

2023年6月15日  
データサイエンティスト集会 in VRC

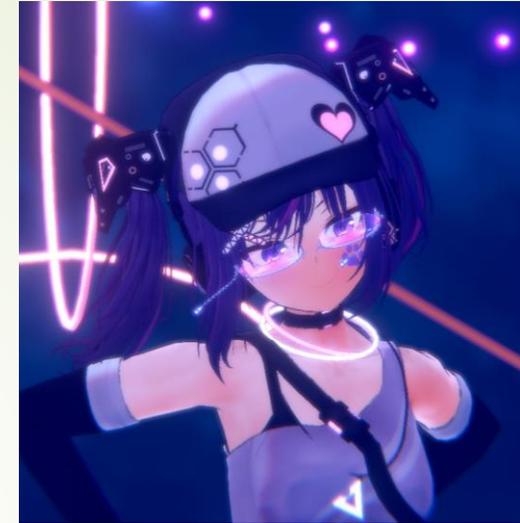
# JTCで役に立つデータ解析ソフト R AnalyticFlowの紹介

ぶんちん

1

# 自己紹介 ぶんちゃん

- ▶ 複合経営が特徴の企業（製造業）に所属
- ▶ データ分析担当者だったが。。。



## 背景

スマートな業界ではなく、

# 泥臭い The 重厚長大

- ▶ データの量はそこまで多くないが、かなり汚い
- ▶ 大規模設備産業なので、その使いこなし（改善）だけで十分な効果額が見込める
- ▶ 機械学習より基礎統計のほうが実は役に立つ場合が多い
- ▶ 物理理論やメカニズムベースの議論が中心で、そちらの専門家が大半

が前提のお話です。

以前のLTで紹介

4

# Orange Data Mining

<https://orangedatamining.com/>

- ▶ ビジュアルプログラミング的にデータ分析や機械学習モデル作成・評価が可能
- ▶ 公式HPから入手すれば無料で使用可能

▶ UIが素晴らしい

- ▶ 初心者は勉強に使おう！
- ▶ 専門家は手抜き・教育に使おう！

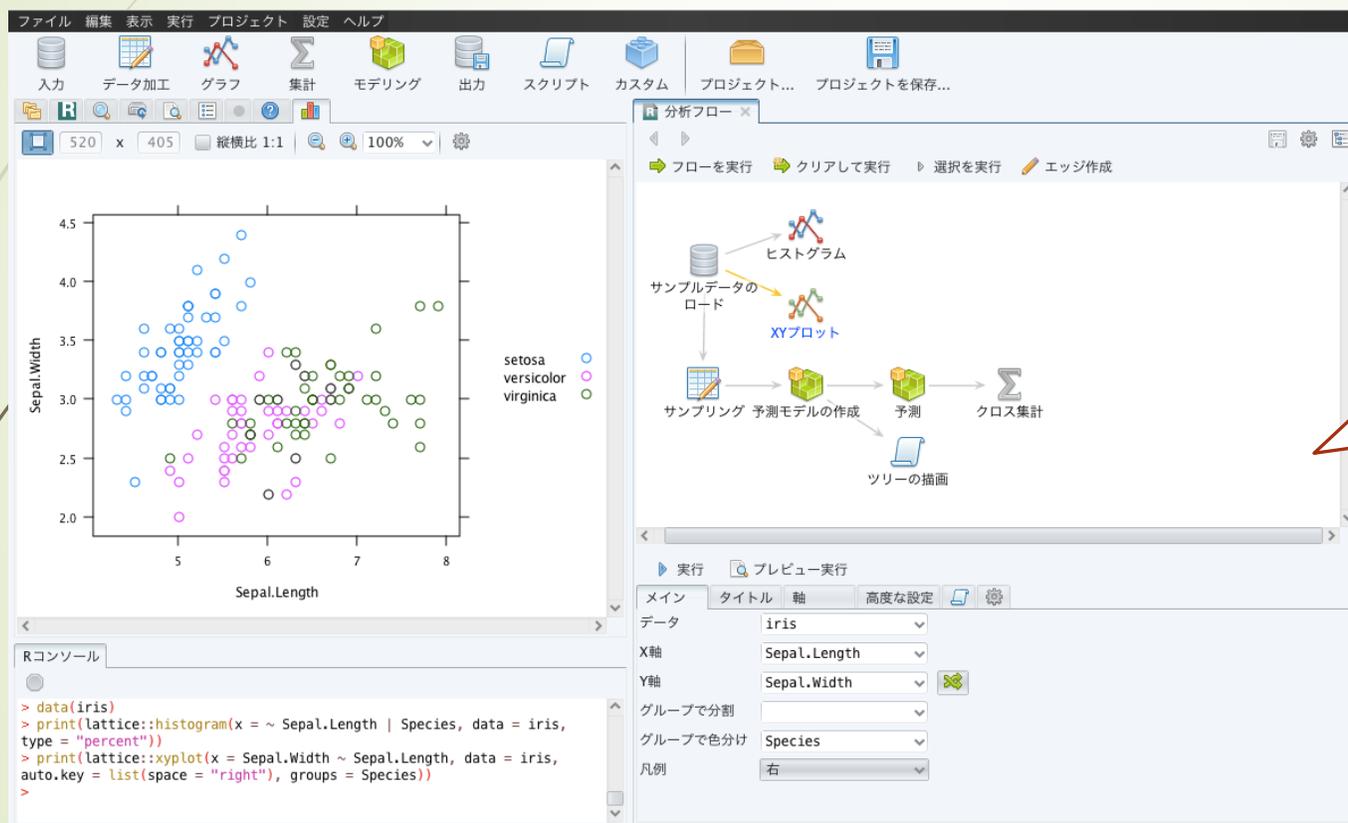
Model	AUC	CA	F1	Prec	Recall	MCC
Random Forest	0.988	0.953	0.953	0.953	0.953	0.930
Gradient Boosting	0.989	0.953	0.953	0.953	0.953	0.930
Tree	0.976	0.960	0.960	0.960	0.960	0.940

**UIが素晴らしいものの、完全に英語なので拒否感を持つ人がいる  
初期実装機能は優秀だが、凝ったデータハンドリングが苦手  
→ JTCの非専門家が使用するツールに向かない**

# R AnalyticFlow

<https://r.analyticflow.com/ja/>

私の仕事（データ分析教育）で  
メインに使っているソフトです



日本で作られたソフトで、  
UIやチュートリアルなど  
全てが日本語対応！

注意)  
最新版のRを使う際には  
CRANを"0-Cloud"に設定

Rベースで動く、GUI操作を基本とする“データ解析ソフトウェア”

# Orange Data Miningと比べて

- ▶ UIがやぼったい
- ▶ 初期実装機能が貧弱
- ▶ データ型の設定や欠損値などでエラーが生じやすい
- ▶ 技術的に高度な機能は特に実装されていない



R AnalyticFlowを使う必要ないのでは？

# R AnalyticFlowの本当の特徴

The screenshot displays the R AnalyticFlow interface. On the left, a scatter plot shows Sepal.Length on the x-axis (ranging from 5 to 8) and Sepal.Width on the y-axis (ranging from 2.0 to 4.5). Data points are colored by species: setosa (blue), versicolor (pink), and virginica (green). On the right, a workflow diagram shows the process: 'サンプルデータのロード' (Load sample data) leads to 'ヒストグラム' (Histogram) and 'XYプロット' (XY plot). 'XYプロット' leads to 'サンプリング' (Sampling), which then leads to '予測モデルの作成' (Create prediction model), '予測' (Prediction), and 'クロス集計' (Cross-tabulation). Below the workflow, there are options for '実行' (Execute) and 'プレビュー実行' (Preview execution). At the bottom, the R console shows the following code:

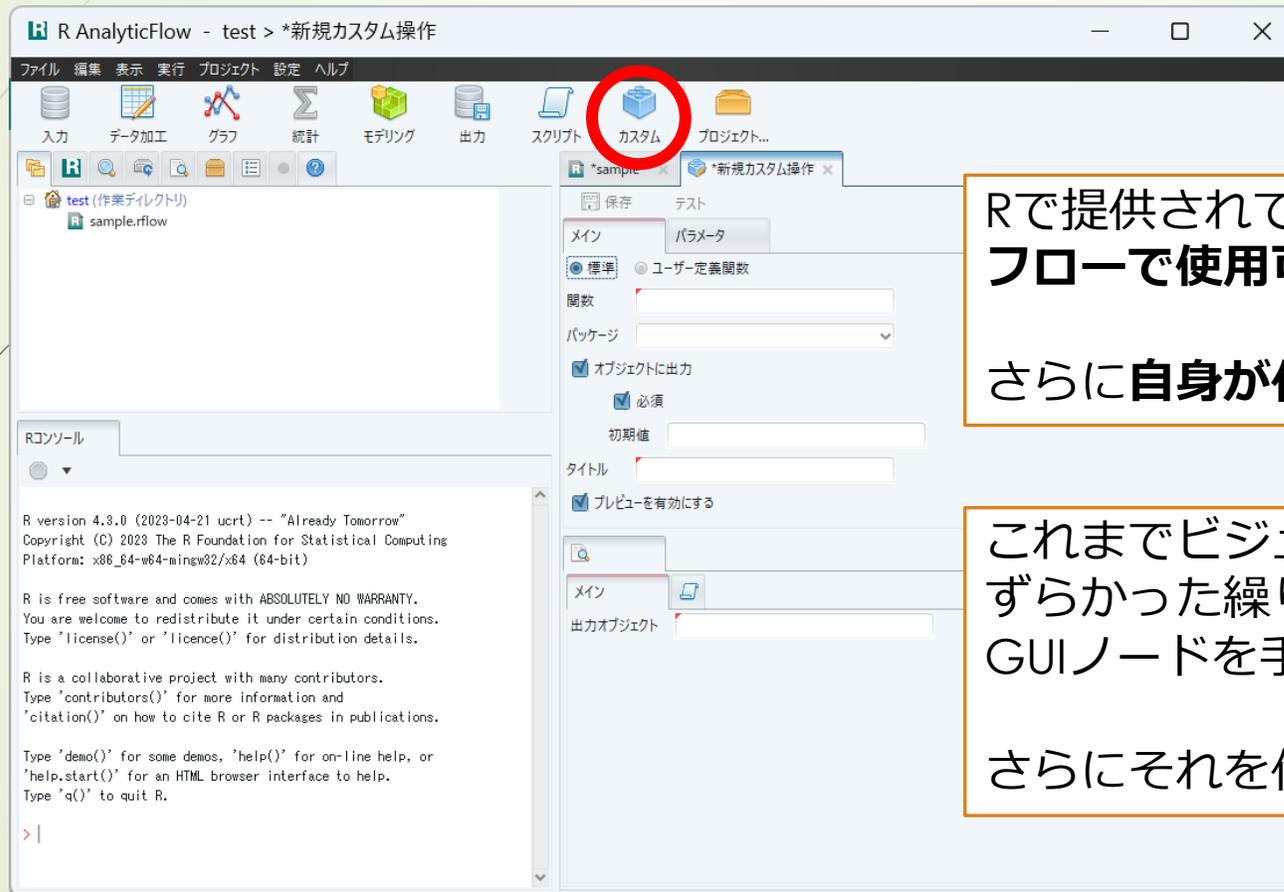
```
> data(iris)
> print(lattice::histogram(x = ~ Sepal.Length | Species, data = iris,
type = "percent"))
> print(lattice::xyplot(x = Sepal.Width ~ Sepal.Length, data = iris,
auto.key = list(space = "right", groups = Species))
>
```

右上で作成した分析フローを基に  
Rのコードが生成されて左下で実行される

フローの途中に自身のコードを挿入可能  
→**全てのRのライブラリを使用可能!**

Rベースで動く、GUI操作を基本とする“~~データ解析ソフトウェア~~”  
簡易的な“開発環境”

# オリジナルのUIノード作成可能



Rで提供されている全ライブラリ・全関数を  
フローで使用可能なGUIを簡易に作成可能！

さらに自身が作成した自作関数でも可能！

これまでビジュアルプログラミングでは扱い  
ずらかった繰り返し機能などの機能を持つ  
GUIノードを手軽に作成可能

さらにそれを他者に簡単に共有可能！

これまでGUI操作系のソフトでは扱いが難しかった組織に対して、  
Rでできる範囲は簡単にGUI作ってなんとかできるデータ分析ツール

## R AnalyticFlowの評価

- ▶ UIがやぼったい、初期実装機能が貧弱といった欠点はある
- ▶ 一方で、Rのコード・ライブラリを全て導入可能であり、それらをGUI操作可能にする機能が整備されていることによる機能の拡張性が優秀
- ▶ UIがすべて日本語！

すでに実力のある人にとっては有用ではないが、**初心者や非専門家で構成されている組織の実力向上に大きく貢献できるソフト**