数据分析、展现与R语言 第11周

2013.04.10

法律声明

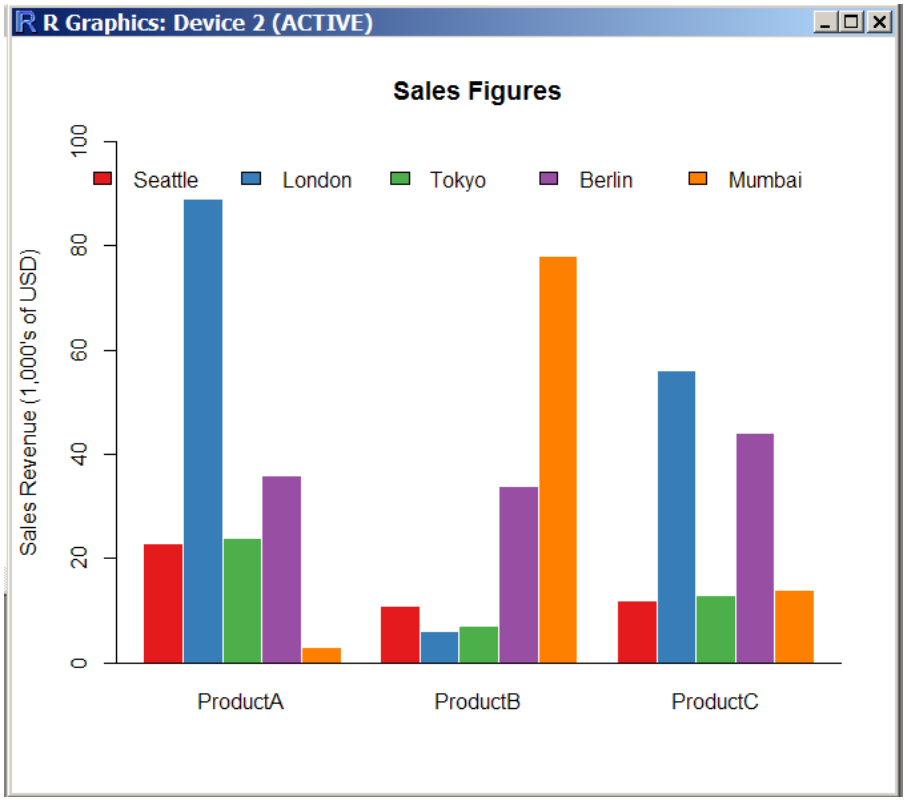
【声明】本视频和幻灯片为炼数成金网络课程的教学资料，所有资料只能在课程内使用，不得在课程以外范围散播，违者将可能被追究法律和经济责任。

课程详情访问炼数成金培训网站

http://edu.dataguru.cn

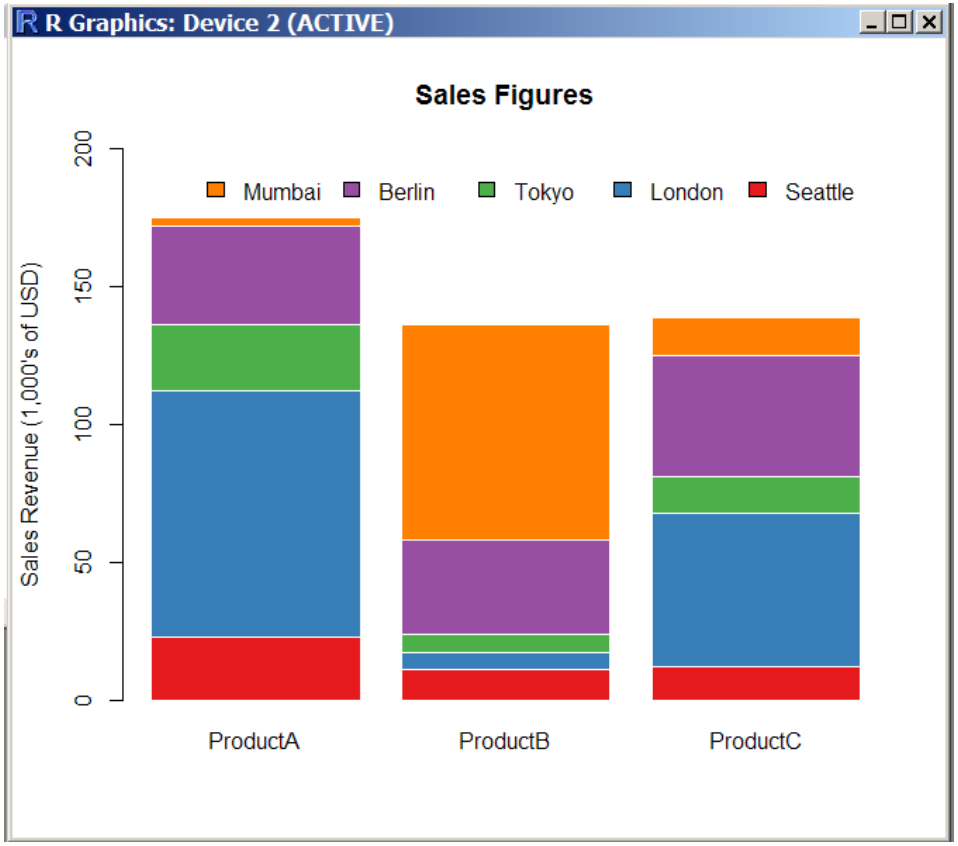
柱形图

```
install.packages("RColorBrewer") #if not already
      installed
library(RColorBrewer)


citysales<-read.csv("citysales.csv")
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")
box(bty="l")
```



堆叠效果

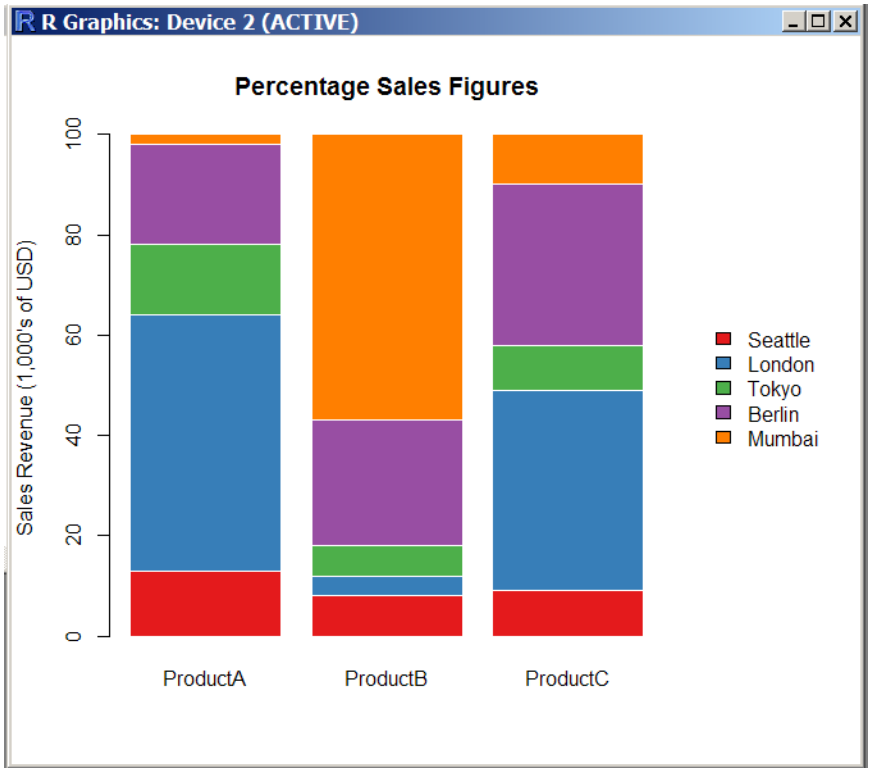
```
citysales<-read.csv("citysales.csv")
barplot(as.matrix(citysales[,2:4]),
legend.text=citysales$City,
args.legend=list(bty="n",horiz=TR
    UE),
col=brewer.pal(5,"Set1"),border="
    white",
ylim=c(0,200),ylab="Sales Revenue
    (1,000's of USD)",
main="Sales Figures")
```



用堆叠效果展示百分比

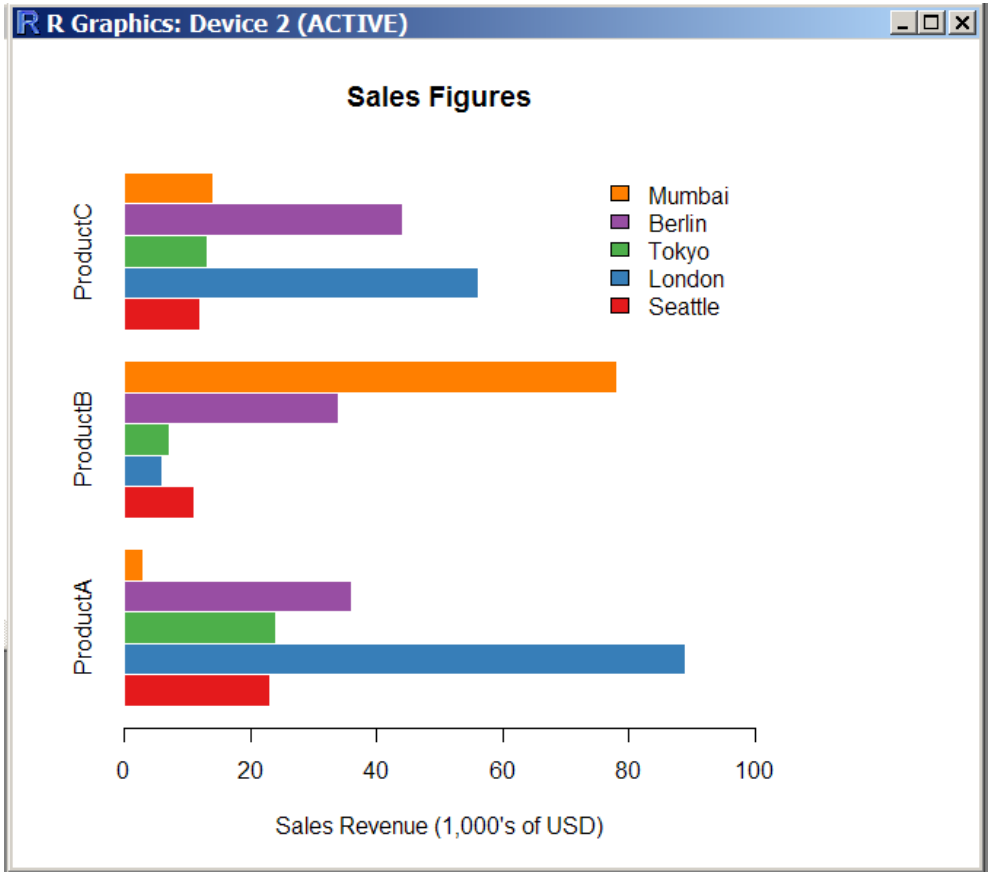
使用第五章数据

```
citysalesperc<-
    read.csv("citysalesperc.csv")
par(mar=c(5,4,4,8),xpd=T)
barplot(as.matrix(citysalesperc[,2:4]),
col=brewer.pal(5,"Set1"),border="white",
ylab="Sales Revenue (1,000's of USD)",
main="Percentage Sales Figures")
legend("right",legend=citysalesperc$City,
    bty="n",
inset=c(-0.3,0),fill=brewer.pal(5,"Set1"))
```



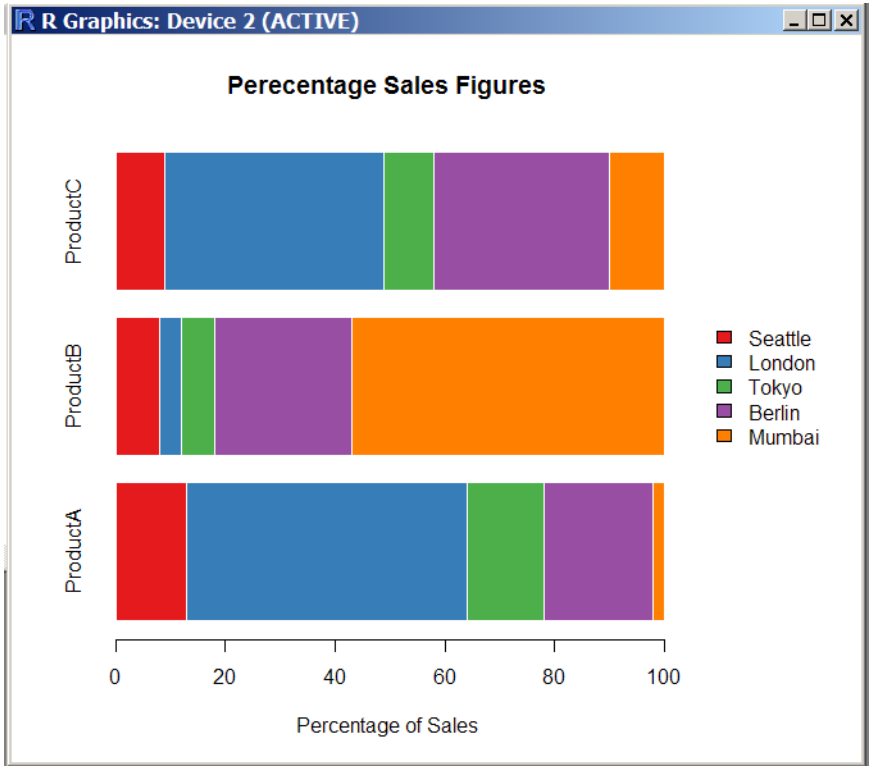
水平方向的柱形图

```
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="w
    hite",
xlim=c(0,100), xlab="Sales Revenue
    (1,000's of USD)",
main="Sales Figures")
```



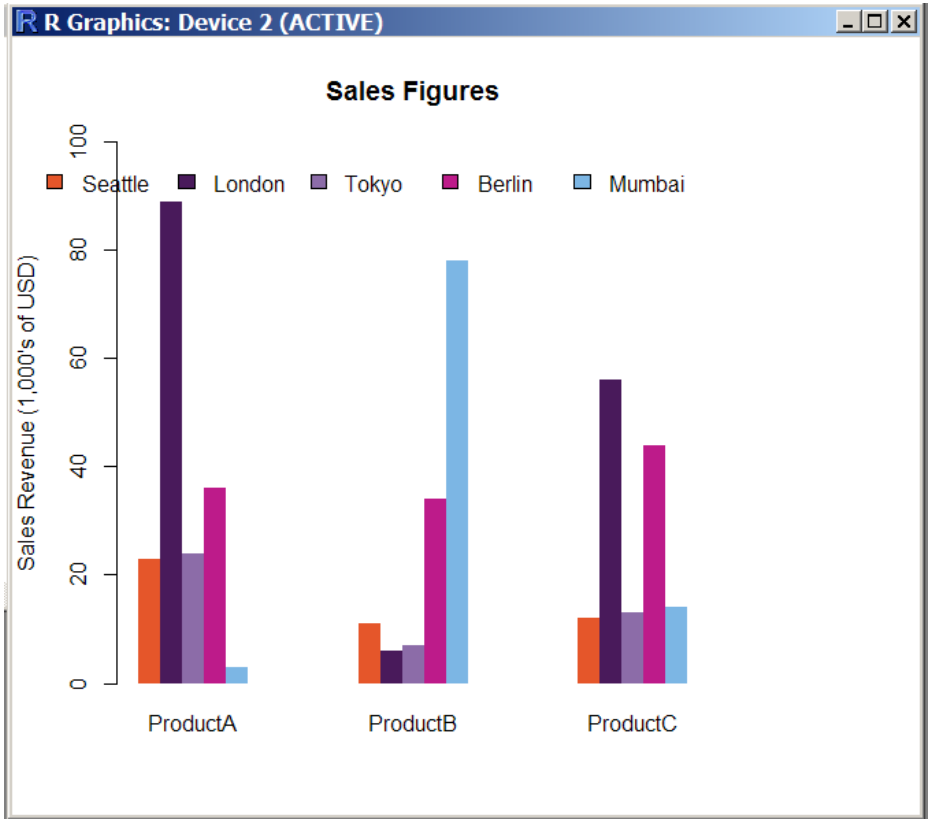
展示百分比的堆叠水平方向柱形图

```
par(mar=c(5,4,4,8),xpd=T)
barplot(as.matrix(citysalesperc[,2:4]),
    horiz=TRUE,
col=brewer.pal(5,"Set1"),border="whi
    te",
xlab="Percentage of Sales",
main="Perecentage Sales Figures")
legend("right",legend=citysalesperc$
    City,bty="n",
inset=c(-
    0.3,0),fill=brewer.pal(5,"Set1"))
```



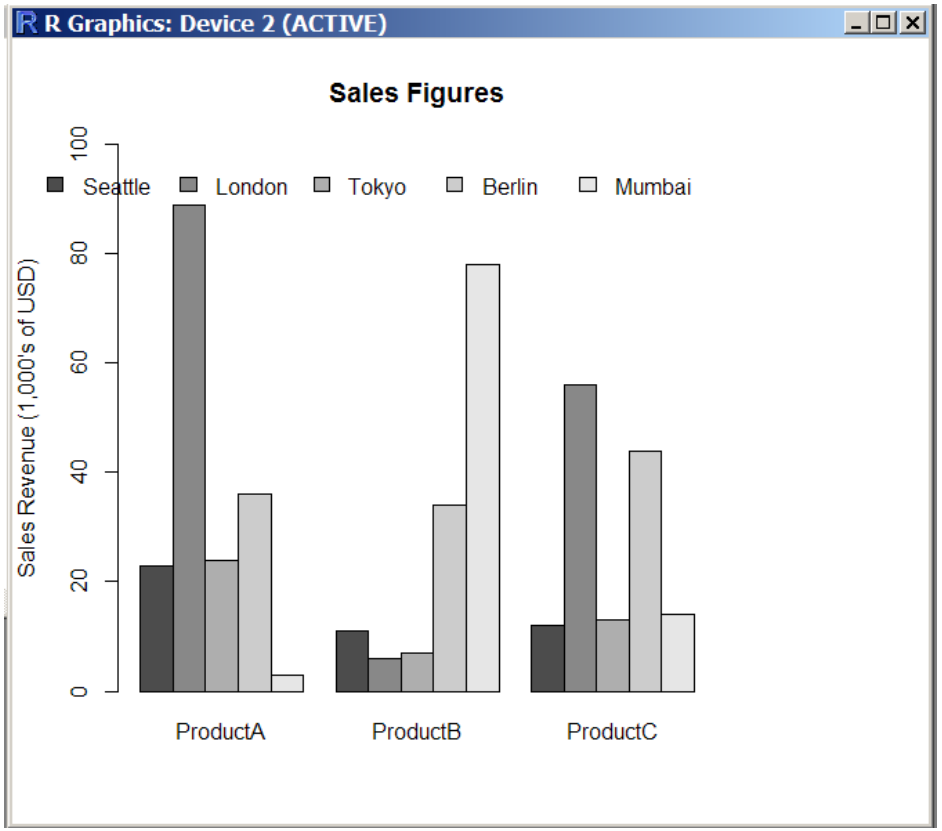
调整柱形图的宽度，间隔和颜色

```
barplot(as.matrix(citysales[,2:4]),
    beside=TRUE,
legend.text=citysales$City,
    args.legend=list(bty="n",horiz=
    T),
col=c("#E5562A","#491A5B","#8C6
    CA8","#BD1B8A","#7CB6E4"),
border=FALSE,space=c(0,5),
ylim=c(0,100),ylab="Sales Revenue
    (1,000's of USD)",
main="Sales Figures")
```



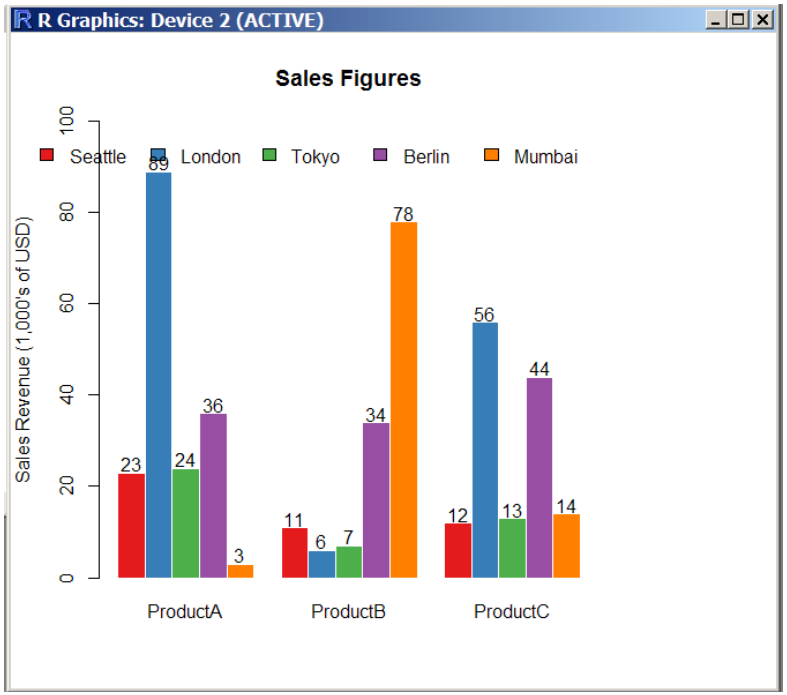
效果对比

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



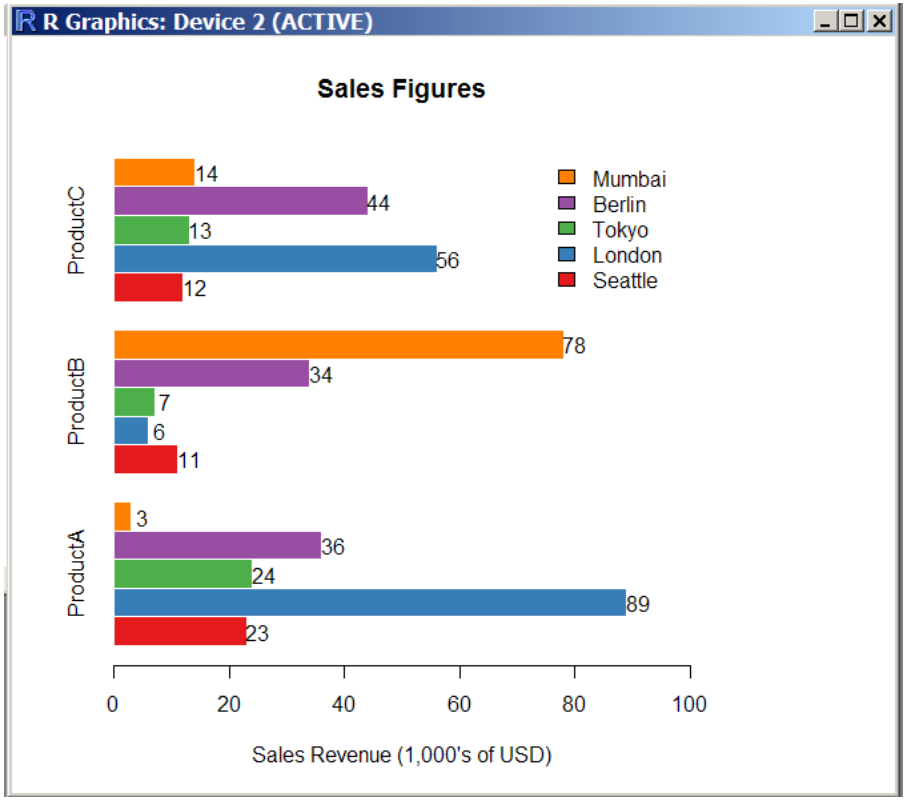
在柱子顶端显示数据

```
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))
```



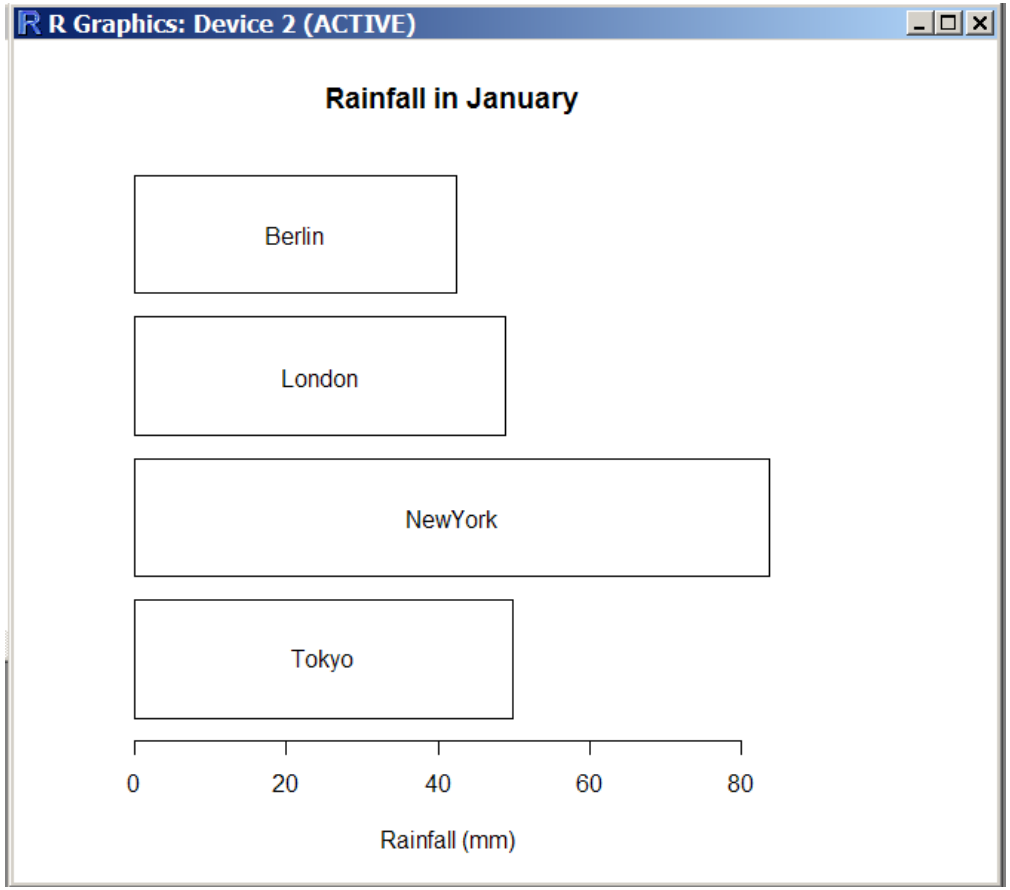
水平柱子旁标注数据

```
y<-barplot(as.matrix(citysales[,2:4]),
    beside=TRUE,horiz=TRUE,
legend.text=citysales$City,args.lege
    nd=list(bty="n"),
col=brewer.pal(5,"Set1"),border="w
    hite",
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))
```



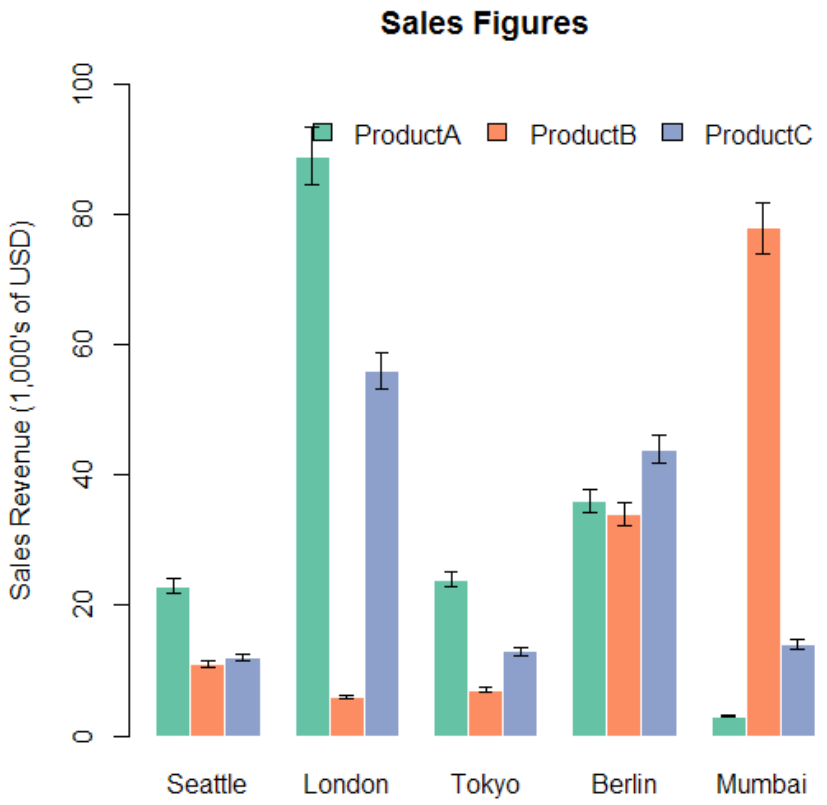
在柱子里面进行标注

```
rain<-read.csv("cityrain.csv")
y<-barplot(as.matrix(rain[1,-
    1]),horiz=T,col="white",
yaxt="n",main=" Rainfall in
    January",xlab="Rainfall
    (mm)")
x<-0.5*rain[1,-1]
text(x,y,colnames(rain[-1]))
```



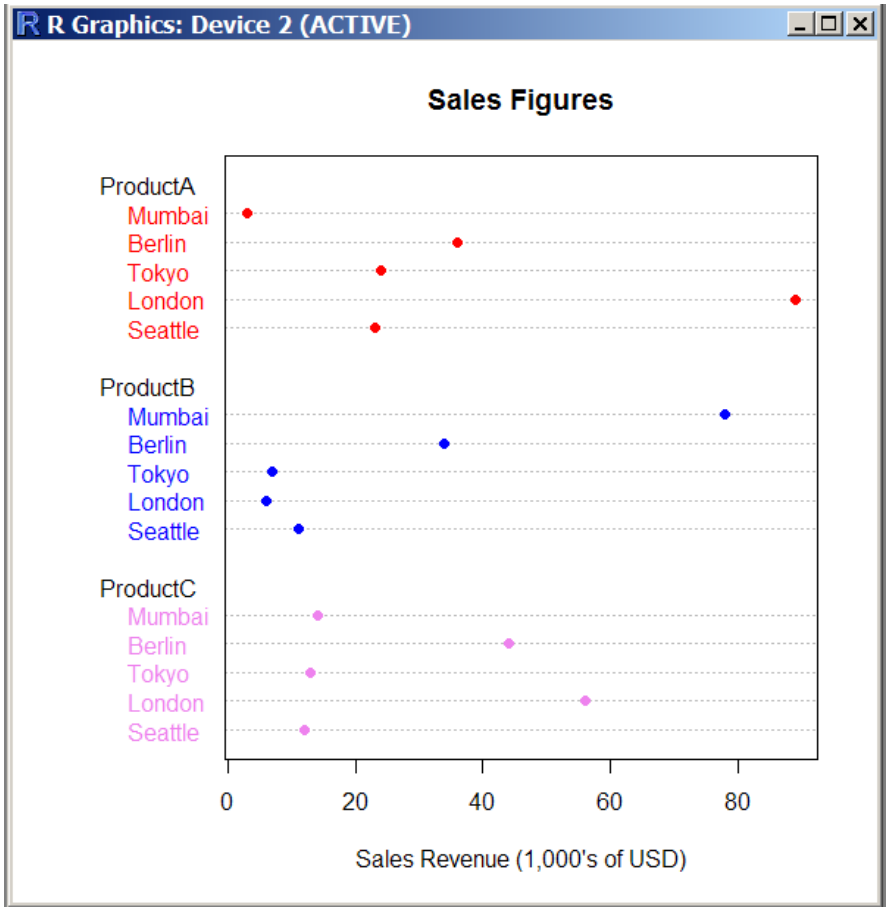
标注误差

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

x<-
    barplot(sales,beside=T,legend.text=rownames(sal
    es),
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)
```



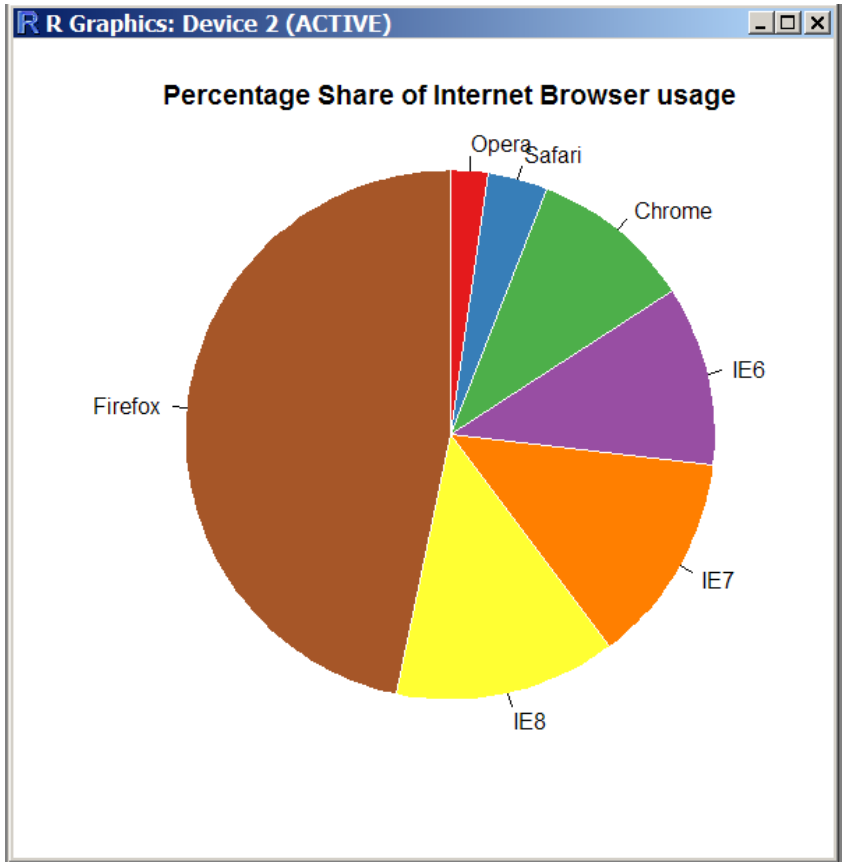
点图

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

sales<-melt(citysales)
sales$color[sales[,2]=="ProductA"] <- "red"
sales$color[sales[,2]=="ProductB"] <- "blue"
sales$color[sales[,2]=="ProductC"] <- "violet"
dotchart(sales[,3],labels=sales$City,groups=sales[,2],
col=sales$color,pch=19,
main="Sales Figures",
xlab="Sales Revenue (1,000's of USD)")
```



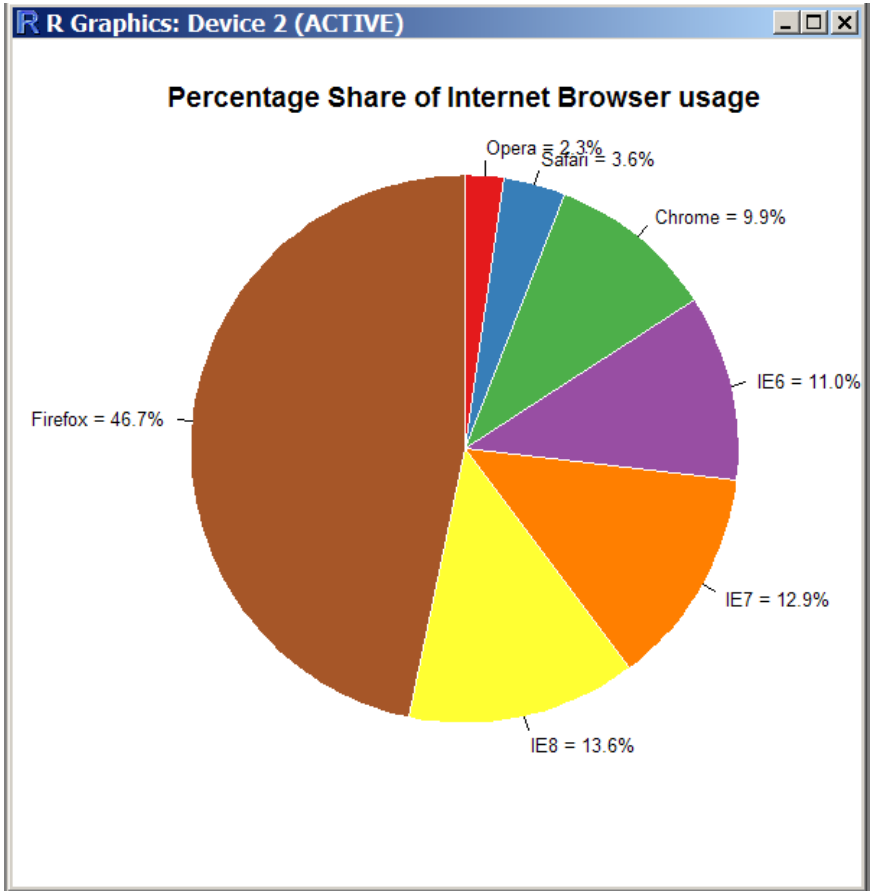
饼图

```
browsers<-
    read.table("browsers.txt",header=TRU
    E)
browsers<-browsers[order(browsers[,2]),]
pie(browsers[,2],
labels=browsers[,1],
clockwise=TRUE,
radius=1,
col=brewer.pal(7,"Set1"),
border="white",
main="Percentage Share of Internet
    Browser usage")
```



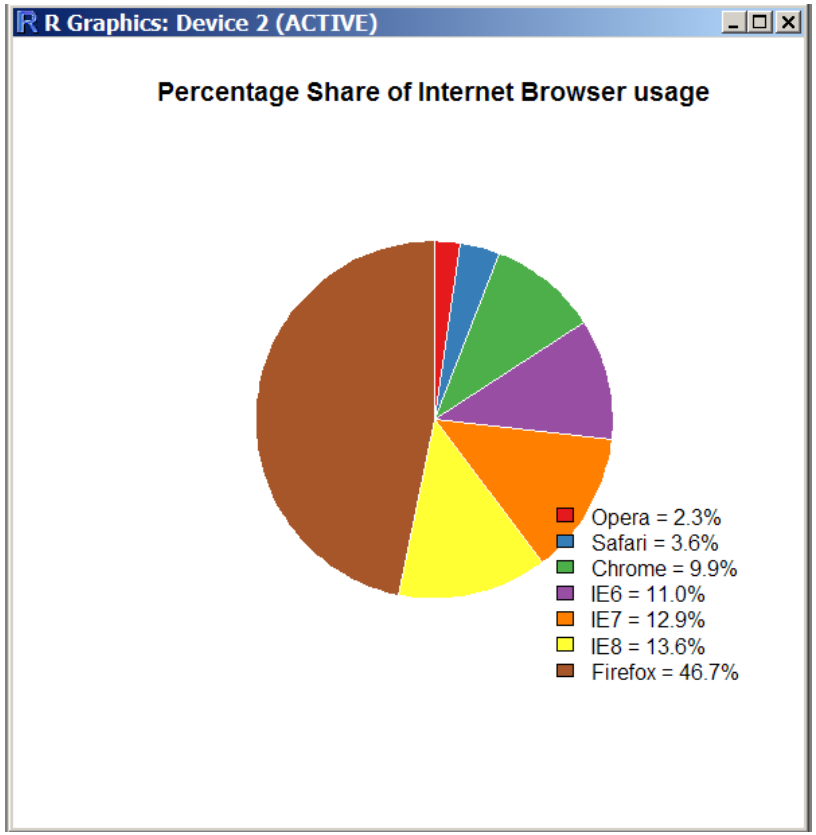
在饼图上标注百分比

```
browsers<-read.table("browsers.txt",header=TRUE)
browsers<-browsers[order(browsers[,2]),]
pielabels <- sprintf("%s = %3.1f%s", browsers[,1],
100*browsers[,2]/sum(browsers[,2]), "%")
pie(browsers[,2],
labels=pielabels,
clockwise=TRUE,
radius=1,
col=brewer.pal(7,"Set1"),
border="white",
cex=0.8,
main="Percentage Share of Internet Browser usage")
```



增加图释

```
browsers<-read.table("browsers.txt",header=TRUE)
browsers<-browsers[order(browsers[,2]),]
pielabels <- sprintf("%s = %3.1f%s", browsers[,1],
100*browsers[,2]/sum(browsers[,2]), "%")
pie(browsers[,2],
labels=NA,
clockwise=TRUE,
col=brewer.pal(7,"Set1"),
border="white",
radius=0.7,
cex=0.8,
main="Percentage Share of Internet Browser usage")
legend("bottomright",legend=pielabels,bty="n",
fill=brewer.pal(7,"Set1"))
```

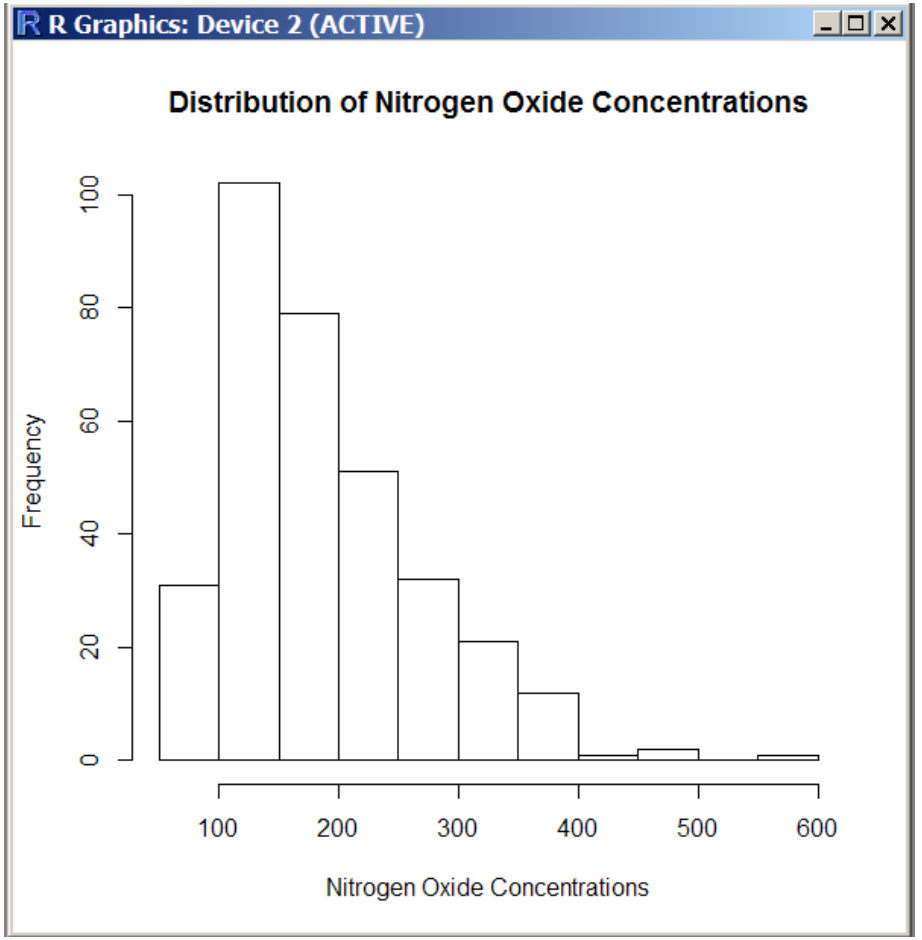


直方图

使用第六章数据

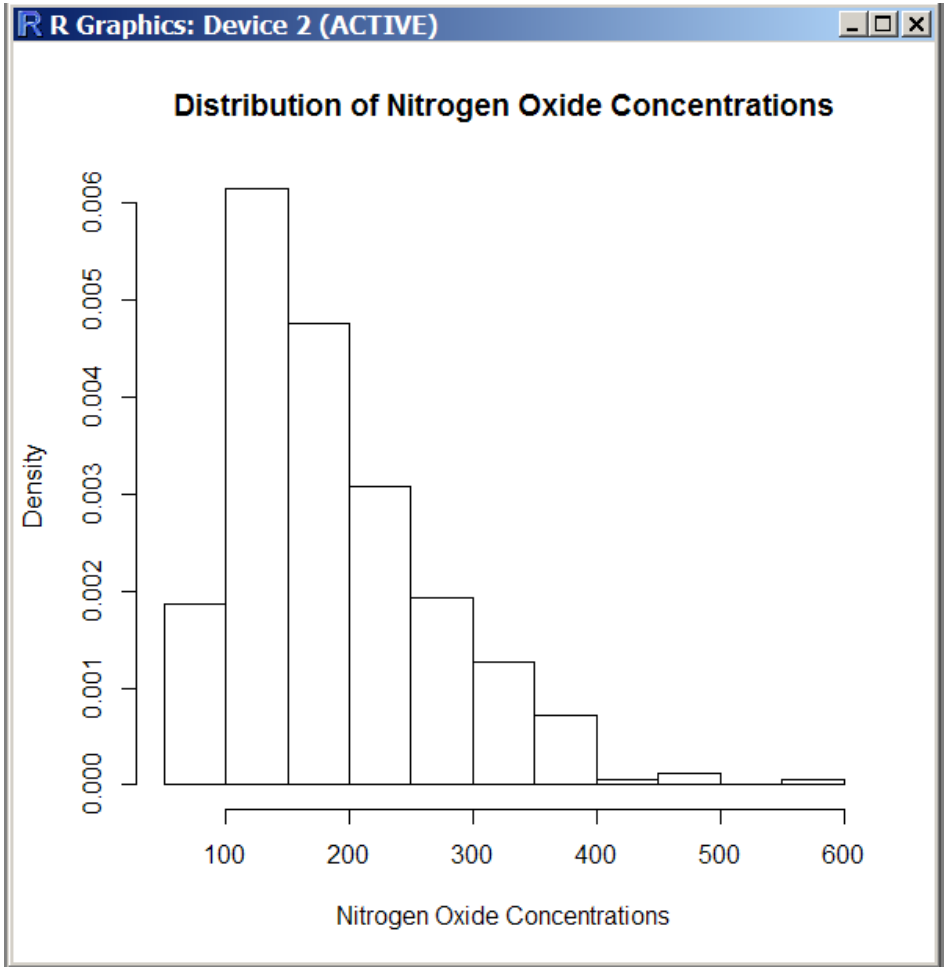
```
air<-read.csv("airpollution.csv")
```

```
hist(air$Nitrogen.Oxides,
xlab="Nitrogen Oxide
    Concentrations",
main="Distribution of Nitrogen
    Oxide Concentrations")
```



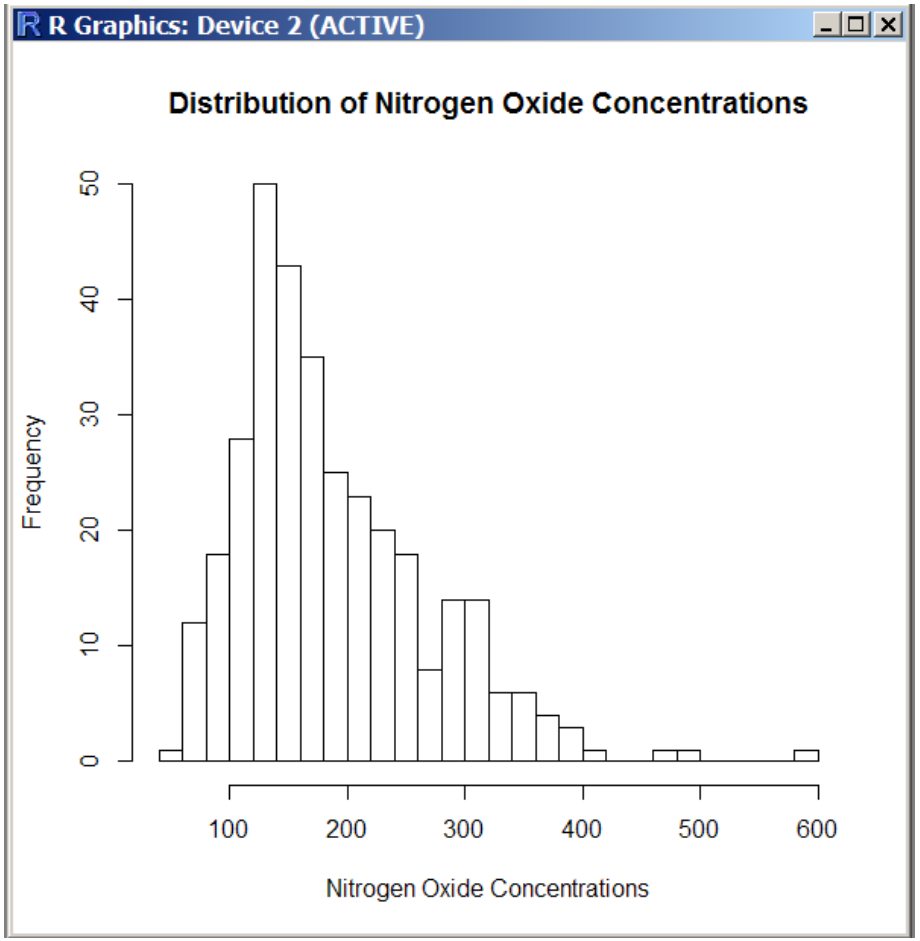
以概率密度显示

```
hist(air$Nitrogen.Oxides,
     freq=FALSE,
xlab="Nitrogen Oxide
     Concentrations",
main="Distribution of Nitrogen
     Oxide Concentrations")
```



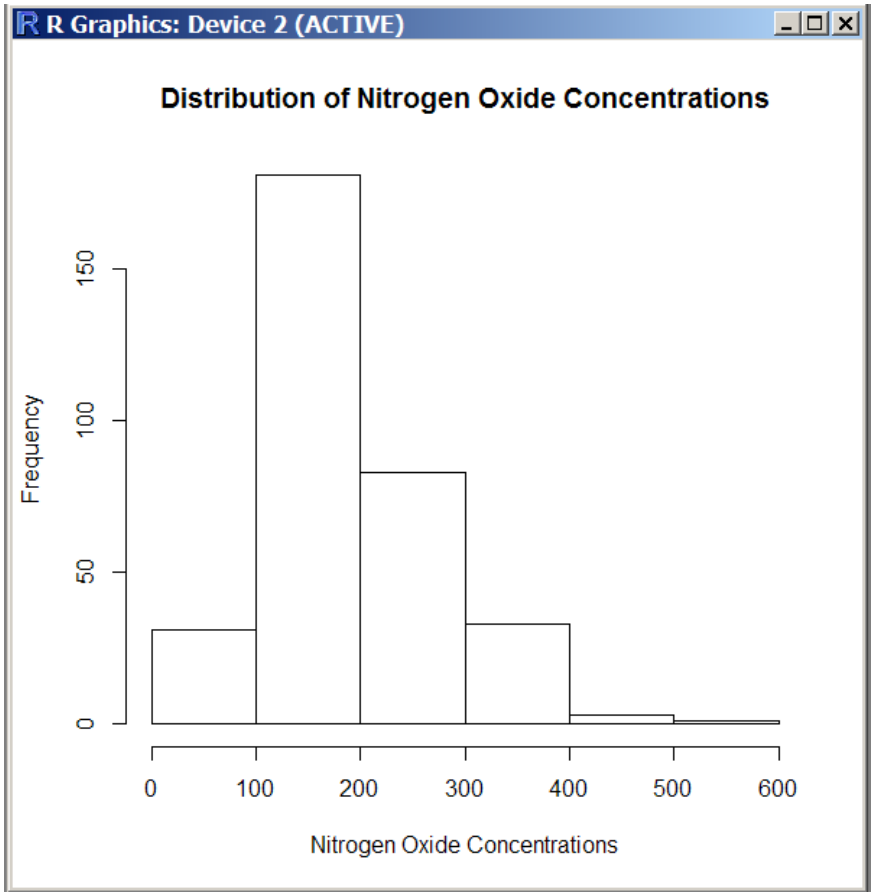
增加breaks

```
hist(air$Nitrogen.Oxides,
breaks=20,xlab="Nitrogen Oxide
    Concentrations",
main="Distribution of Nitrogen
    Oxide Concentrations")
```



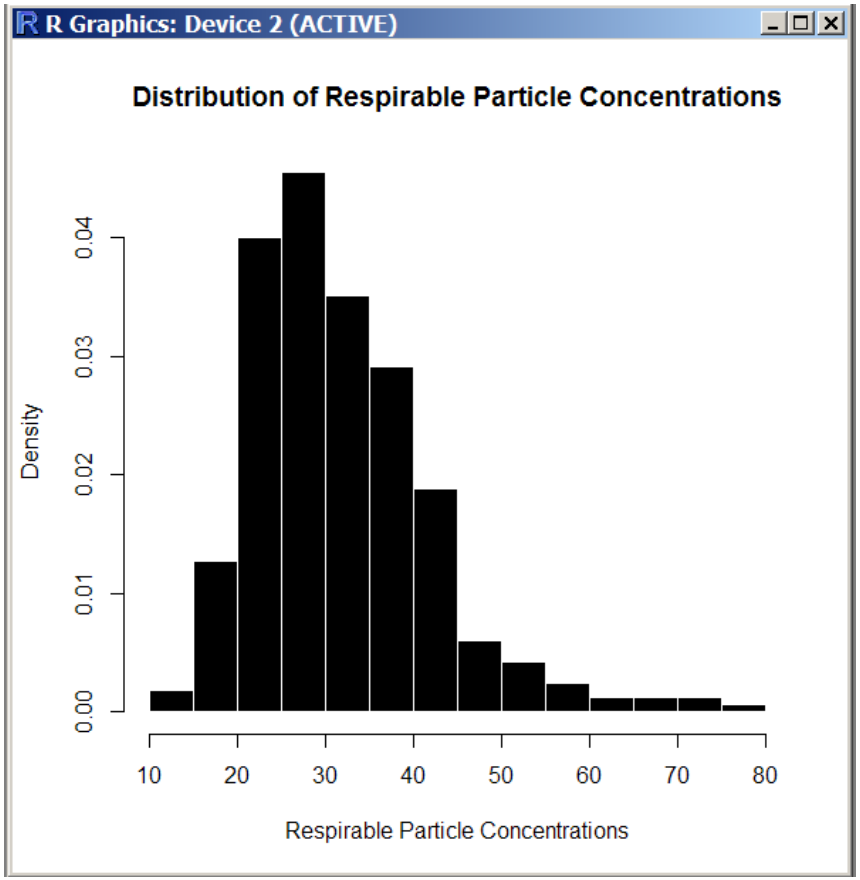
指定breaks范围

```
hist(air$Nitrogen.Oxides,
breaks=c(0,100,200,300,400,500,600),
xlab="Nitrogen Oxide
    Concentrations",
main="Distribution of Nitrogen Oxide
    Concentrations")
```



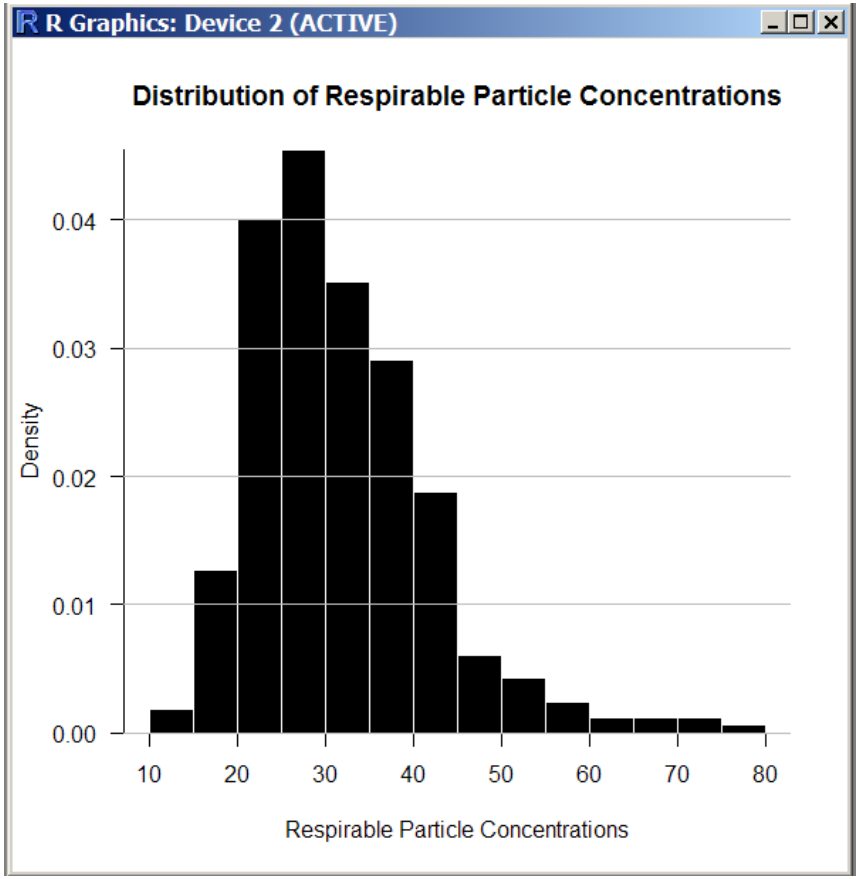
用颜色美化

```
hist(air$Respirable.Particles,
prob=TRUE,col="black",border="wh
    ite",
xlab="Respirable Particle
    Concentrations",
main="Distribution of Respirable
    Particle Concentrations")
```



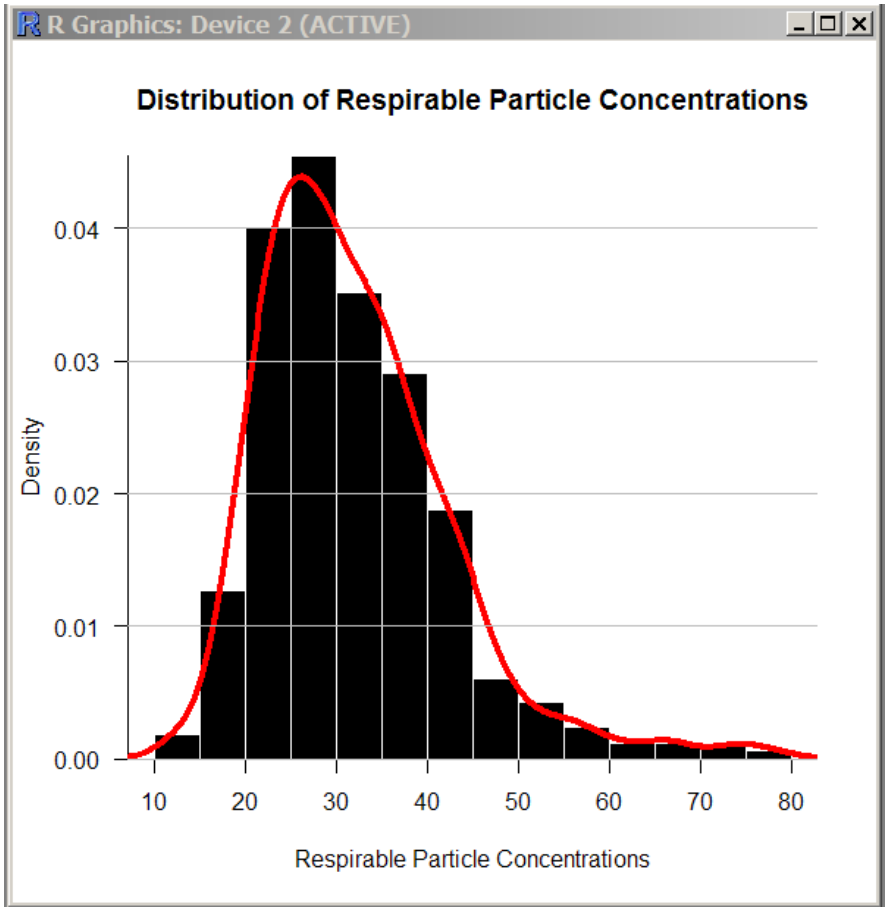
用线条美化

```
par(yaxs="i",las=1)
hist(air$Respirable.Particles,
prob=TRUE,col="black",border="white",
xlab="Respirable Particle Concentrations",
main="Distribution of Respirable Particle Concentrations")
box(bty="l")
grid(nx=NA,ny=NULL,lty=1,lwd=1,col="gray")
```



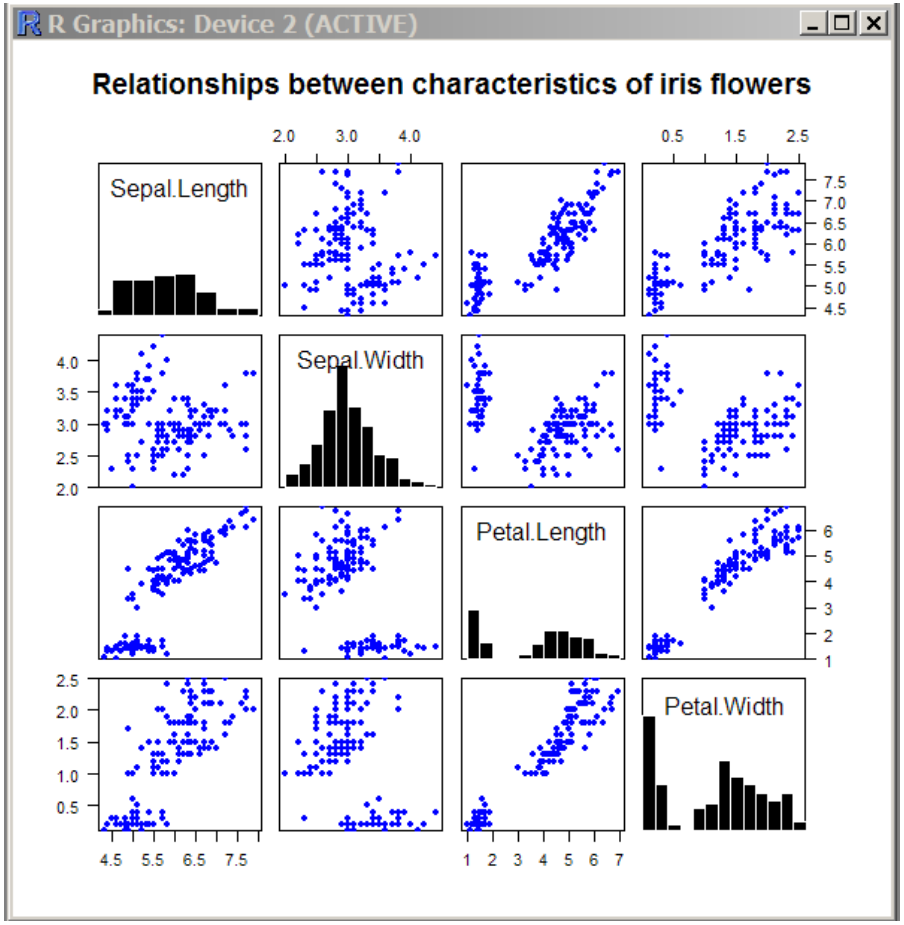
标识密度函数

```
par(yaxs="i",las=1)
hist(air$Respirable.Particles,
prob=TRUE,col="black",border="white",
xlab="Respirable Particle
    Concentrations",
main="Distribution of Respirable Particle
    Concentrations")
box(bty="l")
lines(density(air$Respirable.Particles,na.r
    m=T),col="red",lwd=4)
grid(nx=NA,ny=NULL,lty=1,lwd=1,col="
    gray")
```

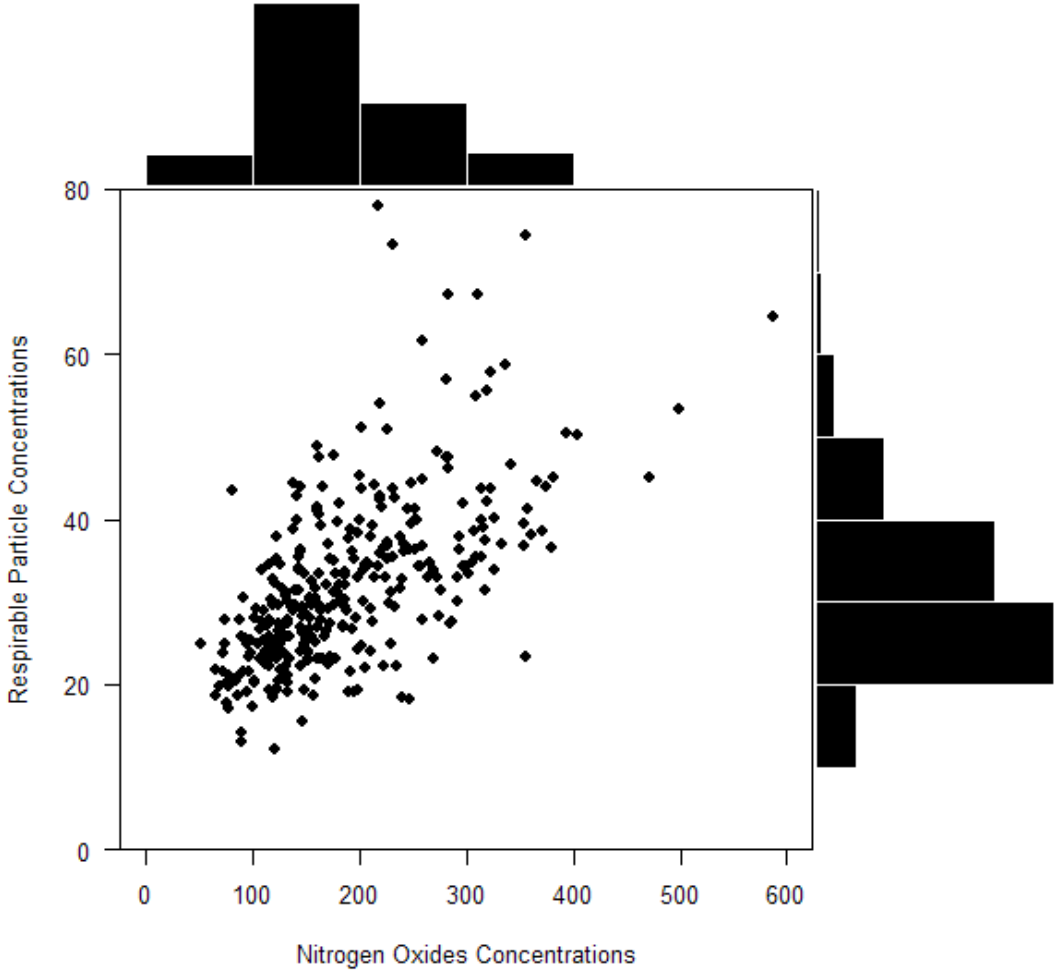


一组直方图

```
panel.hist <- function(x, ...)
{
par(usr = c(par("usr")[1:2], 0, 1.5) )
hist(x,
     prob=TRUE,add=TRUE,col="bla
     ck",border="white")
}
plot(iris[,1:4],
main="Relationships between
     characteristics of iris flowers",
pch=19,col="blue",cex=0.9,
diag.panel=panel.hist)
```



散点图+直方图



2013.04.10

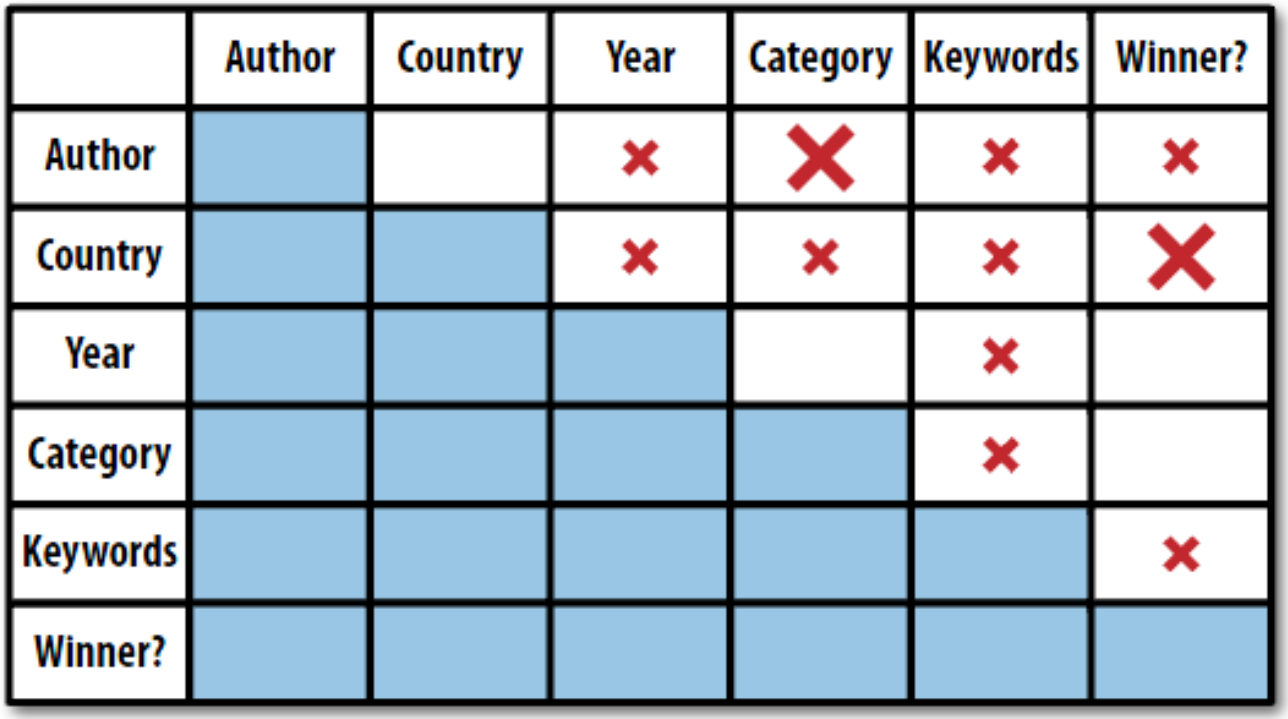
代码

```
#Set up the layout first
layout(matrix(c(2,0,1,3),2,2,byrow=TRUE), widths=c(3,1),
heights=c(1,3), TRUE)
#Make Scatterplot
par(mar=c(5.1,4.1,0.1,0))
plot(air$Respirable.Particles~air$Nitrogen.Oxides,
pch=19,col="black",
xlim=c(0,600),ylim=c(0,80),
xlab="Nitrogen Oxides Concentrations",
ylab="Respirable Particle Concentrations")
#Plot histogram of X variable in the top row
par(mar=c(0,4.1,3,0))
hist(air$Nitrogen.Oxides,
breaks=seq(0,600,100),ann=FALSE,axes=FALSE,
col="black",border="white")
```

代码

```
#Plot histogram of Y variable to the right of the scatterplot
yhist <- hist(air$Respirable.Particles,
breaks=seq(0,80,10),plot=FALSE)
par(mar=c(5.1,0,0.1,1))
barplot(yhist$density,
horiz=TRUE,space=0,axes=FALSE,
col="black",border="white")
```

奥地利电子艺术奖数据可视化分析



	Author	Country	Year	Category	Keywords	Winner?
Author			×	×	×	×
Country			×	×	×	×
Year					×	
Category					×	
Keywords						×
Winner?						

Figure 13-1. *Matrix of initial interest in attribute combinations*

数据情况

Figure 13-2. *First overview of the database contents with dbcounter, a custom nodebox script*

作品年份 vs 作品类别

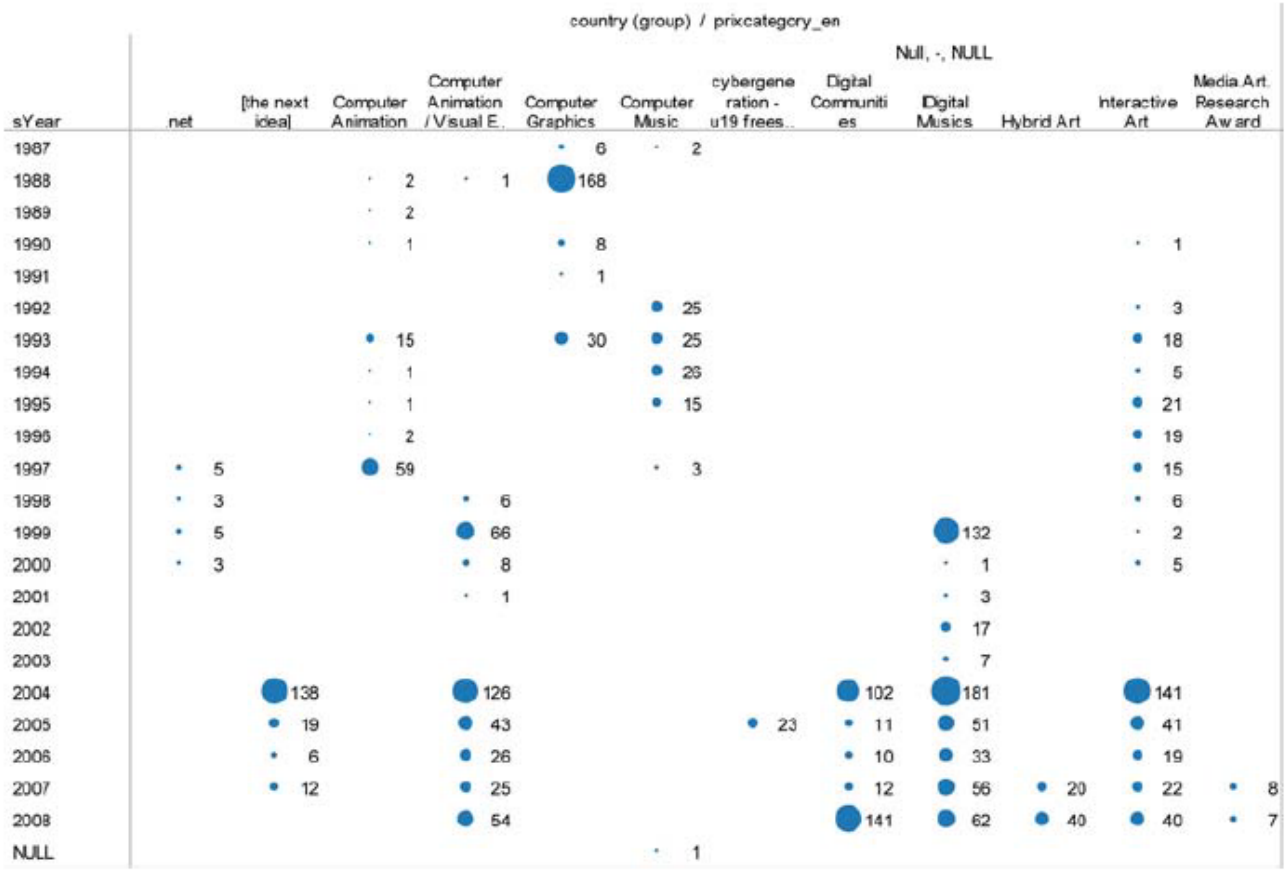


Figure 13-3. *A plot of submissions with missing country information, split up by year and category*

提交机构 vs 作品类别

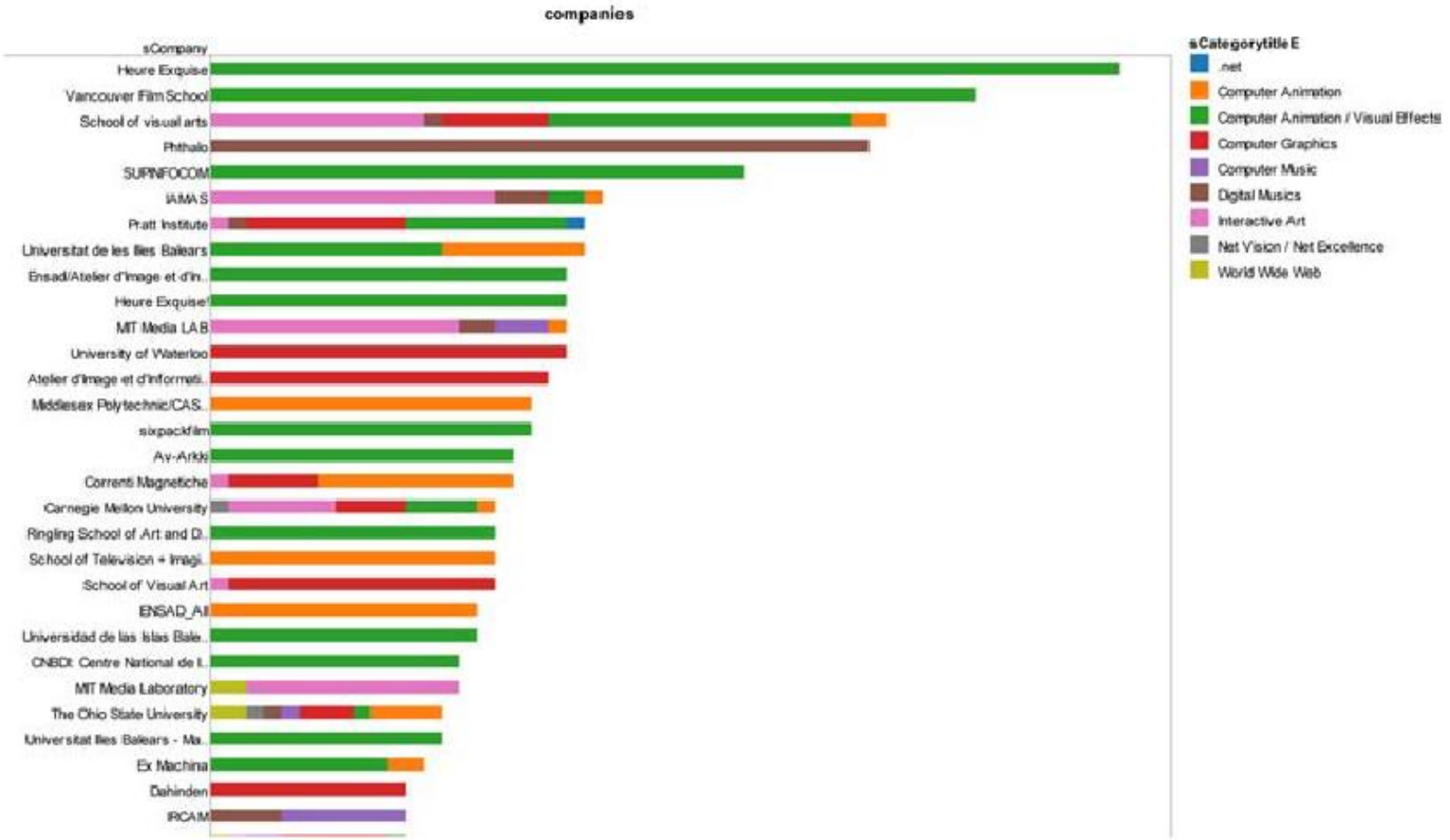


Figure 13-4. *Submissions by company or institution, colored by categories*

地图上的展现

Figure 13-5. *World map with submissions per country, split up by category*

不同国家的获奖情况

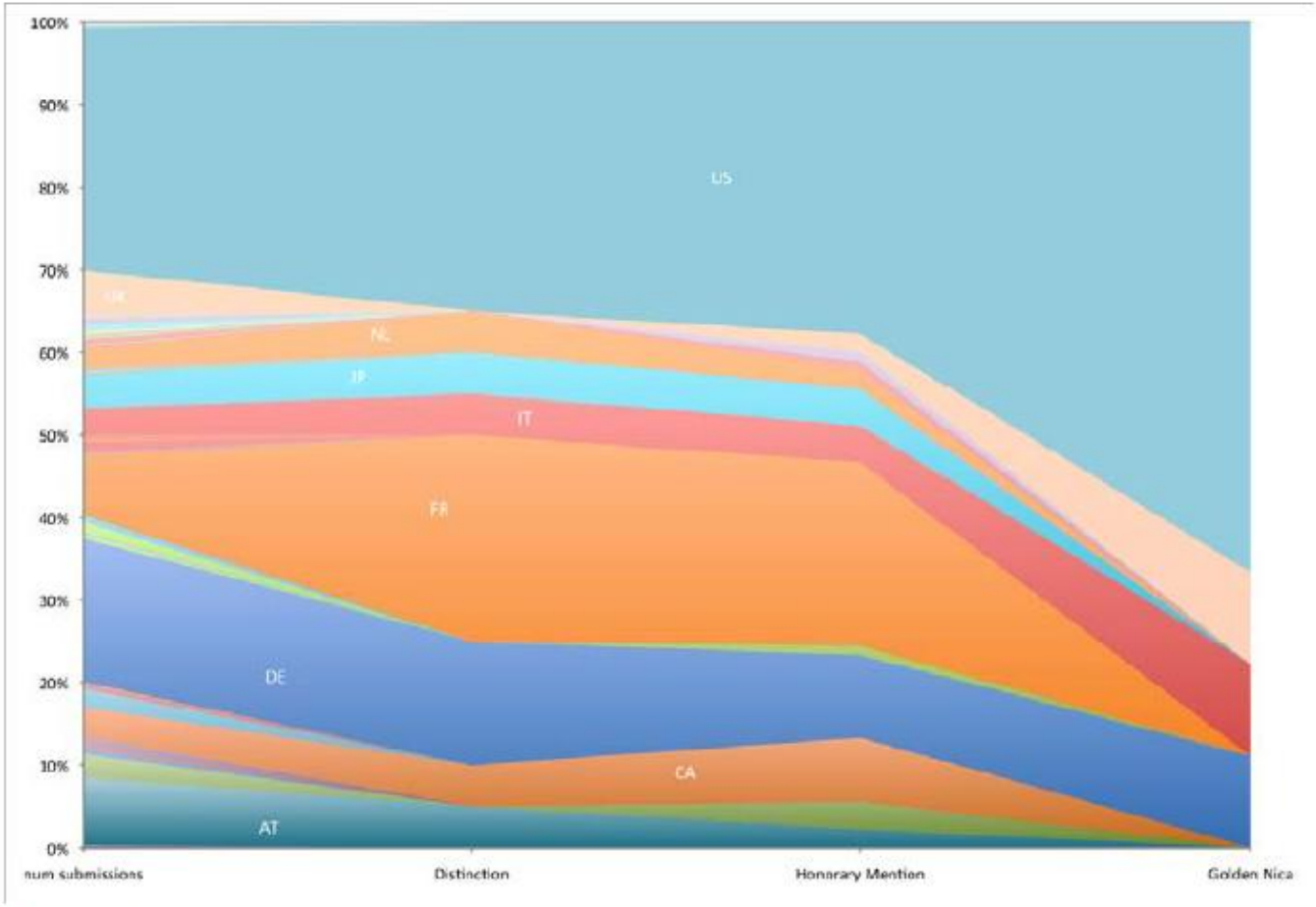


Figure 13-6. *Prizes won by different countries*

2013.04.10

高信息量可视化探索

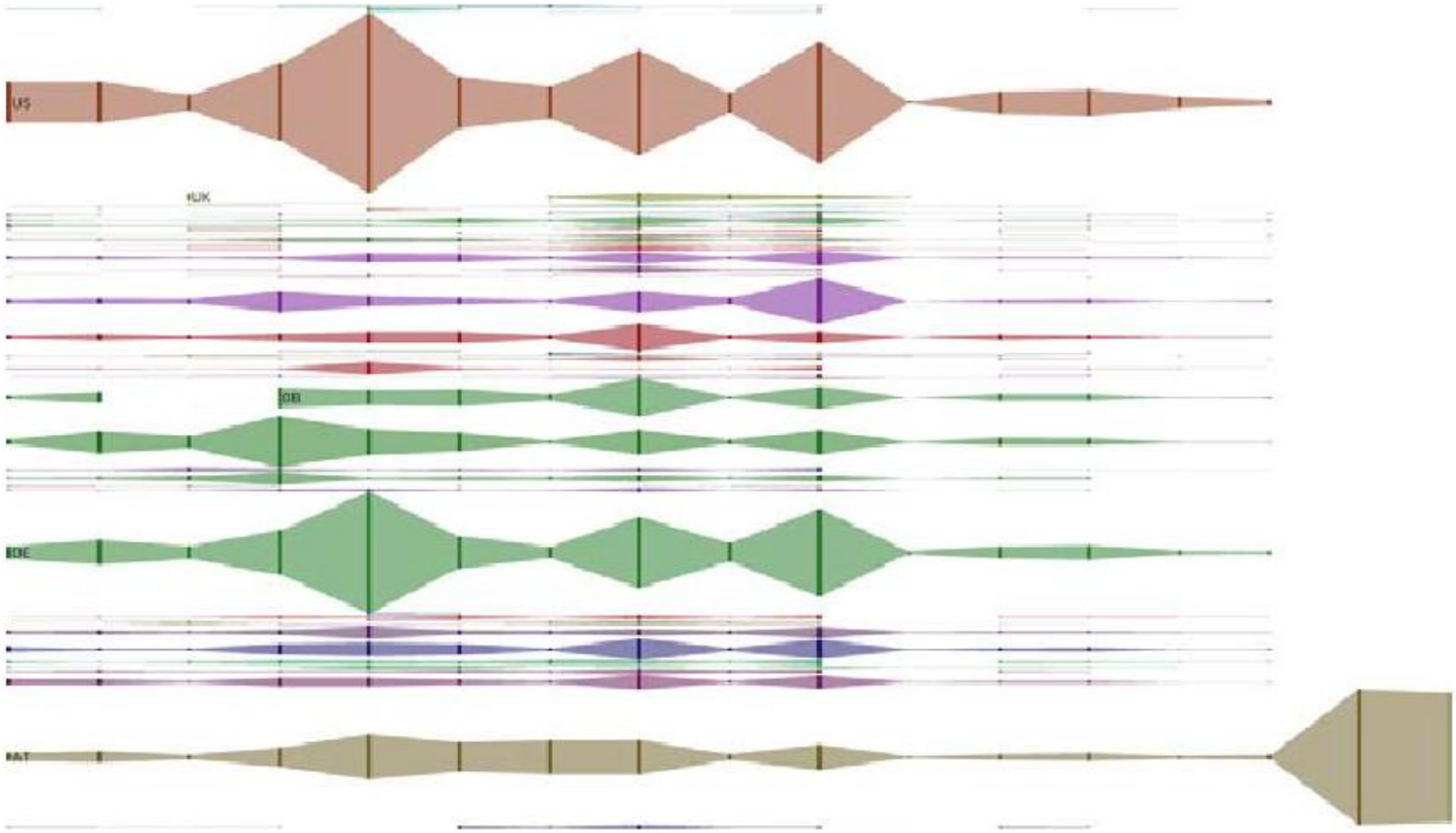


Figure 13-7. *A first attempt at displaying categories by country*

2013.04.10

时间+类别

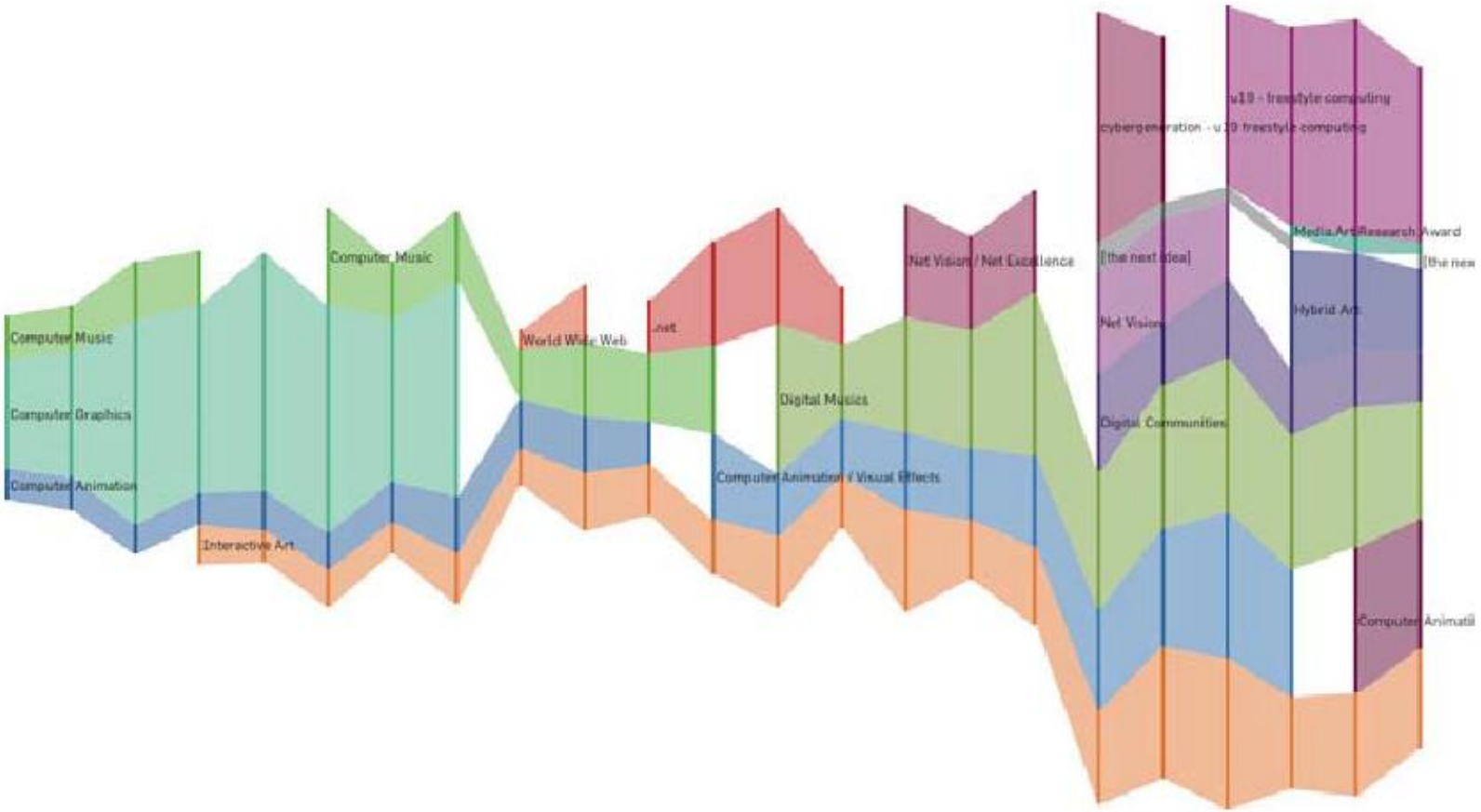


Figure 13-8. *Categories over the years*

像素化显示

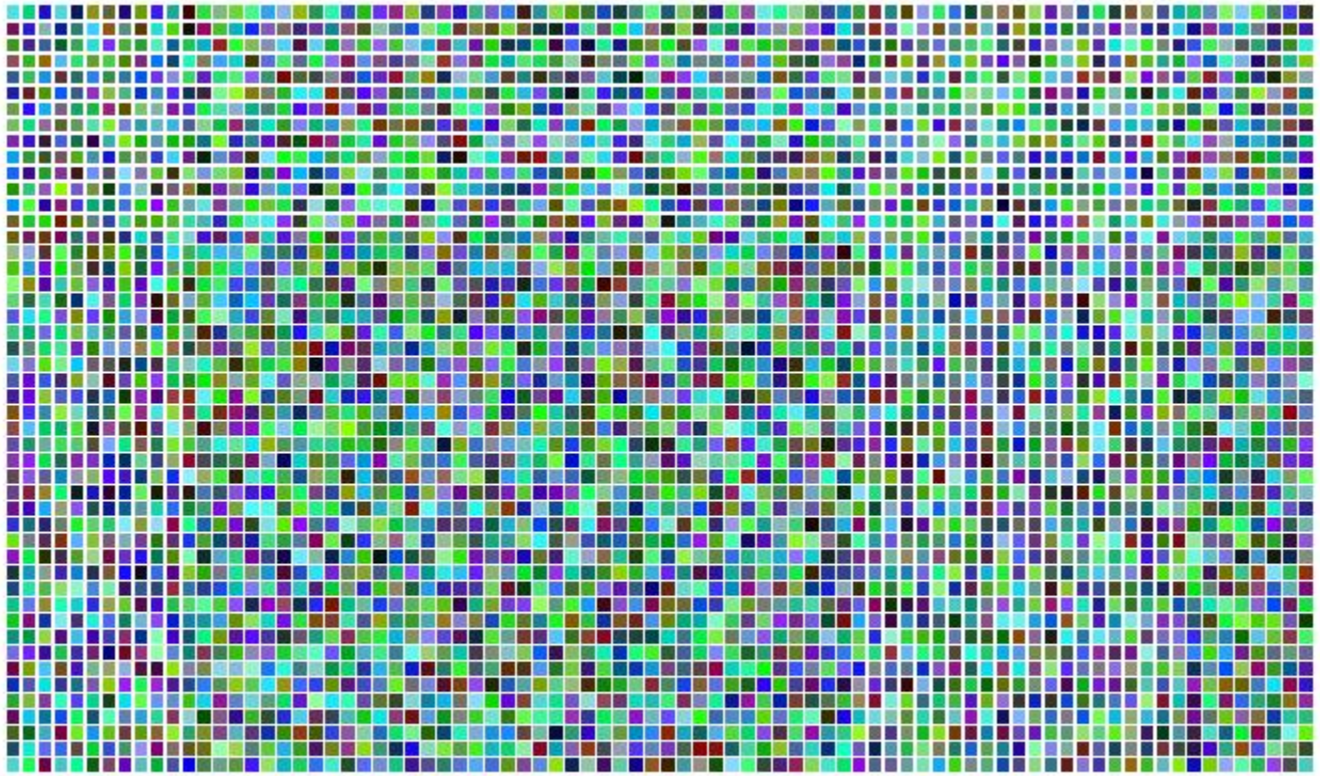


Figure 13-9. *Experimenting with dense pixel displays*

像素化显示

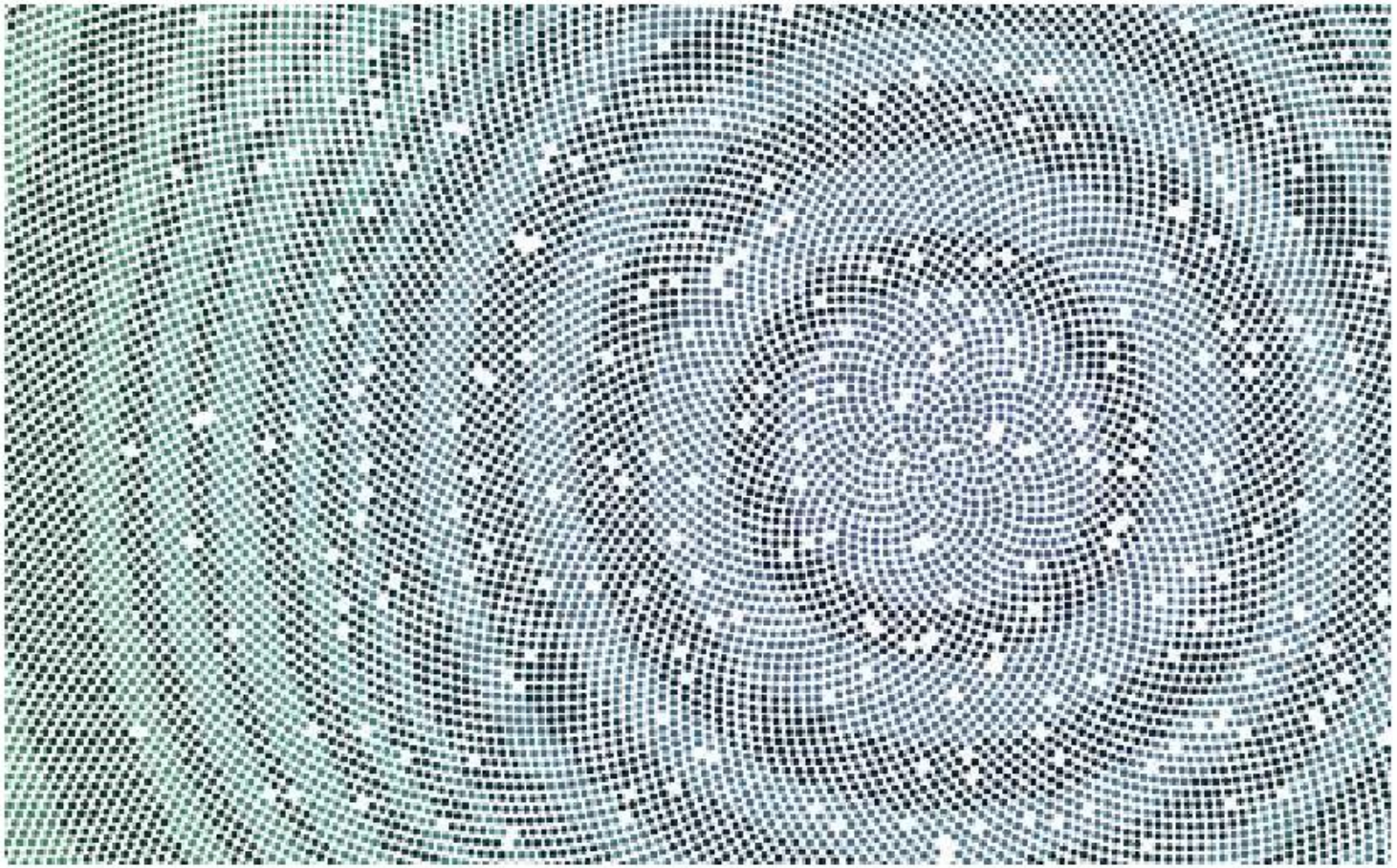


Figure 13-10. *Submissions as dots, packed like sunflower seeds*

2013.04.10

最终作品

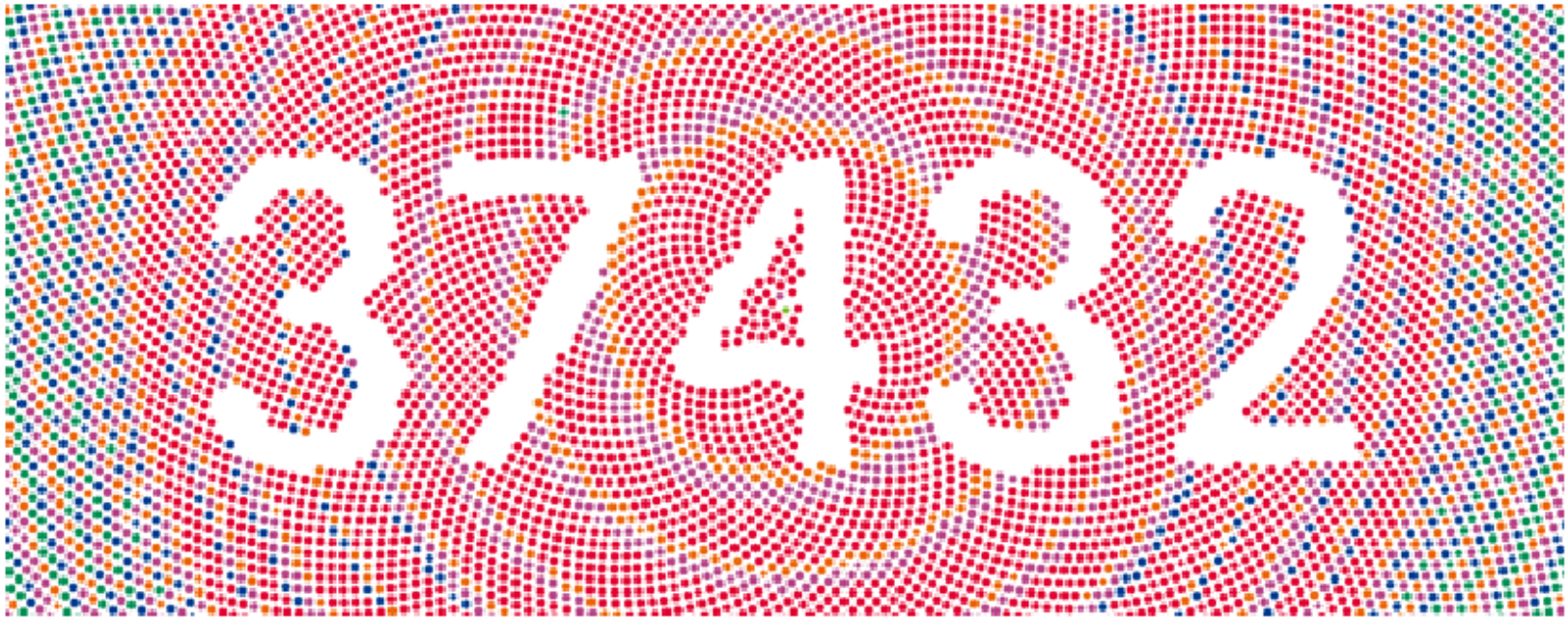
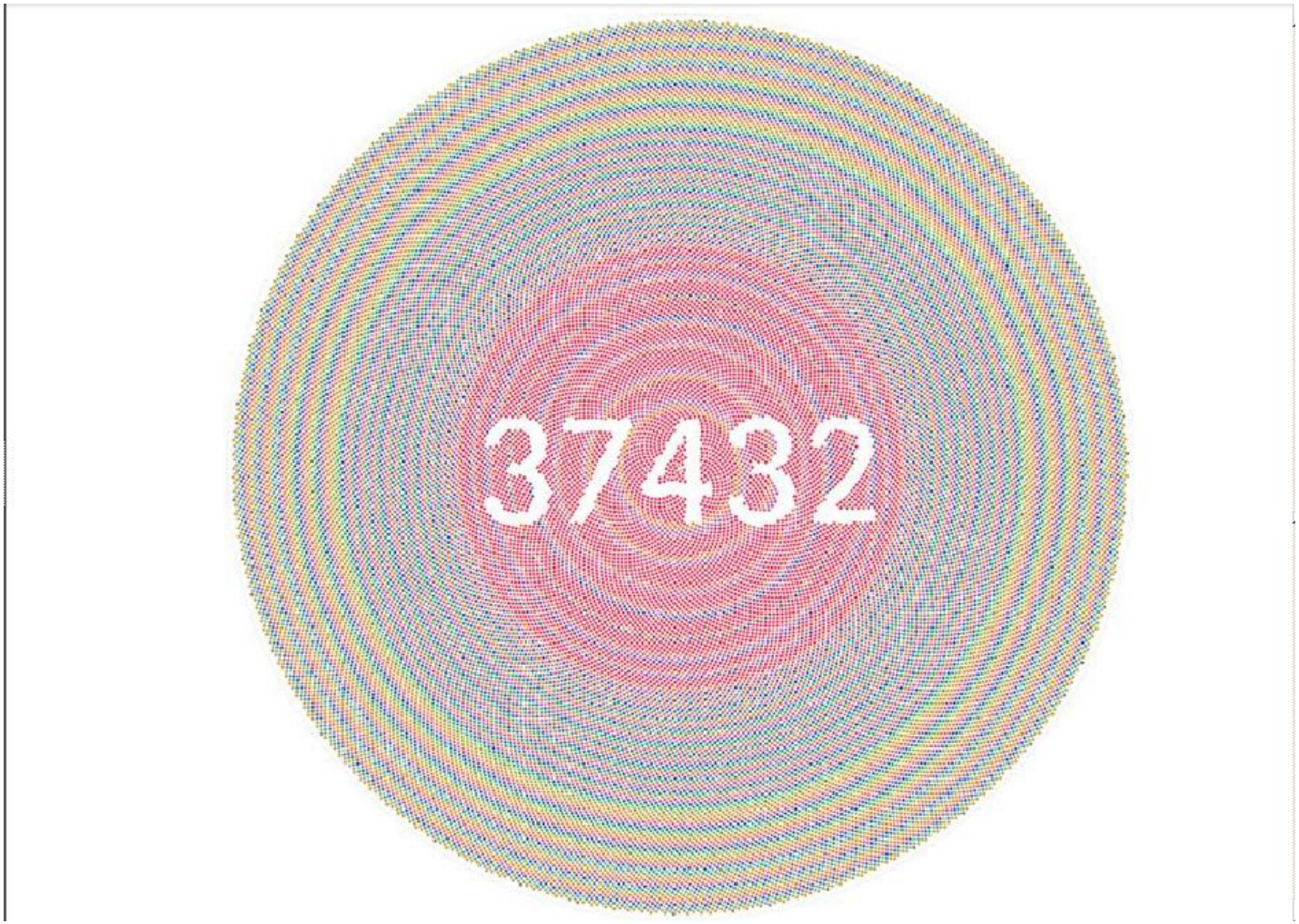


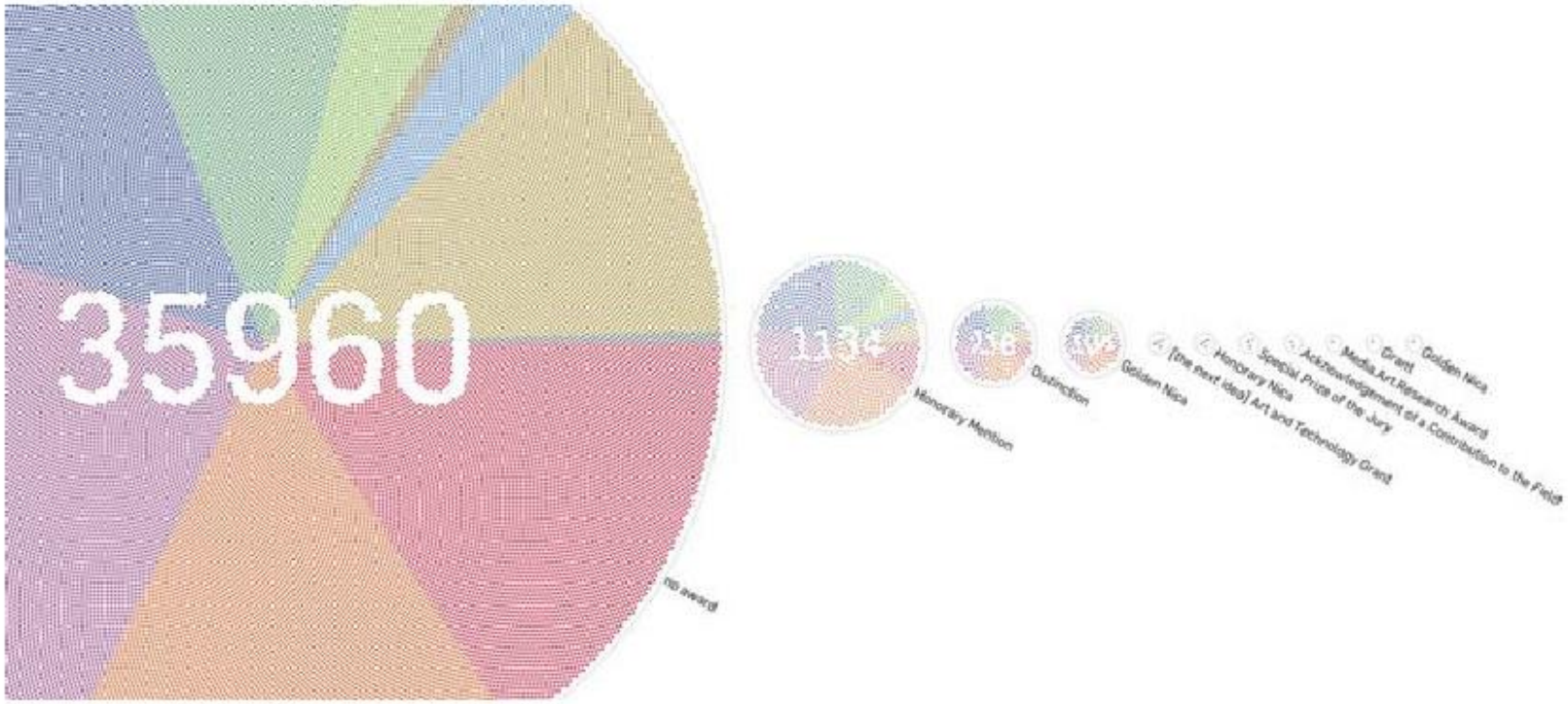
Figure 13-11. *Numbers created by skipping points in the placement sequence*

全部作品



2013.04.10

作品获奖情况


35960
1134
Honorary Mention
Distinction
Golden Nica
[The next idea] Art and Technology Grant
Honorary Nica
Special Prize of the Jury
Acknowledgment of a Contribution to the Field
Grant
Media Art Research Award
Golden Nica

作品所属类别

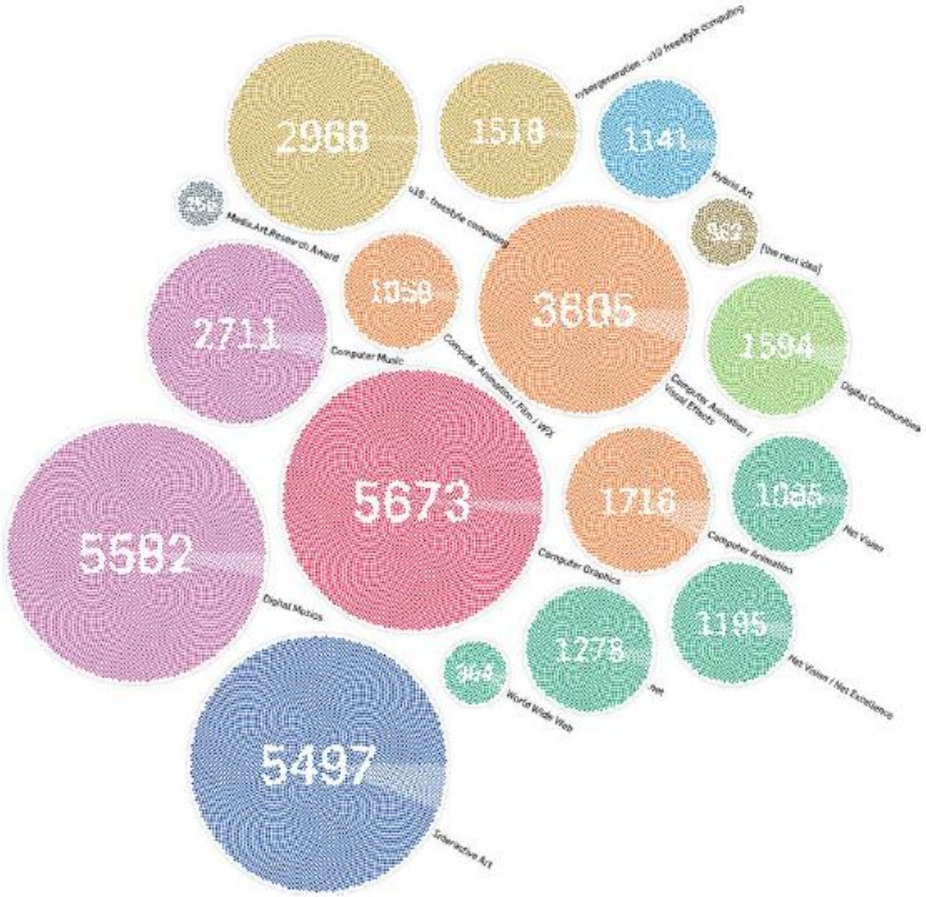


Figure 13-14. *Submissions by category*

2013.04.10

作品所属国家

Figure 13-15. *Submissions by country*

2013.04.10

展出的成品

Figure 13-19. *The poster in the exhibition*

炼数成金逆向收费式网络课程

- Dataguru（炼数成金）是专业数据分析网站，提供教育，媒体，内容，社区，出版，数据分析业务等服务。我们的课程采用新兴的互联网教育形式，独创地发展了逆向收费式网络培训课程模式。既继承传统教育重学习氛围，重竞争压力的特点，同时又发挥互联网的威力打破时空限制，把天南地北志同道合的朋友组织在一起交流学习，使到原先孤立的学习个体组合成有组织的探索力量。并且把原先动辄成千上万的学习成本，直线下降至百元范围，造福大众。我们的目标是：低成本传播高价值知识，构架中国第一的网上知识流转阵地。
- 关于逆向收费式网络的详情，请看我们的培训网站 http://edu.dataguru.cn

Thanks

FAQ时间