Overview of R and RStudio

Spoken Tutorial Project
https://spoken-tutorial.org
National Mission on Education through ICT

http://sakshat.ac.in/

Script: Sudhakar Kumar

Narration: Madhuri Ganapathi

IIT Bombay
15 April 2020

We will learn:

Learning Objectives

We will learn:

- **About R programming language**

Learning Objectives

We will learn:

- ▶ **About R programming language**
- ▶ **Applications of R**

Learning Objectives

We will learn:

- ▶ **About R programming language**
- ▶ **Applications of R**
- ▶ **About RStudio**

Learning Objectives

We will learn:

- **About R programming language**
- **Applications of R**
- **About RStudio**
- **Topics covered in this series**

System Specifications

System Specifications

- **Ubuntu Linux OS version 18.04**

System Specifications

- ▶ **Ubuntu Linux OS version 18.04**
- ▶ **R version 3.4.4**

System Specifications

- **Ubuntu Linux OS version 18.04**
- **R version 3.4.4**
- **RStudio version 1.2.5033**

System Specifications

- ► **Ubuntu Linux OS version 18.04**
- ► **R version 3.4.4**
- ► **RStudio version 1.2.5033**

System Specifications

- ▶ **Ubuntu Linux OS version 18.04**
- ▶ **R version 3.4.4**
- ▶ **RStudio version 1.2.5033**

Install R version 3.2.0 or higher

Prerequisites

To follow this series, the learner should have knowledge of basic Mathematics and Statistics

About R

About R

► **R is a programming language developed by Ross Ihaka and Robert Gentleman in 1993**

About R

- ▶ **R is a programming language developed by Ross Ihaka and Robert Gentleman in 1993**
- ▶ **R is used in statistics, data analysis, and machine learning algorithms**

About R

- ▶ **R is a programming language developed by Ross Ihaka and Robert Gentleman in 1993**

- ▶ **R is used in statistics, data analysis, and machine learning algorithms**

- ▶ **It is a free/libre open source software**

Features of R

Some of its salient features are:

▶ **R provides support for both procedural programming and object-oriented programming**

Features of R

Some of its salient features are:

► **R provides support for both procedural programming and object-oriented programming**

► **R can be integrated with other programming languages like C, C++, Python, Java, etc.**

Features of R

► **R is capable of effective data handling and storage**

Features of R

► **R is capable of effective data handling and storage**

► **R has more than 10,000 packages in its repository**

Features of R

- ▶ **R is capable of effective data handling and storage**
- ▶ **R has more than 10,000 packages in its repository**
- ▶ **R is cross-platform compatible**

Features of R

- ▶ **R facilitates complex operations with vectors, arrays, and data frames**

Features of R

▶ **R facilitates complex operations with vectors, arrays, and data frames**

▶ **R has community support of avid developers**

Applications of R

R is extensively used in:

▶ **Fraud detection especially in the finance and banking sectors**

R is extensively used in:

- ▶ **Fraud detection especially in the finance and banking sectors**
- ▶ **Social media analysis to discover potential customers in online advertising**

Applications of R

- **Healthcare to perform pre-clinical trials and analyze the drug-safety data**

Applications of R

▶ **Healthcare to perform pre-clinical trials and analyze the drug-safety data**

▶ **E-Commerce companies to analyze the purchases made by customers**

Names of the companies that use R

ANZ Bank:

► **ANZ Bank, the fourth largest bank in Australia, uses R for credit risk analysis**

ANZ Bank:

▶ **ANZ Bank, the fourth largest bank in Australia, uses R for credit risk analysis**

▶ **It also uses R to fit models for mortgage loss**

Names of the companies that use R

Facebook:

► **Facebook uses R for exploratory data analysis, experimental analysis, etc.**

Facebook:

► **Facebook uses R for exploratory data analysis, experimental analysis, etc.**

► **It also uses R for analyzing user behavior like status updates and profile pictures**

Names of the companies that use R

Google:

► **Google uses R to calculate return on investment on advertising campaigns**

Google:

► **Google uses R to calculate return on investment on advertising campaigns**

► **It also uses R to improve the efficiency of online advertising**

About RStudio

About RStudio

▶ **RStudio is an integrated development environment for R**

About RStudio

- ▶ **RStudio is an integrated development environment for R**
- ▶ **It includes a console and an editor**

About RStudio

▶ **RStudio is an integrated development environment for R**

▶ **It includes a console and an editor**

▶ **It provides tools for plotting, debugging and workspace management**

Installing R and RStudio

1. **Installing R and RStudio on Linux**
2. **Installing R and RStudio on Windows**

Basics of R and RStudio

1. **Introduction to basics of R**
2. **Introduction to Data Frames in R**
3. **Introduction to RStudio**
4. **Introduction to R script**
5. **Working directories in RStudio**

Basics of Data frames and Lists

Basics of Data frames and Lists

1. **Indexing and Slicing Data Frames**
2. **Creating Matrices using Data Frames**
3. **Operations on Matrices and Data Frames**

Basics of Data frames and Lists

4. **Merging and Importing Data**
5. **Data types and Factors**
6. **Lists and its Operations**

Visualizing Data using plots

Visualizing Data using plots

1. **Plotting Histograms and Pie Chart**
2. **Plotting Bar Charts and Scatter Plot**

Customizing Plots using ggplot2

Customizing Plots using ggplot2

1. **Introduction to ggplot2**
2. **Aesthetic Mapping in ggplot2**

Manipulating Data using dplyr

Manipulating Data using dplyr

1. **Data Manipulation using dplyr Package**
2. **More functions in the dplyr Package**
3. **Pipe Operator**

Using Conditional Statements

1. **Conditional Statements**
2. **Functions in R**

Advanced Level Topics

▶ **We may add a few more advanced level topics in this series**

▶ **Please visit this series at regular intervals for the latest updates**

Summary

We have learnt:

- ▶ **About R programming language**
- ▶ **Applications of R**
- ▶ **About RStudio**
- ▶ **Topics covered in this series**

About the Spoken Tutorial Project

- ▶ Watch the video available at https://spoken-tutorial.org /What_is_a_Spoken_Tutorial

- ▶ It summarises the Spoken Tutorial project

- ▶ If you do not have good bandwidth, you can download and watch it

Spoken Tutorial Workshops

The Spoken Tutorial Project Team

► **Conducts workshops using spoken tutorials**

► **Gives certificates to those who pass an online test**

► **For more details, please write to contact@spoken-tutorial.org**

Forum to answer questions

▶ **Do you have questions in THIS Spoken Tutorial?**

▶ **Please visit**
https://forums.spoken-tutorial.org

▶ **Choose the minute and second where you have the question**

▶ **Explain your question briefly**

▶ **Someone from the FOSSEE team will answer them**

Forum for specific questions

▶ **The Spoken Tutorial forum is for specific questions on this tutorial**

▶ **Please do not post unrelated and general questions on them**

▶ **This will help reduce the clutter**

▶ **With less clutter, we can use these discussions as instructional material**

Forum to answer questions

▶ **Questions not related to the Spoken Tutorial?**

▶ **Do you have general / technical questions on the Software?**

▶ **Please visit the FOSSEE Forum**

https://forums.fossee.in/

▶ **Choose the Software and post your question**

Textbook Companion Project

- ► **The FOSSEE team coordinates coding of solved examples of popular books**

- ► **We give honorarium and certificates to those who do this**

For more details, please visit these sites:

https://r.fossee.in/
https://fossee.in/

Acknowledgements

Spoken Tutorial Project is supported by

- ▶ **National Mission on Education through ICT (NMEICT)**

- ▶ **Pandit Madan Mohan Malaviya National Mission on Teachers and Teaching**

Initiatives of MHRD, Government of India

► **This tutorial is contributed by Sudhakar Kumar and Madhuri Ganapathi, IIT Bombay**