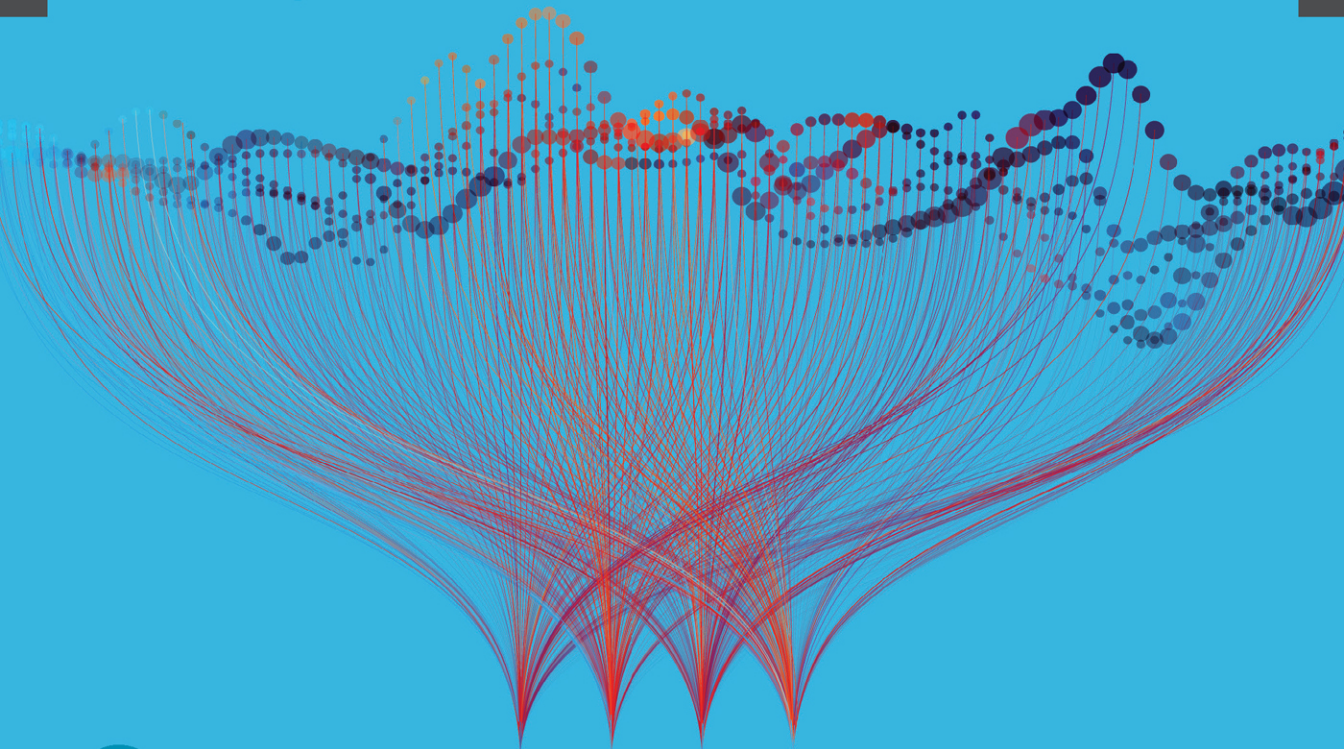


ADDISON WESLEY DATA & ANALYTICS SERIES



# PROGRAMMING SKILLS FOR DATA SCIENCE

Start Writing Code to Wrangle,  
Analyze, and Visualize Data with R



MICHAEL FREEMAN | JOEL ROSS

# Programming Skills for Data Science

---

# Data Science Foundations Tools and Techniques

## Table of Contents

Cover

Half Title

Title Page

Copyright Page

Dedication

Contents

Foreword

Preface

Acknowledgments

About the Authors

I: Getting Started

1 Setting Up Your Computer

1.1 Setting up Command Line Tools

1.2 Installing git

1.3 Creating a GitHub Account

1.4 Selecting a Text Editor

1.5 Downloading the R Language

1.6 Downloading RStudio

2 Using the Command Line

2.1 Accessing the Command Line

2.2 Navigating the File System



# Table of Contents

2.3 Managing Files

2.4 Dealing with Errors

2.5 Directing Output

2.6 Networking Commands

## II: Managing Projects

### 3 Version Control with git and GitHub

3.1 What Is git?

3.2 Conguration and Project Setup

3.3 Tracking Project Changes

3.4 Storing Projects on GitHub

3.5 Accessing Project History

3.6 Ignoring Files from a Project

### 4 Using Markdown for Documentation

4.1 Writing Markdown

4.2 Rendering Markdown

## III: Foundational R Skills

### 5 Introduction to R

5.1 Programming with R

5.2 Running R Code

5.3 Including Comments

5.4 Defining Variables

5.5 Getting Help

### 6 Functions

6.1 What Is a Function?

6.2 Built-in R Functions

6.3 Loading Functions

6.4 Writing Functions

6.5 Using Conditional Statements

### 7 Vectors

# **Table of Contents**

7.1 What Is a Vector?

7.2 Vectorized Operations

7.3 Vector Indices

7.4 Vector Filtering

7.5 Modifying Vectors

## **8 Lists**

8.1 What Is a List?

8.2 Creating Lists

8.3 Accessing List Elements

8.4 Modifying Lists

8.5 Applying Functions to Lists with `lapply()`

## **IV: Data Wrangling**

### **9 Understanding Data**

9.1 The Data Generation Process

9.2 Finding Data

9.3 Types of Data

9.4 Interpreting Data

9.5 Using Data to Answer Questions

### **10 Data Frames**

10.1 What Is a Data Frame?

10.2 Working with Data Frames

10.3 Working with CSV Data

### **11 Manipulating Data with dplyr**

11.1 A Grammar of Data Manipulation

11.2 Core dplyr Functions

11.3 Performing Sequential Operations

11.4 Analyzing Data Frames by Group

11.5 Joining Data Frames Together

11.6 dplyr in Action: Analyzing Flight Data

# **Table of Contents**

## **12 Reshaping Data with tidyr**

12.1 What Is Tidy Data?

12.2 From Columns to Rows: gather()

12.3 From Rows to Columns: spread()

12.4 tidyr in Action: Exploring Educational Statistics

## **13 Accessing Databases**

13.1 An Overview of Relational Databases

13.2 A Taste of SQL

13.3 Accessing a Database from R

## **14 Accessing Web APIs**

14.1 What Is a Web API?

14.2 RESTful Requests

14.3 Accessing Web APIs from R

14.4 Processing JSON Data

14.5 APIs in Action: Finding Cuban Food in Seattle

## **V: Data Visualization**

### **15 Designing Data Visualizations**

15.1 The Purpose of Visualization

15.2 Selecting Visual Layouts

15.3 Choosing Effective Graphical Encodings

15.4 Expressive Data Displays

15.5 Enhancing Aesthetics

### **16 Creating Visualizations with ggplot2**

16.1 A Grammar of Graphics

16.2 Basic Plotting with ggplot2

16.3 Complex Layouts and Customization

16.4 Building Maps

16.5 ggplot2 in Action: Mapping Evictions in San Francisco

### **17 Interactive Visualization in R**

# **Table of Contents**

17.1 The plotly Package

17.2 The rbokeh Package

17.3 The leaflet Package

17.4 Interactive Visualization in Action: Exploring Changes to the City of Seattle

## **VI: Building and Sharing Applications**

### **18 Dynamic Reports with R Markdown**

18.1 Setting Up a Report

18.2 Integrating Markdown and R Code

18.3 Rendering Data and Visualizations in Reports

18.4 Sharing Reports as Websites

18.5 R Markdown in Action: Reporting on Life Expectancy

### **19 Building Interactive Web Applications with Shiny**

19.1 The Shiny Framework

19.2 Designing User Interfaces

19.3 Developing Application Servers

19.4 Publishing Shiny Apps

19.5 Shiny in Action: Visualizing Fatal Police Shootings

### **20 Working Collaboratively**

20.1 Tracking Different Versions of Code with Branches

20.2 Developing Projects Using Feature Branches

20.3 Collaboration Using the Centralized Workflow

20.4 Collaboration Using the Forking Workflow

### **21 Moving Forward**

21.1 Statistical Learning

21.2 Other Programming Languages

21.3 Ethical Responsibilities

## **Index**