

Dealing with Materials Data
Collection, Analysis and Interpretation
Professor MP Gururajan
Department of Metallurgical Engineering and Materials Science
Indian Institute of Technology, Bombay
Lecture 08
Module: Introduction to R

Welcome, this is a course on dealing with materials data, we are going to discuss all aspects of dealing with materials data. We are going to talk about collection of data analysis of data and interpretation of data, even though not necessarily in that order. This course has 2 parts. One is the statistical concepts associated with data analysis and interpretation.

The other one is hands on part which we will do using the R programming language. So, for this course I will be teaching the R part of the course. So, we will take as much as possible data from material science and engineering. And we will try to deal with that using the R programming language and the concepts that are needed to deal with data will be taught by the other instructor Professor Hina Gokhele. So, let us start with the first module.

(Refer Slide Time: 01:22)

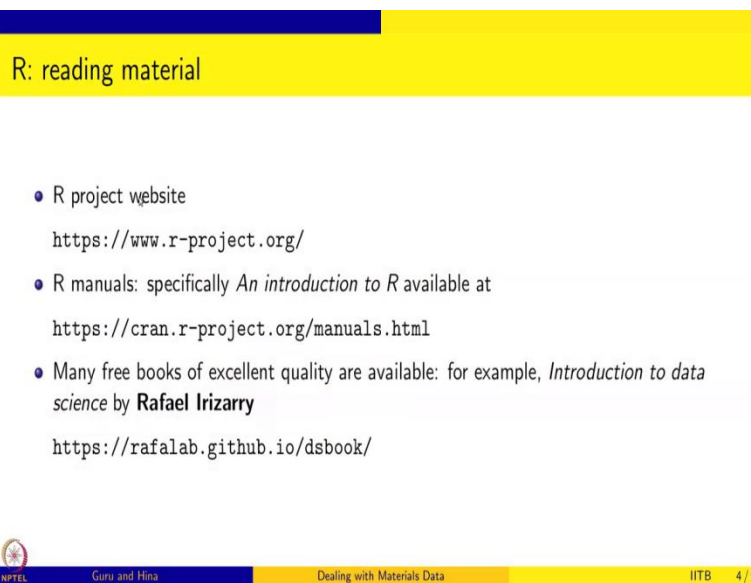
Objectives

- 1 Introduce the R programming language
- 2 Give a hands-on, tutorial introduction to R
- 3 Examples: using materials data (as far as possible)

NPTEL Guru and Hina Dealing with Materials Data IITB 3 / 9


This module is an introduction to R. So, in this module we want to introduce the R programming language and we want to give a hands on tutorial introduction to R and for examples we will use materials data as far as possible. So, there are plenty of reading material for R that is available.

(Refer Slide Time: 01:47)



R: reading material

- R project website
<https://www.r-project.org/>
- R manuals: specifically *An introduction to R* available at
<https://cran.r-project.org/manuals.html>
- Many free books of excellent quality are available: for example, *Introduction to data science* by **Rafael Irizarry**
<https://rafalab.github.io/dsbook/>

 NPTEL Guru and Hina Dealing with Materials Data IITB 4/9

The first one is the R project website <https://www.r-project.org/> this website has plenty of information about the R project itself. And the R manuals are available at the <https://cran.r-project.org/manuals.html> page.

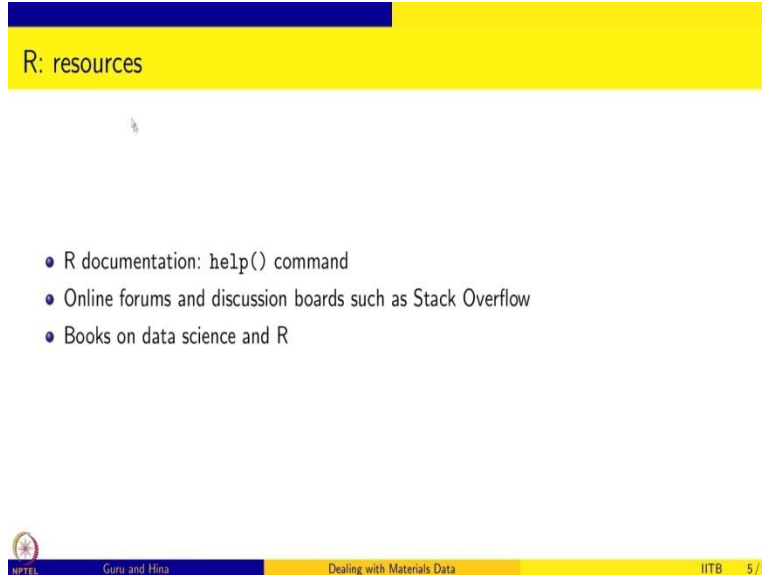
There are many references and some of them are very exhaustive, some of them are meant for developers. But what I have in mind specifically is an introduction to R and this is available at the <https://cran.r-project.org/manuals.html>. There are also many free books of excellent quality that are available if you go to the manuals dot html page for example, there are also user generated manuals there is a link which you can go and look at the material that is available.

And there are many people who have written small introductions are specific things how to get it done using R and this information is available and some of them are of excellent quality and they are also freely available. For example, there is a book by Rafael Irizarry called introduction to data science and you can download this book for free.

If you feel like you can also pay the author for taking the pains for writing this book. But if it is not affordable for you it is not in your budget at this moment, you can freely download this book and use. So there are plenty of reading material that is available online. And some of them are from authentic sources like the R project website or R manuals. So I strongly recommend that you

try to use some of these reading material in addition to the tutorial introduction that you are going to get from this course from these modules.

(Refer Slide Time: 03:41)



R: resources

- R documentation: `help()` command
- Online forums and discussion boards such as Stack Overflow
- Books on data science and R

NPTEL Guru and Hina Dealing with Materials Data IITB 5/9

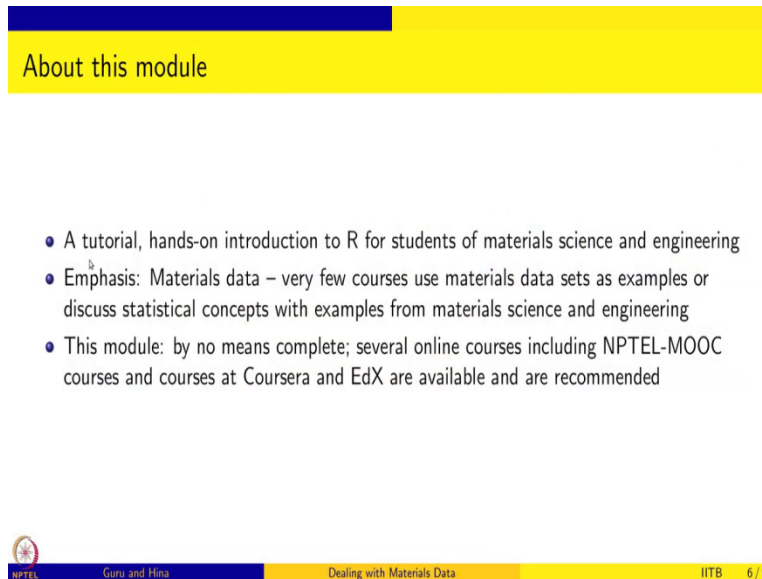
There are also plenty of other R resources. R documentation is one of the best resources that is available and that is always available. So in the R console, you can just look for `help()` and you can get the required help and I am going to show you how to do that in this module and there are lots of online forums and discussion boards such as Stack Overflow, where lots of details are discussed. Of course, these online forums and discussion boards are useful, once you have some familiarity with R and you can understand the solution that is given are the kind of questions that are being asked.

So, in order to understand you may have to have some familiarity with that, if you have then these online forums and discussion boards are very useful, and they can answer specific questions or specific problems that you have, even if it is not part of this course or even if it is not part of any course, then you can actually go find some help and using that you will be able to solve your problems. And of course, there are lots of books on data science and R and we are not going to use any of them in this course.

But those of you who are more data science minded and are interested in applying these concepts to material science and engineering should take a look at some of these books also. So, so, having

said that, there are plenty of resources and the reading material that is available, what is it that we want to do?

(Refer Slide Time: 05:14)



About this module

- A tutorial, hands-on introduction to R for students of materials science and engineering
- Emphasis: Materials data – very few courses use materials data sets as examples or discuss statistical concepts with examples from materials science and engineering
- This module: by no means complete; several online courses including NPTEL-MOOC courses and courses at Coursera and EdX are available and are recommended

NPTEL Guru and Nima Dealing with Materials Data IITB 6 / 9

So in this module, I want to give a tutorial hands on introduction to R and I am having students of material science and engineering in mind. This is because if you go to the R manuals, for example, and the user generated manuals if you look at, many of these manuals are prepared for specific fields.

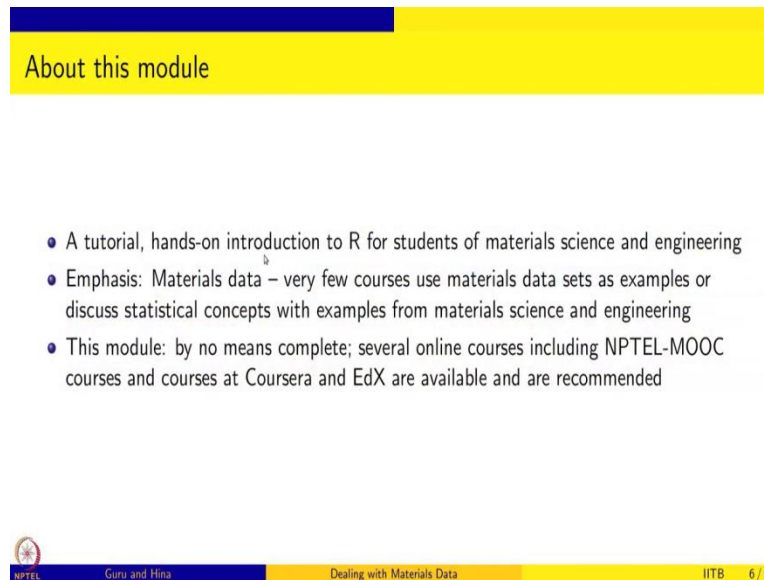
For example, there are people who have looked at social sciences or geography or biology and how to solve problems in these areas in terms of analyzing the data, and interpreting the data and presenting the data and graphically describing the data and things like that, so for that the tutorials are returned, but there is nothing that uses material science and engineering problems as the example case.

So that is the first aim of this module to give a tutorial hands on introduction, its a tutorial introduction because it is no means complete introduction to R there are like I said lots of reading material and other material that is available and so you should take help of those material.

This will just give you whatever is required to solve the problems that we have at this point and how to do it using R so it will be hands on so as much as possible. When I am teaching the course

I will open an R console or R studio window and I will type the commands so that you can also have your laptop or computer with R open and you can also work along with these tutorials. So you will have a hands on experience with R.

(Refer Slide Time: 06:53)



About this module

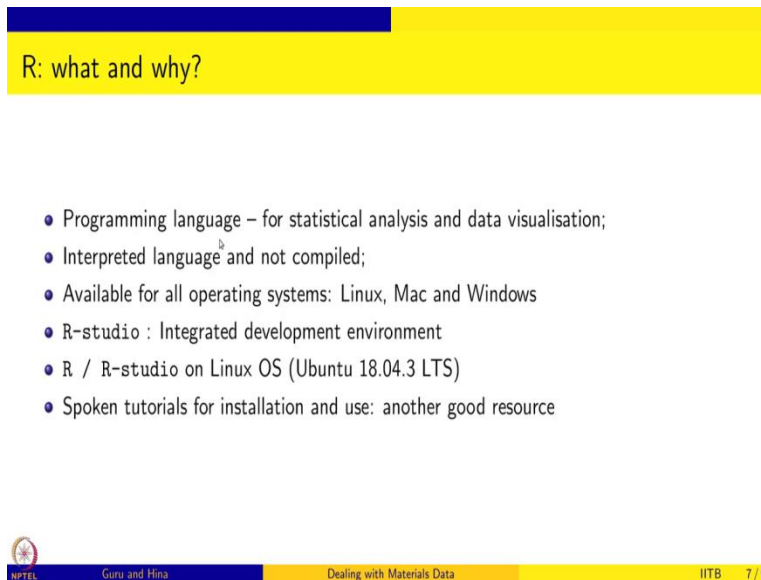
- A tutorial, hands-on introduction to R for students of materials science and engineering
- Emphasis: Materials data – very few courses use materials data sets as examples or discuss statistical concepts with examples from materials science and engineering
- This module: by no means complete; several online courses including NPTEL-MOOC courses and courses at Coursera and EdX are available and are recommended

NPTEL Guru and Hina Dealing with Materials Data IITB 6 / 9

The emphasis is on materials data. There are very few courses that use materials data sets as examples are discussed the statistical concepts with examples from material science and engineering. So, we are going to emphasis on this part and this is also lots of value addition, because to collect these data and to put them in one place and use them as example, will also help you analyze your own data when you generate, because you will be familiar with the similar data and similar kind of analysis that you can do on them. So, the emphasis is A on tutorial hands on introduction B on using materials data. Like I said this module is by no means complete.

There are several online courses including NPTEL MOOC courses, there are 2 or 3 of them from IIT Bombay itself from IIT Kanpur and IIT Madras. So, it will be useful for some of you to for interested in learning more about R to go to some of these courses. In addition, of course there are also courses at Coursera and EdX that are available. And I strongly recommend that as you are learning R in this course, in a tutorial hands on way, you should also try to do some of these other courses and bring in some of the expertise and knowledge on R that you gain from these courses to your own data analysis for material science and engineering.

(Refer Slide Time: 08:29)



R: what and why?

- Programming language – for statistical analysis and data visualisation;
- Interpreted language and not compiled;
- Available for all operating systems: Linux, Mac and Windows
- R-studio : Integrated development environment
- R / R-studio on Linux OS (Ubuntu 18.04.3 LTS)
- Spoken tutorials for installation and use: another good resource

NPTEL Guru and Hina Dealing with Materials Data IITB 7 / 9

So, what is R and why are we using R? R is a programming language specifically meant for statistical analysis and data visualization and this is the part I want to emphasize. Because you will find that programming languages like Python also, for example can do most of what we are doing, but our emphasis is on statistical analysis and interpretation and visualization of data. So R is a programming language which is specifically meant for such statistical analysis.

So we are going to use R its an interpreted language and not compiled by that what we mean is that you can open an R console and you can give commands and you will immediately get the reply or the computation done and the answer given to you. So there is no need like if you are writing a program in C programming language, for example you have to compile get an executable, and that executable has to run to give you the answer that you are seeking. But that is not needed.

In this sense, R is more like GNU octave, or MATLAB if you have used or Scilab there are several such languages that are available and so they are easy to work with and they are also very powerful as you would see. Another advantage of R that you will see is that it is available for all operating systems, specifically for Linux, Mac and Windows.

In this course, I am going to use Linux operating system. But if you have a Mac or Windows you will be easily able to use R on these operating systems also, there is a nice integrated development environment IDE called R studio. And so I am going to be using R studio also in this course. So I am going to use both R console and R studio to give you a flavor of both and to tell you how to

deal with about this of which R studio is more powerful and more complete. It has several pains, I will show it to you. And R is just the console and so it is minimalist, but it is also equally powerful, so it can do things that that R studio can do.

And for example, using R studio is how I have prepared these slides. So R studio can work with things like Latex and produced presentations. And I am also preparing documentation for what I am teaching using R studio. So you will see the documents also prepared using tech. And all of this is done using R studio.

So it is a really integrated development environment. So you can use it as a text editor, you can use it to prepare documents, you can use it to prepare the scripts, you can use it to prepare the script and document together. And then you can decide to use only the script or only generate the document.

So all sorts of possibilities are there. And so you will see some examples of that. So we are not going to spend too much time on how to do that. But like I said, there are enough materials once you are familiar with this for you to go and explore and learn. So specifically in this course, we are going to use R and R studio on the Linux operating system specifically I am using Ubuntu 18 point 04 long term support version there is one more resource which is very useful. This is called spoken tutorials and if you look up online for spoken tutorials for R this is also maintained by somebody from IIT Bombay Professor Khangan (()) (12:19) and his group.

So you will find that, how to install R for example, and how to use and these are spoken tutorials in the sense that you will see instructions being given to you. And so I strongly recommend that you also utilize this resource look it up and learn more about R studio. The other aspects like installation, for example, which is not part of what we are going to do. I am going to assume that you have R, R studio installed on your operating system, and that you are a little bit familiar with how to work with them. And so I am going to just start. So what is the first thing we want to do?

(Refer Slide Time: 13:10)

Alice in wonderland

The White Rabbit put on his spectacles. "Where shall I begin, please your Majesty?" he asked. "Begin at the beginning," the King said gravely, "and go on till you come to the end: then stop."



NPTEL

Guru and Hina

Dealing with Materials Data

IITB 8 / 9

The A to Z of an R session!!

- Create a directory
- Invoke R from the directory
- Check the version of R
- Write Hello, World in R
- Quit R.



NPTEL

Guru and Hina

Dealing with Materials Data

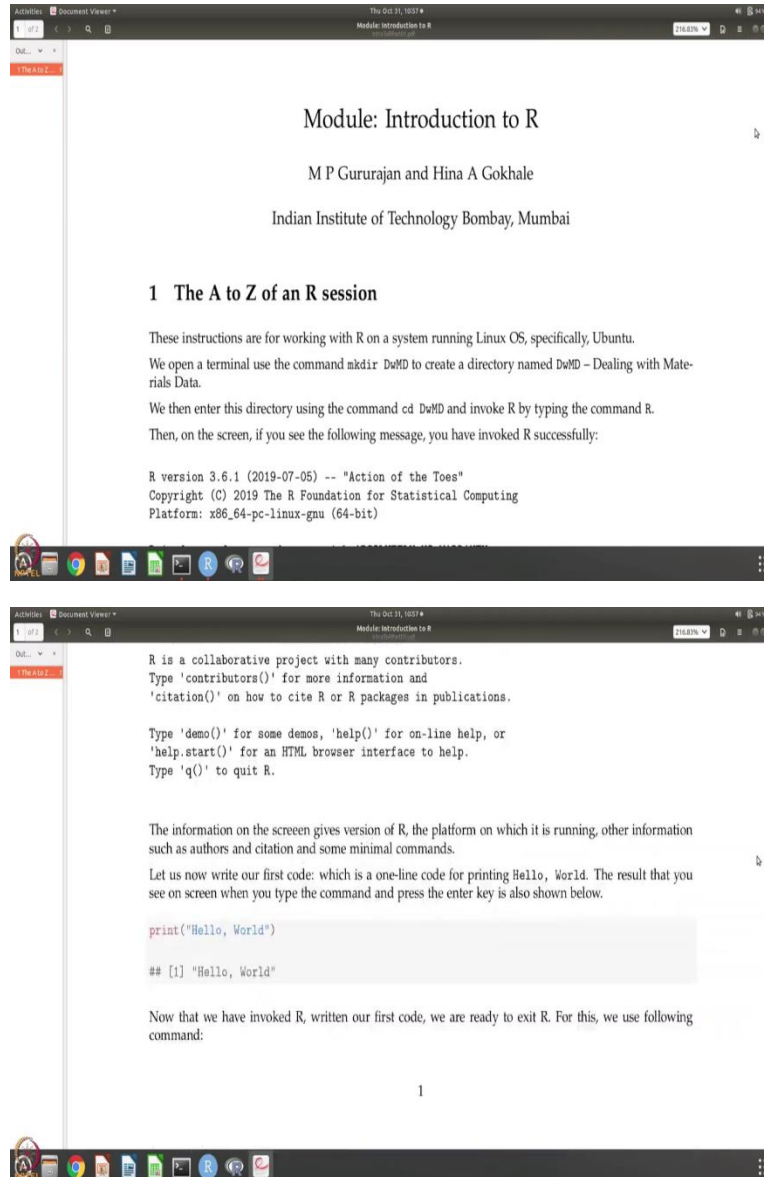
IITB 9 / 9

So here is a quote from Alice in wonderland the white rabbit put on his spectacles. Where shall I begin, please your majesty, he asked, begin at the beginning, the king said gravely, and go on till you come to the end, then stop. So what does that mean? We want to have the first session, I want to call it A to Z of an R session. So we want to create a directory, we want to invoke R from that directory, we want to check the version of R and we want to write the first program which is the hello world program. And then we want to quit R, right.

So if you can go to a directory open R do something close and come out. So that is the first session that we want to do. And that session has a beginning how to invoke R and it has something in the

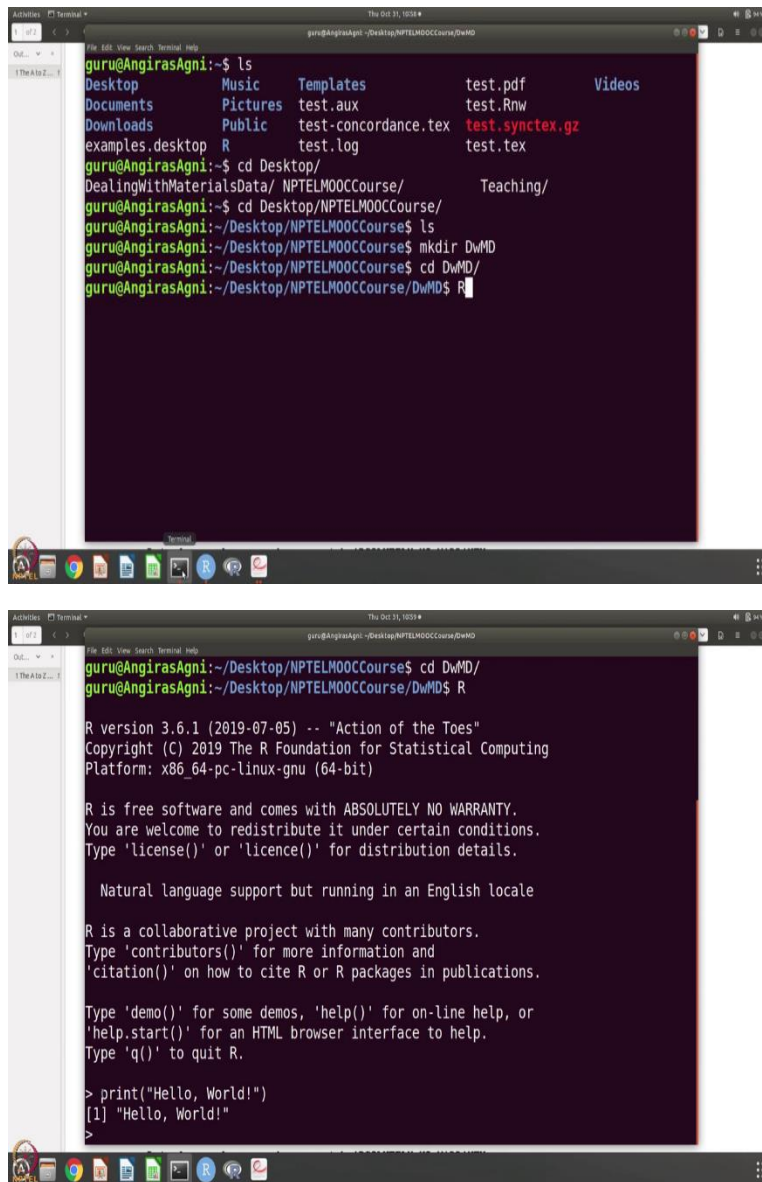
middle how to write your first program. And it also tells you how to end the R session. So this is the very first introductory session I want to have.

(Refer Slide Time: 14:12)



So let us do that. And like I said, have also prepared the notes for these sessions and they are written like this and like I said, they are prepared using the R itself using R studio. So this will be available to you and the MOOC website. But now what we are going to do, so we are going to start doing this.

(Refer Slide Time: 14:37)



The image shows two screenshots of a Linux terminal window. The top screenshot shows the user navigating through the file system. The bottom screenshot shows the user running the R command, which displays the R version and some introductory text.

```
guru@AngirasAgni:~$ ls
Desktop      Music      Templates  test.pdf   Videos
Documents   Pictures  test.aux   test.Rnw
Downloads    Public    test-concordance.tex test.synctex.gz
examples_desktop R         test.log   test.tex

guru@AngirasAgni:~$ cd Desktop/
DealingWithMaterialsData/ NPTELMOOCcourse/ Teaching/
guru@AngirasAgni:~$ cd Desktop/NPTELMOOCcourse/
guru@AngirasAgni:~/Desktop/NPTELMOOCcourse$ ls
guru@AngirasAgni:~/Desktop/NPTELMOOCcourse$ mkdir DwMD
guru@AngirasAgni:~/Desktop/NPTELMOOCcourse$ cd DwMD/
guru@AngirasAgni:~/Desktop/NPTELMOOCcourse/DwMD$ R

R version 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

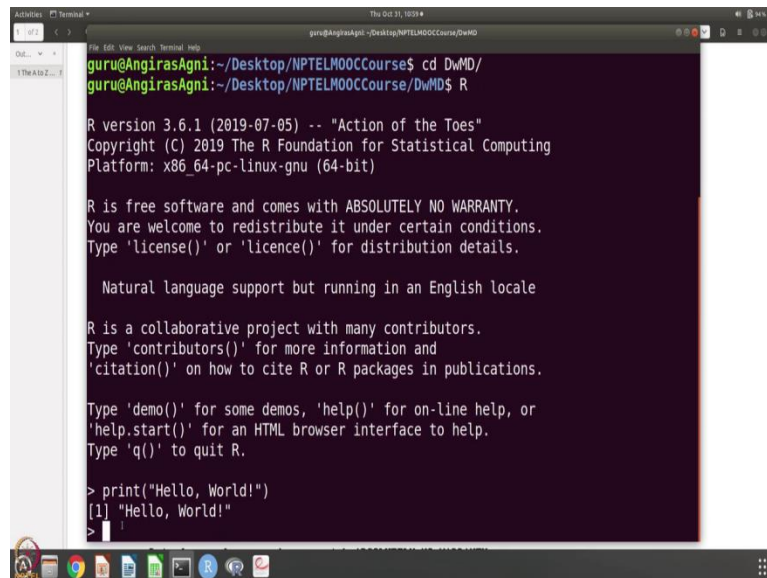
> print("Hello, World!")
[1] "Hello, World!"
```

So what is the first thing we want to do? We want to prepare a directory right? So I want to go to, so I want to make a directory called dealing with materials data. So this is the command in Linux to make directory mkdir, of course if you are using windows machine or even on Linux, if you are on the X windows, you can right click prepare a new folder and name it as dw md and so on.

So let us go to this directory and we want to invoke R from this. So invoking R are getting on R consoles means simply typing R and entering. And when you enter, you see that this is the R version that I am using 3 point 6 point 1. So I strongly recommend that you also use this version so that there will be compatibility in terms of versions what I am doing, and what you will see and

this version of 3 point 6 point 1 is called action of the toes. And so it tells you something about R and we are ready. So we want to write the first program and the first program is of course to print hello world. So let us do that.

(Refer Slide Time: 16:25)



```
guru@AngirasAgni:~/Desktop/NPTELMOOCcourse$ cd DwMD/
guru@AngirasAgni:~/Desktop/NPTELMOOCcourse/DwMD$ R

R version 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

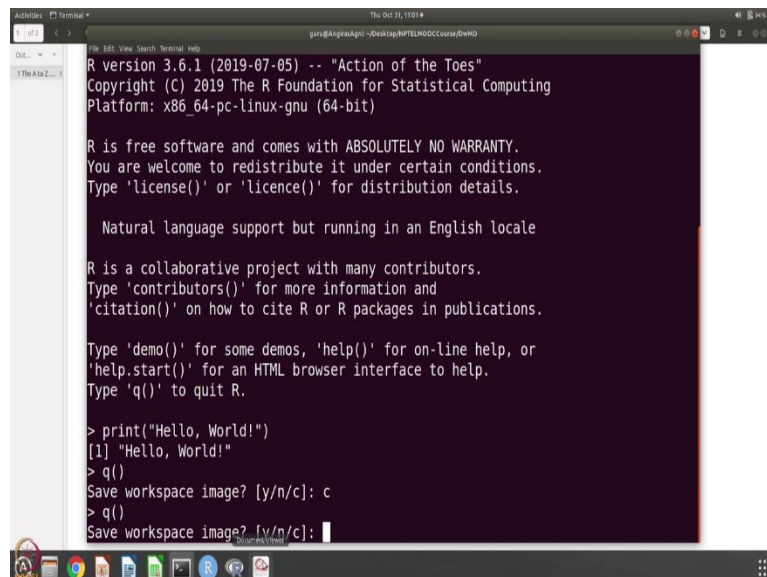
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> print("Hello, World!")
[1] "Hello, World!"
>
```



```
R version 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

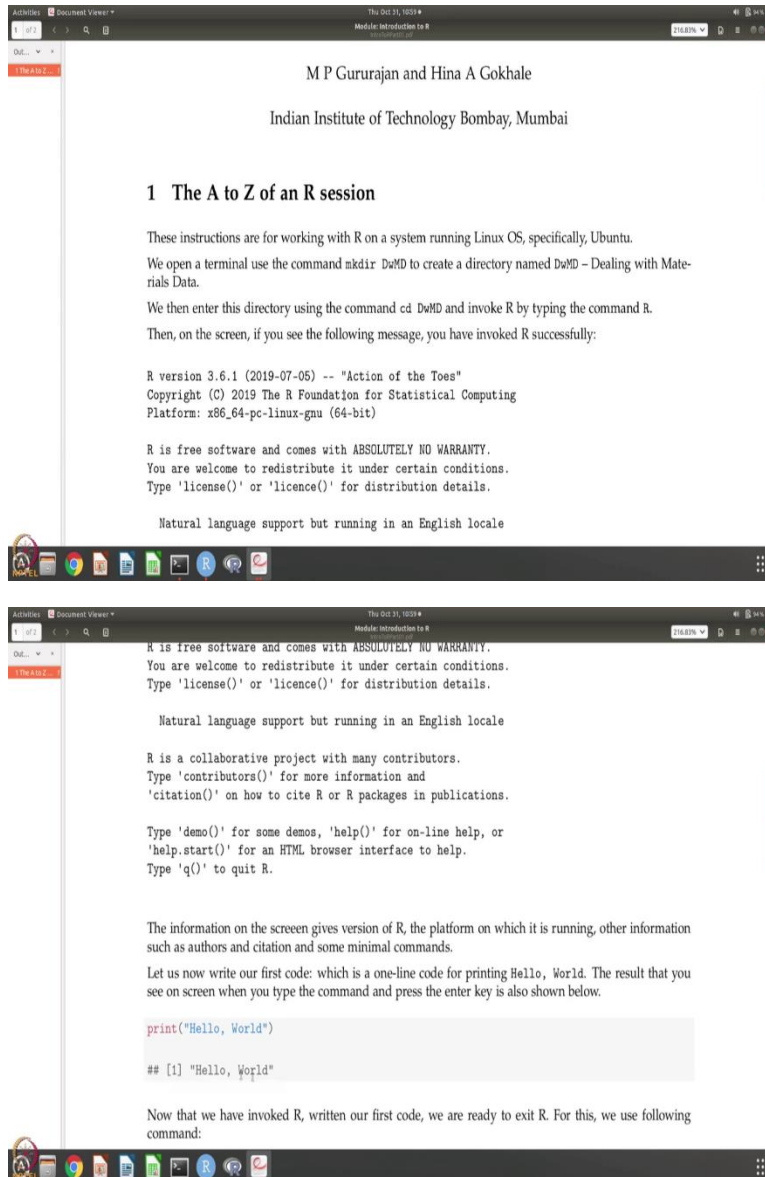
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> print("Hello, World!")
[1] "Hello, World!"
> q()
Save workspace image? [y/n/c]: c
> q()
Save workspace image? [y/n/c]:
```



As you can see, it is rather straightforward in R to print so it is almost like English to print hello world means print hello world right? So we say print hello world. So, there is a parenthesis, right. And what you want to print is given within quote marks and so let us put a bang so hello world. So, let us enter and R gives the answer immediately hello world you said print hello world, so it has printed hello world and this is how it will appear in the documentation also you will see for example, the command that I want to give is given like this.

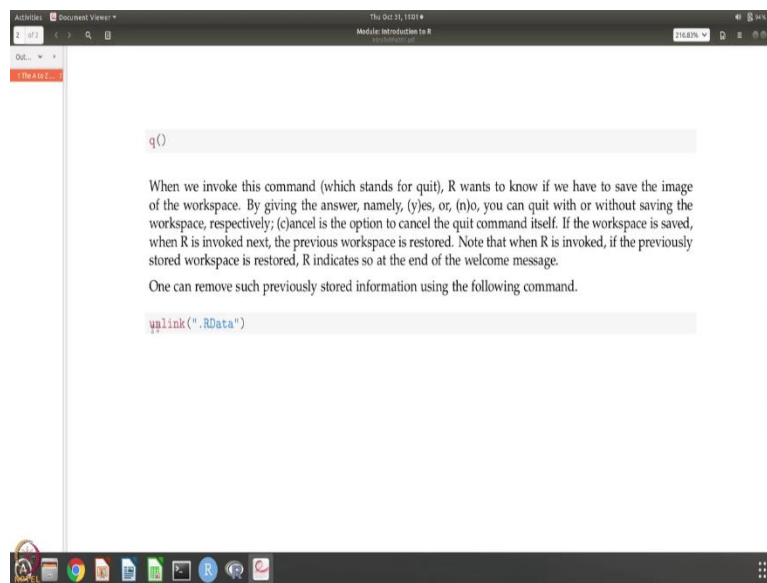
This is the command and the answer that are returns is given with this hash mark in the in these notes. So that is how I have described everything in this notes. So the commands will be shown and sometimes the answers that R will generate will also be shown so you can do it for yourself

and confirm that you are getting the same answer. And this markers, the greater than kind of symbol that you see that is the prompt. So that is never typed and that is already there.

So now that we have done we want to quit and of course R is very helpful. So it tells how to quit type q that is for quit. And that is like a function, so there is a parenthesis but there is no input. And you will say enter and typically R as do you want to save the workspace image. You can say yes, you can say no. And you can say cancel. So if you say cancel, then R does not quit. And if you say yes then it will save the workspace image and quit.

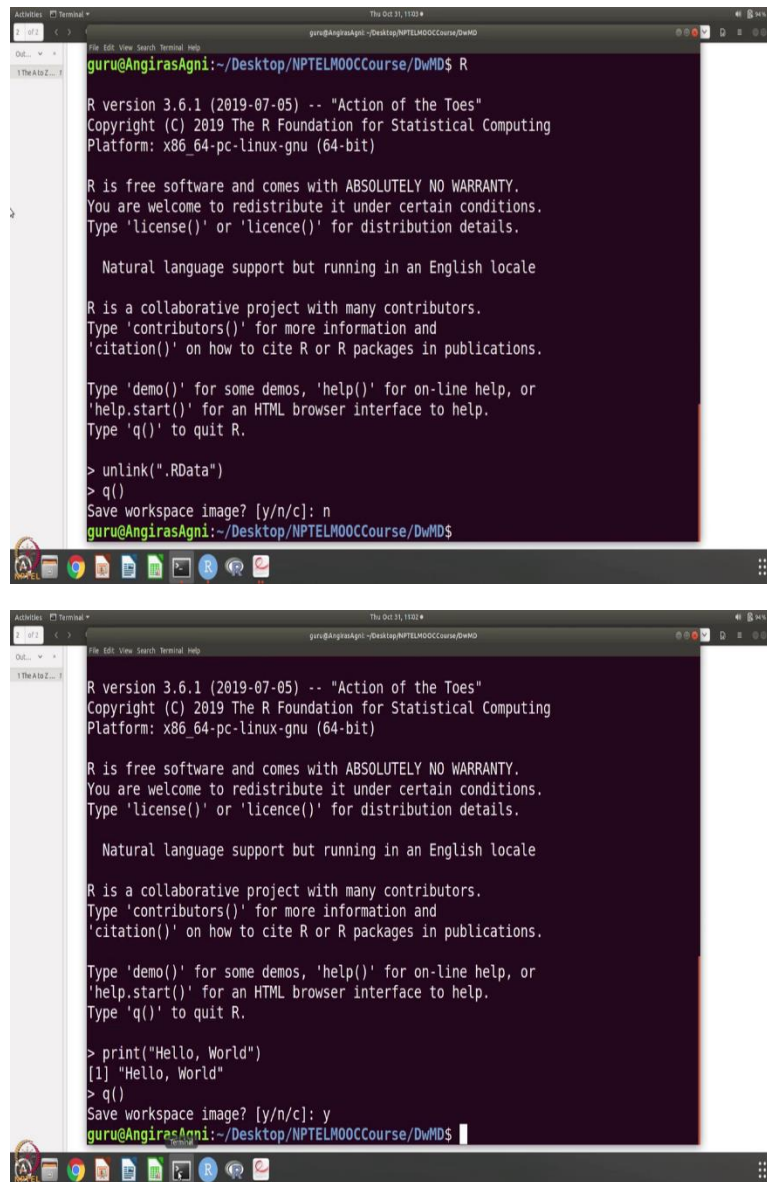
If you say no to not save, and if it saves the workspace image by default you can ask it to load it when you open R next time R sometimes you can load it yourself manually and it is also possible that if you have already have it loaded, then you can also remove those saved information.

(Refer Slide Time: 18:18)



And that is done using this command unlink within quote marks dot R capital Data data, so you have to be very careful the there is the caps and the small letters. So, the commands are sensitive to capitalization. So, you need to use exactly capital R and capital D and the other letters are small.

(Refer Slide Time: 18:48)



```
guru@AngirasAgni:~/Desktop/NPTELMOOCCourse/DwMD$ R
R version 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> unlink(".RData")
> q()
Save workspace image? [y/n/c]: n
guru@AngirasAgni:~/Desktop/NPTELMOOCCourse/DwMD$

R version 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> print("Hello, World")
[1] "Hello, World"
> q()
Save workspace image? [y/n/c]: y
guru@AngirasAgni:~/Desktop/NPTELMOOCCourse/DwMD$
```

So, by using this suppose you said so let us let us say no and I am going to quit and it quits R so this is the session on A to Z of an R session. But suppose if you said yes and It stored and when you opened R right next time when you say R so it does not say anything because we did not save and now lets redo this again let us say print hello world, so I have printed hello world and I say quit and it wants to save the workspace image I say yes. Now if I say R so, so it will it will remember my previous commands like this right.

So, you do not want this to happen. So, you will say unlink dot R data. So, then it will it will not have this information from data, the previous session data from R, so lets do this and no, so this is

the first session. So we opened a session, we wrote the hello world program and we quit our session. Thank you.