

2010.11.27.SAT

Rによるデータ解析ツールの新提案

株式会社 ef-prime
鈴木 了太
suzuki@ef-prime.com

自己紹介

■ 鈴木 了太

- 株式会社ef-prime
 - 代表取締役 / データアナリスト
- 主な仕事
 - データ分析コンサルティング
 - プロジェクトマネージャー
 - ソフトウェアデザイン
 - 研究開発
 - 営業
- Rへの取り組み
 - パッケージpvclustの開発・メンテナンス(個人)
 - 関連学会への参加



自己紹介

■ 株式会社 ef-prime

- 2006年3月設立。所在地は東京都中央区
- 業務内容:
 - ・ 企業向けデータ分析コンサルティング
 - ・ ソフトウェア受託開発
 - ・ データ分析トレーニング
 - ・ その他ソフトウェアの開発、公開
- Rへの取り組み
 - ・ フローチャート式GUI 「R AnalyticFlow」の開発・公開
 - > 開発はチームで行っています
 - ・ R Foundationへの(ささやかな)寄付
 - > Supporting Institutionsに登録



本日の内容

- フローチャート式GUI 「R AnalyticFlow」
 - R AnalyticFlow とは？
 - 基本的な使い方
 - 新機能の紹介
- おまけ



R AnalyticFlow

R AnalyticFlow とは？

■ 概要

- フローチャート形式のR GUI フロントエンド
 - Javaで開発、RとはJRI経由で接続
 - 各種オープンソースソフトウェアで構成
 - > R AnalyticFlow自体もオープンソース(BSDライセンス)
- マルチOS、多言語対応
 - Windows / Linux / MacOS Xに対応
 - 日本語または英語が選択可能
 - 世界各国で利用
 - > ドイツ、フランス、ニュージーランド、ケニヤ、インド…
- 入手方法
 - 以下のサイトからダウンロード

<http://www.ef-prime.com/>



R AnalyticFlow とは？

■ スクリーンショット

The screenshot illustrates the R AnalyticFlow environment, which integrates R's analytical power with a graphical workflow interface.

Top Right: A window titled "R Graphics: Device 2 (ACTIVE)" displays a 4x4 grid of scatter plots for the Iris dataset, comparing Sepal.Length, Sepal.Width, Petal.Length, and Petal.Width across three species (versicolor, virginica, setosa).

Middle Left: The main interface shows a flowchart for an "IrisAnalysis.rflow" project. The workflow starts with "data" (represented by a cylinder icon) pointing to "plot" (triangle icon). "plot" also receives input from "boxplot" (triangle icon). The output of "plot" goes to "text" (triangle icon), which then connects to "xtabs" (cube icon). Along the path, there are nodes for "library" (star icon) and "rp<-rpart" (cube icon). A callout bubble points to this area with the text "分析フローを作成、クリックして実行" (Create a workflow, click to run).

Bottom Left: An "R console" window shows the R code used to generate the analysis. The code includes loading the iris dataset, fitting a rpart model, creating plots, and generating frequency tables.

```

library(rpart)
rp <- rpart(Species ~ ., data = iris)
plot(rp, margin = 0.2, brule = 1)
text(rp, fancy = T, all = T)
pred <- predict(rp, newdata = iris)
xtabs(~ Species + ., data = pred)

```

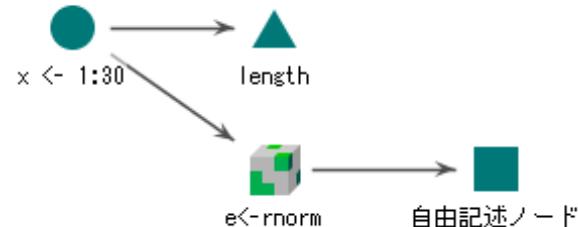
A callout bubble points to the R console with the text "コンソールも利用可能" (Console is also available).

Bottom Right: A callout bubble points to the "text" node in the flowchart with the text "分析結果を出力" (Output analysis results).

R AnalyticFlow とは？

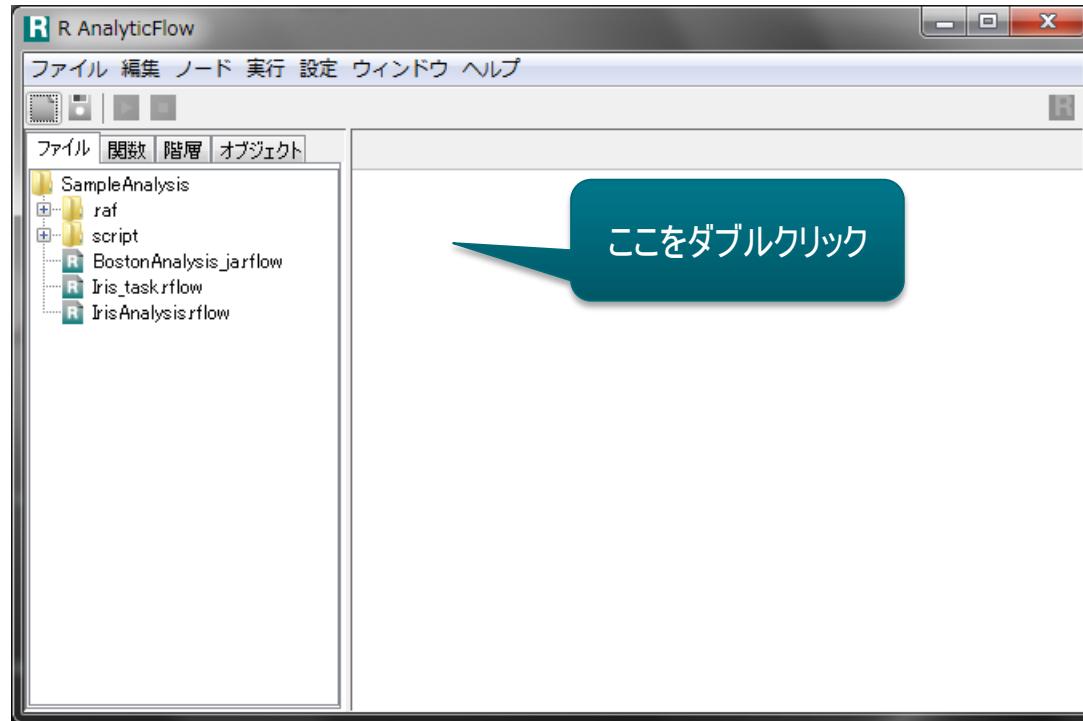
■ 特徴

- 分析過程をフローチャート形式で記述
 - ・ 思考が整理できる
 - > 分析の「本流」と「支流」を分離
 - ・ 作業グループでの共有がしやすい
 - ・ 再実行がカンタン
 - > Rの使い方を知らなくても、右クリックして「実行」するだけ！

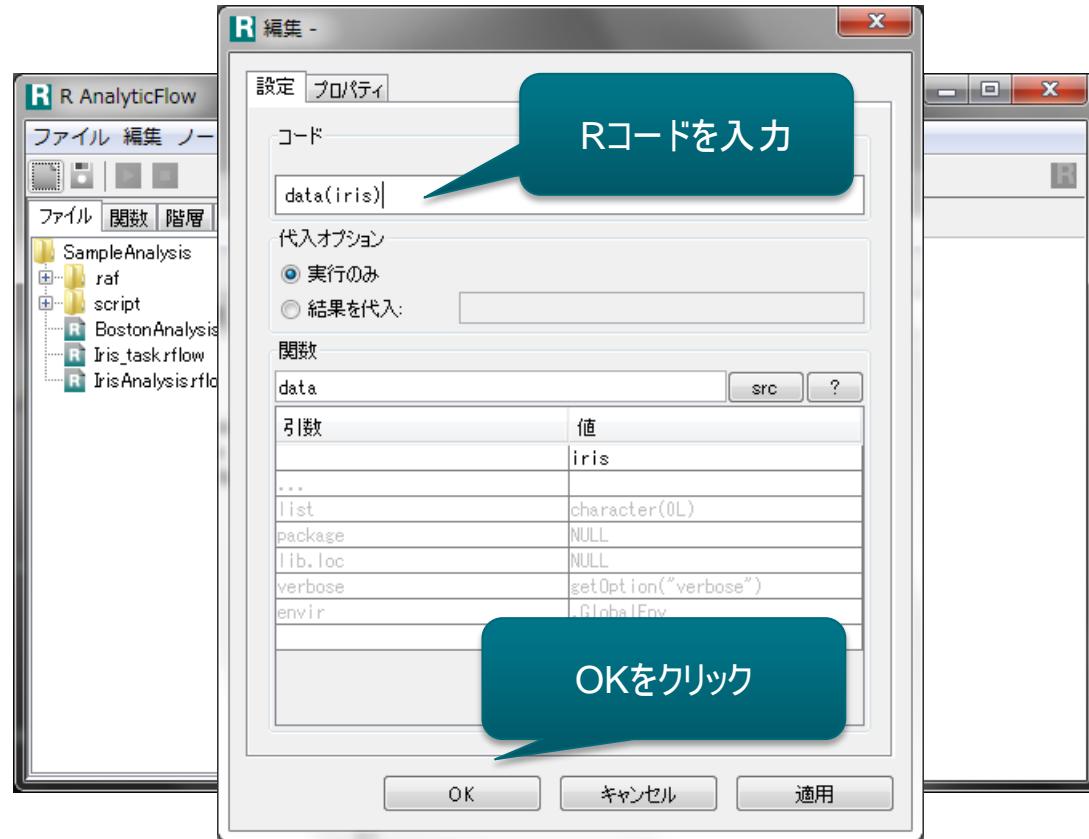


- 小回りが利く
 - ・ コンソールからの実行も可能
 - > ちょっとした確認などはコンソールで
 - > コンソールの実行結果からフローを作ることもできる

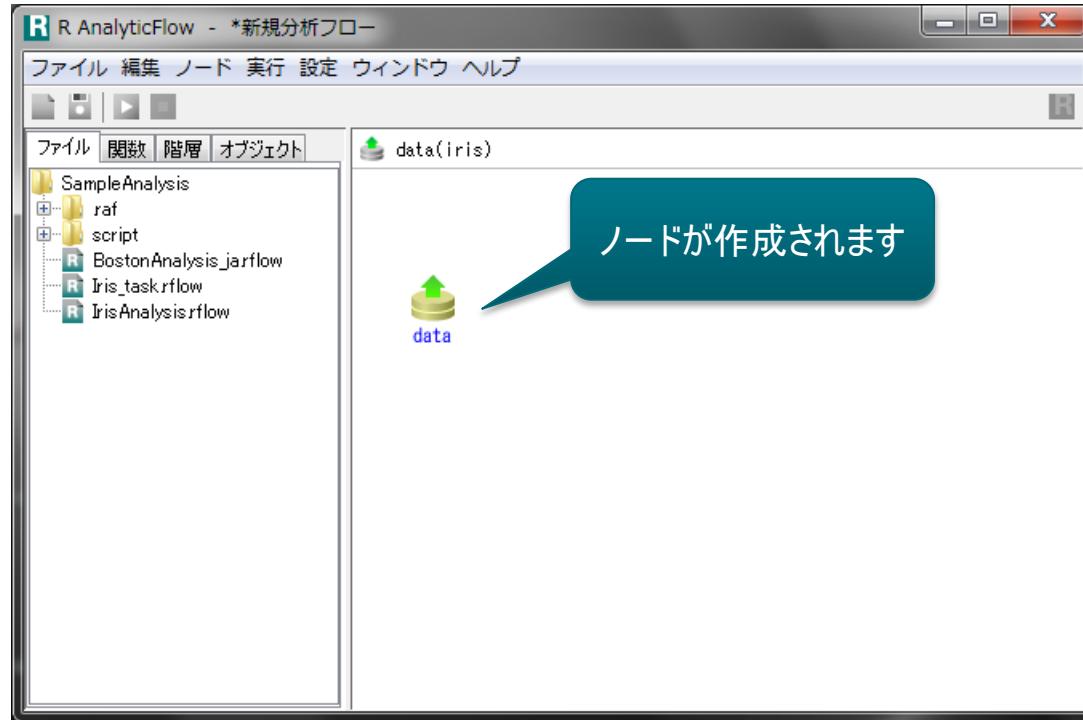
基本的な使い方



基本的な使い方

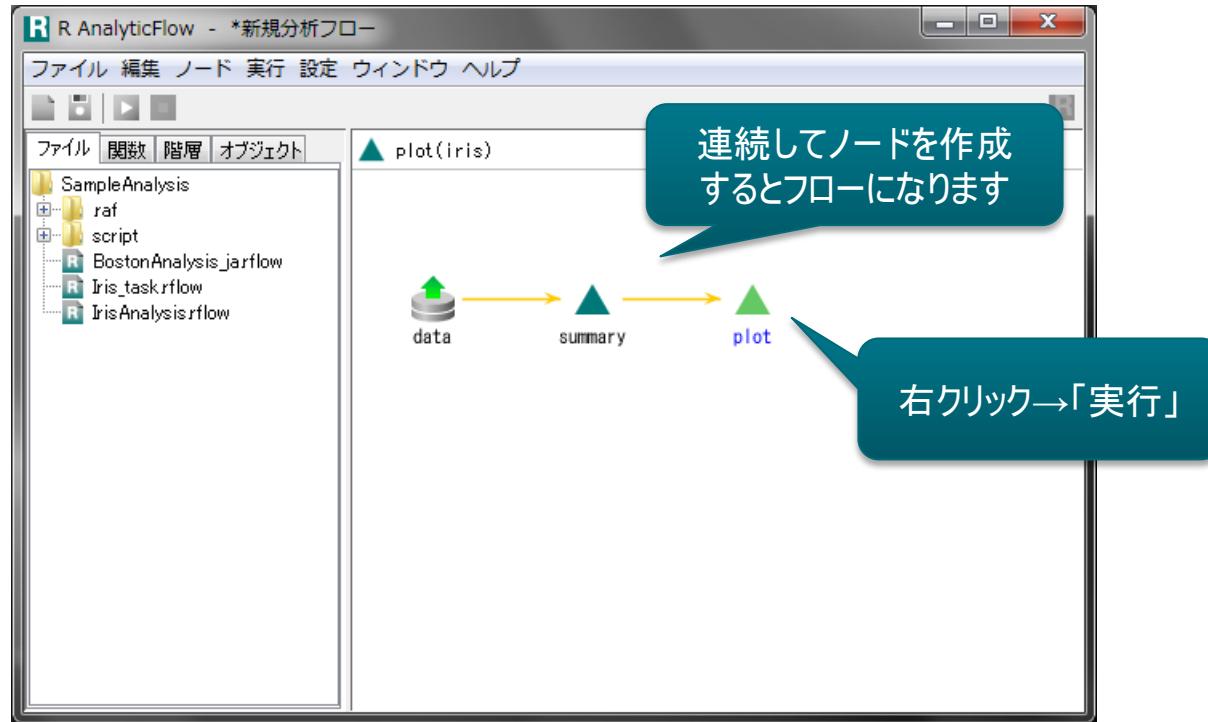


基本的な使い方



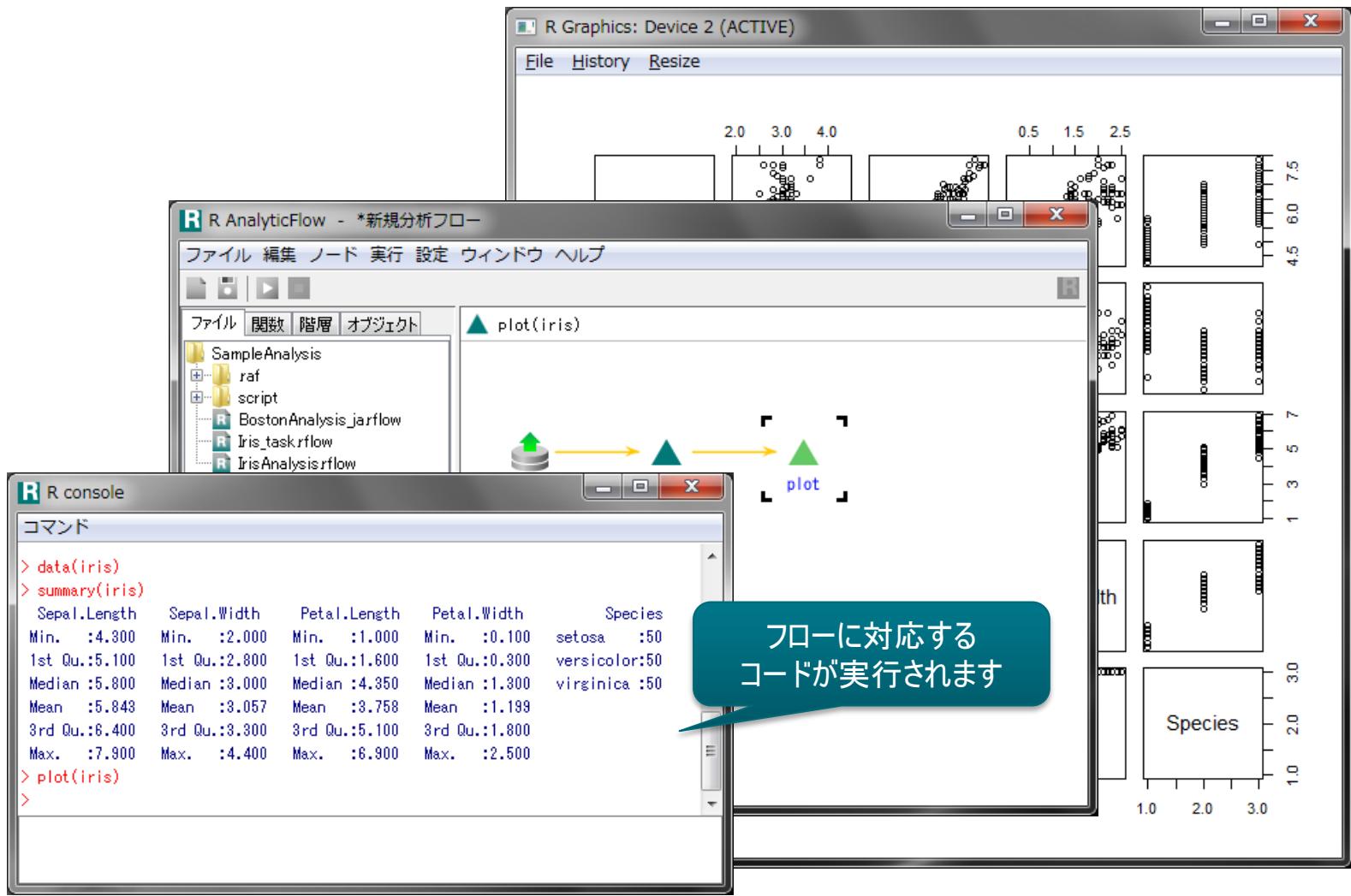
※右クリック→「一行ノードを作成」、またはCtrl + Enterでも
同様にノードを作成することができます。

基本的な使い方



※ノード間の接続は、接続元のノードを選択して接続先のノードを中クリック、またはAlt + クリックでも可能です。エッジ(矢印)は右クリック→「エッジの消去」で消去できます。

基本的な使い方

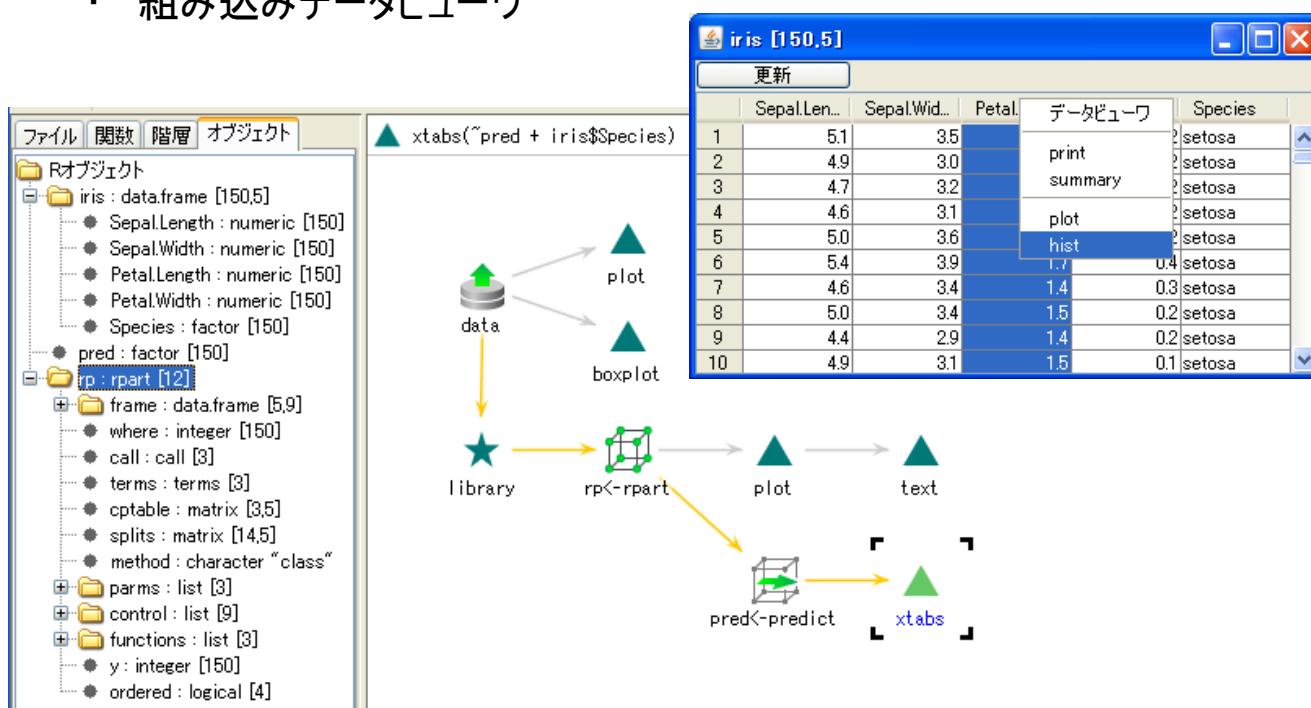


※「クリアして実行」を使うと、実行前にRオブジェクトを全て消去します。

便利な機能

■ オブジェクトブラウザ

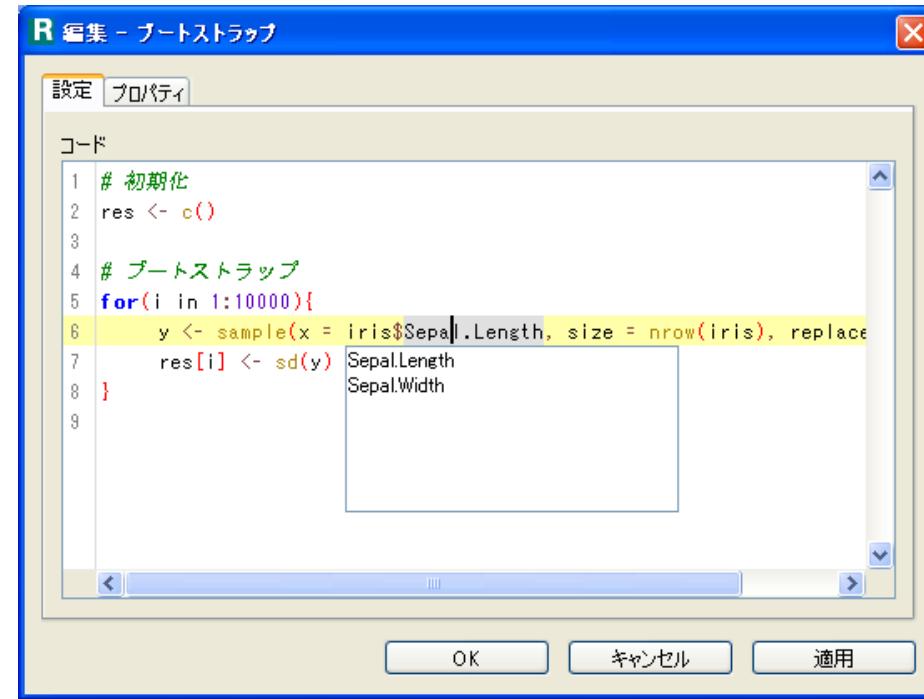
- ワークスペース内のオブジェクトをツリー形式で表示
 - ・ オブジェクト内の階層もツリー形式
 - ・ マウス操作でプロットなども可能
 - ・ 組み込みデータビューア



便利な機能

■ 自由記述ノード

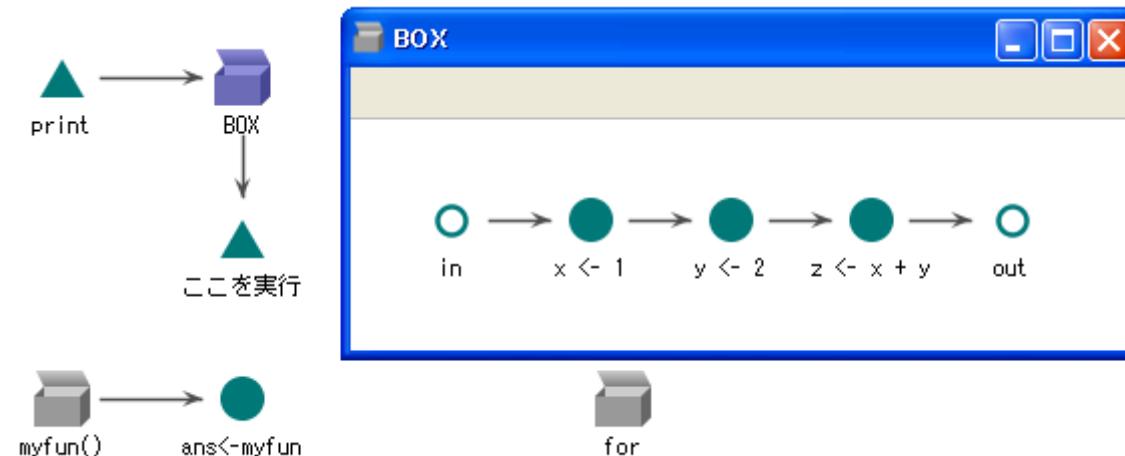
- 複数行のコードを記述可能
- コードの色分け、ハイライトなどが可能なコードエディタ
 - ・ オープンソースのJavaライブラリ「RSyntaxTextArea」により実現



便利な機能

■ ボックス

- 部分フローを格納できる特殊なノード
- 入れ子もOK
- 関数やループを表現可能な「特殊ボックス」



便利な機能

■ キャッシュ

- 実行結果を保存し、次回以降は自動的に結果を呼び出し
 - ・ 時間のかかる計算を何度も行わなくてよい
- 右クリックして設定するだけの簡単操作
 - ・ 上流のコードが変更されたり、実行結果ファイルが書き換えられるとキャッシュは自動的にリセットされる
 - ・ 「この.Rdata、いつ出力したんだっけ？」と気にしなくてよい



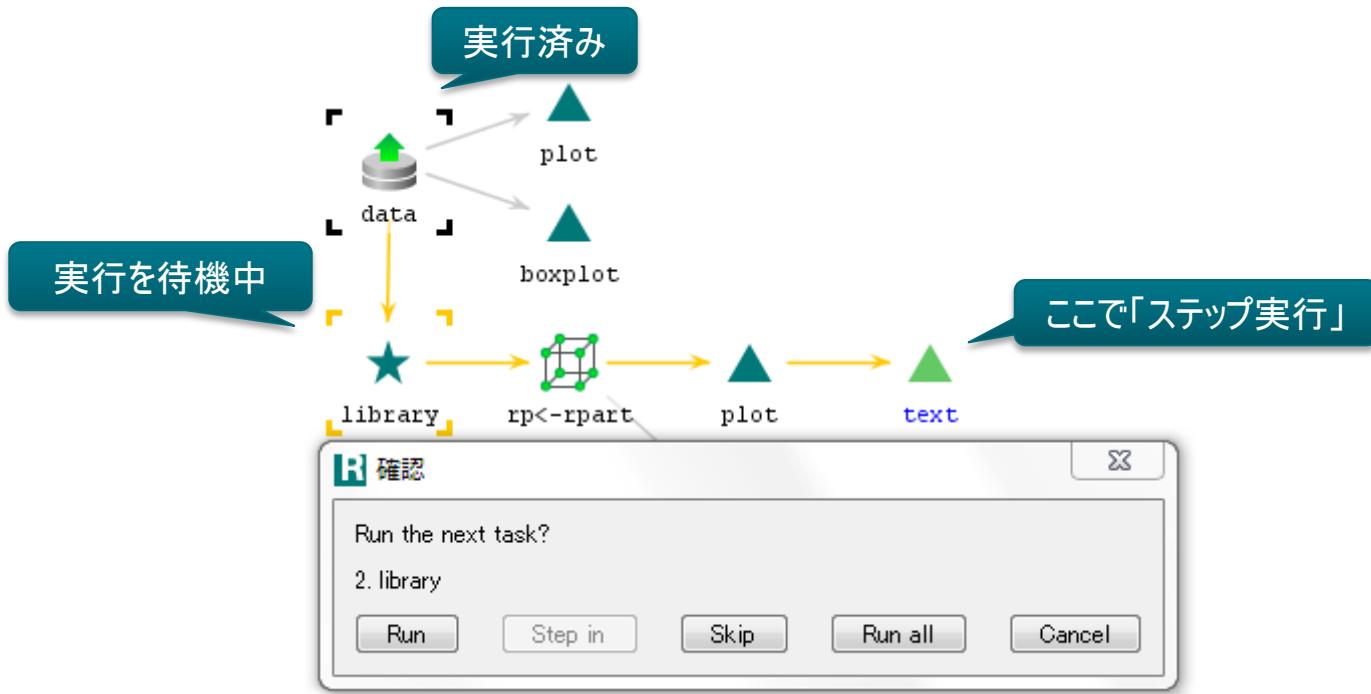
新機能の紹介



新機能の紹介

■ ステップ実行

- ノードをひとつ実行することに一時停止
- 停止中にオブジェクトの表示などの操作も可能
- デバッグに便利



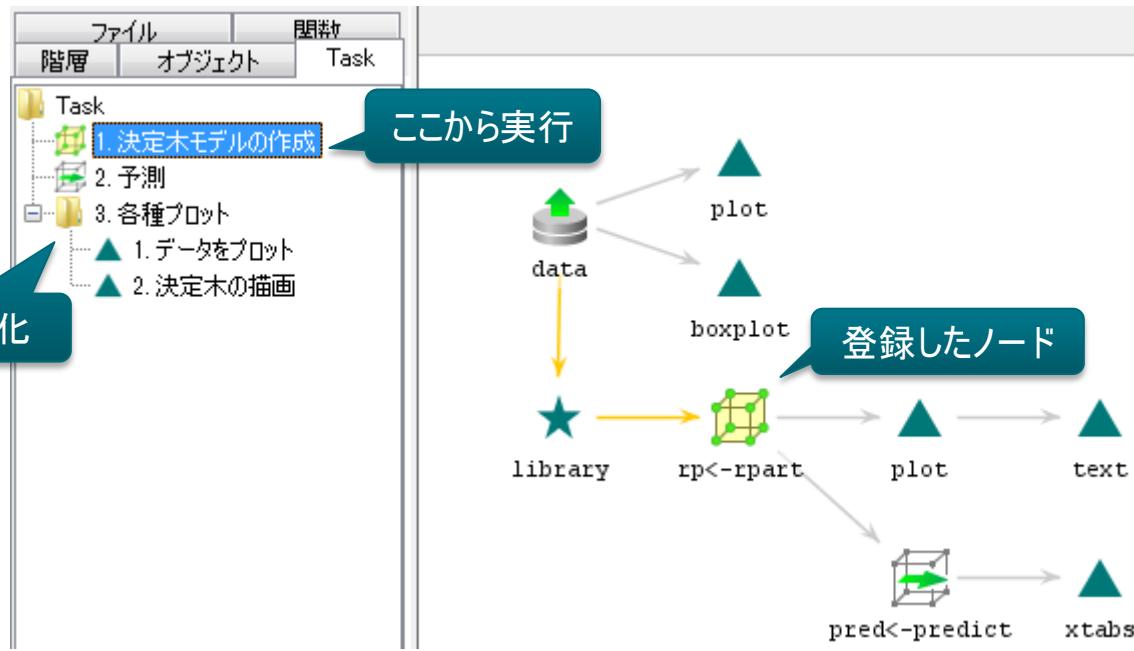
新機能の紹介

■ タスク

- よく実行するノードに名前をつけて登録
- タスク一覧から直接実行できる
- 複数のタスクをグループ化して連続実行
- シェルからのバッチ実行も可能



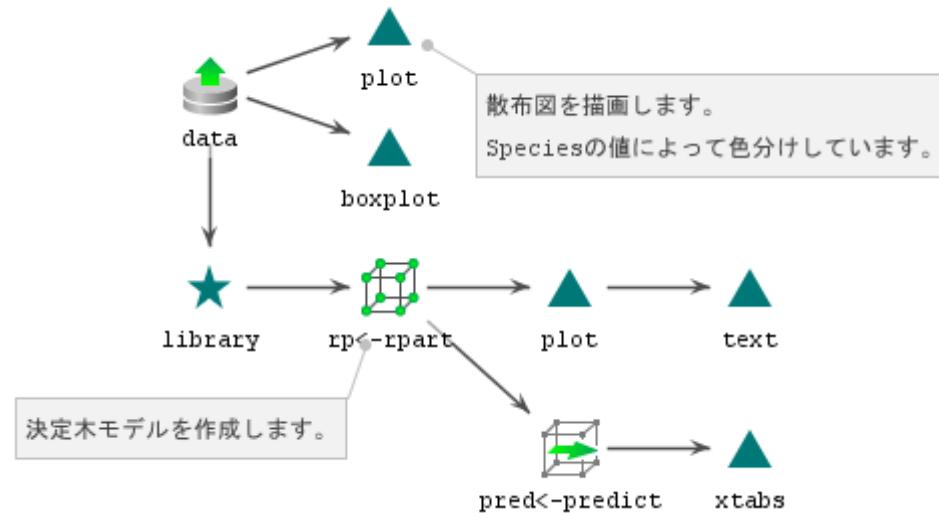
グループ化



新機能の紹介

■ コメント機能の拡張

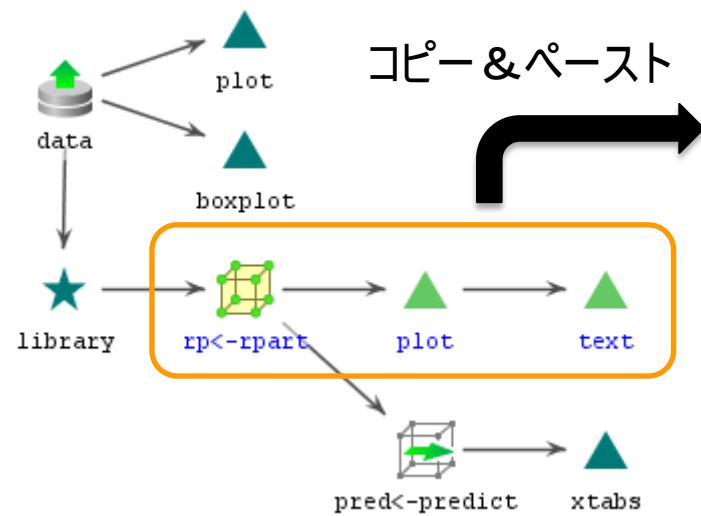
- 従来はマウスオーバーで表示されていたコメントを常に表示することが可能
- 作業メモやチーム内の共有に便利



新機能の紹介

■ エクスポート機能の拡張

- ノードを実行する代わりに、対応するRコードをファイルやクリップボードに出力
- フローをコピーしてRコンソールやテキストエディタに貼り付けると、対応するRコードが貼り付けられる



A screenshot of the R Editor window titled 'R 無題 - Rエディタ'. The menu bar includes 'ファイル', '編集', 'パッケージ', and 'ヘルプ'. The code area contains the following R script:

```
rp <- rpart(Species ~ ., iris)
plot(rp, margin = 0.2, branch = 0.3)
text(rp, fancy = T, all = T, use.n = T)
```

リリース予定

■ バージョン1.0.4

- 変更内容

- R 2.12.xに対応
 - > 2.12.0でのRの仕様変更に対応
 - > Window版はひとまず32bitのみ対応

- 12月上旬リリース予定



■ バージョン1.1.0

- 新機能

- ステップ実行
- タスク機能、バッチ実行
- コメント機能の拡張
- エクスポート機能の拡張

- 完成次第リリース

- できれば今年中…？

更新情報をチェック！

<http://www.ef-prime.com/>

または

「R AnalyticFlow」「ef-prime」で検索



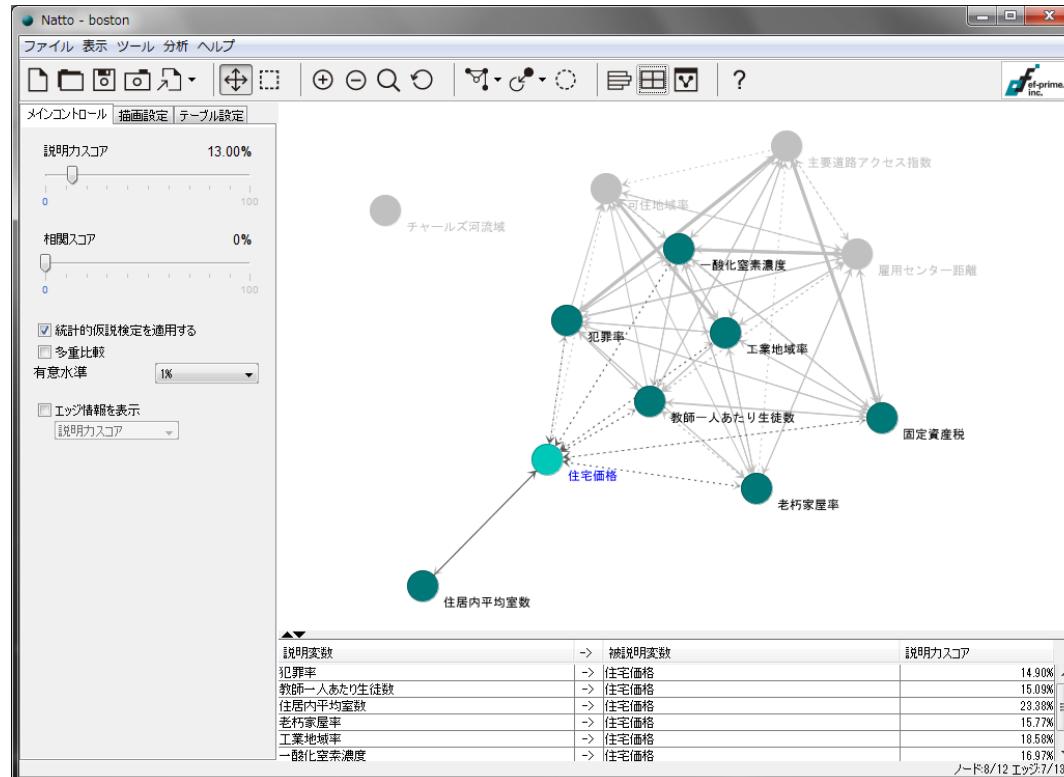
RSSで更新情報を配信しています。
RjpWikiでも情報公開させていただいております。

おまけ

探索的データ解析ツール 「Natto」

■ 特徴

- 変数間の相関関係を検出し、グラフィカルに表示
 - ・ 数値と文字の混合データに対応、欠損を含んでいても可
 - ・ 非線形の相関も検出可能



探索的データ解析ツール「Natto」

■ 特徴

- 数値を自動的に区間に分割し、クロス集計表を作成
 - ・ 相関関係の検出もこの離散化を利用
 - ・ ほか、独立性の検定や相関ルール探索機能など



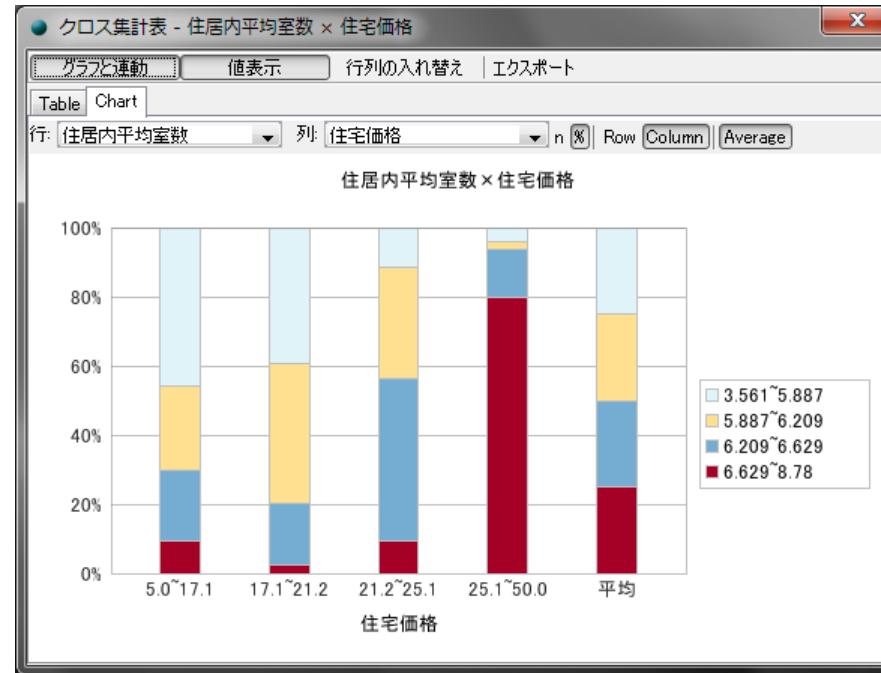
The screenshot shows a window titled "クロス集計表 - 住居内平均室数 × 住宅価格". The window has tabs for "グラフと運動" (Graph and Motion), "値表示" (Value Display), "行列の入れ替え" (Transpose Matrix), and "エクスポート" (Export). Below the tabs, there are dropdown menus for "行" (Row) set to "住居内平均室数" and "列" (Column) set to "住宅価格". There are also buttons for "全体%" (Total%), "行%" (Row %), "列%" (Column %), and "リフト" (Lift).

	5.0~17.1	17.1~21.2	21.2~25.1	25.1~50.0	計
3.561~5.887	58 45.7%	49 38.6%	15 11.8%	5 3.9%	127 100.0%
5.887~6.209	31 24.6%	50 39.7%	42 33.3%	3 2.4%	126 100.0%
6.209~6.629	26 20.5%	22 17.3%	62 48.8%	17 13.4%	127 100.0%
6.629~8.78	12 9.5%	3 2.4%	12 9.5%	99 78.6%	126 100.0%
計	127 25.1%	124 24.5%	131 25.9%	124 24.5%	506 100.0%

探索的データ解析ツール「Natto」

■ 開発中の新機能

- 積み上げ棒グラフ
 - ・ クロス集計をグラフィカルに表示
- エクスポート
 - ・ クロス集計表や積み上げ棒グラフを出力



ソフトウェアの入手＆更新情報

<http://www.ef-prime.com/>

または

「ef-prime」で検索

ご清聴ありがとうございました。

