Introduction to R Software

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

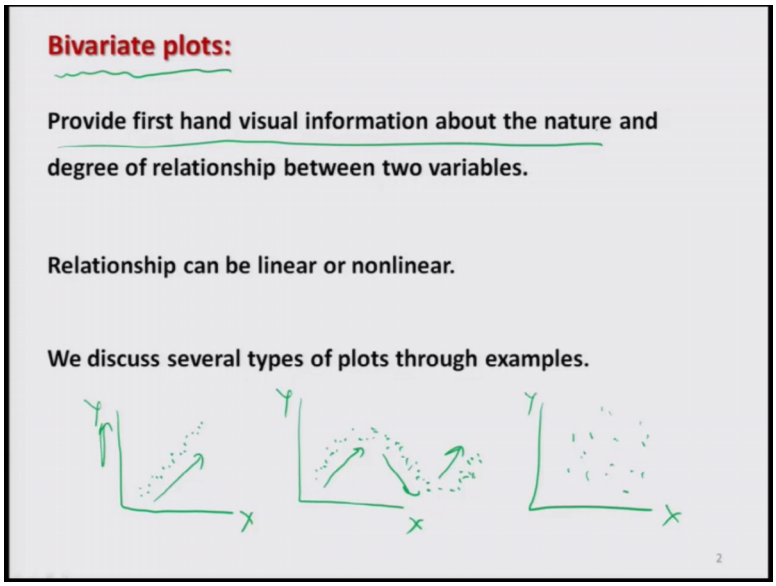
Lecture - 39
Bivariate and Three Dimensional Plots

Welcome to the next lecture in the course Introduction to R Software. In the last couple of lectures, we have considered some statistical function and graphics and usually we have concentrated basically on one variable. That means, we have data only on one variable that is something like height or say weight. Now in this lecture, we are going to consider a situation where we have data on say 2 or 3 variables. And we want to extract the information that is contained inside these data sets.

In order to do it, we have 2 options: first we try to have some information from the graphical procedures and second option is that we try to get this information in some quantified form. For example, we had discussed earlier the interpretation and meaning of a Skewness and courtesies and then we quantify them using the coefficient of a Skewness and coefficient of courtesies. So, similar thing we would like to do with the 2 variates and 3 variates data.

So, in this lecture, we are going to concentrate only on the graphical procedures and we would try to learn how to create various type of graphics when we have data on say 2 or 3 variables and in the next lecture, we will discuss how to quantify them what are the analytical tools which can help us in studying the relationship.

(Refer Slide Time: 02:17)



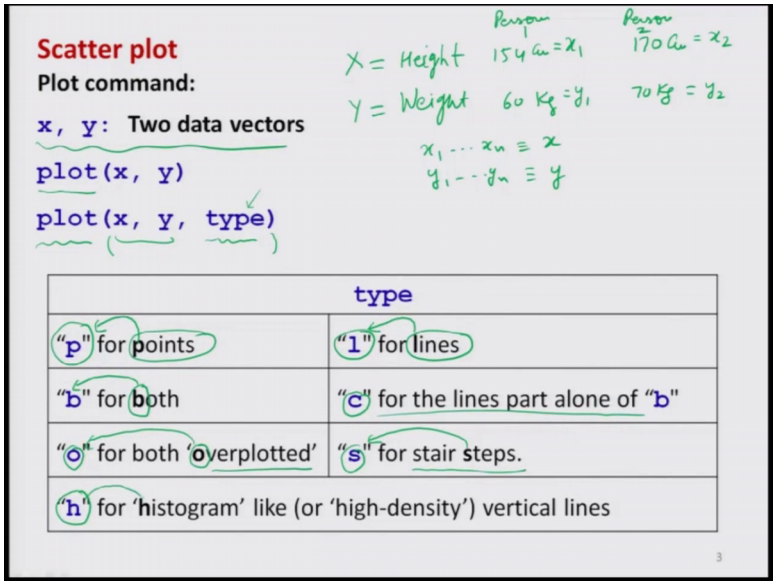
So, we start here in this lecture and we are first going to consider the bivariate plots. Bivariate means there are 2 variables and we want to study what is the type of relationship which they are having what is the mean by this thing for example, if I try to take here say; this a sort of plot or a graphic which is like this and here is my variable x and here y; here I can see that this graph is indicating that the relationship between x and y is increasing that mean as the value of x increases the value of y also increases and similarly, if I say here that if the plot looks like this one. So, I can see here that this relationship is not linear, but it has got some curve which is first increasing then decreasing and then finally, increasing as x increases.

Similarly, if I try to see here this type of relationship between x and y 2 variables I can see that there is no relationship and possibly they are randomly distributed. So, this type of information we can obtain from the bivariate plots, here I am going to consider 2 variables and in this example, I have denoted them by x and y. So, this bivariate plots; they provide the first hand visual information about the nature and degree of relationship between 2 variables what do you mean by degree.

For example if I try to take here say 2 data sets on the same scale say on x and y suppose one is like this and another is like this case on the x axis y axis are the same. So, you can see here that the degree of relationship is here higher the values are very very close to the line, but here the degree of linear relationship is lower than this one. So, we are going to

explore all these aspect and you also have to keep in mind that these relationships can be linear or they can be non-linear also.

(Refer Slide Time: 04:50)



So, we try to take here some example and through that example we will try to see how we can draw different types of curves right one popular simple good command in R to make such bivariate plot is plot; p l o t. And obviously, these are bivariate plots. So, there have to be 2 data vectors what you mean by data vector suppose I take here 2 variables suppose one is here height and another variable is here weight and suppose I take the person number here one; I try to measure its height suppose 154 centimeter and weight suppose 60 kgs and I try to take here person number here 2 and I try to measure his height say 170 centimeter and say weight 70 kg.

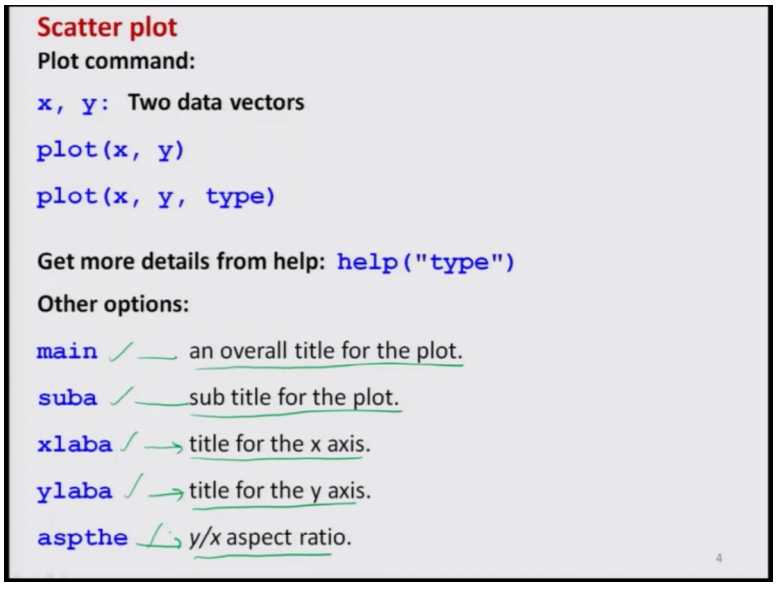
So, this is I can denote this is my here say variable x which is height and similarly, I can denote say another variable weight is denoted by y and this is the first observation. So, this is like here x 1 and this is the first observation on weight. So, I can denoted by here y one and similarly, the height of the second person which is 170 centimeter this is the second observation on the height. So, I can denoted by here x 2 and the weight of the second person which is 70 kg can be denoted as y 2 which is the second observation.

So, based on that we can have here 2 sets of data x 1 x 2 x n and say y 1 y 2 y n and these values are going to be contained inside the data vector x and data vector here y. So, these 2 x and y they are denoting the 2 data vectors. Now I simply have to use here plot x y

and this will give us a plot beside this plot command there are some options which can be given inside the argument and this is called here as a type by controlling the type, I can generate different types of graphics. For example, will I right here p l o t and inside the argument I write the 2 data vectors separated by comma and then for the type I can choose here different types of options; suppose, if I choose say type equal to here p then the graph will be coming in terms of points and if I try to use here see here l type equal to l then the curve will be coming in terms of here lines.

So, you can see here that p is here actually point and this here l is denoting here this l for lines and similarly if I need points and line both, then I have to use here the option both and we use here the option b which is the b of both. And similarly I can use here the option or say type c for the lines that are the part alone of only this here both that is b. And similarly I can use another type here o which is used for both; that means, over plot it. And similarly I can use another type here s this means say stair steps. So, this s is coming over here and here this o is used over here. And similarly I can use a something like a height density vertical lines by using the option here s which is something like for histogram.

(Refer Slide Time: 08:25)

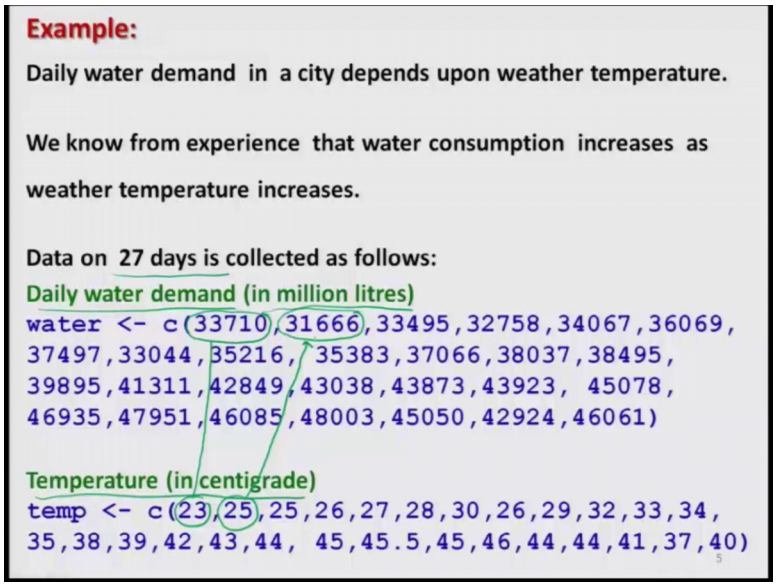


So, one or to take a example and try to explode that what is this type means and how do we obtain this type of plots besides this type command there are some more option. For example, whenever you are trying to make a plot you would try to control the values on

the x axis the label on the x axis similarly the values and labels on the y axis different types of colors. So, there are different options and some of the options I am mentioning here that I can use here the option like here main sub a x lab a y lab a a s p t h e aspthe and these have got different meanings.

For example, when I want to have the title of my choice, then I can use here main, then if I want to have a subtitle for the plot, I can use here the sub a when I want to control the title for the x axis then I can use the option x lab a when I want to control the title on the y axis I can use the option y lab a. And when I want to control the ratio of y and x variables then I can use the command aspthe a s p t h e; right.

(Refer Slide Time: 09:37)



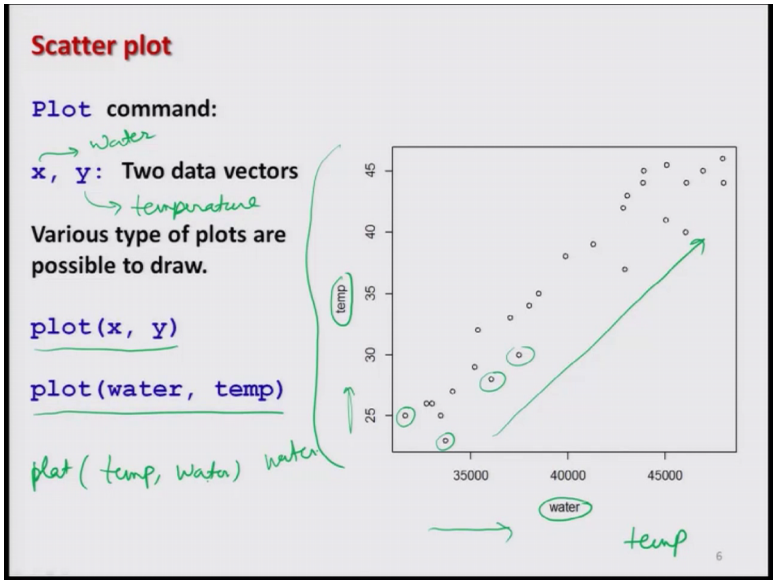
So, why not you take an example and we try to implement all this option and see; what do you really mean by these things. So, in order to understand this plots let me take here an example which has 2 data which are dependent for example, we know that during summer when the temperature is high the demand of water consumption also becomes higher. So, we know by our experience that as the temperature increases that is the weather temperature increases the demand of water also increases.

So, what I have done I have just taken a artificial data and this data is indicating that the demand of water in a particular city is collected on say 27 different days and the weather temperature during day on these days is also collected. So, we have here data on say 27

days in summer and all these 27 values are obtained for say daily water demand that is in million liters and the day temperature in say centigrades.

For example this first value here this is 3370 and the first value and the temperature is 23. So, that is indicating that on the first day the temperature was 23 degree centigrades and on that day the consumption of water was 33710 million liters and similarly, the second value in the water is 31666 and the second value in the temperature is 25 this means on the second day the temperature was 25 degrees and on that day the water consumption was 31666 million liters.

(Refer Slide Time: 11:27)



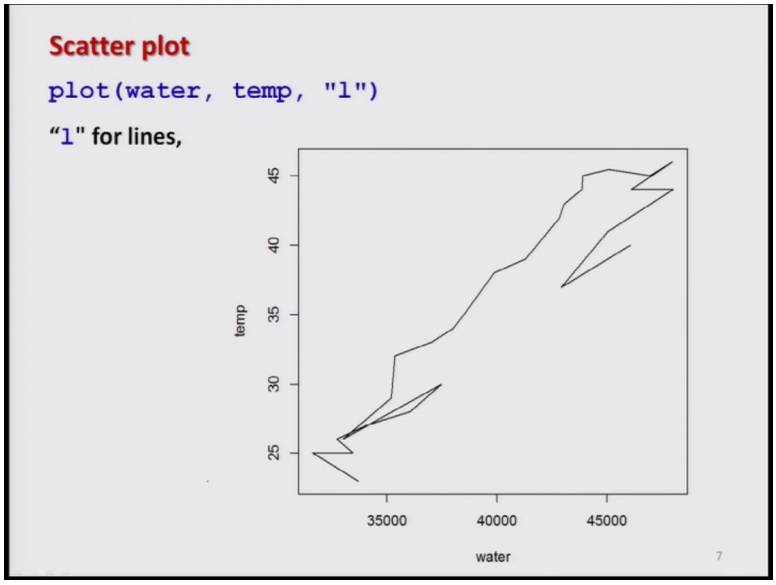
So, we have this type of data for say 27 days, now using this data I try to store them inside the R and I have plot a different types of graphics. So, I am using here say here for the x I am using the water data and for why I am using the data on temperature right these are my 2 data vectors and I use the command here plot x y which is p l o t inside the argument you have to give the 2 variable separated by comma.

So, first I will try to show you all the slides which are the outcomes of different option and then I will try to repeat them on the R console. So, when I try to give the command here plot inside the argument water and temp which are indicating the data values on the water consumption and the temperature of the weather you can see here you are getting here this type of curve here on with; you can see here it is automatically taking the labels as name of the variables water and here temp and here you are trying to get here this

these values as the say small circle very small circle and you can get this type of plot and this plot is indicating that that there is a linear trend in this directions.

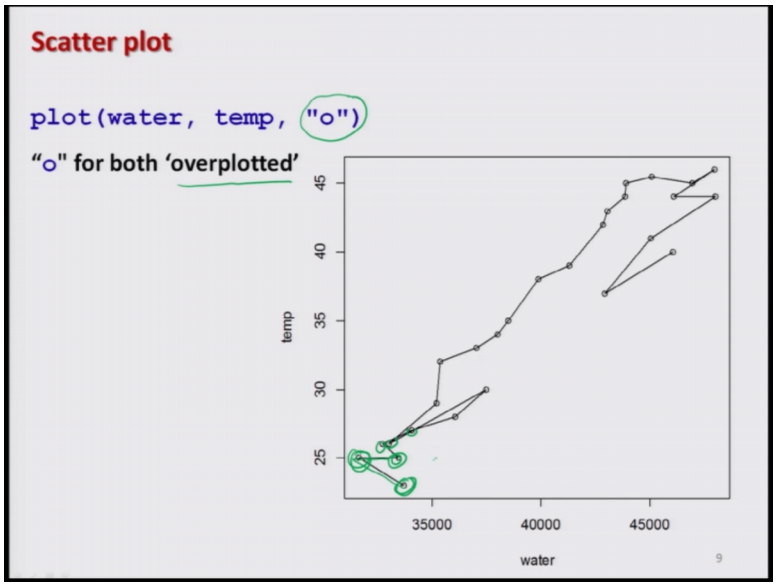
And you can see here that the water consumption is say increasing with respect to here temperature. And if you want to rotate it that you want to bring the temperature on the x axis and you want to bring the water on the y axis; you can plot here plot say here temperature and water whatever you want you can do it. Well, I am not discussing here what is the right or wrong way, I am simply trying to show you how to create a graphic, right.

(Refer Slide Time: 13:18)



Now, in the same data set water and temperature if I use the option here l this l is used for lines you can see here I am getting this type of data set you can see here this data set. And the earlier data set they are the similar data set, but what is happening that in the earlier data set these points have been joined together.

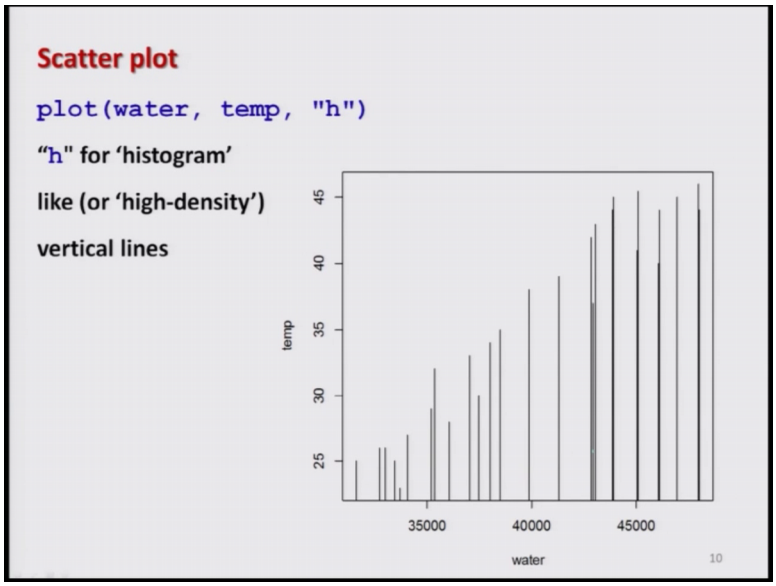
(Refer Slide Time: 13:47)



So, this is a line plot similarly I try to take another option where I want line and point both. So, I use the option here b type is equal to here b and I get here the plot which is indicating here a points as well as here this lines. So, I will get here this type of plot. Similarly, suppose I want to make a plot in which the points are interconnected. So, whatever try to do here that I will use the option o which is mean over plotted and in this case you can see here all these points they are joined together.

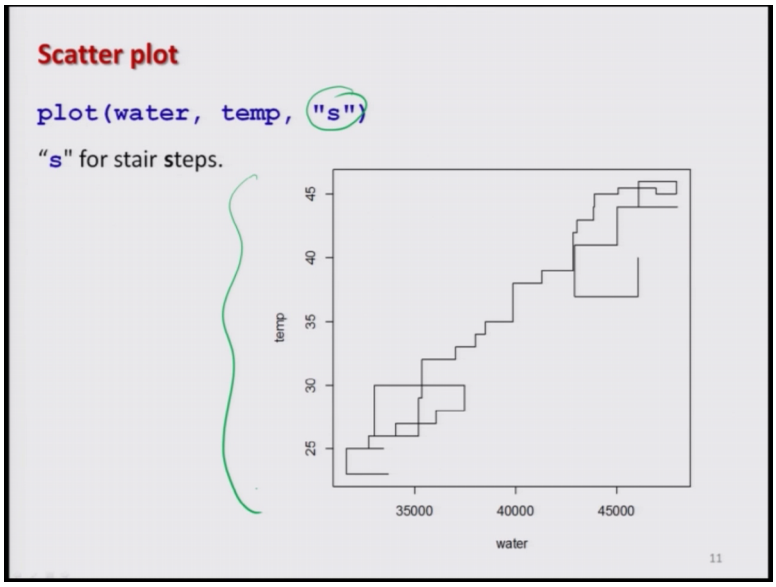
So, now in this plot, we can see that we have here points and the lines are interconnecting them. So, they are over plotting them that is why this is called as over plotted.

(Refer Slide Time: 14:41)



Now in the next thing; suppose, I want to have the graphic in terms of say high density or say histogram. So, I use here option h and I get here this vertical line this is the first observation this is the second observation this is the third observation and so on right.

(Refer Slide Time: 15:03)



And similarly here suppose I want these points as a stair steps. So, I have to use the option here s and as soon as I do; so, I get here; this type of say this curv, right. So, why not to plot it on the R console and try to see whether are we getting the similar outcome or not. So, first I have to enter the data which I already have done so far example here

water is like this and temperature here is like this and now first I would try to plot the same thing.

So, I can copy the same command so that you are confident that is the same thing for example, I can try here with plot water temperature. So, you can see here, here is the curve and I can reduce this window size also. So, that you can see it here more clearly, right and similarly if you want to have here this a plot with here lines. So, you can see here this command how this command is going to change you get here this type of curve.

Now, if I trying to change here for b, you get here this type of curve where we have both lines and points and similarly, if you want to change here the option to be here o that is over plotted you get this type of curve and if you want to change it with say high density histogram line. So, it is like this and similarly if you want to change it with your say stair steps this is something like stair steps type of plot, right.

So, you can see it is not difficult to draw such graphics and now let us come back to our slides. Now after this, we tried to play with this graphics and right tried to take several example where I try to chase the title or some other aspects of the graphic. So, let us try to take the same example; you can see here one thing, before I go further, you can see here that the title on the x axis here is water and the title on the y axis here is temp t e m p and suppose I want to make it more.

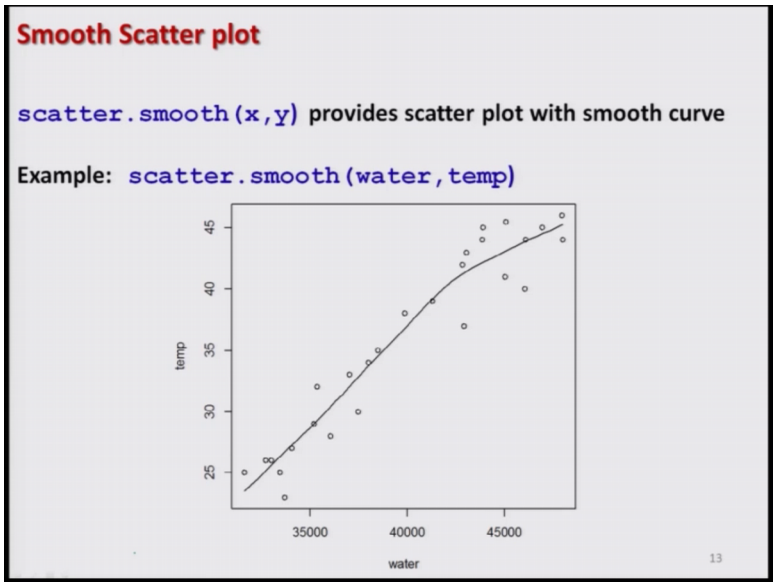
(Refer Slide Time: 17:36)

So, now we do like this I use the same command here plot between water and temperature that is temp, but I am changing my titles on the x axis and in the y axis I want to give here daily water consumption on the x axis and I want to give day temperature on the y axis. So, I try to use here the options x lab and say y lab and whatever title, I want with that has to be given inside the bracket sign this is the only thing which we have to keep in mind and I also want to change the title of the curve if you see in the earlier slide there is no title here nowhere here if you try to see there is no title here, but now I want to have a title of the curve.

So, for that I use the option here main and inside this double quotes, I try to write down that what title; I want daily water consumption versus day temperature, right and when I try to do. So, I get this type of curve you can see here right here I have got the label daily water consumption which is coming from here, then here I am getting here daily temperature which is coming from here y lab this is here and this title that is coming from here.

So, why not to do it on the R console? So, you can see here at this moment we have the earlier curve on the screen. And now I am trying to write down the new command and as soon as you try to say enter you can see here, this things are changing here, you have try to move where my cursor try to look at my cursor this is the title and here on the y axis it is that new title on the x axis there is new title. So, this is how I can play with say these different things.

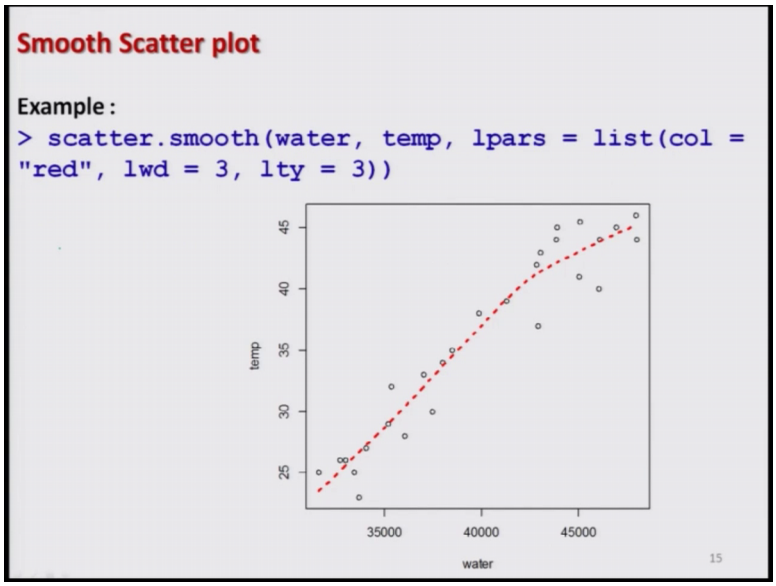
(Refer Slide Time: 19:41)



Now, there is another option which can create different type of a smooth curve and for that we have a command say scatter dot say smooth; that means, a scatter diagram has to be there and we want to join those points with a smooth line. So, this is the function is scattered dot smooth and inside the argument we have to give the data for example, if I try to take the same data on water and temperature. And I try to plot this is scatter smooth you can see here we get this type of curve right here all these points are there, but beside them there is a line which is here which is a smooth line passing through with most of the point or close to that those point. So, why not to try it here on the R console; so, you see here when as soon as I run you get this type of here curve, right.

So, you come back toward slides and try to play with more. So, let us come back towards slide and try to explore some more options in this is scatter is smooth options you see there are several options which are available with this scatter smooth function right means I would say that you please try to go to the help menu and try to look into those details and try to see what more can be done there are different options.

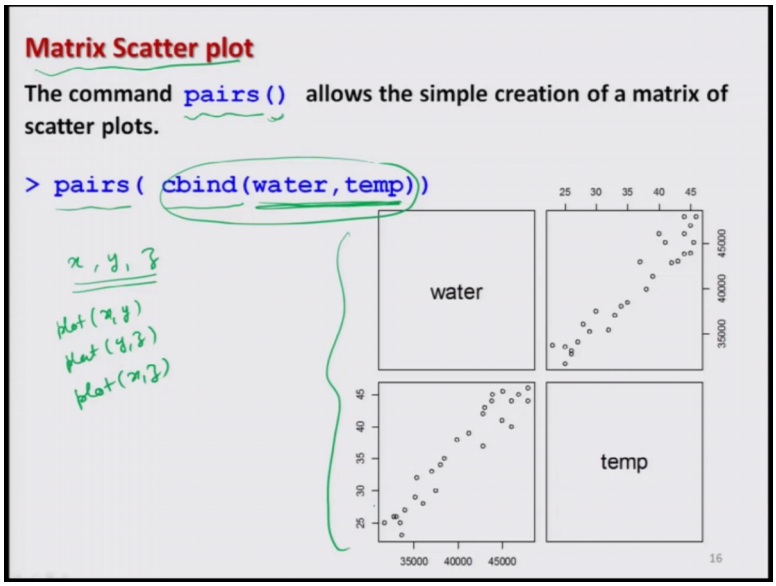
(Refer Slide Time: 21:18)



For example, suppose if you want to change the color or mean their position and everything. For example, here I am using something like I will pass which is equal to list which is column which is giving us the color red and dumb thing lower value and say l w d and l t y. So, what are these things it is difficult for you to remember each and everything, but whenever you are trying to play with the graphics it is important that you try to go into the help menu and trying to see that there are so many options for example, if you can see here there are so many options here say span degree family y lab y lim and there are many things right and depending on your need you try to use them. For example, here I wanted to change the color of the line and I wanted a dotted line. So, you can see here I am able to obtain with these commands. So, if you play more and you experiment more with such options you can learn better than anybody else.

Another option in this plot command is that suppose I want to make more than one plots together. Suppose for example, here I have a 2 variables water and temperature think of a situation where I have got suppose 3 variables say call them as say x, y and z something like water temperature say humidity, right.

(Refer Slide Time: 22:50)



Now, I would like to have a plot in which I can plot all the things together. So, there can be a plot something like plot between x and y plot between y and z and there can be another plot between x and z. So, one option is this I tried to plot them separately or another option is that that I can plot all those plot inside a same graphic that can be achieved by matrix scatter plot and for that we have a command pairs and inside the arguments, we have to specify the data vector, but remember one thing when we are trying to specify the data vector that has to be in the form of a vector.

For example, mans if I want to create a matrix scatter plot of the same data water in temperature then I have to write water and temperature inside argument and then I have to combine them together, the c bind function now becomes a vector quantity and now I can see here pairs inside the argument write down this vector quantity and as soon as I do. So, you can see here I get here a outcome of this type.

Well, here on this screen I am trying to take only 2 variables. So, that you can see clearly what is really happening means if I tried to take a 3 or 4 variables the figure will become very congested and it will be very difficult for me to explain you; what is really happening inside the figure. So, I would request you; you try to take 3 variables for variables and practice with them here I am trying to take this example and I would like to explain you what is the meaning of different graphics you can see here in this graphic we have here four pictures 1, 2, 3 and here 4. So, what is really happening actually in the

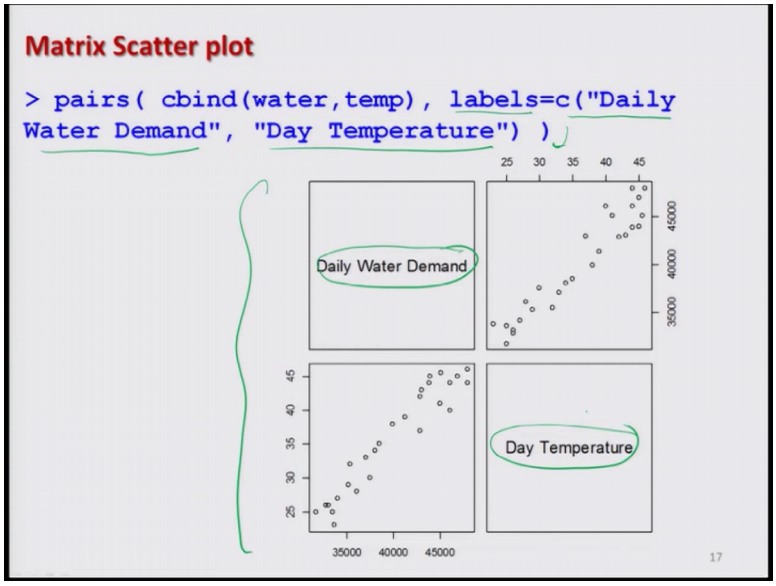
picture number four it is trying to take here water on the x axis and water on the y axis; so which has no meaning.

So, this is simply trying to say here this is water and similarly in the picture number 4; they are trying to take temperature, but more important pictures are on the of diagonal sites which are here and say here. So, you can see here this is a temperature. So, this temperature comes over here and on this axis this is here temperature and here this is here water. So, this water comes over here and this is here water.

So, this is a graphic that is the graphic number 2 is indicating a plot between water and temperature this is a plot between water and temperature now we come to figure number 3, here you can see here first we have to find out what is happening and on the x and y axis. So, to know this thing I will say here this temperature is coming over here. So, on the y axis we have here temperature and this water is coming from here on the x axis. So, on the x axis we have here water.

So, again this is a picture between see here temperature and water. So, means having a picture between water and temperature and temperature and waters with the interpretation remain the same only the graphic is reverted. So, you can see here if you try to compare this figure number 2 and figure number 3 you can see that all the points of are just flipped and this is how we try to make an interpretation in the matrix plot.

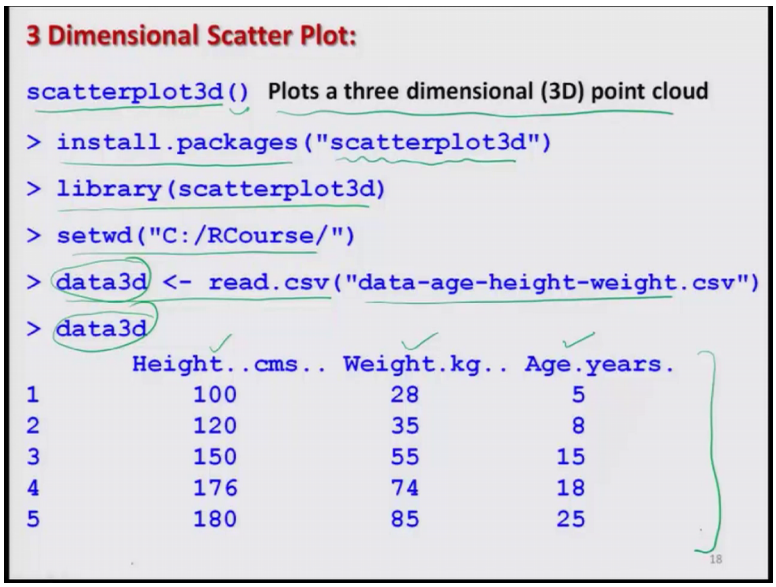
(Refer Slide Time: 26:55)



Now, if you want to make some more changes in the matrix plot you can see here that in this matric plot, we have the title here water and here it is the temperature and suppose, I want to chase them these options are also available. And now I can modify my this pair's function by using the option here labels and inside the combined vector I write down the daily water demand and day temperature which are the desirable labels which I want and if I tried to plot it here, you can get this type of picture which is the same picture, but only this labels are now changed.

So, why not to experiment it on the R console itself and we try to first create a matrix scatter plot you can see here we have no matrix scatter plot now I tried to copy and as soon as I execute it this gives me this type of matrix scatter plot and now when I try to experiment with the second one this is here like this and now in this thing you will see that the names are changed right. So, you can see that it is not really difficult to do all these things.

(Refer Slide Time: 27:57)



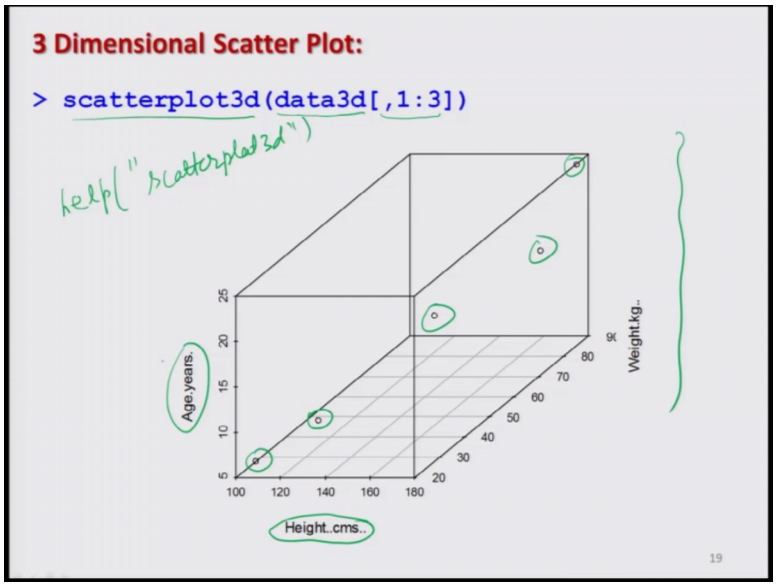
So, here means up to now I have taken the example of 2 dimensional plot. Similarly, we can also create 3 dimensional plot 3 dimensional means there are 3 variables and in order to make 3 dimensional plot in R we have a special function which is called as scatter plot 3 d- s c a t t e r p l o t 3 d and inside the argument we have to give different types of option depending on our need and requirement and this will plot a 3 dimensional point

cloud, but this option is not available in the base package for that you have to download and install a package what is called as say scatter plot 3 d.

So, using the command install dot packages you can install this package, then you can upload this package using the library command, then you need to set your working directory do not forget it and then you have to read the data in order to show you here I have considered a very small data set so that I can illustrate how you can create a 3 dimensional plot. So, I try to take 3 variables here say height weight and age of only five persons and then I have stored that data in say in this file data hyphen age hyphen height hyphen weight dot csv. So, I try to read this csv file by this read dot csv command and I stored them in a data vector 3 d and I try to store this data inside a variable call data 3 d. So, data 3 d looks like this over here.

Well, now I will not be showing you this thing on the R console, but I will request you that you please try yourself then you will understand it more, I am simply giving you ahead the commands the screenshots or the graphics which will come.

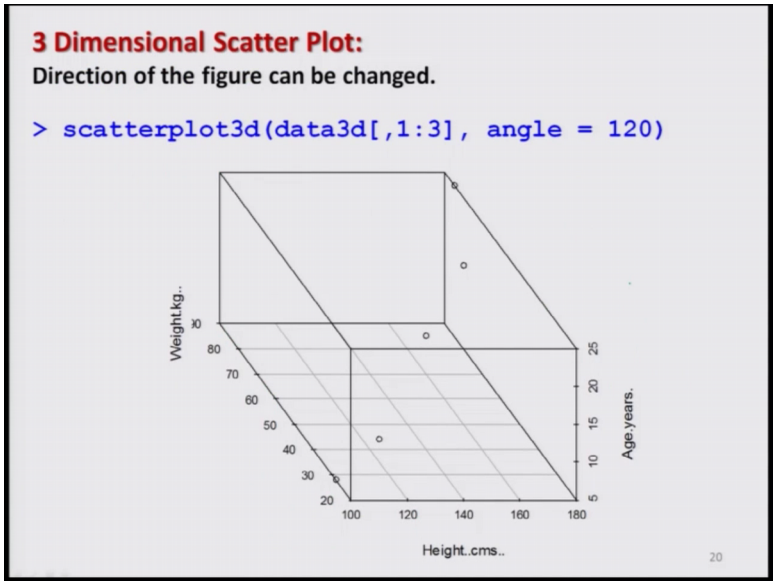
(Refer Slide Time: 30:12)



So, now in case if I try to create this scatter plot 3 d, then the variable name here is data 3 d, but here you can see one special thing I have to write down the data in this format comma and one colon 3 that from one to third column there are various options which are available with this scatter plot 3 d and it is really very difficult for me to explain each and everything here. So, I would request you that you please take the help using the help

scatter plot 3 d and then try to look on the help menu, but my objective here is to show you that these things are possible.

(Refer Slide Time: 31:19)



So, if you try to do. So, here you can see here I get this type of 3 dimensional picture where these are the points which are indicating the five values and here on say x axis, we have a height in centimeters in y axis we have your age and on the z axis we have here weight and similarly in case if you want to change the angle of this 3 dimensional picture that is also possible you simply have to add one more option here; angle is equal 120 and the angle of the figure will be rotated by 120 degrees and similarly, there are several other options which I would like you to explode depending on your need whenever you need them try to understand them and try to implement them.

Some more functions say for this 3 dimensional plot are say contour lines dot charts images mosaic plot perspective plot. So, here I am simply trying to give you here some examples of those functions and whenever you need them you can just use them. Now in the next few slides, I simply want to show you that what other types of graphics are possible, I will not create them here, I will not discuss them here, but I am I simply want to inform you that these things are possible and whenever you want to create them please try to learn and do it. I already have given you the basics that how to understand and how to create graphics.

For example, if you want to create here multiple bar plots that is possible like this one right if you want to create here group box plots; that means, more than one box plots inside the same plot means earlier in our example I had created 2 boxplot separately, but here all the box plots are going to be computed simultaneously, right.

Now, I would like to stop here and I have tried my best to give you a flavor for 2 dimensional plot and 3 dimensional plot now the success depends on you that unless and until you take some data sets and you try to execute this commands you itself; it will be difficult to learn the more you practice more you learn.

So, you try to take some data set and practice them. And we will see you in the next lecture where we are going to discuss that how to quantify the information on bivariate data using the concept of say covariance and correlation.

So, see you in the next lecture, till then goodbye.