

2016.11.27.SUN

# マウス操作でかんたん予測分析

使ってみよう！ R AnalyticFlow

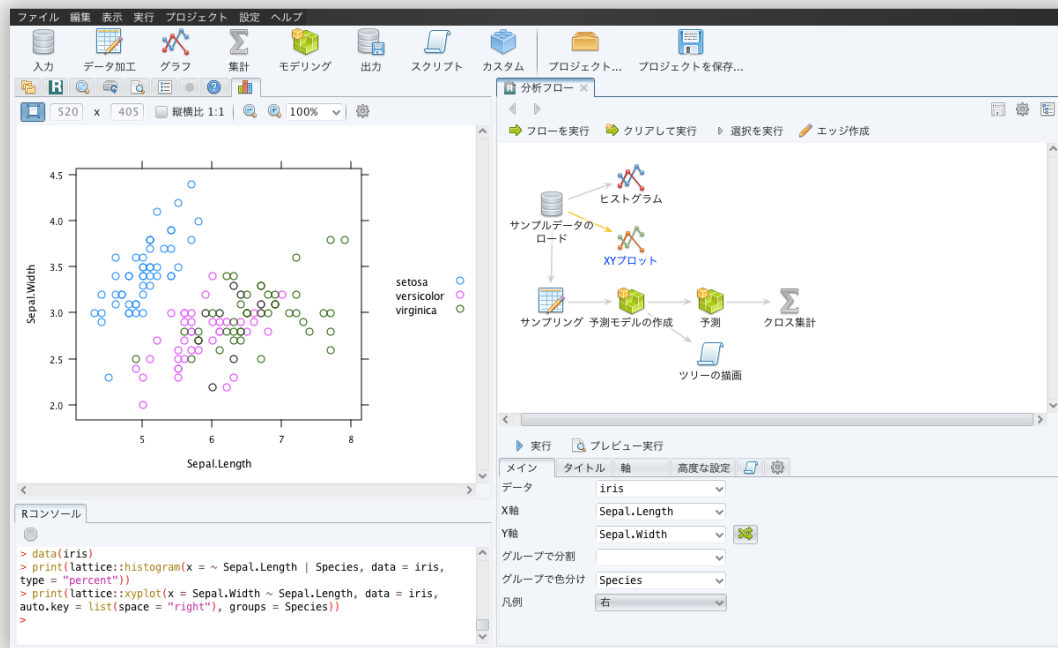
株式会社ef-prime  
鈴木 了太

# 自己紹介

## ■ 鈴木 了太

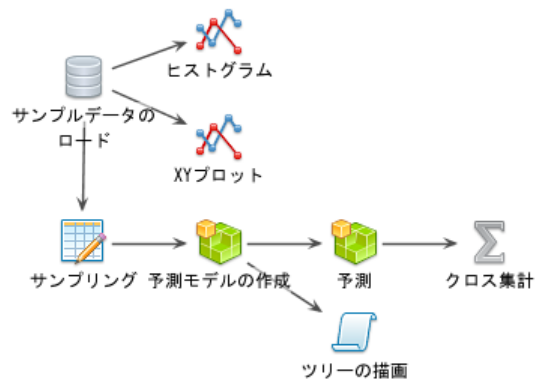
- 株式会社ef-prime代表
  - ・ 企業向けデータ分析コンサルティングなど(2006年～)
  - ・ 分析ツールの開発・公開も
- Rユーザー歴15年くらい
  - ・ 開発CRANパッケージ: pvclust
  - ・ 統数研のRユーザー会やUseR!にも出沒、Japan.R初体験
  - ・ 先日Tokyo.Rに初参加

# データ解析ソフトウェア R AnalyticFlow



# R AnalyticFlowとは

- データ解析のためのR GUI
  - 分析プロセスをワークフローで表現
  - オープンソース
  - Javaで開発、マルチOS対応
    - ・ Windows / Mac / Linux



# 欲しいものを作る

## ■ ビジネスデータの分析

- プログラミングの負担を最小限にし、データと分析プロセスそのものにフォーカス
- 作業者間でのコードの違いを最小化し、チームでの確認・共有をしやすいにする

## ■ その結果...

- Rの知識がなくても分析を始められる

# R AnalyticFlowの画面構成



分析機能呼び出すツールバー

分析フローやスクリプトの表示、編集および実行

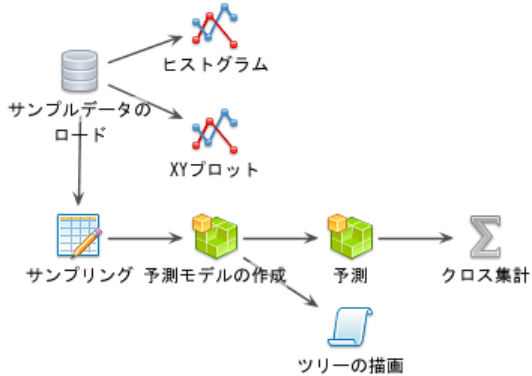
ファイルやRのオブジェクト、グラフィックス、ヘルプなどを表示

Rコンソール。ここで直接処理を実行することも可能

フロー上で選択した処理の詳細を表示、編集、実行

# R AnalyticFlowの特徴

分析フローを作成し、対応するRスクリプトを生成・実行



# 1. データの読み込み

```
data(iris)
```

# 2. 探索的分析

```
plot(iris[, 1:4], col =
as.integer(iris$Species) + 1)
boxplot(Petal.Length ~ Species, data = iris,
col = 3, main = "Petal.Length")
```

# 3. モデリング

```
library(rpart)
rp <- rpart(Species ~ ., iris)
```

# 4. モデルの確認

```
plot(rp, margin = 0.1, branch = 0.3)
text(rp, fancy = T, all = T, use.n = T)
```

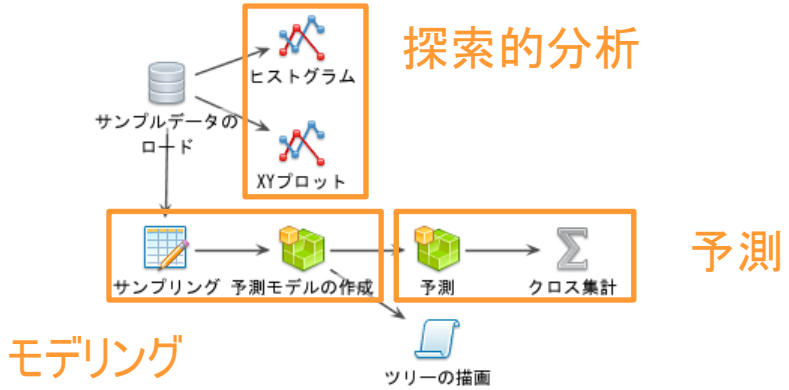
# 5. 予測および評価

```
pred <- predict(rp, type = "class")
xtabs(~pred + iris$Species)
```



# 分析フローの利点

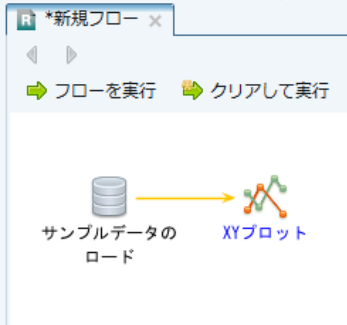
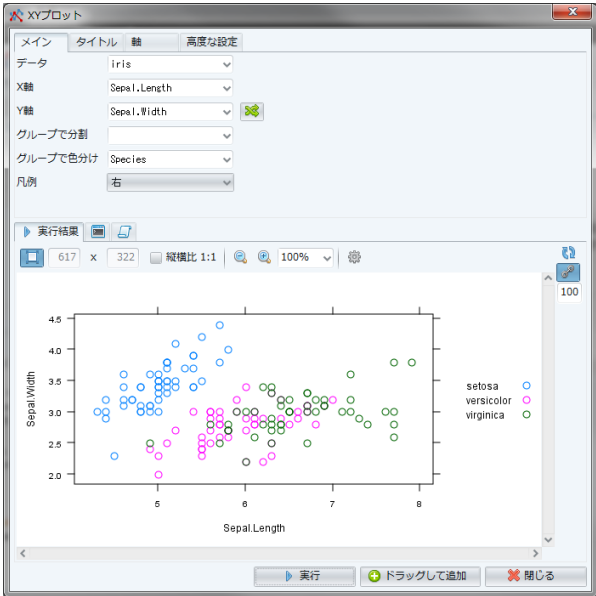
探索的分析からモデリング・予測に至る作業の流れを視覚的に整理して記述することができる





# 充実した分析GUI

簡単なマウス操作でRスクリプトを記述せずに分析が可能  
処理をドラッグしてフローに追加



# Rプログラミングのサポート

任意のRスクリプトを記述してフローに含めることが可能  
ハイライト表示やコード補完、ブレイクポイントの作成など



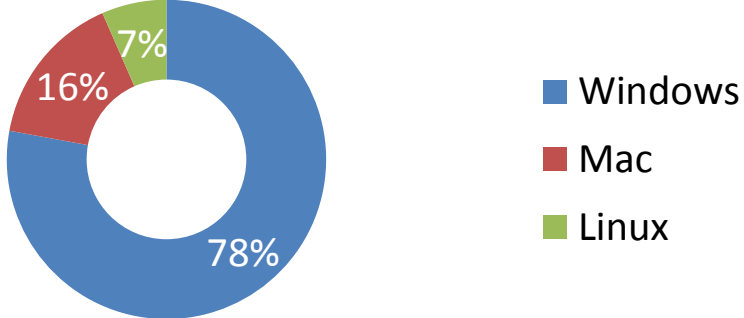
```
1 ## 予測モデルの作成と確認
2
3 # データの読み込み
4 data(iris)
5
6 # サンプリング
7 local({
8   x <- iri
9   smpl <- iris iris.mean iris ... 00, replace = FALSE)
10  assign(x, Ctrl + Spaceで一覧を表示), , drop = FALSE], pos = parent.frame(n = 2))
11  assign(x = "test", value = x[-smpl, , drop = FALSE], pos = parent.frame(n = 2))
12 })
13
14 model <- rpart::rpart(formula = Species ~ ., data = train)
```

# 多言語・マルチOS

表示言語として日本語と英語を選択可能  
世界80ヶ国でダウンロード(2016年11月時点、直近6ヶ月間)  
Windows / Mac / Linux をサポート



- 1. Japan
- 2. United States
- 3. India
- 4. Taiwan
- 5. United Kingdom
- 6. China
- 7. Germany
- 8. Italy
- 9. Spain
- 10. Australia



## 開発状況

### ■ 現バージョン(3.0.6)の特徴

- 多くの処理をGUIのみで実現することで、Rの知識を問わず分析が可能
- プレビューにより結果が見えるインターフェース、プロジェクト管理機能によりデータとプロセスを統合

### ■ 次バージョンの方針

- 実際のデータ分析で必要とされる機能を強化
  - ・ 予測分析、多変量解析、仮説検定など
- 安定性と利便性の向上

# 3.1.0

近日公開予定。安定性の向上に加え、分析機能を大幅に強化。  
予測モデルの選択などより実践的な応用がGUIのみで実現可能に



予測モデルの作成

実行 プレビュー実行 ドラッグして追加

メイン モデル 詳細設定 モデルの評価

多項ロジット  
決定木  
ニューラル

決定木 高度な設定

出力名 決定木

タイプ (自動)

木の剪定 自動 手動

複雑度/パラメータ >= 0.001

モデル詳細

分割前の最小サンプルサイズ 20

分割後の最小サンプルサイズ round(minsplit/3)

クロスバリデーションの分割数 10

木の深さ (最大値) 30

実行結果

- test.eval: list [3]
- train.models: list [3]
- 多項ロジット: multinom
- 決定木: rpart [14]
- ニューラル: nnet.form

print plot plotcp summary str

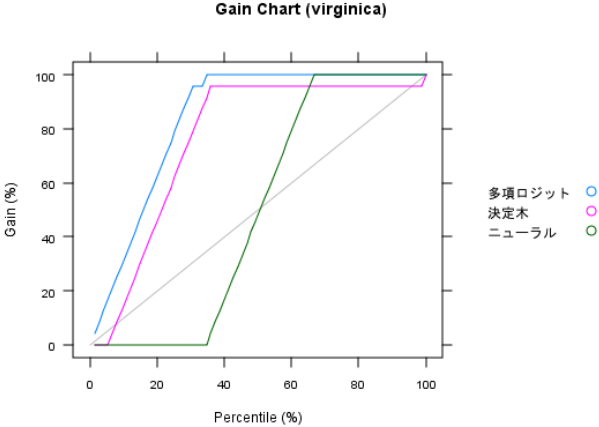
420 x 206 縦横比 1:1 100%

Petal.Length < 2.5

setosa

Petal.Length < 4.75

versicolor virginica



## 3.1の特徴(予定)

### ■ 分析機能の追加

- データ型の設定、欠損値の処理
- 予測分析(複数モデルの作成、モデル選択)
- 仮説検定(t検定、Wilcoxon検定、比率の検定)
- 多変量解析(主成分、クラスター)

### ■ 安定性の向上

- Rプロセスを分離し、道連れのクラッシュを防止

### ■ 利便性の向上

- プロジェクトのインポート・エクスポート
- 外部Rスクリプトファイルをフローに組み込み

## 3.2に向けての構想(一例)

### ■ Rパッケージの利用

- 独自関数のパッケージ化
  - ・ R AnalyticFlowのインターフェースに合わせた関数群
  - ・ より読みやすいコードを生成し、RScript等からも利用可能に
- 外部パッケージの利用
  - ・ データ読み込みにはreadr、ランダムフォレストにrangerなど  
高速で有用な外部パッケージを利用しやすい仕組みを構築

### ■ 課題:ポータビリティの担保

- 生成したスクリプトを外部で実行する際に  
パッケージ不足によるエラーを防ぐための仕組みが必要

ご清聴ありがとうございました



 <http://r.analyticflow.com>

 @efprime\_jp

 R AnalyticFlow