

biostatistics

biostatistics

R

ggplot2

geom_histogram

ggplotで描くヒストグラム

geom_histogram

2017.03.04

1. [geom_histogramの基本的な使い方](#)
2. [複数のヒストグラムを描く方法](#)
3. [ヒストグラムの幅を指定する方法](#)
4. [縦軸を密度表示にする方法](#)
5. [正負で色を塗り分ける方法](#)

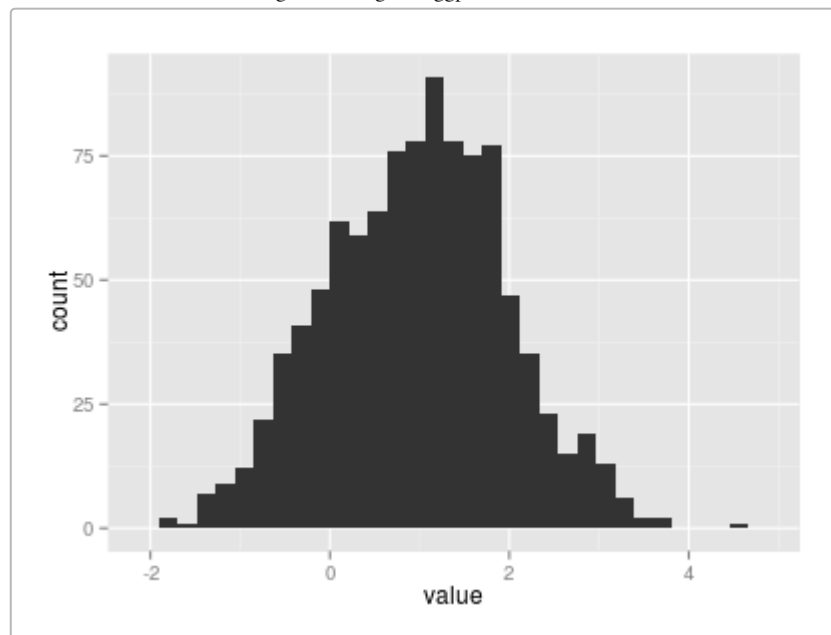
geom_histogram の基本的な使い方

ggplot を利用してヒストグラムを描く簡単な例。

```
library(ggplot2)

#ggplot 用データフレームを作成(データは乱数生成する)
df <- data.frame(value = rnorm(1000, 1, 1))
head(df)
##           value
## 1 -0.3904148
## 2  1.3788577
## 3  1.0617106
## 4  1.2977758
## 5 -0.8127540
## 6  1.5101867

# 描画キャンパス
g <- ggplot(
  df,
  aes(
    x = value
  )
)
g <- g + geom_histogram()
plot(g)
```



複数のヒストグラムを描く方法

ggplot で複数のヒストグラムを重ねて描く方法。

```
library(reshape2)
library(ggplot2)

# サンプルデータを生成
data <- data.frame(
  CDS = rnorm(1000, 20, 5),      # 乱数によりデータ生成
  exons = rnorm(1000, 25, 6),   # 乱数によりデータ生成
  introns = rnorm(1000, 45, 6)  # 乱数によりデータ生成
)

head(data)
##           CDS      exons  introns
## 1 22.37381 20.78157 41.65061
## 2 24.21539 23.83124 33.92928
## 3 17.69864 21.41805 54.18681
## 4 14.51886 27.42381 36.77361
## 5 23.14270 27.32694 37.98734
## 6 20.97776 24.98222 46.58773

# ggplot 用にデータフレームを変換
df <- melt(data)
head(df)
##   variable    value
## 1      CDS 22.37381
## 2      CDS 24.21539
## 3      CDS 17.69864
## 4      CDS 14.51886
## 5      CDS 23.14270
## 6      CDS 20.97776

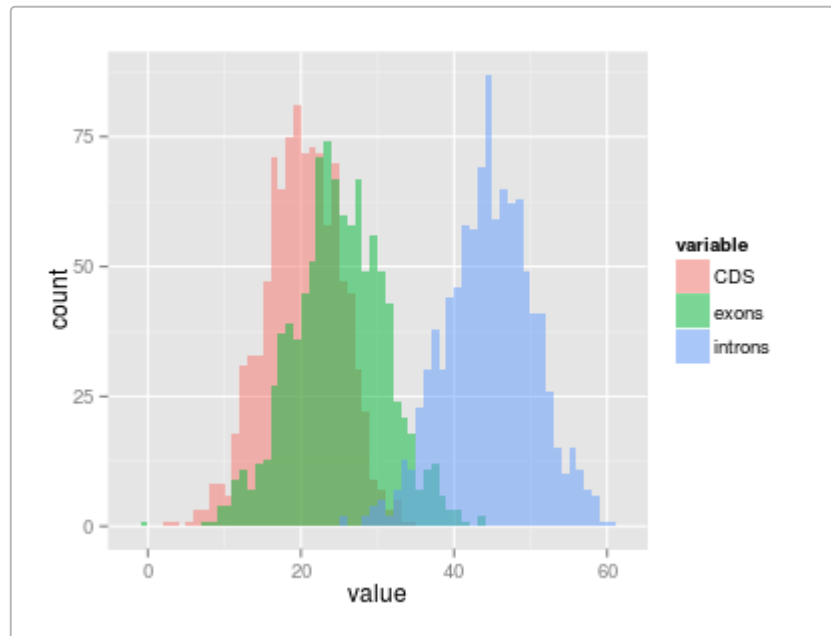
g <- ggplot(
  df,
  aes (
    x = value,          # x で数値データを指定
    fill = variable     # fill でカテゴリーごとに分ける
  )
)
```

```

)
g <- g + geom_histogram(
  alpha = 0.5,          # 透明度を 50% にする
  position = "identity" # カテゴリーごとに描画する
)

plot(g)

```



ヒストグラムを複数描く時は、その並べ方を `position` で指定できる。

```

g <- ggplot(
  df,
  aes (
    x = value,          # x で数値データを指定
    fill = variable     # fill でカテゴリーごとに分ける
  )
)

```

左上の図

```

# position = "stack" # 積み上げて描く
g1 <- g + geom_histogram(
  alpha = 0.5,
  position = "stack"
)
g1 <- g1 + ggtitle("position = stack")
plot(g1)

```

右上の図

```

# position = "fill" # 割合で描く
g2 <- g + geom_histogram(
  alpha = 0.5,
  position = "fill"
)
g2 <- g2 + ggtitle("position = fill")
plot(g2)

```

左下の図

```

# position = "dodge" # 横に並べて描く
g3 <- g + geom_histogram(

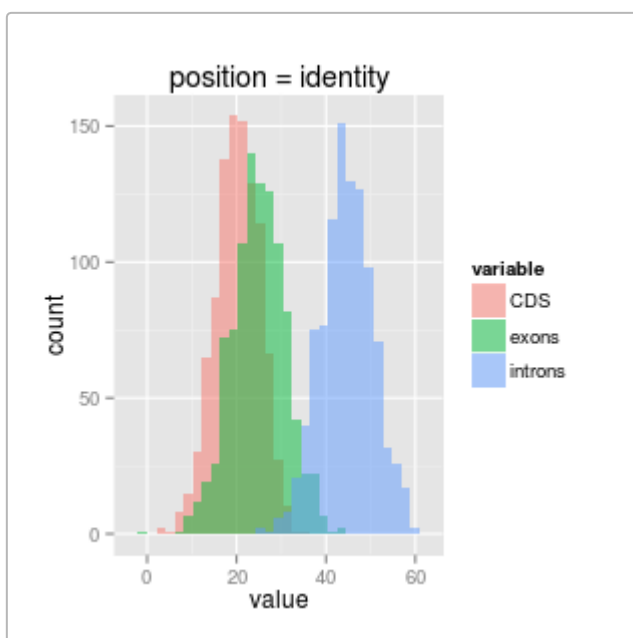
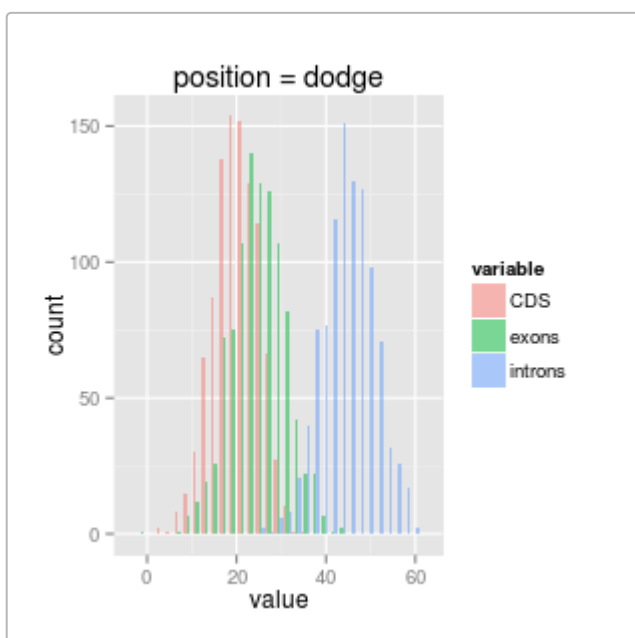
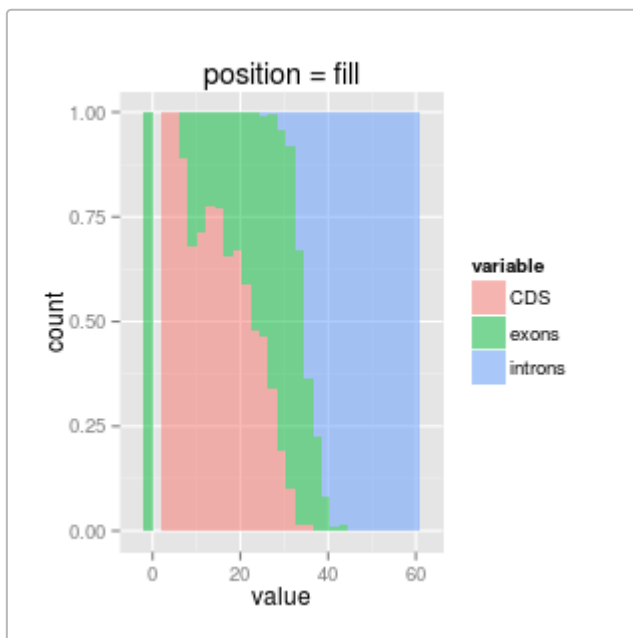
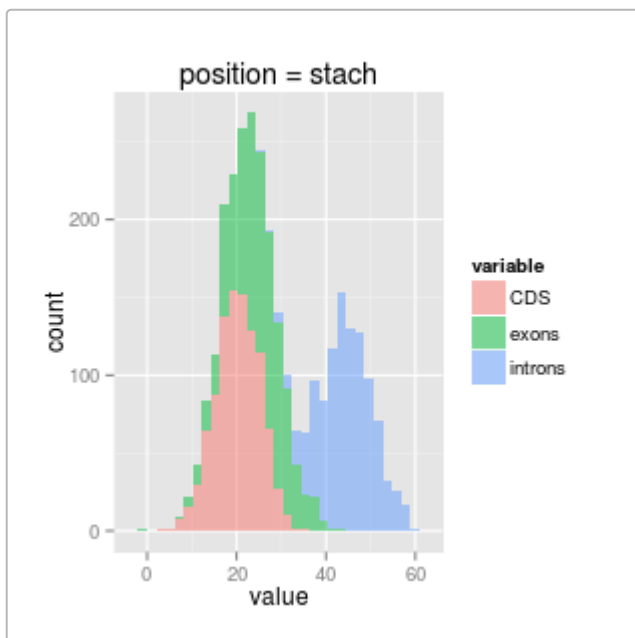
```

```

alpha = 0.5,
position = "dodge"
)
g3 <- g3 + ggtitle("position = dodge")
plot(g3)

# 右下の図
# position = "identity" # 各カテゴリーをそれぞれ独立に描く
g4 <- g + geom_histogram(
  alpha = 0.5,
  position = "identity"
)
g4 <- g4 + ggtitle("position = identity")
plot(g4)

```



ヒストグラムの幅を指定する方法

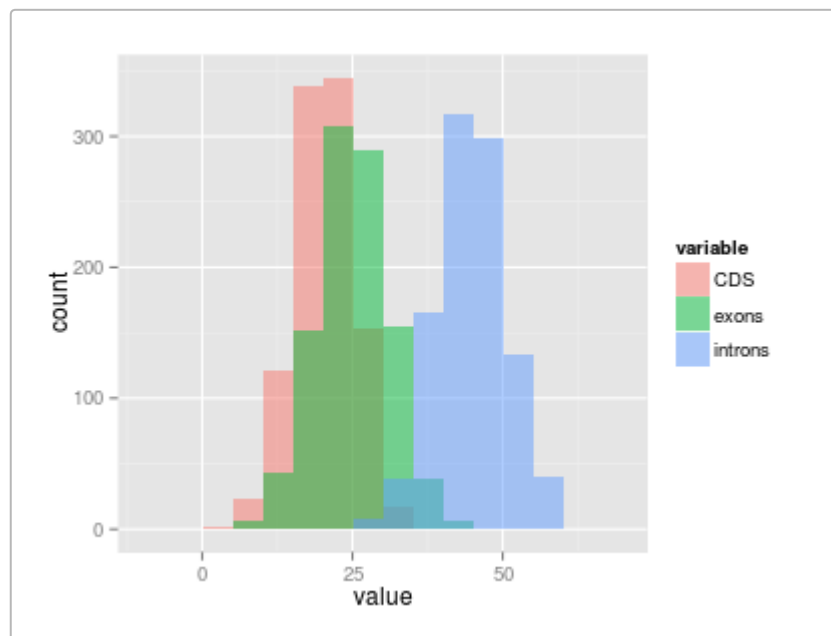
ヒストグラムの幅は `binwidth` オプションで指定できる。

```
library(reshape2)
library(ggplot2)

data <- data.frame(
  CDS = rnorm(1000, 20, 5),
  exons = rnorm(1000, 25, 6),
  introns = rnorm(1000, 45, 6)
)
df <- melt(data)

g <- ggplot(
  df,
  aes (
    x = value,          # xで数値データを指定
    fill = variable     # fillでカテゴリーごとに分ける
  )
)
g <- g + geom_histogram(
  alpha = 0.5,         # 透明度を50%にする
  position = "identity", # カテゴリーごとに描画
  binwidth = 5         # 幅の単位を5とする
)

plot(g)
```



縦軸を密度表示にする方法

縦軸を密度表示にすることもできる。[geom_density](#) のレイヤーを足す。

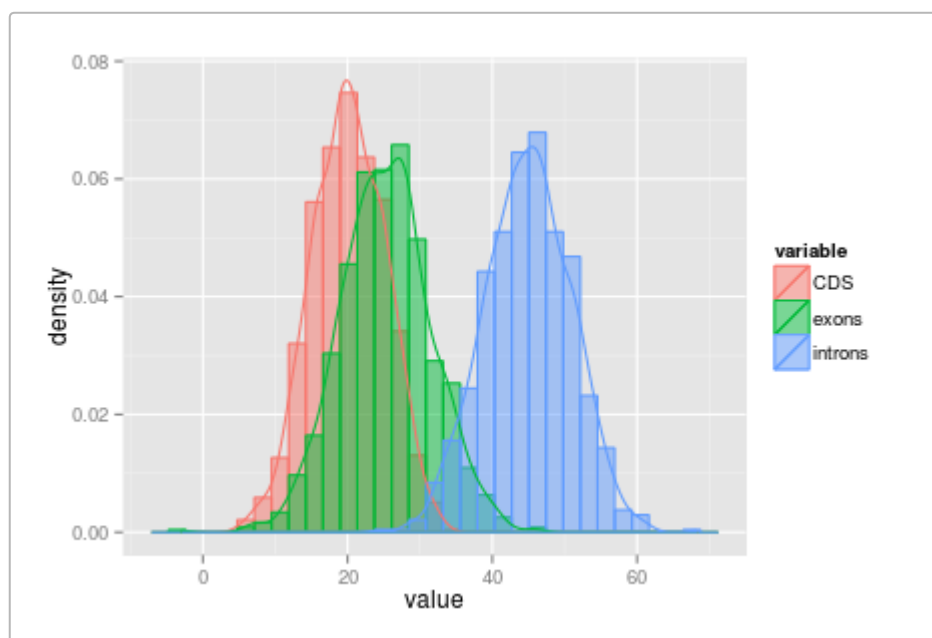
```
library(reshape2)
library(ggplot2)

data <- data.frame(
  CDS = rnorm(1000, 20, 5),
  exons = rnorm(1000, 25, 6),
  introns = rnorm(1000, 45, 6)
)
```

```
df <- melt(data)

g <- ggplot(
  df,
  aes (
    x = value,
    y = ..density.., # 縦軸を密度にする
    fill = variable,
    colour = variable
  )
)
g <- g + geom_histogram( # ヒストグラムを描く
  alpha = 0.5,
  position = "identity"
)
g <- g + geom_density(alpha = 0) # 密度曲線を描く

plot(g)
```



正負で色を塗り分ける方法

正負で色を塗り分けたい場合は、値が正か負のラベル情報を追加してヒストグラムを描けばよい。

```
library(ggplot2)x <- c(rnorm(100, 5, 3), rnorm(100, -3, 4))

df <- data.frame(value = x, isplus = as.logical(x > 0))
head(df)
##      value isplus
## 1  7.589554  TRUE
## 2 -1.603345 FALSE
## 3 10.420338  TRUE
## 4  6.077103  TRUE
## 5  3.945517  TRUE
## 6  3.607557  TRUE

g <- ggplot(df, aes(x = value, fill = isplus))
g <- g + geom_histogram()
plot(g)
```

