まえがき

R は統計計算，データ分析，そしてその結果を可視化するために設計されたプログラミング言語であり，近年，データサイエンスおよび統計学の分野においては最も有名なものとなっている．R によるプログラミングにはデータプロセシングが大きな比重を占めており，これは R に不慣れなユーザにとってはかなり骨の折れる作業でもある．

R は動的言語 (dynamic language) であり，C++ や Java，C#といった厳密な型をもつ言語に比べると非常に柔軟なデータ構造を有する．実際，そのような言語に慣れていた筆者が R を利用し始めた時には，R の挙動は非常に奇妙かつ予測不能で，一貫性のないものに思えたものである．R を利用するようなデータ分析プロジェクトにおいて，多くの時間はモデルの作成ではなく，データクリーニングや探索的分析，その可視化に割かれる．そして，おかしな結果やエラーを吐き出すコードの間違い探しも，我々から多くの時間を奪う．本質的な問題解決とは程遠いプログラミングそのものの問題と戦っている時，とりわけヒントなしに何時間もバグと戦っている時などはイライラして仕方がない．しかし，データ分析プロジェクトにおける経験を積み，R におけるオブジェクトや関数の挙動を理解するにつれ，思ったよりもこの言語は一貫性があり美しく設計されているのではないかと思うようになった．この観点を多くの人と共有したい，これがこの本を執筆するに至った動機である．

本書を通じて，あなたは R をプログラミング言語として広く一貫性をもった形で理解でき，周辺の有用なツールについても知ることができる．また，あなたの生産性を向上させるベストプラクティスやデータ操作についての深い理解も得られるだろう．そして，R を用いてプログラミングを行い，適切な問題解決を実行するために十分な自信も獲得できるものと思われる．

本書がカバーする内容

第 1 章「クイックスタート」では，R に関する基本的な知識や R の環境構築方法，そして RStudio を用いたプログラミングについて学ぶ．第 2 章「基本的なオブジェクト」では，R における基本的なオブジェクトの内容および挙動について学ぶ．第 3 章「作業スペースの管理」では，作業ディレクトリや R 内の環境，そして拡張パッケージの管理方法について学ぶ．第 4 章「基本的な表現式」では，代入式，条件式，ループといった基本的な条件式について学ぶ．第

5 章「基本的なオブジェクトを扱う」では，R におけるオブジェクトを操作するための基本的な関数について学ぶ．第 6 章「文字列を扱う」では，文字列に関連する R のオブジェクトおよび文字列操作のテクニックについて学ぶ．第 7 章「データを扱う」では，実例を交えた形でシンプルな入出力関数について学ぶ．第 8 章「R の内部を覗く」では，遅延評価や環境，関数，レキシカルスコーピングといったトピックを取り上げながら，R における評価の仕組みについて学ぶ．第 9 章「メタプログラミング」では，R の言語オブジェクト (language object) や非標準評価 (non-standard evaluation) を理解する際に役立つメタプログラミングについて学ぶ．第 10 章「オブジェクト指向プログラミング」では，S3，S4，参照クラス，拡張パッケージとして提供されている R6 といった R におけるクラスシステムについて学ぶ．第 11 章「データベース操作」では，SQLite や MySQL といったポピュラーなリレーショナルデータベースを R でどう扱うかに始まり，MongoDB や Redis といった NoSQL 型データベースの操作方法についても学ぶ．第 12 章「データ操作」では，**data.table** パッケージおよび **dplyr** パッケージを用いたデータフレームの操作，そして **rlist** パッケージを用いたリストの操作について学ぶ．第 13 章「ハイパフォーマンスコンピューティング」では，パフォーマンスの話題を扱い，**Rcpp** パッケージの利用といった R における計算速度を高める方法について紹介する．第 14 章「ウェブスクレイピング」では，ウェブページ，CSS セレクタ，XPath の基礎および **rvest** パッケージを用いたウェブスクレイピングについて学ぶ．第 15 章「生産性を高める」では，分析結果のレポーティングやプレゼンテーションを効率よく実行できる RMarkdown，Shiny について学ぶ．

本書のコードを実行するために必要な環境

本書のコードは R3.4.0 日本語版で実行を確認している [1)]．また，開発環境は RStudio が望ましい．

第 11 章のコードを実行するためには MongoDB，Redis の実行環境が必要である．第 13 章における Rcpp のコードを実行する際，Windows の場合は利用している R のバージョンに合わせた Rtools のインストールが，Linux/macOS の場合は gcc のインストールがそれぞれ必要である．

本書の想定読者

本書の読者としては，データ分析プロジェクトにかかわっており生産性を上げたいと思っているものの，プログラミングには不慣れな人を想定している．また同時に，周辺技術やツール，拡張パッケージも交えた形で R を言語として体系立てて理解したい上級者も想定読者に含まれる．

いくつかの章は初心者にとっては難しいため，読み飛ばしてもらっても構わない．だが，これらの章を読んでおくと，R に限らないプログラミングの基本概念やデータ分析における基本

[1)] 訳注：エラーメッセージ，警告メッセージについては和英が混在している形になっているが，これは本実行環境を反映したものとなっている．

概念を知ることができるので，上級者を志すなら一読をお薦めしたい．

本書の表記について

本書では用途に合わせて表記を変えている．データベースのテーブル名，フォルダ（ディレクトリ）名，ファイル拡張子，Twitter のハンドル名は，コードと同様の書体で表記している．また，関数名については `apply()` のように末尾に括弧を付与している．以下にその表記例を示す．

例：`apply()` は配列による入力を受け付け，行列の形で出力する．

本文中のコードは以下のようにインデントを加えて表記しており，出力結果については `#` を 2 つ並べた形で区別している．

```
x <- c(1, 2, 3)
class(x)
## [1] "numeric"
typeof(x)
## [1] "double"
str(x)
## num [1:3] 1 2 3
```

一部のコードでは，重要なポイントを表すために太字で表記している．

```
x <- rnorm(100)
y <- 2 * x + rnorm(100) * 0.5
m <- lm(y ~ x)
coef(m)
```

初出の単語および重要な用語についても太字で表記している．

[2] 訳注：本書の正誤表およびコードに関する補足事項は，GitHub 上のサポートサイトに掲載している．
https://github.com/HOXOMInc/Learning_R_Programming