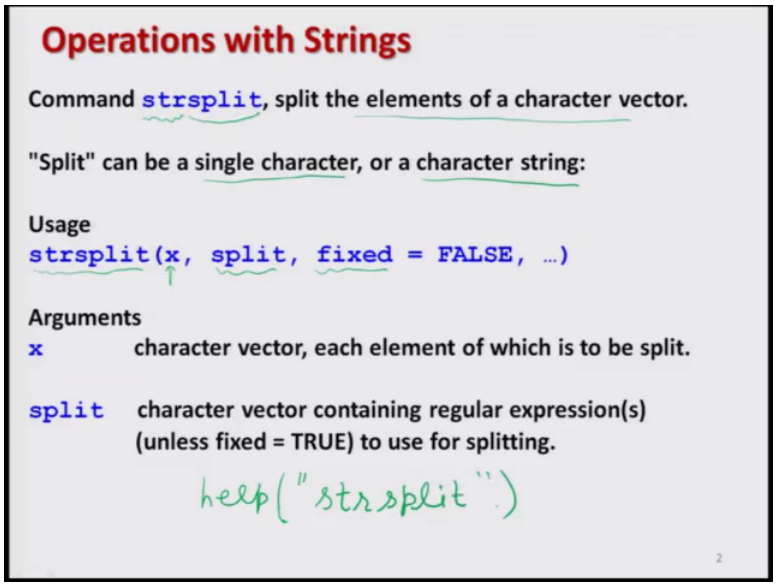
Introduction to R Software
Prof. Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture – 27
Splitting

Welcome to the next lecture on the course; Introduction to R Software. You may recall that in the last 3 lectures, we had dealt with the topics of a displaying and formatting different types of strings and essentially what we have learn in the last 3 lectures that how to join different types of numbers string characters to have a particular type of outcome.

Now in this lecture, we are going opposite. Earlier, we have joined the strings. Now we want to split them. Suppose you are given a sentence that is a character that is a string and you want to break the sentence at different places depending on certain arguments certain logics. So, how to get it done that is what we are going to learn in this lecture.

(Refer Slide Time: 01:14)

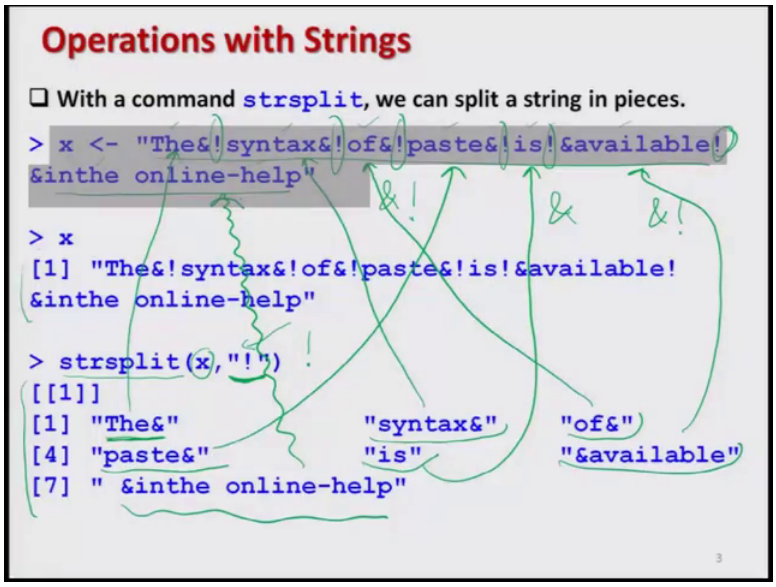


So, in order to split a string in R we have a command what is called as str and split; str means string and split is a split. So, this is string split and this command basically splits the element of a character vector. This can be a single character or this can be a character string. Both are possible the syntax for using string split command is that we try to write all these in a small letter strsplit, then inside the arguments we try to give the variable x

or that string which we want to split and we give it here say split and then we have another options here fixed, but again I would say that if you want to have some more information on this string split; please try to go to help menu and try to type on R console help and then str split you will get more information.

But here I will try to show you certain example and try to say illustrate how these operations can be done.

(Refer Slide Time: 02:35)



So, let me try to take here several examples to explain the use of string split function. So, you see here; I am trying to take here one say sentence from here to here; this sentence looks a little bit is change in the first shot, but if you try to see what I have written is this that the syntax of paste is available in the online help, but intentionally what I have done that I have inserted 2 characters one is here and another is exclamation sign and exclamation sign, this is here, this is here, this is here, this is here, this is here and this is also here.

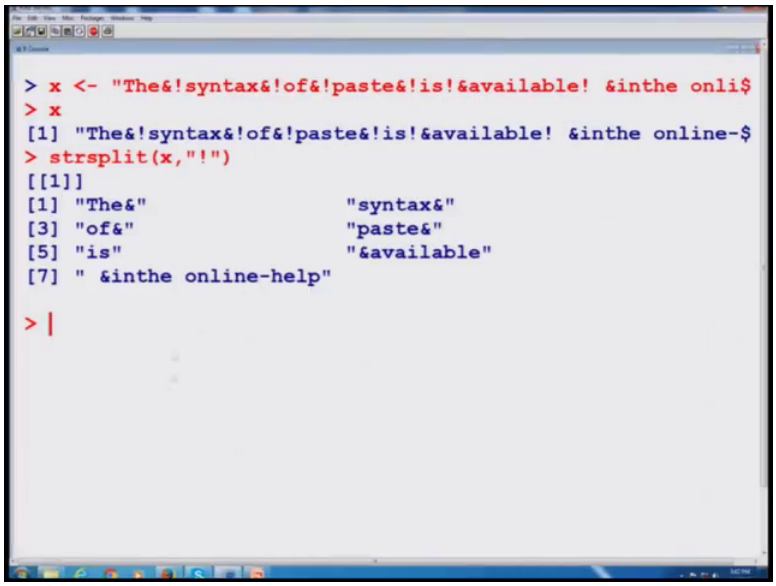
And my objective is that I want to split this sentence based on my request to split the characters using the sign and exclamation sign. So, if you try to see this is the outcome of here x we will try to do it in the R console to see the outcome, but before that let us try to analyze; what we want to do. Suppose I say please try to split the string; what is given inside this x using the separator this exclamation sign. So, now, if you try to see in this

case I try to remove this circle. So, that I can show you; how this split is happening, my separator here; here is this sign. So, I try to find out here where is this sign.

So, you can see here this sign is here, here, here, here, here and here and as soon as I give this command on the R console; I get this outcome and you can see here that this is the first thing; this is coming from here; this is the second split which is coming from here; this is the third split which is coming from here; this is the fourth split which is again coming from here this is the fifth split which is coming from here and this is the sixth split which is coming from here and this is the last split which is coming from here.

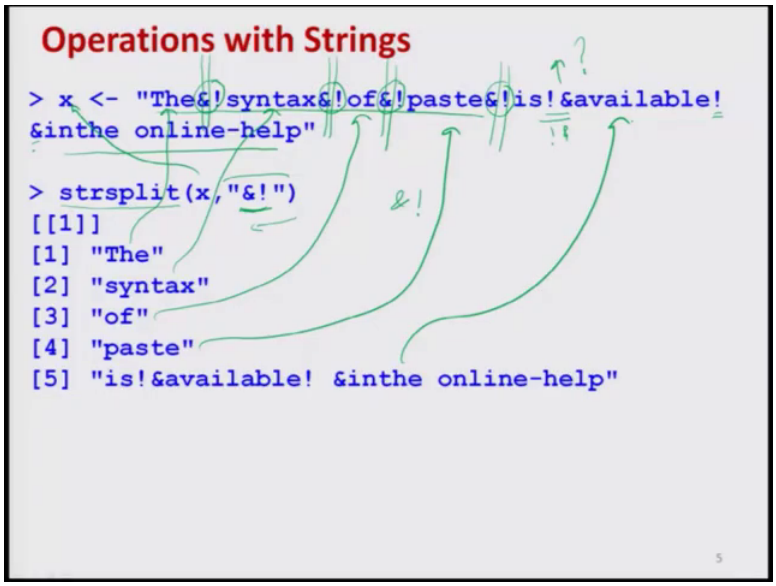
So, you can see now this split command has splitted means our statement in the required way and this is the screenshot of the outcome, but and definitely we would like to do it ourselves to see what is really happening.

(Refer Slide Time: 05:54)



So, this is my statement here you can see this is my statement here x and now I try to use this command see string is split x using the exclamation sign and we get here an outcome like this you can see here very clearly.

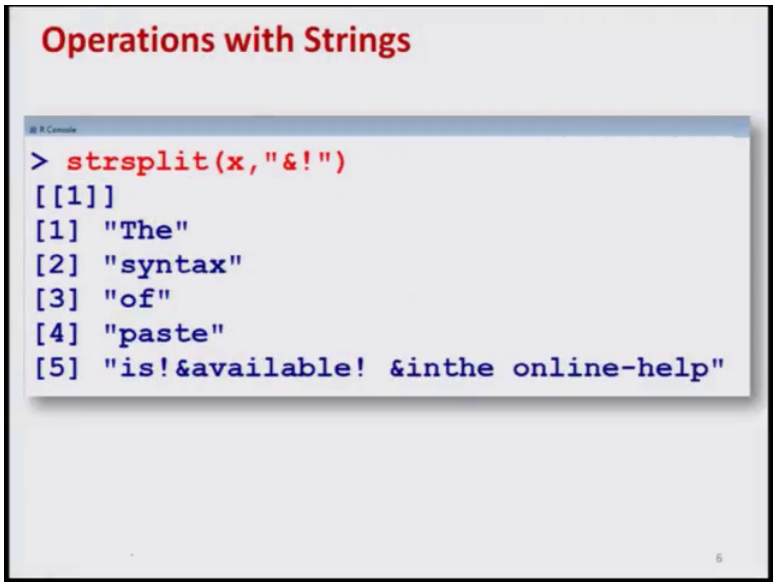
(Refer Slide Time: 06:10)



So, now we take another example using the same syntax, but we try to split the statement at different place and see what happens we consider the same example; what we had done earlier, but now I would like to split this statement at another point and I would like to split at symbol variable; I get an exclamation sign together; first let us try to see where this symbol is occurring in this sentence; this is here, this is here, this is here, this is here is it here, also no; this is opposite, this is exclamation sign and this is the point, this is the sign which we do not want and similarly here also.

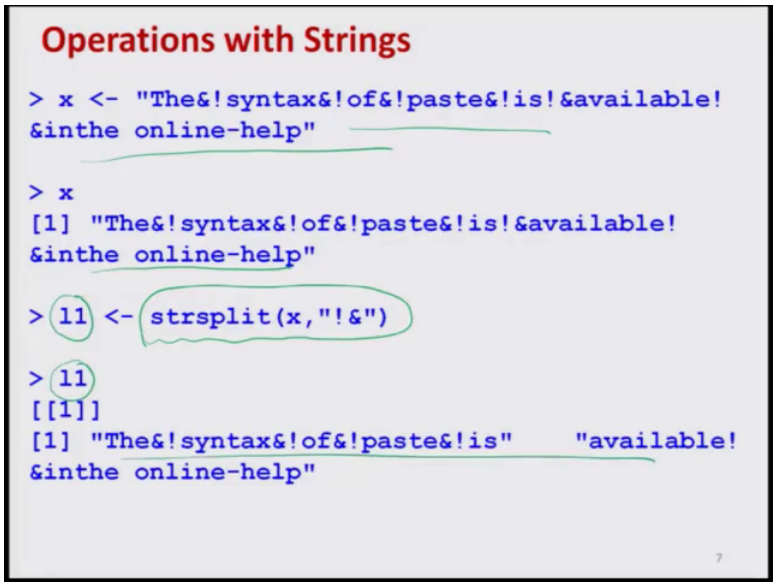
So, when I try to use the command string split of this vector here x using the separator like this one, then ideally; what do you expect the split should happen here, here, here and here and now this is here the outcome that you see here the first is occurring the which is coming from here second is coming from this, here syntax third is coming from here of fourth is coming from here fifth and all this is coming from the remaining part. So, you can see here by using different types of logical operators you can do it and similarly, if we would like to first see; what happens when we try to do it over the R console see like this and was your here x it was something like this.

(Refer Slide Time: 08:14)



Now, suppose I say; suppose I want to split at another symbol wherever is here e you can see here that this is happening. So, depending on the requirement you can just play with this and can get the required outcome now here is the screenshot of the same outcome.

(Refer Slide Time: 08:51)

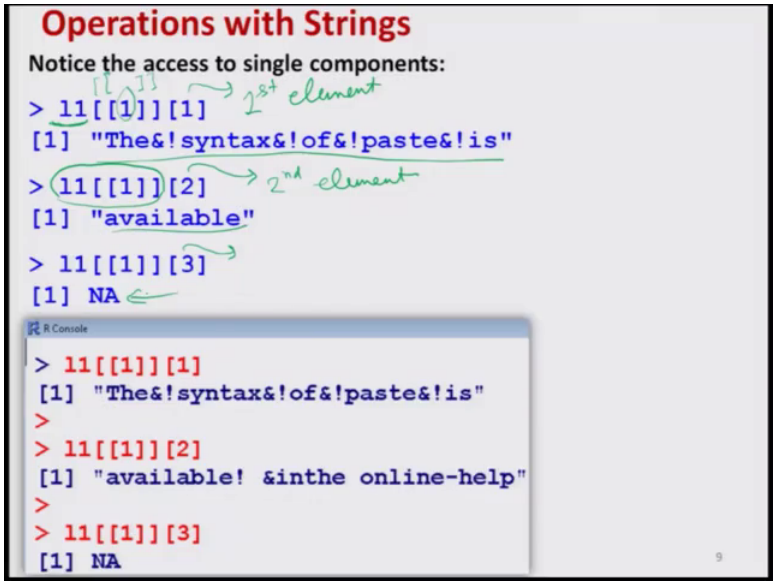


And now, we try to see in the same example that if I want to call a particular element after the string is splitted, then how to do it for example, we had a string that was saying the syntax of paste is available in the online help and we had splitted it at a symbol,

wherever we are observing the symbol and followed by say exclamation sign and then we had here different types of breakups.

Now, I want to call one of the element in those breakups one of the element of that splitted vector; how to get it done. So, if you try to see here this is the same thing what we had done and this is the same outcome and now I am saying here whatever is my here split whatever we have obtained this; I am trying to store inside a vector say l q right now I call see here l 1 and this is like this one right and now.

(Refer Slide Time: 10:11)

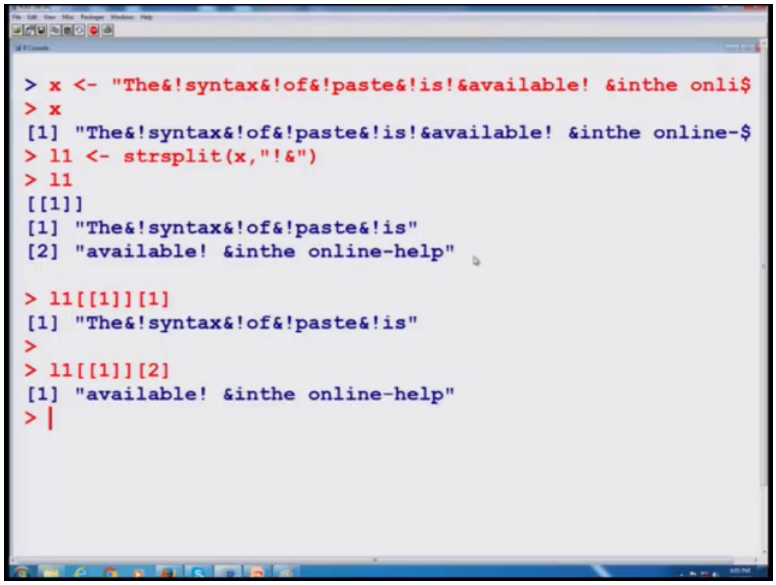


This is the screenshot and now I want to access a single component. So, what I try to do is very simple, but what you have to observe that what is the syntax and how I am using the square brackets. So, I write a down here say l 1 the vector in which I have saved the outcome after the splitting and then inside these 2 square brackets. These are 2 square brackets, then I try to write down the outcome address that we have seen that was 1, I will try to show you again in the R console and then I am calling the first element.

So, this is giving me the first element and the first element was like this one and similarly, when I want to call the second element of the same thing this part remains the same and I change this address here this is second element and second element was available and suppose I also want to have the third element; it is saying not available because there was no third element there were only 2 separations; let us try to see over the R console and see what happens with this thing. So, first let me try to have this

statement over here I try to store this thing first I have to copy this x because now this x is also changed.

(Refer Slide Time: 12:03)



So, you can see here this is my here x and then I am following this l 1 say spread split and the split is coming here as say l 1 and when I see here l 1 I can see here that there are only 2 breakups when I am trying to use this separator this exclamation sign and you see they were occurring here in that second part of the sentence right and now. So, I have here 2 values and in order to remember how to write down you see; I am just simply writing first this l 1, then this address and then if I want to have the first one then this address or the second address if I want the second address.

So, if you see this is very simple to write; I am simply copying from above l 1 then these 2 brackets I am simply copying and then I need suppose I want the first one. So, I try to copy this address and you see here I get the same thing and similarly, if I want to have here 2 the second value which is here this thing then I get the same thing right. Now let us come back to our slide and see one thing what you have to keep in mind that here in this case, I had changed my separator and the separator here is this one earlier separator was say nth and exclamation sign that is why earlier you are getting here 5 see inputs.

But when I when you are trying to split it with this thing, then this is occurring only at 2 play and this separator is appearing only at one place here that is why there is a split only at this place and that is why we are getting here only 2 things; one is this one and say

another is this one right and this you can see this is the outcome the screenshot and that is why when I was saying that I want to have here the third outcome this was telling me not available because this is really not available because there are only 2 splits; how it can give you the third split when it is not available.

So, these are some say small points which now you have to keep in mind and try to understand this minor details and this minor details will help you in getting a better outcome according to the need and wish. So, now, I would try to stop here and in the next lecture, we will try to do something else some more features of strings; till then goodbye.