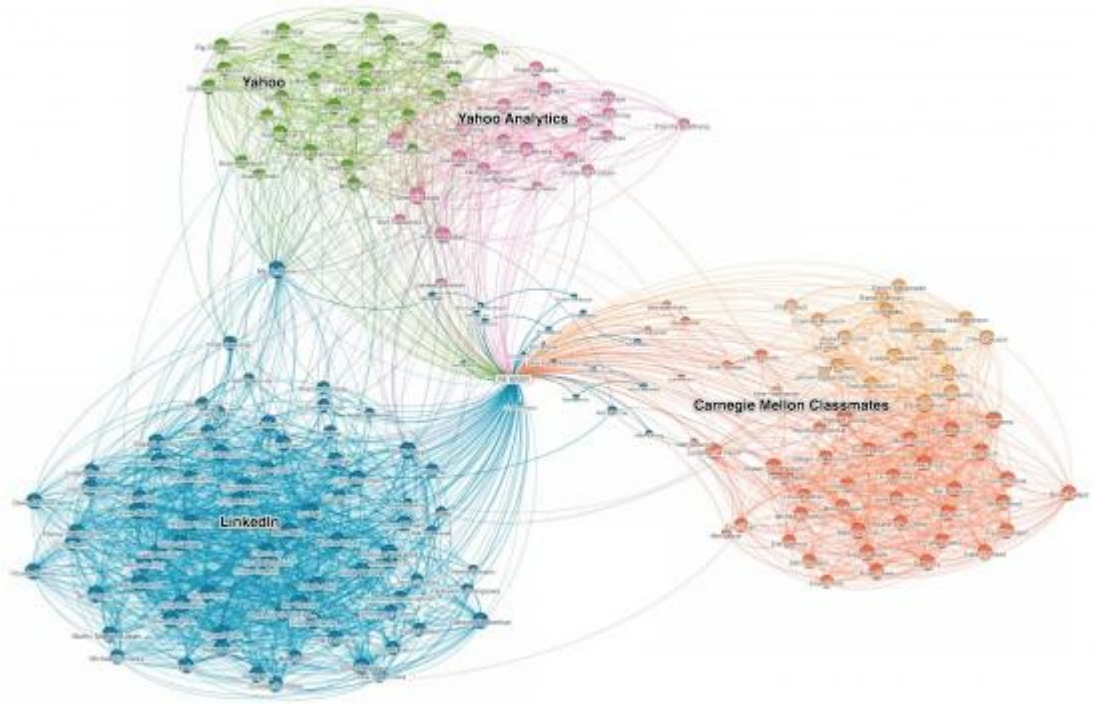
DATAGURU 炼数成金



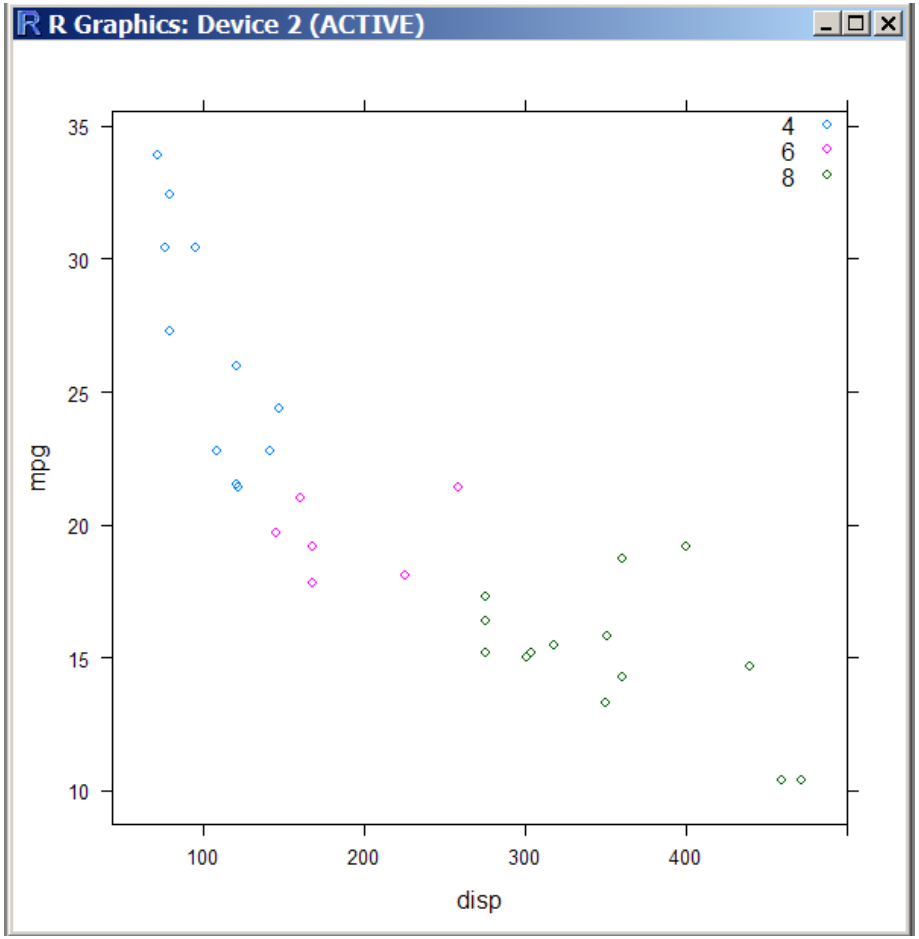
数据分析、展现与R语言 第7周

2013.03.09

利用xyplot()对散点分组

```
library(lattice)
xyplot(mpg~disp,
data=mtcars,
groups=cyl,
auto.key=list(corner=c(1,1)))
```

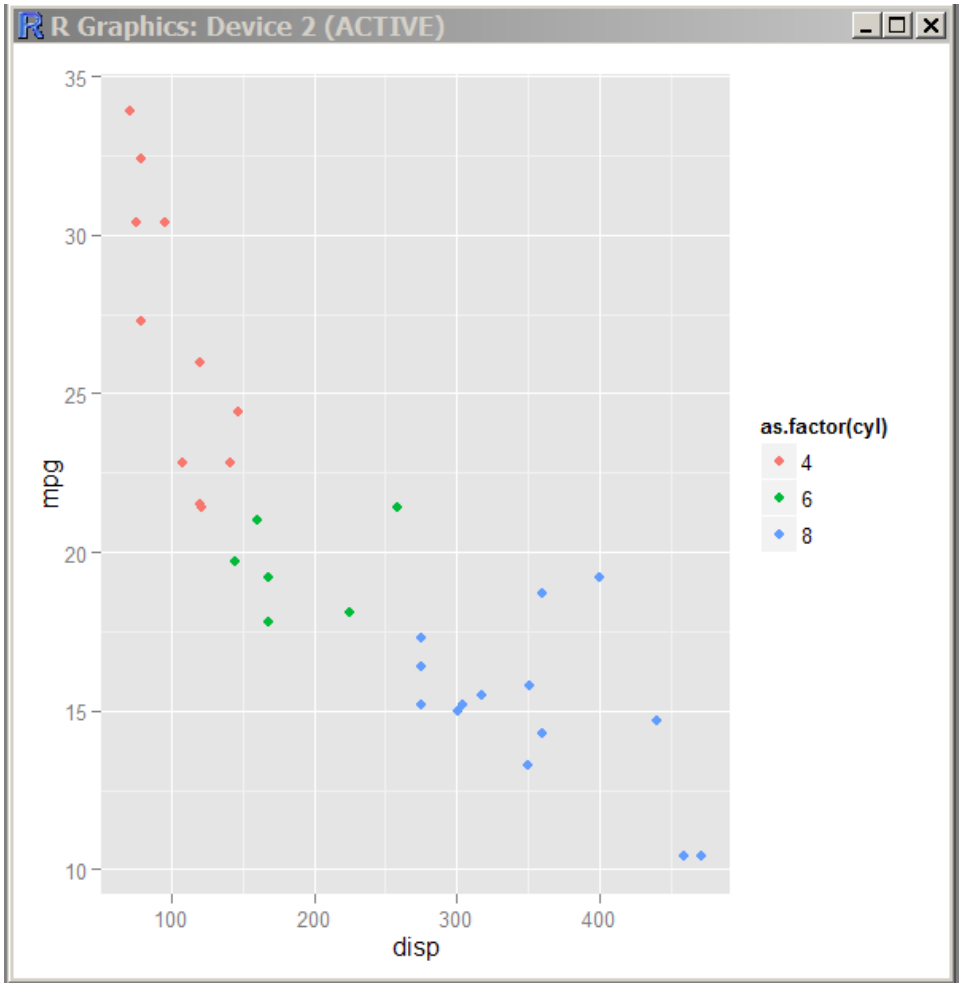
```
> mtcars
                   mpg cyl  disp  hp drat
Mazda RX4         21.0   6 160.0 110 3.90 2.6
Mazda RX4 Wag     21.0   6 160.0 110 3.90 2.8
Datsun 710        22.8   4 108.0  93 3.85 2.3
Hornet 4 Drive    21.4   6 258.0 110 3.08 3.2
Hornet Sportabout 18.7   8 360.0 175 3.15 3.4
Valiant           18.1   6 225.0 105 2.76 3.4
Duster 360        14.3   8 360.0 245 3.21 3.5
Merc 240D         24.4   4 146.7  62 3.69 3.1
Merc 230          22.8   4 140.8  95 3.92 3.1
Merc 280          19.2   6 167.6 123 3.92 3.4
Merc 280C         17.8   6 167.6 123 3.92 3.4
Merc 450SE        16.4   8 275.8 180 3.07 4.0
```



2013.03.09

初试ggplot2包

```
library(ggplot2)
qplot(disp,mpg,data=mtcars,
      col= as.factor(cyl))
```

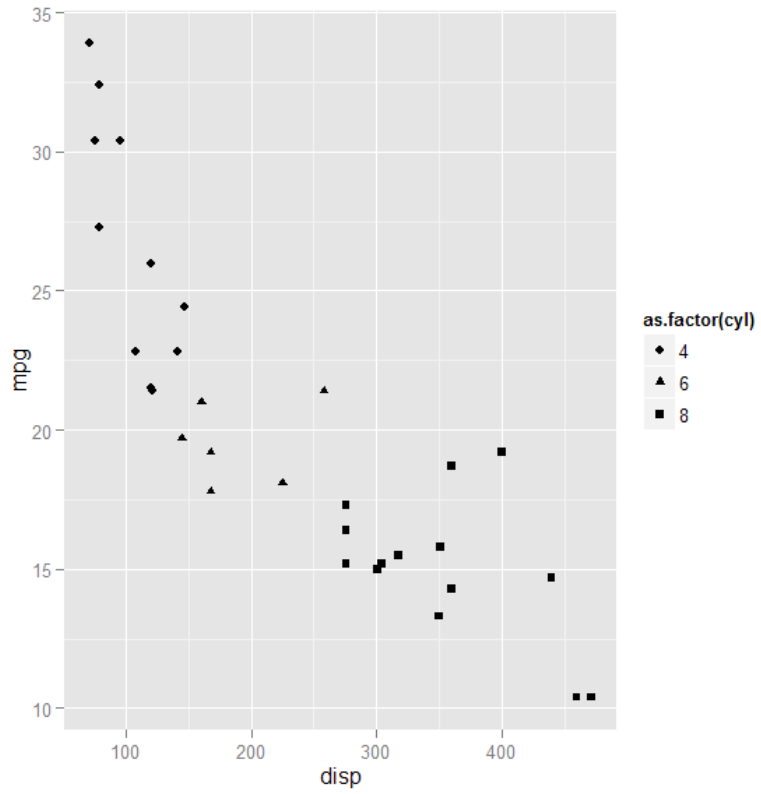


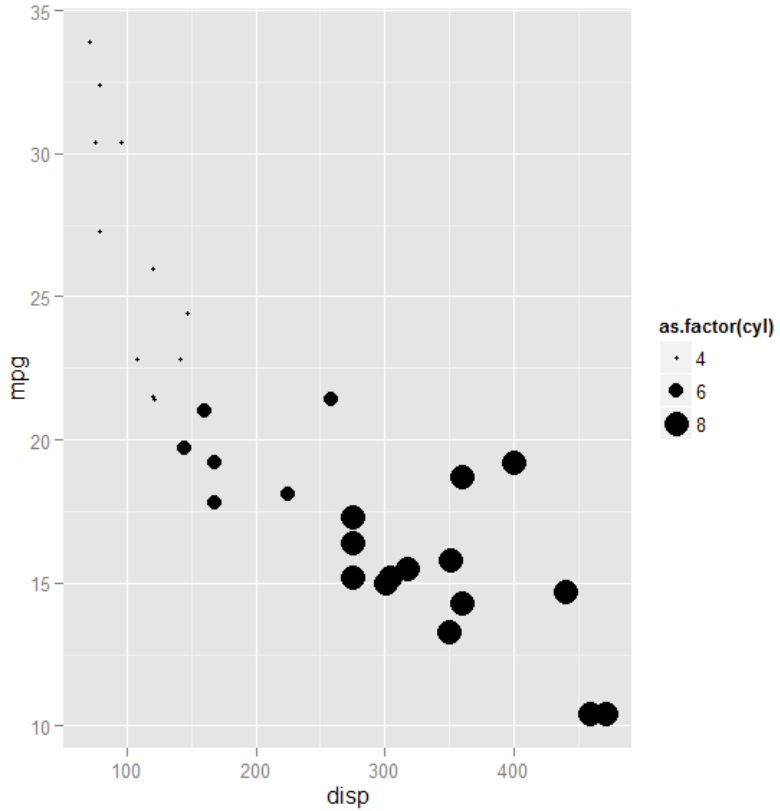
2013.03.09

使用散点形状和大小来表示分组

qplot(disp,mpg,data=mtcars,shape=as.factor(cyl))

qplot(disp,mpg,data=mtcars,size=as.factor(cyl))



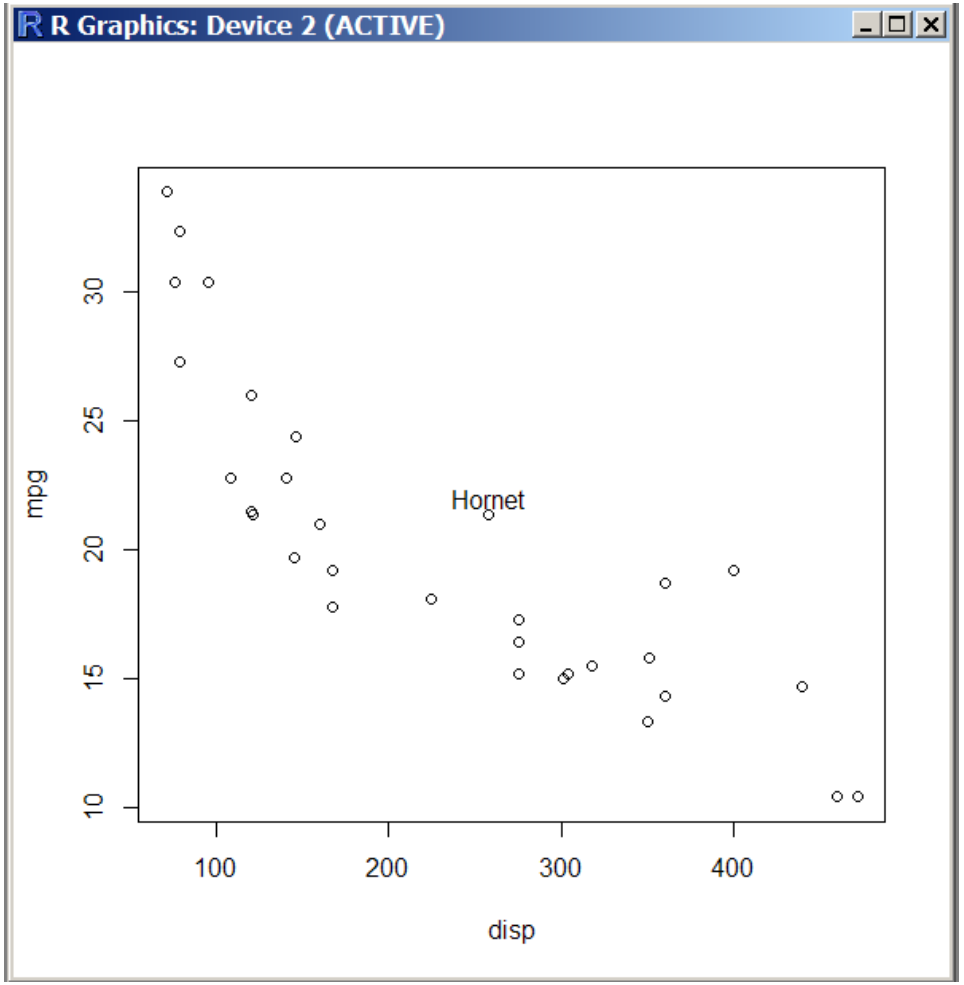


2013.03.09

用text()标识点

plot(mpg~disp, data=mtcars)

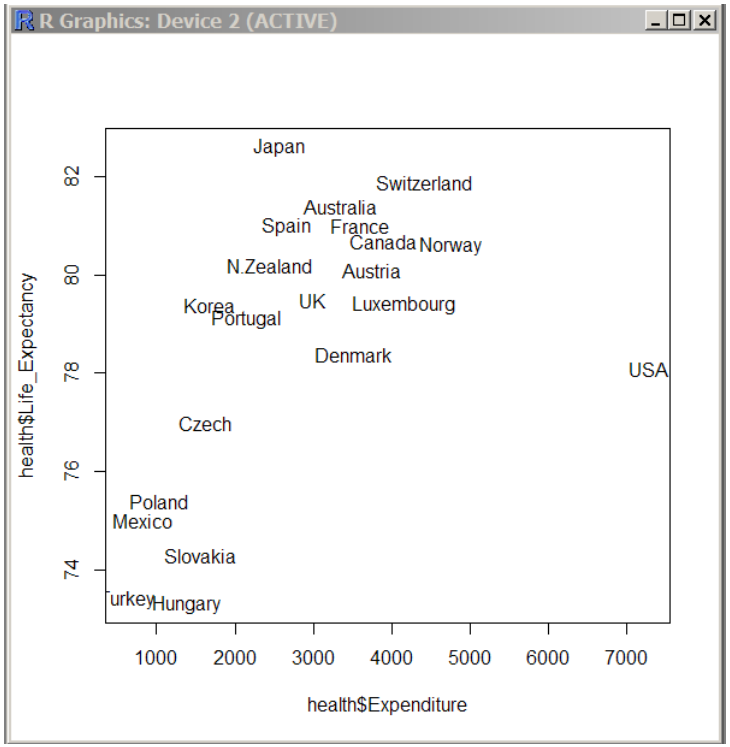
text(258,22,"Hornet")



2013.03.09

用text()标识点

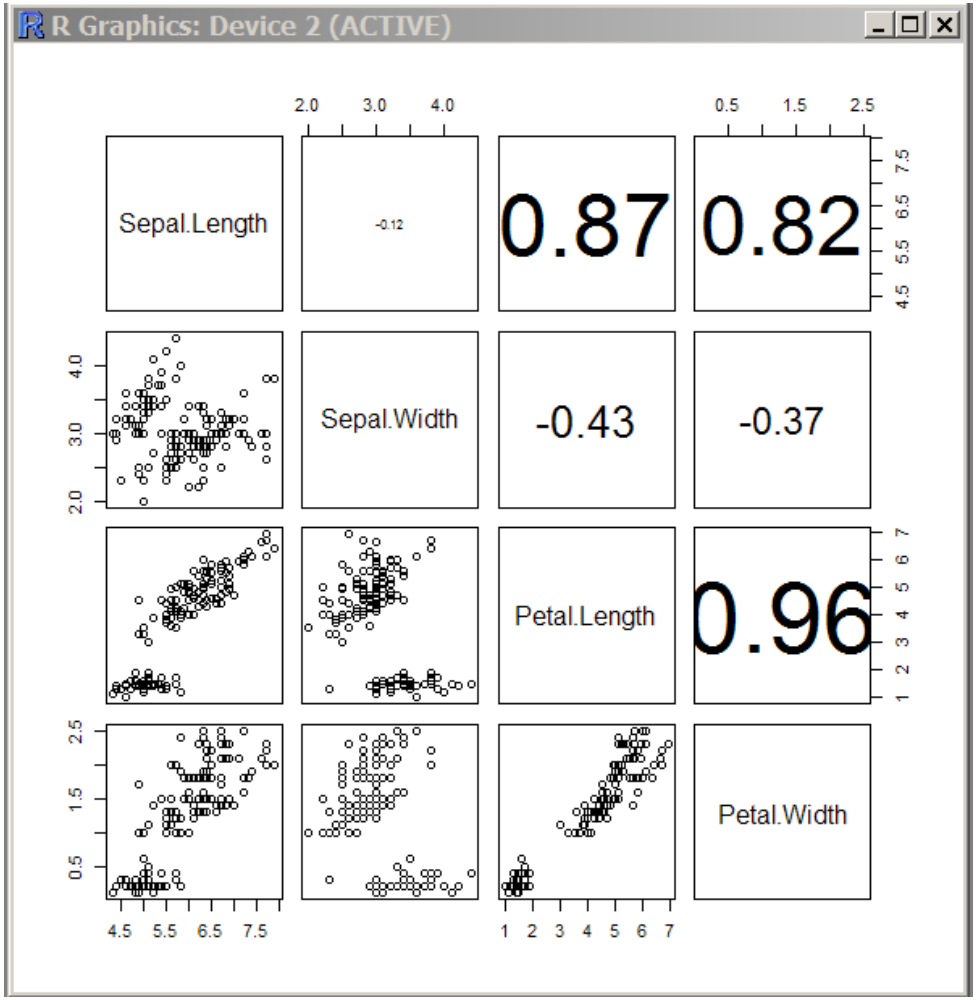
```
health<-read.csv("HealthExpenditure.csv",header=TRUE)
plot(health$Expenditure,health$Life_Expectancy,type="n")
text(health$Expenditure,health$Life_Expectancy,health$Country)
```



2013.03.09

例子：散点图+相关系数

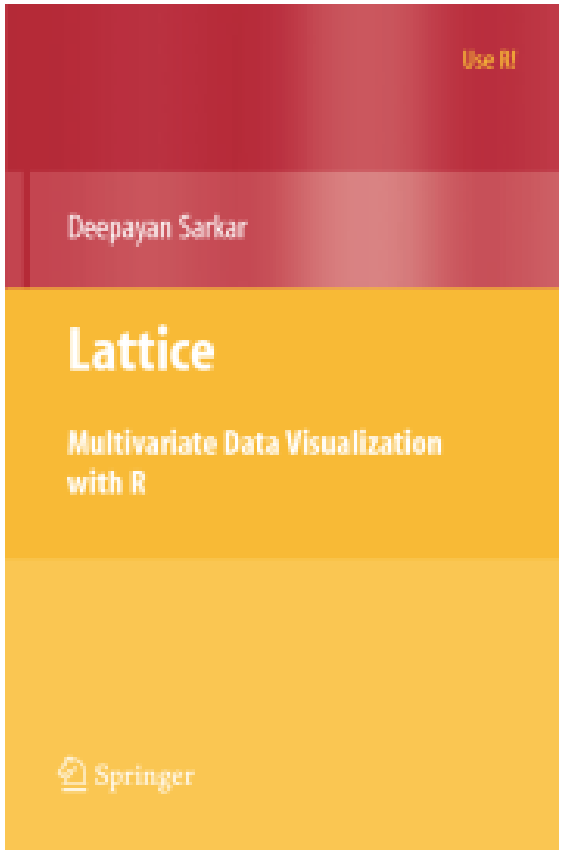
```
panel.cor <- function(x, y, ...)
{
par(usr = c(0, 1, 0, 1))
txt <- as.character(format(cor(x,
    y), digits=2))
text(0.5, 0.5, txt, cex = 6*
    abs(cor(x, y)))
}
pairs(iris[1:4],
    upper.panel=panel.cor)
```



2013.03.09

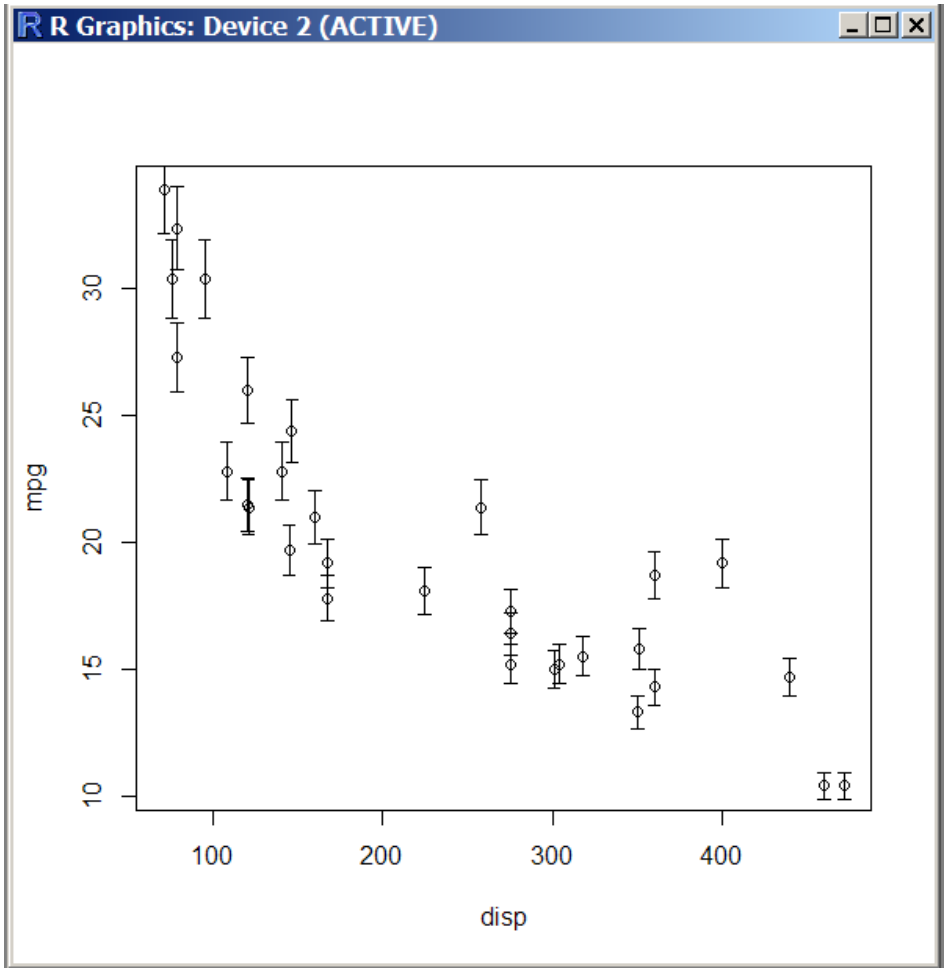
关于lattice包

http://lmdvr.r-forge.r-project.org/figures/figures.html



误差条

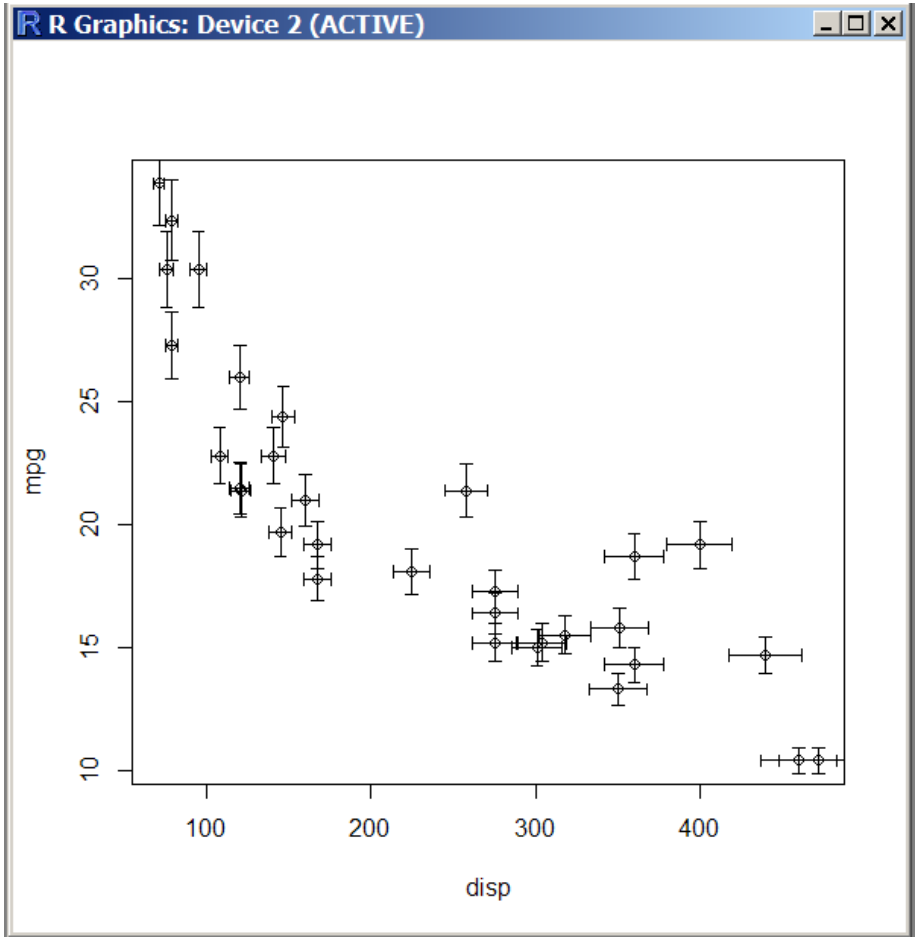
```
plot(mpg~disp,data=mtcars
     )
arrows(x0=mtcars$disp,
y0=mtcars$mpg*0.95,
x1=mtcars$disp,
y1=mtcars$mpg*1.05,
angle=90,
code=3,
length=0.04,
lwd=0.4)
```



2013.03.09

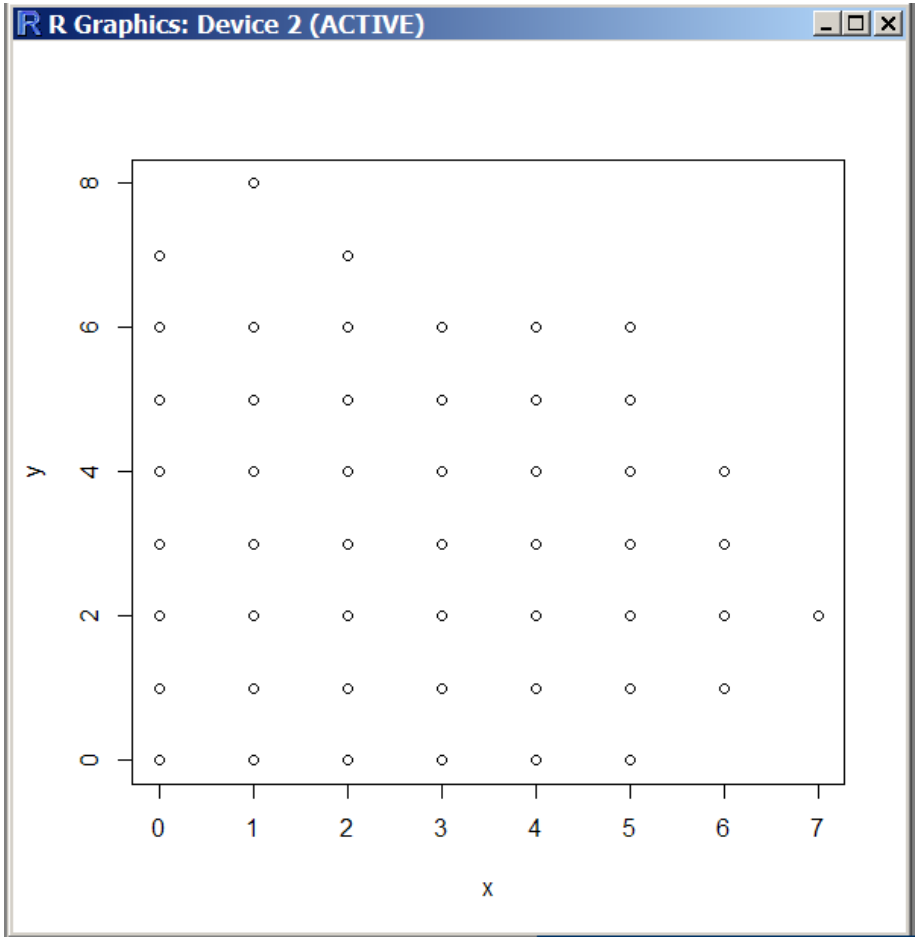
误差条

```
arrows(x0=mtcars$disp*0.95,
y0=mtcars$mpg,
x1=mtcars$disp*1.05,
y1=mtcars$mpg,
angle=90,
code=3,
length=0.04,
lwd=0.4)
```



jitter( )

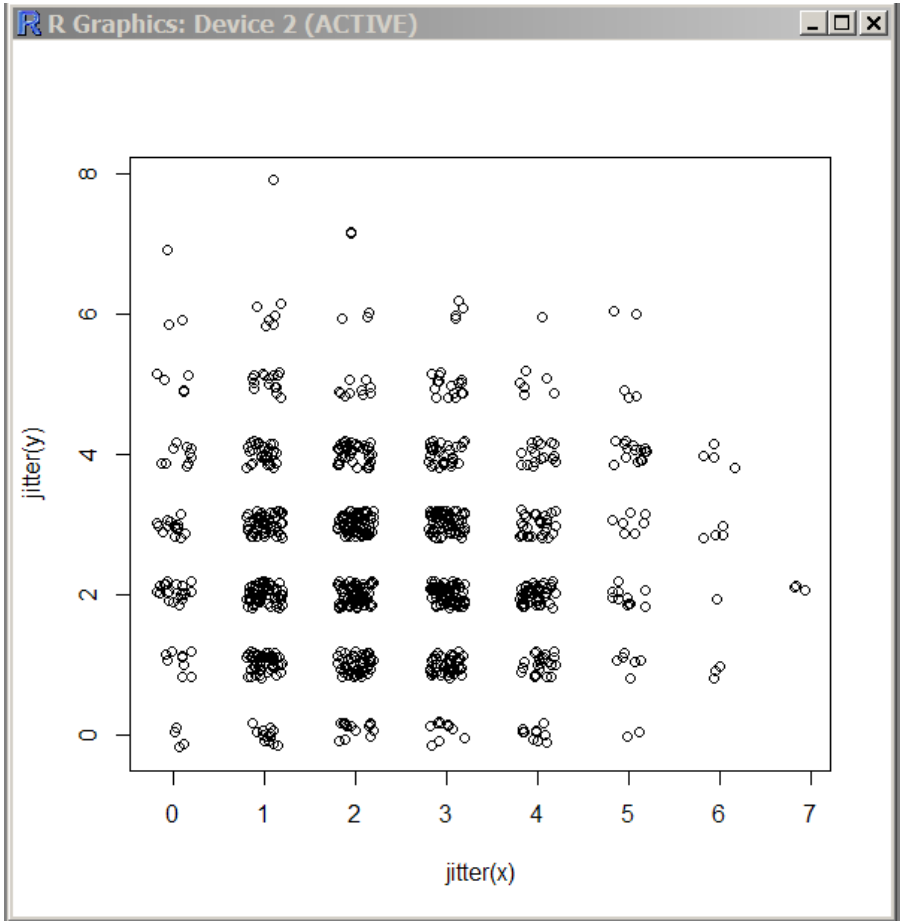
```
x <- rbinom(1000, 10, 0.25)
y <- rbinom(1000, 10, 0.25)
plot(x,y)
```



jitter( )

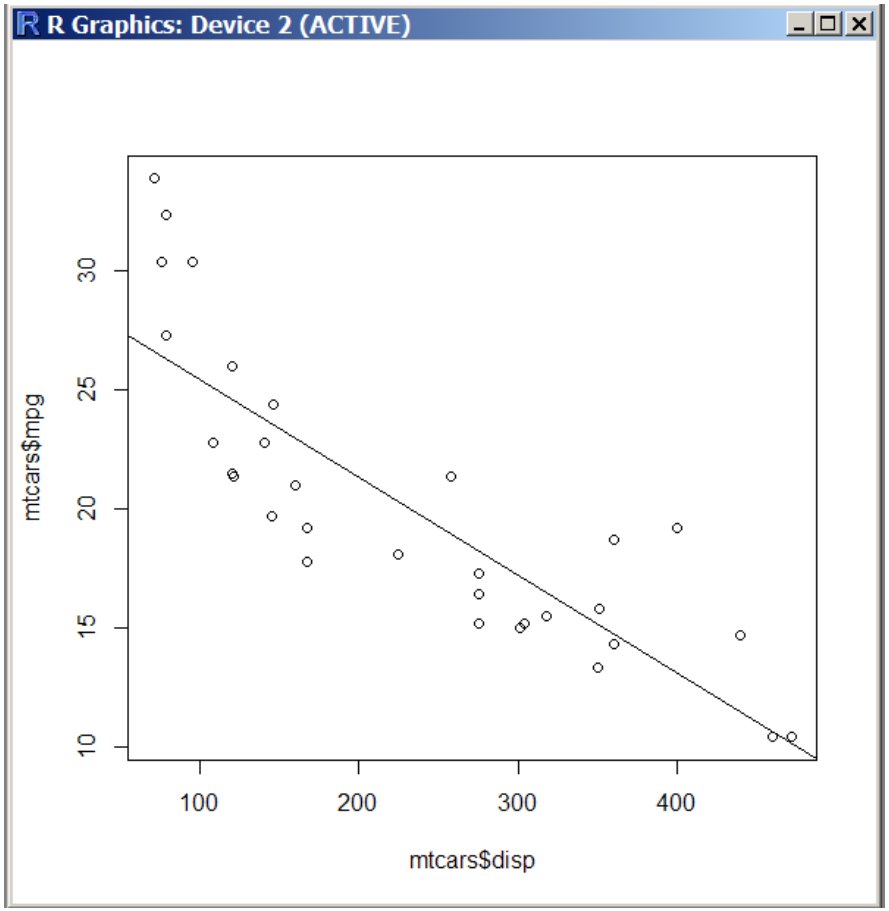
plot(jitter(x), jitter(y))

Jitter函数：给向量加上
少许噪音



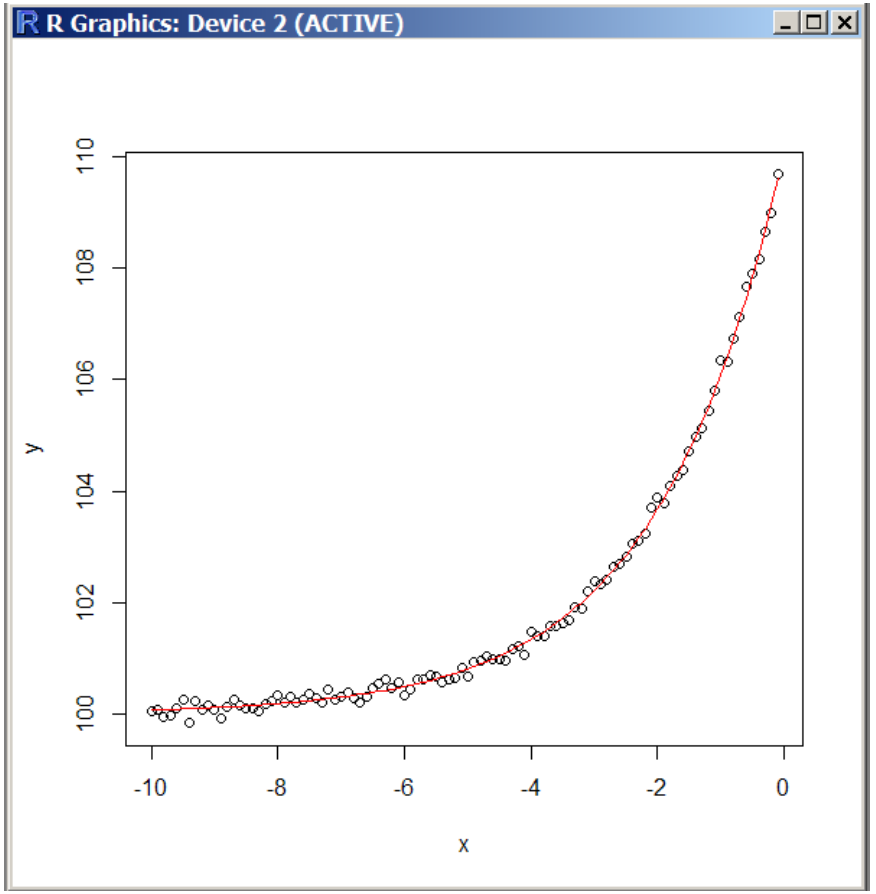
线性模型：画回归直线

```
plot(mtcars$mpg~mtcars$disp)
lmfit<-lm(mtcars$mpg~mtcars$disp)
abline(lmfit)
```



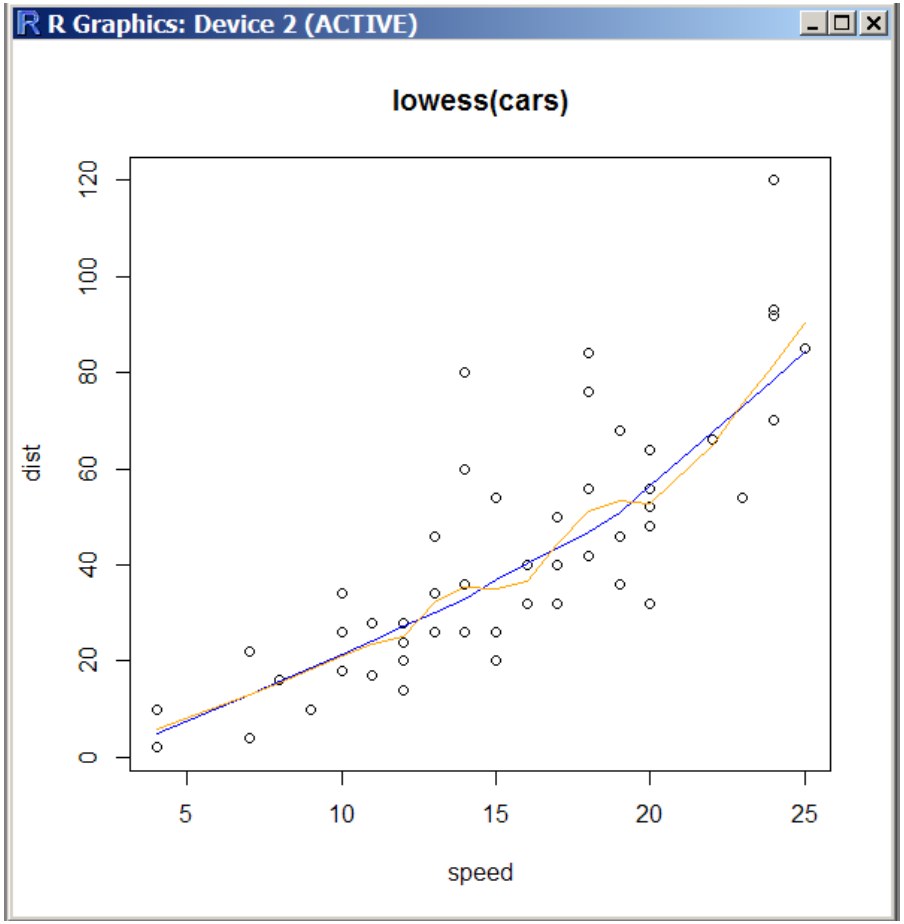
非线性模型的拟合曲线

```
x <- -(1:100)/10
y <- 100 + 10 * exp(x / 2) +
     rnorm(x)/10
nlmod <- nls(y ~ Const + A * exp(B
     * x), trace=TRUE)
plot(x,y)
lines(x, predict(nlmod), col="red")
```



lowess：局部加权回归散点平滑法

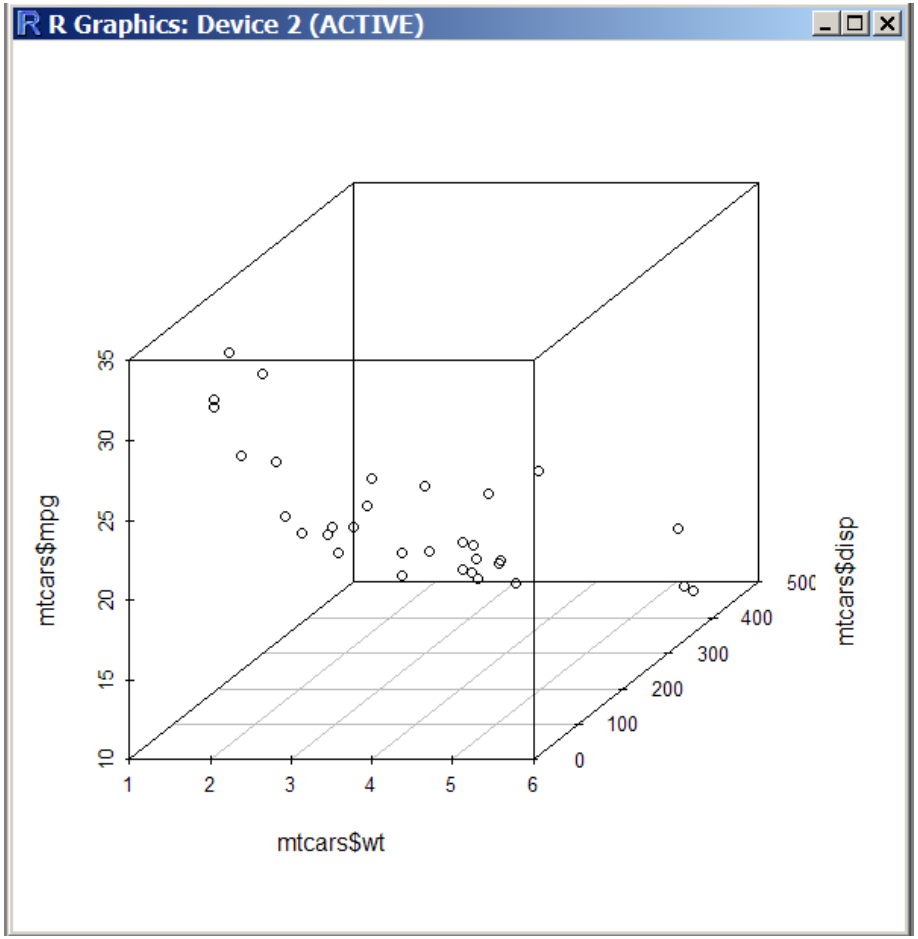
```
plot(cars, main = "lowess(cars)")
lines(lowess(cars), col = "blue")
lines(lowess(cars, f=0.3), col =
    "orange")
```



三维散点图

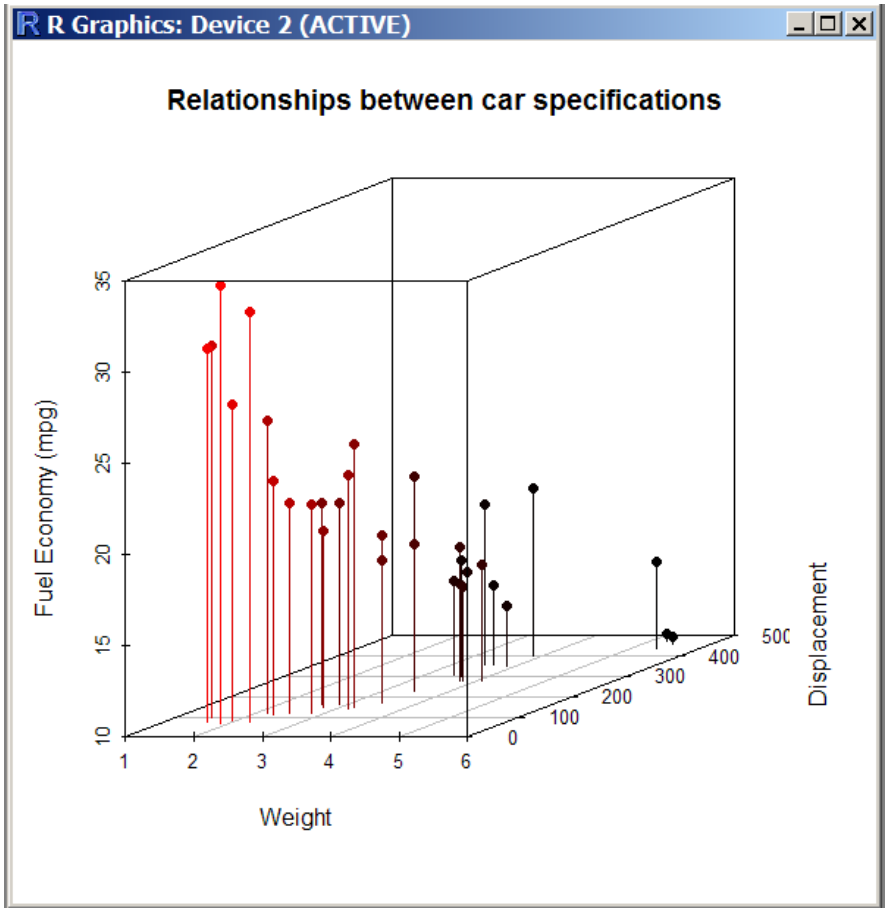
```
install.packages("scatterplot3d")
library(scatterplot3d)

scatterplot3d(x=mtcars$wt,
y=mtcars$disp,
z=mtcars$mpg)
```



三维散点图

```
scatterplot3d(mtcars$wt,mtcars$disp,mtcars$mpg,
pch=16, highlight.3d=TRUE,
angle=20,
xlab="Weight",ylab="Displacement",zlab="Fuel Economy (mpg)",
type="h",
main="Relationships between car specifications")
```



高维数据展现专业扩展包rggobi

install.packages("rggobi")

http://www.ggobi.org/rggobi/

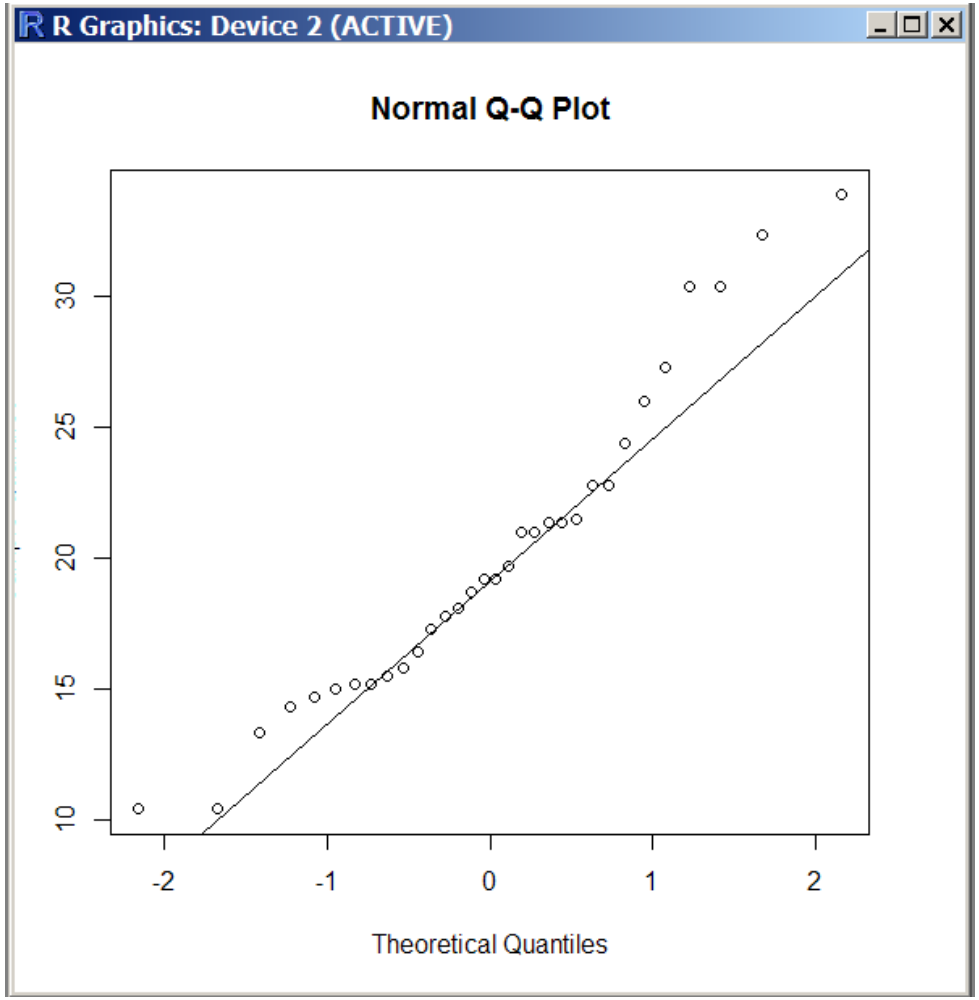
动画demo

http://www.ggobi.org/demos/

QQ图

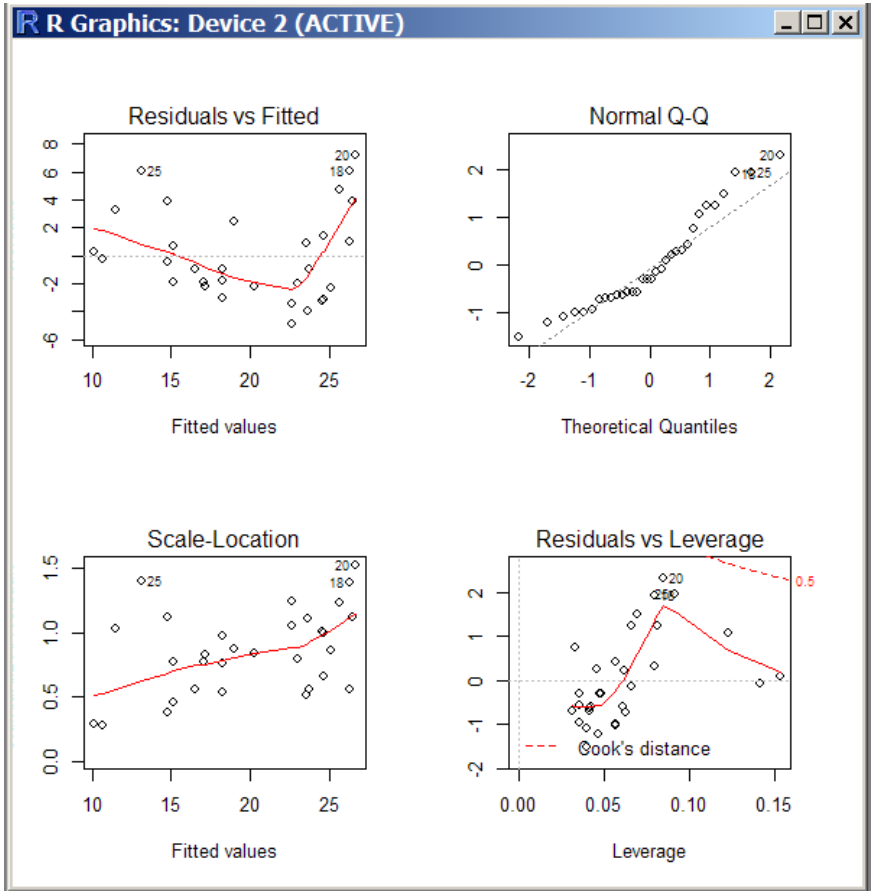
qqnorm(mtcars$mpg)

qqline(mtcars$mpg)



线性回归模型的检验

```
lmfit<-lm(mtcars$mpg~mtcars$disp)
par(mfrow=c(2,2))
plot(lmfit)
```

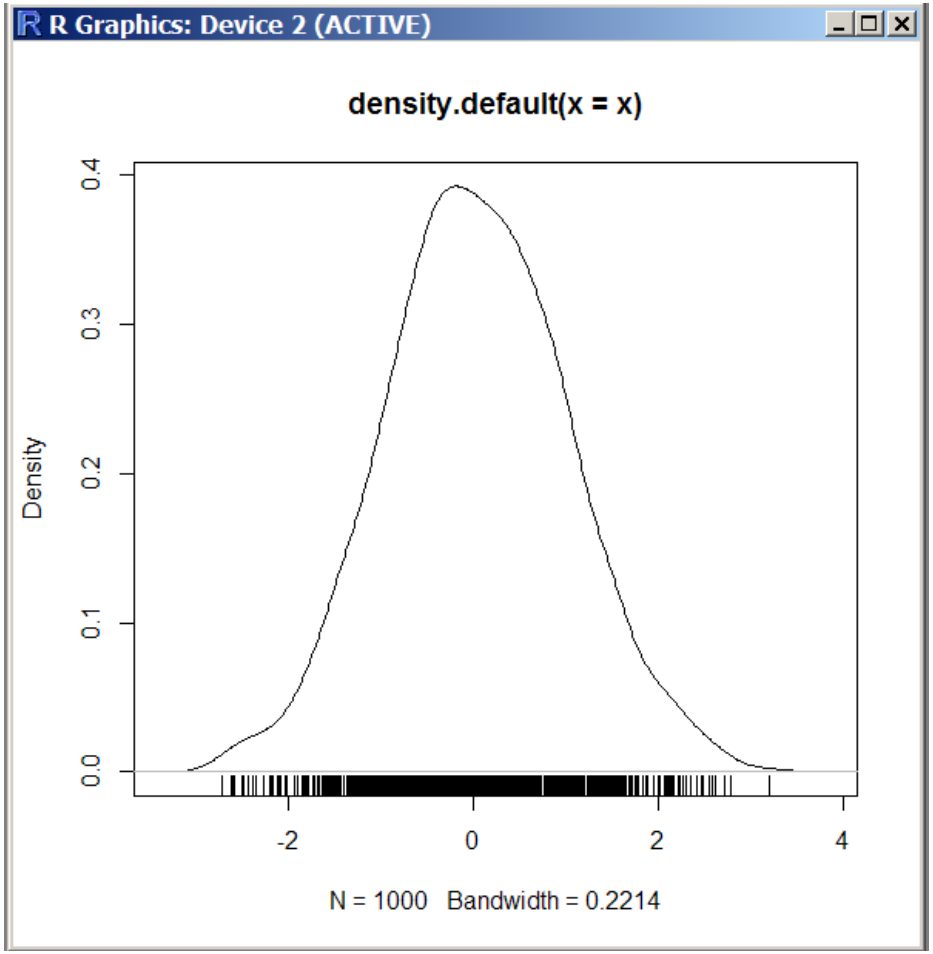


画密度函数

x<-rnorm(1000)

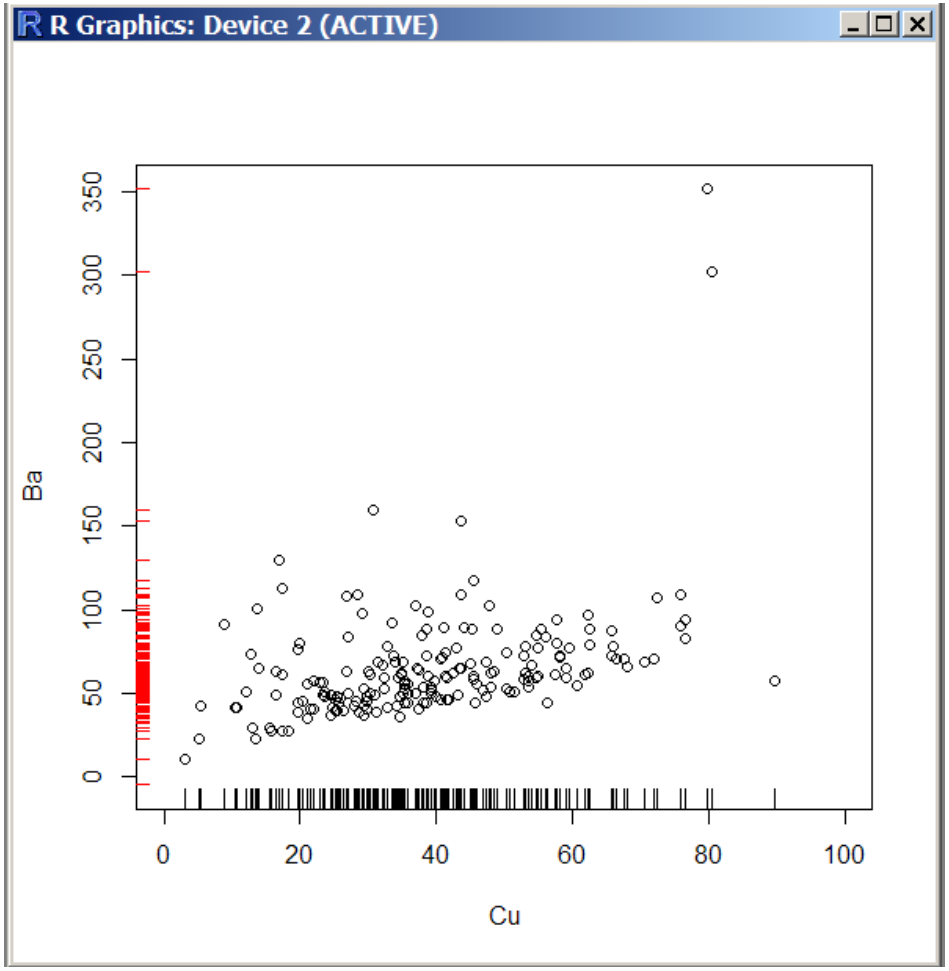
plot(density(x))

rug(x)



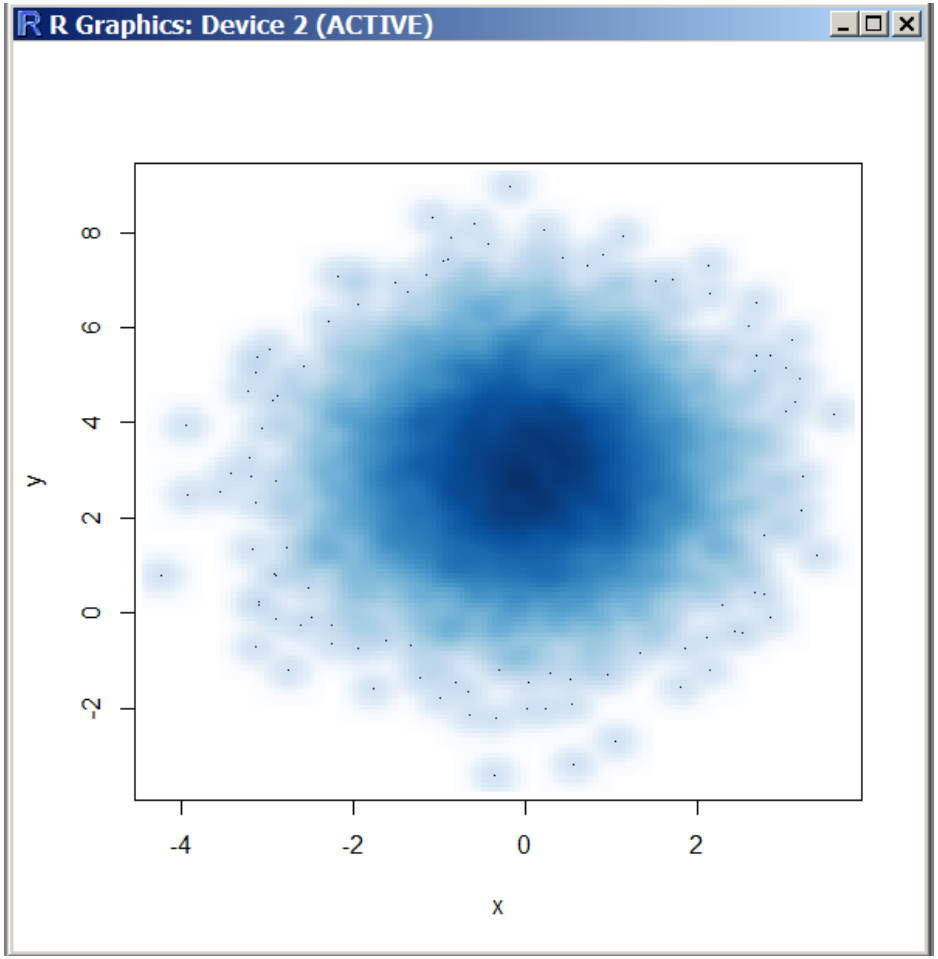
rug( )

```
metals<-read.csv("metals.csv")
plot(Ba~Cu,data=metals,xlim=c
    (0,100))
rug(metals$Cu)
rug(metals$Ba,side=2,col="red"
    ,ticksize=0.02)
```



smoothScatter()

```
n <- 10000
x <- matrix(rnorm(n), ncol=2)
y <- matrix(rnorm(n, mean=3,
    sd=1.5), ncol=2)
smoothScatter(x,y)
```



2013.03.09

早期社交图

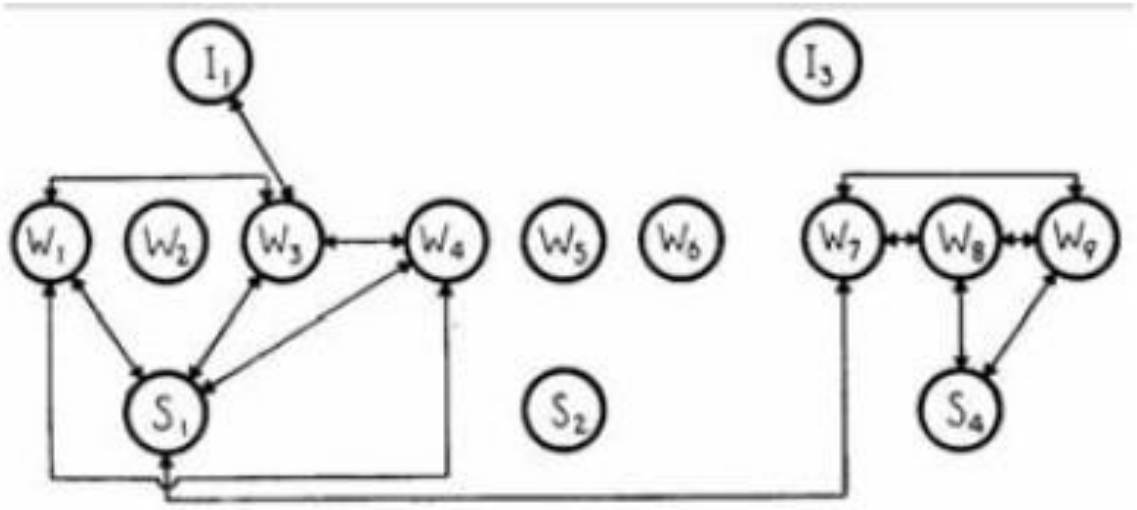


FIGURE 43
FRIENDSHIPS

早期社交图

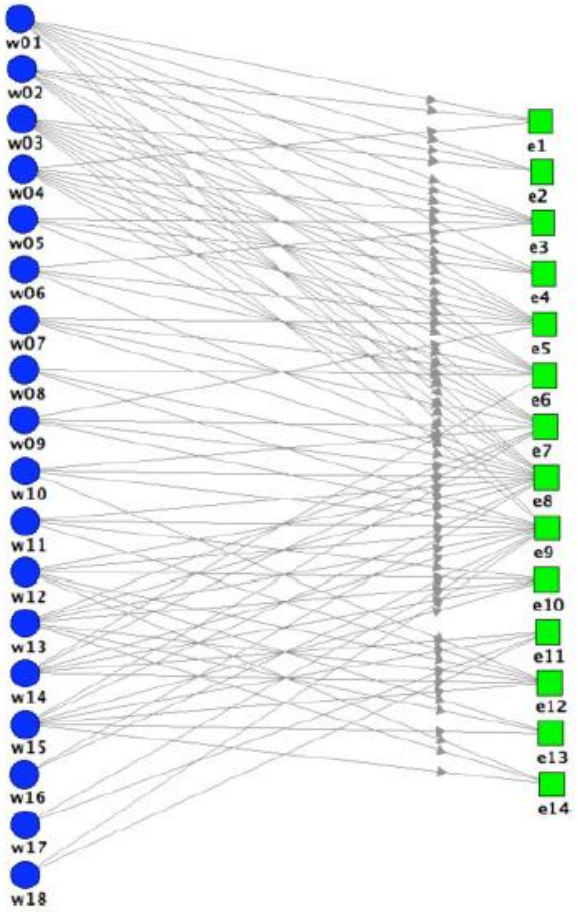


Figure 7-2. *Two-mode view of the Southern Women social event dataset*

2013.03.09

改进：活动布局图

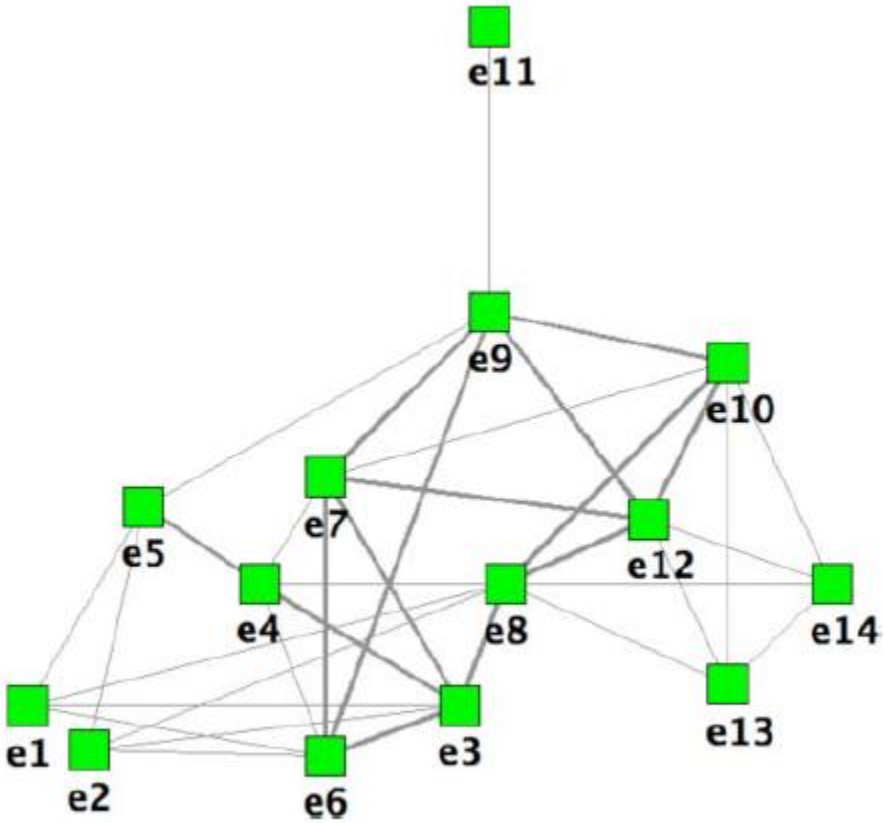
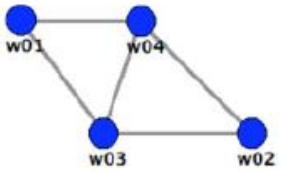


Figure 7-3. *Layout of events based on attendance by people in common*

聚类



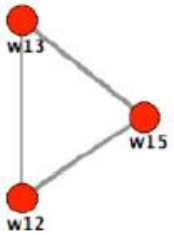


Figure 7-4. *Strongest ties amongst women based on common event attendance*

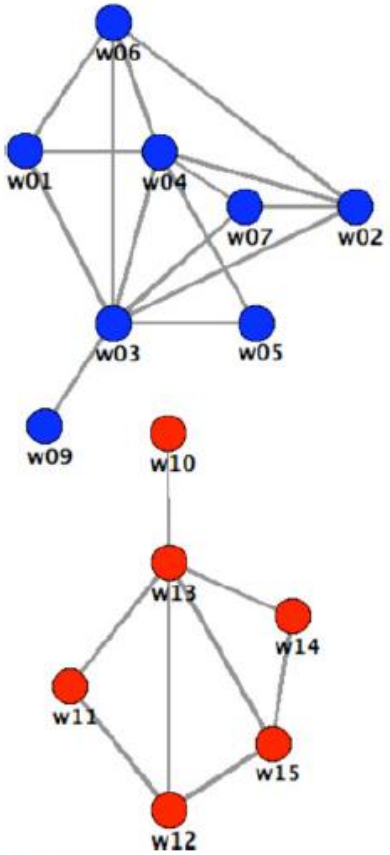


Figure 7-5. *The two strongest link levels between women attending common social events*

2013.03.09

桥接

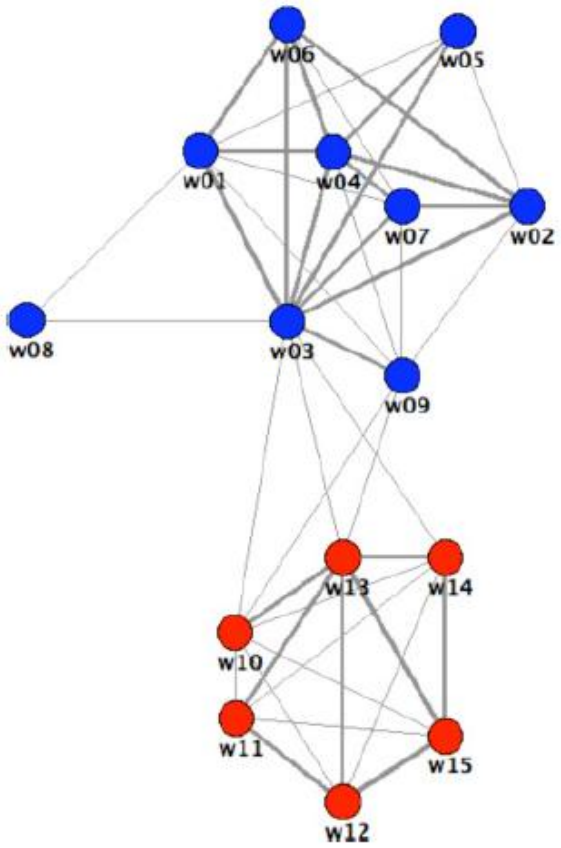


Figure 7-6. *The two groups are bridged with gradual inclusion of weaker ties*

解读社交图

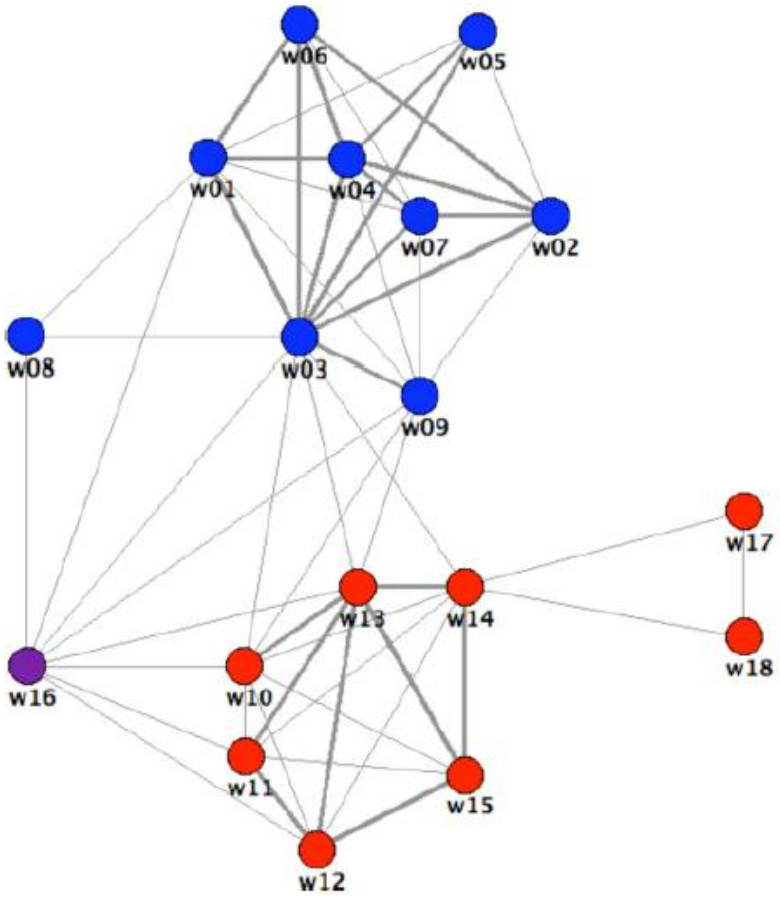
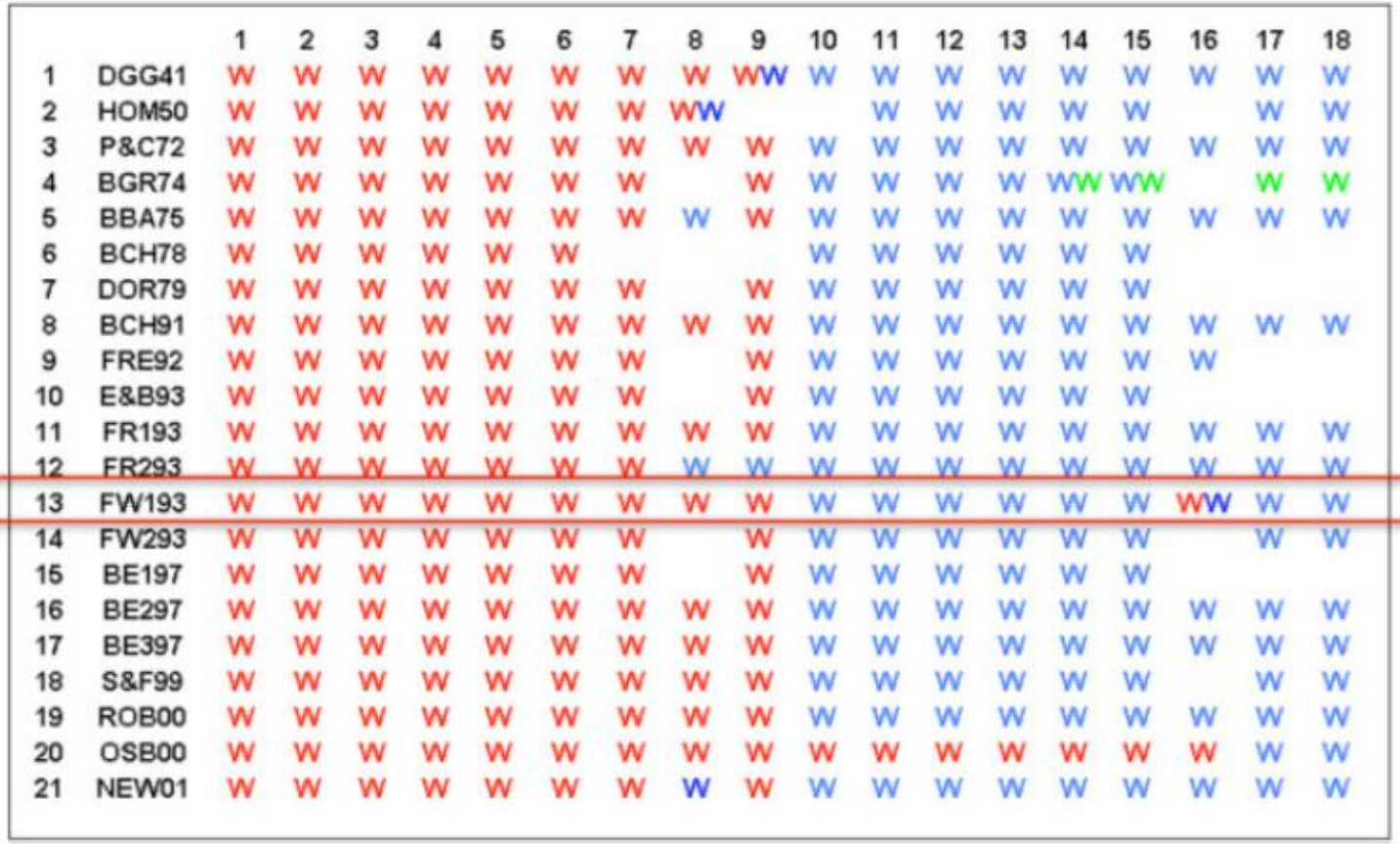


Figure 7-7. *Emergent social graph of women based on common attendance at social events*

2013.03.09

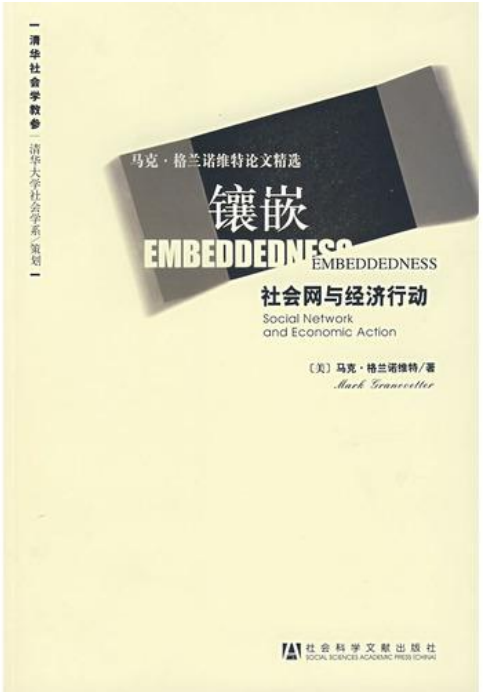
一些研究

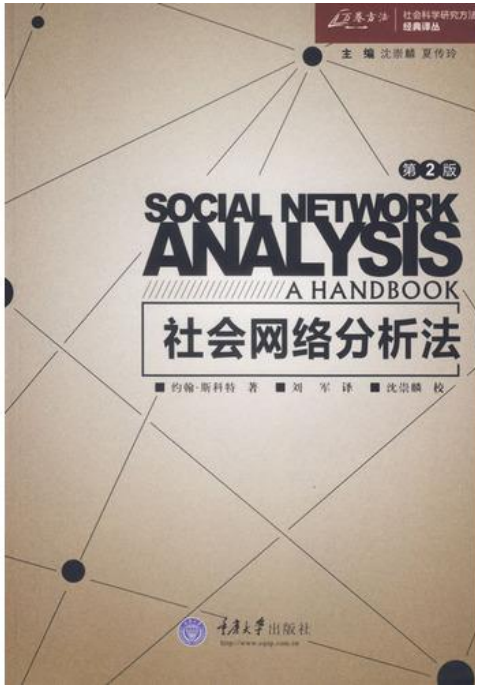


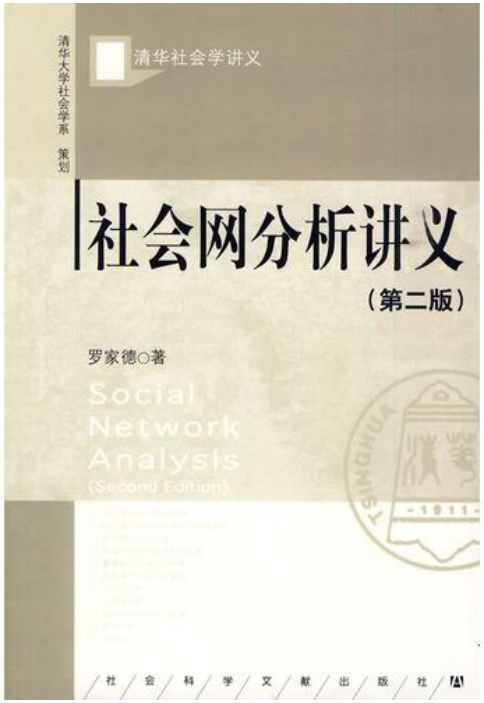
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	DGG41	W	W	W	W	W	W	W	W	WW	W	W	W	W	W	W	W	W	W
2	HOM50	W	W	W	W	W	W	W	WW			W	W	W	W	W		W	W
3	P&C72	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
4	BGR74	W	W	W	W	W	W	W		W	W	W	W	W	WW	WW		W	W
5	BBA75	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
6	BCH78	W	W	W	W	W	W				W	W	W	W	W	W			
7	DOR79	W	W	W	W	W	W	W		W	W	W	W	W	W	W			
8	BCH91	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
9	FRE92	W	W	W	W	W	W	W		W	W	W	W	W	W	W	W		
10	E&B93	W	W	W	W	W	W	W		W	W	W	W	W	W	W			
11	FR193	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
12	FR293	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
13	FW193	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	WW	W	W
14	FW293	W	W	W	W	W	W	W		W	W	W	W	W	W	W		W	W
15	BE197	W	W	W	W	W	W	W		W	W	W	W	W	W	W			
16	BE297	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
17	BE397	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
18	S&F99	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W		W	W
19	ROB00	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
20	OSB00	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
21	NEW01	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W

Figure 7-8. *Results of 21 studies of the Southern Women social event dataset by network scientists (Freeman 2003)*

参考书


清华社会学教参
马克·格兰诺维特论文精选
镶嵌
EMBEDDEDNESS
社会网与经济行动
Social Network
and Economic Action
〔美〕马克·格兰诺维特/著
社会科学文献出版社


第2版
SOCIAL NETWORK
ANALYSIS
A HANDBOOK
社会网络分析法
约翰·斯科特 著
刘 军 译
沈崇麟 校
重庆大学出版社


清华社会学讲义
清华大学社会学系 策划
社会网分析讲义
（第二版）
罗家德○著
Social
Network
Analysis
(Second Edition)
社会科学文献出版社

图书推荐

购买本商品的顾客还买过

R语言初学者指南

¥ 28.80

- 统计建模与R软件
- 多元统计分析及R语言建
- 统计模拟及其R实现
- 时间序列分析及应用（R
- C++ Primer Plus（第6
- 数学之美——架起从数学

更多>>

浏览本商品的顾客还看过

R语言与统计分析

¥ 30.40

- 多元统计分析及R语言建

更多>>

R语言数据操作

包含此书的书单

分享到：

当 当 价：￥27.20

定　　价：¥ 34.00　折扣:80折

顾客评分：★★★★★ 已有184人评价，97.8%推荐

配 送 至：广东广州市海珠区　缺货

作　　者：（美）斯佩克特　著，朱钰，柴文义，张颖　译

出 版 社：西安交通大学出版社

出版时间：2011-7-1

版　次：1	页　数：167	字　数：176000
印刷时间：2011-7-1	开　本：16开	纸　张：胶版纸
印　次：1	I S B N：9787560538730	包　装：平装

我要买：1 件

购　买　　收　藏

送积分：272　收藏人气：203　50000000张优惠券疯抢！！

2013.03.09

亚马逊书籍购买数据形成的社交图

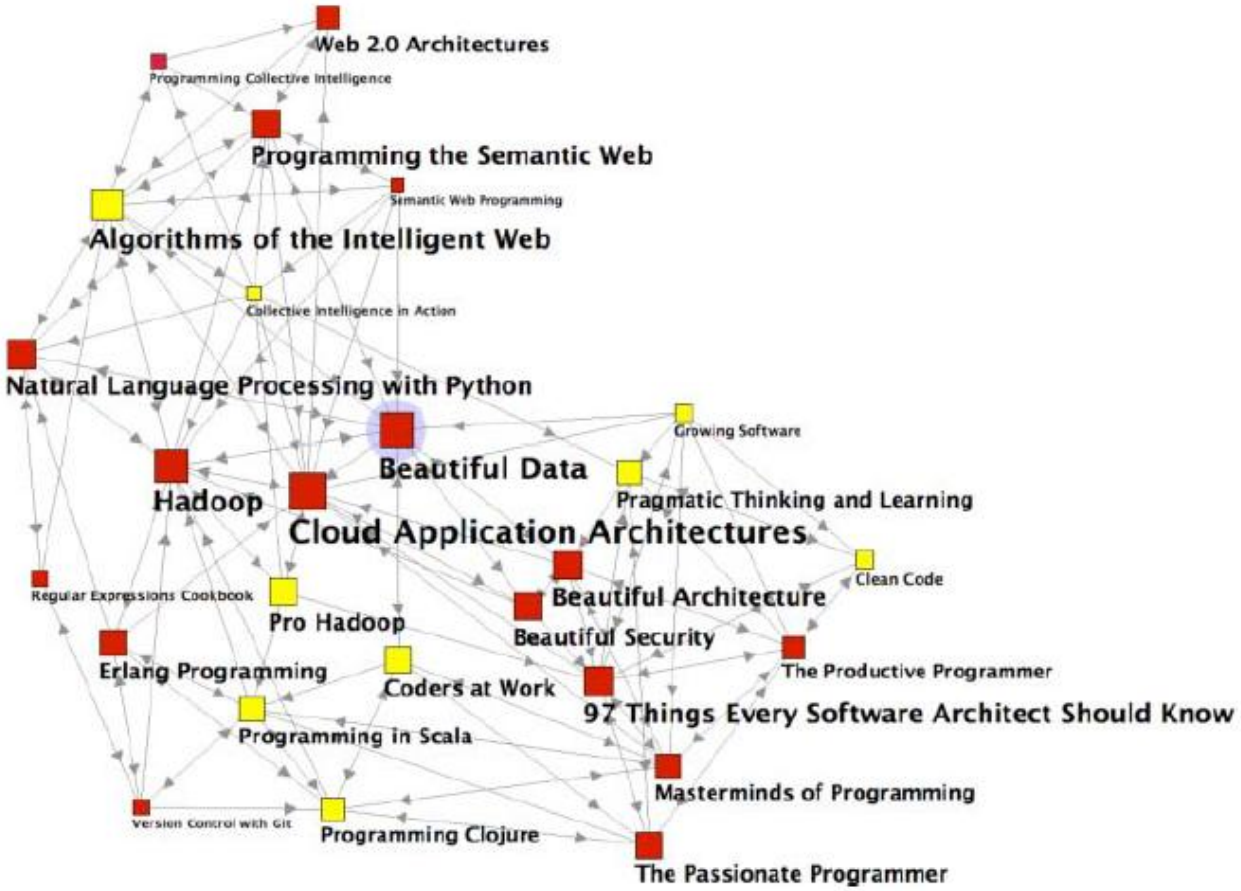


Figure 7-9. *The network neighborhood of books surrounding Beautiful Data*

2013.03.09

政治书与政治阵营

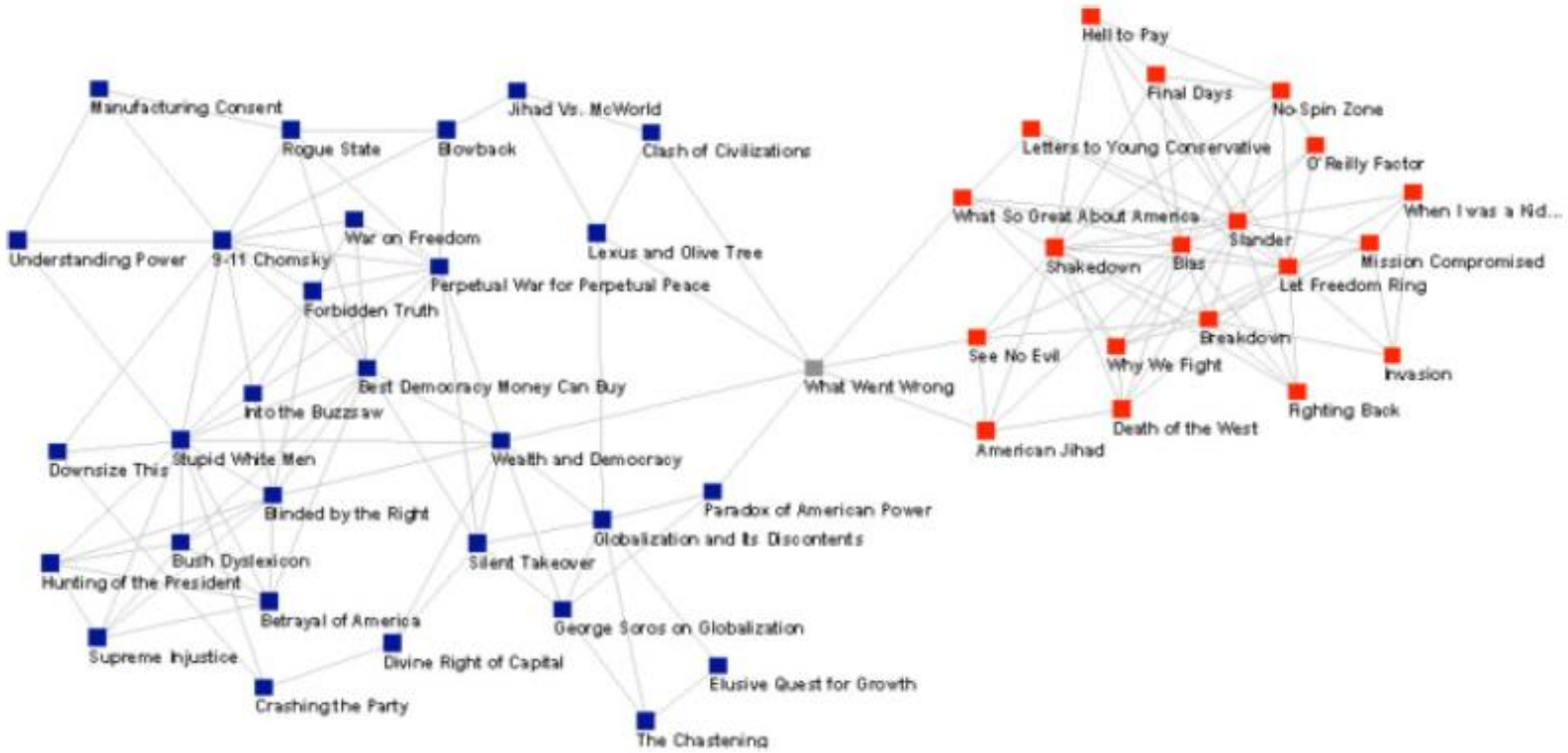


Figure 7-10. *Divide of political books in 2003*

政治书与政治阵营

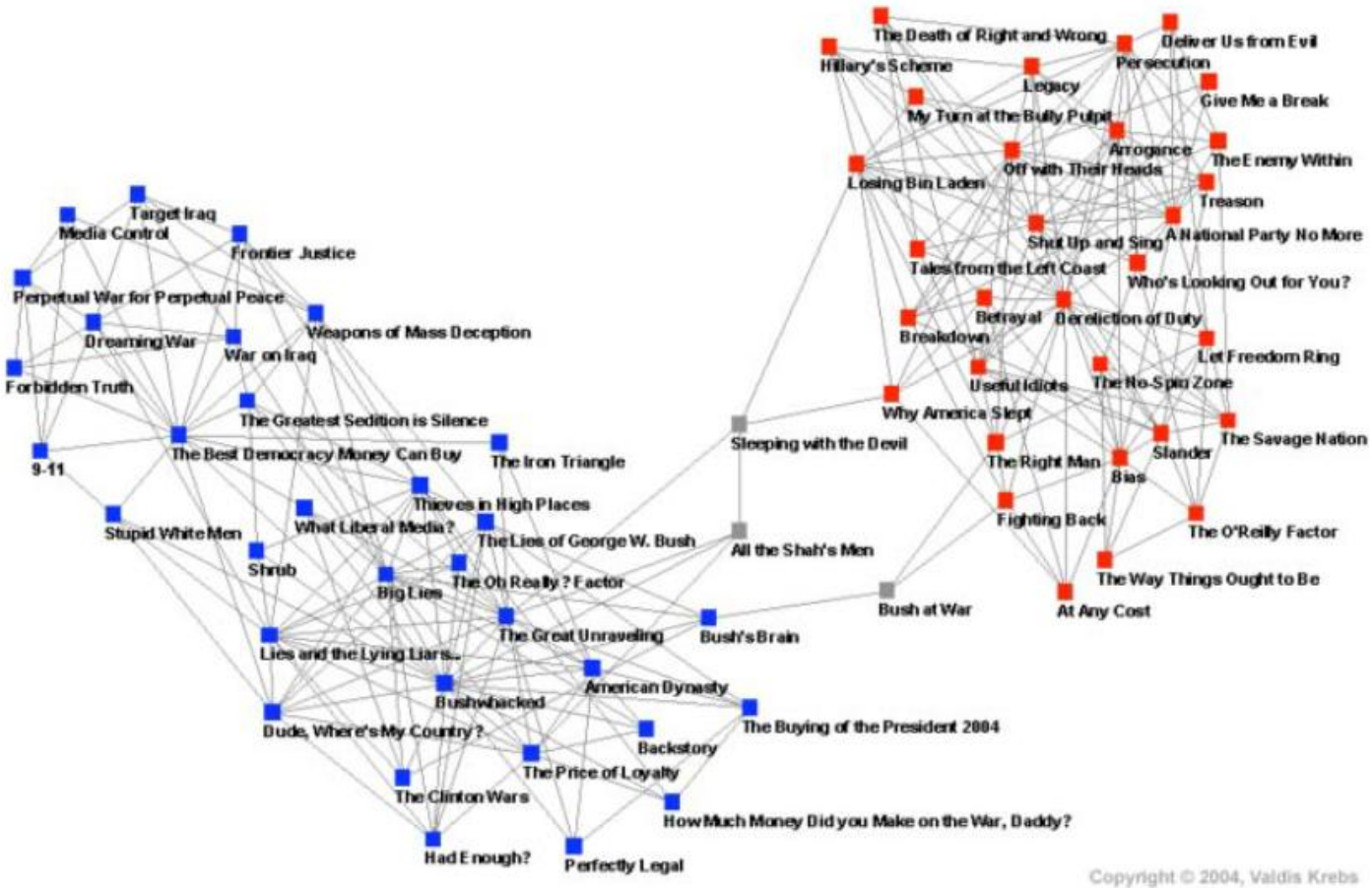


Figure 7-11. *Divide of political books in 2004*

2013.03.09

Thanks

FAQ时间