Introduction to R



**Anish Mahapatra**

1. Introduction

Hello everyone!

Welcome to the course on the Introduction to Data Science with R. The very fact that you registered for this course proves that you’re headed a step in the right direction to start your journey and to enhance your position in the Data science domain.

But before we begin let me tell you what you can expect from this course.

We will primarily be speaking about data science, some of the applications, not all of them since it’s a field with a vast scope across different domains. I will also be introducing you to the statistical coding language **R** which acts as a great tool to actively compute complex statistical analysis in a few simple steps.

The objective that I would like to achieve is to make sure you leave this course a little better versed in R and it’s uses without delving too deep into the technical side of things.

We will be dealing with questions such as What is data? What is R? What are packages? And a whole lot more that will give you a basic understanding of the beautiful field and how to go about it in the right way.

My name is Anish Mahapatra. I’m a Senior Data Scientist with a MSc degree in Data science with over than 5 years of work experience. I currently work at a Fortune 200 Retail Giant and I also am a consultant that’s worked with multiple Fortune 500 clients in the data science domain. A great way to know more about me would be to simple Google my name – Anish Mahapatra.

Without furthor ado, let us now dive into the fundamentals of Data Science and R!

1. What is Data?

It’s a simple question, right? What is Data? Data is a collection of discrete values that convey information. This can be in the form of quantity, quality, fact, statistics, and other basic units of meaning.

The world is driven by data – all the way from your mobile phones, e-commerce, banking and even your cars!

When looking at it from an uber perspective, it is important to understand what are companies spending their money on. The reason Data Science has such high holding for the next couple of decades is that companies are in the process of leveraging their data.

Data can be seen as rows and columns, where we are able to get contextualized information in a concise manner. It can be somethingas simple as your shopping list.

Now, how does this work in the digital world? We can have data in the form of csv files, excel files, a database and on the cloud platform. All of this data flows from different sources and there is great value in understanding what all of it means.

It might be easy to understand data that comprises of 10-20 rows. But, what if we increase that number to a 100 million rows? It becomes difficult to “understand” the data then, because it does not even open in Excel.

This is where a language like R can help. It can help you understand, infer and visualize large volumes of data.

Let’s have a look at What is R in our next session. This is probably a great place to tell you to not fear code. You have gotten this far and it is an achievement, and we will make this as easy to understand as possible. Stick with me.

1. What is R?

R is a great language for everyone to learn as it’s sole purpose is to collect, analyse, interpret and present data. In fact, this is the very definition of statistical analysis.

R is a programming language and environment for statistical computing and graphics to analyze and visualize data.

Every language has associated rules, called syntax that govern it. There are three items that are a part of the syntax of R are

* **Variables**: Objects that can store data
* **Comments**: To make the code easier to understand and more readable
* **Keywords**: Words that are reserved for the compiler of the language. What is a compiler? It converts instructions into machine-level language.

So, why do people actually use R?

As one of the top five programming languages currently. R is used by large companies in disciplines such as Fintech, Retail, Social Media, Healthcare etc. to do the following:

* Data Analytics
* Statstical Inference
* Machine Learning

You can then do things to those objects like:

Perform operations and calculations in a simple way.

Statistical Analysis and Tests

Make tables

Draw plots and visualise data for better understanding.

Objects can be single numbers, characters, vectors of numbers, matrices, multi-dimensional arrays, lists containing different objects and so on. We will use a fantastic IDE1 provided by RStudio. This is free to download, provides some neat features, and crucially, looks the same on all operating systems!

1. The applications of R - (Talk about RStudio)
2. What is an R Script?
3. What are packages in R
4. Commonly used packages in R
5. R for statistical analysis
6. Excel versus R versus Python
7. Data Types in R
8. Coding with R – Installation of R / R Studio
9. How to read a dataset in R

This can also be done manually in Rstudio by choosing the import dataset feature found in the file menu.

1. Data Operations in R



1. Data Wrangling in R
2. Data Wrangling in R
3. Visualization with R
4. Applications of Machine Learning in R
5. AI in R
6. Thank you!

# Author

**Anish Mahapatra**   
*Senior Data Scientist*   
LinkedIn: [https://www.linkedin.com/in/anishmahapatra](about:blank) ([LinkedIn](about:blank))   
Email: anishmahapatra01@gmail.com