

プログラミング言語 R における 2つの mosaic plot と日本語、多言語表示

藤本 一男

概要

本稿は、カテゴリカル・データの可視化に威力を発揮する mosaic plot の活用方法について説明する。統計処理言語 R の普及と並行して、mosaic plot に接触する機会が増えている。しかし、学校教育の中では、このグラフ表示に対する扱いは明確ではないこと、また、多くの人がグラフ表示ソフトとして用いている Excel にはこのグラフ表示機能が組み込みではないこともあり、必ずしも知られている状況にはない。他方、R の利用者の中でも、機能的に興味があり使ってみたものの、日本語表示が「できない」という問題に直面し、普及がすまないと状況もある。そこで本稿では、mosaic plot の特徴を、Friendly 2002 をもとに整理し、実際に、日本語データに対して利用する際に必要になる設定について解説し、事例として「シンプソンのパラドックス」を扱い、mosaic plot でどのように見えるかを提示する。

KEYWORD: mosaic plot、vcd、R、カテゴリカル・データ、日本語、韓国語、シンプソンのパラドックス

1 問題の所在

まず、最初に棒グラフの持っている問題点を整理し、それを手がかりに mosaic plot の優位さをみていきたい。

なお、以下の展開においては、OS がなんであれ、つまり macOS、OSX、Windows、Linux であれ、R の環境で通常の plot 関数のラベル、タイトルで日本語が正常に出力されるように環境設定されていることを前提にしている^(注1)。また、本稿では、R version 3.4.0 を OS X 10.11.6 (El Capitan) の元で使用している。

1.1 シンプソンのパラドックスと帯棒グラフ

「シンプソンのパラドックス」と呼ばれるものがある。ある分割表でのカテゴリ間の割合に注目したときに、部分表での傾向と全体の傾向が矛盾するというものである。

シンプルな例として、Wikipedia で使われている値^(注2)をもとに、その「矛盾」を確かめてみたい。

二人の生徒、A君、B君がおり、二人は、二回の試験を受けた。各回での正解をA君とB君で比較した時に、それぞれでB君の方がA君よりも正解率が高かった。しかし、全体(1、2回を総合して)で比較すると、A君の方が、正解率が高い、というものである。

ここでは、次のような数値が使われている。

1回目 A君、100問中60問正解、B君、10問中9問正解。2回目 A君、10問中1問正解、B君、100問中30問正解。ここで、回数ごとに比較すれば、一回目のA君の正解率：B君の正解率は、 $\frac{60}{100} = 0.6 < \frac{9}{10} = 0.9$ となり、B君の方が高い(0.6 < 0.9)。二回目については、 $\frac{1}{10} = 0.1 : \frac{30}{100} = 0.3$ であり、こちらもB君の方が高い。ここから、1、2回を合計した総合成績もB君の方が正解率が高いように思ってしまう。ところが、これを、1、2回の合計数で比較すると、 $\frac{60+1}{100+10} = 0.55 : \frac{9+30}{10+100} = 0.35$ となり、A君の方が高い。

この矛盾のからくりは、回数ごとの受験者数が100人と10人であることにありそうなのはすぐに察しがつく。しかし、そうした吟味をしないと、1回目と2回目でもともに正解率がA君よりB君が高いのであるから、総合得点でもA君よりもB君が高いと思ってしまうところにこのパラドックス(「錯覚」)が「成り立つ」条件がある。

比率を数値だけでみていくのであれば、各回の比率を回答問題数で加重して比較することで正確に評価することが可能になる。

しかし、ここでは、この「錯覚」にもう少しつきあいたいと思う。実は、**積み上げ棒グラフ**は、こうした錯覚を引きだしてしまう機能的制約を持っている。各棒の幅は等幅で描かれるために、その棒100%に対応する度数(nの値)の大小とは無関係にカテゴリの比率だけが強調されて表示されてしまうからである。それゆえ、カテゴリ比率を比較する目的の積み上げ棒グラフをつくる時は、n=の表記が必須であるし、読み取る際も、このn=を抜きに正確な評価はできないのである。

では、先の例を等幅の積み上げ棒グラフで描いてみよう。図1.1のようになる。1回目(A1, B1)も2回目(A2, B2)も、Bの方が正解率が高い。

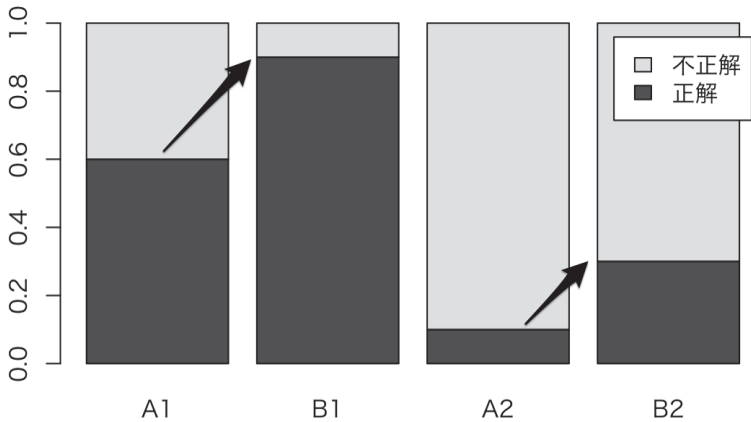


図 1.1 1 回目、2 回目ごとの正解率の比較

次に、表から、1 回目 2 回目を合計して、それをもとに棒グラフを作成してみると図 1.2 のようになる。これで見ると、正解率は $A > B$ である。

この転倒 (矛盾) は、積み上げ棒グラフが情報を「 $n =$ 」をグラフの中に取り込めていないという事情によって引き起こされている。

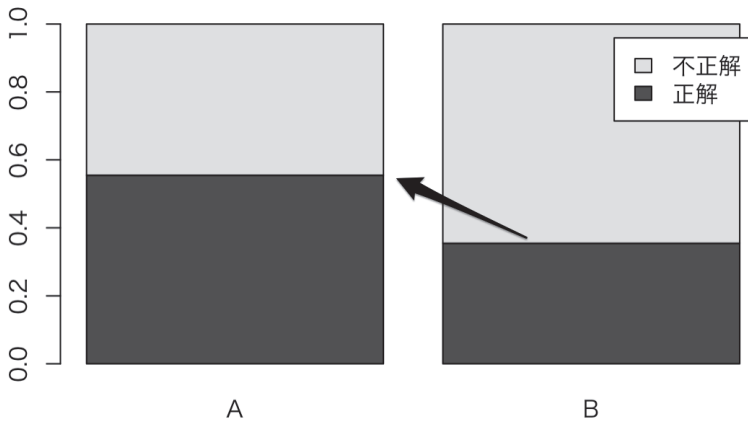


図 1.2 1 回目、2 回目を合計した正解率の比較

1.2 タイルの面積として n の値を取り込む mosaic plot

こうした等幅の積みげ棒グラフの制約を克服すべく活用できるのが、**mosaic plot** である。

mosaic plot は、クロス表のセル度数を、グラフのタイルの面積に対応させて表示する。このグラフ表示は、棒グラフにくらべて計算量が増えるためか、学校教育ではお目にかかることがない^(注3)。しかし、コンピューターが普及し、Rのような統計処理言語が簡単に使えるようになってきた過程で認知されるようになってきた。

本稿は、こうした流れを踏まえ、mosaic plot を活用するためにネックとなっていた日本語使用の問題をクリアし、最後に、今扱った「シンプソンのパラドックス」という錯覚をもたらすデータがどう表示されるかをみていく。

2 量的変数における散布図に対応する質的変数におけるクロス表

データ分析の基本は、ある変数とある変数の関係を見つけることにある。量的変数の場合は、 x 軸と y 軸にそれぞれ注目している変数をわりあてて、それをプロットすることで関係をみることができる。これは散布図 (scatter plot) と呼ばれている。

それに対して、質的変数 (カテゴリカル・データ) の場合は、この散布図に相当するものがクロス表である。それゆえ、カテゴリカル・データの分析の基礎は、このクロス表にある。表の形にまとめられたものを、適切な方法で数量化することで、さらに分析をすすめることができるが、まず必要なのは、手に入れたデータの様子を把握することにある。量的データの場合は、変数ごとに、ヒストグラムを描き、データの分布を確認することができる。カテゴリカルデータの場合は、カテゴリーごとの集計が基本であり、それは、やはりクロス表での集約である。

次に必要になるのは、量的変数においては散布図で確認したような変数間の関係の把握である。そこでまずおこなわれるのは、作成したクロス表の行変数か列変数の一方に注目して、そこに対応する列変数、行変数との関係を図示することにある。ここで、用いられるのが、棒グラフ、なかでも、カテゴリーごとの度数の累積が 100% になるような積み上げ棒グラフである。このグラフによって、注目している行変数の特徴が、列カテゴリーの分布によって読み取られることになる。

しかし、冒頭のシンプソンのパラドックスの例でみたように、棒グラフは、

累積棒グラフであっても、等幅で描かれる。そのために、 $n =$ の記載が必須にはなるものの、グラフにとっては、付加情報の位置であって、グラフそのものの等幅性は依然かわりがない。

そこで、開発されたのが mosaic plot である。

3 mosaic plot の機能的特徴

まずは、mosaic plot による表示を見ていただこう。データは、後に解説するが、HairEyeSex という、髪の色、目の色、性別の 3 元クロス表 (R に default で提供されている) のデータセットである。タイルの面積がセルの度数に対応しているものとして眺めてみれば、各カテゴリーの関係が見えてくる。

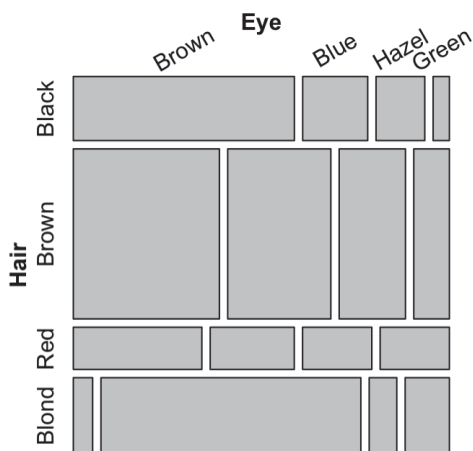


図 3.1 mosaic plot の表示例：HairEyeSex で性別を統合して表示している。

3.1 二つの mosaic plot

さて、ここまでは一般名称として「mosaic plot」という語を用いてきたが、R で使える代表的な **mosaic plot** には、二つのものがある。

一つは、base で提供される **mosaicplot()** である。今一つは、その拡張版として位置付けられている、**mosaic()** である。それぞれ提供されているパッケージを明示して表記すると以下ようになる。

- `graphics::mosaicplot()`
- `vcd::mosaic()`

Rに最初からインストールされているのが、`mosaicplot`であるが、提供されている機能的な充実度から見て `vcd::mosaic()` を利用していきたい。

3.2 mosaic plotの歴史

その `vcd::mosaic()` は、`vcd` というパッケージで提供されているが、このパッケージ誕生をたどると、今日の `mosiac plot` への経緯があきらかとなる。

中心にいるのは、トロント大学の Michael Friendly 教授である。

`Friendly 2002` をもとに整理する。

詳細は、“Advances in Visualizing Categorical Data Using the `vcd`, `gnm` and `vcdExtra` Packages in R CARME 2011”, http://carme2011.agrocampus-ouest.fr/slides/Friendly_Turner_Firth_Zeileis.pdf に記述されている。

— 1981, 1984 Hartigan and Kleiner

n 元の分割表を、度数を「タイル」の面積～セル度数に分解して(再帰的)な `mosaic` 表示として提示。

— 1994 Friendly

`mosaic` 表示と対数線形モデルの結合を開発。これによって、観測度数(面積)と各モデルからの残差を色分けして表示することができることを示す。

— 1995 CARME1995 において最初の発表^(註4)。

— 2000 Friendly “Visualizing Categorical Data” 刊行(SAS Institute)。

— 2002 Friendly 「しかし `mosaic` ライクな表示は、長い歴史を持っている！」

— 2002 ウィーン工科大学とウィーン経済大学で `vcd project` 発足(Kurt Hornik, David Meyer, Achim Zeileis)。 `vcd package` の誕生へ。

4 データの用意：HairEyeColor データ

`Friendly2000` をもとに、データの詳細を確認する。

4.1 基本とするデータ：HairEyeColor

`Hair × Eye` の表は、Snee (1974) によって報告されたデラウェア大学の学生の調査に基づいている。性別による分割は、教育的な目的のために Friendly (1992a) によって追加された^(註5)。

このデータセットは、標準カイ二乗検定、より一般的には対数線形モデル

ング、mosaic plot、sieve diagram、または association plot などのグラフィカルメソッドなど、分割表を分析するためのさまざまな手法を示すのに役立つ。

4.2 書式

3 つの変数で 596 人の観測値をクロス集計した 3 次元配列。変数とそのレベルは以下のとおり。

名前	レベル
1. Hair (髪の色)	Black 黒、Brown 茶色、Red 赤、Gold 金髪
2. Eye (目の色)	Brown 茶色、Blue 青、Hazel 薄茶、Green 緑
3. Sex (性別)	Male 男性、Female 女性

4.3 日本語変数名、カテゴリ名をつける。韓国語変数名、カテゴリ名をつける

さて、次のように英語、日本語、韓国語のデータを準備し、それを、mosaic で使用しながら、パラメータの設定の仕方を確認していこう。

```
library(vcd)
.d <- HairEyeColor
dimnames(.d)
.d_k <- .d_j <- .d
dimnames(.d_j) <- list(髪=c("黒","茶","赤","ブロンド"),
                       眼= c("茶","青","薄茶","緑"),
                       性別= c("男性","女性"))
dimnames(.d_k) <- list(   = c("검정", "차", "빨강", "금발"),
                       = c("차", "파랑", "밝은", "녹색"),
                       = c("남자", "여자"))
```

5 graphics::mosaicplot での表示

最初に、mosaicplot() を用いて、データを表示してみる。

```
mosaicplot(.d, main="HairEyeColor")
```

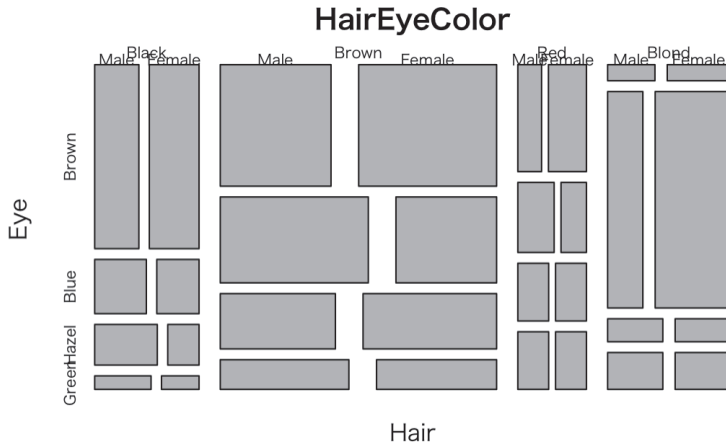


図 5.1 Exhibit 7.1

ここで確認しておきたいのは、データ表での変数の位置と mosaicplot での変数の位置の関係である。なお、.d を、.d_j に変えれば、日本語のデータが図示される。ただし、.d_k を選択してもハンゲル表示はされない^(注6)。

```

, , Sex = Male

      Eye
Hair   Brown Blue Hazel Green
Black   32   11   10    3
Brown   53   50   25   15
Red     10   10    7    7
Blond    3   30    5    8

, , Sex = Female

      Eye
Hair   Brown Blue Hazel Green
Black   36    9    5    2
Brown   66   34   29   14
Red     16    7    7    7
Blond    4   64    5    8

```


このように、データ表では、行に髪の色 (Hair) が位置し、列に、目の色 (Eye) が位置しているにもかかわらず、グラフ上はそれが入れ替わっている。入力する際に `t()` で転置させてもよいが分割方向の設定などの面倒が発生するので解決にはならない。

`vcd::mosaic` は、この `mosaicplot()` の拡張版として位置付けられている。`vcd` の解説を引用しておく。

`mosaic` 表示は、Hartigan and Kleiner (1984) の統計学論文で示唆されており Friendly (1994) によって拡張された。`mosaicplot` は `base` の `graphics` による実装であるが、`mosaic` ははるかに柔軟で拡張可能な `grid` に基づく実装である。`mosaic` は、デフォルトのメソッドと数式インタフェースを持つ汎用関数^(注7)である。ともに、`strucplot` 関数への高水準インタフェースであり、(拡張された) `mosaic` 表示を生成する。独立モデルの指定、ラベル付け、凡例、スペーシング、シェーディング、その他のグラフィカルなパラメータなど、ほとんどの機能がそこに記述される。期待度数の (おそらくより高次の) 表の領域に対応した視覚化を行う。

`mosaic plot` は、正方形の再帰的な垂直および水平分割によって作成された (セルに対応する) タイルで構成されている。各タイルの面積は、前の次元の分割の寸法を考慮して、対応するセル値に対応している。

また、拡張 `mosaic plot` は、特定の対数線形モデルへの適合度合いを視覚化する。特徴的には、これは、タイルの色および/または輪郭で、対応する残差の符号 (\pm)、大きさおよび可能性としての有意性を視覚化する残差に基づくシェーディングによって行われる。レイアウトは非常に柔軟である。

シェーディング、ラベル付け、スペーシング、および凡例の仕様はモジュール化されている (詳細は `strucplot` の文献を参照)。`graphics` の `mosaicplot` 関数とは対照的に、`split` (タイルへの分割) はデフォルトで `structable` の出力に一致するように水平方向から開始する。(https://cran.r-project.org/web/packages/vcd/vcd.pdf の `mosaic` の項より)

6 `vcd::mosaic` での表示

拡張された `mosaic plot` として位置付けされている `vcd::mosaic()` であるが、日本語ラベルを用いるのは「簡単」ではなかった。この `mosaic` を含めて `vcd` のグラフ描画関数が `grid` 系の `graphics` である `strucplot` というフレームワークにもとづいた設定をおこなわなければならないのだが、必ずしもわかりやす

いものではなかったのである^(注8)。

以下に、変数名、カテゴリ名、を日本語で設定した場合の mosaic での表示を示す。

```
my.largs = list(gp_labels = gpar(fontsize = 10, fontfamily = "sans"),
               gp_varnames = gpar(fontsize = 12, fontfamily = "sans"))
my.main <- gpar(fontsize = 14, fontfamily = "sans")
my.sub <- gpar(fontsize = 10, fontfamily = "sans"),
mosaic(.d_j,
       labeling_args=my.largs, ← 本体の日本語表示に関する部分
       main_gp = my.main, sub_gp = my.sub, ← タイトル、サブ・タイトルの日本語表示に関する部分
       rot_labels = c(left = 0, top = 0, right = 0),
       pop = FALSE,
       keep_aspect_ratio = FALSE,
       shade = TRUE
)
```

上記リストで mosaic() を呼ぶ前に、設定している **my.largs** = の設定が、mosaic を含めた strucplot frame work でのパラメータ (gpar) の設定であり、これを、本体 (mosaic の中) で参照している。ここでの fontfamily をハンゲル表示が可能なフォントに設定すれば mosaic でハンゲル表示が可能になる。macOS、OSX であれば、AppleGothic を指定することで、日本語韓国語(そして英語)の混在表示が可能になる。

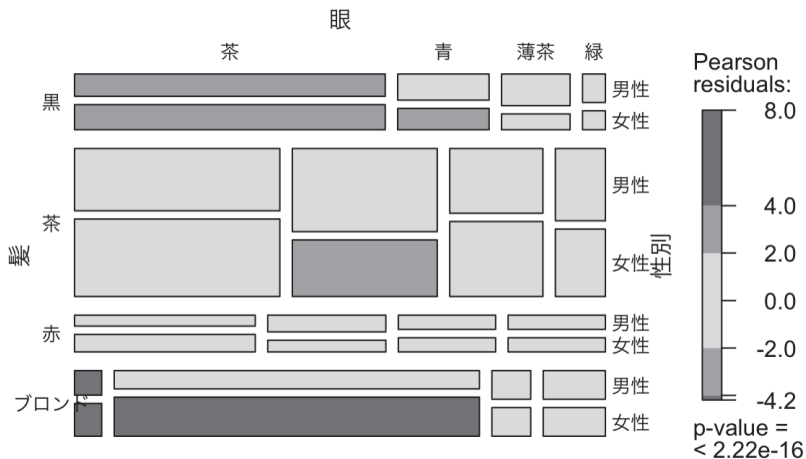


図 6.1 vcd::mosaic で表示した HairEyeSex (日本語)、残差表示付き

```

my.largs.k = list(gp_labels = gpar(fontsize =10, fontfamily = "AppleGothic"),
                 gp_varnames = gpar(fontsize = 12, fontfamily = "AppleGothic"))
my.main = gpar(fontsize = 14,fontfamily = "AppleGothic")
my.sub = gpar(fontsize = 10,fontfamily = "AppleGothic")
mosaic(.d_k,
       labeling_args=my.largs.k,
       rot_labels = c(left = 0, top = 0,right=0),
       main="日本語 한글混在 vcd::mosaic",
       main_gp = my.main,
       )#

```

ここで `mosaic()` で設定したパラメータは、`vcd` が提供する他の関数^(注9) においても共通のフレームワーク、`strucplot` の設定値である。それを以下に簡単に説明する。

7 strucplot framework

`strucplot` framework の解説は、[Meyer2006] に詳説されている。ただし、フレームワークの説明であるため、必ずしもわかりやしくない。ここでは、上述した日本語表示、他の幾つかの機能の設定を記述しておく。それを手がかりに、[Meyer2006] にある `mosaic` 他の表示に関するパラメータを使っているであろう。なお、`strucplot` のパラメータの説明は、`?strucplot` でも表示されるし、また、R documentation ページの `strucplot` を参照しても得られる (<https://www.rdocumentation.org/packages/vcd/versions/1.4-3/topics/strucplot>)。

- `gp_labels`, `gp_varnames` : 上述のソースコードでは、これらのパラメータに、`gpar` (grid パラメータ) の “`fontsize`”、 “`fontfamily`” を指定している。`vcd` 系のグラフィック機能でタイルの表示、また、変数名カテゴリー名に日本語を使用するには、ここで、日本語対応しているフォントファミリーを指定しなくてはならない。日本語韓国語混在表示で用いたフォント “`AppleGothic`” は、その両方の表示が可能なものである。

- `main_gp`, `sub_gp` : これは、それぞれ、`main` 領域 (表題)、`sub` 領域 (副題) で用いる `gp` (grid parameter) である。表題や副題で日本語を使う場合には、ここに `gpar` で設定する。日本語に限らず、欧文であっても、使用するフォントを指定する場合は、ここで指定すればよい。

- `rot_labels` : カテゴリーラベルの文字向きを指定できる。

`rot_labels = c(left = 0, top = 0,right=0)` のようにベクトルで与える。数値は、回転角度である。0 であれば回転なし。45 であれば、45 度傾けて

日本語한글混在 vcd::mosiac

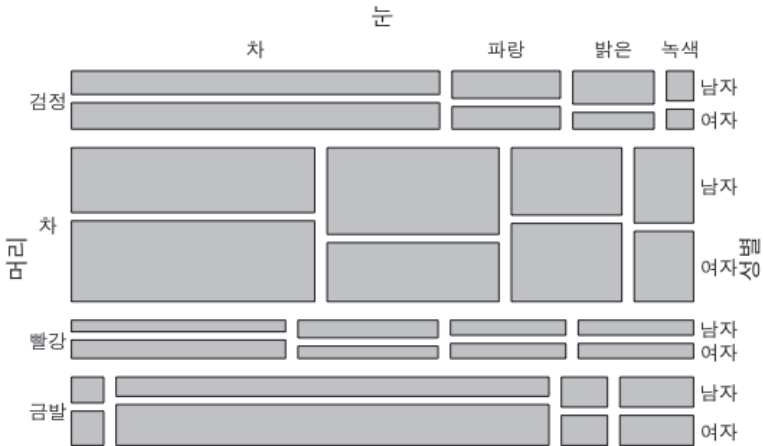


図 6.2 vcd::mosaic で表示した HairEyeSex (日本語韓国語混在)

カテゴリーラベルが表示される。

strucplot framework には、かなり多くのパラメータが用意されている。ただ、ここまでで説明したような使い方さえ理解すればあとは必要に応じて用いていけばよい。

8 vcd::mosaic で拡張された機能を使ってみる

8.1 mosaic でセル度数をタイルに表示する

これを行うには、以下のように、`labeling = labeling_values`を指定する。タイルに、セルの残差を値で表示するには、`labeling = labeling_residuals`を指定する。以下に、変数名、カテゴリ名に日本語を用い、残差で色分けし、右辺のラベル (性別) を 30 度傾斜させて表示するためのコードと結果を示す。

```
mosaic(.d_j,
  margins = c(left = 5,top = 4),
  labeling_args=my.largs,
  rot_labels = c(left = 0, top = 0,right=30),
  keep_aspect_ratio=FALSE,
  gp = shading_Friendly,
  labeling = labeling_residuals)
```

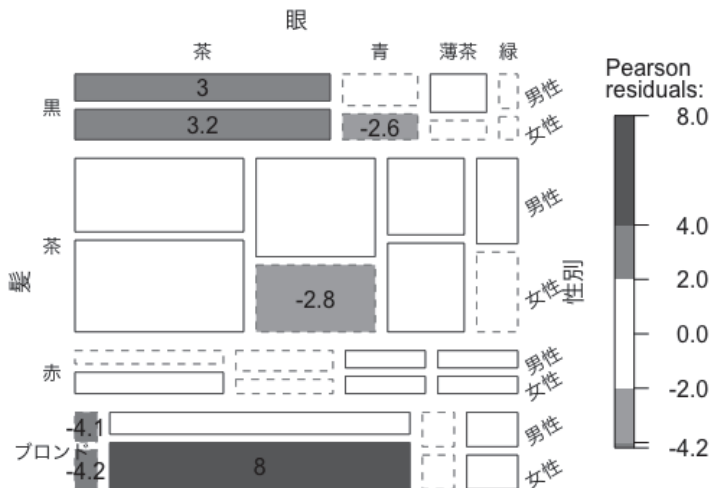


図 8.1 日本語ラベル、残差色分け、タイルに表示、右辺 30 度

8.2 DoubleDecker という mosaic () のカスタマイズ版

vcd が提供する便利な機能に、DoubleDecker がある。これは、2 値の目的変数の結果と独立変数の関係を可視化する機能である。[Meyer2016] には次のような解説がある。

Doubledecker プロットは、1 つのカテゴリ (通常は 2 値) 変数の他のカテゴリ変数への依存を視覚化する。内容は、従属変数 (結果) を除くすべての次元 (他のすべての変数カテゴリ) に対して垂直分割された mosaic plot である。最後の変数は、水平分割、タイル間のスペースなし、およびレベルの個別の色に設定されて視覚化される。

ここでは、HairEyeSex と同じように日本語化した Titanic のデータを表示させてみる。ラベルの日本語表示部分は、mosaic で日本語フォントを指定したのと同じものを用いている。

```
Titanic_J <- Titanic
dimnames(Titanic_J) <- list(Class=c("1 等", "2 等", "3 等", "船員"),
                             性別=c("男性", "女性"),
                             年齢=c("子ども", "大人"),
                             生存=c("No", "Yes"))
doubledecker(Titanic_J, labeling_args=my.largs, main="Titanic_J")
```

Titanic_J

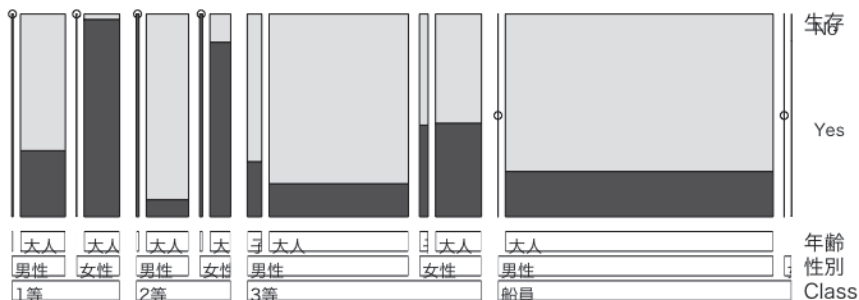


図 8.2 タイタニック号生存率の客室等級、性別による分類

9 実践例

この DoubleDecker プロットを用いることで、冒頭にあげたシンプソンのパラドックスがどのように視覚化されるかみておきたい。積み重ね棒グラフの「等幅」がシンプソンのパラドックスという「錯覚」の誘引であった。ここでは、それを mosaic/DoubleDecker で表示する。

9.1 3元クロス表をつくる

データを準備する。必要になるのは3元クロス表なので、1回目、2回目ごとの2元クロスを2つ用意しそれを array でまとめている。

```
N1 <- matrix(c(60, 40,
              9, 1),byrow=T,2,2)
N2 <- matrix(c(1,9,
              30,70),byrow=T,2,2)
.d <- array(c(N1,N2),dim=c(2,2,2),dimnames=list(生徒=c("A","B"),
                                                解答=c("正解","不正解"),
                                                回数=c("1回目","2回目")))
```

日本語表示に必要な my.largs はこれまで使ってきたものと同じである。

```
STR <- structable(解答 ~ 回数 + 生徒,data=.d)
STR[,2:1]
doubledecker(STR[,2:1],labeling_args=my.largs)
```

行頭に「解答」、行側に「回数・生徒」つまり「回数」との「生徒」で表示。図9.1

このグラフをみれば、 $n =$ に対応する情報がグラフの横幅として表示されているので、それぞれの回ごとには、Bの方が正解率は高くても、正解の面積自体は、高校Aの方が多くことは一目瞭然ではある。それがより明確になるように、これを、高校ごとに並べかえてみる。この並べ替えも、変数の順番の変更で可能なことが、strucplotの強力なところである。

```
STR2 <- structable(解答 ~ 生徒 + 回数,data=.d)
vcd::doubledecker(STR2[,2:1],labeling_args=my.largs)
```

行頭は「解答」、行側は、「生徒・回数」つまり、「生徒別に回を並べて表示。図9.2

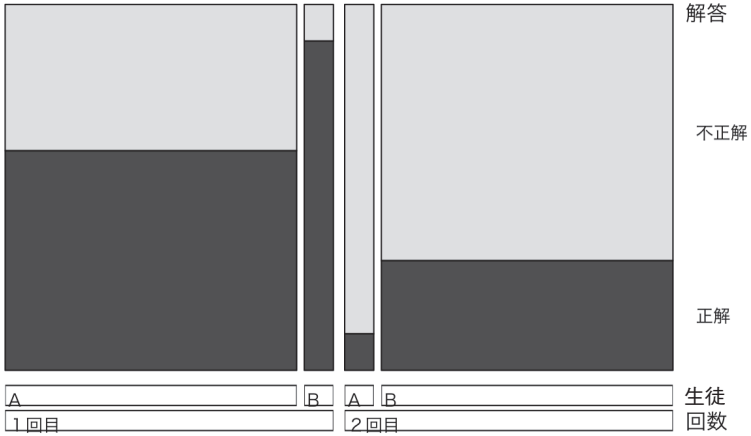


図9.1 1回目、2回目ごとにA校、B校を並べた Doubledecker

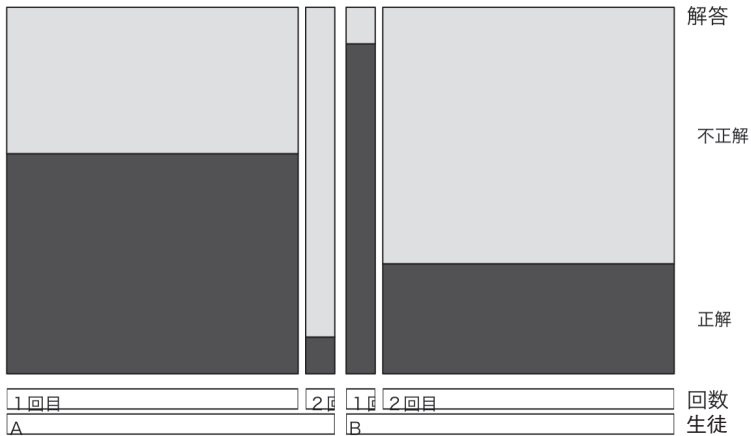


図9.2 高校ごとに並べた正解率

10 まとめ

以上みてきたように、

- mosaic plot は帯棒グラフの表示機能を拡張しながら、カテゴリカルデータを可視化するための基本ツールとなっている。
- 日本語表示の問題は、strucplot framework のパラメータ設定の問題として解決される。
- シンプソンのパラドックのデータを、mosaic (Doubledecker) で表示させれば、矛盾 = 「錯覚」はおこならない。

注

- 1) ネット上に公開されている日本語利用のための font 設定は、以下のページなどを参照。
<http://d.hatena.ne.jp/MikuHatsune/20120428/1335544822>
- 2) <https://ja.wikipedia.org/wiki/シンプソンのパラドックス>
- 3) mosaic plot の「教科書」での扱い。大学の使用される統計学のテキストでは紹介され始めている。たとえば、稲葉由之, 2012, 『プレステップ統計学I 記述統計学』(プレステップシリーズ14)、弘文堂、ppxx のコラム。
- 4) CARME (“Correspondence Analysis and Related Methods”, 対応分析と関連技法に関する国際会議) 会議の歴史は、Greenacre 2017 の「はじめに」にまとめられている。タイトルにあるのは対応分析であるが、その関連技法に、vcd は含まれており、2015 年のナポリ大会では、vcd の中心人物、Michael Friendly が「Data and Visualization」をテーマに記念講演を行っている。
<https://youtu.be/uti5FIFJsN4>
- 5) Friendly2000: 392 によるコメント
 髪の色と目の色との関係は、このようなデータでは驚くほど見通せない。遺伝的に、髪の色、目の色、および肌の色はメラニンに関連しており、少なくとも 2 種類のメラニンが存在し、そのうちの 1 つは赤みがかった (明るい) もの、もう 1 つは黒っぽいことが知られている。皮膚の中より暗い形のメラニンをたくさん持っている人は、おそらく髪と目にそれを持っている。しかし、遺伝的には無関係であるが、人種の中で一緒に起こりうる多数の人種特性が存在する。Snee のデータセットの学生は、人種では分類されていないため、不明だ。髪の色、目の色、性別の間に何らかの関係があるかどうかは、さらに謎である。
- 6) これは今使用しているマシンのシステム言語が日本語だからであり、韓国語の環境であればハングルが表示される。しかし、その時は日本語は表示されない。
- 7) 汎関数については、“Advanced R” chap11 Functionals (邦訳「R 言語徹底解説」11 章 p215「汎関数」)

を参照。

8) vcd::mosaic で日本語がうまく表示できない問題について

- Rjpwiki での質問あり。誰も回答してない。vcd 初心者? (2012-10-23 (火) 17:43:10)

```
http://www.okada.jp/RWiki/?%EF%BC%B1%EF%BC%86%EF%BC%A1%20(%E5%88%9D%E7%B4%9A%E8%80%85%E3%82%B3%E3%83%BC%E3%82%B9)%2F14#sbdd4c23
```

「さらに、ラベルを日本語で表示しようとして、次のように入力すると、

```
dimnames(tab) <- list(サイズ=(c("S", "M", "L"), 性別=c男(" ", 女(" ")))
mosaic(tab, split_vertical=T)
```

以下にエラー `grid.Call.graphics(L_downviewport, name$name, strict):Viewport' cell:`

`サイズ = S' was not found`

とエラーになってしまいます。日本語が通らないようなのですが、何か対処法はありますでしょうか。」

- 豊澤栄治「実践! Web マーケッターのための R 入門顧客属性データとコンバージョンデータを使ってどんどん打ち手を考えよう!【R 入門講座】

<https://markezine.jp/article/detail/20665> (2014/08/28 08:00)

「あれ! ? なんで今回は日本語の変数名ではないのか? とツッコミが入りそうですね。そうなんです。vcd パッケージを使うのですが日本語表示で問題がありまして、英語表記とさせていただきます。環境によってはうまくいかも知れませんが……。お許しください!」

- Winston Chang (石井、河内、瀬戸山、古島)「R グラフィックスブック ggplot2 によるグラフ作成のレシピ集」オラリー・ジャパン、2013 年。なお、「ggplot2 による……」は邦題でのサブタイトル。この「レシピ 13.15 モザイクプロット」を作成するは、vcd::mosaic を紹介している。

このレシピの最後に、「モザイクプロットを作成するその他の関数は、?mosaicplot を参照してください。」(p316) とあるが(原著でも同様)、vcd::mosaic と graphics::mosaicplot は別のものであり、vcd::mosaic の help は、?mosaic である。なお、本書の付録 B「グラフでの日本語フォントの利用」は、windows や OSX 向けの日本語設定の話と、個別 ggplot2 で日本語をつかえるようにする説明だけである。

参考文献

- [Friendly2000] Friendly, M, 2000, “Visualizing Categorical Data”, SAS series in statistical applications, SAS Institute, サポートページ, <http://www.datavis.ca/books/vcd/>
- [Friendly2002] Friendly, M, 2002, “A Brief History of the Mosaic Display”, Journal of Computational and Graphical Statistics, Vol.11, No.1 (Mar., 2002), pp.89-107, American Statistical Association
- [Friendly2016a] Friendly, M, 2016, “Working with categorical data with R and the vcd and vcdExtra

- packages”, Using vcdExtra version 0.7-0 and vcd version 1.4-2,
<https://cran.r-project.org/web/packages/vcdExtra/vignettes/vcd-tutorial.pdf>
- [Friendly. Meyer2016] Friendly. M, Meyer. D, 2016, “Discrete Data Analysis with R”, CRC Press
- [Greenacre2017] Greenacre. M, “Correspondence Analysis in Practice Third Edition”, CRC press
- [Meyer2006] Meyer. D, Zeileis. A, Hornik. K, 2006, “The Strucplot Framework: Visualizing Multi-way Contingency Tables with vcd”, Journal of Statistical Software, October 2006, Volume 17, Issue 3.
<https://www.jstatsoft.org/article/view/v017i03/v17i03.pdf>
- [Meyer2016] Meyer. D, Zeileis. A, Hornik. K, Gerber. F, Friendly. M, 2016, “package ‘vcd’”,
<https://cran.r-project.org/web/packages/vcd/vcd.pdf>
- [Friendly. Meyer2011] Friendly. M, Meyer. D, Zeileis. A, Hornik. K, 2011, “Advances in Visualizing Categorical Data Using the vcd, gnm and vcdExtra Packages in R”, CARME 2011 Rennes, February 9-11, 2011,
<http://www.datavis.ca/papers/adv-vcd-4up.pdf>
- [藤本 2017] 藤本一男, 2017, 「2つの mosaic plot と日本語表示」Tokyo. R#65 における LT 発表資料。slide share で公開。
<https://www.slideshare.net/kazuofujimoto/2mosaic-plot-80084536>
- Wickham, Hadley, 2014 (2016), “Advanced R” (訳: 石田『R 言語徹底解説』共立出版)
- R Core Team (2017). R: A language and environment for statistical computing.
- R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

■ 謝辞

本稿は、Tokyo.R # 65、66 での LT での発表をベースにしています。参加されたみなさま、コメントをくださった方々に感謝いたします。