

R Overview

Eric Snell

2024-05-06

Introduction to R

R is a programming language and software environment widely used for statistical computing and graphics. It provides a wide variety of statistical and graphical techniques, and is highly extensible through packages. R is particularly popular among statisticians and data miners for developing statistical software and data analysis.

Choosing a Development Environment

There are several options for R development environments. Here are three popular choices:

RStudio

RStudio is an integrated development environment (IDE) for R that provides a user-friendly interface for writing, debugging, and running R code. It offers features such as syntax highlighting, code completion, and integrated plotting tools. You can download RStudio from [here](#).

Jupyter Notebooks with R Kernel

Jupyter Notebooks is an open-source web application that allows you to create and share documents containing live code, equations, visualizations, and narrative text. With the R kernel installed, you can write and execute R code directly within Jupyter Notebooks. You can install Jupyter Notebooks with the R kernel using the following command:

```
install.packages('IRkernel')  
IRkernel::installspec()
```

R Console

The R Console is the basic interface for running R code. It comes bundled with the R software and provides a simple command-line interface for interacting with R. While not as feature-rich as RStudio or Jupyter Notebooks, the R Console is lightweight and sufficient for many tasks.

Getting Started with R

Beginner

At the beginner level in R, you'll start by learning the basics of the language and its applications in statistical analysis and data visualization. You'll explore fundamental concepts such as data types, vectors, matrices, and data frames, which are essential for handling and analyzing data. Additionally, you'll become familiar with basic statistical techniques and graphical functions for summarizing and visualizing data. Through hands-on exercises and real-world examples, you'll gain proficiency in tasks such as data import/export, data manipulation, and basic statistical analysis.

- **Coursera Courses:**

- R Programming by Johns Hopkins University.
- Data Science: Foundations using R Specialization by Johns Hopkins University.
- **YouTube Channels/Blogs:**
 - R-bloggers Blog.
 - David Robinson’s YouTube Channel.
- **Projects**
 - Exploratory Data Analysis: Analyze a dataset using descriptive statistics and visualizations to gain insights.
 - Data Visualization: Create visualizations using ggplot2 to effectively communicate data findings.

Intermediate

At the intermediate level in R, you’ll deepen your understanding of statistical analysis and explore more advanced data manipulation and visualization techniques. This stage involves mastering packages like dplyr and tidyr for data wrangling, which allow you to efficiently manipulate and transform datasets. You’ll also learn advanced statistical techniques such as regression analysis, hypothesis testing, and predictive modeling, using packages like glm and caret. Additionally, you’ll explore advanced visualization packages like ggplot2 and plotly to create interactive and publication-quality plots. By honing these skills, you’ll be able to perform sophisticated data analysis tasks and derive actionable insights from complex datasets.

- **Coursera Courses:**
 - Advanced R Programming by Johns Hopkins University.
 - Data Science: Statistics and Machine Learning Specialization by Johns Hopkins University.
- **YouTube Channels/Blogs:**
 - DataCamp’s Blog.
 - RStudio’s YouTube Channel.
- **Projects**
 - Predictive Modeling: Build a machine learning model to predict a target variable using regression or classification techniques.
 - Interactive Visualizations: Create interactive plots using plotly to explore complex datasets and uncover trends.

Advanced

At the advanced level in R, you’ll explore cutting-edge techniques and tools for solving complex data analysis problems and driving business outcomes. This stage involves mastering topics such as deep learning, natural language processing (NLP), and big data analysis, using packages like keras, text2vec, and sparklyr. You’ll also learn about techniques for deploying and scaling data analytics solutions in production environments, ensuring that your organization can derive maximum value from its data assets. Additionally, you’ll explore topics such as reproducible research and version control, using tools like RMarkdown and GitHub to collaborate on data analysis projects and share results with stakeholders. By mastering these advanced skills, you’ll be well-positioned to lead data analytics initiatives within your organization and drive business success through data-driven decision-making.

- **Coursera Courses:**
 - Advanced Machine Learning and Signal Processing by Duke University.
 - Data Science Specialization by Johns Hopkins University.
- **YouTube Channels/Blogs:**
 - RStudio Webinars Channel.
 - RStudio Blog Blog.
- **Projects**
 - Deep Learning: Build a deep learning model to solve a complex problem such as image classification or text generation.
 - Big Data Analysis: Analyze large-scale datasets using distributed computing frameworks like Apache Spark to extract actionable insights.