Foundations of R Software
Prof. Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture - 51
Graphics: Histogram

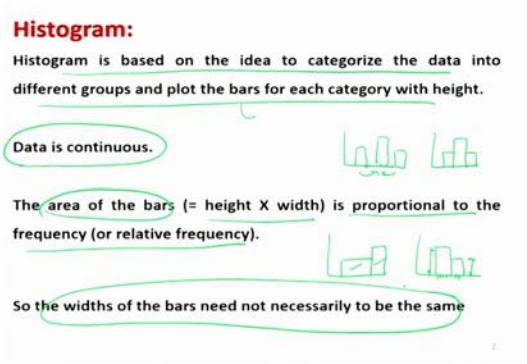
Hello friend welcome to the course foundations of R Software and you can recall that in the last two lectures we were talking about the various types of graphics and we had considered the graphics which are essentially for the categorical variable.

Now, we are going to continue on the same lines in this lecture also and we are going to consider today very popular graphic which is histogram. Histogram also looks like the bar plot, but the main difference between the two is that bar plots and pie diagram equivalently they are used for the categorical data; that means, when you have the data which are in the form of categories.

And histogram is used when your data is continuous means that is the non categorical data, right for example, if I say heights, weight, age etc. So, if you try to take the real values of height weight and age in say centimeters kilograms or say years they are the continuous data and they are the data from the continuous variable.

So, today we are going to learn about how you can create the histogram in the R software. Well what is histogram I know you know better than me. So, we begin our lecture and we try to learn how we can create the histogram.

(Refer Slide Time: 01:33)



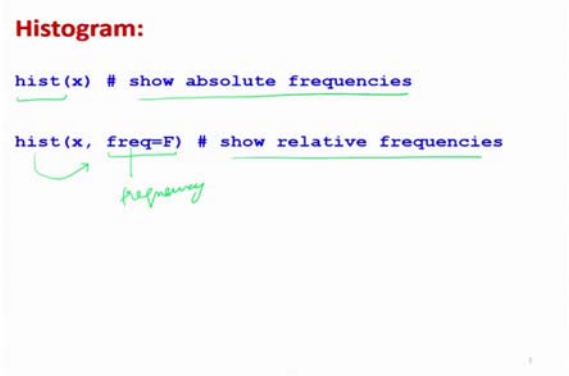
Now, the first question comes over here that how the histogram is considered and how it is different from the bar plot. If you try to have a look the bar plot will look like this and histogram will look like this so; that means, this bars are stacked together, right. So, actually this histogram is based on the basic idea that the data can be categorized into different groups and then their frequencies can counted and bar for each of the category can be created and with certain height, right.

And yeah, but please do not understand that in the histogram the height of the bar is going to proportional to the frequency that is happening only in the case of bar plot. In the bar plot the height of the bar is proportional to the absolute or relative frequency of the data, but that is not the case in the case of histogram. The histograms are constructed when the data is continuous and in histogram the area of the bar is considered.

That means the height of the bar and the width of the bar and this area is proportional to the frequency or the relative frequency. Now if you try to see what really happened? if you try to take here a histogram can be like this one bar is like this and another bar is like this, right that is as such it is correct, but it is not so convenient in practice, right by looking as these two bars you cannot find whose frequencies is more.

So, that is why what people do in practice they will try to keep the width of the bars of the two bars in the histogram to be the same. If this part is same then only the difference will come due to the height and that is why you will see that most of the histograms always have the same weight, right. But as such these widths need not necessarily to be the same always that is what you have to keep in mind.

(Refer Slide Time: 03:26)



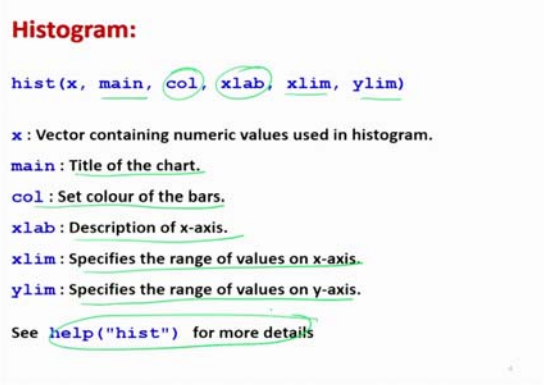
Now, in R software if you want to create the histogram the command here is very simple h i s t. So, if you try to use here h i s t then it will show the absolute frequencies and if you try to use here an option that f r e q the frequency that is the short form for the word frequency that is actually FALSE. Then the histogram will be constructed with respect to the relative frequencies.

One thing I would like to explain you here that how to create the frequency tables from the data in the case of continuous data that is the entirely different topic which I have not covered in the lecture. But I believe that you all have done in the elementary classes and from there if you try to recall you used to write on the frequency table and then you used to write that the frequency classes.

And then you used to count the number of values which are lying in that interval and that used to be the frequency and then you used to consider the midpoint of the class intervals and then you try to create all such graphics. So, well I am not going into that direction for that I will request you please try to go through with some elementary book and try to see that how the frequency tables are created.

And now here in the case of histogram I am talking of the frequencies of the frequencies table in the case of continuous data, right. Because in the case of discrete data you had seen that you had simply counted for example, in the example of gender you simply counted the number of males and females and that had given you the frequency, right.

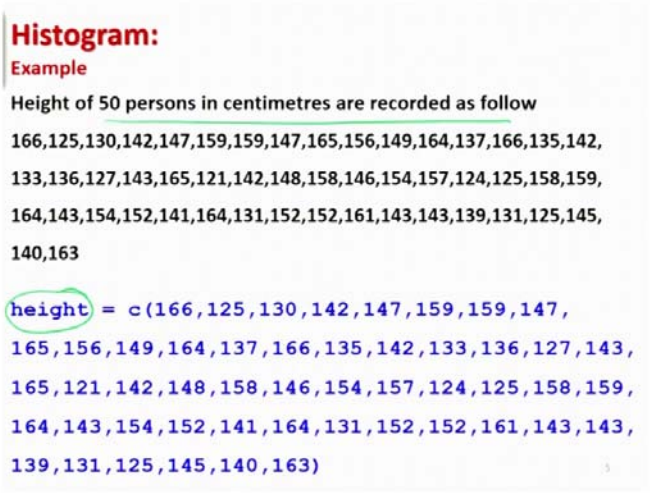
(Refer Slide Time: 04:54)



And then if you try to see when you are trying to use the histogram. Then there are as it happens in all the graphics it also has many options like as say main is for giving the title of the chart, then here col you know this is to choose the colours of the bars and then here x lab which is trying to decide the title on the x axis.

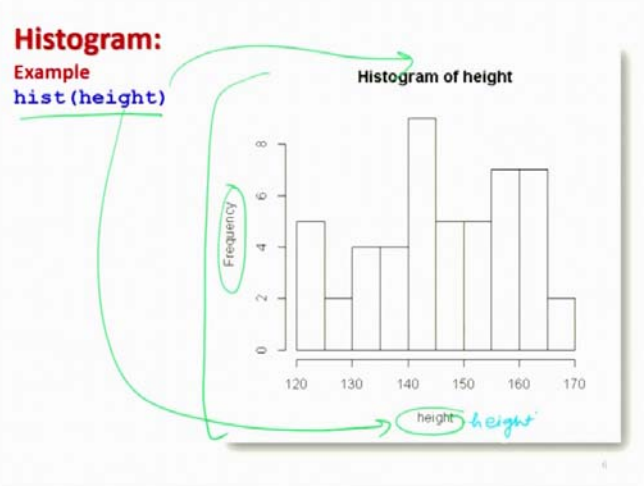
Then similarly you have x lim y lim they are trying to specify the range of the values on the x and y axis. So, they are x limits and y limits etc. So, once again I would strongly recommend you to look into the help in the histogram and try to look for this more details.

(Refer Slide Time: 05:31)



But my thing is there here that I would try to consider here very small example and try to show you that how these things can be created. So, I try to consider here an earlier use data in which we had collected the heights of fifty persons in centimeters and they are stored in the data vector here height like this, right.

(Refer Slide Time: 05:51)

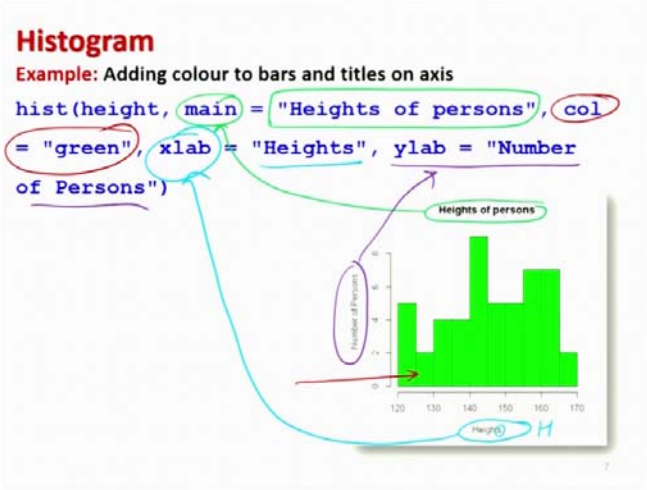


So, now, I will try to create the histogram on this data on height. So, if you try to write down here the command here h i s t and inside the parenthesis. If you simply write here height and now you can see here this is different. Because in the case of bar plot you used to first convert the data into frequency using the command table.

And then the bar plot was used, but in the case of histogram you are simply using the original variable with the command h i s t. So, that is what you have to keep in mind and that is what I always say you have to see that how R is going to work and accordingly you have to choose your values, right so ok. So, if you try to see here this is like here see here now a couple of things will be automated here.

For example it will try to take here the name on the axis directly from the name of the variable and it will try to write down your histogram of height automatically. And then on the y axis there is a frequency, right. So, the data on height has been classified into frequency table and the frequency of those class intervals that has been plotted here, right ok.

(Refer Slide Time: 06:49)

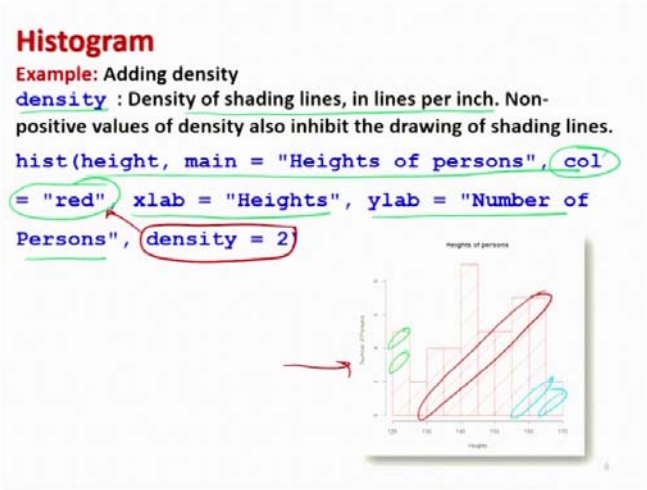


So, now in case if you try to make some alterations in this histogram that is very simple you have to simply use the same type of options which you have used and learned in the earlier lecturer. For example, if you want to make here the title heights of person. So, this can be controlled by the command or option their main and you simply have to write down here heights of persons and similarly if you want to make here the colour of this graphic to be green.

So, we can see here I have used here the option here the c o l is equal to green and you can see the colour of this graphic becomes here green. Similarly if you want to give here this heights here on the x axis then you have to use here the command here x lab is equal to heights yes.

You have to be little bit careful that ok in the earlier graphic also it has used here the height, but it was h e i g h t all in lower case and I have given here the heights where H is in capital letter and it is actually heights s is also added. So, do not get confused that it is coming automatically, right after this you want to give here the say title on the y axis. So, this is given here by here y lab, right. So, this is how you can see it is not a very difficult option to add these options and get this histogram.

(Refer Slide Time: 08:01)



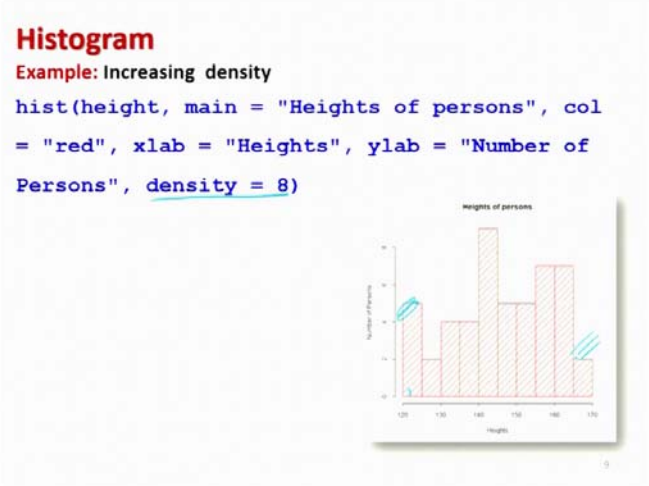
Now I will try to explain you some more options which are available in the histogram and then you have to understand what is really happening. For example, if there is a command there is an option in the histogram that is density d e n s i t y.

Actually this option as the this type of lines that you can see here inside the bars and this density value this controls the shading of the line that is the that how many lines per inch should be there, right. And in case if you try to increase the value of this option here density, right then there will be more number of lines in a square of inch. So, if you try to see here the way I am trying to use it here all other options are just like earlier.

But now I am using here colour is equal to here red. So, this colour is actually going to control the colour of the histogram and the this type of lines, right. So, I try to use here x lab equal to heights y lab equals to is equal to number of persons and only here two changes are made I am trying to add here the option density equal to 2 and the colour here is red.

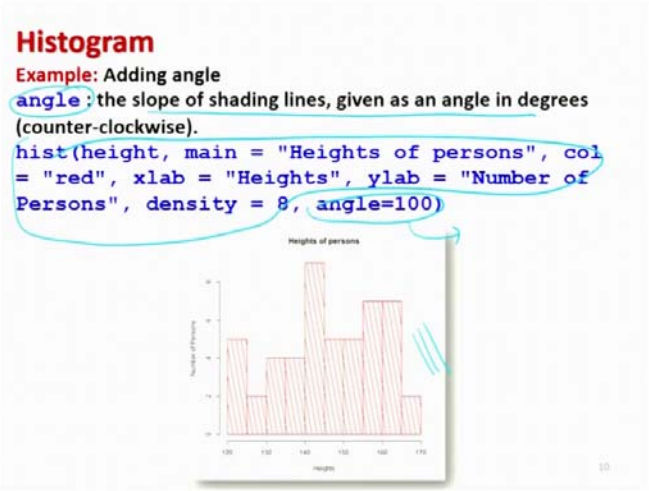
So, if you try to see here the outcome this comes out to be here like this you can see here these types of lines are created, right you can see here like this. Now I try to do here one thing just to show you that what is the interpretation out.

(Refer Slide Time: 09:16)



And how this density actually changes I simply try to increase the value of the density from 2 to 8. And you can see here now this density have become more closer these lines have become more closer to each other, right.

(Refer Slide Time: 09:27)

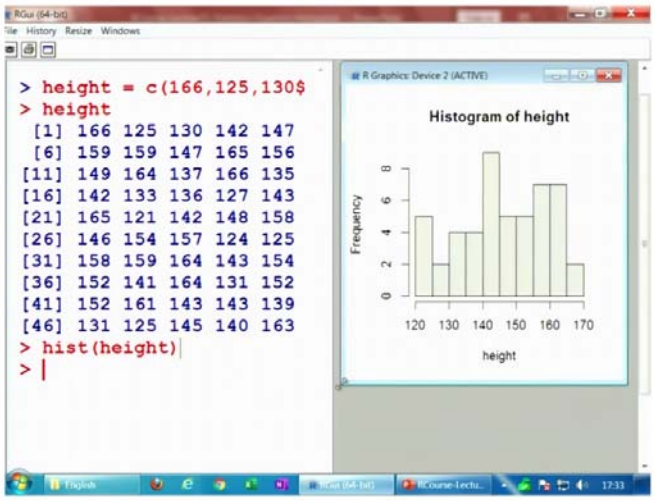


And in this way density if I also try to use here one more option here angle, right there is an option here angle. So, this angle is going to control the slope of the shading lines, right and that is in the counter clockwise direction. So, the entire command is the same

as earlier which I used here in this one and you can see here this is here the line inclination and here this inclination has become just different this is now at the angle of 100 degree.

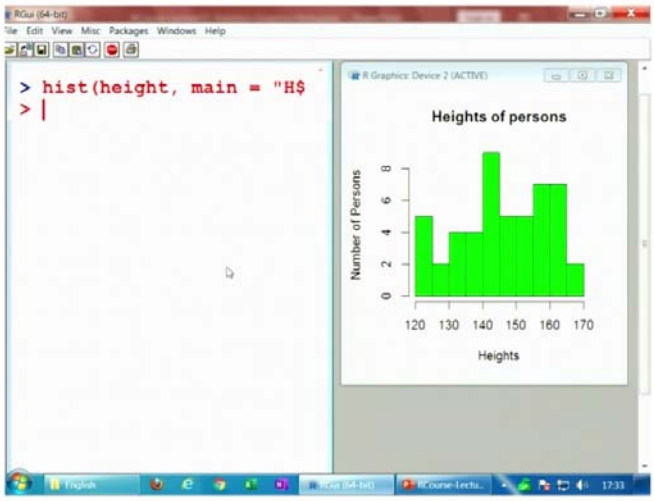
So, these lines are now becoming here like this and earlier the lines were like this. So, you can see here now where you are going to use it that is up to you, but my job was to convince you that ok you can do such jobs very easily in the R software and that is what I am trying to do here, right.

(Refer Slide Time: 10:19)



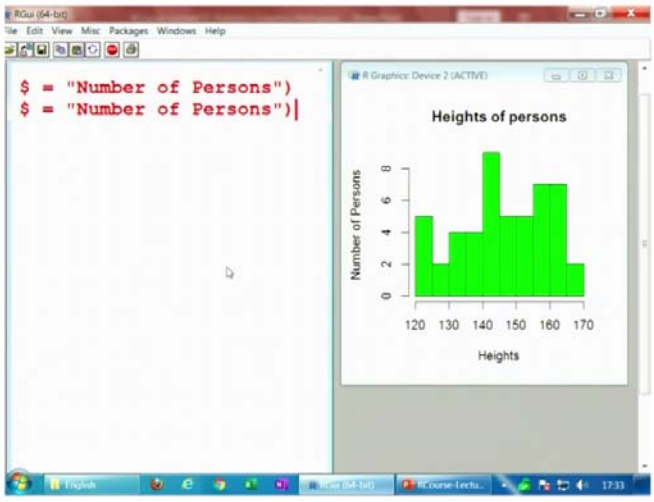
And beside those things I am just trying to make you comfortable in understanding and in believing that R can do many such options which many software can do. So, now, let me try to show you these results in the R console. So, I have prepared here my height data and if I try to use here the command hist h e i g h t you can see here this is my histogram which I have shown you in the beginning, right. Now I try to make here different type of changes so that you can see that these things are working.

(Refer Slide Time: 10:44)

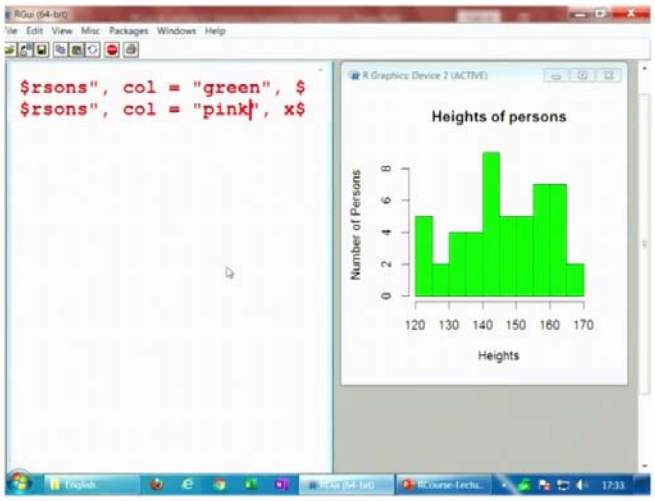


And you can see here now I have changed the colour to be here green.

(Refer Slide Time: 10:47)

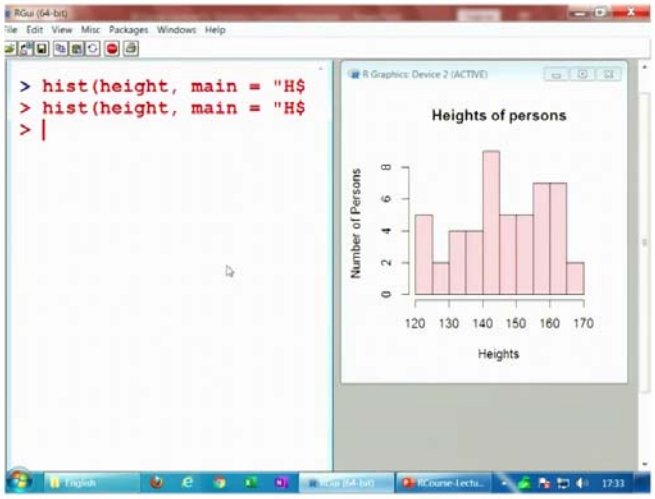


(Refer Slide Time: 10:49)



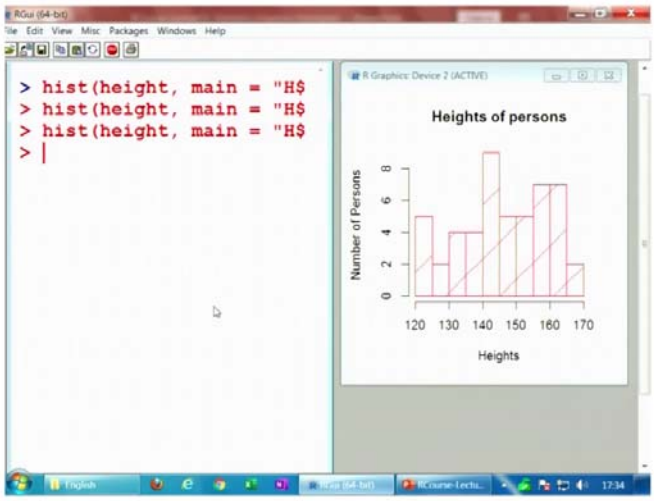
And if you want to change this colour to be for example, here if you want to make it here pink it is here like this if you just and you see I am not change any other command it will become here like this, right.

(Refer Slide Time: 10:57)



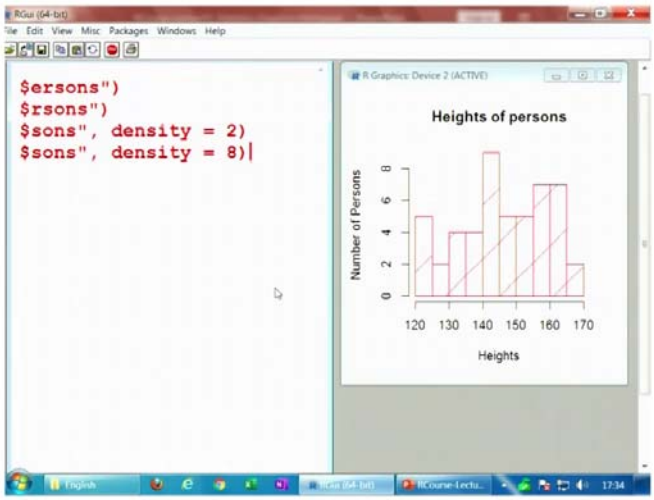
And after that if you want to make here these types of lines using the command density.

(Refer Slide Time: 11:07)



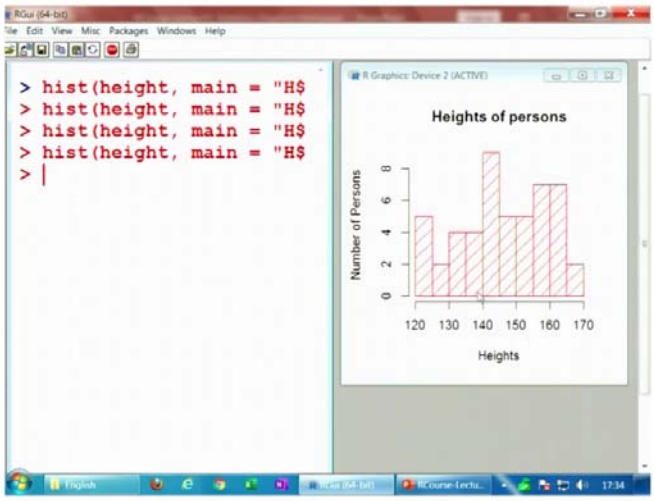
I can change here like this. So, we can see here this density is here like this, right and if you try to use here.

(Refer Slide Time: 11:17)



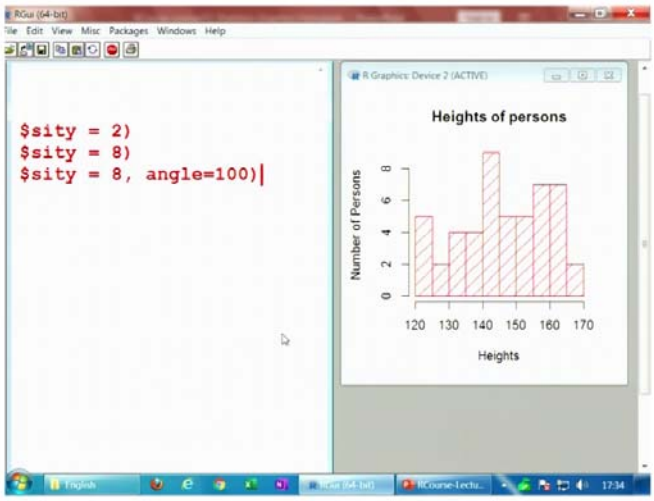
The and if you simply try to change here this value here with density from 2 to 8. I can show you what I am doing.

(Refer Slide Time: 11:20)



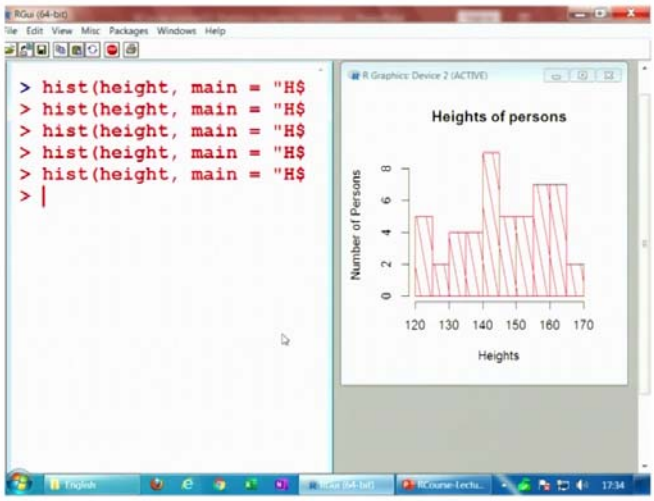
Then you will see here then this lines are becoming more dense. And after that if you want to use here the angle also. So, I can show you here that.

(Refer Slide Time: 11:30)



I will just try to add here the command angle. Angle is equal to 100 and try to see what will happen to the direction of this lines this is now changed, right.

(Refer Slide Time: 11:34)



So, now, that was a pretty simple lecture and now you can see that creating histogram is an easy job and for which you used to take the software from your friends for doing this job. Now, it is in your hand you can do it very easily not only easily. Now you can control each and every aspect of this histogram and you can make it as beautiful as you want as informative as you want that is up to you that will be your capability that how much you have practiced how much you have learnt in the histogram command to make it more impressive.

So, now I will make a very simple request please try to consider the data try to use these options from the help menu. And try to see what happens when you are trying to use a particular command in the hist function and try to see the effect yes at this moment some of these options which they are in the help may not be really helpful.

But try to identify that which are the options which are helpful for you and try to see that earlier the type of histogram which you are creating now can you make it more informative using the R software. So, you try to practice it and I will see you in the next lecture till then good bye.