

RとQuartoではじめるデータサイエンス：データを可視化する《2025》

シラバス

荻谷 千尋

1, Apr, 2025

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。

この授業の大きな目標は、受講生が、教員が提供するデータセットのみならず、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ようになることです。

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

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、テキスト『データ分析のためのデータ可視化入門』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。

評価の割合

- 宿題
 - 授業の感想：30%
 - 演習：30%
- レポート：20%
- プレゼンテーション：10%
- 受講態度：10%

Warning

※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

教科書・参考書補足

教科書

- キーラン・ヒーリー（2021）『データ分析のためのデータ可視化入門』（瓜生真也・江口哲史・三村喬生 訳）、講談社（LINK）

Important

教科書『データ分析のためのデータ可視化入門』は必ず購入してください。購入先はどこでも構いません。セカンドハンドも可。受講者数が少ないため、生協の「教科書」には指定していませんので、ご注意ください（生協で購入する場合は、各自取り寄せ注文すること）。

参考書

- 松村優哉・湯谷啓明・紀ノ定保礼・前田和寛（2021）『改訂2版 Rユーザーのための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（LINK）
- 松村優哉・瓜生真也・吉村広志（2023）『Rユーザーのためのtidymodels「実践」入門：モダンな統計・機械学習モデリングの世界』、技術評論社（LINK）
- 轟亮・杉野勇・平沢和司編『入門・社会調査法〔第4版〕：2ステップで基礎から学ぶ』（LINK）

その他履修上の注意事項や学習上の助言

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に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。

予習に関する指示

事前に教科書を下読することが望ましいです。

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。

ループリック

- ループリック