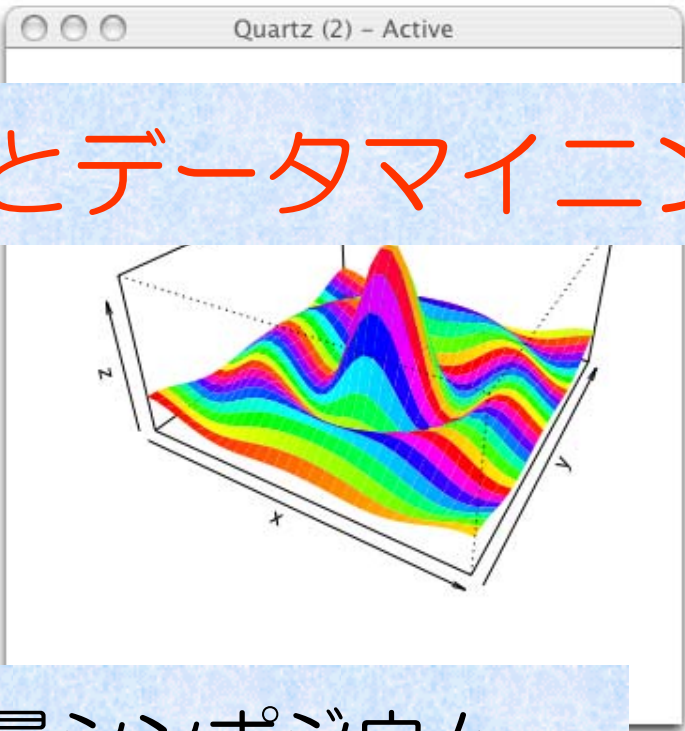


Mac OS X desktop environment showing an R Console window and a Quartz window.

Rによるグラフ作成とデータマイニング

```
> plot(1:10)
> 1+2
[1] 3
> 3+4
[1]
[1]
> 1+
[1]
> 3+
[1]
> ?plot
> x <- seq(-10, 10, length= 30)
> y <- x
> f <- function(x,y) { r <- sqrt(x^2+y^2); 10 * sin(r)/r }
> z <- outer(x, y, f)
> z[is.na(z)] <- 1
> op <- par(bg = "white")
> persp(x, y, z, theta = 30, phi = 30, expand = 0.5, col = "lightblue")
> x <- seq(-10, 10, length= 30)
> y <- x
> f <- function(x,y) { r <- sqrt(x^2+y^2); 10 * sin(r)/r }
> z <- outer(x, y, f)
> z[is.na(z)] <- 1
> op <- par(bg = "white")
> persp(x, y, z, theta = 30, phi = 30, expand = 0.5, col = rainbow(200))
> x <- seq(-10, 10, length= 30)
> y <- x
> f <- function(x,y) { r <- sqrt(x^2+y^2); 10 * sin(r)/r }
> z <- outer(x, y, f)
> z[is.na(z)] <- 1
> op <- par(bg = "white")
> persp(x, y, z, theta = 30, phi = 30, expand = 0.5, border=NA, col = rainbow(50))
> x <- seq(-10, 10, length= 30)
> y <- x
> f <- function(x,y) { r <- sqrt(x^2+y^2); 10 * sin(r)/r }
> z <- outer(x, y, f)
> z[is.na(z)] <- 1
> op <- par(bg = "white")
> persp(x, y, z, theta = 30, phi = 30, expand = 0.5, border=NA, col = rainbow(200))
>
```



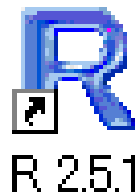
～第 91 回行動計量シンポジウム～

武田薬品工業
舟尾 暢男

本日のメニュー



- R の概要 [2~3分]
- R でデータを読み込む方法 [4~5分]
- グラフとデータマイニング [15~20分]
- データマイニングとは? [7~8分]
- R Commander の紹介 [4~5分]
- 質疑応答など [5分程度?]



```
R Console
ファイル 編集 その他 パッケージ ヘルプ

R version 2.5.1 (2007-08-27)
Copyright (C) 2007 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

Rはフリーソフトウェアであり、「完全に無保証」です。
一定の条件に従えば、自由にこれを再配布することができます。
配布条件の詳細に関しては、'license()'あるいは'licence()'を入力してください。

Rは多くの貢献者による共同プロジェクトです。
詳しくは'contributors()'を入力してください。
また、RやRのパッケージを出版物で引用する際の形式については
'citation()'を入力してください。

'demo()'を入力すればデモをみることができます。
'help()'とすればオンラインヘルプが出ます。
'help.start()'でHTMLブラウザによるヘルプがみられます。
'q()'と入力すればRを終了します。

> |
```

R の概要



- オープンソース&フリーの統計解析用ソフト

【インストール】

- 「[今日からあなたも統計ソフト開発者](#)」で Google 検索 →

【長所】

- 関数電卓, 数値計算, プログラミング, 統計解析, グラフィックスの機能があり, どの機能も充実している
- 機能拡張が容易に行える
- 使用人口が多いので, バグが少なく情報も豊富

【短所】

- EXCEL などの表計算ソフトに比べて GUI (マウス操作) の機能が劣っている ⇒ R の命令をひとつひとつ覚えなければいけない...
- 大規模なデータを扱う場合は多少骨が折れる

本日のメニュー



- R の概要
- R でデータを読み込む方法 ←
 - データフレームとは
 - データフレームの作成
- グラフとデータマイニング
- データマイニングとは？
- R Commander の紹介
- 質疑応答など

データフレームとは



- データマイニングをするには、まずデータの作成から！
 - (R上で) データを手で入力して・・・
 - テキストファイル, EXCEL, SAS などのファイルを読み込み・・・
- Rでデータ解析を行う際は、データフレームという形式にデータを変換することが多い

	A	B	C
1	sex	height	weight
2	F	160	50
3	F	165	65
4	M	170	60
5	M	175	55
6	M	180	70

EXCEL : シート

	SEX	HEIGHT	WEIGHT
1	F	160	50
2	F	165	65
3	M	170	60
4	M	175	55
5	M	180	70

R : データフレーム

	sex	height	weight
1	F	160	50
2	F	165	65
3	M	170	60
4	M	175	55
5	M	180	70

SAS : データセット

データフレームの作成 (手入力)



3つのベクトルから

SEX	HEIGHT	WEIGHT
F	158	51
F	162	55
M	177	72
M	173	57
M	166	64

`data.frame()`



1つのデータフレームを作成

SEX	HEIGHT	WEIGHT
F	158	51
F	162	55
M	177	72
M	173	57
M	166	64

```
> sex      <- c("F", "F", "M", "M", "M")
> height   <- c(158, 162, 177, 173, 166)
> weight   <- c( 51,  55,  72,  57,  64)

> x <- data.frame(SEX=sex, HEIGHT=height, WEIGHT=weight)
```

データフレームの作成 (⇔ テキストファイル)



- 関数 `read.table()` などでテキストファイルからデータを読み込むことができる

```
> x <- read.table("data.txt",  
                  header=T, sep="," )
```

	sex	height	weight
1	F	158	51
2	F	162	55
3	M	177	72
4	M	173	57
5	M	166	64

```
> x <- read.csv("data.txt")
```

```
C:¥  
sex,height,weight  
F,158,51  
F,162,55  
M,177,72  
M,173,57  
M,166,64
```

data.txt

データフレームの作成 (⇔ 他のファイル形式)



- パッケージ **foreign** 等の中には，外部データを読み込むための関数が多数用意されている

関数	用途
<code>data.restore()</code>	S3 Binary ファイルの読み込み
<code>read.dbf()</code>	DBF ファイルの読み込み
<code>read.dta()</code>	Stata バイナリファイルの読み込み
<code>read.epiinfo()</code>	Epi Info ファイルの読み込み
<code>read.mtp()</code>	Minitab Portable Worksheet の読み込み
<code>read.octave()</code>	Octave テキストデータファイルの読み込み
<code>read.spss()</code>	SPSS データファイルの読み込み
<code>read.ssd()</code>	SAS データセット (.sas7bdat) の読み込み
<code>read.systat()</code>	Systat ファイルの読み込み
<code>read.xport()</code>	SAS XPORT ファイルの読み込み
<code>odbcConnect()*</code>	EXCEL, ACCESS, ... ファイルの読み込み

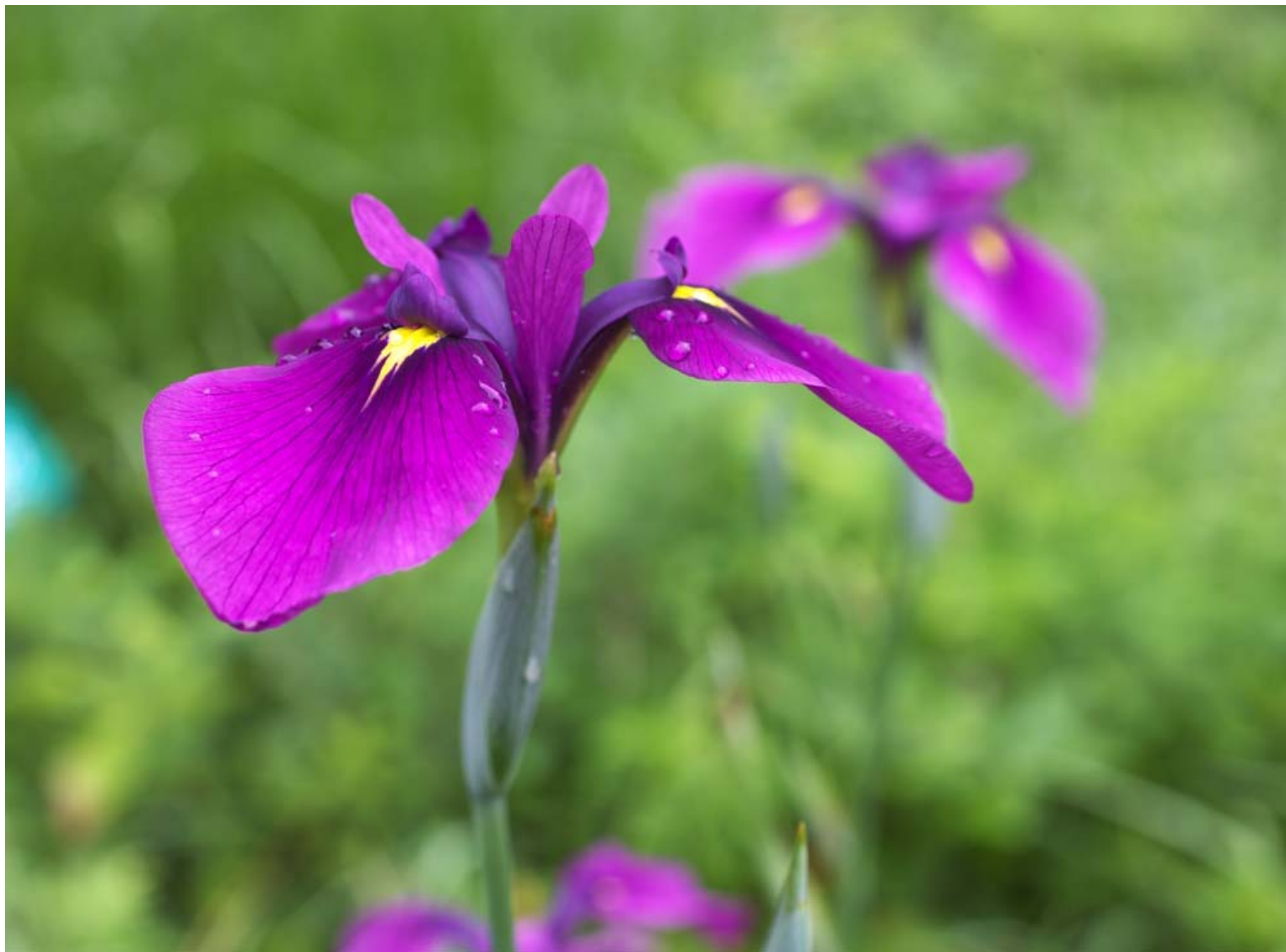
※ はパッケージ「RODBC」の関数，その他はパッケージ「foreign」の関数

本日のメニュー



- R の概要
- R でデータを読み込む方法
- グラフとデータマイニング ←
 - 1 変数の要約を見る
 - 2 変数の関係を見る
 - 層別解析
 - CART
- データマイニングとは？
- R Commander の紹介
- 質疑応答など

グラフとデータマイニング



Graphic by (c)Tomo.Yun (<http://www.yunphoto.net>)

グラフとデータマイニング ⇒ データ「iris」



Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
...

- フィッシャーが判別分析法を紹介するために利用したアヤメの品種分類 (Species: setosa, versicolor, virginica) に関するデータ
⇒ 以下の4変数を説明変数としてアヤメの種類を判別しようとした
 - アヤメのがくの長さ (Sepal.Length)
 - アヤメのがくの幅 (Sepal.Width)
 - アヤメの花弁の長さ (Petal.Length)
 - アヤメの花弁の幅 (Petal.Width)

グラフとデータマイニング ⇒ データ「iris」



Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
...

- データを眺めてもよく分からない... ⇒ **データを要約する!**

グラフとデータマイニング ⇒ 1変数の要約



Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
...

■ アヤメのがくの長さ (Sepal.Length) の特徴をつかむには...

1. 数値による要約 ⇒ 要約統計量を求める

2. グラフによる要約 ⇒ ヒストグラムを作成する

3. 層別して要約統計量やヒストグラム

グラフとデータマイニング ⇒ 要約統計量



```
> x <- iris # データ iris の読み込み
> summary(x$Sepal.Length) # 要約統計量を算出する
  Min. 1st Qu. Median Mean 3rd Qu. Max.
 4.300  5.100  5.800 5.843  6.400  7.900
```

一番小さい値
(最小値)

真ん中の値
(中央値, 平均値)

一番大きい値
(最大値)

- 要約統計量を眺めてもよく分からない... ⇒ データをグラフにする!

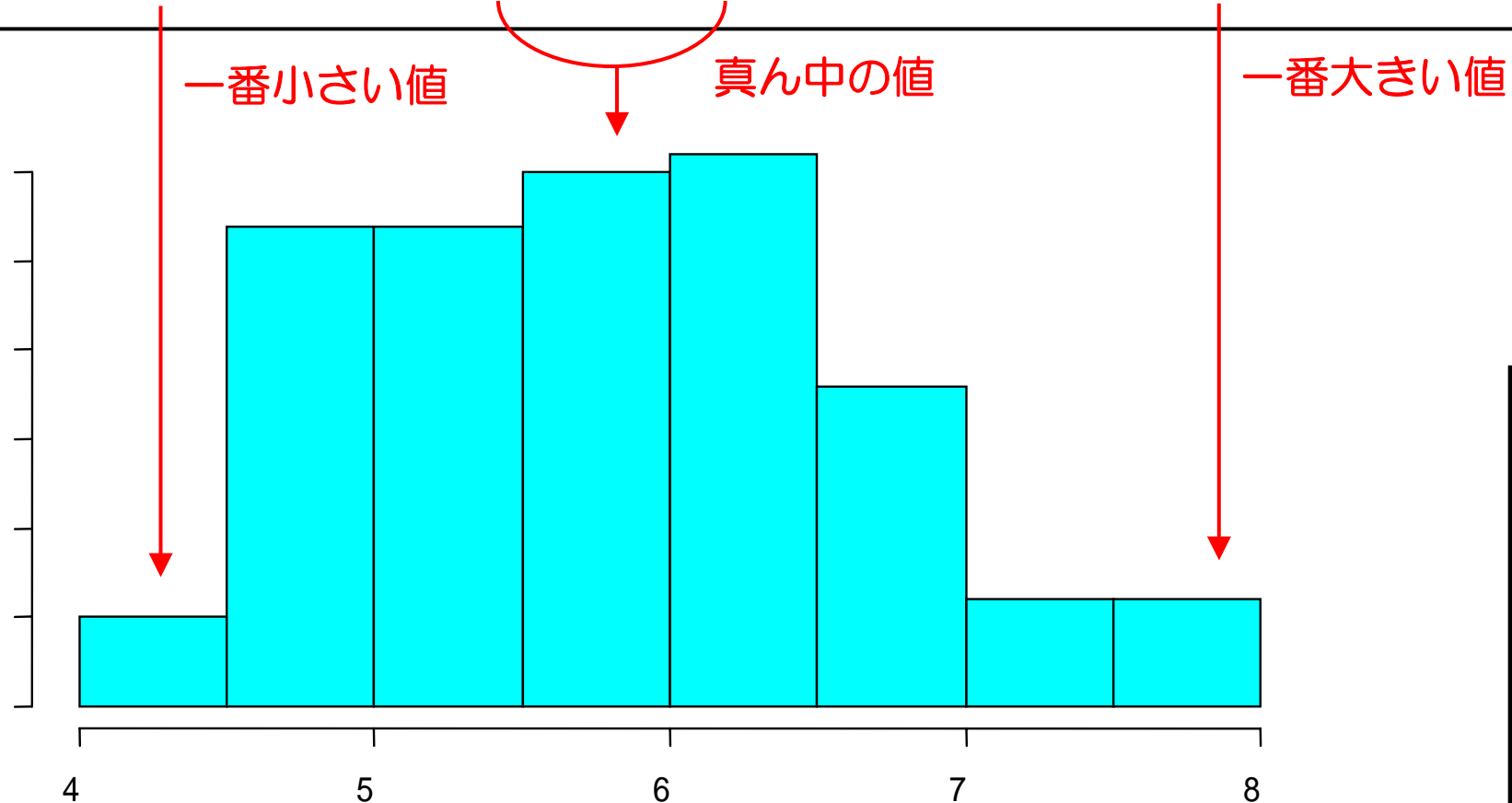
(注) 要約統計量を算出するのも大事なお仕事です^^;

グラフとデータマイニング ⇒ ヒストグラム



```
> x <- iris # データ iris の読み込み
> hist(x$Sepal.Length) # ヒストグラムを作成する
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
4.300	5.100	5.800	5.843	6.400	7.900

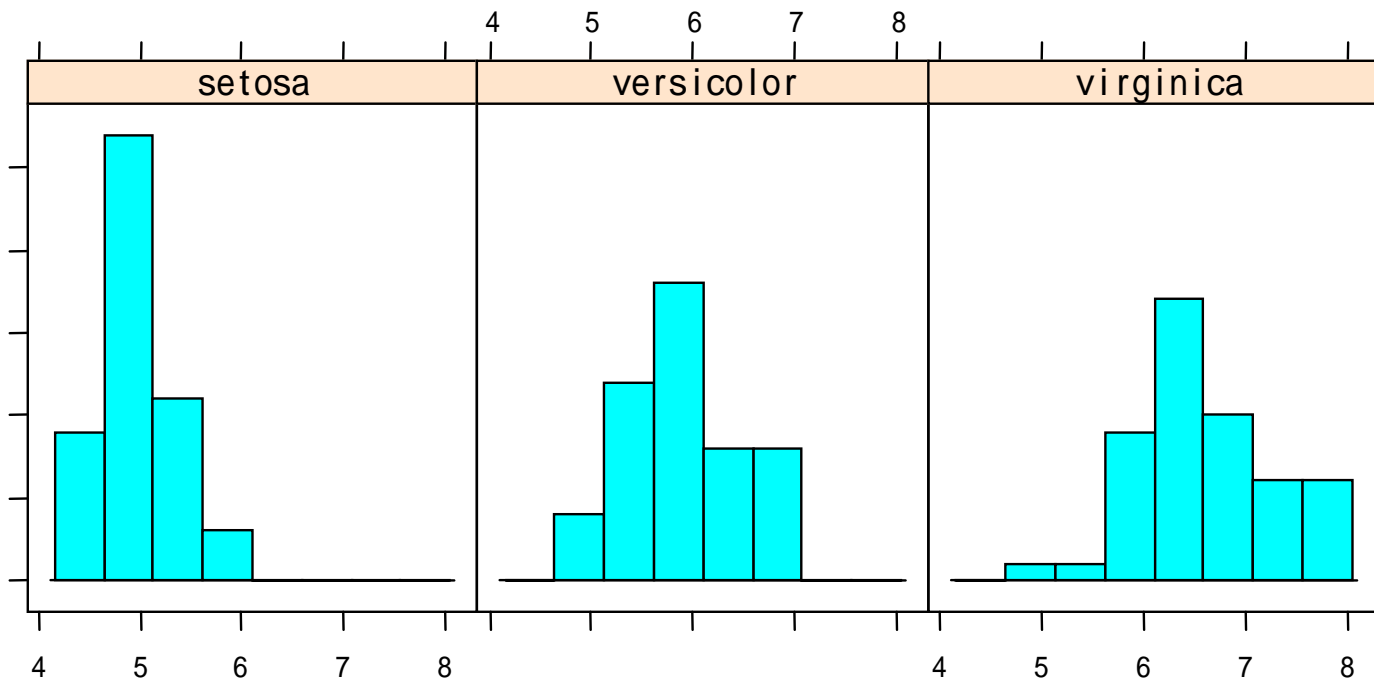


一目瞭然！

グラフとデータマイニング ⇒ 層別解析



```
> library(lattice)           # おまじない  
> hist(x$Sepal.Length)      # 層別ヒストグラム
```



- Setosa : がくが短い
 - Verginica : がくが長い
- ⇒ 層別すると特徴が浮き出る！

一目瞭然！

グラフとデータマイニング ⇒ 2変数の関係



Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
...

- アヤメの花弁の幅 (Petal.Width) と花弁の長さ (Petal.Length) の関係を見る場合は...
 1. 数値による要約 ⇒ 相関係数を求める
 2. グラフによる要約 ⇒ 散布図を描く
 3. 層別してグラフ (散布図) を描く

グラフとデータマイニング ⇒ 2変数の関係

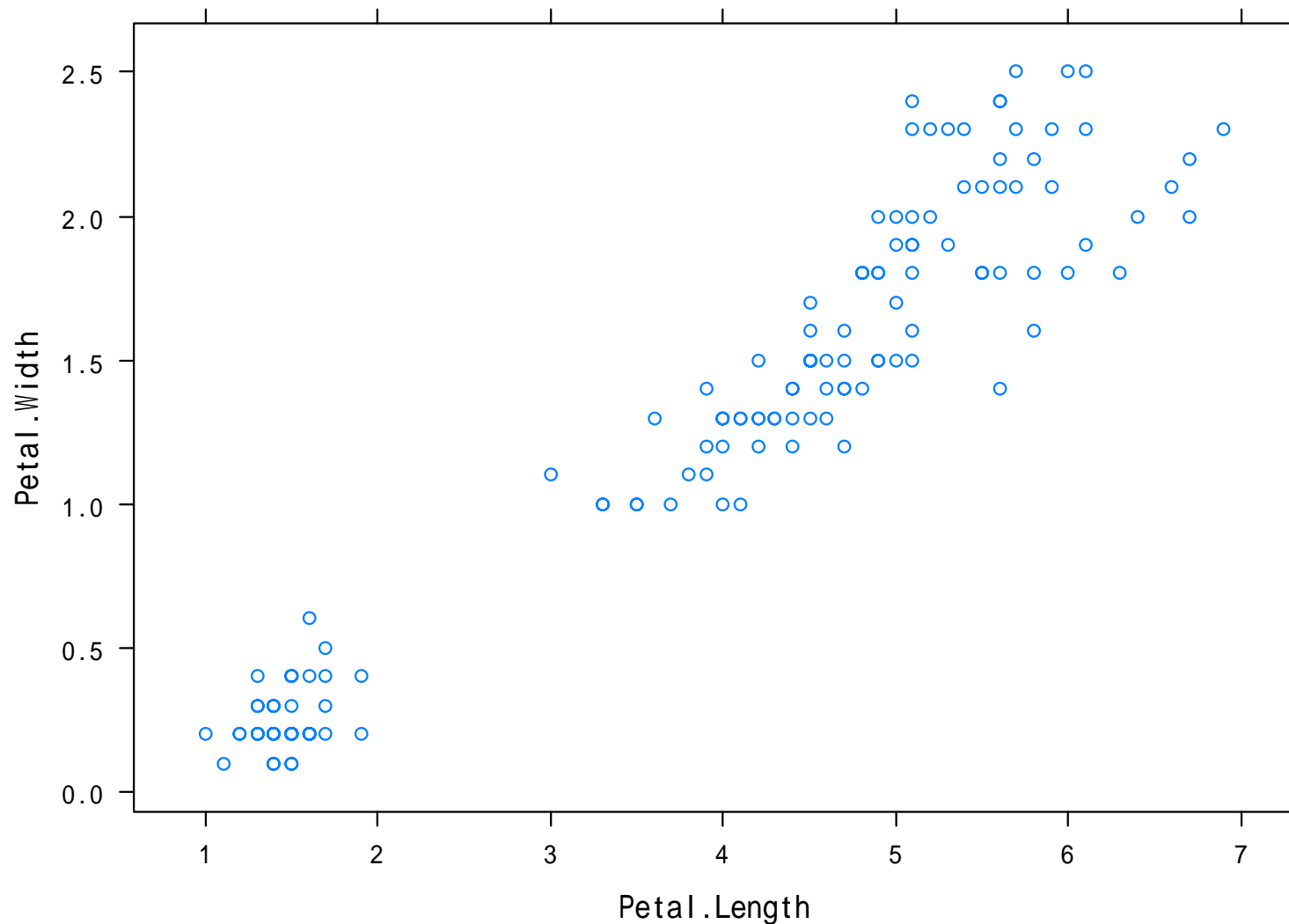


```
> # Petal.Width と Petal.Length の相関係数を算出する
> cor(Petal.Width, Petal.Length)
[1] 0.9628654
```

- 花卉の幅 (Petal.Width) と長さ (Petal.Length) の関係を調べるには
 1. 相関係数を算出する ⇒ よく分からない...
 2. グラフを描く ⇒ ちょっと分かる
 3. 層別にグラフを描く ⇒ 非常によく分かる!

```
> # Petal.Width と Petal.Length の散布図を描く
> library(lattice)
> xyplot(x$Petal.Width ~ x$Petal.Length, ps=20)
> # Petal.Width と Petal.Length の層別散布図を描く
> xyplot(x$Petal.Width ~ x$Petal.Length,
+       col=as.integer(Species), pch=as.integer(Species), ps=20)
```

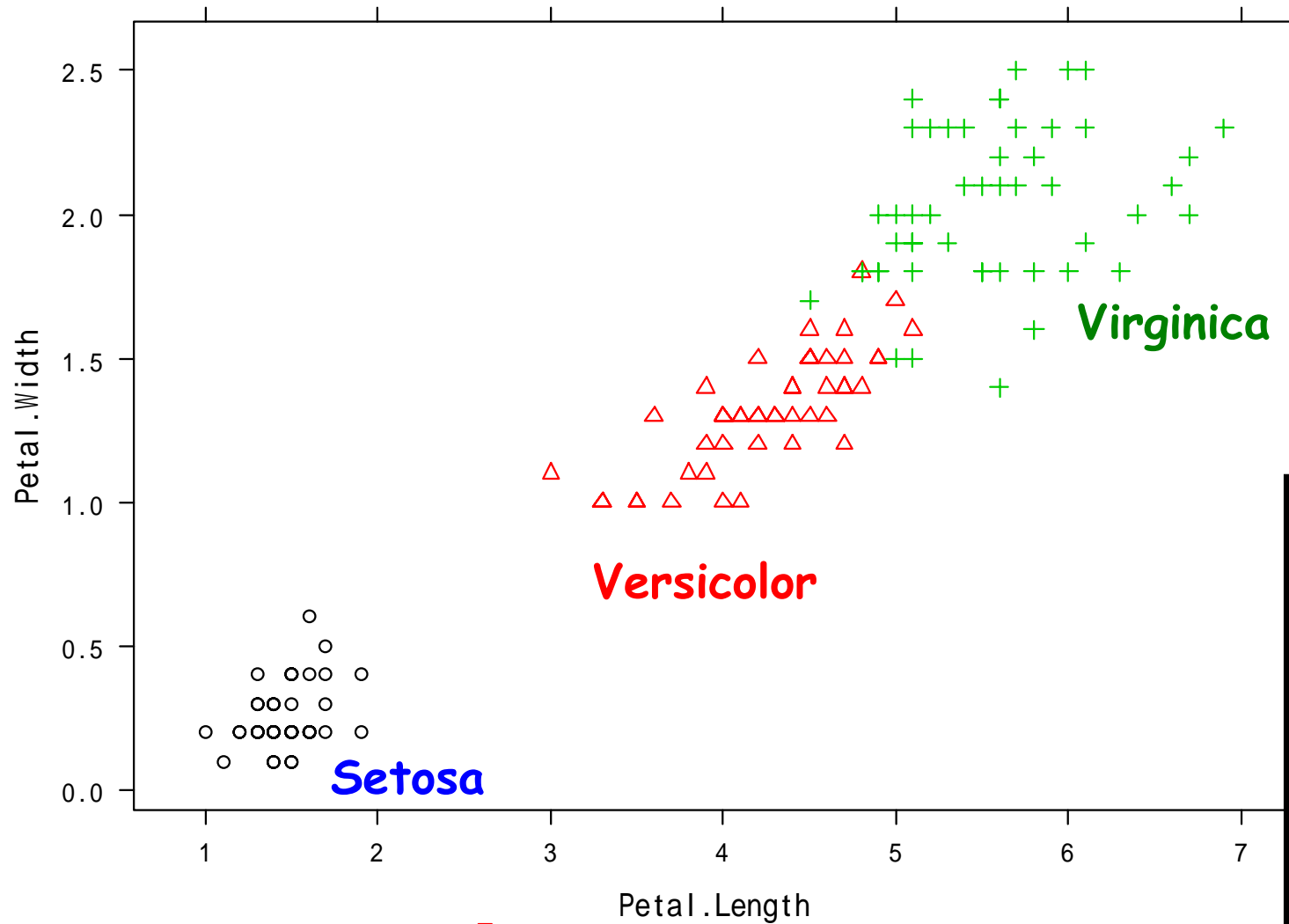
グラフとデータマイニング ⇒ 2変数の関係



- Petal.Width と Petal.Length の関係は右肩上がり ⇒ ひと目で分かる！

一目瞭然！

グラフとデータマイニング ⇒ 層別解析



- Setosa : 左下に分布
 - Verginica : 右上に分布
- ⇒ 層別すると特徴が浮き出る！

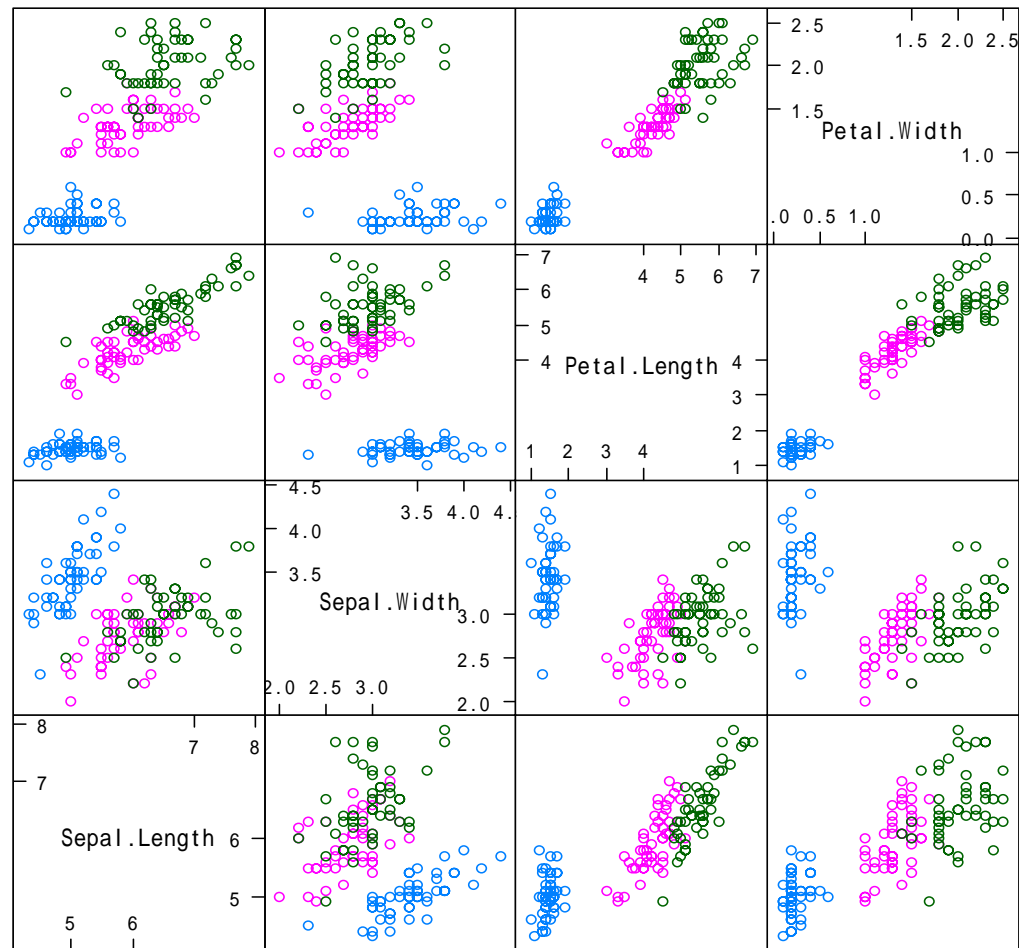
一目瞭然！

重要なのは・・・



層別解析!

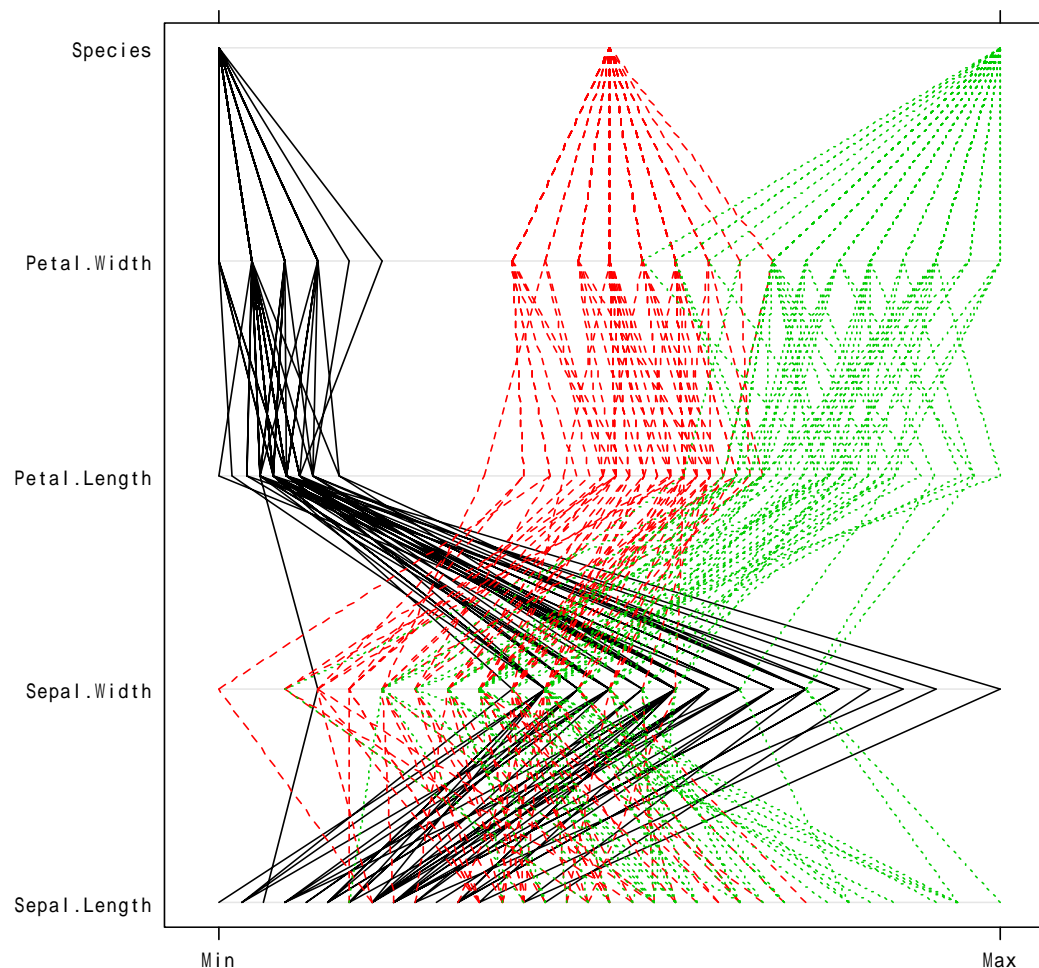
グラフとデータマイニング ⇒ その他



Scatter Plot Matrix

```
> library(lattice)
> splom(~ x[,1:4], groups=Species, data=x) # group=グループを表す変数
```

グラフとデータマイニング ⇒ その他



```
> library(lattice)
> parallel(x, col=as.integer(x$Species), # parallel(データフレーム名)
+         lty=as.integer(x$Species))
```

グラフとデータマイニング ⇒ その他



```
> x <- iris
> cor(x[, -5], method = "spearman") # 相関係数
      Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length  1.0000000 -0.1667777  0.8818981  0.8342888
Sepal.Width   -0.1667777  1.0000000 -0.3096351 -0.2890317
Petal.Length  0.8818981 -0.3096351  1.0000000  0.9376668
Petal.Width   0.8342888 -0.2890317  0.9376668  1.0000000

> symnum(abs(cor(x[-5])), cutpoints=c(0, 0.2, 0.4, 0.6, 0.8, 1),
+        symbols=c(" ", ".", "-", "+", "*")) # 相関係数
      S.L S.W P.L P.W
Sepal.Length *      *      *
Sepal.Width   *      -      .
Petal.Length *      -      *      *
Petal.Width  *      .      *      *
attr(,"legend")
[1] 0 ' ' 0.2 '.' 0.4 '-' 0.6 '+' 0.8 '*' 1
```


CART (Classification and Regression Trees) について



Graphic by (c)Tomo.Yun (<http://www.yunphoto.net>)

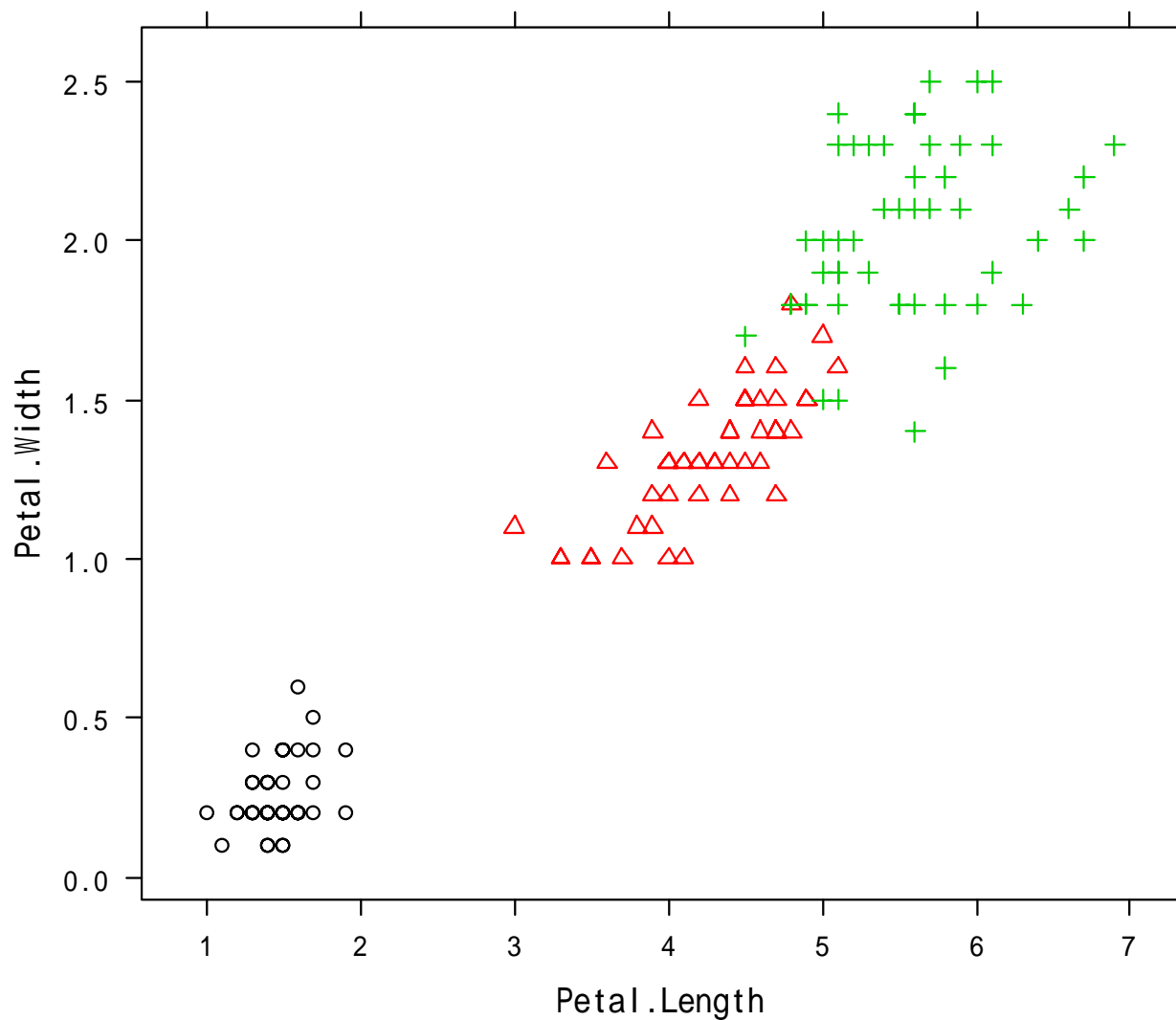
グラフとデータマイニング ⇒ 分類・予測



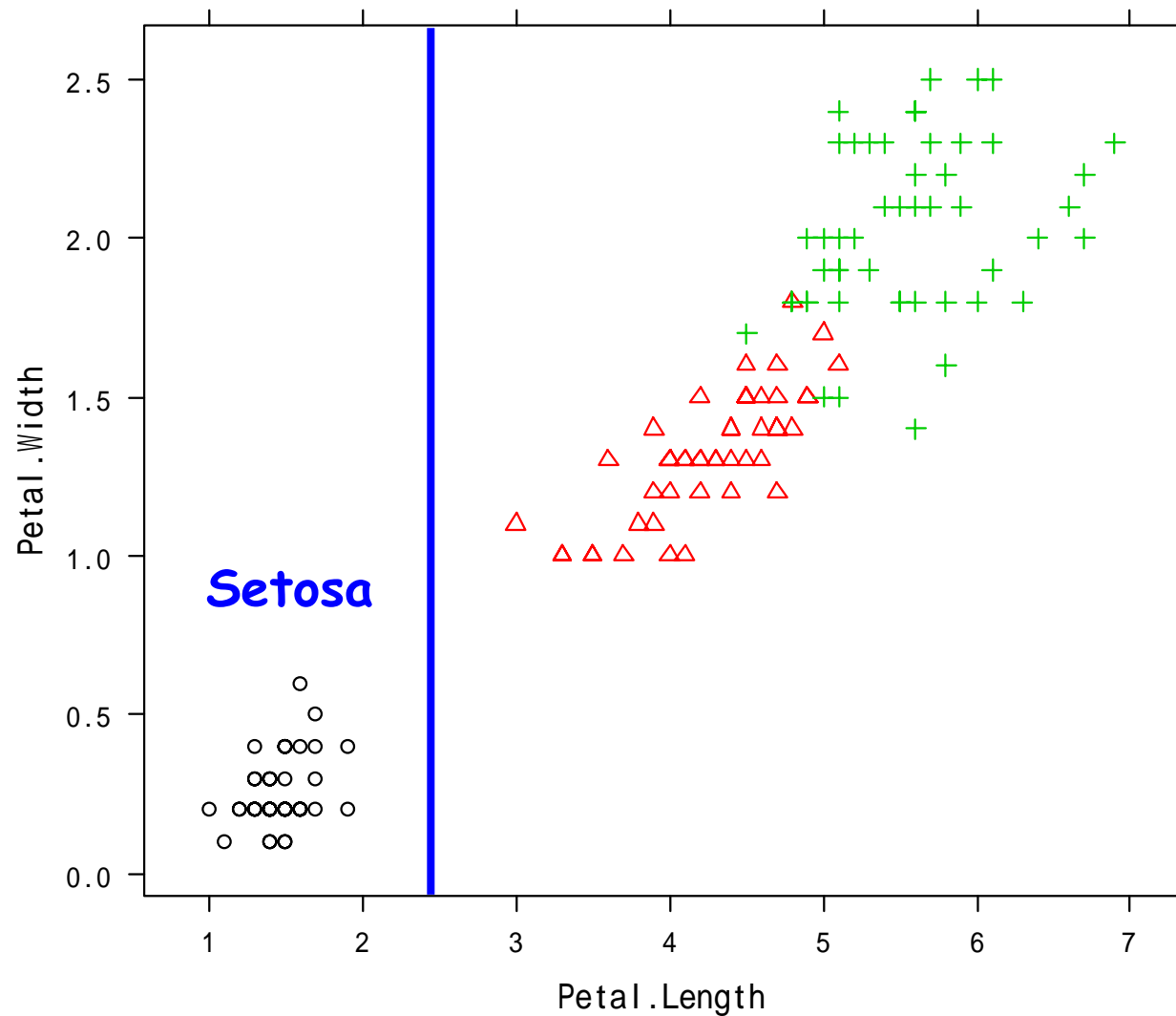
Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
...

- ここまでは「アヤメの種類 (Species) で層別するのが重要！」というお話でした
- 今度は逆に「他の変数からアヤメの種類を予測する」ことは出来る？
- 例えば「花弁の長さ (Petal.Length) が●●以下ならば setosa 」のような分類ルールを作ることは出来る？

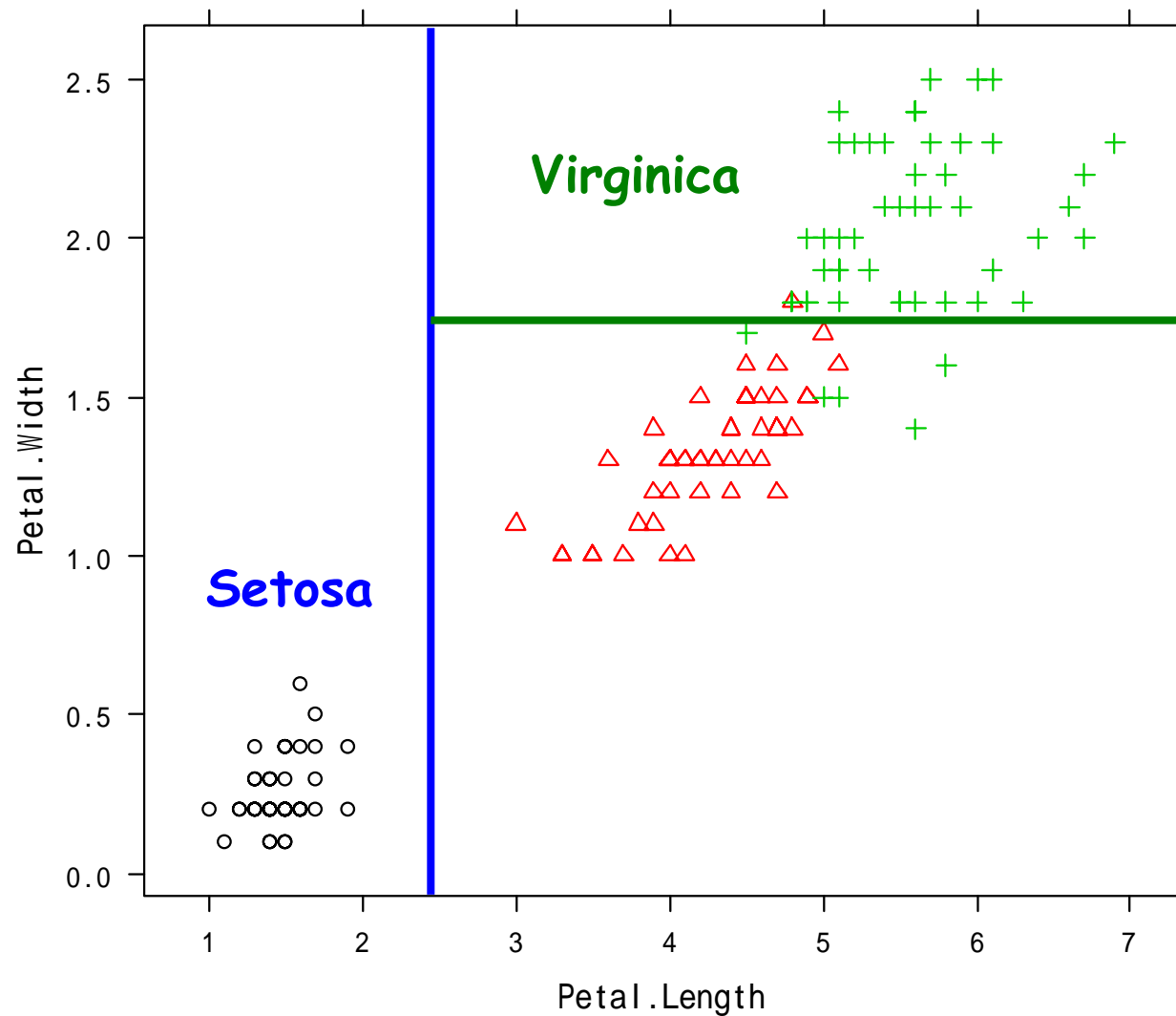
グラフとデータマイニング ⇒ 分類・予測



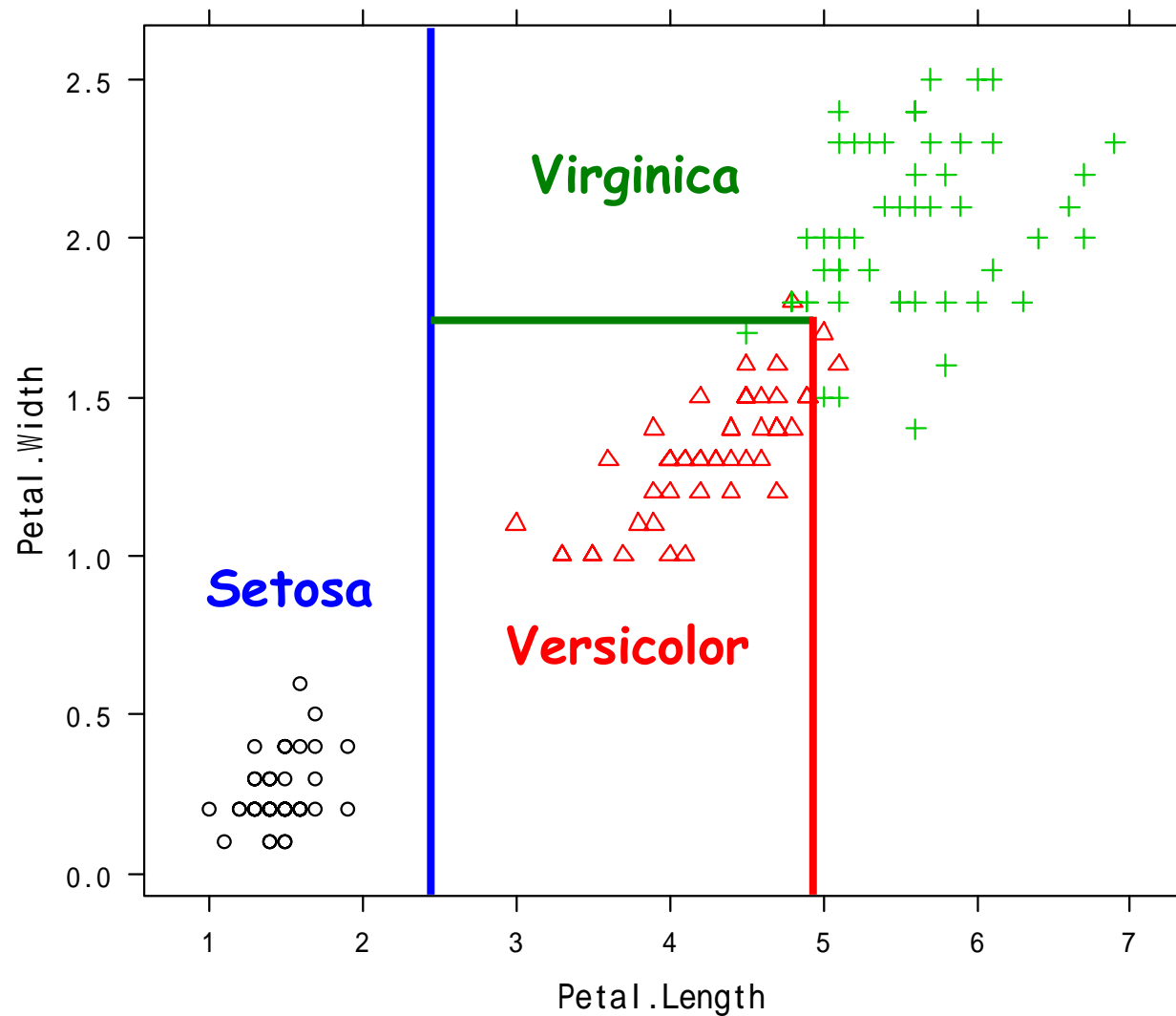
グラフとデータマイニング ⇒ 分類・予測



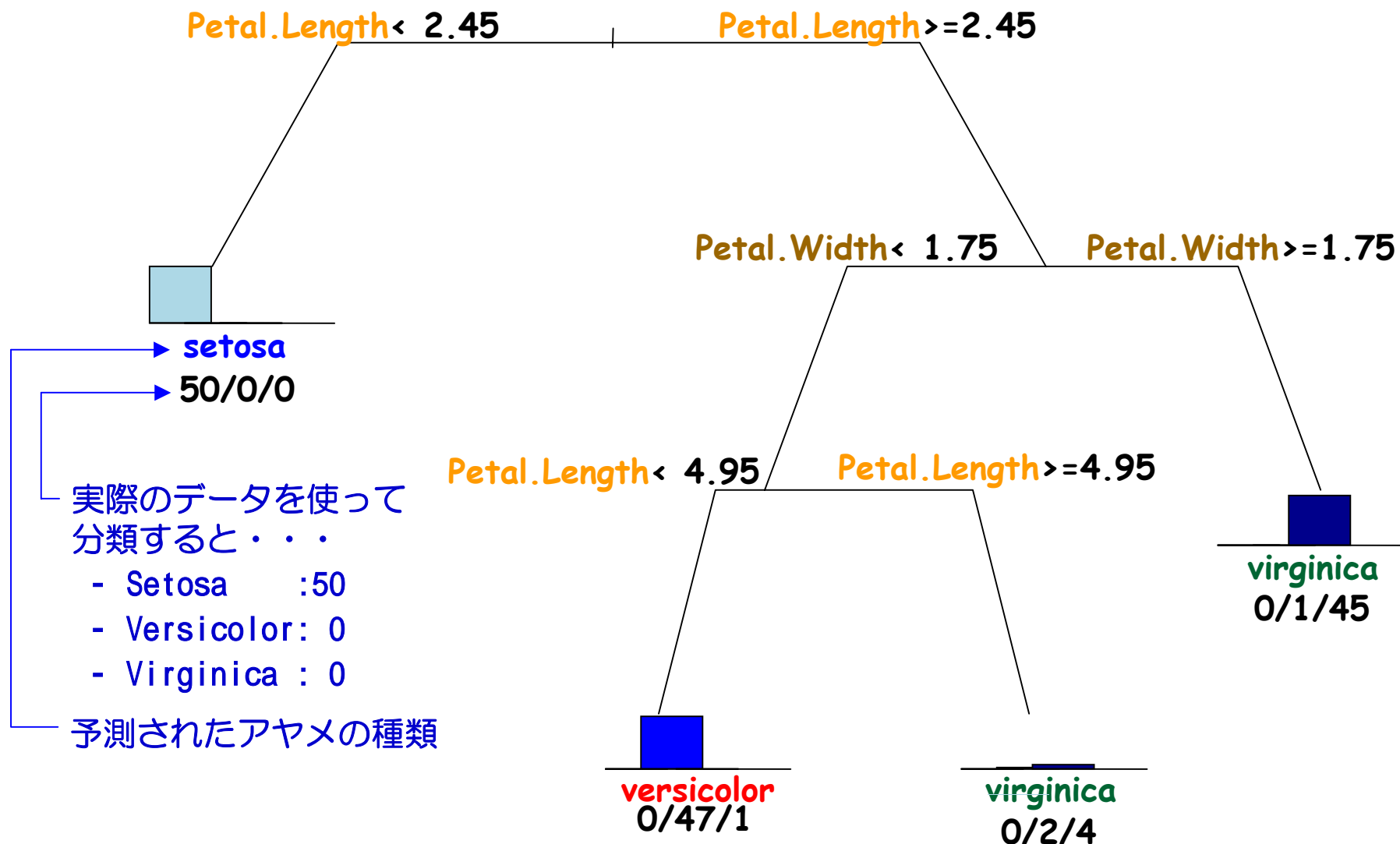
グラフとデータマイニング ⇒ 分類・予測



グラフとデータマイニング ⇒ 分類・予測



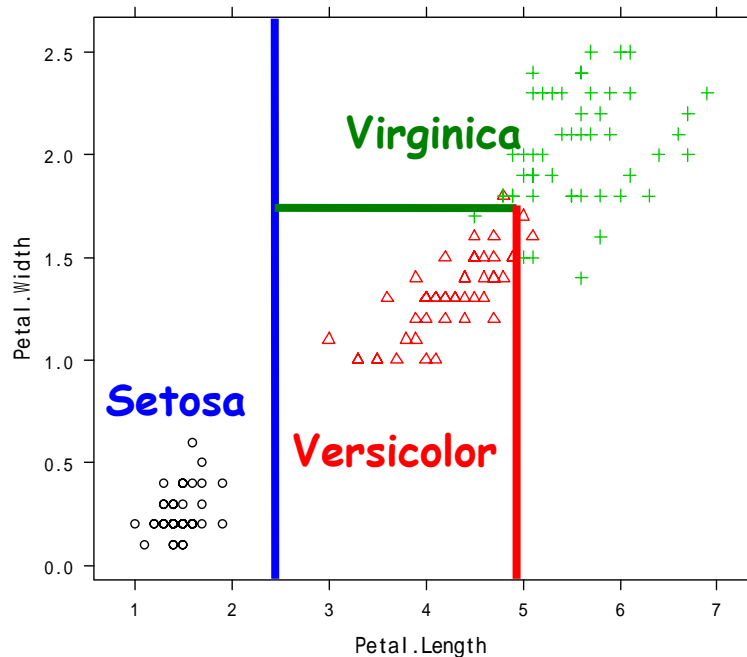
前頁のスライドをルール化 ⇒ 分類木 (CART)



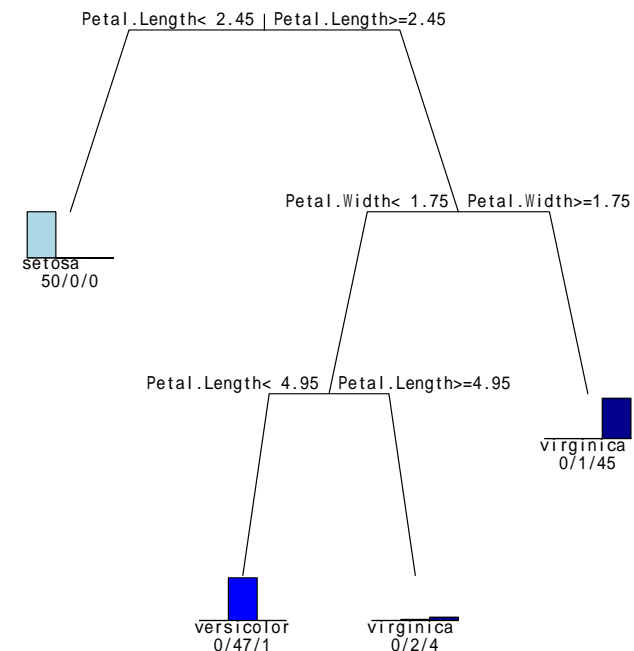
CART (Classification and Regression Trees) とは?



- あるルールに従ってデータを分け，分類や予測を行う
- 目的変数がカテゴリ：分類木 ← この iris の例！
- 目的変数が連続変数：回帰木 ← 次のスライド



ルール化！
→
(分類・予測)



```
> library(mvpart)
> result <- rpart(Species ~ ., data=x)
> plot.new() # デバイスを開く
```

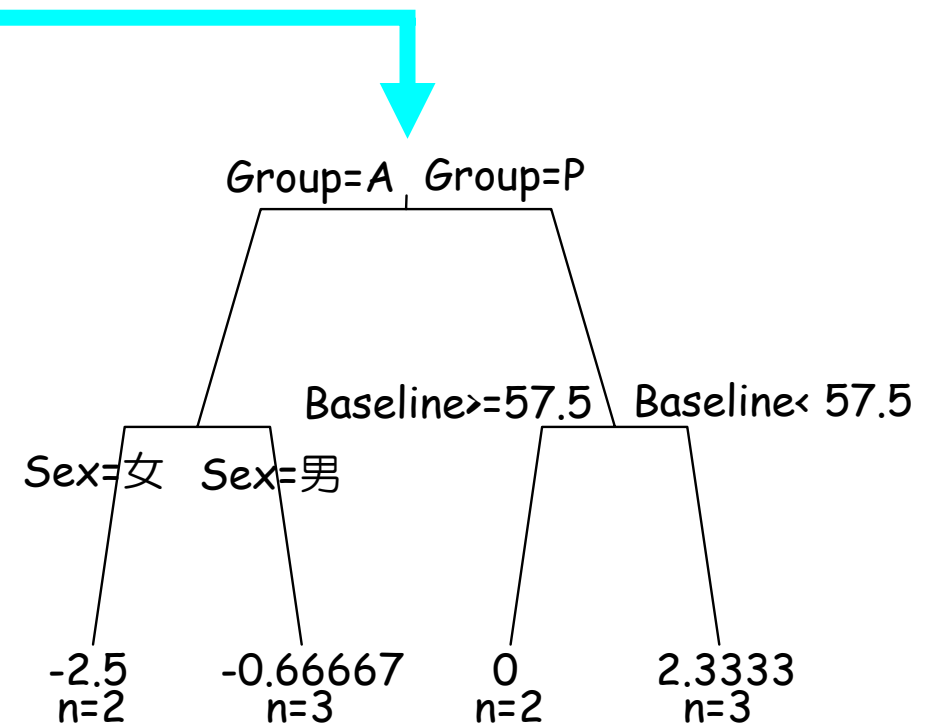
```
> par(xpd=T) # 文字が切れないための命令
> plot(result, uniform=T, branch=0.7)
> text(result, use.n=T, all.leaves=F)
```




回帰木の例

- 男女 10 人にやせ薬 (A) と偽薬 (P) を飲んでもらう
- 飲みはじめの体重を測り, 1 ヶ月飲み続けた後, 再度測定
- 体重の変化量 (kg) を目的変数として回帰木を作成

Difference (体重の変化)	Group (薬剤)	Sex (性別)	Baseline (前値)
1	A	Man	55
-1	A	Man	65
-2	A	Man	70
-2	A	Woman	45
-3	A	Woman	50
3	P	Woman	50
2	P	Woman	55
2	P	Woman	55
1	P	Man	60
-1	P	Man	65



(各カテゴリの平均値と例数が表示される)

CART 以外の分析・分類・予測手法



■ 回帰分析（共分散分析・ロジスティック回帰）

お気楽に，ざっくりと分析する場合・・・

■ ニューラル・ネットワーク

非線形的なデータ構造にも対応，分類ルール不明

■ サポートベクターマシン

CART のような直線的な分類だけではなく，曲線的な分類が出来る

「判別」が目的なので，説明変数が目的変数に対してどのような影響を与えているかは不明

■ 集団学習（バギング，ランダムフォレストなど）

□ **バギング**：多数の CART を生成して多数決（分類の場合）

or 平均（回帰の場合）

□ **ランダムフォレスト**：与えられたデータからいくつかの変数を

ランダム抽出し（ブートストラップサンプル）CART を作成

⇒ 多数回くり返して多数決（分類の場合）or 平均（回帰の場合）

■ アソシエーション・ルール（相関分析）：ビールと紙おむつ

本日のメニュー



- R の概要
- R でデータを読み込む方法
- グラフとデータマイニング
- データマイニングとは? ←
 - データマイニングとは
 - 医薬品医療機器総合機構の例
 - シグナルとシグナル検出
- R Commander の紹介
- 質疑応答など

データマイニングとは？



- データマイニングをそのまま日本語に訳すと
「データから宝石（鉱物）を掘り当てる」となる
⇒ データという鉱山から「統計的な手法」などを施すことで
「宝石」を見つけるのが目的
- 「宝石」とは, データが巨大であるがゆえに古典的な
統計手法を施すだけでは発見できなかったような
「有益なルール」「意外なパターン」「意思決定に
繋がる情報」のことをさす
⇒ 例えば **CART** の場合は「分類ルール」が「宝石」となる
- 間違っているかもしれないが, 個人的には「グラフ化」や
「層別解析」も立派なデータマイニングだと思う

データから情報発掘

三菱総研など 経験・勘頼み改善

大量のデータから有用な情報を発掘するデータマイニングの手法を使って医薬品の副作用を見つけた。新システムを、厚生労働省所管の独立行政法人・医薬品医療機器総合機構と三菱総合研究所が開発した。今年夏に試験運用を始める。担当者の目に頼る現在のやり方を改善し、複数の薬の飲み合わせなどによる副作用の被害を事前に把握するのに役立つ。


データマイニングはビジネス分野で普及している。この手法でスーパーの販売データを分析、紙おむつと缶ビールを同時に購入する顧客が多いことが分かり、同じ棚に並べたら販売量が増えたなどの成功例が知られている。

薬事法に基づいて同機構が製薬会社や医療機関から集めた副作用情報を利用する。新システムでは、医薬品の分子構造や主成分、患者の病状、年齢層、性別、副作用の症状といった情報から、因果関係のありそうな医薬品と副作用の組み合わせなどを見つける。性別・年齢層に特有な副作用や複数の医薬品の併用による副作用を事前に発見するのに役立つと期待している。

試作したシステムを業務で使えるように改良する作業を進め、今年夏に試験運用を開始。一年間ほど改良点などを調べ、二〇〇九年度に業務で使う予定だ。

同機構に集まる副作用情報は年間で約二万五千件。海外からの情報も含めると六万―七万件ある。これまでは十人ほどの担当者が情報を選別しているが、経験と勘に基づいている現状が改善されるという。

医薬品医療機器総合機構って？



独立行政法人
PMDA 医薬品医療機器総合機構 Pharmaceuticals and Medical Devices Agency

<http://www.pmda.go.jp/>

より有効で、より安全な医薬品や医療機器をより早く皆様へ提供するとともに、
医薬品の副作用などによる健康被害に対して迅速な救済を行います。

- 日本の厚生労働省が管轄している独立行政法人
- 医薬品・医療機器の「審査関連業務」や「安全対策業務」などを行っている
- 「新薬の候補」が出来た時に「薬として認めて下さい」と申請するところ

参考文献（医薬品医療機器総合機構 HP より）



1. 平成18年度データマイニング手法の導入に関する検討概要について
http://www.info.pmda.go.jp/kyoten_iyaku/file/dm-gaiyou18.pdf
2. データマイニング手法の導入に関する検討結果報告書（平成19年3月）
http://www.info.pmda.go.jp/kyoten_iyaku/file/dm-report18.pdf
3. 平成17年度データマイニング手法の導入に関する検討概要について
http://www.info.pmda.go.jp/kyoten_iyaku/file/dm-donyu1807.pdf
4. データマイニング手法の検討を行うための支援業務報告書（平成18年3月）
http://www.info.pmda.go.jp/kyoten_iyaku/file/data060728.pdf
5. データマイニング手法の導入に関する検討内容について
http://www.info.pmda.go.jp/kyoten_iyaku/file/dm-donyu.pdf
6. データマイニング手法の検討を行うための支援業務報告書（平成17年3月）
http://www.info.pmda.go.jp/kyoten_iyaku/file/data050720.pdf

データマイニングとは？（三菱総研）



- 参考文献 6 では，ある手法が「データマイニング」であるかどうかを判断する基準を紹介している（以下引用）

判別基準	内容
【定義】 未知かつ有用な知識を発見できている	結果として業務効率，成果が向上している
【データ量】 大量のデータを分析している	人手では見切れないようなデータを分析している
【手法】 データマイニング手法（ツール）を使っている	決定木，ニューラルネット，相関分析を使用している，あるいは市販ソフトを利用している
【自動的】 半自動的な分析により知識（ルール）が抽出されている	データをツール（ソフト）に入れると知識が抽出される



シグナルとシグナル検出

■ シグナル：

それまで知られなかったか、不完全にしか証拠付けられていなかった有害事象と薬との因果関係に関する情報
(WHO の定義)

⇒ データマイニングの「宝石」にあたるもの！

■ シグナル検出：

「詳細な調査を必要とする自発報告の発見およびその優先順位付けを行うこと」

(国立保健医療科学院 藤田利治先生の定義)

■ 医薬品医療機器総合機構では「ROR」「GPS」

「BCPNN」などをシグナル検出の指標とする予定

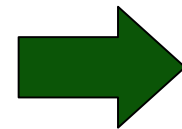
⇒ 時間の関係上、本日は「ROR」のみ紹介・・・

ROR (Reporting Odds Ratios) とは？



	副作用 1	...	副作用 k	合計
医薬品 1	n_{11}	...	n_{1k}	n_{1k}
:	:		:	:
医薬品 m	n_{m1}	...	n_{mp}	n_{1k}
合計	n_{+1}		n_{+k}	n_{++}

クロス表
の作成



	副作用 A	副作用 (その他)	合計
医薬品 B	n_{11}	n_{12}	n_{1+}
医薬品 (その他)	n_{21}	n_{22}	n_{2+}
合計	n_{+1}	n_{+2}	n_{++}

※ 赤枠と青枠の非比例性を検討する

- $ROR = (n_{11} / n_{12}) \div (n_{21} / n_{22})$
- $95\%CI = \exp\{ \ln(ROR) \pm 1.96(1/n_{11} + 1/n_{12} + 1/n_{21} + 1/n_{22})^{1/2} \}$
- シグナルあり！ = 95%CI の下限が 1 より大きい場合
- 併用薬との関係を考慮した上での算出も？
- 性別や年齢別などの「層別」によるシグナル検出も予定？

本日のメニュー



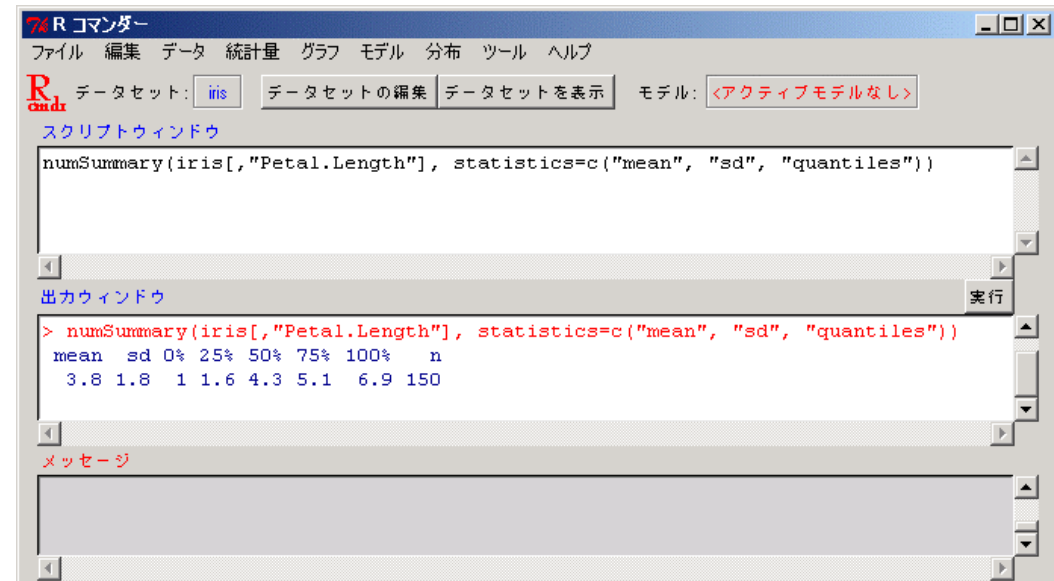
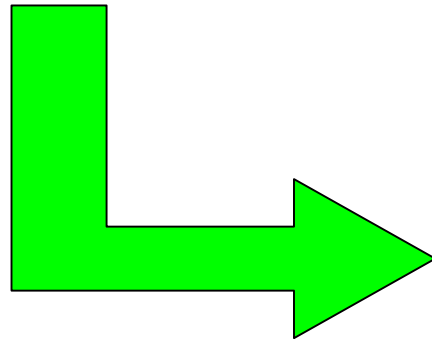
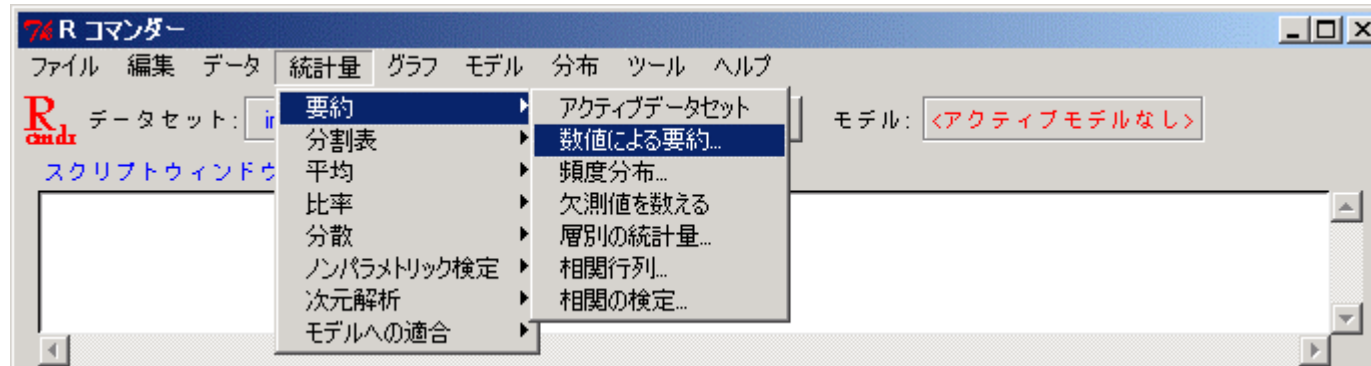
- R の概要
- R でデータを読み込む方法
- グラフとデータマイニング
- データマイニングとは？
- **R Commander の紹介 ←**
 - ★ R をコマンド入力で使いたくない方向け
 - R Commander とは？
 - R Commander の機能紹介
- 質疑応答など

R Commander とは？



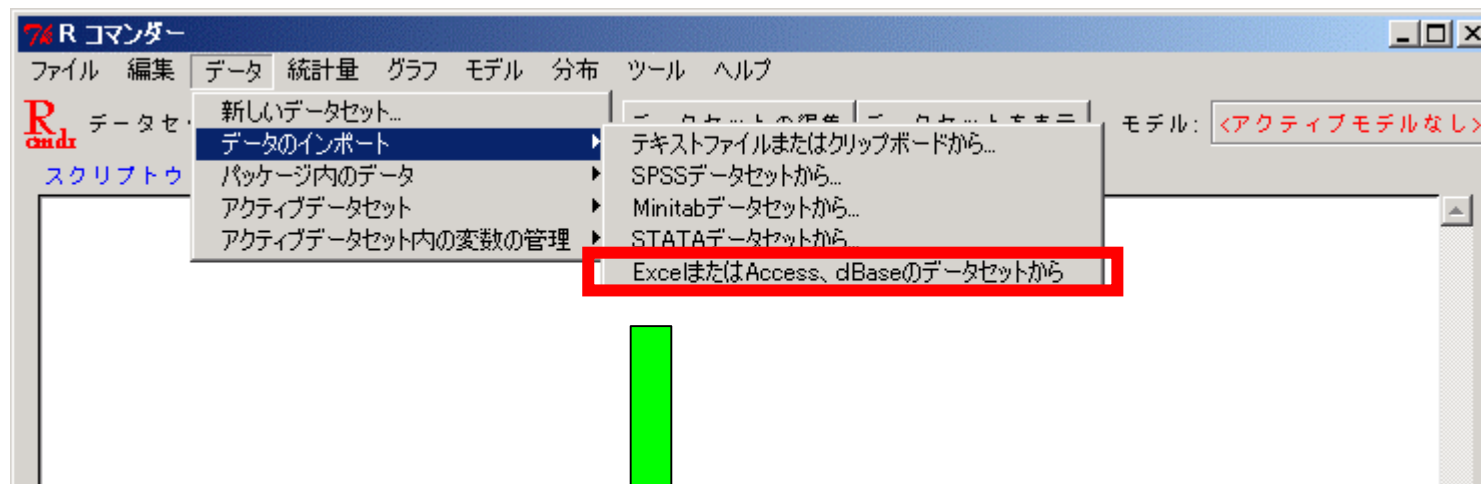
- ここまで紹介した解析は、全て **R** で実行できます！
- が、GUI（マウス操作）の機能が劣っており、**R** の命令をひとつひとつ覚えなければいけません...（⇒ **R** の短所）
- そこで **R Commander** の登場！マウス操作で **R** が使えます！（**R** の命令を覚えなくても **R** の出力が得られます！）
- インストール方法は「[今日からあなたも統計ソフト開発者](#)」で Google 検索 →
- 使用方法は「[R Commander](#)」で Google 検索 ⇒ 私の解説 PDF あり
- より詳しい情報は「**R Commander** ハンドブック（舟尾；九天社）」をご覧ください♪

R Commander とは？



- ① メニューから機能を選択
- ② スクリプトウィンドウには実行した R のコマンドが出力される
- ③ 出力ウィンドウには、実行結果が出力される
- ④ メッセージにはエラーや警告が出力される

R Commander の機能紹介（データの読み込み）

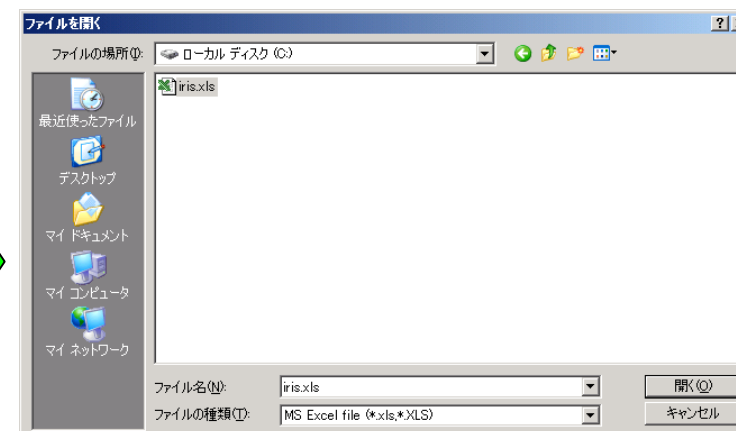


EXCELやAccessファイルの
データを読み込むときは、

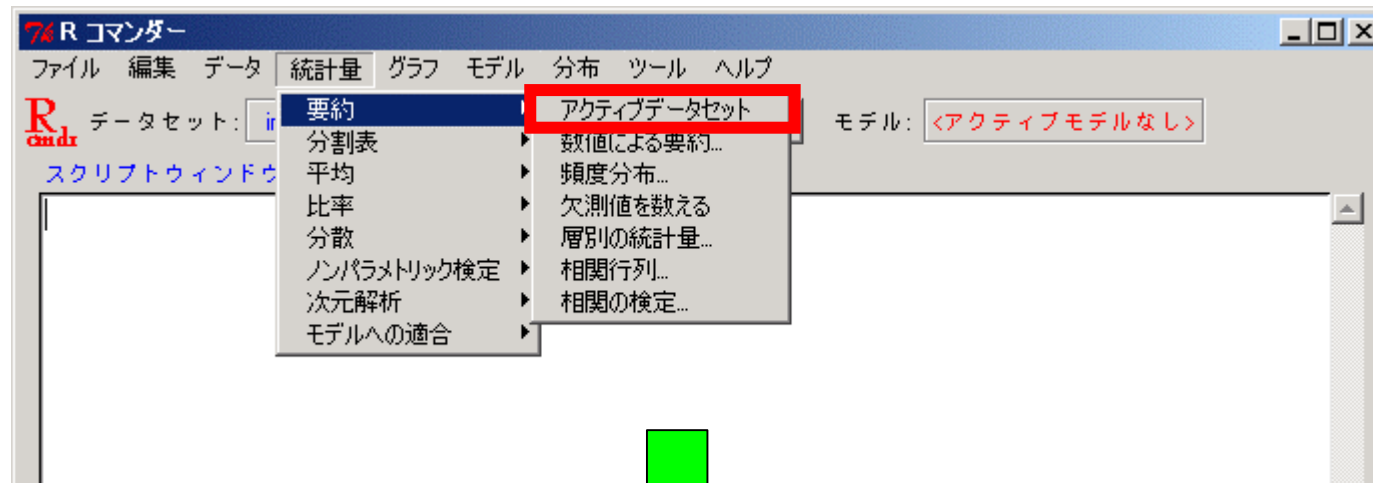
- ① 上記画像のメニューを
選択してデータセット名を入力



- ② 読み込むファイル
を選択する



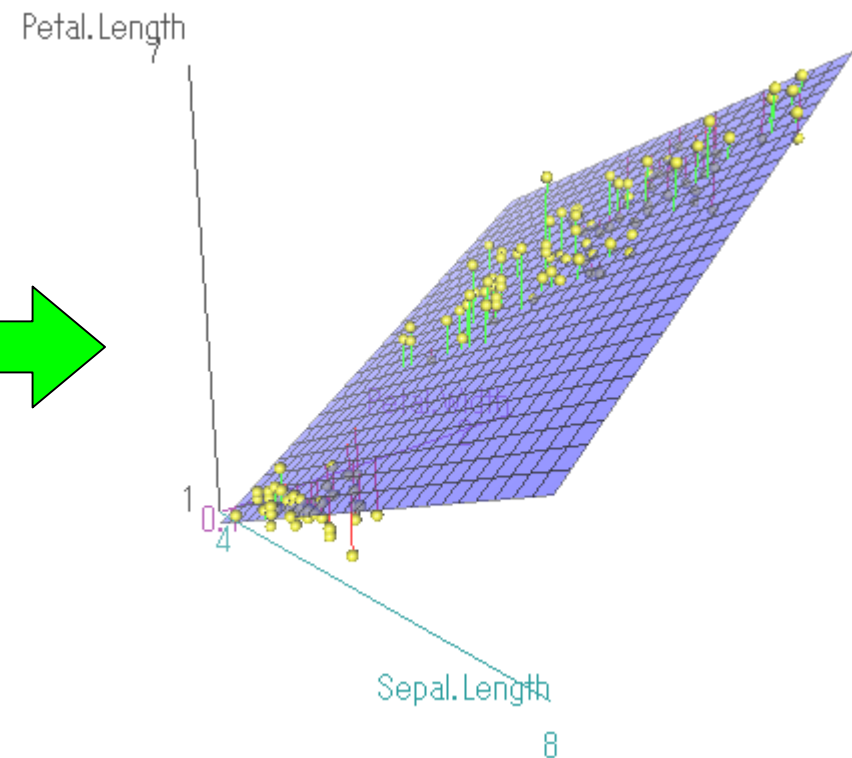
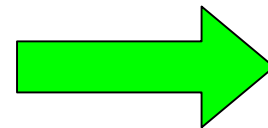
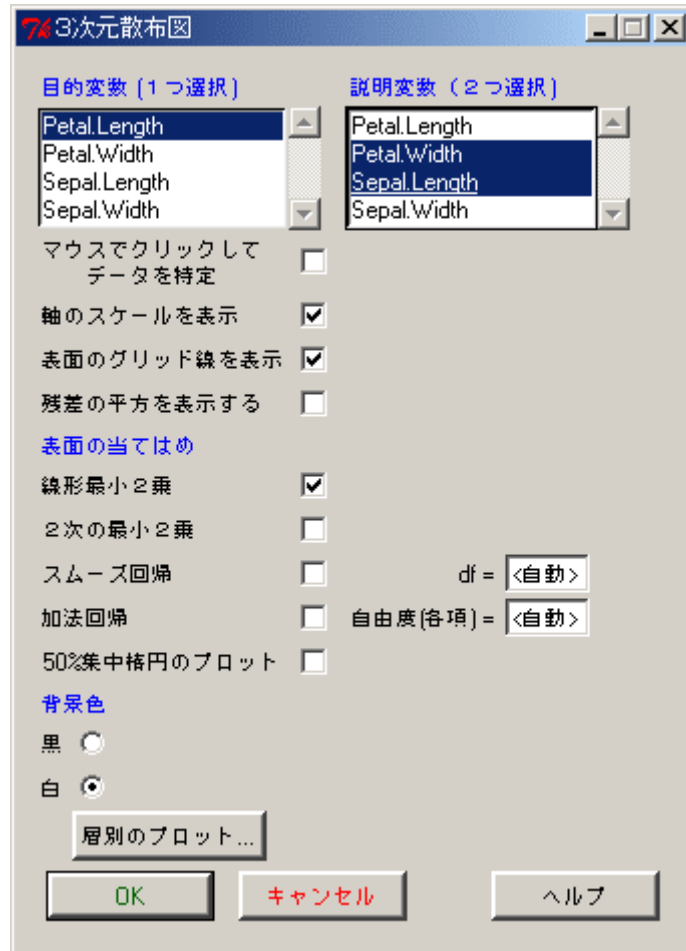
R Commander の機能紹介 (要約統計量)



データセットの要約統計量が出力される

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
Min. :4.30	Min. :2.00	Min. :1.00	Min. :0.1	setosa :50
1st Qu.:5.10	1st Qu.:2.80	1st Qu.:1.60	1st Qu.:0.3	versicolor:50
Median :5.80	Median :3.00	Median :4.35	Median :1.3	virginica :50
Mean :5.84	Mean :3.06	Mean :3.76	Mean :1.2	
3rd Qu.:6.40	3rd Qu.:3.30	3rd Qu.:5.10	3rd Qu.:1.8	
Max. :7.90	Max. :4.40	Max. :6.90	Max. :2.5	

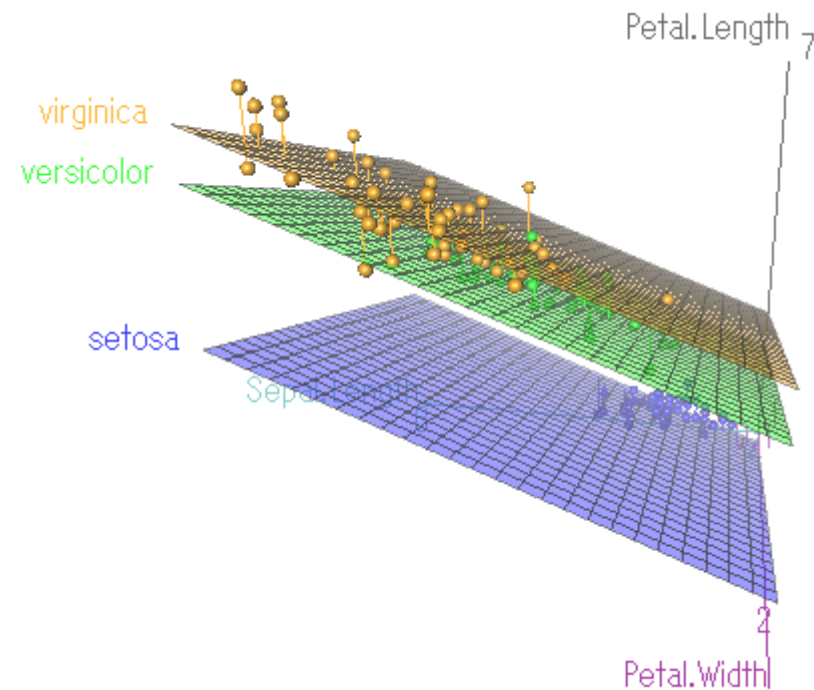
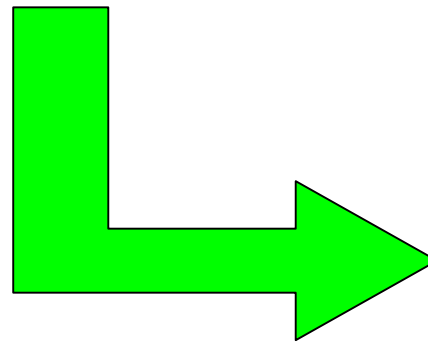
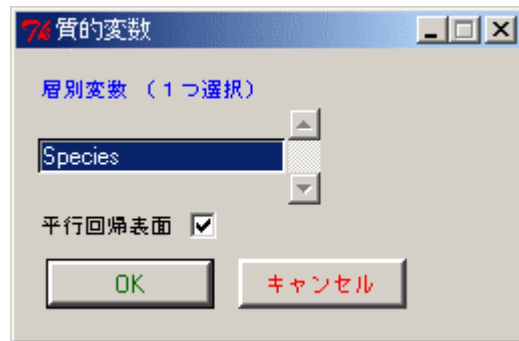
R Commander の機能紹介 (3Dグラフ)



目的変数を1つ，説明変数を2つ指定する

マウスでグラフを動かすことができる！

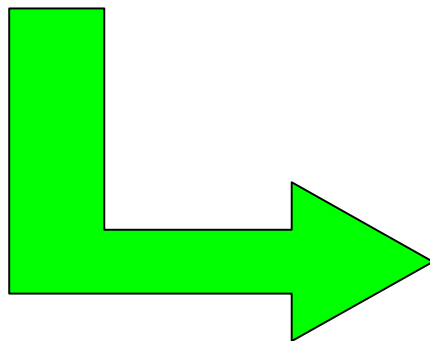
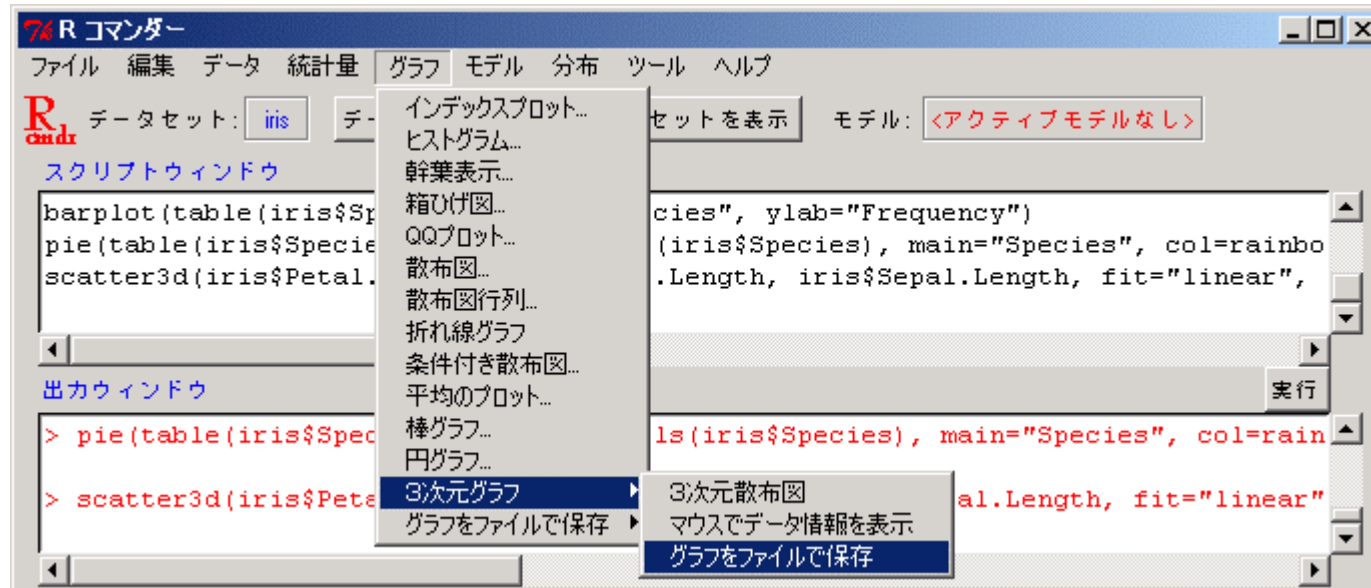
R Commander の機能紹介 (層別化)



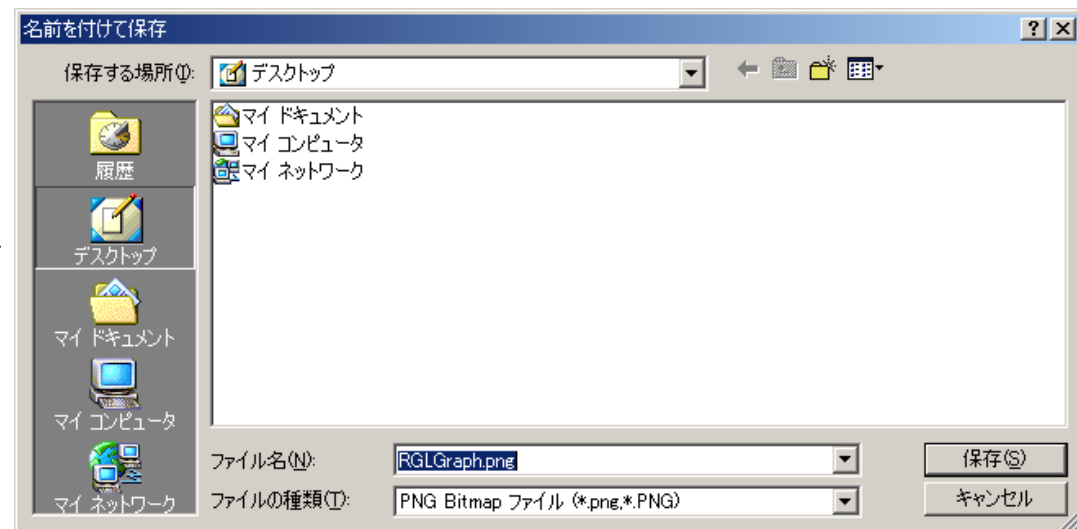
前のページの画面で「層別のプロット」を選択することで
カテゴリ変数で層別したグラフを出力

マウスでグラフを動かすことができる！

R Commander の機能紹介 (グラフの保存)



描いたグラフを保存
することが出来る



参考文献



- 「よくわかる多変量解析の基本と仕組み」
山口和範, 高橋淳一, 竹内光悦 (秀和システム 2004)
- 「R によるデータサイエンス」 金 民哲 (森北出版 2007)
- 「グラフはこう読む! 悪魔の技法」 牧野 武文 (三修社 2005)
- 「医薬品医療機器総合機構 HP」の資料 (スライド 39 頁に一覧)
- 「臨床試験ハンドブック 21章」 丹後 俊郎, 上坂 浩之
(朝倉書店 2006)
- 「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」
舟尾 暢男 (九天社 2005)
- 「R で学ぶデータマイニング 1」 熊谷悦生, 舟尾暢男 (九天社 2007)
- 「R Commander ハンドブック」 舟尾 暢男 (九天社 2007)

本日のメニュー

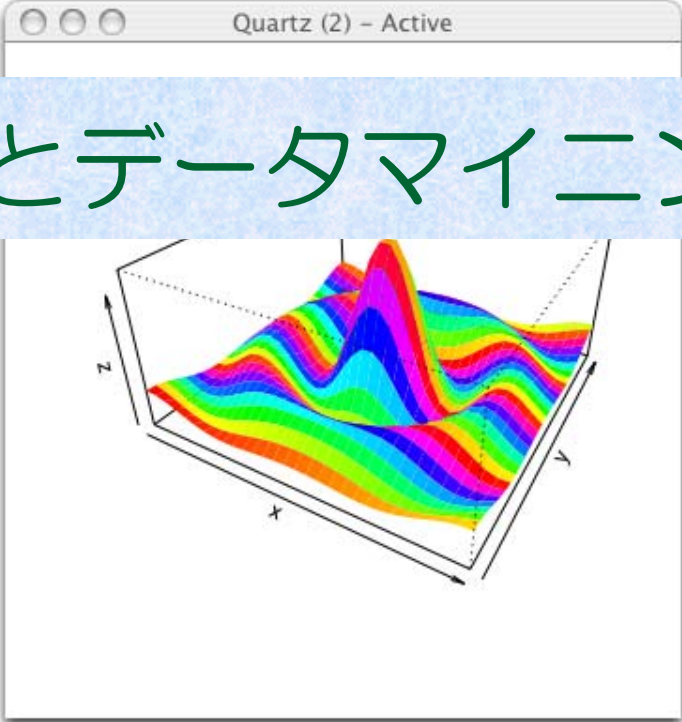


- R の概要
- R でデータを読み込む方法
- グラフとデータマイニング
- データマイニングとは？
- R Commander の紹介
- 質疑応答など

Mac OS X desktop environment showing an R Console window and a Quartz window displaying a 3D surface plot.

Rによるグラフ作成とデータマイニング

```
> plot(1:10)
> 1+2
[1] 3
> 3+4
[1]
> 1+
[1]
> 3+
[1]
> ?plot
> x <- seq(-10, 10, length= 30)
> y <- x
> f <- function(x,y) { r <- sqrt(x^2+y^2); 10 * sin(r)/r }
> z <- outer(x, y, f)
> z[is.na(z)] <- 1
> op <- par(bg = "white")
> persp(x, y, z, theta = 30, phi = 30, expand = 0.5, col = "lightblue")
> x <- seq(-10, 10, length= 30)
> y <- x
> f <- function(x,y) { r <- sqrt(x^2+y^2); 10 * sin(r)/r }
> z <- outer(x, y, f)
> z[is.na(z)] <- 1
> op <- par(bg = "white")
> persp(x, y, z, theta = 30, phi = 30, expand = 0.5, col = rainbow(200))
> x <- seq(-10, 10, length= 30)
> y <- x
> f <- function(x,y) { r <- sqrt(x^2+y^2); 10 * sin(r)/r }
> z <- outer(x, y, f)
> z[is.na(z)] <- 1
> op <- par(bg = "white")
> persp(x, y, z, theta = 30, phi = 30, expand = 0.5, border=NA, col = rainbow(50))
> x <- seq(-10, 10, length= 30)
> y <- x
> f <- function(x,y) { r <- sqrt(x^2+y^2); 10 * sin(r)/r }
> z <- outer(x, y, f)
> z[is.na(z)] <- 1
> op <- par(bg = "white")
> persp(x, y, z, theta = 30, phi = 30, expand = 0.5, border=NA, col = rainbow(200))
>
```



終