Introduction to R



**Anish Mahapatra**

Contents

[1. Introduction 3](#_Toc120372053)

[2. What is Data? 4](#_Toc120372054)

[3. What is R? 4](#_Toc120372055)

[4. The applications of R - (Talk about RStudio) 5](#_Toc120372056)

[5. What is an R Script? 7](#_Toc120372057)

[6. What are packages in R 7](#_Toc120372058)

[7. Commonly used packages in R 7](#_Toc120372059)

[8. R for statistical analysis 7](#_Toc120372060)

[9. Excel versus R versus Python 7](#_Toc120372061)

[10. Data Types in R 7](#_Toc120372062)

[11. Coding with R – Installation of R / R Studio 7](#_Toc120372063)

[12. How to read a dataset in R 7](#_Toc120372064)

[13. Data Operations in R 8](#_Toc120372065)

[14. Data Wrangling in R 9](#_Toc120372066)

[15. Data Wrangling in R 9](#_Toc120372067)

[16. Visualization with R 9](#_Toc120372068)

[17. Applications of Machine Learning in R 9](#_Toc120372069)

[18. AI in R 9](#_Toc120372070)

[19. Thank you! 9](#_Toc120372071)

[20. Author 11](#_Toc120372072)

[21. Archive 11](#_Toc120372073)

# 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!

# 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.

# 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:

1. **Variables**: Objects that can store data
2. **Comments**: To make the code easier to understand and more readable
3. **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

There is a good reason why a language becomes one of the most widely used in the world. Some of the reasons for R programming to be one of the op languages are:

* **Open-source free language**: R and it’s suite of tools is completely free to download and use
* **Platform Independent**: Whether you have a mac, windows, linux or even a cloud server, R can be run on all platforms with ease
* **Leading tool for machine learning**: R programming is one of the leading tools for machine learning, statstics and data analysis. It is one of the most requested language in the Data Science job market.
* **Flexibility**: R can be integrated with many other languages, and this makes it even more useful, irrespective of the language that is used
* **Community Support**: Any questions that you have with R can easily be answered with a Google search. This is because there are a lot of people that support the community.

Alright Anish, you seem to have sold me on the potential of R, not just in the job market, but also in the search for great Data Science jobs. So, how do we actually use it?

I’m glad you asked. In the next session, we will talk more about some of the applications of R, before we deep-dive and look at how you can get started in under ten minutes!

# The applications of R - (Talk about RStudio)

Hi everyone, welcome back to this session, where we will discuss the applications of R. At any point if the complexity of data increases, we need ways that we can interpret, interface and analyze the data. There are use-cases in multiple domains such as:

* **Research & Academics**: Since R Programming is used primarily for statistical computing, it is a great resource for students & teachers to leverage
* **Information Technology (IT) Sector**: The IT sector uses R to enable businesses to build statistical, computational tools, data handling and business intelligence.   
    
  There are use-cases such as the analysis of web activity, recommendation systems, marketing campaign analysis and even matchmaking systems
* **Banking & Finance**: Risk management of assets, providing loans to potential customers, upselling financial & banking products are some use-cases where R can be tremendously helpful.   
     
  It can also be used for fraud detection, loan stress modelling, loan stress test simulation and a lot of other forms of risk analytics.
* **E-Commerce**: E-commerce uses the potential of R to work on use-cases such as marketing mix modelling, cross-selling, targeted advertising, sales and financial data modelling.
* **Healthcare**: R is greatly used in healthcare for a plethora of operations. It can be used to do analysis in the fiels of Genertics, bioinformatics, epidemiology, and, drug discovery. It can also be used to understand the statstical understanding of the success of clinical trials.

Alright, great, it’s good to know that it is used in a lot of verticals across multiple industries. Now, the question is are we getting closer to understanding how can I get started with R? Yes, we are getting close, but we are not just there yet.

An IDE or integrated development environment is a software to build code, which combines developer tools into a single interface. Luckily for us, the IDE for R, called RStudio is a all-in-one solution to quickly get started with R!

In the next section, we will learn more about the core concept of how we will set out to build fantastic things with R – using R scripts!

# What is an R Script?

# What are packages in R

# Commonly used packages in R

# R for statistical analysis

# Excel versus R versus Python

# Data Types in R

# Coding with R – Installation of R / R Studio

# 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.

# Data Operations in R



# Data Wrangling in R

# Data Wrangling in R

# Visualization with R

# Applications of Machine Learning in R

# AI in R

# Thank you!



# Author

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

# Archive

1. Introduction
2. What is Data?
3. What is R?
4. The applications of R - (Talk about RStudio)
5. What is an R Script?
6. What are packages in R
7. Commonly used packages in R
8. R for statistical analysis
9. Excel versus R versus Python
10. Data Types in R
11. Coding with R – Installation of R / R Studio
12. How to read a dataset in R
13. Data Operations in R
14. Data Wrangling in R
15. Visualization with R
16. Applications of Machine Learning in R
17. AI in R
18. Thank you!