

R と Quarto ではじめるデータサイエンス

苅谷千尋

kariyach@staff.kanazawa-u.ac.jp

Q1, 2026

I. 授業の主題

データサイエンスは、膨大なデータを分析する、あるいは個々のデータを結合しビッグデータにして解析することが必要です。そのためには Microsoft Excel のような GUI (グラフィカルユーザーインターフェース) に依拠するアプリケーションではなく、コード中心のプログラム言語が適しています。研究分野を問わず、海外の大学や研究機関において、「R」(後述) や「Python」のスキルが重視される傾向にあるのはこのためです。日本の研究機関は、全般的に、このような海外のトレンドに追いついていません。

この授業は、統計プログラム言語「R」の基本的な機能と、Quarto (markdown にもとづく簡易入力と R の出力結果を Pandoc にもとづいて出版する Posit 社の推奨する新しい出版システム) を学習するものです。R 言語はプログラム言語のなかでは比較的習得が容易であり、社会科学から自然科学にいたるまで幅広い分野の研究者が利用しています。

この授業は統計学、確率、多変量解析の授業ではありません。統計学的な出力やその妥当性の検討よりも、数値の出力結果を可視化する (ggplot2 パッケージ) 方法に重点を置きます。わかりやすい図を簡単に作成する、同じコードを繰り返し利用することで作業量を減らす、また、再現可能性を高めることも重視されます。

専攻、専門分野は問いません。実験や研究ですでに手元にあるデータを可視化したい、あるいは、データを可視化することに関心をもてる方の受講を歓迎いたします。プログラム言語の初心者であっても受講に問題はありません。この授業では、Chat GPT のような生成 AI を補助的に使うことを積極的に推奨することで、初学者やプログラム言語に苦手意識をもつ学生を支援します。

本授業の目標は、受講生が教員の提供するデータセットの中から自ら関心のあるものを選択し、目的に応じた適切な図を作成できるようになることです。教員提供以外のデータセットを用いて作図することも可能です。その場合、目的に応じた適切な図が作成できていれば、加点の対象とします。

II. 学修目標 (到達目標)

1. 受講生が、R 言語の基本的な操作 (データの読み込み、加工、抽出、結合など) ができる
2. 受講生が、ggplot2 パッケージを使い、基本的な図 (棒グラフ; ヒストグラム; 箱ひげ図; 散布図; 折れ線グラフ) を作成できる
3. 受講生が、自分の関心のあるテーマにかかわるデータから適切な図を作成できる

III. 授業概要

授業は、講義と演習の形式でおこないます。

講義の前半は、教員が用意した授業教材を中心に、コードの意味や記述方法、動作、実行結果などを例示、解説します。後半の演習は、1. 各自所有のノート PC を使って演習をおこないますので、必ずノート PC を持参してください。テキストの例題と合わせて、2. 自分の関心のあるテーマにかかわるデータセットでも、同様のコードを実践してください。

授業教材はウェブサイト (<https://kariyach.github.io/teaching/RandQuarto/>) で公開します。このウェブサイトは Quarto と GitHub を用いて作成されています。

IV. 講義スケジュール

1. イントロダクション
 - ・ データを見る；R と R Studio のインストール；プロジェクトの作成；base R と Tidyverse
2. R の基本的な操作方法 (1)
 - ・ データセット；データの読み込み；データの型；データの構造；データの集計；再現可能性 (アウトプット)；生成 AI
3. 可視化 (1)
 - ・ ggplot2；プロットとレイヤー
4. R の基本的な操作方法 (2)
 - ・ Tidyverse；パイプ演算子；データの加工；データの抽出；記述統計量；データの結合；欠損値
5. 可視化 (2)
 - ・ 5 Named Graphs (5NG)：棒グラフ；ヒストグラム；箱ひげ図；散布図；折れ線グラフ
6. 可視化 (3)
 - ・ プロットを整える；色とラベル；横持ちデータと縦持ちデータ；チャンクオプション
7. 可視化 (4)
 - ・ 最終成果物; YAML; CSS; テキストラベル
8. 可視化 (5) とプレゼンテーション
 - ・ 自作関数；ループ処理；プレゼンテーション

V. 評価方法と割合

1. 評価方法

- ・ 次項の項目及び割合で総合評価し、次のとおり判定します。
 - ・ 「S (達成度 90%~100%)」、「A (同 80%~90%未満)」、「B (同 70%~80%未満)」、「C (同 60%~70%未満)」を合格とし、「不可 (同 60%未満)」を不合格とする。(標準評価方法)

2. 評価の割合

項目	割合
宿題：授業の感想	30%
宿題：演習	30%
レポート	30%
プレゼンテーション	30%
受講態度	10%

! Important

- ・ ※授業には 3 分の 2 以上の出席を必要とします
- ・ ※評価基準は、学修目標、授業別ルーブリックに準じます

VI. 授業時間外の学修に関する指示

- ・ 予習に関する指示
 - ・ 事前に教材を下読みして下さい。
- ・ 復習に関する指示
 - ・ 毎回の授業の感想を書くことを宿題とします。また、毎回の授業で学習したコードやパッケージを各自の PC で再実践してください。これも宿題とします。この他、各自、最終成果物のレポート作成に向けて、努力して下さい。

VII. テキスト・参考書

1. テキスト

- ・ 使用しません。教材は適宜、授業サイト (<https://kariyach.github.io/teaching/RandQuarto/>) を通して配布します。

2. 参考書

- ・ キーラン・ヒーリー著、瓜生真也・江口哲史・三村喬生訳『データ分析のためのデータ可視化入門』(講談社、2021)
- ・ 松村優哉・湯谷啓明・紀ノ定保礼『改訂 2 版 R ユーザのための RStudio 「実践」 入門：tidyverse によるモダンな分析フローの世界』(技術評論社、2021)

- ・ 松村優哉・瓜生真也・吉村広志『R ユーザのための tidymodels[実践]入門：モダンな統計・機械学習モデリングの世界』

VIII. オフィスアワー等（学生からの質問への対応方法等）

- ・ 随時、対応します。授業の際、またはメールによる事前連絡にて、日時を調整します。

IX. その他

- ・ 履修条件：ありません
- ・ 適正人数：10名
- ・ 受講者調整方法：応募者多数の場合は抽選とします。

その他履修上の注意事項や学修上の助言

R 言語に初めて触れる方で受講を検討している方は、まず「統計言語「R」の神はなぜ無償で貢献したのか：オープンコミュニティで活躍する“新人類”の誕生」（日経ビジネス・電子版）を読んでみるとよいでしょう。また検索サイトに検索ワードとして「ggplot」と「興味のある専門領域」を入力してみると、この授業のイメージを掴めると思います。

R および R Studio は無料で利用できます。この他、この授業で使用するソフトウェア、パッケージ、データに利用料はかかりません（かかる費用は教科書代のみです）。

持参する PC は、Windows でも Mac でもかまいませんが、R Studio をインストールできる環境が必要です。[公式サイト](#)から確認してください。ここ数年、販売された PC であれば問題なく動作するはずです。

X. 特記事項

この授業は「データサイエンス特別プログラム」の対象科目です。同プログラムについては[リンク](#)を参照下さい。