

R と RStudio の使い方

芳賀敏郎 (2011) 医薬品開発のための統計解析 第1部 基礎
4 相関・回帰
4.1 散布図

テキストと利用上の注意

●テキスト

芳賀敏郎（2011）医薬品開発のための統計解析

第1部 基礎 改訂版、サイエンティスト社、p.275

（サイトへアップすることに対して、サイエンティスト社の了解を得ています）

●Rによる解析事例を紹介

R スクリプトの出力結果を紹介します（PDFの注釈にスピーカーノートがあります）

R スクリプト（文字コードUTF-8に設定）を、このサイトから[ダウンロード](#)できます

R スクリプトを [Compile Report] することにより、Word または HTML で見ることが出来ます

R と RStudio の設定と基本的な使い方は「[R と RStudio の使い方](#)」を参照してください

R の出力結果の見方は、テキストとそれを解説した [PDF ファイル](#) を参照してください

グラフ表示は、解析手段として、必要最小限の表現に止めています

●自己責任で利用

上記のことを理解した上で、自己責任により利用してください

第1部 基礎

- 1. 統計の基礎
 - 1.1 宝くじの期待値と分散、1.2 サイコロの目の数の期待値と分散
 - 1.3 分散の加法性・中心極限定理・正規分布、1.4 統計的推測、1.5 モデル
- 2. 1組のデータの解析
 - 2.1 データの特徴の記述、2.2 データのグラフ表示と外れ値
 - 2.3 対数変換と対数正規分布、2.4 平均に関する推測（母標準偏差 σ 既知）
 - 2.5 分散に関する推測、2.6 平均に関する推測（母標準偏差 σ 未知）
- 3. 2組のデータの解析
 - 3.1 データのグラフ化、3.2 平均値の差の t 検定、3.3 分散の違いの検定
 - 3.4 分散が異なる場合の平均値の差の比較
 - 3.5 対応のある場合の平均値の差の t 検定、3.6 検出力と n の決め方
 - 3.7 ノンパラメトリック検定
- 4. 相関・回帰
 - 4.1 散布図、4.2 相関係数、4.3 回帰モデルとモデルの推定
 - 4.4 誤差を考慮した推定、4.5 回帰分析適用上の諸問題



散布図

●表示4.1.6 JMP による散布図と層別散布図

スクリプトファイル：Green1-4-1a.R

利用した関数：plot

方法

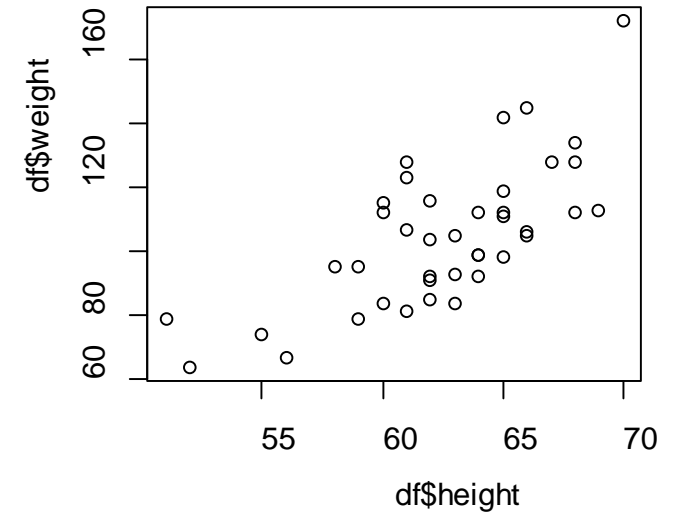
スクリプト

```
df <- read_excel("Green1-4.xlsx",
                 sheet = "4-big-class")
df <- data.frame(df)
df$sex <- factor(df$sex) # 因子型に指定

plot(x = df$height, # x のデータ
     y = df$weight) # y のデータ
```

データフレーム名\$列名

簡単な散布図



読み込んだデータ

```
> df
##      name age sex height weight
## 1  KATIE  12  F    59     95
## 2  LOUISE  12  F    61    123
## 3   JANE  12  F    55     74
## .....
## 39  KIRK  17  M    68    134
## 40 LAWRENCE 17  M    70    172
```

●表示4.1.6 JMP による散布図と層別散布図

スクリプトファイル：Green1-4-1a.R

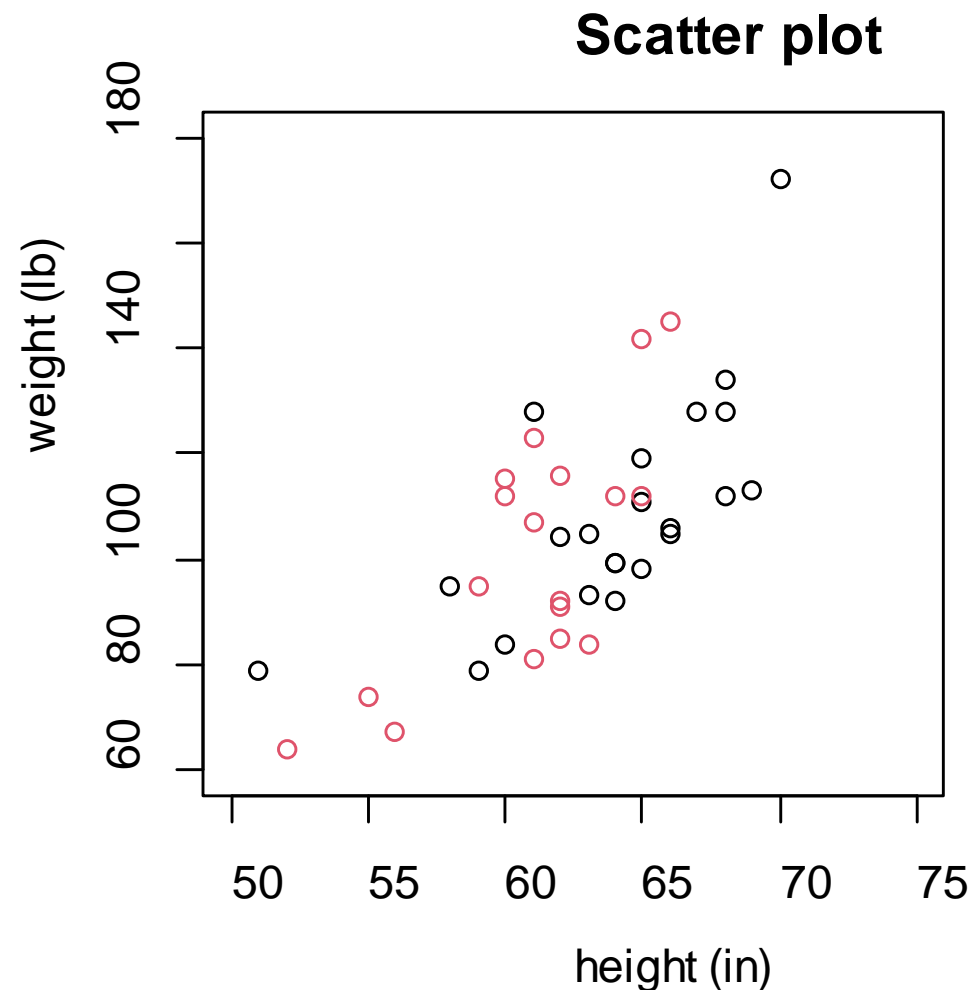
利用した関数：plot、ifelse

方法：男女をifelse 文で区別して表示

(3種類以上：dplyr::case_when 第3部 [第§2.4](#) 参照)

```
sex <- ifelse(df$sex == "M", 1, 2) # 色指定に利用
```

```
plot(x = df$height, # x のデータ
     y = df$weight, # y のデータ
     type = "p", # プロットの種類
     col = sex, # 色指定、1は黒, 2は赤
     pch = 1, # マークの種類指定
     xlim = c(50, 75), # x 軸の最小値と最大値
     ylim = c(60, 180), # y 軸の最小値と最大値
     xlab = "height (in)", # x 軸のラベル
     ylab = "weight (lb)", # y 軸のラベル
     main = "Scatter plot") # タイトル
```



●表示4.1.6 JMP による散布図と層別散布図

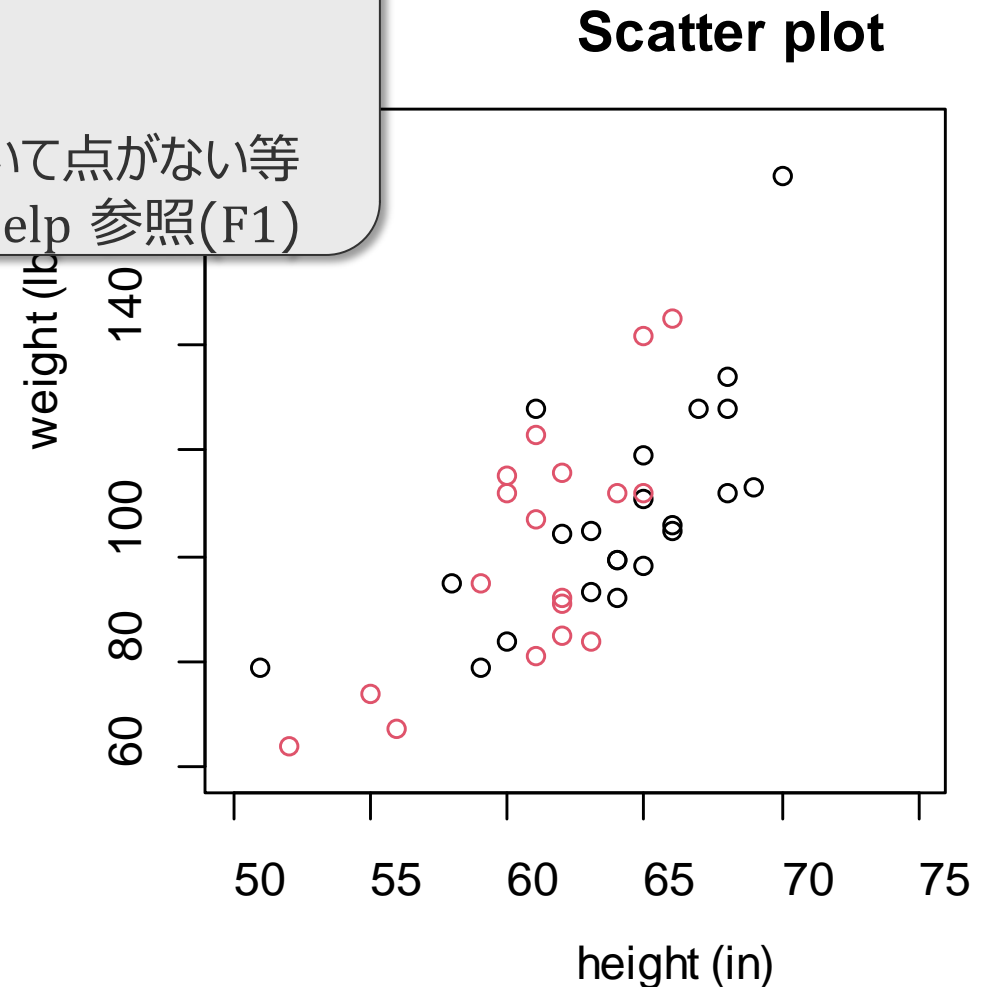
スクリプトファイル：Green1-4-1a.R

利用した関数：plot

方法：プロットのタイプ

```
sex <- ifelse(df$sex == "M", 1, 2) # 色指定に  
  
plot(x = df$height, # x のデータ  
     y = df$weight, # y のデータ  
     type = "p", # プロットの種類  
     col = sex, # 色指定、1は黒, 2は赤  
     pch = 1, # マークの種類の指定  
     xlim = c(50, 75), # x 軸の最小値と最大値  
     ylim = c(60, 180), # y 軸の最小値と最大値  
     xlab = "height (in)", # x 軸のラベル  
     ylab = "weight (lb)", # y 軸のラベル  
     main = "Scatter plot") # タイトル
```

"p" : 点 (デフォルト)
"l" : 線
"b" : 点と線
"c" : "b"において点がない等
plot 関数の Help 参照(F1)



●表示4.1.6 JMP による散布図と層別散布図

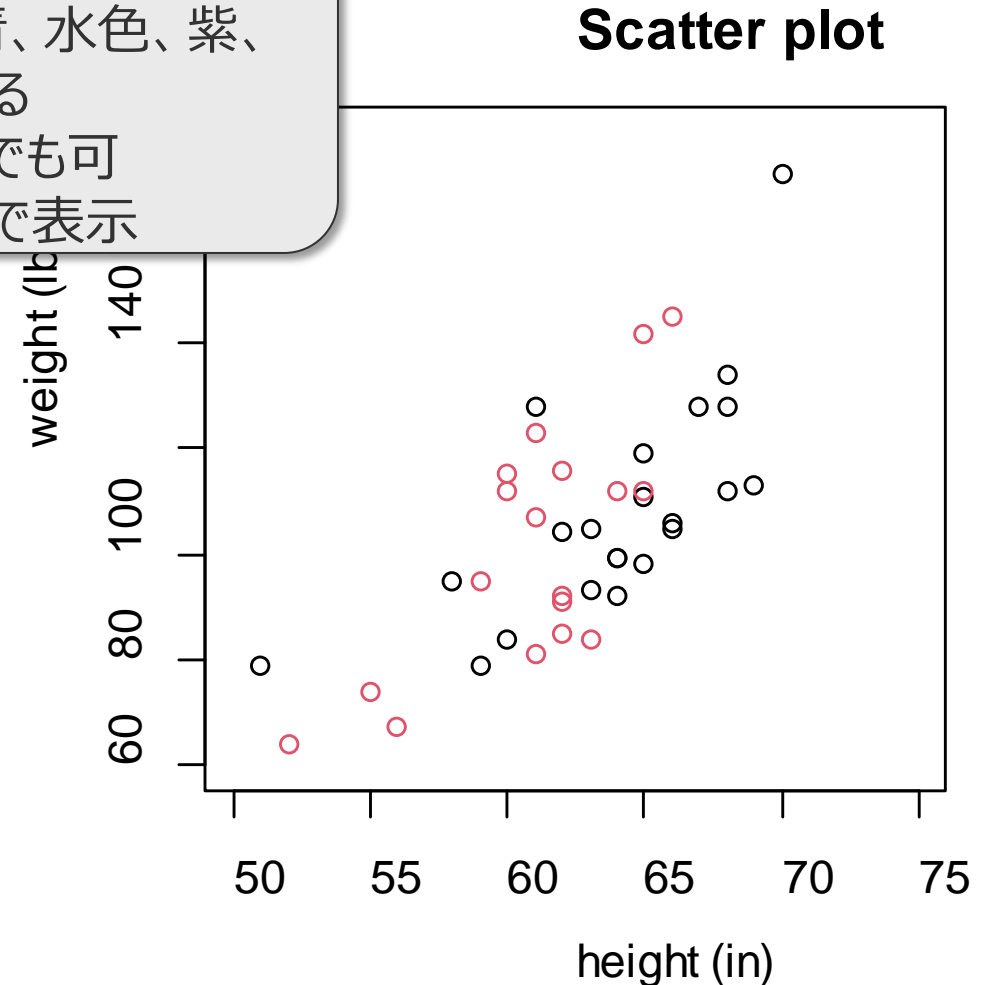
スクリプトファイル：Green1-4-1a.R

利用した関数：plot

方法：色の指定

```
sex <- ifelse(df$sex == "M", 1, 2) # 色指定に  
  
plot(x = df$height, # x のデータ  
     y = df$weight, # y のデータ  
     type = "p", # プロットの種類  
     col = sex, # 色指定、1は黒, 2は赤  
     pch = 1, # マークの種類  
     xlim = c(50, 75), # x 軸の最小値と最大値  
     ylim = c(60, 180), # y 軸の最小値と最大値  
     xlab = "height (in)", # x 軸のラベル  
     ylab = "weight (lb)", # y 軸のラベル  
     main = "Scatter plot") # タイトル
```

1～8の順に
黒、赤、緑、青、水色、紫、
黄、灰色になる
col = "black"でも可
colors 関数で表示



●表示4.1.6 JMP による散布図と層別散布図

スクリプトファイル：Green1-4-1a.R

利用した関数：plot

方法：マークの指定

```
sex <- ifelse(df$sex == "M", 1, 2) # 色指定に  
  
plot(x = df$height, # x のデータ  
     y = df$weight, # y のデータ  
     type = "p", # プロットの種  
     col = sex, # 色指定、1は黒, 2は赤  
     pch = 1, # マークの種類  
     xlim = c(50, 75), # x 軸の最小値と最大値  
     ylim = c(60, 180), # y 軸の最小値と最大値  
     xlab = "height (in)", # x 軸のラベル  
     ylab = "weight (lb)", # y 軸のラベル  
     main = "Scatter plot") # タイトル
```

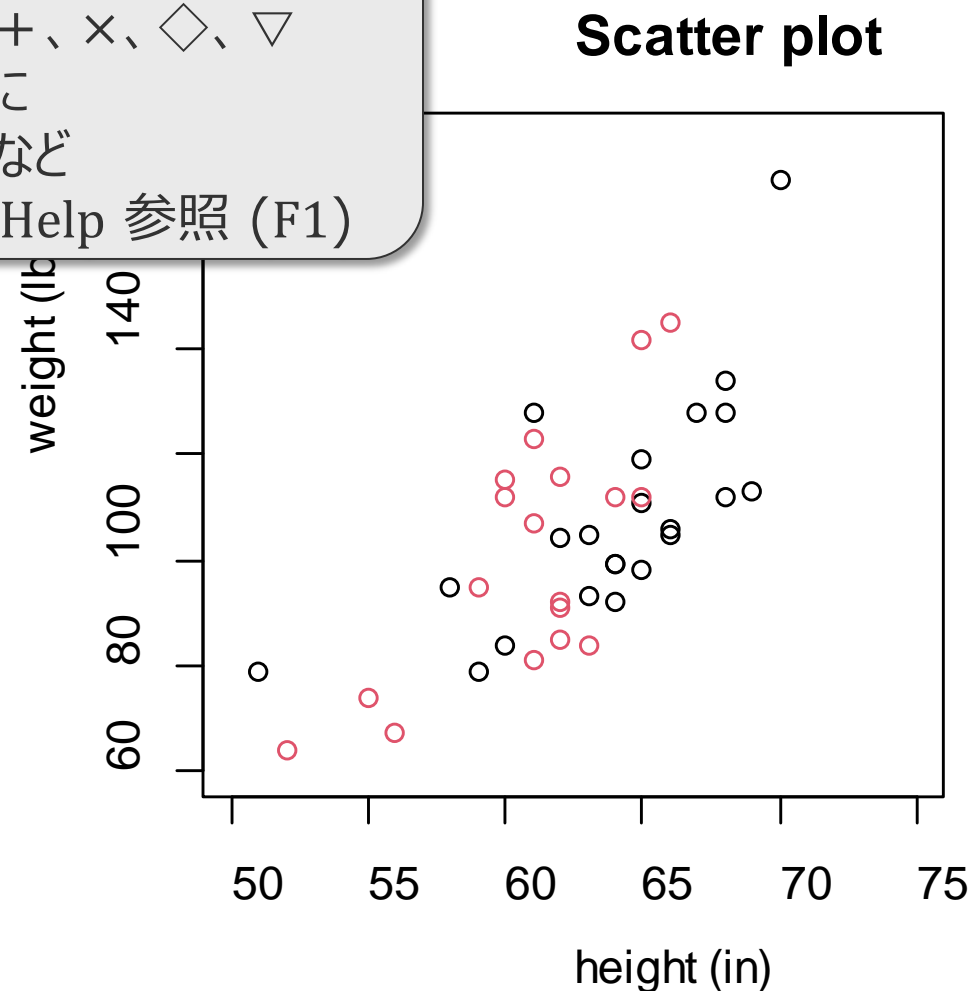
0~6 の順に

□、○、△、+、×、◇、▽

15~17 の順に

■、●、▲ など

points 関数のHelp 参照 (F1)



●表示4.1.6 JMP による散布図と層別散布図

スクリプトファイル：Green1-4-1a.R

利用した関数：plot、legend、dplyr::case_when

方法：3項目以上への拡張

```
sex <- ifelse(df$sex == "M", 1, 2)
```

```
plot(.....  
  col = sex,  
  pch = 1,  
  ...)
```

入れ子で3項目以上に拡張可
ifelse(条件, ifelse(, ,), ...)

```
> sex  
[1] 2 2 2 2 2 1 1 1 2 2 2 1 1 1 1 2 2  
[18] 2 2 2 1 1 1 1 1 1 2 2 1 1 1 1 1  
[35] 2 2 1 2 1 1
```

3項目以上に対応

```
library(dplyr)  
  
sex <- case_when(  
  df$sex == "M" ~ 1,  
  df$sex == "F" ~ 2,  
  df$sex == "X" ~ 3)  
plot(.....  
  col = sex,  
  pch = sex,  
  ...)
```

因子型

```
df$sex <- factor(df$sex)
```

簡便

```
plot(.....  
  col = c(1, 2, 3)[df$sex],  
  pch = c(1, 2, 3)[df$sex],  
  ...)
```

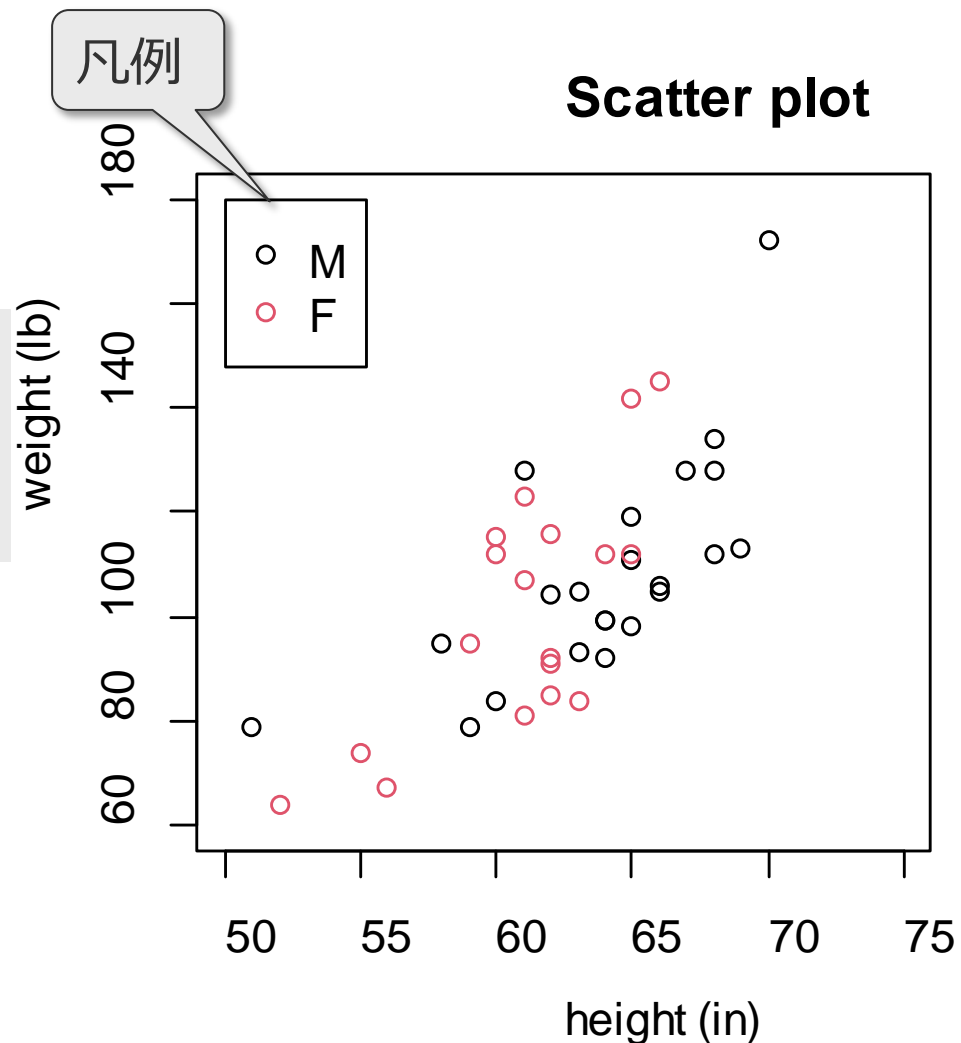
●表示4.1.6 JMP による散布図と層別散布図

スクリプトファイル：Green1-4-1a.R

利用した関数：plot、legend

方法：凡例の表示

```
legend(x = 50, y = 180, # 凡例の位置  
       legend = c("M", "F"), # 凡例に表示される文字  
       col = c(1, 2), # 凡例に表示される色  
       pch = 1) # 凡例に表示されるマーク
```



●表示4.1.6 JMP による散布図と層別散布図

スクリプトファイル：Green1-4-1a.R

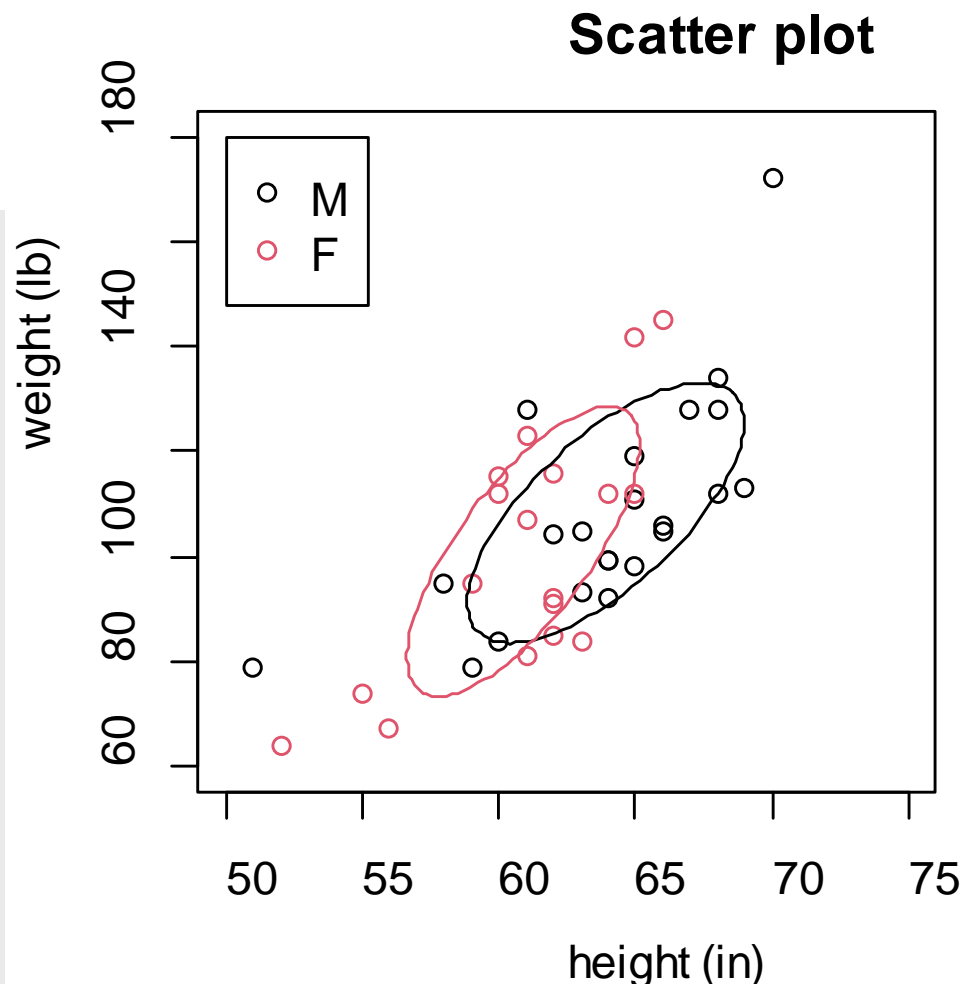
利用した関数：plot、subset、polygon、ellipse::ellipse

方法：確率楕円の表示

```
df_m <- subset(df, sex == "M",  
               select = c(height, weight))  
df_f <- subset(df, sex == "F",  
               select = c(height, weight))
```

分散共分散行列

```
polygon(  
  ellipse(cov(df_m), # 楕円のデータ  
          centre = colMeans(df_m), # 楕円を描く中心  
          level = 0.50), # 確率楕円の信頼率  
  lty = 1, # 線の種類 1は実線  
  lwd = 1, # 線の太さデフォルト  
  border = 1) # 境界線の色
```



●表示4.1.6 JMP による散布図と層別散布図

スクリプトファイル：Green1-4-1a.R

利用した関数：plot、subset、polygon、

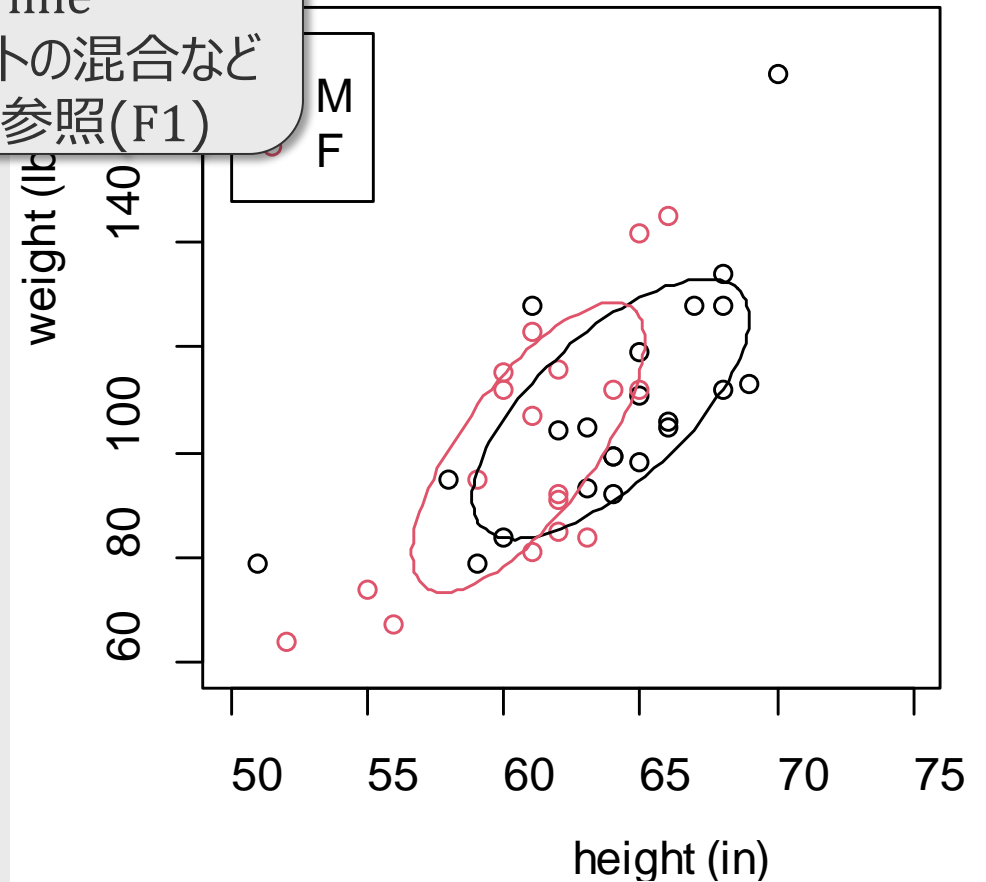
方法：確率楕円の線種と太さの指定

```
df_m <- subset(df, sex == "M",
               select = c(height, weight))
df_f <- subset(df, sex == "F",
               select = c(height, weight))
```

```
polygon(
  ellipse(cov(df_m),
          centre = colMeans(df_m),
          level = 0.50),
  lty = 1,
  lwd = 1,
  border = 1)
# 楕円のデータ
# 楕円を描く中心
# 確率楕円の信頼率
# 線の種類 1は実線
# 線の太さデフォルト
# 境界線の色
```

- 1 : 実線 solid line
 - 2 : 破線 dashed line
 - 3 : 点線 doted line
 - 4 : ダッシュとドットの混合など
- par 関数のHelp参照(F1)

Scatter plot



- 表示4.1.6 JMP による散布図と層別散布図 [次節](#) を参照

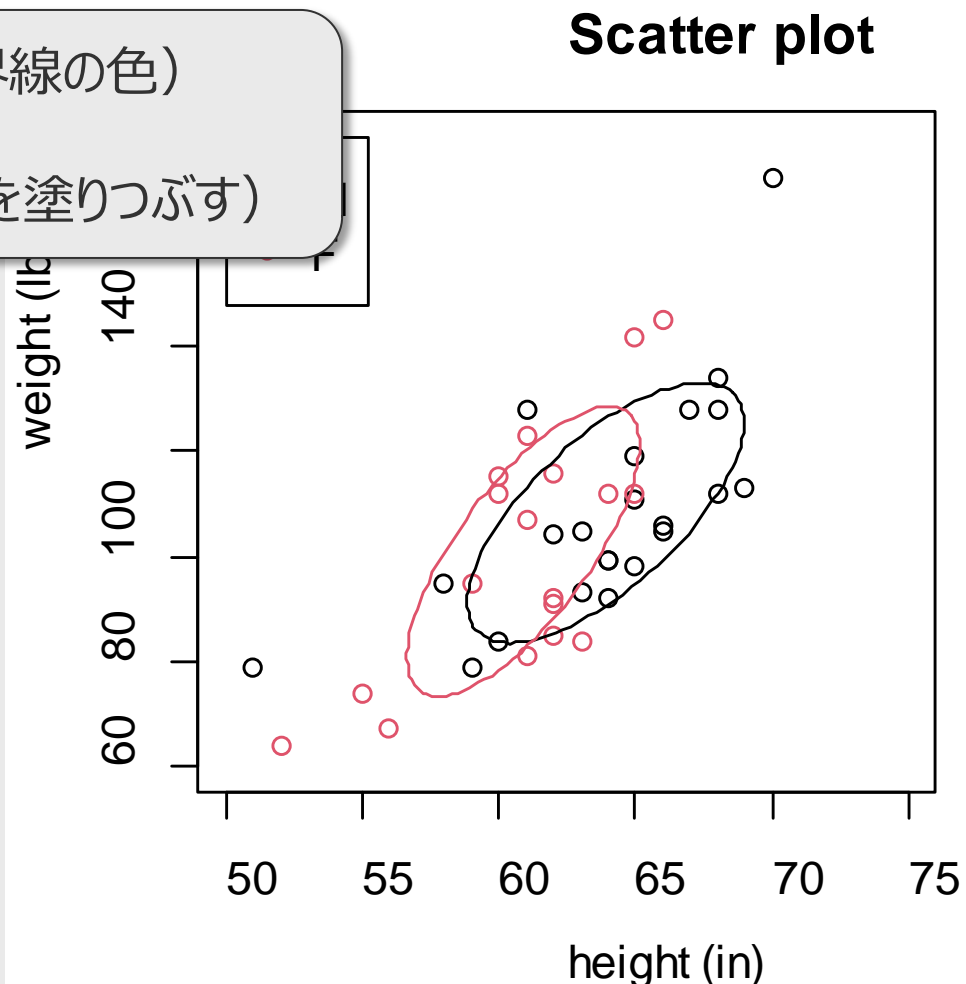
スクリプトファイル：Green1-4-1a.R

利用した関数：plot、subset、polygon、

方法：確率楕円の色の指定

```
df_m <- subset(df, sex == "M",  
               select = c(height, weight))  
df_f <- subset(df, sex == "F",  
               select = c(height, weight))  
  
polygon(  
  ellipse(cov(df_m),  
          centre = colMeans(df_m), # 楕円のデータ中心  
          level = 0.50), # 確率楕円の信頼率  
  lty = 1, # 線の種類 1は実線  
  lwd = 1, # 線の太さデフォルト  
  border = 1) # 境界線の色
```

楕円の色（境界線の色）
（引数 col は
楕円内を塗りつぶす）



- 表示4.1.6 JMP による散布図と層別散布図

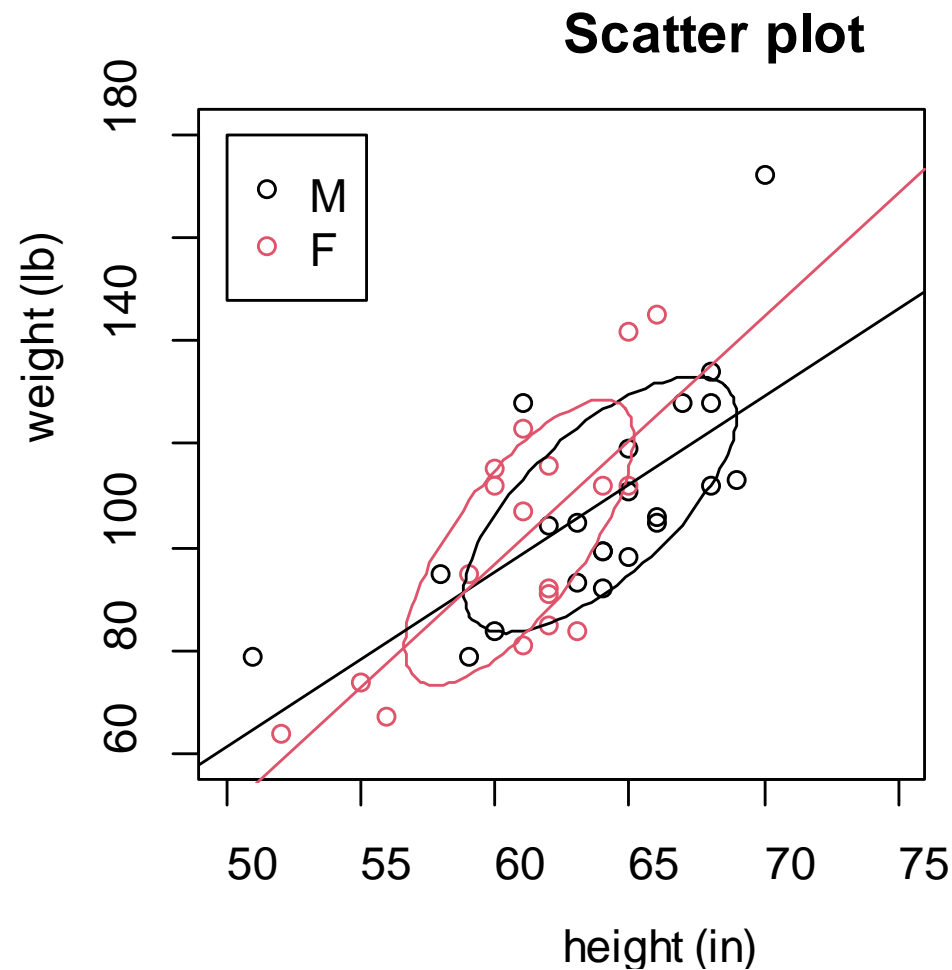
スクリプトファイル：Green1-4-1a.R

利用した関数：plot、lm、abline

方法：回帰直線を表示

```
abline(lm(weight ~ height, data = df_m), col = 1)  
abline(lm(weight ~ height, data = df_f), col = 2)
```

lm 関数は回帰分析を行う
[§4.4](#) を参照



- 表示4.1.6 JMP による散布図と層別散布図

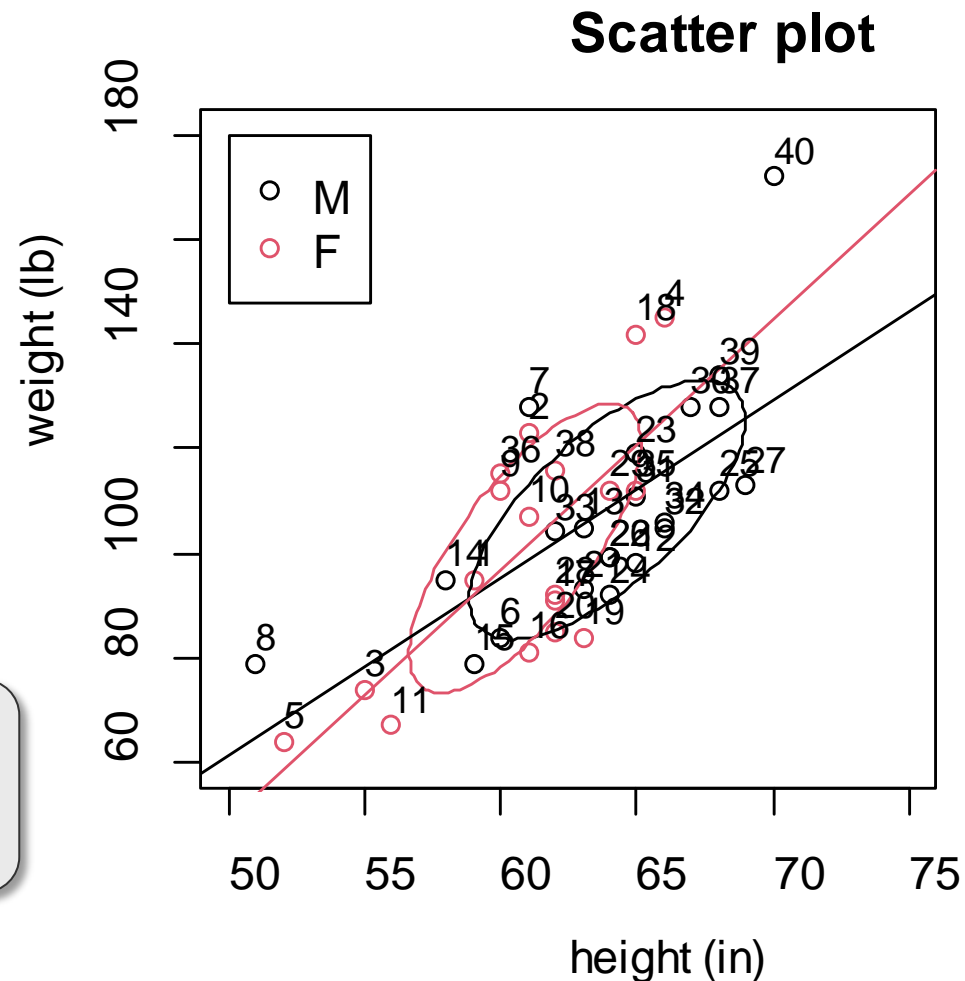
スクリプトファイル：Green1-4-1a.R

利用した関数：plot、text

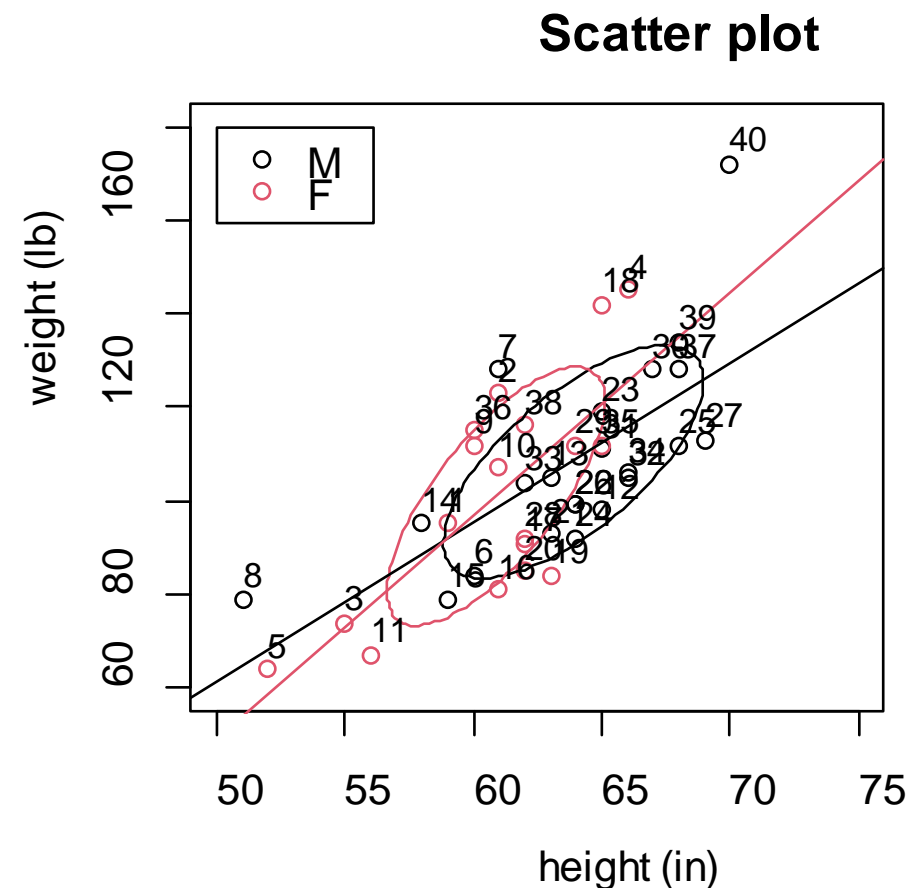
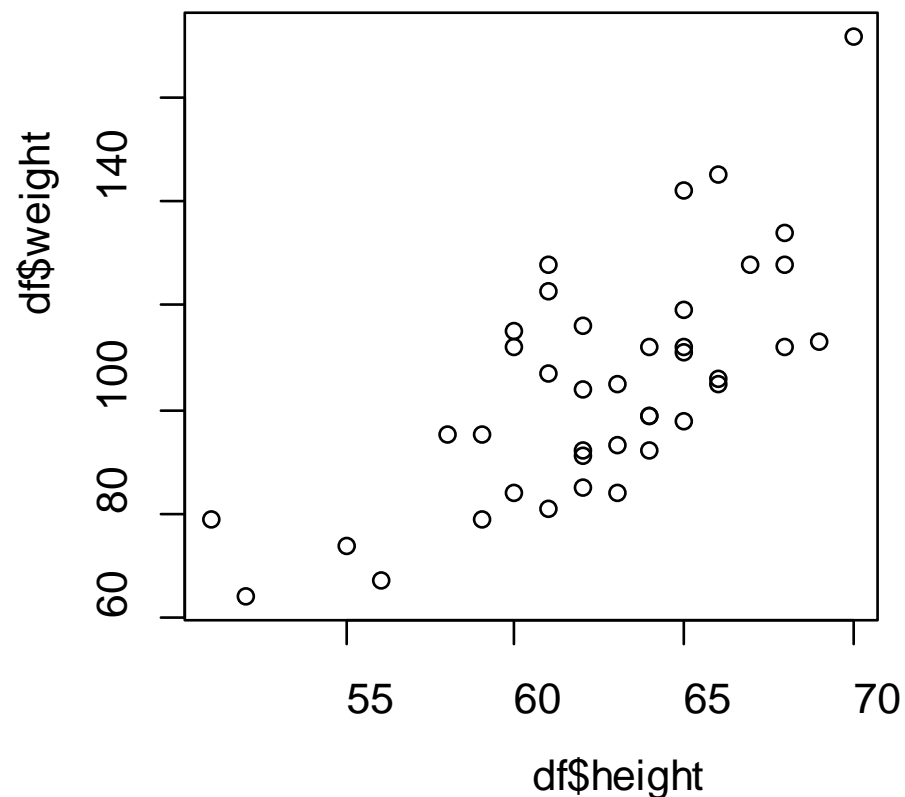
方法：個々の値のラベルを表示

```
text(x = df$height, y = df$weight,  
     labels = 1:40,  
     cex = 0.8,  
     adj = c(0, -0.5))
```

labels = df\$name とすると
name 列の文字列が表示



- 表示4.1.6 JMP による散布図と層別散布図
スクリプトファイル：Green1-4-1a.R
方法：画面を分割して複数のグラフを同時に表示



- 画面分割：複数のグラフを表示

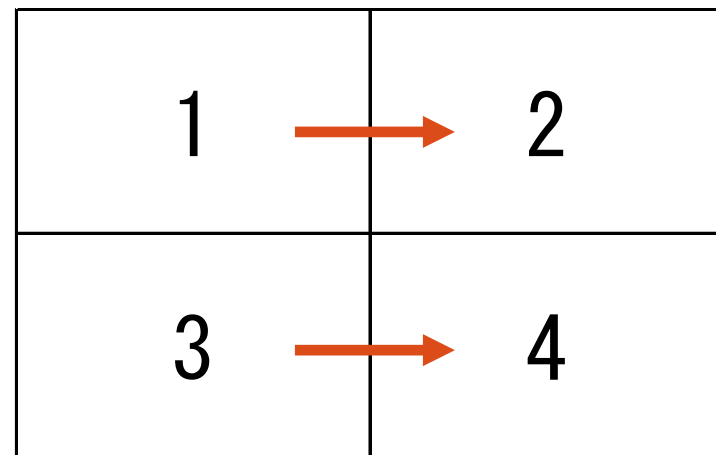
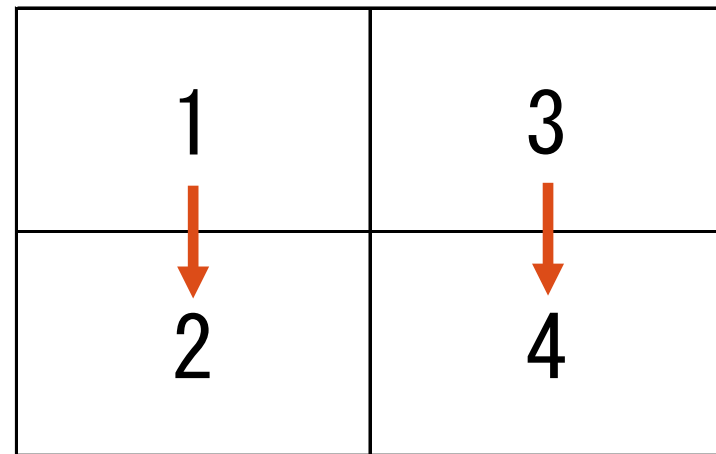
スクリプトファイル：Green1-4-1b.R

利用した関数：par、mfcol、mfrow

方法

```
par(mfcol = c(2, 2)) # 画面を2行2列に分割  
plot(1)              # 列方向に配置 (右上の図)  
plot(2)  
plot(3)  
plot(4)
```

```
par(mfrow = c(2, 2)) # 画面を2行2列に分割  
plot(1)              # 行方向に配置 (右下の図)  
plot(2)  
plot(3)  
plot(4)
```



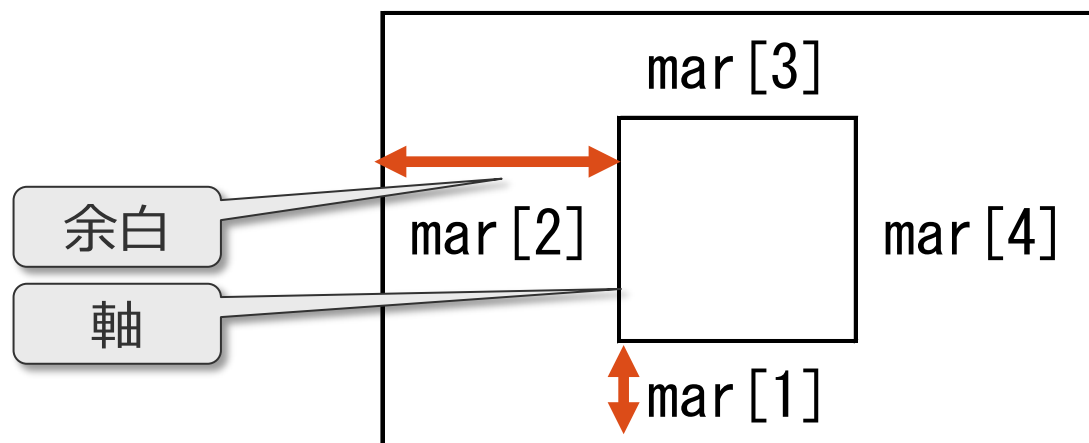
● 余白の設定

スクリプトファイル：Green1-4-1b.R

利用した関数：par、mar

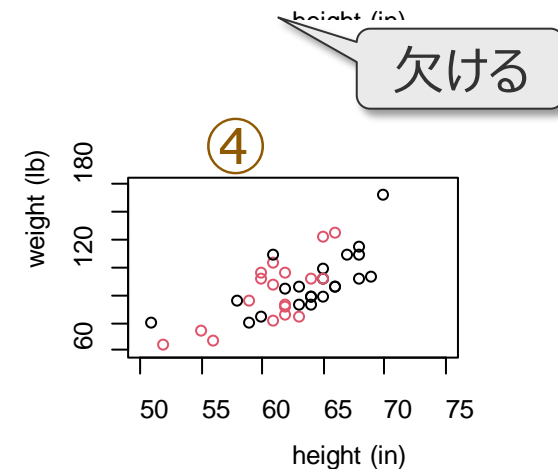
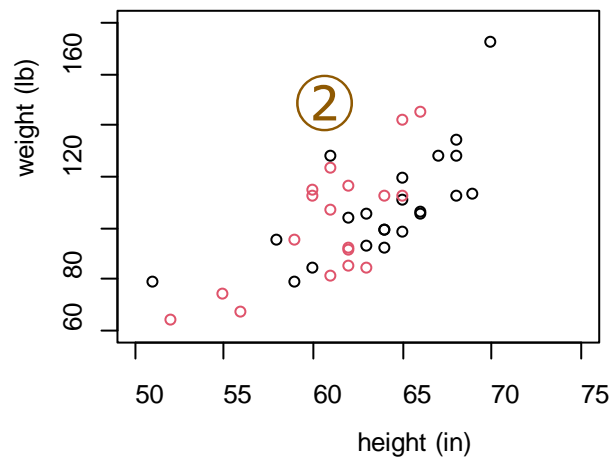
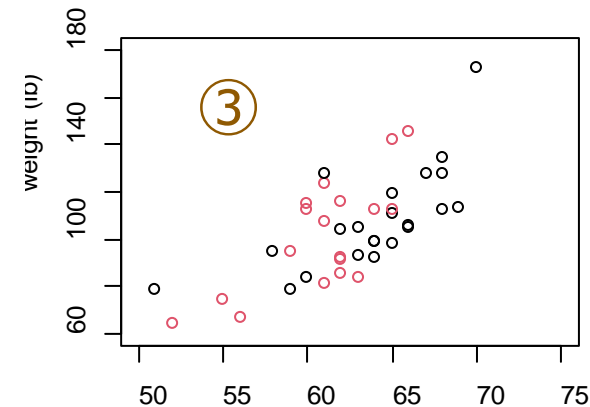
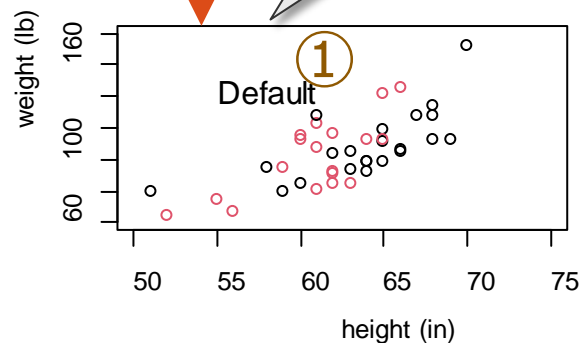
方法

```
par(mar = c(5, 4, 4, 2) + 0.1) # ①デフォルト  
par(mar = c(4, 4, 1, 1) + 0.1) # ②  
par(mar = c(3, 3, 3, 3)) # ③  
par(mar = c(5, 5, 5, 5)) # ④
```



やや間延び

軸



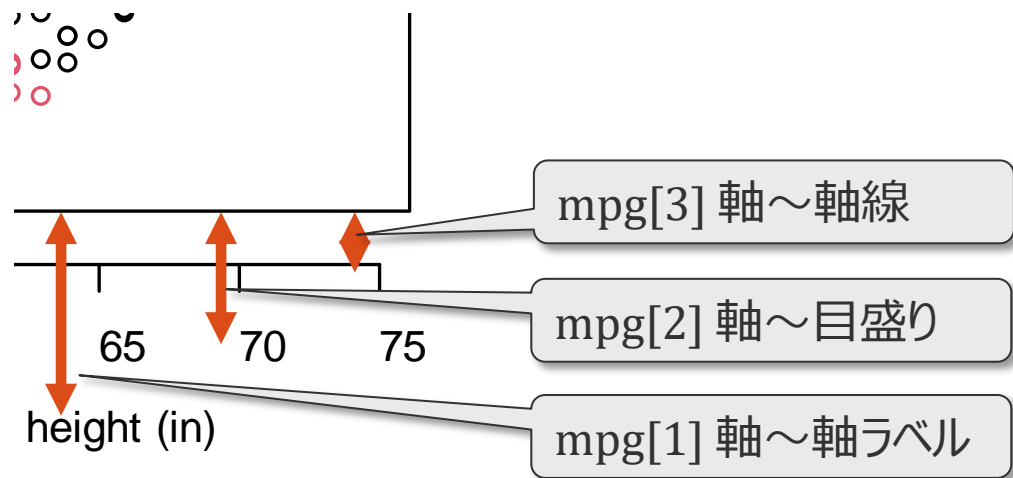
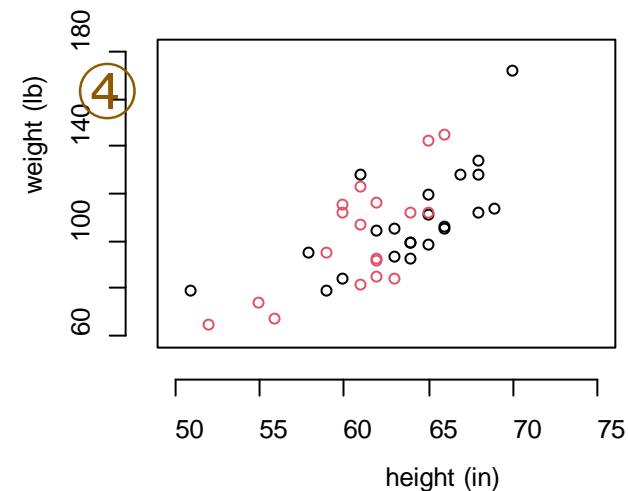
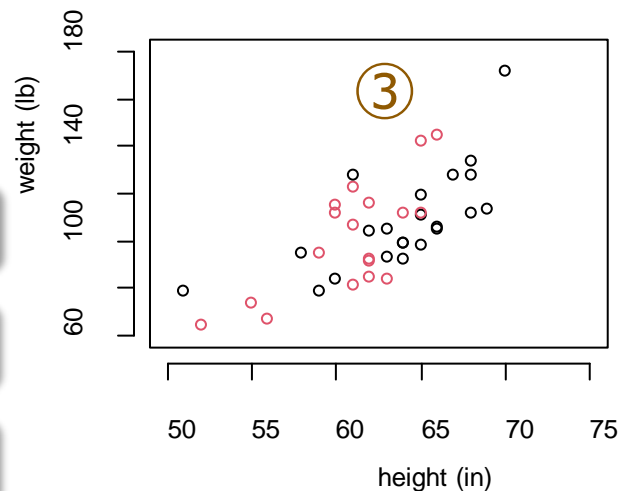
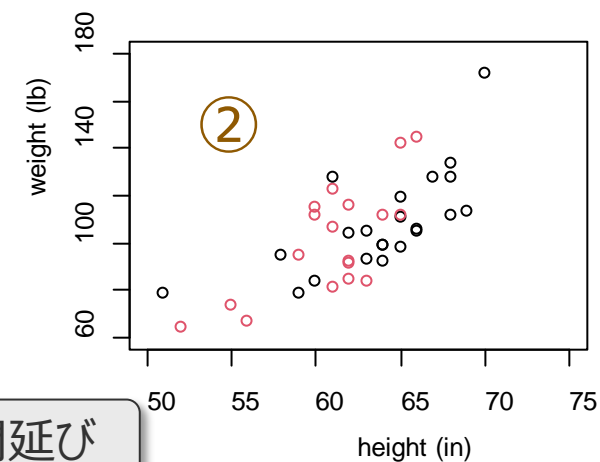
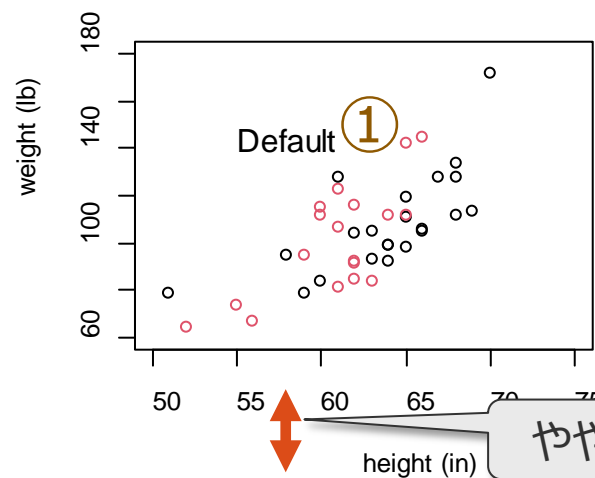
● 軸ラベルの位置

スクリプトファイル：Green1-4-1c.R

利用した関数：par、mgp

方法

```
par(mgp = c(3, 1, 0))      # ①デフォルト  
par(mgp = c(2.5, 1, 0))   # ②  
par(mgp = c(3.5, 1.5, 0.5)) # ③  
par(mgp = c(3.5, 2, 1))   # ④
```





- 作成 片瀬雅彦
- 作成時期 2021年8月16日
- 改訂 2022年1月10日