

Genome_Omics_Final_Assignment

このマニュアルではRのグラフィックツールであるggplot2を用いたグラフ作成の基本を初心者向けに解説します。

ggplot2でグラフを作製するには、グラフ上の一つの点に対応するデータを以下のように行毎にまとめます。

例として、10個のコントロールと10個のサンプルの蛍光強度を測定したデータセットを用いて以下進めていきます。

ここではエクセルで作製したデータセットを読み込んでいます。データを読み込んだ時にはCell_Typeは”文字型”として読み込まれるので、”要素型”に変換しておきます。

```
library(ggplot2)
#Import dataset
data<- read.delim("~/Desktop/Genome_Omics_class/final_assignment.csv")
data
```

```
##      Intensity Cell_Type
## 1         8.45   control
## 2         9.80   control
## 3         8.80   control
## 4        10.41   control
## 5         7.37   control
## 6        11.75   control
## 7        12.50   control
## 8         8.95   control
## 9        12.33   control
## 10        10.51   sample
## 11         7.22   sample
## 12         5.62   sample
## 13         6.93   sample
## 14         7.13   sample
## 15         8.33   sample
## 16         9.76   sample
## 17         4.88   sample
## 18        10.48   sample
## 19        10.92   sample
## 20        10.84   sample
```

```
#Change the class of the second column to "factor" class
data$Cell_Type<- factor(data$Cell_Type, levels=c("control","sample"))
```

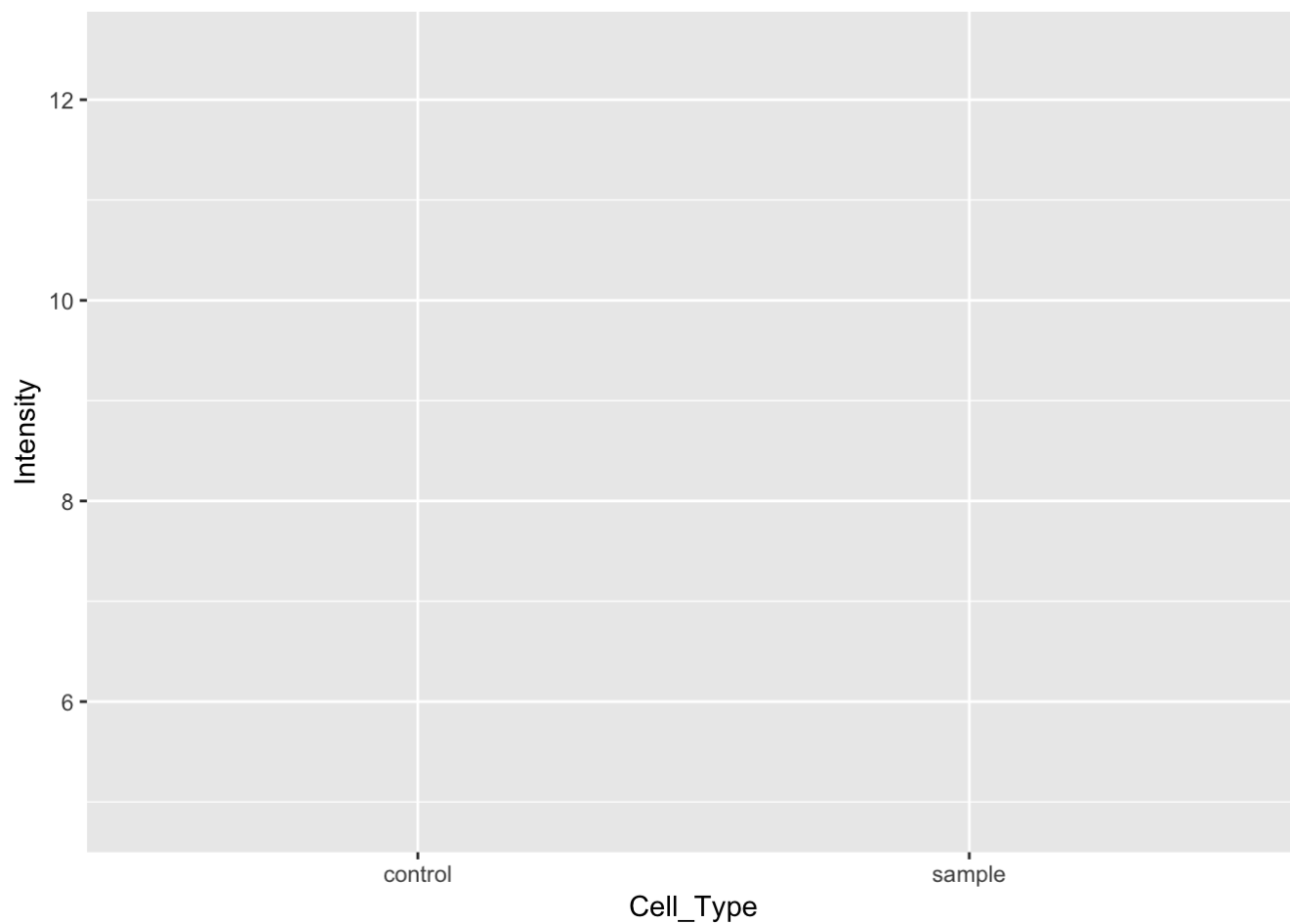
ここからグラフを書いていきます。

ggplot2では基本的に図を重ね合わせていくことで作図を行います。

”+”をつけることでグラフ内に図形を足し合わせていくことが可能です。

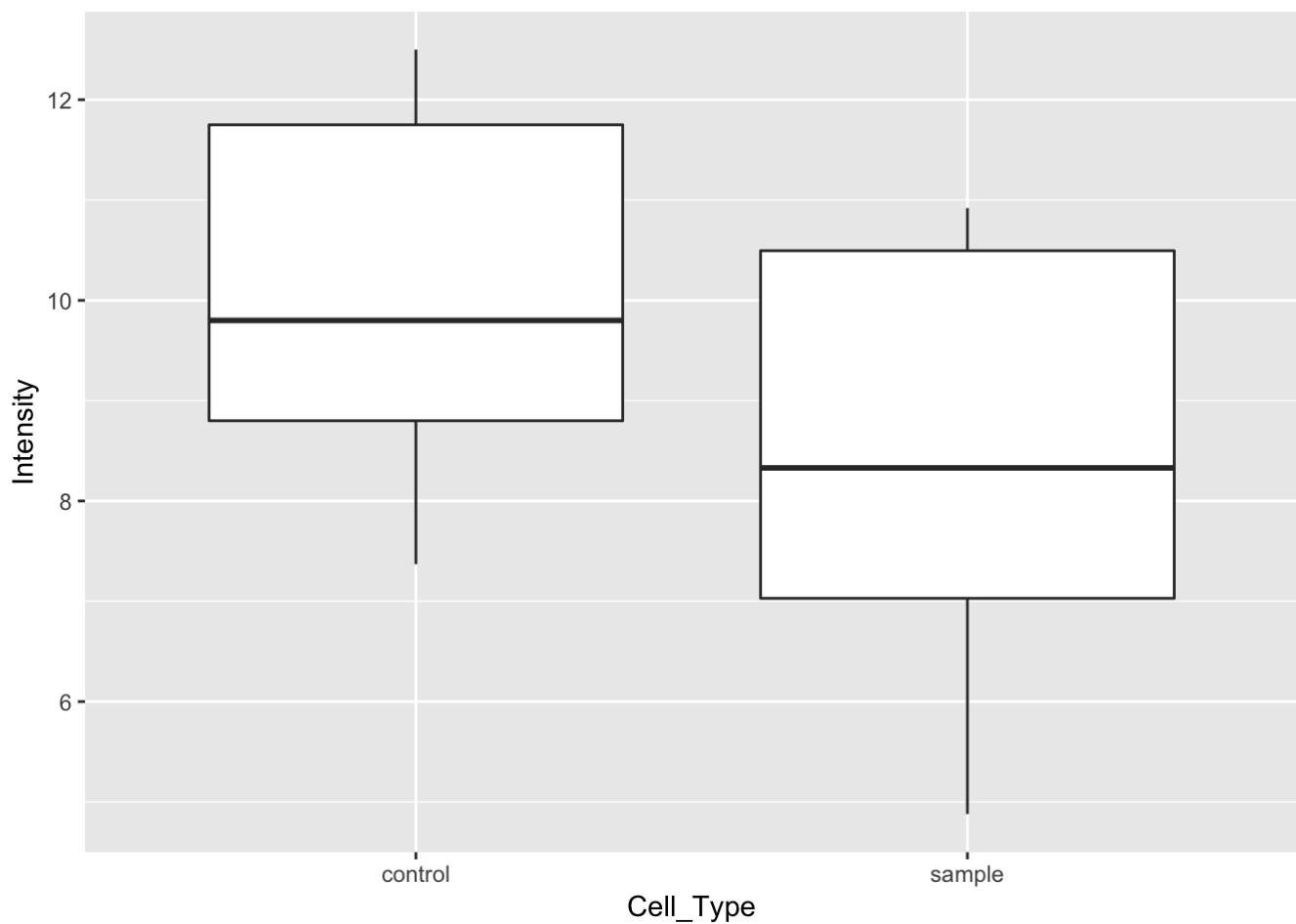
まずはデータ、X軸、Y軸を指定します。

```
p <- ggplot(data, aes(x=Cell_Type, y=Intensity))
p
```



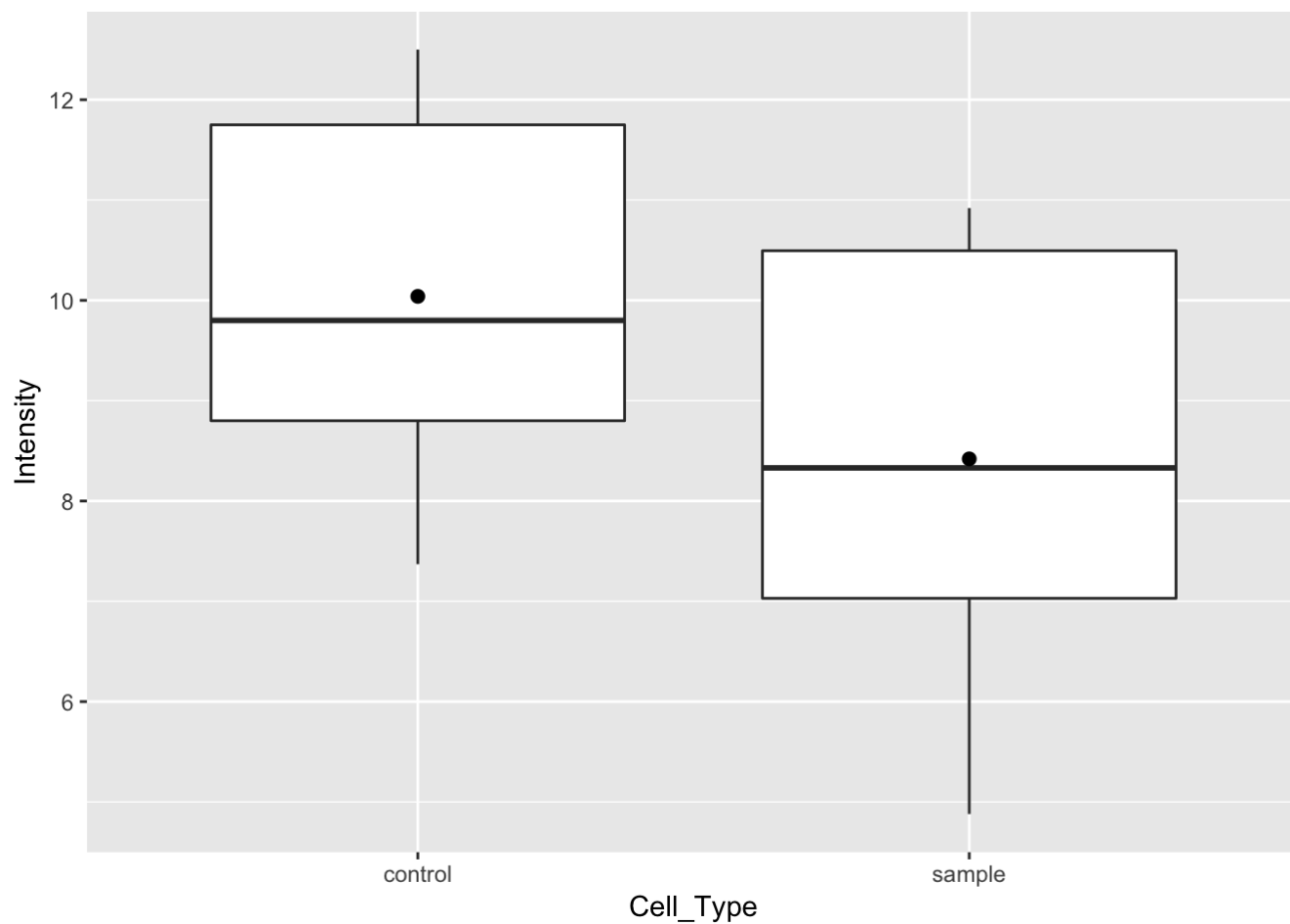
基本となる箱ひげ図を描いてみます。

```
p <- p + geom_boxplot()  
p
```



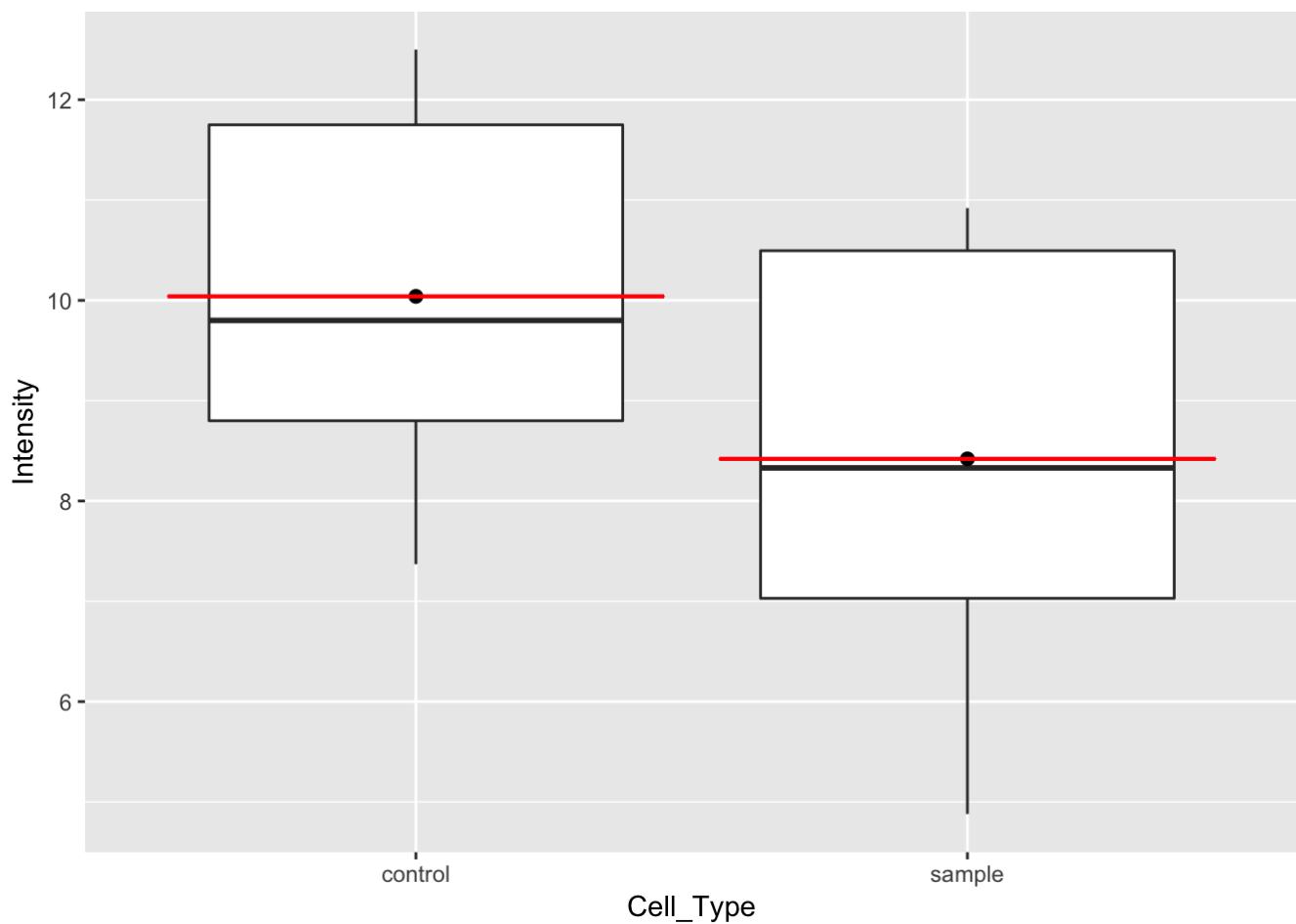
箱ひげ図内に平均値を表す点を書き加えたい場合は以下のように行います。

```
p <- p + stat_summary(fun.y=mean, geom="point", size=2)
p
```



線として書き加えることも可能です。

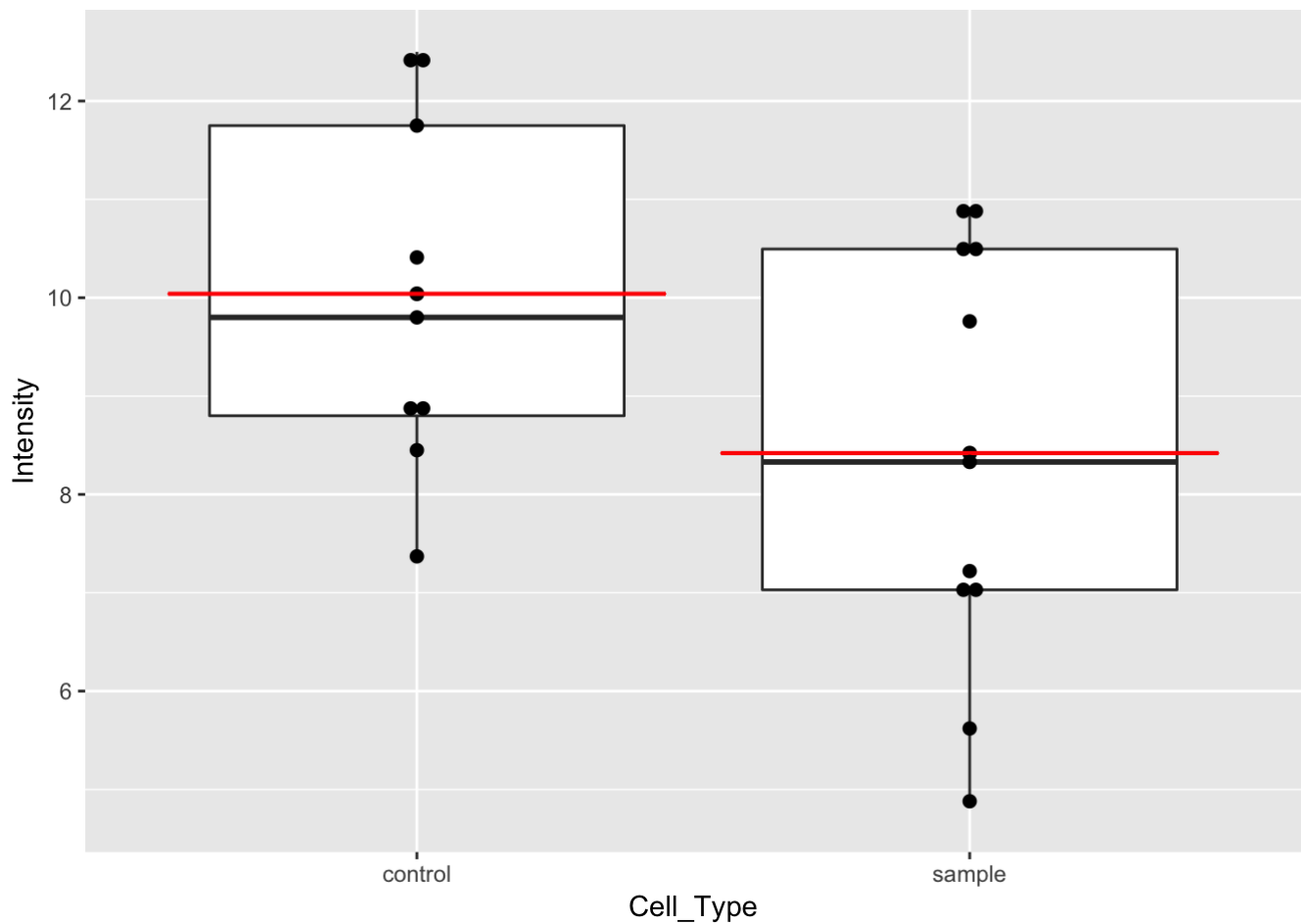
```
p <-p + stat_summary(fun.y = mean, fun.ymin = mean, fun.ymax = mean, colour = "red", size = 0.3, geom = "crossbar", show.legend = F)
p
```



このサンプルの分布を見るために箱ひげ図の中にドットプロットを追加することも出来ます。

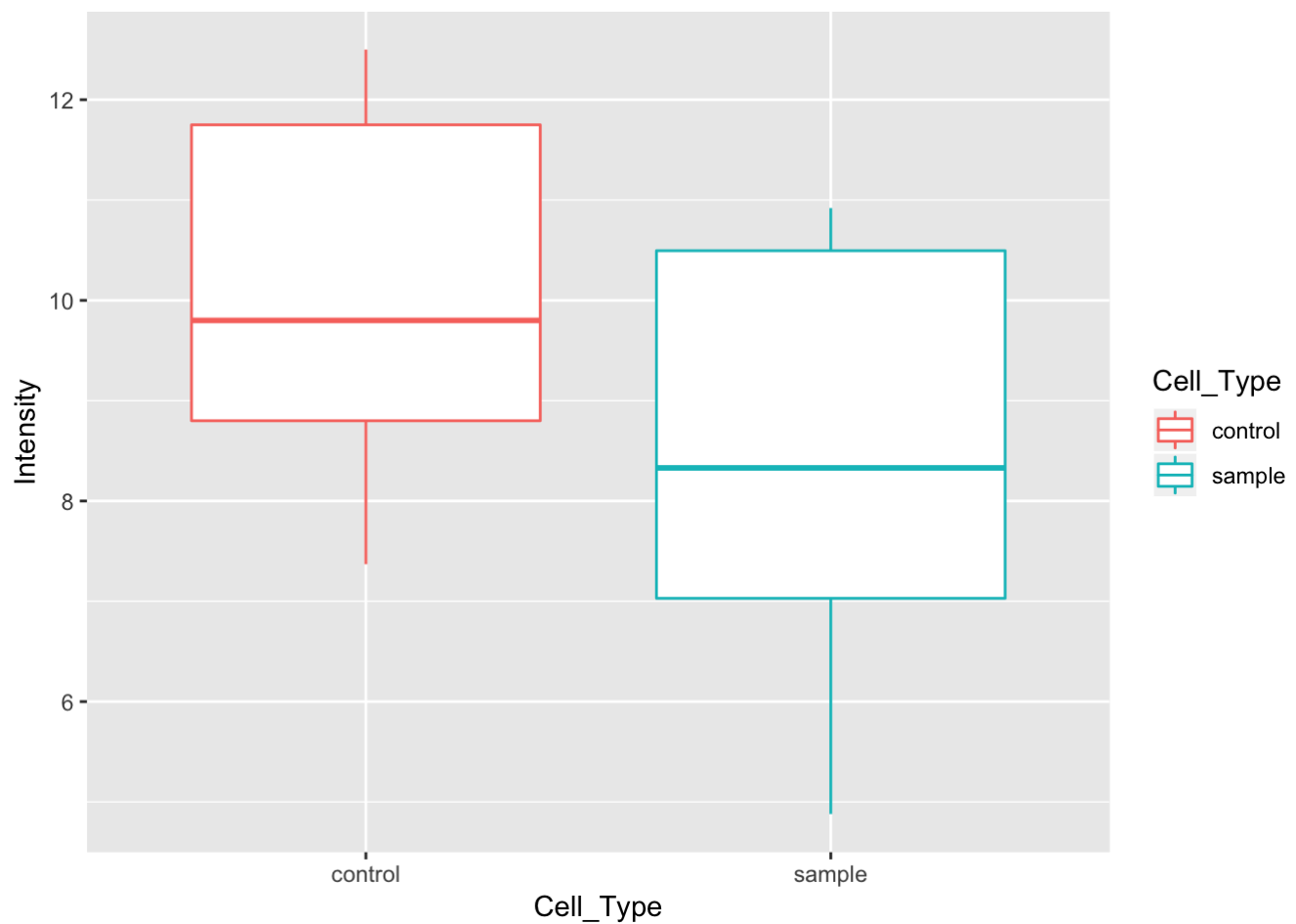
```
p<- p + geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5)
p
```

```
## `stat_bindot()` using `bins = 30`. Pick better value with `binwidth`.
```



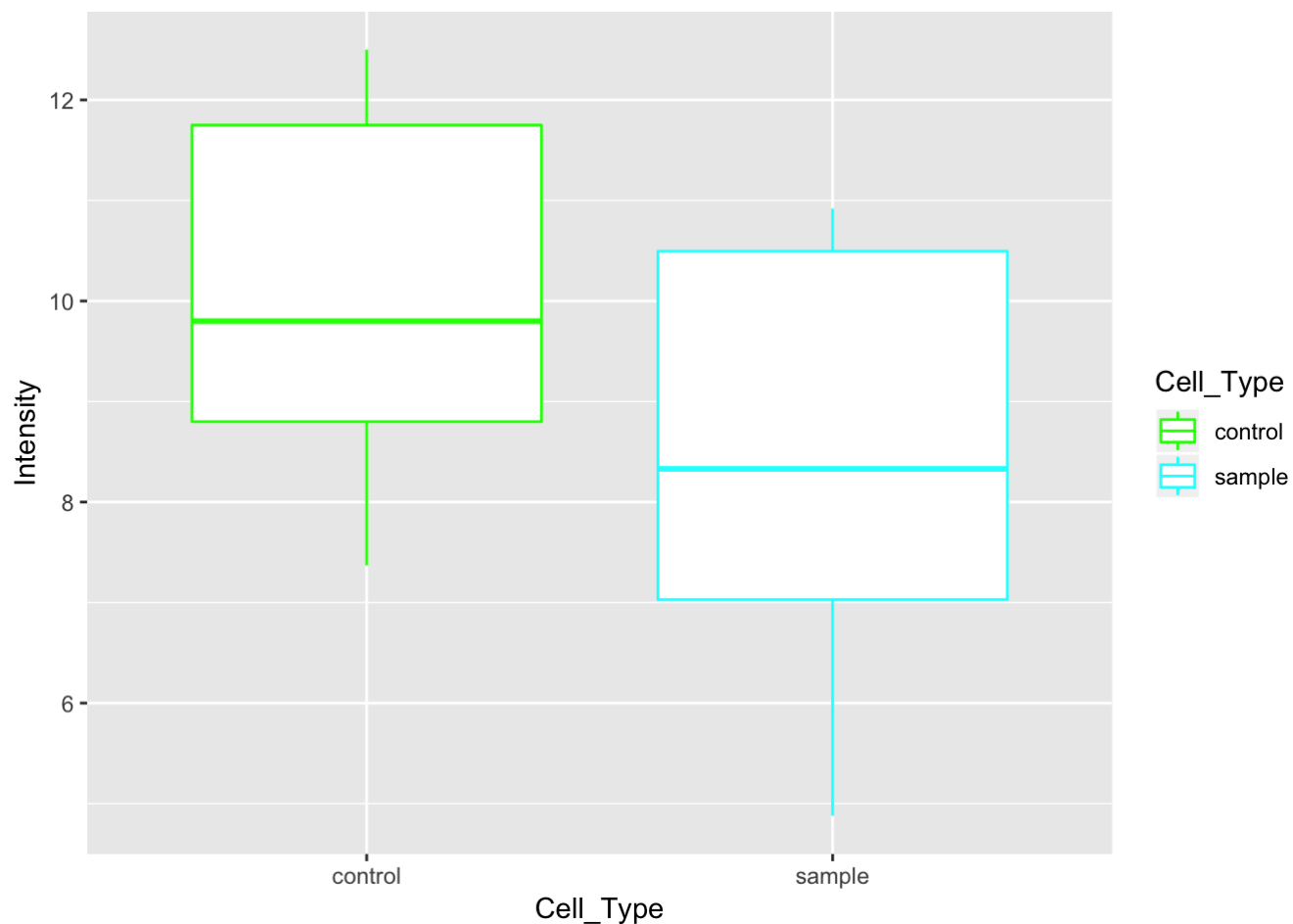
また箱毎の色を変えるには以下のようにします。

```
p <- ggplot(data, aes(x=Cell_Type, y=Intensity, color= Cell_Type)) +  
  geom_boxplot()  
p
```



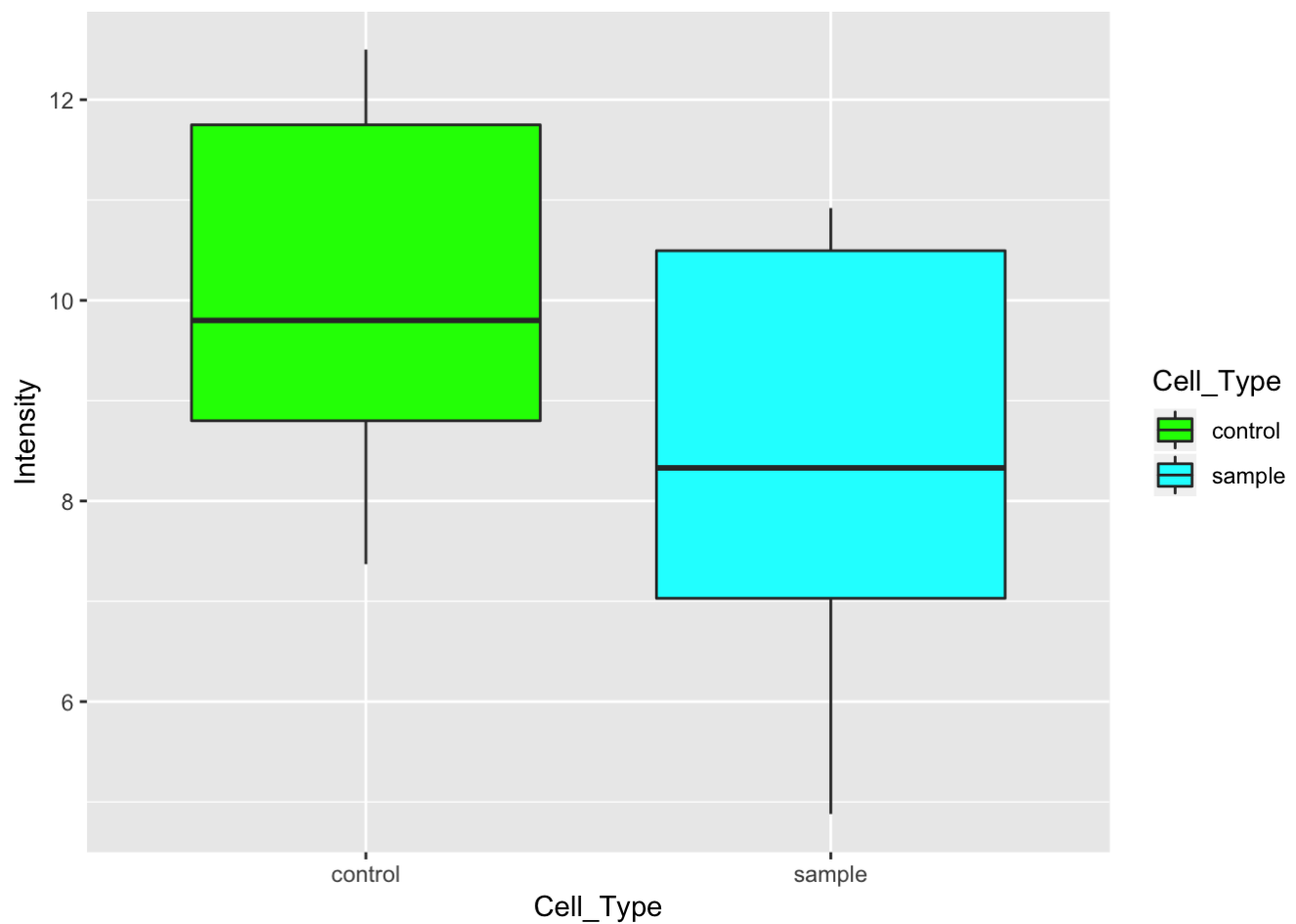
マニュアルでRGBカラーを指定することも可能です。

```
p <- p+scale_color_manual(values=c("#00FF00", "#00FFFF"))  
p
```



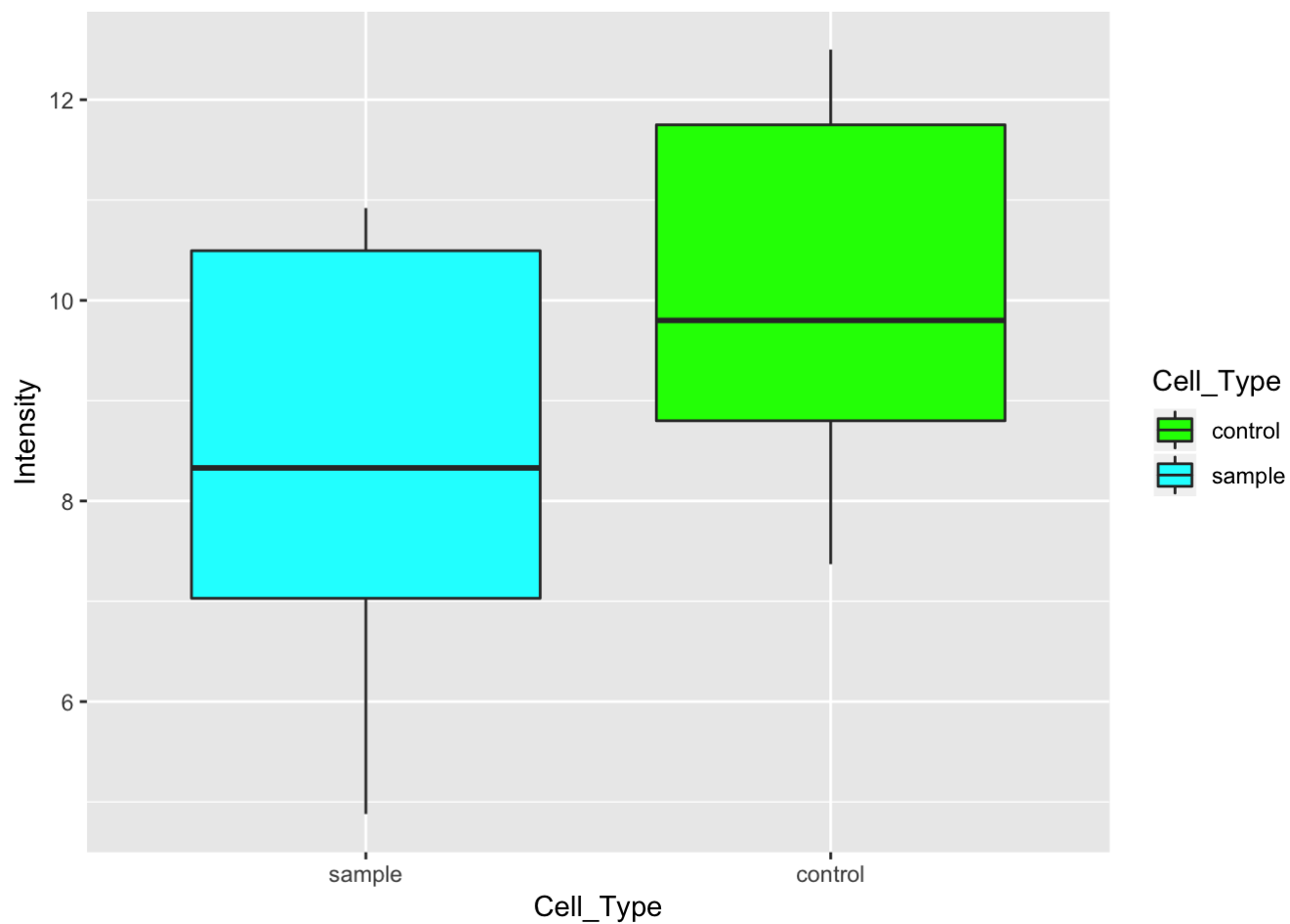
また箱の中の塗りつぶしの色は"fill"を指定することで変更します。ここでは同時にマニュアルで色を指定しています。

```
p <- ggplot(data, aes(x=Cell_Type, y=Intensity, fill= Cell_Type)) +  
  geom_boxplot() +  
  scale_fill_manual(values=c("#00FF00", "#00FFFF"))  
p
```

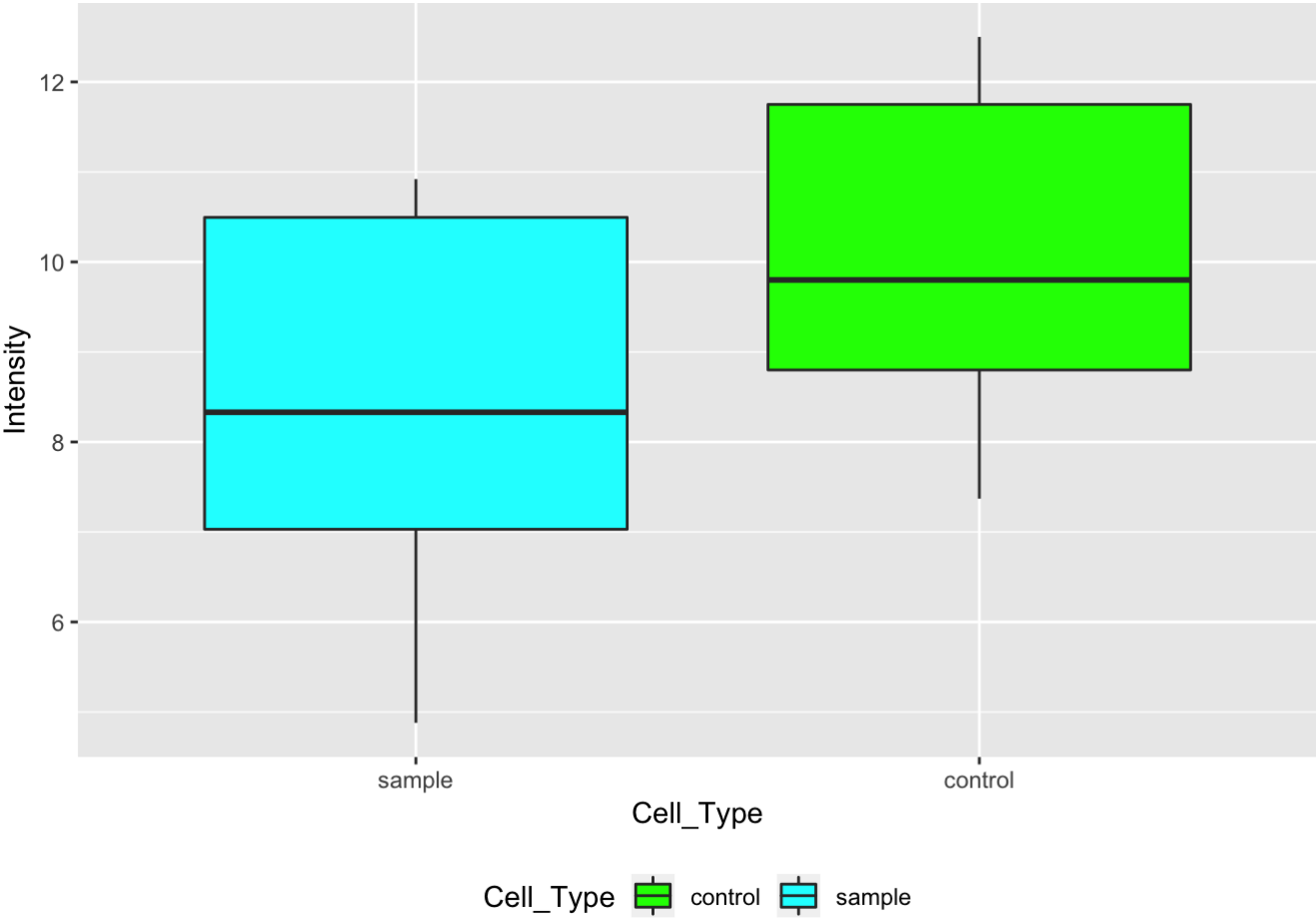
X軸の順番を入れ替えるには以下のように行います。

```
p <- p + scale_x_discrete(limits=c("sample", "control"))  
p
```

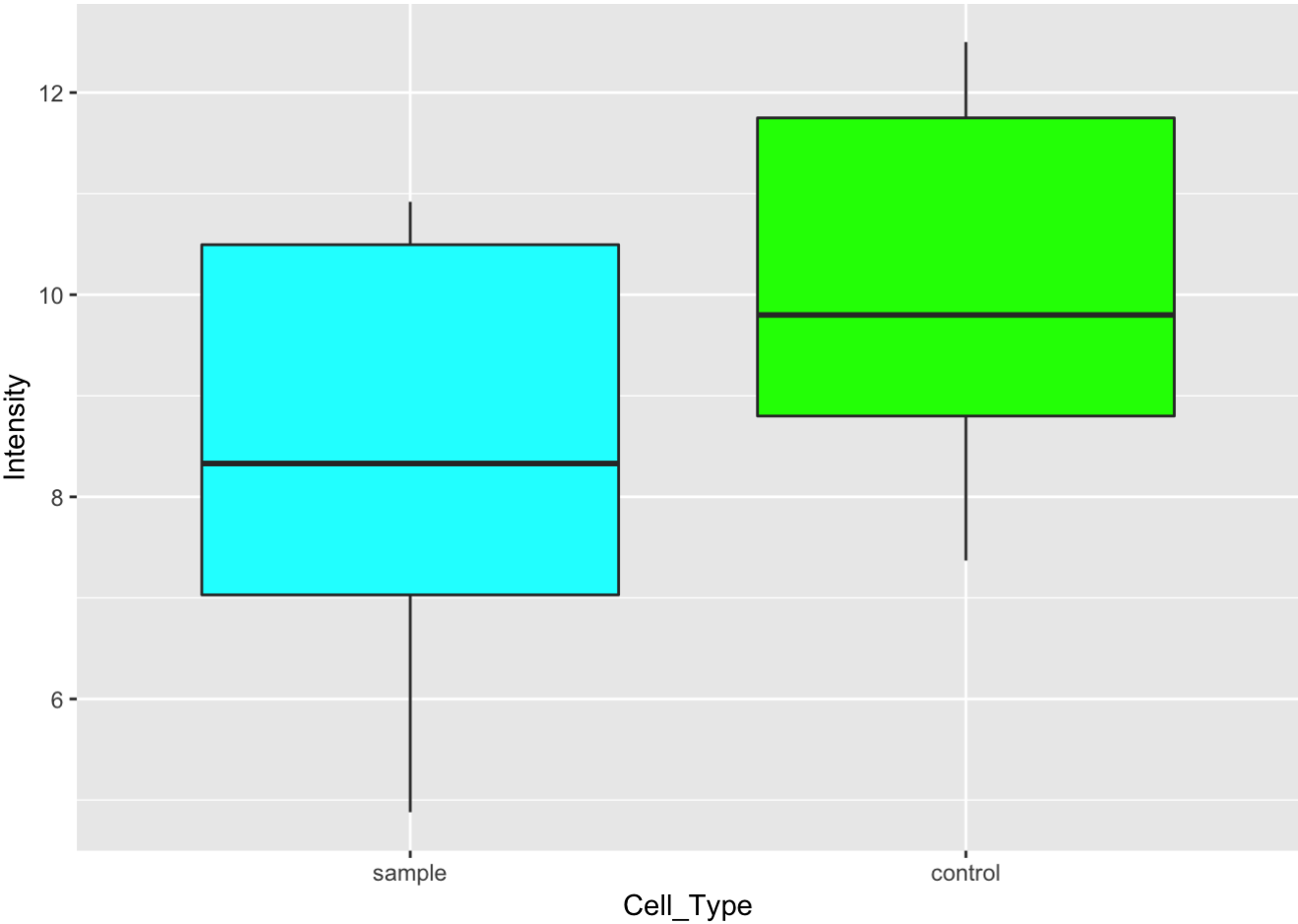


レジェンドの位置は上下左右に変えることも出来ます。"none"を指定することで書かないことも出来ます。

```
p <- p + theme(legend.position="bottom") # The allowed values for the arguments legend.position are :  
  "left", "top", "right", "bottom".  
p
```



```
p<- p + theme(legend.position="none")
p
```

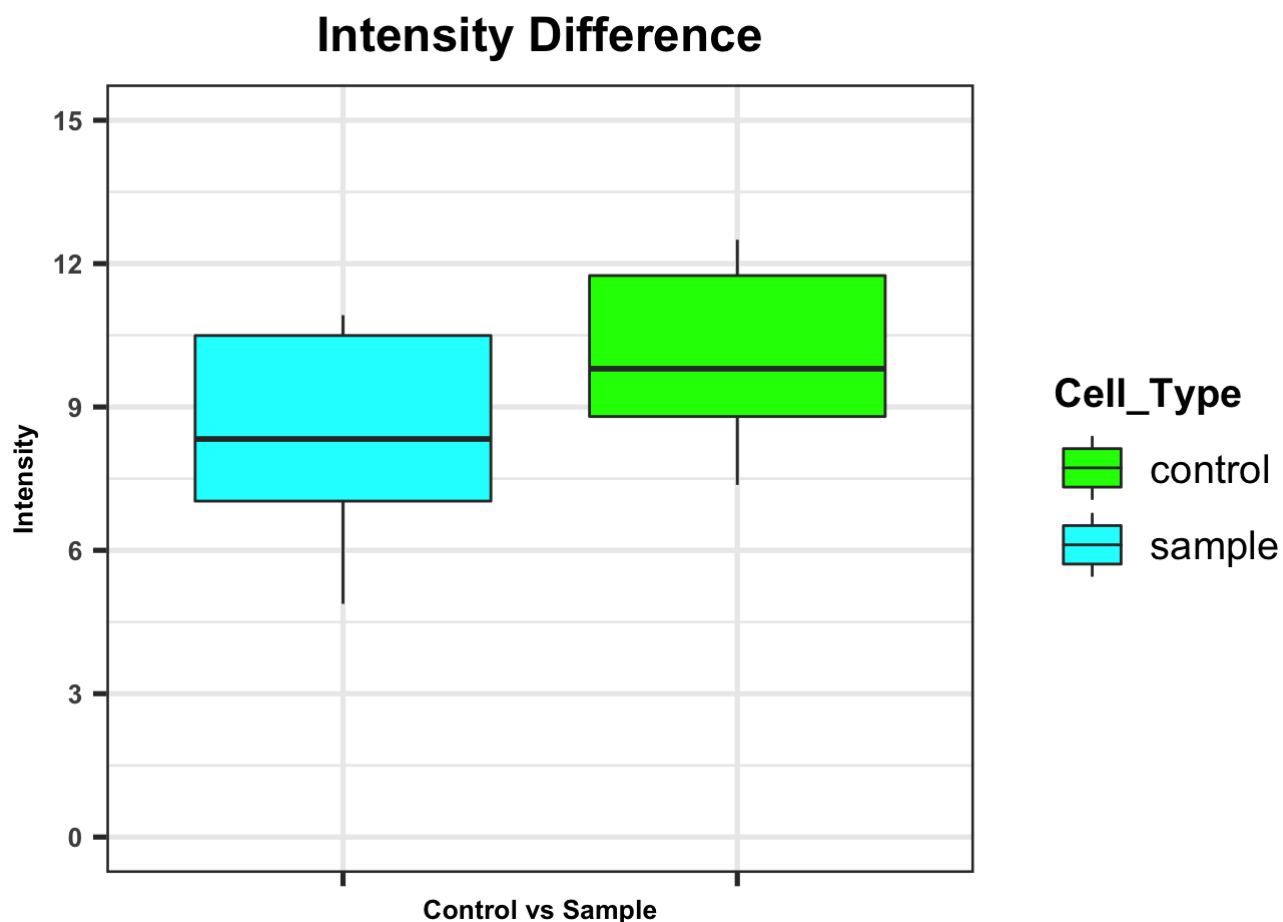


軸のタイトル名、軸の目盛り、軸の文字の大きさ、タイトルの文字の大きさ、レジェンドの文字の大きさなども調整出来ます。

文字は`element_blank()`を指定することで非表示にも出来ます。

`theme_bw()`では背景のデザインを調節できます。

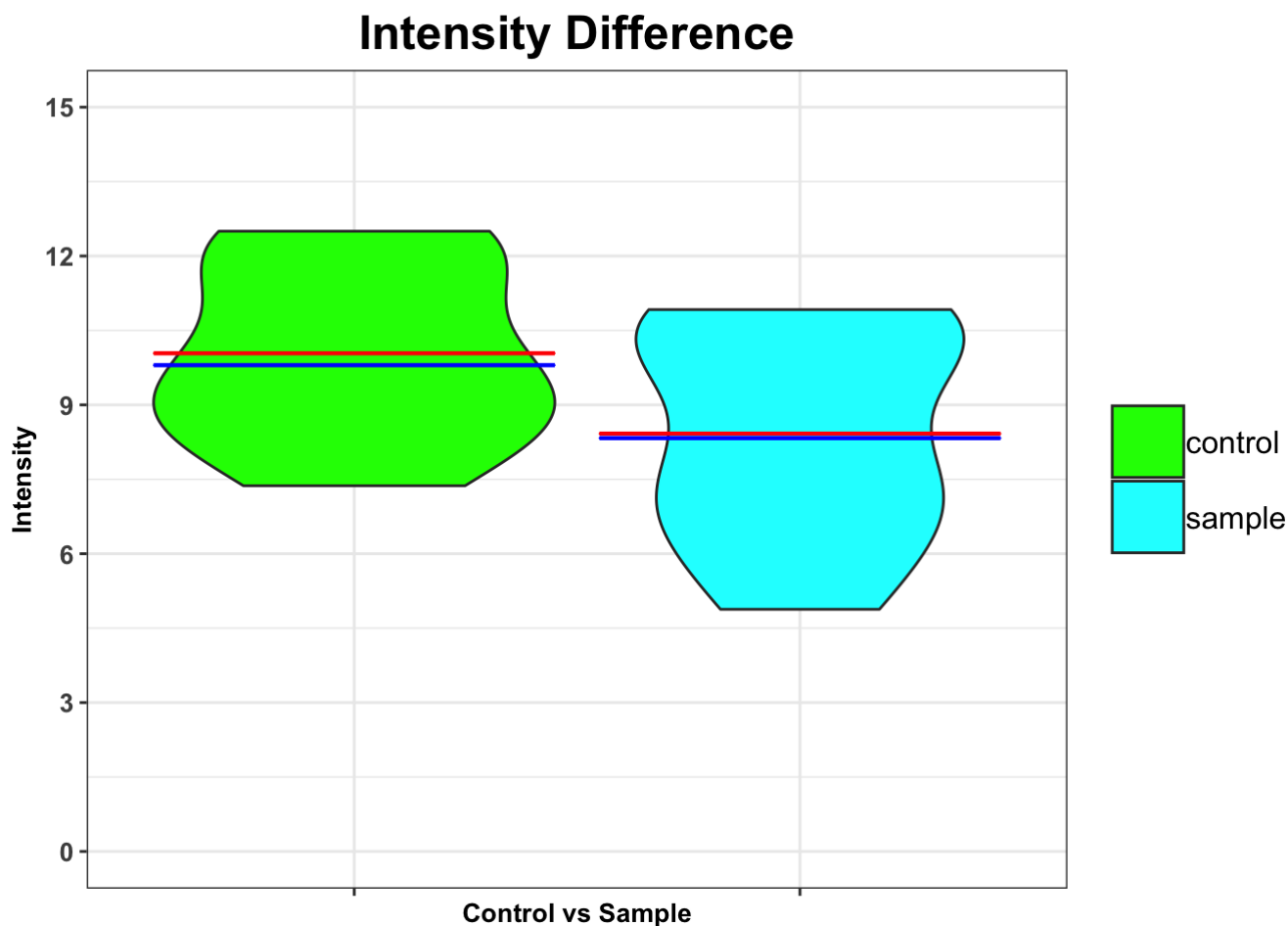
```
p <- p + labs(title="Intensity Difference", x="Control vs Sample", y = "Intensity") +  
  theme_bw(base_size=20, base_family='') +  
  scale_y_continuous(breaks=seq(0, 15, 3), limits = c(0, 15)) +  
  theme(axis.text=element_text(size=10, face="bold"),  
        axis.title=element_text(size=10, face="bold"),  
        title =element_text(size=15, face=' bold'),  
        plot.title = element_text(hjust = 0.5),  
        axis.text.x=element_blank(),  
        legend.key.size = unit(2, "line"),  
        legend.text = element_text(size=15))  
p
```



最後にバイオリンプロットを書きます。

今までのオプションを組み合わせでレジェンドのサイズなども調節します。

```
p<- ggplot(data, aes(x=Cell_Type, y=Intensity, fill=Cell_Type)) +
  geom_violin() +
  stat_summary(fun.y = mean, fun.ymin = mean, fun.ymax = mean, colour = "red", size = 0.3, geom = "crossbar", show.legend = F) +
  stat_summary(fun.y = median, fun.ymin = median, fun.ymax = median, colour = "blue", size = 0.3, geom = "crossbar", show.legend = F) +
  scale_y_continuous(breaks=seq(0, 15, 3), limits = c(0, 15)) +
  theme_bw(base_size=11, base_family='') +
  scale_fill_manual(values=c("#00FF00", "#00FFFF")) +
  labs(title="Intensity Difference", x="Control vs Sample", y = "Intensity") +
  theme(axis.text=element_text(size=10, face="bold"),
        axis.title=element_text(size=10, face="bold"),
        title =element_text(size=15, face='bold'),
        plot.title = element_text(hjust = 0.5),
        axis.text.x=element_blank(),
        legend.key.size = unit(2, "line"), legend.text = element_text(size=12), legend.title = element_blank())
p
```



ここではggplot2を用いた作図の基本を扱いました。

ggplot2は多くのオプションがあり、細かい作図を行うことが可能なため強力な作図ツールと考えられます。

一年間の統計学の授業でRをメインに勉強し簡単な作図は出来るレベルになりました。

今後はより複雑なデータを扱うことを目標に勉強を続けていきたいと思います。