

Multivariate Procedures with R

Prof. Shalabh

Department of Mathematics and Statistics

IIT Kanpur

Week – 05

Lecture – 19

Three Dimensional Plots, Star Plots and Chernoff Faces

Hello friend, welcome to the course Multivariate Procedure with R. So, you can recall that in the last couple of lectures we have considered the univariate and bivariate graphics which are essentially trying to handle only two variables at a time. And in case if you want to consider more than two variables then one option was that try to consider them pair wise that means two variables at a time. And this is how you can do it, but there are options where you can consider the graphics in three dimensional or you can consider a greater number of variables without increasing the dimension of the graphic. So, in this lecture we are going to consider such topics. For example, we will be talking about the three-dimensional plot and then I will try to explain you how you can use it for drawing different types of statistical inferences.

I will try to show you about the star plots and turn off faces. And I will try to show you that star plots and turn off faces they can take care of more than three variables also, but in a two dimension. So, let us begin our lecture and try to understand these things.

So, now if you try to see what is the three dimension? Three dimension means it is like this x, y and here z. Our eyes can see in only in three dimensions not more than three dimensions we can see, but on our slides, we can make only the two dimensions, but through some way I can create here a graphic which may look like a three dimension. For example, if you try to see this is rectangle or square. Now, if I want to make it into three dimension you can see here is still, I am in the two dimensions, but now if you try to look at this graphic it looks like as if you are viewing in the three dimensions. So, we are going to consider this type of here using the command `scatter plot 3D`.

And if you try to see here how I have written here this is scatter plot 3D. It is `scatter plot 3d`. In order to create such a three-dimensional scatter plot we need to use a special package which is a scatter plot 3D. So, you know that if you want to install a package you can use the command here `install.packages` and within the double quotes inside the

parenthesis you can write down `scatterplot3d` and it will be installed on your computer. After that you have to upload it.

So, you will write down here `library(scatterplot3d)`. And now I want to illustrate here the creation of picture in three dimensions. So, for that I try to create here a data sets on say three variables and I have artificially chosen them such that they are interrelated. For example, if I say age, weight and height of a person they are not independent. As the age increases the height and weight also increases up to certain age.

So, this type of data I have considered here and I want to create here a three-dimensional scatter plot so that I can see that how these three variables are behaving together. So, in order to create the three-dimensional scatter plot the first requirement is to express the data in a combined way and for that we use here the concept of data frame. So, we create here a data frame by using the command `data.frame` and inside the parenthesis the three-variable height, weight and age and I try to store it in a variable name `data3D`. So, if you try to see here this will look like this. So, now this is my here data.

So, now I use here the command here `scatterplot3d` and inside the parenthesis I try to use this data frame name `data3D` and inside the brackets I try to write here comma and `1:3` which is actually 1, 2, 3. And if you see you will get here this type of graphic. We can see here height is on this axis, age is here, weight is here. So, there are basically here three dimensional 1, 2 and here 3. And if you try to see if you visually try to inspect it these are the points which are here.

So, you can see that how these points are located in a three-dimensional plane. And similarly, if you try to rotate the angle. So, for that if I try to add here an option here angle is equal to 120. So, this graphic is rotated by 120 degrees. So, you can see here now this points are lying here like this. Now the next question come how you can modify it or how you can think that this will do a good job. Here I am trying to give you one idea. Now you can see here using this option `angle` the three-dimensional picture is rotated. Now in case if you try to make here a loop and try to rotate this picture by changing the angle. So, now this picture can be rotated in different directions etcetera.

And when this picture is revolving then possibly you try to see that how these points which are lying inside the three-dimensional picture are changing. And then you can see that whether they are in a plane look like a straight line or they are in some non-linear curve etcetera etcetera. So, by experimenting on this type of a three-dimensional plot you can get different type of information in different ways, but definitely it depends on your imaginative power that how would you like to think to present the same picture in a different way and how you can retrieve different type of information from here. So, now let me try to show you this thing on the R software. So, that you get here more confident.

So, if I you try to see here that I have to first load here the library and I already have copied this data height weight and age and now I have to create here a data frame right. So, if you try to create it here it will look like data 3D right. And now in case if you try to make this command well, I can share with you that I already had installed this package scatter plot 3D on my computer that is why it is giving you here only I am using the command library right. So, you can see here this type of command is coming here. Now in case if you want to change it then it will be like here this.

So, now the angle is rotated by 120 degrees. So, now if you try to rotate this angle possibly this if you try to for example, if you try to make it say here 180 degree you can see here this is changed if you want to make it here say 280 degree it is here change if you want to make it here suppose here 320 degree it is here change. So, now if you try to make it inside a loop then possibly you will get a you can rotate it in the desired direction and can get different type of information right. So, let me close this thing clear the screen for our next command and we come back to our slides ok. So, now the next plot which we are going to consider is called as a star plots have you seen the stars like this one.

So, this type of picture will come that is why it is called as star plots and star plots are very useful when we want to display the multivariate data. What is the difference between multivariate data and multiple data? Suppose if I say here, I have the data on say here height, weight, age on say person number 1. So, if you try to see now, I am trying to write down here the height, weight and age on person number 1. Now, I try to choose another person here person number here 2, I try to write down the height, weight and age of person number 2. So, this is my here data first data this is my here second data right.

So, this is actually multivariate data that means, you are trying to write down here variable like as here the height, weight and here age. So, now this is a 3 cross 1 vector ok. We are going to talk about this type of multivariate random variable soon right. So, this is the type of data that on that one single data set consist of more than one variable and when you try to plot all of them together then only the data set is complete. So, in star plots each star represents a single observation.

For example, if I see here for person number 1 there will be 1 star and person number 2 there will be 1 more star like this. And this type of plots is used to examine the relative values of a single data point and to locate similar point or say dissimilar points. You can think that when you have multivariate data the nature of the data depends on different types of values. And when you try to inspect it, if you try to think about different values, it is very difficult to conclude what we really want to do. And we always want to see ok what are the points which are similar, which are the point which are dissimilar and then we would like to see that what is the relative values of a single data point that how height, weight and age are placed for person number 1 in comparison to the person number 2.

So, these types of queries can be solved from the star plots. So, these star plots are generated in a multi plot format and if you have more than one observation. So, that many stars on are plotted on each page and each star represents one observation. So, if you have 20 observations there will be 20-star plots on a one page right. And the construct of star plot is that a star plot has got a radius and this radius is actually a sequence of equiangular spokes you know.

For example, you have seen the bicycle, bicycle has a wheel which has spokes right. If you try to see the bicycle look like this means like this and you have here spokes to maintain the wheel. So, and each of this spoke represent one of the variables. And this spoke mean that the data values for each of the spoke they are connected by a line. And when you try to for example, if you try to join these lines these are the spokes.

If you want to join these lines like this one like this one then it looks like a star. That is why it is a starlike structure and this is called as a star plots. The data length of a spoke is proportional to the magnitude of the variable for the data point relative to the maximum magnitude of the variable across all data points. So, that is why by looking at the size of the star plots you can get an idea about the magnitude also. So, in R software the star plots can be created by using the command `stars` and inside the parenthesis you have to give the data.

The data can be here in the matrix format or data frame and each row of that matrix or the data frame will be will create one plot right. So, each row of this data matrix or data frame of `X` will create one star plot. And after that there are many options here say `full` is equal to 2, `scale` is equal to 2, `radius`, `labels`, `locations`, `X limit`, `Y limit` etc. So, once again I would request you that you please try to look into your help menu and try to see how you can do it. So, now let me try to take an example to explain you here how the star plots can be created.

I am considering here the same example which we considered earlier and these variables are height, weight and age. The reason being I want to create here the plots different plots on the same data set so that you can compare how are you going to get different types of information from different types of plots. So, I try to create here the data frame of this height, weight and age which is here like this `data3d`. Now, this is here the data set right. So, now if you try to see here, I use here the command here `stars data3d` and you can see here you are getting here this type of graphic right.

So, each of the spoke that is trying to indicate one variable. For example, this is one spoke, this is one spoke, this is one spoke and after that all such points they are joined by these lines and that is why they are trying to create a star like structure right. So, this is how the stars will look like and suppose now if I try to add here you can see here there are here 3 lines 1, 2 and 3 and suppose here my data becomes here 4 dimensional.

Suppose I try to add here one more variable here income. Yes, a hypothetical data so please consider in that way.

Now, I try to create modify my here data frame by adding here height, weight, age and income which is now here data3d1 which looks here like this. Now, in case if you try to plot make the star plots this will be your here stars data3d1 and it will look like here this. So, you can see here that this and this picture they are different. Now, you have here 4 spokes 1, 2, 3 and here 4 and these points they are joined here by these lines. So, now if you similarly start increasing the dimension of your data then you will get different type of structure but they will look more or less like the stars.

So, let me try to create here this picture for you. So, if you see I already had created this data frame on my R computer. So, you can see here data3d right and now I try to create here my stars you can see here these are here and now in case if I try to add here one more variable see here income and I try to create here this new data vector here. So, it will be modified here like this you can see here. And now if I try to create here the stars with data3d1 it will look like this.

So, you can see here it is not a very difficult thing to create these star plots, but surely interpreting it in the right way will require some experience and practice. Because you see as soon as you are trying to change the dimension of the data from 3 to 4 the structure is completely changed. And now you have to see that how are you going to interpret this structure and how are you going to understand that for example, if I try to look into this graphic that how you can how are you going to understand by looking at this picture that what is the interpretation and how it is going to interpret the data right. So, this requires some experience ok. So, now we try to understand one more concept which is very interesting.

First, we try to understand what are we trying to do. Well in case if I try to create here two faces suppose here like this and here like this you can see here, they are the same. But now if I make here two two you know minds here one is here like this and one here is like here this. What information do you get? If you try to see eyes, nose etcetera they are the same. And similarly, if I try to create here the same face which is here like this and like this and like this you get here a different type of information here.

So, now do you think that these three faces they are trying to give you a different type of information? Answer is yes. And how they are differing? Just by using different types of structures of eyes, nose, mouth etcetera you can create different types of expression. If you try to see this face this looks like a happy face and this look like a sad face and this possibly looks like a happiest face. So, if you try to see if your data is good as per your definition well in control and if the same data set can be expressed in a face structure like this one. And in case if the data is bad it is not fulfilling your condition then if the data

can be represented in this type of face which is showing a sad face then do you think that just by looking at the faces you can judge whether the data is good or bad or even what data is trying to inform us.

And in face we have different types of these parameters which can be controlled by eyes, nose, mouth, ears for example, if I make it ears like this one or with this one. So, they will give you different types of information and the beauty is that the way God has created us by looking at the face we have the capability or say inherent capability to just many things. You have seen that in case you can very easily find whether your friend is happy or not today. You can see in a look at the faces and can decide whether the person is crying or person is laughing right. So, this was the idea which was thought to use to represent the multivariate data.

Multivariate data means each of the facial expression can control one variable and now face has so many variables. So, at least some finite number of variables can be controlled just by plotting this two-dimensional plot which is called as Chernoff faces. Why this is called like this? Because actually these faces were devised or proposed by professor Hermann Chernoff and he proposed to use this cartoon like faces to represent the points in the K dimension. Definitely K is going to be some finite value. And he noted that the human mind's ability to decode hosts of meaning from people's minutes expression.

So, he thought to leverage it to encode data into cartoon faces for statistical visualization. It is correct. Means if you try to see that human being can make various types of faces and by looking at them, we can immediately judge what the person is trying to think, happy, sad, not so happy etcetera etcetera. And these expressions are given only by the facial features like eyes, nose, eyebrows, ears etcetera. So, if I try to map the facial features to multiple variables with size, orientation, shape, color, placement, possibly they can potentially represent the different attributes of a single observation.

For example, if I take the example of health, the health depends on several factors, height, weight, age, blood pressure, blood sugar, different things. So, now if a person is healthy do you think that the face of the person should look like a happy face and if the person is not healthy then the face of the person should look like a sad face. So, this is how we can create the multivariate graphics. So, the main use of this face graph is to enhance the user ability to detect similarities between different items. Why? We sometime make a statement ok, his face is very similar to someone.

For example, his father's face or mother's face. How do you judge it? You just try to look at the two faces and you immediately take a call that whether the faces are similar or not. So, this is our ability. God has given us this ability even the person is not so qualified irrespective of the education one can very easily identify the faces and can judge what person is trying to think and how the person is feeling. So, this concept is used to detect

the similarity or dissimilarity among different data points in a multivariate setup. Another advantage of faces is that ease of monitoring the sensitivity of variables to each other, fast identification of key differentiating dimensions and to detect the longitudinal trends.

Definitely I am not going to explain to you each and everything here, but I will try to show you here how you can create such faces. But after that if you try to create different types of data set and try to create different types of faces, you can very easily obtain them. Well, means you know that when we are going into the multivariate direction the dimensions are increasing. So, it is the level of difficulty is going to be more. So, implementing the visualizations of this 10 of faces is challenging.

The data must be normalized and often binned in order to convey meaning. How to normalize the data is an issue is a topic which we are going to learn in the forthcoming lectures. And sometime if you see that the data that there are some data set which are very very close to each other then we try to combine them in the single bin. But how to do it that is also a challenge, but with an experience we can always learn it.

Sometime the scaling and normalization can be complicated. For example, if there are different variables of different kinds of data, some categorical variables, some continuous variables, some discrete variable. But anyway when we are trying to move into the multivariate direction we should be mentally prepared to handle such complications and such challenges. And I believe that as a human being if we cannot do it then who else will do it. So, now we learn how to create such faces in the R software.

For that we need a special package here `apl` pack. For that you can use the command here `install.packages` and it will be installed and then you have to load it using the library command. And after that the command is `faces` `faces` and inside the parenthesis you have to give the data. So, if you try to go into the help menu then you will find here different types of options which are available here and so like a scale, by row, mean, color eyes and then means color hair, color face etc. So, I am not going into these details, but I would request you all to look into the help menu and try to understand it.

And this data what you are going to enter here it has to be in the format of say matrix or data frame and each row of the matrix or data frame will indicate one set of variable. For example, in this set of faces you can use 15 variables at a time and these variables are indicating different parameters like height of face, width of face, shape of face, height of mouth, width of mouth, curve of smile, height of eyes, width of eyes, height of hair, width of hair, styling of hair, height of nose, width of nose, width of ears and height of ears. So, these are the order in which you will get the outcome. So, in case if you have 15 variables then there is no issue, but if you have say less number of variables then the data is repeated for remaining type of parameters.

So, let me try to take here the example. So, once again I will try to take the same example of height, weight and age and this data has been stored in a data frame data 3D. Now, I try to create here the this I try to execute here the command faces data 3D. So, as soon as you execute it here you will get here this type of outcome right. So, you can see here it is trying to tell you here what are your variables and how this command faces has modified it. For example, height is now modified to that is indicated by height of face, weight is indicated by the width of face, age is indicated by the structure of face.

Now, because we have not used all the 15 variables, so now this is getting repeated right and all these different features facial expressions are being reflected using the data what is stored in height, weight and age right. So, well my objective was here to illustrate you how to do it, but definitely I would request you to take a data which has 15 such variables and try to create this graph and after that you will see that this type of picture will come. You can see here these are different types of data sets, different types of colors and you can see here what is happening. So, you have here you can see here these are here 5 rows for 1, 2, 3, 4, 5 and then you have here such 5 faces right.

So, let me try to show you this outcome in the R console. So, I already have means install this package on my laptop, but you need to install it, but I need to upload it surely. So, if you try to see let me clear the screen. So, you can see here I have uploaded this package and if you try to see here this is your here data which you have stored here data3D right. So, data3D is here and then I will see here faces(data3D).

So, data3D is here already here. Now, if I see here faces of data3D. So, you can see here there are 5 operation 1, 2, 3, 4, 5, 5 sets of observation in 5 rows and you can see here this type of this outcome is here which I shown you on the screen on the slides and these are the different graphics of different faces which are coming here and if you try to vary the data values the structure of nodes etcetera there that will also get actually changed right. So, you can see that it is not very difficult to create such graphics in the R software and beside those things there are several other functions like contour line, dot chart, image, mosaic plot, perspective plot there are many other options in the R software which are available right. So, now we come to an end to this lecture and we will also stop here with the description of various types of graphics. So, now you can see in this lecture we have considered star plots and this 10 of faces which are capable of handling more than 3 data variables. But once again I will say if you try to increase the dimension of the data different types of complexities will automatically enter into your analysis.

They will enter, but it is not difficult to understand them it is not difficult to solve them. The only thing what you need is practice and experience and experience can be gained only by solving more number of issues, more number of problems and getting more experience. You can see here now for example, in the case of Chernoff faces it is handling 15 variables. Now, to understand that which data is going to change how, how

the nose is going to look like if I try to change increase the data, how the eyebrows are going to look like in case if I reduce the data, how the hairs are going to look like if I try to increase the data or decrease the data.

These are different aspects which you have to learn yourself. I can give you here this guidance, but unless and until you try to create a data set and try to vary those different values and try to compare how the faces are changing and similarly how the stars are changing you will not be able to understand the relationship between data and the graphic. And finally, what are you going to do in real life? You will have a huge data set, you will have no time or no capability to look into the individual values, but you will try to create such plots like as cerna faces or star plots. Now, looking at those faces or this star plots you have to diagnose what your data is trying to speak, what your data is trying to inform you and this is the beauty of graphics. For example, you can handle up to 15 variables without any problem just by looking at the faces for which the god has given us the inherent capability. So, now towards the end of this chapter on graphics I would say that I have considered here only a limited number of graphics, but there is a long list.

I cannot make you expert in all the graphics in the given time, but surely I can help you in gaining the confidence. So, that now after this if you want to learn any other graphics you can look into the help menu, you can look into the R software and then I am sure that you will be able to plot them successfully and then you will have an expertise, you will develop an expertise to look into the graphic and then retrieve the information from the given set of data. So, you try to take an example, try to practice it, gain some experience and be an expert data scientist and I will see you in the next lecture till then goodbye. Thank you.