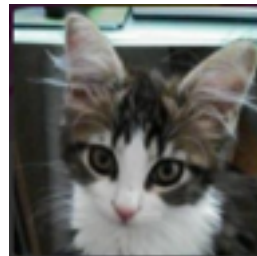


LIFE & INFO WOMEN



@marimiya_clc

宮本 真理

株式会社キアゲン
グローバルインフォマティクスソリューションズ&サポート
マネージャー

MEMBERS

私たちのチームを紹介します！

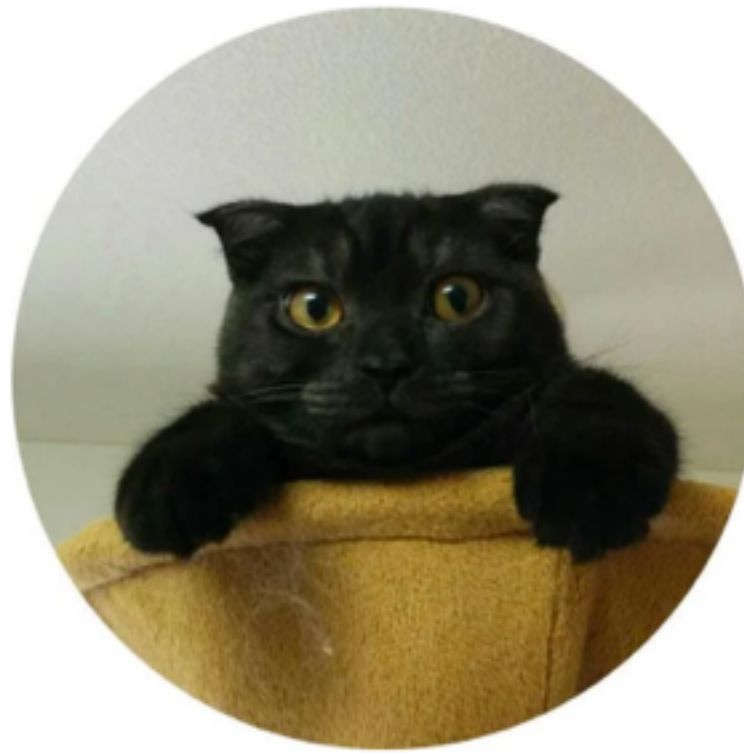


marimiya

Founder

バイオインフォのお仕事しています。アルゴリズムや統計が好きで、最近ではビジュアライゼーションの事をよく考えます。
元気女子でバイオインフォを少しでも便利でキレイに。

Let's make a change!



yokonaga

Founder

バイオテクノロジーカンパニーで働いています。昔はサカナとバイオインフォを使って遺伝性疾患の研究をしていました！
バイオとインフォと猫でライフを豊かにすることがモットーです。



ayaKT

Founder

Microbiome Specialist wannabe 仕事でもプライベートでもウンと格闘中〜☆



自己紹介

バイオラッド 営業

マイクロソフト マーケティング

Rosetta Biosoftware

APACテクニカルアカウント
マネージャー

CLCバイオジャパン

シニアフィールドアプリケーションサイエン
ティスト



キアゲン
アドバンストゲノミクス
マネージャー

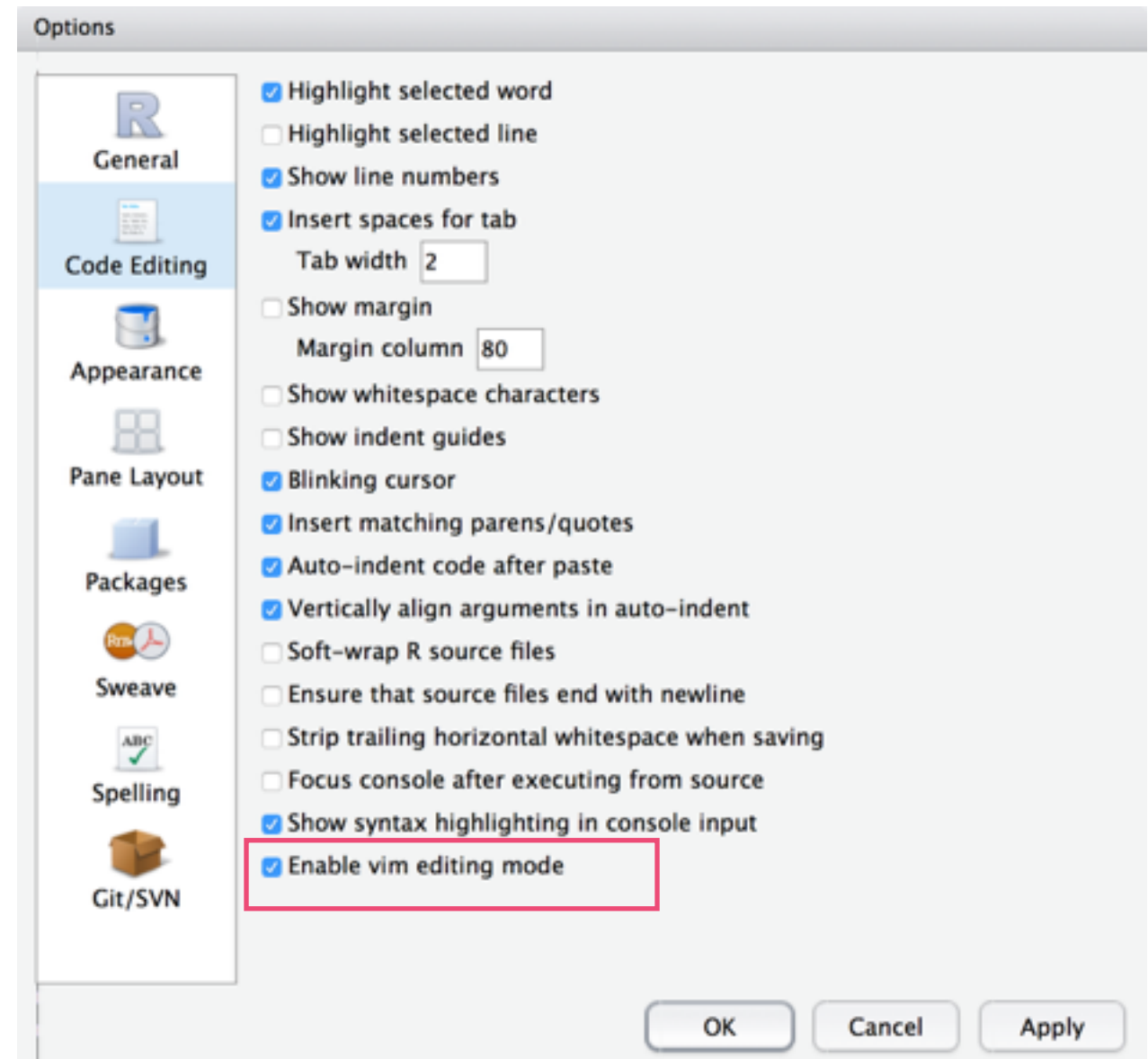
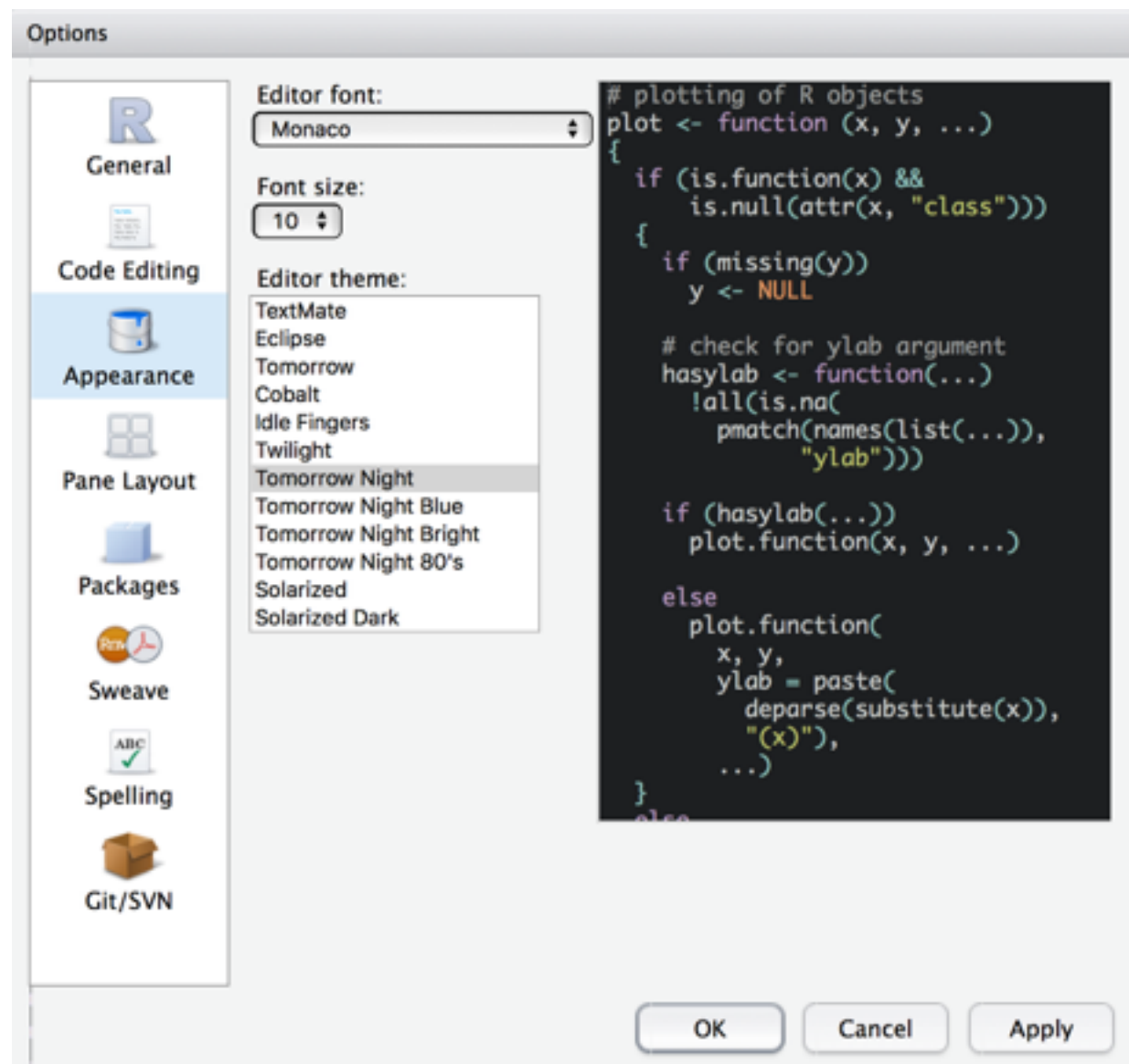
北里大学 生物科学科

久留米大学 バイオ統計学群

総合研究大学院大学
統計科学専攻

RStudioも

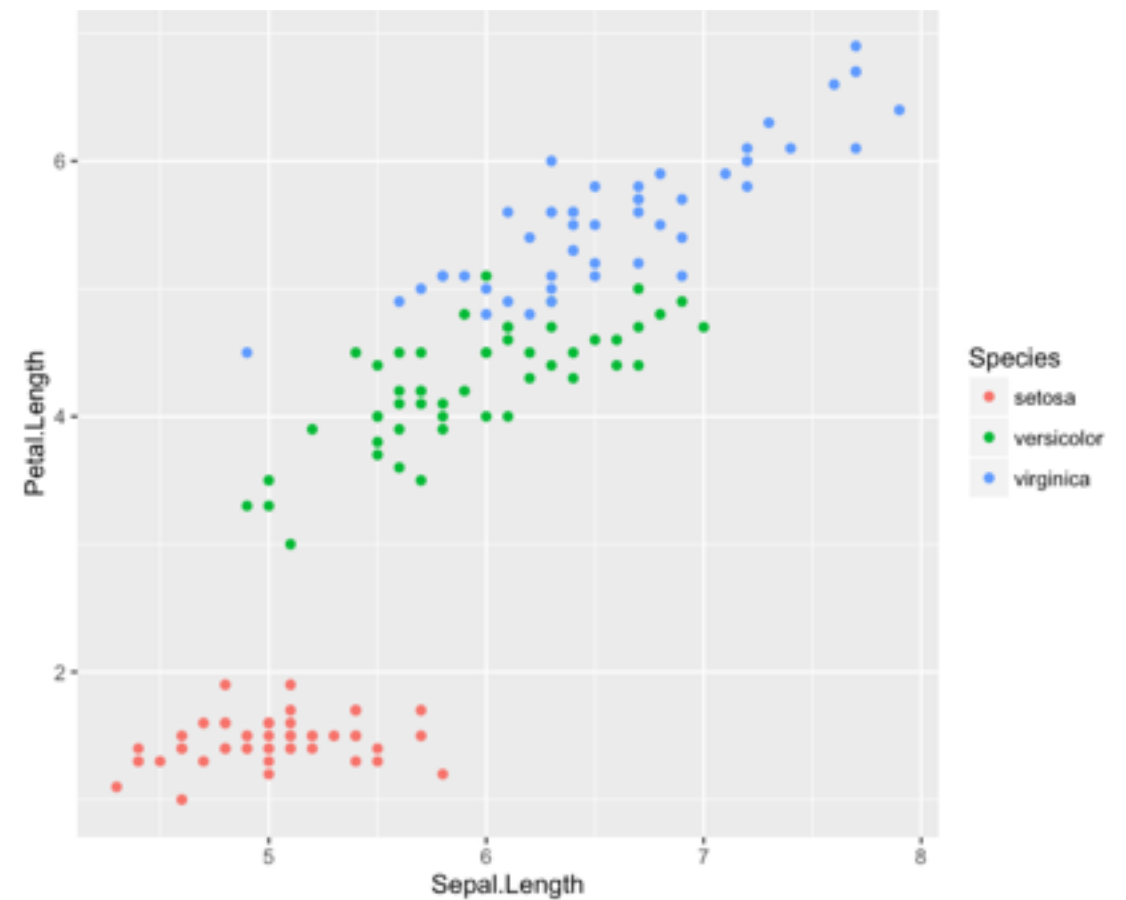
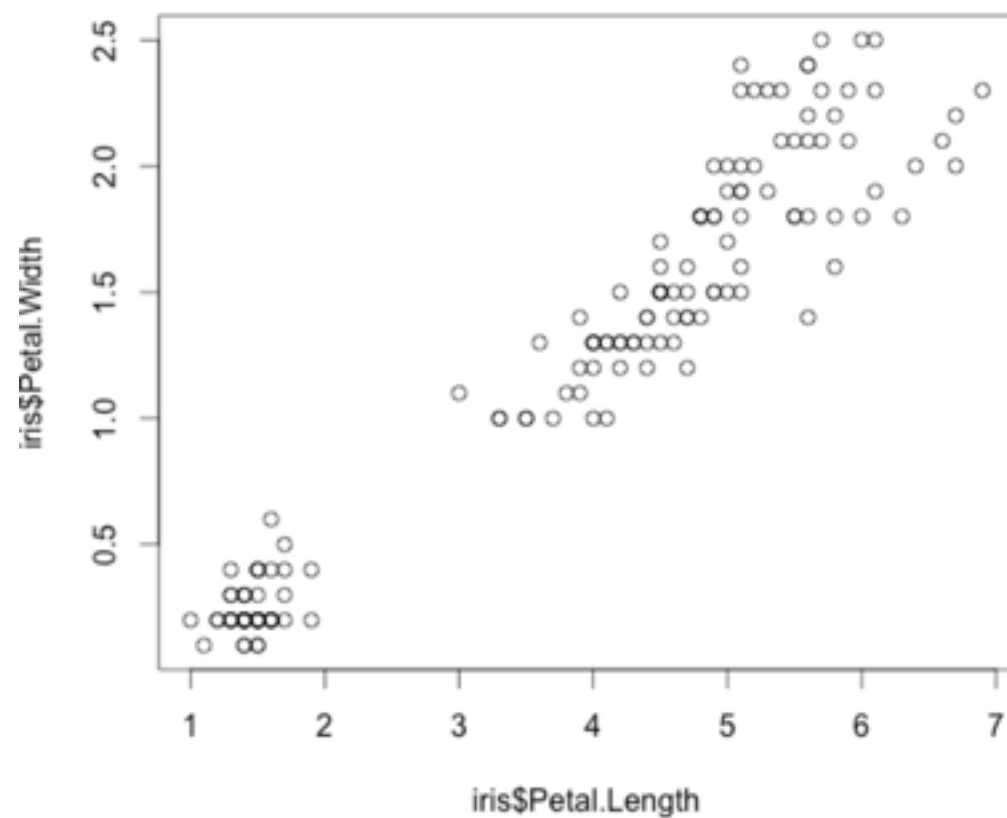
Tools > Global options



ggplot2とは？

Grammer of Graphics

綺麗なプロットが描けるRのライブラリ



ggplot2とは？

見た目が大事な時もある



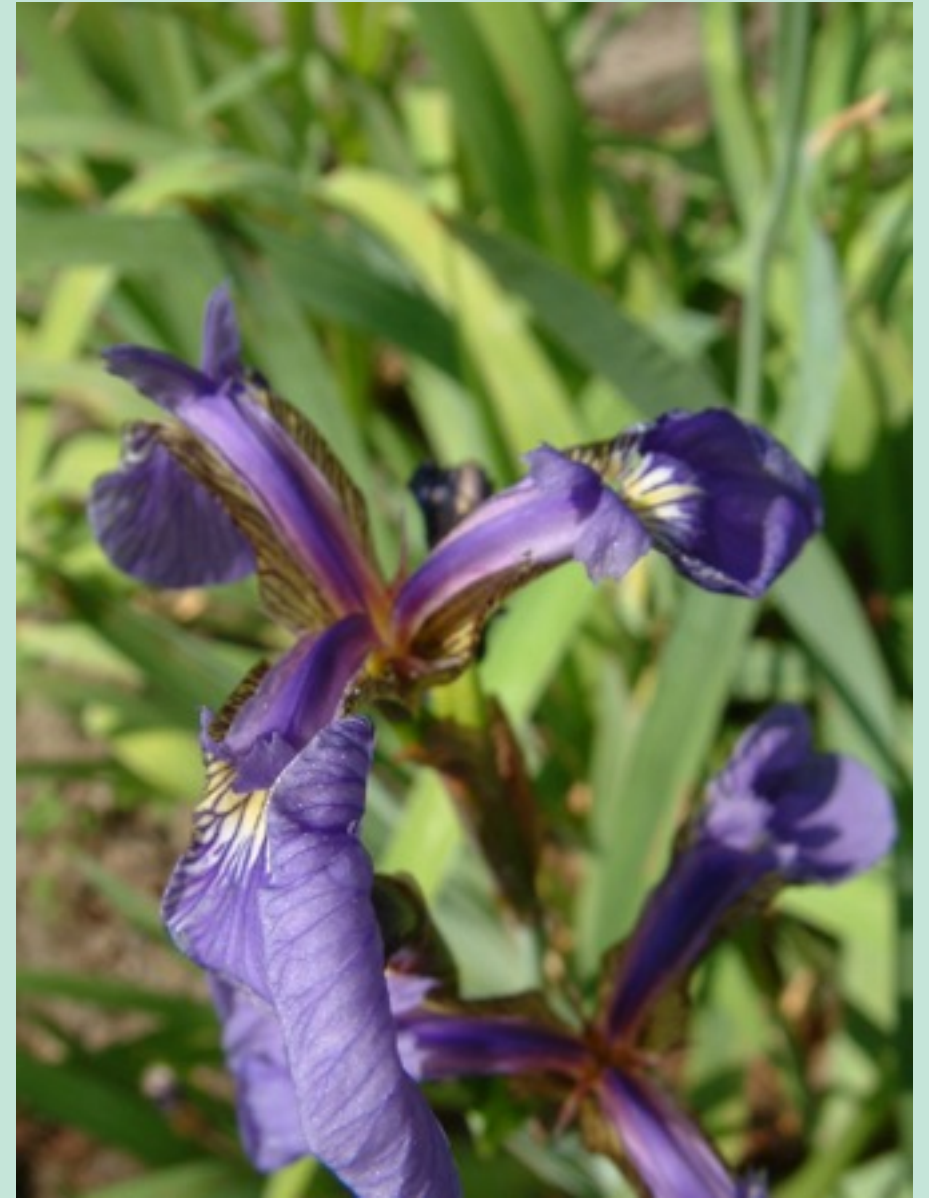
TPOに合わせて使い分け

iris (アヤメ)のデータ

- ・有名なデータセット
- ・data(iris)で利用可能
- ・3種類のアヤメのデータをAndersonがデータを収集しFisher が数値データとして発表



versicolor



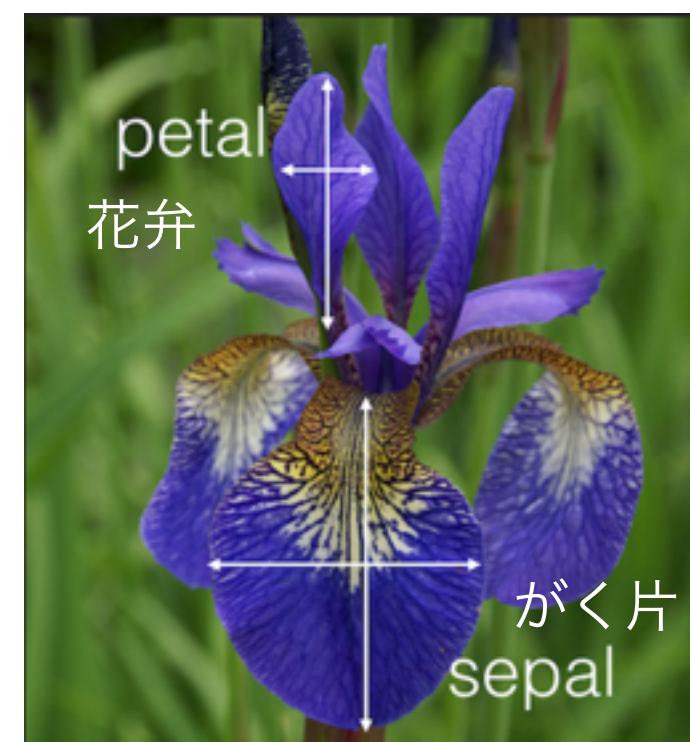
setosa

virginica



iris (アヤメ)のデータ

- ・がく片の長さ と 幅
- ・花弁の長さ と 幅
- ・それぞれのアヤメの種類



Sepal length	Sepal width	Petal length	Petal width	Species
5.1	3.5	1.4	0.2	I. setosa
4.9	3	1.4	0.2	I. setosa
4.7	3.2	1.3	0.2	I. setosa
4.6	3.1	1.5	0.2	I. setosa
5	3.6	1.4	0.2	I. setosa
5.4	3.9	1.7	0.4	I. setosa

Rコマンド練習

```
data(iris)    #irisデータの呼び出し  
head(iris)    #データの先頭部分を表示
```

```
> head(iris)  
  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
```

Rコマンド練習

`dim(iris)` #データのサイズ

`summary(iris)` #データの概要（平均や最小・最大）

```
> dim(iris)
[1] 150   5
> summary(iris)
  Sepal.Length   Sepal.Width   Petal.Length   Petal.Width
Min.      :4.300   Min.      :2.000   Min.      :1.000   Min.      :0.100
1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
Median :5.800   Median :3.000   Median :4.350   Median :1.300
Mean    :5.843   Mean    :3.057   Mean    :3.758   Mean    :1.199
3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
Max.    :7.900   Max.    :4.400   Max.    :6.900   Max.    :2.500
   Species
setosa      :50
versicolor:50
virginica   :50
```


基本の文法

1. qplot()

```
qplot(data=diet,x=day,y=calorie)
```

2. ggplot()

```
ggplot(data=diet,aes(x=day,y=calorie)) +  
geom_line()
```

2のggplot関数を使いレイヤーのような重ね書きをしていく事でより細かな設定ができるようになる

今日は2についてのお話

基本の文法

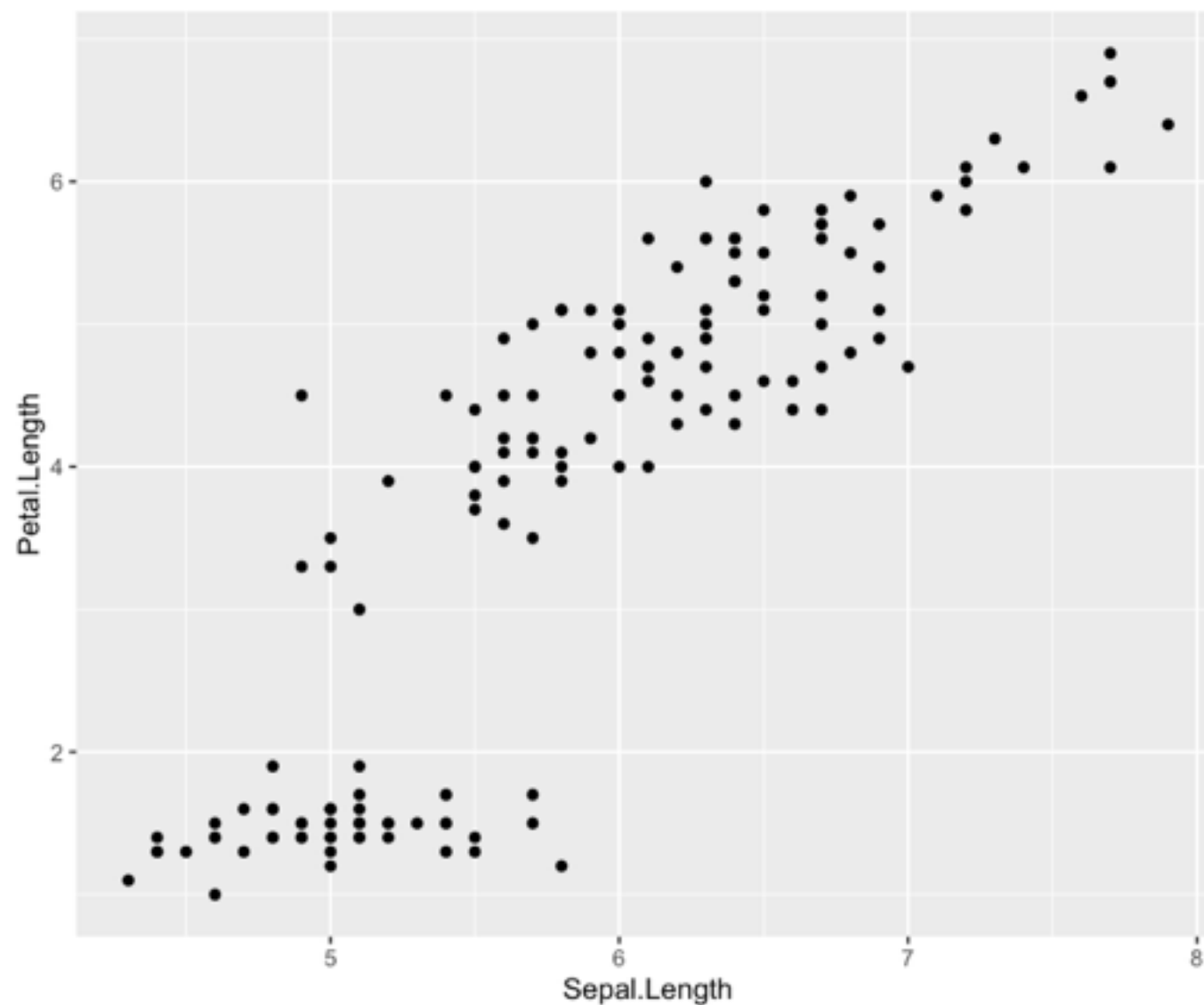
Plotを構成する要素

- ・ `data`: 使用するデータ
- ・ `mapping`: 変数と視覚的要素のマッピング
- ・ `geom`: グラフの種類
- ・ `stat`: 統計値を使った描画
- ・ `position`: 位置情報など

これらを適宜組み合わせ、土台となるデータに追加していくイメージ

ggplot2 基礎

```
g <- ggplot(data = iris, aes(x = Sepal.Length, y =  
Petal.Length))  
g + geom_point()
```

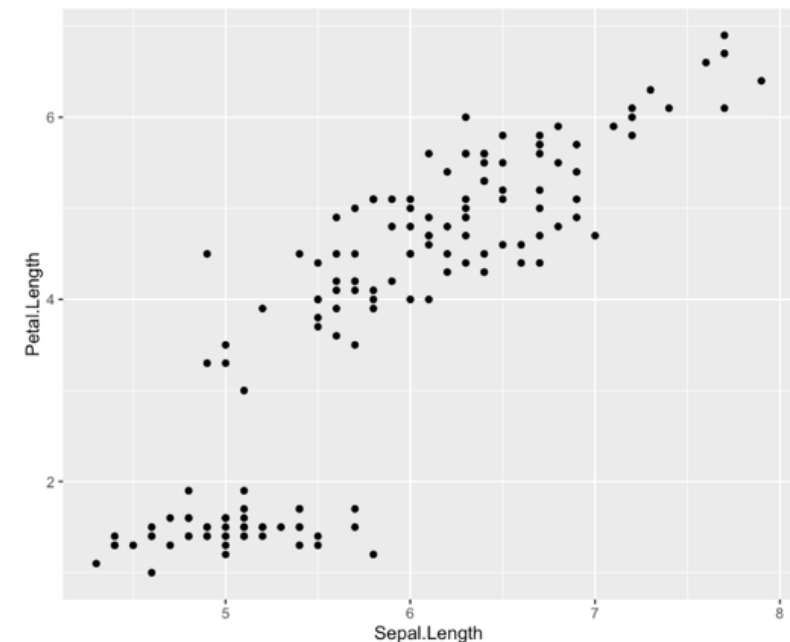


ggplot2 基礎

データの指定
x軸に使う変数の指定
y軸に使う変数の指定

```
g <- ggplot(data = iris, aes(x = Sepal.Length, y =  
Petal.Length))  
g + geom_point()
```

グラフの種類



`aes(x, y, ...)`

Generate aesthetic mappings that describe how variables in the data are mapped to visual properties (aesthetics) of geoms. This function also standardise aesthetic names by performs partial name matching, converting color to colour, and old style R names to ggplot names (eg. pch to shape, cex to size)

Press F1 for additional help

aes = aesthetic attribute (審美的属性)。変な名前ピンと来ないので、単に八重洲からyを抜いたアエスとか、ゴロ合わせで覚えましょう。
描画したいx軸とy軸なんかを指定と覚えましょう。このaesでデータをマッピングしている

```
g <- ggplot(data = iris, aes(x = Sepal.Length, y =  
Petal.Length))  
g + geom_point()
```

```
g <- ggplot(iris, aes(Sepal.Length, Petal.Length))  
g + geom_point()
```

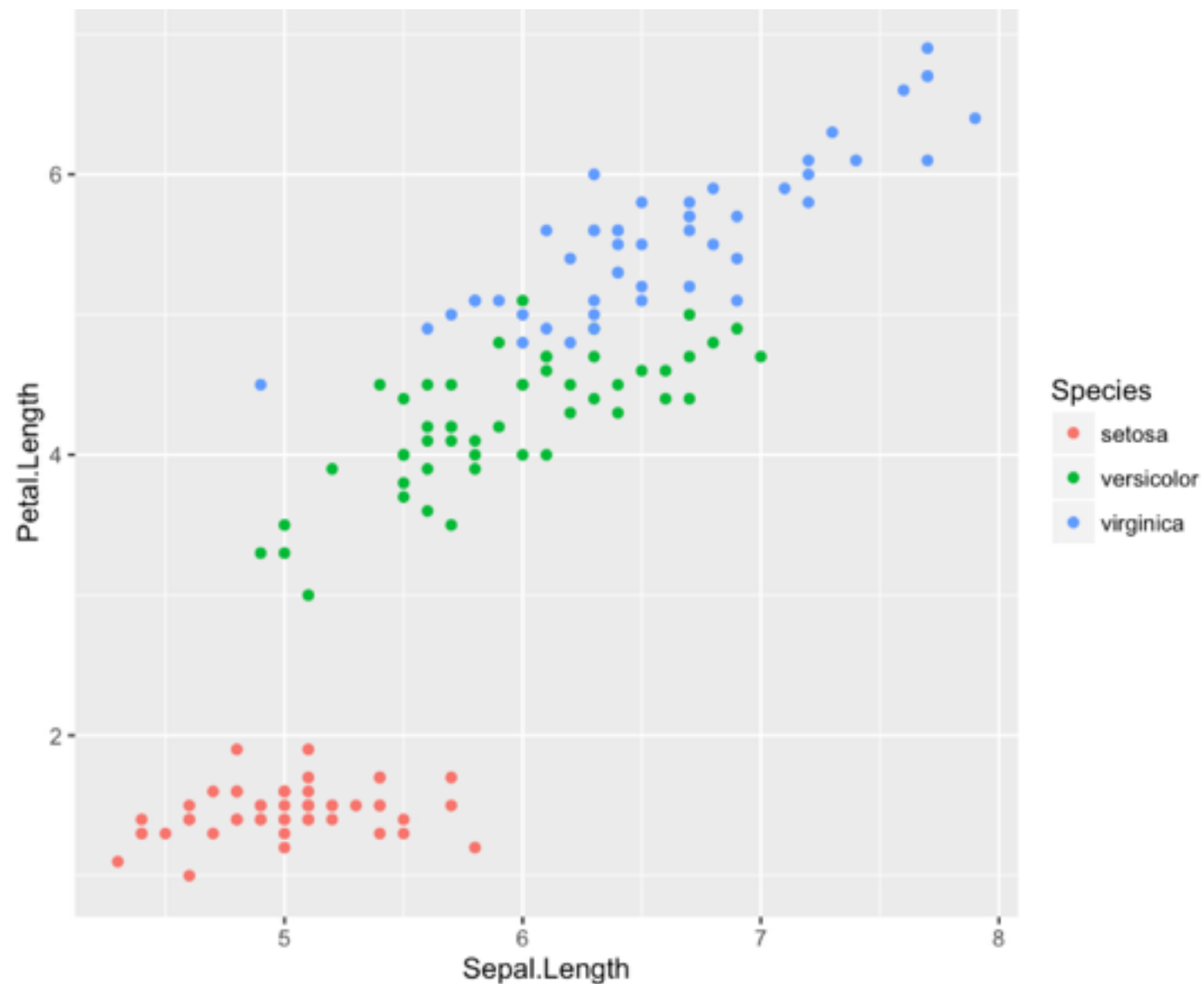
```
ggplot(iris, aes(Sepal.Length, Petal.Length)) +  
geom_point()
```

```
ggplot(iris, aes(Sepal.Length, Petal.Length)) + geom_point()
```

とも書けます

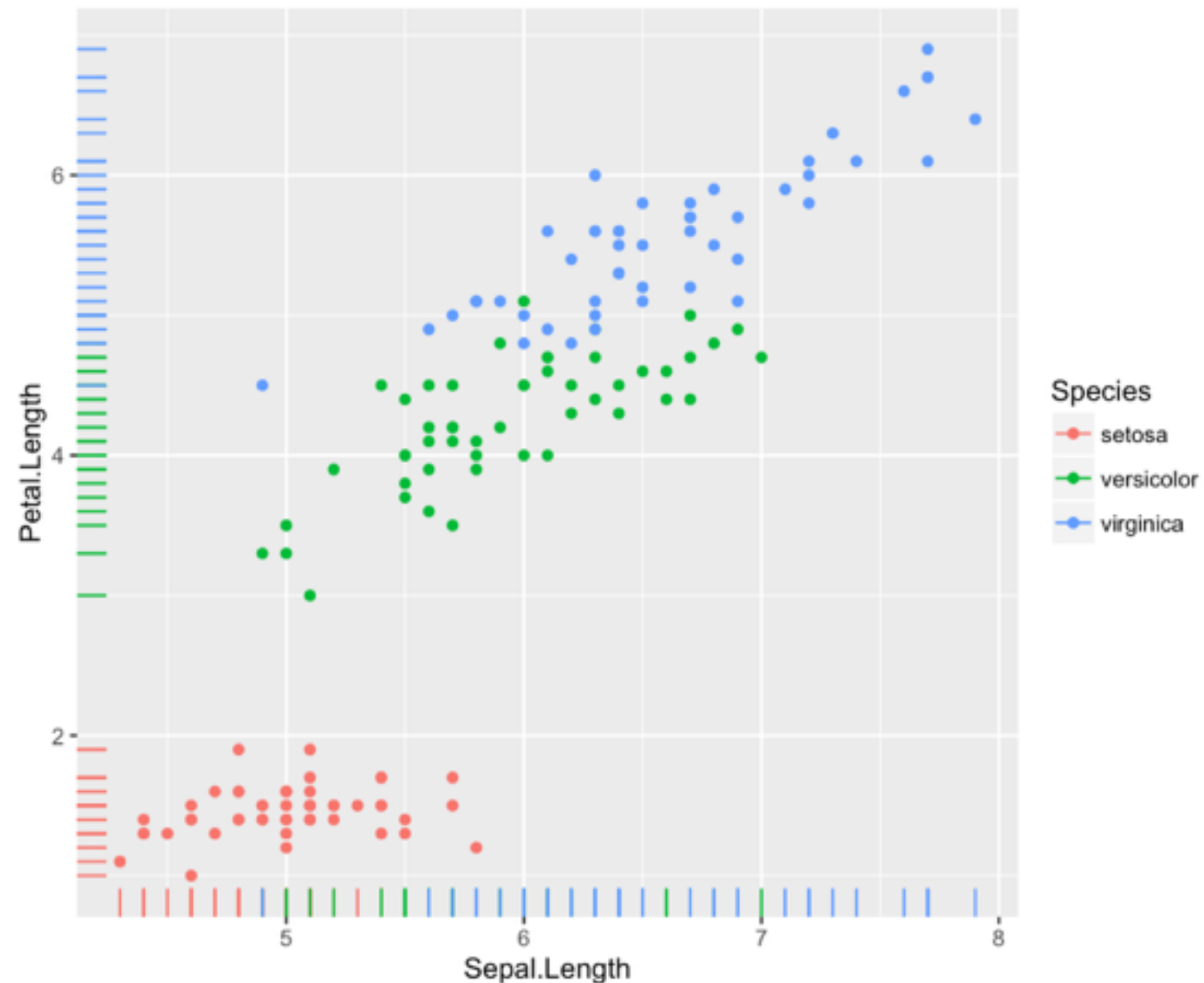
ggplot2 描画例

```
g<-  
ggplot(iris,aes(Sepal.Length,Petal.Length,color=Species))  
# 種類別に色を付ける,  
g + geom_point()
```



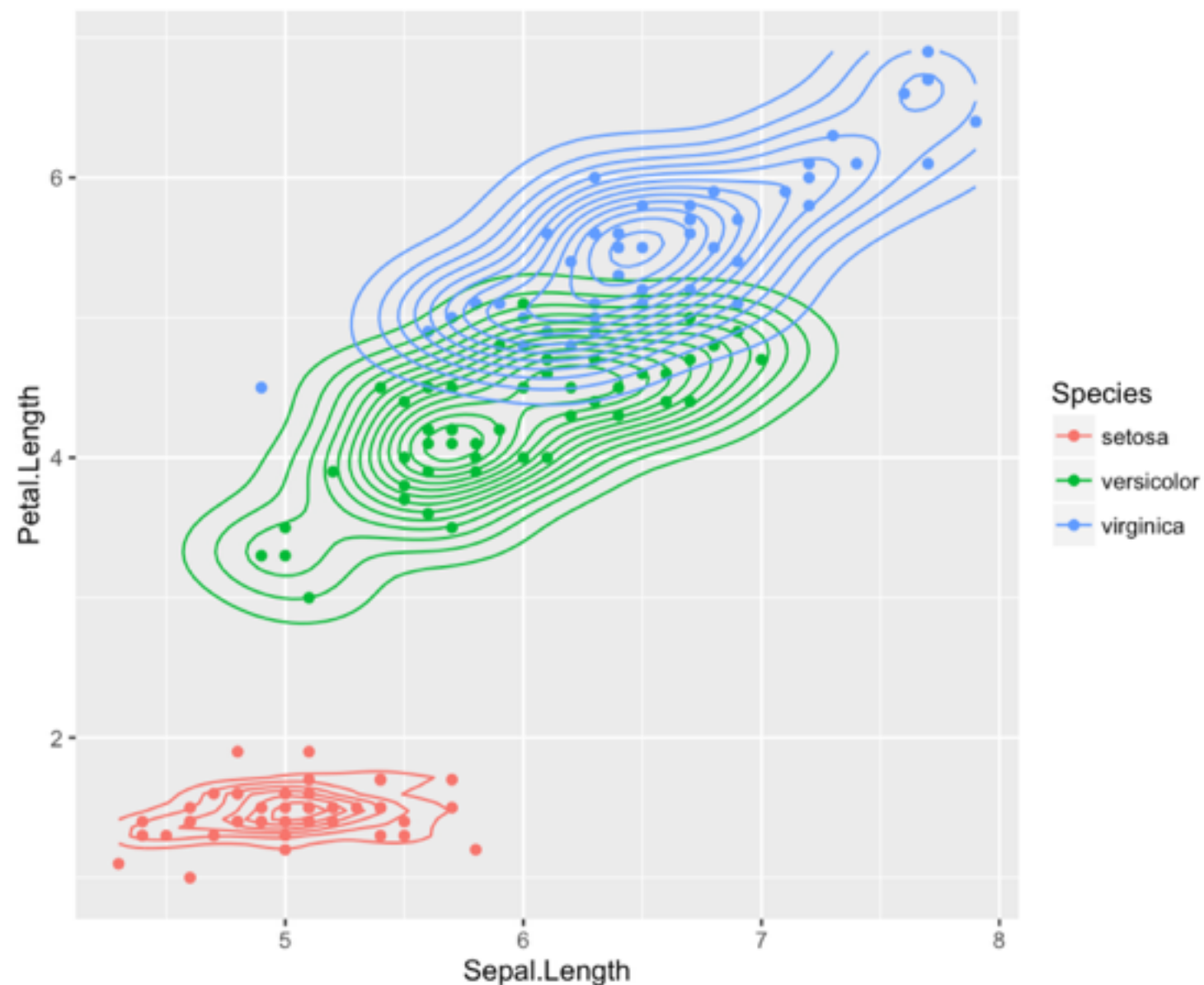
ggplot2 描画例

```
g<-  
ggplot(iris,aes(Sepal.Length,Petal.Length,color=Species))  
g + geom_point() + geom_rug() #ラグプロットを追加
```



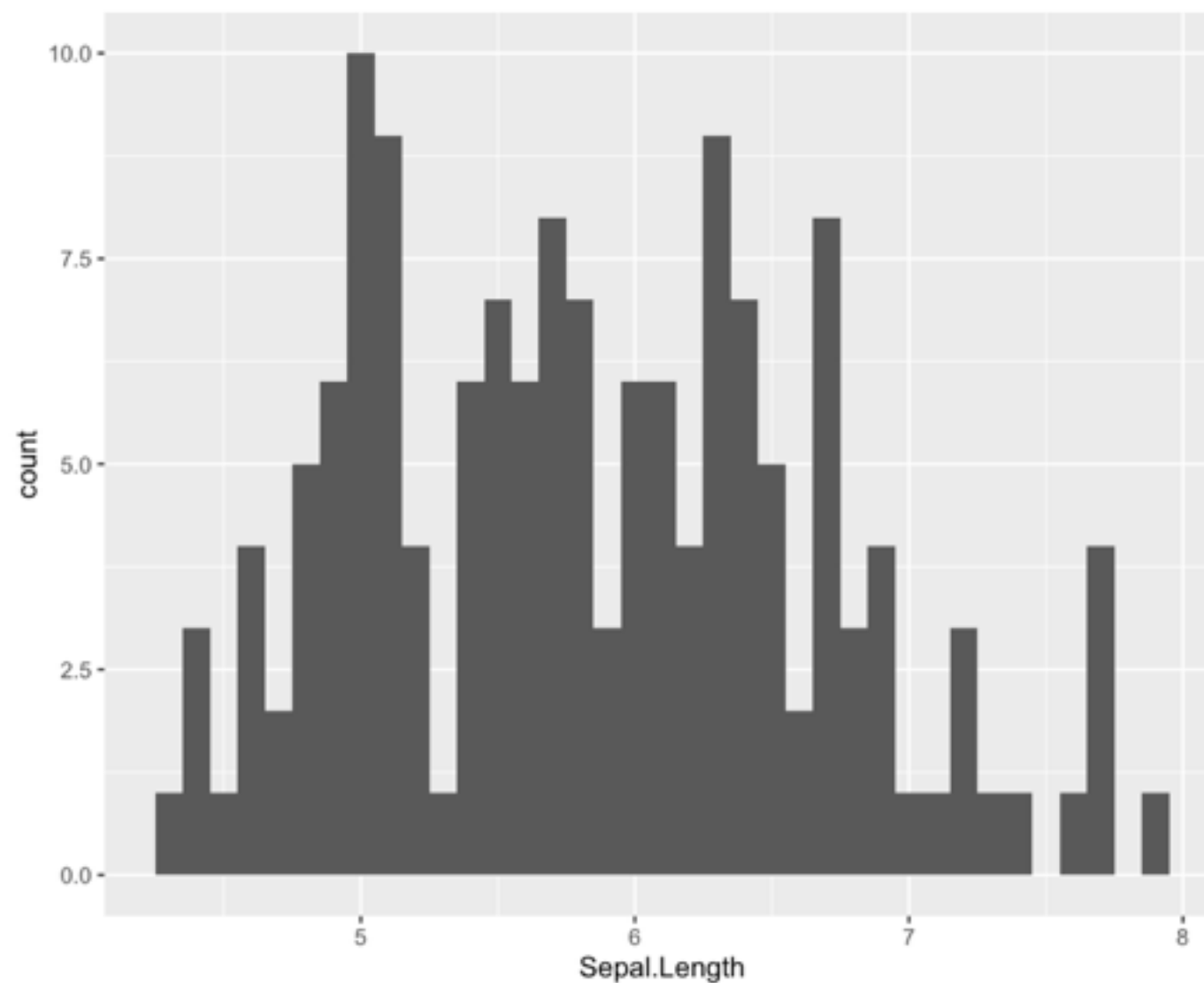
ggplot2 描画例：統計値を使った描画

```
g<-  
ggplot(iris,aes(Sepal.Length,Petal.Length,color=Species))  
g + geom_point() + geom_density2d()  
# 2次元密度分布を描くための関数が用意されている
```



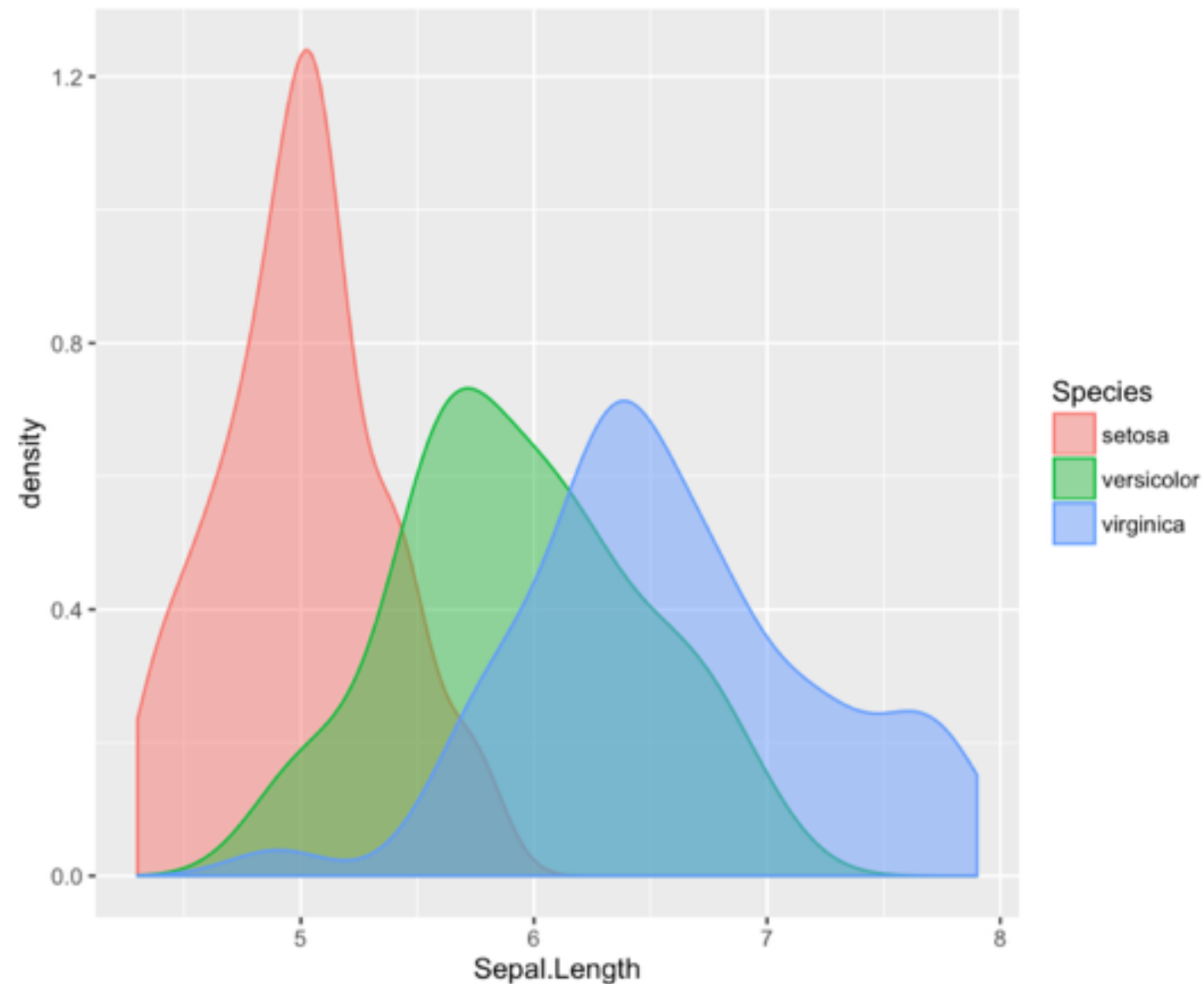
ggplot2 描画例：1変数のプロット

```
g <- ggplot(iris, aes(x=Sepal.Length))  
g + geom_histogram(binwidth=0.1) #がく片のヒストグラム
```



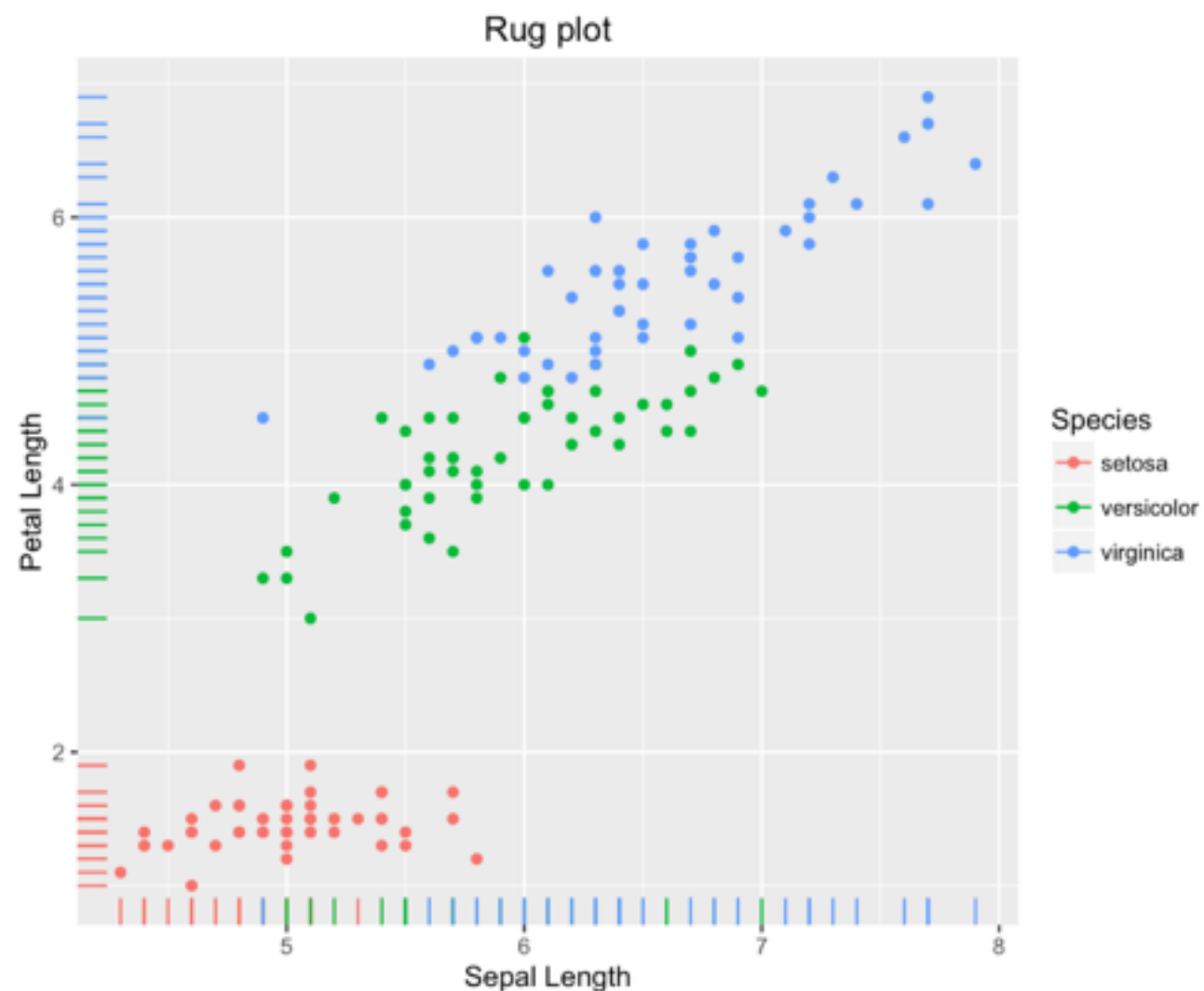
ggplot2 描画例

```
g <-  
ggplot(iris, aes(x=Sepal.Length, col=Species, fill=Species))  
g + geom_density(alpha=0.5)
```



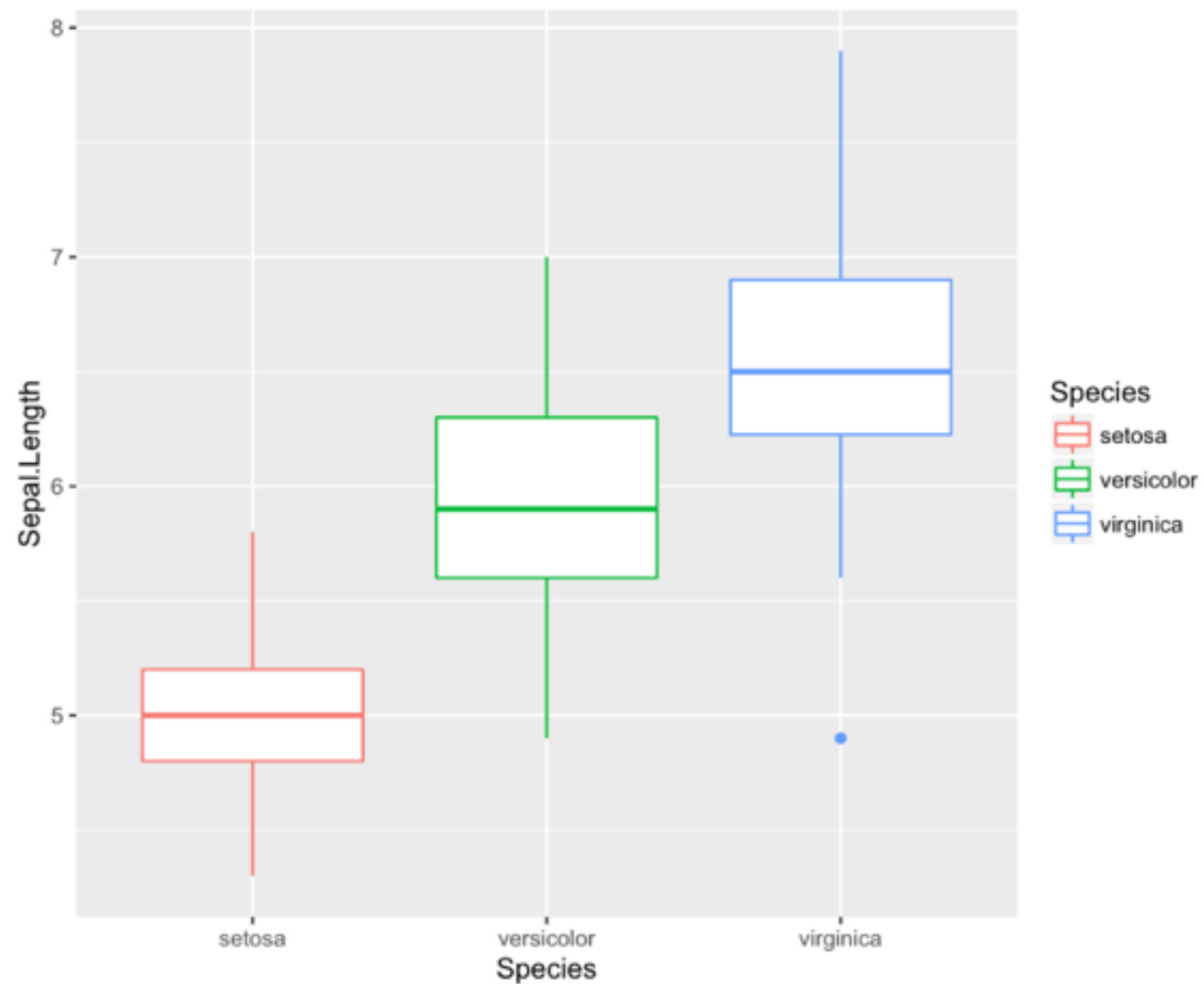
ggplot2 描画例: タイトルやx軸、y軸ラベル

```
g<-  
ggplot(iris,aes(Sepal.Length,Petal.Length,color=Species))  
g + geom_point() + geom_rug() +  
  ggtitle("Rug plot") + xlab("Sepal Length") +  
  ylab("Petal Length")
```



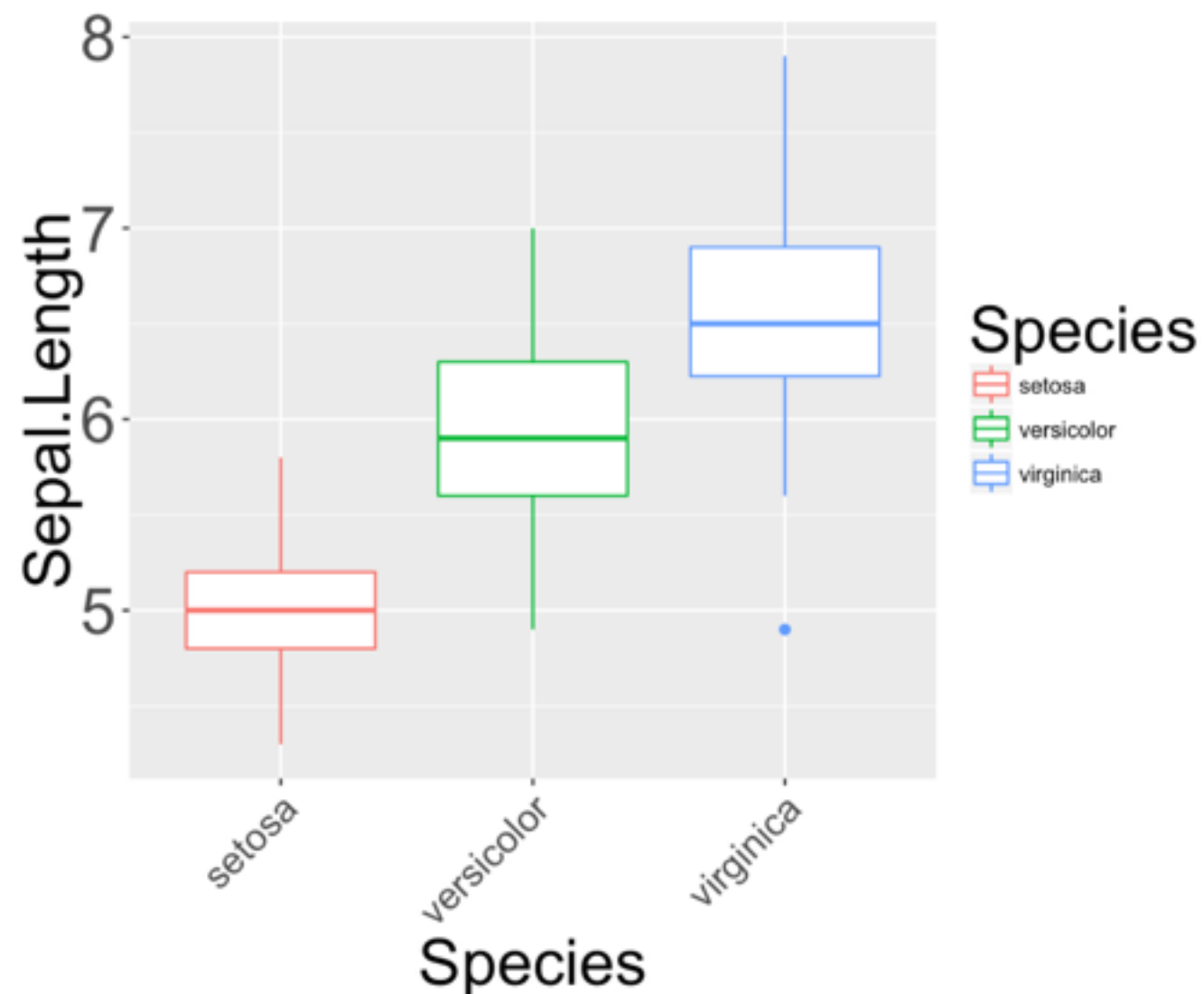
ggplot2 描画例

```
g <- ggplot(iris, aes(Species, Sepal.Length, color=Species))  
g + geom_boxplot() # geom_boxplot() に変えるだけ
```



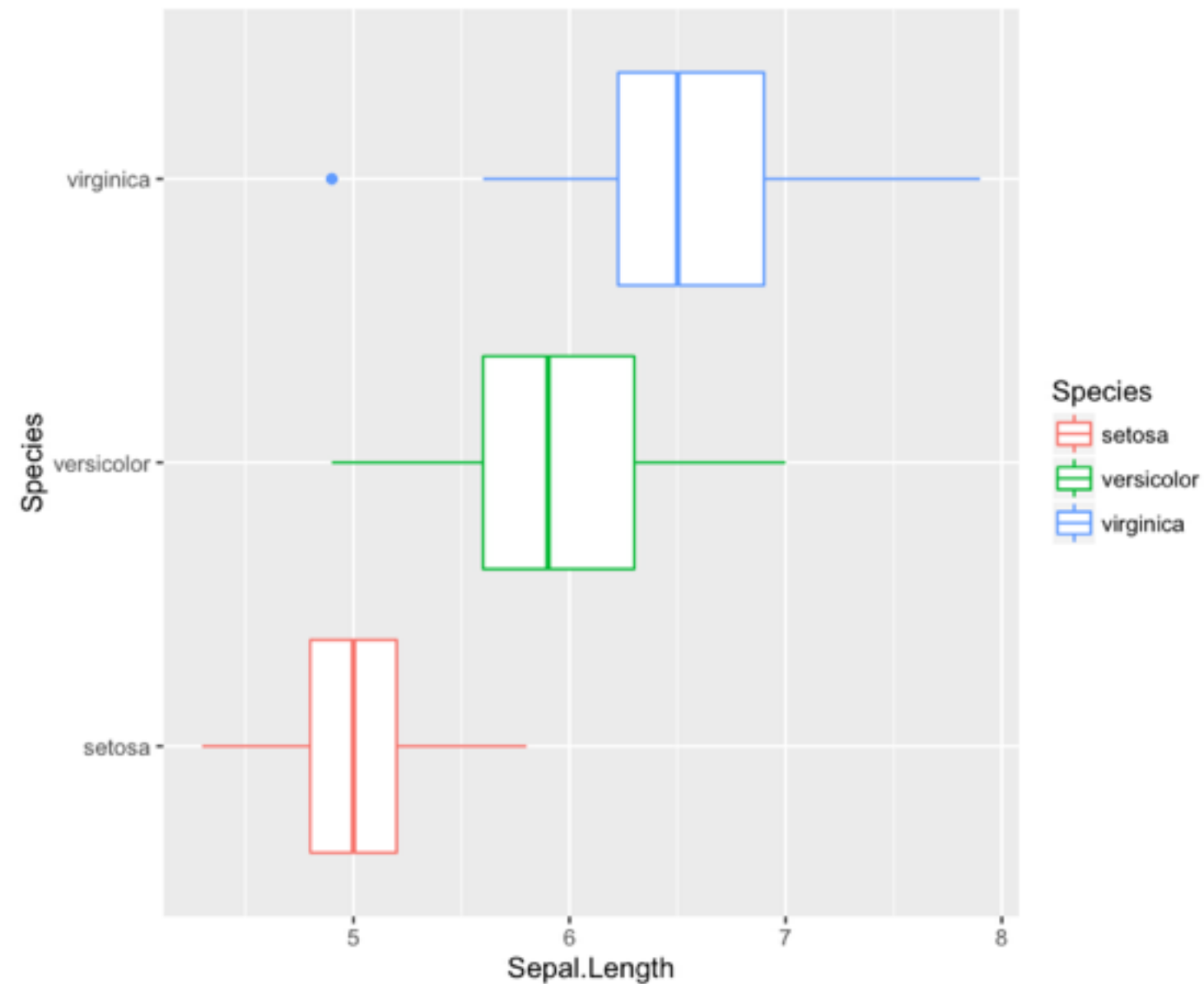
ggplot2 描画例

```
ggplot(iris, aes(Species, Sepal.Length, color=Species)) +  
  geom_boxplot() + theme(axis.text.y=element_text(size=25),  
    axis.text.x=element_text(angle=45, hjust=1, size=15),  
    title=element_text(size=25))
```



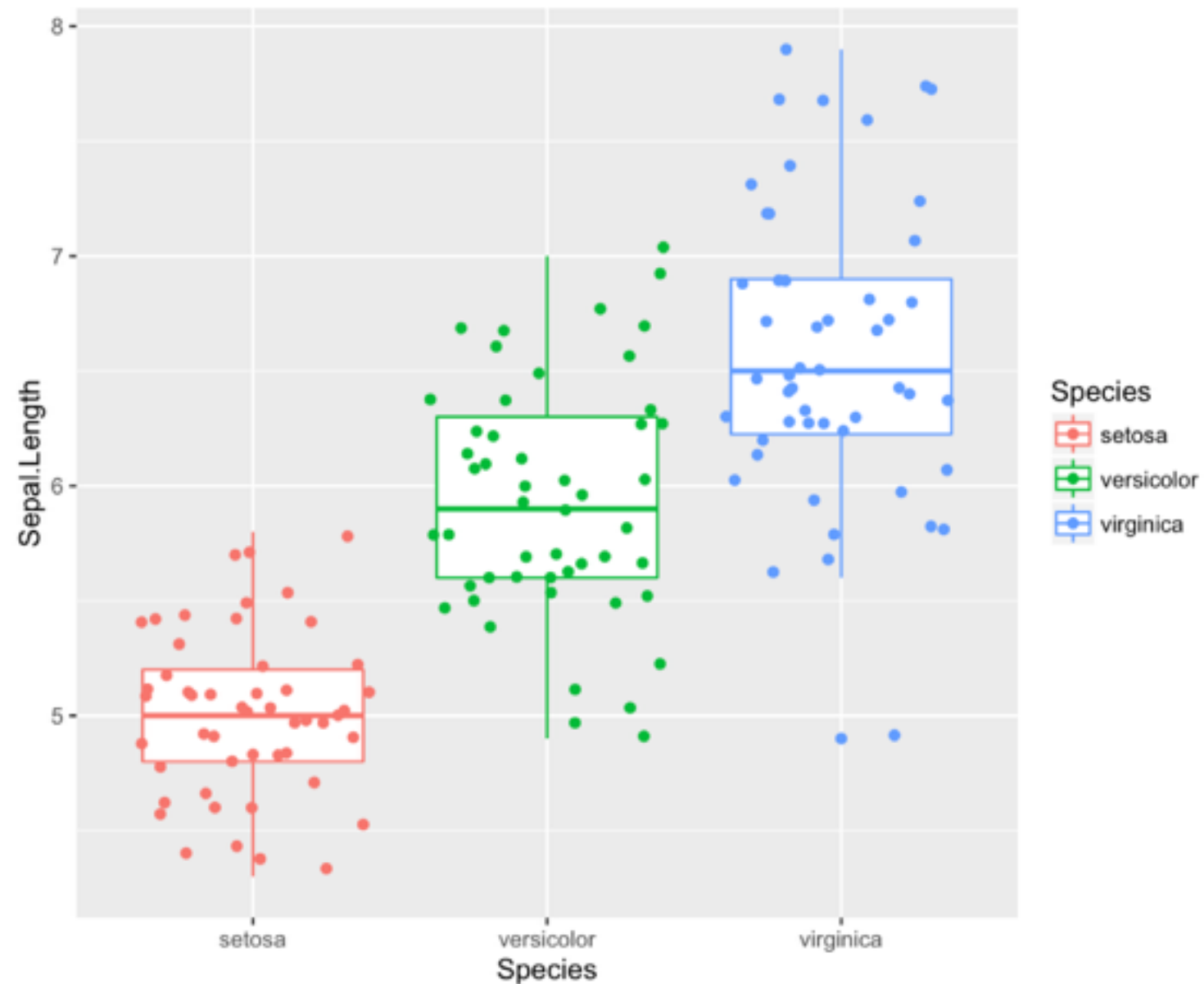
ggplot2 描画例

```
g <- ggplot(iris, aes(Species, Sepal.Length, color=Species))  
g + geom_boxplot() + coord_flip() # 横に変換
```



ggplot2 描画例

```
g <- ggplot(iris, aes(Species, Sepal.Length, color=Species))  
g + geom_boxplot() + geom_jitter()
```

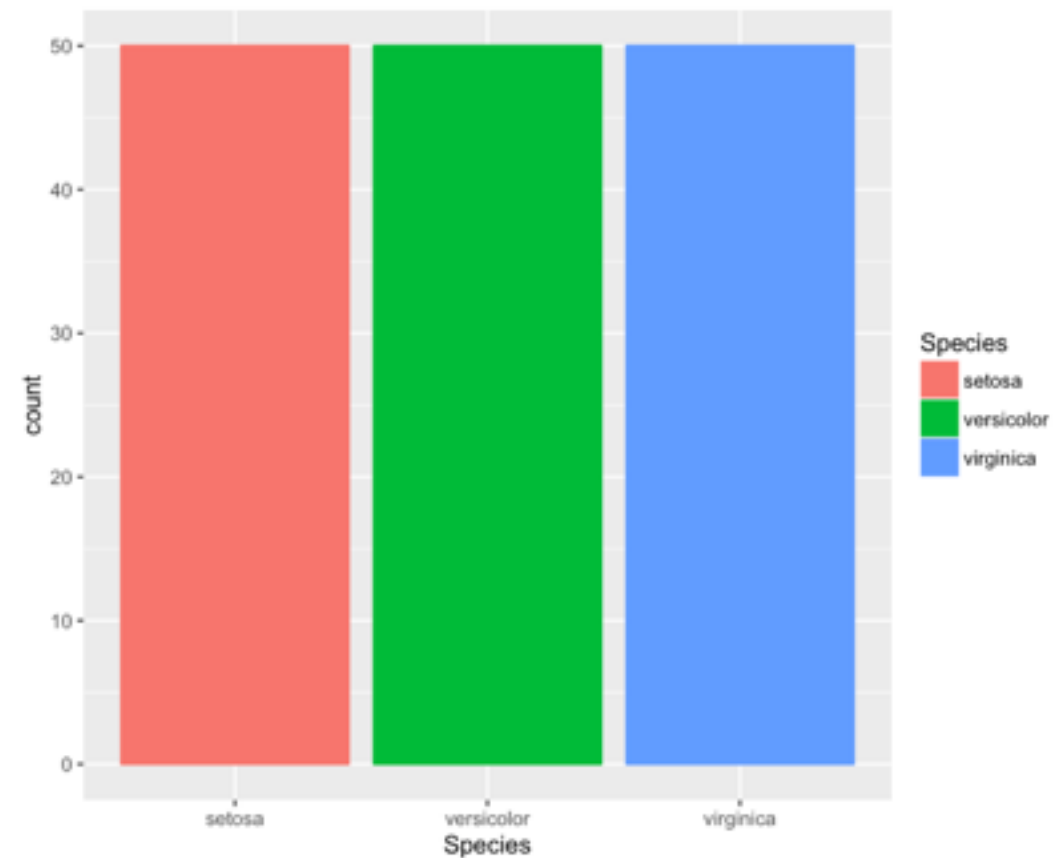
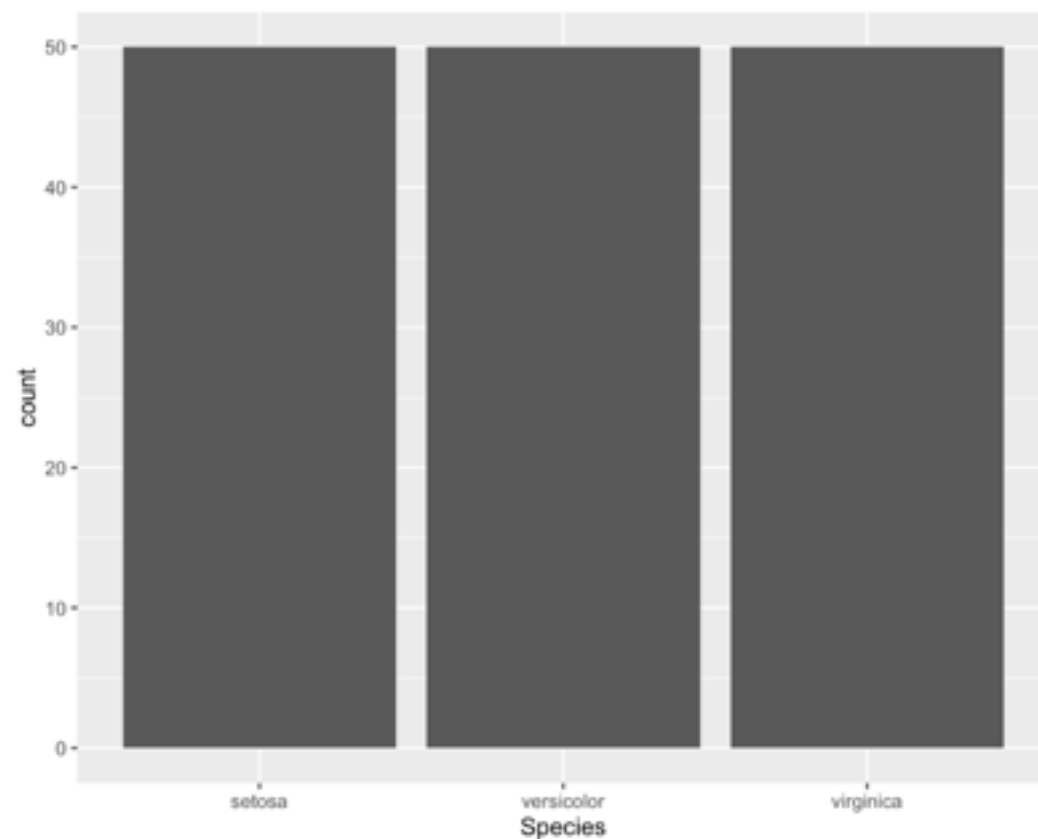


ggplot2 描画例

```
g <- ggplot(iris, aes(Species))  
g + geom_bar()
```

colは枠線、fillはぬりつぶす色

```
g <- ggplot(iris, aes(Species, col=Species, fill=Species))  
g + geom_bar()
```

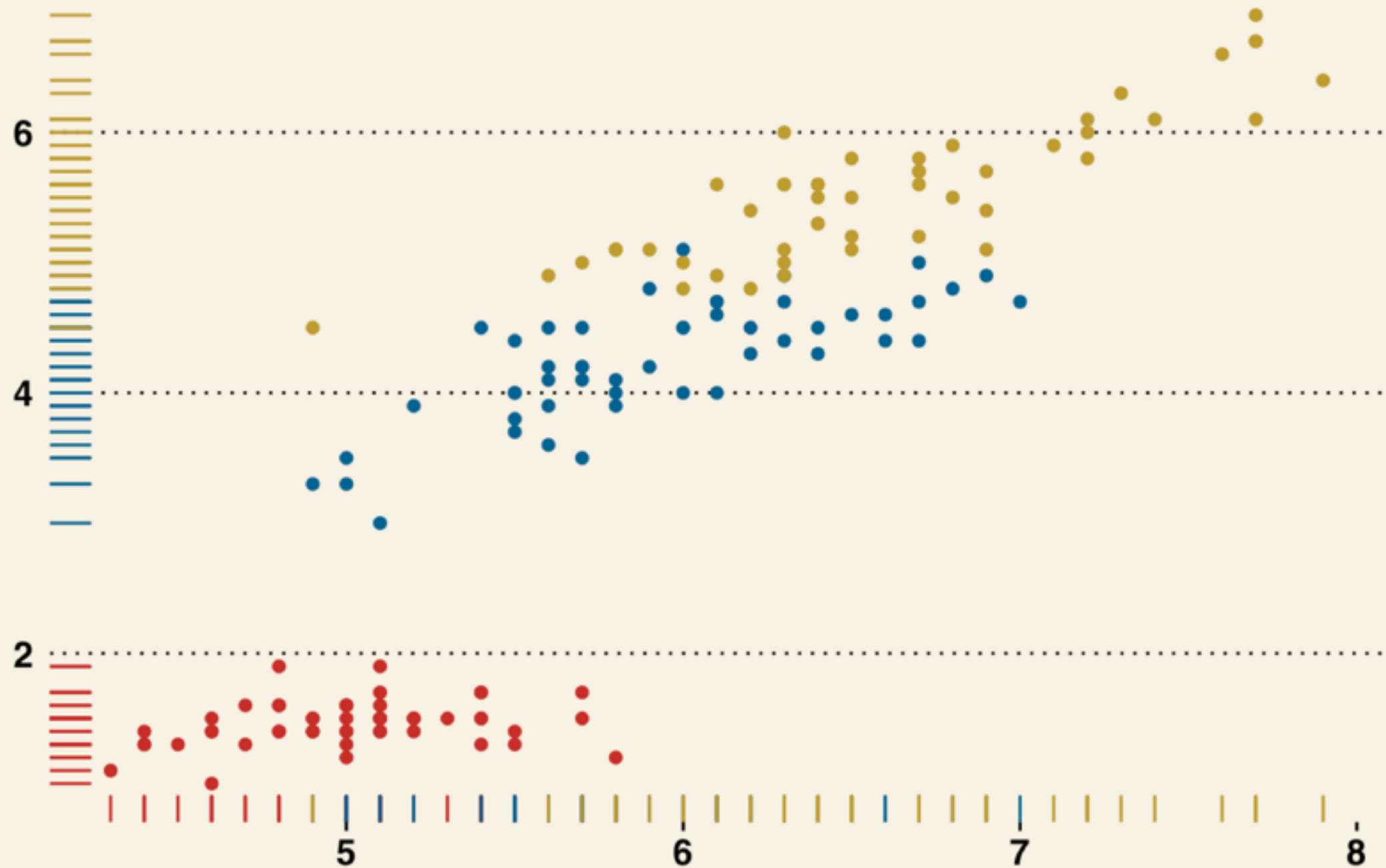


ggplot2 描画例: テンプレート (ggtheme)

```
install.packages("ggthemes")  
library(ggthemes)  
g<-  
ggplot(iris,aes(Sepal.Length,Petal.Length,color=Species))  
g + geom_point() + geom_rug() + ggtitle("Rug plot") +  
  xlab("Sepal Length") + ylab("Petal Length") +  
  theme_wsj() + # Wall street journal  
  scale_colour_wsj() # ポイントの色もWSJ風になる
```

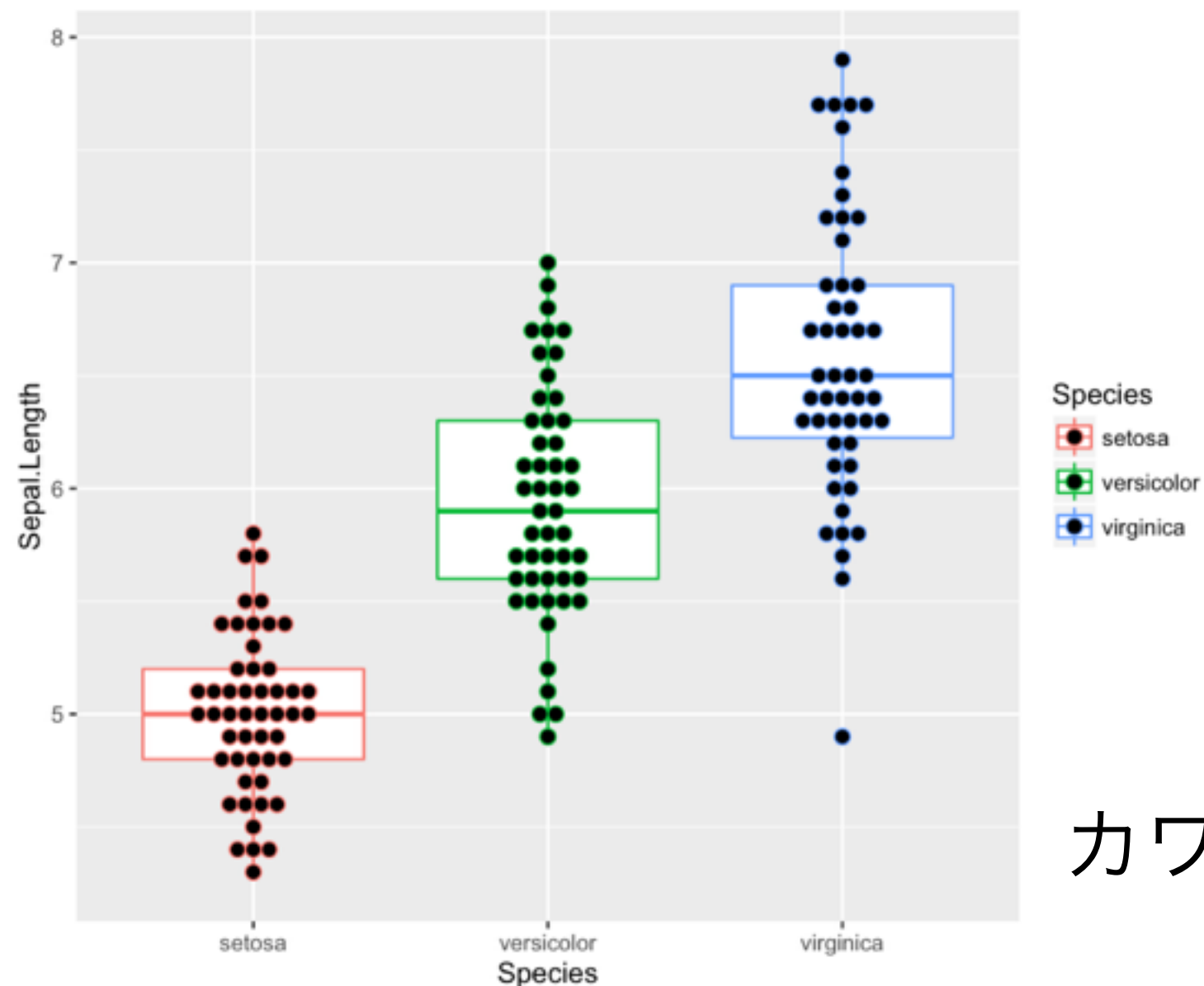
Rug plot

Species ● setosa ● versicolor ● virginica



ggplot2 描画例: ありがちな落とし穴

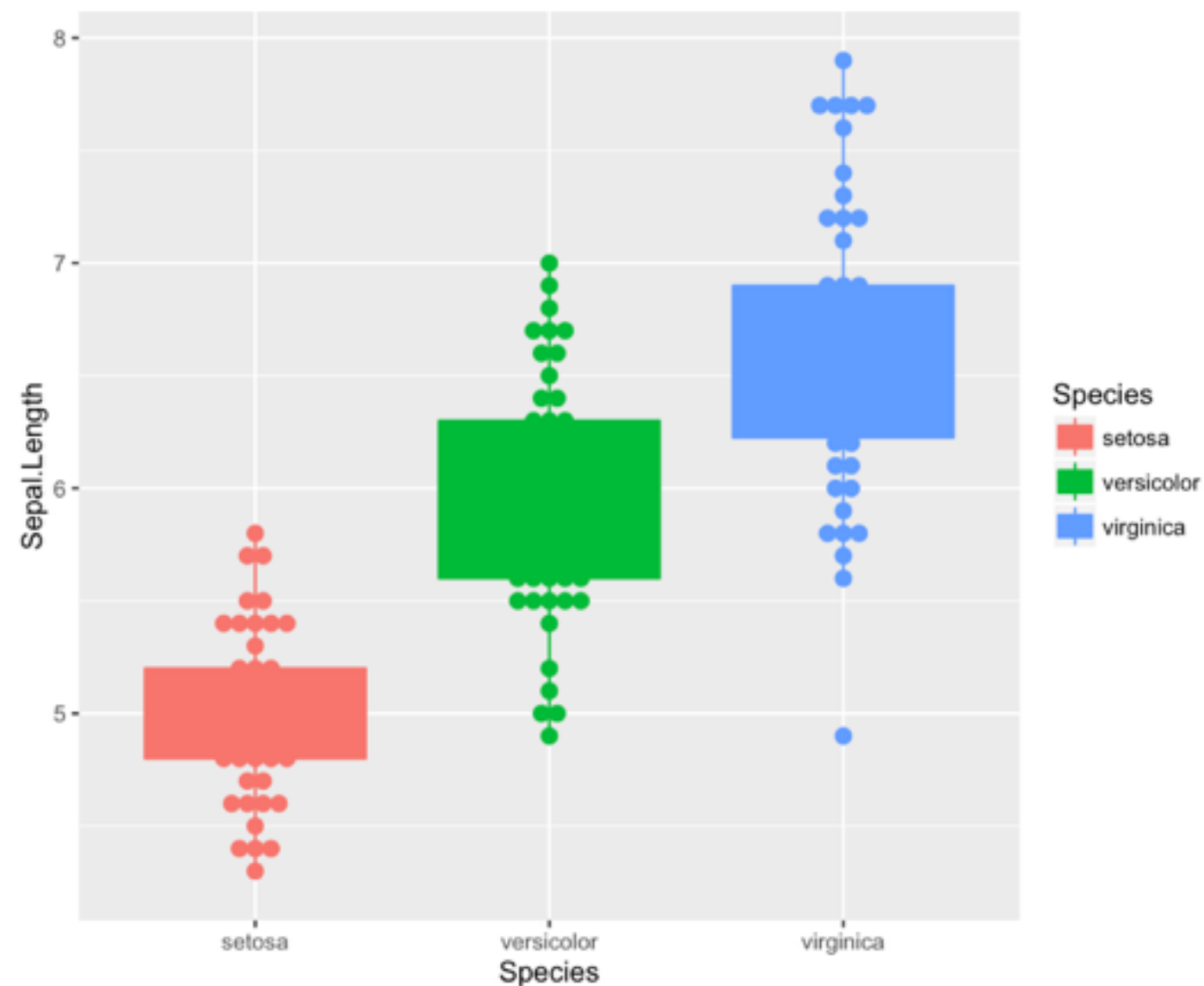
```
g<- ggplot(iris,aes(Species,Sepal.Length,color=Species))  
g + geom_boxplot() + geom_dotplot(binwidth=  
0.07,binaxis="y",stackdir="center")
```



カワイくない！

ggplot2 描画例: ありがちな落とし穴

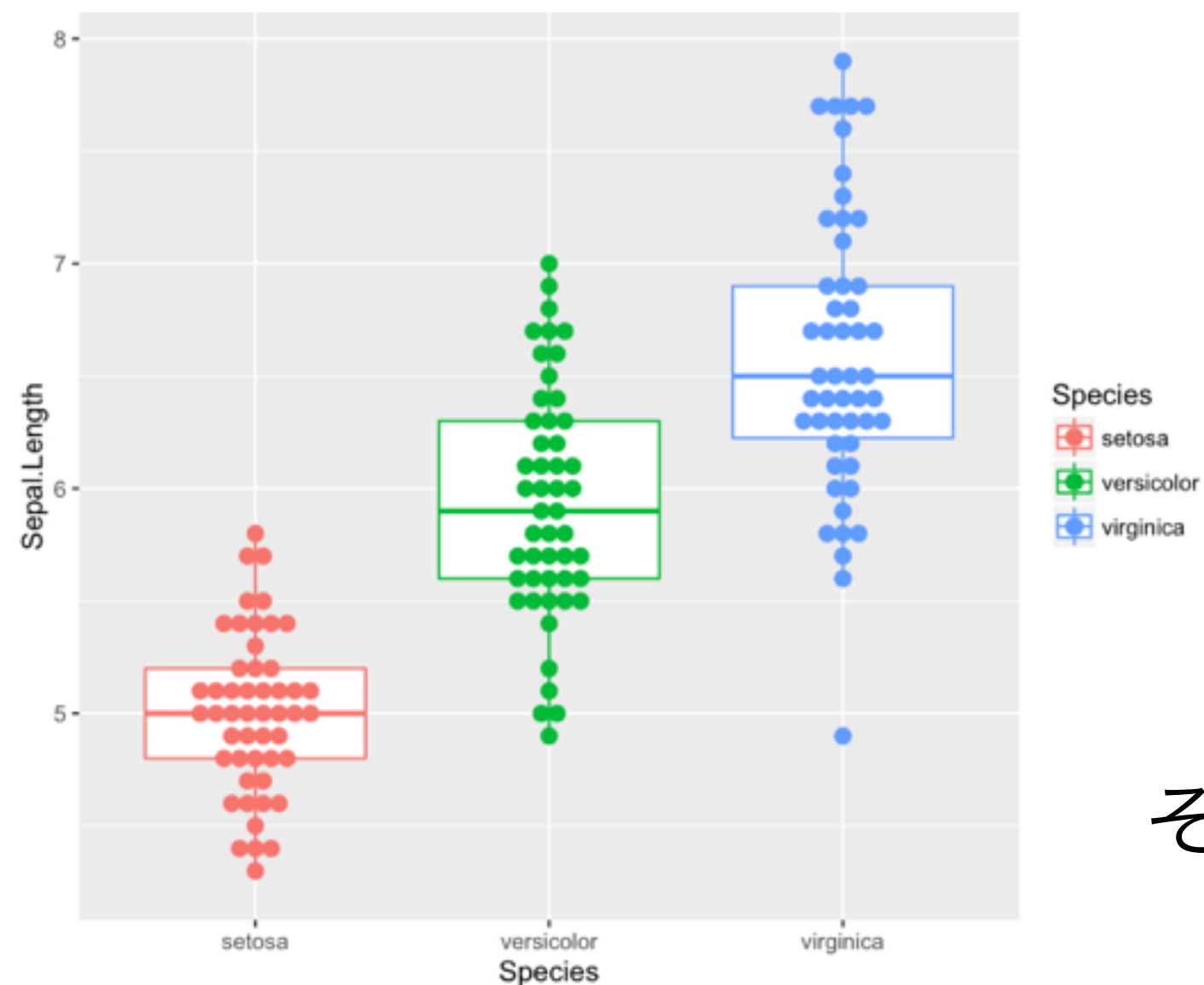
```
g<-  
ggplot(iris,aes(Species,Sepal.Length,color=Species,fill=Species))  
g + geom_boxplot() + geom_dotplot(binwidth=  
0.07,binaxis="y",stackdir="center")
```



違う！

ggplot2 描画例: ありがちな落とし穴

```
g<- ggplot(iris,aes(Species,Sepal.Length,color=Species))  
g + geom_boxplot() +  
geom_dotplot(aes(fill=Species),binwidth=  
0.07,binaxis="y",stackdir="center")
```



そうこれ！

もっと勉強したい！(私もですが・・・)

- ・Google で検索すると色々出てくるので、それを1行ずつ実行して描画を確認。
- ・かなりの事が既にWebで解決済み。英語で質問をそのまま打てば、だいたい誰かがQAサイトに質問してる。

