

# ggplot2 逆引き集

*@kazutan*

*2016-11-12*



# Contents

<b>1</b>	<b>ggplot2 で同一グラフに 2 変数の折れ線グラフを描きたい</b>	<b>5</b>
1.1	Q . . . . .	5
1.2	A . . . . .	5
1.3	参考 . . . . .	7
<b>2</b>	<b>ggplot2 で積み重ね棒グラフに値を表示させたい</b>	<b>9</b>
2.1	Q . . . . .	9
2.2	A . . . . .	10
2.3	補足 . . . . .	11
2.4	参考 . . . . .	12

これは Qiita で公開されている `ggplot2` 逆引きの記事を集めたものです。

またこれは `bookdown` パッケージのテストを兼ねています。うまく動くといいなあ…。



# Chapter 1

## ggplot2で同一グラフに2変数の折れ線グラフを描きたい

### 1.1 Q

このようなデータが手元にあります:

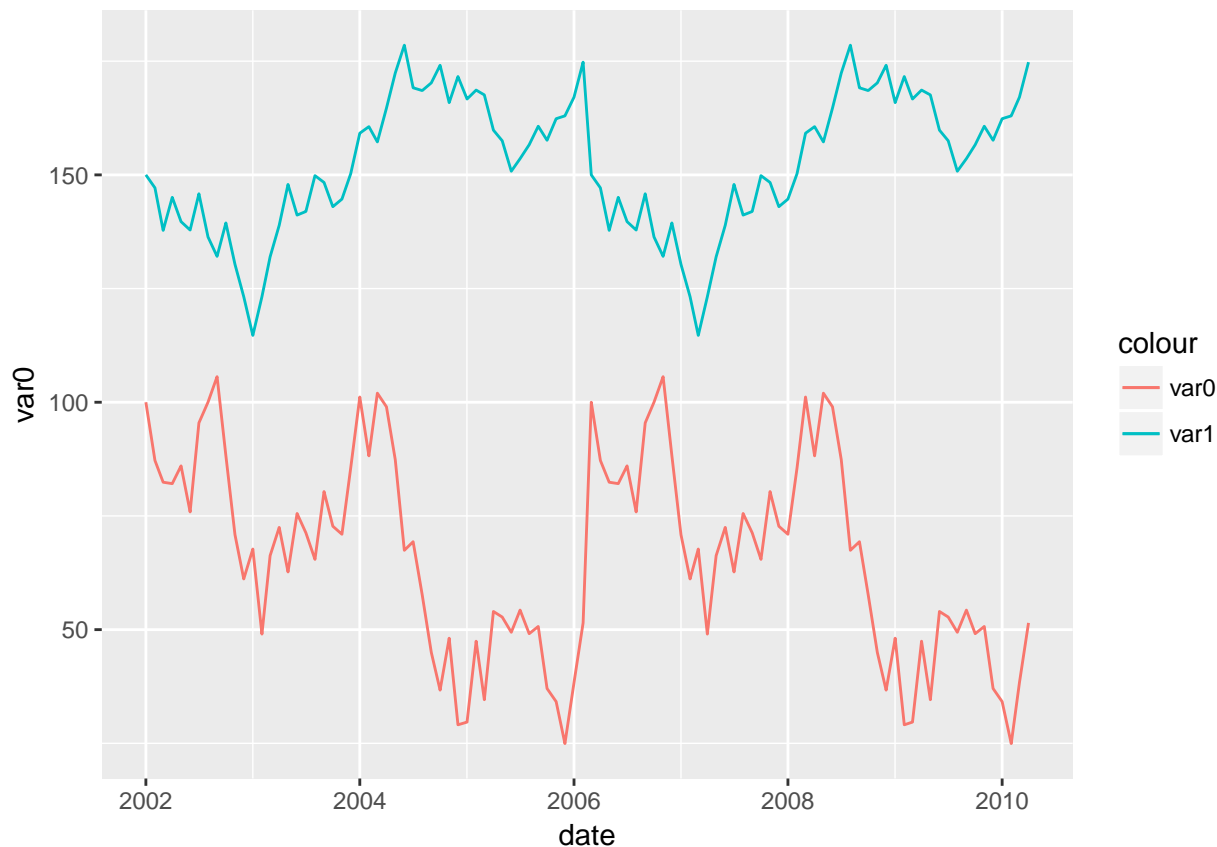
```
test_data <- data.frame(  
  var0 = 100 + c(0, cumsum(runif(49, -20, 20))),  
  var1 = 150 + c(0, cumsum(runif(49, -10, 10))),  
  date = seq.Date(as.Date("2002-01-01"), by="1 month", length.out=100))
```

この時系列変数 var0 と var1 の両方共を、date を x 軸にして ggplot2 でどうやったら描けますか? できれば var0 と var1 の色を変えて、さらに凡例も付けれたら嬉しいです。

### 1.2 A

もし変数が少ないのであれば、マニュアルで別々に作成ビルドアップできますよ:

```
library(ggplot2)  
ggplot(test_data, aes(date)) +  
  geom_line(aes(y = var0, colour = "var0")) +  
  geom_line(aes(y = var1, colour = "var1"))
```



一般的なアプローチとしては、`tidyr` パッケージを利用してデータを縦型 (long format) に変換していく方法があります:

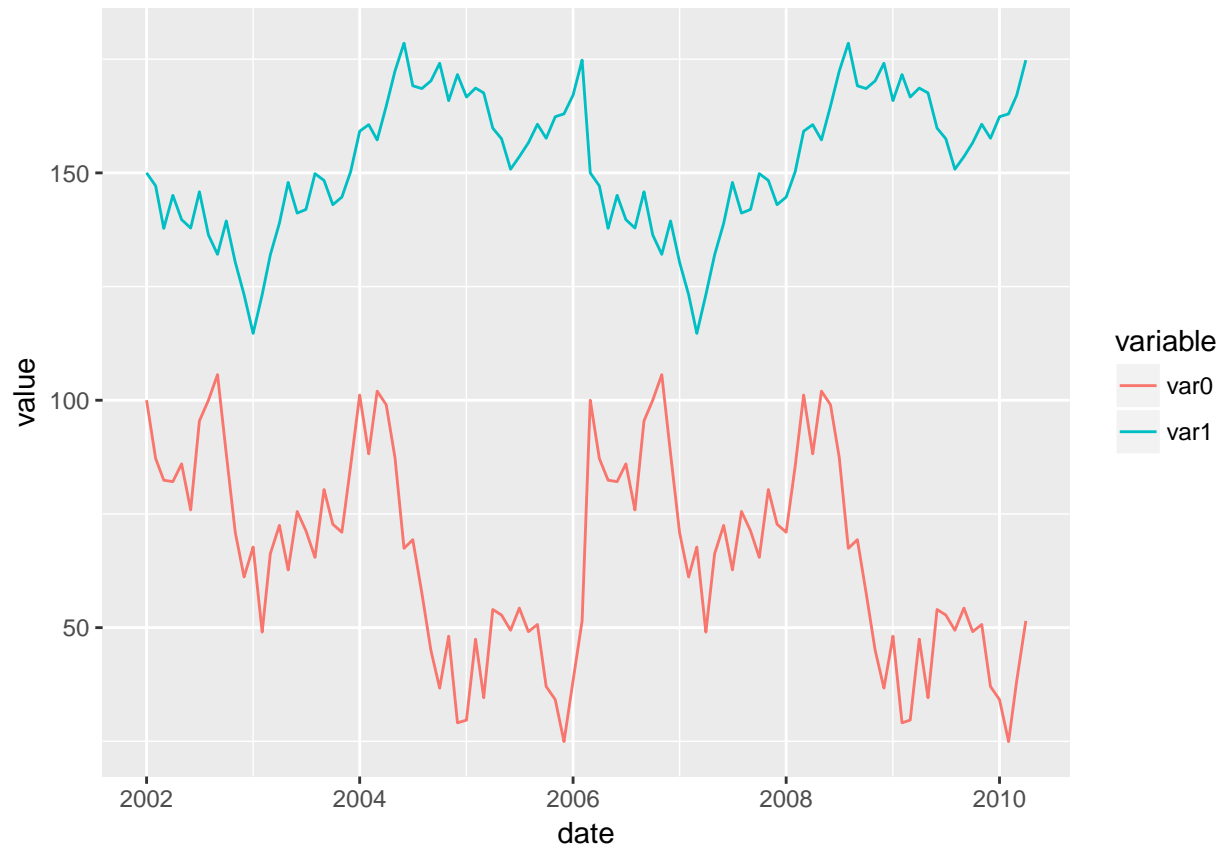
```
library(tidyr)
library(ggplot2)

test_data_long <- tidyr::gather(test_data, key="variable", value = value, -date) # 縦型に変換

knitr::kable(head(test_data_long, 6))
```

date	variable	value
2002-01-01	var0	100.00000
2002-02-01	var0	87.20176
2002-03-01	var0	82.41162
2002-04-01	var0	82.10215
2002-05-01	var0	86.00951
2002-06-01	var0	75.88203

```
ggplot(data=test_data_long, aes(x=date, y=value, colour=variable)) +
  geom_line()
```



データを縦型のデータに変換し、var0 と var1 を分けるための変数で色分けを指定すればこのようになります。

## 1.3 参考

この記事は、StackOverflow に投稿された以下の記事をベースに、コードを一部改変して翻訳して作成しました:  
- r - Plotting two variables as lines using ggplot2 on the same graph - Stack Overflow

関連ドキュメント: - geom\_line. ggplot2 0.9.3.1





## Chapter 2

# ggplot2で積み重ね棒グラフに値を表示させたい

2016/04/12 追記: ggplot2 v2.1.0 に対応するため、一部コードを修正しました

本記事は ggplot2 逆引きプロジェクトの一環として、Stack Overflow の下記記事を翻訳したものです。

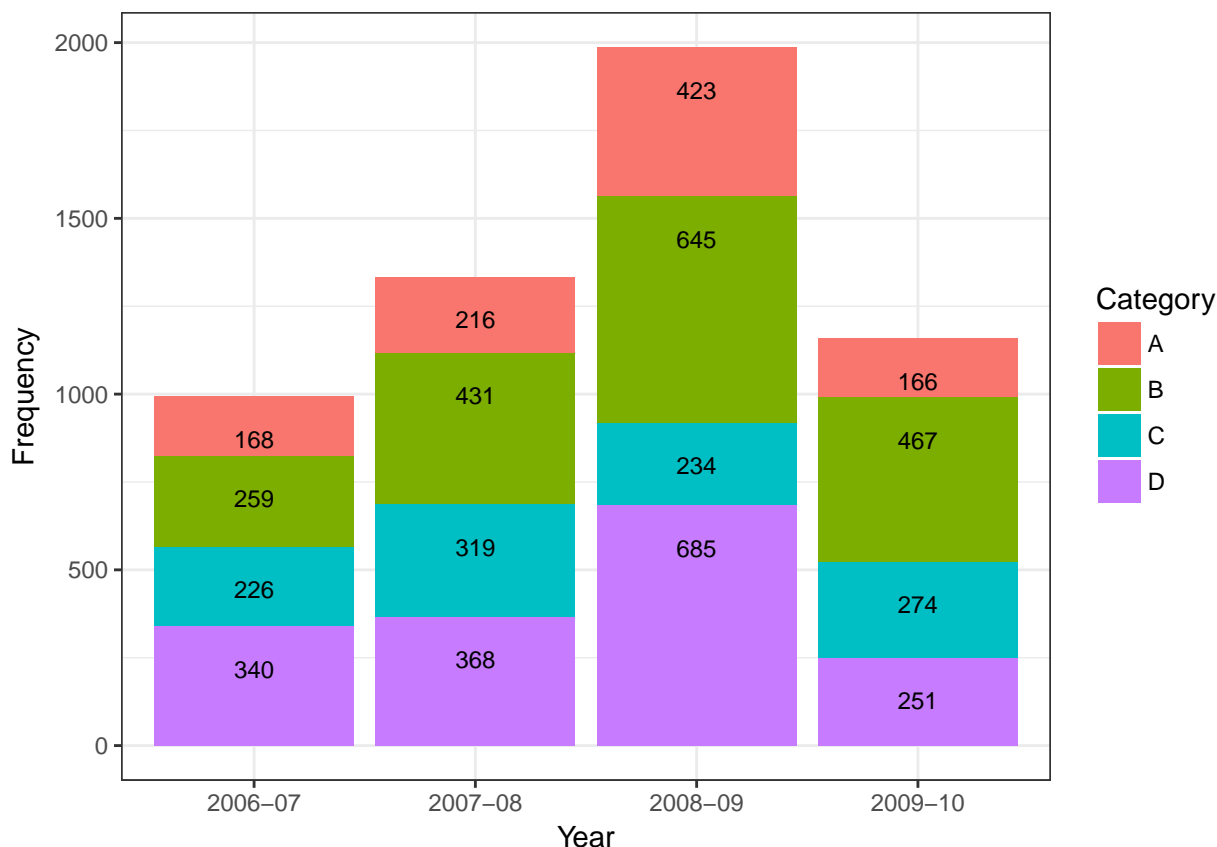
r - Showing data values on stacked bar chart in ggplot2 - Stack Overflow

## 2.1 Q

ggplot2 で積み重ね棒グラフに値を重ねて表示したいです。私が試したのは以下のコードです:

2016/04/12 追記: ggplot2 v2.1.0 で `qplot()` の仕様が変更されているようでしたので、`ggplot()` で描くように修正しました

```
Year      <- c(rep(c("2006-07", "2007-08", "2008-09", "2009-10"), each = 4))
Category  <- c(rep(c("A", "B", "C", "D"), times = 4))
Frequency <- c(168, 259, 226, 340, 216, 431, 319, 368, 423, 645, 234, 685, 166, 467, 274, 251)
Data      <- data.frame(Year, Category, Frequency)
library(ggplot2)
p <- ggplot(Data, aes(Year, Frequency, fill = Category)) +
  geom_bar(stat = "identity") +
  theme_bw()
p + geom_text(aes(label = Frequency), size = 3, hjust = 0.5, vjust = 3, position = "stack")
```



でもこのデータの値を、各部位の中央に配置したいのです。お願いします。

## 2.2 A

一つのやり方として、各バーのミッドポイントを算出するというのがあります:

```
library(plyr)
library(ggplot2)

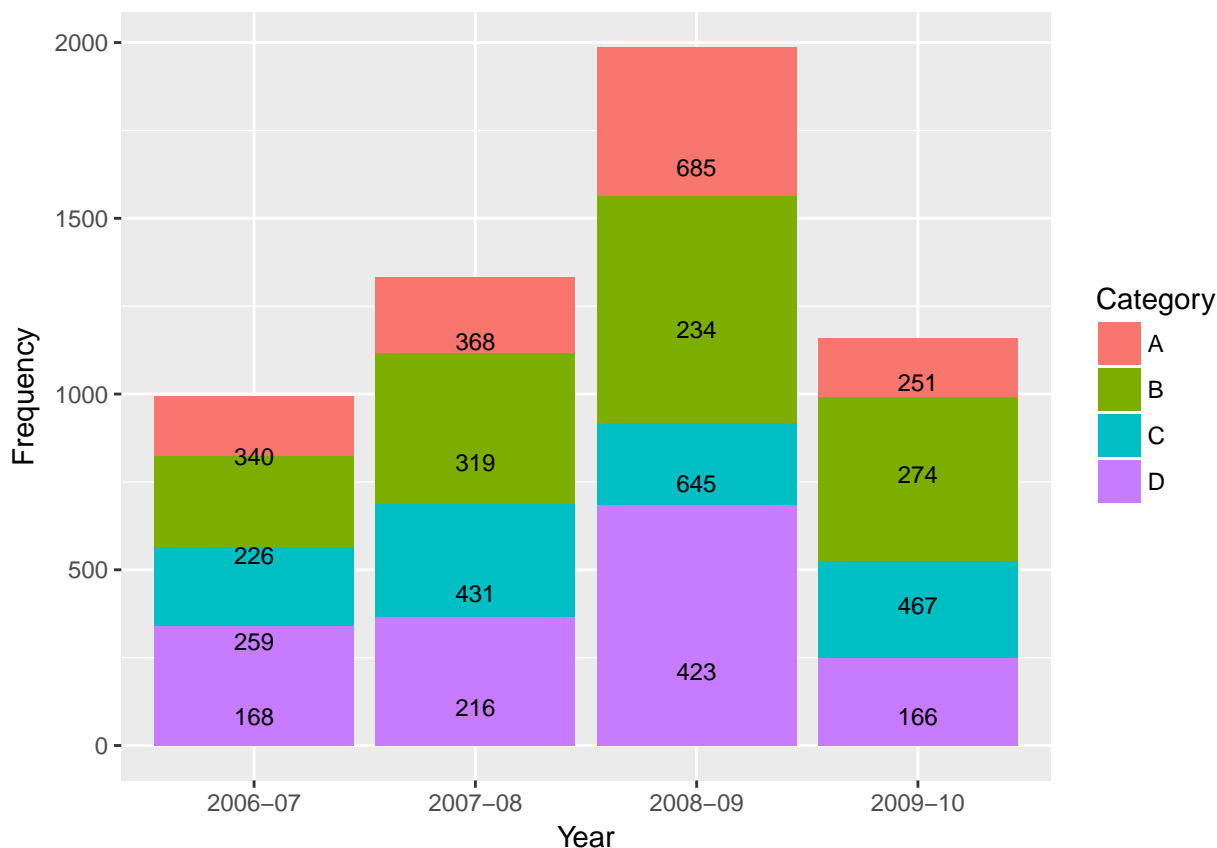
# calculate midpoints of bars (simplified using comment by @DWin)
Data <- ddply(Data, .(Year), transform, pos = cumsum(Frequency) - (0.5 * Frequency))

# plot bars and add text
p <- ggplot(Data, aes(x = Year, y = Frequency)) +
  geom_bar(aes(fill = Category), stat="identity") +
  geom_text(aes(label = Frequency, y = pos), size = 3)

library(plyr)
library(ggplot2)
Year <- c(rep(c("2006-07", "2007-08", "2008-09", "2009-10"), each = 4))
Category <- c(rep(c("A", "B", "C", "D"), times = 4))
Frequency <- c(168, 259, 226, 340, 216, 431, 319, 368, 423, 645, 234, 685, 166, 467, 274, 251)
Data <- data.frame(Year, Category, Frequency)

# calculate midpoints of bars (simplified using comment by @DWin)
Data <- ddply(Data, .(Year), transform, pos = cumsum(Frequency) - (0.5 * Frequency))
```

```
# plot bars and add text
p <- ggplot(Data, aes(x = Year, y = Frequency)) +
  geom_bar(aes(fill = Category), stat="identity") +
  geom_text(aes(label = Frequency, y = pos), size = 3)
p
```

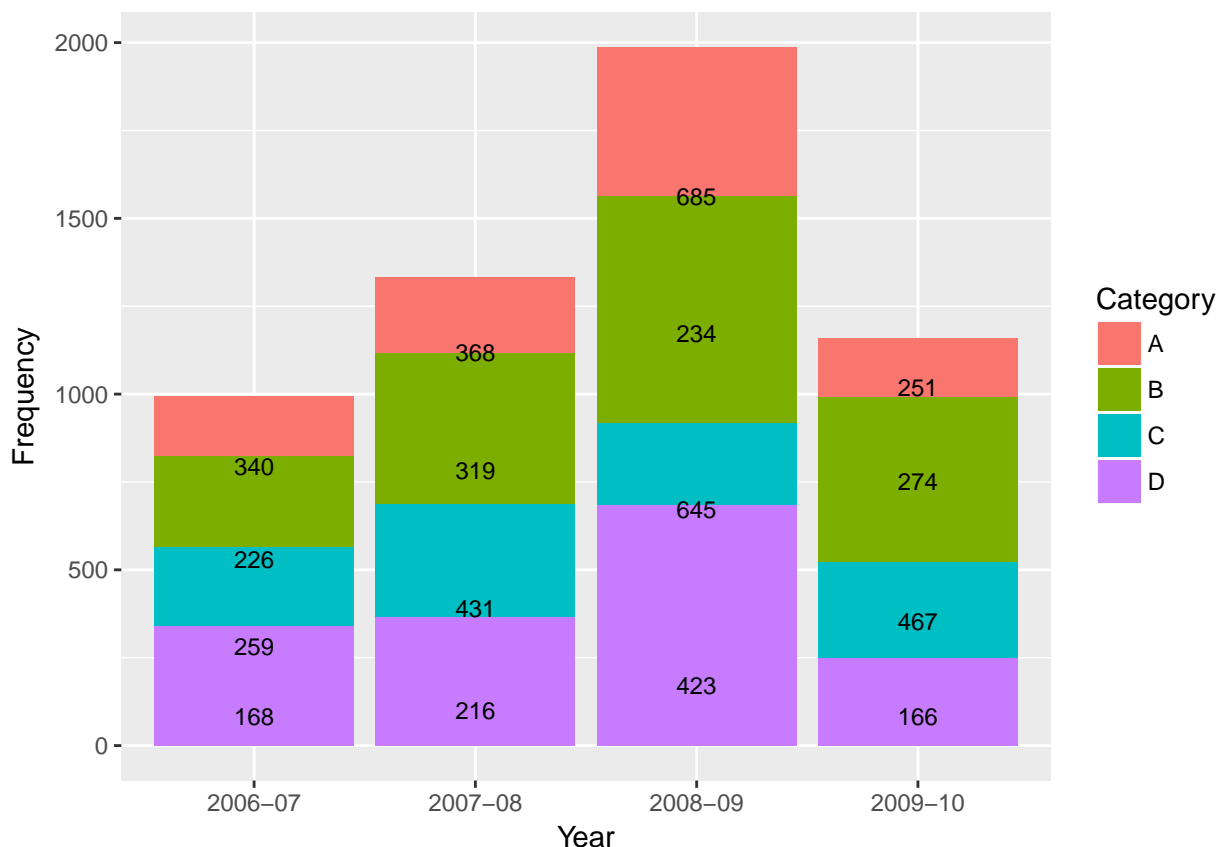


データフレームにあらかじめ値を配置する高さ (y 値) を算出して変数として追加し、`geom_text(aes(label = Frequency, y = pos))` とその変数を指定すれば、そのラベルは算出した位置に配置されます。

## 2.3 補足

上記の A の方法は、データフレームに新たに変数を追加することとなります。それを避けて (だいたい) 真ん中にくるようにできればいい、というのであれば、以下のようなコードでも可能です:

```
p <- ggplot(Data, aes(x = Year, y = Frequency)) +
  geom_bar(aes(fill = Category), stat="identity") +
  geom_text(aes(label = Frequency, size = 3, position = "stack", vjust = Frequency/75))
p
```



これは、`geom-text` の `vjust` に対して設定を加えています。このように `y` 値の変数を定数で補正したものを `vjust =` に投げると、割合的にずれてくれるようになります。

ただこの方法ですと、その補正する定数をどう設定したらど真ん中にくるか、はっきりとはわかりません。詳しくは [RPubs - vjust のテスト](#) をご覧ください。

また、この記事の元となっている [StackOverflow](#) の記事には、Hadley Wickham 氏による以下のコメントがあったことも付記します：

Please don't. It's not a good idea to try and improve a confusing visualisation by adding text labels. Either make a table or use a better display of the data. – hadley Jul 12 '11 at 1:10

## 2.4 参考

- [r - Showing data values on stacked bar chart in ggplot2 - Stack Overflow](#)
- [RPubs - vjust のテスト](#)
- [ggplot2 逆引き- ggplot2 で geom\\_text\(\) を使って集合棒グラフにラベルを付けたい #rstatsj - Qiita](#)