Noc19-Ma14

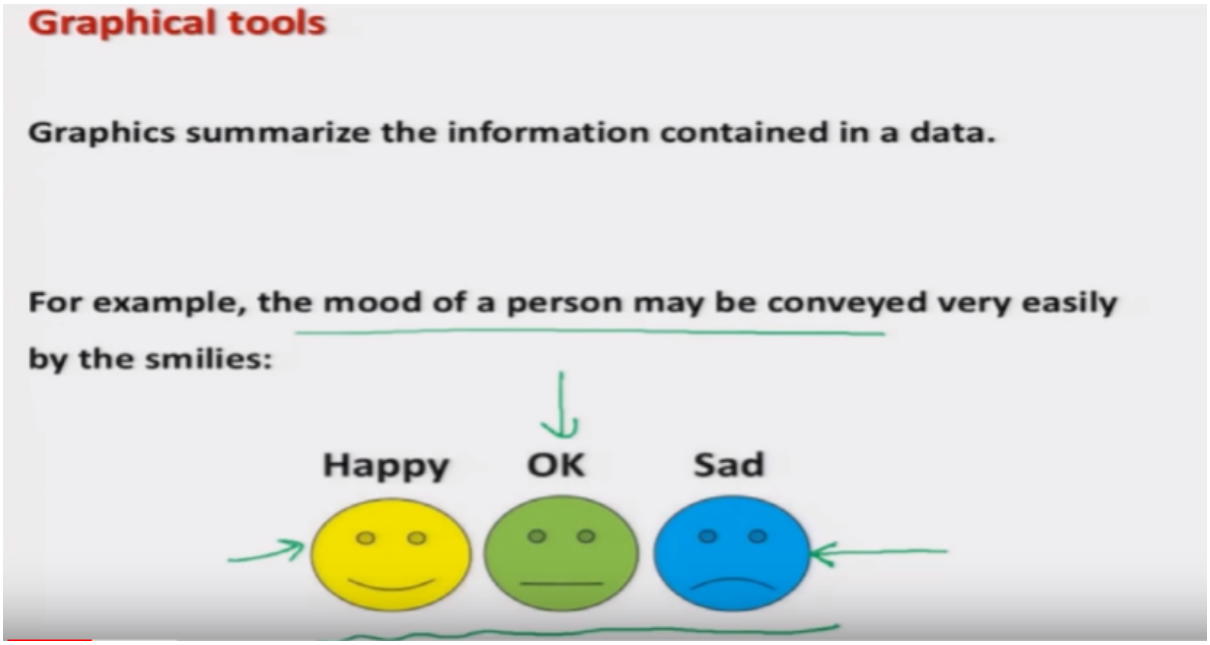
Lecture -10

Graphic and Plots - Bar Diagram

Welcome to the next lecture on the course descriptions that stick with our software. Up to now what we have done? We have considered the aspect of frequency distribution. In the last couple of lectures and for that we had two introductory lectures, where we had learned that once we get the data in our hand, then there are two types of tools that can be applied, one is graphical tool and another is analytical tool. so in the last two lectures where we had done the frequency distribution that was the first step when you would like to make an arrangement with your data so that the data is compatible to be exposed to the graphical and analytical tools, so now in this lecture and in the next couple of lectures first I will try to target at the graphical tools and after that I will continue with the analytical tool. so now the first question comes why should we use the graphical tools? so we know, that graphics are very easy to understand and that is why, we take the help of graphics to extract and to

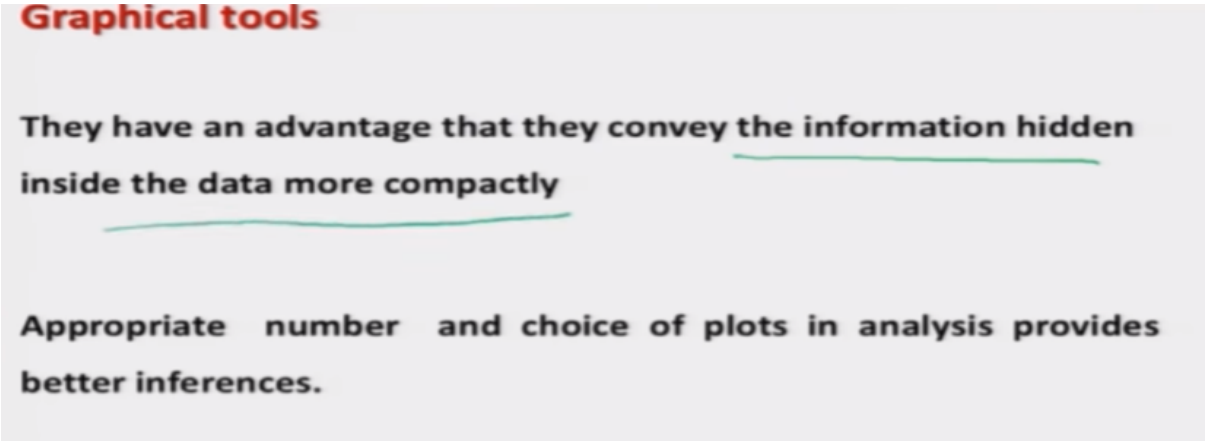
understand the information which is contained inside the data. now we are going to use the information extracted from graphical tools as as well as analytical tools together, so now firstly let me try to explain you why do we need this graphical tool.

Refer Slide Time :(02: 25)



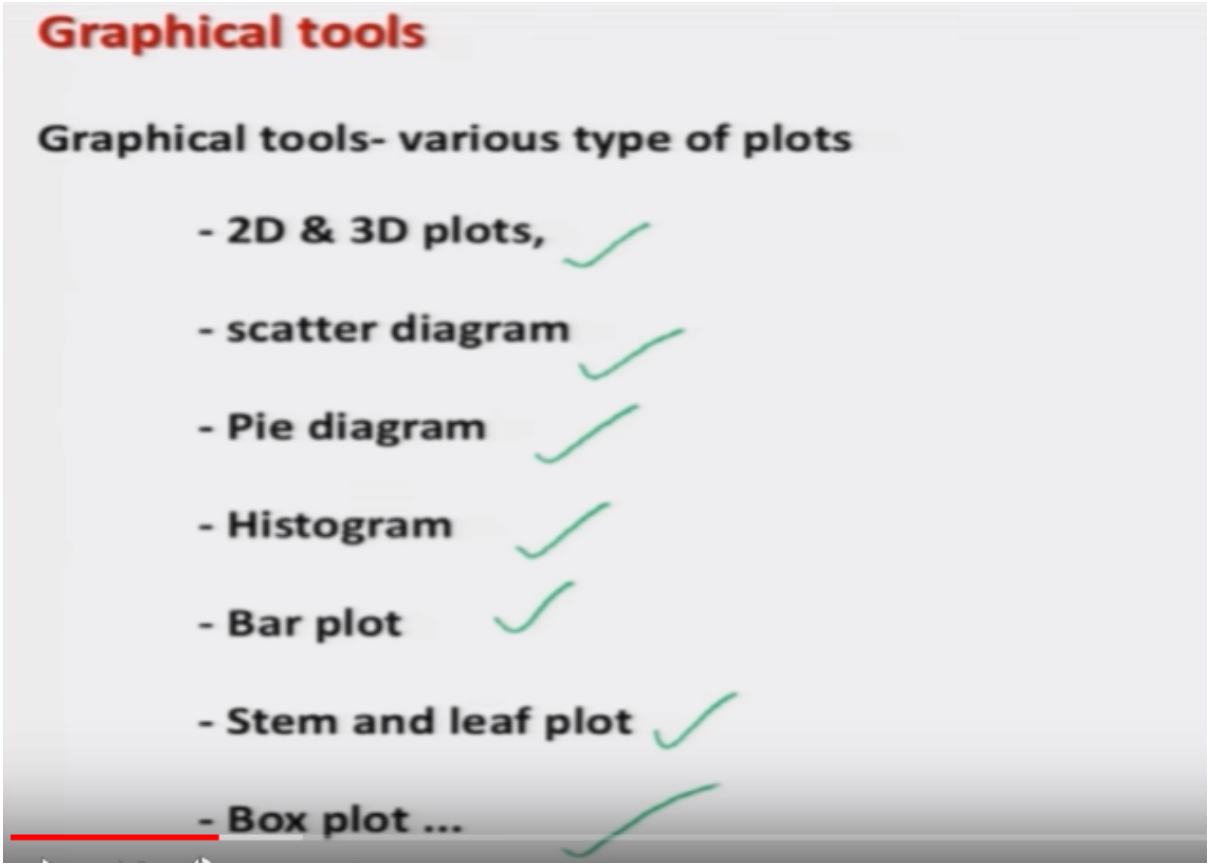
so now if I say suppose, I want to convey that a person is happy or sad, well you can explain in say several sentences that how the happiness will look like or how the face of a sad person will look like and so on, but in case if you try to use the Smiley's, what do you see here, that the mood of a person is very easily conveyed by these three Smiley's, just by looking at this structure means anybody can say very quickly, that this is indicating the happiness and this one blue one means anybody can say after looking at this face that the face is sad and in the green face in the middle one means anybody can have a look and can say very easily that this is reflecting where the person is okay. Similar is the information that is expressed in graphics mode. as we had discussed that when we start in start six we have only a sample of data and the size of the data can be very small or it can be very very large and each and every data contains some information, but we want to have the information in some combined way and that is why our first target was frequency distribution, now once again we are trying to combine the information in some graphical way and you would like to condense the information in the form of a graphic, so that we can have some idea about the information that is hidden inside the data.

Refer Slide Time :(03: 56)



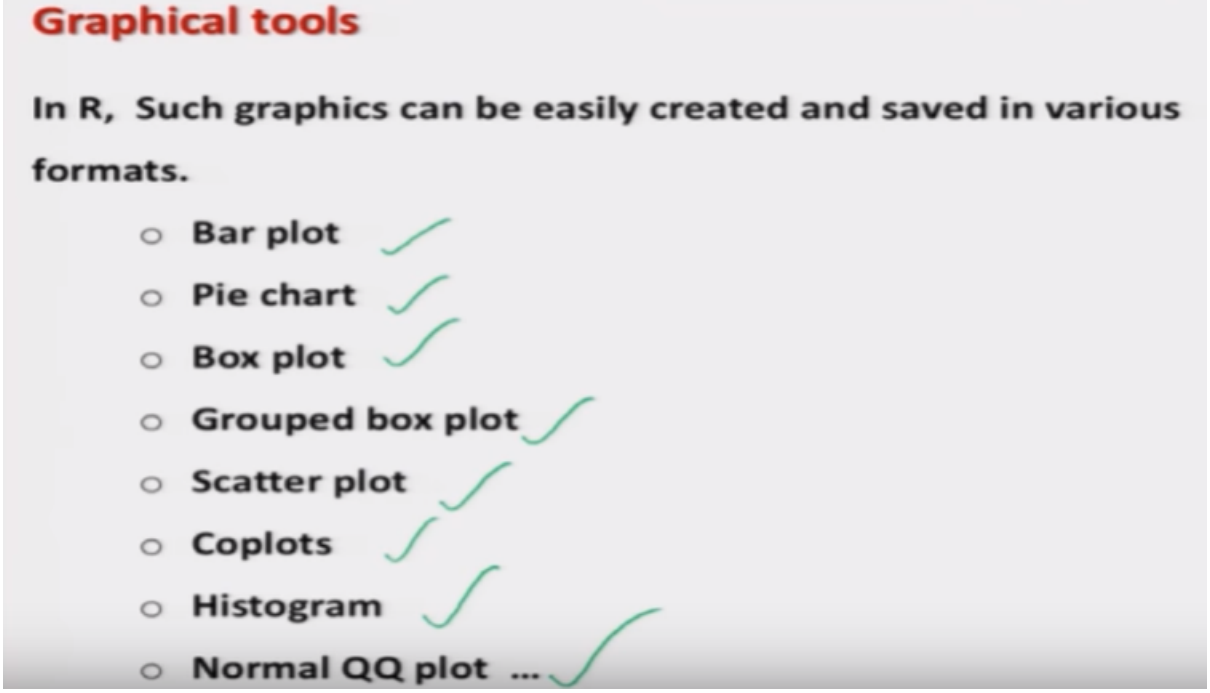
So there are various types of advantages the graphics can explain the hidden information very compactly and very quickly, which is very easy to under standby a common person. nobody needs much stronger knowledge of mathematics and statistics to understand behaviour of a smiley face or the behaviour of a curve, one can easily understand it, so that is the advantage. once again, I would say that whenever you are trying to conduct any statically analysis there are various types of graphics that can be used, but sometimes there is a myth that unless and until you use more number of graphics the analysis is not good, rather, I have heard people saying that that the goodness of a statistical report, depends on the number of graphics, higher the number of graphics better is the report, well this is wrong, means if somebody has some problem, it does not mean that if the doctor gives more medicine, then the doctor is good, the doctor is good if he gives the appropriate medicine inappropriate quantity, same is the message in the statistics also. we have to use appropriate graphics and say appropriate number of graphics also right and the use of appropriate graphics and Cypriot number of graphics will only give us the correct information in a more fruitful way. so in static there are various types of graphical tools, that can be used.

Refer Slide Time :(05: 35)



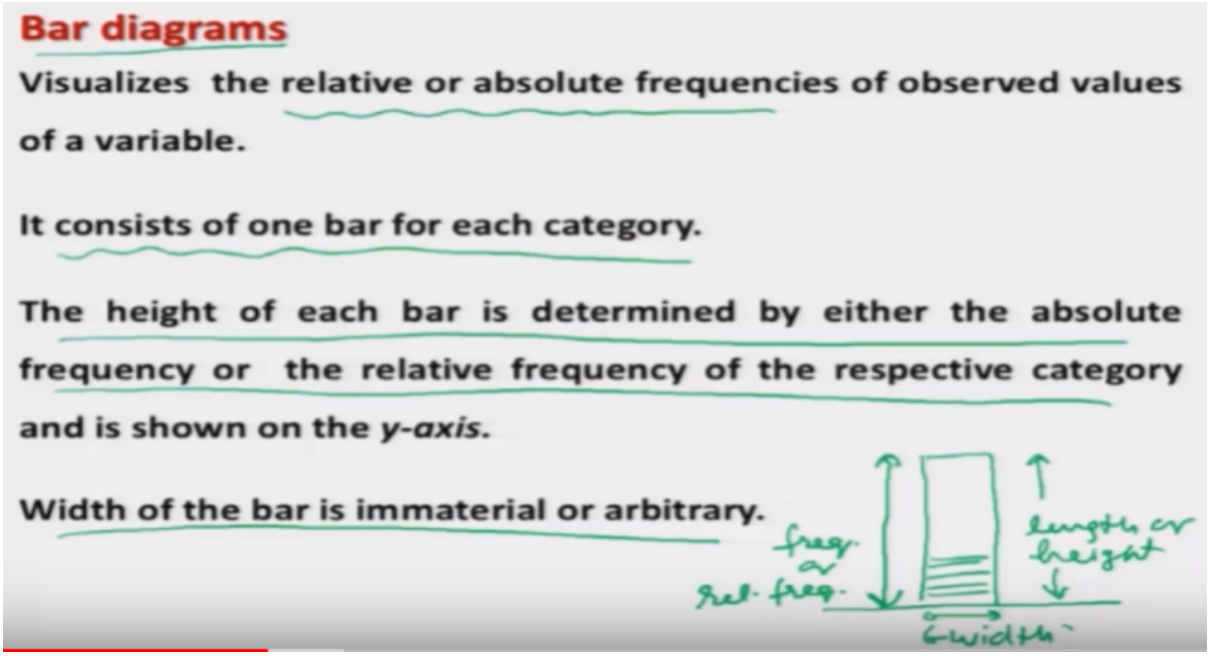
there are two dimension plot, three dimension plot, you know a scatter diagram, there is pie diagram, histogram, bar plots, stem and leaf plot and there are box plot and there are many, many more and particularly in the advent of soft wares, these graphics have become very popular because they are very easy to create and they can be created in a very small time actually. Similarly just like all other software's, this R also as a capability to create the graphics, not only create the graphics, but it also gives you an option to save the graphics in different mode, like as Postscript format, jpg format, PDF format and so on. So in R there is a long list of graphics that can be created, for example.

Refer Slide Time :(06: 29)



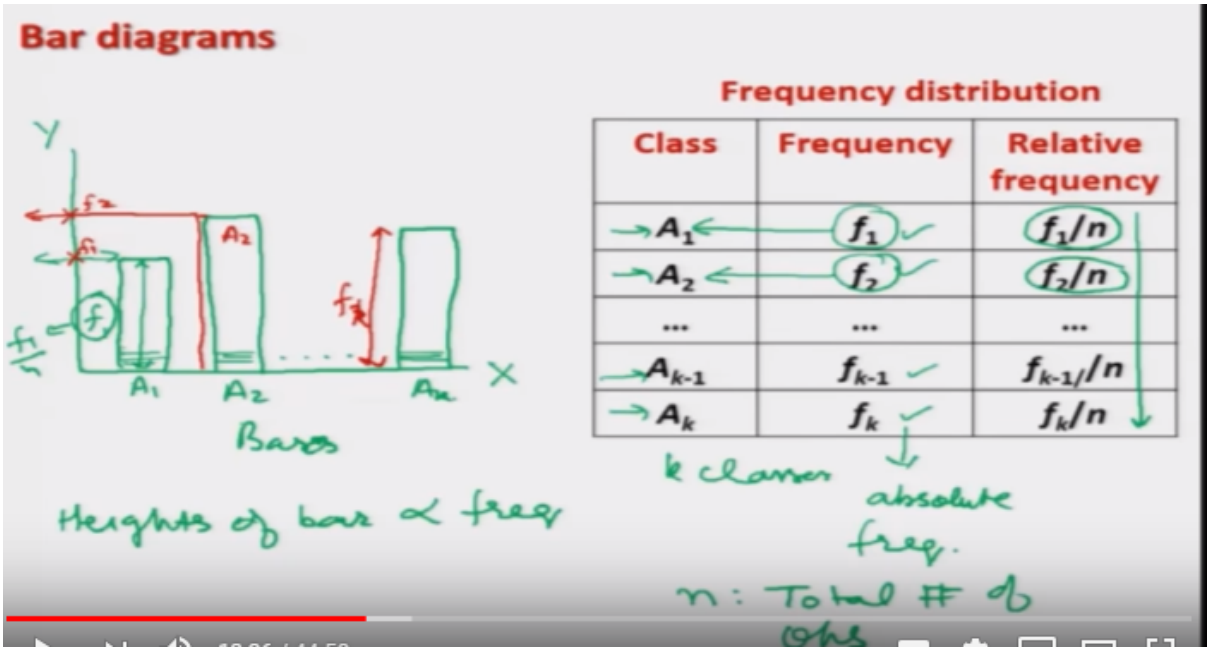
same plots, which you have learnt earlier like a, bar plot, pie chart, box plot, group box plot, a scatter plot, Coplots, histogram, normal QQ plot and there is a long list, all sort of two dimension, three dimension, coloured. There are many, many possibilities and it simply depends on your capability that, how many graphics you can learn. well I am going to explain here some selected type of graphics, my idea is not to teach you the graphics, my idea is that, I will try to show you, how one can create the graphics in our software and what are the different options, which are available and then I will try to give you here several example and I believe that after that you will be confident enough to learn how to create any other graphics according to your need.

Refer Slide Time :(07: 34)



Well, so what we try to start herewith one of the very basic graph this is called bar diagrams. So this bar diagram is essentially used to visualize the relative or absolute frequencies of the data or the relative or absolute frequencies of the values that are observed for a variable. And this bar diagram, this will consist of one bar for each category and one of the characteristic or one of the very important characteristic, of the bar diagram is the height of the bar, height of the bar is simply proportional to the frequency or to the relative frequency, right? and the height of each bar is determined by the absolute frequency, or the relative frequency of the respective class and this height is shown on the y-axis whenever we are trying to create a bar, like this then this bar has two things, one is the width of the bar, this is here and another is the length or height of the bar. So what we have to keep in mind that when we are trying to consider the bar diagram then this width, this width is not important and width of the bar is in material and this can be chosen arbitrarily. only this length this is import and this is going to represent the frequency say absolute frequency or relative frequency but one thing I would like to emphasize here that many times or rather most other times you will see that, whenever we are trying to create the bar diagram, the width of the bars are taken to be the same, that is just because the graph should look nice and one should not get an impression, who does not know the theory of bar diagram, he should not get confused, that why the widths are so different, so that is the only reason, right? so now in case we have a frequency distribution of a discrete data or a qualitative data that is converted into some numerical value, through some proxy variables, like a yesterday, we had given the values, tastes, like a sweet will be denoted by one salty will be denoted by two and so on, so similar to that,

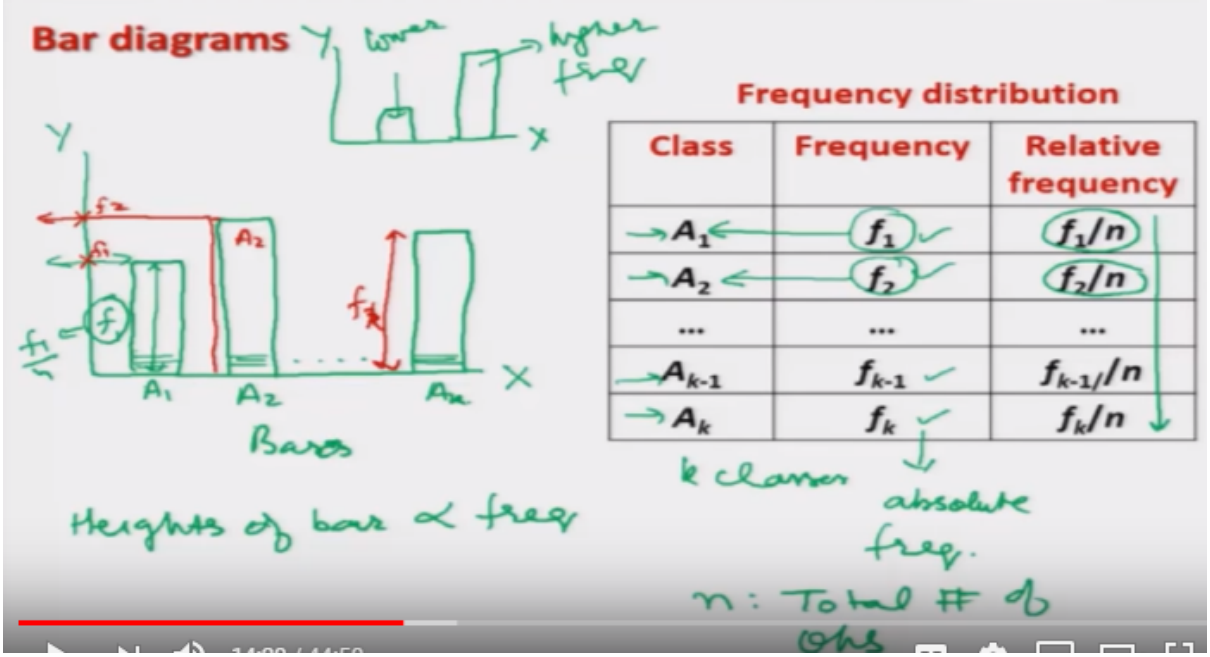
Refer Slide Time :(10: 16)



Now we assume that we have this type of frequency distribution, where these are my classes, A1, A2, AK minus 1, AK. so there are altogether k classes and f1, f2, fK minus one, fK, they are the frequencies or they are the absolute frequencies so they are simply going to represent that f1, number of values belongs to class A1, F2 number of values belong to class A2 and so on and once they are divided by the total number of observations, which is denoted by here, 'n', total number of observations, then this third column, this is giving as the value of relative frequency. So F1upon n, this is the relative frequency of class A1, F2 upon n, is the relative frequency of class A2 and so on. so now suppose I want to create here a bar diagram, then this is the basic philosophy or these are the basic fundamentals, that first I need to create here, X and y-axis and on x-axis, I need to create here the bars. so for example, if I say here I have a class here, A1, so this class A1 is going to be denoted by here some bar like this, so this will denote here class A1 and similarly class A2 can be denoted by here another bar, say here A2 and so on, they will be here AK, so width of these bars like is this, this they can be same or they can be different, right? but it is always advised to have the equal width so that the bar looks better and now if you try to see the height of this bar, on the y-axis this is somewhere here, so this is going to represent the frequency or absolute frequency of the class A1 if I try to consider here Class A2 here then this height, this height is here somewhere here and this is going to denote here point F2, this is point here F1 and this is at Point F2 and similarly here this height here is proportional is actually see here FK. so you can see here that the heights of the bar, of bar is proportional to the frequency, now instead of frequency one can also use the relative frequency, so in that case the F1 is going to be changed by F1 upon N and so on so these heights are going to represent only the relative frequency. So now I have to option I have this bar and the height of the bar is F1 or say F1 by n that is the absolute frequency or the relative frequency. The advantage of using

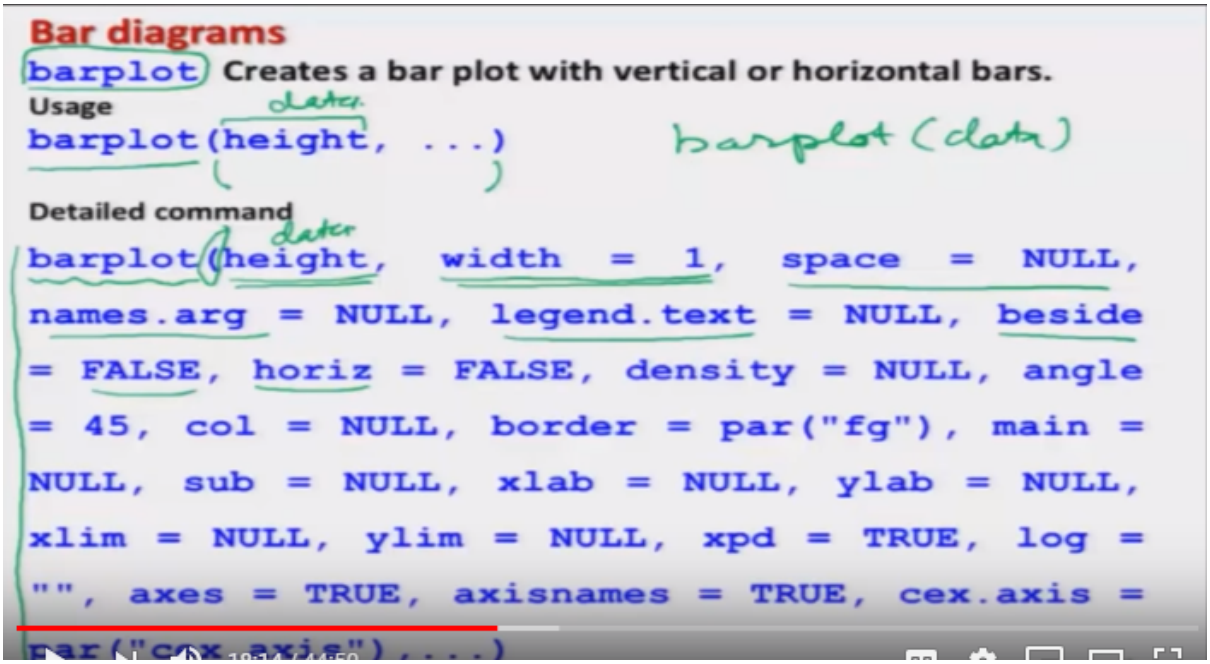
the relative frequency is that, that the maximum value of the relative frequency is always 1. So it becomes easier to compare the heights of the bar.

Refer Slide Time :(14: 08)



So in case if you see a diagram like this one, small bar and a higher bar, by looking at this diagram I can always say that this has a lower frequency and this has a higher frequency. suppose if I say these bars are going to indicate the number of shirts sold in a shop on a given day, so I have known to shops and sales are represented by the height of the bar, so by looking at the height of the bar, I can very easily conclude that which of the shop is selling more number of shirts, right? Now the question is this how to create this a bar diagram or say bar plot on, on the R software.

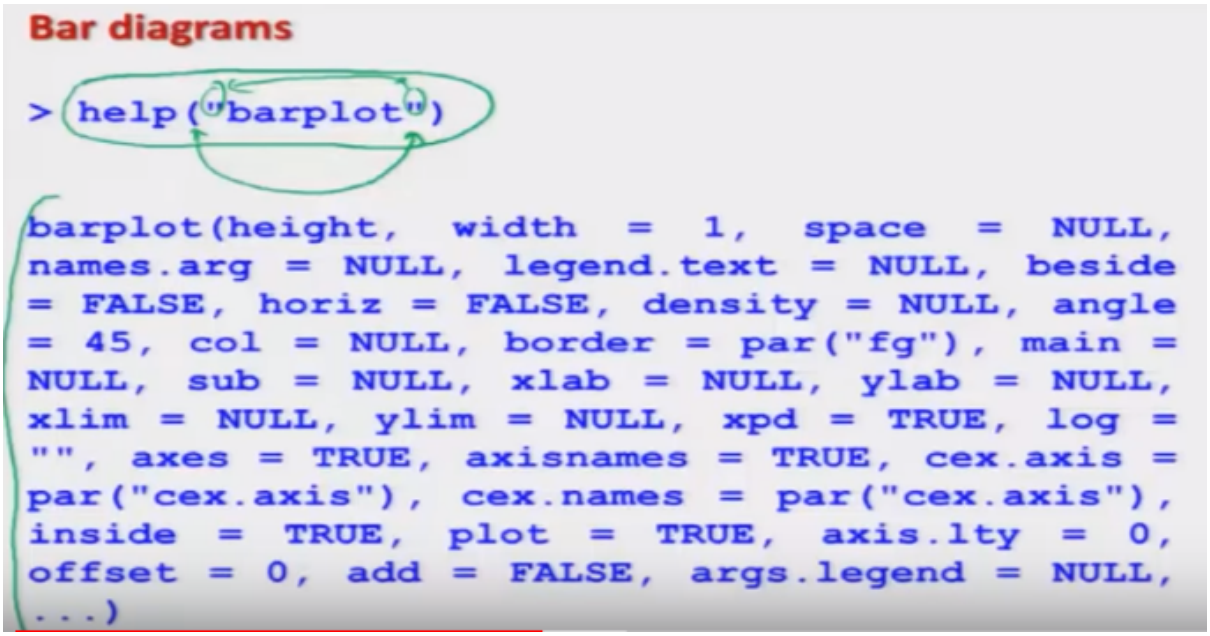
Refer Slide Time :(14: 56)



In R software we have a command here, bar plot. This bar plot helps us in the construction of bar diagram. When we try to construct a graph, then there are many, many parameters and you would like to handle those parameters so that you can get the outcome in the required format, which is more suitable to understand and for that we have different parameters. for example if you want to create a graphics then there will be x-axis there will be y-axis so you would like to put some desired labels on your x axis and y-axis that what they are representing, you need to control the width of the bar, different bars are representing different things, so you would like to add information inside the graph that which of the bar is indicating what you would like to give different colours to the bar and there are many things that you can do. So when we try to use this command bar plots, then I have two options in a simple language I will say, simple bar diagram and in an in simple way I will call the other option, bar diagram with more options. so if you simply want to create a bar diagram, then I need to use this command bar plot and inside the arguments, I simply have to give here the data inside a variable called here as a height and that will work, so if I simply try to type, bar plot and say here data vector this will give us a simple bar diagram but suppose you want to modify it, you want to improve it, so that it looks better, then in that case the detail command of the bar plot is here like this and you can see here this is the command here bar plot and this is here the argument, first value is nothing but the height, which is here the data second is here which is equal to one so that is going to control the width of the bar similarly there is here a space, there is here names dot arg, legend dot text, besides equal to false horizon-- and soon you can see here that this is a long list so next question comes, how you will learn all these things? I would suggest you simpler option is to take the help on the command bar plot. Because it is practically impossible to keep all the commands always in your mind, so best option is this, R is free software that will always be available with you. try to look into the help menu of bar plot and there you will see that the interpretation of each and every parameter,

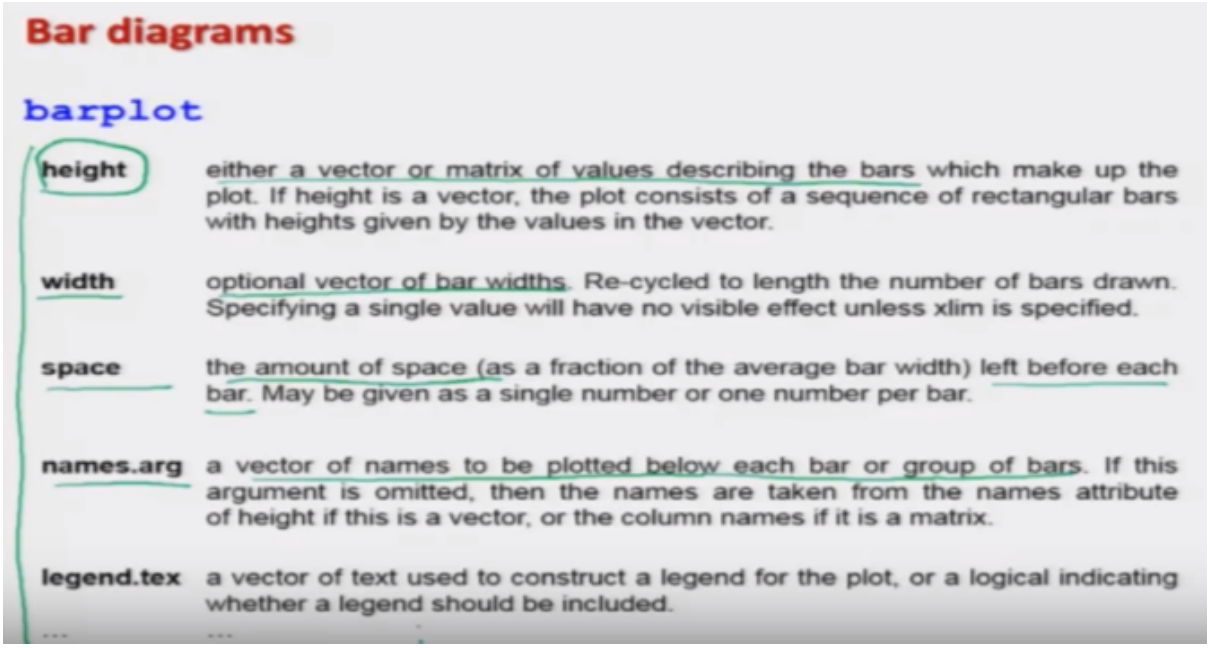
that is expressed inside the argument, that is very well explained there, so whenever you need, you simply try to read that part and execute it, to make you comfortable to make you understand, I will try to take care some option and I will try to add them one by one, through two examples.

Refer Slide Time :(18: 33)



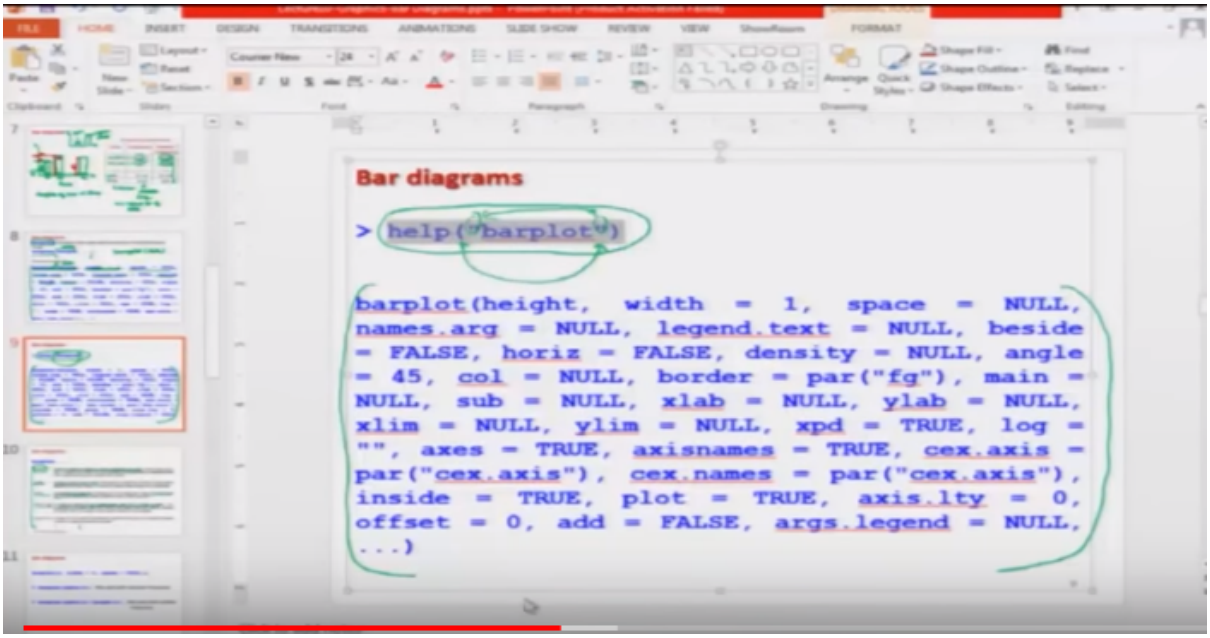
But before that in case if you really want to know or if you want to take the help from the bar plot then how to do it, in order to do it you simply have to use here the command help and inside the arguments, within the double quotes, you need to write bar plot, this syntax is not only for the bar plot, but this is for all the graphics, also all the commands, this is one of the way to obtain the detail help and now I will try to show you on the console, R console that how this happens and how do you get all this information but before that please try to have the loop on this slide and the next slide. You will get the same information whatever is mentioned here, you can see this is our detailed knowledge.

Refer Slide Time :(19: 33)



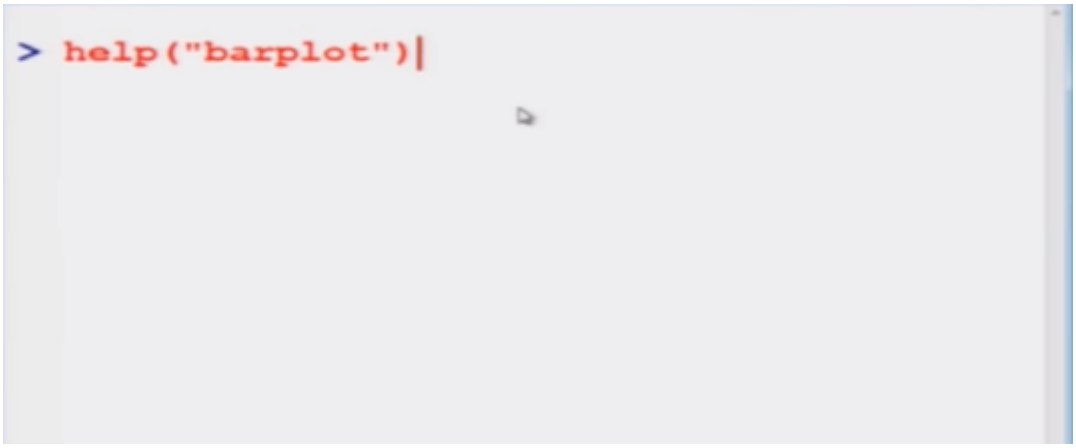
And in the next slide you have all the details, like as what is represented by height, either a vector or a matrix of values describing the bars and so on? What is here the width? this is an optional vector of bar widths space, the amount of space left before each bar names dot Arg, this is a vector of the names to be plotted below each bar or the groups of bar and so on? And there will be a long list. so what you need to-do, you simply need to read this help bar and this will be your best feature to understand what is really happening and based on that you can do that right?

Refer Slide Time :(20: 16)



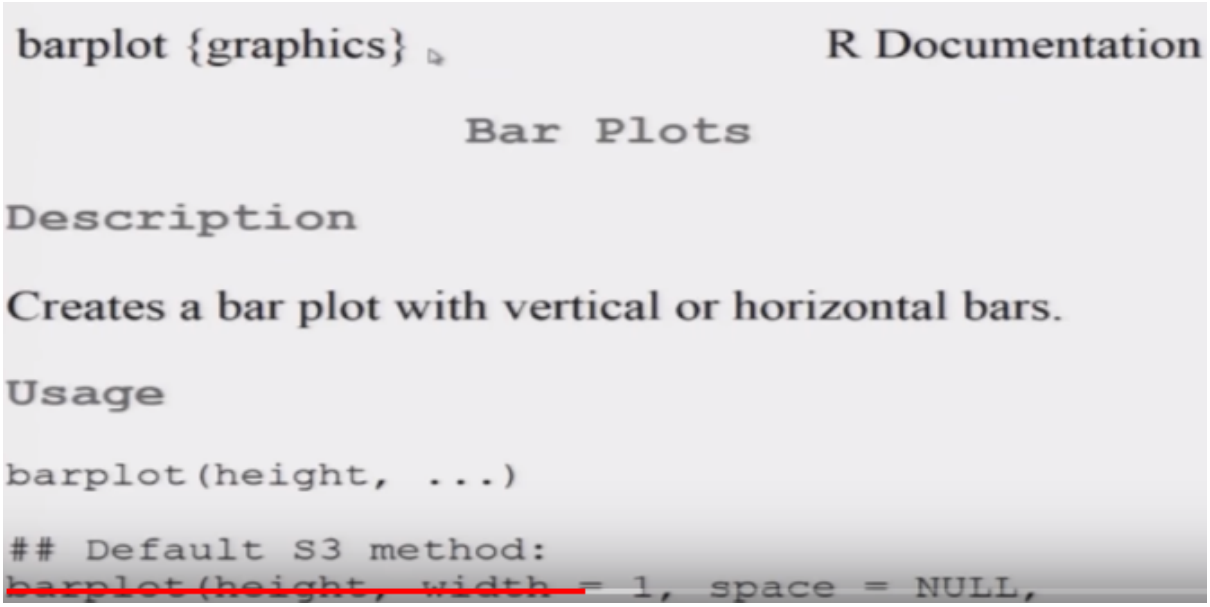
So you will see here, now I try to copy and paste this command on the R console.

Refer Slide Time :(20: 23)



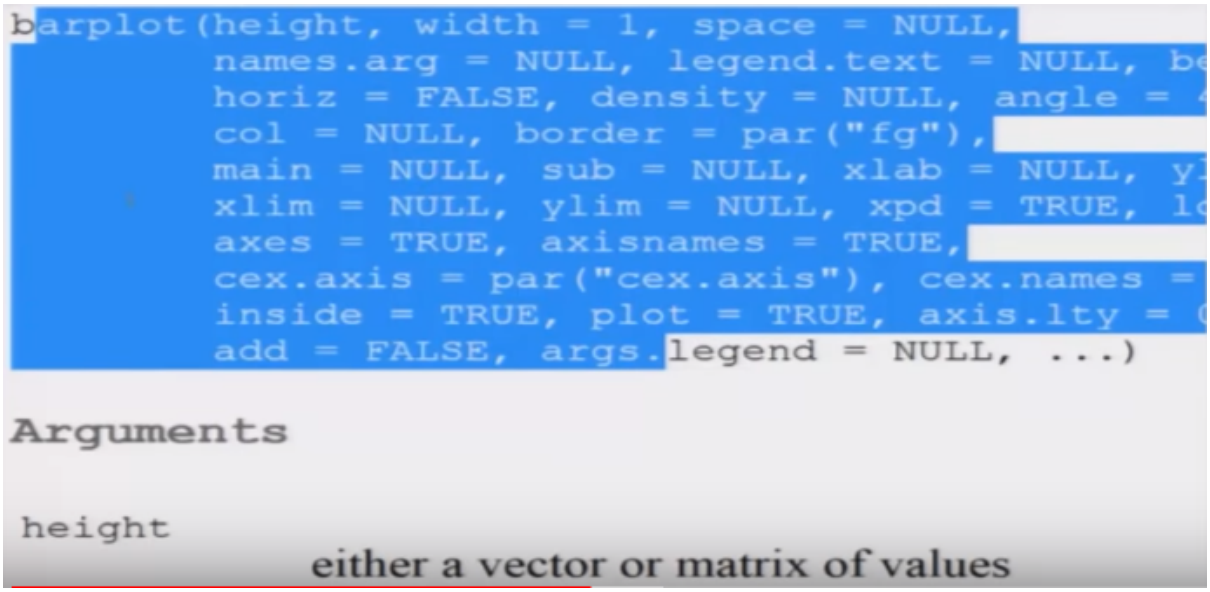
You can see here as soon as I will press enter, this will come on the internet and a website containing the help on this bar plot will be opened. this is one possible way, so in order to use this type of help you need to have an internet connection or else you can also go to the help menu were here inside the R software and then you can have a detailed thing, but here as soon as I say here enter you can see here what is happening.

Refer Slide Time :(20: 53)



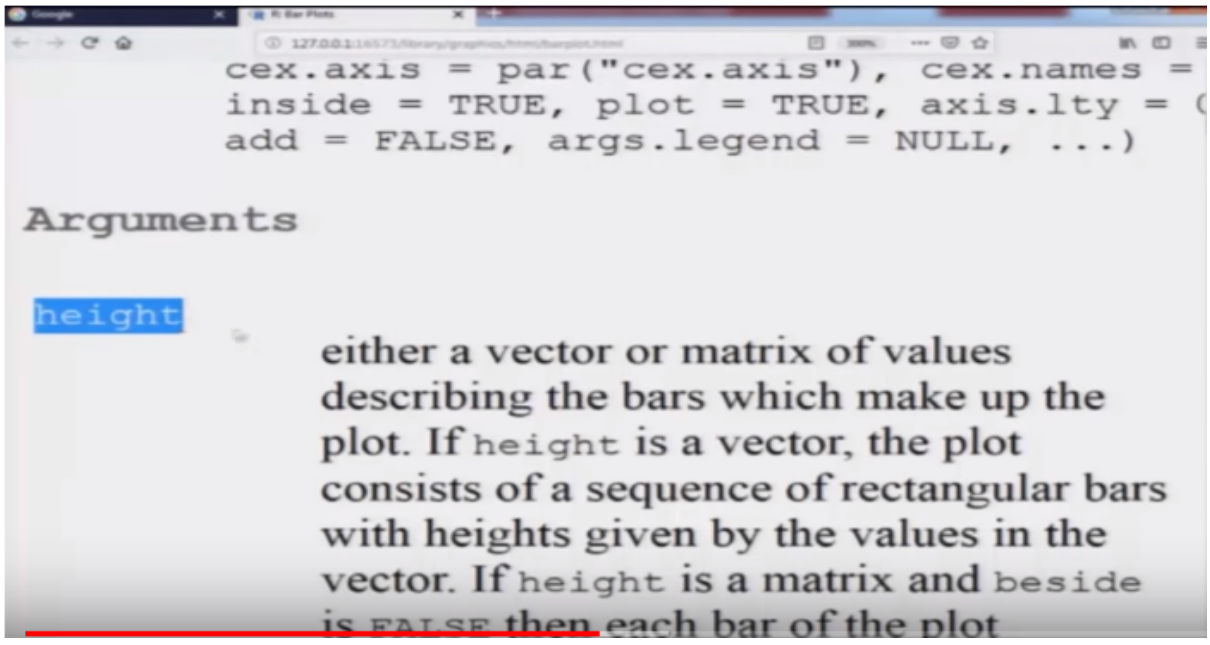
you can see here that this internet site, is opened here and actually this site or this command has taken you directly to the R server and on the R server you have the latest help whatever is documented, that is available for you, so you can see the advantage of this R. you are getting the best possible help over here.

Refer Slide Time :(21: 29)



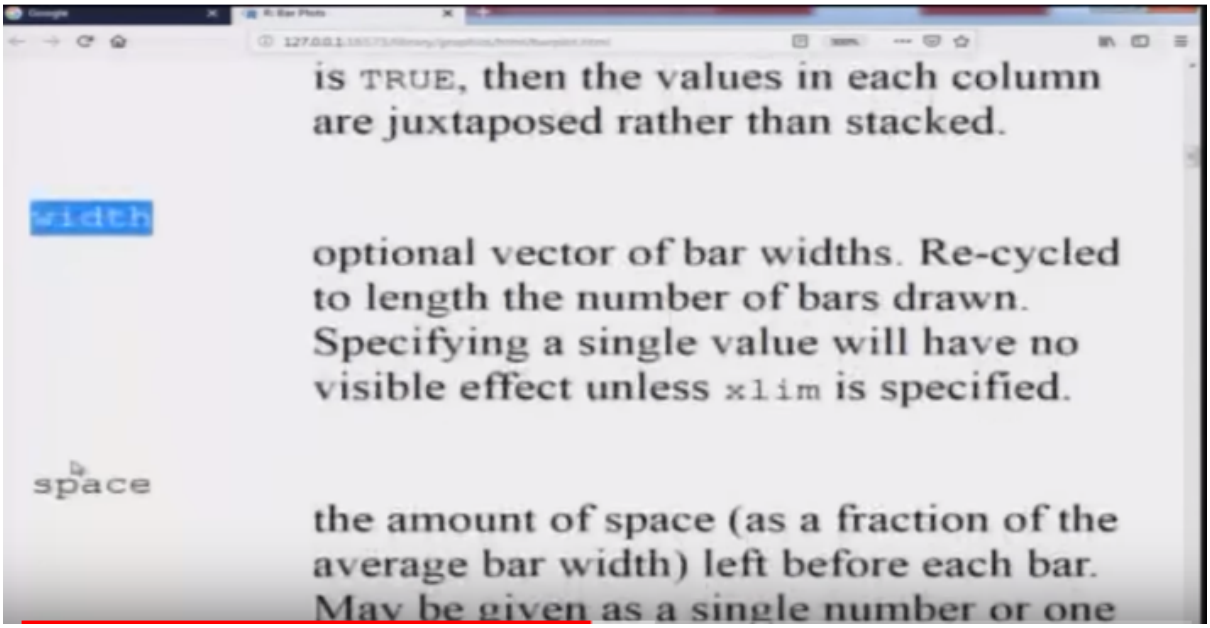
And you can see here if I try to scroll down you can see here that this part is the same part which I had copied and pasted here

Refer Slide Time :(21: 39)



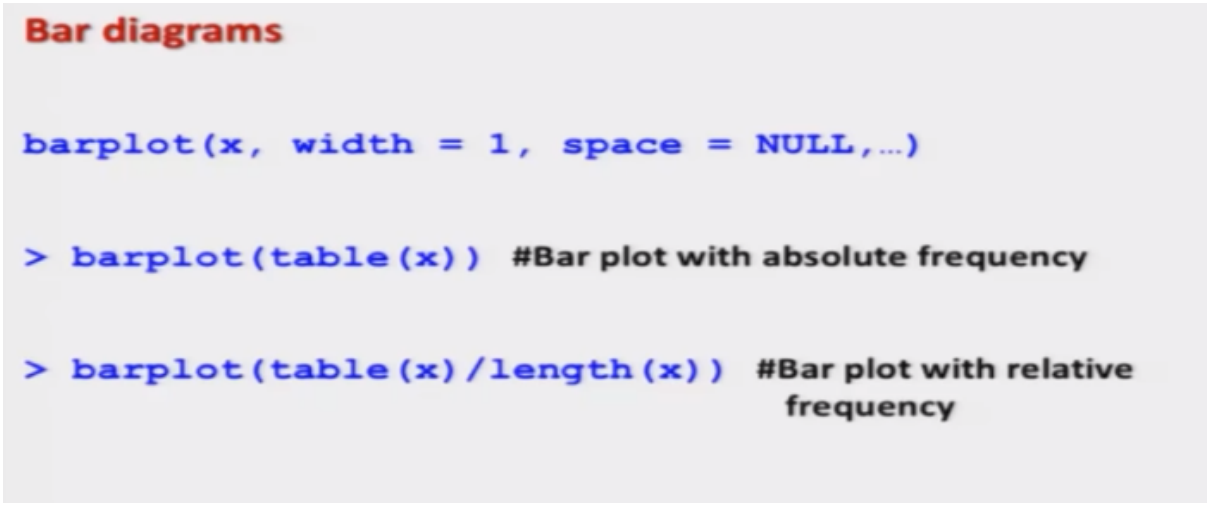
And if you move further you can see here that there are different here argument like as here, height and height is trying to give you the all this information something like this

Refer Slide Time :(21: 45)



and if you scroll it more there is an detailed interpretation for the width and detail interpretation for the space and so on you can see here this is a long list so now it depends on your capability that how much here you want to learn and how beautiful or how informative graphic you want to create, right? And you can see here I have simply copied and pasted this thing just to give you an idea right? So now one very important aspect in bar plot whenever you want to construct the bar plot, please decide you are constructing the bar plot on what? On the individual data or on the categories, the answer is this, we want to create a bar plot on the categories, for example, in case if we have data, which has been categorized in two categories one and two, suppose you have got hundred data values, would you like to plot those hundred values or

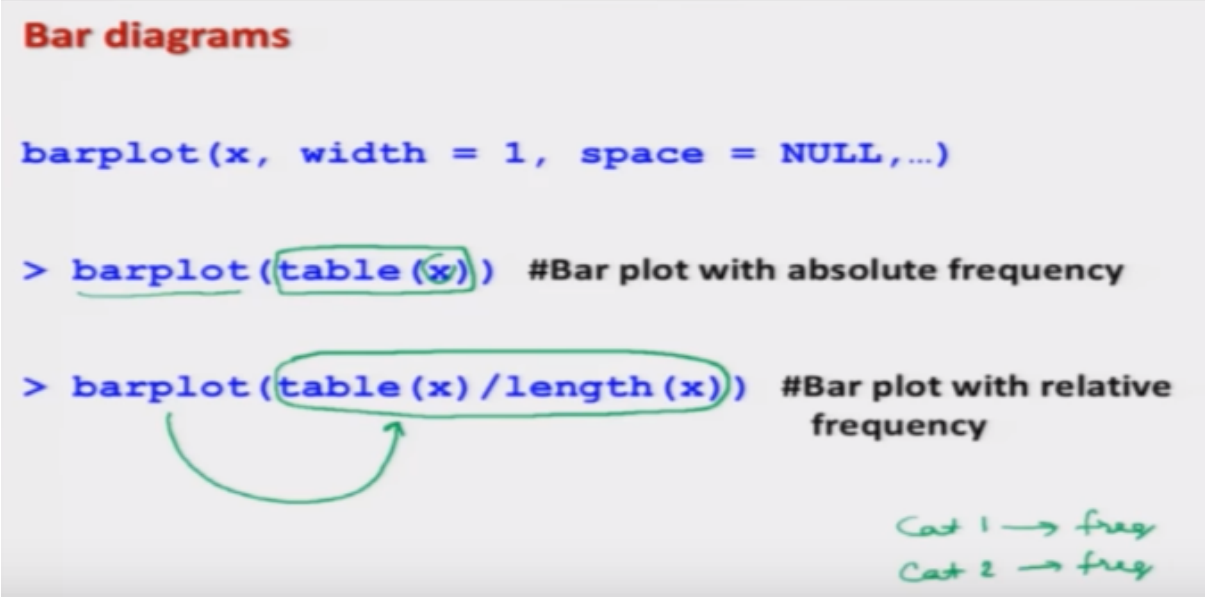
Refer Slide Time :(23: 02)



You would first translate those hundred values into two categories? category 1, category 2 and based on that you will have here their frequencies or the absolute frequencies and actually, we would like to

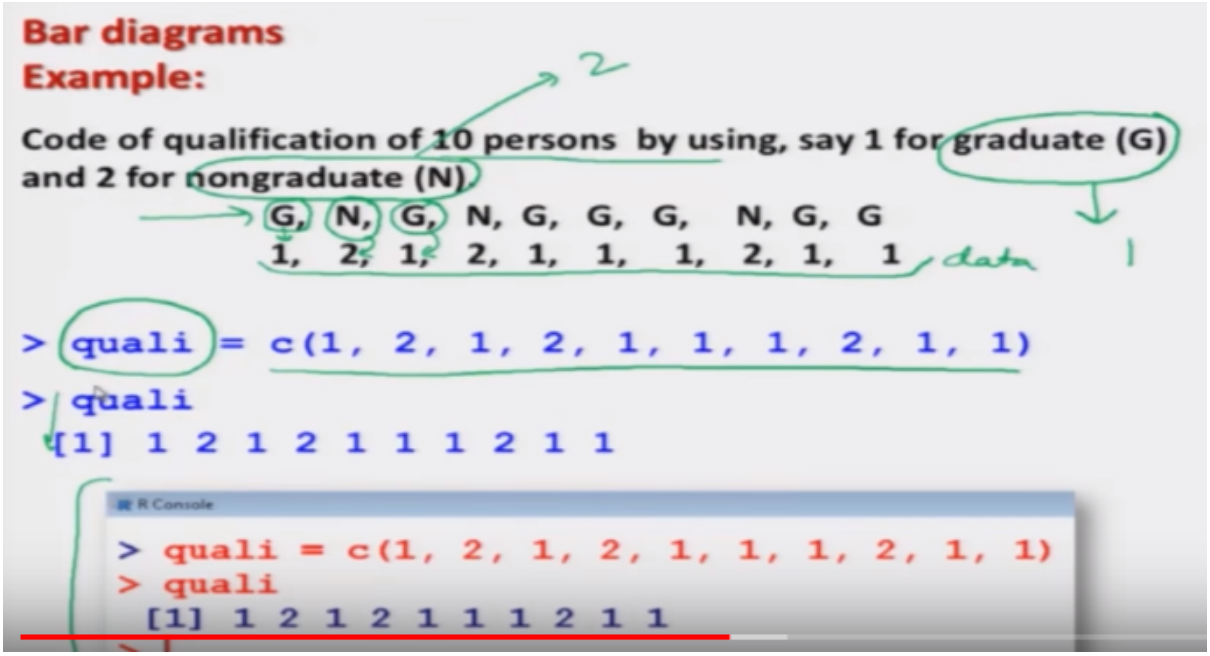
plot the frequencies, so first you need to input your data in the form of a frequency and the frequency can be obtained by using the command table. So that is why suppose I have some data vector here X.

Refer Slide Time :(23: 31)



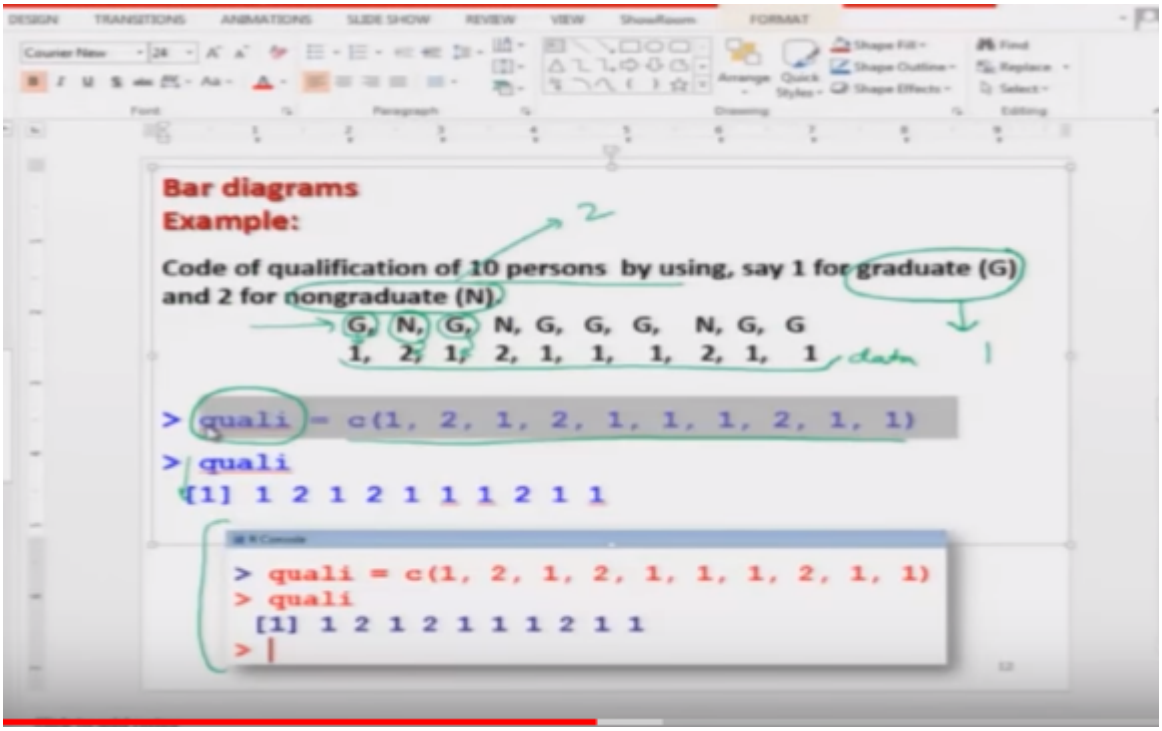
So if I want to create the bar plot using this command, bar plot then first I need to transform this data into a frequency table from, using the command table and then I have to create the bar plots, similarly if you don't want to use the absolute value and if you want to use a relative frequency, then in that case the same command has to be transformed and in this case what I will do that now I will try to operate the command bar plot over the table divided by length of X, we have learned that once we try to use this command, this will give us a frequency table frequency distribution using the relative frequency, so now I would try to create a bar plot on the data which is given by table X divided by length of X.

Refer Slide Time :(24: 42)



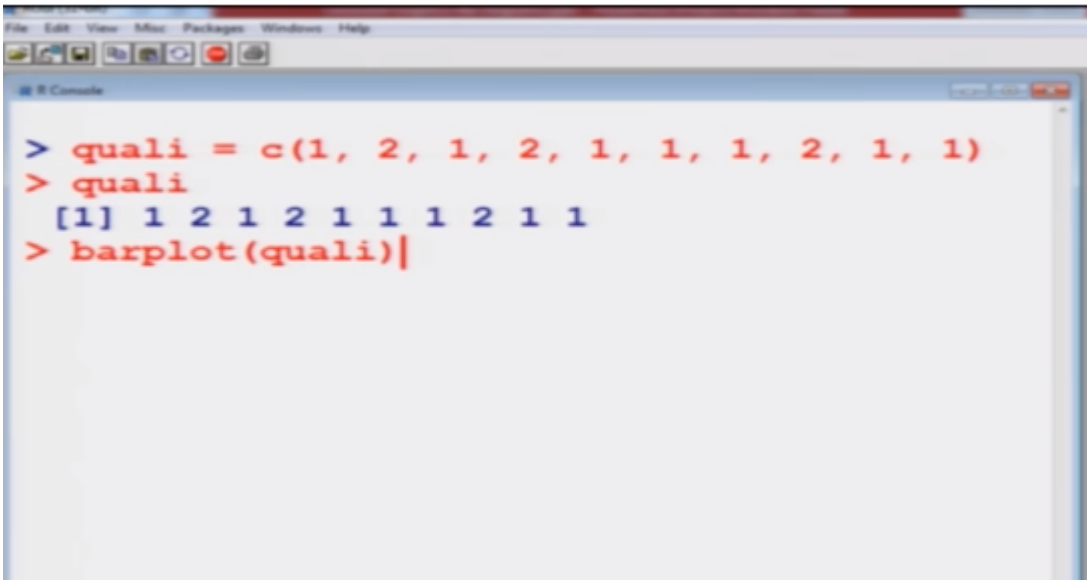
So why not to consider here a simple example and we try to see how these things are happening, so now let me take a very simple example, so I have data often persons and we have recorded whether the person is graduate or non graduate, so in case of a person is graduate, then it is it denoted by here G and if the person is non graduate it is denoted by here N. now so we have this dictator here G and G and and so on. So as we had discussed earlier this data cannot be exposed to an aesthetical tool, we need to convert it into a number, so what we have done? we are giving this graduate a value 1 and a non graduate a value 2, so now in this case the person is graduate, so I am giving it here value 1, then the second person is non graduate I am giving it value here 2, third person is graduate, I am giving here the value 1and so on. so now I have a data which is here 1, 2, 1, 2, ,1 1, 1, 2, double 1. this is my here data, on which I would like to create my bar plot. So I try to store this data using the C command here, in variable here ‘quali’, which is a I have used a short form of qualification. So you can see here I have entered this data and this will look like this and this is here with the the screenshot. Well I will try to show you on the R console also. So let us first try to go to the R console.

Refer Slide Time :(26: 22)



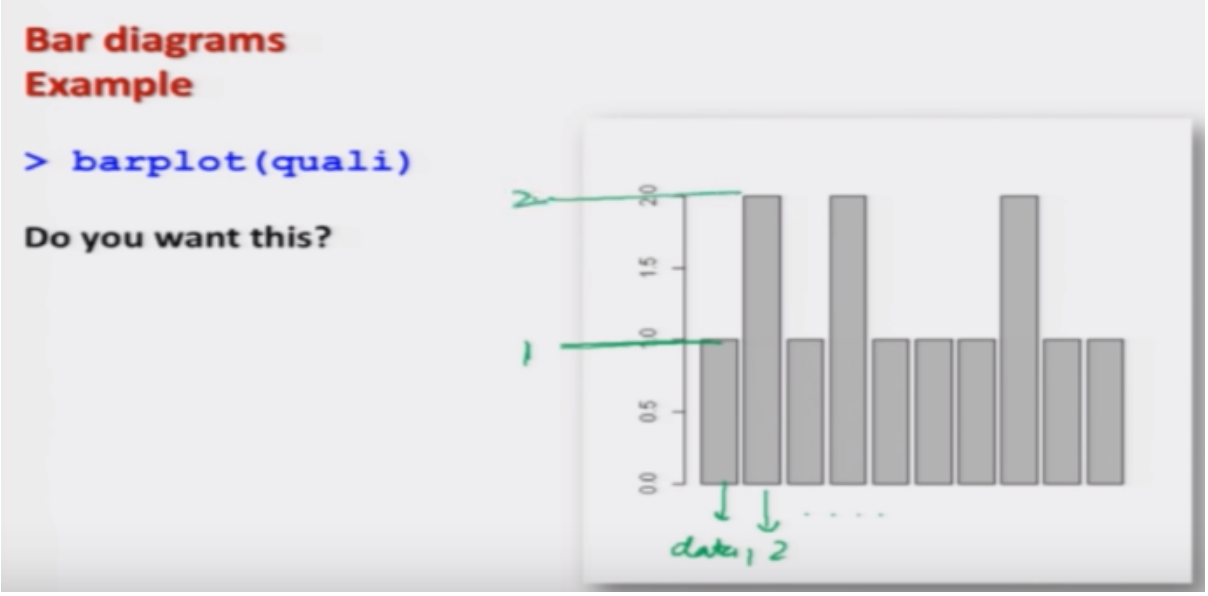
First let me create this data vector over here, right?

Refer Slide Time :(26: 25)



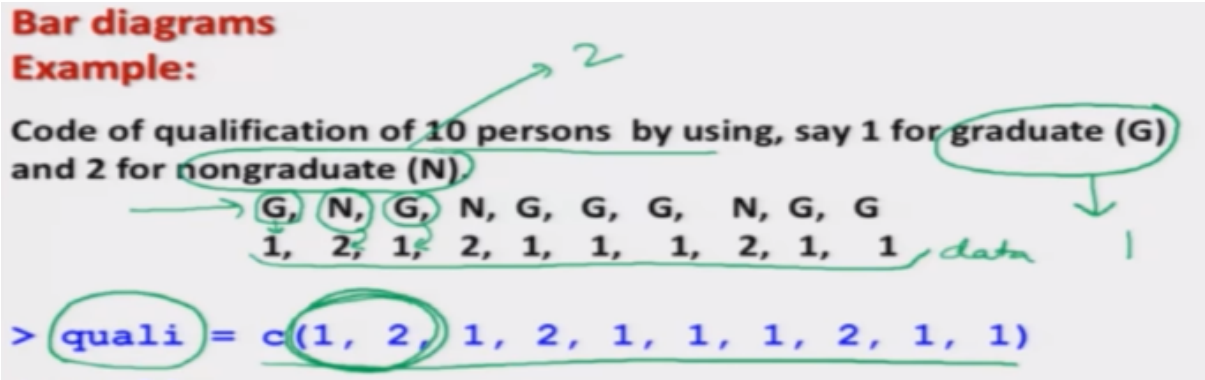
So you can see here this is data, now please try to observe what is really happening. Now I will say here I have been told to use the command bar plot on this data to get the bar plots, so I try to use here bar plot and this quali,

Refer Slide Time :(27: 01)



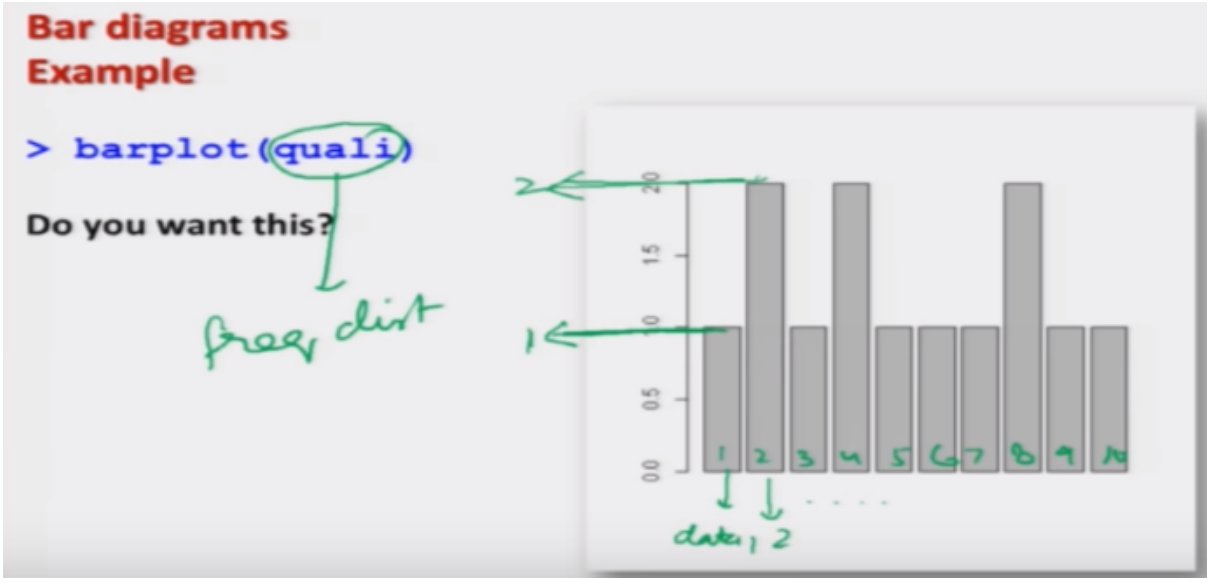
Or now if you try to observe what are you getting here do you think exactly do you want it this thing for example I will just copied and Pierce this graph over here if you try to look at this graphic you would realize that no, this is not matching what we wanted because this is giving me a graph of data 1, 2 and so on, right? This is giving me only here two values 1 and here 2.

Refer Slide Time :(27: 22)



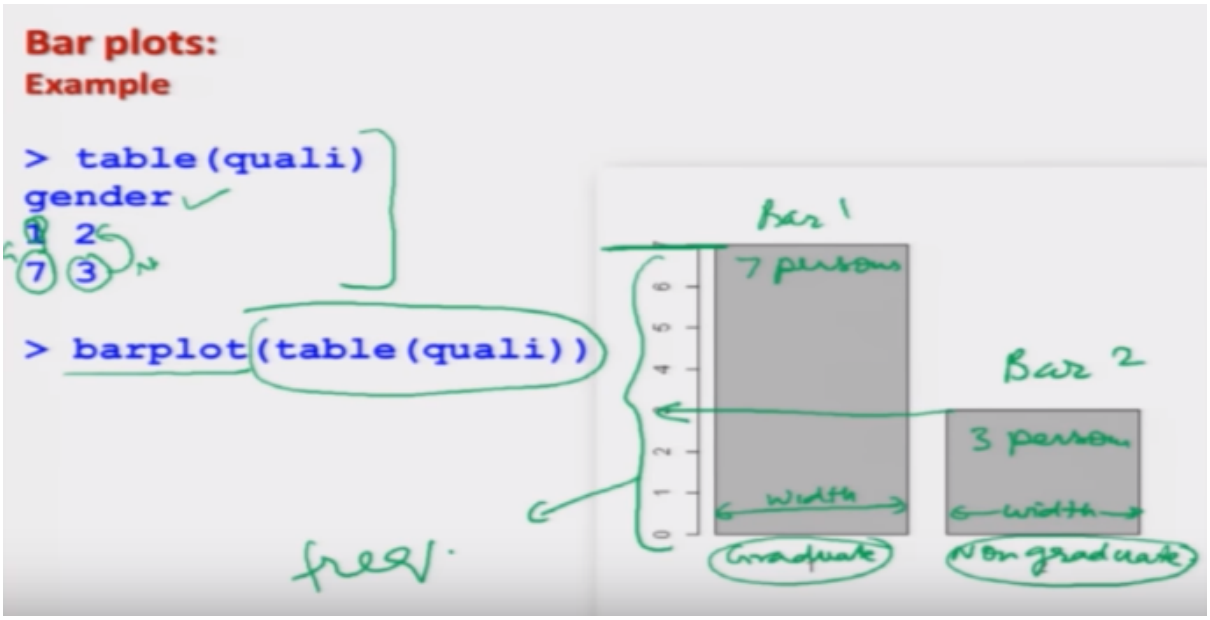
So if you try to look at that data first 2 people this is here 1 & 2

Refer Slide Time :(27: 27)



And this is giving here first person here 1 and second person here to be 2 and then there are such 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, what is this? we didn't wanted this thing, well this is the very common mistake what we always make while making this bar plots or say pie chart that I will show you later on that we directly expose the original data set, but what we have to do here I need to first create the frequency distribution of this variable here quali and then I need to give here the command, bar plot

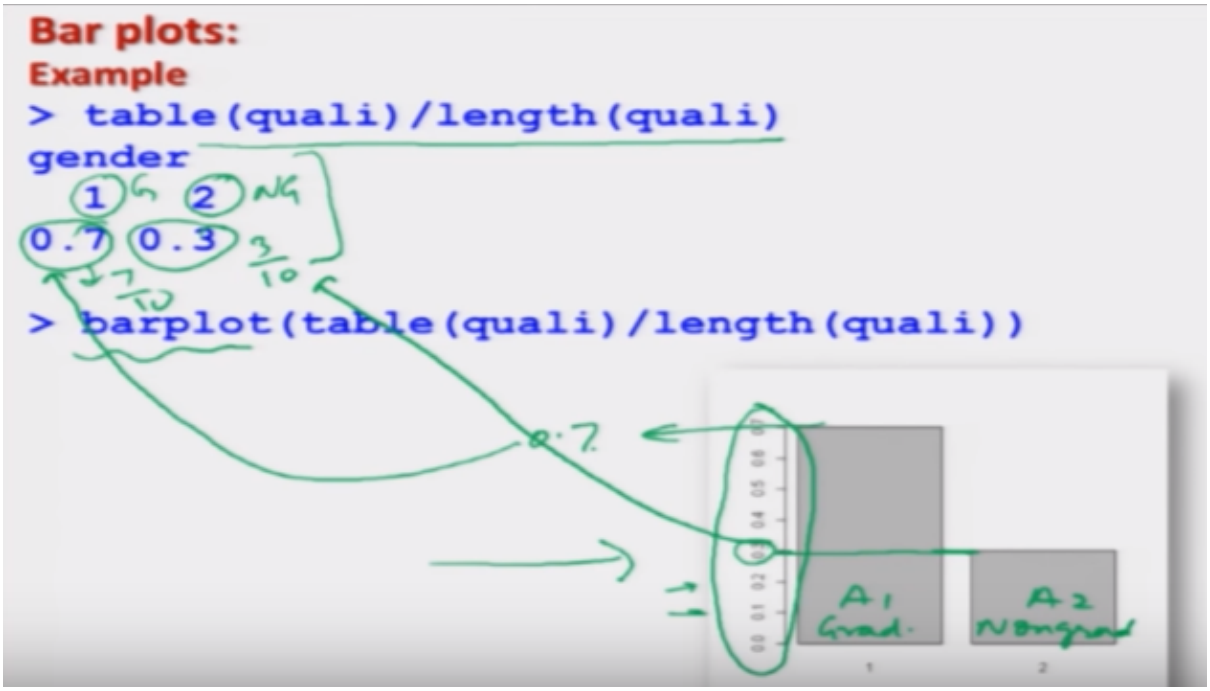
Refer Slide Time :(28: 09)



And now this is what I'm trying to do here. you can see here now first I try to create the data of the frequency table, using the cable command on the data quali, this comes out to be here like this gender and then 1, 2 and here 7, 3, so that simply indicating that there are 7 percent which are graduate and there are three percent which are non graduate and then I'm trying to use the command bar plot on this data table quali and now I am getting here this bar chart, you can see here this is bar number one and

this is here bar number two. so this bar number 1, that is representing the graduates and this is and the bar number2, this is representing the non graduates and by looking here you can see here at this height, this is indicating that there are seven percents, the absolute frequency here is seven and similarly here if you try to see this is here three, so there are three persons, so by looking at this bar diagram I can very easily conclude that there are seven percent which are graduate and there are three percent which are non graduates and. in case if you see what is indicating here? this is here no nothing but the frequency or the absolute frequency and the width of this bars, with of these two bars, say this is here width and this is here the width this is the width here they can be arbitrary, but anyway as I said, it is nice to have the bars with equal width, further easy to understand graphics, right? So this is how we try to create this thing. Now I will try to show you on the

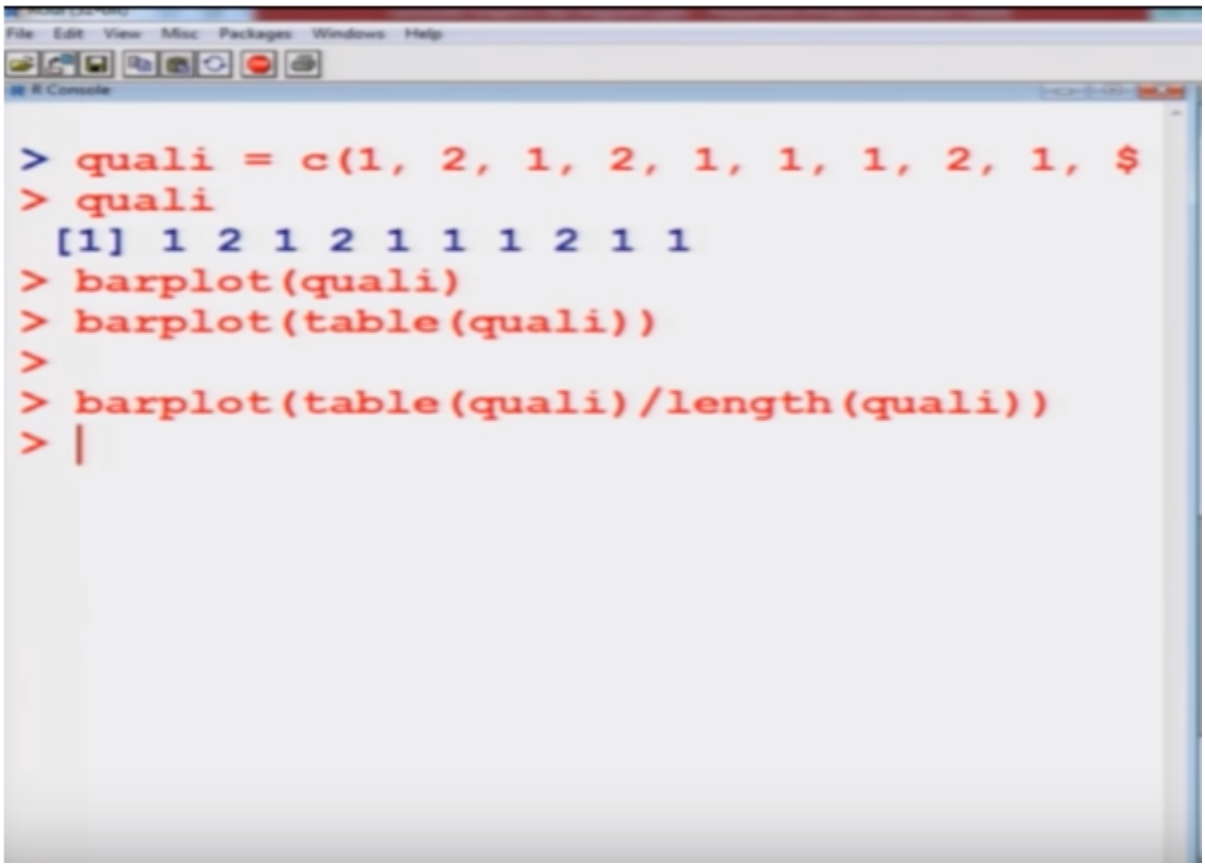
Refer Slide Time :(28: 09)



R console but before that, let me also show you here that in case if you want to create this bar diagram using the relative frequency, how are you going to do it? so first I try to create the frequency distribution, with the relative frequency, so you can see here, once I try to execute the table command on quali divided by length of quali, I get here this type of frequency distribution, where this is indicating graduate category, this is giving me no graduate category and this point seven is actually seven upon 10 and then 0.3here is 3 upon 10, which is indicating the relative frequency of the two classes and when I try to execute the bar plot command over this data, then I get this type of graph, but now you can see here means again that this is my class A1 of say graduates and this is my hair class A2 of non graduate, so this graph is exactly the same as the earlier one, but there is a difference here on the y-axis. now you can see here these values are0.0, 0.1, 0.2 and so on and this is here the height which is indicating 0.7 and this 0.7 is the same thing here like this and this height of second bar

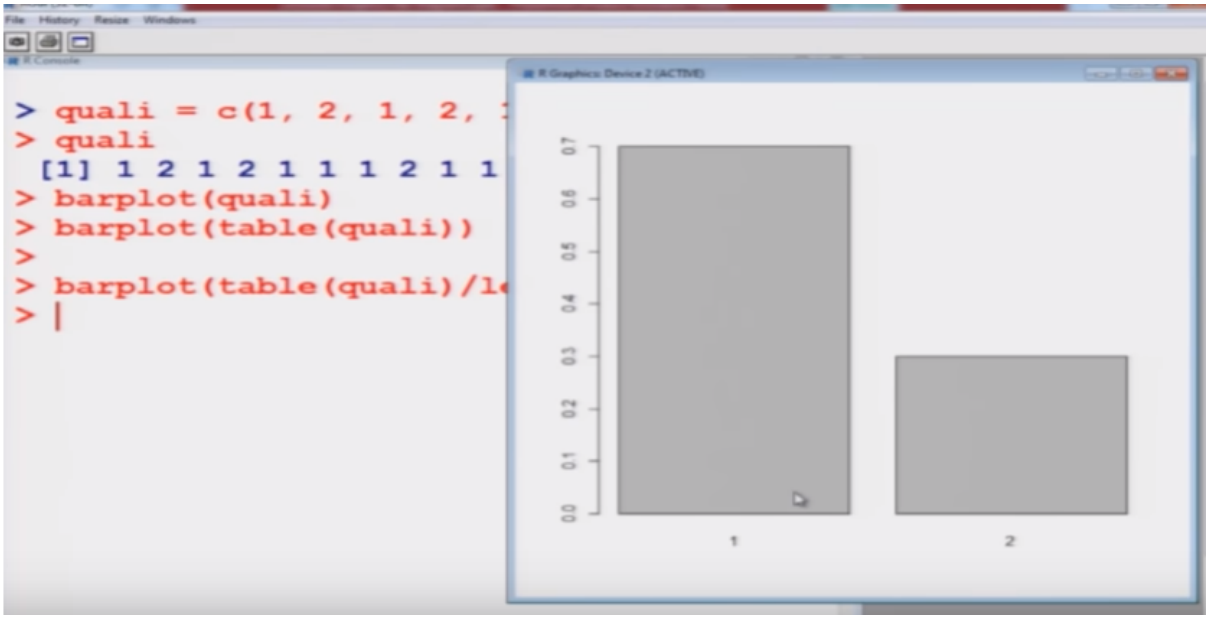
A2 you can see here this is 0.3 and this 0.3 is the same thing which is coming here. So this is how we try to create the bar lots.

Refer Slide Time :(31: 47)



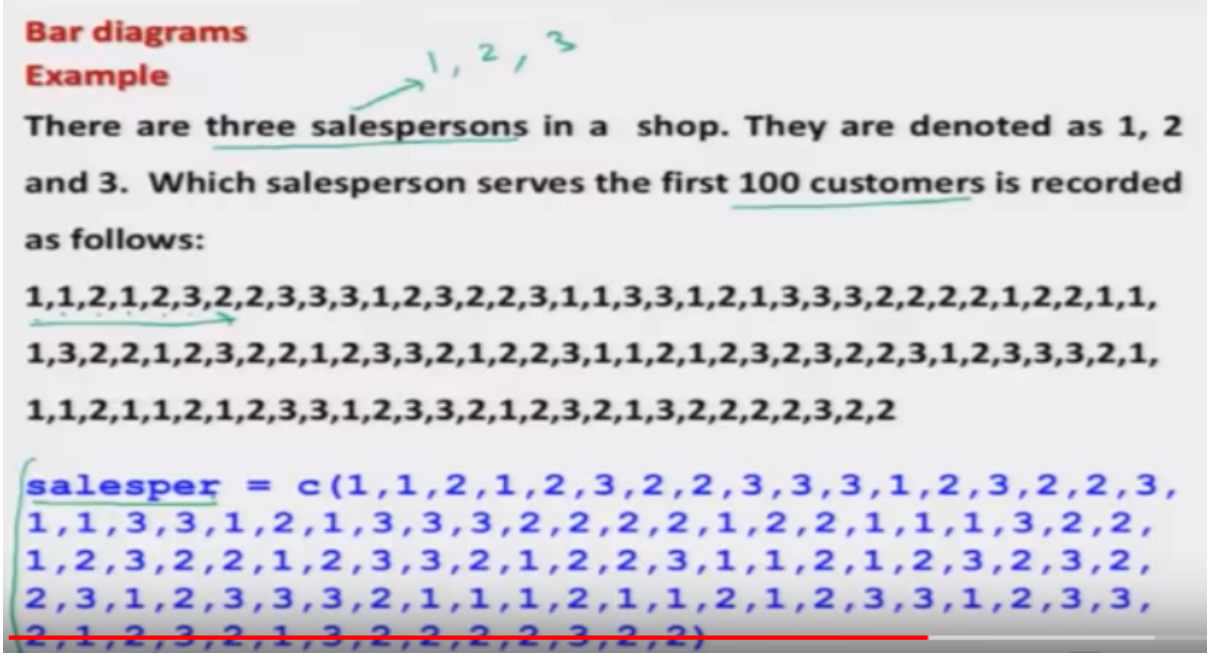
But now let me try to show you on the R console also, how you are going to do it. So now I would try to correct it and I would say I would like to make a bar plot of the table quali, so if you try to see it here quali and then this you can see here now, this is your here the graphic and this is here in the on the y-axis, you can see here where my cursor is this is 0, 1, 2, 3, 4, 5, 6, 7, this is indicating the frequency and if you try to use the same command with the length of quali, you can see here the graph.

Refer Slide Time :(32: 31)



The graph remains the same but now you can see here that that on the y-axis these frequencies have now changed and now they are relative frequencies. So the height of this bar is simply proportional to the relative frequency.

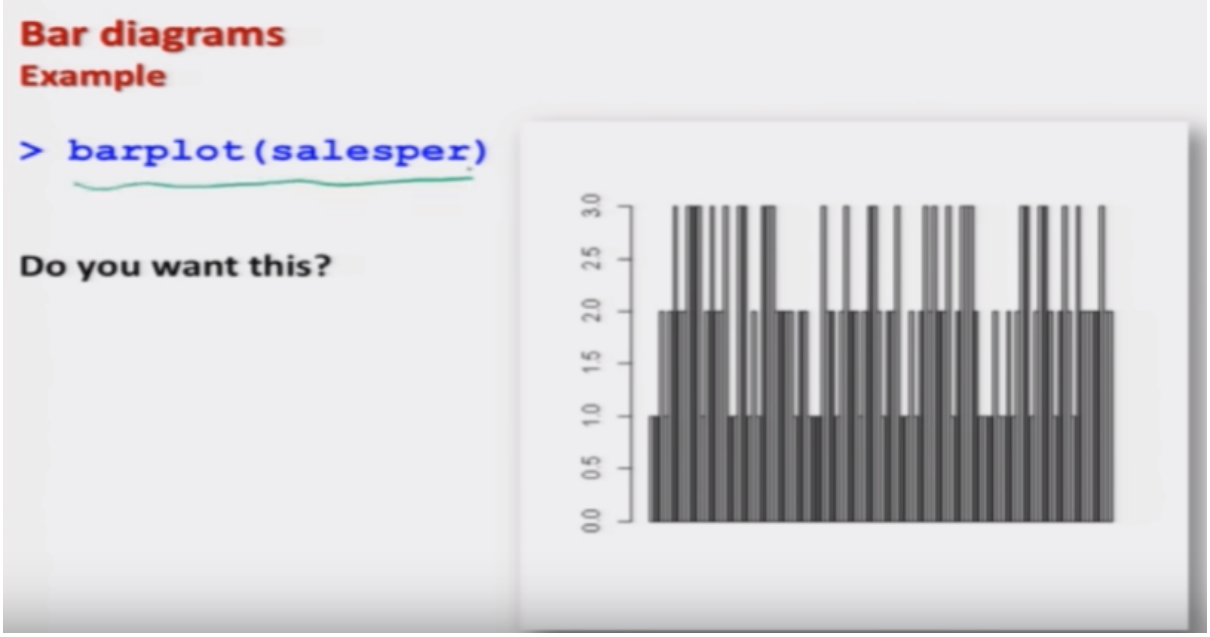
Refer Slide Time :(32: 44)



So this is how we try to create the bar diagram. Now can we take here one more example to make you understand better and so I have collected the data of say hundred customer, these hundred customers are visiting a shop and there are three sales persons and these three sales persons are indicated by numbers 12 & 3, salesman number one, salesman number two and salesman number three and when these customers entered into the shop, they are attended by a particular sales person. So now which sales person has attended which of the customer? this data is recorded for the first hundred customers

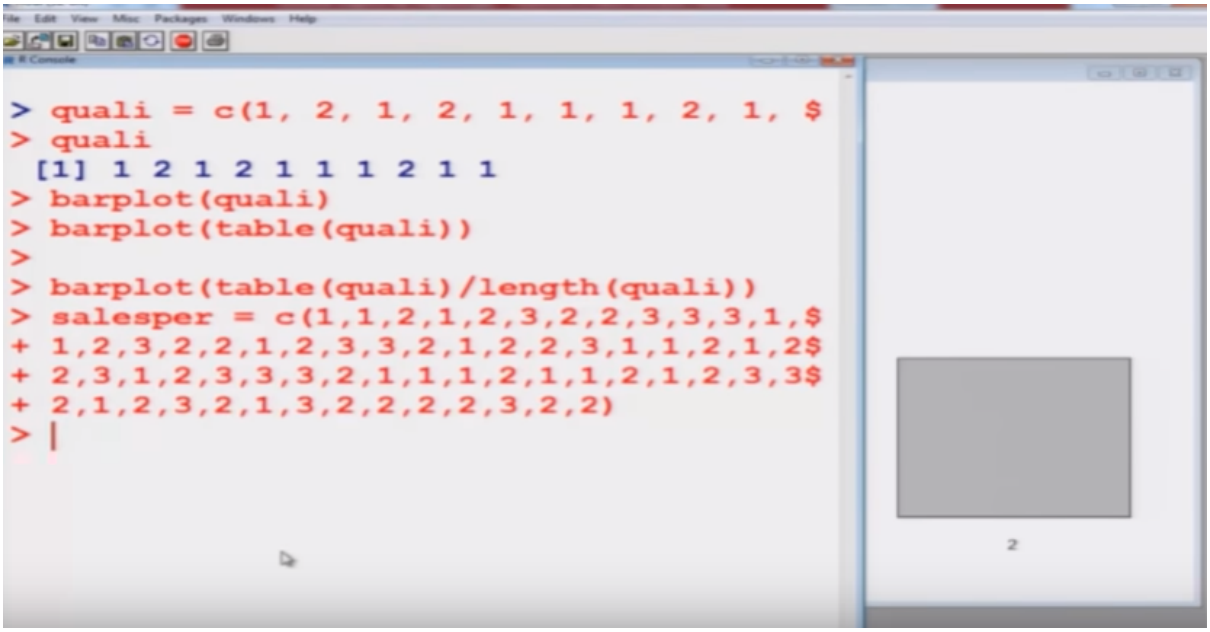
entering into the shop and so this data you can see here this consist of the numbers 1 2 & 3 only. 1, 1, 2, 1, 2, 3 and so on, so now you can see here by this small data set, set consisting of hundred values you can see this is only a say, no numbers numbers and numbers they are not giving you any fruitful information, so my first attempt will be to make some suitable graphic to have this information. So I would try to first attempt to create a bar diagram. So I would like to store all this data into a data vector head called ‘sales per’, which is a short form a ‘sales person”, alright?

Refer Slide Time :(34: 09)



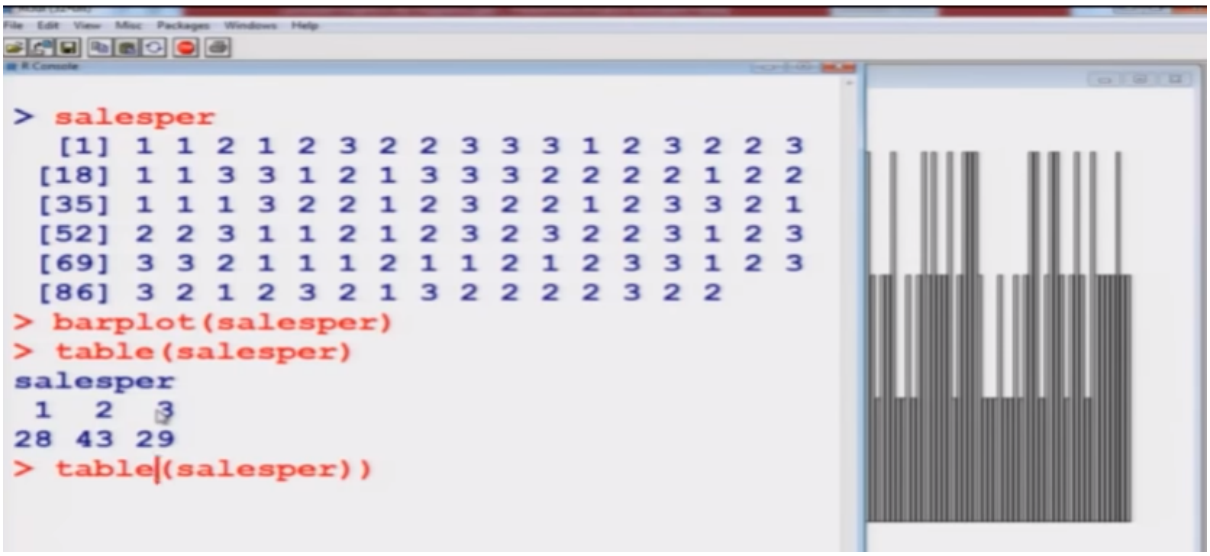
And now, after this if you try to make the same mistake here if you try to make it here bar plot of here sales per you will not get here the correct information, right?

Refer Slide Time :(34: 22)



So firstly let me try to show you these things on the, our console also and we will move together. So let me first try to create this data vector.

Refer Slide Time :(34: 22)

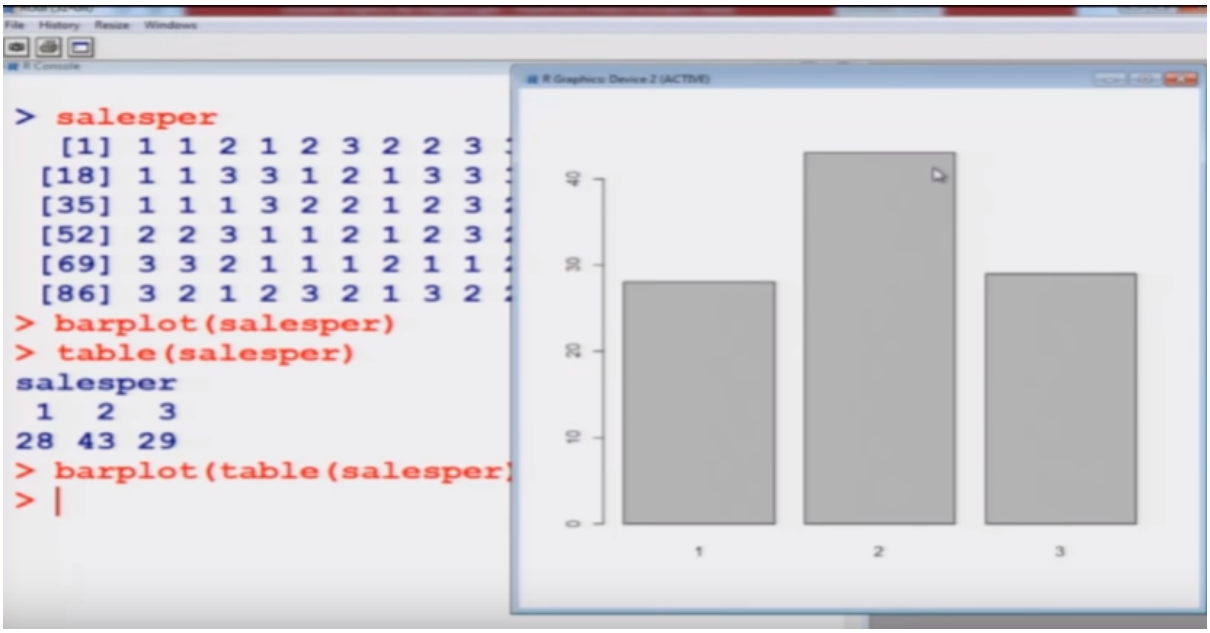


So I have created this data vector and you can see here now this is my data, right? and now if I try to create here a bar plot, of this thing, you can see here, you are getting a graph like this one, definitely you don't want this thing and yeah I accept that I have made here a mistake,

Because that correct option is this I don't have to use the data directly, but I have to first create the the frequency table. so I would first try to create the frequency table, you can see here looking at this hundred data values you are not betting much information but by looking at these three values, you are getting the information, that there are 28 customers, which are attended by the salesperson number

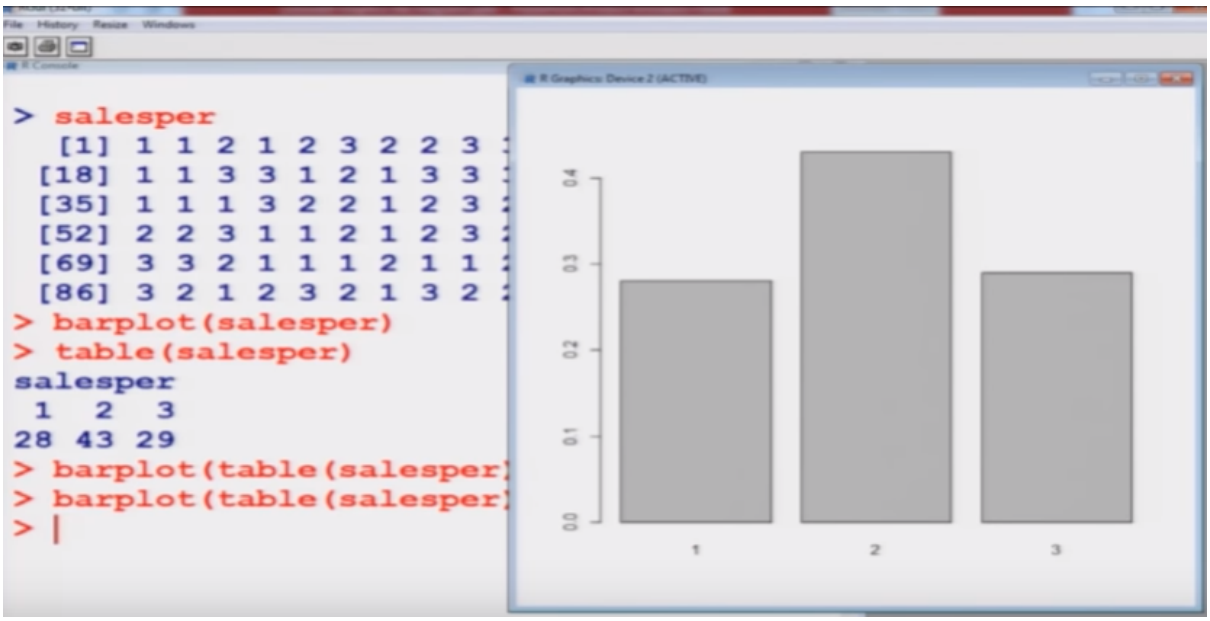
one, there are 43 customers, which are attended by salesperson two, and there are 29 customers which are attended by salesperson three, right?

Refer Slide Time :(35: 40)



Now I would try to create a bar plot over this thing. So I try to write down here the command bar plot and inside the bracket I have to get the data and you can see here that this is giving me this bar plot, right?

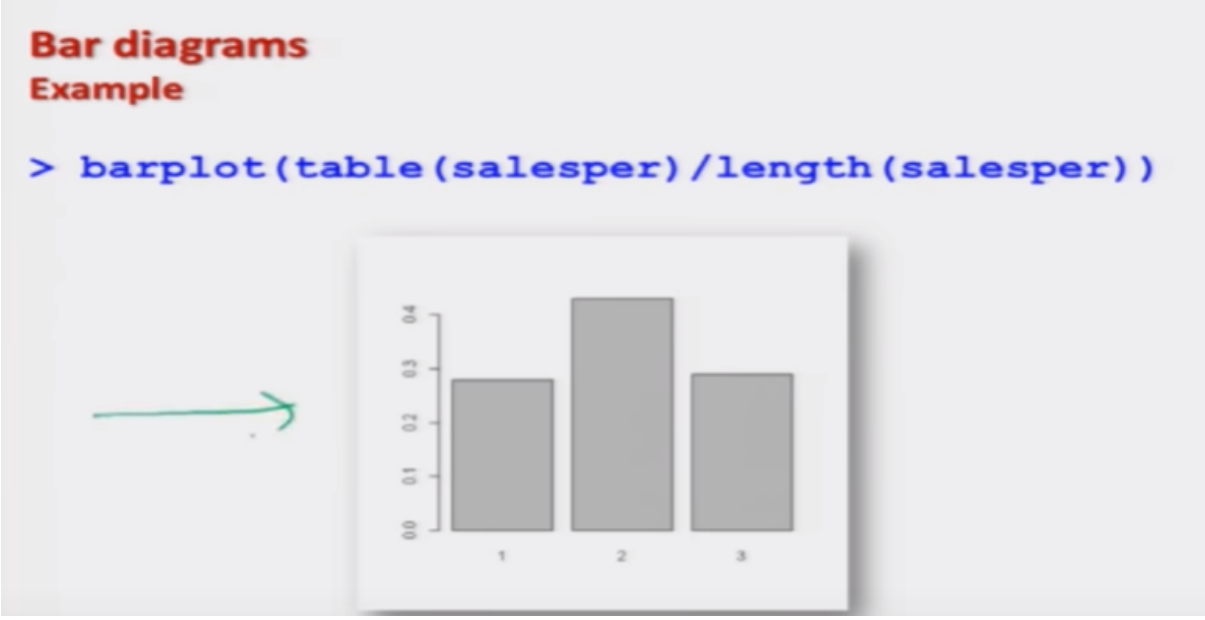
Refer Slide Time :(36: 01)



and similarly in case if you want to create this bar plot with respect to the relative frequency, then I need to modify the data vector as by length of salesperson, data vector and you can see here, now in

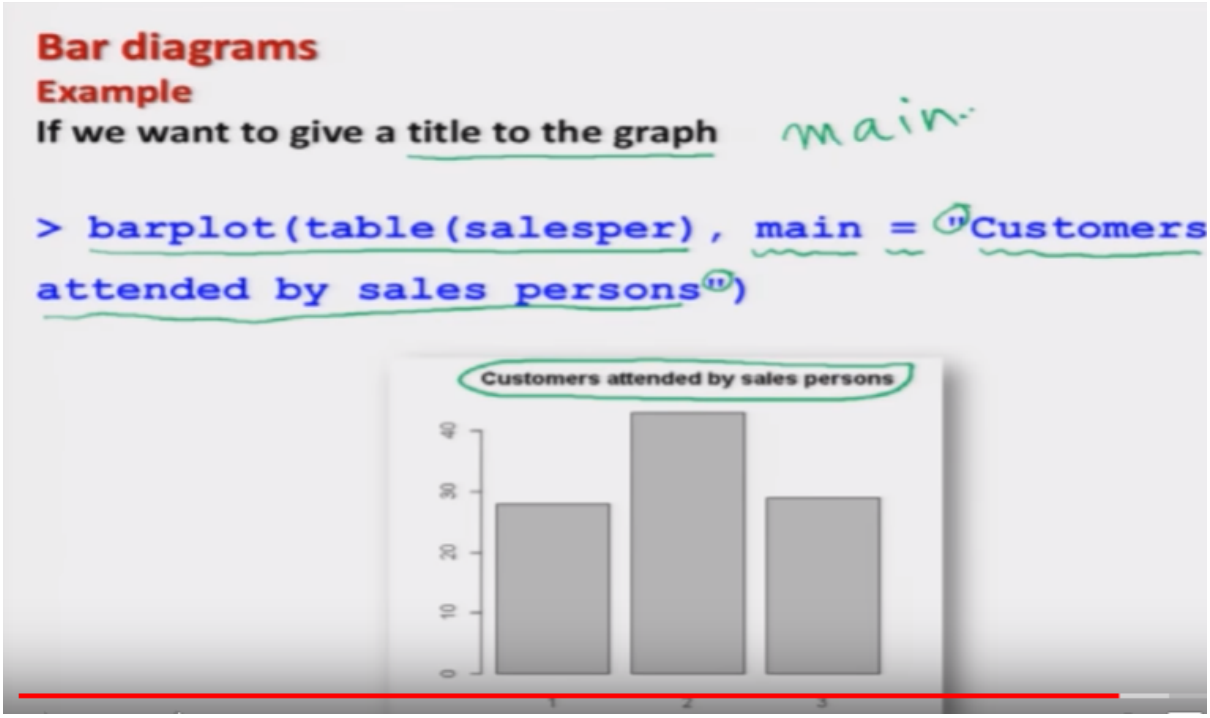
this case the frequency that was earlier as 10, 20, 30, 40 and so on, this frequency on they-axis is now changed to a relative frequency, right?

Refer Slide Time :(36: 28)



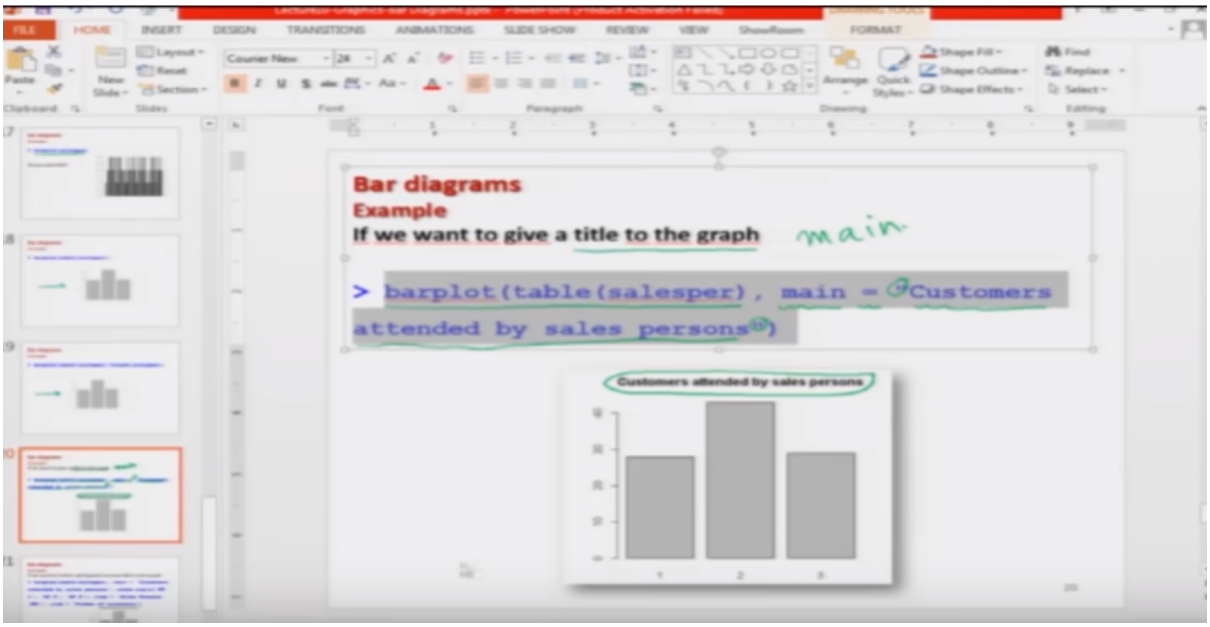
So the same thing I have copied and pasted here for your understanding and then I would try to do something more on this thing, so you can see here that this is the same bar plot, that we have just obtained and similarly in case if you want to have this bar plot with respect to the relative frequency this is obtained here.

Refer Slide Time :(36: 50)



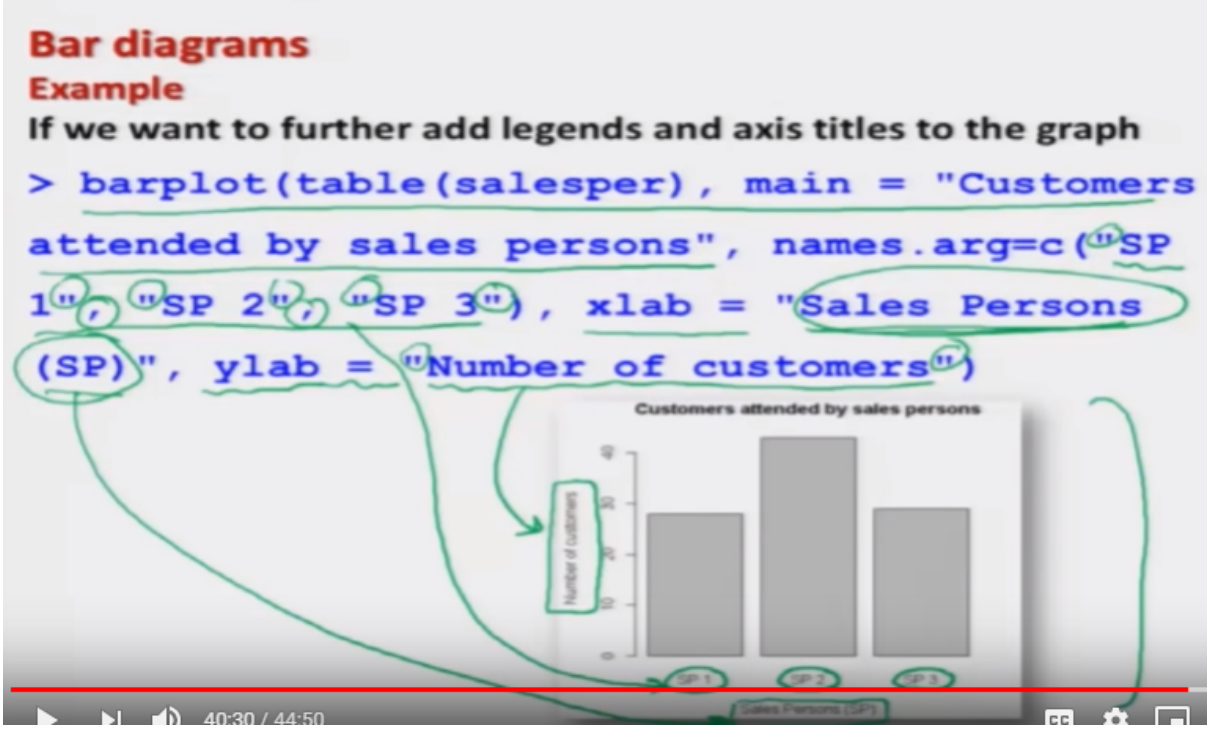
And now, now I would try to add some features to this plot and suppose I want to give here at title, so now you can see here this command, this command is the same what we had used earlier. But now inorder to add a title on the graph, I amusing here a command main, equal to this and inside this double quotes, I am writing the title what I want. customers attended by sales person and if you try to execute it, you will get here the same graph but there you can observe that you are getting this title, so the moral of the story is this, that in case if you want to add a title on the graph, use the command main.

Refer Slide Time :(37: 45)



And I would try to show you it on the, our console also and you can see here now we have this title added, right?

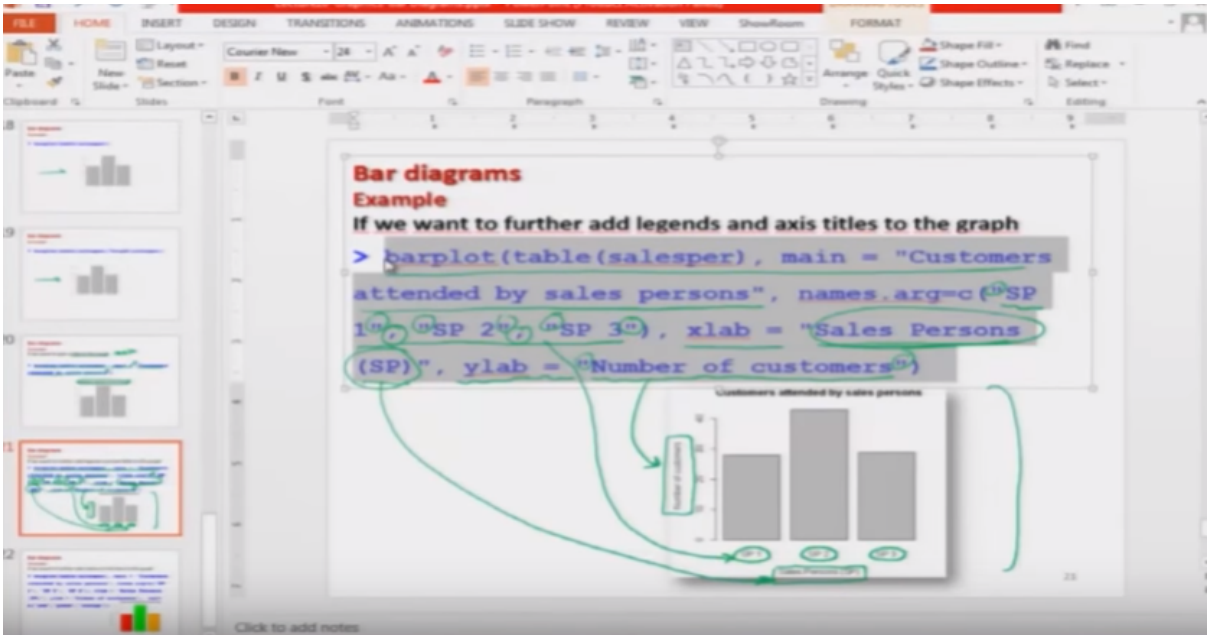
Refer Slide Time :(38: 08)



Now let me try to do some more changes in this graph. suppose I want to add here some legions and title to this bar chart, so first you try to look at this bar diagram, you can see here now I have added here a legion to the bar number one, as SP 1 that is salesperson one, Legion two bar number to, SP 2, salesperson two, Legion to, bar number3 SP 3 and here I have added a title on the x-axis and here I have added a title on the y-axis. So now we need to see that how these things can be done. well, I will tell you the very simple way go to the help menu, try to see what are the different parameters that can do the job and they will also explain you how this values have to be given, some time they are in some number some time they are inside the double course, please try to look from the help menu and try to use it. for example here if you try to see this bar plot table and main command they are from my earlier plot, now I want to add here three names sp1, sp2 and sp3, now this help tells me that please try to give the names of the bar inside the double quotes, whatever you want, so I'm trying to give you give here three names sp1, sp2, sp3, inside the double quotes and which are separated by comma, this is the format and when I want to give a title on the x-axis, I have to use the command X lab equal to and inside the double quotes I have to give whatever I want to write, for example I have given here sales person and inside and the bracket SP and this entire thing I want to print on the x-axis, so you can see here this is coming here and this sp1 sp2, sp3, this is coming here and, similarly if I want to

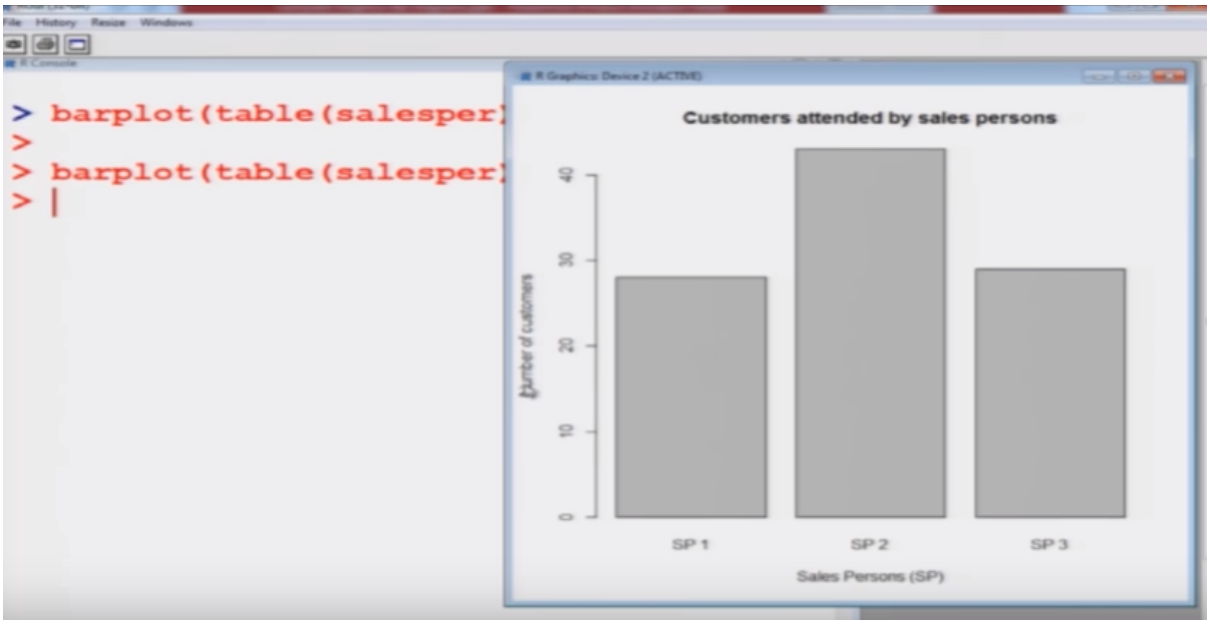
give a title on the y-axis and then the command is Ylab. So I have to write Ylab, equal to and whatever I want to give the title on the y-axis I have to write it inside the double quotes and you can see here I have written number of customers and this is here.

Refer Slide Time :(40: 32)



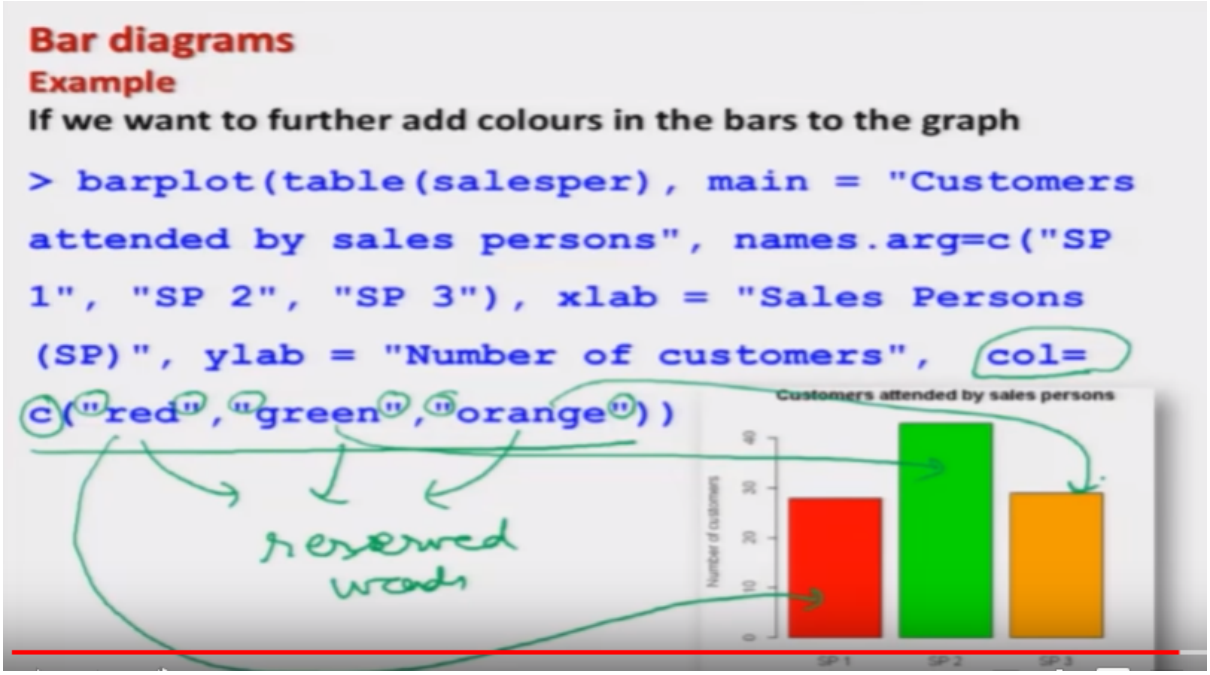
So I can execute this thing on the, our console also and you can see the outcome which you are getting over here, right?

Refer Slide Time :(40: 39)



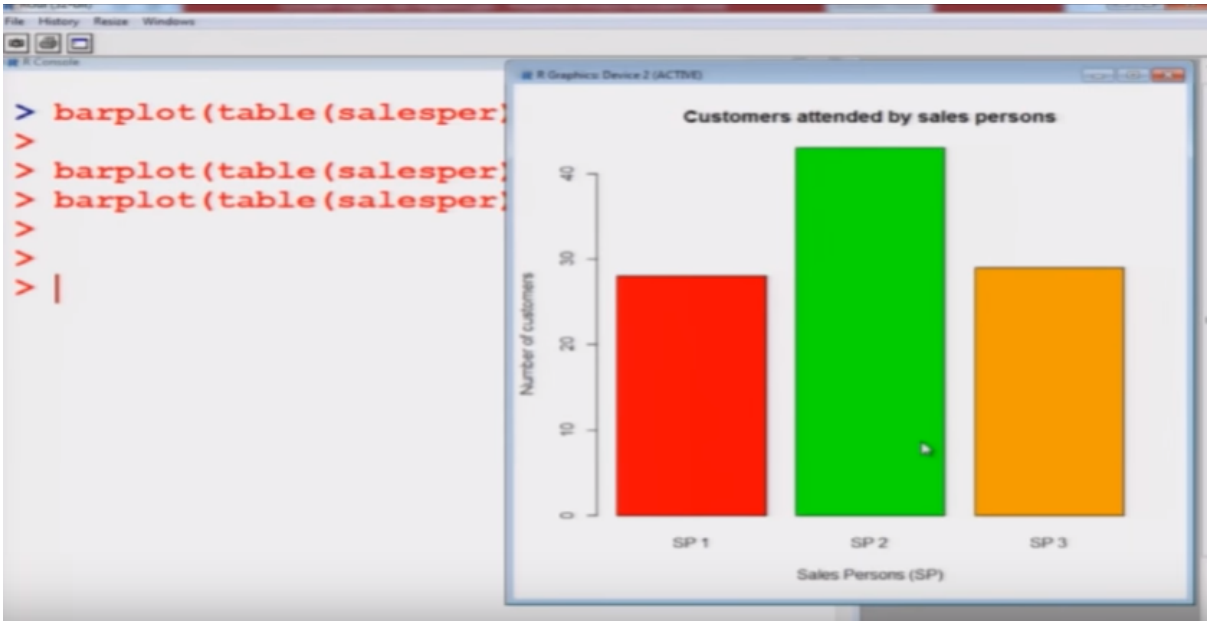
you can see here there is no title no regents over here and now I'm trying to execute it and you can see here that this labels sp1, sp2, sp3 and titles on X and y-axis they are added over here.

Refer Slide Time :(40: 57)



Now suppose I want to add some colours and I want to make this bars of different shades, different colours right? So in order to add the colours I will have to use one more command this is here CoL and rest of the command that is simply the same as in the last slide. now if you try to see in order to and here colours I am giving here CoL equal to then I am trying to give here three colours, red, green and orange and these colours are given inside the double quotes and remember one thing, red, green and orange, they are the reserved words, reserved words means, once you try to write down red, green, orange the R will also understand them red green and orange and there is a list of different types of colours which are understands, so you simply have to give these colours inside the bracket using the C command and the same order will be followed in your bar diagram also, for example here in case of here red, this is red is coming here in the first, second place is green this is coming here green, third place is orange, so orange is coming over here, right?

Refer Slide Time :(42: 26)



So we kind of try to do the same thing on the R console and let us try to see what do we obtain? So you can see here now the, this is the same graphics but now as soon as I execute it these colours are added right? now similarly you can add many more things into the graphic to make it more informative and more useful and there is a long list, but definitely it is not possible for me to take all the issues in this lecture, but I believe that I have given you sufficient background to understand, that it is not difficult at all, to add different types of parameters and it is not difficult at all to create nice graphics, using the R software which is completely free. So you are creating all these graphics just for free. Now only thing is this yes you need to spend some time in learning it, but definitely, in case if you have to spend some time in learning and you are saving a lot of money in from buying an expensive software to create all this graphic, I think is not a bad bargain. So you please try to have a look try to take different datasets and try to create more graphics and in the next turn I will try to take more graphics and I will try to explain you all these features. So you practice, enjoy it and I will see you in the next lecture again.
Till then good bye