

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《2024》

役立つサイト

荻谷 千尋

12, Jun, 2024

Note

随時更新します

Warning

以下のサイトは有益ですが、今となっては古い情報も交じっています。まずは指定の教科書をしっかりと理解することから始めましょう。いずれ教科書も古くなりますが、教科書内においては、相互に矛盾する記述はないため、理解しやすいです。

0. 教科書

- 《訳書》『データ分析のためのデータ可視化入門』
- 《原著》Data Visualization: A practical introduction
- 《書評》『データ分析のためのデータ可視化入門』と最近の R グラフィックスパッケージ事情

1. 基本

Note

1から5までの基本的な内容は、この授業内で理解して欲しい範囲です（3の(5)と(6)は除く）

1. 全般

- Posit Cheatsheets
- 私たちのR：ベストプラクティスの探求
- 実験データの解析はエクセルではなくRを使った方がその研究は成功する
- AirbnbのデータサイエンティストはなぜRが好きなのか？
- Statistics Globe
- Tidyverse Skills for Data Science -Introduction to the pubh package

2. 初学者

- R初学者のためのtidyverse100本ノック
- R初学者からよく質問されること
- R Studio: Beginners
- Allison Horst

3. ggplot

(1) ggplot

- ggplot2(公式サイト)
- ggplot2 extensions - gallery（公式サイト）
- The R Graph Gallery
- RStudio - tidyverse, ggplot2を用いたデータ可視化の用例集
- 統計学2: ggplot2入門
- ggplot2の辞書
- 図から逆引きするgeom_*関数 (ggplot2)
- ggplot2における 軸関連のコントロール～逆引き的ミニまとめ～

- 【R】 ggplot2で相関行列ヒートマップを描く
- 複数の図をまとめて表示する際に便利なタグ
- ggplot2のstat_*()関数についてのまとめ

(2) アノテーション

- ggplot2のアノテーション技法
- ggplot2を使って、注釈を入れる-1
- FAQ: Annotation
- Beautifully Annotated: Enhancing your ggplots with Text
- Rの{gghighlight}で折れ線グラフをいろいろな条件でハイライトするグループを決定したいメモ

(3) カラーパレット

- ggsci
- fishualize - Paracanthurus_hepatus
- R Color Palettes
- colorblindr
- カラーパレットで色を指定する (RColorBrewer, viridis, ggsciパッケージ) [R]

(4) テーマ

- ggthemes

(5) ラッパー

- ggmosaic (公式サイト)

(6) アニメーション

- 【R, gganimate】 COVID-19感染者数Top 5をアニメーションで可視化する

4. テーブル

- Introduction to Creating gt Tables (gt公式サイト)
- 【R】 ガチできれいなTableを作れるgtパッケージとは?
- 【R】 データ要約ガチ勢のためのgtsummaryで表を書こう
- 回帰モデルの結果をキレイに表示する (gtsummaryパッケージtbl_regression関数) [R]

5. 出力方法

(1) 総論

- R Markdown クックブック
- .qmd を超お手軽完全無料パブリッシュできる Quarto Pub を試してみる
- Quarto Pub

(2) pdf

- TeX Live
- TeX Live ガイド (日本語)
- TinyTeX

(3) ウェブサイト

- Create Your Website with Quarto

(4) ダッシュボード

- How to create a dashboard using Quarto
- Welcome to the Shiny Gallery!
- R Shiny Dashboard Portfolio Project | 1973 USArrests Dashboard | R Shiny Tutorial for Beginners
- How the “Clusterbuster” Shiny App Helps Hundreds of Doctors and Epidemiologists Battle COVID-19 in the Netherlands
- Our Big Mac index shows how burger prices differ across borders

6. マップ

- leafletではじめるRによる地図プロット はじめに
- Leaflet Map View
- Leaflet Maps with CSV Data ・ 公式サイト
- ggmap ・ 公式サイト
- Thematic maps with cartography ・ 公式サイト
- Rで国土地理院 基盤地図情報データを扱う
- ggplot2 で沖縄をずらして日本地図を描きたい
- [小ネタ] R でコロプレス図 (色分け地図) をなるべく簡単に描く

7. コミュニティ

- r-wakalang へようこそ
- Tokyo.R
- R-bloggers

8. インストール他

- WindowsでのR/RStudioのインストール
- localeを日本語(ja_JP.UTF-8)にする
- ものすごくどうでもいい覚え書き: RStudioのエディタ ・ テーマ総ざらえ
- Modern CSV ・ 公式

II. 研究分野別

1. 社会科学

- sjPlot
- 社会学のデータ分析でよくある処理
- 総選挙データの可視化
- quanteda：衆議院外務委員会の議事録
- 西山 他『計量経済学』のためのR
- Introduction to Econometrics with R
- RパッケージtidyquantとExploratoryで始める0からのファイナンス計量分析
- Julia Silge

2. 医学・生物学

- 疫学や公衆衛生に関わる業務のためのR
- 医学統計でよく使うRパッケージ
- Rtigraphではじめる生物ネットワーク解析
- 【RNA-seq】RNA-seq解析を徹底的に解説！Part2^{発現変動解析(DEG解析)}
- Bioconductor — Genomicデータ解析ツール群
- Qiime2とRを用いて菌叢の機能予測をする
- Interfacing Seurat with the R tidy universe
- Introduction to ‘gggenes’
- Prevalence of SARS-CoV-2 Variants of Concern in Aotearoa New Zealand

3. 心理学

- 無料統計ソフトRで心理学

4. 自然科学

- Democratizing Historical Weather Data Analysis with R
- 気象庁提供の潮位表、過去の気象データをRで読み込む
- 「昔はこんなに暑くなかった」をR言語で可視化する

5. スポーツ

- Using the R and the Tidyverse to analyze Joel Embiid’s Player Stats

Note

この他、専門分野ごとに多くの専門書が出版されています

III. 分析手法

1. ネットワーク分析

- bob3’s blog
- The Jose M Sallan static website

2. テキストマイニング

- RとMeCabによる日本語テキストマイニングの前処理
- quanteda（公式）

3. コレスポンデンス分析

- コレスポンデンス分析の同時布置図は本当に使えないのか？

4. クラスター分析

Hierarchical Clustering（公式サイト）

IV. データ入手

1. 日本政府統計の総合窓口（e-stat）

- Rで日本の統計データを効率的に取得しよう
- みんなのデータサイエンス：e-Stat
- Rでe-statからxls（xlsx）ファイルをダウンロードするときうまくいかなかったときに対処したメモ

V. 前処理（データ加工）

1. 前処理全般

- Rで読むExcelファイル
- 複数のファイルを一度に読みこむ方法
- Pivoting your tables with Tidyr: Part I
- 「Rによるデータクリーニング実践」をやってみた
- 君の知らない物語(あまり知られていないtidyverseの便利な機能)
- Style guide
- 住所や年号、漢数字のデータ操作を楽にするRパッケージをCRANに登録しました
- Rでデータフレーム ⇔ リスト の変換

2. 時間処理

- lubridate（公式サイト）
- 【R前処理講座32】{lubridate}：時間処理【tidyverse】
- lubridateパッケージ入門

3. 同一処理

- Rで `for` ループで大量に `{ggplot2}` でグラフを書きたい（`purrr::map()` も添えて）

VI モデル

- Tidymodelsをシンプルに使う
- 【R】tidymodelsとworkflowを中心に～機械学習のフレームワーク～(その1)
- tidymodelsとDALEXによるtidyで解釈可能な機械学習
- tidymodels（公式サイト）
- R言語の基礎 #04、データの回帰分析・相関分析(diamondsデータ)
- purrrとbroomで複数の回帰モデルを効率的に管理する
- 回帰分析ではlm()ではなくestimatr::lm_robust()を使おう / TokyoR100
- 固定効果モデルについて / パネルデータ分析
 - 【R】固定効果モデルの推定がめっちゃ速いパッケージ { fixest }
 - ggfixest

VII. データセット

- palmerpenguins（公式サイト）
- Tokoy.R #99 パーマーステーションのペンギンたち #1
- Basic data analysis with palmerpenguins
- Penguins Dataset Overview — iris alternative