

# 応用プログラミング3 第4回 GUIによる データ分析

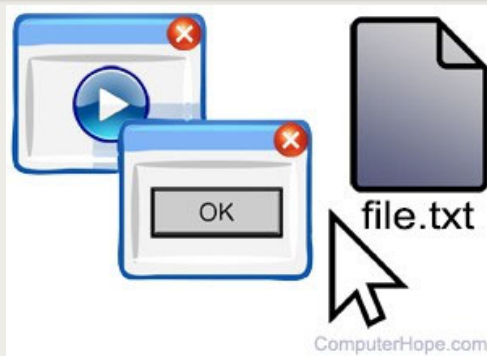
専修大学ネットワーク情報学部  
田中健太

# 1. GUIによるデータ分析

1

第4回の講義では、データ分析をGUIで行うことができるRadiantというパッケージを中心に紹介します。はじめに、一般論としてのGUIについて確認しておきましょう。

## 1.1 CUIとGUI



<https://www.computerhope.com/jargon/g/gui.htm>

```
$ cd /home/user/dir
```

- **GUI: Graphical User Interface**

- マウス
- アイコン
- フォルダー

- **CUI: Character based User Interface**

- キーボード
- コマンド
- ディレクトリ

2

コンピューターを操作する上でのインターフェースには、大きく分けてCUIとGUIという2つがあります。CUI (Character based User Interface) は、名前の通り画面に表示される文字を読み取り、キーボードでコマンドを入力し、コンピューターに指示をします。1960年代に普及 (それ以前は紙テープやパンチカードなど) し、現在でもWindowsのコマンドプロンプトやPowerShell、macOSのターミナルアプリ (アプリ内で起動するシェル) などCUIが提供されており、特にIT系のビジネス現場などでは、CUIでサーバーやソフトウェアを設定し、運用することが多くあります。

一方、GUI (Graphical User Interface) は、画面に現実の作業環境のメタファーとしてデスクトップが表示され、そこに配置されたさまざまなアイコンやメニューをマウスで操作します。1980年代にパーソナルコンピューターに搭載され、一般化しました。現在のコンピューターやスマートフォン、コンビニのセルフレジや駅の券売機など、画面に表示される要素をマウスや指で操作するものは、すべてGUIです。

CUIとGUI、どちらが使いやすいかは個人の好みやコンピューター・装置の利用形態によります。GUIはITに詳しくない人でも直感的に操作できますが、一方で操作を指示したり教育を提供する場合、「この画面の右上にあるいくつかのアイコンのうち、この模様のものをダブルクリックして…」と説明が冗長になります。また、何度も同じ操作を繰り返す際も面倒です。CUIは、やるべきことを文字として示せますので、伝達が容易です。また、事前に操作 (コマンド) を書き出し、推敲することでミスを防げます。そして、コマンドを書き溜めておいて、繰り返し処理を行う際にはコピーアンドペーストで貼り付ければよいので、再現性も高くなります。

## 1.2 世間一般のニーズ



筆者による雑コラ

- そもそも私たちがやりたいことはデータ分析であり、プログラミングではない
- GUIでできるなら、その方が楽（分析に集中できる）
- ただしGUIでは再現性が課題になる
- また、用意されたメニュー以上のことはできない

3

ここでは、特にデータ分析をする場合について考えてみます。世の中にはさまざまなデータ分析ソフトウェアがあり、そのインターフェースもCUI、GUIさまざまです。一般に、非データサイエンス人材（悪い意味ではなく、ビジネス現場でデータから得られる知見を活用する立場）が利用するソフトウェアはGUI、データサイエンス人材が利用するソフトウェアはCUIであることが多いです。高度なデータ分析の結果を評価するには、その過程を検証する必要があるため、プログラムとして一覧でき、再現性のあるCUIのほうが適していると考えられます。

そのため、現状では「簡単なデータ分析はGUI、高度なデータ分析はCUI」というようになっています。しかし、実際にはデータ分析の難しさとインターフェースは切り分けるべきもので、高度なデータ分析もGUIで容易にできるなら、その方が良いわけです。

つまり、「再現性があって高度なデータ分析がGUIでできるソフトウェア」が求められています。今回の講義では、それに当てはまると考えられるRのパッケージを紹介します。

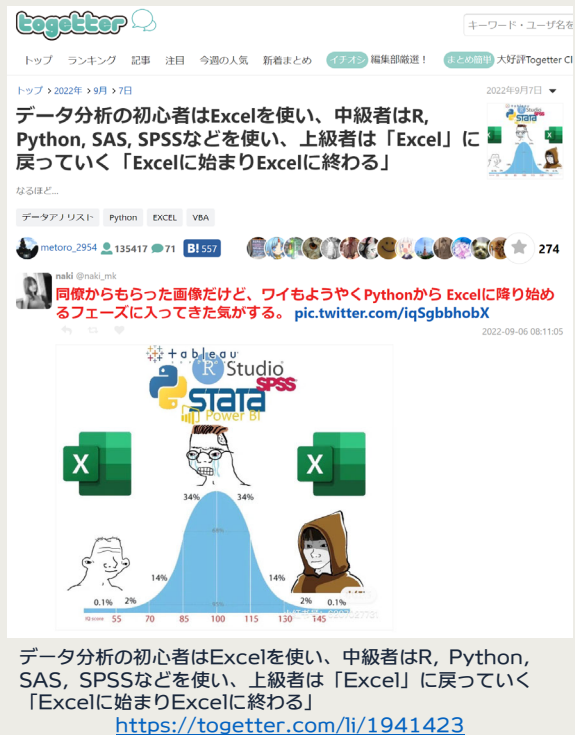
## 2. さまざまな GUIデータ分析ツール

4

この節では、GUIでデータ分析ができるツールを取り上げ、その利用場面や特徴を紹介します。

## 2.1 Microsoft Excel

- 使いこなせば、文系学部  
の卒論からビジネスまで、  
大抵の分析はできる
- アドインを活用すれば  
さらに分析・可視化の幅が  
広がる
- ただし、計画的にセル、  
シートを使わないと、  
再現性が極めて悪い
- また、バグが多かった  
昔の経験から、毛嫌いの  
エンジニア、教員も多い



5

Microsoft Excel (以下、Excel) は、世界で最も広く使われているデータ分析ソフトウェアと言えるでしょう。(説明するまでもないですが) 表構造のデータを「シート」に入力し、「セル」にさまざまな計算式を入力することでデータの集計や演算ができます。また、セルのデータをもとにグラフィックスを作成し、データの特徴を可視化できます。

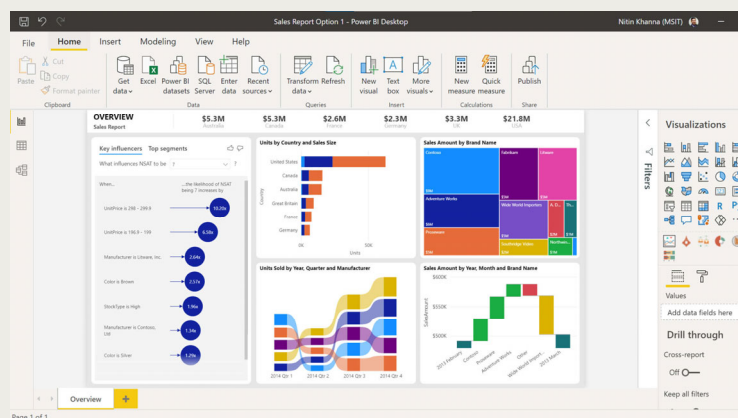
Excelは随時バージョンアップされ、高度な分析機能やグラフィックスが実装され、Web APIなどとの連携も進み、高機能化しています。データ分析の基本から中級(何をもって、というのはありますが)は、Excelで十分実行できます。また、Excelは"ほぼ全ての"パソコンにインストールされており、データや分析結果を社内や分析チーム内で共有しやすいという利点もあります。特に、非データサイエンス人材に分析結果を活用してもらおうと考えるのであれば、「誰でも使える」Excelに落とし込むことも重要です(上図参照)。

一方、機械学習などの高度な手法には対応していないため、有償のアドインを導入するか、セルに数式を入力して実行する必要があります。セルに数式を入力する場合、式の妥当性をチェックしたり、ユーザーの操作によって「壊れない」ようにセルをロックしたり、ユーザーへの十分な説明が必要です。

- 参考1: エクセル統計アドイン <https://bellcurve.jp/ex/>
- 参考2: XLSTATアドイン <https://www.xlstat.com/ja/>
- 参考3: Prediction Labsアドイン <https://predictionlaboratory.com/>
- 参考4: HAD (関西学院大学清水教授によるExcel VBAアプリ) <https://norimune.net/had>

## 2.2 Microsoft Power BI

- 無償で利用できるBI (Business Intelligence) ツール
- 近年、BIはデータの可視化という意味で使われる
- さまざまな形式のデータを読み込み、多様なグラフィックスを作成できる
- RやPythonと連携し、モデリング、予測を行うこともできる



<https://powerbi.microsoft.com/ja-jp/>

6

データを可視化し、特徴を理解することを目的とするアプローチをBI (Business Intelligence) と呼びます。1980年代から存在する言葉で、当初は今で言う「データ活用」全体を指すものでしたが、その後データサイエンスやAIなどの言葉が出てきたことで、現在はデータの可視化を指すようになっています。

BIツールは、Excelよりも豊富なグラフィックス機能を有し、GUIでダッシュボードと呼ばれるような、さまざまなグラフを組み合わせたアウトプットを作成できます。ビジネスサイド (営業担当者や管理職、経営層など) はダッシュボードを見て、現状を把握し、次にとるアクションを意思決定します。

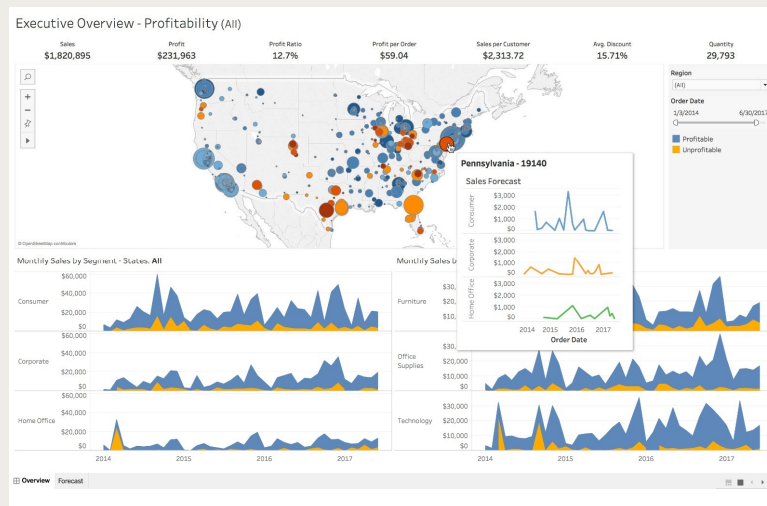
ここでは、いくつかのBIツールを紹介しますが、はじめにMicrosoftがWindows版を無償で公開しているPower BIを取り上げます。Excelファイルやさまざまなデータベース、Webサービスからデータを読み込み、GUIで容易に可視化できます。また、標準でRやPythonと連携する機能があり、RやPythonで高度な分析を行った結果を、Power BI上で可視化することができます。

ただし、どういうわけか (Microsoft製品なのに) 起動が遅かったり、(Microsoft製品なのに) 作成したグラフをPowerPointやWordにコピーアンドペーストできなかったり、不便なところもあります。



## 2.3 Tableau

- 商用BIツールで（おそらく）最もシェアが大きい
- PC、モバイル、サーバーそれぞれの環境で容易にダッシュボードを作成、共有できる



<https://www.tableau.com/ja-jp>

7

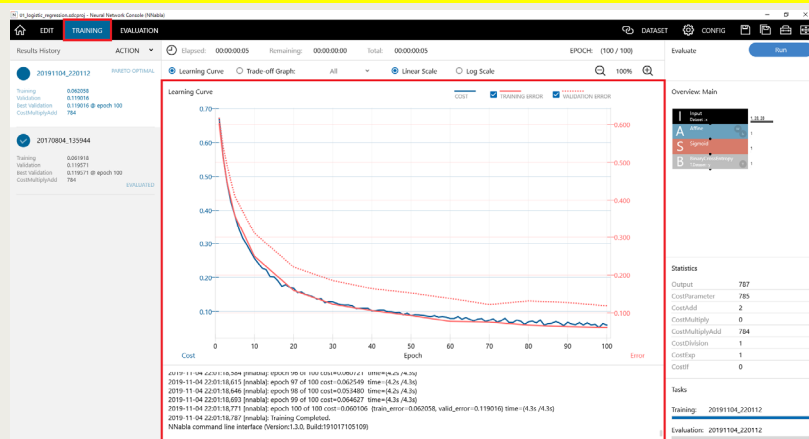
次に、Power BIと同等以上に広く使われているBIツールであるTableau（タブロー）を紹介します。基本的には、Power BIと同様に、さまざまなデータソースに対応し、GUIでダッシュボードを作成、共有できます。また、モバイルアプリも提供されており、タブレットなどでダッシュボードを表示し、顧客先などでデータをもとに議論することもできます。

Tableau Public ( <https://public.tableau.com/app/discover> ) に、ユーザーによって作成されたダッシュボードが多数公開されています。また、Tableau Publicのデスクトップ版アプリを無償で利用することもできます。



## 2.4 Neural Network Console

- ・ ソニーが開発、公開したGUIでニューラルネットワークモデルの開発ができるツール
- ・ クラウド版は有償、Windowsローカル版は無償
- ・ **ディープラーニングモデルをマウス操作で構築できる**



<https://dl.sony.com/ja/>

8

近年では、ディープラーニングもGUIでできるようになっています。ソニーが開発、公開しているNeural Network Consoleは、パソコンまたはクラウド上で、ドラッグアンドドロップを中心としたGUIでニューラルネットワークを構築できます。Windows版は無償で利用できます。従来、ニューラルネットワークやその発展形のディープラーニングは、高度な数学的知識をもとに、プログラムとして実装することが求められましたが、本ソフトウェアでは、(もちろん理論の理解なしに使うことはできませんが) 部品を配置してつなげるだけで、一般的な表構造データや、画像、音声などに対応したネットワークを構築できます。

上記のようにWindows版は無償で利用できますが、パソコンで動かす以上、計算能力には限界があります。複雑なネットワークを構築したい場合は、有償のクラウド版を利用することになるでしょう。

### 3. RにおけるGUI分析環境

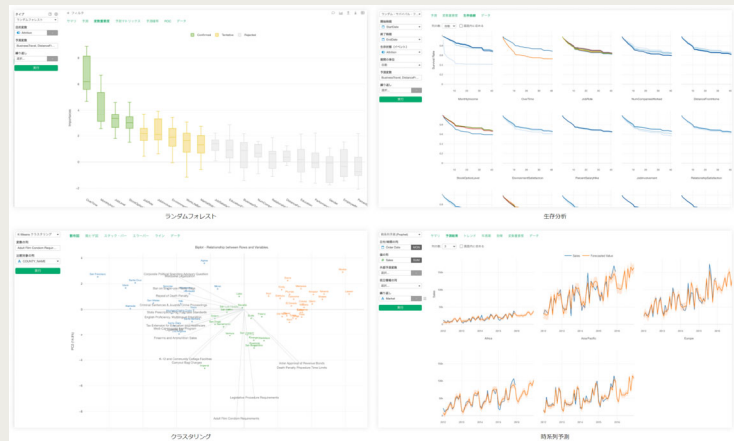
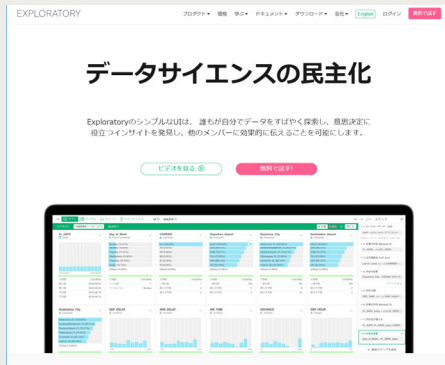
9

ここまで、R以外の、GUIで利用できる分析ツールを紹介しました。他にも、SPSSや国内ベンダーが開発するソフトウェアなど、もはやGUIで大抵の分析はできるようになっています。

それはRにおいても同じです。この授業では、基本的にR「言語」のプログラミングを学びますが、Rにおいても、さまざまなデータの加工や集計、高度な分析をGUIで実行するためのパッケージが複数開発され、利用されています。また、Rをバックエンド（分析エンジン）として利用し、優れたGUIを提供するソフトウェアも提供されています。この節では、そのようなRにおけるGUI環境を紹介します。

## 3.1 Exploratory

- ・ シリコンバレーで日本人が創業した Exploratory社の製品
- ・ **Rをバックエンドに、高度なGUIを提供する**
- ・ 学生は無償で利用でき、日本語の情報も豊富に入手できる



<https://ja.exploratory.io/>

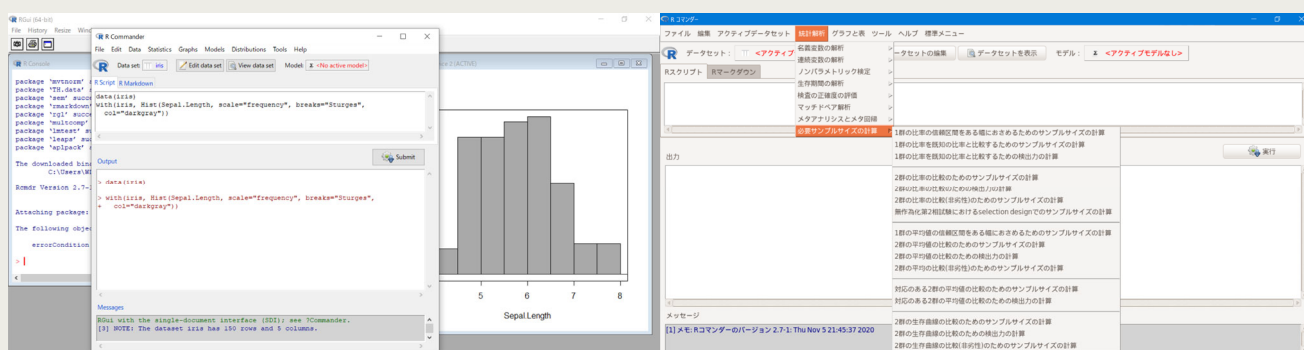
10

Exploratoryは、Rをバックエンドとして利用し、高度で使いやすいGUIを提供するソフトウェアです。アメリカ・シリコンバレーで日本人が創業したExploratory社が開発、提供しています。データの読み込みから複雑な加工、可視化、高度な機械学習やテキストマイニング、結果のレポート・共有まで、一般的なデータ分析の需要をほぼカバーする高機能が特徴です。また、日本人スタッフが中心に開発しているため、メニューやマニュアルを含め完全に日本語化されており、ビジネス用途でも利用しやすい(みんな学校で英語を学んだはずなのに、社会人になると英単語が出てくるだけで拒絶する人たちが多くいます)ソフトウェアです。

なお、ビジネスで利用するためには年間ライセンスを購入する必要があり、やや高額です。学生は無償でフル機能を利用できるライセンスがあります。

## 3.2 Rcmdr / EZR

- 主に統計学的アプローチの分析手法をカバーしている
- EZRは日本の医学者による医療統計分野を強化したプラグイン



<https://socialsciences.mcmaster.ca/jfox/Misc/Rcmdr/>

<https://www.iichi.ac.jp/saitama-sct/SaitamaHP.files/statmed.html>

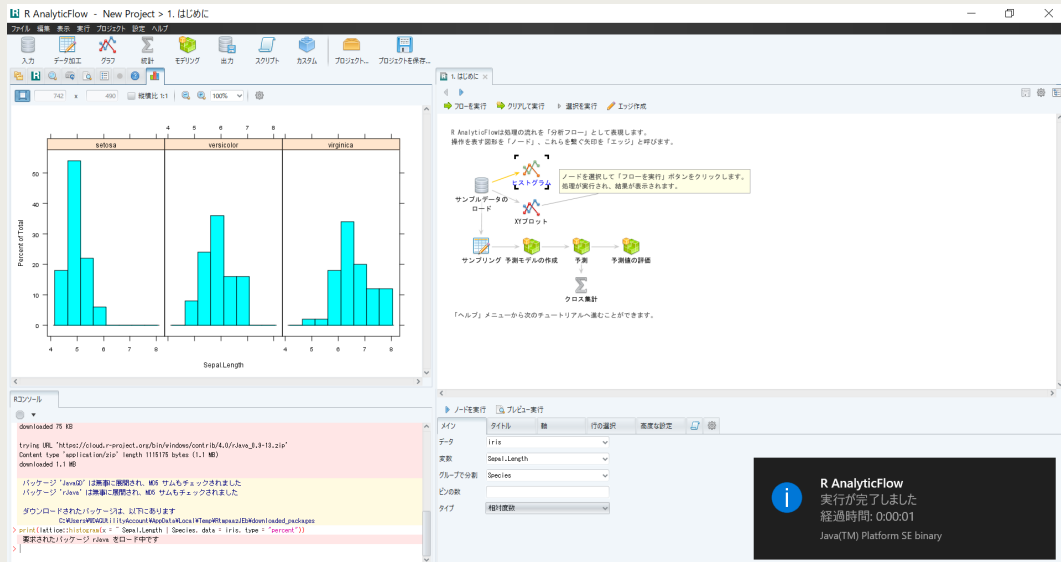
ここで紹介するのは、Rのパッケージとして提供されているGUIです。海外で2000年代はじめから開発されているRcmdr (Rコマンダー) は、データの加工や統計的検定、一部の機械学習 (一般化線形モデルやクラスター分析) に対応しています。歴史があるために、現代のソフトウェアのようなアイコンとドラッグアンドドロップでの操作ではなく、文字で表示されるメニューをクリックしていくインターフェースです。歴史があるので、日本語化もしっかりとされています。

Rcmdrをベースに、医療統計 (治療法や薬の効果を検証するなど) において用いられる分析手法を追加したパッケージがEZR (イーザー) です。日本の医学者が開発、公開しています。

いずれも、GUIで操作した結果をRプログラムとして出力できるので、結果の再現性は担保できます。なお、基本的にRの組み込み (標準) 関数の範囲にのみ対応しているので、tidyverseなどモダンなパッケージを利用することはできません。

## 3.3 R AnalyticFlow

- 日本のdef-prime社によるGUI
- マウスで部品を並べて分析フローを作成、実行できる



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

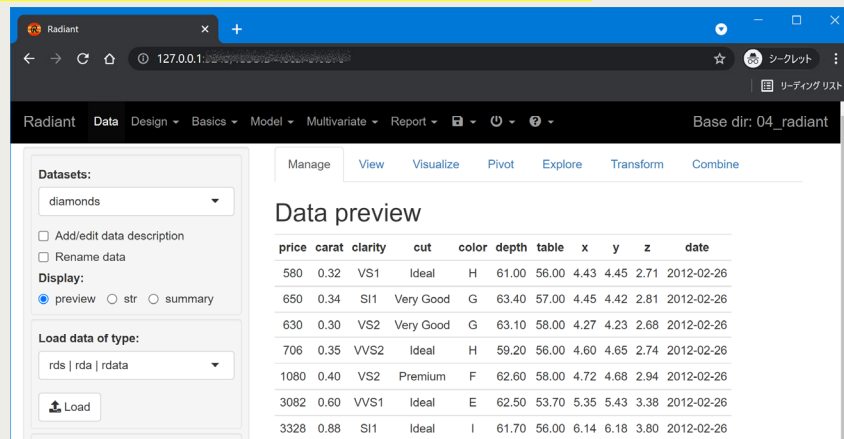
12

R AnalyticFlowは、日本のdef-prime社が開発、公開しているソフトウェアで、前述の Neural Network Consoleのように、部品を並べて分析フローを作成、実行できます。データの読み込みから加工、モデリング、評価までプログラミングせずに進めることができます。Java言語で開発されており、Windows、macOS、Linuxのいずれの環境でも利用できます。

部品の細かい動作については、R言語でプログラミングすることも可能です。また、最新版ではggplot2などモダンなパッケージにも対応しています。

## 3.4 Radiant

- Vincent Nijs氏によるRパッケージ
- Webブラウザで操作する "データ分析環境"
- データの加工から集計、可視化、モデリング、レポートまで「なんでも」できる



<https://radiant-rstats.github.io/docs/>

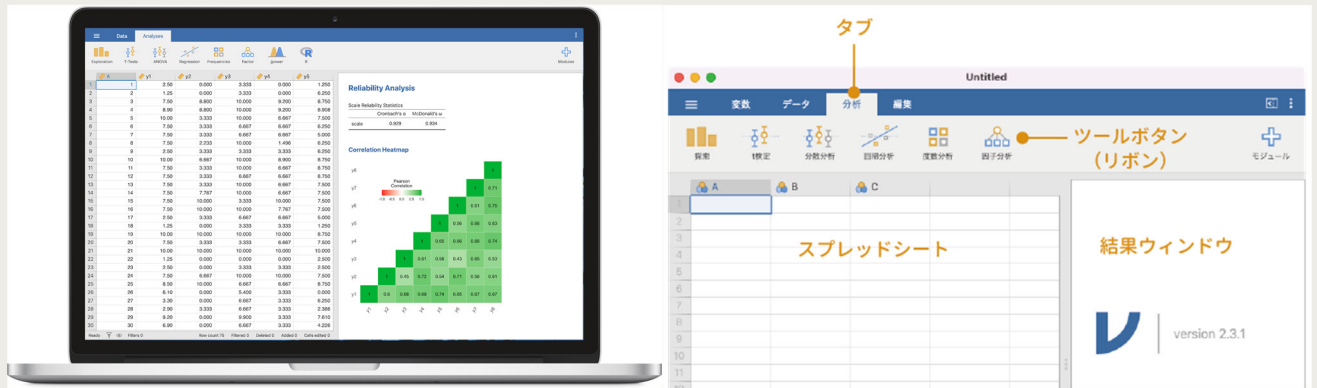
13

Radiantは、カリフォルニア大学サンディエゴ校 (UCSD) のVincent Nijs氏が開発、公開しているパッケージです。第2回の講義で紹介したWebアプリケーションフレームワークShinyを使って開発されており、データの読み込みから加工、探索的データ解析 (EDA)、機械学習までデータ分析のニーズの大半をカバーしています。また、操作の結果をR Markdownのレポートして出力できることも特徴で、結果の再現性を担保したり、非Radiantユーザーとのやり取りをしやすくします。

今回の講義では、Radiantを取り上げ、利用法を紹介します。

## 3.5 jamovi

- The jamovi projectによる「統計解析プラットフォーム」
- 市販の統計ソフトウェアに似た使いやすいUI
- "jamoviで学ぶ心理統計" や "jamovi完全攻略ガイド" という、優れた日本語ドキュメントがある



<https://www.jamovi.org/>

14

jamoviは、The jamovi projectが開発、公開しているソフトウェアです。バックエンドでRを使用し、インターフェースはSPSSなど有償の統計ソフトウェアのような外見になっています。

データの読み込みから加工、統計学的な分析まで、人文社会系でのデータ分析に必要な機能がカバーされています。決定木やランダムフォレスト、ニューラルネットワークなどの機械学習も、無償のモジュール（拡張機能）を追加することで対応します。

jamoviはRとは独立しており（Rを組み込んでいない）、Rがインストールされていない環境でも動作します。インターフェースは最も洗練されているように思いますので、「卒業研究でデータ分析を『しなければいけなくなった』」人には適しているでしょう。

- 参考1: jamoviで学ぶ心理統計 <https://bookdown.org/sbtseiji/lswjamovi/>
- 参考2: jamovi完全攻略ガイド [https://bookdown.org/sbtseiji/jamovi\\_complete\\_guide/](https://bookdown.org/sbtseiji/jamovi_complete_guide/)



## 4. Radiantによるデータ分析

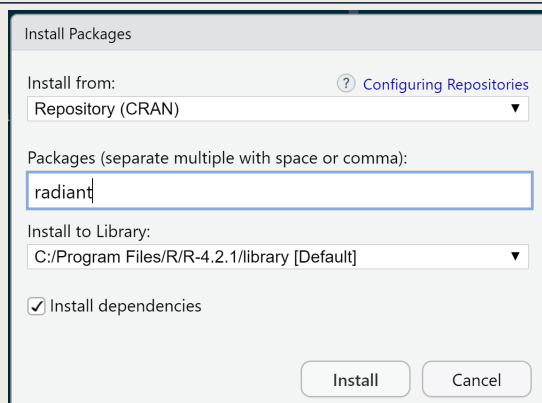
15

ここから、前節で紹介したRadiantパッケージの使用方法を紹介します。Posit Cloudの場合、Radiantのウィンドウがブラウザのポップアップとして開きます。最近のブラウザは標準でポップアップブロックが有効になっていることが多いので、<https://posit.cloud>からのポップアップを許可するよう設定する必要があります。

## 4.1 Radiantのインストール

- RadiantはCRANに登録されている
- `install.packages("radiant")` でインストールする
- macOS、Linuxではソースからのコンパイルに  
凄まじい時間がかかる

```
install.packages("radiant")
```



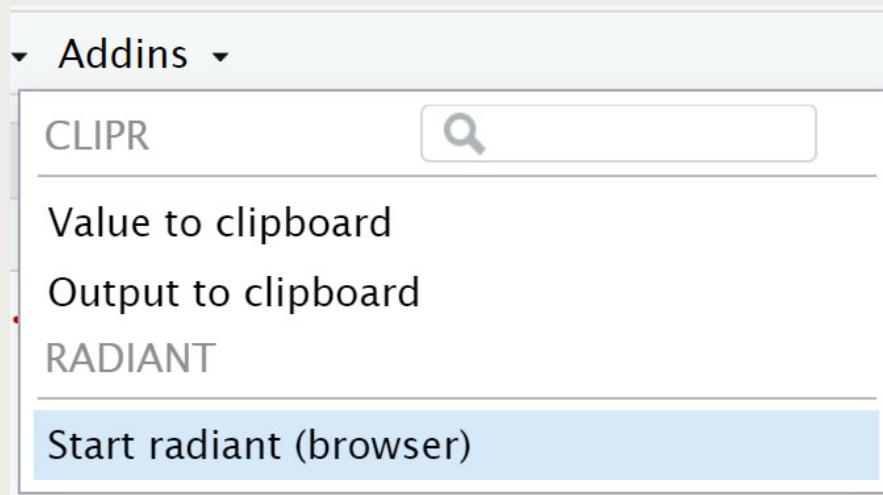
16

RadiantパッケージはCRANに登録されているため、インストールは `install.packages("radiant")` とするだけです。依存関係にあるパッケージが多いため、WindowsやPosit Cloudでのバイナリ（コンパイル済み）パッケージのインストールであっても、時間がかかります。また、macOSやLinux環境では、ソースパッケージをダウンロードし、コンパイルしてインストールするため、かなりの時間がかかります。

## 4.2 Radiantの起動

- Radiantパッケージの `radiant()` 関数を実行する
- RStudioでは、"Addins" にメニューが登録されている

```
radiant::radiant()
```



17

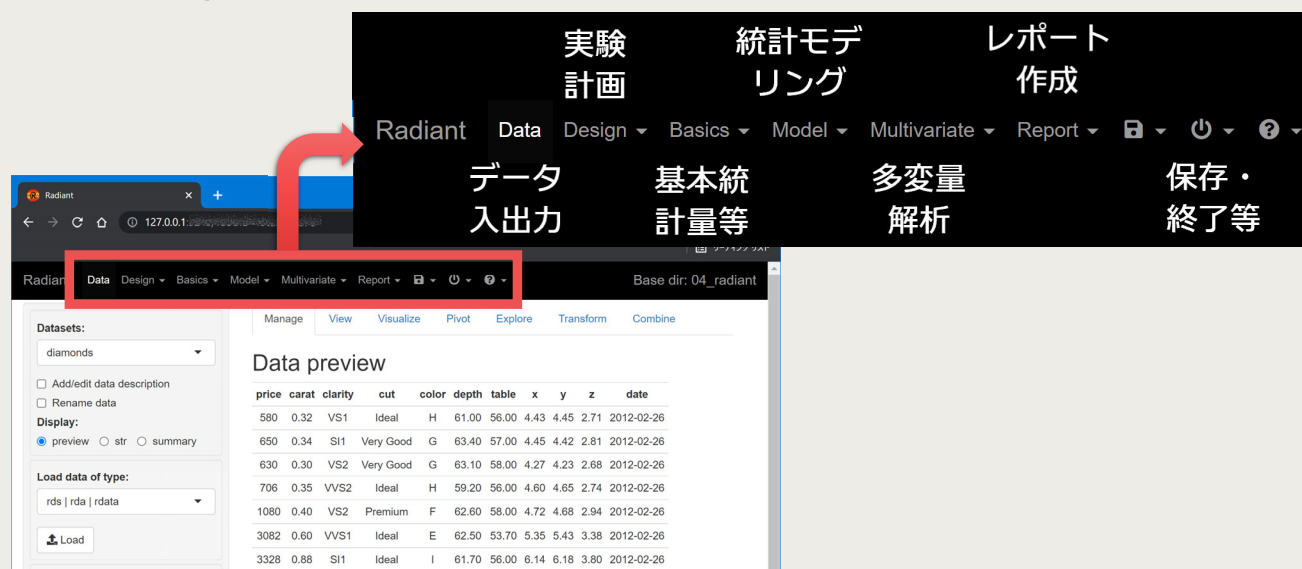
Radiantを起動する方法は複数あります。まず、コンソールで `radiant()` 関数を実行すると、Shinyアプリが起動し、OS既定のブラウザでウィンドウが開きます。また、既定のブラウザとは別のブラウザで開きたい場合は、`radiant_url()` 関数で、アプリにアクセスするためのURLが表示されます。RStudioを使用している場合、`radiant_viewer()` 関数で右下ペインの "Viewer" タブで、`radiant_window()` 関数でRStudio内の別ウィンドウとして起動します。

また、RStudioでは、メニューからRadiantを起動できるよう、アドインが追加されます。画面上部の "Addins" メニューから "Start Radiant (browser)" を選択するとブラウザでRadiantが開きます。

他に、R / RStudioを起動してからさらにRadiantを起動するのが面倒だ、という場合は、`launcher()` 関数を実行すると、デスクトップにRadiantを起動するショートカットを作成できます。

## 4.3 画面構成

- Radiantはブラウザ上で操作が完結する
- 上部のメニューバーからデータの読み込みや可視化、分析を進めていく

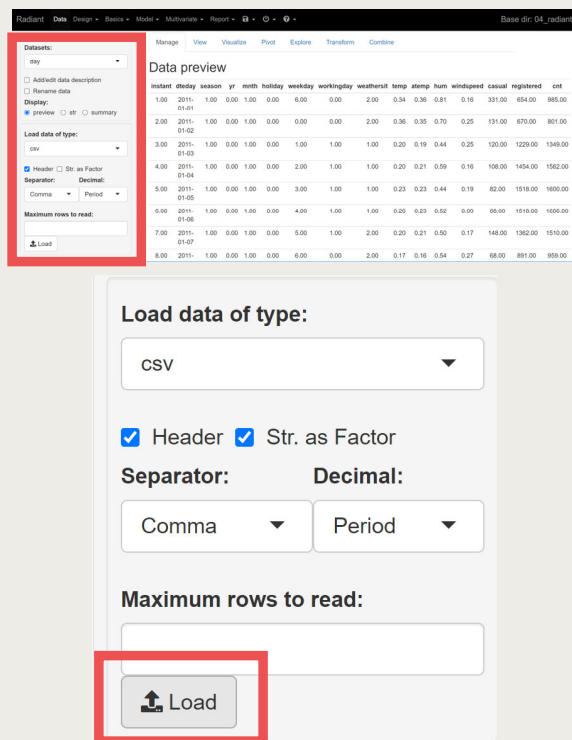


18

はじめに、Radiantの画面構成を紹介します。Radiantが起動すると、サンプルデータが読み込まれた "Data" メニュー画面が表示されます。ここから、画面上部のメニューバーに表示されたさまざまな項目を選択し、操作していきます。

次のページ以降で、それぞれのメニューの機能を紹介していきます。

## 4.4 データの読み込み



- "Data" メニューからCSVファイルなどを読み込める
- クリップボード経由の貼り付けもできる
- Radiant起動前に作成したオブジェクトも読み込める
- 文字コードはUTF-8限定

19

Radiant起動時に表示される "Data" メニューでは、ファイルなどからのデータの読み込み、データ表示、可視化、集計、加工などさまざまな操作ができます。

まず、データの読み込み操作について紹介します。画面左側のメニューから、ファイル形式などを指定します。Excel形式 (.xlsx) には対応していませんが、CSVファイルやR独自のデータ形式 (.rds など)、Radiant起動前にR上で作成したデータフレームなどを読み込むことができます。また、他のアプリケーションやWebページで「コピー」したデータをクリップボード経由で読み込むこともできます。読み込む際に、区切り文字や先頭から読み込む行数を設定できます。

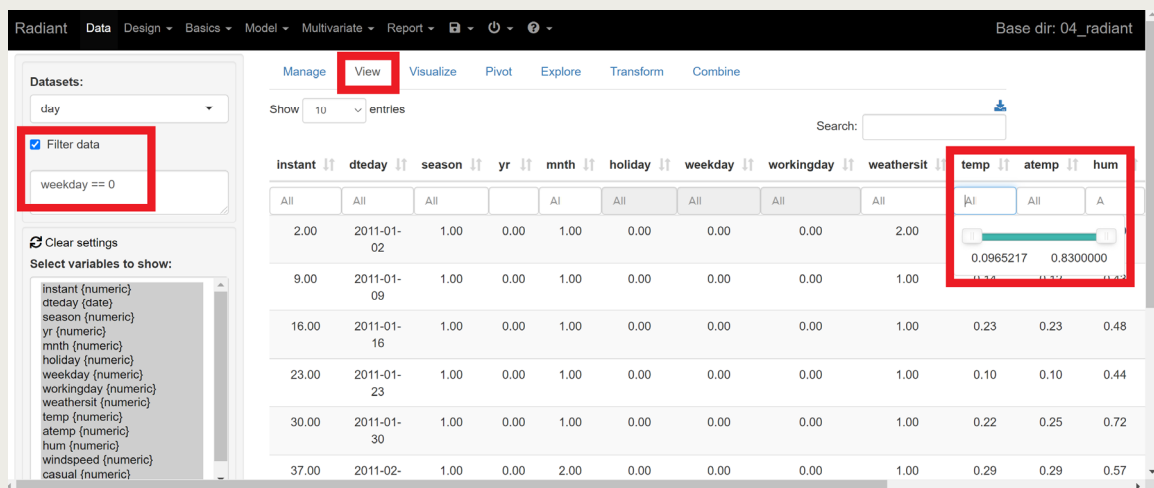
事前に別のパッケージを使い、Excelファイルやデータベースから分析したいデータをデータフレームに読み込んでおけば、Radiantでも利用可能です。

なお、テキストファイルの文字コードはUTF-8にのみ対応しています。

起動時には、サンプルデータとして組み込みのdatasetsパッケージに含まれる Titanic データと、ggplot2パッケージに含まれる diamonds データが読み込まれています。他のデータセットを使用するには、"Datasets" メニューのドロップダウンリストから選択します。

## 4.5 データの観察 (1) "View" タブ

- "Data" - "View" メニューでデータ全体の確認ができる
- 並び替え、フィルタリングもできる



20

ここから、"Data" メニューの詳細な機能について紹介していきます。"Data" メニューの中にサブメニューとして "Manage", "View", "Visualize", "Pivot", "Explore", "Transform", "Combine" という項目があります。

はじめに、"View" タブではデータ全体を一覧表示できます。表示する列の選択、条件式を指定してのフィルタリングもできます。

## 4.6 データの観察 (2) "Explore" タブ

- "Data" - "Explore" メニューで探索的にデータの特徴を観察できる
- 列を選択し、任意の基本統計量を算出したり、グループ化して差異を確認できる

タイプ	variable	n_obs	mean	min	max	sd	n_missing
一般酒	アルコール分	632	15.306	12.700	17.300	0.749	11
吟醸酒	アルコール分	132	15.830	14.200	18.100	0.758	4
本醸造酒	アルコール分	86	15.501	14.400	16.400	0.422	4
純米酒	アルコール分	123	15.457	13.900	16.900	0.632	1
NA	アルコール分	25	15.480	14.000	16.800	0.712	0

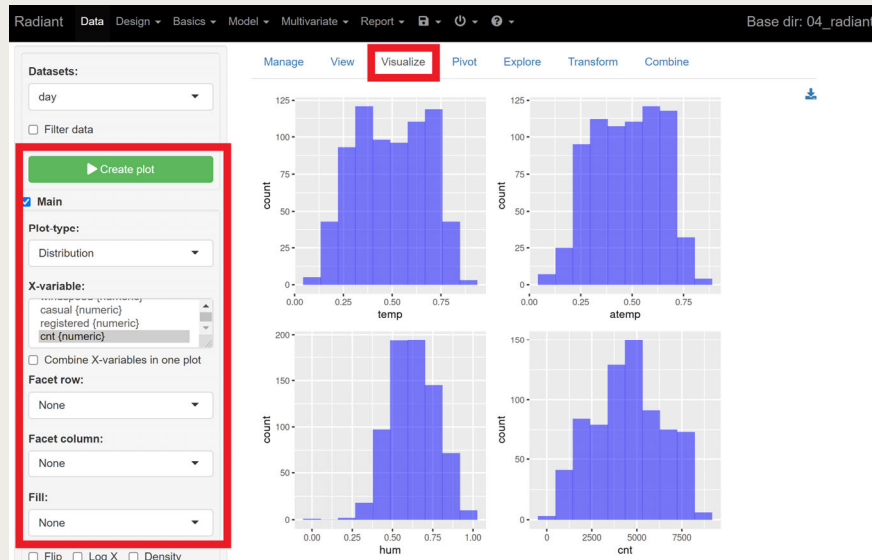
21

"Explore" タブでは、データに対してさまざまな集計、加工を適用し探索的に観察できます。列を選択し、"Group by" でグループ化変数を指定できます。"Apply function(s)" でデータ数や平均などの基本統計量、欠損値の数など、適用する集計処理を選択します。



## 4.7 データの観察 (3) "Visualize" タブ

- グラフィックスもGUIで手軽に作成できる
- "View" タブや "Explore" タブでフィルタリング、集計した結果も使用できる



22

"Visualize" タブでは、さまざまなグラフィックスをGUIで簡単に作成できます。これだけでも、Radiantを使う価値があると言えるでしょう。

まず、"Plot type" メニューでグラフィックスの種類を選択します。英語表記ですが、以下のグラフィックスを作成できます。

- Distribution: ヒストグラム
- Density: 密度関数
- Scatter: 散布図
- Line: 折れ線グラフ
- Bar: 棒グラフ
- Box: 箱ひげ図

また、"Facet row" または "Facet column" に変数 (列) 名を選択すると、その変数でグループ化したグラフィックスを縦 (row) または横 (column) に並べて出力します。"Fill" を指定して1枚のグラフィックスの中でグループを塗り分けることもできます。

グラフィックスのタイトルや軸ラベル、見た目の細かい調整は下部の "Labels" と "Style" の項目で設定できます。

作成したグラフィックスは、画面右側に表示されるダウンロードアイコンから画像としてダウンロードできます。

## 4.8 データの加工

- 標準化、リコード、欠損値の処理、型変換などの加工を列ごとに行える
- 多少の条件式は書くが、R言語を意識する必要はない

The screenshot shows the RStudio interface with the 'Transform' tab selected. The 'Select variables' section is highlighted in red, showing a list of variables including 'instant', 'dteday', 'season', 'yr', 'mnth', 'holiday', 'weekday', 'workingday', 'weathersit', 'temp', 'atemp', 'hum', 'windspeed', and 'casual'. The 'Transformation type' section is also highlighted in red, showing options like 'None (summarize)', 'Bin', 'Change type', 'Normalize', 'Recode', and 'Remove/reorder levels'. The main data table shows 7 rows of data with columns: instant, dteday, season, yr, mnth, holiday, weekday, workingday, weathersit, temp, atemp, hum, windspeed, and casual. The 'Transform command log' at the bottom shows the command: '1 ## change variable type', '2 day <- mutate\_at(day, .vars = vars(season), .funs = as\_factor)', '3'.

23

"Transform" タブでは、データの列ごとにさまざまな加工を適用できます。列を選択し、適用したい加工を指定します。連続データの離散化 ("Bin" / "Recode") や標準化 ("Normalize")、型変換 ("Change type") など「前処理」として行うことの多い処理がカバーされています。

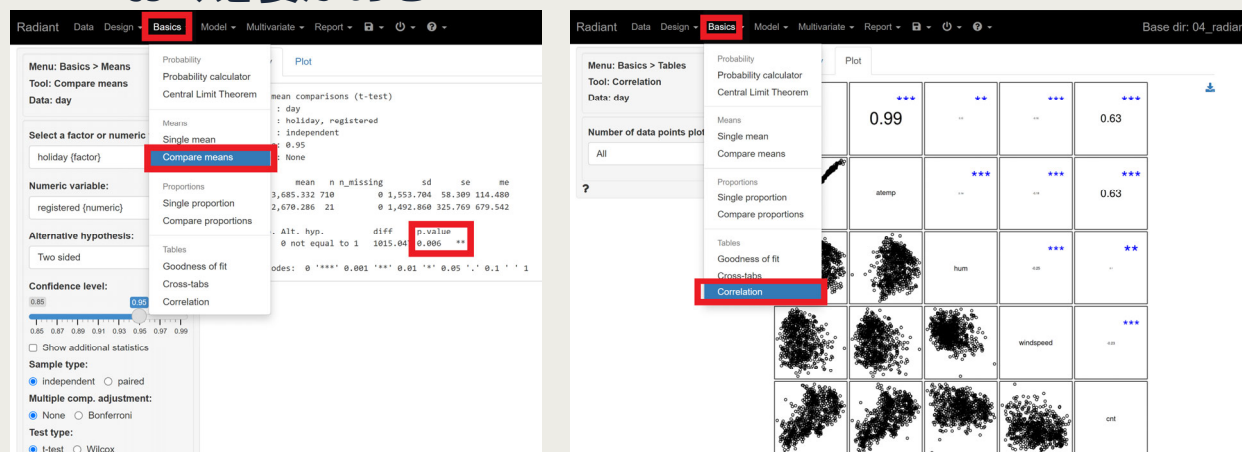
"Recode" では、Yes / No などの文字を1 / 0に変換する、100点満点の数値のうち、80以上を "A"、70点以上を "B" という文字に変換する、といったことができます。この際、例えば 'Yes' = 1; 'No' = 0 や lo:80 = 'A'; lo:70 = 'B' といったような条件式を記述します。式は左から順に評価されるので、記述の順番に注意してください。内部では、carパッケージの Recode() 関数が呼び出されているようで、関数のマニュアルに lo や hi など条件式の書式が解説されています。

参考: Recode() 関数マニュアル <https://rdr.io/cran/car/man/recode.html>

加工後のデータ (列) は、新しい名前データフレームに追加したり、元の列に上書きすることができます。最後に、下部の "Store" ボタンを押さないで保存されないので注意してください。

## 4.9 統計学的分析

- "Basic" メニューで統計的検定や相関分析など、基本的な統計学的分析ができる
- 平均値の差の検定 ( $t$ 検定) は "Compare Means" でできるが、グループ化変数をfactor型に変換しておく必要がある



24

ここからは、"Basic" メニューで利用できる機能を紹介します。"Basic" メニューでは、主に古典的な確率統計学のアプローチに基づく分析が実行できます。具体的には、 $t$ 検定や二項検定、 $\chi^2$  検定 (前回配布した参考資料を参照)、相関分析などができます。

それぞれ、対象となる1つまたは2つの変数を指定し、有意水準や片側・両側検定の設定などを行います。結果は、数値中心の出力と "Plot" タブで出力されるグラフィックスで確認できます。

なお、 $t$ 検定については、3群以上の組み合わせについても比較可能で、検定の多重性の問題に対処するBonferroniの補正を適用できます。ただ、このやり方は「分散分析をせずに直接多重比較をしている」ように思われ、統計学のスタンダードな手順とは異なるので注意が必要です。(この文を書くために少し調べると、このあたりも最近はガイドラインが変わってきているようです。)

## 4.10 機械学習

- 一般化線形モデル、決定木、ランダムフォレスト、勾配ブースティング、ニューラルネットワークなどの機械学習アルゴリズムも使用できる
- 事前に "Transform" タブでデータを学習用と検証用に分割しておけば、精度評価もできる
- "Multivariate" メニューでクラスタリングやマーケティング領域の分析ができる



25

次に、"Models" メニューで実行できる、機械学習モデリングについて紹介します。データに目的変数と説明変数（特徴量）が含まれている場合、その関係をさまざまなアルゴリズムを用いてモデルとして表現できます。線形回帰やロジスティック回帰などの線形モデルやツリーベースのアルゴリズム、ニューラルネットワークなどが利用できます。

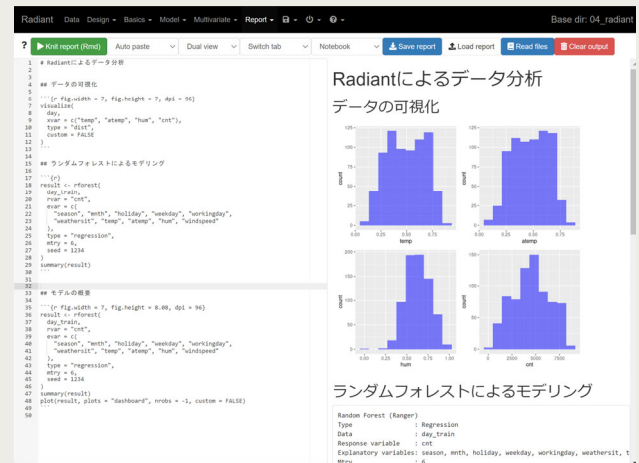
また、あらかじめデータを学習（モデル作成）用と検証・評価用に分割しておけば、一方のデータでモデルを作成し、もう一方のデータを予測して精度評価を行うこともできます。

作成したモデルを使い、"Predict" タブで予測することができます。また、"Plot" タブでは、モデルの当てはまりをグラフィックスで確認できます。"Dashboard" を選択すると、グラフィックスをまとめて出力できます。

なお、決定木については、分類（Classification）の際に目的変数が2値でなければなりません（Radiant以外では、もちろん3クラス以上の多値分類ができます）。

## 4.11 レポートिंग

- 研究であれビジネスであれ分析結果はレポートしなければならない
- Radiantの操作手順は裏でスクリプトとして記録されていて、各ページのレポートボタンを押すと、R Markdownファイルが作成される
- PDF出力はLaTeX環境でエラーになるが、Word, PowerPoint, HTMLは問題なく出力できる



26

最後に、これもRadiantの便利な点として、ここまでGUIで行ってきた操作を、R Markdownファイルとして出力し、レポートに変換できることがあります。Radiantの各メニュー画面には、左側にペンと紙のようなアイコン (📝) があります。これをクリックすると、"Report" メニューに遷移し、GUI操作の内容を、R Markdownのチャックとして出力します。

Radiant起動以降に行った操作は記録されており、例えばファイルからデータを読み込んだあと、欠損値を除去し、特定の列の数値を標準化し、列名を変更した、といった加工のプロセスも、すべてチャックに出力されます。分析の流れを記録するためには、"Data" メニューの "Manage" タブから順にアイコンをクリックしていくか、ばらばらに出力されたチャックをカットアンドペーストで並べ替える必要があります。

なお、Radiantの操作と対応するRプログラムは、Radiantパッケージで定義された関数を多用したものであるため、一般的なプログラムとは少し異なり、一見しての解読が難しいものになります。

作成したR Markdownドキュメントは、"Report" メニュー画面上部のボタンで、各種フォーマットに変換して出力できます。ただし、テンプレートなどは使用できないので、本格的なレポートを作成する場合には、.Rmd ファイルとしてダウンロードし、ローカルのR環境で変換したほうが良いかもしれません。

## 5. まとめ

## 5.1 今日の内容

---

- GUIデータ分析ツール
- Radiantによるデータ分析
  - 起動
  - データの読み込み
  - データの観察
  - モデリング
  - レポーティング



## 5.2 次回までの課題

---

- Classroomに掲示した 04\_radiant\_exercise.pdf の指示に従ってRadiantを操作し、HTMLレポートを作成してください
- 作成したファイルは、ファイル一覧でチェックを入れ、[more] メニューから [Export] を選択し、[Download] ボタンを押してダウンロードしてください
- 今回の提出物はHTMLファイルです
- ダウンロードしたファイルを、Classroomの課題ページから提出してください
- 提出期限: 2023-10-23 23:59