



Rベースの統計解析ツールVisual R Platform ～オープンソースソフトウェアによる柔軟な解析～

株式会社NTTデータ数理システム
データマイニング部
関川 拓也



Visual R Platform (VRP) とは

NTT DATA

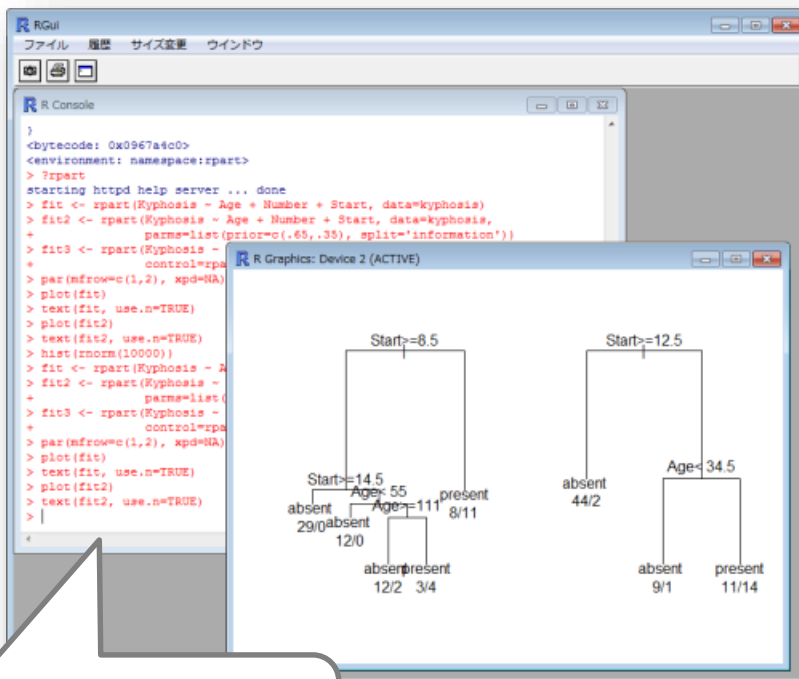
NTT DATA Mathematical Systems Inc.

- **ビジネス力**: 課題を解決する
- **データサイエンス力**: 分析結果を読み取る
- **データエンジニアリング力**: 統計、プログラム

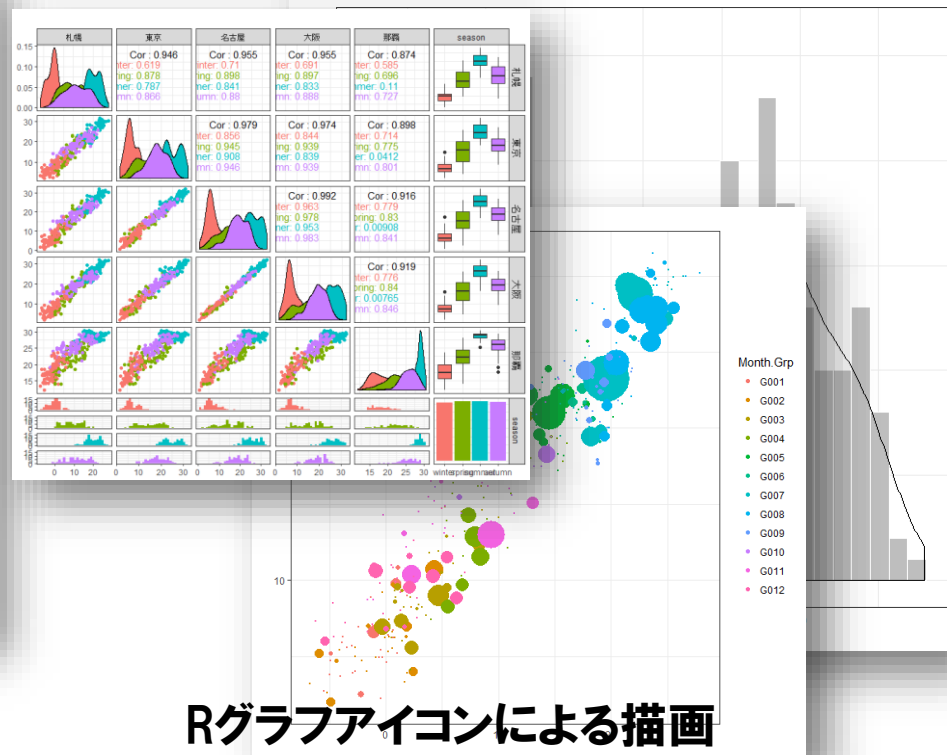
最も大切な「ビジネス力」を持った人材が「データエンジニアリング力」を身に着ける有効な手段が

統計ソフトウェア

フリーの統計ツール



「R言語」(コマンド)
入力による実行



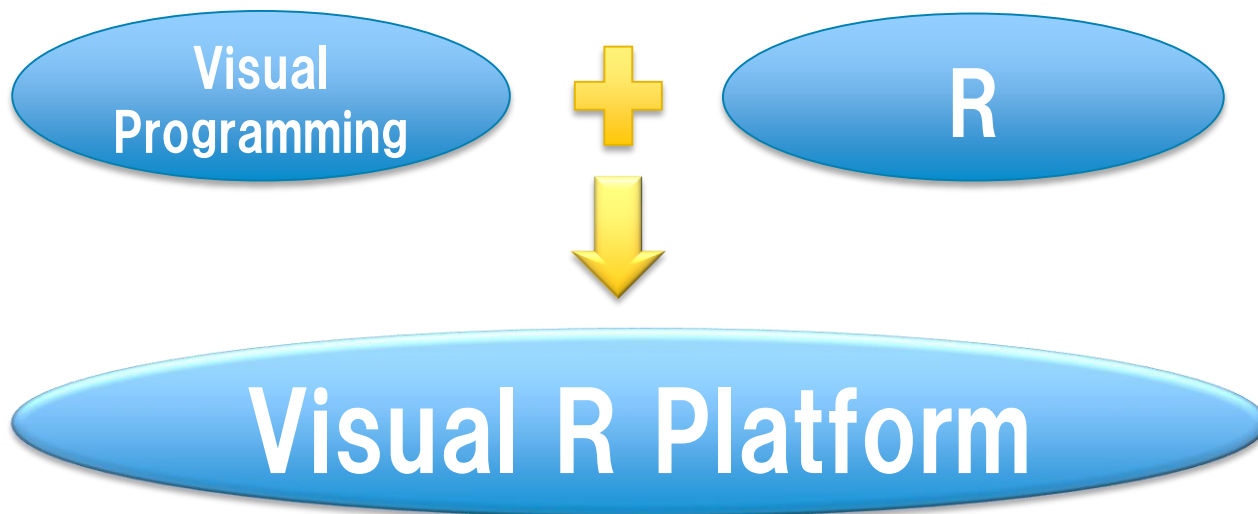
Rグラフアイコンによる描画

- **世界トップレベルのユーザー数**、米国政府機関等でも採用されています。
- **強力な機能**(統計解析、高速な関数群、豊富なグラフ機能)が利用できます。
- 膨大な拡張ライブラリ(CRANネットワーク)が日々強化され、**最新の分析アルゴリズム**を利用できます。
- **オープンソース・コミュニティ**の強力なバックアップがごございます。

- 今や、統計ソフトのデファクトスタンダードとなった**R**
- しかし、Rは実は プログラミング言語・・・
- もっと手軽に統計の分析・学習はできないものか・・・？

■ Rベースのビジュアル分析ツール

- 分析ツールの世界的標準Rの豊富な資産を利用
- VMS等で定評あるビジュアルプログラミング環境



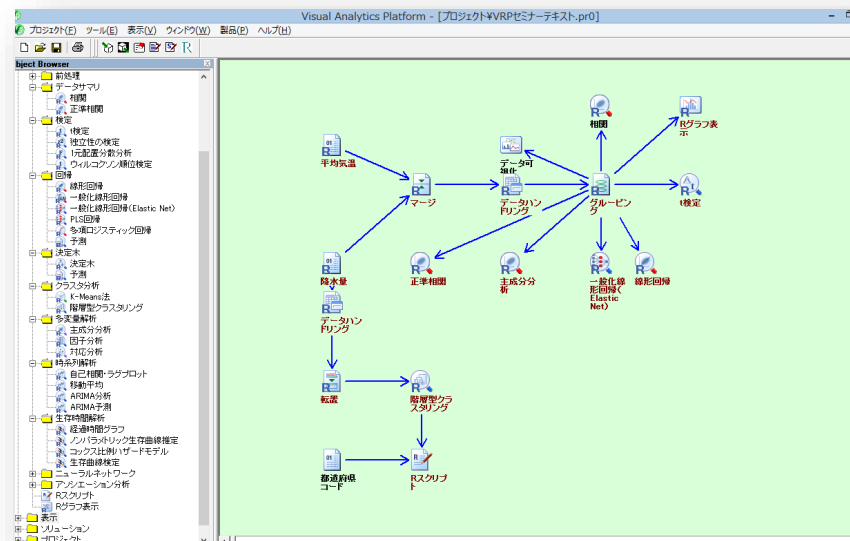
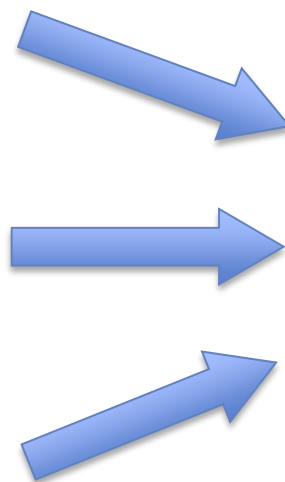
- アイコンが分析の単位を表す
 - パラメータをダイアログで入力するだけで即実行
- アイコン間のリンクで処理フローを表現(コーディング不要)
- Rの関数を知らなくても分析が可能
 - 「どう書くか」ではなく「何を分析するか」に集中できる

The screenshot displays the Visual Analytics Platform interface. On the left is a 'Project Browser' with a tree view of analysis tasks. The main workspace shows a workflow diagram with icons for '平均気温', '降水量', 'データ可視化', 'データリンク', '相関', 'Rグラフ表示', 'R検定', '線形回帰', '一般化線形回帰 (Elastic Net)', '主成分分析', '正準相関', '決定木', 'クラス分析', 't-Test', '階層型クラスター分析', '多変量解析', '特異値分解', '自己相関ラゲプロット', '移動平均', 'ARIMA分析', 'ARIMA予測', '生存時間解析', '経路分析', 'リンク生存曲線推定', 'コックス比例リゲードモデル', '生存曲線推定', 'ニューラルネットワーク', 'アンサンブル分析', 'Rスク립ト', 'Rグラフ表示', '表示', 'シミュレーション', and 'プロジェクト'. A dialog box titled '独立性の検定' (Independence Test) is open on the right, showing options for '検定方法' (Bivariate Chi-square, Fisher's Exact), '入力形式' (Cross-tabulation, Univariate), '対象列指定' (Target variables: 大都市圏, 年間収入.Grp), and 'オプション' (Bivariate Chi-square: Yates' correction, Fisher's Exact: Monte Carlo simulation).

■ データを簡単にVRPに取り込めます

- ファイル (csv, excel他) をドラッグ & ドロップ
- 取り込み時のオプション指定もラクラク

■ 各種DBからのデータ取り込みもサポート

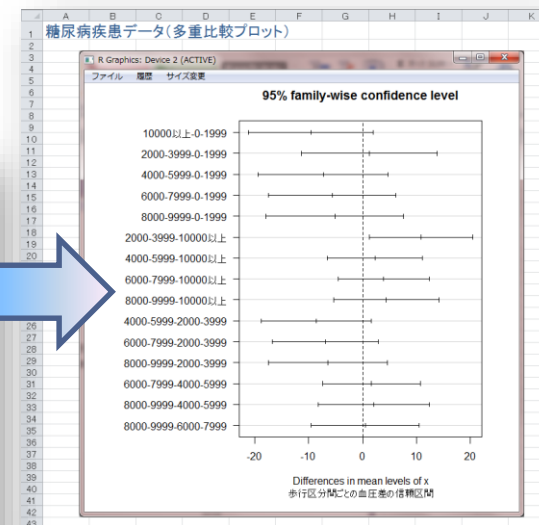
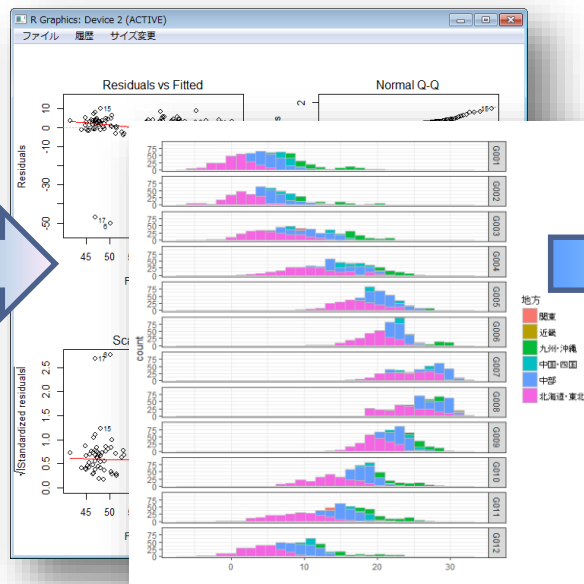
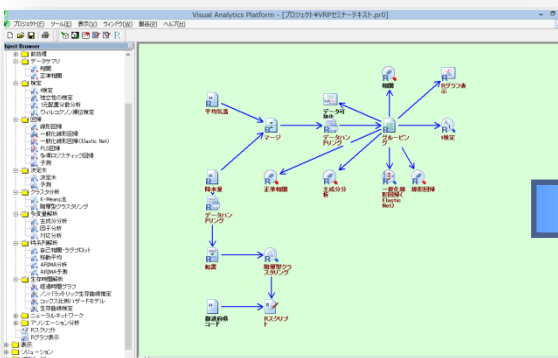


■ ワンクリックで視覚化

- Rグラフ表示アイコンで見栄えのするグラフを簡単に作成
- オブジェクトビューによる分析結果の視覚化

■ レポート、文書への貼り付け

- 各種の画像ファイルに変換可能。資料作成にもご利用いただけます



Excel, Word,
Powerpoint等

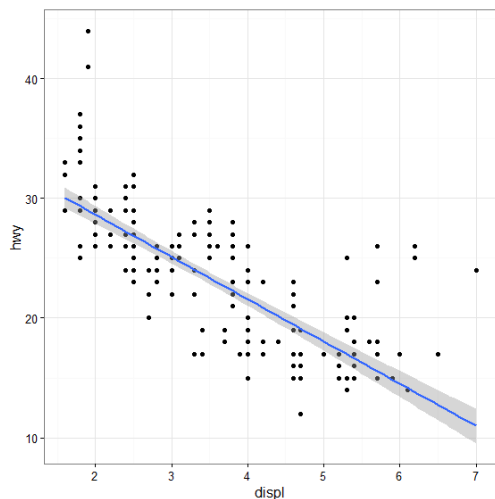
■ VRP上のグラフ表示アイコン

- 美しく、見栄えのするグラフを簡単に作成
- 通常のVRPアイコンと同様に、Rソースエクスポート機能でR言語のソースコードに変換可能

グラフの種類と主な機能

散布図	散布図を描く。回帰直線等のプロットも可能
折れ線	複数個の系列を同時にプロット
箱ひげ	カテゴリごとに分類してプロットする機能
対散布図	相関係数、箱ひげ図、ヒストグラムも同時に描画
ヒストグラム	カテゴリごとのプロット機能有り
棒グラフ	カテゴリごとのプロット機能有り
円グラフ	Pieチャートの描画
hi-lowグラフ	株価変動などに使われるグラフ

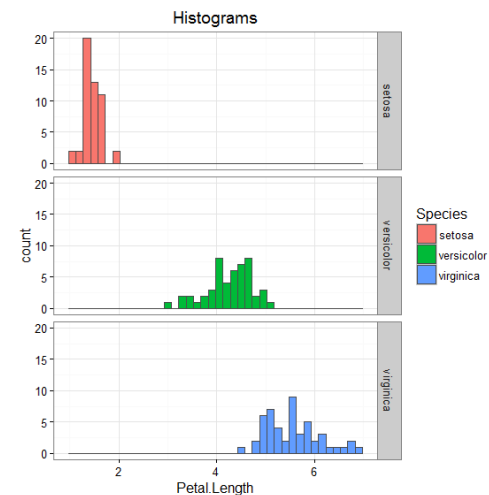
Rグラフ表示アイコンのプロット例



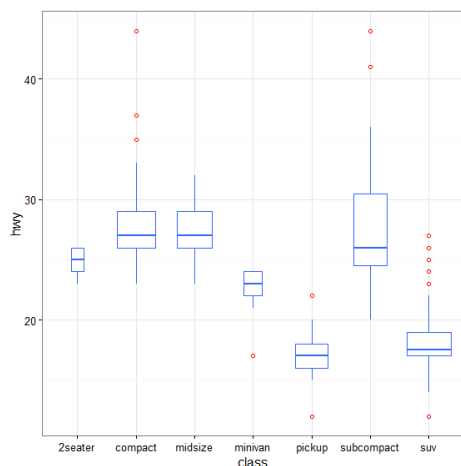
散布図(回帰直線付き)



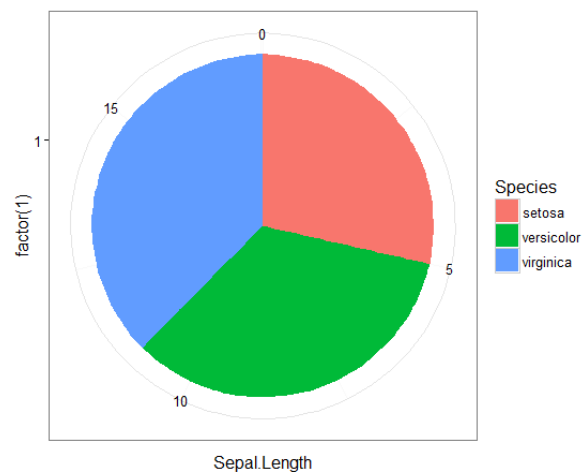
対散布図



ヒストグラム



箱ひげ図



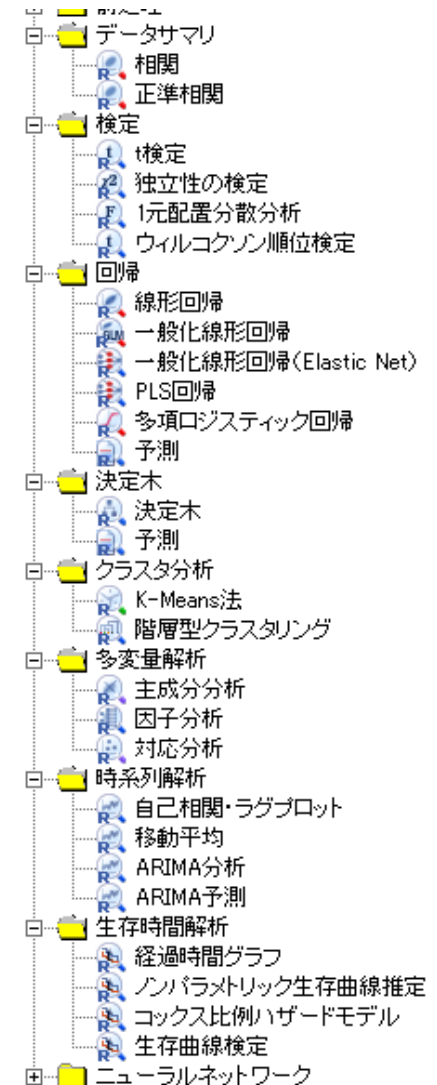
円グラフ

統計解析アイコン

オブジェクトブラウザ内の 統計解析アイコン一覧

- 基本的な統計解析機能は、既存のアイコンを利用
 - 検定、線形回帰、時系列解析など
- アイコン化されていない機能は、Rスクリプトアイコンを利用
 - 並列処理や要望に応じた機能追加（Rライブラリ、階層ベイズ、他機械学習）が利用可能
 - Rのグラフ関数も利用可能

R の持つ豊富な統計計算リソース、日々増加している機能やパッケージをフルに活用することができます



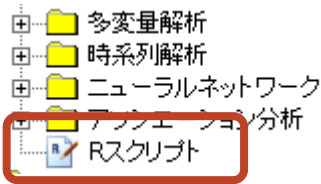
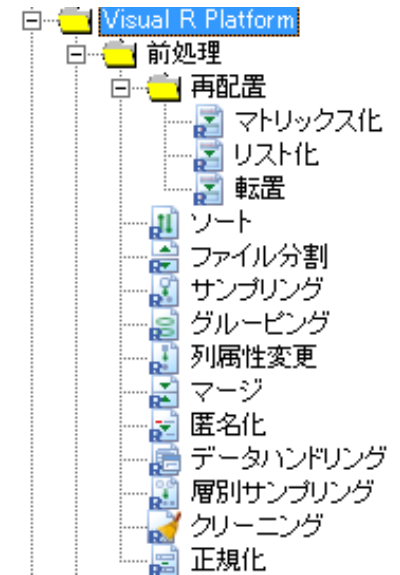
前処理アイコン

- 基本的な前処理加工、グラフ作成機能は、既存のアイコンを利用
 - マージやグルーピング、サンプリングなど
 - 平均や分散などの集計機能は「データハンドリング」アイコンで利用

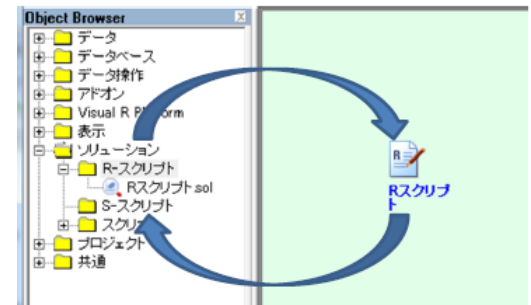
Rスクリプトアイコン

- アイコン以外の処理は、Rスクリプトアイコンを利用
 - Rスクリプトアイコンでカスタム対応も可能

オブジェクトブラウザ内の前処理アイコン一覧



作成したRスクリプトアイコンは、アイコン化して再利用可能!



- Rソースエクスポート…プロジェクトの内容をR言語のスクリプトとして出力

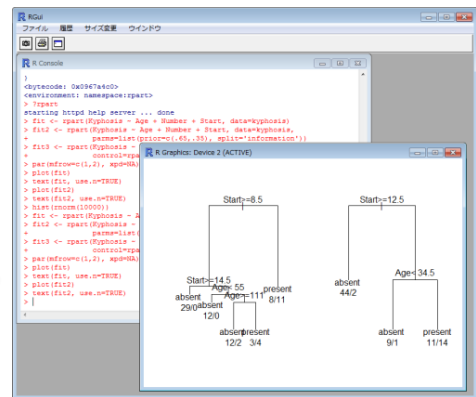
右クリックメニューから
Rソースエクスポートを実行

Rスクリプトを自動生成

Rがインストールされていれば
ほとんどの環境で実行可能！
(VRP不要)

右クリック

```
#####  
# VRP専用スクリプトロード  
#####  
if (exists("vrpEnv")) {  
  vrpEnv=new.env()  
  source("VRP.lib.R", local=vrpEnv)  
  attach(vrpEnv)  
}  
#####  
# 実行部  
#####  
# R観客データ  
vrp1 = vrp.RData(rdataFile="R観客データ_1.RData")  
# サンプルング  
vrp6 = vrp.RSampling(input=vrp1[["noName"]], target=c("性別","年代","職業","利用時間",  
"利用回数"), n.samp="100%", method="先頭から", key=NULL, seed=NA,  
n.inter=0)  
# ファイル分割  
vrp4 = vrp.RDivide(input=vrp6[["result"]], rate=c(7.3), seed.auto=F, seed=0)  
# 線形回帰  
vrp5 = vrp.RLinearModel(x=vrp4[["learn"]], dependentColName = c("収益"),  
independentColName = c("性別","職業","利用時間","利用回数"), constantTerm = TRUE,  
weightColName = NULL, dir = "both")  
print("result is stored in vrp5")
```



活用方法の幅が広がる！

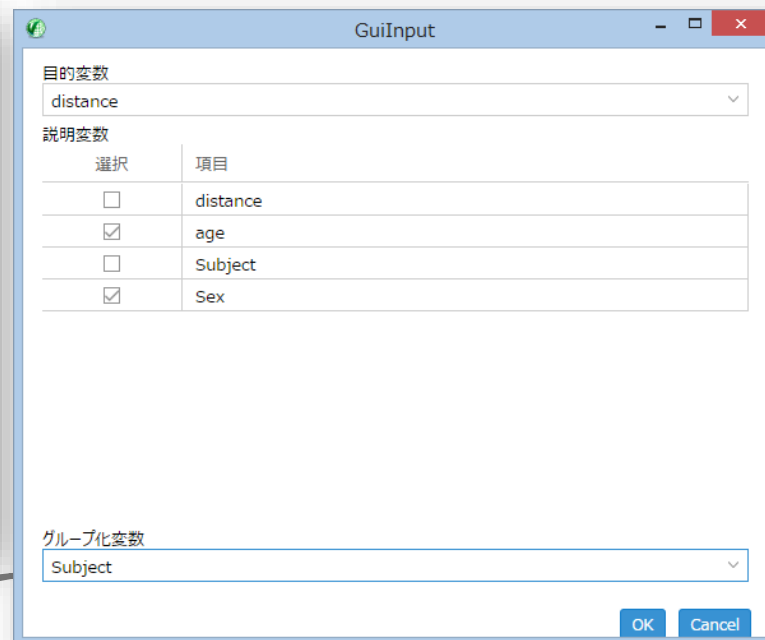
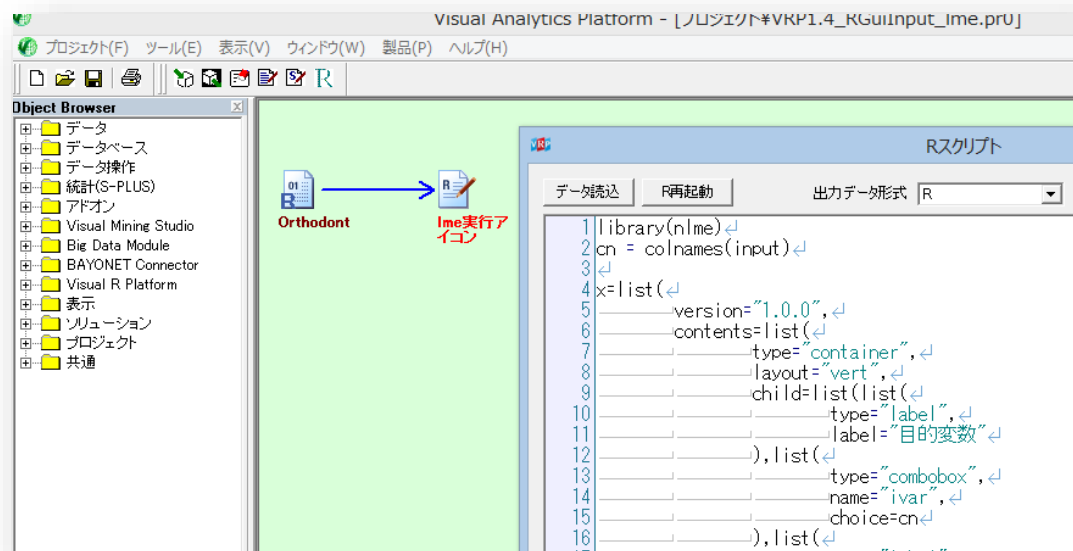
- ✓ 分析処理の記録
- ✓ スクリプトの共有
- ✓ Linux上での定型処理化 etc...



- 簡易ユーザーインターフェース作成機能:簡易なスクリプト記述で標準アイコン同様のダイアログを表示、値の指定が可能に

スクリプトアイコンを編集し

実行すると、ユーザーインターフェース起動。パラメータ指定などがリスト、コンボボックスで実現

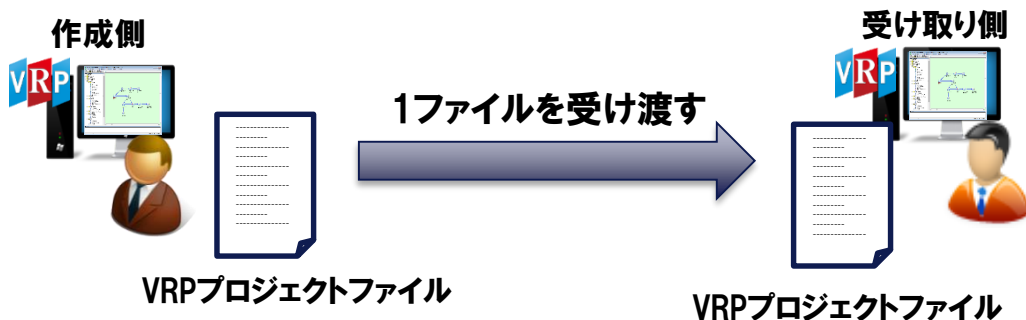


分析手法の幅が広がる！

- ✓ 専門性の高い分析用のカスタマイズを行ったアイコン作成が可能

分析環境の共有

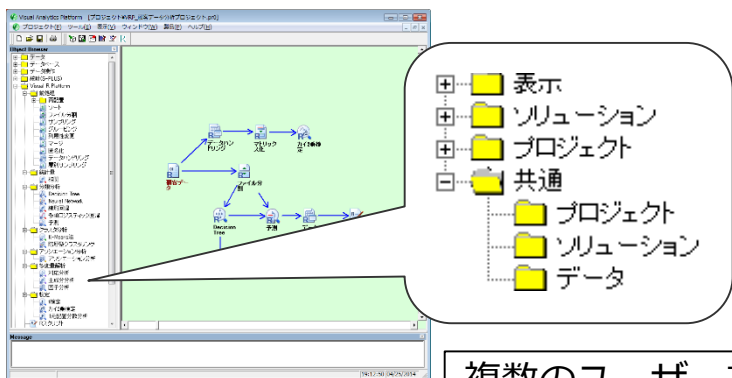
- 複数名でご利用の場合 その1…分析フローをファイル化して受け渡す



こんな時に有効！

- ✓ 作成したプロジェクトを異なるマシンの間で交換したい
- ✓ 遠く離れた人に、添付ファイルの形でプロジェクトを渡したい
- ✓ 作成したプロジェクトのバックアップを取りたい

- 複数名でご利用の場合 その2…共有フォルダにプロジェクトファイルを保存する

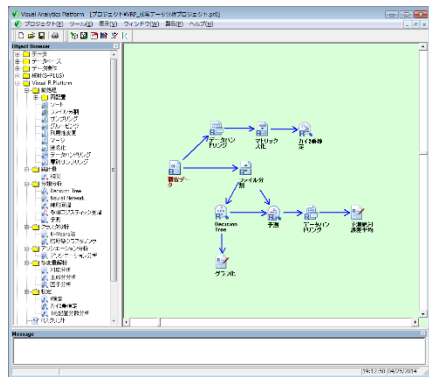


こんな時に有効！

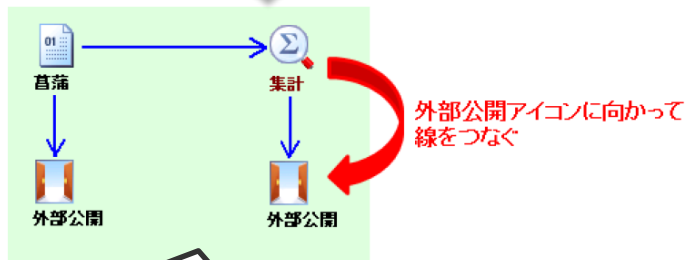
- ✓ ユーザ共通のネットワークサーバに接続できる環境
- ✓ 作成したデータ、スクリプトの交換を手間なく実施したい

複数のユーザーで共有可能な領域を作成
データも分析フローも簡単に共有できます

- 繰り返し頻繁に行う処理を定型 (バッチ) 処理化



VRPプロジェクト作成後



プロジェクト上で右クリック「外部公開」を選択し、バッチ実行時に内容を変更したいアイコンと外部公開アイコンを矢印でつなぎます。

こんな時に有効！

- ✓ 他システムから計算エンジンとして呼び出したい
- ✓ 日常的に繰り返す定型処理をスケジュールを組んで実行したい
- ✓ 複数ファイルに対して同じ処理を実行しておき、結果だけを確認したい

全処理流れの実行
全処理アイコンの結果削除
Web分析メニュー設定
インポート
エクスポート
バッチプログラム作成
全アイコン削除
元に戻す(Ctrl-Z)

バッチプログラムを作成し、設定、実行します

日次バッチ、月次バッチなど、レポート機能としても活用可能！グラフ画像の出力も可能！



VRP 1.6の新機能

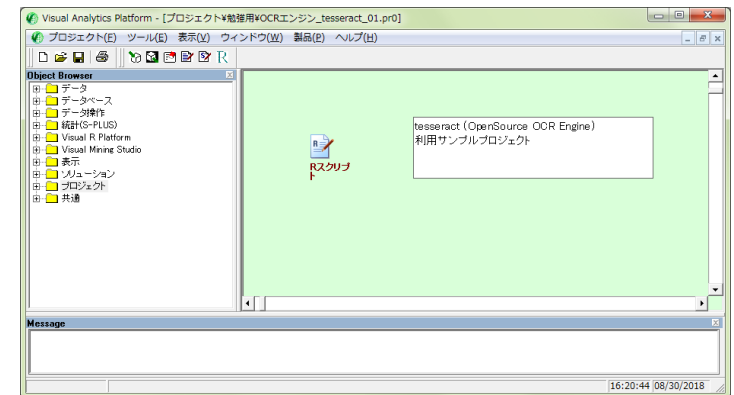
- Rパッケージ サンプルプロジェクト ダウンローダー
 - 利便性の高いRパッケージを手軽に利用できるVRP用プロジェクトをご提供
- RGuiInput用GUI部品拡充
- 新データビューア (gv3) への対応強化
- 多数のアイコンについて操作性、機能の両面を強化

- CRANで提供されているRパッケージを利用したサンプルプロジェクトをご提供
- ユーザーフレンドリーなGUIで、最新のRパッケージの機能を手軽に利用できます
- 各種Rパッケージで提供されている関数群の、実用的サンプルコードとしてもお使いいただけます



Sample
Project
Download

.prjファイル



VRPに.prjファイルを
インポートして利用

選択	パッケージ名	カテゴリ	機能
<input type="checkbox"/>	igraph	グラフ	ネットワークグラフ図
<input type="checkbox"/>	xgboost	分析	ブースティング
<input type="checkbox"/>	caret	分析	機械学習
<input type="checkbox"/>	mice	データ処理	多重欠損補完
<input checked="" type="checkbox"/>	hexbin	グラフ	2次元密度グラフ
<input type="checkbox"/>	mclust	モデル作成	混合正規分布によるモデル作成
<input type="checkbox"/>	dbscan	分類	密度ベース(density based)クラスタリング
<input type="checkbox"/>	tesseract	文字認識	文字認識(OCR)エンジン
<input type="checkbox"/>	knitr	文書ツール	R markdown でのレポート生成
<input type="checkbox"/>	sqldf	データ処理	SQL文でデータ加工
<input type="checkbox"/>	RODBC	入出力	データベース接続

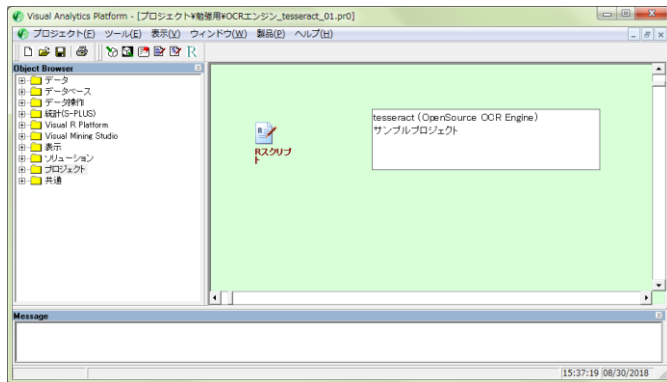
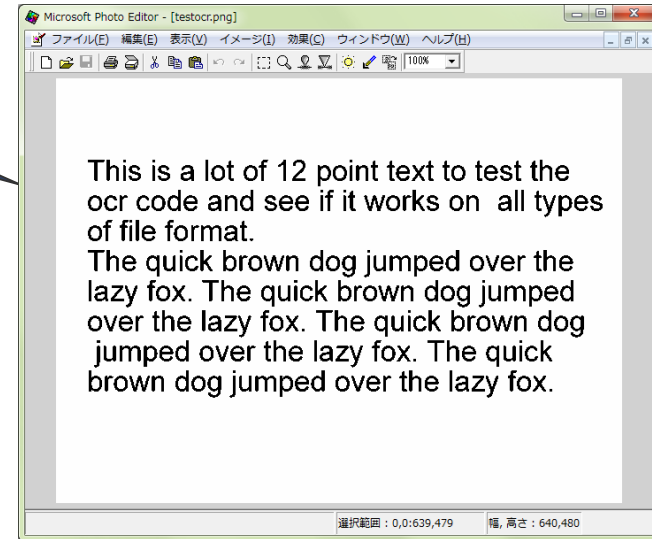
散布図を点の集合の密度により色を変えて描画する。巨大データの散布状態を視覚化したい場合に便利

[VRPサンプルプロジェクトをダウンロード](#)

サンプルプロジェクトご提供対象のRパッケージは続々追加予定です

- Hewlett-Packard/Googleで開発されたオープンソースのOCR(画像文字認識)エンジン
- 画像ファイルを認識し、テキスト(文字列)に変換

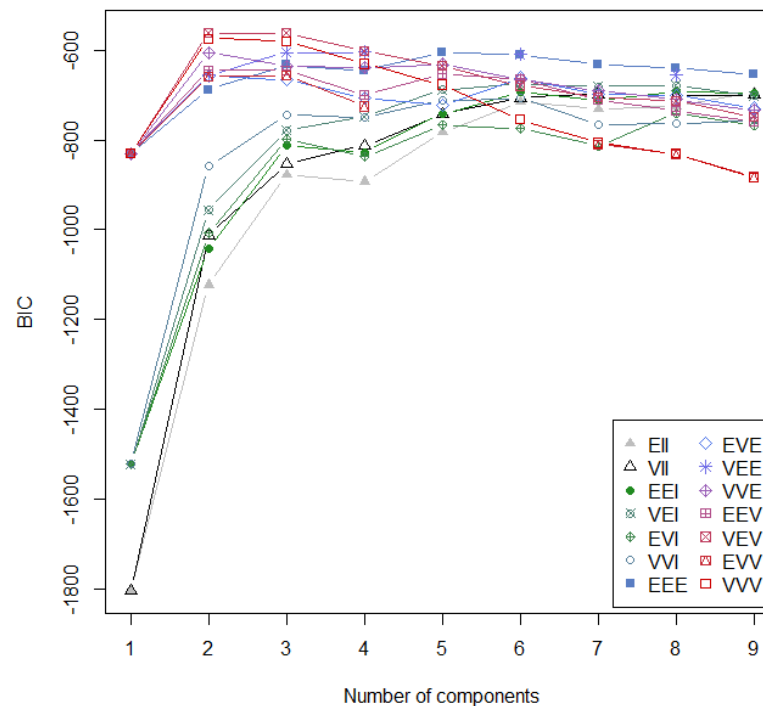
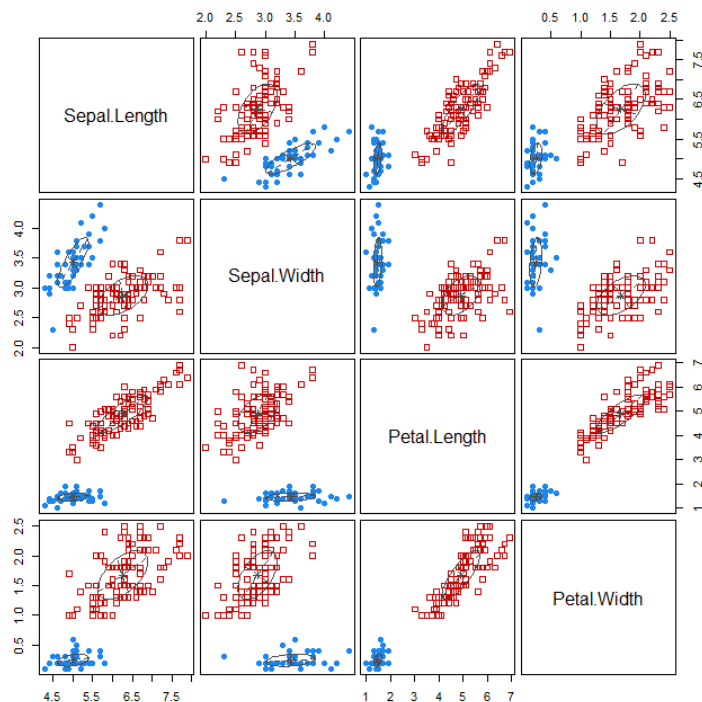
画像ファイル



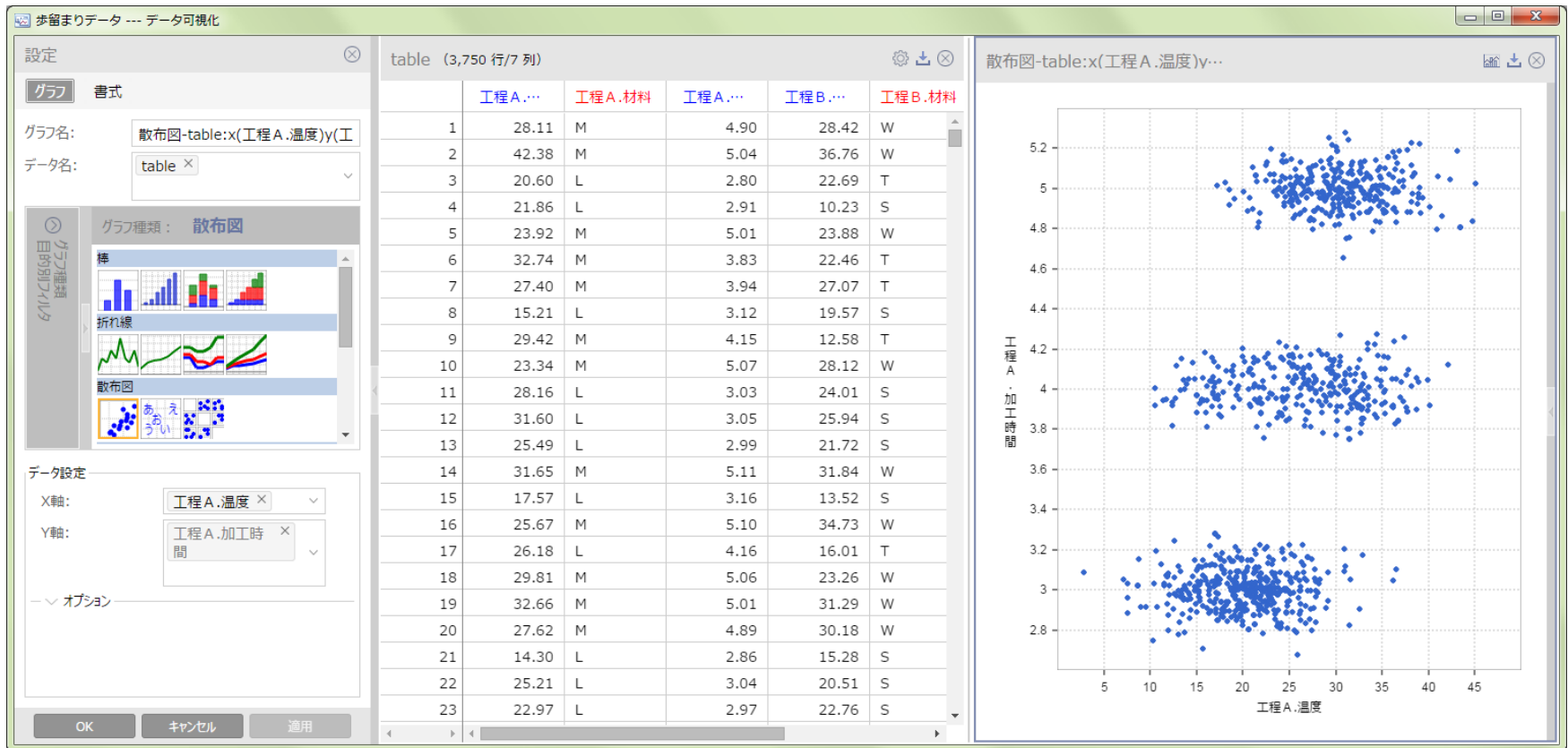
生成テキスト

This is a lot of 12 point text to test the cor code and see if it works on all types of file format.
The quick brown dog jumped over the lazy fox. The quick brown dog jumped over the lazy fox. The quick brown dog jumped over the lazy fox. The quick brown dog jumped over the lazy fox. The quick brown dog jumped over the lazy fox.

- 混合正規分布モデル (Gaussian Mixture Model) によるクラスタリング
 - サンプル点が各クラスに属する確率が求まる(ソフトクラスタリング)
 - 外れ値検出モデルとしても使える
 - 評価指標 (BIC) が算出される
 - 最適なクラスタ数の決定に使える



■ 行名を持つdata.frame(.RDataファイル)の表示に対応



- 従来のGUI部品に加え、新たなGUI部品を多数追加しました

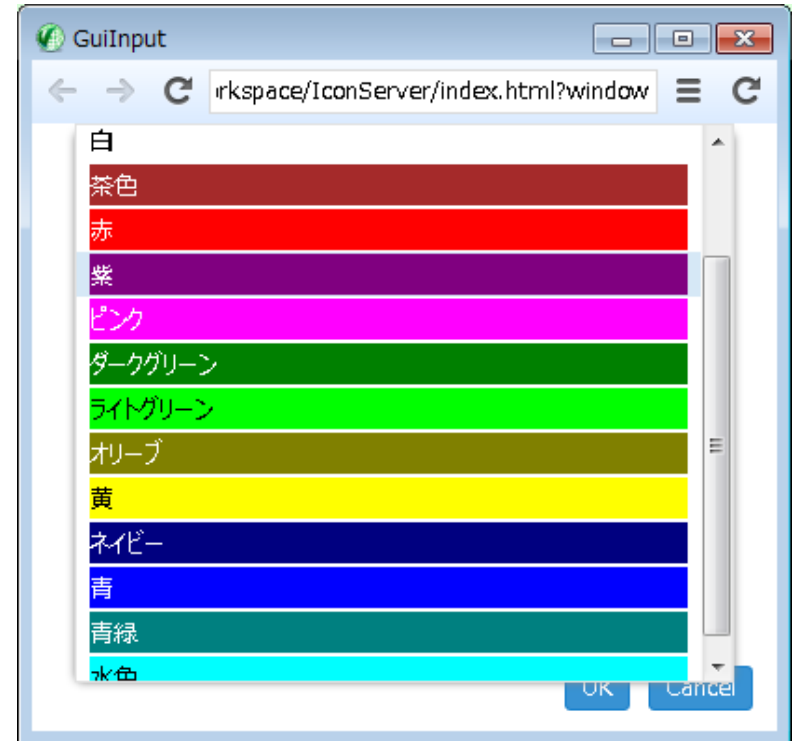
追加コントロール	内容
<code>vap.textarea</code>	テキスト表示・編集(複数行対応)
<code>vap.randomseed</code>	乱数シード設定
<code>vap.scaling</code>	スケーリングパラメータ指定
<code>vap.sampling</code>	サンプリングパラメータ指定
<code>vap.colorCombobox</code>	色選択コンボボックス
<code>vap.NAFill</code>	欠損補填値指定
<code>vap.NAOpt</code>	欠損処理方法指定



**サンプリングパラメータ指定
(vap.sampling)**



**欠損処理方法指定
(vap.NAOpt)**



**色選択
(vap.colorCombobox)**

■ モダンなUIへ刷新し、操作性が向上しました

前処理アイコン	分析アイコン
ソート	決定木
サンプリング	因子分析
クリーニング	対応分析
グルーピング	自己相関 - ラグプロット
列属性変更	アソシエーション分析
正規化	
匿名化	
層別サンプリング	

サンプリング

パラメータ設定 オプション

サンプリング方法: 先頭から

数: 100 割合(%)で指定

乱数の初期値

自動

手動 0

期間: 1

選択

列名	列属性
種類	カテゴリ
がく長	整数
がく幅	整数

非選択

列名	列属性
花びら長	整数
花びら幅	整数

OK Cancel

アソシエーション分析

列名	キー列	分析対象列
顧客ID	<input type="checkbox"/>	<input type="checkbox"/>
年齢	<input type="checkbox"/>	<input type="checkbox"/>
性別	<input type="checkbox"/>	<input type="checkbox"/>
職業	<input type="checkbox"/>	<input type="checkbox"/>
月収	<input type="checkbox"/>	<input type="checkbox"/>
学歴	<input type="checkbox"/>	<input type="checkbox"/>
クレーム回数	<input type="checkbox"/>	<input type="checkbox"/>
利用時間	<input type="checkbox"/>	<input type="checkbox"/>
収益	<input type="checkbox"/>	<input type="checkbox"/>

分類種類: アソシエーションルール

最低サポート: 0.1

最低信頼度: 0.1

最小ルール長: 2

最大ルール長: 2

OK Cancel

匿名化

スケール変換($y=A*x+Const$)

変換スケール(A): 1

定数(Const): 0

乱数の初期値

自動

手動 0

	型
	カテゴリ
	整数
	整数
	整数
	整数

OK Cancel

決定木

目的変数

変数名: 収益

分割指標: anova

複雑度係数: 0.001

分割対象とするノードの最小の観測数: 20

Leafに含まれる最小の観測数: 5

クロスバリデーションの回数: 10

決定木の最大の深さ: 30

説明変数

	変数名	型	説明変数
1	顧客ID	整数	<input type="checkbox"/>
2	年齢	整数	<input checked="" type="checkbox"/>
3	性別	カテゴリ	<input checked="" type="checkbox"/>
4	職業	カテゴリ	<input checked="" type="checkbox"/>
5	月収	整数	<input checked="" type="checkbox"/>
6	学歴	カテゴリ	<input checked="" type="checkbox"/>
7	クレーム回数	整数	<input checked="" type="checkbox"/>
8	利用時間	整数	<input checked="" type="checkbox"/>
9	収益	実数	<input checked="" type="checkbox"/>

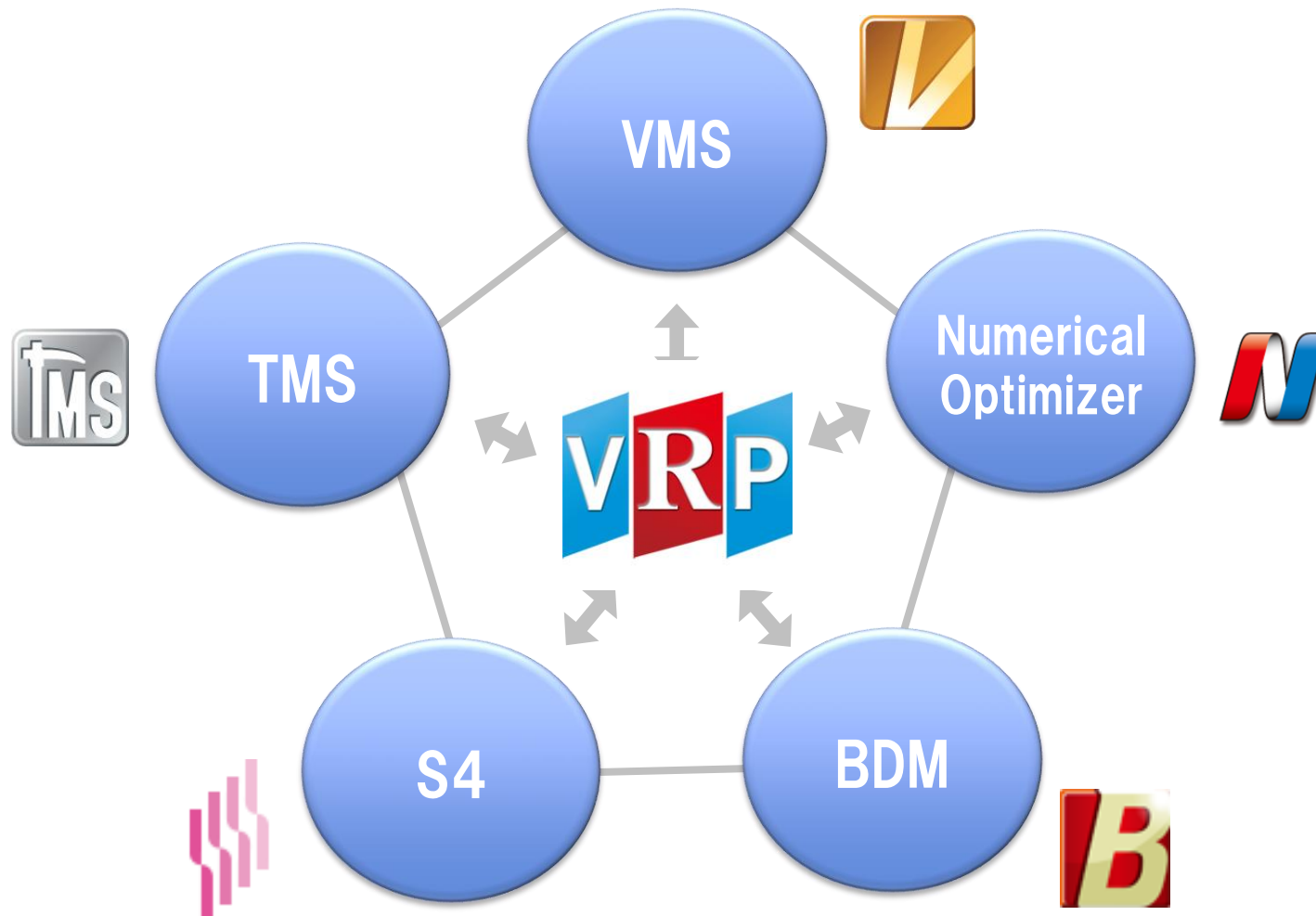
変数: 9個

OK Cancel

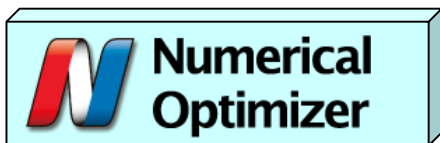


数理システム製品間の連携

- VMS, TMS等の数理システム他製品との連携が可能
 - 複数製品の連携で、さらに高度・多様な分析を



製品連携例 (1) : RNUOPT + VRP

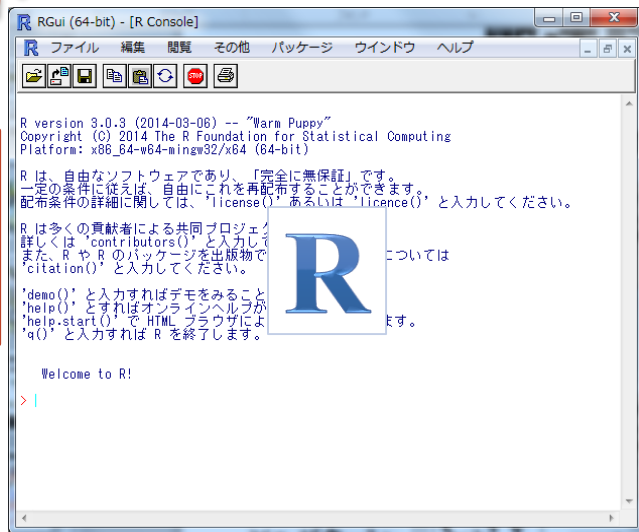


RNUOPTとは？

Numerical Optimizer をエンジンとした
最適化機能を提供する R のパッケージ

R のデータ加工、統計解析機能と
Numerical Optimizer での最適化機能が融合

R のインターフェースで
Numerical Optimizer
を使用できます



もちろん
Visual R Platform
からも使用可能！

Visual R Platform

```
Rスクリプト
```

データ読み込 R再起動 出力データ形式 R 最前面に表示 HELP

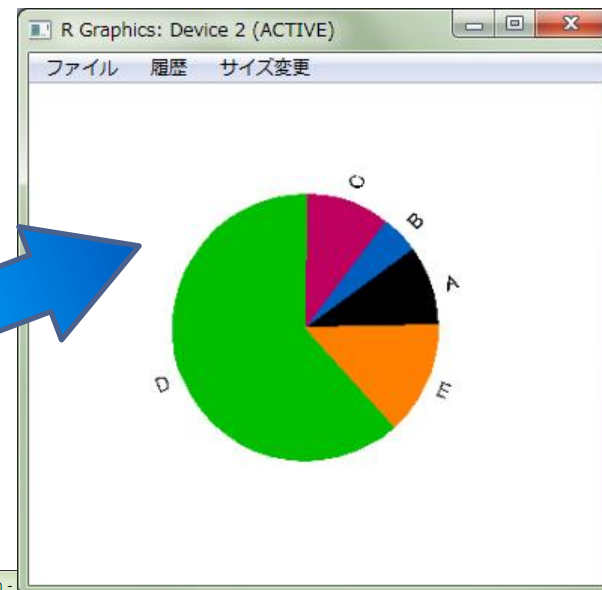
```
1 library(RNUOPT)
2
3 Marko <- function(Q,rbar,rmin)
4 {
5   Asset <- Set()
6   Q <- Parameter(index=dprod(Asset,Asset),Q)
7   rbar <- Parameter(index=Asset,rbar)
8   i <- Element(set=Asset)
9   j <- Element(set=Asset)
10  x <- Variable(index=Asset)
11  risk <- Objective(type="minimize")
12  risk ~ Sum(x[j]*Q[i,j]*x[i],i,j)
13  Sum(rbar[j]*x[j],j) >= rmin
14  Sum(x[j],j) == 1
15  x[j] >= 0
16 }
17
18 s <- System(model=Marko,Q,rbar,0.01)
19 sol <- solve(s)
20
21 x <- as.array(current(sys,x))
22 pie(1000*x,names=dimnames(x)[[1],inner=1.2)
```

入力データ名
input

出力結果名
result

1行実行

保存 Cancel



**Rスクリプトで
RNUOPTパッケージを利用！
結果の可視化も容易！**

Visual Analytics Platform

プロジェクト(E) ツール(E) 表示(V) ウィンドウ(W) 製品(P) ヘルプ(H)

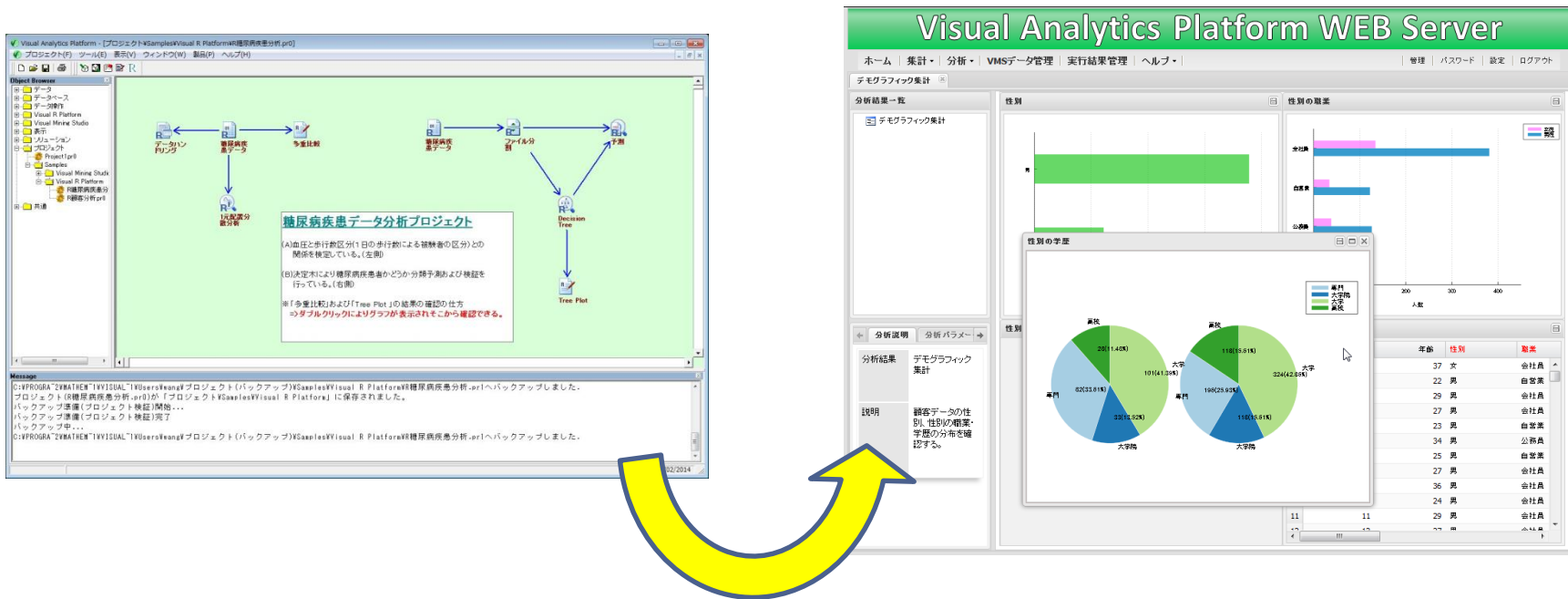
- 統計
- Visual R Platform
- 前処理
- 統計量
- 分類分析
- クラスタ分析
- アンジェーション分析
- 多変量解析
- 検定
- 時系列分析
- Rスクリプト

Rスクリプト

Message

20:12:17 11/20/2014

- VAP Web Serverとの連携により、VRPで作成したプロジェクトを、そのままWebアプリケーションとして公開できます



VRPプロジェクトをWebで公開

NTTデータ 数理システム 営業部またはデータマイニング部まで

R全般のサポート、
コンサルティング、Rによる
システム開発も承ります
何でもご相談ください！

NTT DATA

株式会社NTTデータ 数理システム

TEL : 03 - 3358 - 6681

FAX : 03 - 3358 - 1727

【URL】 <http://www.msi.co.jp>

【E-mail】 vrp-info@msi.co.jp