Data Analysis Tools and Practice(Using R)

2020.03.19

R基本图形II

北京大學 软件与微电子学院
School of Software and Microelectronics, Peking University

Huiping Sun(孙惠平)
sunhp@ss.pku.edu.cn

上次课程内容回顾

- 图形函数：
  - ✴ plot(); barplot(); pie(); hist(); boxplot();
- 图形参数：
  - ✴ col; font; pch; cex; lty; lwd; xlab; ylab; xlim; ylim; type; main; horiz; beside;
- 图例函数：
  - ✴ legend(location, title, legend, …);
- 图形组合：
  - ✴ par(); layout();
- 其余函数：
  - ✴ title(); abline(); line(); text(); mtext();

作业讲解

练习－0015

- 模拟产生100个学号（1300022001到1300022100）
- 模拟产生三个科目的成绩，要求第一科最大值99，最小值70；第二科平均值81，sd=7，最大值100；第三科平均值83，sd＝18，最大值100
- 把学号和三科成绩组成一个数据框，显示数据框内容
- 求每个学生的总分、平均分
- 针对三科成绩、总分、平均分，分别做饼图、直方图、条形图，箱线图
- 分别用par和layout把多个图放在一个图中显示：同一个数据的不同类的图形，不同数据的同一类，不同数据的不同图形

练习－0016

- 某校测的19名学生的四项指标：性别、年龄、身高（cm）、体重（磅），具体见0016_student.CSV，要求：
  - 绘出体重对于身高的散点图
  - 绘出不同性别情况下，体重与身高的散点图
  - 绘出不同年龄段的体重与身高的散点图
  - 绘出不同性别和不同年龄段的体重与身高的散点图

- 0016_height01.txt，画直方图
- 0016_height02.txt，画箱式图
- 0016_marriage.txt，画散点图
- 0016_language.txt，画条形图（母语和日常使用）
- 0016_language.txt，画饼图（世界主要语种使用人数比例）

练习 – 0017

- 从0017_grade.csv中读取两班成绩
- 计算每个班级的均值和标准方差
- 计算每个人的标准化成绩，添加到数据中，写到0017grade.txt中
- 分别画出来两班成绩和标准成绩的箱线图
- 在一张图中画出两班成绩和标准成绩的箱线图

练习 – 0018

```
plot(rnorm(1000),col=“red")
```

- 使用上面的语句，练习颜色的各种表示方法
- 使用Par和layout函数，分别显示不同颜色的多个图形组合，2*2，3*3，1*1*2*3等

课件第12页，citysales.csv

- 输入现有代码，看显示结果
- 用rainbow、top.colors、cm.colors、gray、terrian.colors替换heat.colors，看执行效果
- 练习课件第23页的颜色参数
- 添加图例

练习 – 0019

cityrain.csv

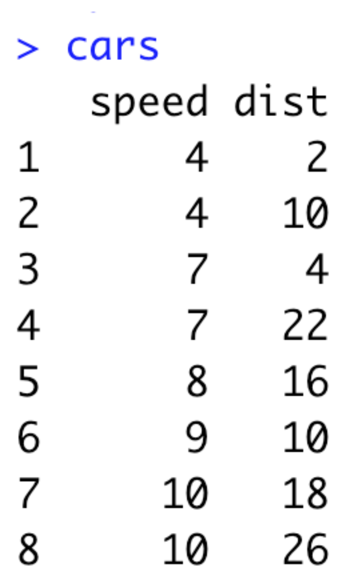
- 用不同颜色画出不同城市的线图
- 用不同符号画出不同城市的线图
- 用不同颜色画出不同城市的散点图
- 用不同符号画出不同城市的散点图
- 分别加上图例
- 用par和layout把前面四个图放在一张图中，分别为2*2，1*4，1+2+1

图形

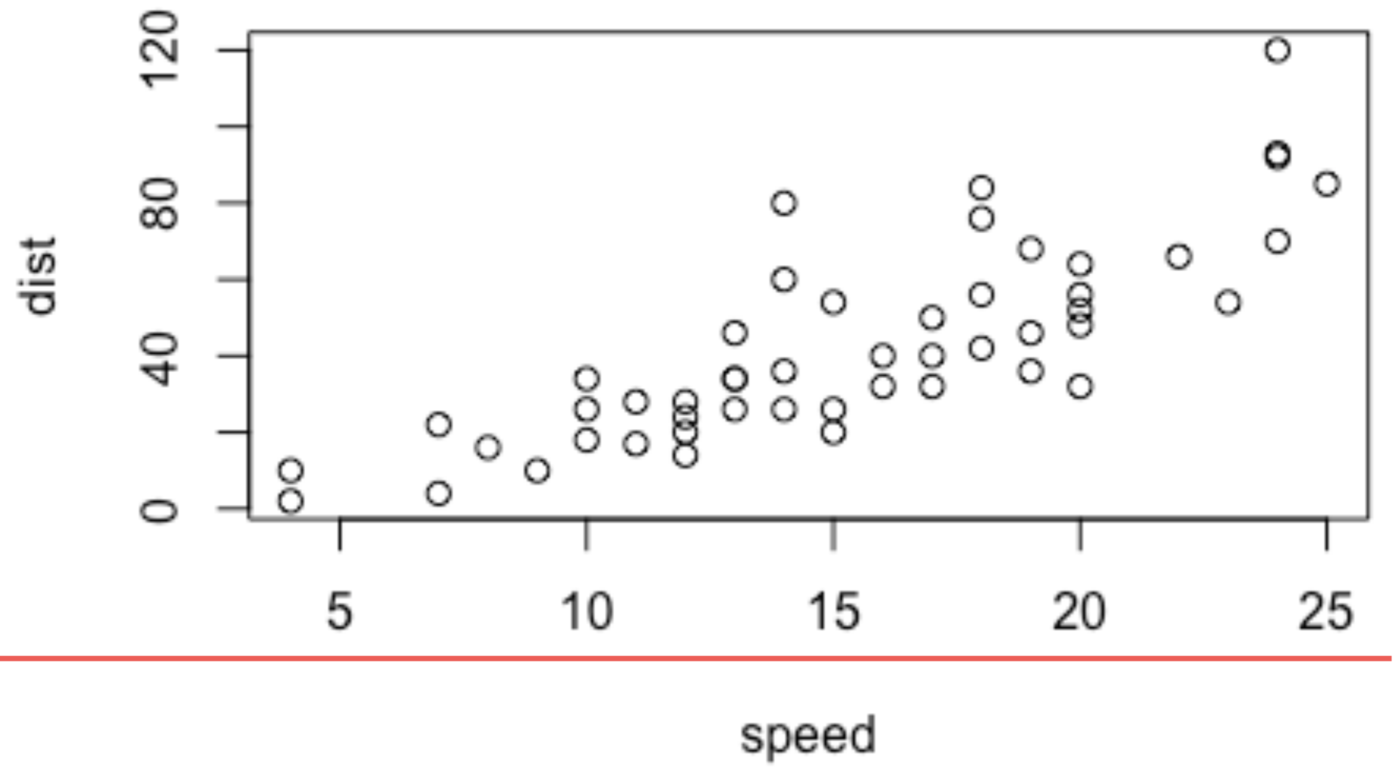
R Cookbook

散点图

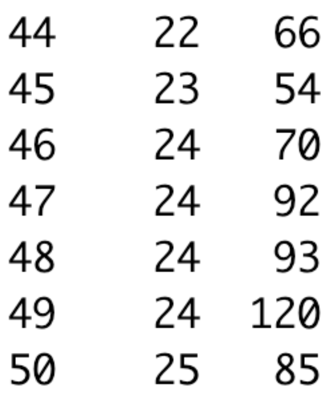
```
plot(cars)
```



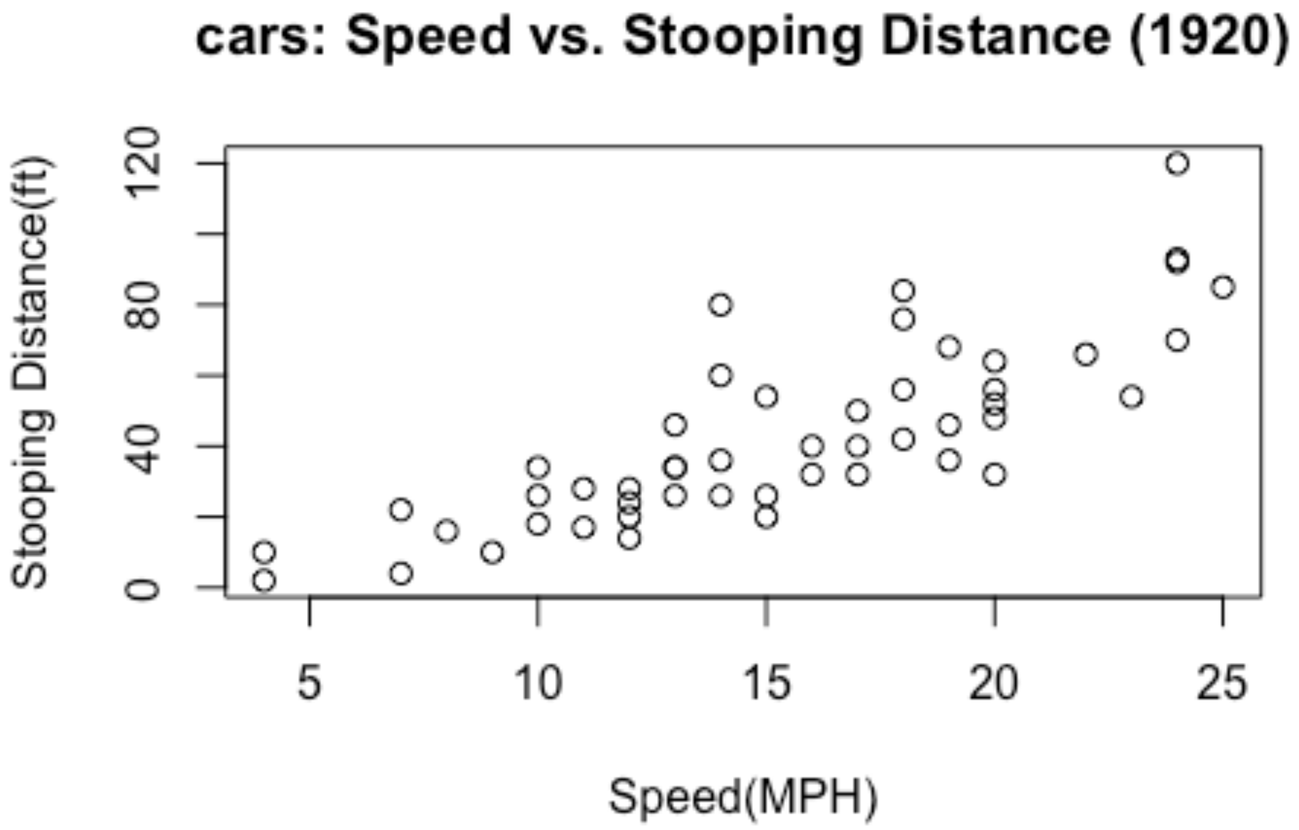
```
> cars
  speed dist
1     4    2
2     4   10
3     7    4
4     7   22
5     8   16
6     9   10
7    10   18
8    10   26
```



```
plot(cars,
    main = "cars: Speed vs. Stooping Distance (1920)",
    xlab = "Speed(MPH)",
    ylab = "Stooping Distance(ft)")
```



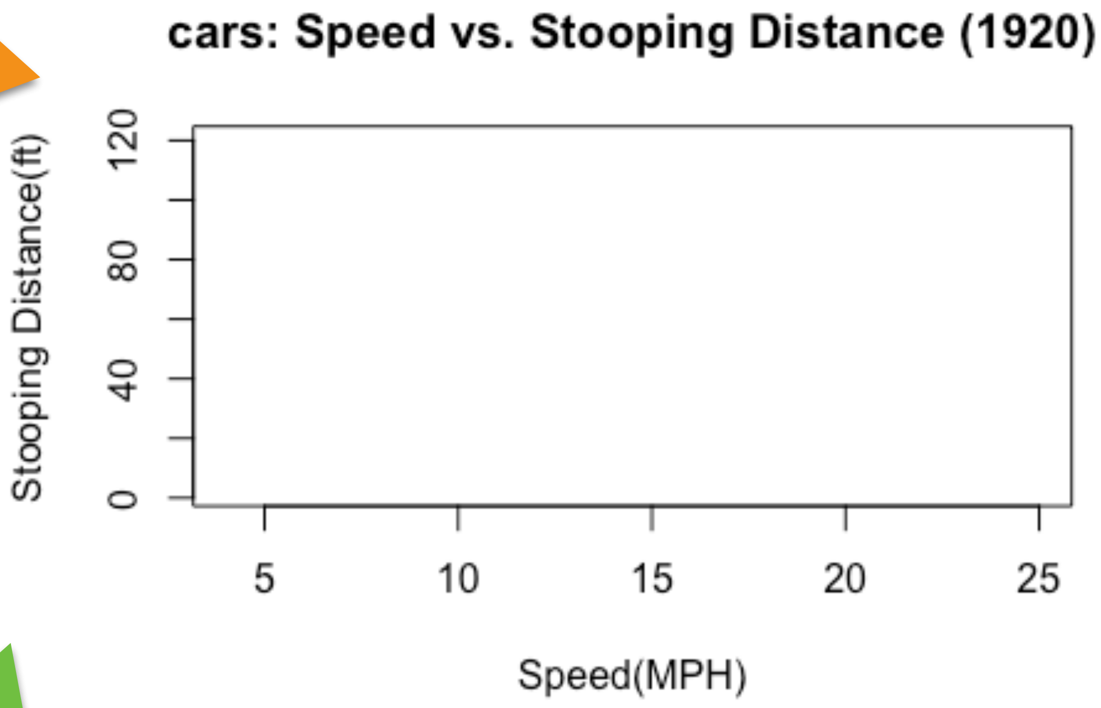
```
44    22   66
45    23   54
46    24   70
47    24   92
48    24   93
49    24  120
50    25   85
```

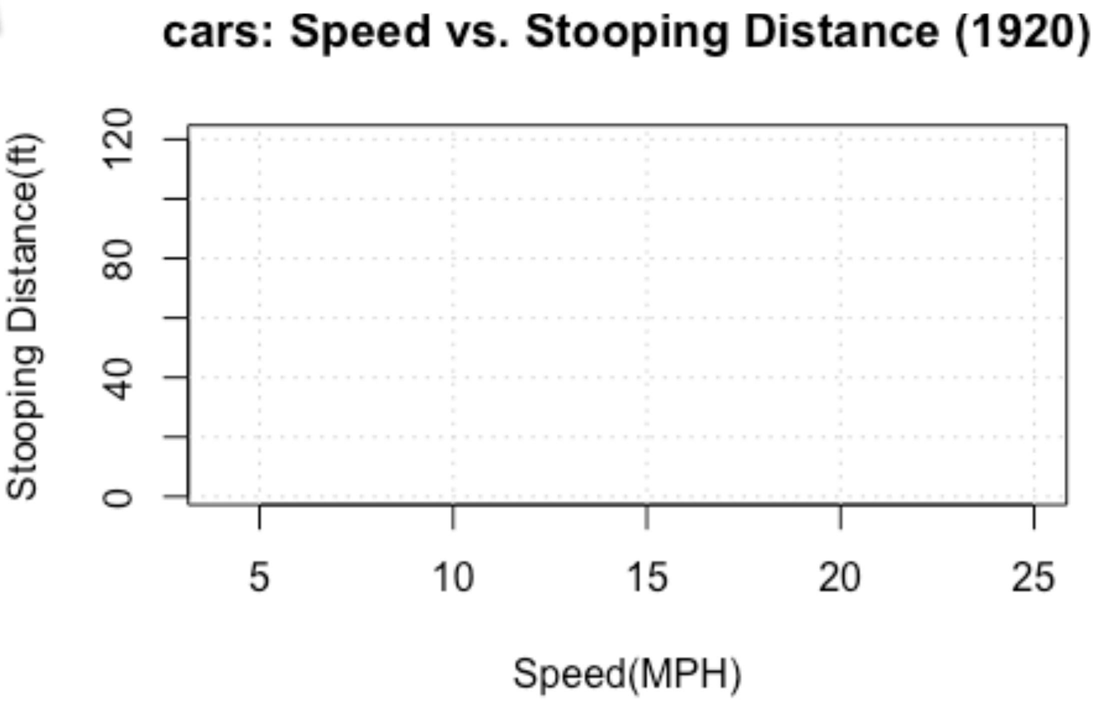


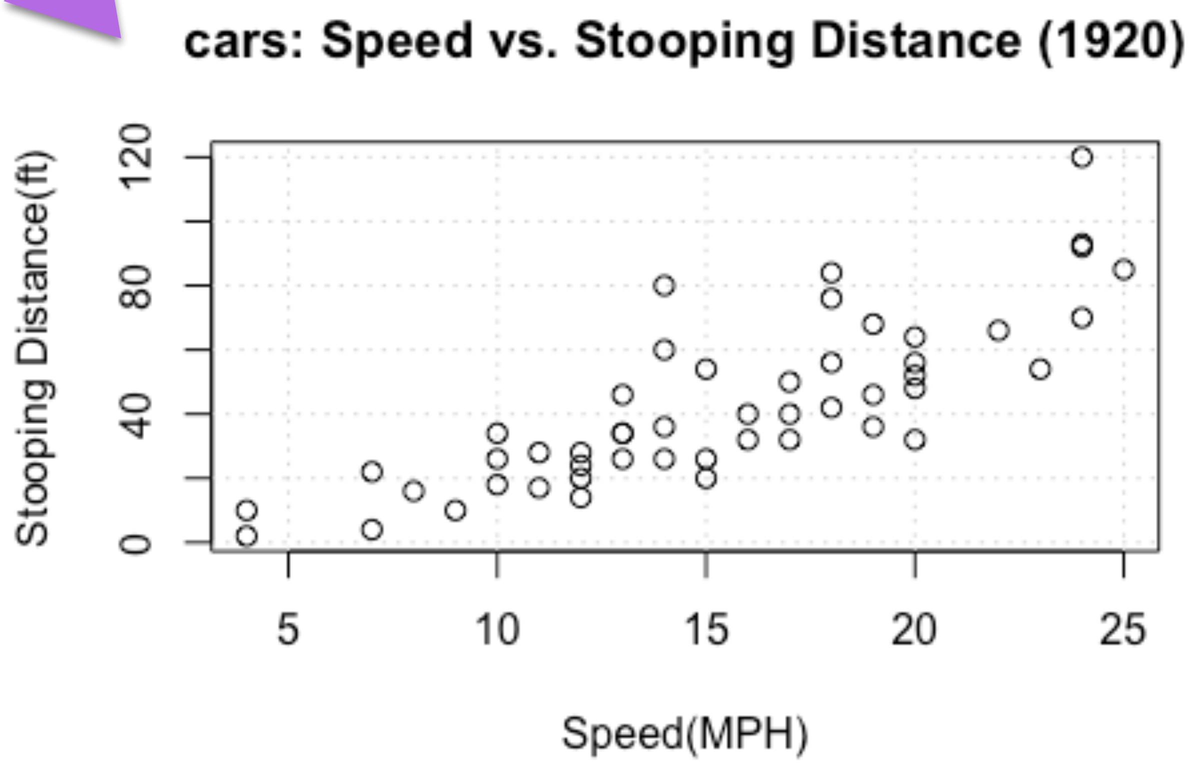
散点图

```
plot(cars,
    main = "cars: Speed vs. Stooping Distance (1920)",
    xlab = "Speed(MPH)",
    ylab = "Stooping Distance(ft)",
    type = “n")
grid()
points(cars)
```

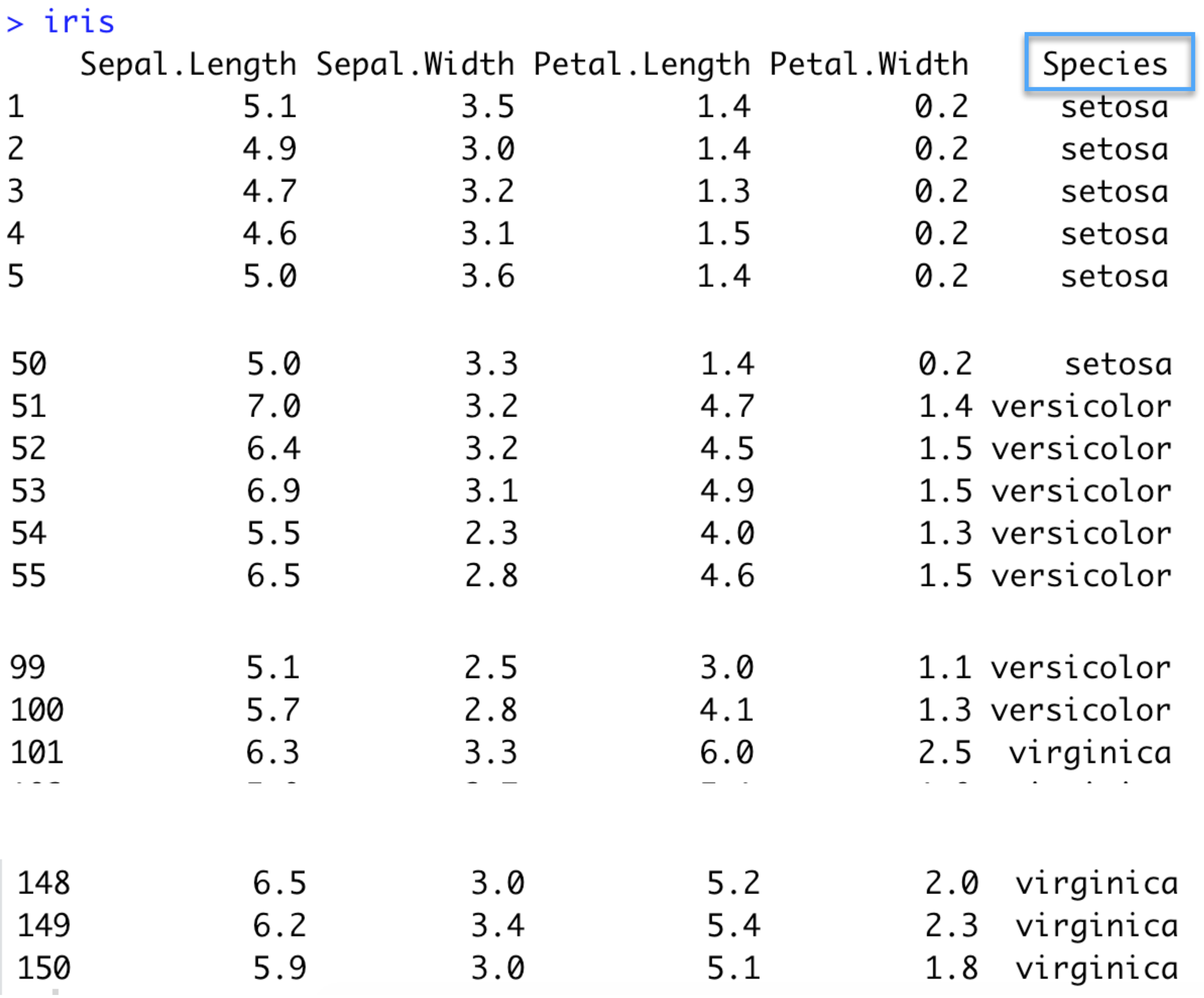
低级函数







多组散点图

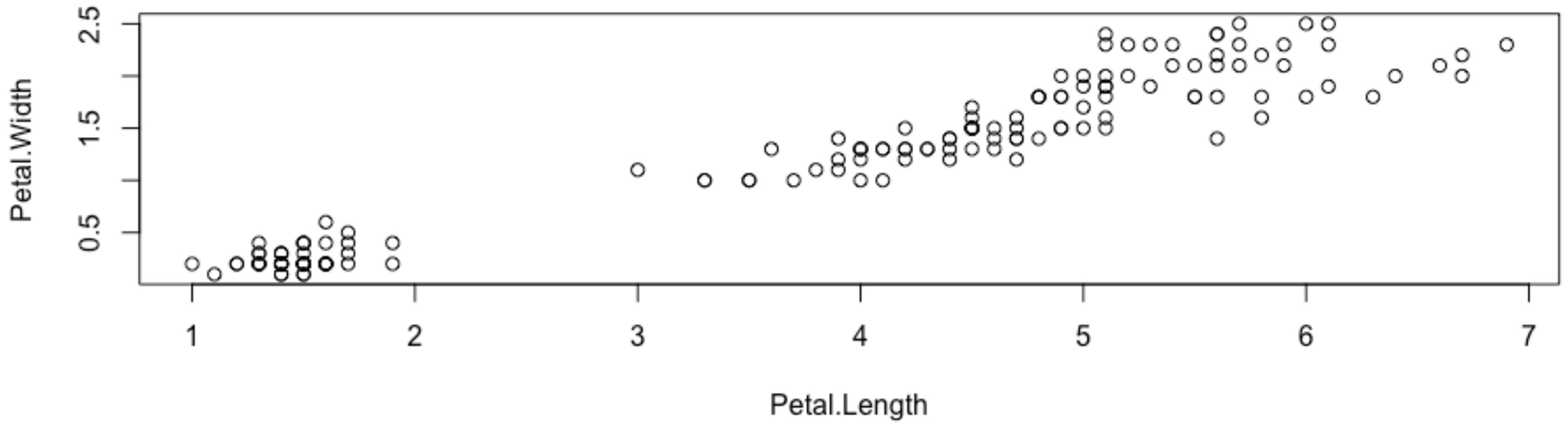


```
> iris
    Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
1            5.1         3.5          1.4         0.2     setosa
2            4.9         3.0          1.4         0.2     setosa
3            4.7         3.2          1.3         0.2     setosa
4            4.6         3.1          1.5         0.2     setosa
5            5.0         3.6          1.4         0.2     setosa

50           5.0         3.3          1.4         0.2     setosa
51           7.0         3.2          4.7         1.4 versicolor
52           6.4         3.2          4.5         1.5 versicolor
53           6.9         3.1          4.9         1.5 versicolor
54           5.5         2.3          4.0         1.3 versicolor
55           6.5         2.8          4.6         1.5 versicolor

99           5.1         2.5          3.0         1.1 versicolor
100          5.7         2.8          4.1         1.3 versicolor
101          6.3         3.3          6.0         2.5  virginica

148          6.5         3.0          5.2         2.0  virginica
149          6.2         3.4          5.4         2.3  virginica
150          5.9         3.0          5.1         1.8  virginica
```

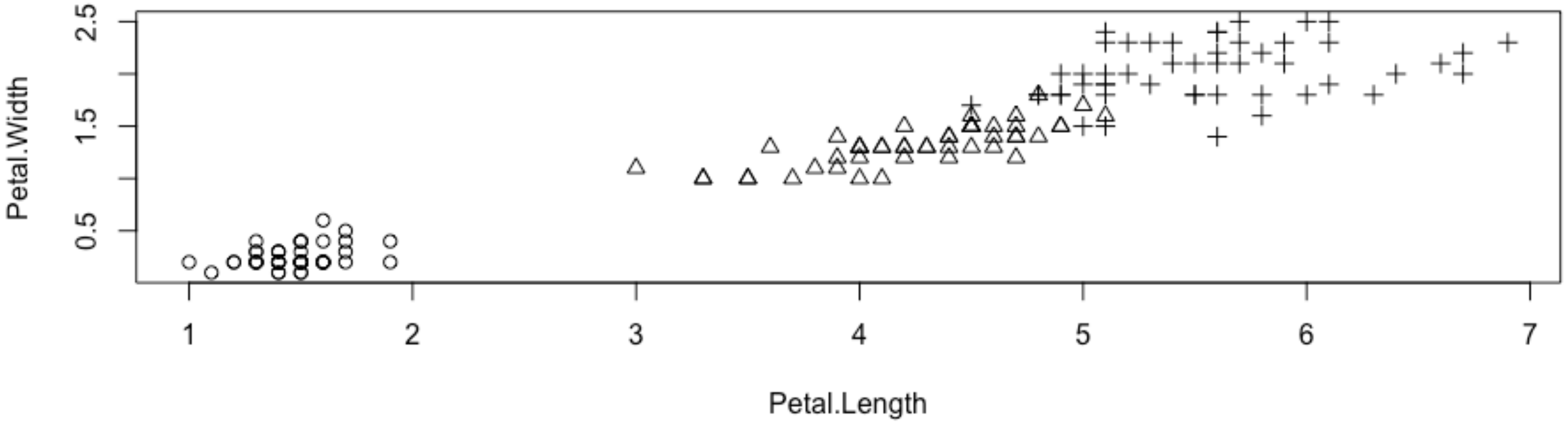
因子

多组散点图

```
with(iris,plot(Petal.Length,Petal.Width))
```

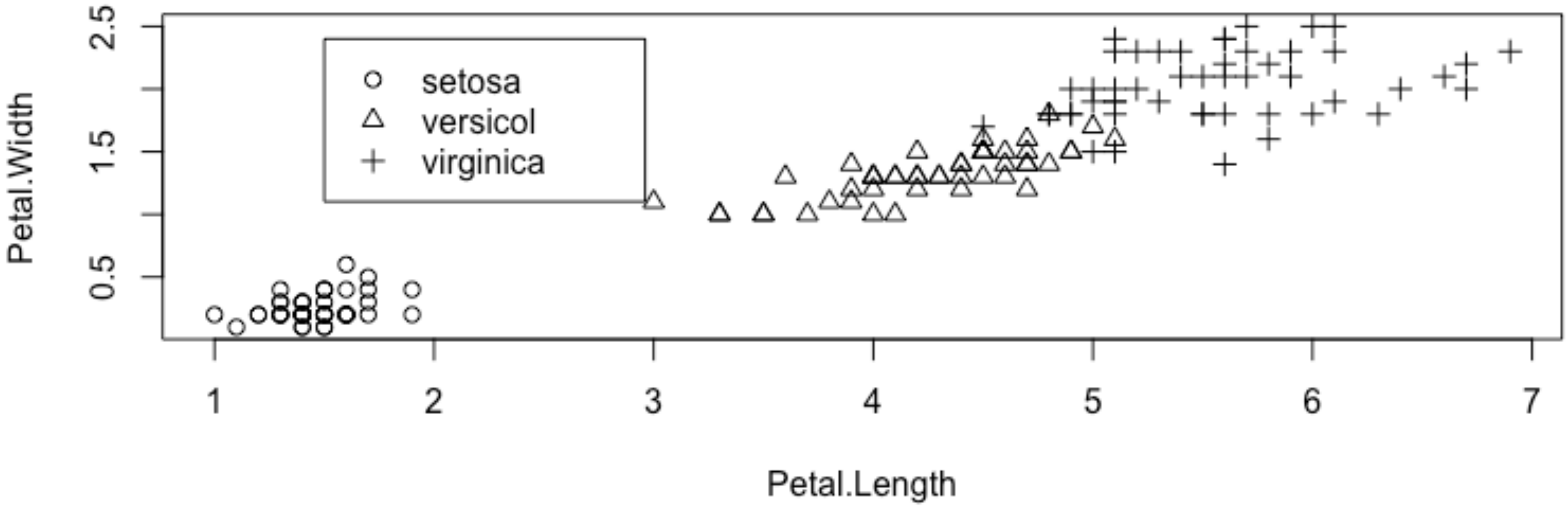


```
with(iris,plot(Petal.Length,Petal.Width,pch=as.integer(Species)))
```



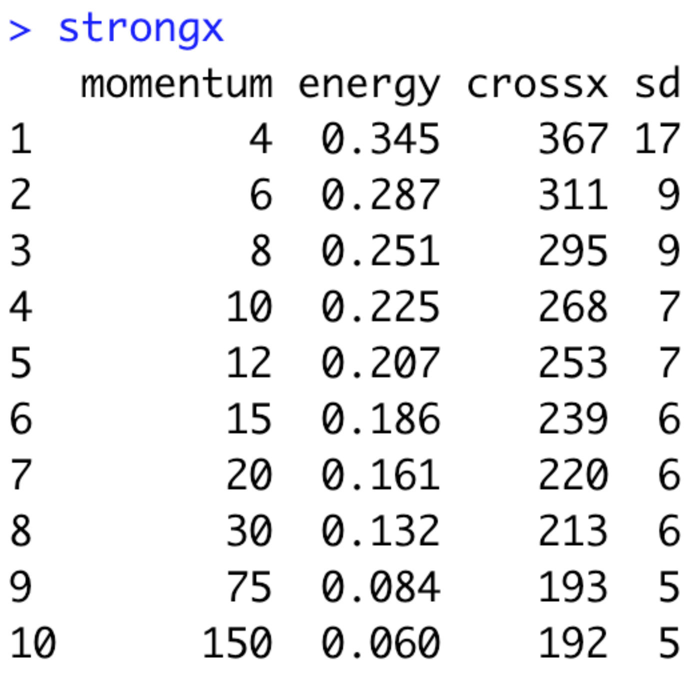
图例

```
legend(1.5, 2.4, c("setosa", "versicol", "virginica"), pch = 1:3)

f <- factor(iris$Species)
with(iris,plot(Petal.Length, Petal.Width, pch=as.integer(Species)))
legend(1.5, 2.4, as.character(levels(f)), pch = 1:3)
```

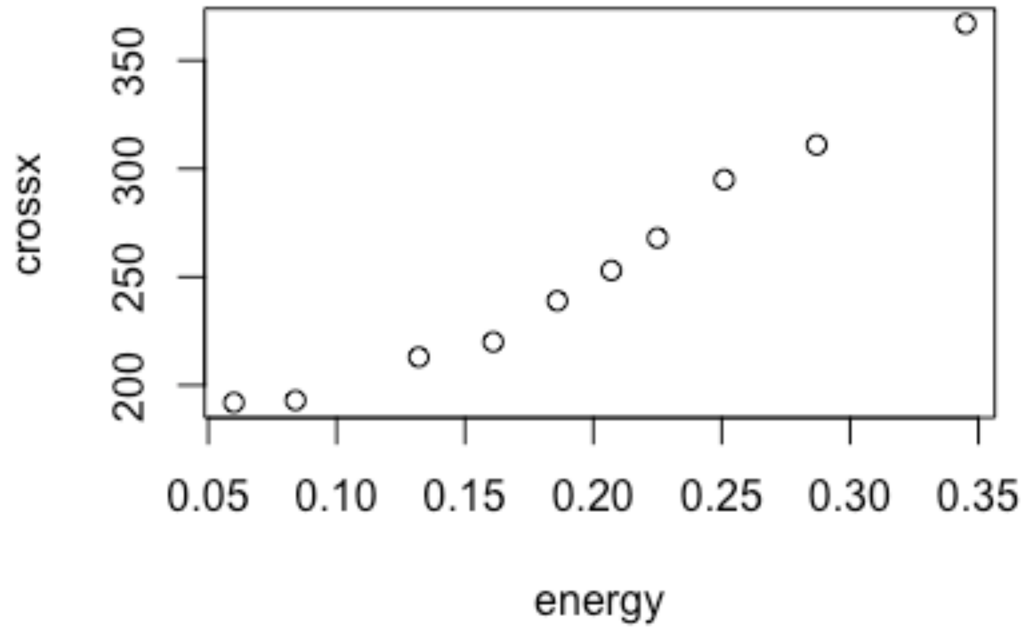


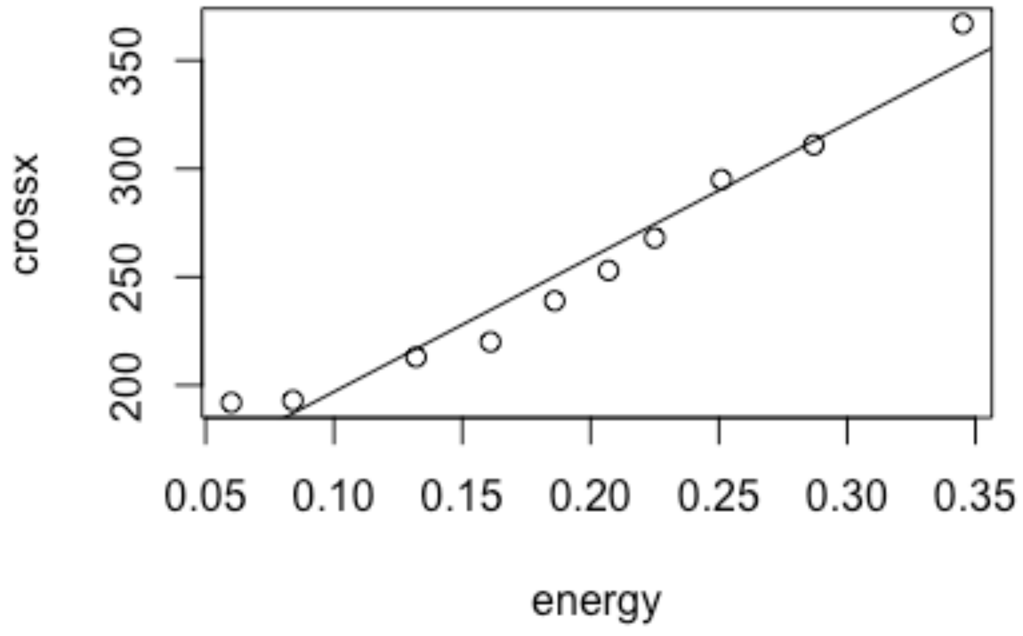
回归线

```
strongx
m <- lm(crossx ~ energy, data = strongx)

plot(crossx ~ energy, data = strongx)
abline(m)
```



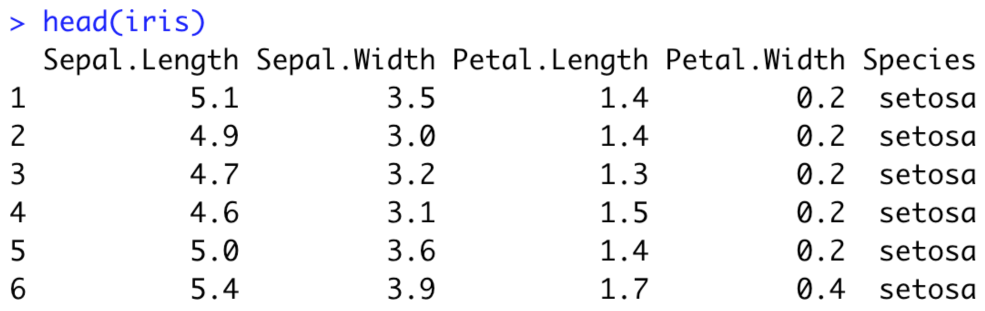
```
> strongx
   momentum energy crossx sd
1         4  0.345    367 17
2         6  0.287    311  9
3         8  0.251    295  9
4        10  0.225    268  7
5        12  0.207    253  7
6        15  0.186    239  6
7        20  0.161    220  6
8        30  0.132    213  6
9        75  0.084    193  5
10      150  0.060    192  5
```





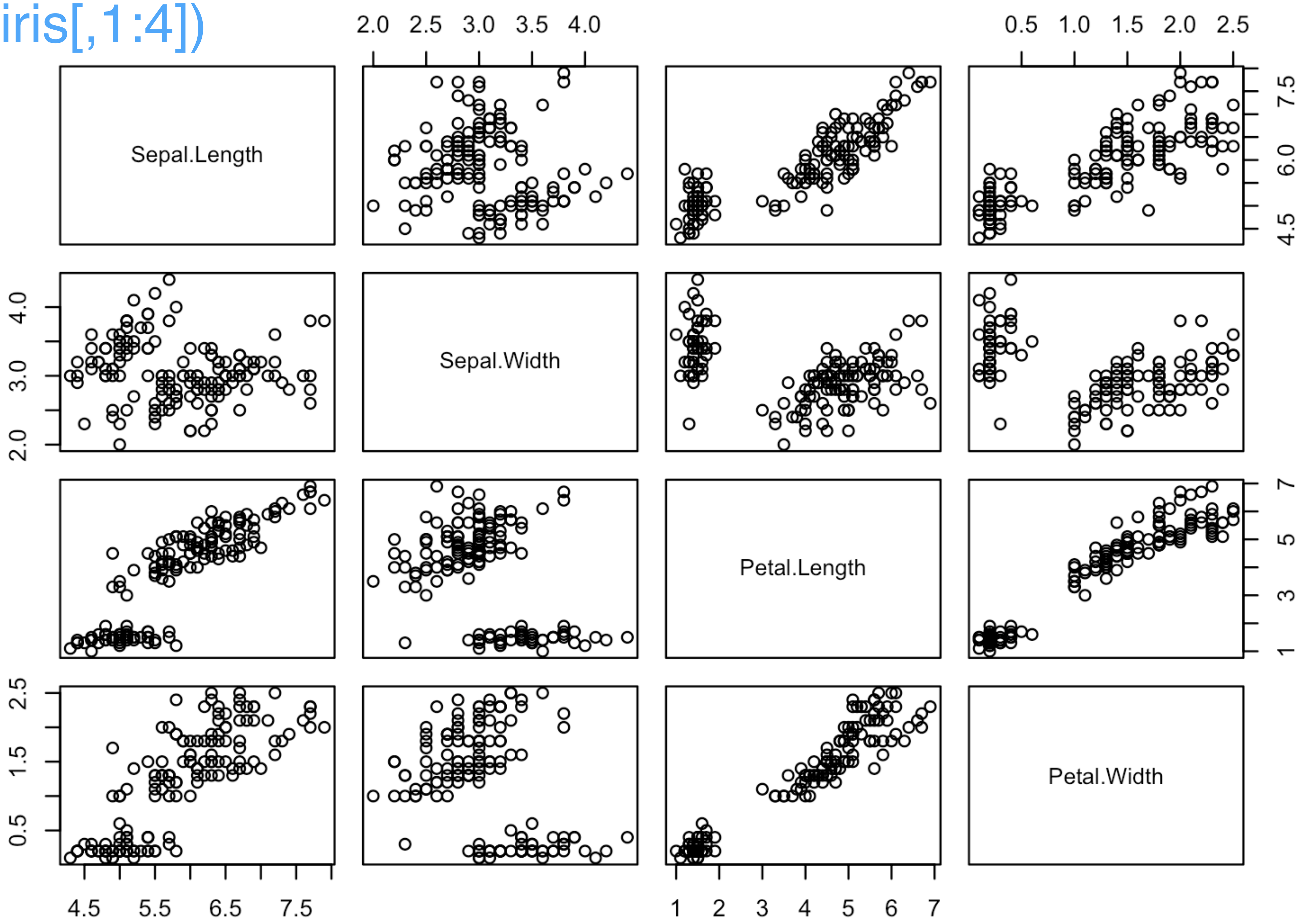
多变量散点图

head(iris)

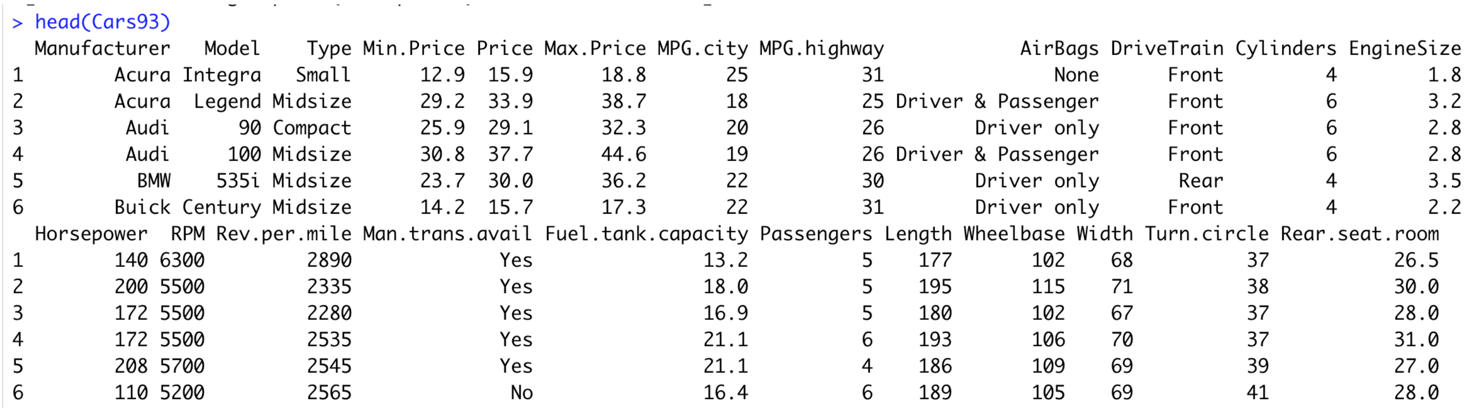


```
> head(iris)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
6          5.4         3.9          1.7         0.4  setosa
```

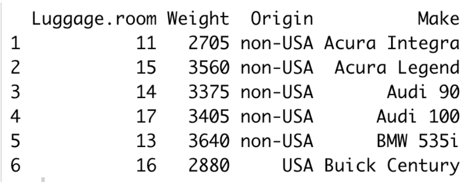
plot(iris[,1:4])



每个因子水平的散点图



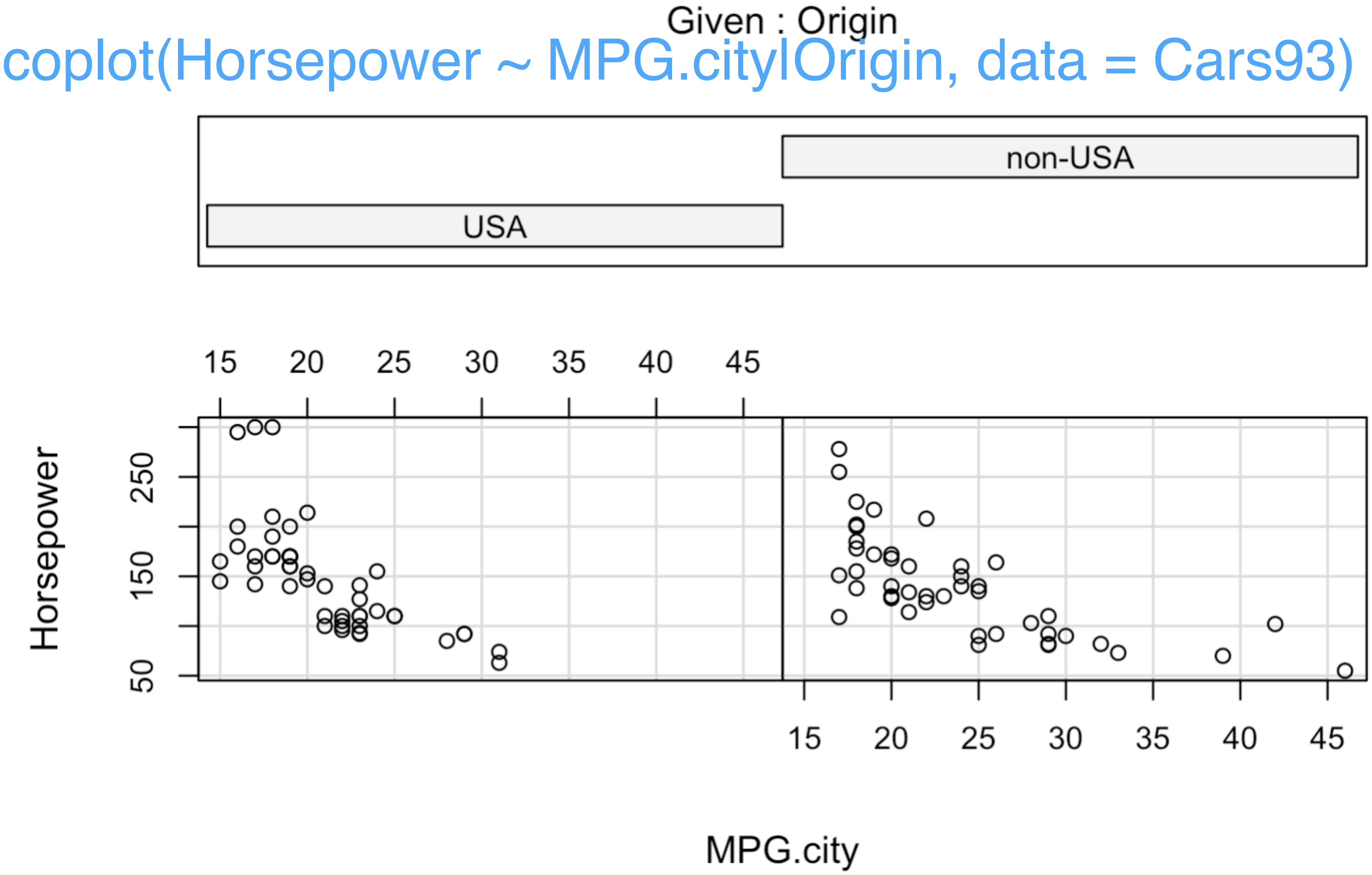
```
> head(Cars93)
  Manufacturer   Model    Type Min.Price Price Max.Price MPG.city MPG.highway            AirBags DriveTrain Cylinders EngineSize
1        Acura Integra   Small      12.9  15.9      18.8       25          31               None      Front         4        1.8
2        Acura  Legend Midsize      29.2  33.9      38.7       18          25 Driver & Passenger      Front         6        3.2
3         Audi      90 Compact      25.9  29.1      32.3       20          26        Driver only      Front         6        2.8
4         Audi     100 Midsize      30.8  37.7      44.6       19          26 Driver & Passenger      Front         6        2.8
5          BMW    535i Midsize      23.7  30.0      36.2       22          30        Driver only       Rear         4        3.5
6        Buick Century Midsize      14.2  15.7      17.3       22          31        Driver only      Front         4        2.2
  Horsepower  RPM Rev.per.mile Man.trans.avail Fuel.tank.capacity Passengers Length Wheelbase Width Turn.circle Rear.seat.room
1        140 6300         2890             Yes               13.2          5    177       102    68          37           26.5
2        200 5500         2335             Yes               18.0          5    195       115    71          38           30.0
3        172 5500         2280             Yes               16.9          5    180       102    67          37           28.0
4        172 5500         2535             Yes               21.1          6    193       106    70          37           31.0
5        208 5700         2545             Yes               21.1          4    186       109    69          39           27.0
6        110 5200         2565              No               16.4          6    189       105    69          41           28.0
```



```
  Luggage.room Weight  Origin           Make
1           11   2705 non-USA Acura Integra
2           15   3560 non-USA  Acura Legend
3           14   3375 non-USA        Audi 90
4           17   3405 non-USA       Audi 100
5           13   3640 non-USA      BMW 535i
6           16   2880     USA Buick Century
```

head(Cars93)

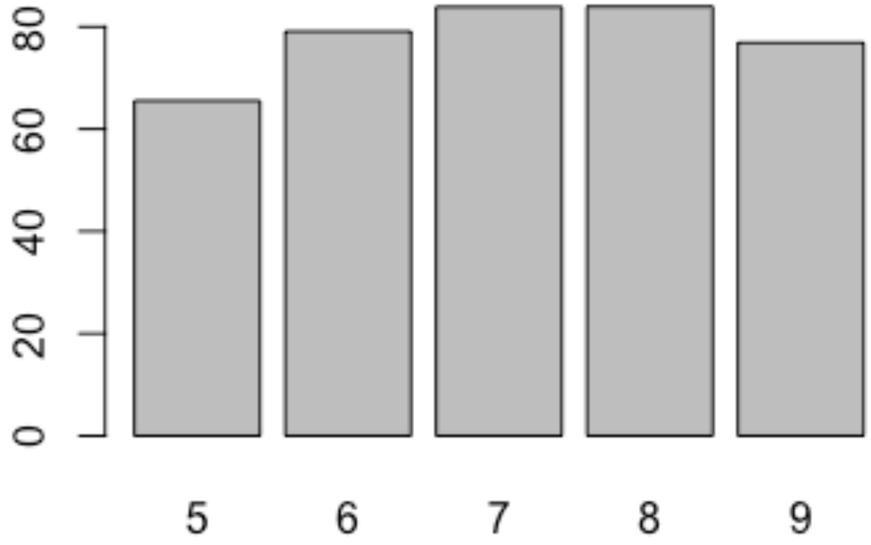
coplot(Horsepower ~ MPG.city|Origin, data = Cars93)



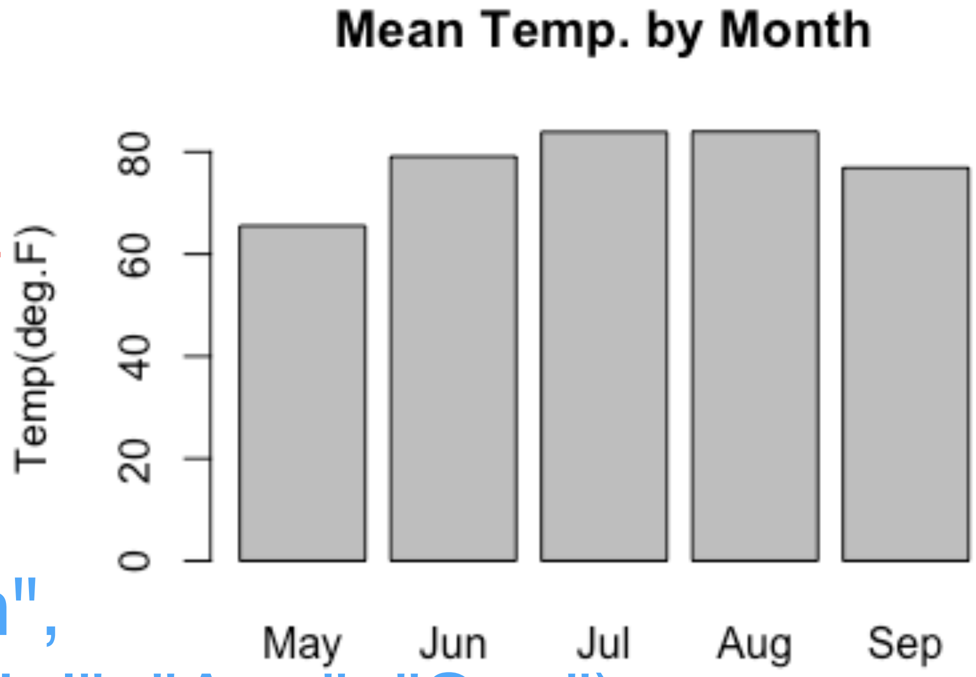
条形图

```
> head(airquality)
  Ozone Solar.R Wind Temp Month Day
1    41     190  7.4   67     5   1
2    36     118  8.0   72     5   2
3    12     149 12.6   74     5   3
4    18     313 11.5   62     5   4
5    NA      NA 14.3   56     5   5
6    28      NA 14.9   66     5   6
```

```
barplot(height)
```



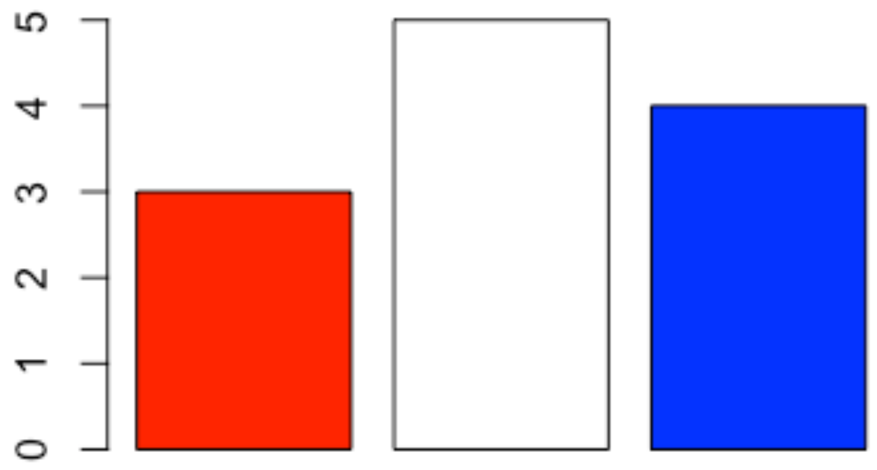
```
> height <- tapply(airquality$Temp, airquality$Month, mean)
> height
       5        6        7        8        9 
65.54839 79.10000 83.90323 83.96774 76.90000 
```



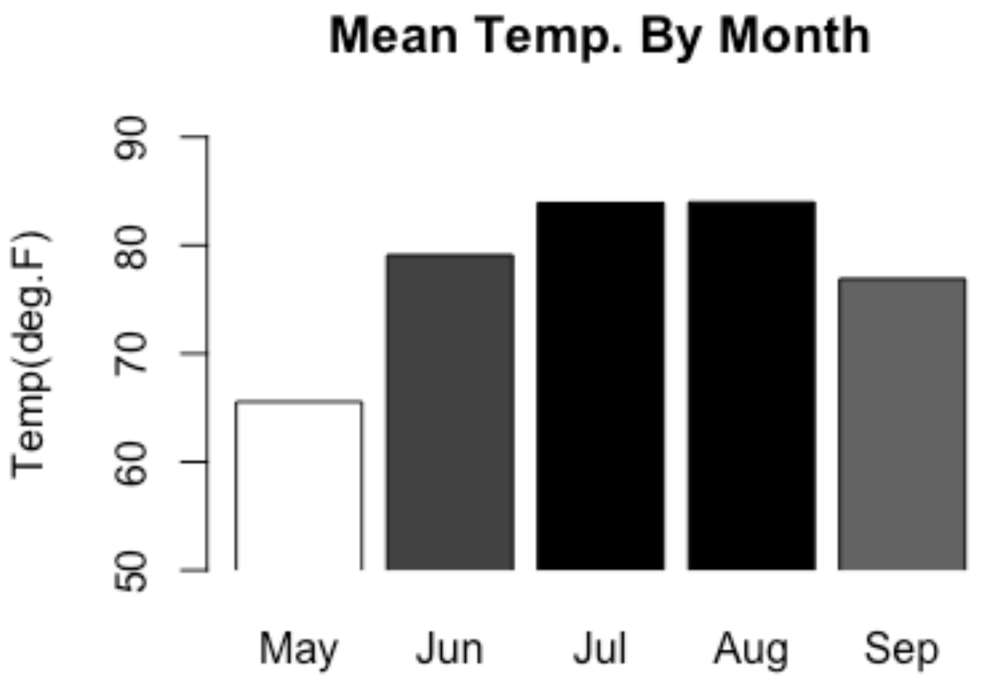
```
barplot(height,
        main = "Mean Temp. by Month",
        names.arg = c("May", "Jun", "Jul", "Aug", "Sep"),
        ylab = "Temp(deg.F)")
```

条形图上色

```
barplot(c(3,5,4),col = c("red","white","blue"))
```

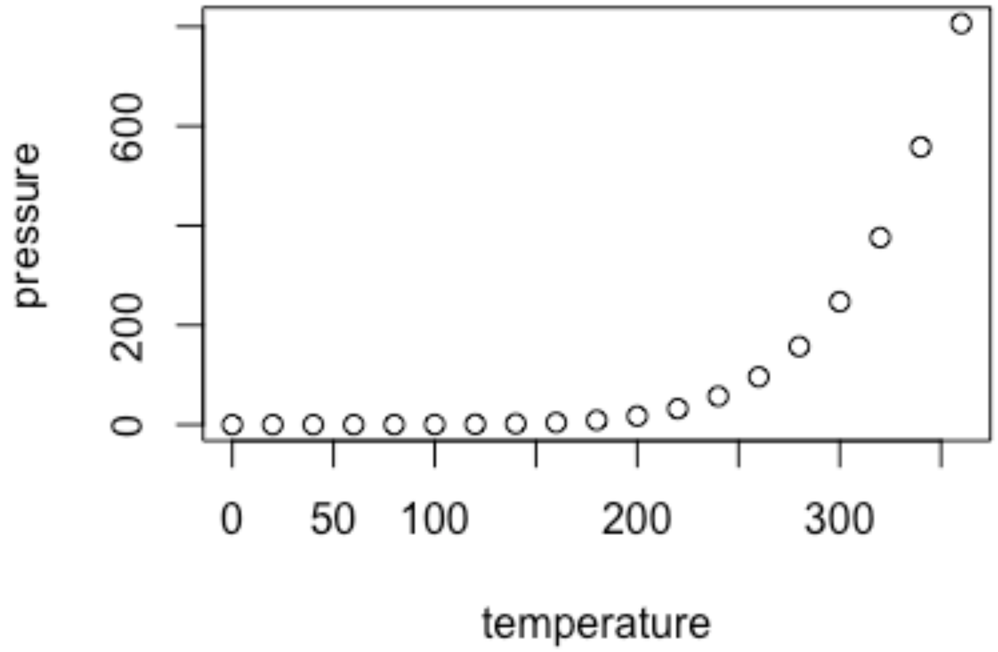


```
rel.hts <- (height - min(height)) / (max(height) - min(height))
grays <- gray(1 - rel.hts)
barplot(height,col = grays,ylim = c(50, 90), xpd = FALSE,main = "Mean
Temp. By Month",names.arg = c("May", "Jun", "Jul", "Aug", "Sep"),ylab =
"Temp(deg.F)")
```

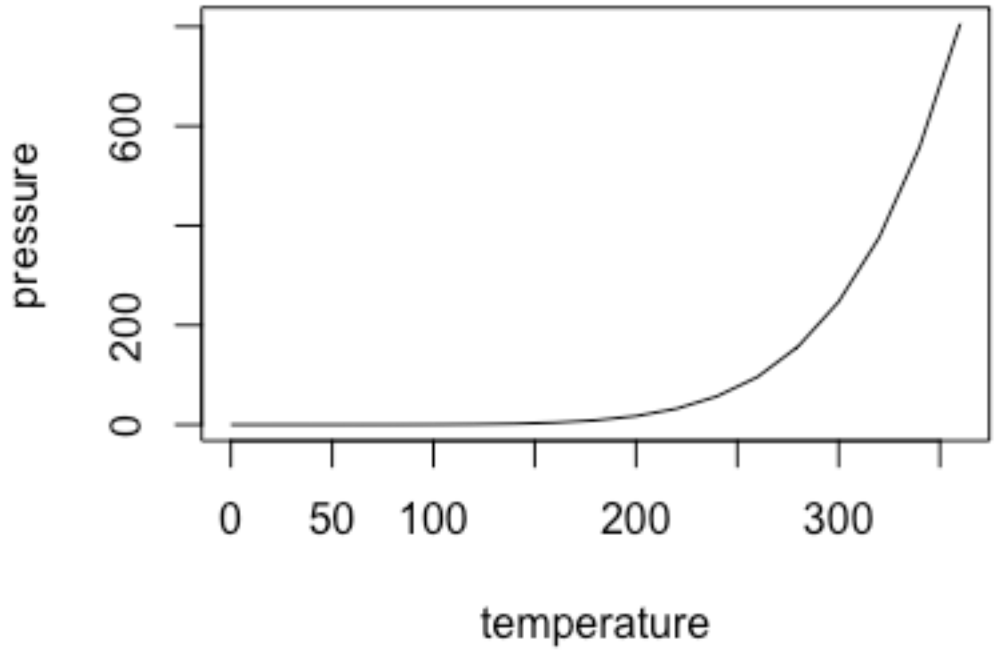


线图

plot(pressure)

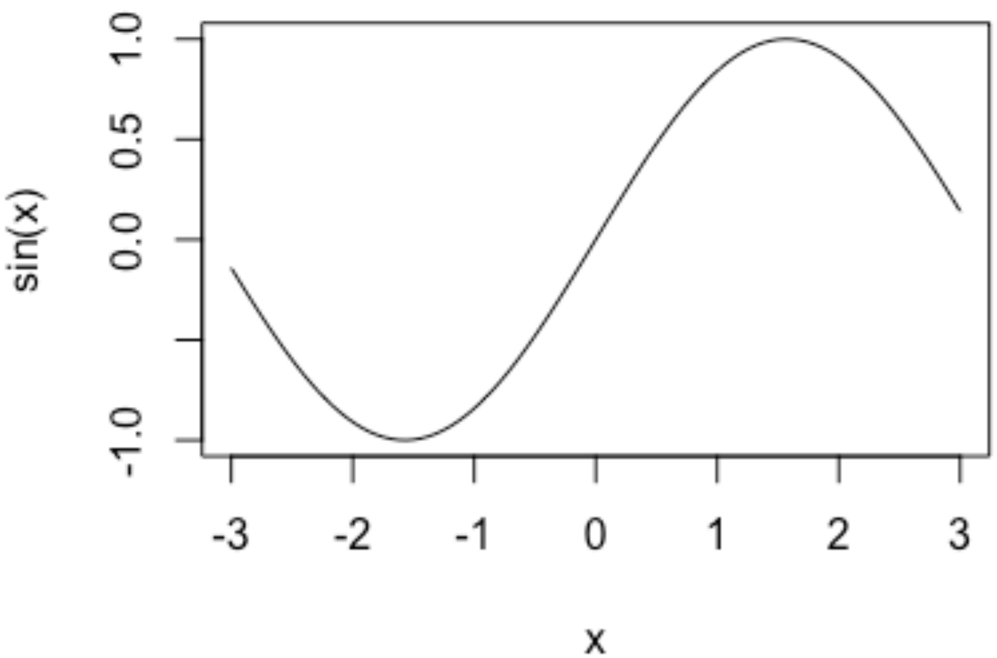


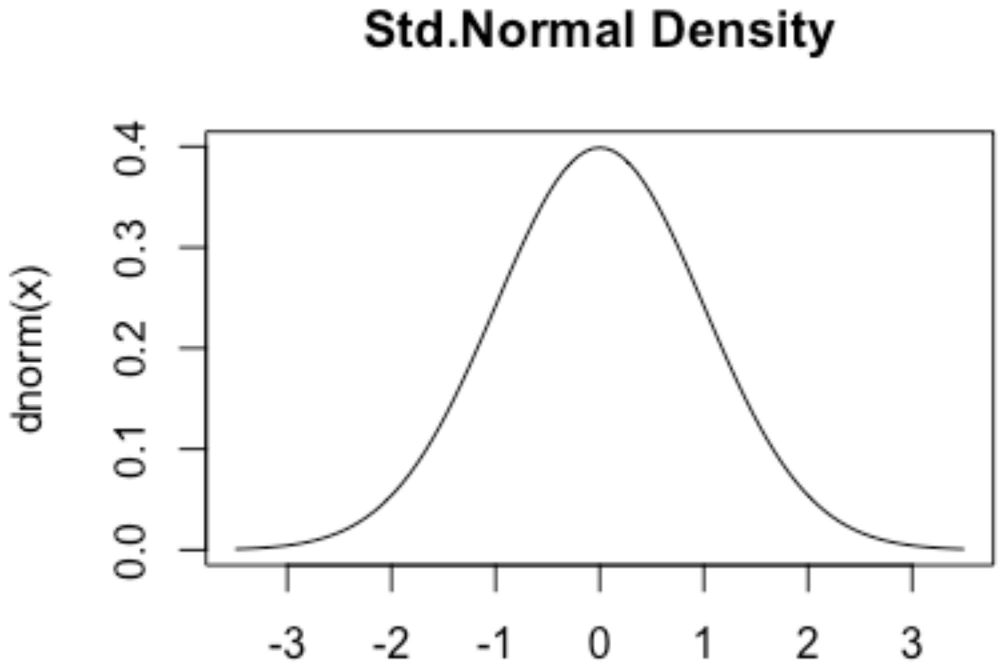
plot(pressure, type = "l")



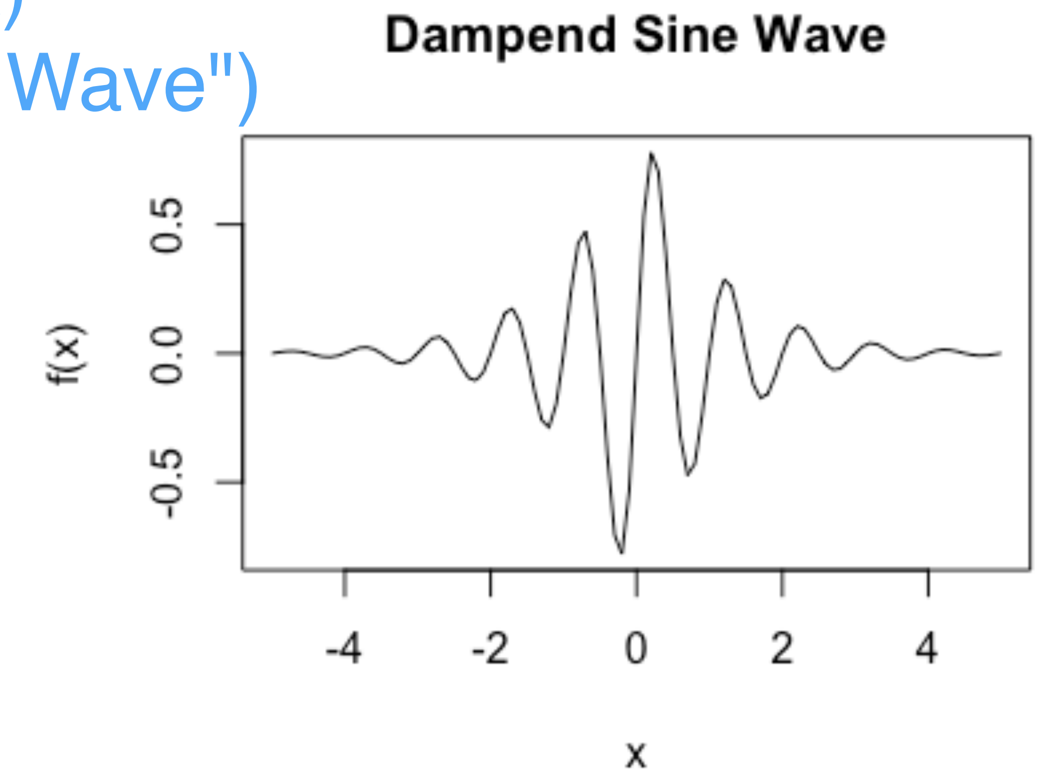
绘制函数

```
curve(sin, -3, 3)        curve(dnorm, -3.5, +3.5, main="Std.Normal Density"
```





```
f <- function(x) exp(-abs(x)) * sin(2*pi*x)
curve(f, -5, +5, main = "Dampend Sine Wave")
```



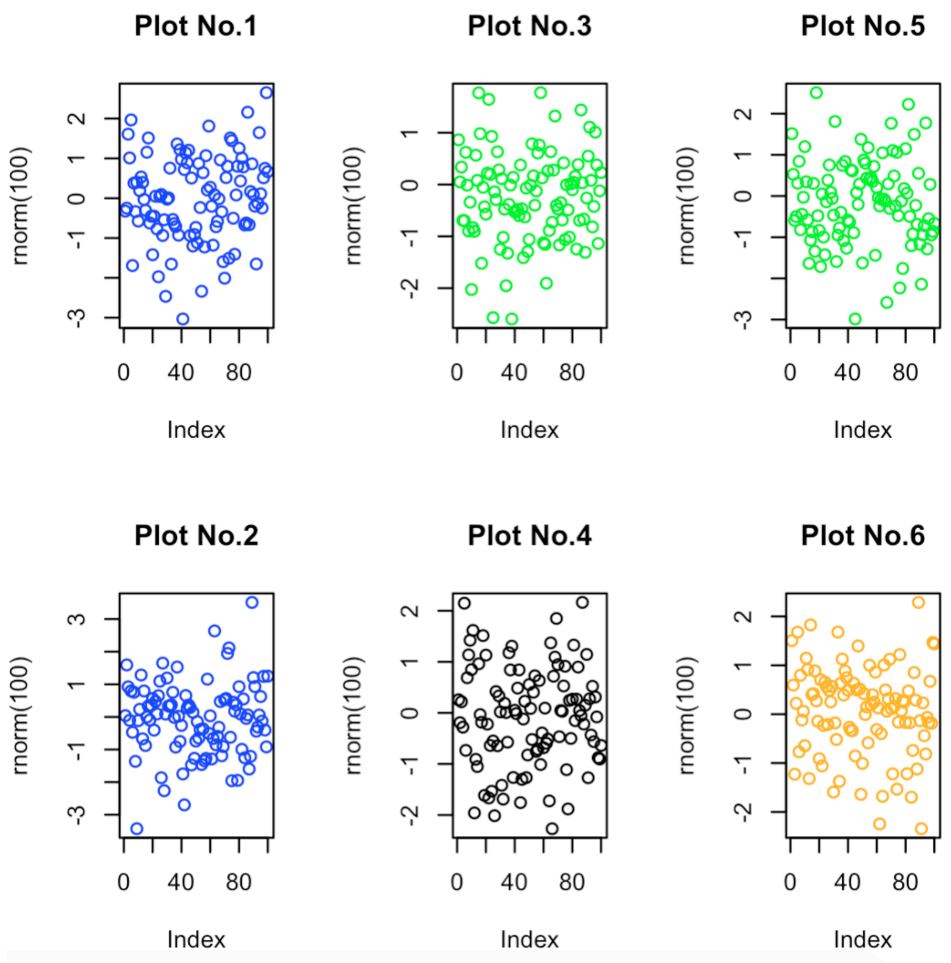
图形控制

R Graphs Cookbook

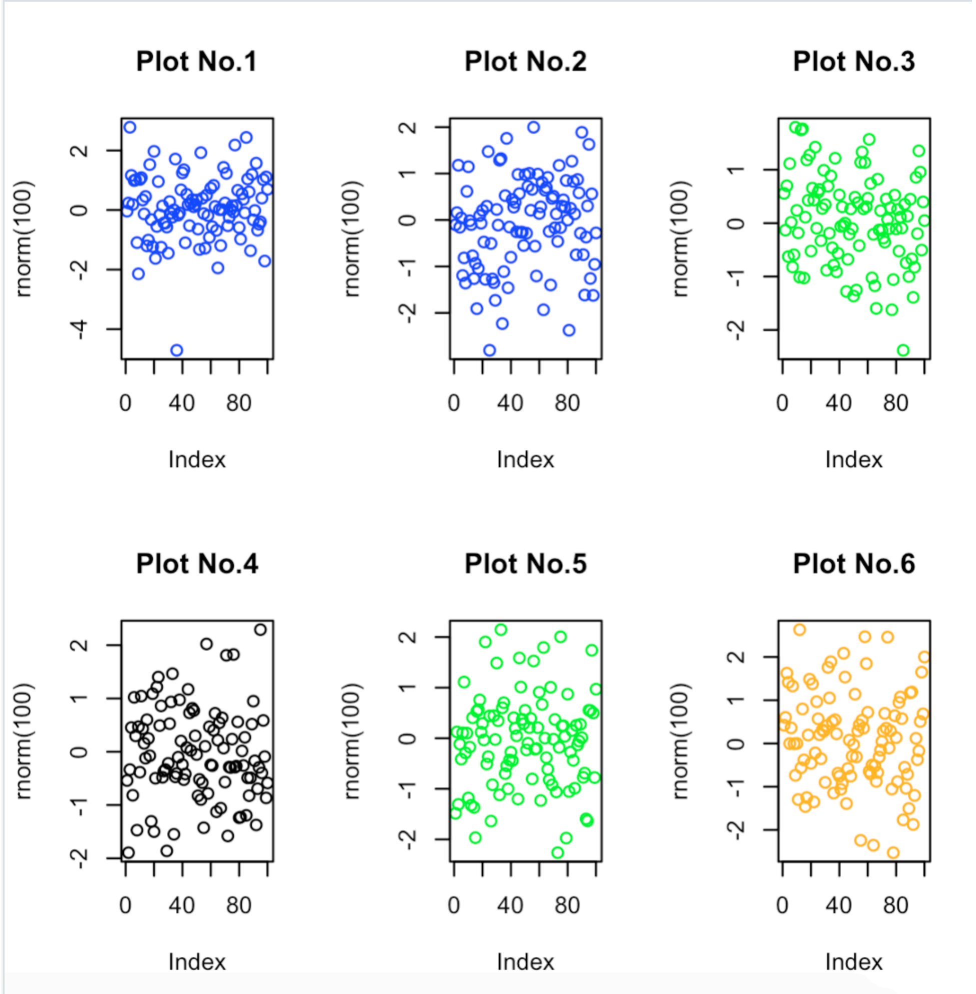
CH1,CH4, CH5@ R Graphs Cookbook

一页显示多个图形

```
par(mfrow=c(2,3))
plot(rnorm(100),col="blue",main="Plot No.1")
plot(rnorm(100),col="blue",main="Plot No.2")
plot(rnorm(100),col="green",main="Plot No.3")
plot(rnorm(100),col="black",main="Plot No.4")
plot(rnorm(100),col="green",main="Plot No.5")
plot(rnorm(100),col="orange",main="Plot No.6")
```

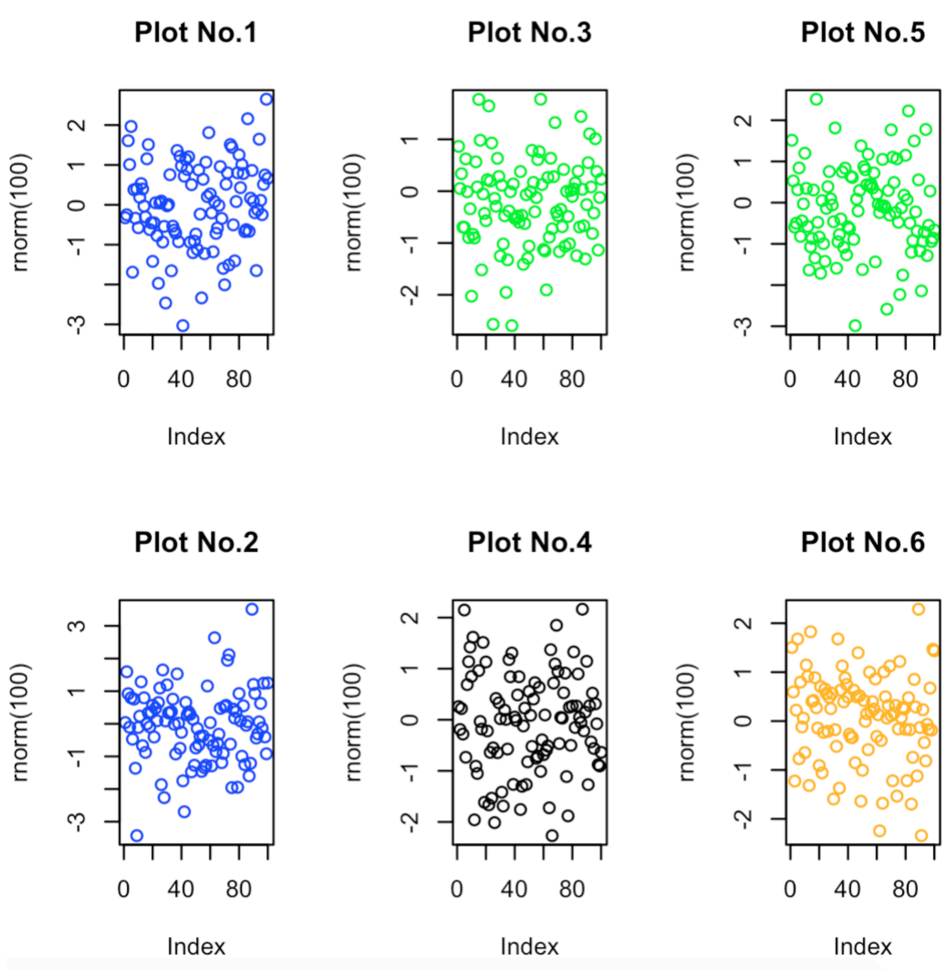


```
par(mfcol=c(2,3))
plot(rnorm(100),col="blue",main="Plot No.1")
plot(rnorm(100),col="blue",main="Plot No.2")
plot(rnorm(100),col="green",main="Plot No.3")
plot(rnorm(100),col="black",main="Plot No.4")
plot(rnorm(100),col="green",main="Plot No.5")
plot(rnorm(100),col="orange",main="Plot No.6")
```

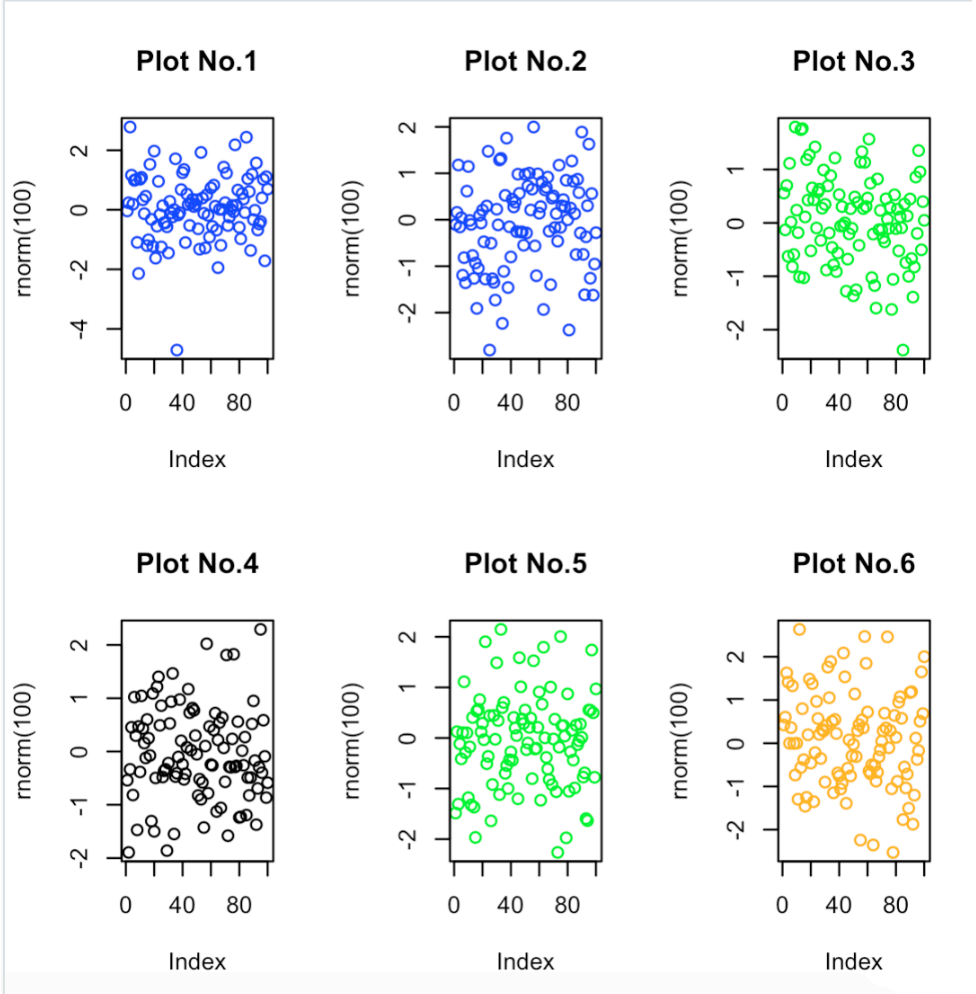


一页显示多个图形

```
par(mfrow=c(2,3))
plot(rnorm(100),col="blue",main="Plot No.1")
plot(rnorm(100),col="blue",main="Plot No.2")
plot(rnorm(100),col="green",main="Plot No.3")
plot(rnorm(100),col="black",main="Plot No.4")
plot(rnorm(100),col="green",main="Plot No.5")
plot(rnorm(100),col="orange",main="Plot No.6")
```

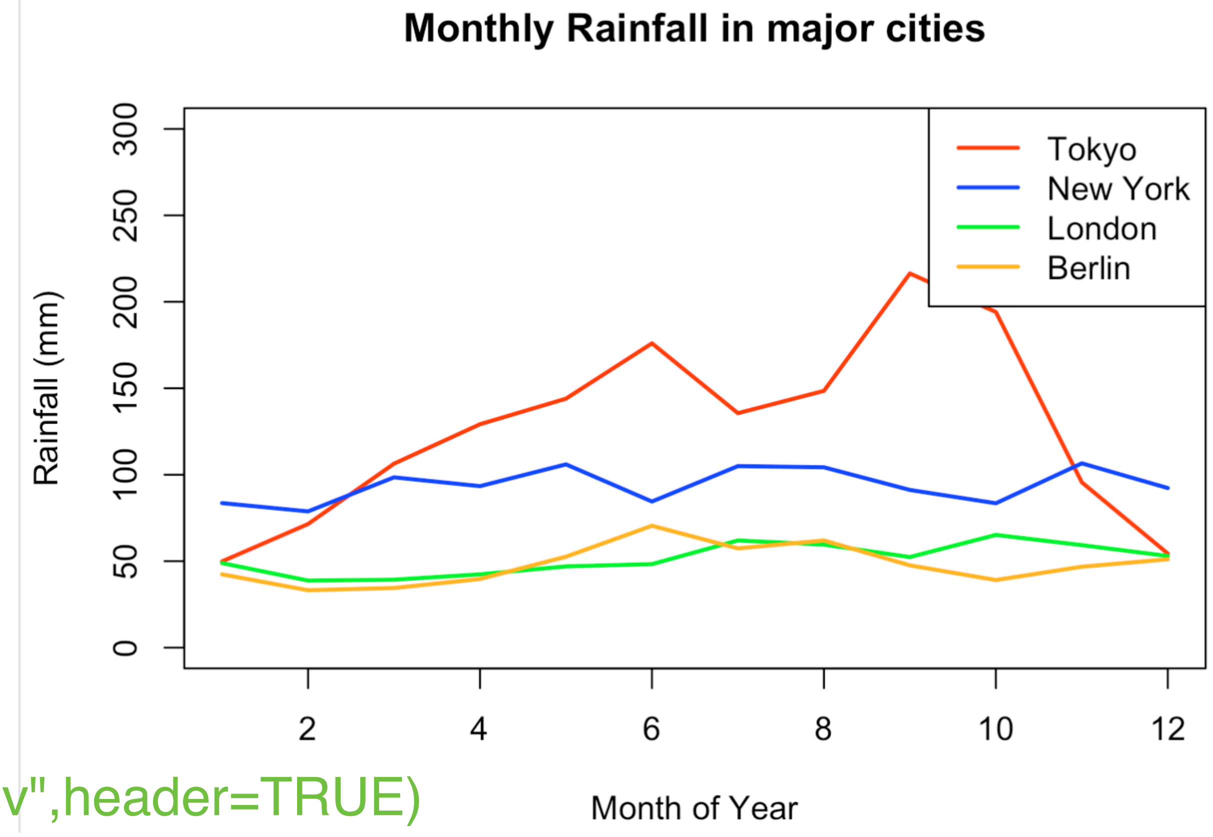


```
par(mfcol=c(2,3))
plot(rnorm(100),col="blue",main="Plot No.1")
plot(rnorm(100),col="blue",main="Plot No.2")
plot(rnorm(100),col="green",main="Plot No.3")
plot(rnorm(100),col="black",main="Plot No.4")
plot(rnorm(100),col="green",main="Plot No.5")
plot(rnorm(100),col="orange",main="Plot No.6")
```



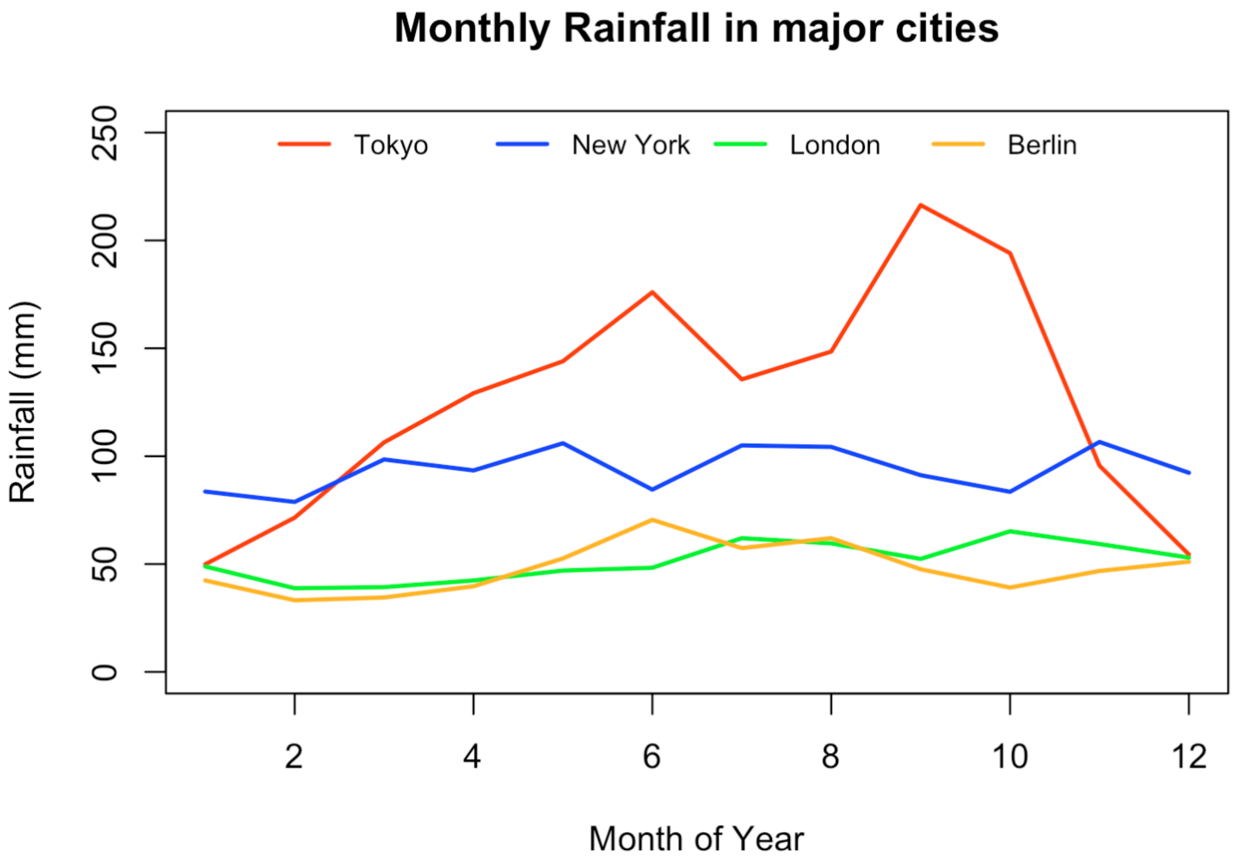
图例 – 基本例子

```
plot(rain$Tokyo,type="l",col="red",
     ylim=c(0,300),
     main="Monthly Rainfall in major cities",
     xlab="Month of Year",ylab="Rainfall (mm)",lwd=2)
lines(rain$NewYork,type="l",col="blue",lwd=2)
lines(rain$London,type="l",col="green",lwd=2)
lines(rain$Berlin,type="l",col="orange",lwd=2)

legend("topright",
       legend=c("Tokyo","New York","London","Berlin"),
       col=c("red","blue","green","orange"),
       lty=1,lwd=2)
```



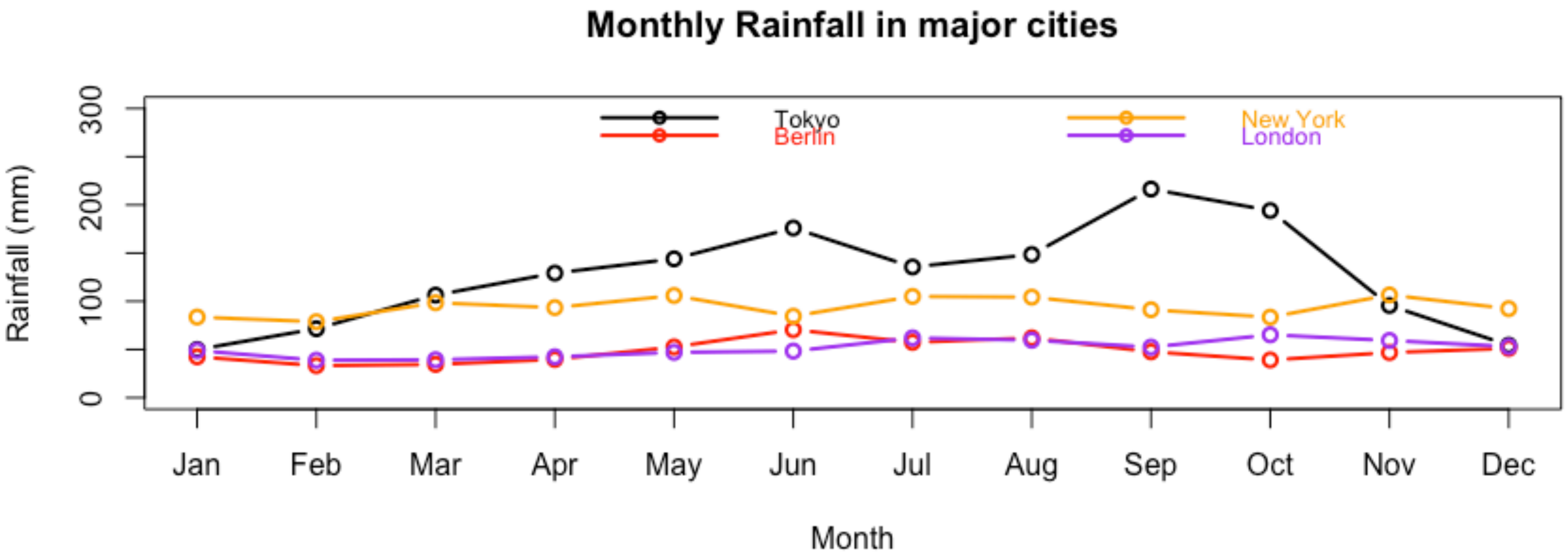
```
rain<-read.csv("cityrain.csv",header=TRUE)
```

```
plot(rain$Tokyo,type="l",col="red",
     ylim=c(0,250),
     main="Monthly Rainfall in major cities",
     xlab="Month of Year",ylab="Rainfall (mm)", lwd=2)
lines(rain$NewYork,type="l",col="blue",lwd=2)
lines(rain$London,type="l",col="green",lwd=2)
lines(rain$Berlin,type="l",col="orange",lwd=2)

legend("top",
       legend=c("Tokyo","New York","London","Berlin"),
       ncol=4,cex=0.8,bty="n",
       col=c("red","blue","green","orange"),
       lty=1,lwd=2)
```



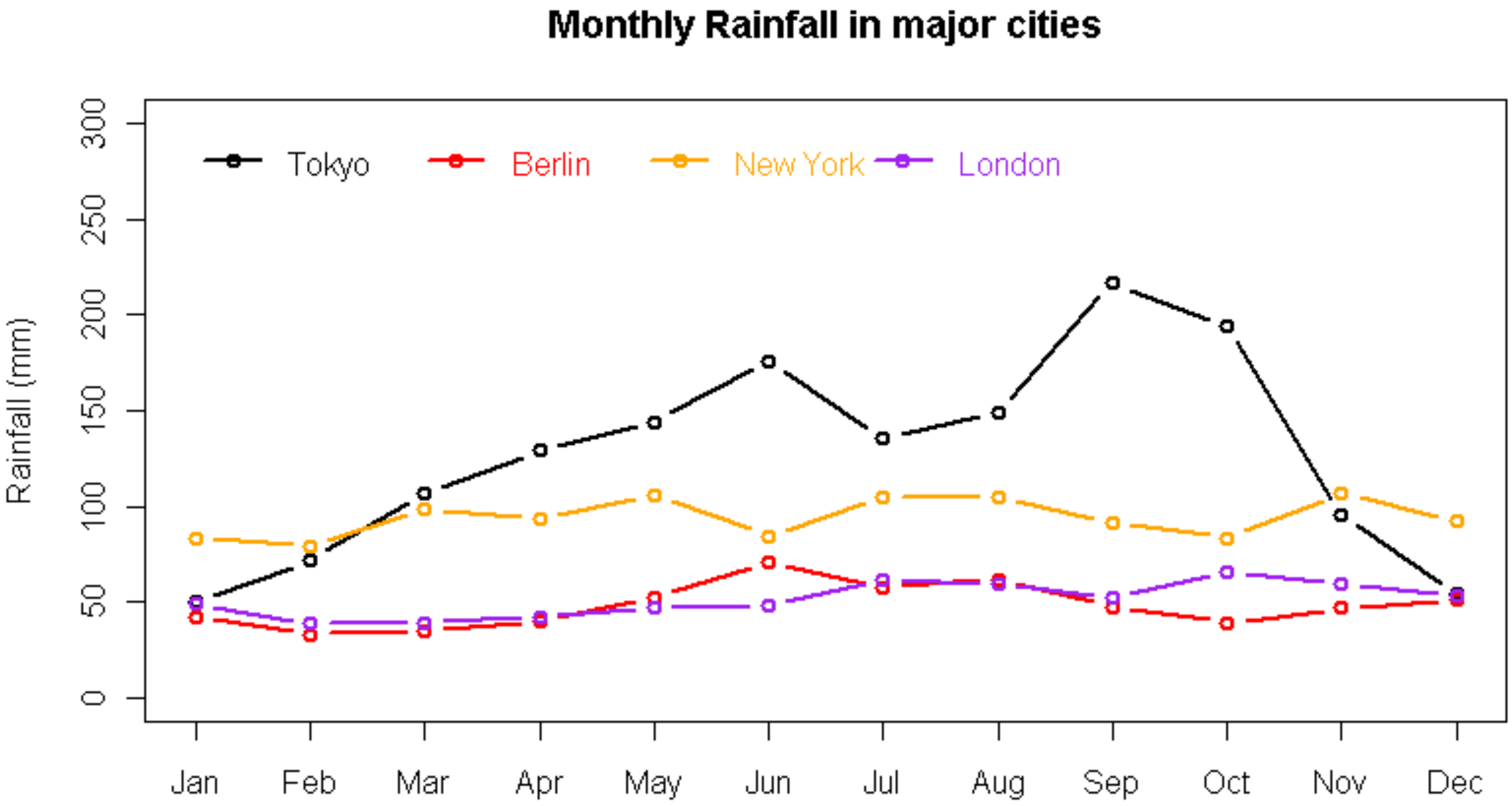
图例 – 使用坐标定位

```
rain<-read.csv("cityrain.csv")
plot(rain$Tokyo,type="b",lwd=2, xaxt="n",ylim=c(0,300),col="black",xlab="Month",ylab="Rainfall
(mm)",main="Monthly Rainfall in major cities")
axis(1,at=1:length(rain$Month),labels=rain$Month)
lines(rain$Berlin,col="red",type="b",lwd=2)
lines(rain$NewYork,col="orange",type="b",lwd=2)
lines(rain$London,col="purple",type="b",lwd=2)

legend("topright",legend=c("Tokyo","Berlin","New York”, “London”), lty=1, lwd=2, pch=21,
col=c(“black","red","orange","purple"), ncol=2, bty="n",cex=0.8, text.col=c(“black","red","orange","purple"),
inset=0.01)
```



图例 – 使用坐标定位

```
legend(1,300,legend=c("Tokyo","Berlin","New York","London"),
      lty=1,lwd=2,pch=21,col=c("black","red","orange","purple"),
      horiz=TRUE,bty="n",bg="yellow",cex=1,
      text.col=c("black","red","orange","purple"))
```



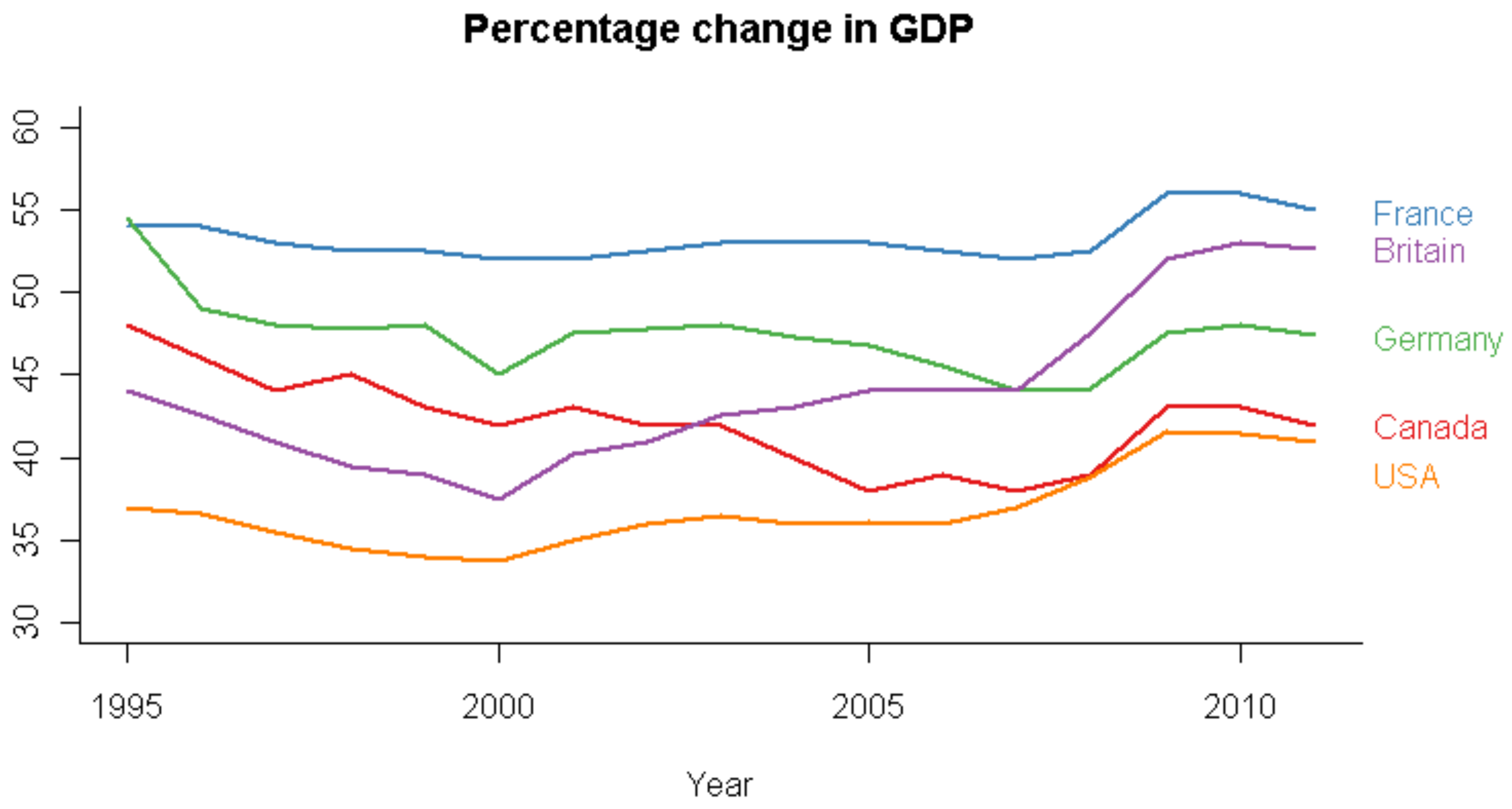
图例 – 边界标记

```
gdp<-read.table("gdp_long.txt",header=T)

library(RColorBrewer)
pal<-brewer.pal(5,"Set1")

par(mar=par()$mar+c(0,0,0,2),bty="l")

plot(Canada~Year,data=gdp,type="l",lwd=2,lty=1,ylim=c(30,60),col=pal[1],main="Percentage change in GDP",ylab="")
mtext(side=4,at=gdp$Canada[length(gdp$Canada)],text="Canada",col=pal[1],line=0.3,las=2)

lines(gdp$France~gdp$Year,col=pal[2],lwd=2)
mtext(side=4,at=gdp$France[length(gdp$France)],text="France",col=pal[2],line=0.3,las=2)

lines(gdp$Germany~gdp$Year,col=pal[3],lwd=2)
mtext(side=4,at=gdp$Germany[length(gdp$Germany)],text="Germany",col=pal[3],line=0.3,las=2)

lines(gdp$Britain~gdp$Year,col=pal[4],lwd=2)
mtext(side=4,at=gdp$Britain[length(gdp$Britain)],text="Britain",col=pal[4],line=0.3,las=2)

lines(gdp$USA~gdp$Year,col=pal[5],lwd=2)
mtext(side=4,at=gdp$USA[length(gdp$USA)]-2,text="USA",col=pal[5],line=0.3,las=2)
```

- o
- l
- 7
- c
- u
-]

- side
- 1,2,3,4

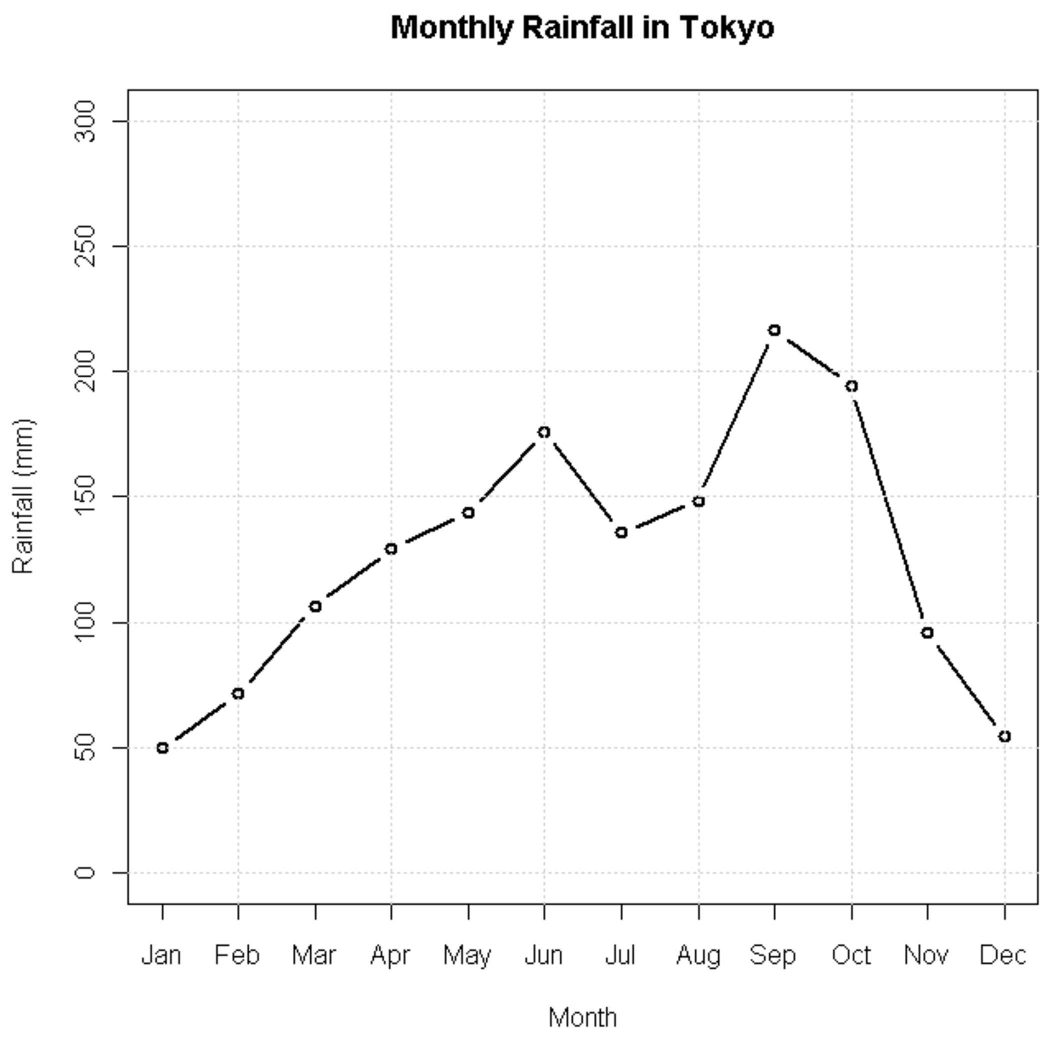
图例 – 边界标记


Percentage change in GDP
60
55
50
45
40
35
30
1995
2000
2005
2010
Year
France
Britain
Germany
Canada
USA

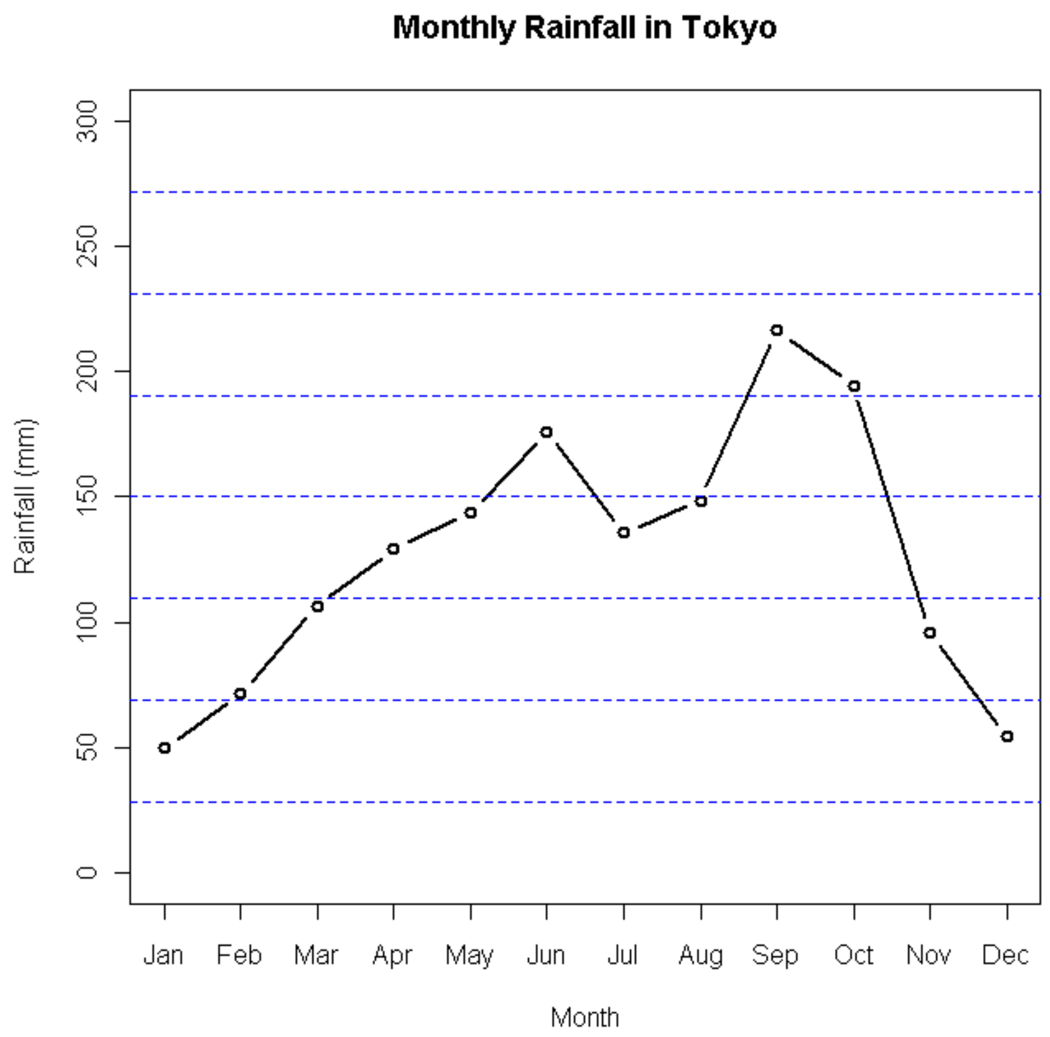
折线图 – 网格图

```
rain<-read.csv(“cityrain.csv")
plot(rain$Tokyo,type=“b",lwd=2, xaxt=“n”,ylim=c(0,300),col=“black", xlab=“Month",
      ylab="Rainfall (mm)",main="Monthly Rainfall in Tokyo")
axis(1,at=1:length(rain$Month),labels=rain$Month)
```

```
grid()
```



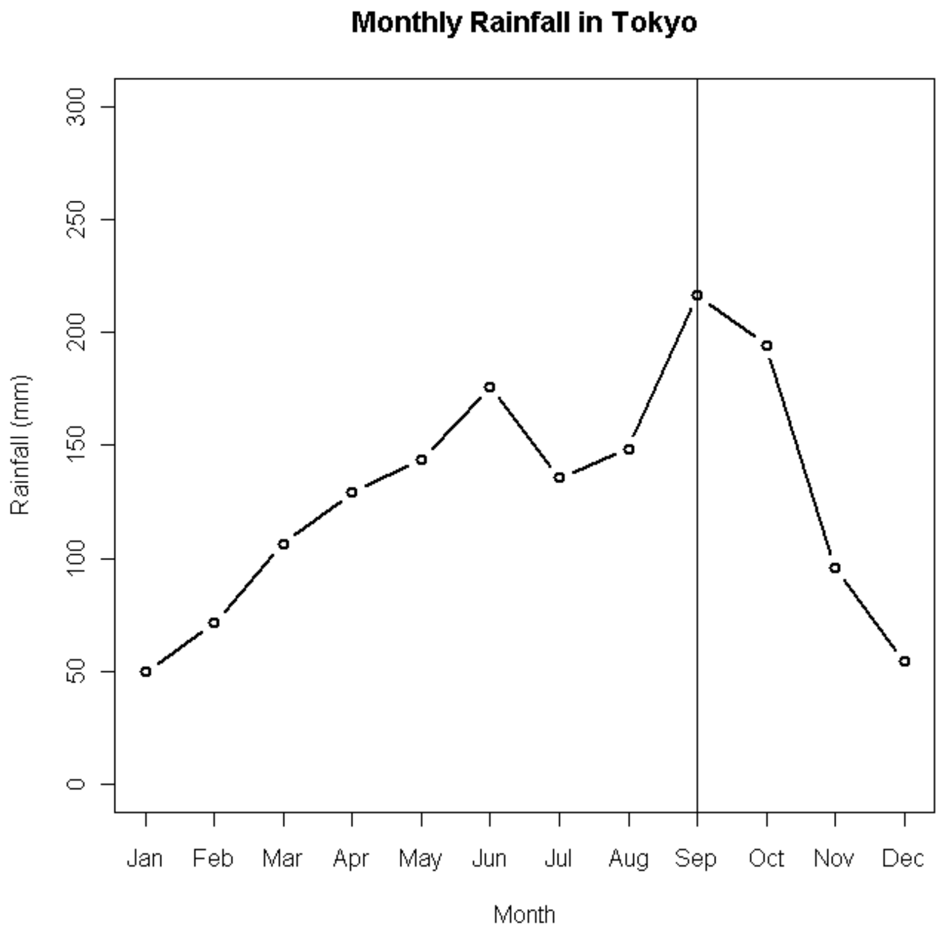
```
grid(nx=NA, ny=8, lwd=1,lty=2,col="blue")
```



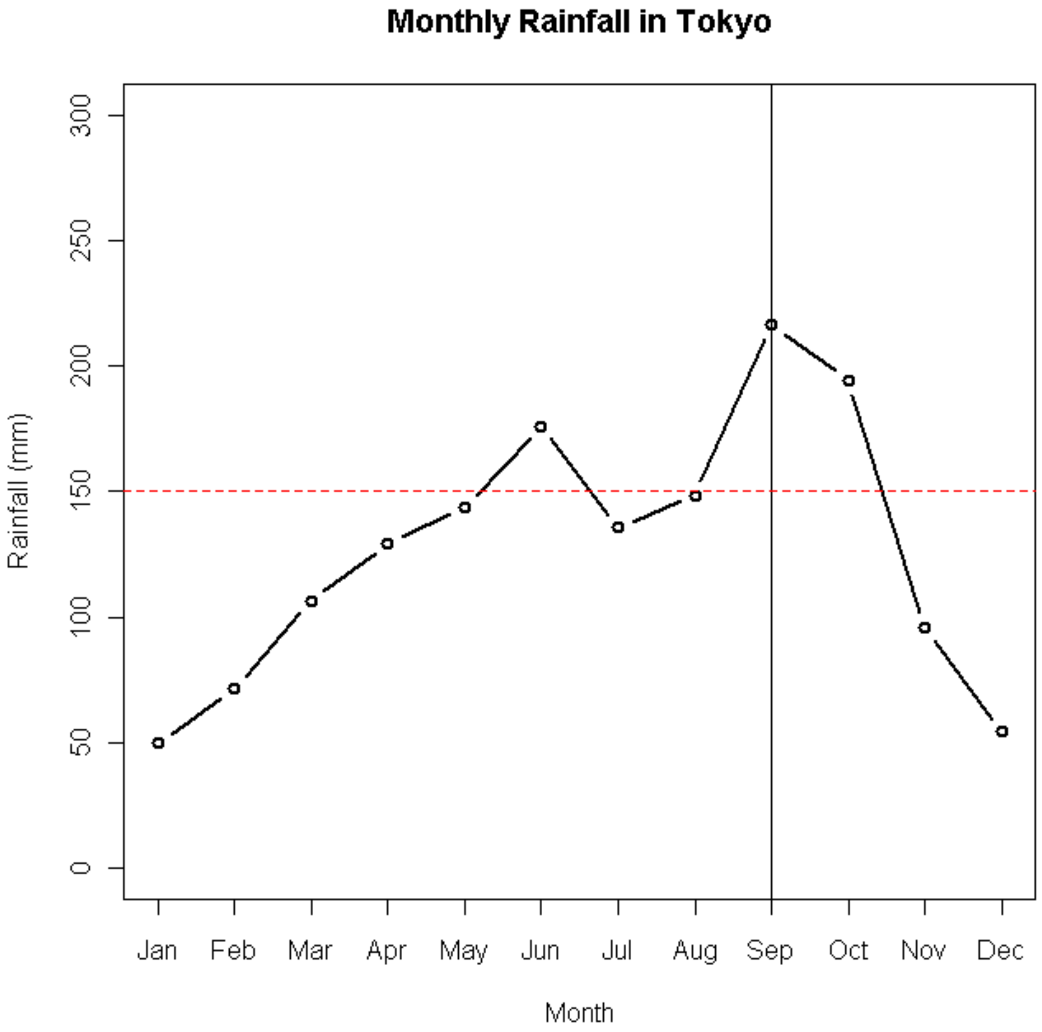
折线图 – 特殊线

```
rain<-read.csv(“cityrain.csv")
plot(rain$Tokyo,type=“b",lwd=2, xaxt=“n”,ylim=c(0,300),col=“black", xlab=“Month",
      ylab="Rainfall (mm)",main="Monthly Rainfall in Tokyo")
axis(1,at=1:length(rain$Month),labels=rain$Month)
```

```
abline(v=9)
```

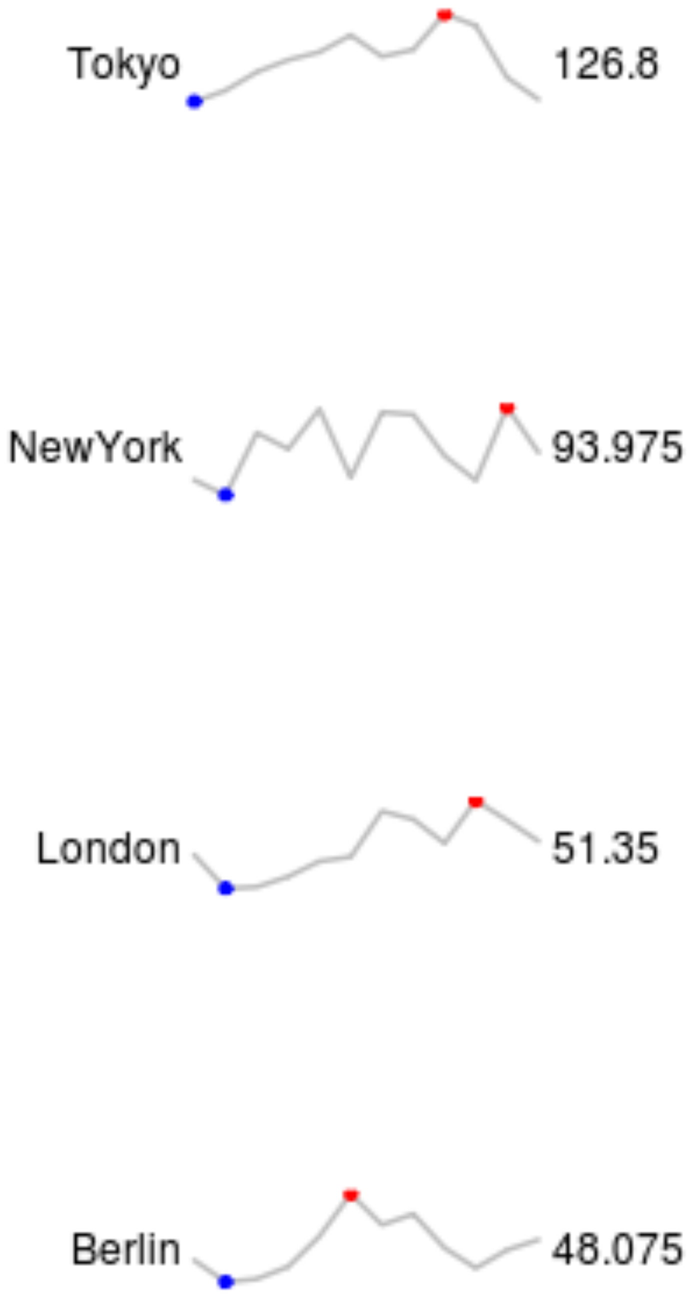


```
abline(h=150,col="red",lty=2)
```



折线图 – 波形线

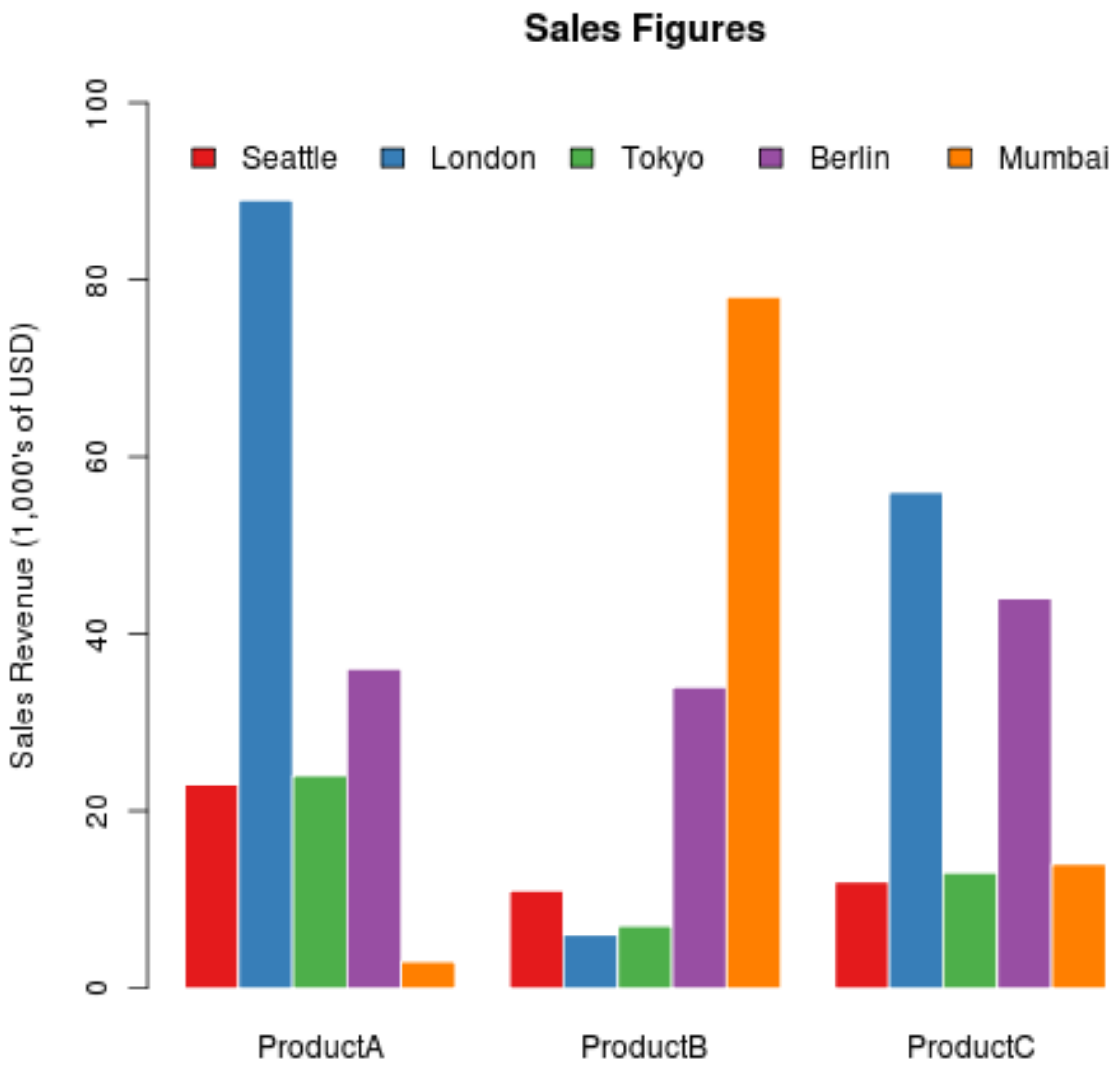
```
rain <- read.csv("cityrain.csv")
par(mfrow=c(4,1),mar=c(5,7,4,2),omi=c(0.2,2,0.2,2))
for(i in 2:5)
{

plot(rain[,i],ann=FALSE,axes=FALSE,type="l",col="g
ray",lwd=2)

mtext(side=2,at=mean(rain[,i]),names(rain[i]),las=2,c
ol="black")

mtext(side=4,at=mean(rain[,i]),mean(rain[i]),las=2,c
ol="black")

points(which.min(rain[,i]),min(rain[,i]),pch=19,col="bl
ue")

points(which.max(rain[,i]),max(rain[,i]),pch=19,col="r
ed")
}
```

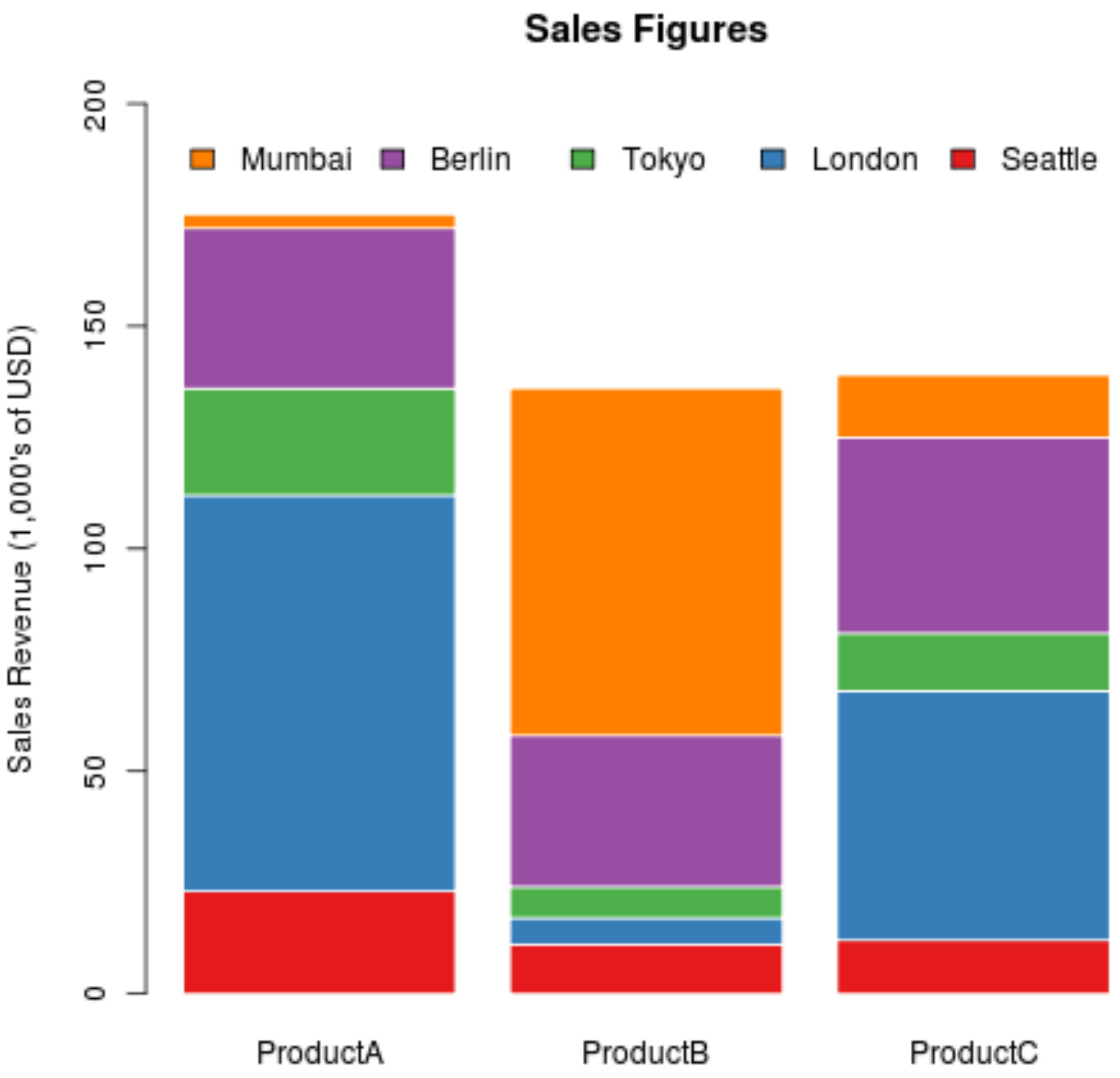


条形图 – 堆积

```
citysales<-read.csv("citysales.csv")

barplot(as.matrix(citysales[,2:4]), beside=TRUE,egend.text=citysales$City,
        args.legend=list(bty="n",horiz=TRUE),col=brewer.pal(5,"Set1"),
        border="white",ylim=c(0,100),ylab="Sales Revenue (1,000's of USD)",main="Sales Figures")

box(bty="l")
```

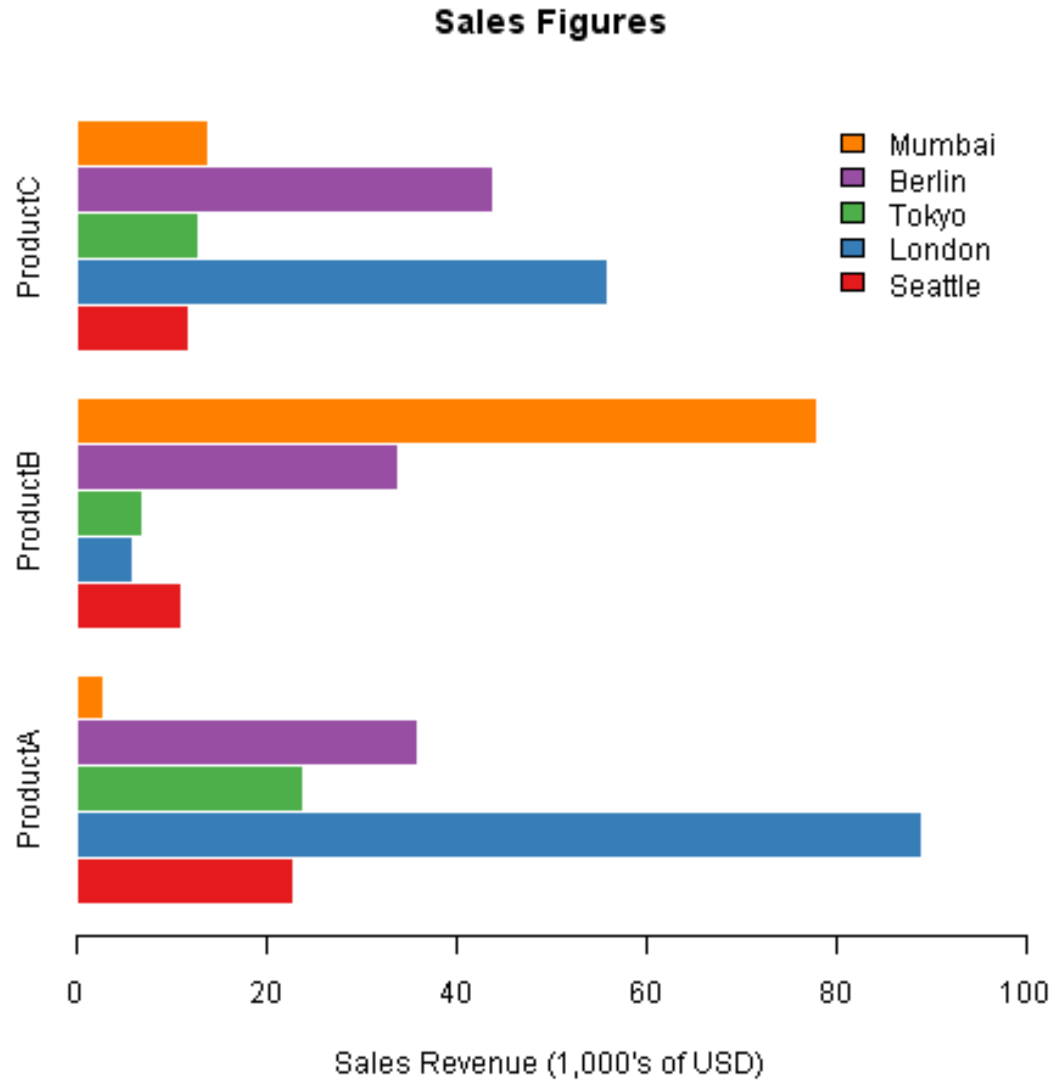
矩阵

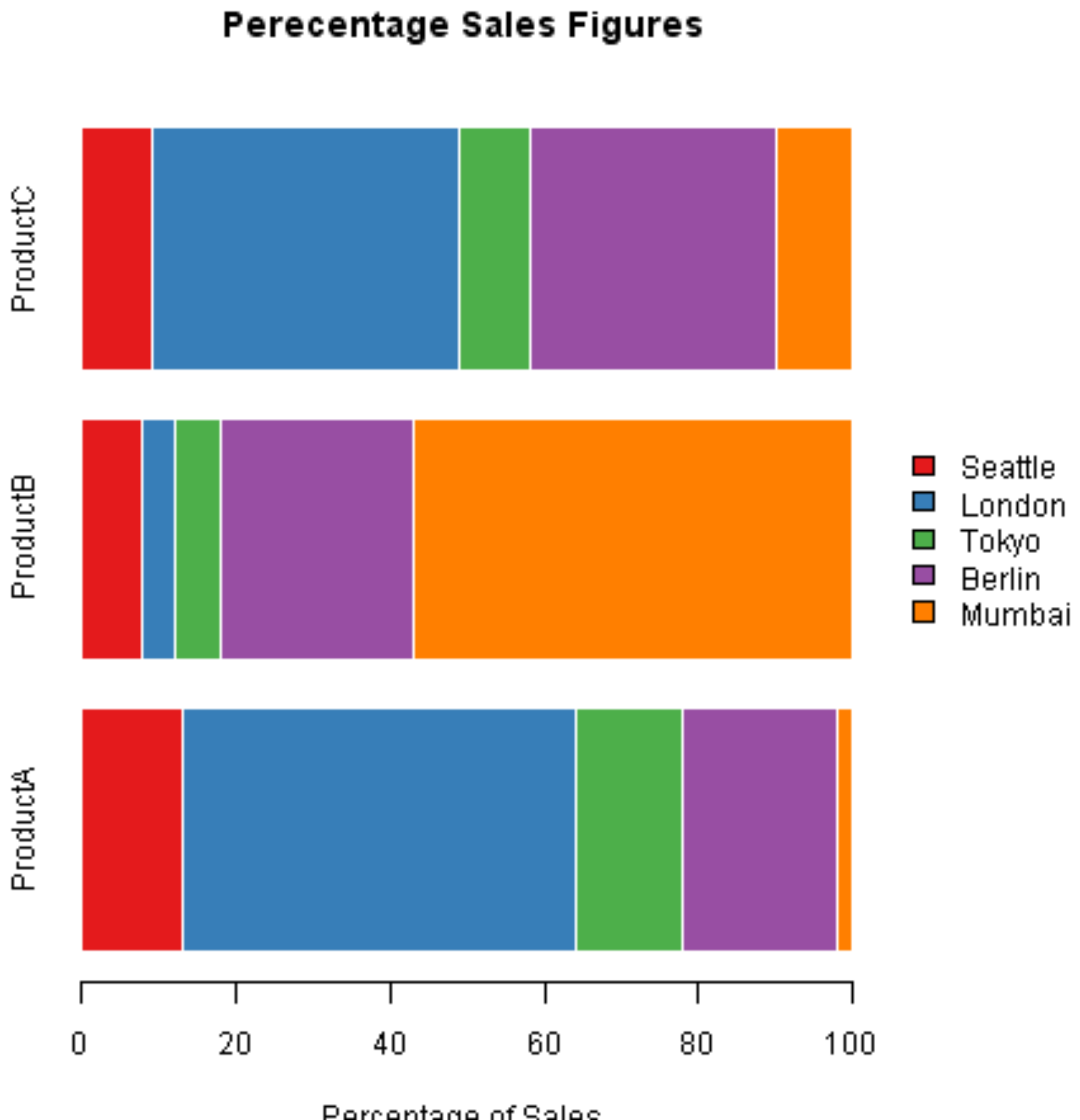




条形图 – 方向

```
barplot(as.matrix(citysales[,2:4]), beside=TRUE,horiz=TRUE,
        legend.text=citysales$City,
        args.legend=list(bty="n"),col=brewer.pal(5,"Set1"), border="white",
        xlim=c(0,100),
        xlab="Sales Revenue (1,000's of USD)",main="Sales Figures")
```

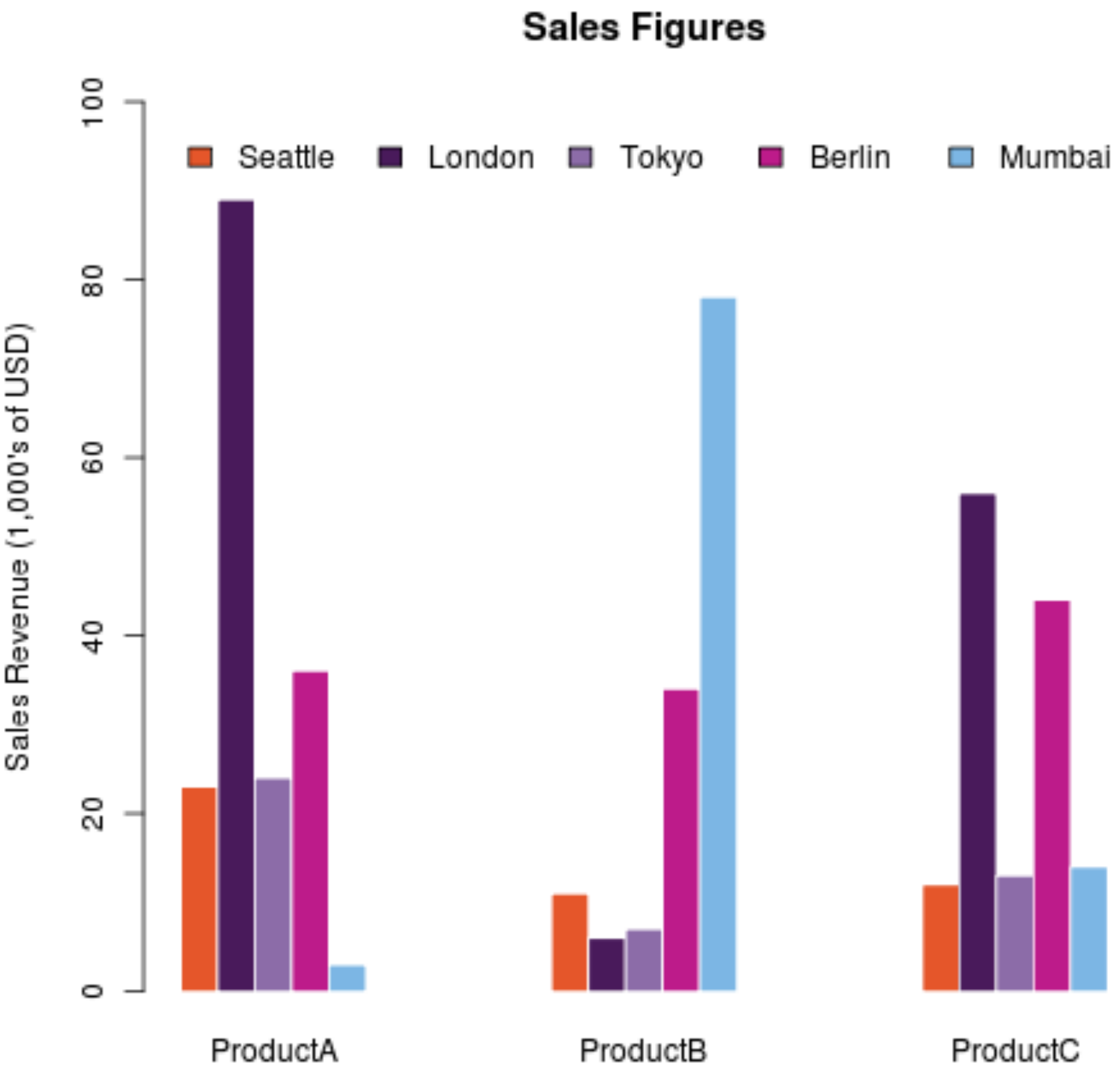


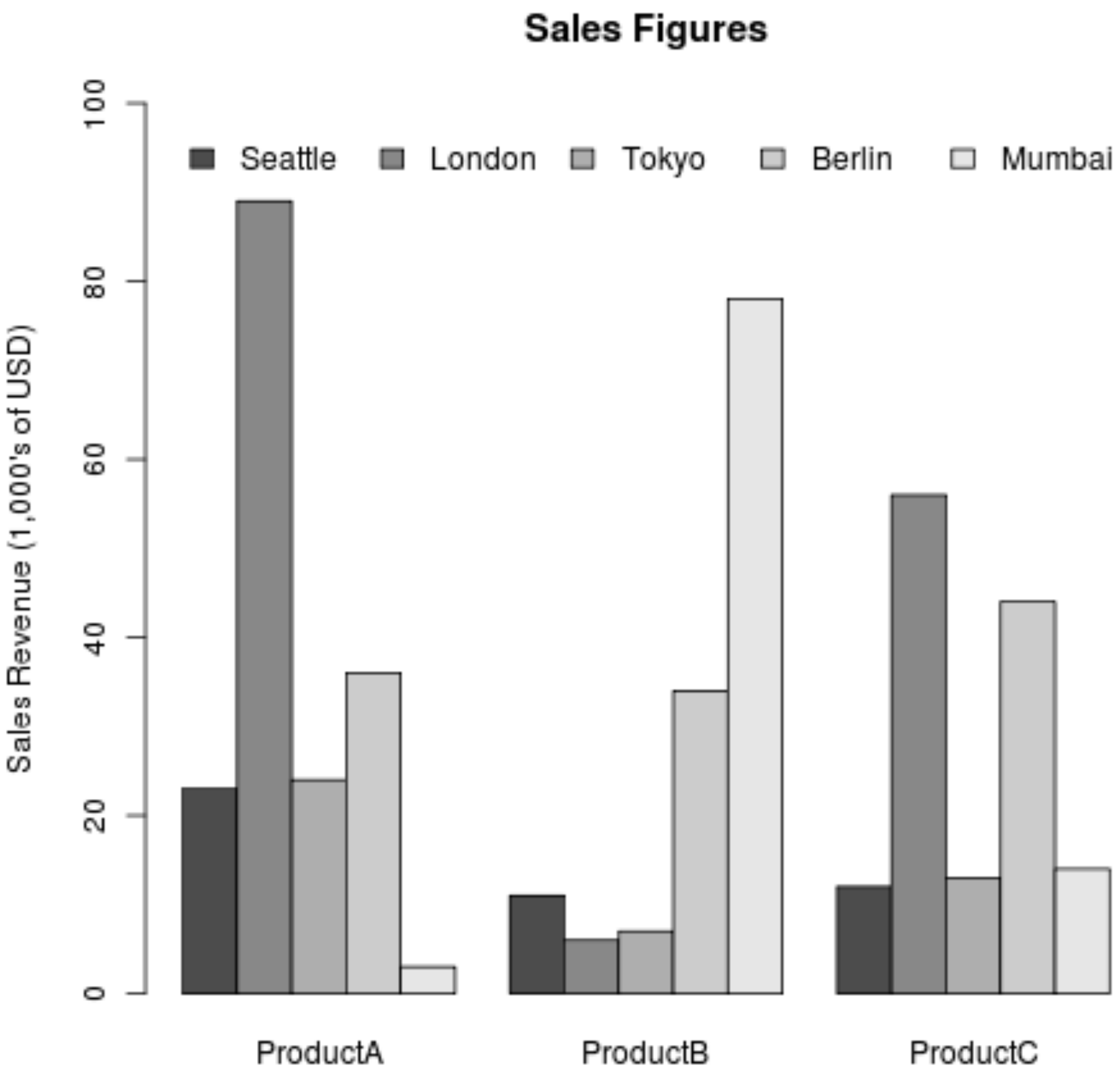


条形图 – 宽度、颜色、边界

```
barplot(as.matrix(citysales[,2:4]), beside=TRUE,
        legend.text=citysales$City,
        args.legend=list(bty="n",horiz=T),
col=c("#E5562A","#491A5B","#8C6CA8","#BD1B8A","
#7CB6E4"),
        border=FALSE,space=c(0,5),ylim=c(0,100),
        ylab="Sales Revenue (1,000's of USD)",
        main="Sales Figures")
```

```
barplot(as.matrix(citysales[,2:4]), beside=T,
        legend.text=citysales$City,
        args.legend=list(bty="n",horiz=T),
        ylim=c(0,100),
        ylab="Sales Revenue (1,000's of USD)",
        main="Sales Figures")
```

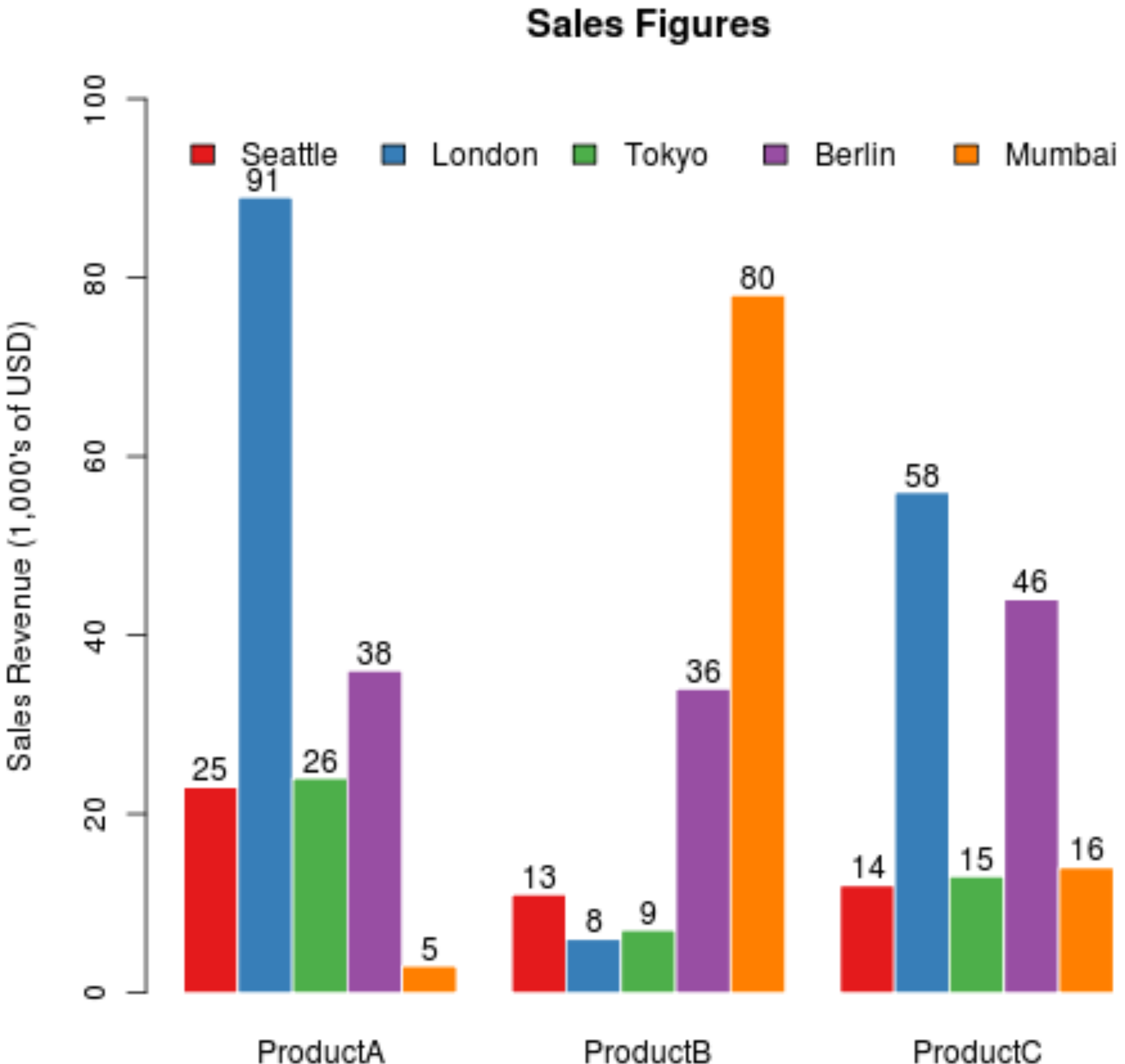


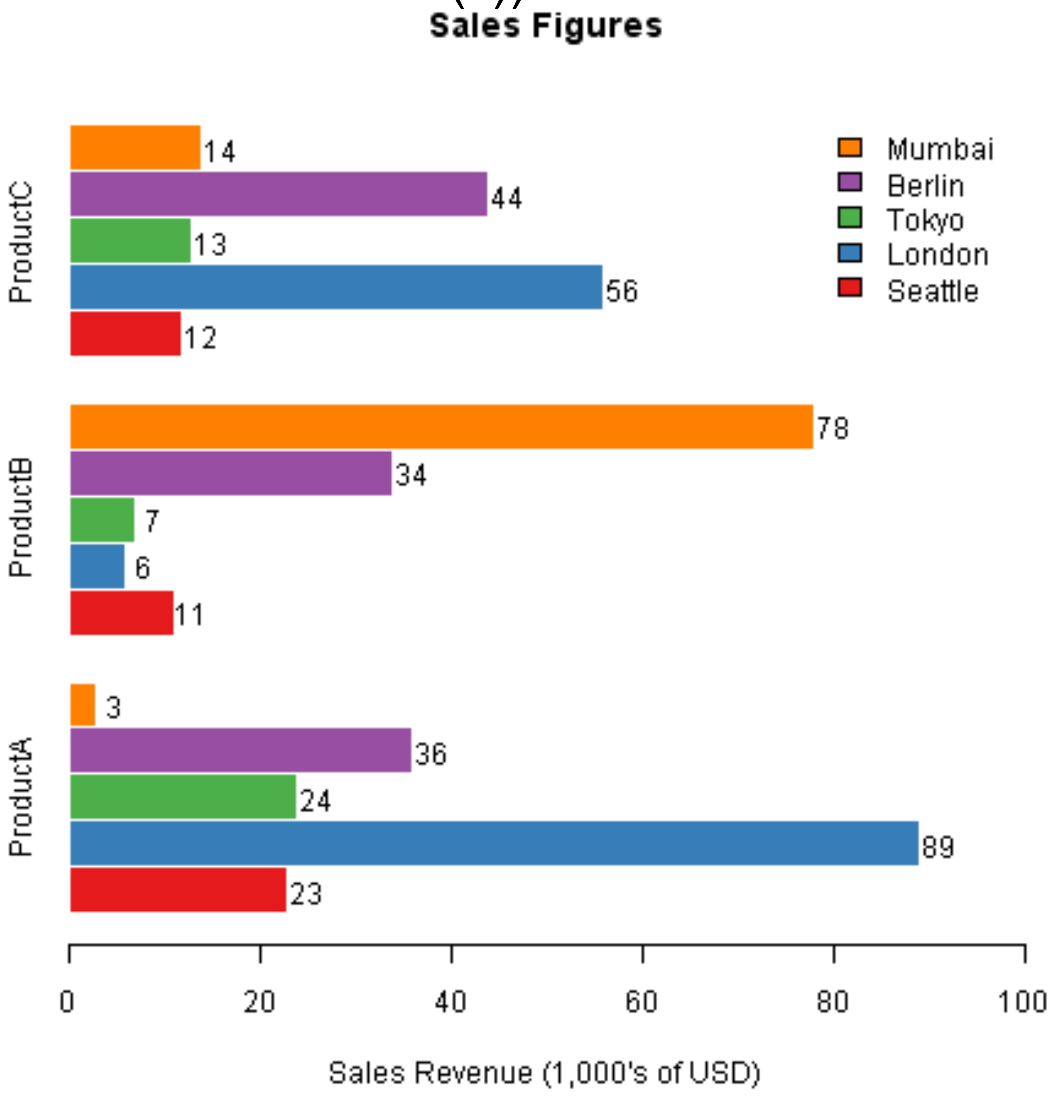


条形图 – 显示数字

```
x<-barplot(as.matrix(citysales[,2:4]), beside=TRUE,
        legend.text=citysales$City,
        args.legend=list(bty="n",horiz=TRUE),
        col=brewer.pal(5,"Set1"),
        border="white",ylim=c(0,100),
        ylab="Sales Revenue (1,000's of USD)",
        main="Sales Figures")
y<-as.matrix(citysales[,2:4])
text(x,y+2,labels=as.character(y))
```

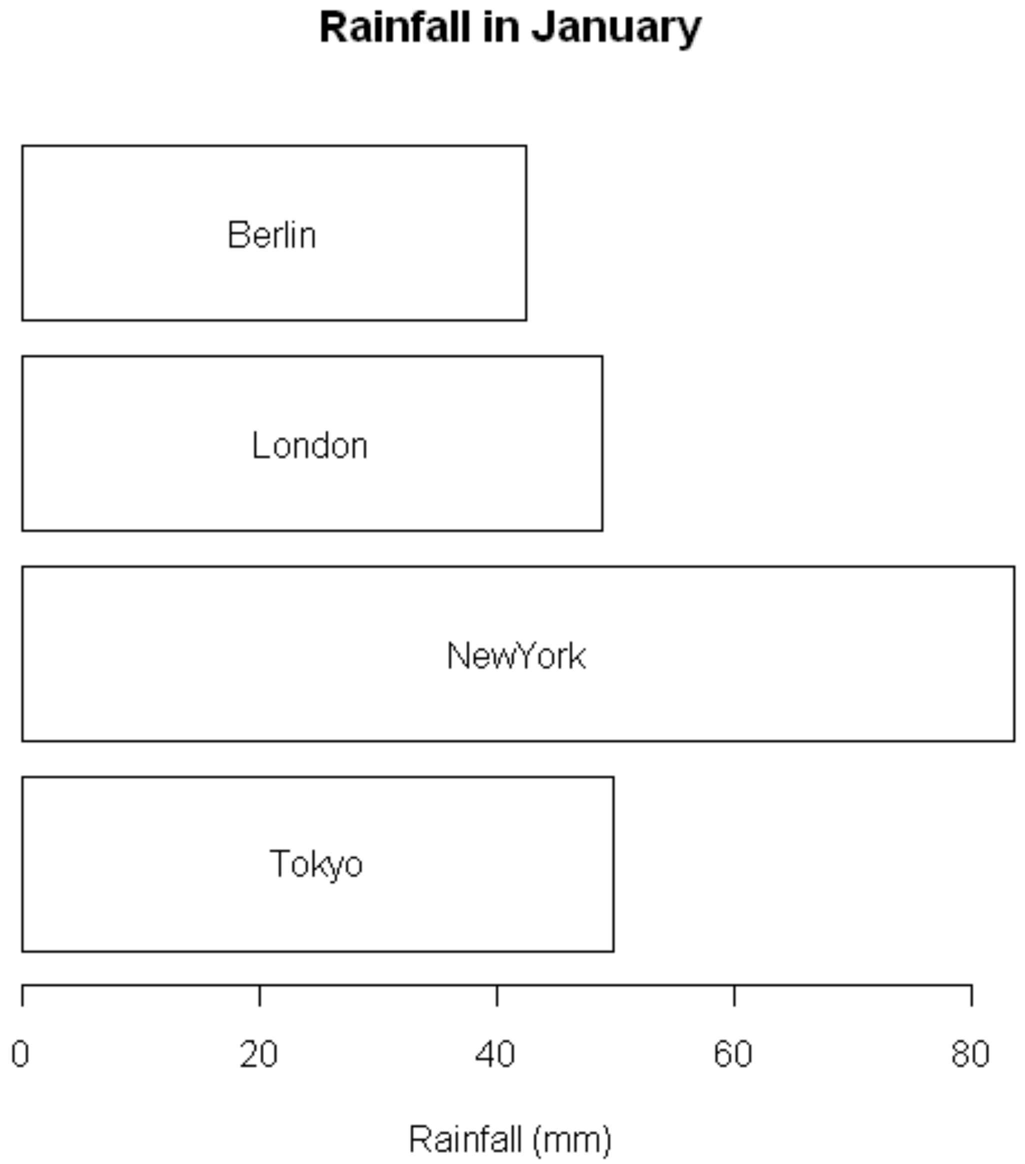
```
y<-barplot(as.matrix(citysales[,2:4]), beside=TRUE,horiz=TRUE,
        legend.text=citysales$City,
        args.legend=list(bty=“n"), col=brewer.pal(5,"Set1"),
        border=“white", xlim=c(0,100),
        xlab="Sales Revenue (1,000's of USD)",
        main="Sales Figures")
x<-as.matrix(citysales[,2:4])
text(x+2,y,labels=as.character(x))
```





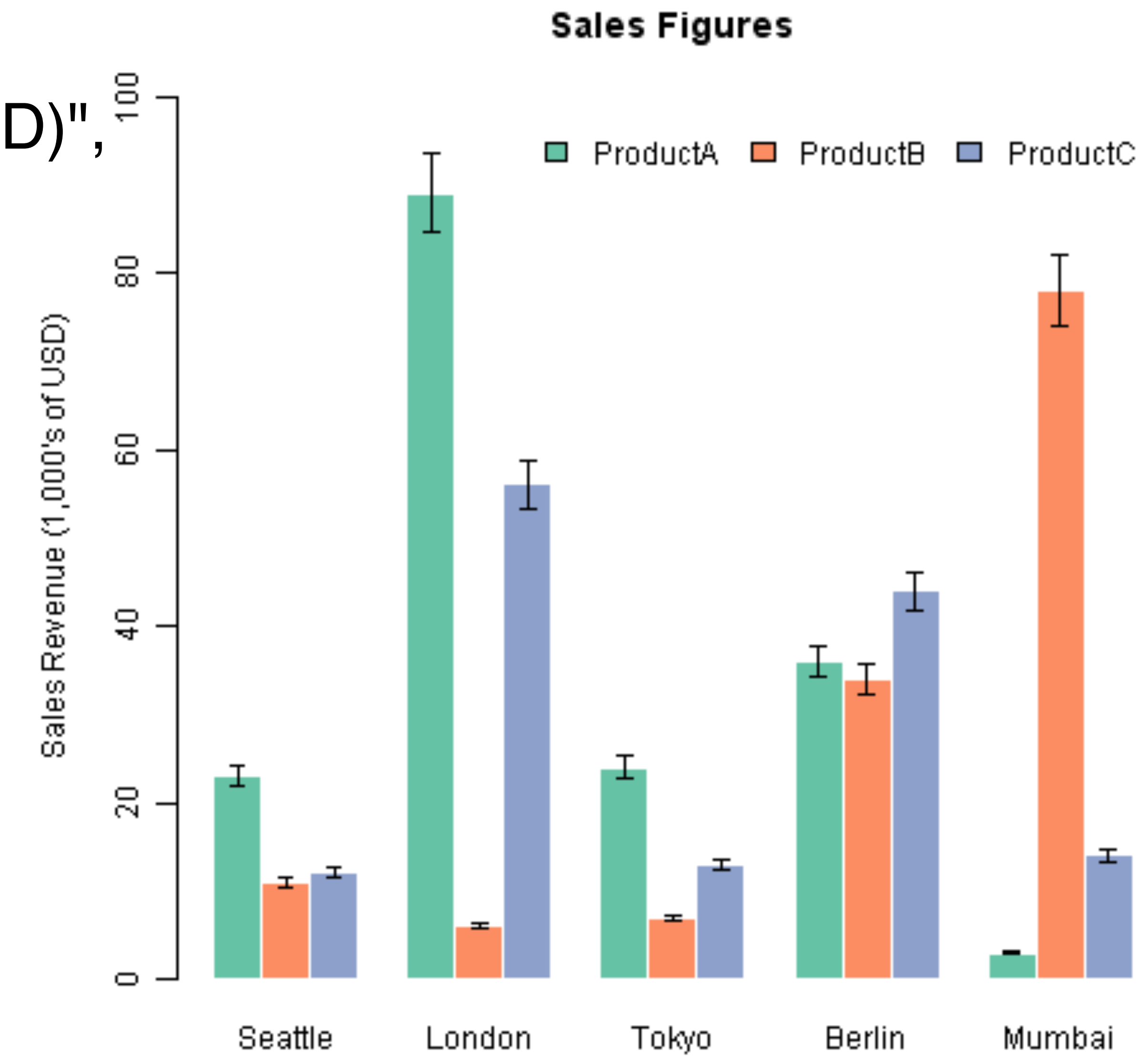
条形图 – 显示标记

```
rain<-read.csv("cityrain.csv")

y<-barplot(as.matrix(rain[1,-1]),horiz=T,col="white",yaxt="n",
        main="Monthly Rainfall in Major CitiesJanuary",
        xlab="Rainfall (mm)")

x<-0.5*rain[1,-1]
text(x,y,colnames(rain[-1]))
```



条形图 – 增加误差线

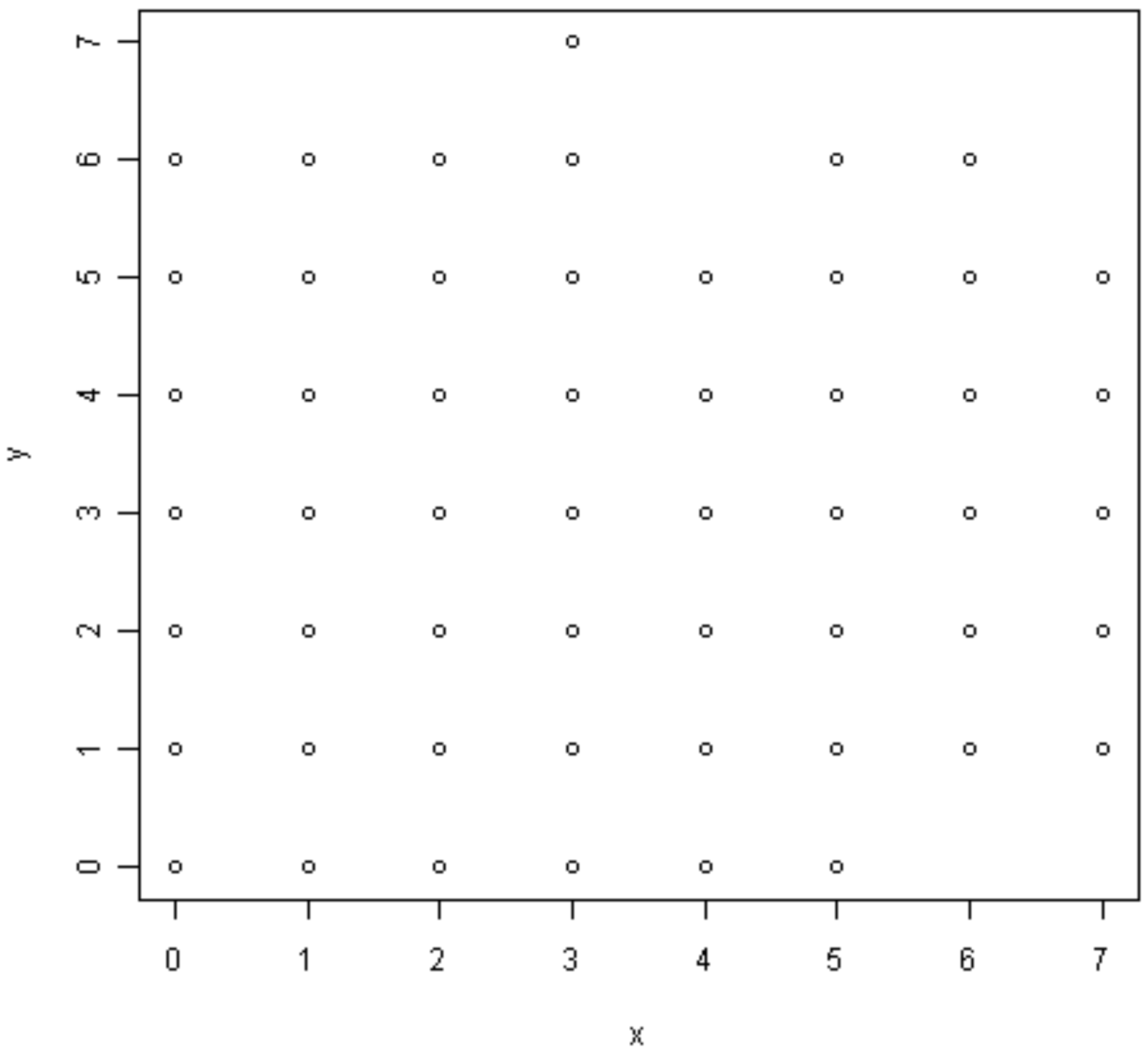
```
sales<-t(as.matrix(citysales[,-1]))
colnames(sales)<-citysales[,1]

x<-barplot(sales,beside=T,legend.text=rownames(sales),
          args.legend=list(bty="n",horiz=T),
          col=brewer.pal(3,”Set2"),
         border="white",ylim=c(0,100),
          ylab="Sales Revenue (1,000's of USD)",
          main="Sales Figures")

arrows(x0=x,
      y0=sales*0.95,
      x1=x,
      y1=sales*1.05,
      angle=90,
      code=3,
      length=0.04,
      lwd=0.4)
```

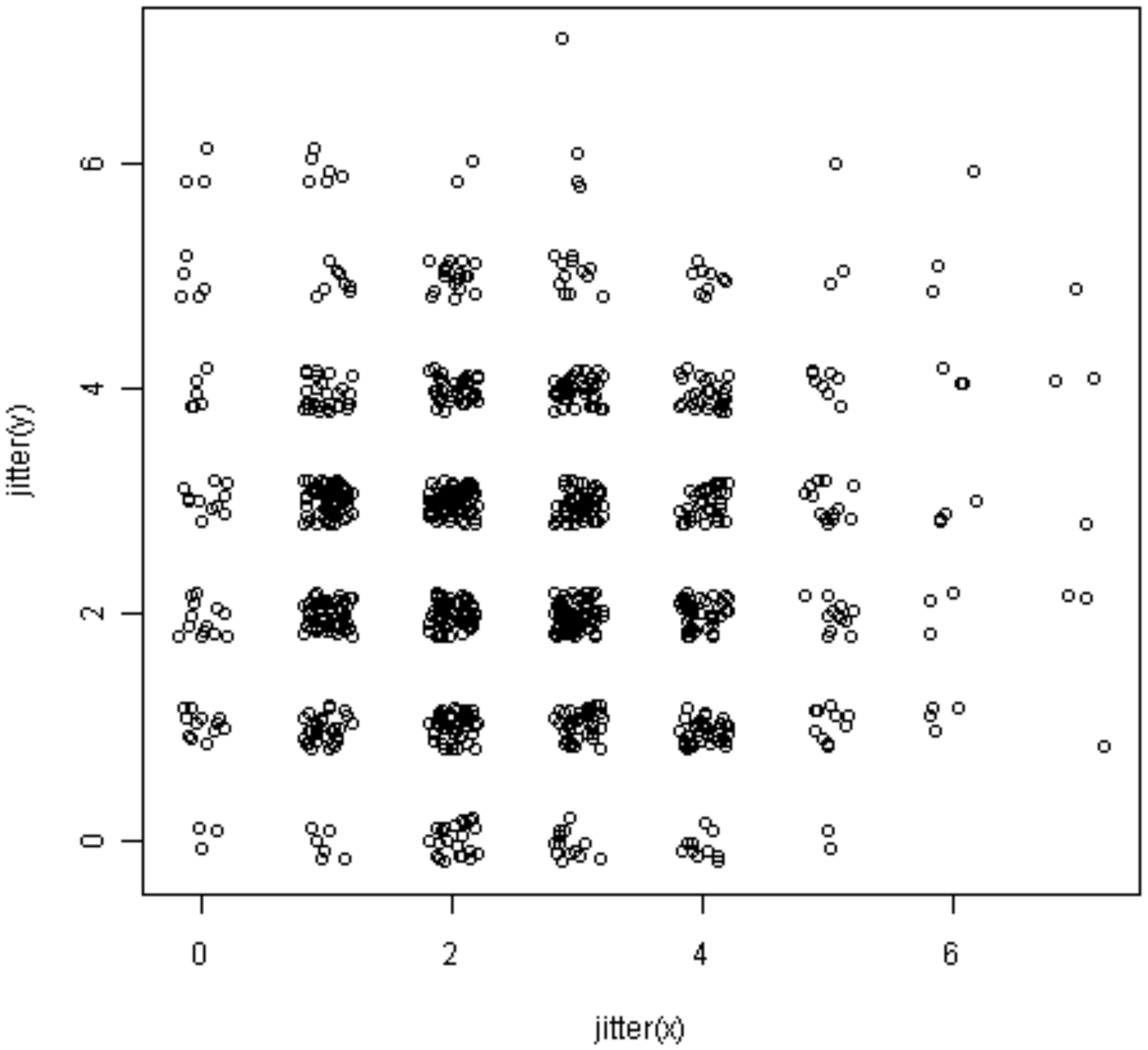


散点图 – 增加抖动

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

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





提问时间！

孙惠平

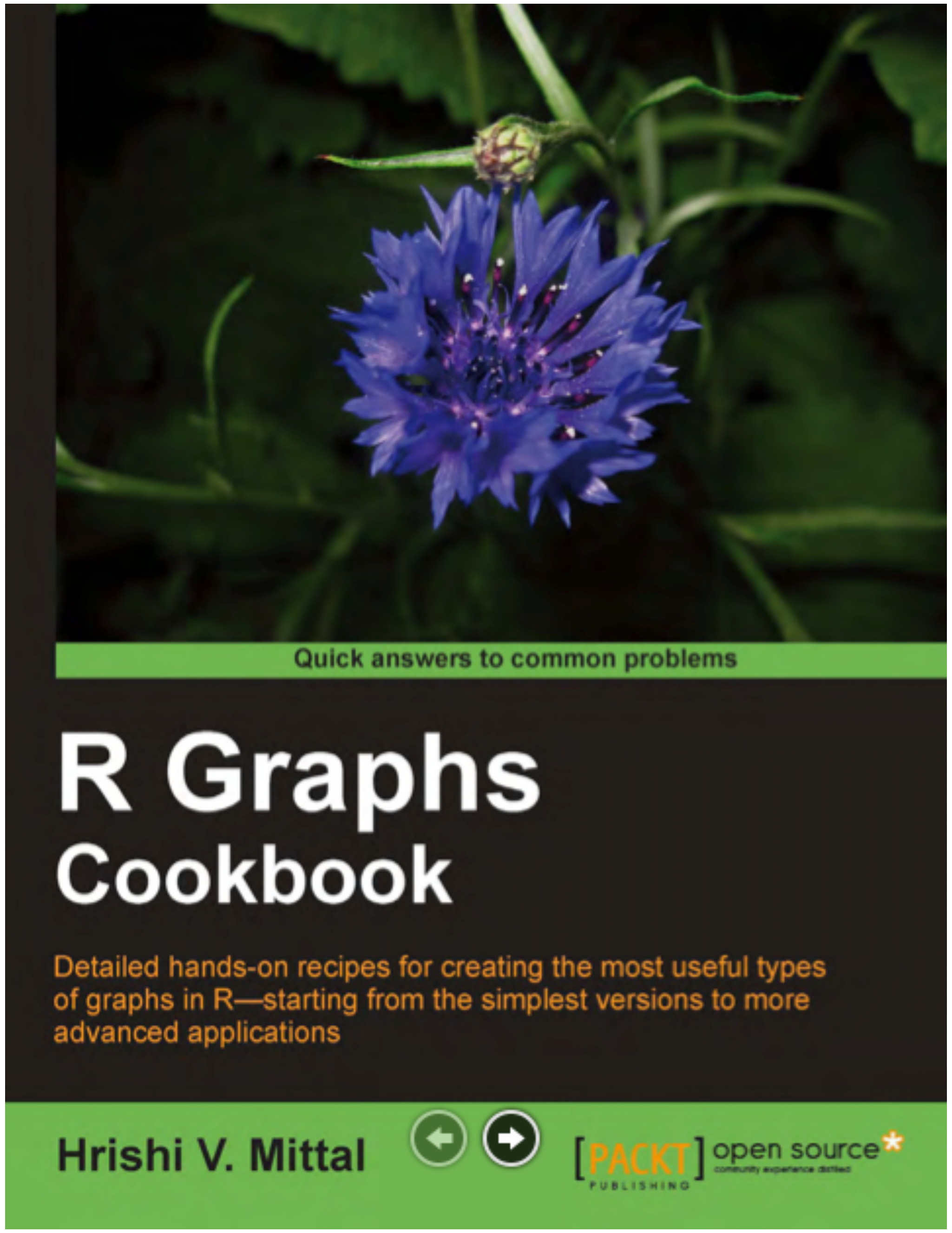
sunhp@ss.pku.edu.cn

练习

R Graphics II

练习－0021

第3、6章


Quick answers to common problems
R Graphs
Cookbook
Detailed hands-on recipes for creating the most useful types of graphs in R—starting from the simplest versions to more advanced applications
Hrishi V. Mittal
[PACKT] open source
PUBLISHING

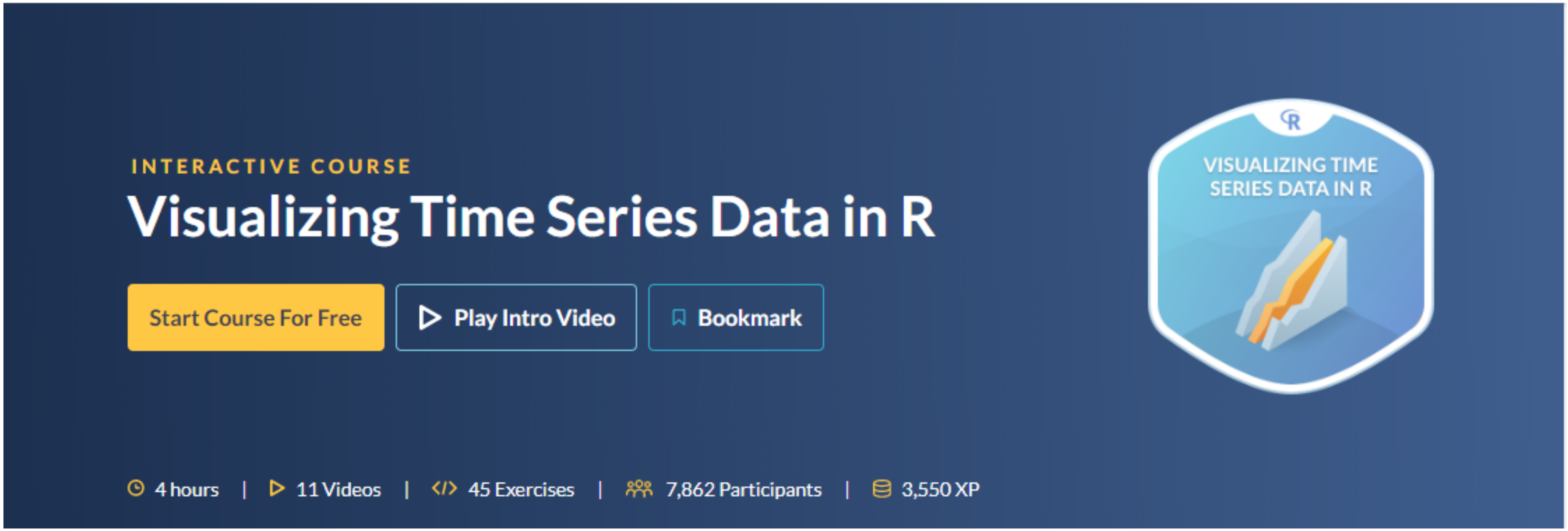
第1－6章：看完！！

练习 - 0022

- gdp_long.txt
- 做折线图（网格、特殊线，图例的不同位置）
- 条形图（正常、堆积、横向、颜色宽度等、显示数字、误差线）

- cityrain.csv
- 做折线图（边界标注，slide，mar和bty的含义）

练习 – 0023



INTERACTIVE COURSE

Visualizing Time Series Data in R

4 hours | 11 Videos | 45 Exercises | 7,862 Participants | 3,550 XP

提交方式和上节课一样！

https://www.datacamp.com/courses

谢谢！

孙惠平

sunhp@ss.pku.edu.cn