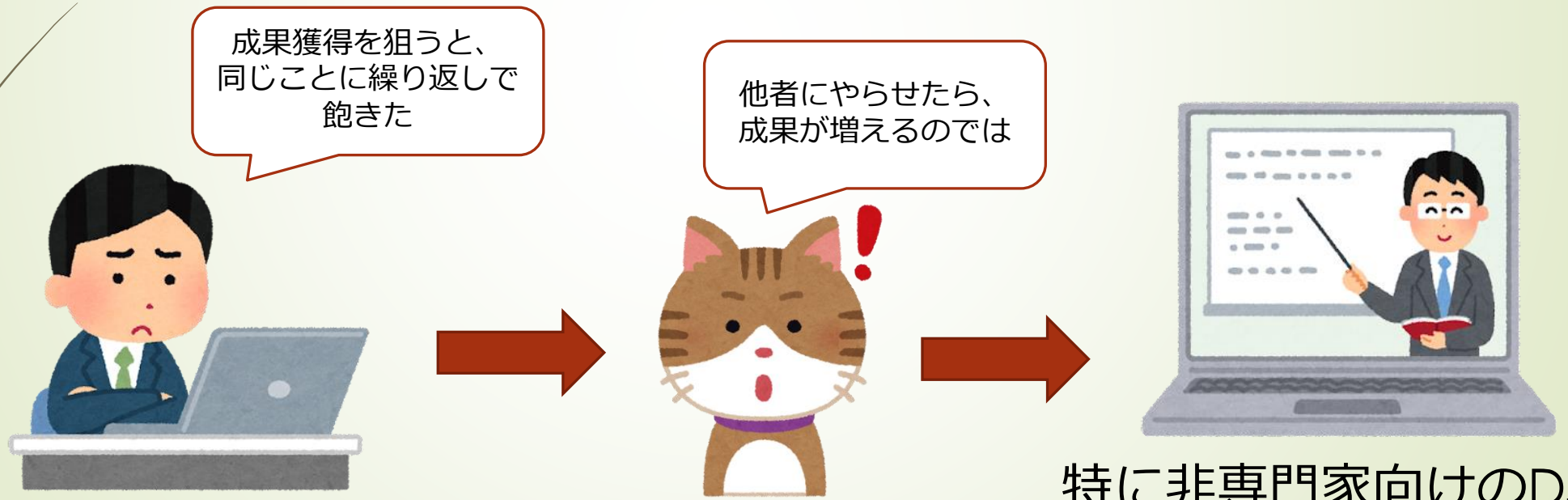
2023年6月1日
データサイエンティスト集会 in VRC

実務で使える Orange Data Miningの便利な機能

ぶんちん

自己紹介　ぶんちん

- 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。。。



特に非専門家向けのDS教育

Orange Data Mining

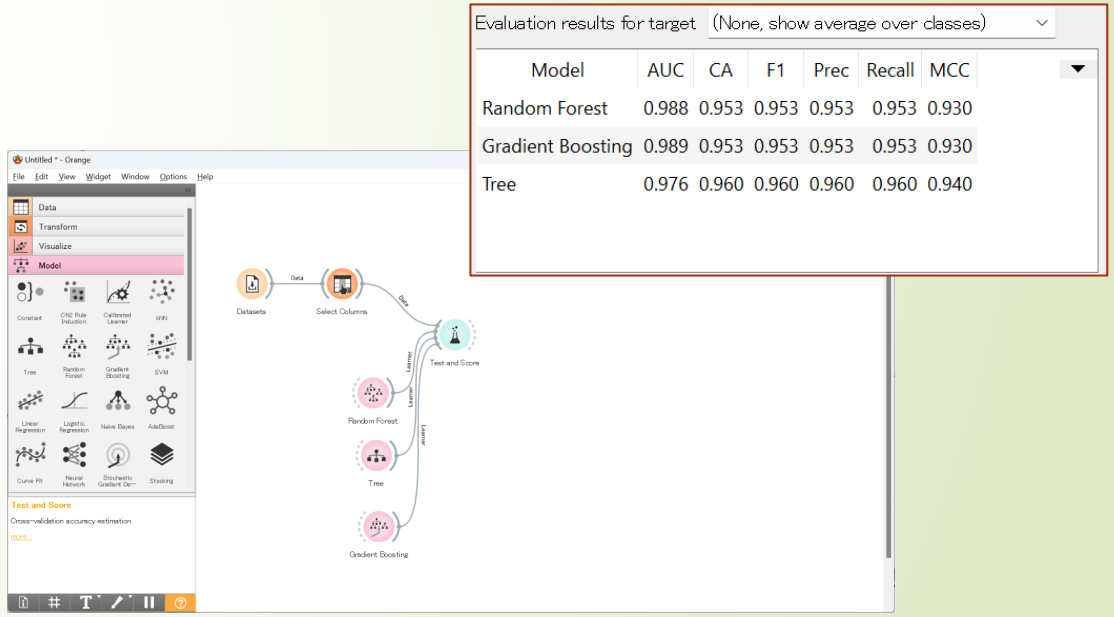
https://orangedatamining.com/

- 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！

基本的な使い方は前回紹介
スライド＆動画公開してます



GUI操作で分析・モデル作成が可能

便利な機能紹介

Orangeには様々な便利な機能が実装されています。

今回はその中から、ちょっと見ただけでは気づきそうにないけど便利なものを中心に紹介します。

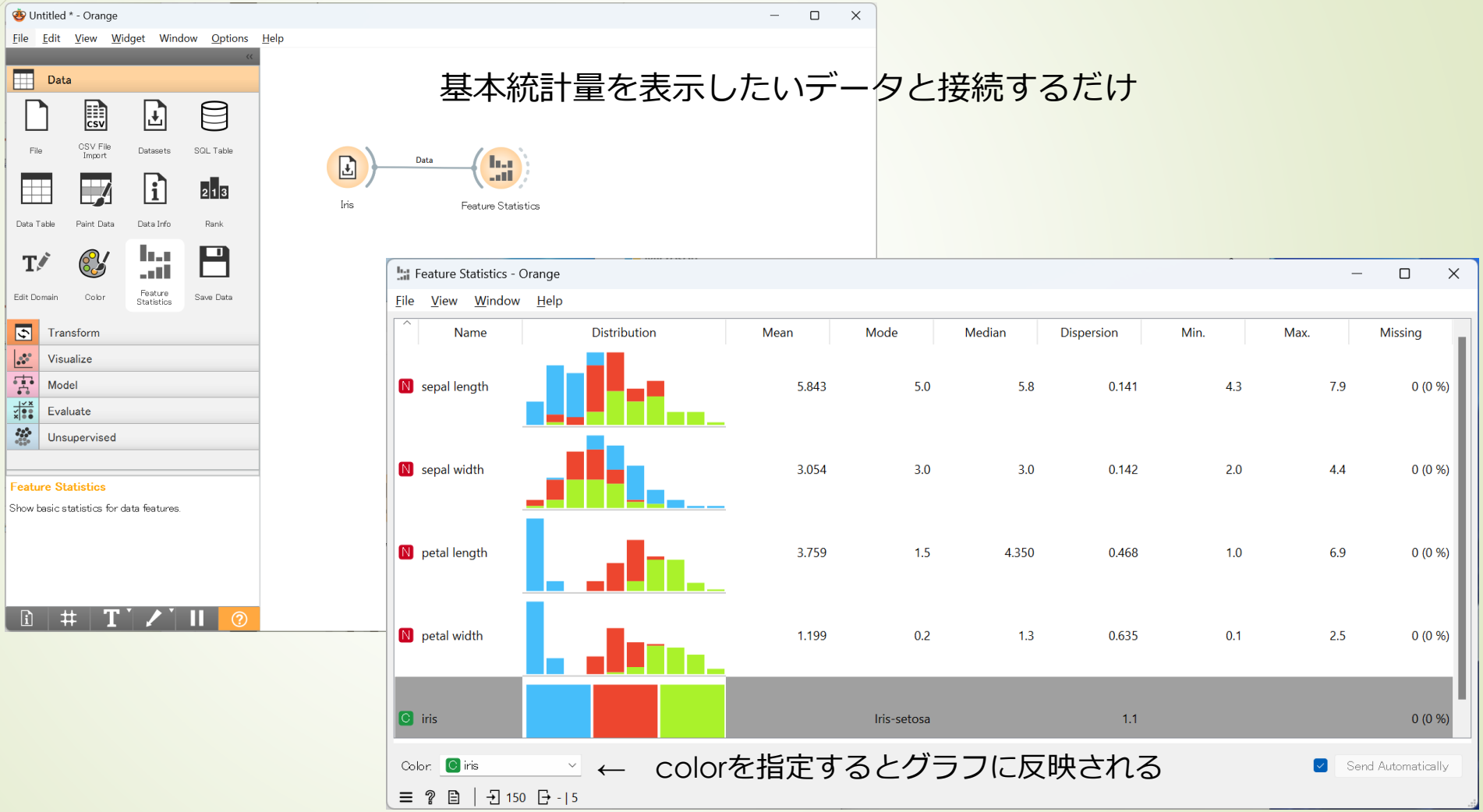
- 入門者向け：誰もが使える便利な機能
- 中級者向け：機械学習についてある知識がある人向けの機能
- アドオン　：特定の領域に特化した追加機能

入門者向け

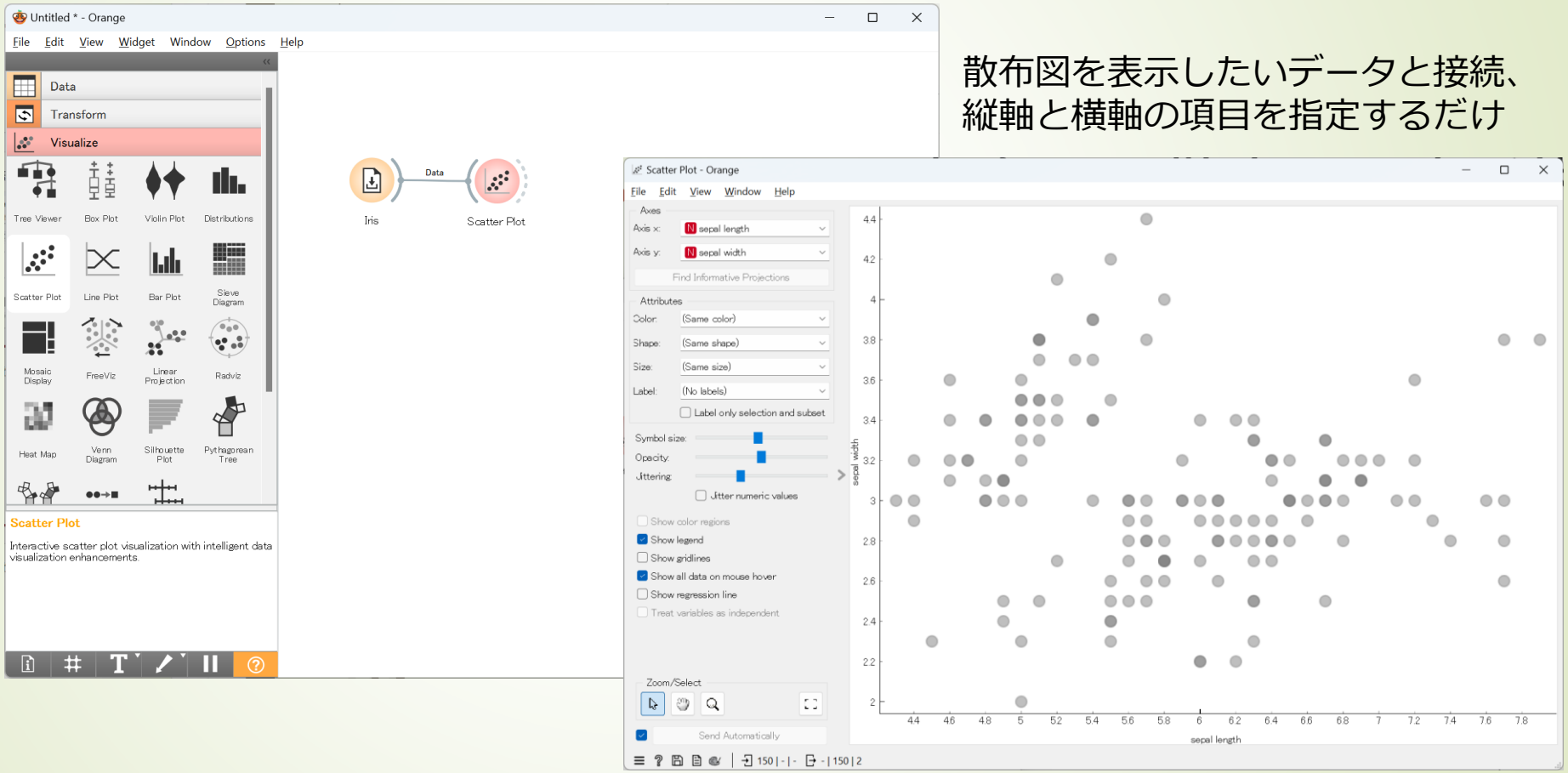
- 基本統計量の一括出力
- 散布図の便利機能×２

入門者向け：基本統計量の一括出力

基本統計量を表示したいデータと接続するだけ



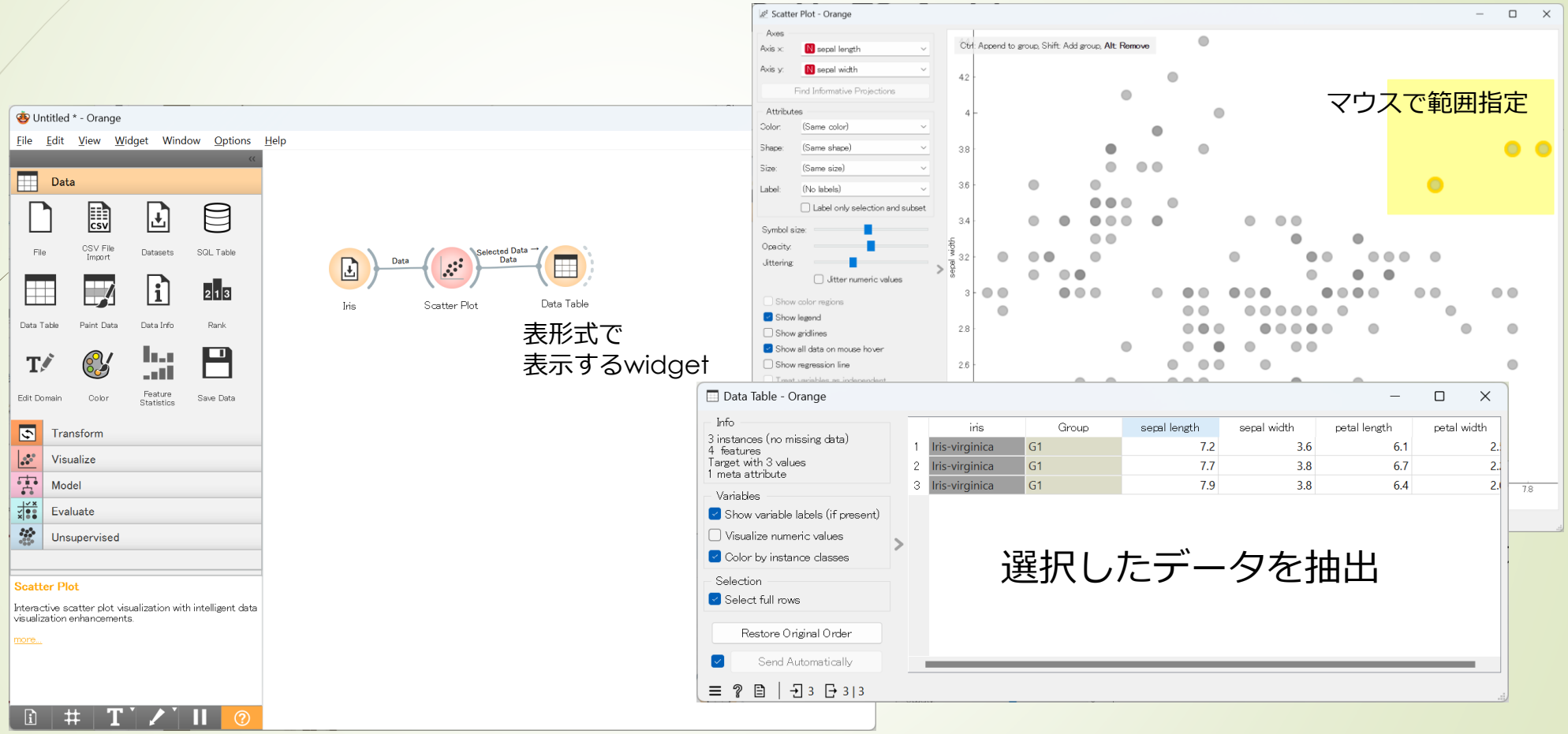
入門者向け：散布図の便利機能 1



散布図を表示したいデータと接続、
縦軸と横軸の項目を指定するだけ

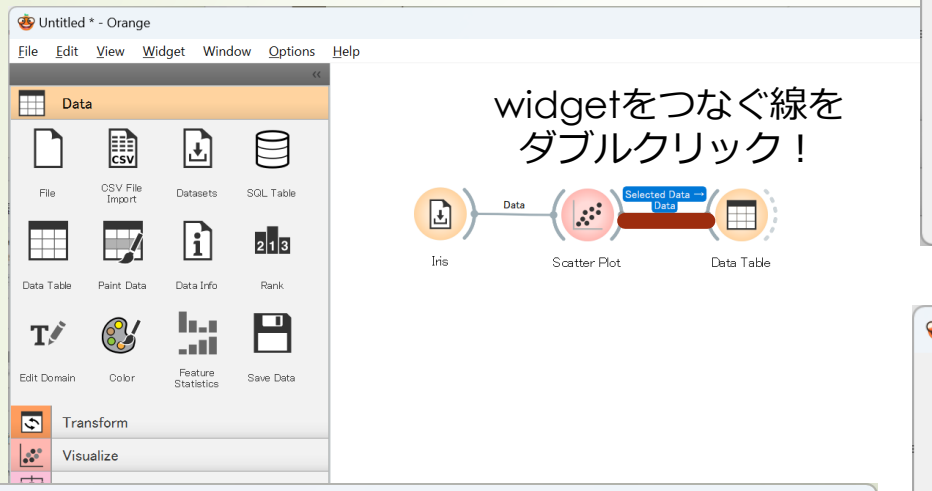
これだけだと普通すぎて面白くないですよね？

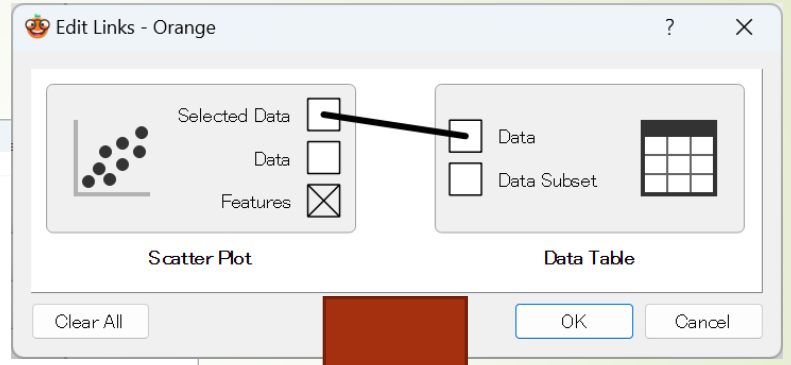
入門者向け：散布図の便利機能 1



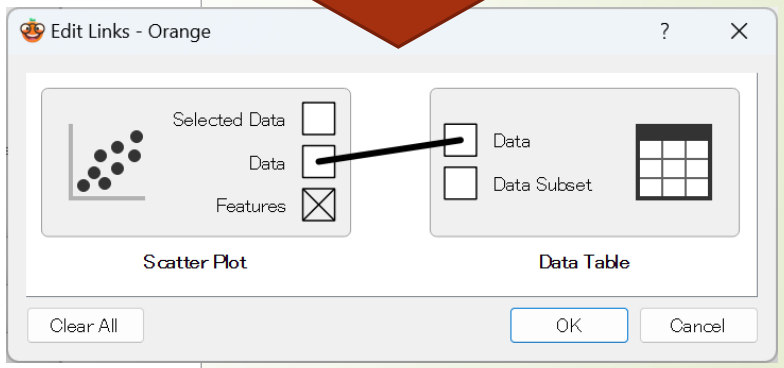
散布図どころかグラフのwidgetに限らず、
全てのデータ可視化機能からGUIでデータ選択が可能

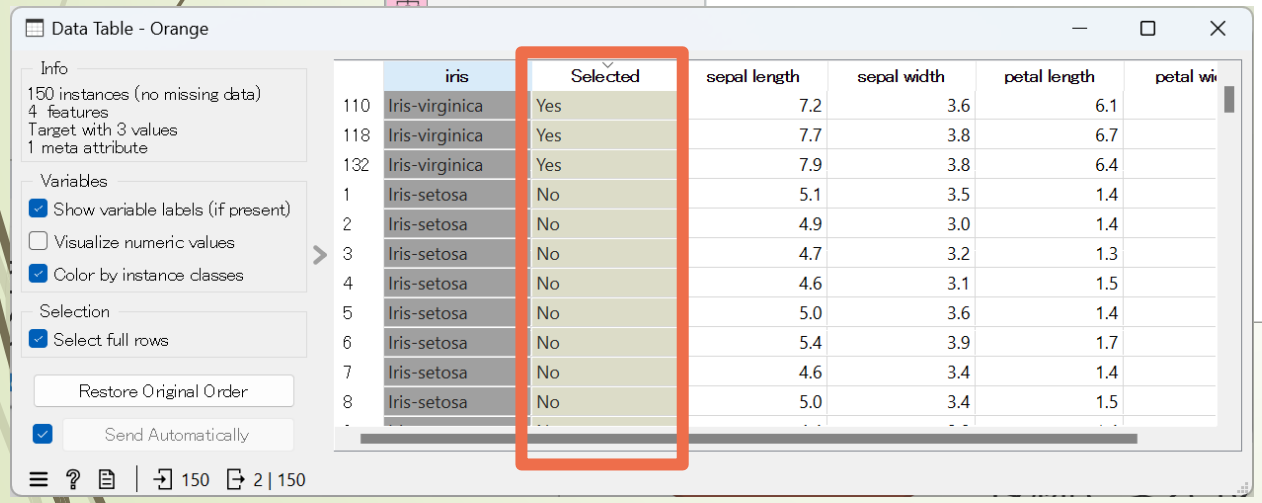
入門者向け：散布図の便利機能 1





接続を変えると





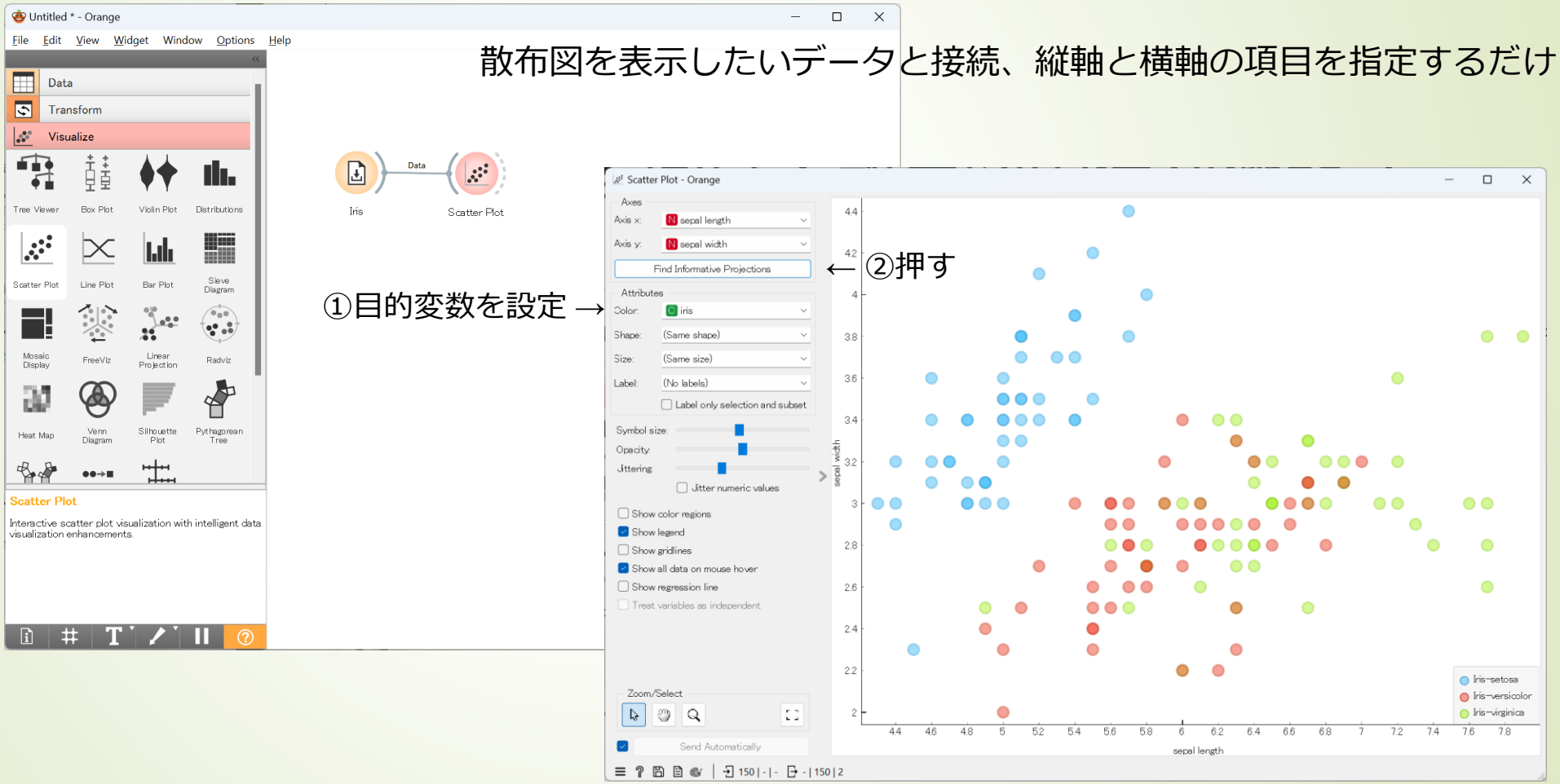
選択有無のフラグ情報を取得可能

この機能を使ってフィルターかければ、
マウス操作で異常値の除去などが可能
即時、他の分析結果に反映できる！

注意：恣意的なデータ選択はやめよう

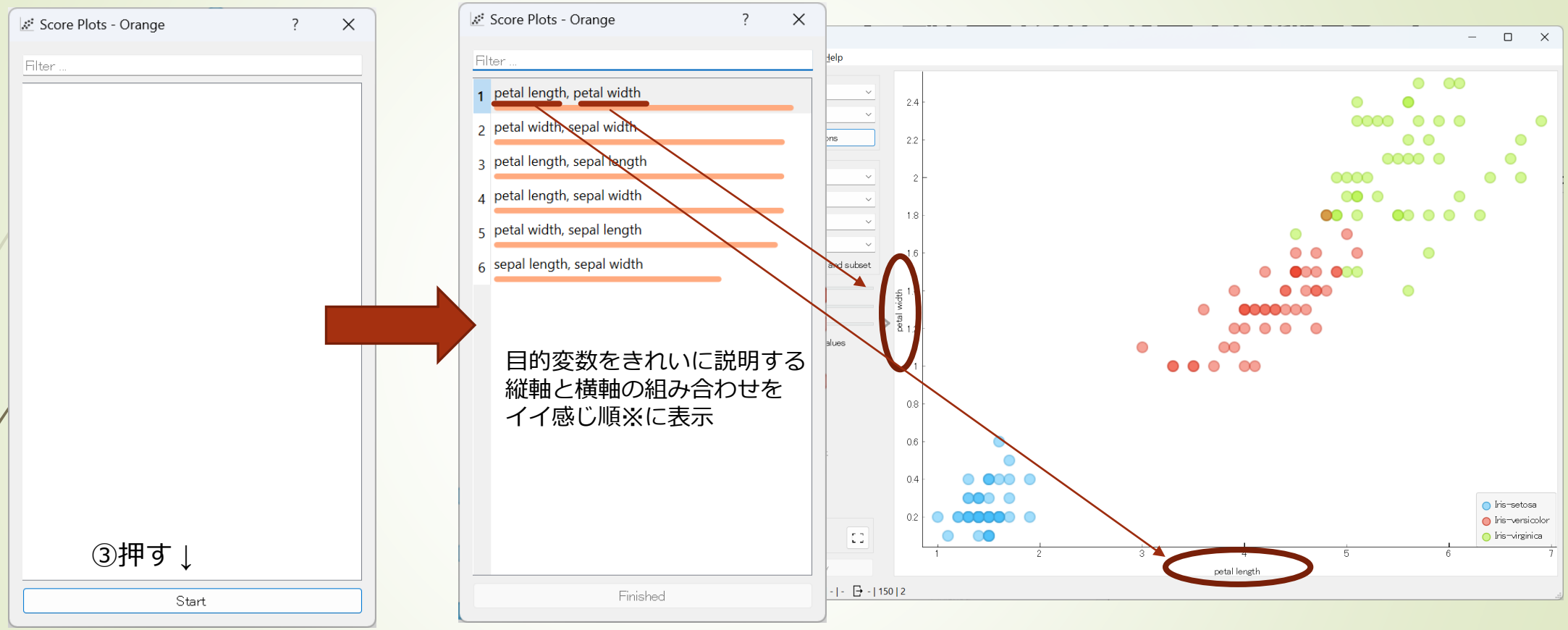
入門者向け：散布図の便利機能２

散布図を表示したいデータと接続、縦軸と横軸の項目を指定するだけ



これだけだと普通すぎて面白くないですよね？

入門者向け：散布図の便利機能 2



※イイ感じ順　の具体的な内容
knn(k=10)で全ての特徴量の組み合わせでモデル作成・評価
精度（オレンジ色のバー）の良い順に表示する

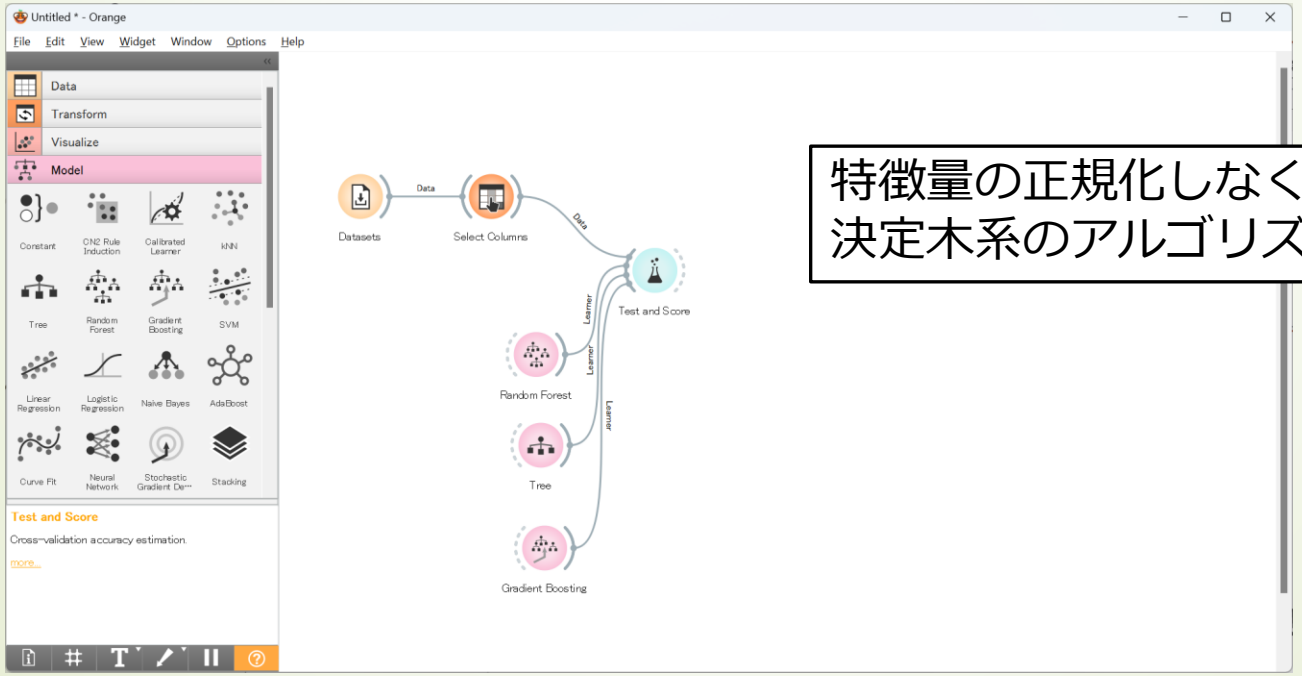
カテゴリ変数や相関係数には表れない相関を扱える
非専門家向けのデータ可視化に便利！

中級者向け

- モデル作成のデータの前処理
- 異常データの除去（異常検知）

中級者向け：モデル作成のデータ前処理

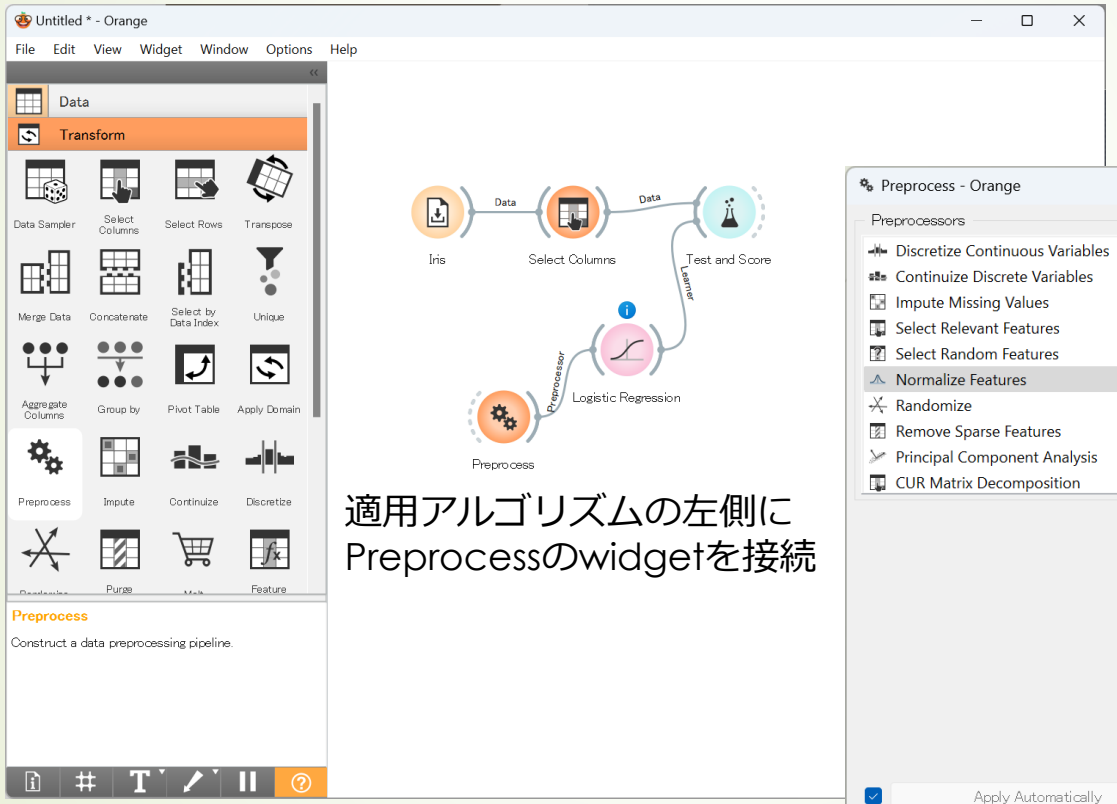
- 前回の資料、説明をシンプルにするため、適切に機械学習モデルを作成するために恣意的なことをしていました。



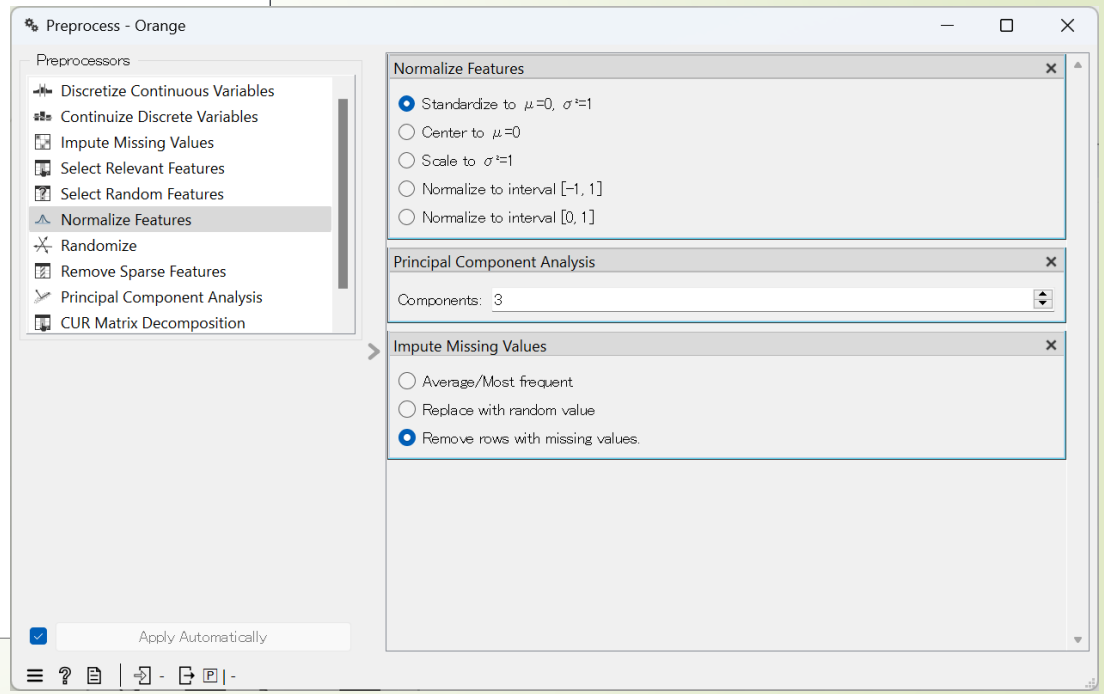
特徴量の正規化しなくても問題ない
決定木系のアルゴリズムを選択

正規化をはじめ、便利なデータ前処理機能を説明します

中級者向け：モデル作成のデータ前処理

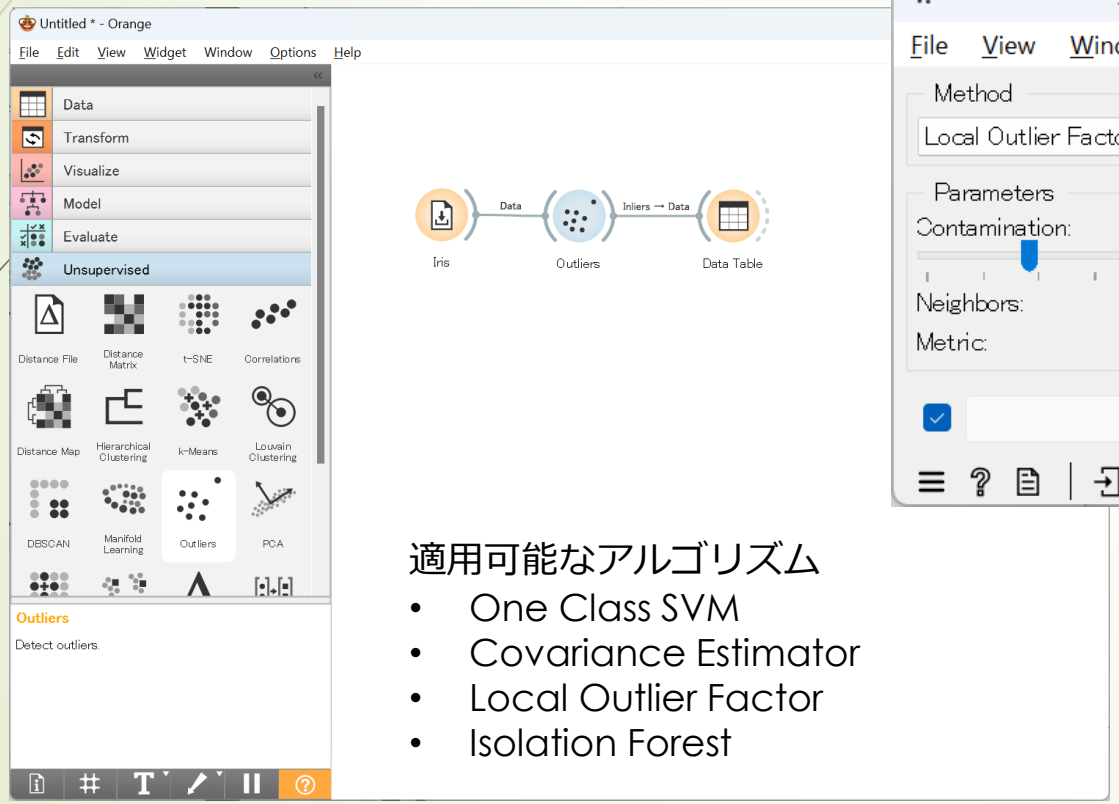


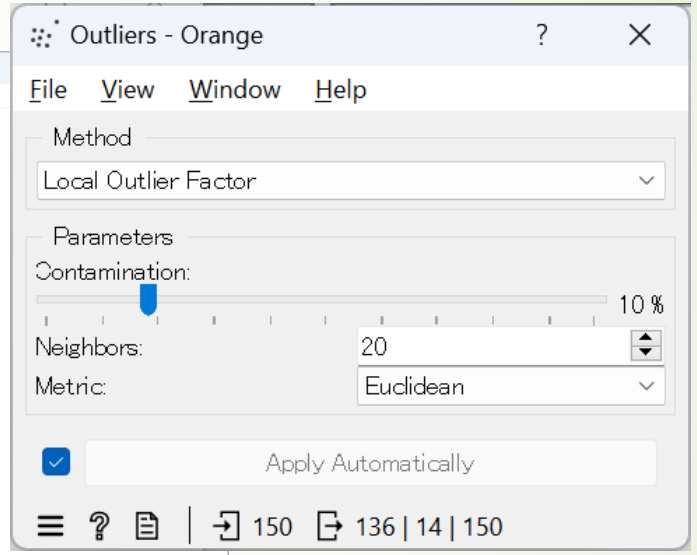
適用アルゴリズムの左側に
Preprocessのwidgetを接続



モデル作成前に実施したい前処理を
順に加えて設定するだけでOK！

中級者向け：異常データの除去（異常検知）





適用可能なアルゴリズム

- One Class SVM
- Covariance Estimator
- Local Outlier Factor
- Isolation Forest

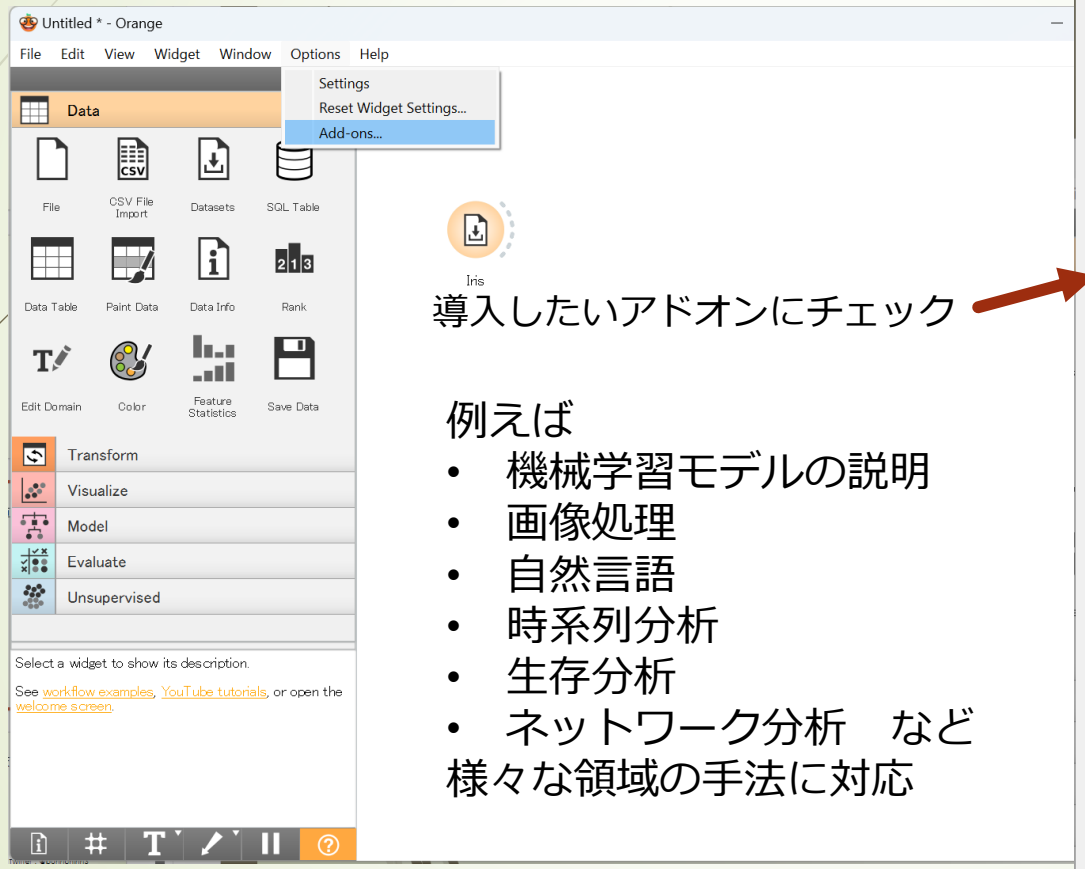
異常検知アルゴリズムを使い、
一定比率の異常値を簡単に除去可能

前述のwidgetの接続を変えれば、
逆に異常データの抽出も可能

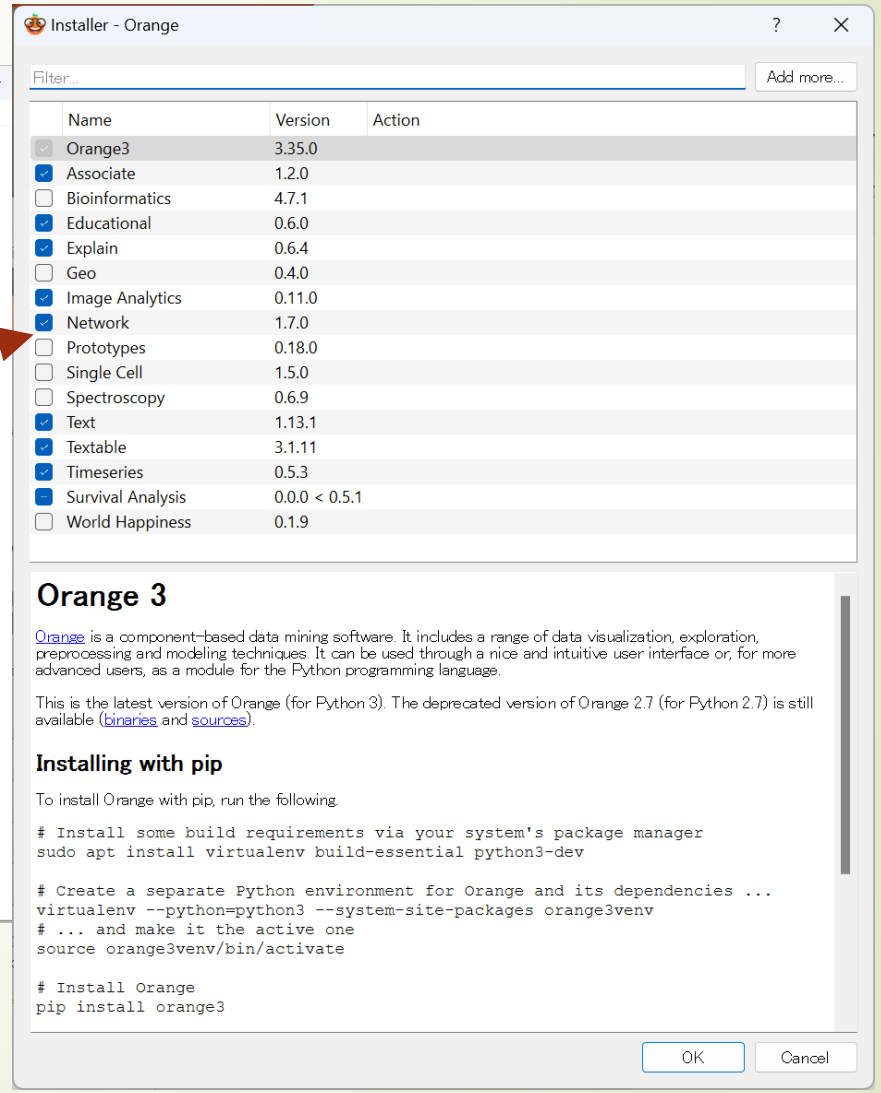
アドオン

- ad-on（追加機能）の導入方法
- 特徴量重要度の算出

アドオン：ad-on（追加機能）の導入方法



導入したいアドオンにチェック



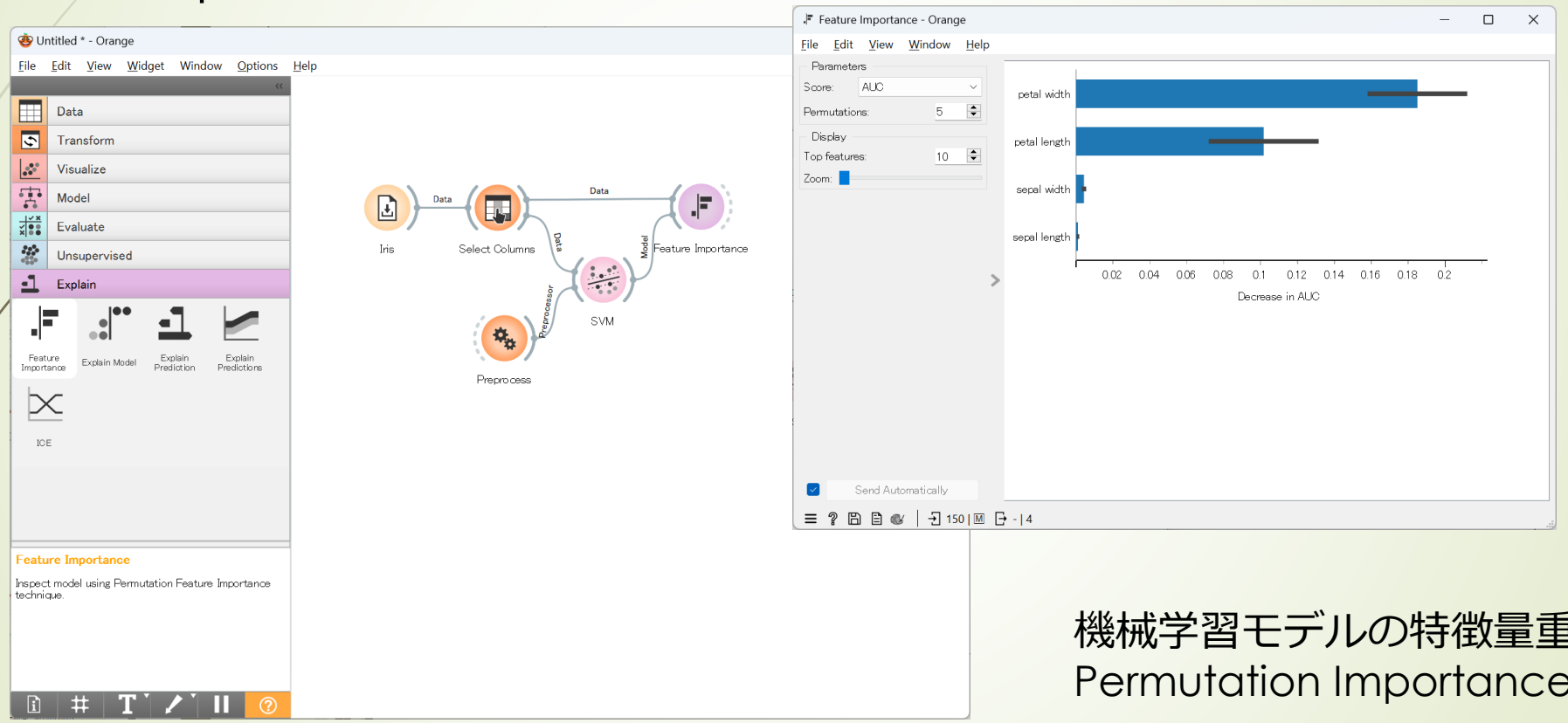
例えば
- 機械学習モデルの説明
- 画像処理
- 自然言語
- 時系列分析
- 生存分析
- ネットワーク分析　など

様々な領域の手法に対応

必要に応じて導入してください
あまり導入しすぎると起動が重くなるので注意

アドオン：特徴量重要度の算出

Explainのad-on



機械学習モデルの特徴量重要度を
Permutation Importanceで評価

補足）Permutation Importanceを使ってモデルがどの特徴量から学習したかを定量化する
https://www.datarobot.com/jp/blog/permutation-importance/

ご清聴、ありがとうございました。

他にも話したいネタがたくさんあります

- 超初心者向け機械学習の考え方
- 組織の基礎レベル向上 ノーコード分析の紹介
- データ分析プロジェクトの進め方ネタ
- あまり知られていない良書紹介　など

Twitter：@bunnchinn3

今後もLTでいろんなお話をしていきたいです。

どれにするかtwitterでアンケートを考えているので、投票してもらえると嬉しいです。

詳しい内容については、個別に対応するのでお気軽にお声がけください。