Science Communication: Research Productivity and Data Analytics using Open Source Software

Mohit Garg

Central Library

IIT Delhi

Week: 04

Lecture 11 : Installation of R

Dear Learners, Welcome again to this fourth week of the course. I hope you are enjoying the course. In the previous week we have discussed different data sources, where we have extracted the data from different sources. I hope you are able to extract data from some of the data sources. Now we have the dataset with us, we have to do the analysis. So, to do the analysis on this data we need a software.

So, for that purpose we will be primarily using the software R and its different packages. We will also discuss some of the other visualization tools in the upcoming lectures. So, this week we will be learning about R aiming for the following outcomes.

First we will start with understanding the fundamental principles of R that includes the basic syntax and structure of R. So, what are the different rules and the conventions we have to follow to use them in the R environment. So, next we will discuss the functionality of RStudio IDE, what exactly IDE is and why we need the IDE for doing the analysis using R.

Then we will know what are the different data types in R such as numerics, characters, or logical. After that we will understand various objects in R including vectors, dataframe, matrices, lists and arrays and different manipulations on these. In the end, we will understand how to use the packages in the R, what exactly that package is and how to install those packages. Then how to call those packages for the analysis. and we will also see how to debug the issue in R independently. how you can use the help feature in R for solving many of the problems that you will be facing during the analysis.

So, I understand that some of you may be worried about learning R. So, it is perfectly normal to have concerns when you are stepping into learning new things. I too faced many of the issues when I started learning R. So, do not worry I will be covering each

part from the basics with lots of examples. So, you can also try the same thing in your machines. So, let us start the first lecture of this week where we will be learning about R.

R is a popular environment used by academicians, scholars, industry professionals for statistical computation and graphical visualization. It was created by two statisticians Rose Ihaka and Robert Gentleman at the University of Auckland, New Zealand in the early 1990s. But the stable version was released in 2000. It is an open source software and it is freely available for downloading. After the downloading there is no cost involved in using the software. One of the key reasons for its popularity is its extensibility with the support of package systems.

So, it has a large number of task specific packages that you can use for your analysis. A package is just a collection of code written for the specific goal. We will be learning about packages in more detail in the coming lectures. The sustainability of any open source project depends on the active participation and the engagement of the community. So, R has a big and active community of researchers, statisticians, professors, data scientists and developers having common interests and interacting with each other to share knowledge for the development of R and its packages.

So, this fosters collaboration and innovation in the R community. R is a platform agnostic software. that means it is a platform independent software. You can use the R either in the Windows or in the Mac or in the Ubuntu or Linux based machines. And the beauty is that the core functionality of the syntax of R is exact and consistent in all the operating systems. So, say for example if you are calculating the average in R that is installed in Windows and then you are doing the same computation in R installed in Mac and then you are doing the same thing in the R that is installed in Linux based machine. So, it will be the same thing. You don't have to worry about what platform you have. So, the whole syntax is the same.

So, however there are slight differences in system specific commands like say for example Windows uses backslashes for the file path. But Linux based machines use the forward-slashes. So, that kind of difference you will see in using the command. But other than these the syntax will be the same in all the three machines. So, don't worry if you don't have any idea about what system specific commands are and what exactly the path is. We will be discussing these things in detail when we will be discussing how to set the directory and how to change the directory in the R.

Following are the system requirements for installation and use of R software. So, there is no specific requirement for using the R software. It will run smoothly on any modern computing system. To install and use R software you need a computing system either a desktop or a laptop. R is available for all three types of operating system Windows, Linux, Mac.

So, you need not to worry about the operating system. Whatever operating system you have, it is available. Then installation of R doesn't require much space. The full installation is in a few hundreds of MB. In the context of this course any machine that has 4 GB of RAM is well sufficient for doing the analysis. Generally a high space of system configuration is required for computing for doing the large scale data analysis.

So, Internet will only be required during the installation of R and then installation of packages or the required dataset. After the installation of the packages and the required dataset there is no such requirement of the Internet. So, you can say that the Internet is only a one time requirement for installing the package. After that you can do the analysis without having the Internet.

So, the first version was launched in 2000 and after that at regular intervals there are many versions that have been released by the R community. and this is particular you will be required when for example you are using some packages that are compatible with some specific version. So, it is always preferable to use the latest stable version for doing the analysis.

So, let's start the installation of R. So, to install the R you have to open the URL it is www.cran.r-project.org in the browser and select the appropriate link based on the operating system you are currently using. So, this CRAN is a kind of repository that hosts R and its packages along with the documentation. All the source code of R along with the documentation and its different versions are released on this CRAN site which is known as the Comprehensive R Archive Network.

So, if you have a Linux based machine like Debian, Fedora, Red Hat, Ubuntu click here by selecting the appropriate machine type and follow the command listed on these bases. So, say for example if you have Ubuntu based operating system in your machine. So, you just click here. So, after that you open the terminal in the machine, and then follow these instructions and after that the R will be installed in that machine. So, if you have a Mac operating system, then click on this Mac URL and download the installer and after downloading run the installer and follow the guideline.

This is how you can install the R in the Mac operating system. If you have a Windows operating system then you have to click on this URL download R for Windows. and if you click on this. So, after this you have to click here install R for the first time. As soon as you click here this file will be downloaded and it will be saved in your download directory. So, for me like say for example this is the download directory and this is the file is downloaded. So, once the file is downloaded you run this file with the administrator right and install this R application with the default setting.

Once you finish these all processes a desktop icon will be created. Also, you can search the R in the Windows applications. So, if you search here like say for example here I will

just click here. So, yeah it is now in my machine. So, this is how you can install R in your machine. So, now I am assuming that you all have installed R in your system. Still if you have any issue, or if you are not able to install R in your machine please let us know in the discussion forum. We will help you in installing the R.

If you see the interface here. So, this is the default interface of R and here you will find that this particular 4.3.2 is the version, then this is the date this is the 31st October 2023. So, this is the version date. then you will find this 'Eye-holes'. 'Eye-holes' is a version name. So, this is a common practice in many of the open source projects to give some attractive names for each of the versions for making each version as a distinctive identity.

So, if you see here in the same paragraph copyright details are given. So, this particular version has a copyright 2023 and this shows that this is for 64 bit. So, in the next para it shows that R is a free software. Of course as I said that R is a free software you can see on screen also that R is a free software. But this software is associated under some license. So, you can check that license using the command license. So, we will see what exactly that license is. So, license if we see that this license of using R is distributed under the terms of GNU General Public License, either version 2 or version 3.

If you want to know more details about the license that R is using, you can check on this URL. So, on this URL you can go and check the license. Then also this is a GNU general public license version 2.0 or so. And in the last if you will see in the license, it is quoted that "Share and Enjoy". So, you use this software freely without worrying about any kind of cost. Now if you see in the next para, it is written that natural language support. But running in an English locale.

So, language support is there. But this particular thing is in English. Then if you see that R is a collaborative project with many contributors. Yes, it is a community driven software as I mentioned and for any open source project the community is more important. So, this particular software is also developed by various contributors. If you want to see what all are the contributors for developing this software you can use this function.

So, we will see who all could contribute to this particular software. So, we'll just see contributors. As soon as you run this command a new tab will open and it will provide the details of all the contributors. So, here the details of the contributors are given. And it is shown here that R was initially written by Robert Gentlemen and Rose Ihaka at the University of Oakland. And after that it has a large group which consists of many members. So, here all the contributors of R have given. So, if you see this particular function citation. The citation function later will help in letting you know how to cite this particular software. If you are using this software, how you can cite this software in your research paper, or you're in or the thesis or whatever the document it is.

So, if you run this command citation and it is showing that if you are citing R in publication you have to use this particular reference that R core team 2023. If you see that the software is community driven. So, there is no one particular author name there, it's the R core team. So, the R core team means all contributors of R. Then this whole details are given the title and all those things. A BibTex entry is also provided by those who use the latex. So, a BibTex entry is also provided for the latex user.

So, this is how you can cite the R in your publications or the thesis document. and like whenever you will be doing the analysis you will not be only using R you will be using different packages also. So, to see the citation of that package you can use the package name in the citation function. We will see that thing also when we will be using the packages. But this is a recommendation that whenever you are using R and any of its packages always cite those things in your publications.

After that if you see on this that the next para is about type demo for some demo and help. So, R comes with some of the default demos. So, that will help you to execute the commands in R. So, if you want to see the demo you can use the function demo. So, say for example. So, you can try this like a demo. So, once you run the command demo it will open this new tab and this is how you can take advantage of the demo.

Then there is another function if you see here is Help and this particular function will be very helpful for whenever you will be using the R. because many times like whenever you are installing the package or you are stuck somewhere. So, this particular function will help you in solving the problem. So, if you run this Help as soon as you run this function, a page is open on the localhost and all those things are given. So, this page is basically documentation. So, the whole documentation of R. if you want to know any of the commands and their arguments you can see here. Then if you see here the next thing is this quit. So, for quitting or for closing this R you can use the function this Q. So, whenever you run the command it will pop up whether you want to save this workspace or not whatever the work you have done it will be saved. And if you click on no it would not be saved.

So, it's a good practice that whenever you are closing a particular session in R always save that session. So, if you want to see in future what analysis you have done, what are the commands you have used. So, it will help you in knowing the previous thing also. Then of course you can quit the R by clicking on simply this closed tab and again it will ask you whether you want to save workspace or not. But one thing if you have noticed that first in this case when we have explicitly mentioned the command Q() and then we have closed the application like we close other applications. So, here automatically commands come. So, whatever the operations we will be doing in R it's a command base. So, it's a good practice that whatever operation you do you do it through command. So, this will help you in using this software very efficiently.

Now we will see the menu bar of R. It has like these seven options on this menu bar of R. So, we will just see what exactly. So, this menu bar has the functionalities, which helps in interacting with the R. So, say for example this file menu has these many options. So, like if you want to create a new script you can click on here like if you click here this editor will be opened or if you have an existing script and you want to open it. So, you can do that then functionality helps in loading or saving the workspace. then like that we can also load the history whatever the commands that we have used previously or we want to save the history. we can do that by clicking here as soon as we click on save it will like this window will be open. and then we can name the file and that history will be saved. As a beginner I will recommend that whenever you are running any command just save the history. So, that you will also know that all commands you have previously you have used then you see here that this is the functionality which helps you in changing the directory. So, say for example by default setting it's in the document directory. But you can change the directory to any folder. Okay. So, we will see exactly which directory is. Okay. But you can change the directory whatever you want. Okay. say for example now presently this is a document. But you want to change to downloads. Okay. So, we can do that by clicking here and now the directory will be changed. Okay. We click here now if you see R directories downloads. Okay. And how this changing directory helps we'll see in the upcoming lectures. Okay. Then it has the functionality of printing this. So, if you click here the whole thing can be printed then you can save it to a file. And of course this is the option to exit the software. Then in the second menu it has multiple options like copying the code or pasting it. So, these three are important functions. So, first is to clear the console. So, if you click clear here. So, this whole console is clear. So, what are the details there, all cleaned up.

Now if you see, the next functionality is data editor. So, if you have a dataset and if you want to edit the dataset you can use this particular option. So, using this data editor option you can edit the data frame. So, for that you have to name the data frame and then a new tab will open. and then you can edit the dataset. This is an important option to change how the GUI will look like. So, this is a graphical user interface of R. So, if you click here it will show multiple options. So, you can change the font style then you can change the color like say for example if you this is a background if you do it to say for example antique white. Okay. And then it's a normal text to maybe. So, gray. Okay. So, I am changing normal text to gray then user text by default is red let's change this to purple. Okay.

So, see this is how this is changed and this is how the color and all those things can be done by using this particular option. then the third menu is view if you click here. So, here it will show that this particular version is. So, this is known as the status bar and if you unclick this. So, this whole toolbar will be gone. Okay. So, this will basically help to add the toolbar in the interface. So, these are the options available in the toolbar. So, the

toolbar opens the script. But we have an option in file then it has an osption of loading workspace. This is again a functionality of the file menu then saving the workspace again. like instead of going to file and then doing the things you can just simply click on here and you can either open the script or save the workspace. then we have copy paste and then if you want to do copy and paste. then we have an option to stop the computation. So, for example you if you have a large dataset and you want to and it's running and it's not stopping and you want to stop that particular computation. So, if you click here it will be stopped then in the last there is an option to print the details here. Okay. then in the fourth menu. So, these are the different functionalities provided here. So, what exactly is important here is this word completion and the file name completion. So, what exactly is it? So, say for example by default these will be checked in down. Okay. But if you like to say for example if you do like this. and if you write anything like we have used the contributor command. Okay. So, it will not complete that thing. Okay. So, what I am doing is writing the country and then pressing the tab. But it's not showing anything. But if you do this and now if you press the tab it will automatically show some suggestions. Okay. So, it is showing that there are two commands available with the which are starting with contrib. So, if I write contribute it will automatically complete contributors. Okay. So, this is how it helps in completing the command. Okay.

We will be seeing throughout this week and whenever we will be using the area that is very important for doing the analysis. Okay. then we have like this important functionality of listing the object whatever things you are doing in R it considers it as an object. So, what are the variables and what are the data you have created. So, if you want to list all the objects you can just simply click here. So, it is showing that these many objects are there. then if you want to remove all objects you can just simply click here it will ask first whether you really want to remove all the objects. So, I am saying no.

now if you want to see the search path. So, this is how you can check the search path of the packages. So, this is how you can list the search path now the fifth option is packages and packages are the packages are very important in R. So, there are various packages available for doing the analysis. So, you can just load the package here then you can set the CRAN mirror. So, CRAN mirrors are basically a kind of site from where you want to install the packages. So, you can select any of these mirrors and then you put an. Okay. and then after that you can select the repository also like this. This CRAN is the repository. There are other repositories that are also there for installing the packages. Okay. So, from what repository you want to install the package. So, this is how you can select the repository and then you can install the package by clicking here. So, once you click on installed packages this tab will open and these are the like also for the first thing to install the package is selecting the mirror what I discussed there. So, say for example if I do select the by default one. Okay. then it will ask what packages you want to install. Okay. So, these are the different packages there. But as per your requirement, what are

the analysis you are doing you just click in down and that package will be installed. Then we have this menu of what the interface will look like. whether it is in the cascade or the original say for example if I put it horizontally. it will look like this and if I do cascading. So, it will look like this. Okay. and in the end this is the Help menu which is very useful throughout the course and whenever you are using it you will be required to help a lot. So, it provides some frequently asked questions on the R website. and then it has some manuals also. if you want to see the manuals in R. you can that is also provided here. So, for that you have to click here. So, this is the manual and introduction to R and this is the whole content that will be useful for you what exactly R environment is and all those things are there. you will be surprised to know that you haven't downloaded this particular manual or thing. Okay. So, all those things that come with default whenever you have installed are. Okay.

So, this will be a very handy tool for you whenever you will be using the art and we will be discussing in detail how help functions can be a useful assistant for debugging yourself for solving any of the issues whenever you are stuck to any of the points.

Okay. So, this is all about the interface of R, what exactly the different menus are there, what are the different functionality you can use for changing either the interface or you want to save the workspace, or you want to install the packages.

Okay. So, now we will move to the first command in R. So, say for example I am using this 2+2. as soon as I hit the enter it gives me the output 4. So, that means it is working fine. So, this is how you can use other automatic operations in R. and please do try other arithmetic operations like subtraction division multiplication with some values. and please do share with us what all two numbers you have selected for doing the multiplication, subtraction and division. Please post all your responses in the discussion forums. So, this is the end of lecture 1 on working with R. I hope you have got the understanding of how to install the R and then what are the different options available in the R interface itself which you can use during the analysis. So, we will meet in the next lecture to extend our journey of learning R. Thank you.