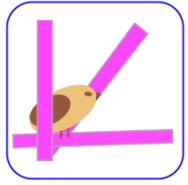


TRADのいま

データサイエンスコンソーシアム, 慶應義塾大学

柴田 里程



TRAD

- TextilePlot, R and DandD
 - TextilePlot
 - データの視覚表示
 - R
 - 計算エンジン
 - Data and Description
 - データの XML ファイル

データサイエンスにおけるソフトウェアの役割

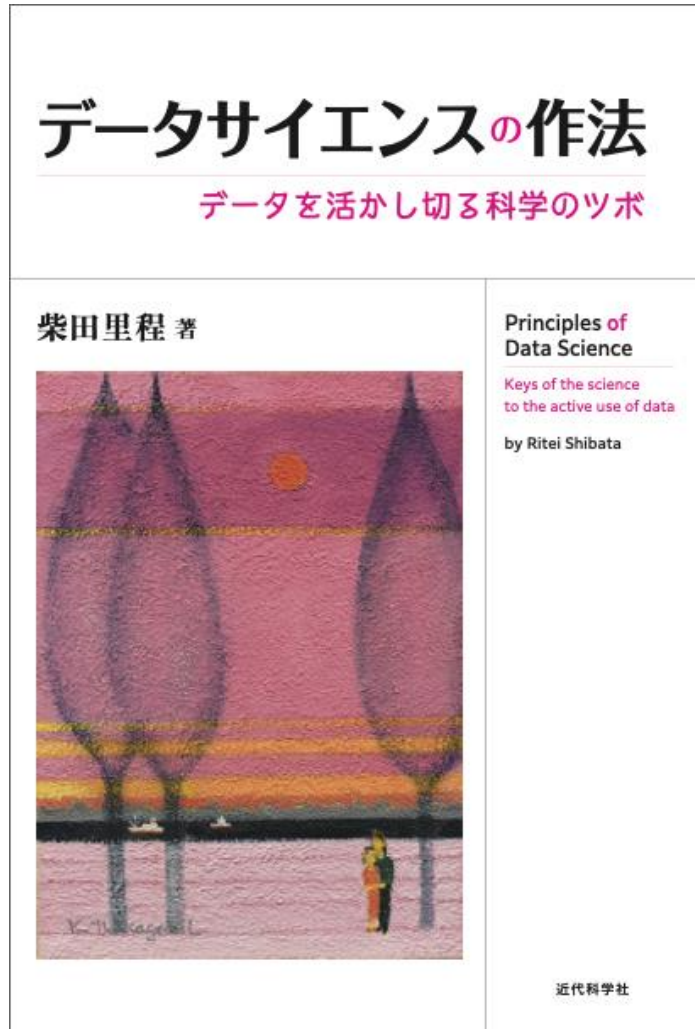
- 研究

- データサイエンスの骨格
- 課題の具象化
- 実証

- 実践

- Feel and Play
- 大規模, 大量データ
- 拡張性
- 汎用性

ソフトウェアへの姿勢



カミヤシロ

5つ星のうち3.0 TRADとは
2021年6月18日に日本でレビュー済み
Amazonで購入

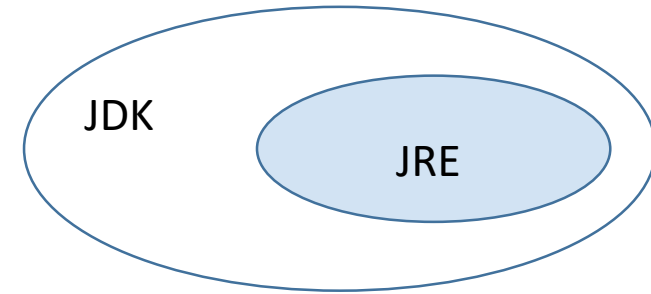
TRADというソフトで出来ることはパラレル・コーディネート・プロット（並行座標プロット）です。少しRに慣れていれば自分でも描けます。それより、簡潔に書かれた各章の「まとめ」が役立ちます。厚紙の「下敷き」のような形式でおまけにして欲しかったです。

できる, できない

TRAD の基本的な視覚表示であるテキストスタイルプロットは単なる並行座標プロットではありません. 確かに, 数値変量だけのときの単なる並行座標プロットなら R でも簡単に描けますが, 非数値変量も含む場合には, まずその値を軸上にどう配置するかから始めなくてはなりません. テキスタイルプロットは, 各軸の位置尺度の適切な選択や非数値変量の値の配置の問題を水平線規準を導入することで解決しています.

また, データテーブルのテキストスタイルプロットによる視覚表示は TRAD のほんの表面にすぎません. データと人間とのさまざまな対話で必要となる様々な機能の実装がメインです. このような機能の GUI を R 上に実装するのはほとんど不可能といってもよいでしょう. 大量大規模データもカバーするとしたら, なおさらです.

Java



- 汎用なオブジェクト指向言語
 - 世界中で広く用いられている(唯一の)言語
 - 家電製品にも組み込まれている本格的な言語
 - OS に依存しない汎用言語
 - オープンソース, 大量のライブラリー
- JDK → Oracle JDK → openJDK
 - TRAD プロジェクトもこの春に openJDK 16 に移行
 - 必要最小限の JAVA 実行環境(JRE)が作れるようになったため, あらかじめ JREをインストールする必要はなくなった (JRE 単独のダウンロードできなくなった)
 - Oracle JDK 8 : 200M → TRADに含まれる JRE: 45M

JRI

- Java R Interface
 - Java から R を呼び出すための Java プログラム
 - すでに開発は終了しているので, 自らメンテナンスする必要あり
 - 毎年春に行われる R のメジャーアップデートに合わせて修正
 - R 4.0.x → R 4.1.x (2021)
 - R のバグにも対処
 - R の日本語環境の不備にも対応

オープンソフトウェアを取り巻く環境

- オープンソフトウェア

- 誰でも自由に使える無償ソフト(野良ソフト)

- S → R → TRAD

- PHP, AWK, Java Script, Python

- UNIX → LINUX

- 逆風

- 周辺(大学, 研究機関)の無理解

- セキュリティーの強化

- 署名

- 認証 (Notalization)

- 怪しげなソフトウェアだという噂の流布

- 特定のサイト(マイクロソフトストア, アップルストア)からのダウンロードに制限

TRAD は実用にも耐える Feel and Play オープンソフト

- ダウンロード
 - <https://datascience.jp/TRAD.html>
 - Windows 用, MacOS 用
 - R は最新版まで対応
- 研究成果の集積
 - 最新の研究成果を反映するため, 日々アップデート
 - 随時ユーザからの意見を取り入れ改良
 - 詳細なヘルプ (RoboHelp) 12M
 - 多言語対応
 - ソースコード 64,353行

