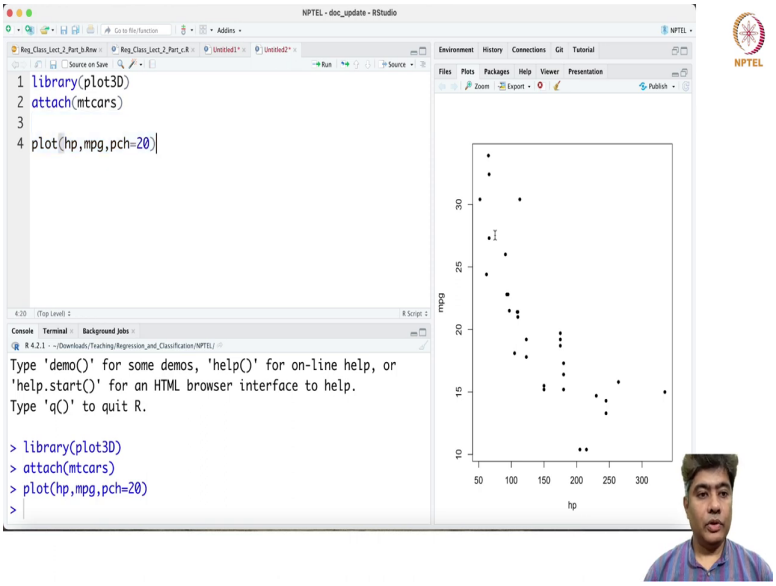
Predictive Analytics - Regression and Classification
Prof. Sourish Das
Department of Mathematics
Chennai Mathematical Institute

Lecture - 09
Hand-on with R Part -3

Welcome back to the lecture 2 of Regression Classification course, I am going to discuss now or I am going to demonstrate how three dimension histogram looks like.

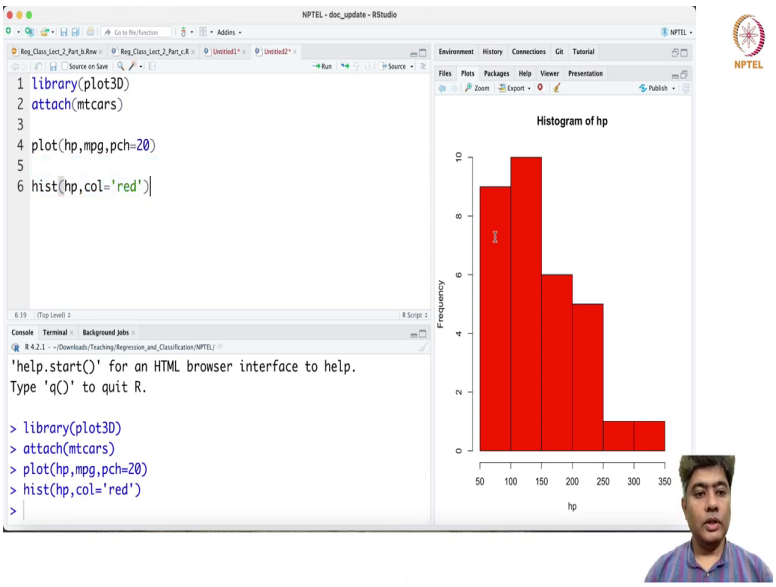
(Refer Slide Time: 00:30)



So, we are going to consider the 3D plot, plot3D and then we are going to consider the empty curves data set and now if you draw the scatter plots of say horsepower versus miles per gallon and pch equal to 20. So, on the x axis we are drawing the horsepower, on the y axis we are drawing the miles per gallon.

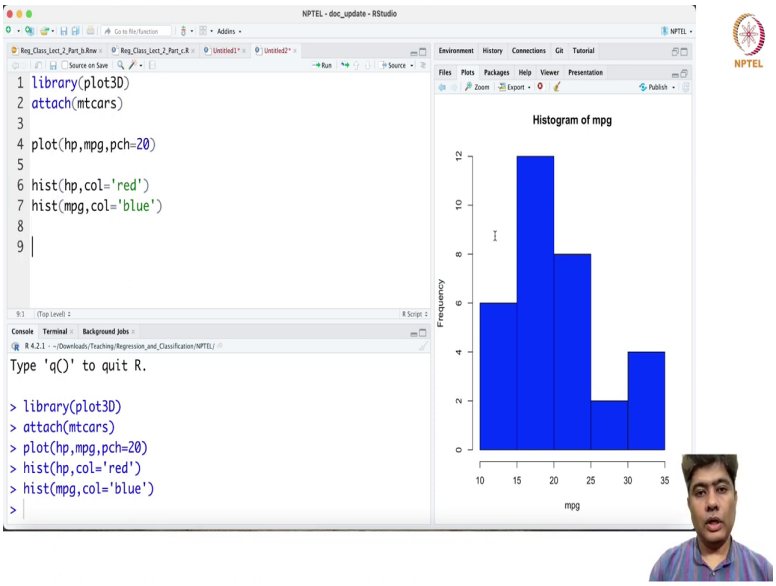
And we would like to draw histogram for horsepower we had like to draw histogram for the miles per gallon. And then we want to draw a joint histogram which will be a 3D histogram for horsepower and miles per gallon together.

(Refer Slide Time: 01:49)



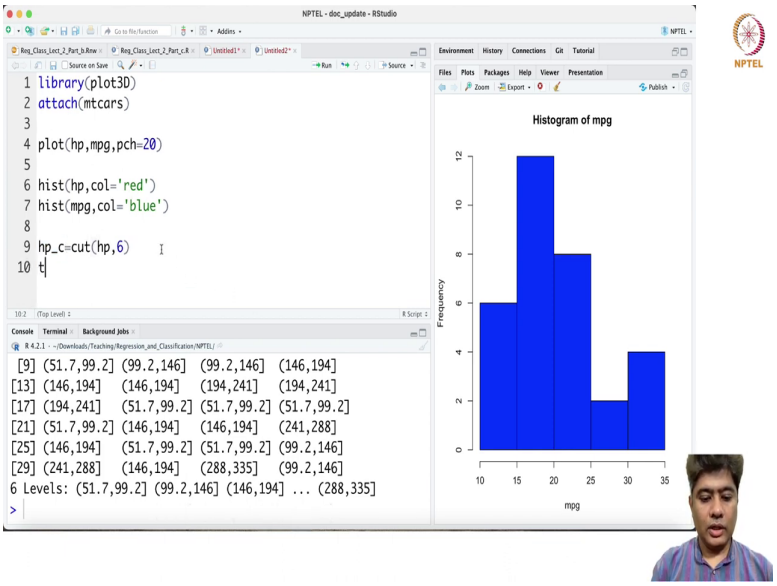
So, drawing a histogram for horsepower is pretty straightforward, if you just say hist; so, it will draw a histogram for the horsepower.

(Refer Slide Time: 02:08)



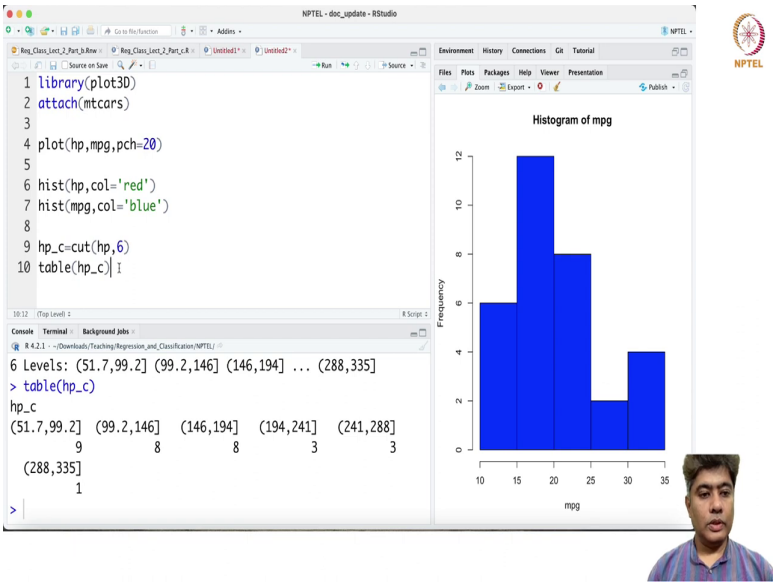
Similarly, you can draw a histogram for miles per gallon you can choose maybe a slightly different color maybe blue and so, this is miles per gallon. Now, what I am going to do I am going to first draw a joint table.

(Refer Slide Time: 02:37)



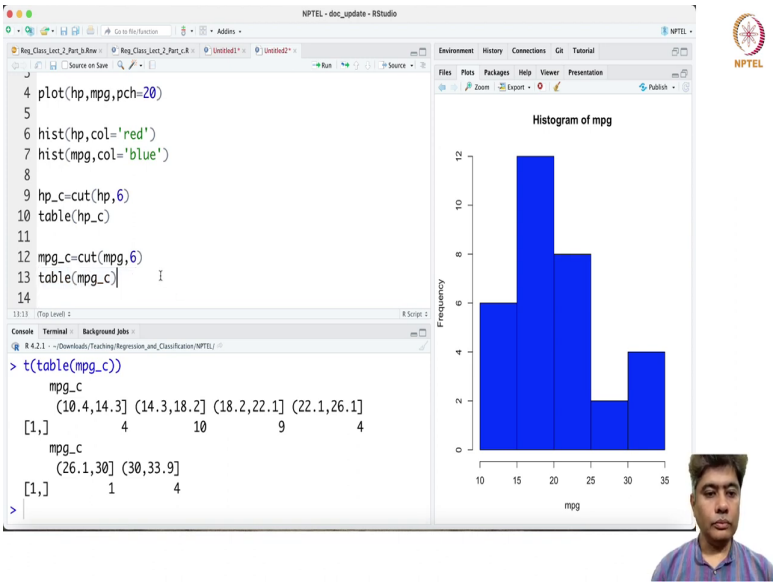
So, first you have to do a cut say cut hp it maybe 6; so, now, it will give me all the variables values which are between these different cuts.

(Refer Slide Time: 03:00)



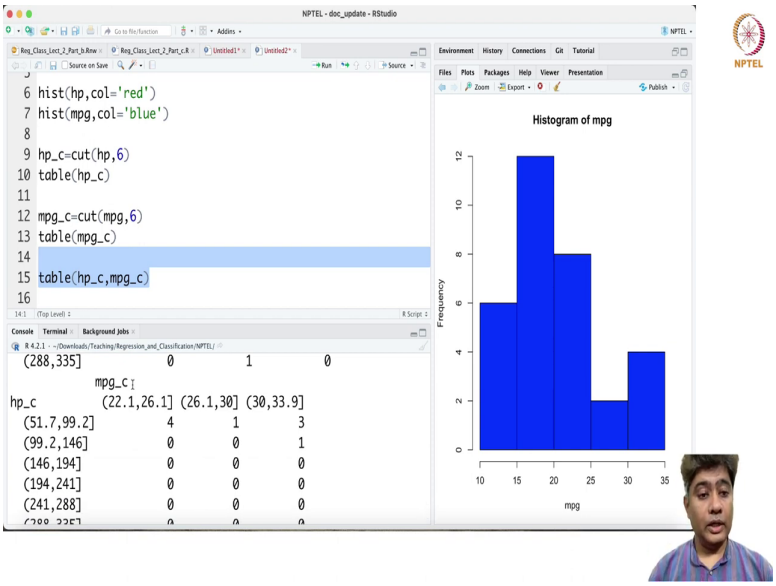
And if I just do a table of hp c, this will give me the frequency distribution of horsepower between the variables.

(Refer Slide Time: 03:21)



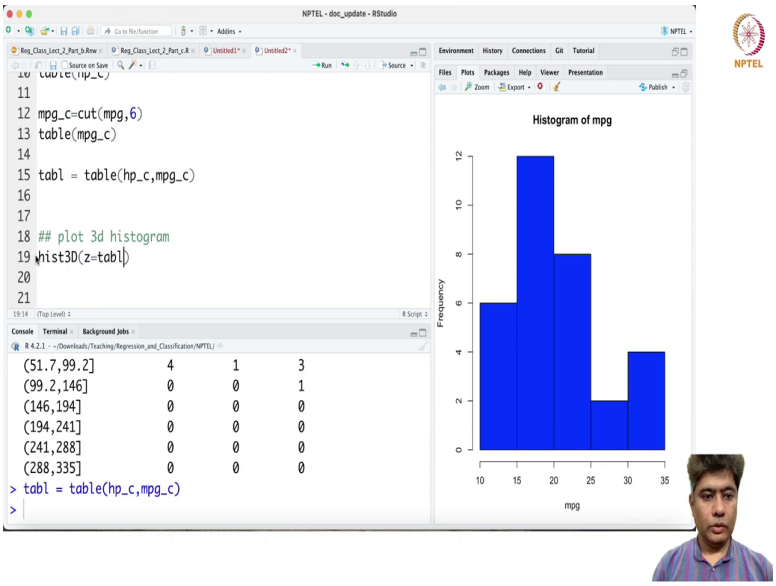
Now, similarly we can do a miles per gallon cut c u t, mpg again with 6 and again if we do a label of cut. So, this will give us our my frequency table of miles per gallon; so, if you just transpose it.

(Refer Slide Time: 03:58)



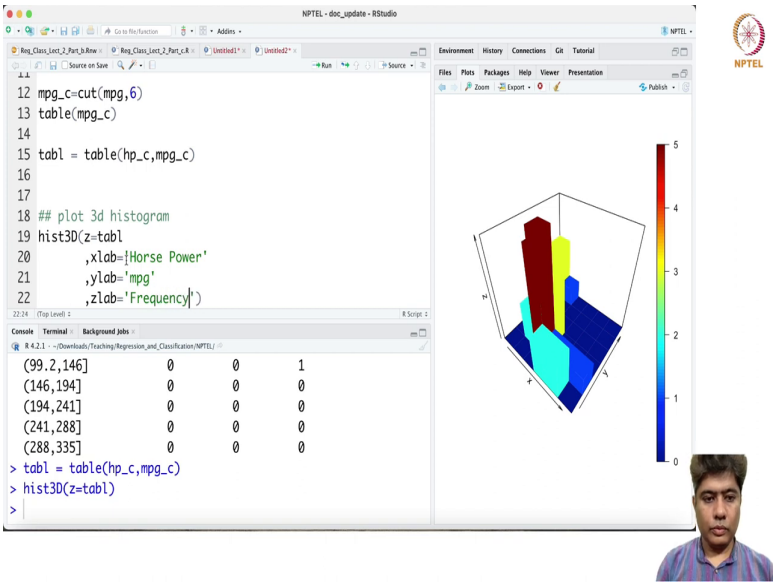
Now, we can joint create a joint table of horsepower versus miles per gallon. So, this gives us a joint table of horsepower versus miles per gallon.

(Refer Slide Time: 04:20)



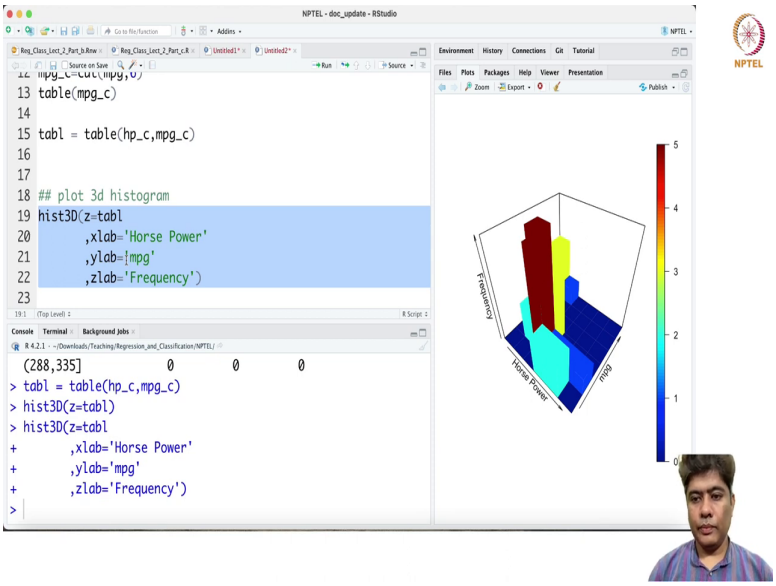
And let us call the tabl plot 3d histogram no plot 3d histogram is hist3D z equal to we have to just select table features.

(Refer Slide Time: 05:23)

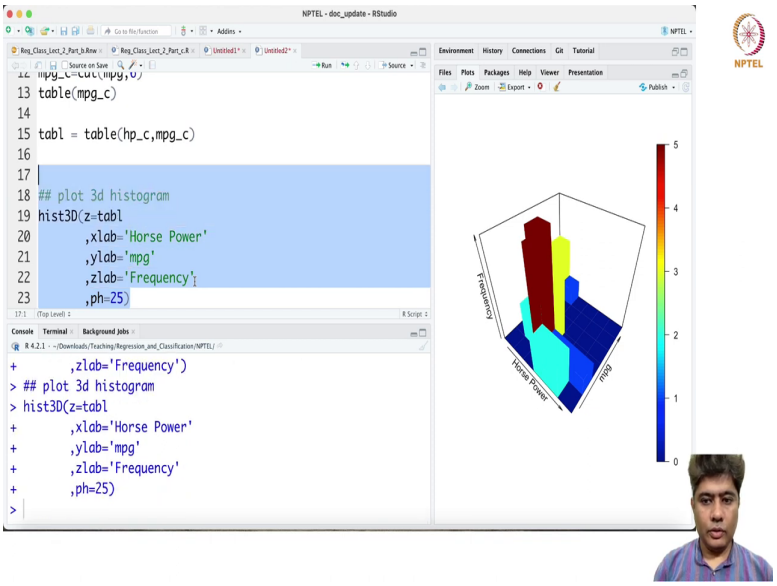


So, this is our 3D histogram, but we have to say which is what; so, we have to say xlab equal to Horse Power ylab is miles per gallon and zlab is equal to frequency ok.

(Refer Slide Time: 06:07)

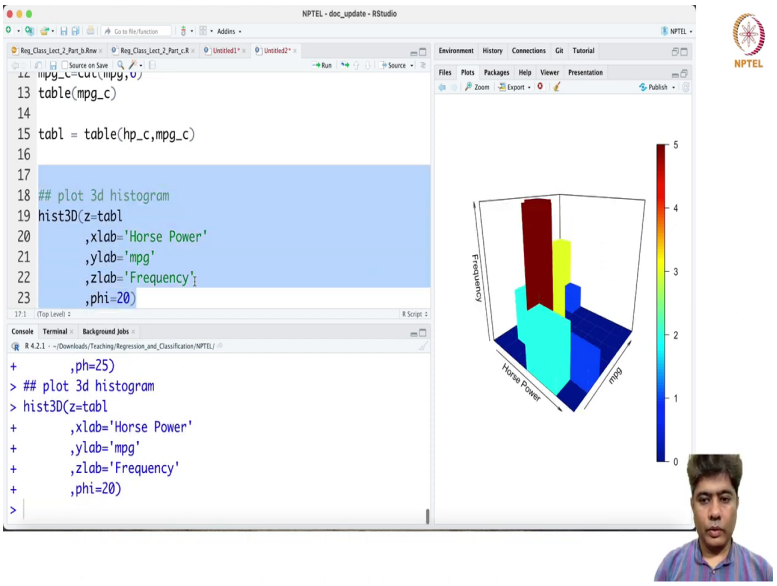


(Refer Slide Time: 06:13)



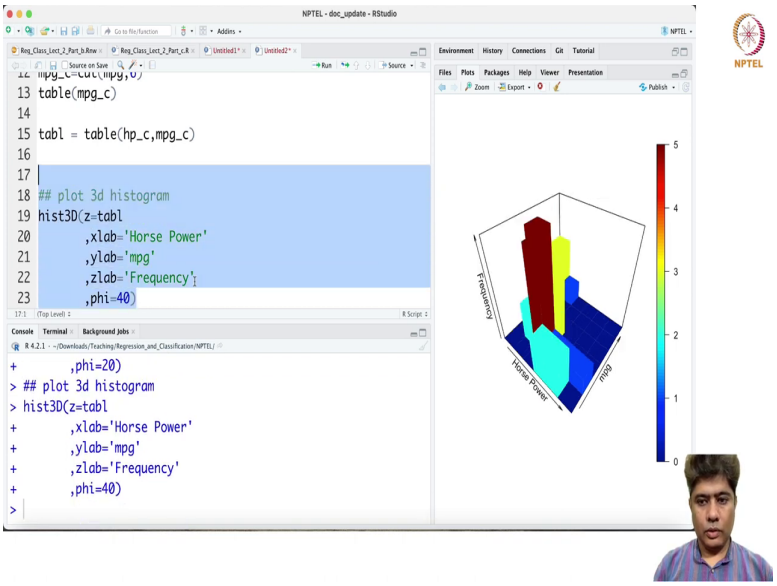
So, then you can choose ph to be 25 maybe then it will create a bit of a 5.

(Refer Slide Time: 06:24)



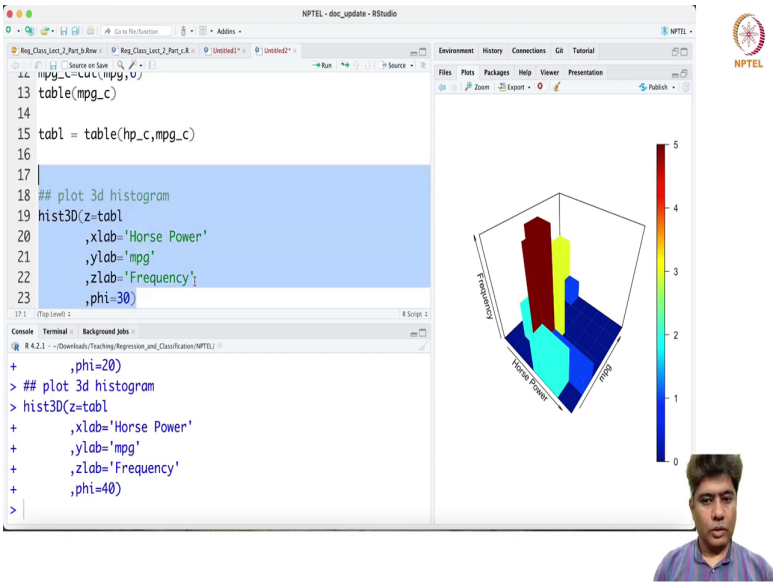
Or sorry ph equal to 25 20.

(Refer Slide Time: 06:34)



And if you choose maybe 40 30.

(Refer Slide Time: 06:41)



So, this gives you a idea about the joint distribution of miles per gallon and horsepower. So, that is how we are going to draw the you get a sense of what is the joint distribution of two variables. And now we are going back to theory about how to what is the regression line from a joint distribution.