

Statistical graphics with R

2011/6/7th

KANETA Taku



データの読み込み



コピーして

	G	H	I	J	K	L	M
#Lgs	q1	q2	q3	q4	q5	q6	
2	3	3	2	3	3		
2	4	3	2	4	3		
2	3	3	3	4	2		
2	4	3	3	4	3		
2	3	3	2	4	3		
2	3	2	3	3	3		
2	4	4	4	3	2		
4	3	4	2	2	2		
2	4	4	3	3	3		

```
> q <- read.delim("clipboard")
```

- 非常に便利でイチオシです





石橋を叩いてみる

- データ形式を確認する



```
> q
```

	q1	q2	q3
1	3	3	2
2	4	3	2
3	3	3	3

```
> str(q)
```

```
'data.frame':  
16 obs. of 21  
variables:  
 $ q1 : int 3 4 3...
```

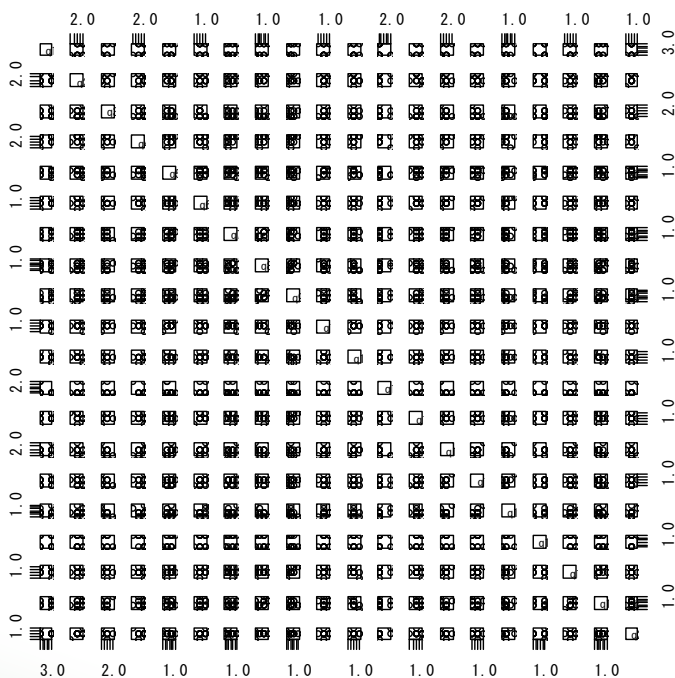
```
> edit(q)
```

R データエディタ				
	q1	q2	q3	q4
1	3	3	2	3
2	4	3	2	4
3	3	3	3	4
4	4	3	3	4



いいからplot() だ？

-何がなんだか、分からない

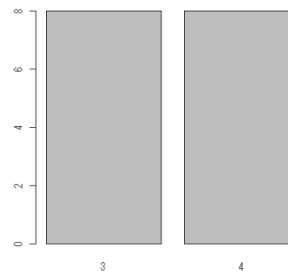
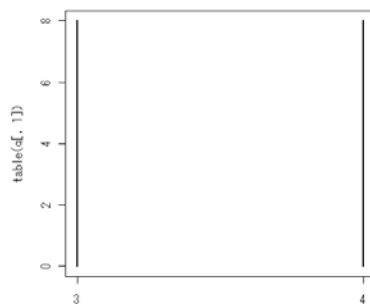
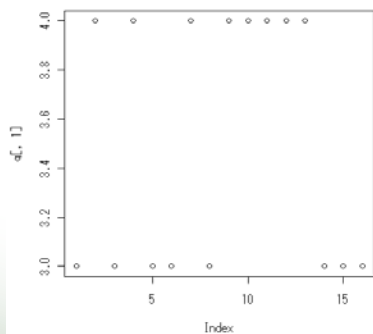




now plotting

- 焦らず急がず一つずつ

```
> plot(q[,1])  
> plot(table(q[,1]))  
> barplot(table(q[,1]))
```

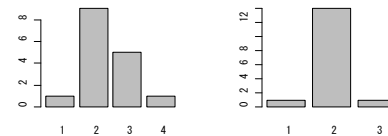
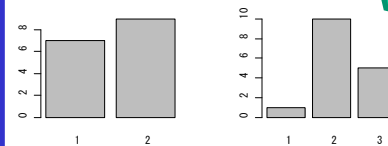




It's future sight –ループ処理

- 一度に全てプロットしてしまえる

```
> par(mfrow=c(5, 4))  
> for (i in 1:length(q)) {  
  barplot(table(q[,i]))  
  i <- i+1  
}
```





ハワイ大学院生と東京外国語大学院生で.....

最も意見が分かれた質問は？



大学を示す列を追加する

- ここではtufsという列を追加し、1, 2 で属性を割り振ることにする。

- 追加後再度読み込み

```
> q <- read.delim("clipboard")
```



q20	tufs
2	2
2	2
2	2
2	2
2	2



tufs or not tufs

- tufs, hawaiiそれぞれのサブセットを作成

```
> tufs <- subset(q, tufs==1)  
> hawaii <- subset(q, tufs==2)
```



- 各項目について平均の差を求めてみる

```
> gap <- mean(tufs)-mean(hawaii)
```





gap in the box

- 平均の差が大きい順に並び替えてみる

```
> order(gap, decreasing=TRUE)
```

```
[1] 7 4 2 11 8 1
```





gap in the box

- 平均の差が大きい順に並び替えてみる

```
> order(gap, decreasing=TRUE)
```

```
[1] 7 4 2 11 8 1
```



- q7が最も差が大きい ⇒ 意識が違うのは

Q 7. Vocabulary words are the most important part of learning a new language



クロスタブ集計

- Rでクロスタブ集計してみる

```
> (tab1 <- xtabs(~q7+tufs, data=q))
```



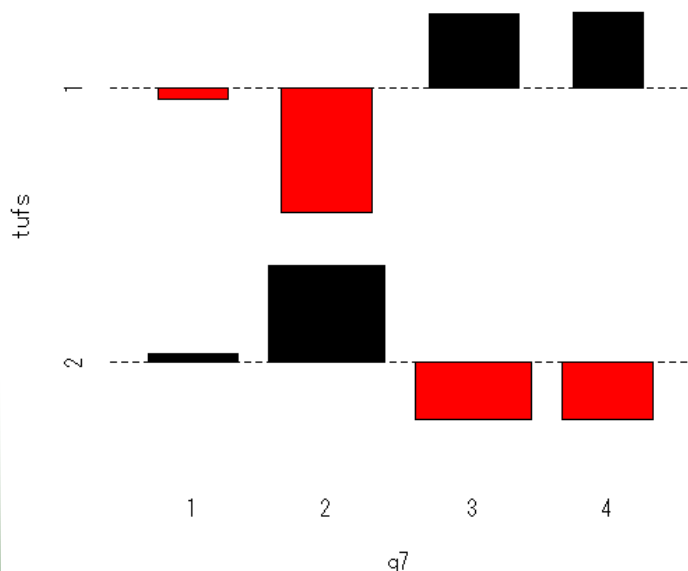
	tufs	
q7	tufs	hawaii
1	1	2
2	0	5
3	3	2
4	2	1



クロスタブをグラフィカルに

- Cohen-Friendlyの連関プロットを描く

```
> assocplot(tab1)
```





Practice makes perfect

1. Exercise 1

- 同様の手順で、最も意識の差がなかった質問項目を見つけて、集計する

2. Exercise 2

- 教育経験のあり/なしを基準にして、最も意識差の大きい項目を見つけ、集計する