrete data, continuous data and so on, so histogram also does the same thing what a bar diagram or a pie chart does, but the difference is this bar diagram and pie diagrams they are essentially for that categorical variable where the values are indicated by some numbers representing the category, but histogram is for continuous data, so histogram also does the same thing that it first try to categorize the data into different groups, and then it plots the bars for each category and in this case the data is always continuous or I would say that, whenever the data is continuous please plot histogram.

Now there is a difference between the bar plot and histogram, you may recall in case of bar diagram I had told you that the height of the bar is simply proportional to the frequency or relative frequency, width of the bar is immaterial, so we don't bother about it which has no interpretation, but this does not happen in the case of histogram.

The size of the bar is essentially proportional to the area of the bars in case of histogram, so essentially the area of the bar is given by the height of the bar and width of the bar which have to be multiplied, so now in this case you can see here that the bars in the histogram had to be controlled with respect to height and width both, you will notice in most of the cases the width of the bar is kept the same in case of histogram, but the reason for this is just to make it simple to understand, means if you have 2 bars and if you have to compare with their area whereas if you have 2 bars where you have to compare them only with respect to the length or the height of the bar because the width is same, then which is more convenient, obviously the length or the height of the bar is more easy to compare than the area of the bar, okay.

(Refer Slide Time: 18:28)

## Histogram

**Frequency distribution**

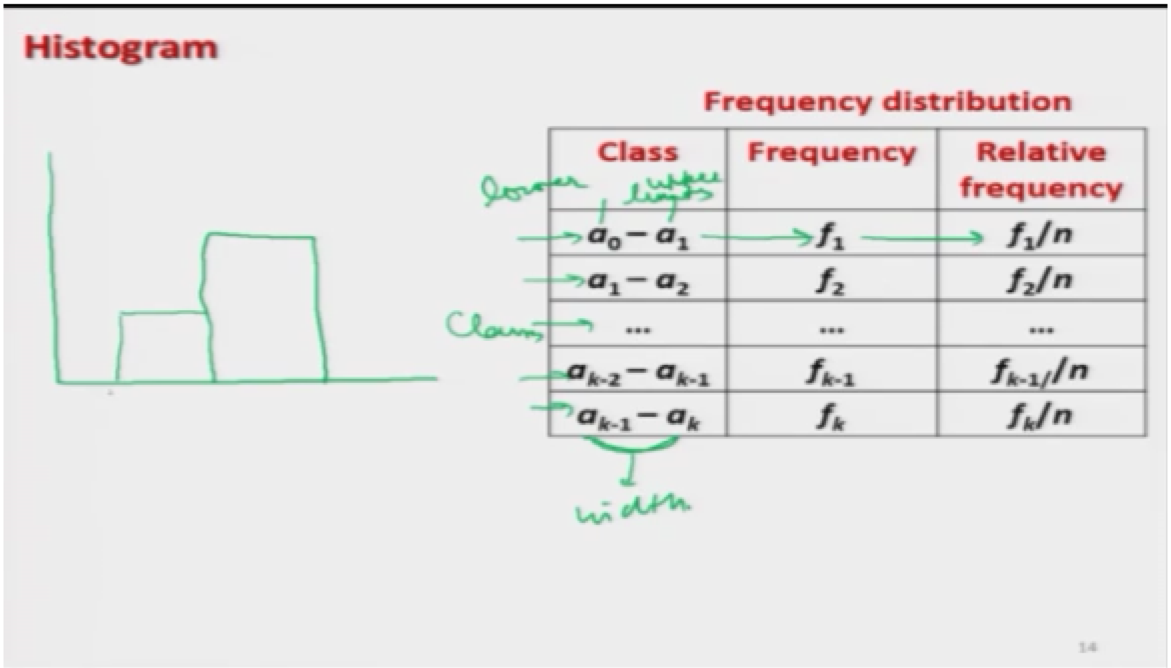
| Class | Frequency | Relative frequency |
|---|---|---|
| $a_0 - a_1$ | $f_1$ | $f_1/n$ |
| $a_1 - a_2$ | $f_2$ | $f_2/n$ |
| ... | ... | ... |
| $a_{k-2} - a_{k-1}$ | $f_{k-1}$ | $f_{k-1}/n$ |
| $a_{k-1} - a_k$ | $f_k$ | $f_k/n$ |

14

Now let us try to see how you are going to create the histogram based on the frequency distribution, you may recall that we had some data, some continuous data and we had discussed that these data is divided into different classes, and those classes have lower limits and upper limits, and this is called the class interval.

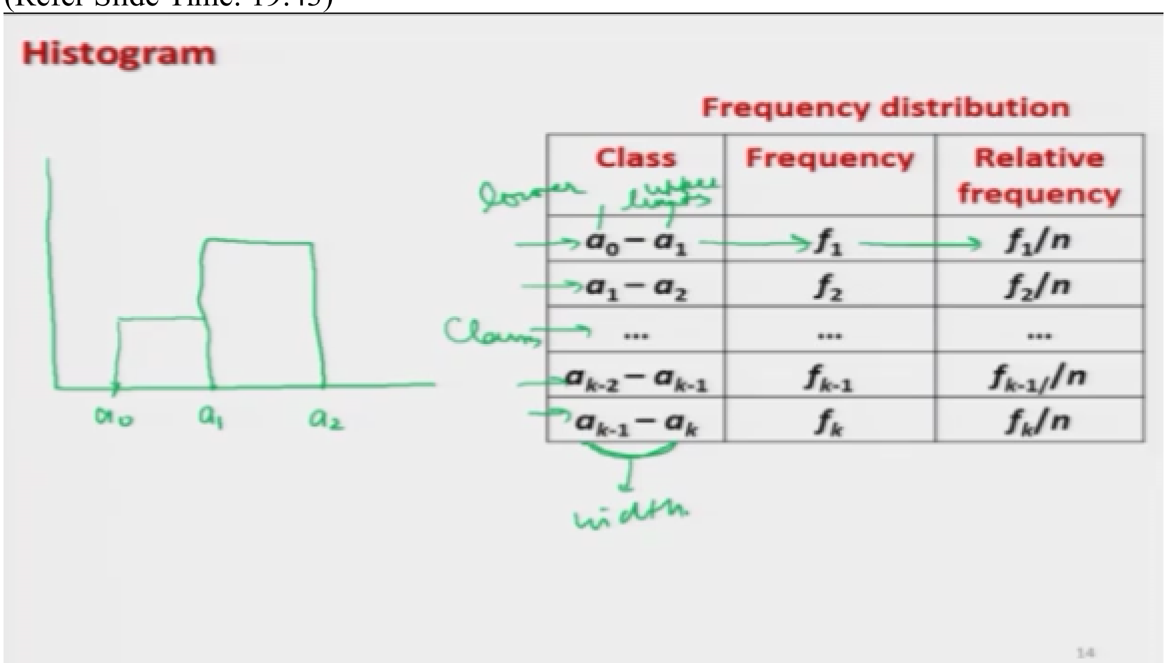
And the size of the interval that will provide us the width, and when every class will have some frequency or the relative frequency which is the number of values which are belonging to that particular category.

So now if you try to understand the construction of a histogram what we really try to do? That we tried to create here 2 bars,
(Refer Slide Time: 19:35)



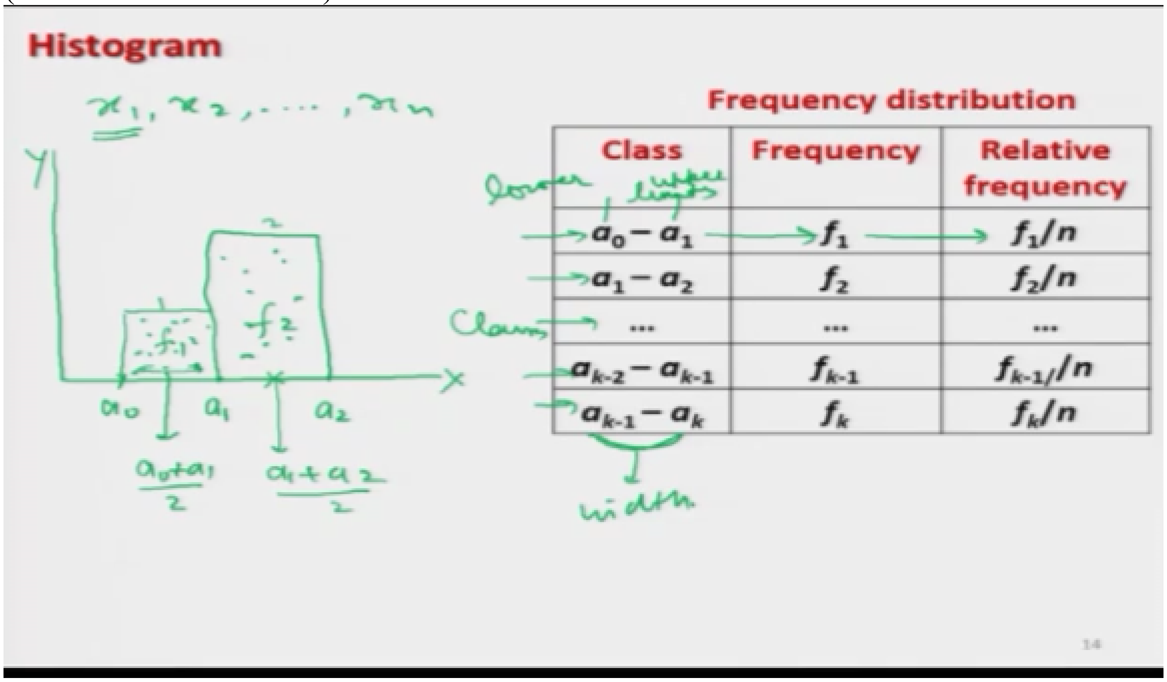

say this will have the limits for example this value will be your A0, this will be your here A1, and this will be your here A2.

(Refer Slide Time: 19:43)

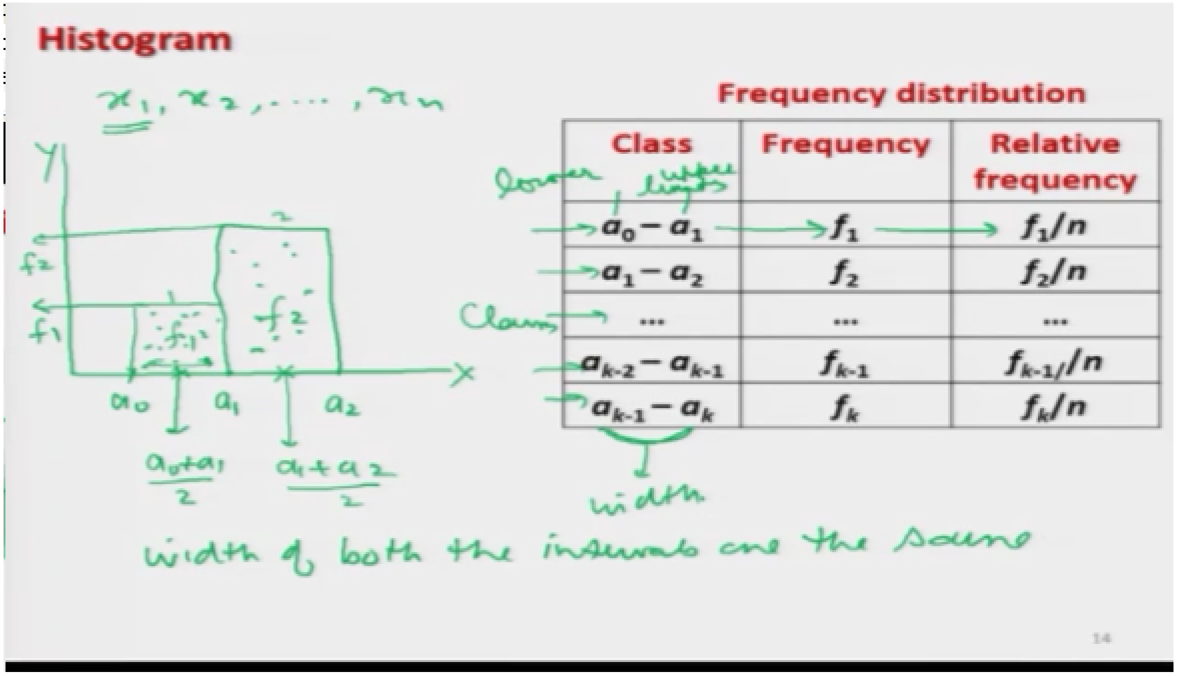



Now we have got the data X1, X2, XN, suppose N values are there, now I try to see where this X1 belongs to category A1 or category A2 or to the class interval 1 or class interval 2, this is my class interval 1 and this is here 2, suppose this belongs to X1, so suppose its value on the X axis inside this bar lies somewhere here, and now I take the second value suppose this values lies over here, third values which lies over here, fourth value which lies over here, fifth value here and so on some F1 number of values will be lying inside the bar 1, and similarly here F2 number of values will be lying inside the bar 2.

So one thing what we do that we assume that all this values which are spread around the mid value, mid value is to determine by this A0 + A1/2 for the category 1, and for this interval, for the second interval A1 to A2 the mid value will be A1+A2 divided by 2, so what I'll try to assume here that all the values are concentrated in the mid value.
(Refer Slide Time: 21:17)

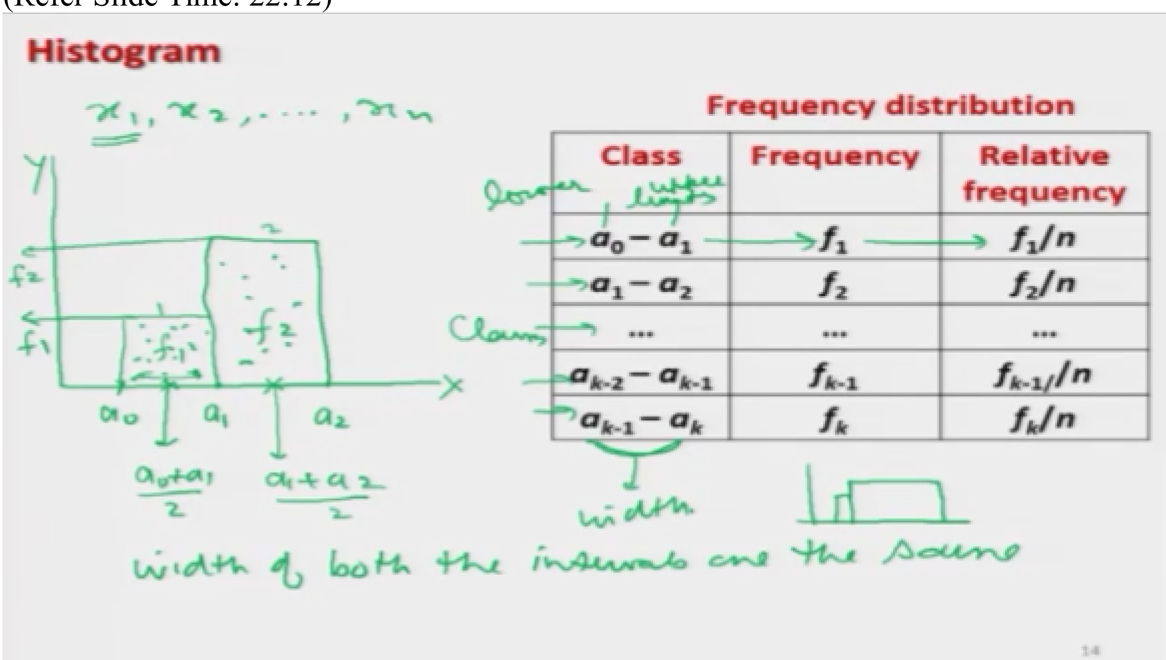



So what I'll try to see here that the frequency of the class interval 1 A0 to A1 is F1, so assuming that all the values are at one place, I'll try to make it here the frequency F1, and similarly the height of this one will become here F2, and I'm assuming that the width of both the intervals are the same, so this is how the histogram is constructed.
(Refer Slide Time: 21:57)




Now obviously in case if you try to create here a histogram something like which is so thin and another is so big,

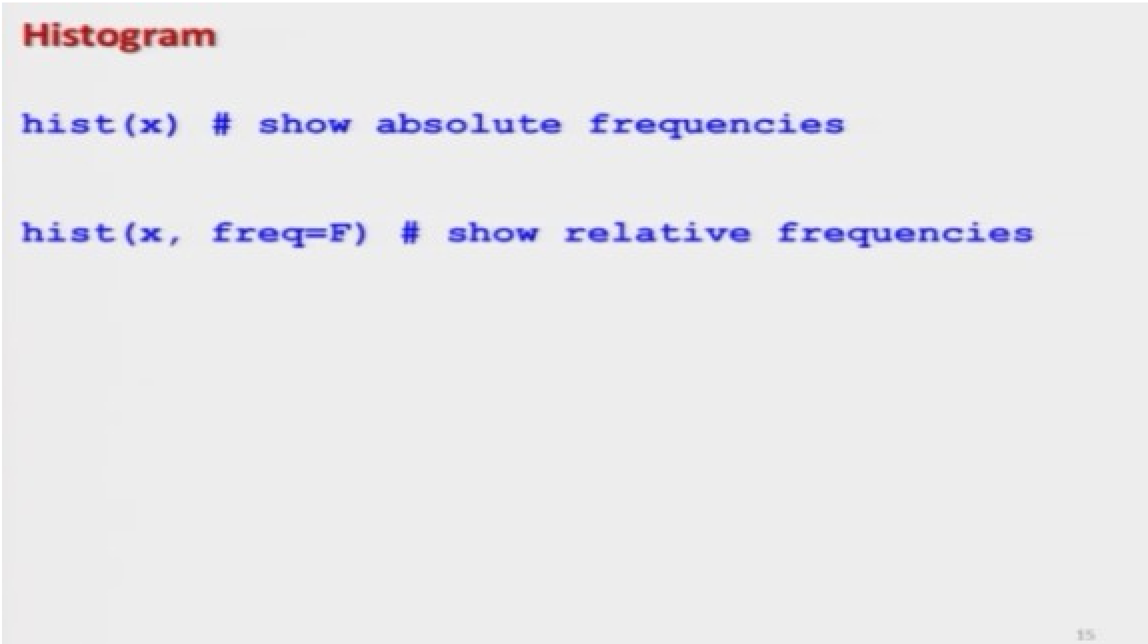
(Refer Slide Time: 22:12)




this is not so convenient to compared the two areas, so that is why it is emphasized that for all practical purposes the widths of the bars are kept the same.

And now instead of this frequency I can also have here relative frequency F1/N and say F2/N, but it depends on the need and requirement what we really want to do.

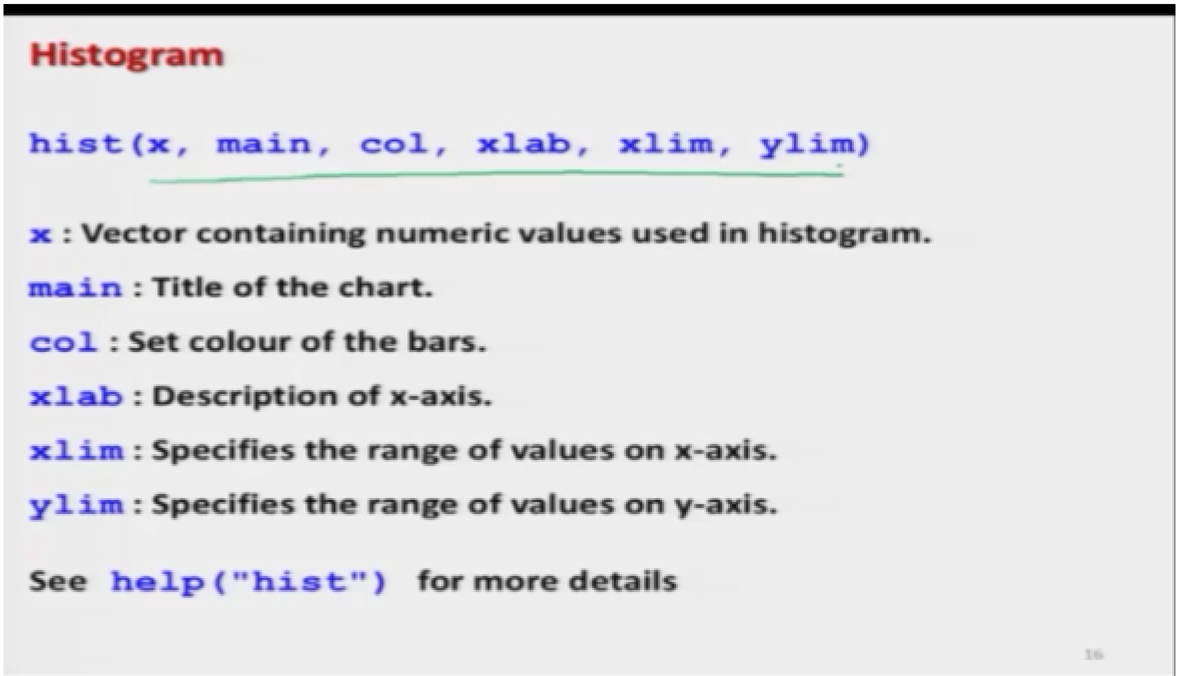
(Refer Slide Time: 22:40)




Now in R software the histograms are constructed by the command HIST and inside the X which is, we have to give the data vector, and you will see that in this case you don't need to create the table, histogram function or the function HIST will automatically create the frequency table and then it will create the bars, so this is different than in the case bar diagram or the pie diagrams.

Now in histogram I have two options, histogram can be created using the absolute frequencies or it can be created increasing the relative frequencies, so in case if I want to use the absolute frequency then there is no issue this command HIST will take care of it and that is the default choice, but in case if you want to create the histogram using the relative frequencies then you have to add here one more parameter FREQ = capital F that means frequency = false, so as soon as you give the frequency to be false the function HIST will automatically considered that the function has to considered the relative frequency for the construction of the bars.

Now in case if you want to improve your histogram as we have done in the case of bar chart, pie chart and so on,
(Refer Slide Time: 24:20)




there are some more choices of parameters which can be given inside the arguments, so obviously this here X this is going to determine the data vector, the numerical values for which the histogram has to be constructed.

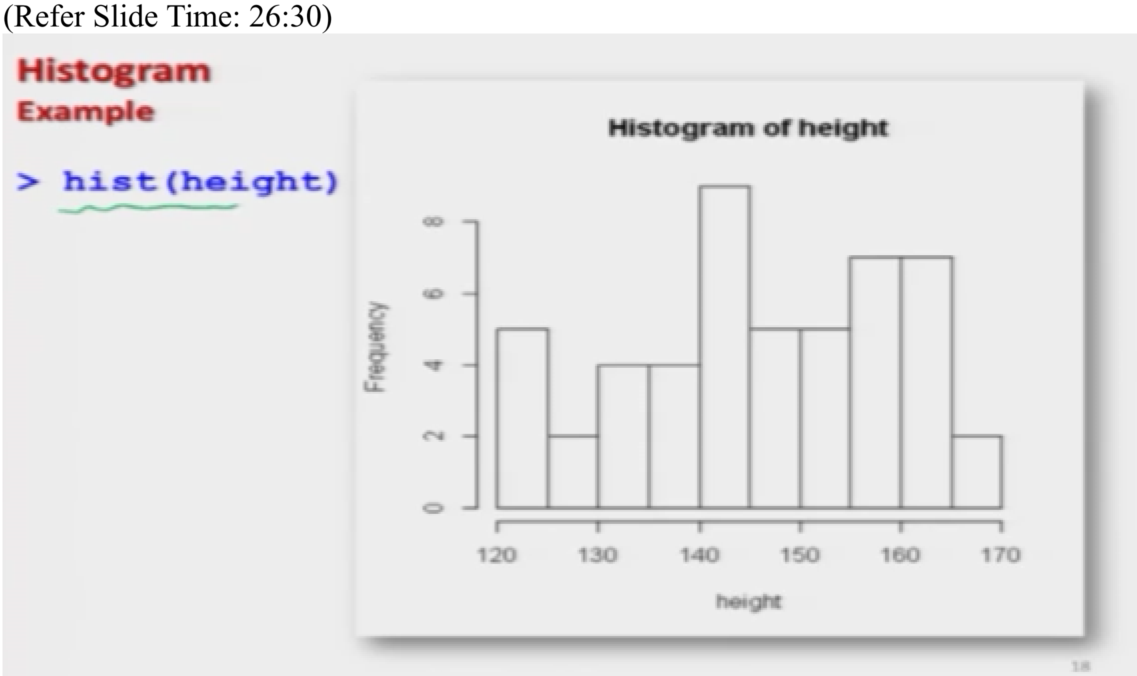
Now in case if we want to give the title of the chart then this is controlled by main, MAIN, in case if you want to change the colours of the bar then we have to use the parameter COL, in case if you want to add any description on the X axis then we have to use X lab, and in case if you want to control the limits on the X axis then you have to use the command X lim.

And similarly in case if you want to control the limits on the Y axis also then you can use it here Y lim, and there are more options but I would suggest you that you please try to look into the help using the command, help HIST inside the double quotes, inside the arguments, that will give you more information.

(Refer Slide Time: 25:40)

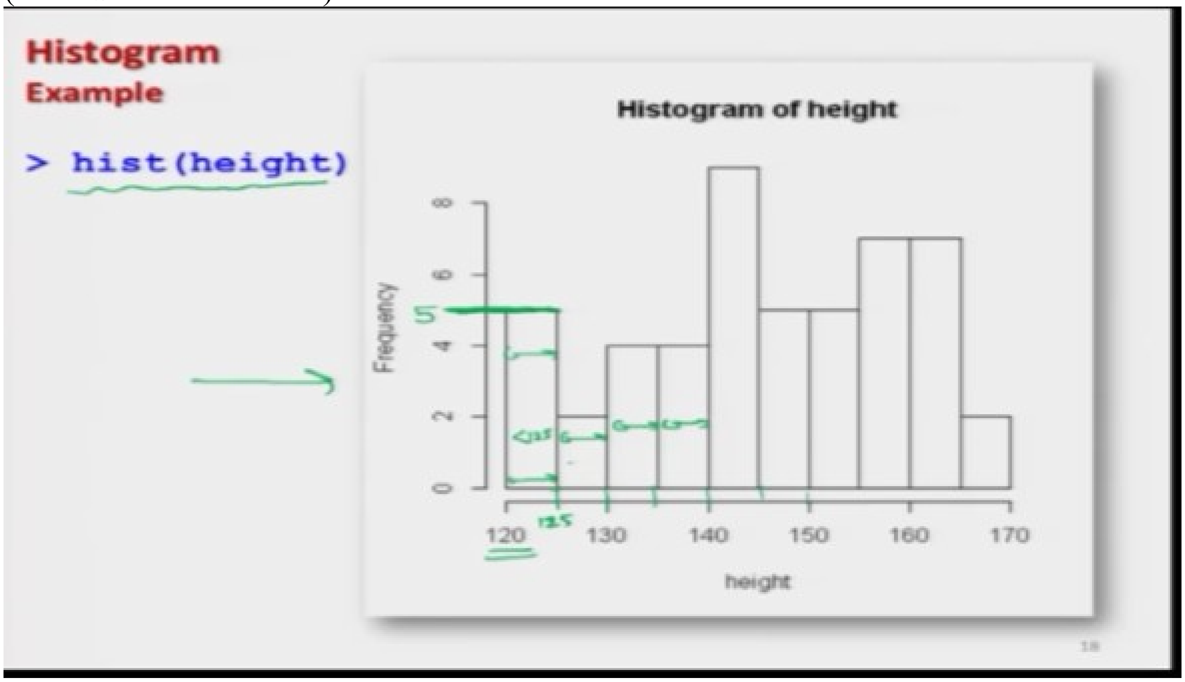

Now let me take herein example to show you the construction of histogram, so herein this example we have the heights of 50 persons recorded in centimeters, now you can look in these values, do you think that way? In the first glance are you getting any information whatever is contained inside the data? It is very difficult and that is the advantage of using the histogram that it will try to reveal the information contained inside this set of data, so I tried to stored all this data into a variable here height using the command here C,

(Refer Slide Time: 26:30)




and after this if I try to use here the command HIST over the variable height HEIGHT we get here this type of data, so you can see here this is trying to give us the intervals here 120 and 125, then here is 130, 135, 140 and so on, and once I said what are the values which are

contained inside this bar, so all those values which are less than 125 they are stored in this bar, and I can look at the height of the bar which is here, so since the width of each of the bar this, this, this, this and so on they are the same, so by looking at this value I can say that there are 5 values which are smaller than 125.
(Refer Slide Time: 27:30)



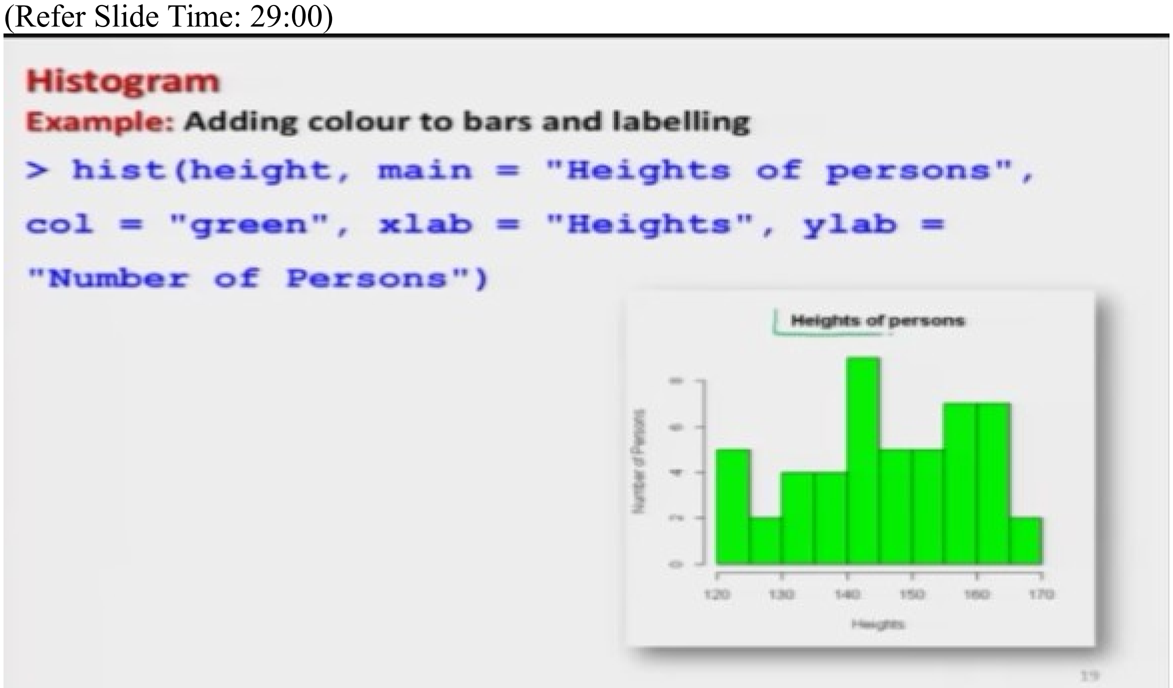

Similarly in case if I try to look at this interval, the frequency here is 2, so I can say here there are 2 values which are lying between 125 and 130.

Similarly in case if I try to look at this interval, this is starting from 155 to 160 and the frequency here is say close to 7,
(Refer Slide Time: 28:00)


so that is indicating that there are 7 values between 155 and 160 and also this is the same height of the next bar which is here, so these two bars they have the same frequency, so I can say that the number of persons having the heights 155 centimeter to 100 centimeters, and 160 centimeter to 165 centimeter they are the same, so this type of information is revealed from this type of graphics.

Now in case if you want to improve the look by adding colours or by adding say legends or controlling the limits you can use the parameters, and how to use those parameters I will try to show you here some of them,

(Refer Slide Time: 29:00)



but I would request you to have a look on the help and then try to see. So for example here I am trying to give the calculate of the chart as say heights of person, and I have changed the colour, colour of the bars and on the X axis I am trying to give here a legend say heights or title, heights, on the Y axis I'm trying to give here the title number of person.

So in order to get a graph like this one I simply have to add the parameters inside the hist command, so I'm trying to use here the command here main, heights = heights of persons, so that is going to give me the outcome of title of the graph.

And similarly this green colour, this is controlled by this command COL, so I'm trying to give here COL = green inside the double quotes, and this title on the X axis heights that is going to be controlled by X lab, so I'm trying to give the name of the height inside the double quotes.

And similarly the name on the Y axis is controlled by Y lab which once again I'm trying to give it inside the double quotes. And similarly if you try to add some more parameters over here, you can make it more informative depending on your choice, depending on your wish, depending on your requirement, so now I stop here in this lecture and once again I would request you that you please try to choose some dataset from the book which are continuous and try to create this histogram, and similarly you try to practice for the 3 dimensional pie diagram and try to use different types of parameter, try to give them different values for example I have shown you that one, that when we try to use the explode is equal to 0.2 to 0.8 then how much is the separation between the two, so that will give you a more idea that how the graphics can be controlled and or how the graphics can be made more informative.

Similarly in the histogram also there are some other parameters which we have not used here, but I would request you to have a look on the help menu and try to see how they are used, and try to experiment them, so keep on practicing and we will see you in the next lecture, till then good bye.

**Acknowledgement**
**Ministry of Human Resource & Development**

**Prof. Satyaki Roy**
**Co-coordinator, IIT Kanpur**

**NPTEL Team**
**Sanjay Pal**
**Bharat Lal**
**Ashish Singh**
**Badal Pradhan**
**Tapobrata Das**
**K. K. Mishra**
**Ashutosh Gairola**
**Puneet Kumar Bajpai**
**Bhadro Rao**
**Shikha Gupta**

**Aradhana Singh Rajawat**
**Sweta**
**Nipa Sikdar**
**Anupam Mishra**
**Ram Chandra**
**Manoj Shrivastava**
**Dilip Tripathi**
**Padam Shukla**
**Sharwan K Verma**
**Sanjay Mishra**
**Shubham Rawat**
**Santosh Nayak**
**Praduyman Singh Chauhan**
**Mahendra Singh Rawat**
**Tushar Srivastava**
**Uzair Siddiqui**
**Lalty Dutta**
**Murali Krishnan**
**Ganesh Rana**
**Ajay Kanaujia**
**Ashwani Srivastava**
**M.L. Benerjee**

**an IIT Kanpur Production**