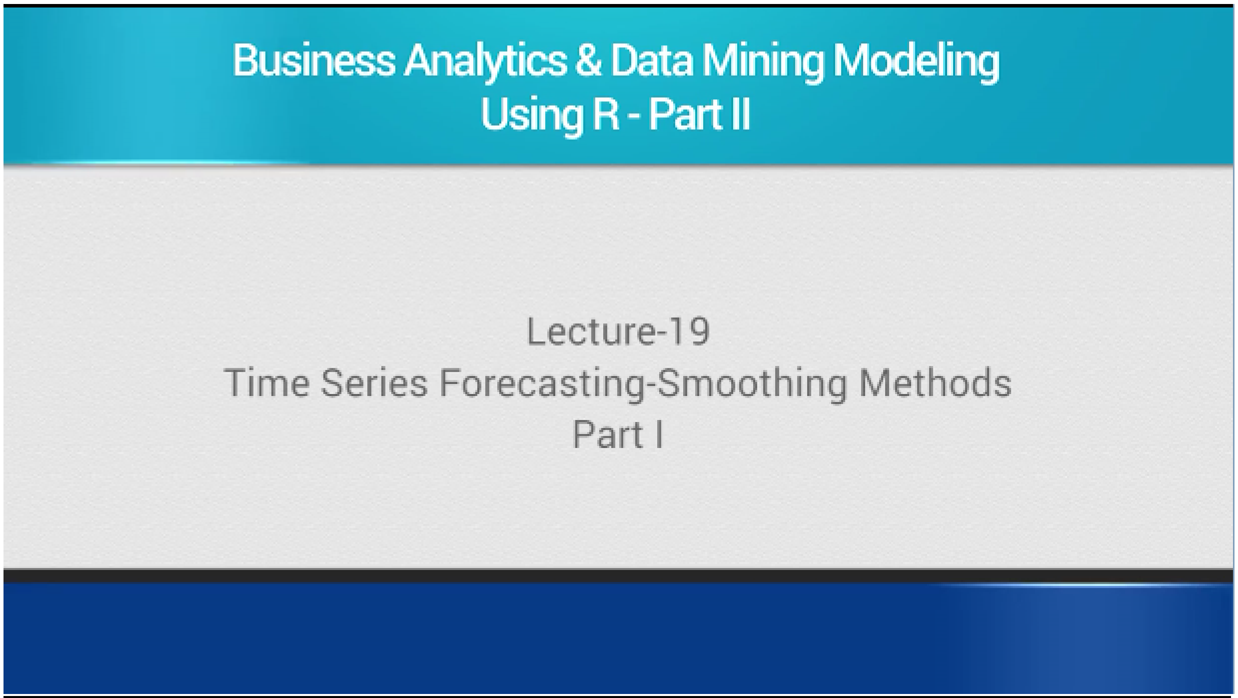
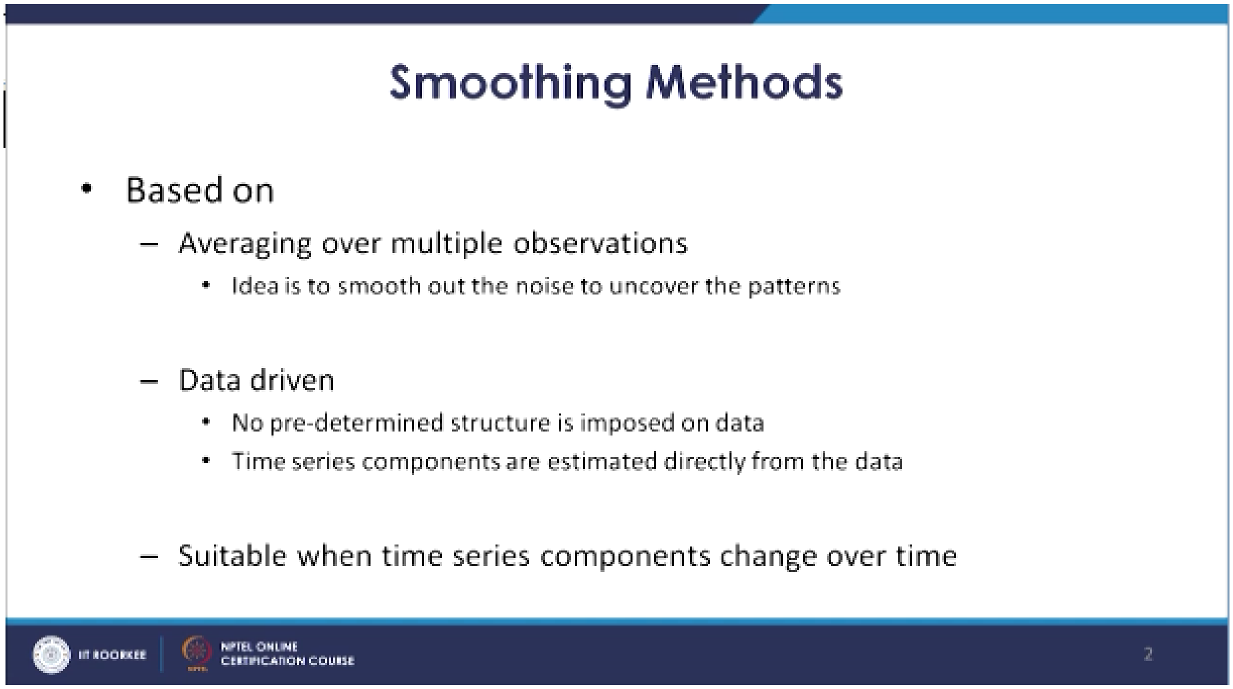
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
NPTEL
NPTEL ONLINE CERTIFICATION COURSE
Business Analytics & Data Mining Modeling
Using R – Part II
Lecture-19
Time Series Forecasting – Smoothing Methods
Part 1
With
Dr. Gaurav Dixit
Department of Management Studies
Indian Institute of Technology Roorkee



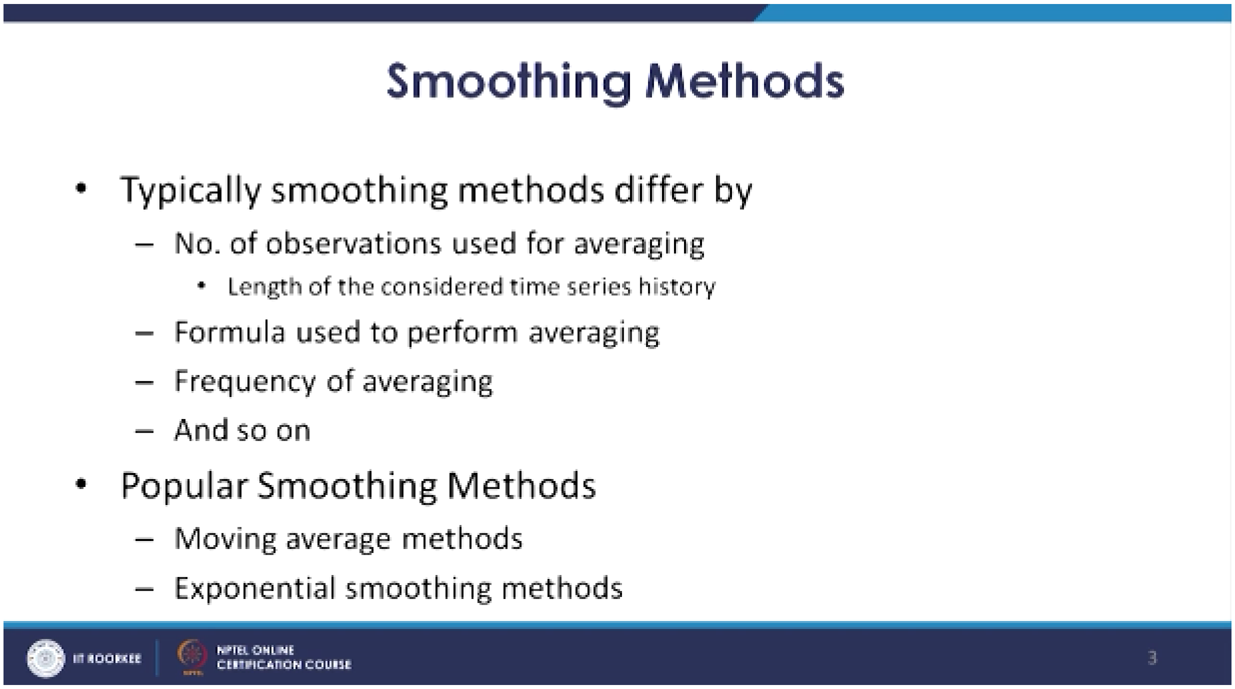
Welcome to the course Business Analytics and Data Mining Modeling Using R – Part 2, so this is the last topic of this particular module time series forecasting, and so this particular topic is on smoothing methods, so we have talked about different you know types of methods which are popular for time series forecasting, so we are now going to start the last you know type of method that is quite popular for time series forecasting, this is smoothing methods, so let's start our discussion.



So smoothing methods they are typically based on averaging over multiple observations and the main idea is to smooth out the noise to uncover the patterns, so we consider a number of observations and because we are looking to uncover different time series component, it could be trend or seasonality and you know, so we would like to average out some of the noise and then use this the average series for our forecasting purpose, so idea is to smooth out the noise to uncover the patterns and averaging over multiple observation is typically done.

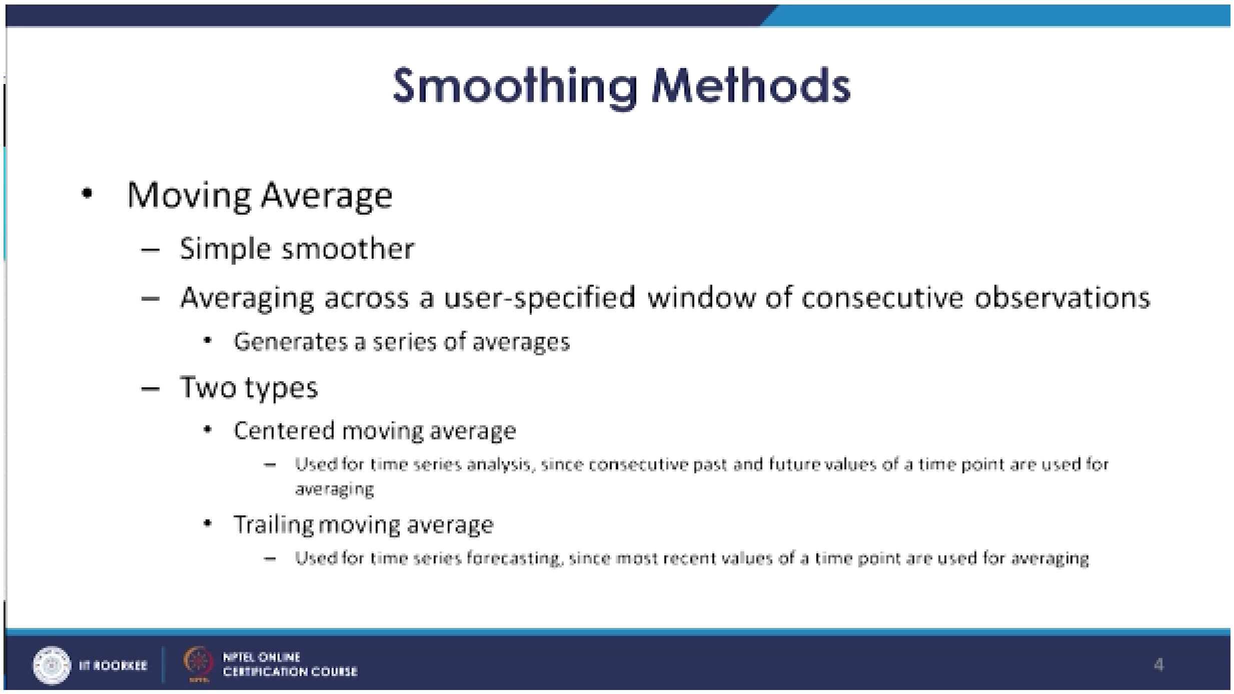
These smoothing methods are data driven so we don’t impose any pre-determined structure, so there is no structural modeling, any functional form of relationship that is assumed and implemented in smoothing methods, it is totally data driven so the model building, we learn from the data itself the different components of time series they are learned from the data itself and then they are used for forecasting purposes.

Suitable when time series components change over time, so typically regression based forecasting method that we have discussed in previous few lectures so that is more suitable when there is you know that the components they all, they don’t change quite frequently and therefore we can structurally model them which is why use the regression models, however if these time series components change quite often then these regression based forecasting method they might not remain useful and therefore smoothing methods are the one which can be used in such a scenario.



Main advantage being that since they are data driven so they can adopt and learn from the data and understand these changes in the time series component, so typically the smoothing methods differ by number of observations used for averaging, so the number of observation because typically the smoothing methods they are you know typically used for short term forecasting, because if the time series components they change quite often then this short term forecasting is the you know the toughest you know task that is to be done in forecasting time series, and the number of observation that are to be used for this purpose and specifically for averaging which is part of the smoothing methods, so that itself you know creates a different, creates a different types of smoothing methods, so length of considered time series history so that is one variation that we see in different smoothing methods, and the second one is formula used to perform averaging, so what is the formula? How do we go about doing this averaging, whether we just take a simple averaged or if we take a weighted average so how this averaging is to be done, so the formula that is being used for this averaging operation so that can change from method to method, then the frequency of averaging.

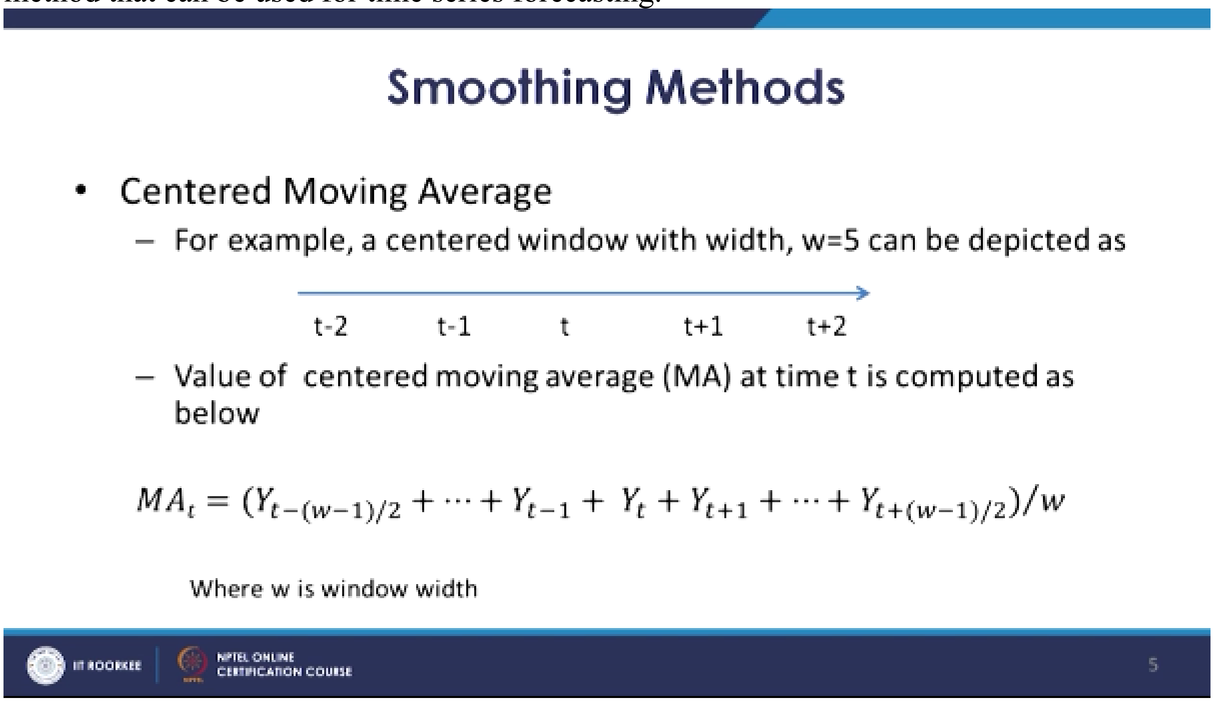
How often we go about doing this averaging, so that will also be there and so on, so you can see all these aspects, the different smoothing methods and few things that we discussed number of observation, formula used for averaging, frequency of averaging, so these are based on the data, so these you know competition are to be done on the data and you know there could be different smoothing methods depending on how these things are changed, so we are going to cover two popular smoothing methods, so first one is moving average method, the second one is exponential smoothing method, so we'll discuss one by one, so let's start with moving average method.



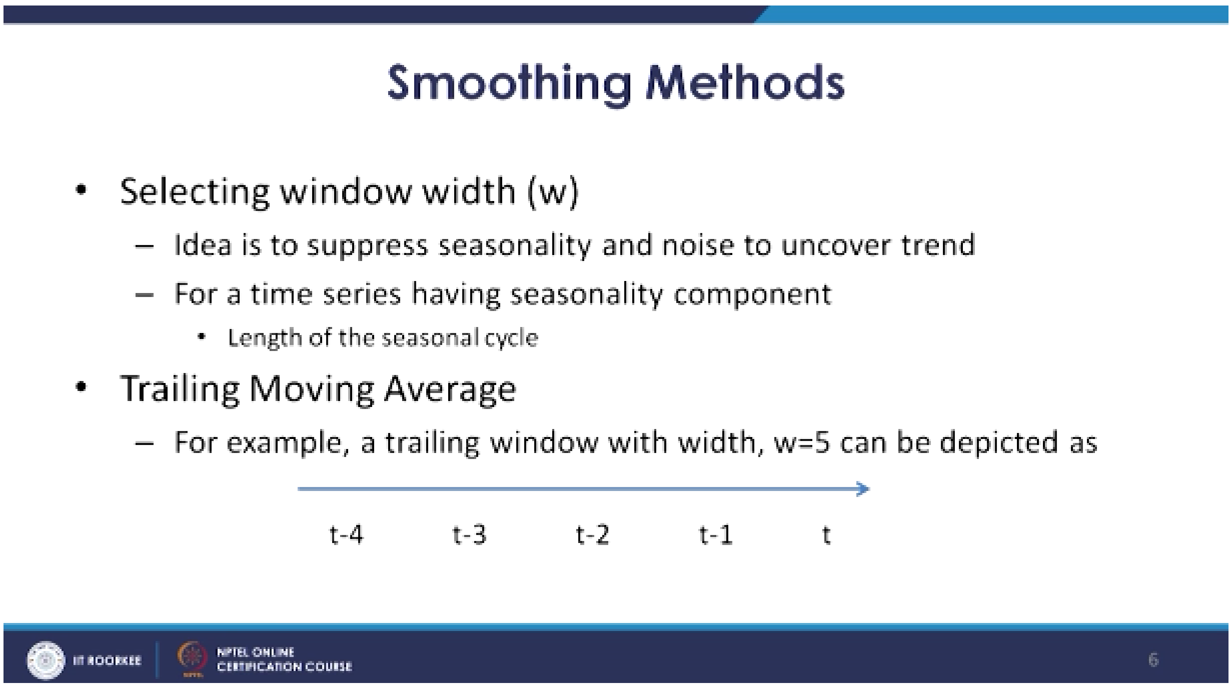
So moving average this is you know simple smoother, the simplest kind of smoothing that can be done on a time series, so this is mainly based on averaging across the user specified window of consecutive observations, so you know user has to specify in terms of the number of observation which are going to be useful for forecasting or analysis purpose, and of course that is, that can be determined by doing some trial run and error or some experimentation, so that and again domain knowledge and experience with the time series forecasting and analysis that is going to help the user in terms of selecting a particular you know window for this averaging, so once this window is known to us, suitable window size is known to us then averaging operation that is part of moving average that is going to be performed on those consecutive observations.

So what if this averaging operation essential does is it generates a series of averages, so we get a time series which is based on these averages value, so again there are two types of moving average methods that are popular, so first one is centered moving average, so in this particular method you know this typically this particular method is used for time series analysis, and not for forecasting, why? That is also mentioned here since consecutive past and future values of a time point are used for averaging, so we are going to use past and future values for a given time T, if we are also going to include T+1 and T+2 which are future values with respect to a given point T in this averaging process then of course this method cannot be used for the forecasting, because in forecasting scenario the average values are not known, right, otherwise if they are known then there is no point of forecasting, so centered moving average is typically used for time series analysis where we try to understand the time series analyze its understand its component and behavior, so in that sense the moving average can be good providers, good mechanism to create visualizations and you know understand the time series, so past and future values they are taken in this particular method, centered moving average for the averaging calculations.

The second type of moving average approach is trailing moving average, so this is the method which is actually used for time series forecasting and in this case most recent values, so if we look at you know if we are time point T, then we look at the most recent values at T-1 and T-2, T-3 and so on, and these most recent values are then taken for averaging, of course as we have talked about it is the user which is going to specify or you know window of these consecutive observations or you know this window can also be determined by, can be determined by experimentation as we'll be discussing in this lecture itself, so trailing moving average is the method that can be used for time series forecasting.

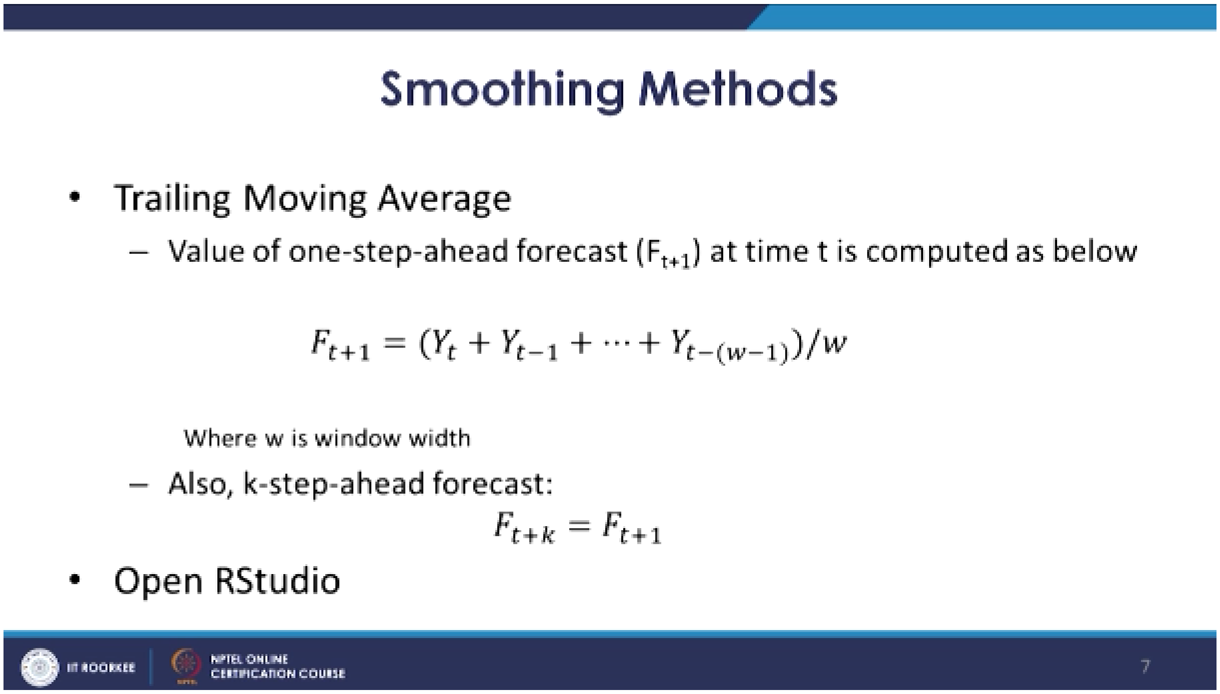


So let's understand these two you know moving average methods in a bit more detail, so first one centered moving average, if we take this example a centered window with width and a 5, W = 5, so it can be depicted in this fashion you can see this arrow is depicting this you know time series so we can see T-2, T-1, T and then T+1 and T+2, so T+1 and T+2 are the future values and T-2 and T-1 are the past values, so now these, because the window size is 5, so these 5 observation, these 5 consecutive observations are going to be taken for the averaging operation, so you can look at the second point, value of centered moving average MA at time T is computed as below, so here you can see Y, so this is the generalized equation that has been written, so for any you know time T the moving average can be computed like this Y and this T- within parenthesis W-1 divided by 2 then so on, YT-1 then YT, then YT+1 and so on, all this is divided by the window size W, so if you look at this we are simply you know computing the average for a number of consecutive values, so we decide on W, it take those consecutive values and take the average of it, and that average is being used as the moving average value, you can also see where W is the window width.



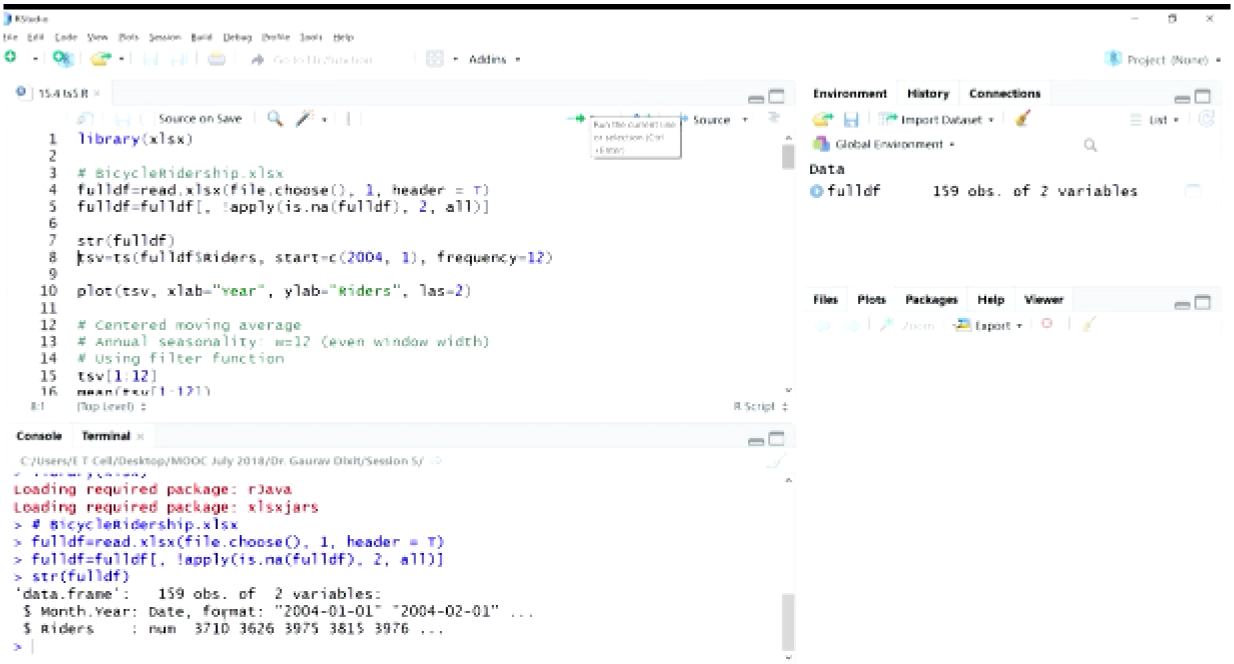
Now how do we select you know window width? So we would be discussing this particular thing quite often in this lecture, so the main idea is to suppress seasonality and noise to uncover you know trend, so if that is the idea you know then for a time series having seasonality component, length of the seasonal cycle could be one good you know, you know rule of thumb to select a particular window size. We would be again discussing you know how do we you know select a window width, few more details in the lecture.

Now let's move to the you know trailing moving average which is the method which can be used for the forecasting, so again if we take the same example, we take the trailing window with width W=5, so it can be depicted in this fashion you can see here this is the you know time T and then we have time T-1 the most recent value, then T-2, T-3, and T-4, so you can see it is the you know most recent values are the past values which are going to be used, past consecutive values which are going to be used in computing you know moving average, therefore this particular method can be used for time series forecasting.

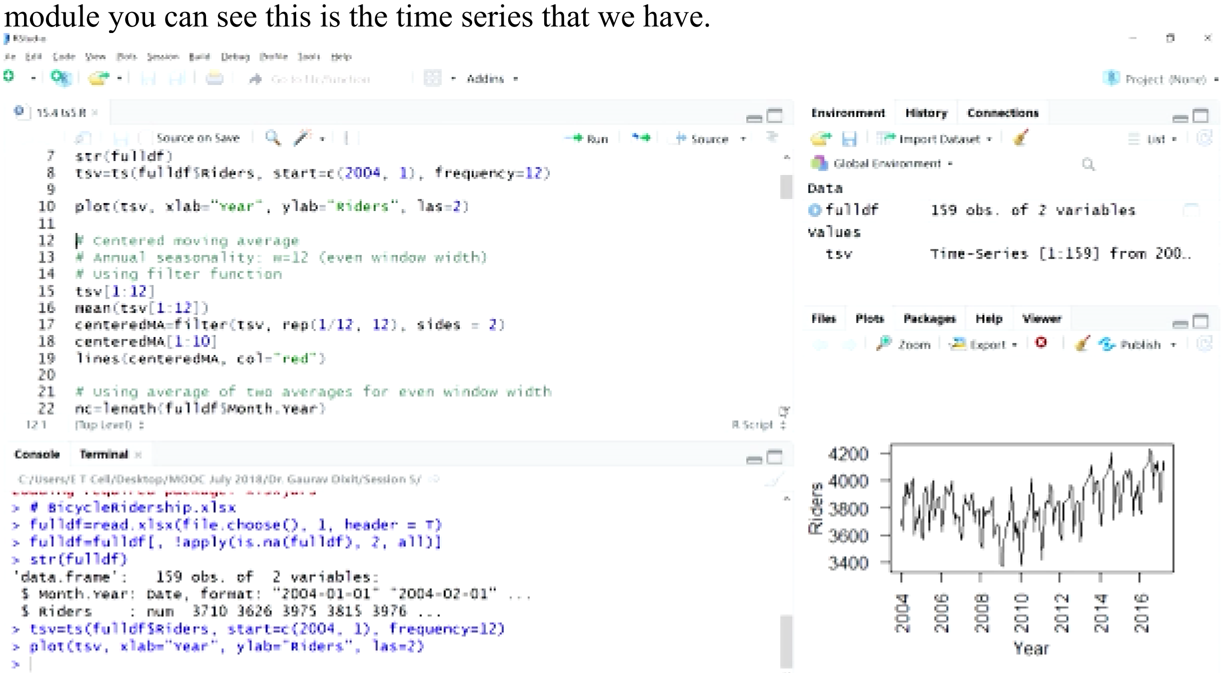


If you are interested in the formula how this is going to be computed, so this is how we can write it, so value of 1 step ahead forecast FT+1, so using those you know values till time T we are going to, we want to forecast 1 step ahead in the future, so this is represented by, indicated by capital FT+1, so this can be you know computed as below, so you can see FT+1 = YT + YT-1 up to the you know YT – within parenthesis W-1 divided by W, so again this is also in a way simple average calculation that is being done, so we need to select a window size and those number of observations are going to be taken and that average is going to be the forecasted value for the next.

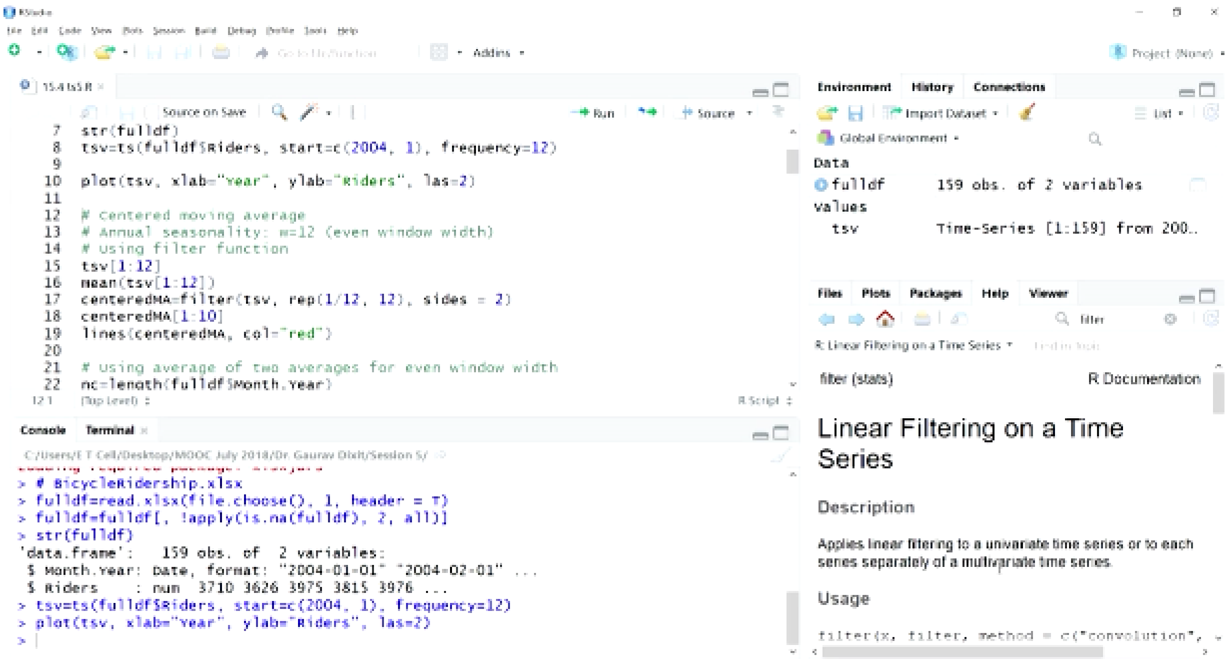
So the previous few values and they are being averaged out and simply that averaged value is being given as the forecast for the next point, so this is how it is being done, if you are interested in K step ahead forecast then FT+K equal to, is going to be equal to FT+1 because we are assuming at this point that we have the values up to point you know time T, so therefore for any you know, even for K value more than 1 we have the same information, therefore the best forecast that can be done is also going to be the same, so FT+K also essentially becomes FT+1, so this is the trailing moving average which can be used for time series forecasting, so let's open R studio and we'll do few exercise to understand this concept in a few more detail.



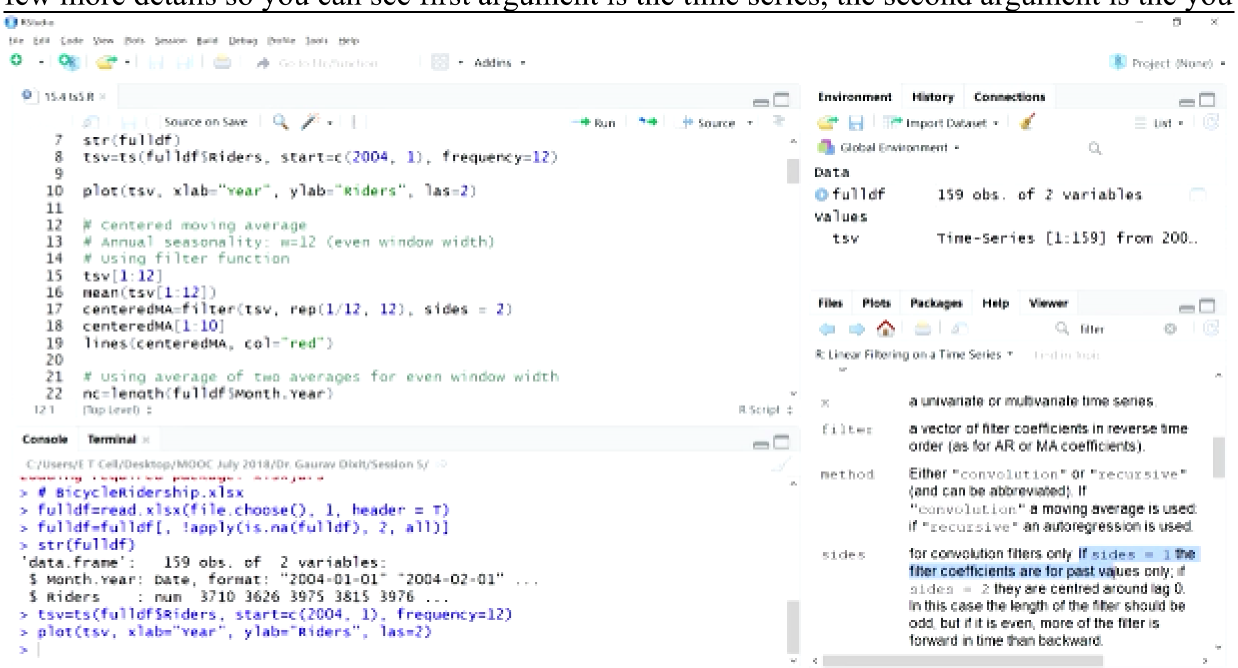
So if we look at, so what we are going to do is we'll first load this library, again the same dataset we are going to use bicycle ridership, so let's import this dataset, so you can see in the environment section 159 observations, 2 variables, let's remove NA columns if there are any, so this is the data frame that we have, structure of the data frame, we have the month information and the riders information, so let's create a time series object from this, so TS is the function for this, so let's run this code and time series object as you can see in the environment section it has been created here, let's again create a plot, so the same dataset that we have been using in this module you can see this is the time series that we have.



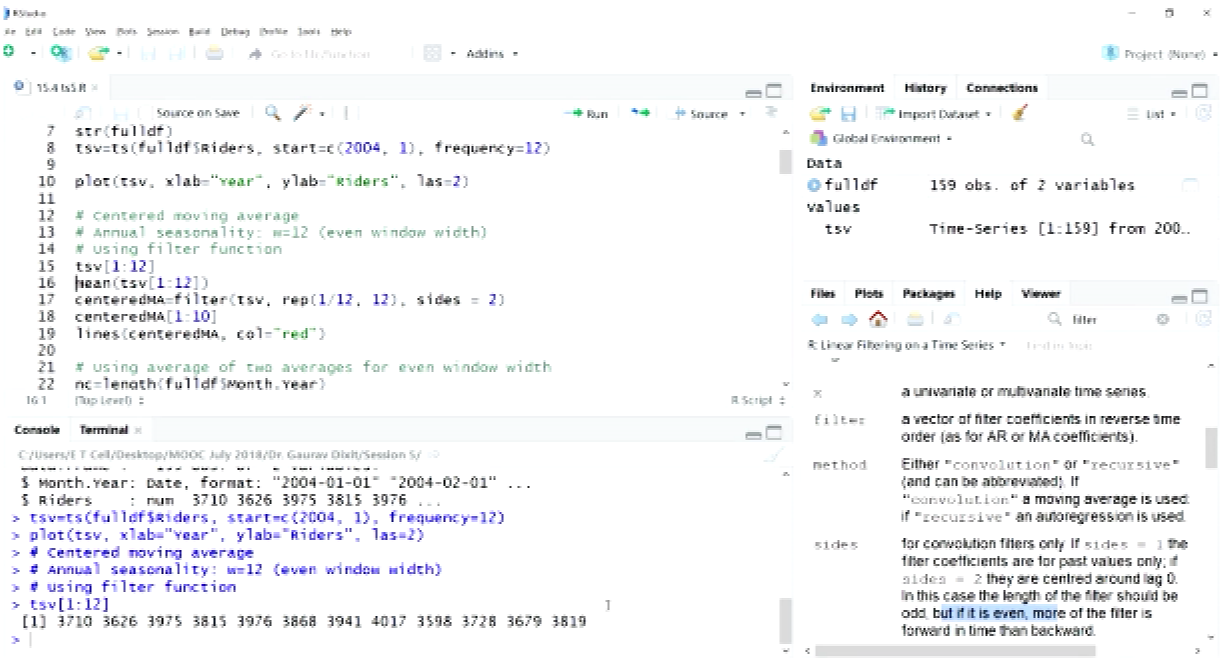
Now what we are going to do is we will first implement the centered moving average and in this we'll take this you know annual seasonality we already know about this time series, so we know that there is annual seasonality W12, so we are taking that as the window width for this example and we are going to use this filter function to implement our centered moving average, so let's find out a few more detail about this particular function, so let's open the help section, let's have a look at the filter function.



So we can see this function, linear filtering on a time series applies linear filtering into univariate time series or to each series separately of a multivariate time series, so let's look at few more details so you can see first argument is the time series, the second argument is the you

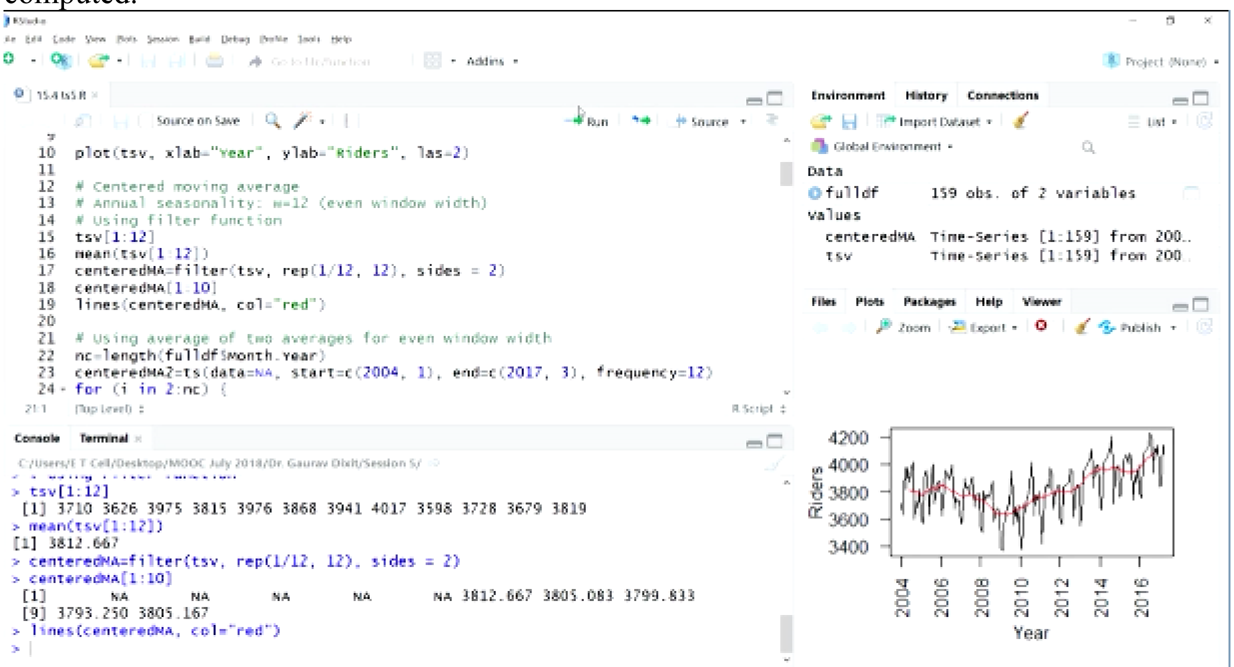


know filter, so then if you are interested in few more details for example this argument sides, so this argument is important for us right now, so this is if the sides argument is 1, then filter coefficients are past values only, so you can see if the you know this K scenario is mainly applicable for trailing moving average, but right now we are going to implement the centered moving average, so therefore we need to specify sides 2, which is for centered around lag 0, so again there could be 2 scenarios here, in this case the length of the filters would be odd, but if it even more of the filter is forward in time then backward, so if the window size is an odd value then there is no problem, everything is going to be you know centered around lag 0, so if it is T and the window size is 5, then we are going to have the you know T-2, T-1 and then you know T+1 and T+2 no problem, but if the you know window size is even, then how do we you know compute the average, so because the you know if you take you know T and so there could be one scenario that you take, you know if the window size let's say window size is even and it is, let's say it is 4 so you can either take T-2, T-1 T and T+1, so it is just one value which is into the future, so this is one way to compute average or you can take T-1 T and then T+1 and T+2, so there are two values in the future and just one value in the past, so that you know average is going to be you know more you know futuristic, so the more you know, of future value is going to be accounted in the model.

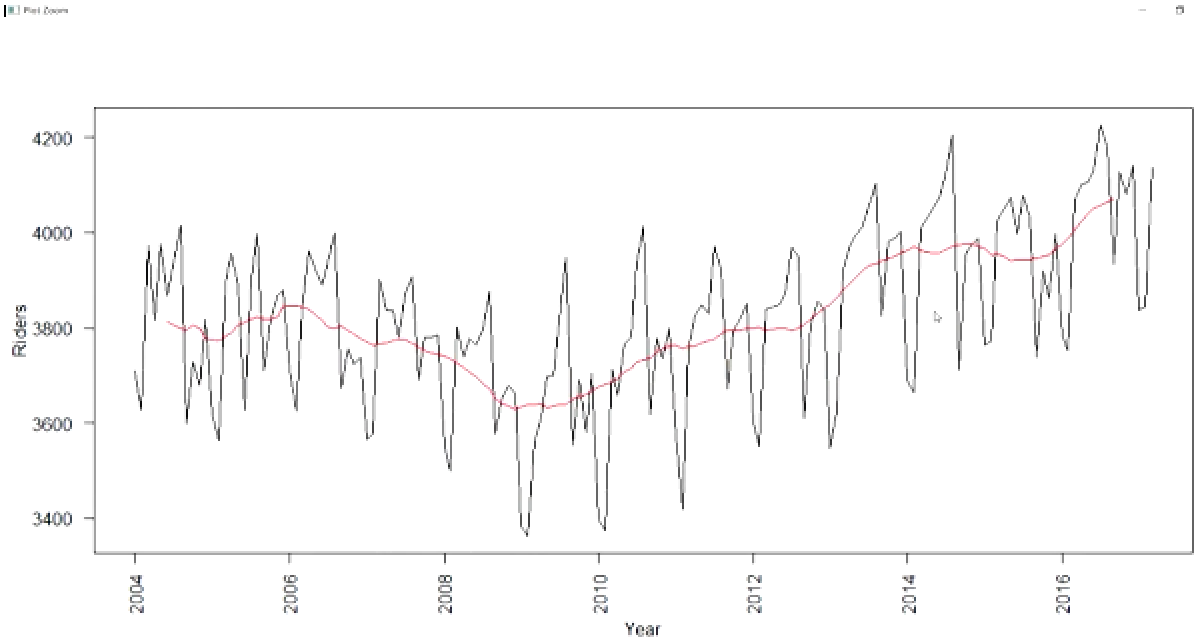


So how do we, so in this case the filter function it is going to take the second case that means if the window size is 4, so T-1 T and T+1 and T+2 is going to be the scenario for computing the average, so this implementation of filter function is essentially going to do this, so we should be aware about how these computations are going to be performed, so if we look at the first 12 values of the time series, the number of riders so these are the values, now because the window size is 12, so we need 12 observations to compute average, right, so and we are implementing centered moving average so the you know, and specifically the this has the window so as we already know it is even value and we already see, what the filter function does, so the first value is going to be come at the 6th point, so at this point so 5 observation in the past and 6 observations into the future they are going to be considered for computing this moving average, and that value first such value is going to be computed at point number 6, because for the previous 5 points we don't have submission information, so those values are going to be NA, so first meaningful value is going to be created at point number 6 and so on, right, reason being as I said 5 values we need in the past and then the remaining can easily be found out you know, 6 values can easily be found out in the future, so from this fashion these values are going to be computed.

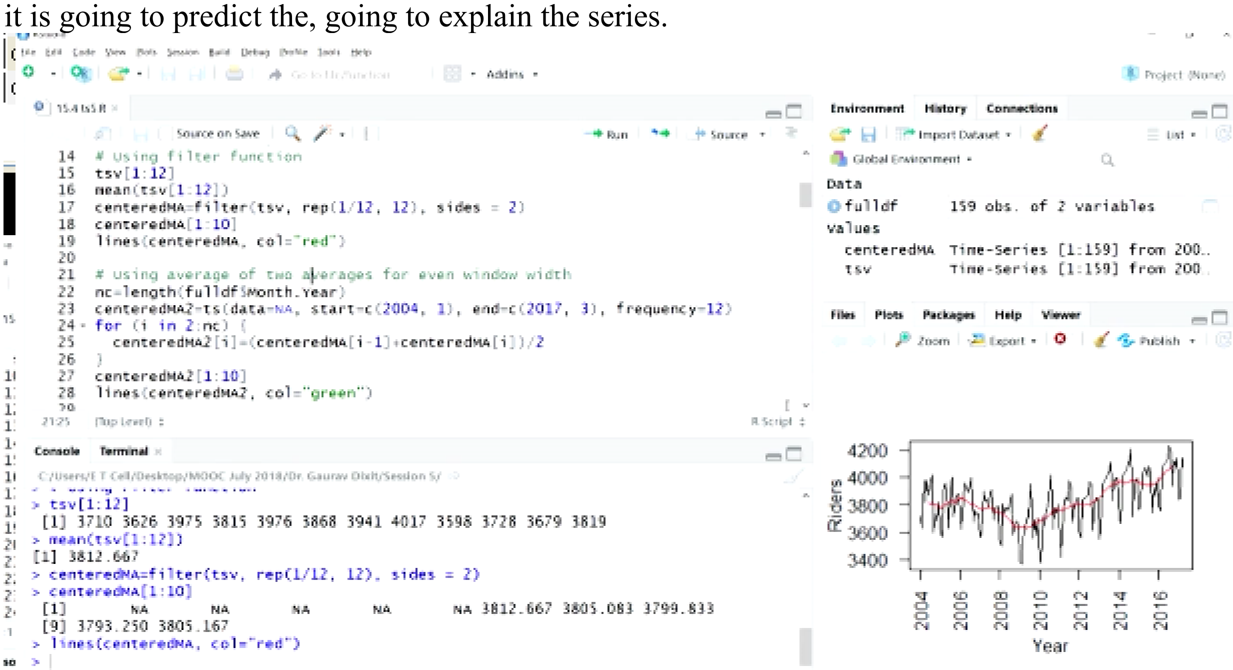
If you look at the mean of these values, so since these are the first observation and for the first value, these are the values which are going to be used, so the value is going to be, come out to be this much, the mean is this is the, going to be the you know value for this moving average, so you can use the filter function to obtain these values for the whole series, so we'll get the you know averaged you know series, time series, so we can use the filter function, then the time series object and then we are using the filter, so you can see the filter is 1/12, so essentially because the window size is 12, so in this fashion we can light and we'll get this, so essentially if you look at this particular you know second argument essentially we are specifying the weight for different you know values and that weight is nothing but 1/12 because essentially we want to compute the simple average, that side is 2, so let's compute this, and let's have a look at the first you know 10 values of this, so you can see it is the 6 value which has been computed, other values are NA because information is not known and then so on, another values have been computed.



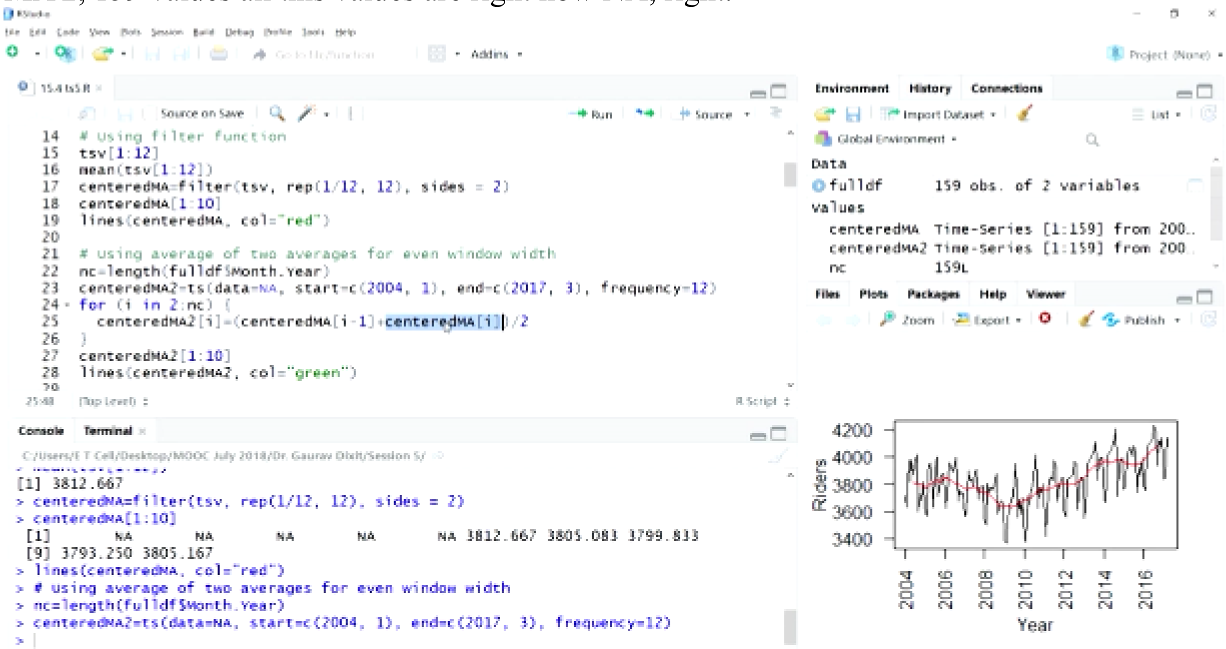
Similarly at the tale side of this you know vector also we'll get some NA's. Now what we can do is in the plot that we had we can add this moving average that we have just computed so you can see here in the plot, so redline has been created, so this redline has been created using this centered moving average approach, so we can see how this particular line, so this is seems to be



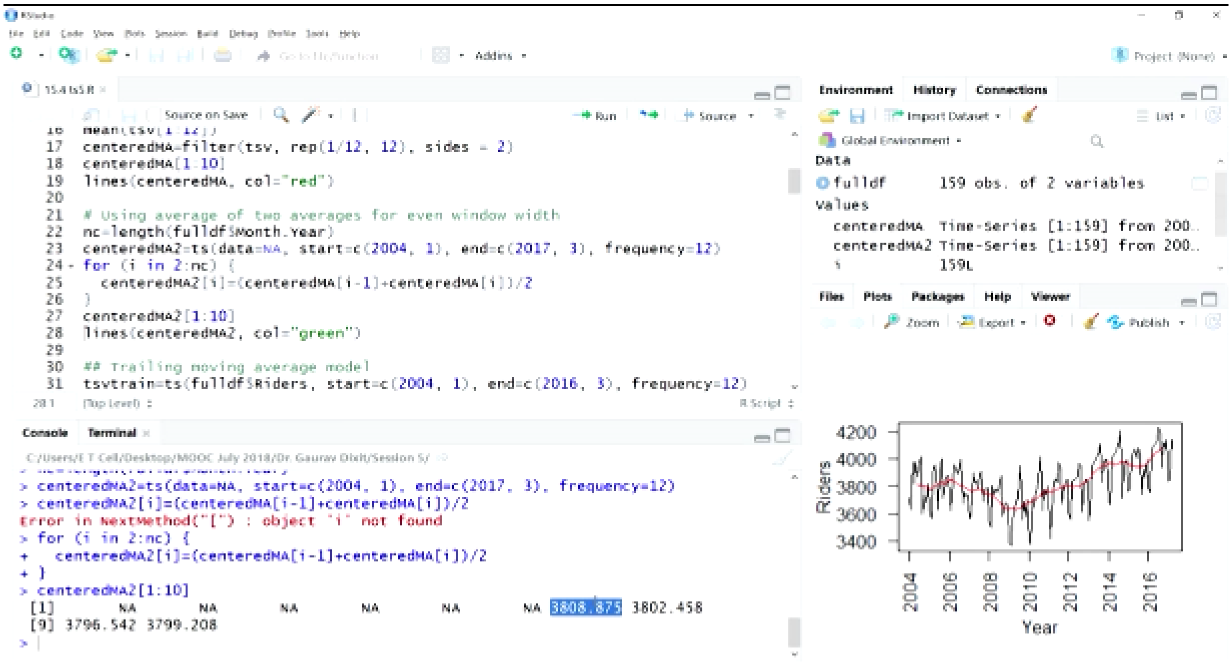
you know U shaped kind of curve but you would see this is not exactly U shaped like we did in the regression based forecasting, weightage was the perfectly U shaped, we did the quadrating modeling however in this you can see that in this centered moving average, it is in this fashion it is going to predict the, going to explain the series.

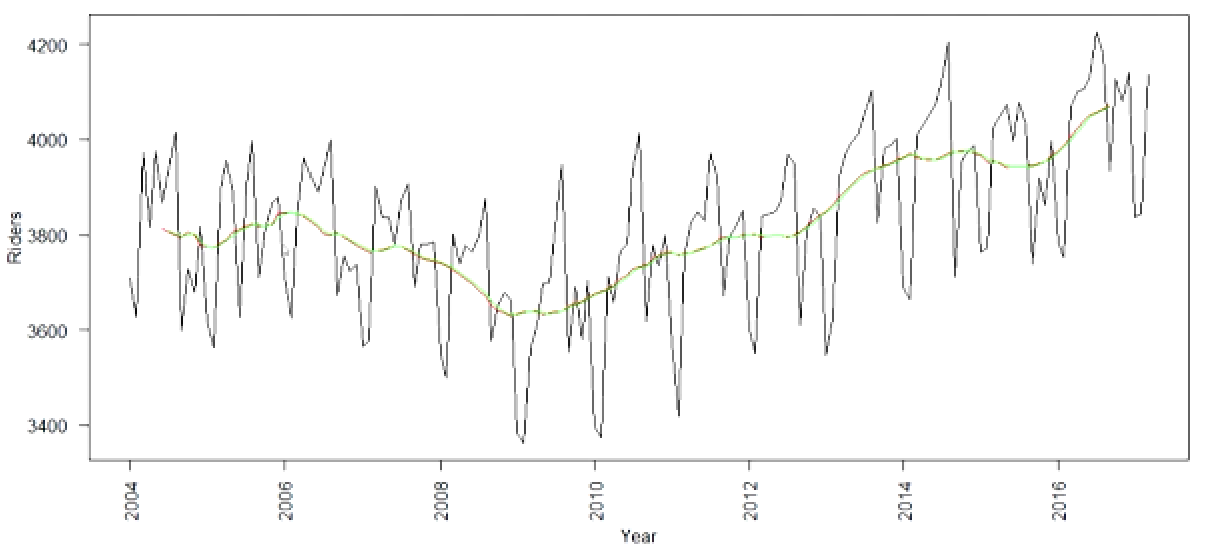


Now this averaging as we talked about the moving average methods, they can differ in the fashion the average is being computed, so we are going to you know take another, we are going to implement another approach for taking average, specifically this is applicable for even window width which is the case now, so first we'll capture this length, number of observations and then what we are going to do is we'll initialize a time series in this fashion so all the values are going to be NA, so they are going to have you know these many NA's values equal to the number of observation that we have in the series, so this has been initiated you can see centered MA 2, 159 values all this values are right now NA, right.

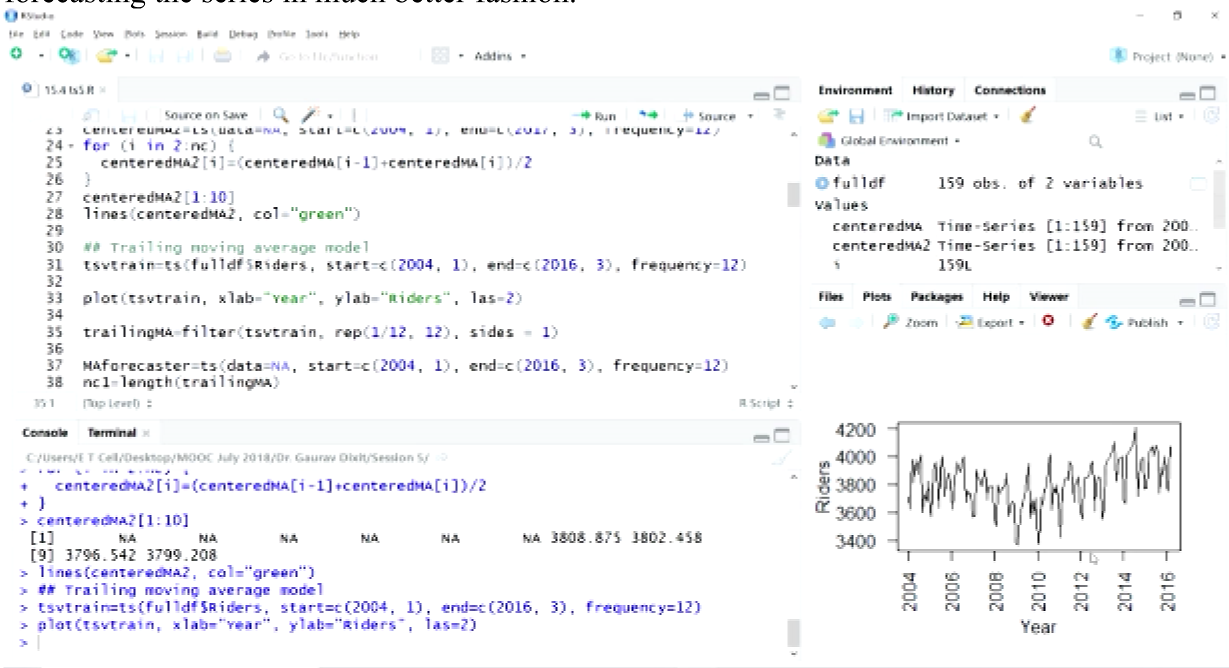


Now we are going to run a loop right, so this loop if you look at this this is taking the centered MA value which we have already computed I-1 and I so what this essentially going to do is we'll take you know, for example if the width was 4 then we are going to take first the average value for you know T-2, T-1, T and T+1 and this value is going to be computed by this, and then we are going to take this value will you know the second value will give us the value for T-1, T, T+1, T+2, so in that sense we'll get the 2 values which have already been computed and we'll take the average of these two values, and that value is something that we are taking at the, as the final you know average, right, so this loop will give us a new way of average time series values, so we'll run this and then have a look at the first 10 observation, so you can see the first value is now not at the sixth point, but at the seventh point, right, design is also quite clear because the way this value has been computed, and the values comes out to be 3808.875.

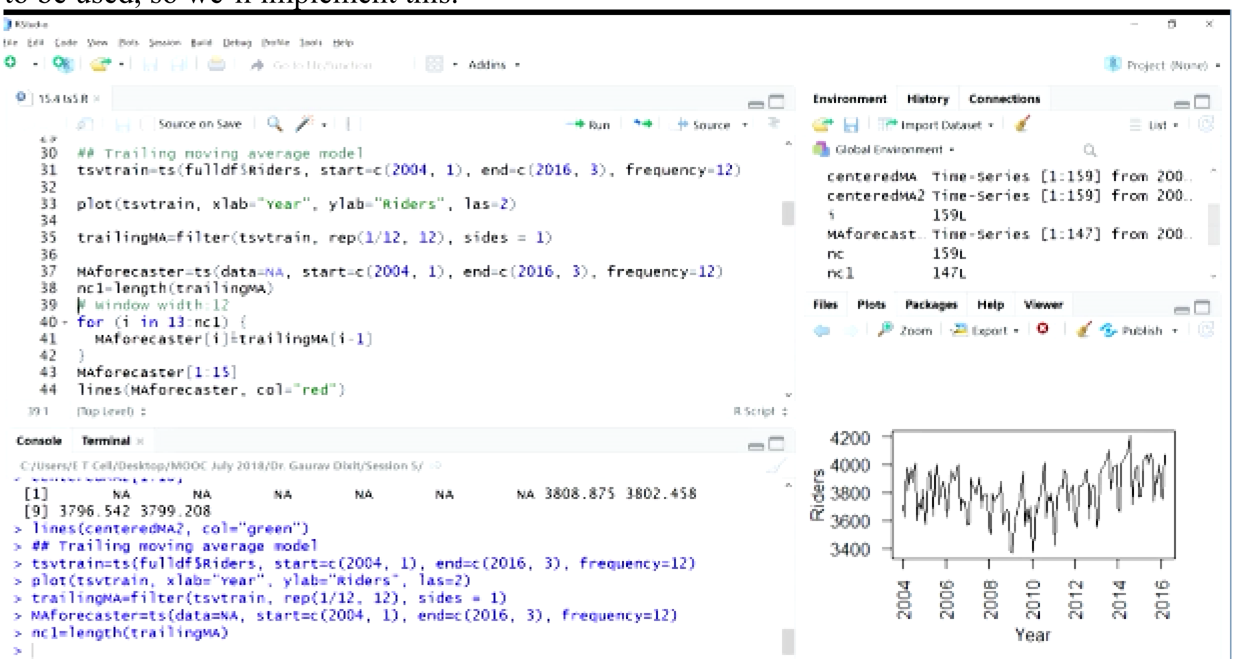




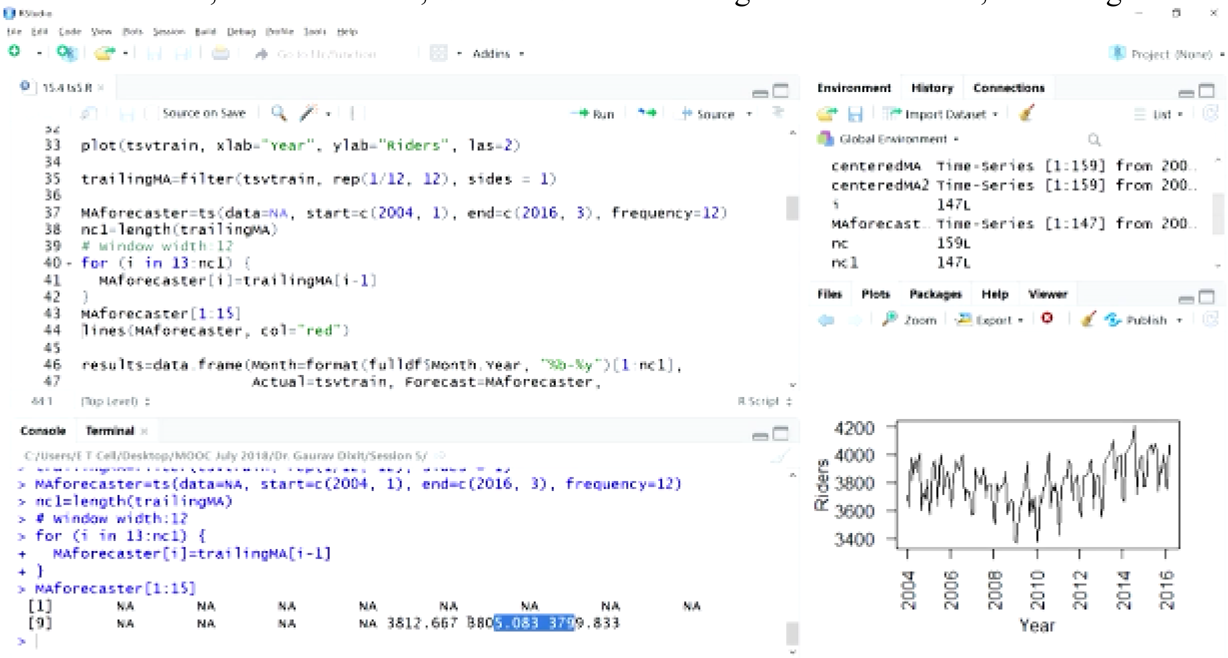
Now again we can add this line also here in the plot, so you can see a green line has been added to the plot and you see the red line the previous centered moving average and the second approach for centered moving average they are you know there is slight difference sometimes red line is you know above green line, sometimes the other way around, right, so because of the way computation has happened these values are differing, so it is important the way you are defining you know your, how you want to take you know average and because later on which one is doing better, of course it can be checked if it is for a forecasting model then this case this is for the, this is centered moving average so this is to visualize the series, however if it was for the trailing moving average which is use for forecasting we could have compared the performance and seen which averaging method is helping us in predicting the series better and forecasting the series in much better fashion.



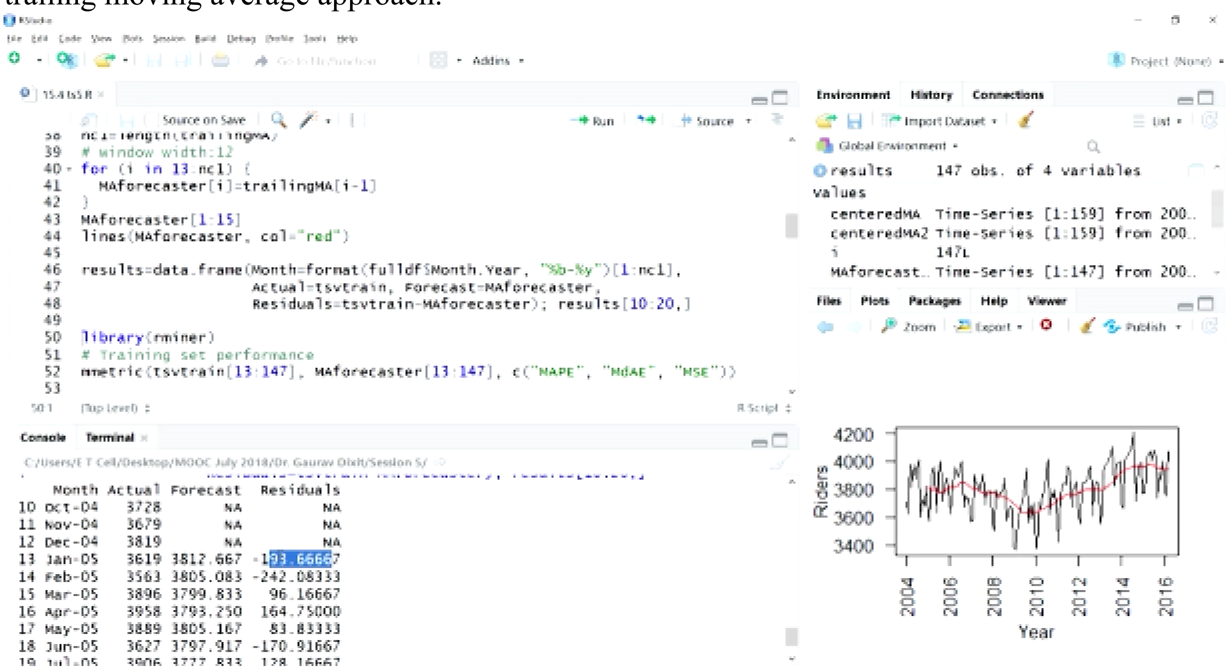
So this brings us to our the next approach, trailing moving average, so now we'll implement this, so let's create this you know TSV, so we are taking these many observations which we want to be part of the training you know set, so let's compute this time series vector, let's plot this, so this is the plot so these are the observation from 2004 to 2016 which are going to be part of this model, train per set.So now again we are using the filter function, right and in the same fashion we are going to, this argument has changed sides is 1, so only the past values are going to be used, so we'll implement this.



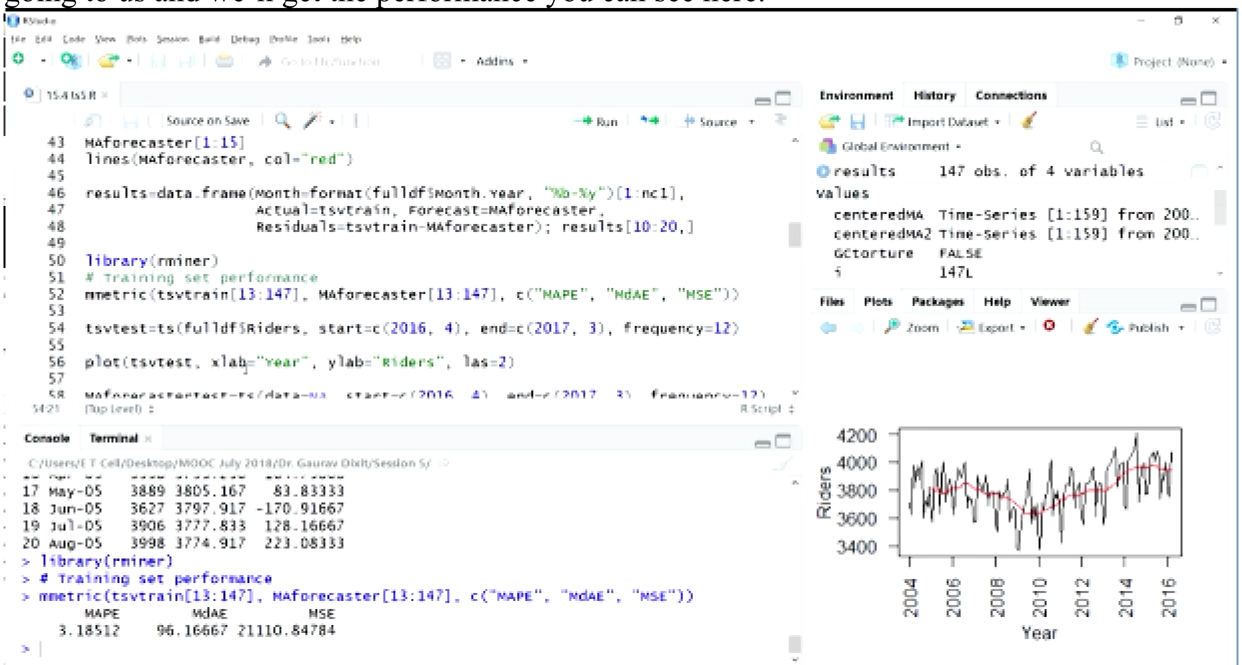
Now again we are going to initialize this particular variable and we'll capture the length and now we are going to you know apply on the test set so you can see you know we are going to apply from 13 NC1, so we are going to compute these values MA forecaster using this trailing moving average approach, so essentially we are computing the forecasted values for the trailing set observations, so let's run this, window width is here again is the same 12, so we'll get these



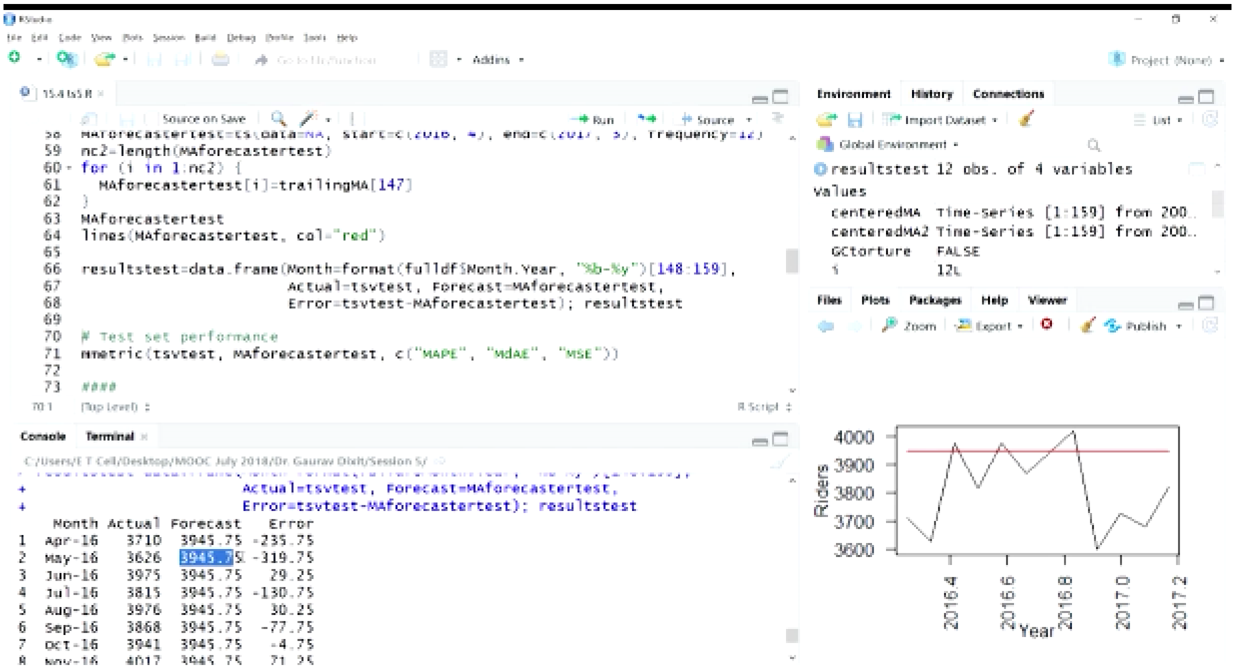
values you can see, so these are the forecasted values, so these have been computed using the trailing moving average approach.



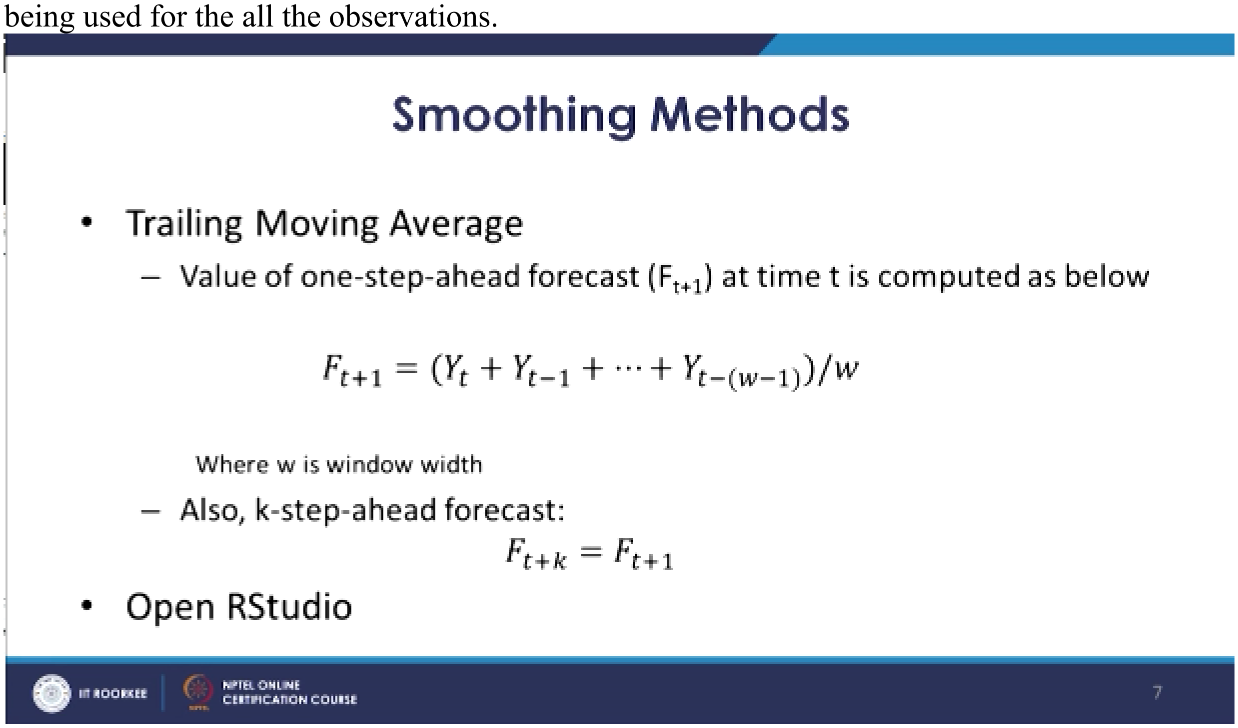
Now we can again plot this, you can say this is the plot, so if this is the plot for, plot using the trailing moving average, and this is one approach that can be used for forecasting, so we are interested in looking at the values, forecasted values, actual values forecasted values and the error, so we can see here, so you can see up to this point there are going to be no forecast in residual then from this point onwards we'll have, we'll start to have the forecasted values and the residuals also, so in this fashion we can implement this model, if you want to find out, check the performance of this trailing you know, trailing moving average model we can do this, let's load the R miner library, and lets you know call this function M metric you can see actual values first, TSV train, and then the forecasted values and these are the three metrics that we are going to us and we'll get the performance you can see here.



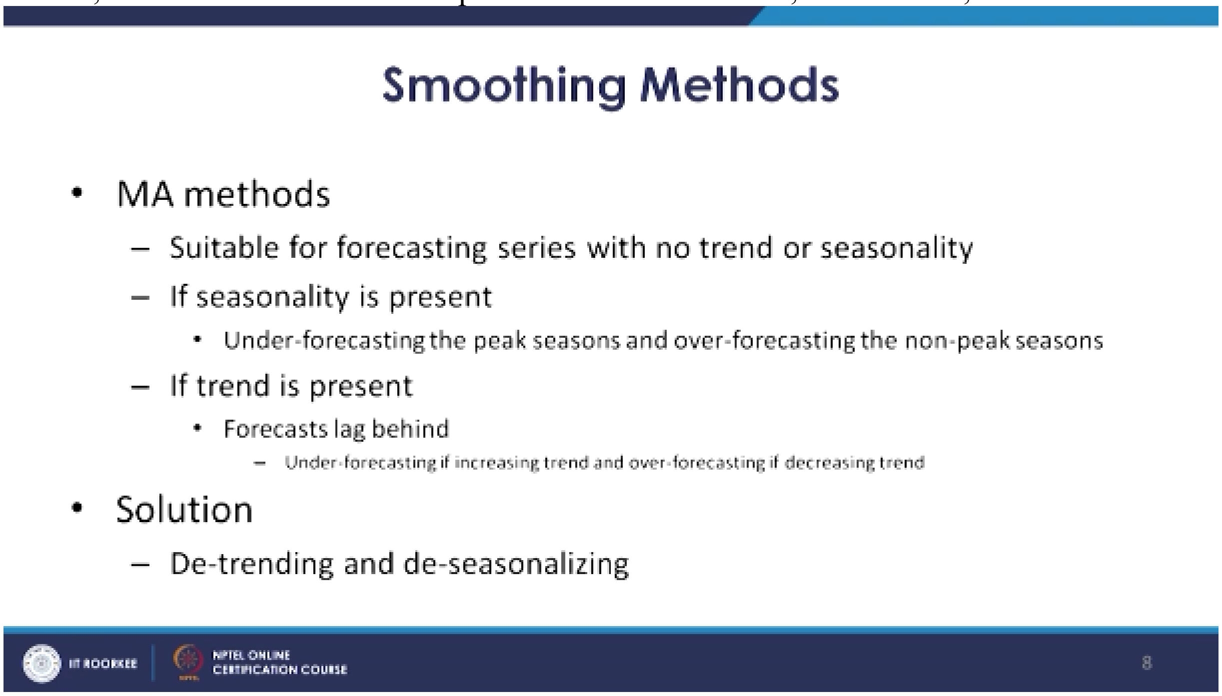
Similarly you know if we are looking to you know plot this, so we can do the same you can see 2016 to 2017 we can again create a time series object, we can plot this and we can apply the model here MA forecaster test so we can apply, we can initialize this, so all the NA values we can initialize capture NC length and then again run this loop to get the forecasted values for the test set and we can add here in the plot, so you can see the same value has been used for all the you know, all the points, all the points in the you know test dataset, it is the same forecast which is applicable, because in this model we'll get only just one value, because no future value is known after you know this, after this point 147 so for the one step ahead, two step ahead, three step ahead all those forecast it is the same value that is going to be used.



So we can have a look at this actual value on forecasted value you can see, it is the same forecast which is being used here as we have discussed before, again performance also for the test set also we can check, so you can see this is the performance we can see here, if we want to compare the performance, so we see the performance there is going to much you know wards in the test dataset you can see MAPE value is 3.18, MDAE value is 96, so if we look at the test set performance you can see this is much worse, MDAE value is on the higher side, MAPE and MSE all these values are on the higher side, so mainly because the same forecasted value is being used for the all the observations.



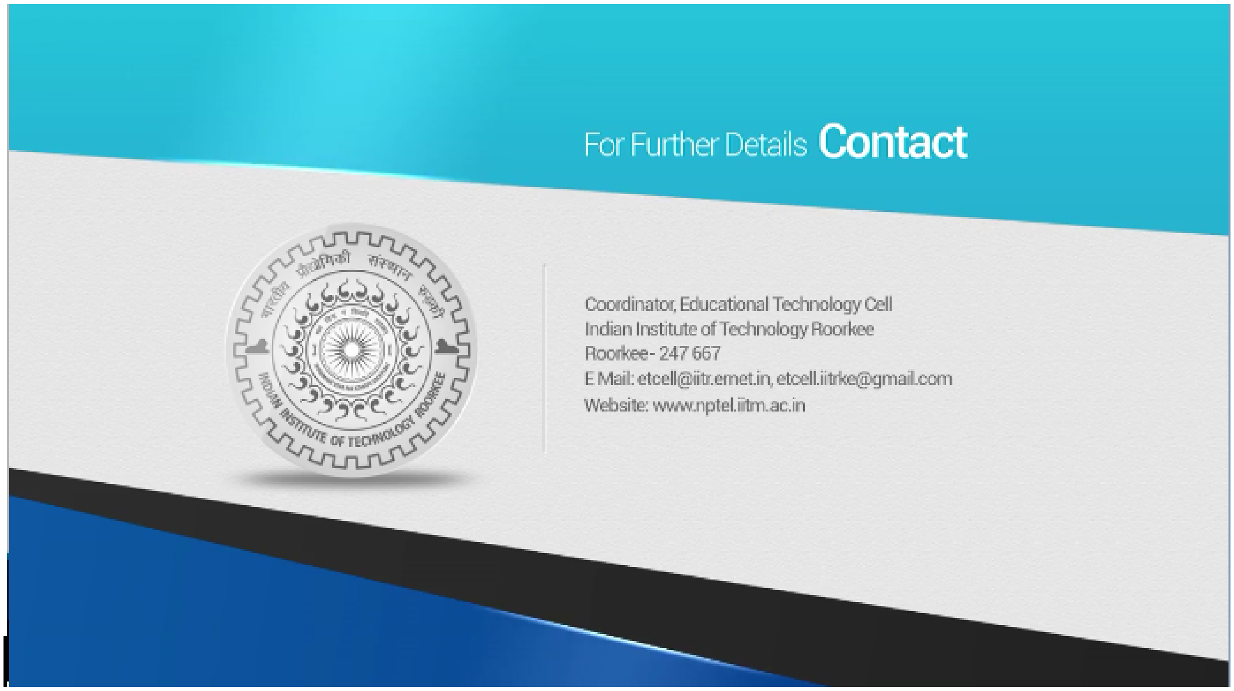
So let's go back, so this is how so till now what we have discussed we have been able to talk about centered moving average and trailing moving average, we also saw how we can go about computing and implementing these models for you know, for our time series bicycle riders dataset, so let’s discuss a few more points about MA methods, so as we see, as we have seen till



now that these are suitable for forecasting series with no trend or seasonality, so you know because we are essentially averaging the series and essentially you know these methods are typically used for short term forecasting, so it is much better if we have a de-trended and de-seasonalized series and we apply moving average method on that de-trended and de-seasonalized series, then short term forecast can actually be improved using these methods, so typically these methods are suitable for forecasting series which doesn’t has a you know trend or seasonality, this is something that we’ll do in the next lecture, well use the residual series that we have build in the previous lectures and apply MA methods on that particular series.

So if seasonality is present what is going to happen is under forecasting the peak seasons and over forecasting the non-peak seasons is something that is going to happen, so that is why you know this is not suitable if the series has a seasonality component, if trend is also present then forecast will lag behind that means under forecasting if there is increasing trend and over forecasting if there is a decrease in trend, so you know essentially these moving average methods can be used to you know forecast or understand the level and noise, so essentially what you are trying to do is you are trying to forecast the level of the time series and you are trying to average out the noise, however if the trend and seasonality component are present then the forecasting of this level can be you know take ahead and above performance will go down, therefore it is these methods are typically used when we have already de-trended and de-seasonalize the series and then these methods can be used to forecast the level of the series and by you know averaging out the noise.

So we’ll stop here, and in the next lecture we’ll talk about how do we go about de-trending and de-seasonalizing a particular time series. Thank you.



For Further Details Contact
Coordinator Educational Technology Cell
Indian Institute of Technology Roorkee
Roorkee – 247 667
E Mail:- etcell@iitr.ernet.in, iitrke@gmail.com
Website: www.nptel.iitm.ac.in

Acknowledgement

Prof. Ajit Kumar Chaturvedi
Director, IIT Roorkee

NPTEL Coordinator

IIT Roorkee
Prof. B. K Gandhi

Subject Expert

Dr. Gaurav Dixit
Department of Management Studies
IIT Roorkee

Produced by

Mohan Raj.S

Graphics

Binoy V.P

Web Team

Dr. Nibedita Bisoyi
Neetesh Kumar
Jitender Kumar
Vivek Kumar
Dharamveer Singh
Gaurav Kumar
An educational Technology cell