

Advanced Algorithmic Trading and Portfolio Management

Prof. Abhinava Tripathi

Department of Management Sciences

Indian Institute of Technology, Kanpur

Lecture 3, Week 1

We will discuss the R implementation of plots and charts. The basic building blocks of these plots and charts we have already discussed in the fundamentals of our module. Now we will apply those building blocks and create charts and plots that are very aesthetic in nature. As a starting point, we will set the working directory first. We will go to the session set working directory. Choose a working directory.

We will set the working directory where we want to save the plots and read the data. So, we will copy this `setwd` command as we have been doing at the beginning itself. So, in the future whenever we want to use it, we can simply run this command and we need not remember the location. So, we will start with the data visualization.

Now as a first example, we will take gasoline price data. We will import data set from this import data set option from Excel and we find that there is a gasoline price which we will select. We will import the data. A certain set of commands is run. So, we will copy these commands and again for future reference we will copy and paste these commands from the console to our R script. So, whenever we need them in our future applications, we already have them ready. So, we have the data which is read. If you want to see the data a very easy way to do that is just type `head weekly gasoline prices`. So, we can see the head of the data. It contains weekly dates and gasoline prices in dollars per gallon.

Let's start by plotting this data. We already have some understanding of plotting. So, we will write the command as follows. Plot command weekly gasoline prices which is the series to be plotted. Then we want the plot to be of type L line plot we want.

Then we want the central heading to be time series. This is the central heading. We can change the name as per the requirement. Then column we want as read and type LWD line width as 2. So, we have seen this argument.

We will set the line width to 2. Then we can run this command and notice a very nice-looking plot has been produced. You can zoom it. The screen can be zoomed in. You can see the plot.

The number of permutation combinations. For example, you can export this plot save it as PDF save it as an image or you can copy it to a clipboard. Adjust the dimensions of the plot as you may wish and then copy and paste it into your Word document or ppt or any value you would like. As a next step, a more efficient way to do plotting in R is to use the library, Zoo. For time series like we have just seen Zoo package provides nice plotting functionalities.

So, we will make use of the library Zoo and we will convert this original series into a Zoo series. It's very easy in R. We will use the Zoo function. Now we have to specify the series we want to convert. So, the series to be converted is prices and you also need to give the time element.

So, the time element is a date. So, we will also use that. So, notice I am giving the price series along with the date dimension and if I run this command, it will be converted into Zoo series. If you do not have Zoo installed, you need to go to Tools install packages and then type Zoo to install this functionality. If you already have Zoo installed like I did then you simply need to run the library Zoo command to add the Zoo functionality in your current working environment.

Now once, you have added a very easy way to plot this or first we will first convert this Zoo series into quarterly. Currently, it is on a weekly basis and is slightly noisy. There are a lot of fluctuations as you can see in the plot. So what we will do is we will convert it into a quarterly series and with Zoo functionality that is very easily available we can use as. a yearquarter so it will be easily converted into a quarterly series.

Now if I run this WGP1 which is the new series that is now a quarterly series. So, if you want to see the plot it is very easy, I can simply use the previous plot command itself and we can I will just plot both of these commands. So, in the quarterly series, I need to simply type WGP1, and I will use the green color probably. What I will do is I will try to superimpose both put these graphs together. So, for example I will mfro command I will use mfro, this mfro helps plot multiple plots in the same window I need par mfro equal to c two commas one.

So, with this, I can plot two plots in the same window. Now along first, I will plot the original plot which is a weekly series. First, I need to run this par mfro command then plot the original plot, and then the next plot or even better I will do something else I will time series weekly time series quarterly, and let's plot them now. Notice the plots I will just zoom the plots and you will notice that the quarterly one is much smoother and gives clear trends while the one which is in red which is weekly is much more fluctuating which is on the expected lines. So, with this, we have understood the simple plots we can also practice a little bit with bar plots, and in the next set of videos we will see bar plot

application in more detail but as a simple example of bar plots, I can simply type bar plots.

I need to specify the series so I will use the quarterly series because it is less noisy. I will type the y-axis as quarterly prices. It's very easy quarterly prices. I also need to give the x-axis, so my x-lab is equal to date, and I need to give the color, the color is blue. I want the graph to be blue.

I need to give the central axis. You will notice that the same features are used whether plot or bar plot. This is the power of R. These are just building blocks that once you learn you can apply in different sets of plots like bar plots, invert plots, histograms, and so on. I will be giving the border red.

Now we are good to go. I will just plot it and zoom it for you. Okay, I need to do the par info as 1, 1 otherwise it will be plotting expecting to plot. So, we will do par mfrow 1, 1 which will tell R that now only one plot on one window. Now I will run the command and once I run the command notice the plot very nicely shaped bar plot for quarterly prices is plotted and the boundaries are red actual bars are in blue and it gives a lot of good information about the price. So, this is the simple bar plot.

In the next set of videos, we will learn some more plotting with different data sets. We will visualize the economic outlooks of different countries around the world. In this video, we will make use of another data set which is the nation's data set. We will go to the import data set as we did earlier and then import the nation's data.

Click on import. The data will be imported. A brief visual examination of data can be done on the screen. We will import this. It is also viewed on the R studio which seems to be correct here.

So, we will keep this data. We will copy and paste these commands for future use that have been used to import the data set. They will be very useful. Now if you want to check the data you can simply use the head command head nations. Let's see what this is. So, the data carries member nations' status with the UN, their different PGAP per capita GDP trade deficit rating, and their outlooks for different member nations.

So, let's start by first telling R that this outlook variable is a factor variable. So, this is basically the rating outlook ratings we can see here in the rating column. The outlook is stable positive or negative depending upon the outlook of the rating agency. They assign these values table positive negative. So, we will tell R that this variable is actually a factor variable.

We have discussed a lot what is factors variable. We will tell R that this is a factor variable. We can use it as `factor` or simply `factor` command to tell R that this is a factor variable. Now that we have explicitly stated it as a factor.

Let's check the levels and of this factor variable. So, we find that there are three levels negative positive, and stable. As a starting exercise, we make a very simple bar plot. We need to table nations' dollar outlook, and we will first plot the main barplot. So, if we plot it, we can see a very nice-looking bar plot with a grey color all the three levels of negative, positive, and stable frequency are plotted.

You can also visualize you can directly not only visualize you can see the frequencies with this table function you can see. So negative 11 positive 10 is table 39 as you can see in the console window. You can rotate this plot instead of this you can use `horizontal` as true. So, if I do that horizontally as true the graph will be tilted. So that those permutation combinations you can do.

Let's do a little bit more aesthetically improved graph now. So, for example, I can use this same command and I can add some more features starting with let's say I can change the color let's change the color equal to `rainbow` maybe since there are three we will use `rainbow(3)` then probably we'll use I can use you can give any name to y-axis so probably I can give y-axis you can change it to anything that you may find more suitable. Similarly, I can give the x lab as an x-axis that can also be changed as may be suitable. You can set the border also let me set the border as maybe green is how it appears.

You can change it later as well. So, let's just run it. So, notice a very nice and interesting looking graph appearing y y-axis x x-axis labels, and everything. So, I can change it for example I can change the y-axis to `frequency` which sounds more suitable here, and the x-axis to the outlook. Central heading, I can change to `rating outlook`. Let's see how it appears if I run this a very nice looking and very aesthetic plot has appeared.

Now on this chart, you can add legends for example if you want to add legends you can simply use this `legend` command and you can put this maybe the top left seem to be slightly empty so I can put the legend on the top left then I can specify what are these legends so I have negative legends negative one value is negative one is positive and the next is stable. Now the fill also will follow the same fill `rainbow` which is `rainbow(3)` so we'll make it more consistent and we'll plot it and then we'll see how it appears so if I run this it seems nice we can zoom it you can see it is quite nice although if you wanted to change the shape and size of this legend you can use this `cex` command, `cex` as maybe if you want to decrease it you can use it `0.4` although it may not look very nice but let's see so it's quite small actually if you want to make it large maybe you can make it too large with `6` equal to `2` and it seems to be too large actually although when I zoom it seems okay it's too large so we'll make `6` equal to `1` probably that would be more appropriate

here and yes it seems to be so here a very nice and aesthetic looking chart is appearing which we can copy paste and use in different purposes.

In this video we'll visualize frequency distribution and relative frequency distribution using soft drink data again we can go to import data select Excel file since it is an Excel file we'll go to soft drink data file cold drink data again the data will be visible on our Excel import and we can see it's a sort of name of cold drinks like coca-cola diet coke and so on we'll import it a view will be appearing on R studio window we can see what is the data how does it look now that we have imported the data we can again copy paste these commands that are appearing in our console it can be useful in future now first probably we will do the head cold drink data and change the name to soft drink to make it more easy to use soft drink now let's see the summary of data, summary(soft drink) data you can put the head to see what is inside this data so you have Coca-Cola, diet coke, Pepsi and so on now as a starting point what we'll do is we'll convert into a tabular form so for that I need to just modify a little bit I'll use C bind dot data and we'll see the impact C bind so we are converting it into a data frame and inside data frame first let me show you what is happening so I am using the table function table function soft drink so it will create a sort of table with frequency of different brands so coca-cola diet coke and so on and when I put this table function inside C bind data frame or data frame will be created in this data frame you have soft drink with two columns of drink and frequency now this kind of data is very suitable for plotting on bar plot so let's try to first compute the relative frequency of the data so soft drink a dollar and we'll create a new variable as we have already seen which is called relative frequency, in the relative frequency we have soft drink dollar frequency divided by sum of the dollar frequency which will give basically the relative frequency sort of probability of some sort so we compute that now we'll plot this bar plot we'll make you the bar plot command soft drink dollar relative frequency now we need to give the names to names dot arg and it's quite easy we can simply use the name column which is soft drink dollar soft drink which is the name and we need to specify the color one way to do is just to explicitly specify the color we make use of numbers R recognises numbers as colors as well so we'll make use of five colors because there are five code so if I run this a very nice looking plot has appeared we can zoom it so this is the plot now if you want to play around with the colors a very interesting way is to use rainbow colors so in color I'll use rainbow 5 then it will provide me rainbow 5 very nice looking colors will be appearing as we can see here you can also shift the graph by making this horizontal argument as TRUE then the graph will become horizontal I'll make a separate command for instance so you can make horizontal equal to true and the graph will be tilted now another way to improve this graph would be to add legend so first I'll go to the previous plot before adding the legend I would like to print it in the vertical form now a good legend will make or improve the aesthetic aspect

of this plot so I'll add the legend that is quite easy so legend it seems probably let me fix the top right you can place the legend wherever you find more convenient let's put it at top right legend top right seem to be slightly empty now in the legend I would be adding short forms of these cold drink to make it more suitable CK for Coke, DCK for Diet Coke, pep for dr. pepper, PSI for Pepsi and SPR for sprite so this will be my legend and then I need to just add some fill colors so probably add a lot of fill of rainbow 5 which I have already used so I'll use rainbow 5 for fill so that color coding remains consistent and I can see it is slightly and it's nicely looking but slightly overlapping so I can change the six probably I can decrease the six to zero point five maybe it will be a slightly smaller plot and smaller legend and yes it is but doesn't look so good so maybe earlier one was better so I'll use it as one which is the original thing and looks nice all the bars are recognized also we can add pie plots to represent the same data we'll start with the pie plot let's see how it works a very simple pie plot is available in basic R functionality you can use pie then related since we are plotting the relative frequency thing we'll use the data which is relative frequency data again we need to add the labels as we did earlier labels will remain same so the labels will remain same I'll use the same labels I need to give a central heading let's give it as main equal to soft drink you can change it as may be convenient and then you need to specify the colors probably I'll use colors as rainbow 5 you can choose any color give explicit names red blue green or use numeric coding 1 2 3 4 5 I am using rainbow 5 so now I'll just plot this and a very nice pie chart has appeared we can see that this pie chart you can export it to PDF or do as you may wish, another way to represent this thing would be in a different manner would be to use the same pie plot but in the labels rather than using the name of the cold drink I'll use their distribution the percentage share so for that I'll do a slight little bit of trick what I'm doing I'll show you so I need to create the percentage share which I can do it like this I'll use space 0 for adding few characters I'll use the round command so that the percentage is in two digits so I'll use the round thing I'll multiply hundred with the relative frequency relative frequency is the sort of probability that we computed earlier and I'll specify that we want only two up to two decimals then we'll add a percentage here because we want to show that this percentage number since we are multiplying with 100 and let's see what we have produced here so if I run this command you notice the percentage shares are appearing now so I'll copy this and in the labels I'll use this I'll use this as a label you can see here what command I have put I have used the label as those percentages and I'll run this notice now labels are appearing but again the problem here is that we have lost the original names of the cold drink so we need to add a separate legend to show the label so I'll add separate legend here in the legend probably top right would be appropriate so I'll use top right you can use top right, top left or any place that you may find convenient again I need to specify the label so I'll copy the original label from here that I have already put and you not type it again so I'll use the original label now I can specify the cex as one if not appropriate will change it again the color coding

will make it consistent as rainbow 5 and let's see what happens so if I add the legend they are slightly appearing on the side of it probably sorry I need to add fill equal to well slightly modified little and yes so now it looks very nice the legends are added and everything is fine legend size also is see fine if you want to decrease the size of the legend little bit you can use 0.8 here probably that would be more appropriate let me try it and yes it is slightly appropriate so this is how we plot the pie chart there's one more addition that you can do with the help of plotrix library you if you have not installed you can since I've already installed I can simply go to my library function and add this plotrix if you have not you can go to install packages select plotrix and then install it since I have already done it I will directly add library plotrix and add the plotrix functionality to my current working map the advantage and benefit of plotrix is that I can add pie 3d pie charts syntax is quite similar I simply copy paste the original thing which I did earlier so I'll simply copy paste I'll simply copy paste the original argument that I use in the pie function it will remain identical and I'll run it and you notice a very nice 3d pie chart has been produced I only need to add legend I'll use the same command for legend no changes I'll use the same command for legend and copy paste it here a very nice legend will be produced so it looks slightly less sharp so probably I'll decrease the okay so too small so you can play around with the size to see whether it is visible or not probably top left let's try top left let's see what appears okay so here it appears to be fine so I'll increase the size to one okay so now it is appearing so we can see the legend also now on our 3d plot so this is how you can produce 3d plots also.

Now that we have understood bar plot and pie chart we'll try to understand plotting histogram in R as a starting point we'll create a random variable sampling observations from 50 to 100 with the 10,000 size and with replacement so this is our variable which ranges from 50 to 100 with 10,000 observations let's see the summary of this variable that we have created so it ranges from 50 to 100 as we thought of now as a very basic histogram plotting to start with we write this hist command var we can give xlab as frequency we are very familiar with these notations ylab as variable maybe or you can give it any name as you may find suitable it would be nice to give color as maybe rainbow 20 depending upon number of bars more or less colors will maybe use, border I can use green so for example here we have if we have 10,000 observations and you have given breaks and now this is a very important argument break this tells are how many breaks are to be created so if there are 10,000 observations and you give 100 breaks then each bucket or each frequency bucket will carry 100 observations so let's give 100 breaks, breaks equal to 100 and that will also determine how many colors will be used and we can see here a very nice histogram you can zoom it see different color codings now this break argument is very critical let me show you its application so what I'll do is I'll create space for three plots one comma three and I will change the break so what I'll do is I'll create histogram with different breaks so 100 breaks and the next one I'll use with 10 breaks and the other one I'll use with single break and we'll notice how it works

so I'll just enlarge the plotting window a little bit and let me use it so `par mfrow = c(1, 1)` I missed one argument now first plot 100, 10, 1 and notice how the distribution changes as I change the breaks so it becomes more and more homogeneous because at an extreme case is one break so all the observations become part of one bucket so only one bar while another extreme case is hundred breaks so each bucket carries hundred observations and the middle case is 10 breaks now we look at some of the plotting arguments we'll again use `par mfrow = c(1, 1)` and we'll add some features in this plotting so we'll use `power.info = c(1, 1)` to plot single plot with hundred breaks but now we'll try to use make use of some of the more arguments starting with central heading so for example you can add a central heading to this name earlier there was no central heading probably we'll use a central heading called histogram that is one heading you can also increase or decrease the size of this central heading with the argument `cex.main` maybe two so enlarge this you can also change the font if you want to change the font you can use font different fonts let's use four you can also change the axis sizes so let's say axis three which will increase the size of axis you can change the color, label colors to green axis labels maybe green central heading also you can change so probably if I want to use central heading to red I can use that red then you can use the axis the numeric values that are plotted on axis you want them to be certain color maybe black I can use that and the size of axis maybe one so with this with these arguments I can run this okay I need to use `par mfrow = c(1, 1)` and now it is produced so you can see the central heading is red, axis labels are green, axis numeric values are black as we have plotted you can change all these things for example it seems that `cex` labels quite small so I can increase the size of notice the labels at the axis have increased `cex` is quite large so I want to decrease that I'll put at one so now the numeric values will be decreased so you can play around with these permutation combinations

Now many times you need to plot rather multiple plots in one window or you need to or also sometimes you need to add numeric labels on these plots so for example I'll make use of the original this plot command `hist` command what did if you remember when you plotted this we had ten breaks so we can see those that is the plot we got because of this so this is the plot that we got and again put it so this is the plot now I want to save this plot object if I run this command the plot object is saved in a variable `M` now I want to make use make use of this variable `M` in order to give labels so I want to use the midpoints of these buckets so each bucket will have a midpoint which we want to use that will be our X coordinate Y coordinate is counts this is our Y counts and the labels we want you want to set the labels as counts only so the counts will be our labels we want a numeric counts and we would position them with this argument you can adjust the position it's just that this position I found more suitable and optimum for this kind of plotting notice this so you have numeric labels associated with each bar which look nice it gives you some exact idea of the size of the bars the `M` dollar mid is the X coordinate, `M` dollar count is the Y coordinate on XY axis because X axis is frequency and Y axis is

variable value so we have set the coordinates and then labels as counts and we could add the labels.

Another interesting thing that you can do is plotting two different data or two different series on this histogram let's see how to do this so again we'll start by simple by setting the margin that is very important here because we are plotting two plots we need to give sufficient margin so we will use this par margin equal to see I mean slightly adjust the margin you can play around with these numbers to see how it fits in your case so I'm setting the margins of the plot once I set the margin I'll plot a very basic R plot let me plot that so I'll plot the histogram and let me plot the histogram I'll make use of the original command so this is the histogram I'll probably set the main heading also here so main heading as histogram and let us plot this so like I said you can play around with this maybe I should use single color maybe here color equal to black or rainbow so this is my this seems to be appropriate I will not change it now you can play around for example you can increase the margin to let's say 0.5 notice how it changes the entire thing or if you want to further increase the margin I will increase on all sides if you want to increase or decrease margin or particular side you can do it right here for example here if I do it 2 it adjust the margin so you can play around with the margin here a lot of ways so I'll keep the original thing in that then I need to start plotting but in the same plot so I'll use par new equal to T so that means new plot will be created on the same original plot and then I want to maybe I want to add density so I'll add the density of this var variable line width I'll specify we know all these features line width 4 maybe the line should be in black color so I'll specify the color to be black I will not since already there are X and Y axis I will not specify any axis labels so I'll keep them empty I'll create my axis as false I'll keep Y labels as empty already I've specified that and then central heading also as empty so this is the new plot and notice how it is plotted okay so I missed the par new T running that so I now if I run it the superimposing density plot is plotted which gives some idea of the density next I need to add the axis on the right side so R recognizes that this axis line is 4 so this will be the fourth axis for R the secondary axis I'll specify the I'll add the color as black and let's add the labels as well and as black so a very nice looking axis is produced we can see that the axis is produced here now if you want to add the axis name you can add this M text argument and here I'll put cell density I need to specify that it has to be added at side equal to 4 then maybe color again I will stick with the black color notice but it is slightly less suitably placed so I need to shift on the right little bit maybe line argument specified to still not well shifted so probably that 4 yes now it is well shifted so I'll create the plot again I'll against run the all the commands from beginning and I'll get the nice looking plot here so this way I could plot or superimpose two plots in the same bar plot and make good use of that

We'll improve our bar plots and try to produce side by side and stack bar plots so we'll try to produce side by side or grouped and stacked bar plots to enrich with more information so for this we'll use meal data set we go to the environment import data set and we have a meal data set that we are going to use so in this meal data set we'll import it before preview data is provided so we'll just copy these commands as we have done earlier command that are appearing on our console window now we'll just have a brief look at the meal data so it seems that the data provides different restaurant how many restaurants under different pricing and quality rating data that is provided here so first let's do a basic group chart or stack chart plotting group or stack chart it's quite easy so first we need to specify the arguments in the bar plot which is we need to specify in the matrix form so this plotting requires matrix form so meal and will be plotting starting from row number one to three which will specify that we need to plot the only three rows good very good and excellent these three quality ratings and on the column side we'll specify column two to six which is ten dollar to nineteen twenty to twenty nine so on so forth up to total values so that will not then we'll specify the color, color as rainbow four since we have four columns we'll use rainbow four, for plotting border as white, x lab equal to price and we can specify y lab as number of restaurant now you can add many more arguments as we have already seen in some of the previous exercises but for now in the interest of brevity we'll stick with these and when I plot them notice a very nice looking bar plot which is stacked or group each bar carries all the four all the three quality ratings: good, very good, excellent, now in order to enrich this graph we need to add the legend so that it conveys the full information we'll add legend probably on the top left it seems to be slightly empty so I'll add the legend on top left I need to specify the labels for the legend and we'll use the names of those so we'll use the names meal dollar quality rating we need to add quality rating as level and three quality ratings we have which we'll use as the rating if I run this command notice three ratings will be produced which is good very good and excellent then cex equal to one which will specify the size of this label we can change that later we'll maintain the color coding with rainbow three so three here also we'll plot the chart and then three: good, very good, excellent so we have the three color codings and a very nice looking chart is produced now often you may want to produce a chart which is that side by side and stack so for that you need to modify the arguments a little bit will provide them so I can add beside equal to T beside default was false so we change it to T and then use the same exactly same set of commands to produce the side by side on stack you can also produce the in a horizontal fashion so if you want to produce in horizontal fashion you simply need to write one argument which is as we did earlier horizon equal to T and you can print in horizontal manner also so this was bar plotting in this video we saw how to plot bar plots in a stacked or group by group manner or stacked nature.