

RcmdrPlugin.DAtoolsforKids の使い方

関西大学商学部 荒木孝治

2010年11月19日

目次

1	R コマンダーとは	2
1.1	R コマンダーのしくみ	2
1.2	データの切り出し	6
1.3	プラグイン	9
2	教育用プラグイン RcmdrPlugin.DAToolsforKids	11
2.1	折れ線グラフ	13
2.2	棒グラフ	16
2.3	人口ピラミッド	18
2.4	世界地図	20
2.5	日本地図	23
2.6	地形図	25
2.7	ヒストグラム	27
2.8	ドットチャート	30
3	今後の課題	31

1 R コマンダーとは

R Commander(Rcmdr , R コマンダー)は ,カナダの McMaster University の John Fox 教授が開発・管理している拡張パッケージである。本章では、R コマンダーの仕組みを概説し、R コマンダーへのプラグインの形での教育用パッケージとして開発したプロトタイプの説明を行う。

1.1 R コマンダーのしくみ

R は、基本パッケージ (base package) と拡張パッケージ (contributed packages) から構成される。基本パッケージは、R 本体をダウンロードしてインストールした際、インストールされるものである。これに対して拡張パッケージは、基本パッケージが持たない機能を利用できるように機能を拡張するものであり、様々な人々が開発して公開している。CRAN (The Comprehensive R Archive Network; <http://cran.r-project.org/>) で公開されているものだけでも約 2000 ある。R コマンダーもこの拡張パッケージ (以下、パッケージという) の 1 つで、メニュー方式 (GUI : Graphical User Interface) で R の機能を利用できるようにするものである。

1.1.1 R コマンダーの起動

R をインストールすると、R のアイコンがデスクトップに作成される*1。これをダブルクリックすると、R を起動することができる。R を起動すると、R Console (R コンソール) が表示される (図 1)。

パッケージの起動は、R Console でプロンプト (>) の右に次の形で入力する*2。

```
>library(パッケージ名) Enter
```

R コマンダーを起動するには、R Console に

```
>library(Rcmdr)
```

*1 R および R コマンダー、他の拡張パッケージのインストール方法については、荒木のホームページ (<http://www.ec.kansai-u.ac.jp/user/arakit/stat-math.html>) 参照。

*2 R 本体の起動とともに、R コマンダーやそのプラグインを自動的に起動するように設定することもできる [2]。

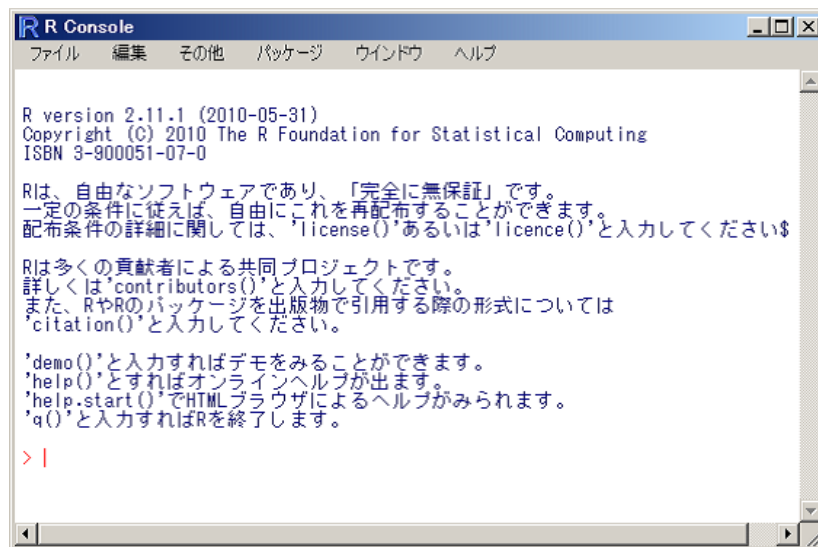


図 1 R Console

を入力する。すると、図 2 に示す R コマンダーのウィンドウが表示される*3。これは、上から [メニューバー] [ツールバー] [スクリプトウィンドウ] [出力ウィンドウ] [メッセージ] 欄から構成される。

R コマンダーを起動した直後では、[データセット] 欄の右には「<アクティブデータセットなし>」が、[モデル] 欄の右には「<アクティブモデルなし>」が表示されている。R コマンダーの諸機能は、基本的にこうしたアクティブデータセットやアクティブモデルに対して適用される。

[スクリプトウィンドウ] の役割は次のとおりである。R コマンダーでは、さまざまな機能がメニューやボタンに埋め込まれている。そのため、利用者は面倒なコマンド (スクリプト) を入力しなくても良く、R の初心者にとって利用時の負担が軽減される。このとき、メニューが実行しているコマンドを [スクリプトウィンドウ] に表示するよう設計されているため、実行したメニューに対応するスクリプトを確認することができる。すると、同じような分析・操作を繰り返す場合、ウィンドウに表示されているスクリプトを修正しながら利用すると、メニュー操作を行わなくて済むという利点を持つ。

1.1.2 スクリプトウィンドウの利用

例えば、C ドライブの work フォルダに「dat1.csv」、「dat2.csv」、「dat3.csv」という 3 つのカンマ区切り (CSV : comma separated value) ファイルがあ

*3 R 本体の起動とともに、R コマンダーやそのプラグインを自動的に起動するように設定することもできる [2]。

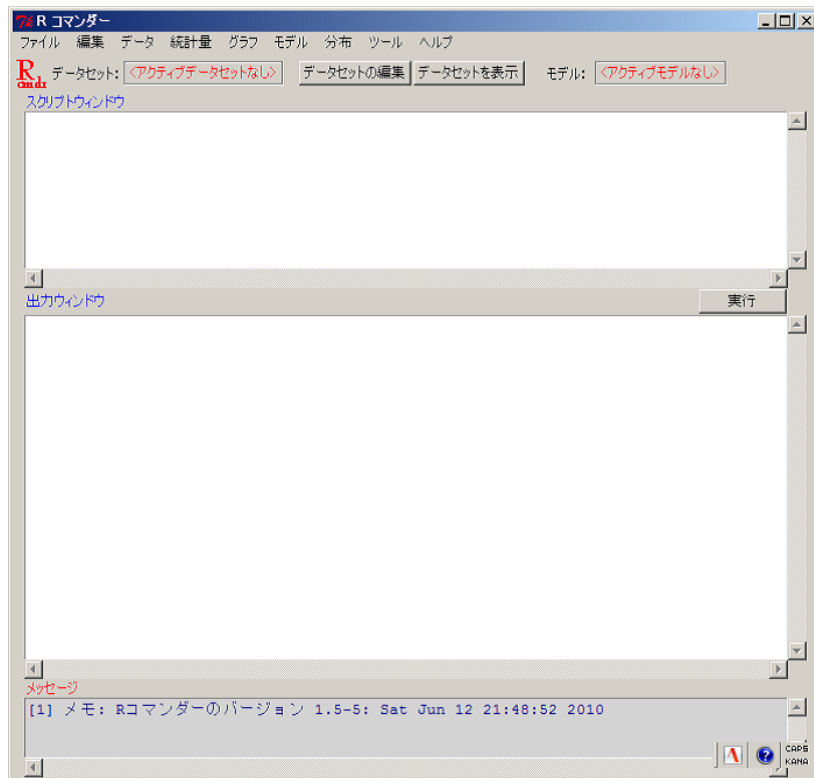


図 2 R コマンダーのウィンドウ

り、これらを連続して読み込みたいとする。このときまず、ファイル dat1.csv を R コマンダーの《データ》▶《データのインポート》▶《テキストファイルまたはクリップボードから...》を利用して読み込む。図 2 に示すダイアログボックスが表示されるので、[データセット名を入力] 欄に読み込んだデータにつけたい名前を入力する（デフォルトでは「Dataset」）。また、[フィールドの区切り記号] を「カンマ」に切り替えて 。

上記の作業が終了すると、[スクリプトウィンドウ] に次のスクリプトが表示されている。

```
dat1 <- read.table("C:/work/dat1.csv", header=TRUE, sep=",",
  na.strings="NA", dec=".", strip.white=TRUE)
```

これは、“関数 read.table() を利用して、フォルダ「C:/work/」にある CSV 形式のファイル「dat1.csv」を読み込み、それに「dat1」という名前をつけよ”という命令である。「<-」は、右辺で行った操作の結果に左辺の名前をつけて保存することを意味し、等号「=」を用いてもよい*4。「header = TRUE」は、ファイルの 1 行目に変数名（ヘッダー）があり、「sep=","」はデータの区切

*4 本稿では、自分でコマンドを入力する場合、「=」を用いる。

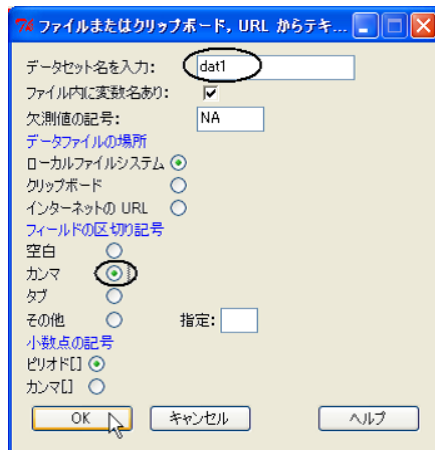


図 3 データの読み込み

り記号 (separator) がカンマ (,) 「na.strings = "NA"」は、欠測値は記号「NA」(Not Available : 利用不可) で入力されており、「dec = "."」は、小数点 (decimal point) の記号がピリオド (.) であること等を意味する。

次に、スクリプト中に 2 ヶ所ある「dat1」を「dat2」に変更して実行をクリックする*5。次に、2 ヶ所の「dat2」を「dat3」に変更して実行をクリックする。これらにより 3 つのファイルの読み込みが終了する。結果が次々と [出力ウィンドウ] に表示されている (図 4 参照)。

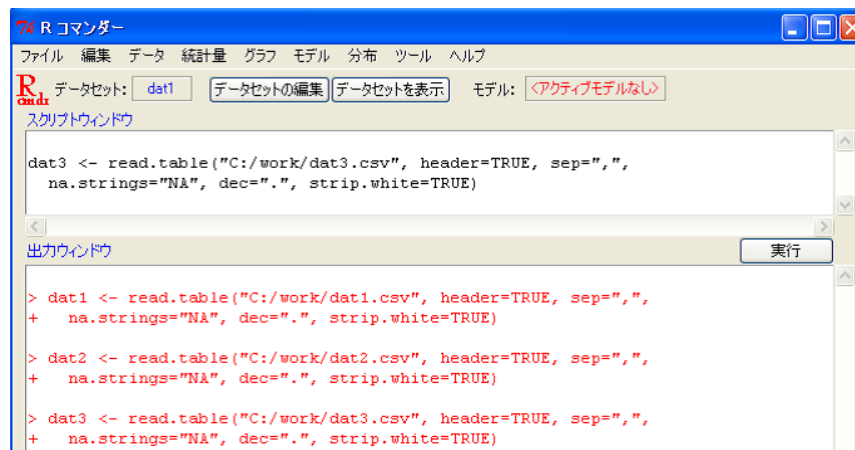


図 4 データの読み込み

*5 スクリプトの行にカーソルがある状態でクリックすること。スクリプトが複数行にわたる場合は、全体をマウスでドラッグして指定した後、実行をクリック。

1.1.3 データのハンドリング

(a) パッケージ内のデータセットのアクティブ化

Rにはたくさんのデータセットが用意されており、これらを用いて手法の利用の仕方を学ぶことができる。パッケージに用意されているデータを R コマンドーに読み込み、アクティブにしたい場合、次のようにする。

R コマンドーの《データ》 《パッケージ内のデータ》 《アタッチされたパッケージからデータセットを読み込む...》を選択する。ダイアログボックスの [データセットを入力] 欄にデータセット名を入力し、。例えば、パッケージ lattice にあるデータセット singer を読み込むには、のように、[データセット名を入力] 欄に「singer」と入力し、。なお、[パッケージ] 欄に表示されているパッケージが、現在読み込まれているものである。読み込むことができた場合、アクティブデータセット欄に「singer」が表示される ()。この欄に表示されているデータセットをアクティブデータセットという。R コマンドーの分析メニューは、このデータセットに対して適用される。

同様に、パッケージ datasets にあるデータセット volcano を読み込んでおく。なお、volcano を読み込むと、[メッセージ] 欄に「警告: volcano has been coerced to a data frame」が表示されるが、これは、volcano が matrix という形式 (クラスという) のデータセットであることに対して、R コマンドーではデータフレーム (data frame) という形式しかアクティブデータセットとして扱えないため、この形式に変換したというメッセージである。

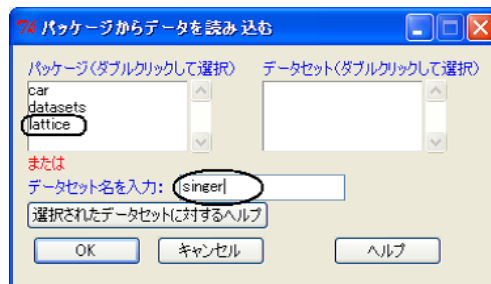


図 5 パッケージからデータを読み込む



図 6 [データセット] 欄

1.2 データの切り出し

あやめデータ (データセット名 : iris) を用いて、データセットからデータの一部を切り出す方法を簡単に見る。まず、iris データセットを関数 data() を用いて利用可能にする。

```
R Console  
> data(iris)
```

データセット iris の構造 (structure) を見るには、次のようにする。

```
R Console
> names(iris)      # どんな変数を含むか
[1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width" "Species"
> str(iris)        # 構造はどうなっているか
'data.frame':   150 obs. of  5 variables:
 $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species     : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 ...
```

関数 `names()` はオブジェクト(ここではデータセット)が持つ名前 (name) を表示する関数であり、関数 `str()` はオブジェクトの構造 (structure) を表示する関数である*⁶。関数 `str()` の出力の 1 行目は、このオブジェクト (iris) は data.frame (データフレーム) であり、データ数は 150 (obs.: observations)、変数 (variables) は 5 つあることを意味する。\$ Sepal.Length の行は、これは数値変数 (numeric) であり、データとして 5.1, 4.9, ... となっていることを意味する。\$ Species の行は、これは因子 (Factor) で、水準 (levels) として 3 水準 ("setosa", "versicolor", ... 等) があることを示す*⁷。

データセットの最初の数行を表示するには、関数 `head()` を利用するとよい。

```
R Console
> head(iris)      # 標準では最初の 6 行を表示する
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1           5.1           3.5           1.4           0.2 setosa
2           4.9           3.0           1.4           0.2 setosa
3           4.7           3.2           1.3           0.2 setosa
4           4.6           3.1           1.5           0.2 setosa
5           5.0           3.6           1.4           0.2 setosa
6           5.4           3.9           1.7           0.4 setosa
```

データセット iris の 1 列目 (Sepal.Length) のデータのみを切り出すには、
データセット名\$変数名
の形で指定する。例えば、次のようにする。

*⁶ R では、ヘルプ情報が充実している。関数やデータセットに関する情報を知りたい場合、`help("関数名")` または `help("データセット名")` を R Console に入力して参照するとよい。

```
R Console
> help(names)
> ?names      # ?でもよい
> help(iris)
```

*⁷ 大きく分けて、変数には、数値変数と因子がある。csv ファイルからデータを読み込む場合、記号で入力されたデータの変数は、Factor として取り扱われる。

R Console

```
> iris$Sepal.Length
 [1] 5.1 4.9 4.7 4.6 5.0 5.4 4.6 5.0 4.4 4.9 5.4 4.8 4.8 4.3 5.8
 [16] 5.7 5.4 5.1 5.7 5.1 5.4 5.1 4.6 5.1 4.8 5.0 5.0 5.2 5.2 4.7
 . . . . .
```

R Console に表示されている記号 [1] は、すぐ右にあるデータ (5.1) が 1 番目のデータであり、[16] は、すぐ右にあるデータ (5.7) が 16 番目のデータであることを示す。なお、関数 `attach()` を利用すると、データセット名を書かずに、変数名単独でデータを利用できる。

R Console

```
> Sepal.Length # 直接変数を利用できない。
エラー: オブジェクト "Sepal.Length" は存在しません
> attach(iris) # そのためには、あらかじめ attach() しておく
> Sepal.Length
 [1] 5.1 4.9 4.7 4.6 5.0 5.4 4.6 5.0 4.4 4.9 5.4 4.8 4.8 4.3 5.8
 [16] 5.7 5.4 5.1 5.7 5.1 5.4 5.1 4.6 5.1 4.8 5.0 5.0 5.2 5.2 4.7
 . . . . .
```

別の方法として、

データセット名 [行インデックス, 列インデックス]

という形でデータセットの一部を切り出すこともできる。例えば `iris$Sepal.Length` と `iris[,1]` は同じである*⁸。データセットのこうした切り出しのイメージは、図 7 のようになる。なお、第 1, 2, 3 列といった複数列を取り出したい場合、これらに関数 `c()` で結合して「`c(1,2,3)`」のように指定する (連番の場合は、「`1:3`」でよい)。また、マイナスのインデックスを指定すると、その行または列以外を抽出することができる。

R Console

```
> iris[,c(1,2,3)] # iris[,1:3] でもよい
  Sepal.Length Sepal.Width Petal.Length
1           5.1           3.5           1.4
2           4.9           3.0           1.4
3           4.7           3.2           1.3
4           4.6           3.1           1.5
. . . . .
> iris[,-c(1,2,3)] # 1,2,3 列以外の列
  Petal.Width Species
1           0.2   setosa
2           0.2   setosa
3           0.2   setosa
4           0.2   setosa
. . . . .
```

*⁸ 行インデックスを指定しない場合は、全ての行が利用され、列インデックスを指定しない場合は、全ての列が利用される。

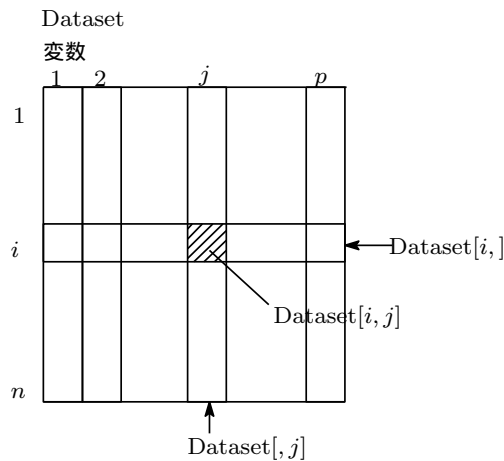


図 7 データの切り出し

メモ

データセットの一部を切り出すためのこうした表記法は、R が作成するグラフの中でも自動的に利用されることが多いので、知っておくと便利である。

データセットからデータを切り出し、新たなデータセットを作り出す場合も上記の方法でできるが、R コマンドの機能を利用して可能である。例えば、Species が "versicolor" と "virginica" の 2 群のみからなるデータセットを作るには、次のようにする。

《データ》▶《アクティブデータセット》▶《アクティブデータセットの部分集合を抽出》を選択する。表示されたダイアログボックスで、必要な変数を指定し（今は全部）、[部分集合の表現]で「!(Species=="setosa")」*⁹、[新しいデータセットの名前]に「iris2g」と指定し（図 8）、。作成されたデータセット iris2g がアクティブになり、データセット欄に表示される（図 9）。

1.3 プラグイン

R コマンドは、利用者が独自に機能・メニューを追加できるように設計されている。例えば、フレームを用意して、その中にラベルやチェックボックス、ラジオボタン、リストボックス、コマンドボタンなどを配置したダイアログボックスを作成することができる。

しかし、この方法には欠点がある。それは、R コマンドのメニューを変更

*⁹ 「!」は否定を意味する。また、「==」であることに注意。「iris\$Species=="versicolor" | iris\$Species=="virginica"」でもよいが、長くなるので面倒。「|」は、「または」を意味する。

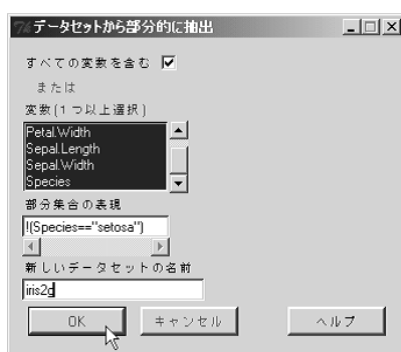


図 8 データセットの部分集合を抽出

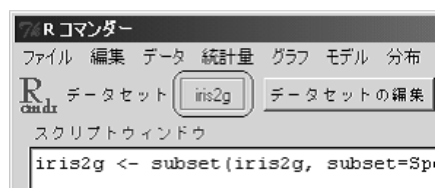


図 9 アクティブデータセット

したり追加機能を含む R の関数ファイルを配置したりという作業を、利用者が自ら行う必要があるからである。この欠点を補うために、様々な機能を自動的に設定するスクリプトを追加機能の作成者が用意することが考えられる。例えば、多変量解析のパッケージである FactoMineR では、当初そうしたスクリプトを提供していた。しかし、これでも、R コマンダーがバージョンアップされるたびにそのスクリプトを利用しなければならない。

R コマンダーのバージョン 1.3-0 以降 (2010 年 6 月現在は、バージョン 1.5-5) プラグイン (Plug-In) という機能が新たに追加された [11]。これは R コマンダーへの機能の追加をパッケージの形で実現する狙いを持つ。プラグインは、通常のパッケージと同じように、R コマンダーとは独立した形でインストールすることができ、利用時に R コマンダーの中で起動させればよいという利点を持つ。R と R コマンダー、プラグインの関係は図 10 のようになる。荒木 [4] は、こうした機能を利用し、統計的品質管理 (SQC) の標準的な手法に関して GUI での分析を可能とする活動の結果をまとめたものである。また、その成果が RcmdrPlugin.QCtools である*10。

現在公開されているプラグインは、19 個ある*11。代表的なものを次に示す。

- RcmdrPlugin.FactoMineR 多変量解析用パッケージ FactoMineR に対するグラフィカル・インターフェース
- RcmdrPlugin.TeachingDemo 教育デモ用パッケージ TeachingDemo に対するグラフィカル・インターフェース
- RcmdrPlugin.HH Heiberger and Holland[12] のコンパニオンパッケージ HH に対するグラフィカル・インターフェース

*10 荒木のホームページ(<http://www.ec.kansai-u.ac.jp/user/arakit/>)よりダウンロード可。

*11 2010 年 6 月 20 日現在。以下、同様。

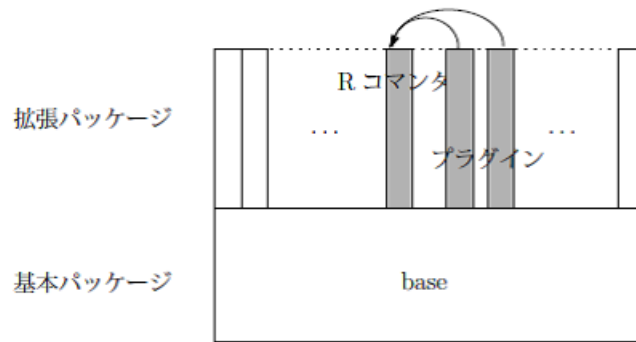


図 10 プラグインの仕組み

- RcmdrPlugin.IPSUR Kerns[13] のコンパニオンパッケージ

RcmdrPlugin.TeachingDemo は、プラグインの開発法を例示するために、別に開発されていたパッケージ TeachingDemo[19] を利用して R コマンダーの開発者である Fox 教授により作成された。そのため現在のところ、TeachingDemo が持つ機能の一部を利用しているのみである。これらをさらに取り入れたり、目的に応じて必要な機能を追加したりすることにより、教育や学習利用のデモンストレーションやシミュレーション用パッケージを作成することができる。

2 教育用プラグイン RcmdrPlugin.DAToolsforKids

R コマンダーのとしての教育用プラグインとしては、すでに述べたように RcmdrPlugin.TeachingDemo がある。これは主に、中心極限定理や大数の法則、信頼区間、検出力、回帰分析、相関等の確率論・統計学における重要な概念を視覚的に学ぶためのシミュレーション用のパッケージであり、データ分析用ではない (図 11)。

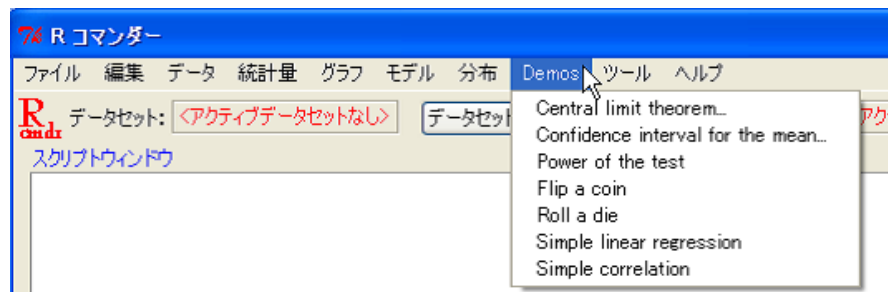


図 11 RcmdrPlugin.TeachingDemo のメニュー

R コマンダーではすでに、図化の手法として、《グラフ》メニューの中に、インデックスプロット、ヒストグラム、幹葉表示、散布図、散布図行列、条件つき散布図、折れ線グラフ、ドットチャート（点図）、棒グラフ、円グラフ、3次元グラフ等の作成が可能である（図 12）。しかし、これらの手法の実装のレベルは、データ分析に十分なものもあれば不十分なものもある。特に、層別（分けて描くこと）ができないと致命的である（層別が可能なのは、散布図、散布図行列、箱ひげ図）。

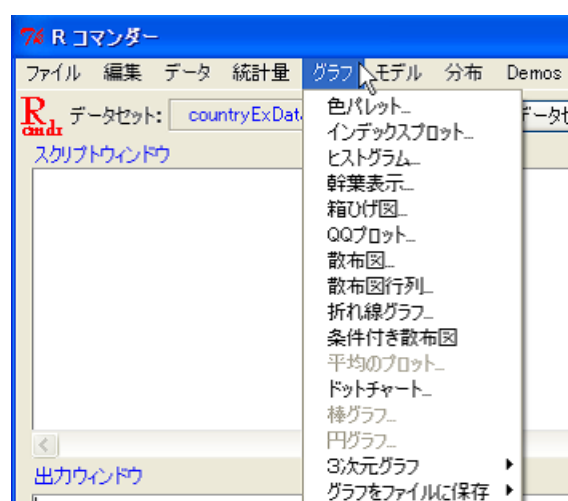


図 12 R コマンダーのグラフメニュー

こうした問題を念頭において、品質管理での利用を目指して開発したのが、RcmdrPlugin.QCtools である。これを利用すると、層別のヒストグラムや管理図、レーダーチャート、パレート図等の作成が可能となる。しかし、これらは品質管理における「QC 七つ道具」と総称されるツール群であり、小学校・中学校・高等学校での利用を考えると、不要な機能・不足する機能がある。そのため、利用できるものは利用し、不足する機能は R コマンダーのプラグインとして実現する必要がある。例えば、散布図、散布図行列、条件つき散布図、箱ひげ図は、R コマンダーの機能を、折れ線グラフ、棒グラフ、ヒストグラムは、RcmdrPlugin.QCtools のものを利用することができる。

統計的手法は、現実のデータで実践するとともに役に立つことを体験させる必要がある。小学校・中学校・高等学校の教育においては、他の教科で利用できることを示すことが有益である。理科や社会科がその有力候補となる。すると、他に必要な機能としては、地理関連（世界地図、日本地図、等高線図、透視図）、人口ピラミッド、ドットチャート（点グラフ、線グラフ、棒グラフ）が考えられる。これらに、RcmdrPlugin.QCtools から折れ線グラ

フ、棒グラフ、ヒストグラムを移植して、1つのパッケージにまとめると便利である。また、開発のベースとして、高品位の図を作成することができるパッケージ `lattice` を利用した [17, 18]。それが R コマンドーのプラグイン `RcmdrPlugin.DAToolsforKids` (仮称) である。

作成したプラグインのメニューを図 13 に示す。以下、この機能の主なものの利用法を説明する。

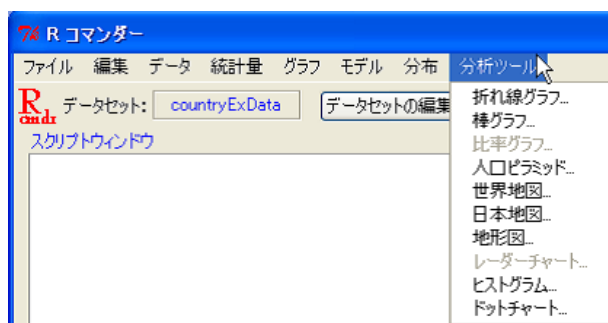


図 13 RcmdrPlugin.DAToolsforKids のメニュー

2.1 折れ線グラフ

2.1.1 データ形式

図 14 は、岡山県の小学校・中学校・高等学校の 1990 年度から 2009 年度までの児童・生徒数のデータである*¹²。表計算ソフトウェアを用いてこのような形でデータファイルを作成しておく。

	A	B	C	D
1	年度	小学校	中学校	高等学校
2	1990年度	146508	85684	95577
3	1991年度	143773	82097	93487
4	1992年度	141487	79124	89928
5	1993年度	139317	76343	85937
6	1994年度	137151	73743	83117
7	1995年度	133824	72626	79951
8	1996年度	129278	72293	76869
9	1997年度	124989	72215	73494
10	1998年度	121361	70925	71986
11	1999年度	118268	68579	71527
12	2000年度	115712	66296	70897
13	2001年度	114393	63888	69285
14	2002年度	113246	61654	66589
15	2003年度	113254	59520	64232
16	2004年度	113437	58074	61888
17	2005年度	113763	57229	59841
18	2006年度	113676	56526	57783
19	2007年度	113207	56747	56113
20	2008年度	113373	56400	55315
21	2009年度	112290	56989	54655

図 14 折れ線グラフのデータの形式

*¹² 岡山県のホームページ (<http://www.pref.okayama.jp>) より取得。以下、岡山県のデータに関しては同様。

R コマンドーより《データ》▶《データのインポート》▶《テキストファイルまたはクリップボードから ...》を選択する (図 15)

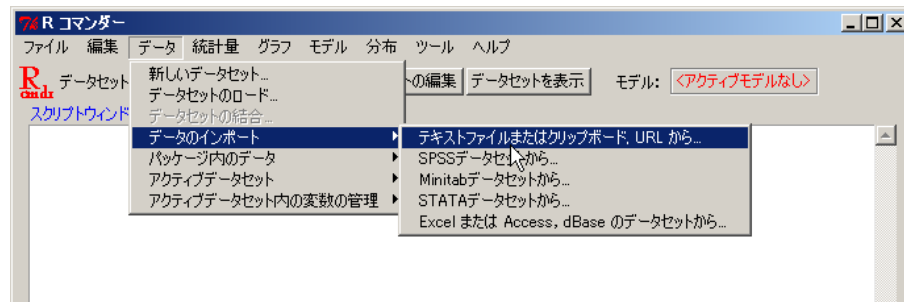


図 15 データの読み込み

データのインポート (読み込み) のダイアログボックス (図 16) で、[フィールドの区切り記号] を「カンマ」に切り替えて、*¹³。次に、データファイルを指定して。読み込みが終了すると、[データセット] 欄に読み込んだデータセット名 (今の場合、「Dataset」) が表示される (図 17)。また、読み込みに利用したスクリプトが [スクリプトウィンドウ] に表示される。

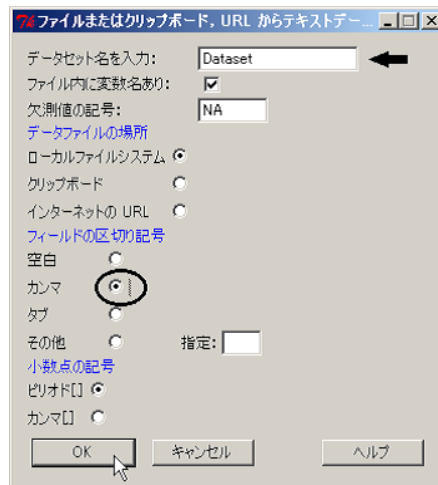


図 16 データのインポートのダイアログボックス

*¹³ なお、このとき、読み込んだデータセット名はデフォルトでは「Dataset」となる。これを
変更したい場合は、[データセットを入力] 欄に入力する。

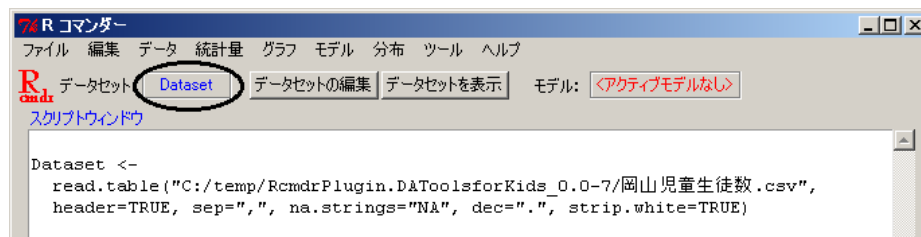


図 17 アクティブデータセット

なお、読み込んだ後、**データセットを表示**をクリックして、正しく読めているかどうかを確認すること。確認できれば、**x** ボタンでウィンドウを閉じる。

	年度	小学校	中学校	高等学校
1	1990年度	146508	85684	95577
2	1991年度	143773	82097	93487
3	1992年度	141487	79124	89928
4	1993年度	139317	76343	85937
5	1994年度	137151	73743	83117
6	1995年度	133824	72626	79957
7	1996年度	129278	72293	76868
8	1997年度	124989	72215	73494
9	1998年度	121361	70925	71984
10	1999年度	118268	68579	71527
11	2000年度	115712	66296	70897
12	2001年度	114393	63888	69284
13	2002年度	113246	61654	66588
14	2003年度	113254	59520	64232
15	2004年度	113437	58074	61887
16	2005年度	113763	57229	59841
17	2006年度	113676	56526	57783
18	2007年度	113207	56747	56113
19	2008年度	113373	56400	55314
20	2009年度	112290	56989	54654

図 18 データのインポートの確認

2.1.2 折れ線グラフのダイアログボックス

《分析ツール》▶《折れ線グラフ...》を選択すると、図 19 のダイアログボックスが表示される。折れ線に描きたい [数値変数] を選択し (複数の選択可) 「x 軸ラベルの変数」(横軸のラベルに利用) を必要なら指定し、**OK**。図 20 に示す折れ線グラフが表示される。

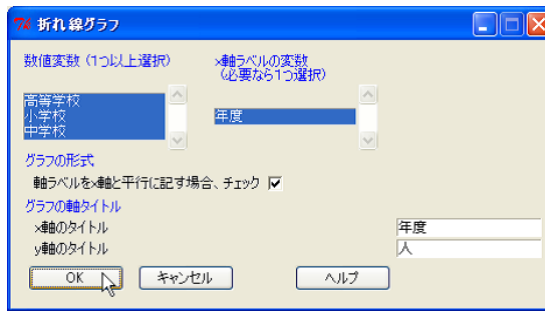


図 19 折れ線グラフのダイアログボックス

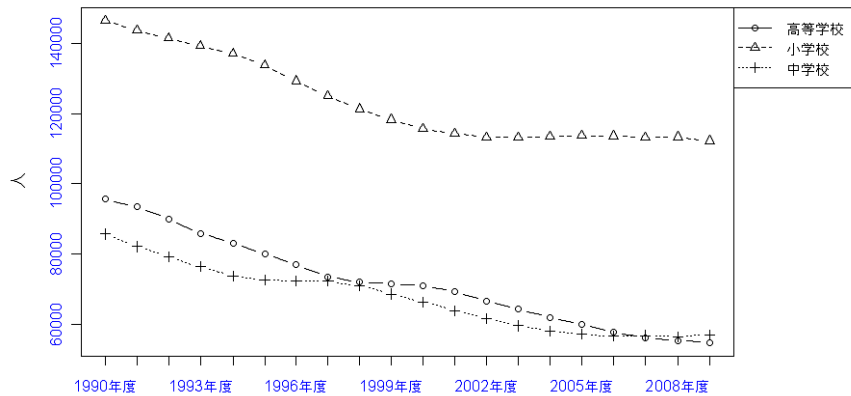


図 20 岡山県の児童・生徒数の推移

2.2 棒グラフ

2.2.1 データ形式

図 21 は、岡山県の月別（1～12月）および場所別（岡山・津山・千屋）の降水量データである。表計算ソフトウェアを用いてこのような形でデータファイルを作成し、R コマンドーに読み込む。

	A	B	C	D
1	月	岡山	津山	千屋
2	1月	38	53	121.5
3	2月	109	116.5	120
4	3月	58.5	79.5	124
5	4月	76.5	115.5	155
6	5月	30	63	55
7	6月	73.5	203	239
8	7月	259	326	407
9	8月	159	204	56.5
10	9月	56.5	83	63.5
11	10月	57	50	119
12	11月	127	116	212.5
13	12月	30	50.5	94

図 21 岡山県の月別・場所別降水量データ

2.2.2 ダイアログボックス

《分析ツール》▶《棒グラフ...》を選択すると、棒グラフのダイアログボックス(図 22)が表示される。棒グラフに描きたい[数値変数]を1つ以上選択し、「項目名の変数」(横軸のラベルに利用)を必要なら選択し、**OK**。図 23 に示す棒グラフが表示される。棒グラフの形式を変更することにより、積み重ね棒グラフ(24)を描くことができる。

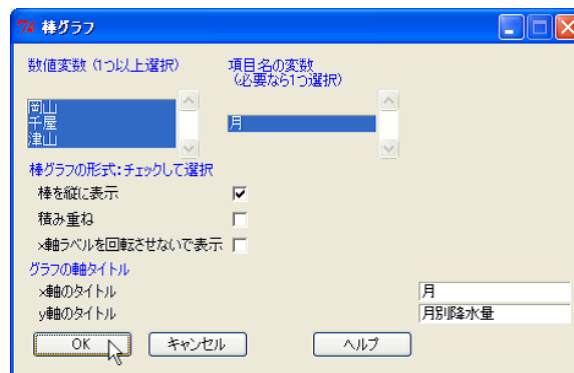


図 22 棒グラフのダイアログボックス

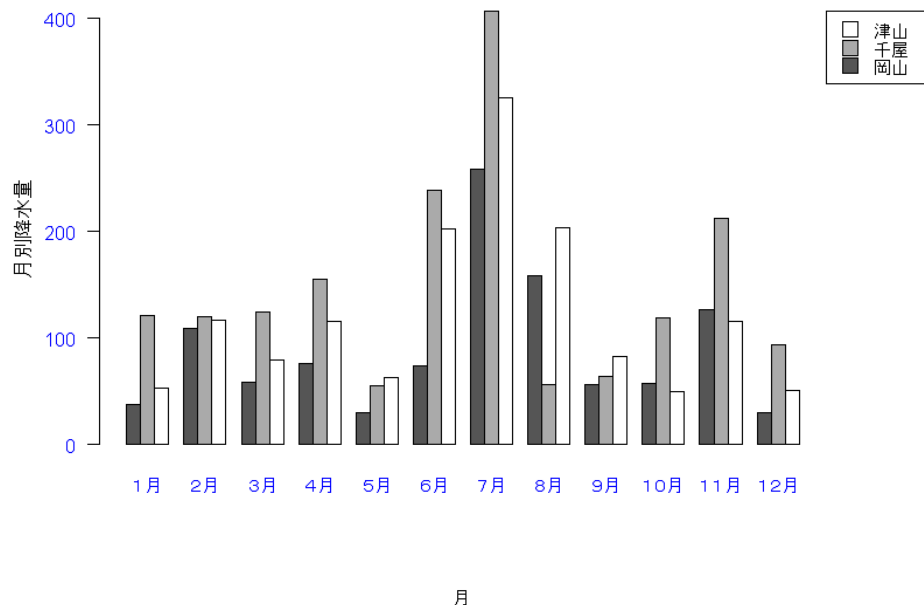


図 23 岡山県の月別・場所別降水量の棒グラフ

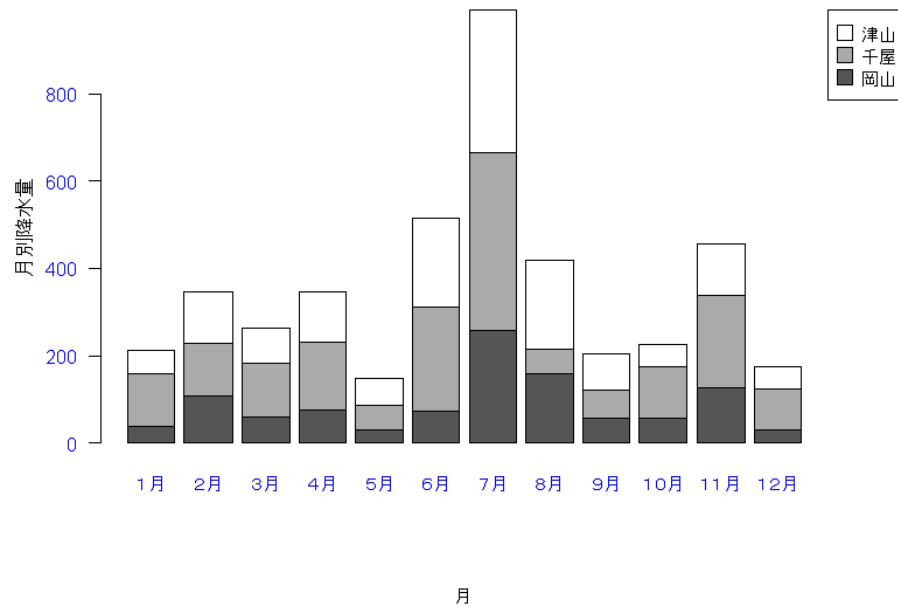


図 24 岡山県の月別・場所別降水量の積み重ね棒グラフ

2.3 人口ピラミッド

人口ピラミッドは、社会科でよく使われる図である。一般にも、少子高齢化の問題を議論するときによく用いられている。R ではパッケージ `pyramid`[15] により作成可能である。本メニューは、これが持つ機能の一部をプラグインに組み込んだものである。

2.3.1 データ形式

大正 9 年と平成 17 年の兵庫県の年齢別人口データを利用して、説明する^{*14}。図 25 に示すように、表計算ソフトで年齢の列、男性の人口の列、女性の人口の列を入力し、CSV 形式で保存する。保存したファイルを、R コマンドの《データ》▶《データのインポート》▶《テキストファイルまたはクリップボードから...》を選択して読み込む。《データセットを表示》をクリックすると、図 26 が表示される。

*14 データは兵庫県のホームページ (http://web.pref.hyogo.jp/ac08/ac08.1_000000220.html) より取得 (兵庫県のデータに関しては、以下同様)。なお、100 歳は 100 歳以上の人数である。また、不明者 (20,677 名) は含まず、総数 5,590,601 より不明者を除いた人数である。

2.3.2 ダイアログボックス

《分析ツール》▶《人口ピラミッド...》を選択すると、図 25 のダイアログボックスが表示される。左に描く変数(男:大正9年)、右に描く変数(女:大正9年)、年齢の変数(年齢)を指定し、図のタイトルを入力して(兵庫県の人口ピラミッド¥n 大正9年*15)、**OK**。図 26 の人口ピラミッドが表示される。

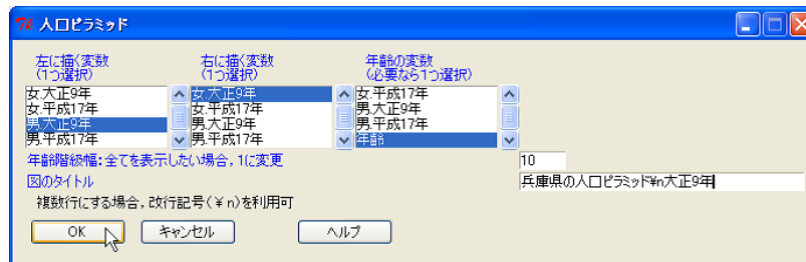


図 25 人口ピラミッドのダイアログボックス

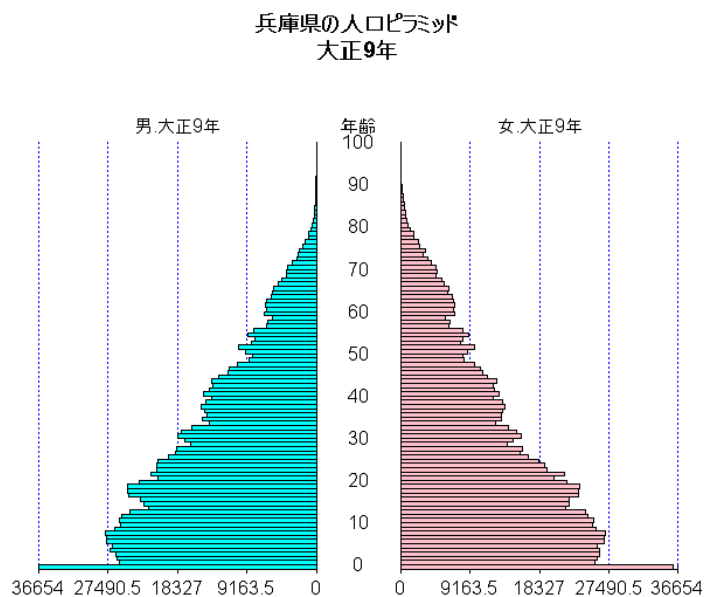


図 26 兵庫県の大正9年の人口ピラミッド

同様に、平成17年の人口ピラミッドを作成すると、図 27 のようになる。

*15 「¥n」をタイトルに入力すると、¥n 以降が改行されて表示される。

兵庫県的人口ピラミッド
平成17年

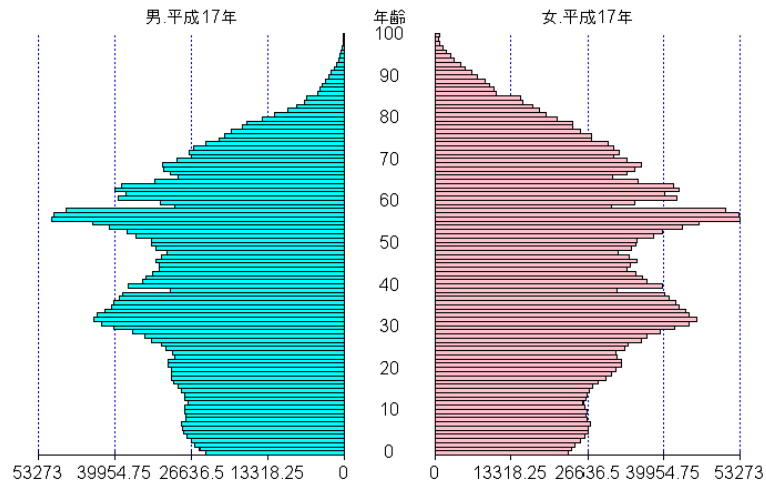


図 27 兵庫県平成 17 年の人口ピラミッド

なお、図 28 に示すように年齢が階級別になっている場合でも、人口ピラミッドを作成することができる（図 29。データは岡山県の平成 17 年の男女別年齢階級別人口）。

	A	B	C
1	年齢	男	女
2	0～4歳	45123	42814
3	5～9歳	48080	45843
4	10～14歳	48208	45675
5	15～19歳	52668	51902
6	20～24歳	54173	55529
7	25～29歳	58884	59063
8	30～34歳	69840	70442
9	35～39歳	57717	59784
10	40～44歳	54583	56208
11	45～49歳	56544	58228
12	50～54歳	64246	64603
13	55～59歳	78031	80631
14	60～64歳	64590	68652
15	65～69歳	53247	58639
16	70～74歳	50786	60967
17	75～79歳	40032	53477
18	80～84歳	22364	39950
19	85歳～	15511	42081

図 28 階級データ

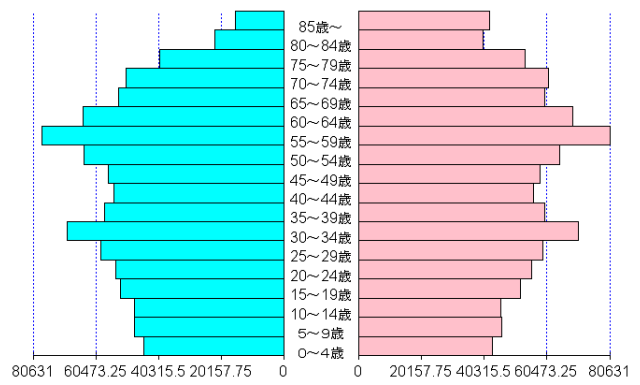


図 29 年齢階級別人口ピラミッド

2.4 世界地図

パッケージ rworldmap は、国と国レベルのユーザーデータを関連づけて地図を作成するためのパッケージである [20, 21]。本メニューは、rworldmap の一部の機能を GUI で利用可能にしたものである。

2.4.1 データ形式

パッケージ rworldmap にあるデータセット countryExData を利用して説明する。これは、149 カ国の 2008 年の環境評価指数 (EPI: Environment Performance Index) 関連データで*16、Population2005 (2005 年の人口)、BIODIVERSITY (生物多様性)、CO2IND_pt (industrial carbon intensity)、WATQI_pt (水質)、AIR_E (大気汚染) 等のデータが含まれる。データの一部を表示すると、図 30 のようになる。

	Country	EPI_regions	GEO_subregion	Population2005	GDP_cap
1	Angola	Sub-Saharan Africa	Southern Africa	15941.4	
2	Albania	Central and Eastern Europ	Central Europe	3129.7	
3		Middle East and North Africa	Arabian Peninsula	4495.8	
4	Argentina	Latin America and Caribbe	South America	38747.2	
5	Armenia	Middle East and North Africa	Eastern Europe	3016.3	
6	Australia	East Asia and the Pacific	Australia + New Zealand	20155.1	
7	Austria	Europe	Western Europe	8189.4	
8	Azerbaijan	Central and Eastern Europ	Eastern Europe	8410.8	
9	Burundi	Sub-Saharan Africa	Eastern Africa	7547.5	
10	Belgium	Europe	Western Europe	10419.1	
11	Benin	Sub-Saharan Africa	Western Africa	8438.9	
12	Burkina Faso	Sub-Saharan Africa	Western Africa	13227.8	
13	Bangladesh	South Asia	South Asia	141822.3	
14	Bulgaria	Central and Eastern Europ	Central Europe	7726.0	
15	na	Central and Eastern Europ	Central Europe	3907.1	
16	Belarus	Central and Eastern Europ	Eastern Europe	9755.1	

図 30 データセット countryExData

2.4.2 ダイアログボックス

データセットを読み込んだ後、《分析ツール》▶《世界地図...》を選択すると、図 31 に示すダイアログボックスが表示される。地図に描く変数 (BIODIVERSITY) を選択し、地域 (世界) を選択して [OK]。作成された生物多様性の世界地図を図 32 に示す。多様性の大きさは、地図の下の凡例に示されているように、色の種類およびその濃淡により表現される。

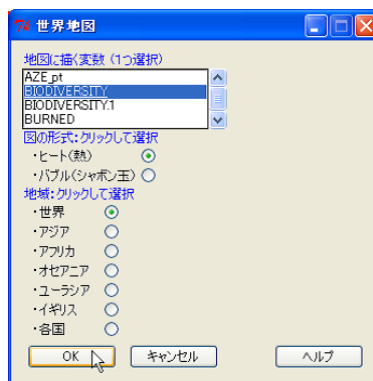


図 31 世界地図のダイアログボックス

*16 エール大学とコロンビア大学の調査による。元のデータは、<http://epi.yale.edu/> または <http://sedac.ciesin.columbia.edu/es/epi/> よりダウンロード可能。

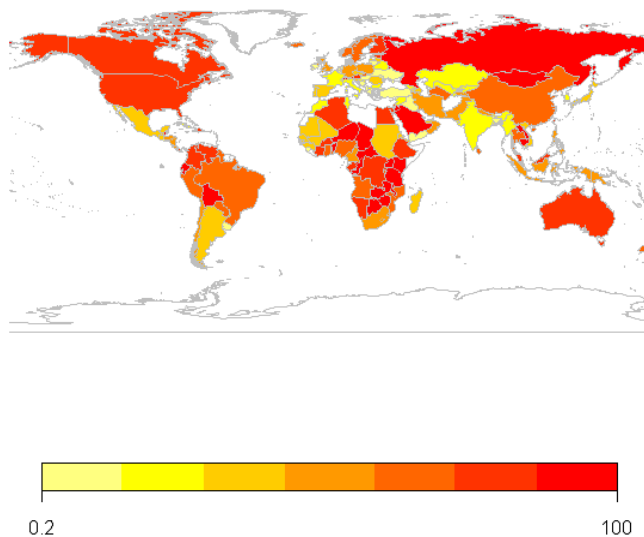


図 32 世界の生物多様性の地図

地域を「アジア」に、図の形式を「ヒート(熱)」に変更すると、図 33、「バブル(シャボン玉)」を選択すると図 34 のようになる。なお、現在のところ、各国は選択できない。

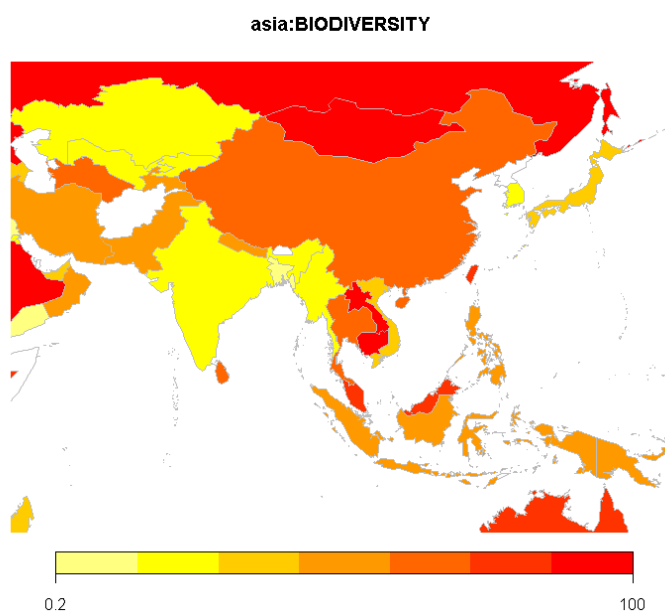


図 33 アジアの生物多様性の地図 - ヒート

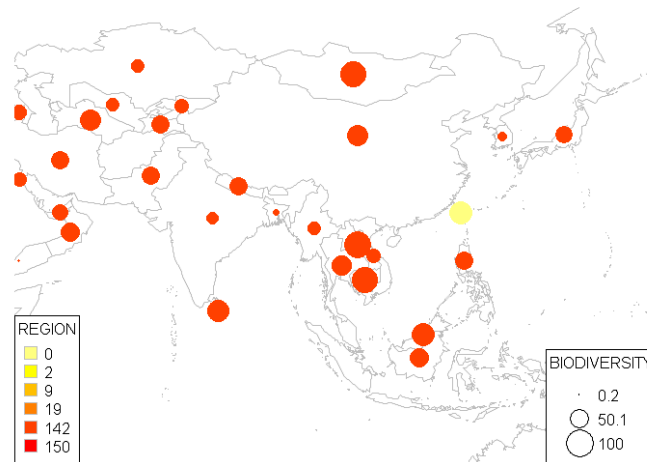


図 34 アジアの生物多様性の地図 - バブル

2.5 日本地図

このメニューにより、都道府県の市町村で区分した地図を作成することがきる。また、市町村別データを地図上にヒートマップとして表示することができる [8]。日本地図を作成するには、ESRI ジャパン (株) が無償で配布している全国市区町村界データを利用する。このデータ (japan_ver62.zip) は、ESRI ジャパン (株) のホームページ (http://www.esrij.com/products/gis_data/japanshp/japanshp.html) よりダウンロード可能である*17。本データに関する「注意」および「著作権と使用上の注意」に関しては、このサイトで確認のこと。

2.5.1 データ形式

地図に表示したいデータを図 35 のような形で用意する。JCODE の列 (列名を JCODE とすること) を含む必要があることに注意。なお、データは兵庫県のものである。

*17 ファイル japan_ver62.zip をダウンロードし、解凍する。解凍したファイル全てを c ドライブにフォルダ GISdata にコピーしておくこと。

	A	B	C	D	E	F	G
1	JCODE	CITY1	PREF	面積	世帯数	人口	老年人口比率
2	28101	東灘区	兵庫県	30.36	92289	208299	17.1
3	28102	灘区	兵庫県	32.4	62972	129598	21
4	28105	兵庫区	兵庫県	14.54	54538	108006	25.9
5	28106	長田区	兵庫県	11.48	47590	102084	26.6
6	28107	須磨区	兵庫県	28.91	70418	168478	20.8
7	28108	垂水区	兵庫県	28.02	93525	220201	21.5
8	28109	北区	兵庫県	240.31	86085	226428	19

図 35 日本地図用のデータ - 兵庫県

2.5.2 ダイアログボックス

データを読み込み、《分析ツール》▶《日本地図...》より図 36 に示すダイアログボックスを開く。地図に描く都道府県名を入力し、ヒートマップに用いる変数を指定して **OK** をクリックすると、地図が表示される。市町村名の表示・非表示の選択、軸の表示・非表示、凡例の位置の選択が可能である。例えば、兵庫県の老年人口比率の地図は、図 37 のようになる。凡例に示されているように、地図に描くデータを階級に分け、それらを種類・濃淡を用いて色に対応づけて描かれている。

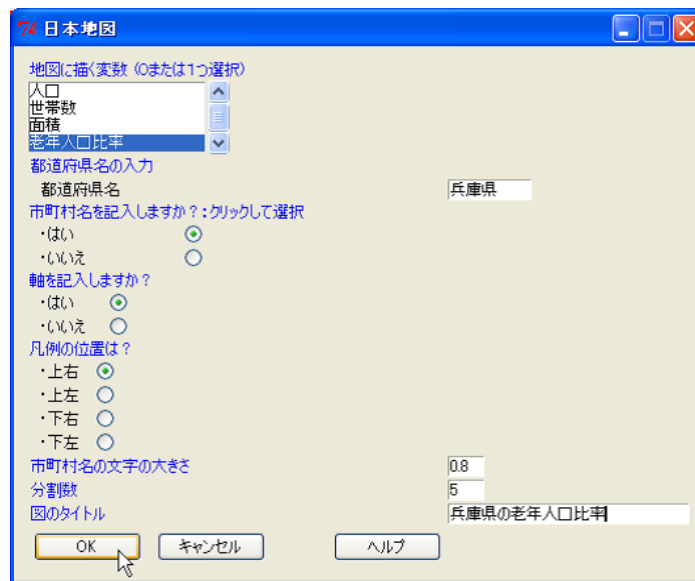


図 36 日本地図のダイアログボックス

兵庫県の老年人口比率

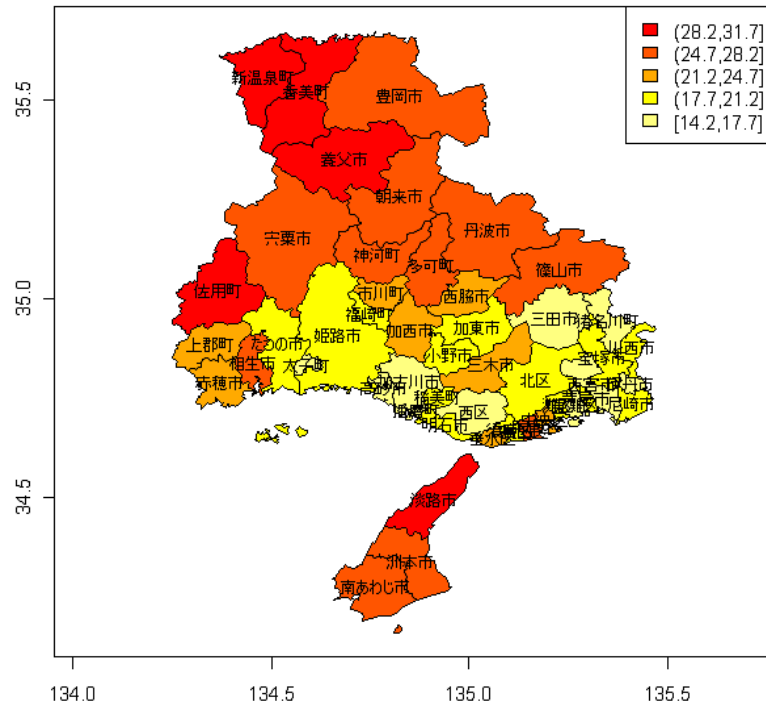


図 37 兵庫県の老年人口比率の地図

2.6 地形図

2.6.1 データ形式

パッケージ datasets にあるデータ volcano を利用して説明する。これはニュージーランドにある Maunga Whau (Mt. Eden) の地形データである。10m × 10m のグリッドで計測されたデータで、87 行 61 列の行列形式となっている。これを R コマンドに読み込んでアクティブにするには、R コマンドの《データ》▶《パッケージ内のデータ》▶《アタッチされたパッケージからデータセットを読み込む...》を選択し(図 39) [データセットを入力] 欄に「volcano」を入力して OK (図 39) . [データセットを表示] をクリックすると、図 40 が表示される。

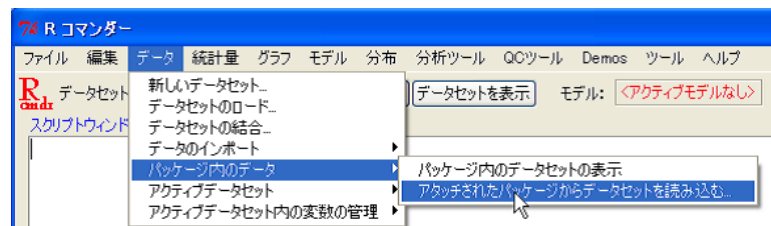


図 38 アタッチされたパッケージからのデータセットの読み込み

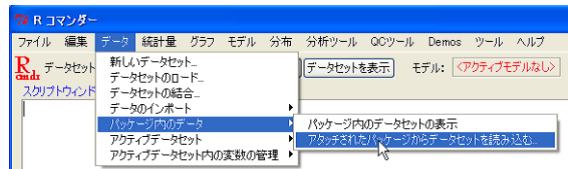


図 39 パッケージからデータを読み込む

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12
1	100	100	101	101	101	101	101	100	100	100	101	101
2	101	101	102	102	102	102	102	101	101	101	102	102
3	102	102	103	103	103	103	103	102	102	102	103	103
4	103	103	104	104	104	104	104	103	103	103	104	104
5	104	104	105	105	105	105	105	104	104	104	105	105
6	105	105	105	106	106	106	106	105	105	105	106	106
7	105	106	106	107	107	107	107	106	106	106	107	107
8	106	107	107	108	108	108	108	107	107	107	108	108
9	107	108	108	109	109	109	109	108	108	108	109	109
10	108	109	109	110	110	110	110	109	109	109	110	110
11	109	110	110	111	111	111	111	110	110	110	111	111
12	110	110	111	113	112	111	113	112	112	114	116	119
13	110	111	113	115	114	113	114	114	115	117	119	121

図 40 データセット volcano の表示

2.6.2 ダイアログボックス

《分析ツール》▶《地形図...》より、図 41 のダイアログボックスが表示される。そのまま **OK** をクリックすると、図 42 に示すレベルプロットが表示される。また、等高線図、ワイヤフレーム図を作成すると、それぞれ図 43,44 のようになる。

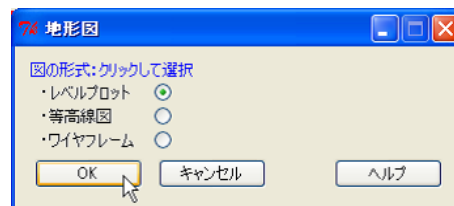


図 41 地形図のダイアログボックス

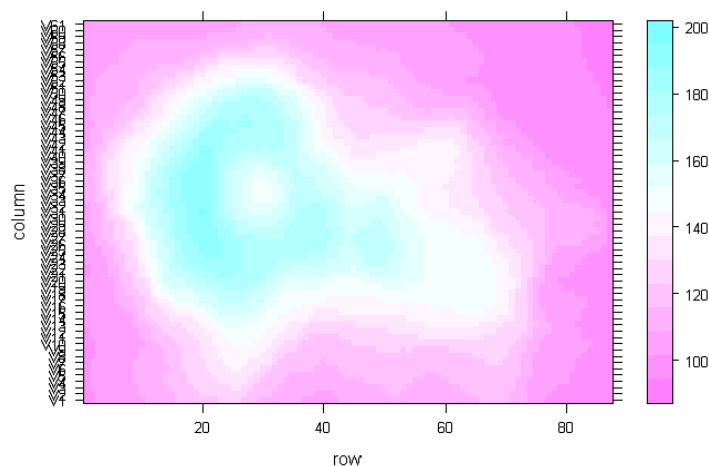


図 42 データセット volcano のレベルプロット

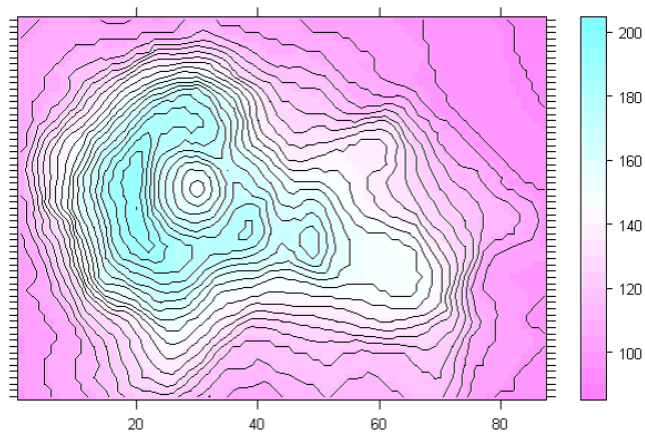


図 43 データセット volcano の等高線図

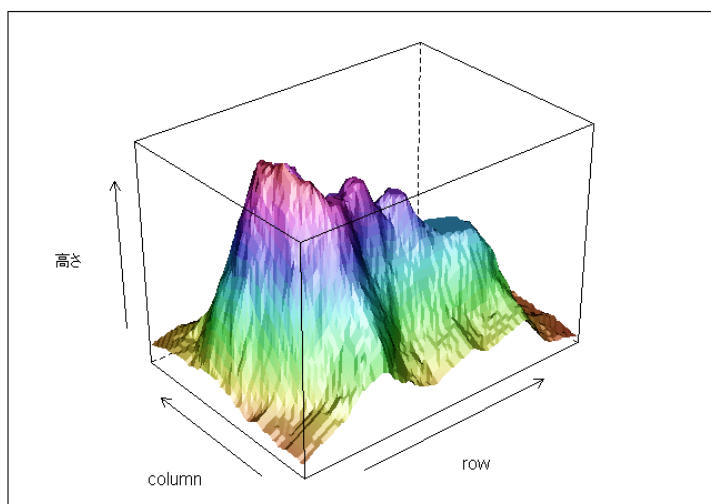


図 44 データセット volcano のワイヤフレーム図

2.7 ヒストグラム

パッケージ `lattice` にあるデータセット `singer` を利用して、ヒストグラムの利用法を説明する。すでに説明したように、このデータは歌手の身長データで、数値変数 `height` (身長; 単位は inch^{*18}) と質的変数 `voice.part` を持つ。

^{*18} 1 インチは 2.54cm。75 インチで約 190cm。

voice.part の水準は 8 つある。それらは Bass 2 , Bass 1 , Tenor 2 , Tenor 1 , Alto 2 , Alto 1 , Soprano 2 , Soprano 1 で*19、この順に音のレベルが高くなる。

2.7.1 データ形式

データセットを R コマンドーに読み込んで表示すると、図 45 のようになる。

	height	voice.part
1	64	Soprano 1
2	62	Soprano 1
3	66	Soprano 1
4	65	Soprano 1
5	60	Soprano 1
6	61	Soprano 1
7	65	Soprano 1
8	66	Soprano 1
9	65	Soprano 1
10	63	Soprano 1
11	67	Soprano 1
12	65	Soprano 1
13	62	Soprano 1
14	65	Soprano 1
15	68	Soprano 1

図 45 singer データの表示

2.7.2 ダイアログボックス

データセット singer をアクティブにし、《分析ツール》▶《ヒストグラム...》を選択すると、図 46 (左) のダイアログボックスが表示される。このまま、**OK** をクリックすると、height データ全体、つまり歌手全員のヒストグラムが表示される (図 47)。

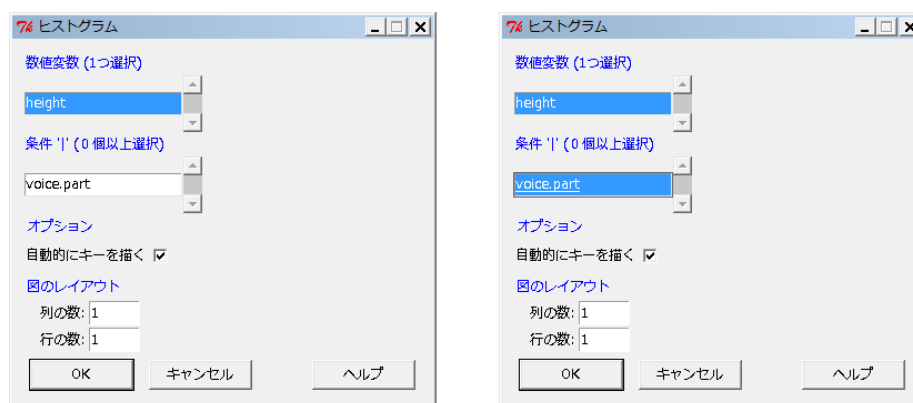


図 46 ヒストグラムのダイアログボックス

*19 Bass : バス、Tenor : テノール、Alto : アルト、Soprano : ソプラノ。

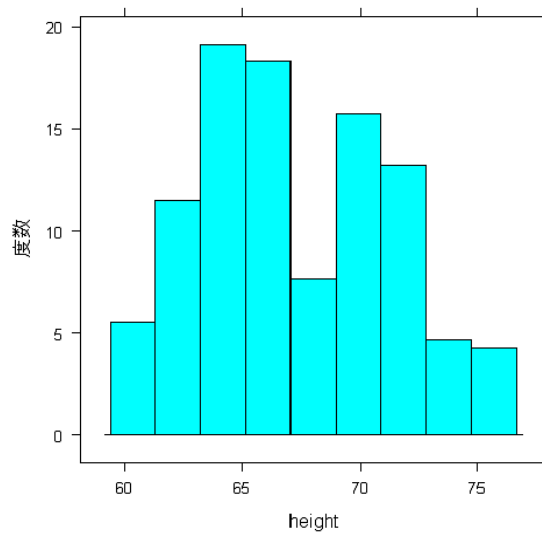


図 47 全体のヒストグラム

ヒストグラムのダイアログボックスで、条件づけの変数に「voice.part」を選択すると(図 46 (右))、voice.part 別のヒストグラムが、下左から右上にかけて(3行3列に)表示される(図 48)。Bass 2 から Soprano 1 に変化するにつれて、身長が高くなっていることがわかる。

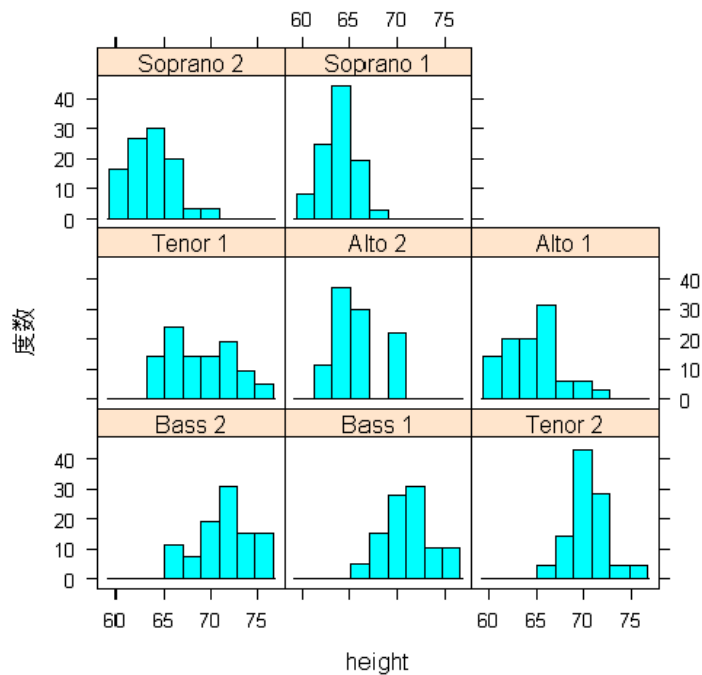


図 48 ボイスパート別(層別)のヒストグラム

ヒストグラムでは、基本的に縦に表示すると比較しやすくなる。そのため、図 49 のようにダイアログボックスの [レイアウト] 欄の列数に「1」を、行数に「8」を入力して（半角数字で）OK をクリックすると、図 50 が表示される（列数を「1」とすることに注意）。この方が一般に、傾向を読み取り易い。

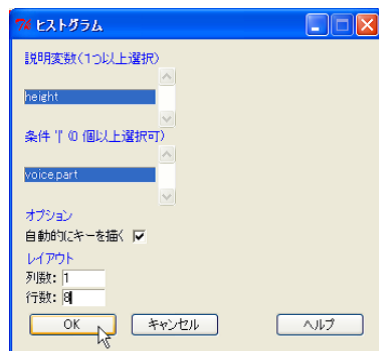


図 49 ダイアログボックスでレイアウトを指定

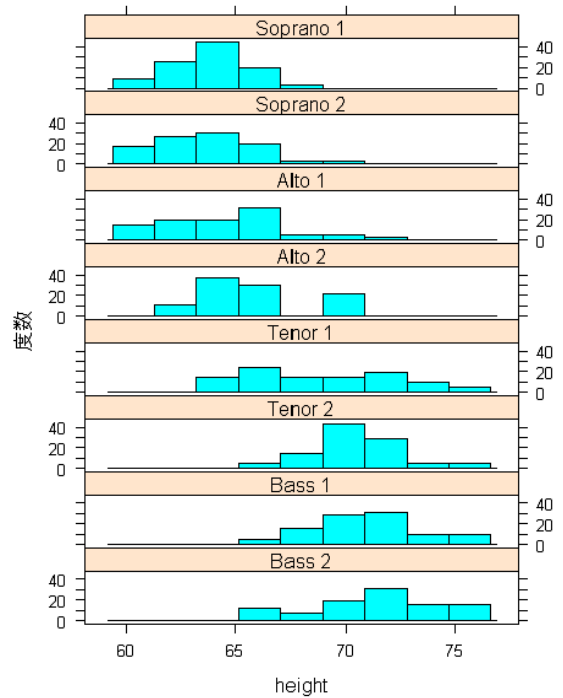


図 50 8 行 1 列表示のヒストグラム

参考 このデータを箱ひげ図で分析することもできる。R コマンドの《グラフ》▶《箱ひげ図...》を選択すると、図 51 のダイアログボックスが表示される。層別のプロットをクリックし、[質的変数] に「voice.part」を指定して OK。再び、[箱ひげ図] のダイアログボックスで OK をクリックすると、図 52 に示す層別の箱ひげ図が表示される。

2.8 ドットチャート

2.8.1 データ形式

岡山県の 1970 年から 2005 年の産業別就業者数データを用いて説明する。データは、図 53 に示すように行列形式で準備する。

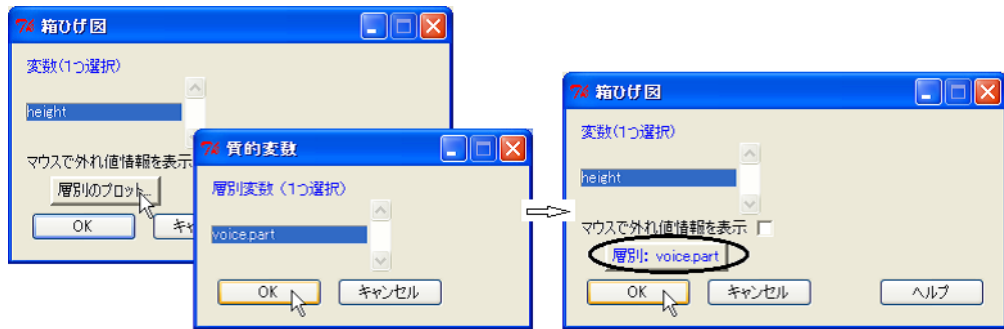


図 51 箱ひげ図のダイアログボックス 層別変数の指定

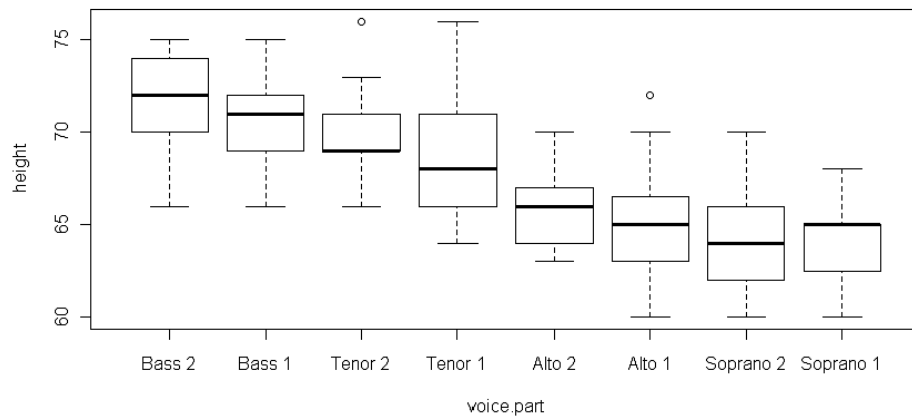


図 52 データセット singer のボイスパート別の箱ひげ図

2.8.2 ダイアログボックス

上記のデータを読み込み、《分析ツール》▶《ヒストグラム...》を選択すると図 54 に示すドットチャートのダイアログボックスが表示される。図の形式を、例えば「点」を選択して **OK** をクリックすると、図 55 のドットチャートを作成することができる。また、「棒」および「一つの図に」を選択すると、図 56 に示す棒グラフを作成することができる。

	第1次産業	第2次産業	第3次産業
1970年	238448	322971	370058
1975年	152419	336950	420059
1980年	122088	338168	463841
1985年	109366	340234	486802
1990年	86017	347743	518825
1995年	77875	344069	565228
2000年	62358	309353	576947
2005年	59677	272414	586459

図 53 行列形式のデータ

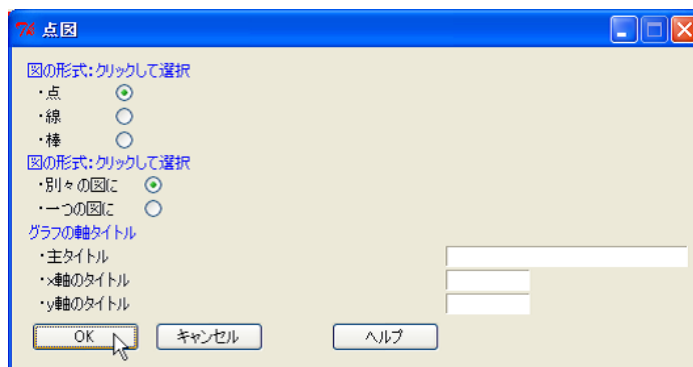


図 54 ドットチャートのダイアログボックス

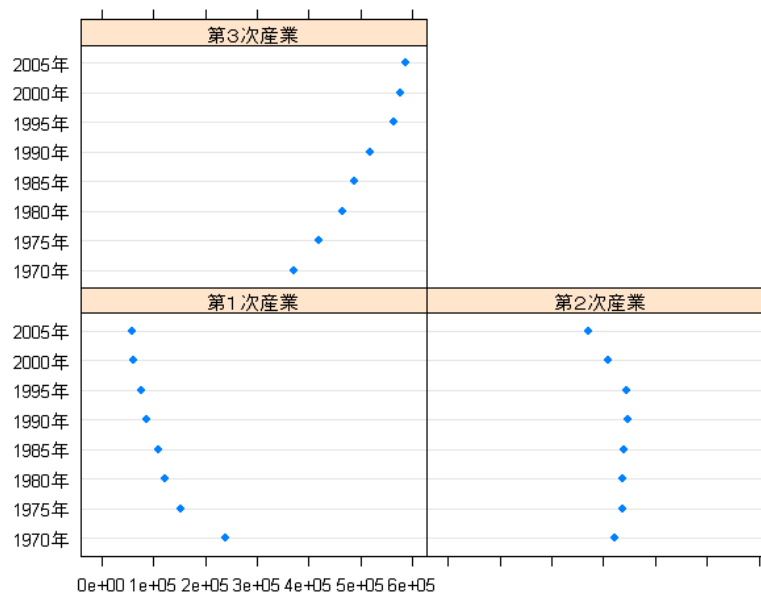


図 55 ドットチャート

3 今後の課題

教育用パッケージには2種類ある。1つはデータ分析用パッケージであり、もうひとつは、確率・統計の考え方を学習するためのパッケージである。後者はパッケージ TeachingDemo や RcmdrPlugin.TeachingDemo でかなりのも

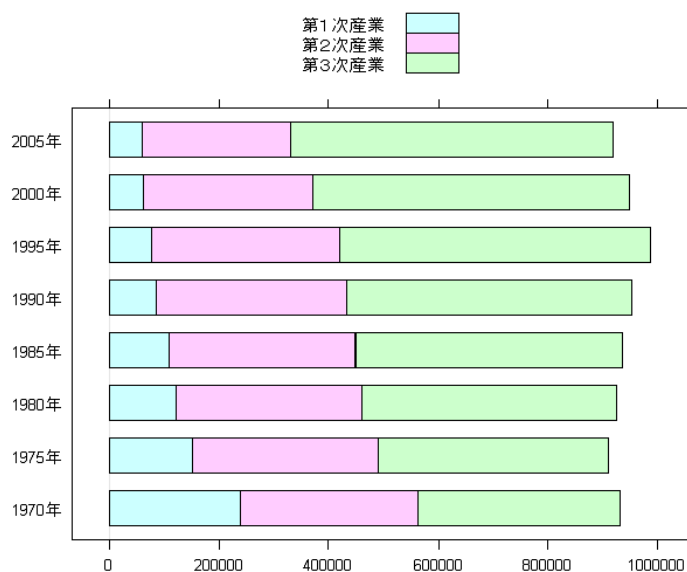


図 56 棒グラフ

のが実現している。また、パッケージ animation により、アニメーションを作成して、シミュレーションをウェブで実行することが可能となる。すでに中心極限定理や大数の法則等のシミュレーションが可能である。

データ分析用パッケージとしては、本稿で荒木が開発している RcmdrPlugin.DAToolsforKids を提示した。このプラグインでは現在、小学校・中学校・高等学校で必要となる度数多角形や複合グラフ（折れ線グラフと棒グラフの混合グラフ）を作成することはできないが、今後、実装していく予定である。また、折れ線グラフ、棒グラフもパッケージ lattice ベースに変更することにより、より高品質なグラフの作成が可能となる。

本パッケージで本質的に欠如しているのは、インタラクティブな機能である。これをパッケージ lattice を基礎として実装することは難しい。そこで、例えば、パッケージ playwith を利用して、インタラクティブに図を作成することができるパッケージを作成することが考えられる。しかし、これにはまず、playwith を日本語化する必要がある。他に、rggobi および GGobi (<http://www.ggobi.org/>) をベースとすることも考えられる。

他の重要な問題として、利用するデータの整備がある。これには、できるだけ実際のデータを用いることにより、統計的思考やツールが現実の問題の把握や解決に役立つことを示すことが重要である。パッケージ datasets を用いると様々なデータが利用可能となる。詳細についてはヘルプで知ることができる（図 57）。これらも日本語化すると教室での利用がしやすくなる。

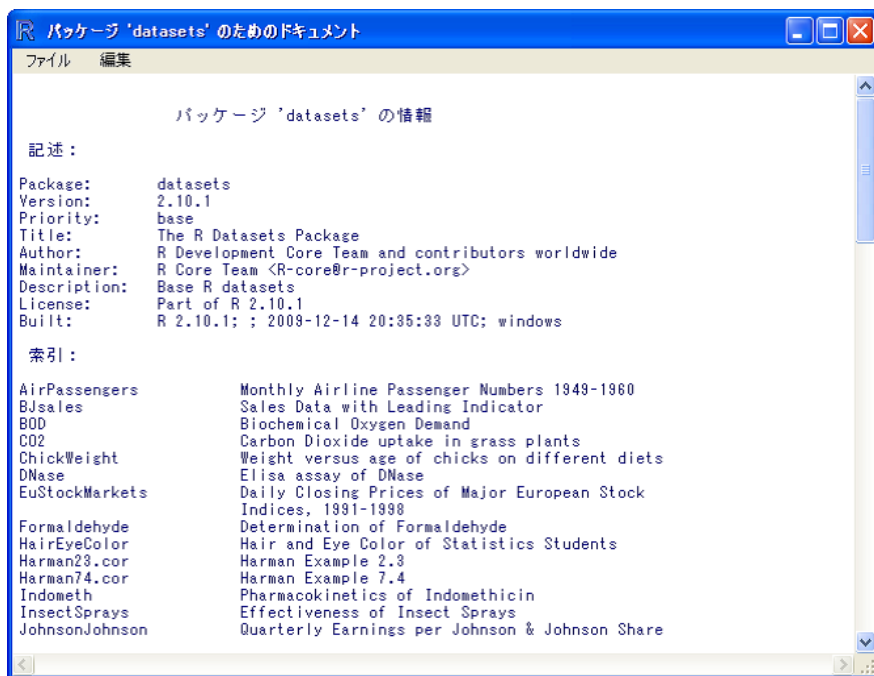


図 57 datasets のヘルプ

参考文献

- [1] 青木繁伸 (2009) 『R による統計解析』 オーム社
- [2] 荒木孝治編著 (2007) 『R と R コマンダーではじめる多変量解析』 日科技連出版社
- [3] 荒木孝治編著 (2009) 『フリーソフトウェア R による統計的品質管理入門 第 2 版』 日科技連出版社
- [4] 荒木孝治編著 (2010) 『R と R コマンダーではじめる事件計画法』 日科技連出版社 (12 月出版予定)
- [5] 金明哲 (2007) 『R によるデータサイエンス』 森北出版
- [6] 舟尾暢男 (2004) 『The R tips - データ解析環境 R の基本技・グラフィックス活用集 第 2 版』 オーム社
- [7] 舟尾暢男 (2009) 『R Commander ハンドブック』 オーム社
- [8] 中澤港 (2007) 『R による保健医療データ解析演習』 ピアソン・エデュケーション
- [9] 間瀬茂 (2007) 『R プログラミングマニュアル』 数理工学社
- [10] Fox, J. (2006a) *An R and S-Plus Companion to Applied Regression*, Sage Books
- [11] Fox, J. (2006b) Getting Started with the R Commander (荒木孝治訳 (2008) 「R コマンダー入門」), パッケージ Rcmdr に付属
- [12] Heiberger, Richard M. and Holland, Burt (2004). *Statistical Analysis and Data Display: An Intermediate Course with Examples in S-Plus, R, and SAS. Springer Texts in Statistics*. Springer.
- [13] Kerns, G. Jay (2010). *Introduction to Probability and Statistics Using R*. このフリー版は、PDF ファイルとして <http://www.lulu.com/product/file-download/introduction-to-probability-and-statistics-using-r/6274176> よりダウンロード可能。書籍版についても本サイトを参照。
- [14] Murrell, P. (2007) *R Graphics*, Chapman & Hall/CRC (久保拓也訳 (2009) 『R グラフィックス R で思いどおりのグラフを作図するために』 共立出版 (株))
- [15] Nakazawa, M. (2010) Package ‘pyramid’, <http://cran.r-project.org/>
- [16] R Development Core Team (2009) R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 3-900051-07-0, <http://www.R-project.org/>

- [17] Sarkar, D. (2008) *Lattice - Multivariate Data Visualization with R*, Springer (石田基広・石田和枝訳 (2009) 『R グラフィックス自由自在』シュプリンガー・ジャパン)
- [18] Sarkar, D. (2009) Package ‘lattice’, <http://cran.r-project.org/>
- [19] Snow, G. (2009) Package ‘TeachingDemos’, <http://cran.r-project.org/>
- [20] South, A. (2010a) Package ‘rworldmap’, <http://cran.r-project.org/>
- [21] South, A. (2010b) rworldmap vignette, <http://cran.r-project.org/>
- [22] Venables, W. N. and B. D. Ripley (2002) *Modern Applied Statistics with S* 4th ed., Springer (伊藤幹夫・大津泰介・戸瀬信之・中東雅樹・丸山文綱・和田龍磨訳 (2009) 『S-Plus による統計解析 第2版』シュプリンガー・ジャパン, 原著第4版の翻訳)
- [23] Verzani, J. (2005) *Using R for Introductory Statistics*, Chapman & Hall/CRC
-
- [24] 荒木のホームページ <http://www.ec.kansai-u.ac.jp/user/arakit/R.html/>
- [25] CRAN (The Comprehensive R Archive Network) <http://www.R-project.org/>
- [26] RjpWiki <http://www.okada.jp.org/RWiki/>