Data Analysis Tools and Practice(Using R)

2018.04.10

R基本图形II

北京大学 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

课堂测试时间

课堂测试06

先用电脑完成
30分钟 然后誊抄纸上

- 1、数据集alpe_d_huez2描述了环法自行车赛期间Alpe d'Huez赛段的最快时间，以及关于年份和吸毒指控的背景信息。绘制出车手最快时间的分布。使用a）直方图和b）箱线图显示它们。

- 2、mtcars是datasets包中的数据集。请使用str()函数了解这个数据集的构成，并输出数据集，然后按要求画图：

  ✳ a. 我们要设置一个蓝色背景和红色的点或线。 我们应该使用什么命令

  ✳ b. 画出cyl和mpg关系的散点图，并将结果输出为plot.png，要求输出为白底，360px*360px,点的大小为72

- 3、obama_vs_mccain数据集描述了2008年美国总统选举中的各州投票信息，以及关于收入，失业，种族和宗教的背景信息。

  ✳ a. 画出收入Income和参加选举比例Turnout之间的关系的散点图。提示：Turnout存在Na值。

  ✳ b. 将上述图形点的形状为黑色实心三角形(17)

  ✳ c. 数据集中有一个因子类型的列regions,请画出每个地区region下的收入Income和参加选举比例Turnout之间的关系的散点图。要求设置布局为5列，行优先。

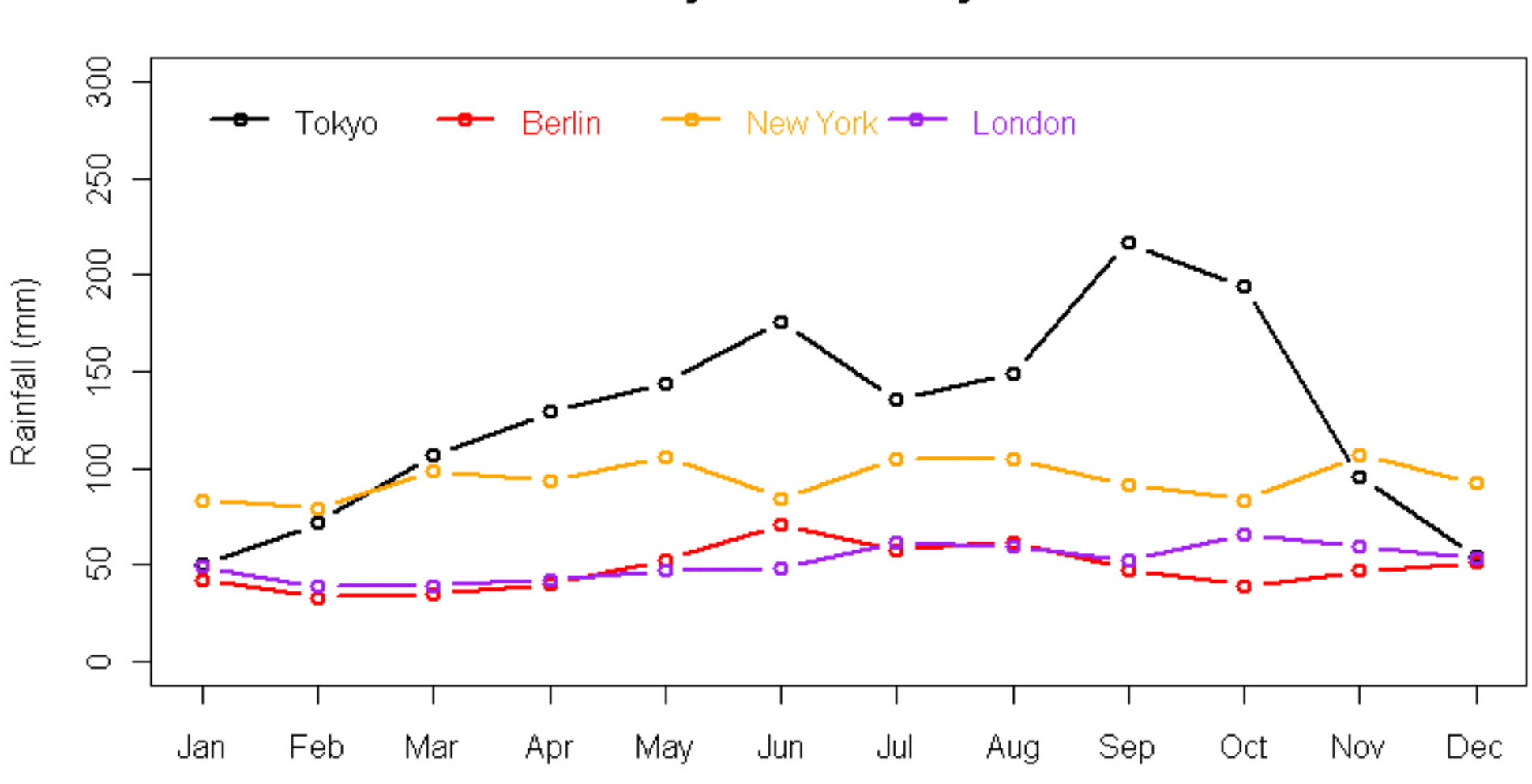
上次课程内容回顾

- 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();

图形控制

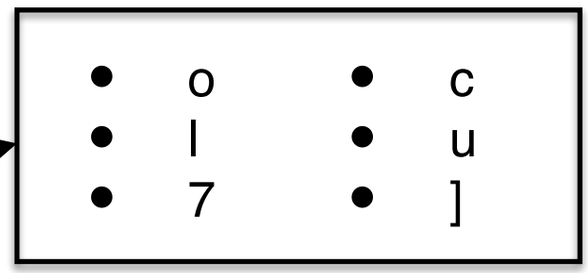
图例 - 使用坐标定位

```
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"))
```



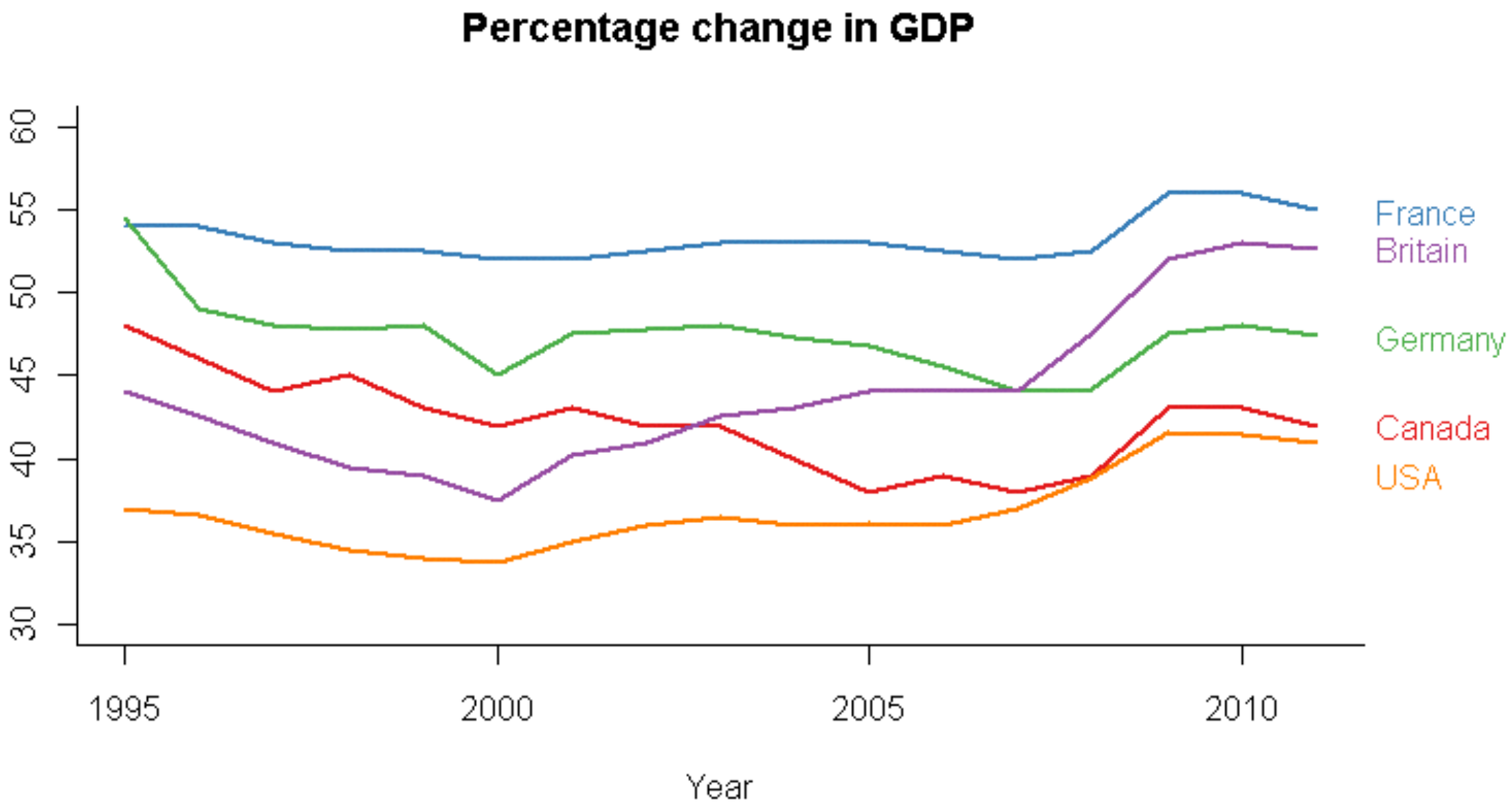
图例 - 边界标记

```
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)
```



- side
- 1,2,3,4

图例 – 边界标记

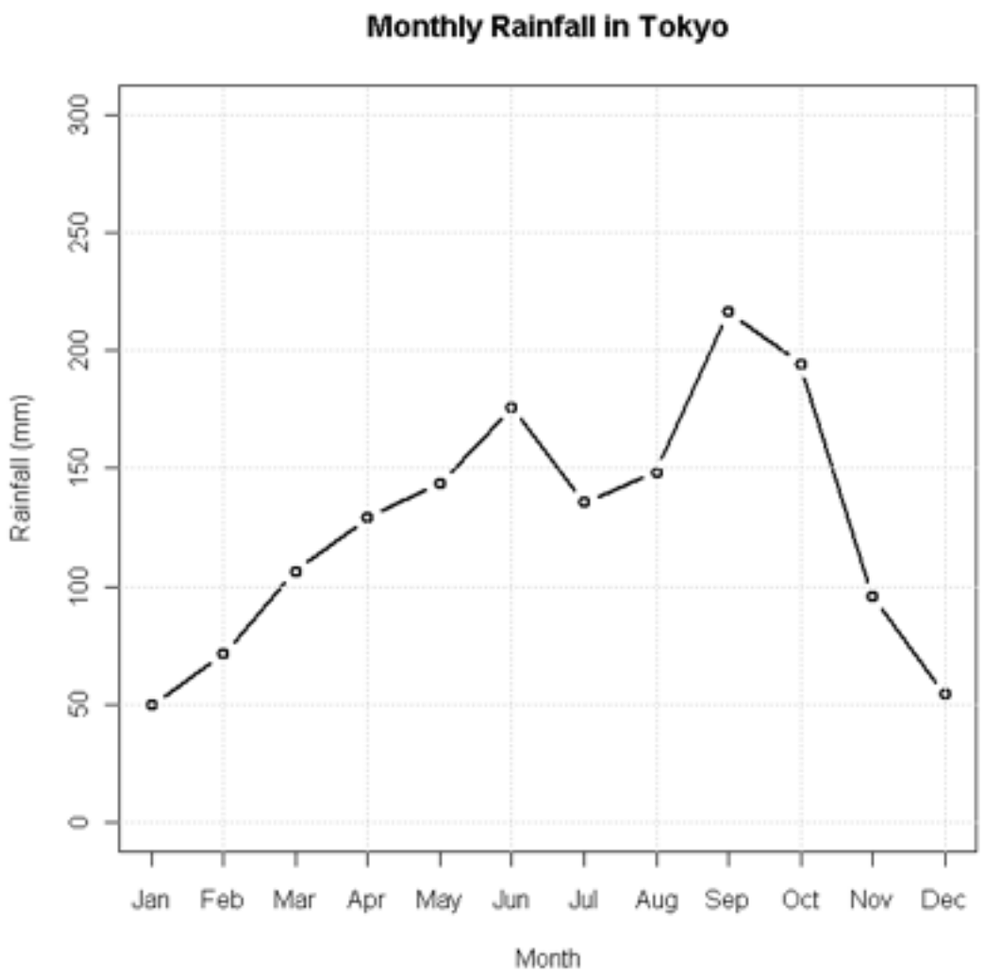

Percentage change in GDP
30
35
40
45
50
55
60
1995
2000
2005
2010
Year
France
Britain
Germany
Canada
USA

折线图 - 网格图

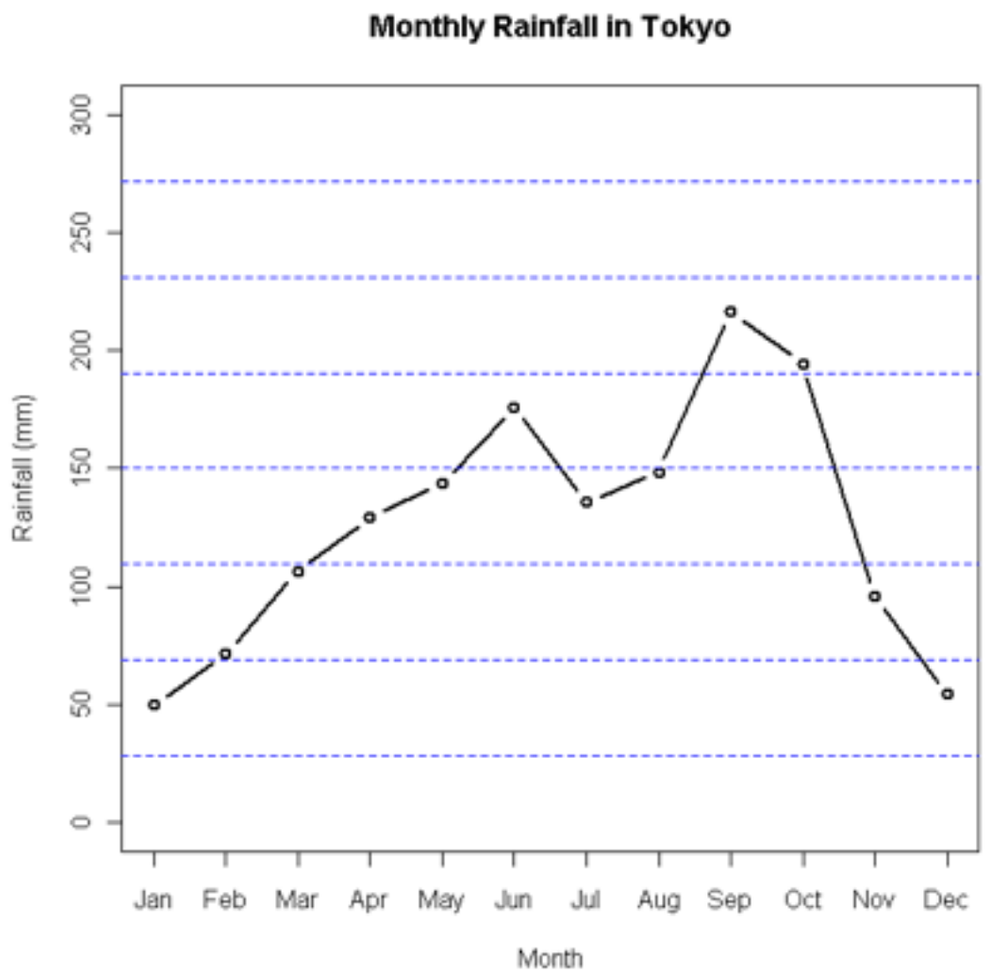
```
rain<-read.csv(“cityrain.csv")
plot(rain$Tokyo,type=“b",lwd=2, xaxt=“n”,ylim=c(0,300),col=“black", xlab=“Month",
      ylab="Rainfall (mm)",main="Monthly Rainfall in Tokyo")
axis(1,at=1:length(rain$Month),labels=rain$Month)
```

---

grid()



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

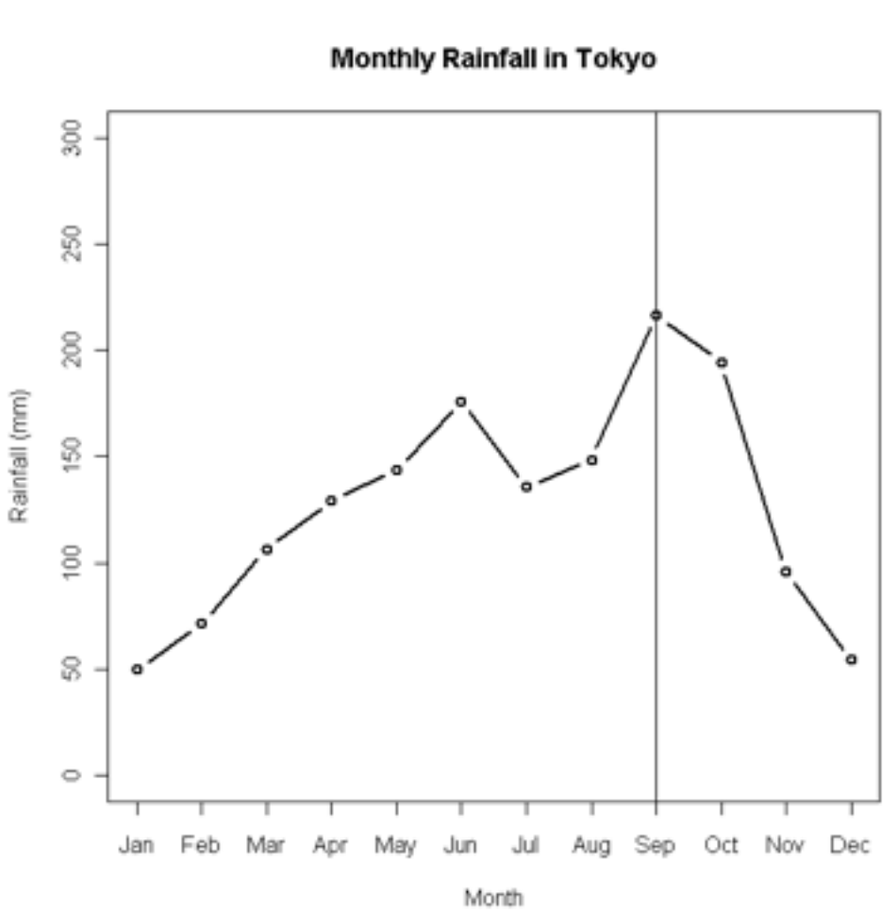


折线图 – 特殊线

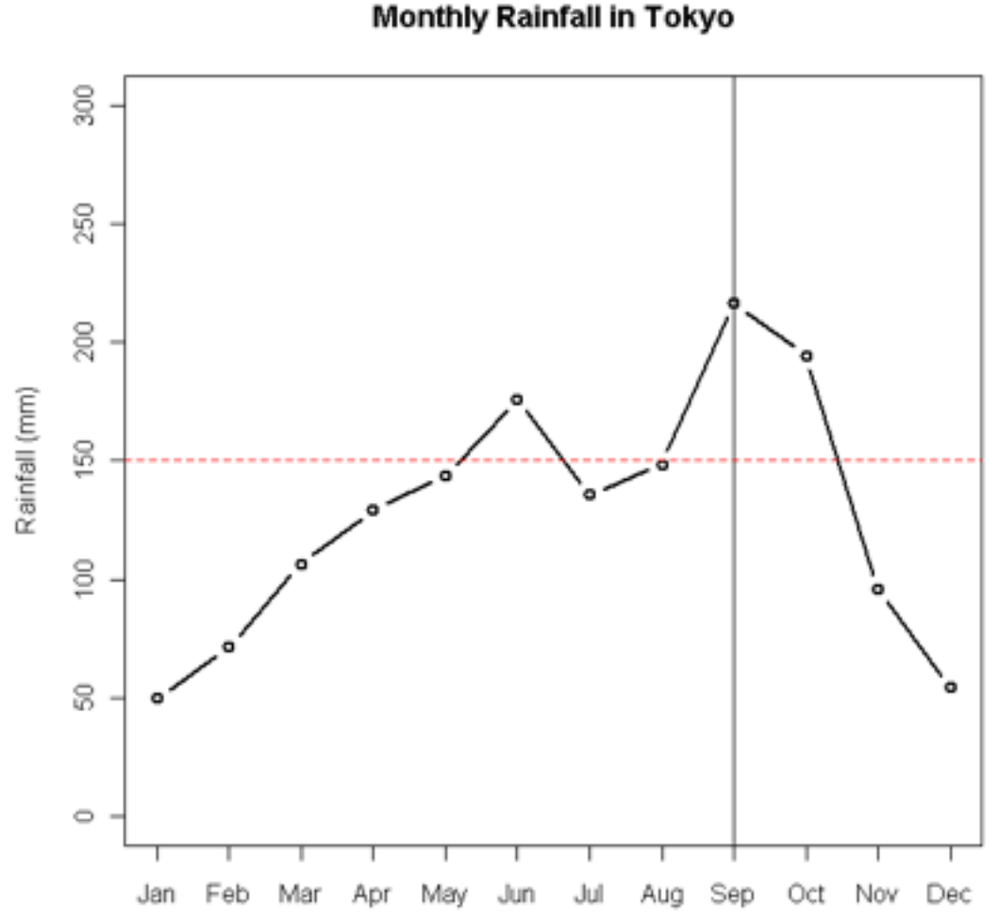
```
rain<-read.csv(“cityrain.csv")
plot(rain$Tokyo,type=“b",lwd=2, xaxt=“n”,ylim=c(0,300),col=“black", xlab=“Month",
      ylab="Rainfall (mm)",main="Monthly Rainfall in Tokyo")
axis(1,at=1:length(rain$Month),labels=rain$Month)
```

---

abline(v=9)



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

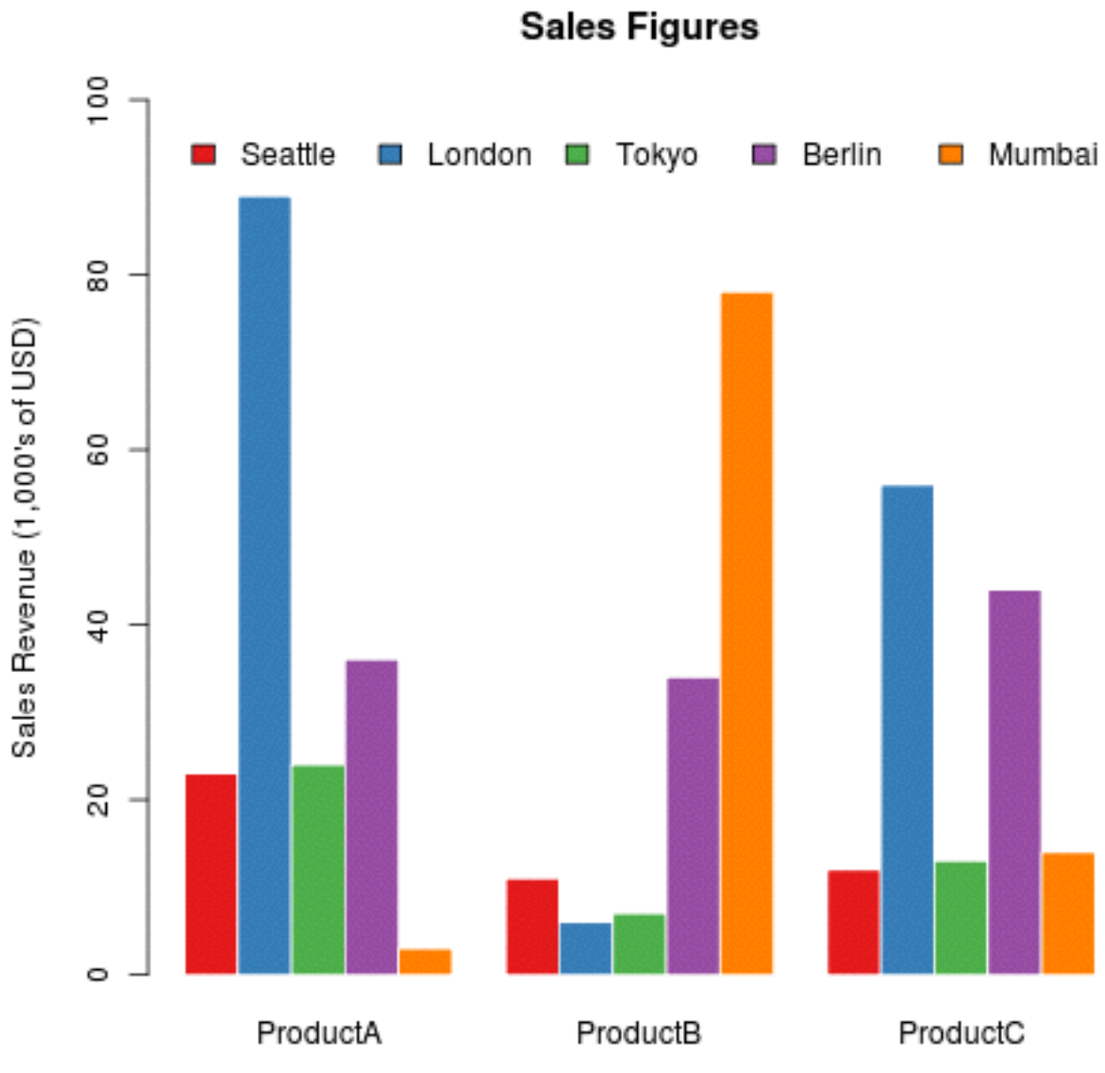


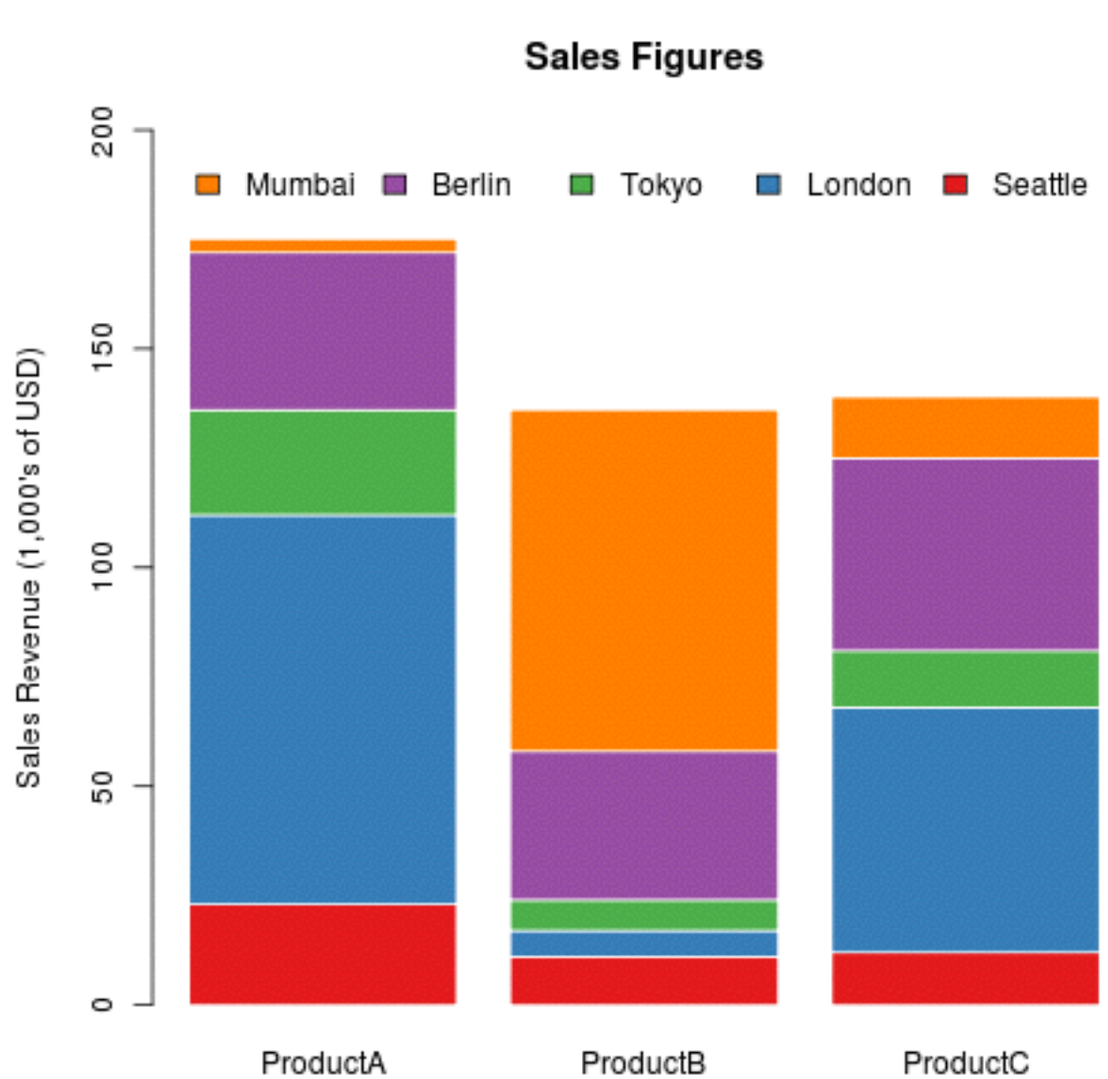
折线图 – 波形线

```
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

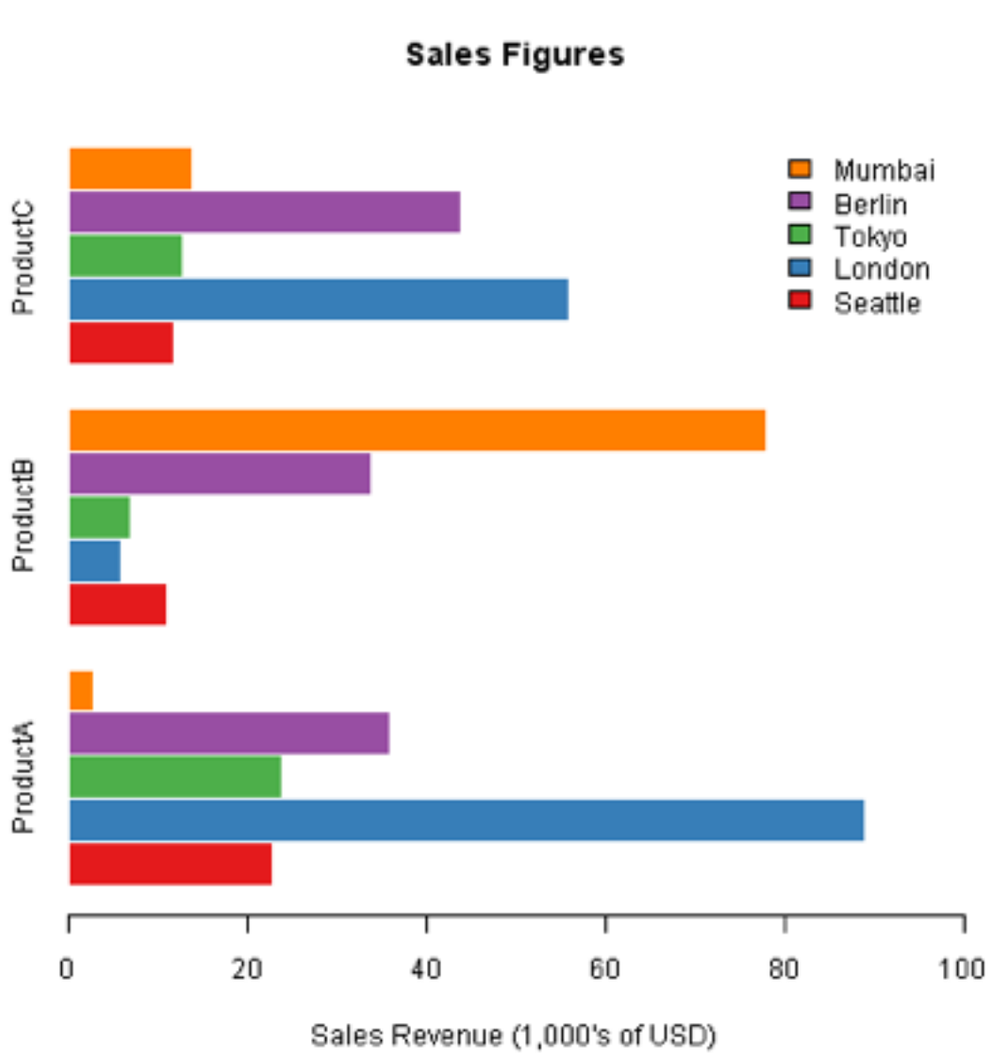
```
citysales<-read.csv("citysales.csv")

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")
```

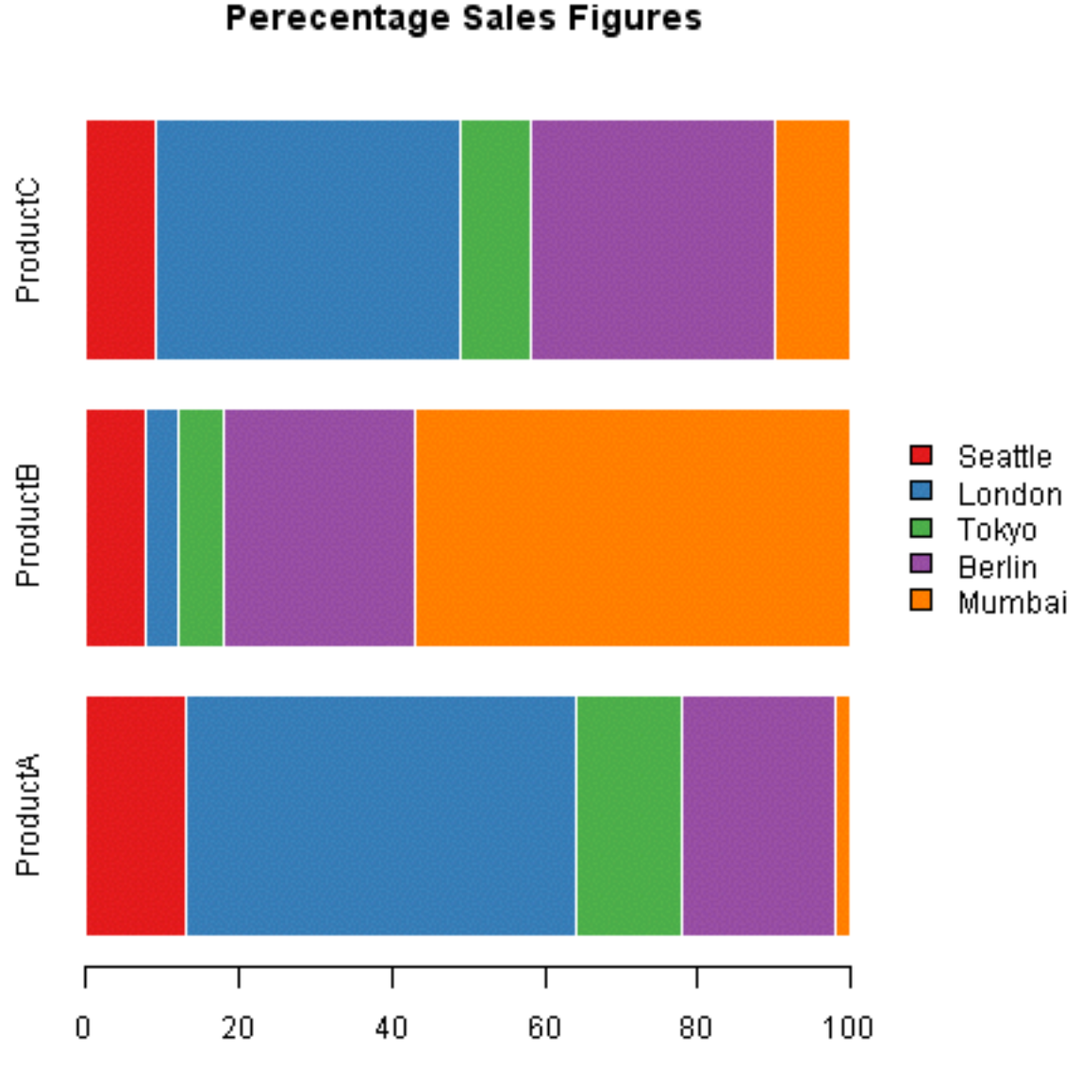




条形图 - 方向

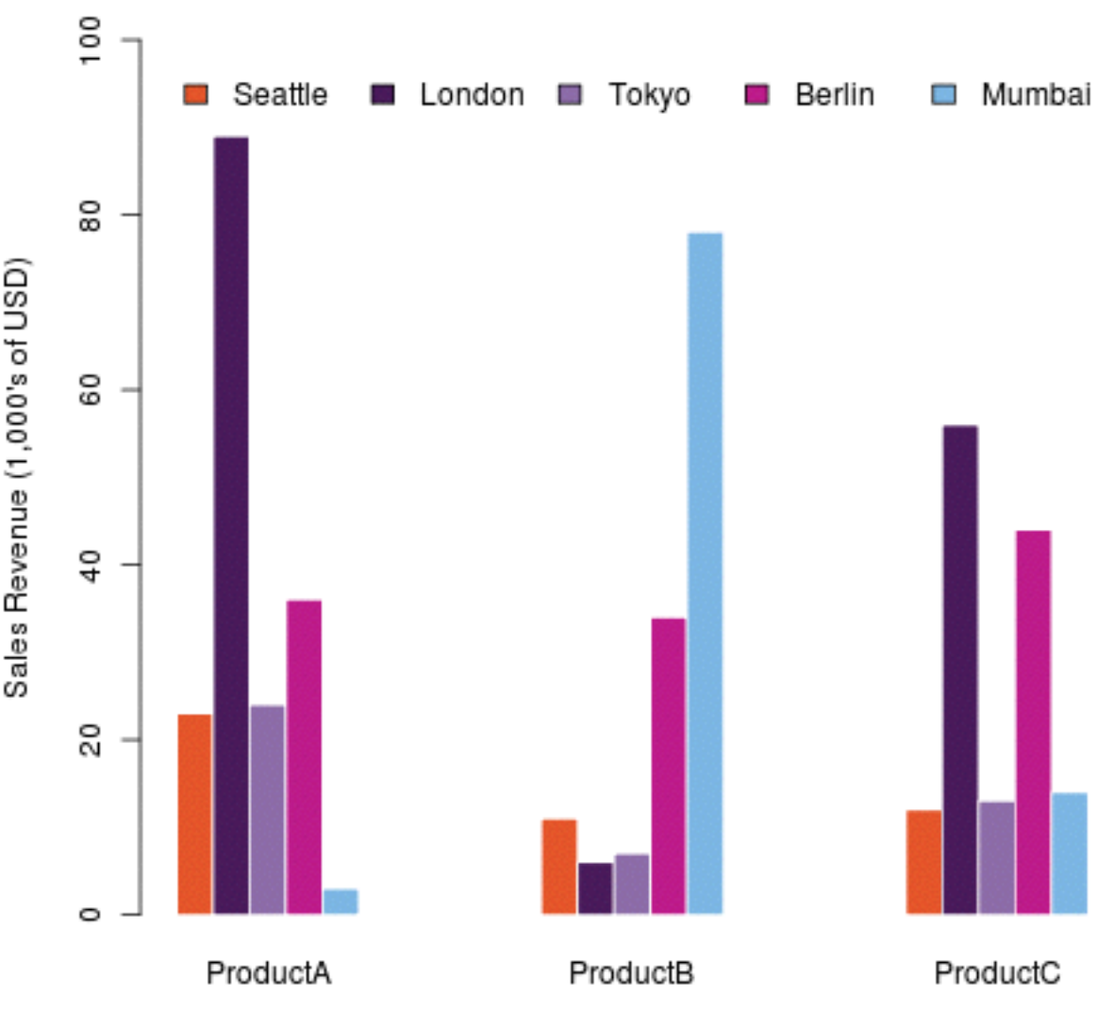
```
barplot(as.matrix(citysales[,2:4]), beside=TRUE,horiz=TRUE,
        legend.text=citysales$City,
        args.legend=list(bty="n"),col=brewer.pal(5,"Set1"), border="white",
        xlim=c(0,100),
        xlab="Sales Revenue (1,000's of USD)",main="Sales Figures")
```



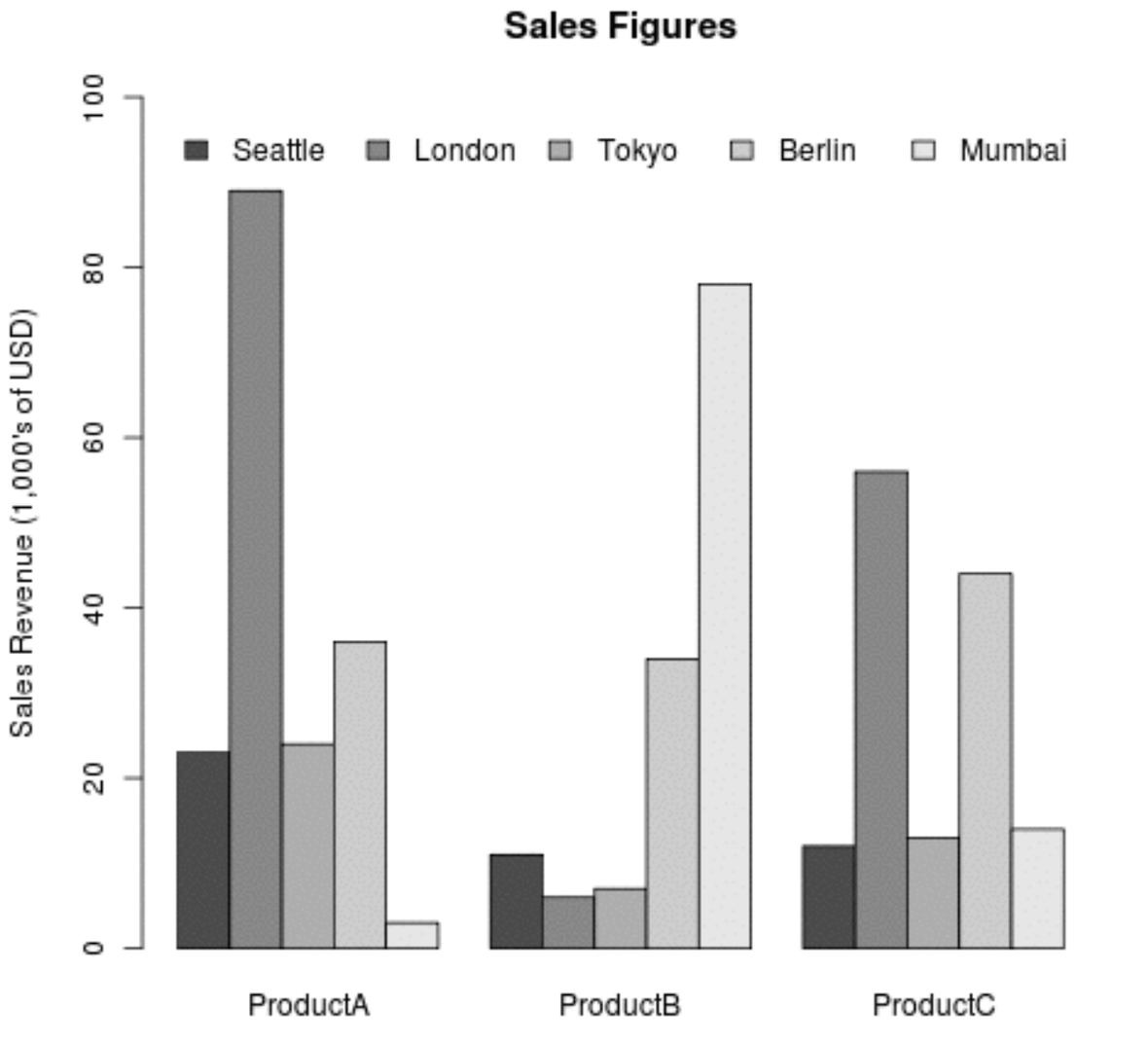


条形图 - 宽度、颜色、边界

```
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")
```

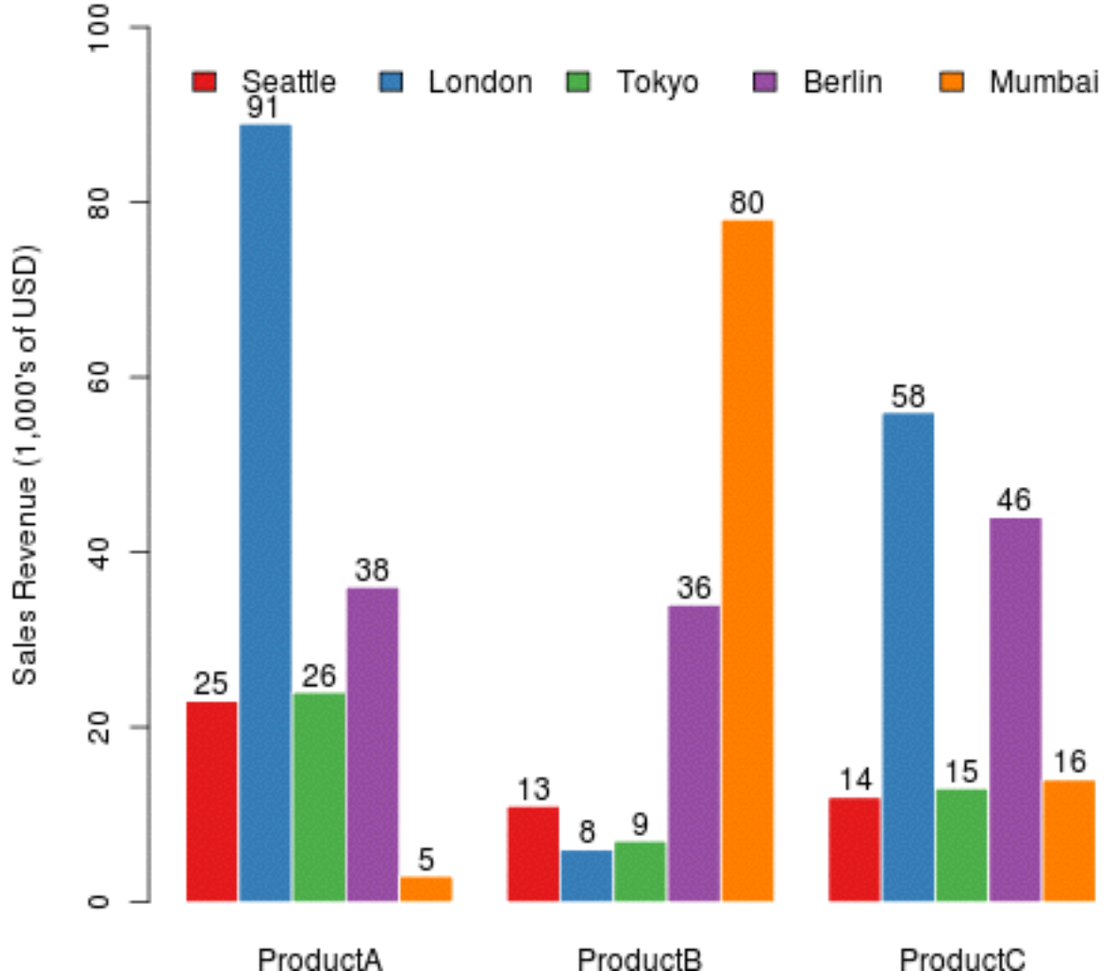



```
barplot(as.matrix(citysales[,2:4]), beside=T,
        legend.text=citysales$City,
        args.legend=list(bty="n",horiz=T),
        ylim=c(0,100),
        ylab="Sales Revenue (1,000's of USD)",
        main="Sales Figures")
```

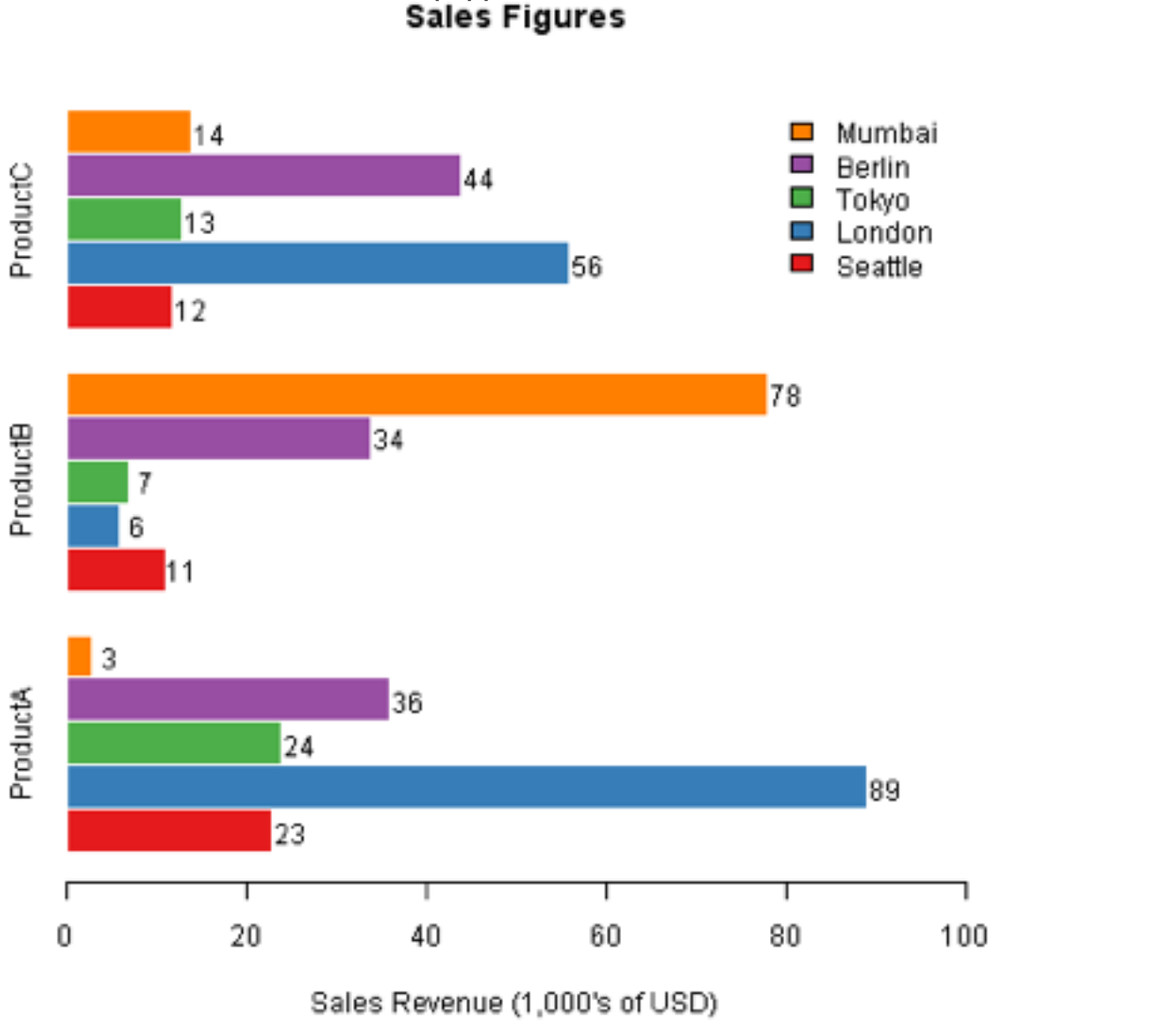


条形图 - 显示数字

```
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))
```

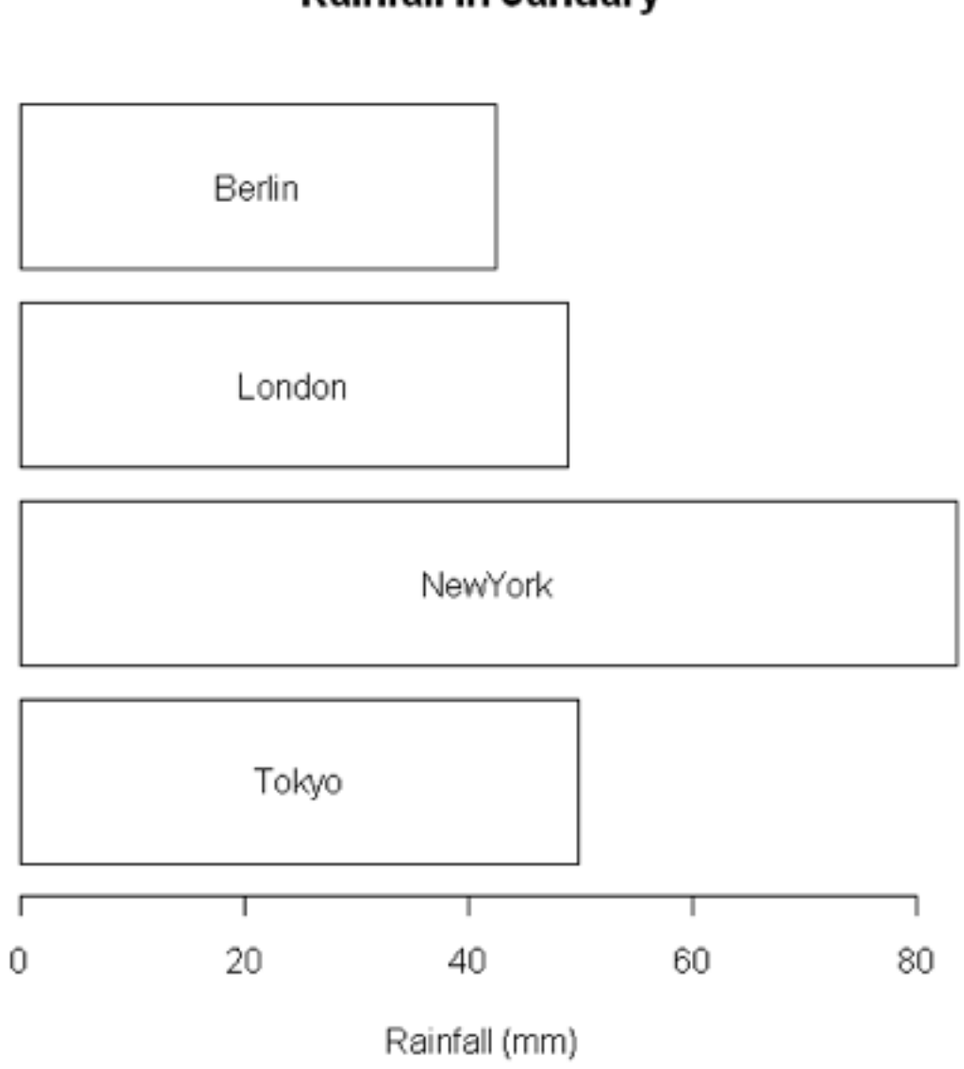



```
y<-barplot(as.matrix(citysales[,2:4]), beside=TRUE,horiz=TRUE,
        legend.text=citysales$City,
        args.legend=list(bty="n"), col=brewer.pal(5,"Set1"),
        border="white", xlim=c(0,100),
        xlab="Sales Revenue (1,000's of USD)",
        main="Sales Figures")
x<-as.matrix(citysales[,2:4])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



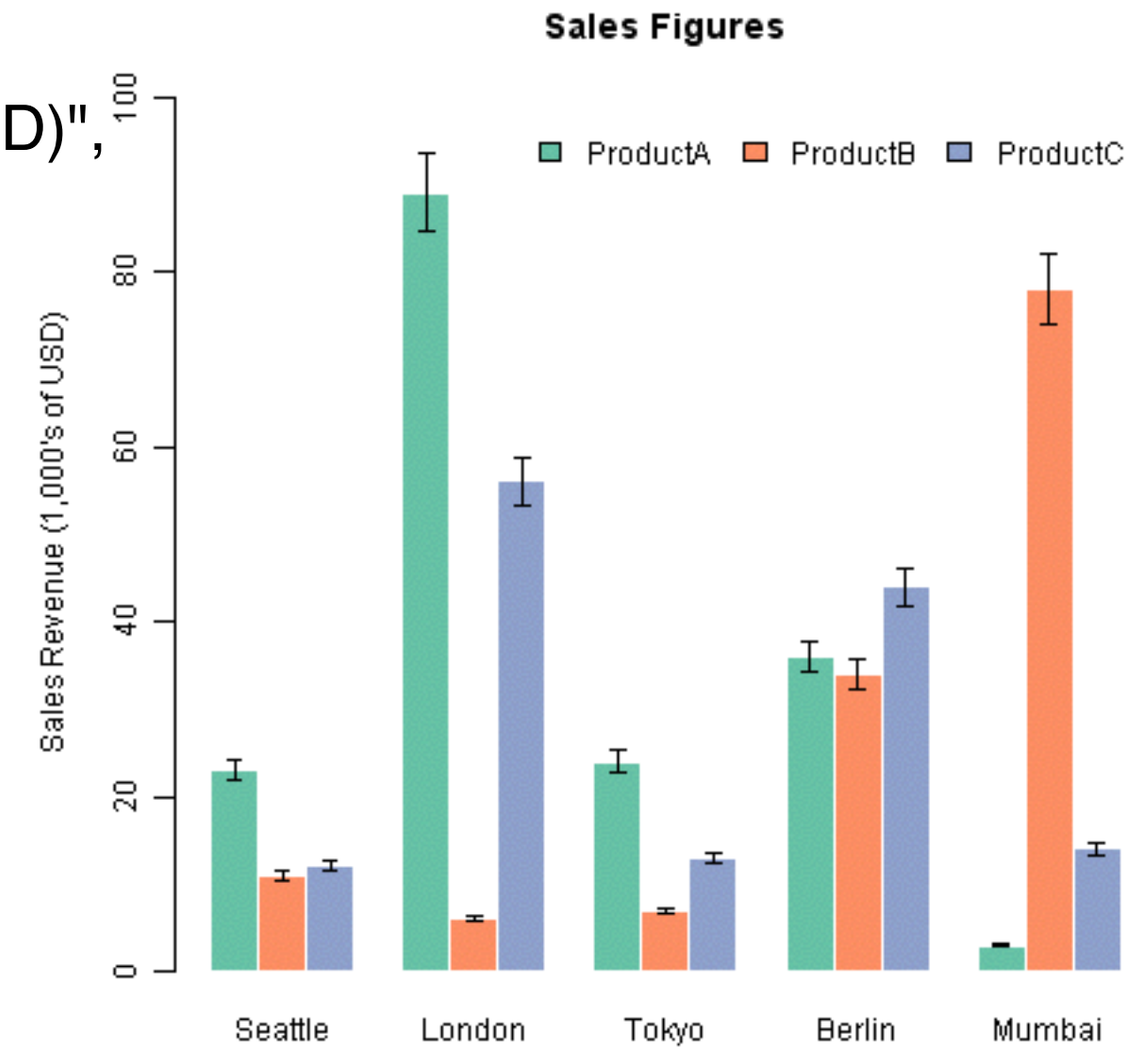
条形图 - 显示标记

```
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]))
```



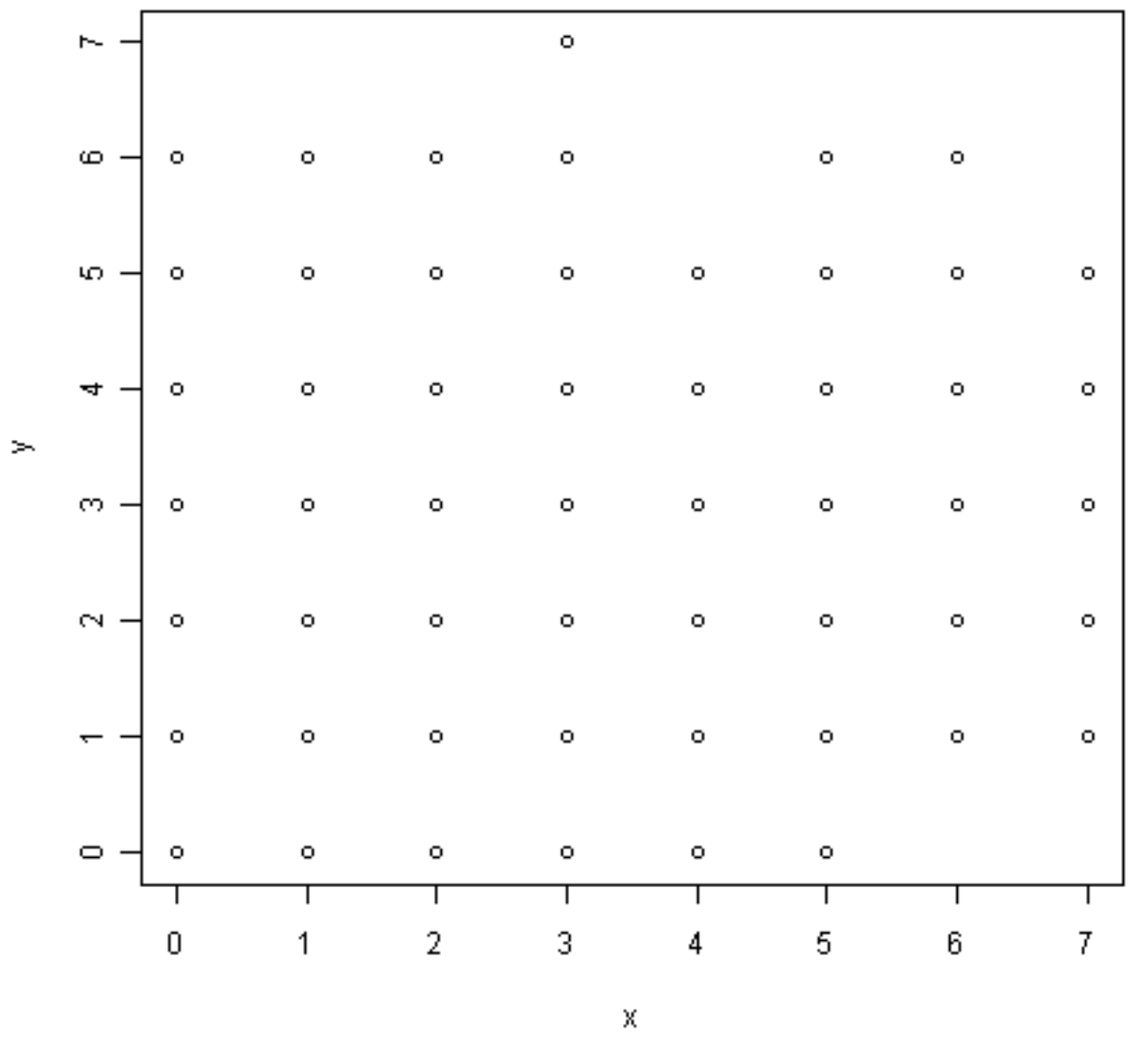
条形图 – 增加误差线

```
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)
```

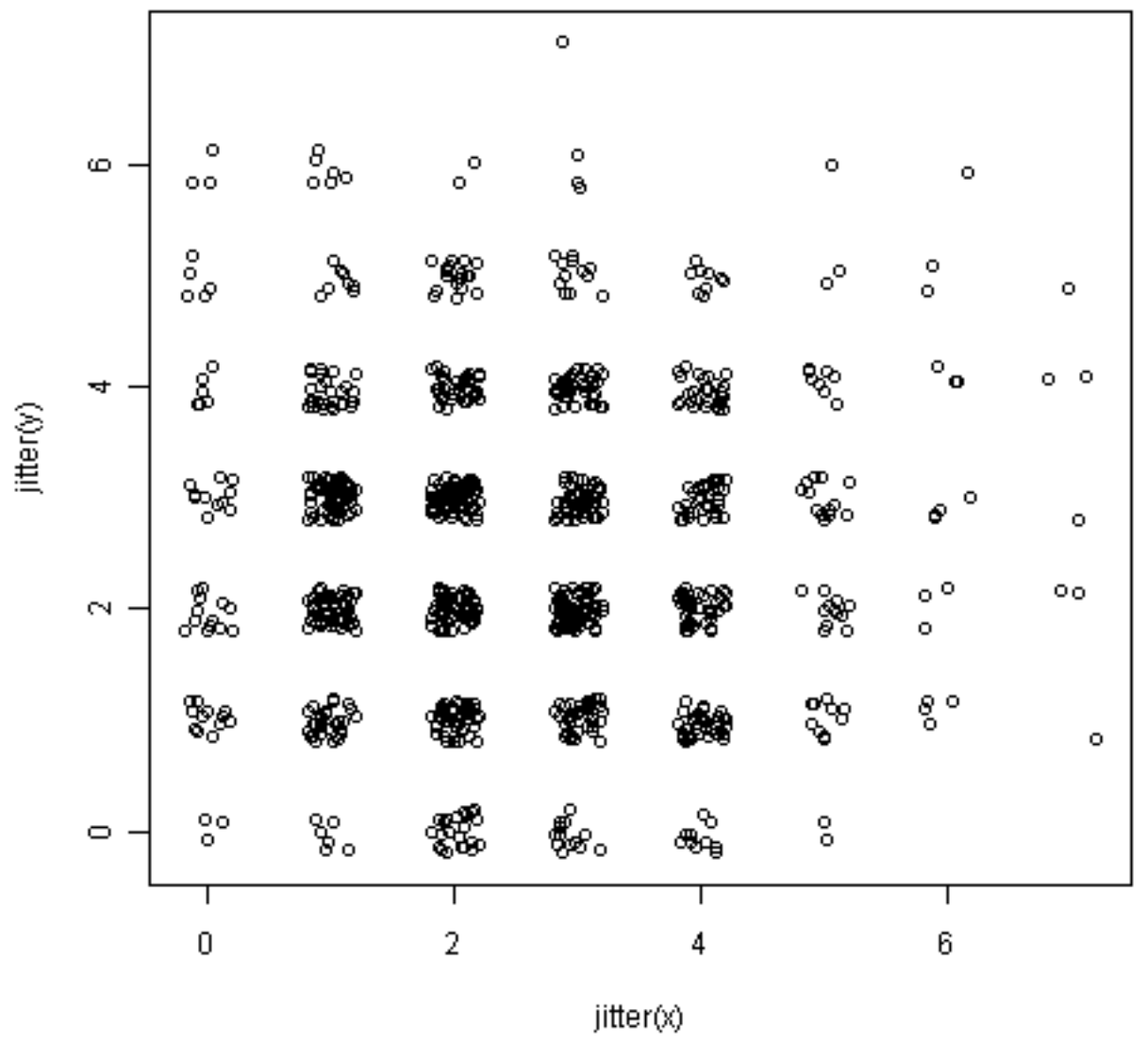


散点图 – 增加抖动

```
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

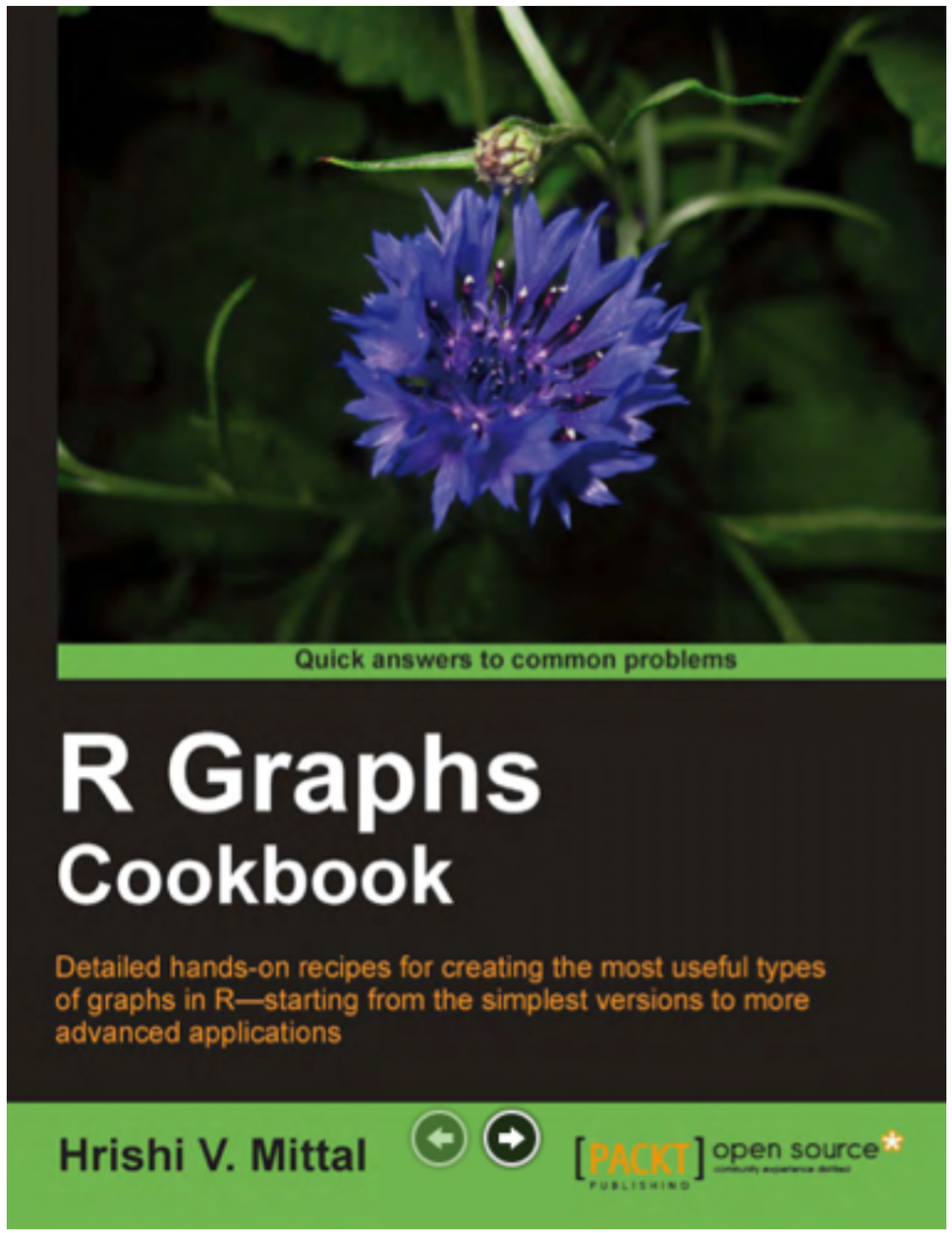
练习

练习 - 0027

install_course_github(“pkussdatanalysis","C7_Base_Graph_02")

练习 - 0028

第3、6章



第1－6章：看完！！

练习 – 0029

- 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）

课程项目

- 选择一个R的扩展包，做10分钟的课堂介绍，包括包的作用，示例，2道习题；
- 一般情况一组4人以内，组团自愿；
- 包的选择可以检索官方网站，也可以搜索。

Available Packages

Currently, the CRAN package repository features 10338 available packages.

Table of available packages, sorted by date of publication

Table of available packages, sorted by name

Installation of Packages

Please type `help("INSTALL")` or `help("install.packages")` in R for information on how to install packages fi

CRAN Task Views allow you to browse packages by topic and provide tools to automatically install all packages

Package Check Results

All packages are tested regularly on machines running Debian GNU/Linux, Fedora, OS X, Solaris and Windows

The results are summarized in the check summary (some timings are also available). Additional details for Wind

Writing Your Own Packages

The manual Writing R Extensions (also contained in the R base sources) explains how to write new packages an

Repository Policies

The manual CRAN Repository Policy [PDF] describes the policies in place for the CRAN package repository.

https://cran.r-project.org/web/views/

Bayesian	Bayesian Inference
ChemPhys	Chemometrics and Computational Physics
ClinicalTrials	Clinical Trial Design, Monitoring, and Analysis
Cluster	Cluster Analysis & Finite Mixture Models
DifferentialEquations	Differential Equations
Distributions	Probability Distributions
Econometrics	Econometrics
Environmetrics	Analysis of Ecological and Environmental Data
ExperimentalDesign	Design of Experiments (DoE) & Analysis of Experimental Data
ExtremeValue	Extreme Value Analysis
Finance	Empirical Finance
Genetics	Statistical Genetics
Graphics	Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization
HighPerformanceComputing	High Performance and Parallel Computing with R
MachineLearning	Machine Learning & Statistical Learning
MedicalImaging	Medical Image Analysis
MetaAnalysis	Meta-Analysis
Multivariate	Multivariate Statistics
NaturalLanguageProcessing	Natural Language Processing
NumericalMathematics	Numerical Mathematics
OfficialStatistics	Official Statistics & Survey Methodology
Optimization	Optimization and Mathematical Programming
Pharmacokinetics	Analysis of Pharmacokinetic Data
Phylogenetics	Phylogenetics, Especially Comparative Methods
Psychometrics	Psychometric Models and Methods
ReproducibleResearch	Reproducible Research
Robust	Robust Statistical Methods
SocialSciences	Statistics for the Social Sciences
Spatial	Analysis of Spatial Data
SpatioTemporal	Handling and Analyzing Spatio-Temporal Data
Survival	Survival Analysis
TimeSeries	Time Series Analysis
WebTechnologies	Web Technologies and Services
gR	gRaphical Models in R

https://cran.r-project.org/web/packages/

谢谢！

孙惠平

sunhp@ss.pku.edu.cn