Chapter 3 – Data Visualization 数据可视化

Instructor: Zach Zhizhong ZHOU, Shanghai Jiao Tong University

主讲教师：周志中，上海交通大学

ggplot2 Elegant Graphics for Data Analysis

Hadley Wickham

R Graphics Cookbook Winston Chang

安装软件

□下载并安装R：http://mirrors.ustc.edu.cn/CRAN/

□下载并安装R Studio：http://www.rstudio.com/products/rstudio/download/

□安装之后运行R Studio，选择菜单Tools，Install Packages…，然后输入ggplot2，R Studio会自动安装ggplot2。

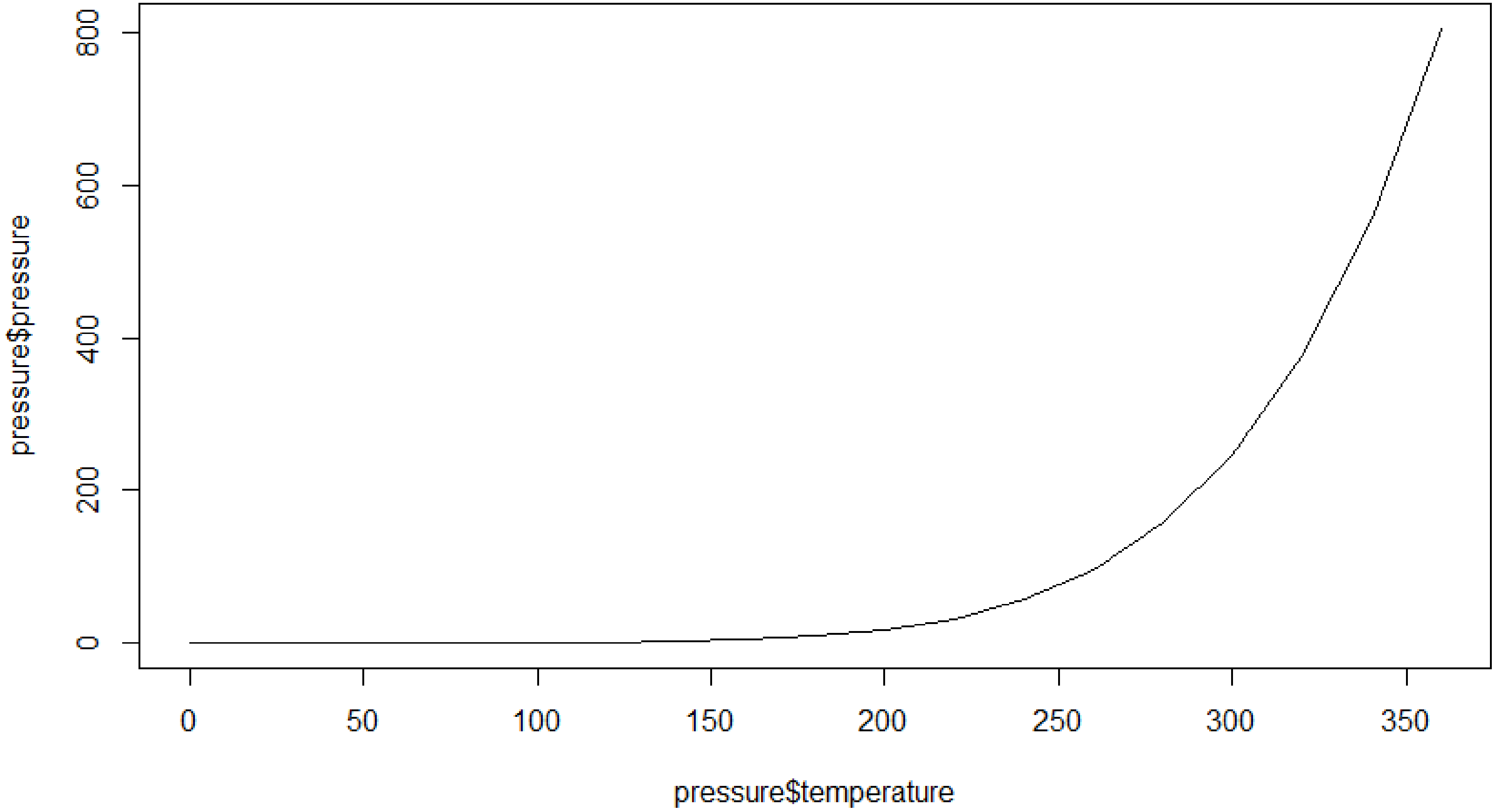
□安装之后点击Packages菜单，然后选择ggplot2。

□用同样的方法安装gcookbook package。

□在命令行输入data()，可看到所有的datasets，其中包含在ggplot2中的datasets有8个。

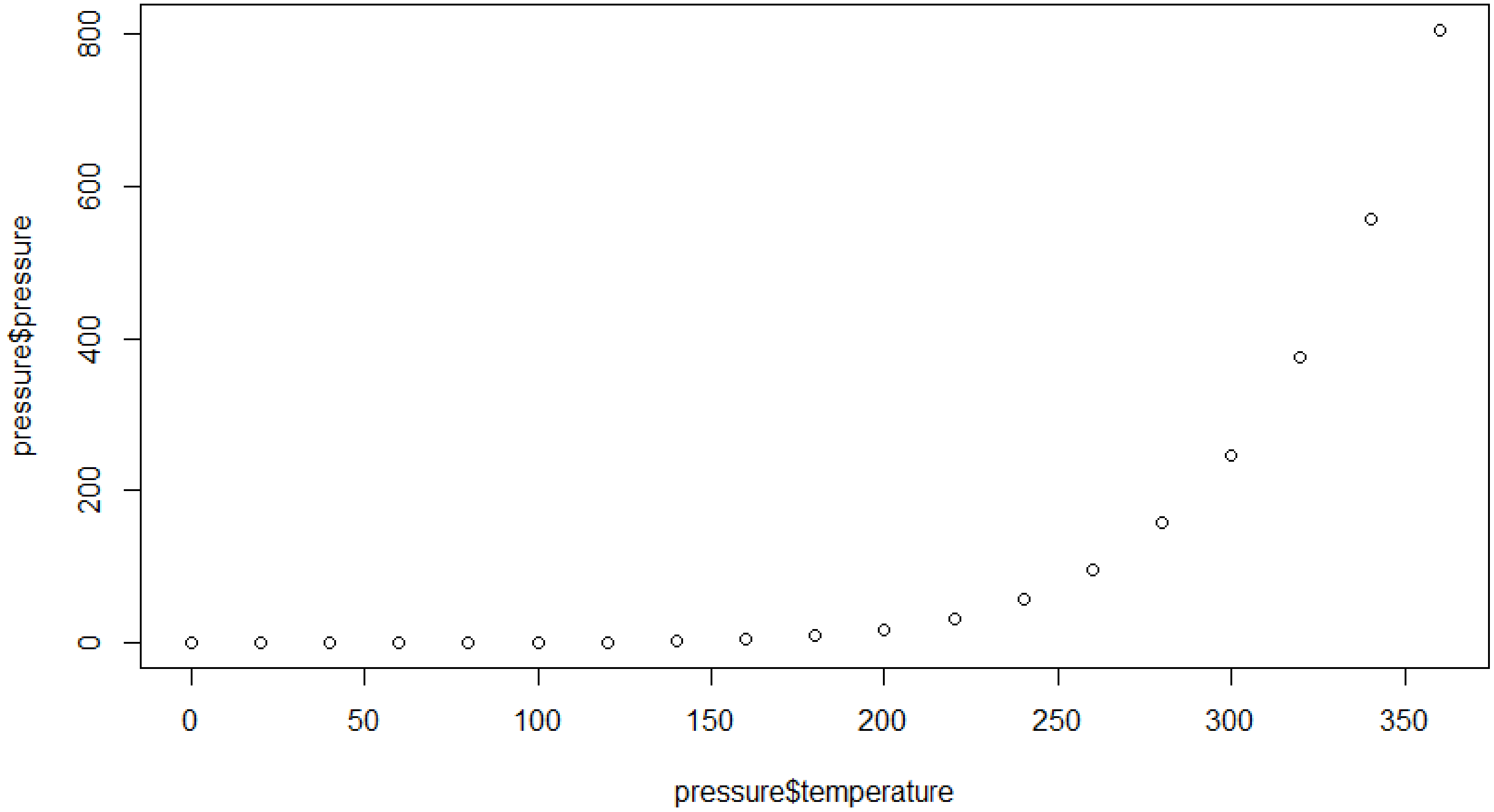
□输入View(diamonds)可看到名为diamonds的dataset的前1000个记录值。

Line Graph 线图



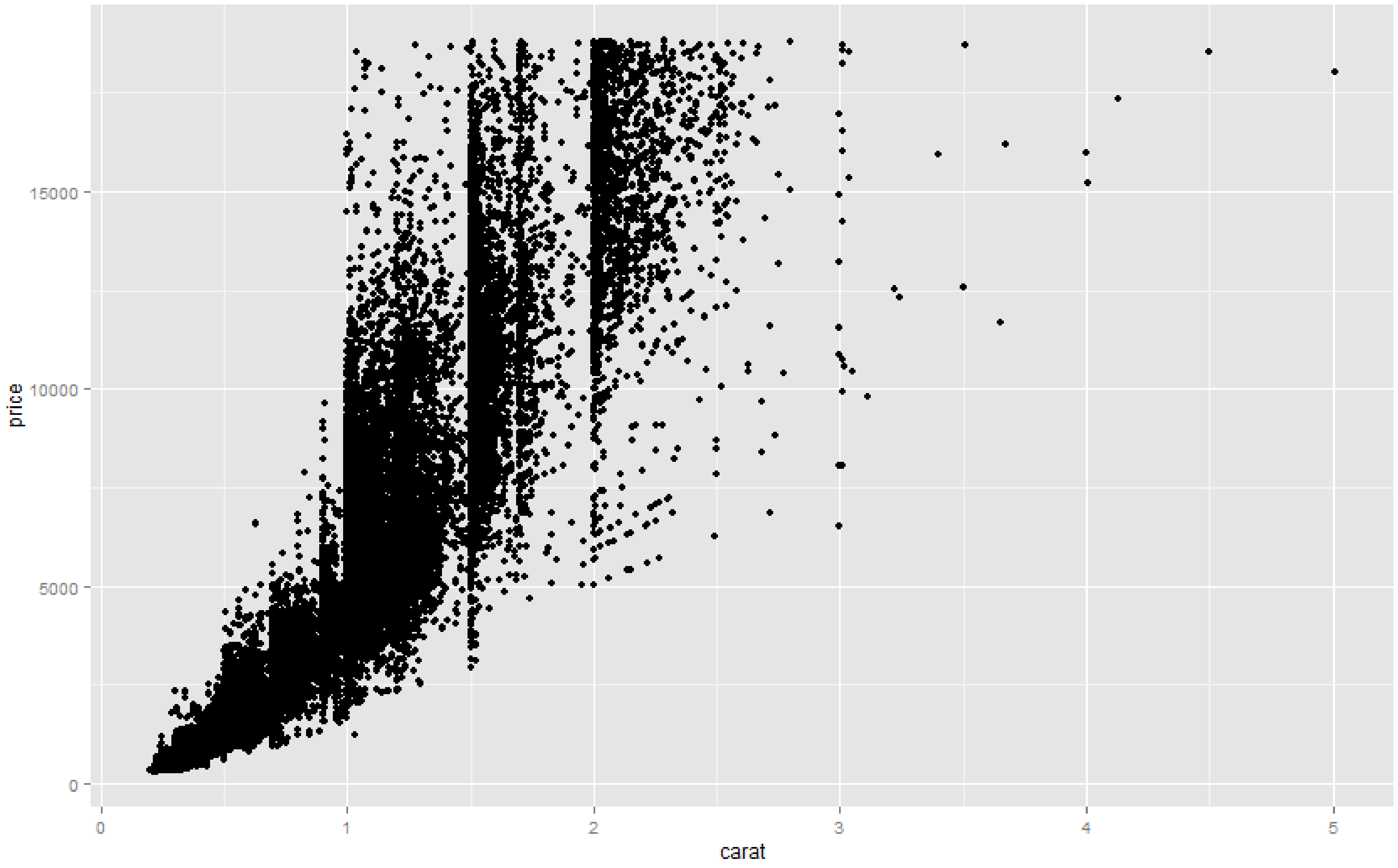
```
plot(pressure$temperature, pressure$pressure, type="l")
```

Scatterplot 散点图



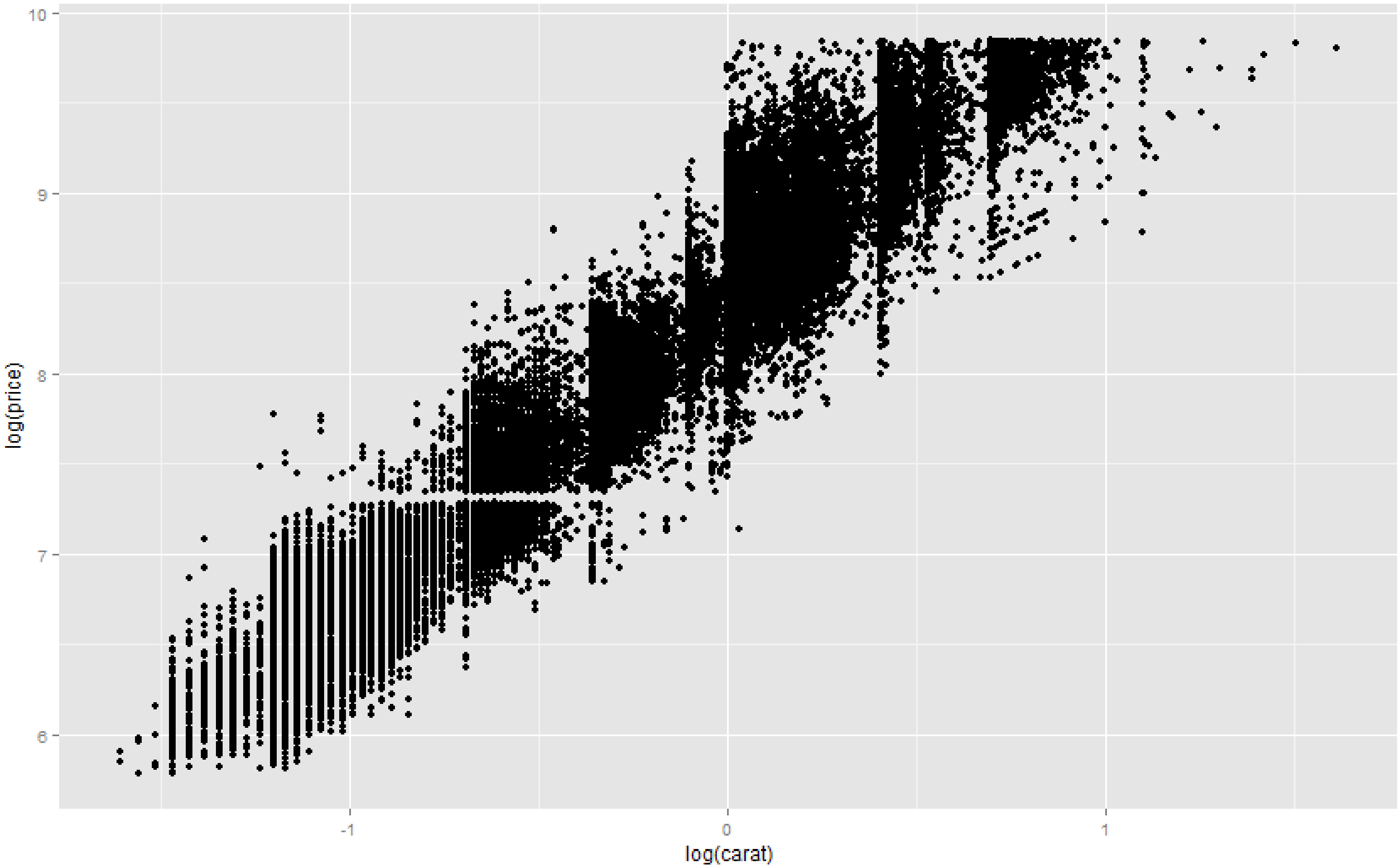
plot(pressure$temperature, pressure$pressure)

Scatterplot 散点图



qplot(carat, price, data = diamonds)

Rescaling to Log Scale



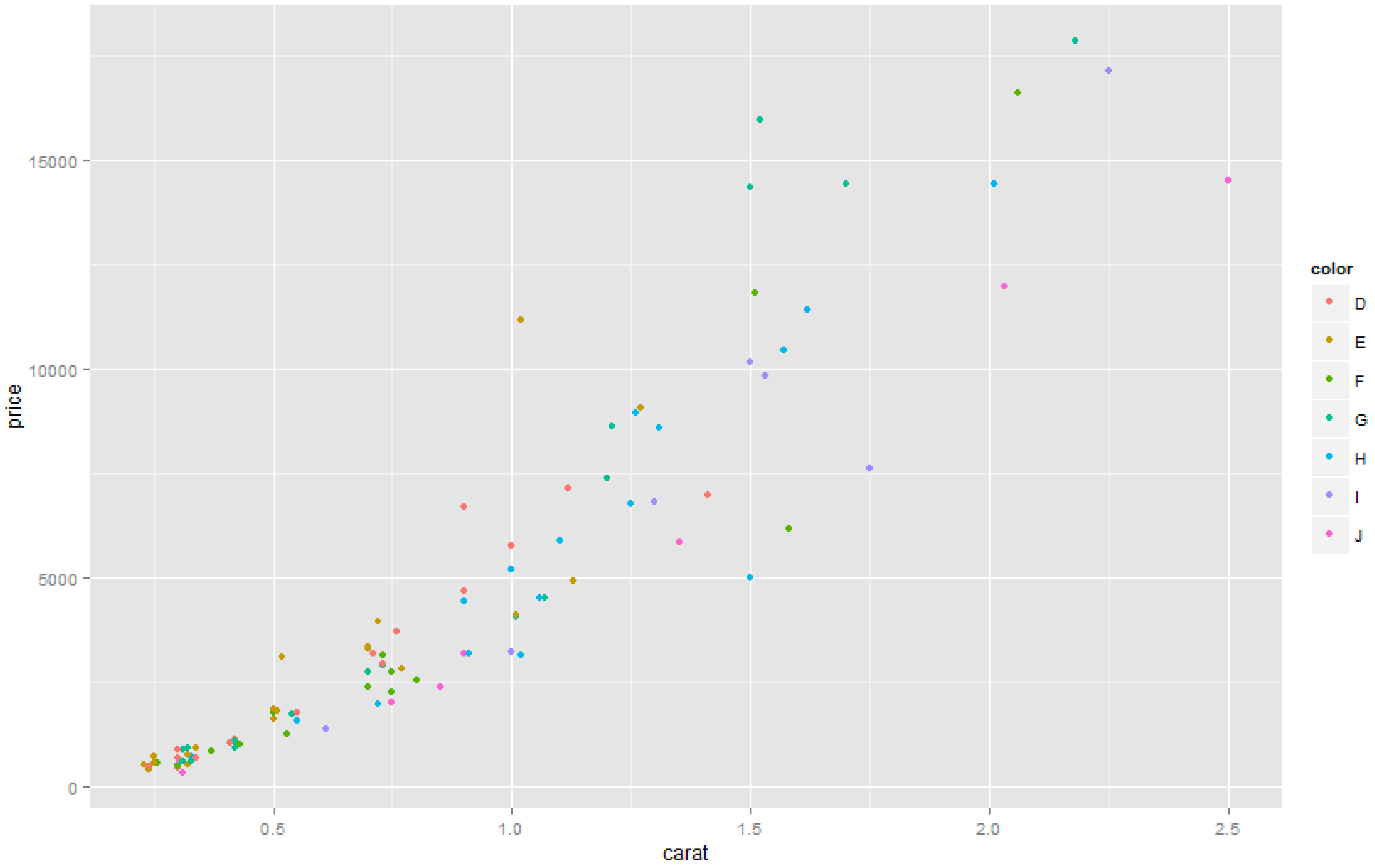
qplot(log(carat), log(price), data = diamonds)

Sampling 抽样

```
> set.seed(1400) # Make the sample reproducible 随机抽样之前先设好种子的数值，可确保每次运行的时候随机抽的数据都是相同的。这有助于保证结果的可重复性便于他人检查结果正确性。如果不设置种子数值，则每次运行随机抽出来的数据都与上次运行时抽出的数据不同。这在正式进行随机抽样时使用。
```

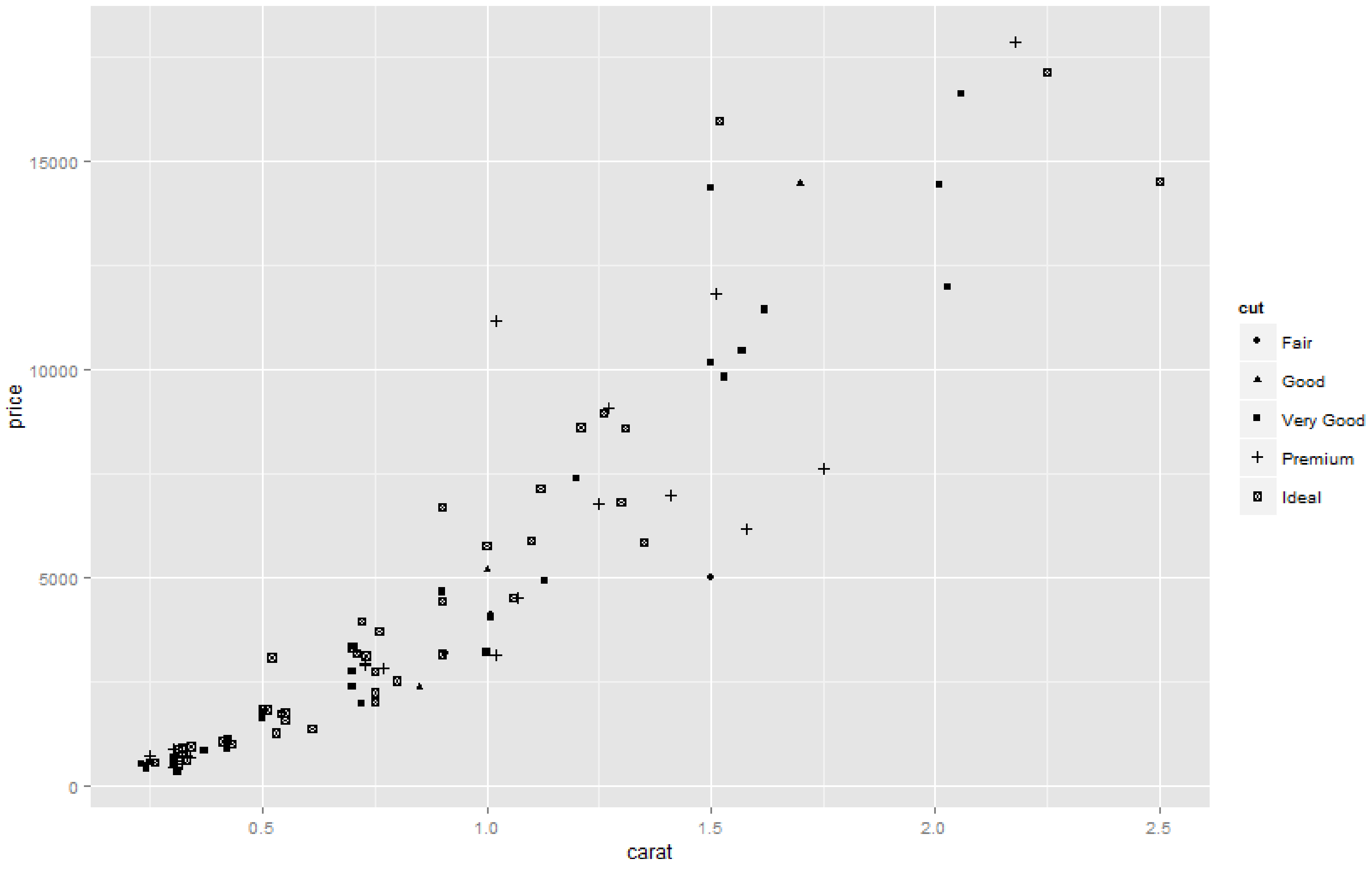
```
> dsmall <- diamonds[sample(nrow(diamonds), 100), ] #从diamonds数据集随机抽出100个数值存入dsmall中。
```

Scatterplot 散点图



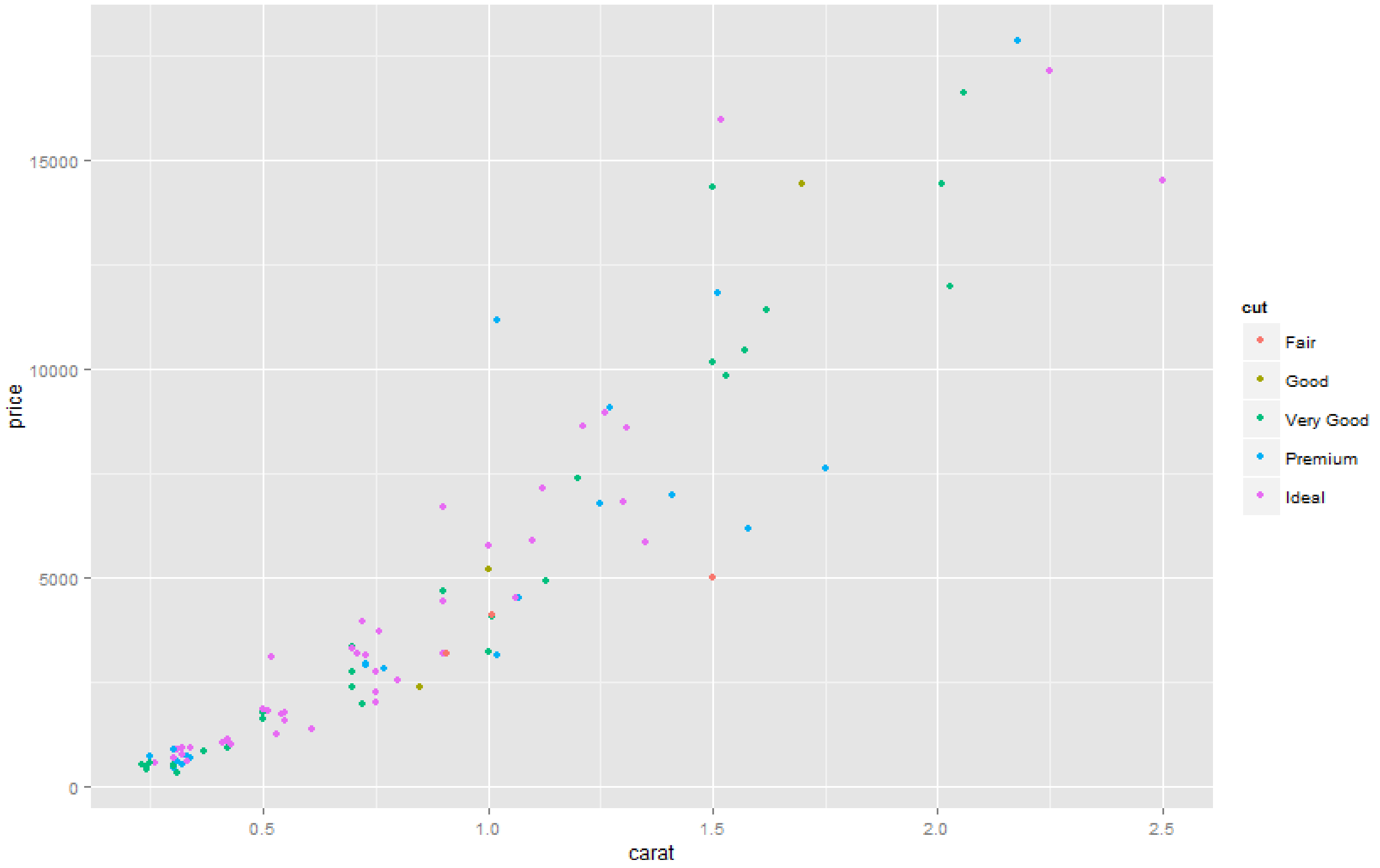
qplot(carat, price, data = dsmall, colour = color)

Scatterplot 散点图



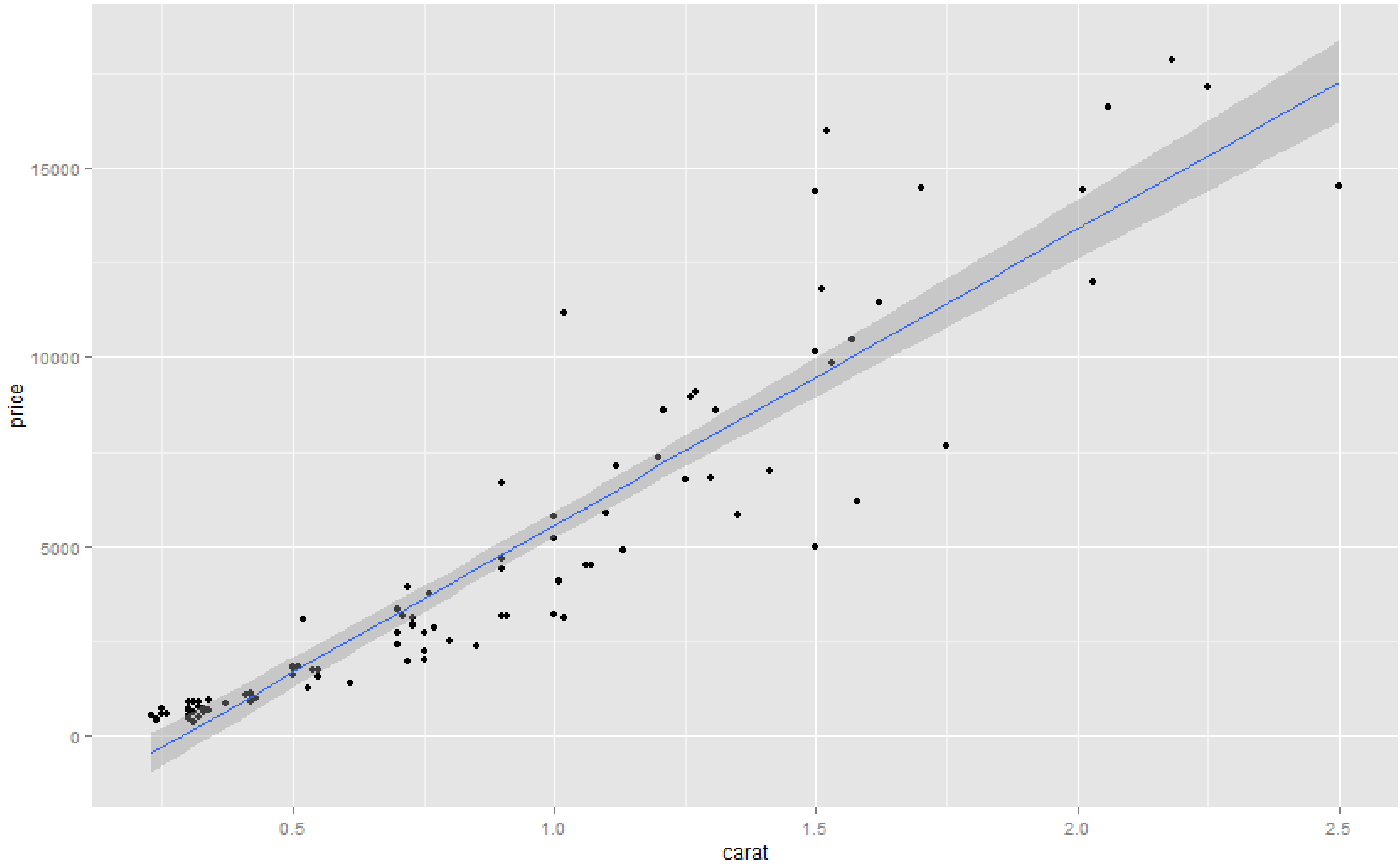
```
qplot(carat, price, data = dsmall, shape = cut)
```

Scatterplot 散点图



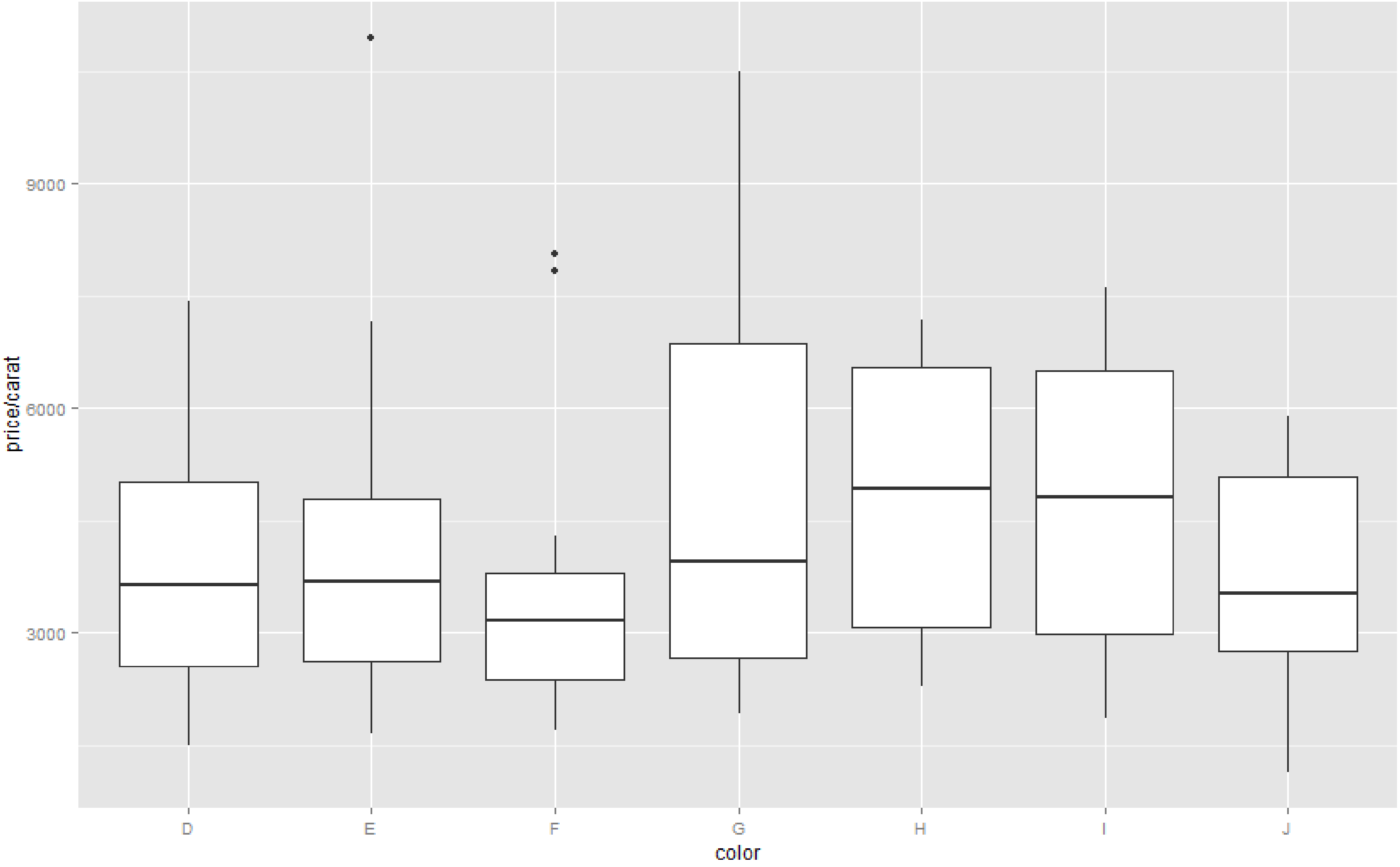
qplot(carat, price, data = dsmall, colour = cut)

直线拟合



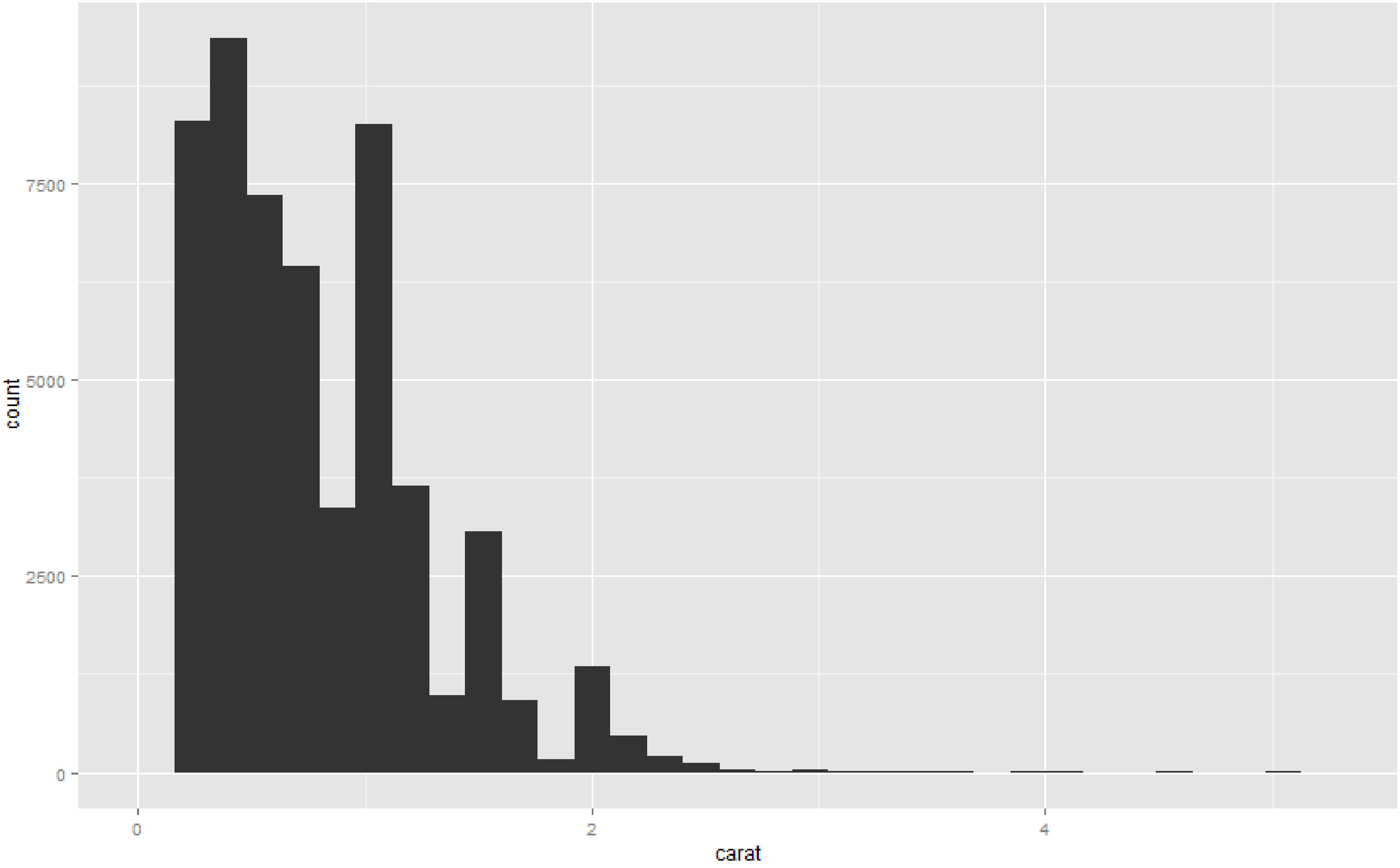
qplot(carat, price, data = dsmall, geom = c("point", "smooth"), method = "lm")
#Geom: geometric object

Boxplot 盒状图



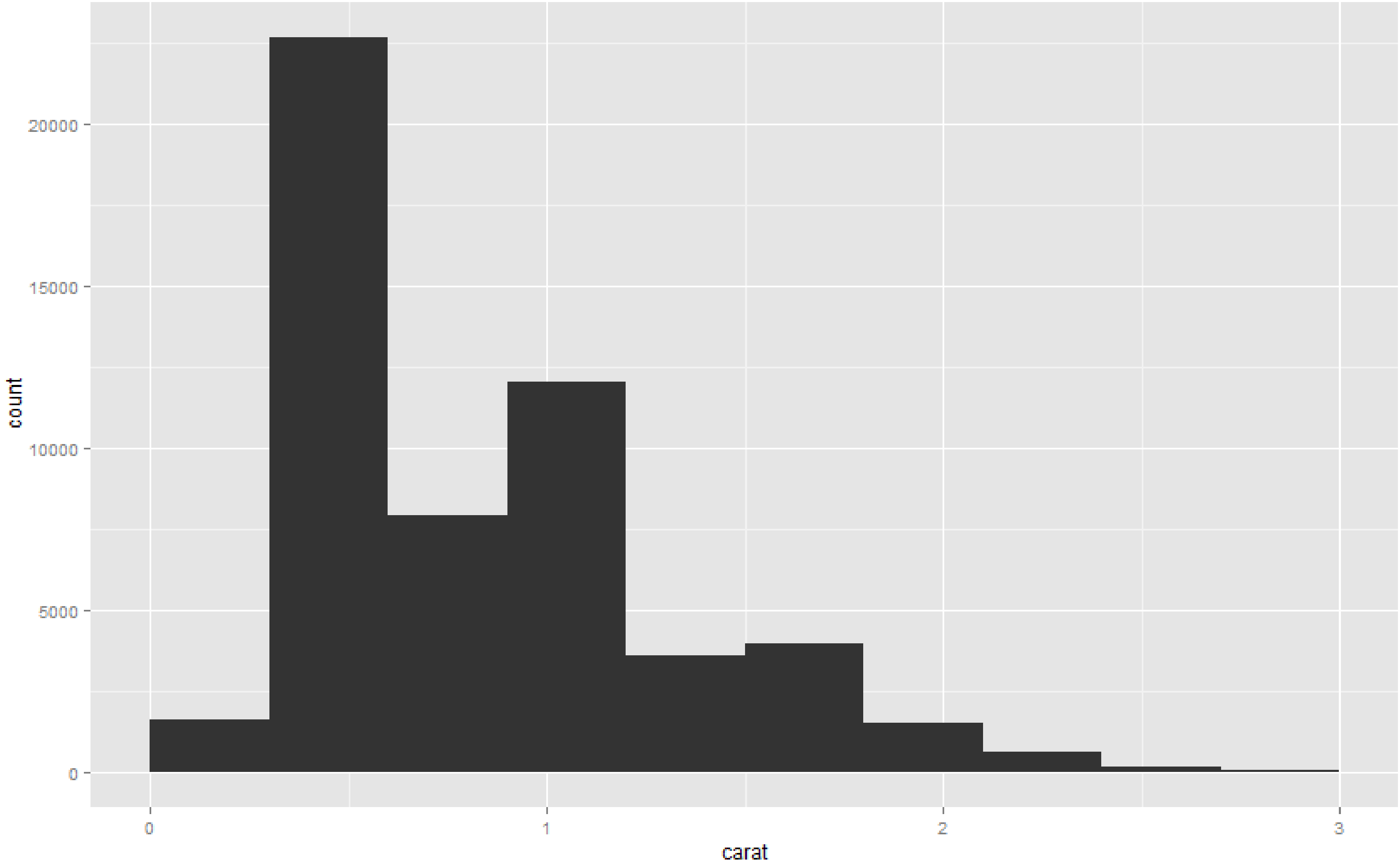
qplot(color,price/carat, data = dsmall, geom = c("boxplot"))

Histograms 直方图



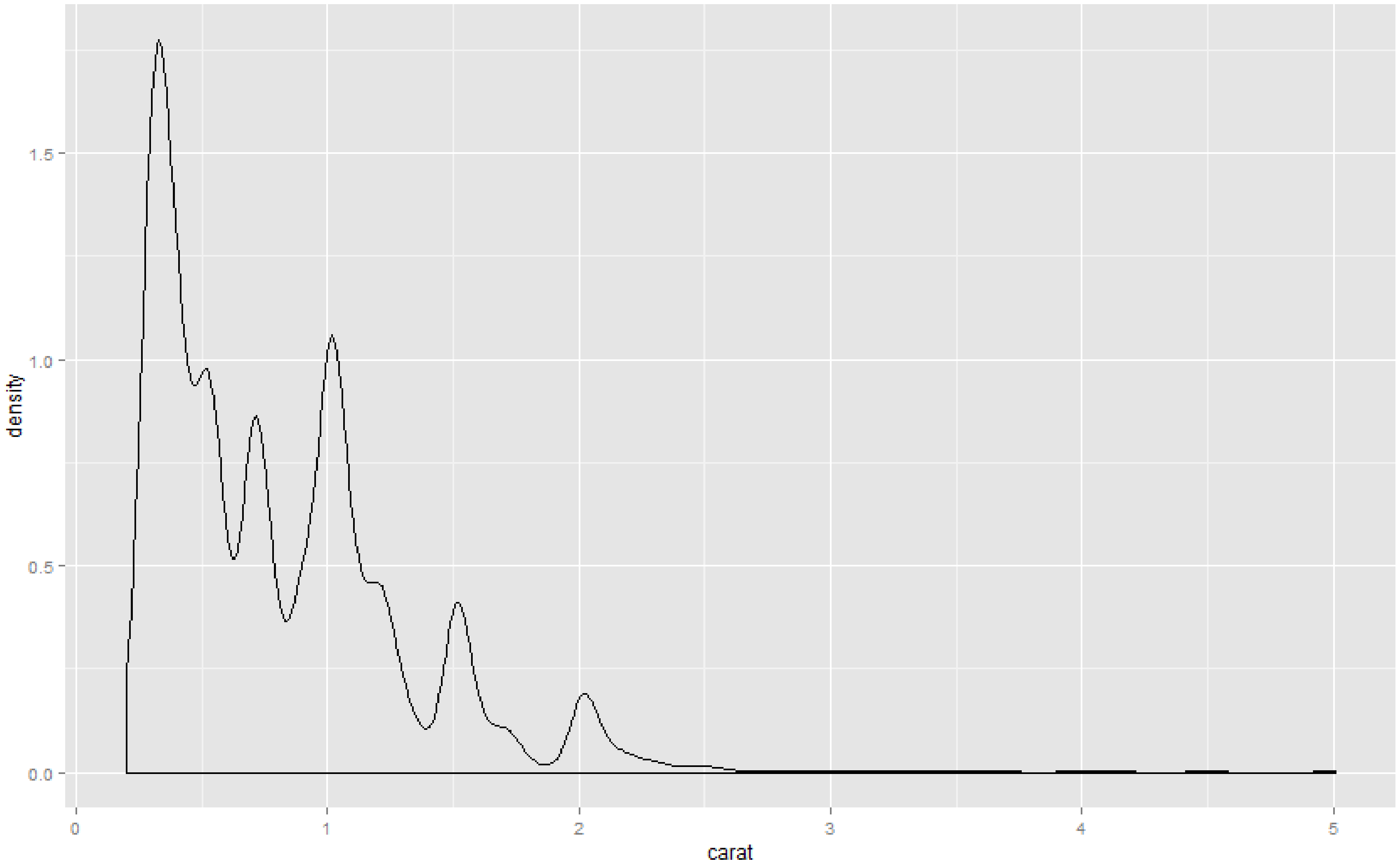
qplot(carat, data = diamonds, geom = "histogram")

Histograms 直方图



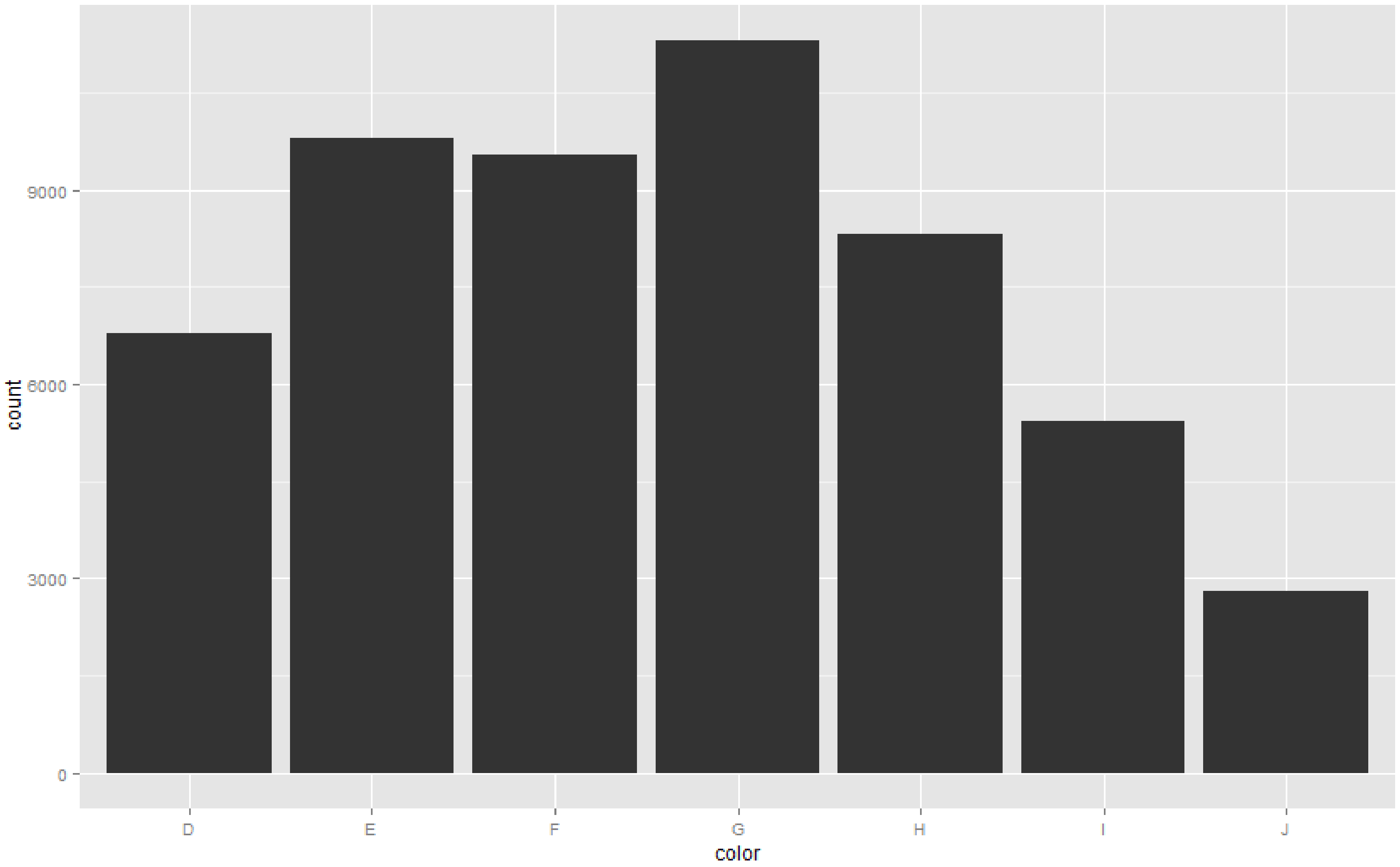
qplot(carat, data = diamonds, geom = "histogram", binwidth = 0.3, xlim = c(0,3))

Density Plot 密度图



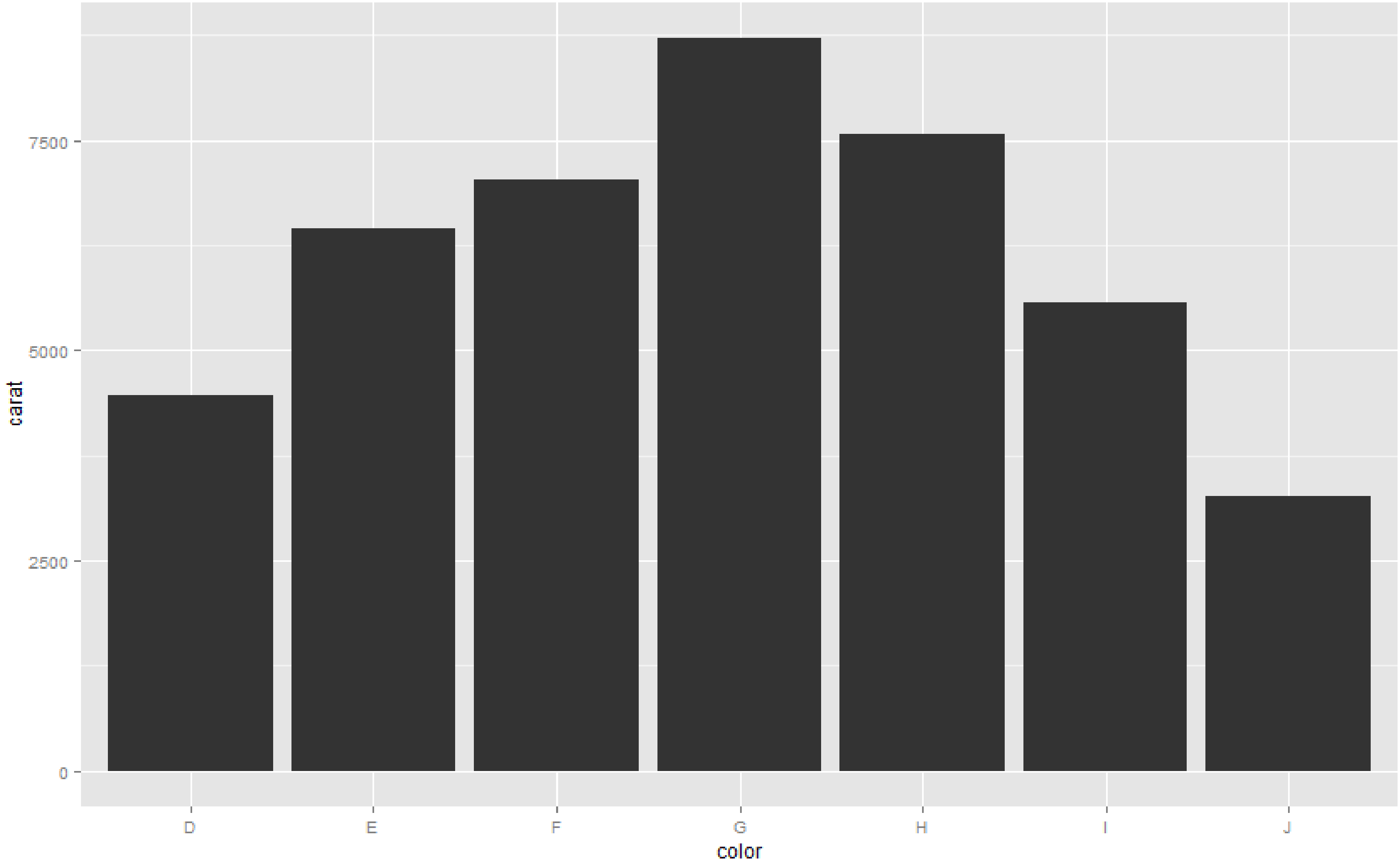
```
qplot(carat, data = diamonds, geom = "density")
```

Bar Chart 柱状图



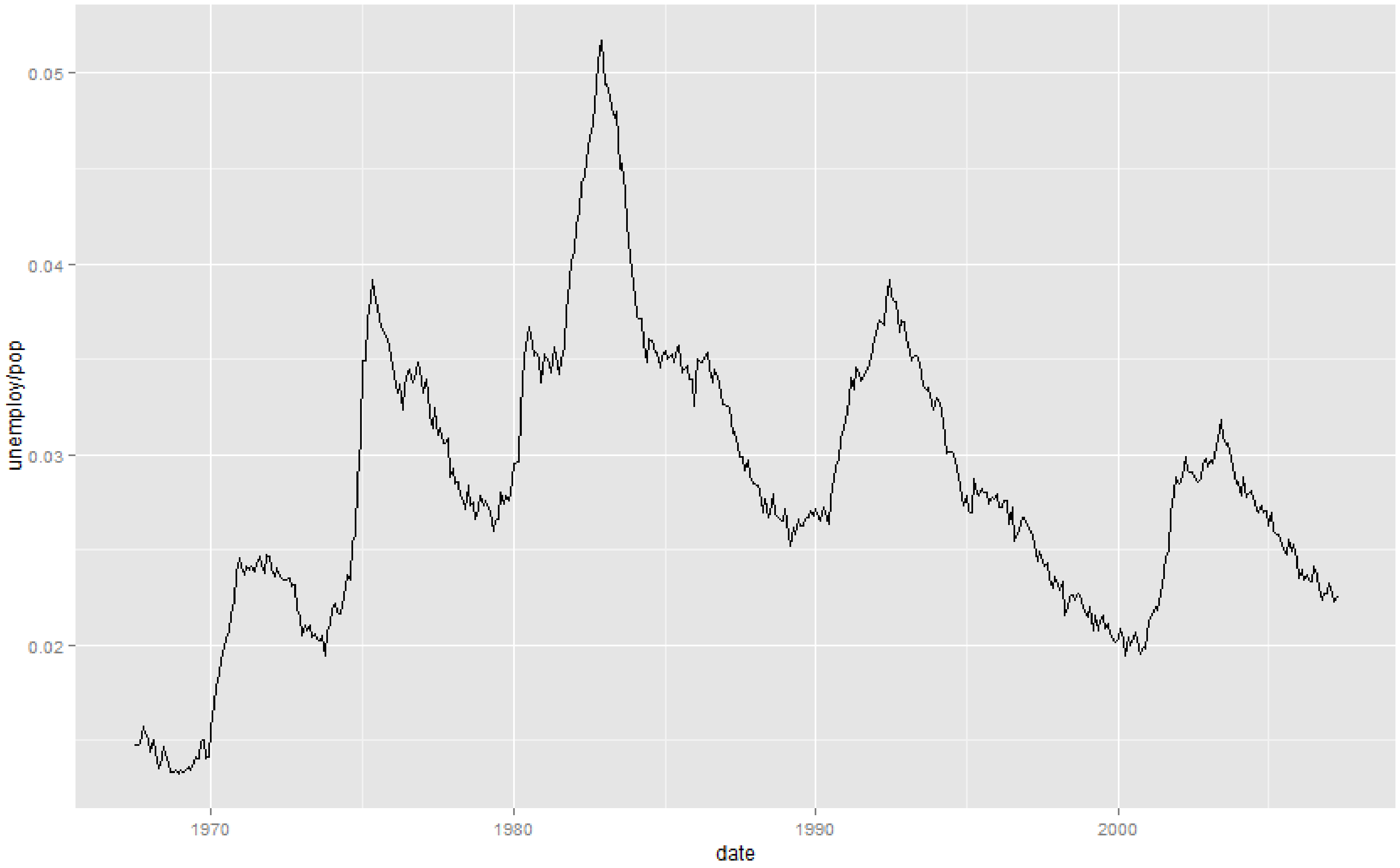
qplot(color, data = diamonds, geom = "bar")

Bar Chart 柱状图



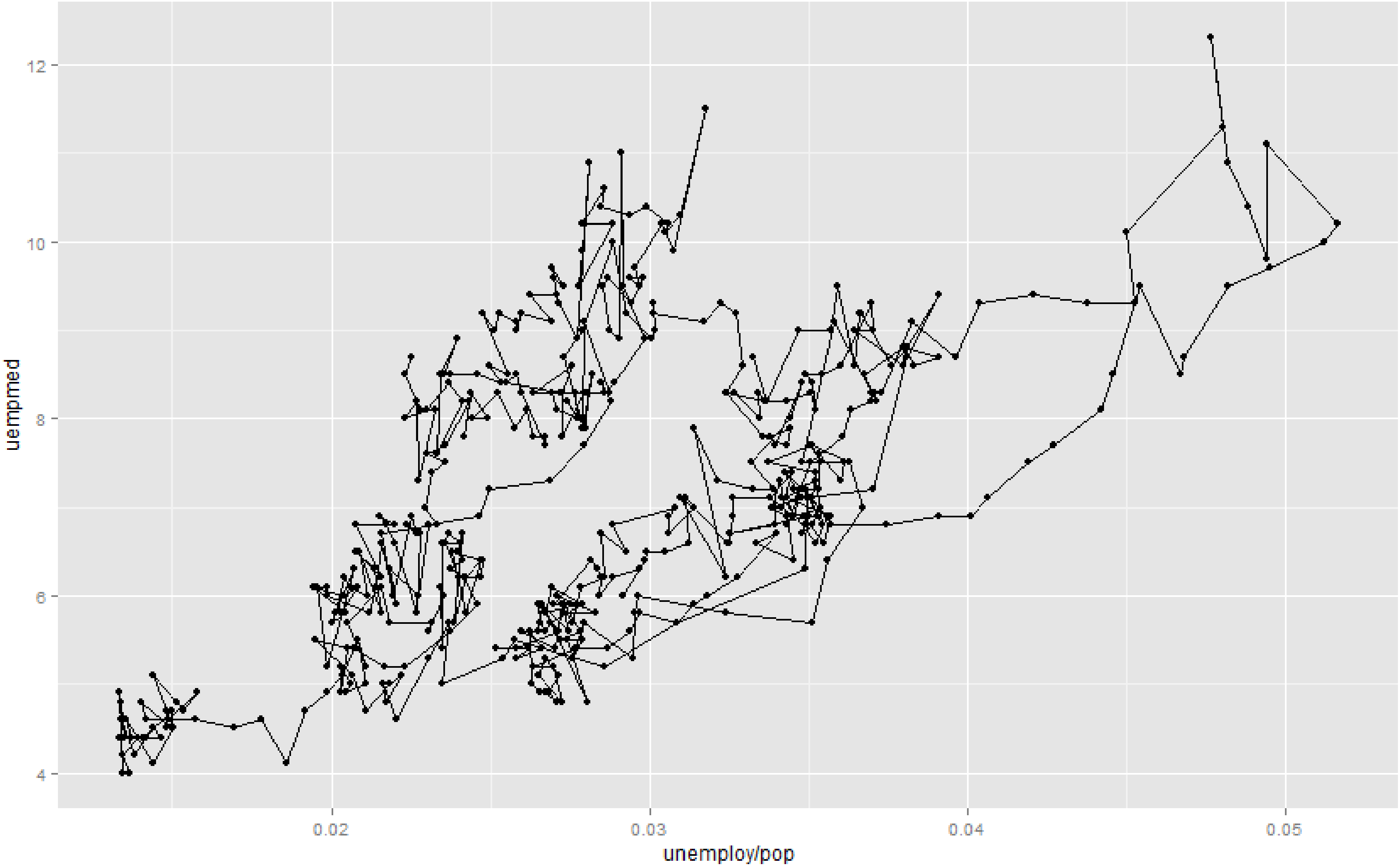
qplot(color, data = diamonds, geom = "bar", weight = carat)+scale_y_continuous("carat")

Line Plot 曲线图



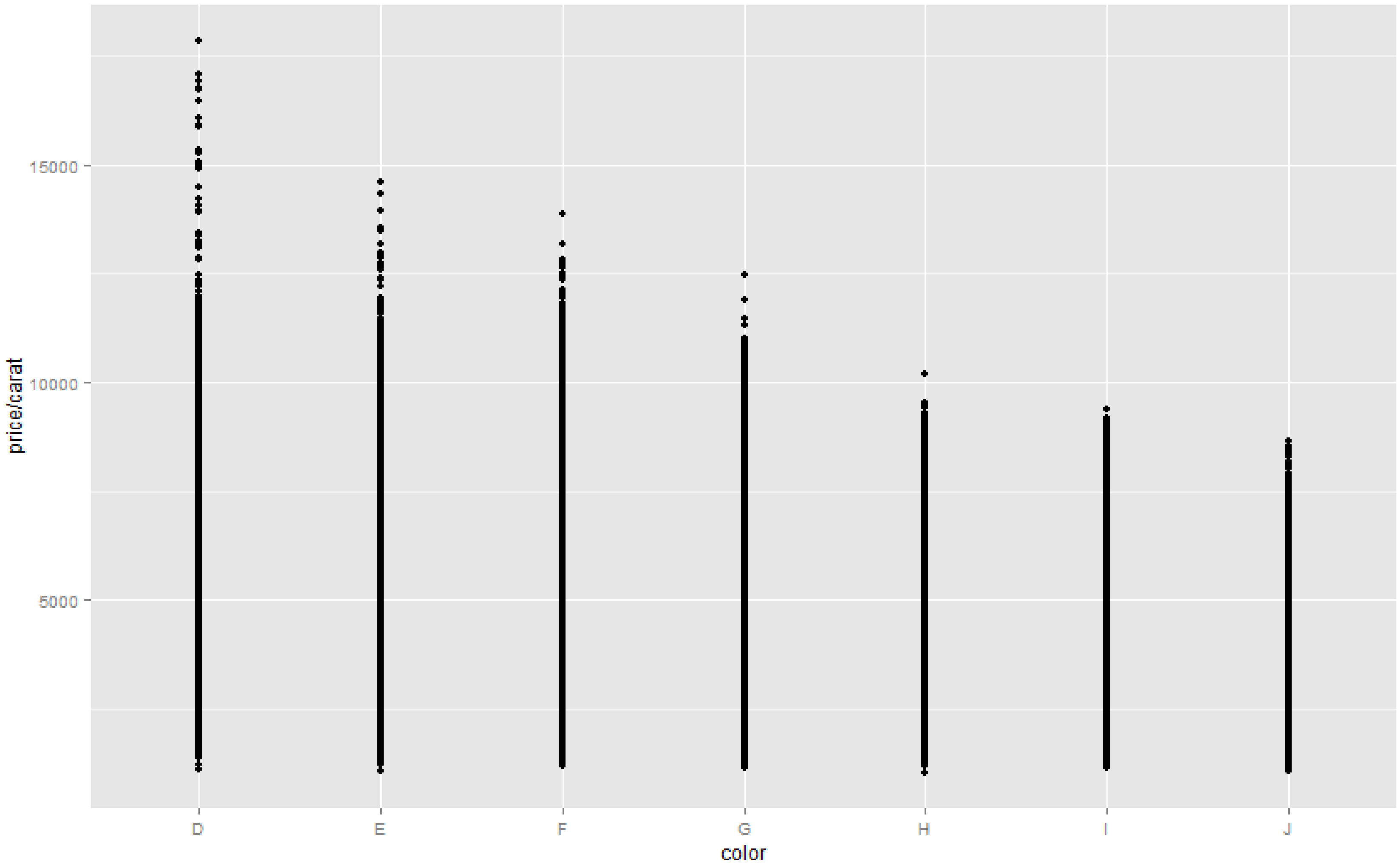
qplot(date, unemploy / pop, data = economics, geom = "line")

Path Plot 路径图



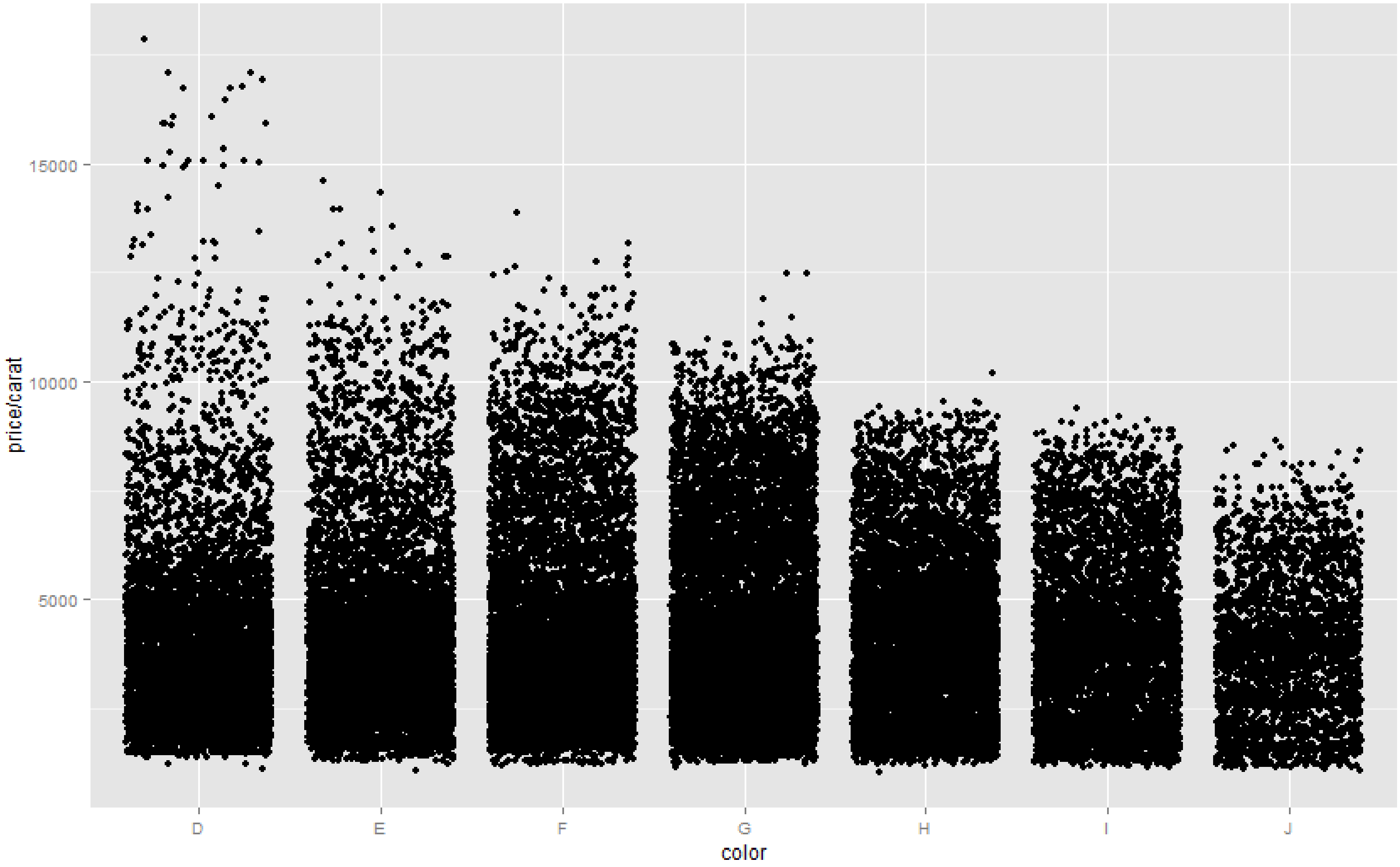
qplot(unemploy / pop, uempmed, data = economics,geom = c("point", "path"))

No Jittering



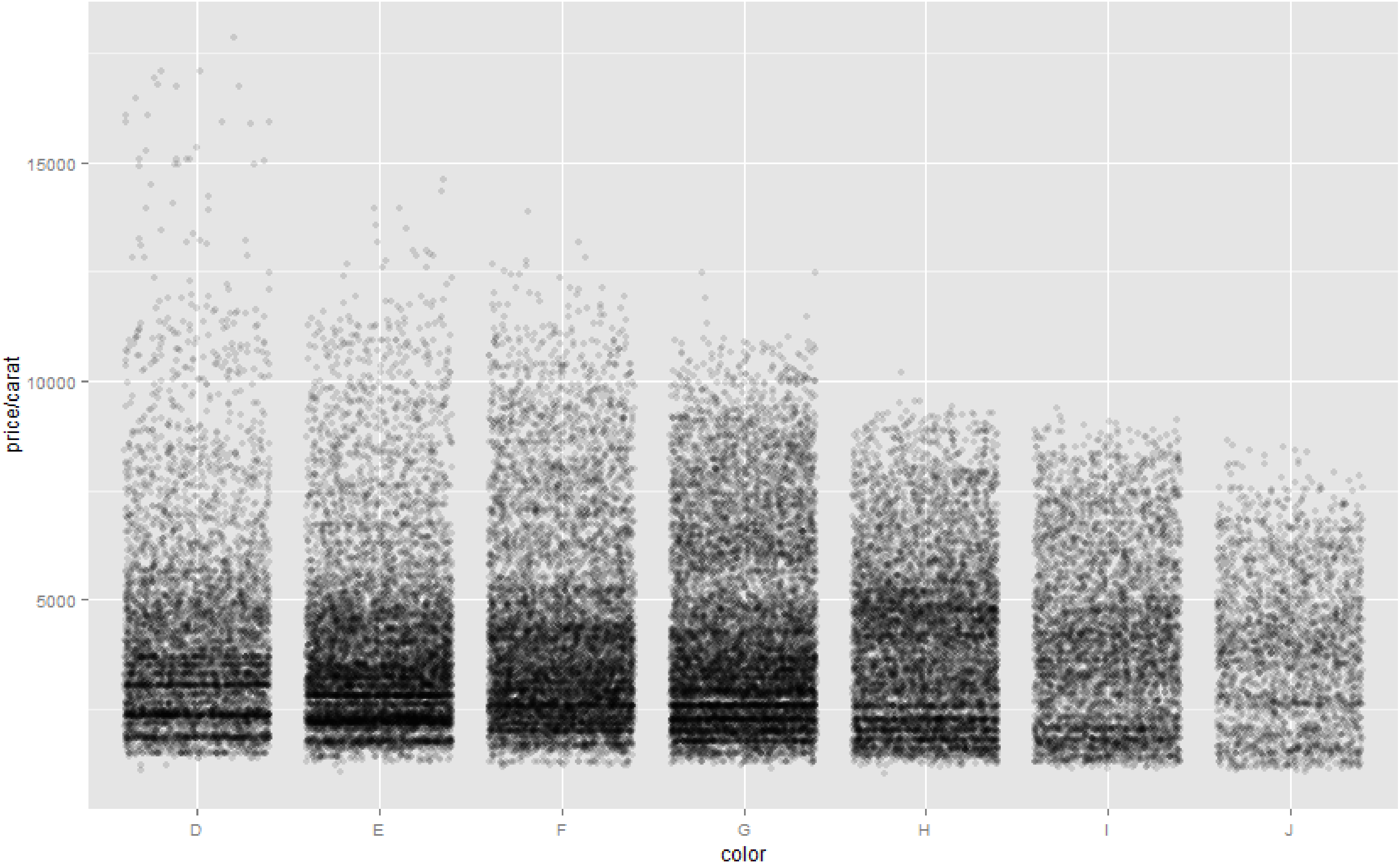
```
qplot(color, price / carat, data = diamonds)
```

Jittering



qplot(color, price / carat, data = diamonds,geom="jitter")

Jittering

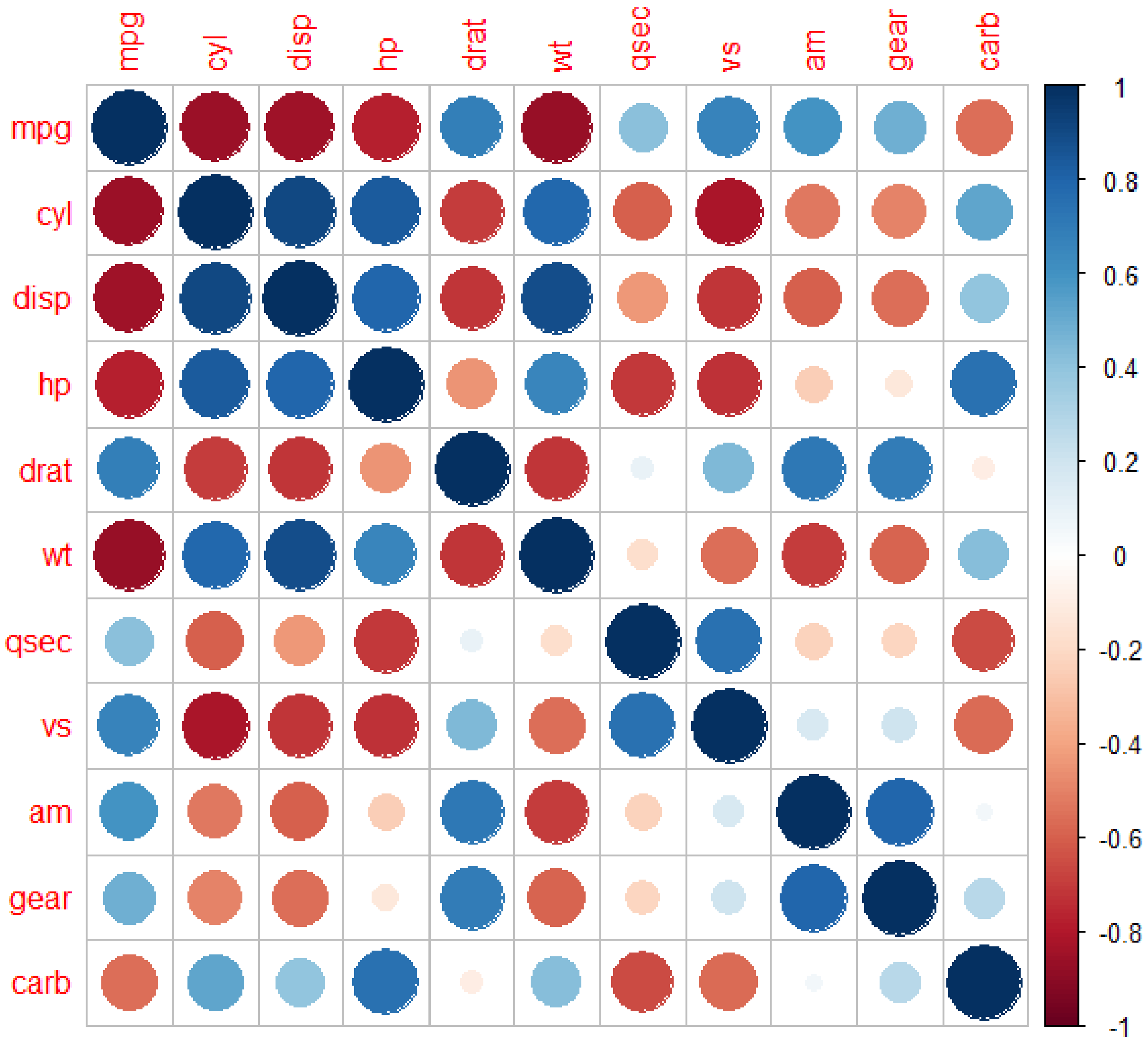


qplot(color, price / carat, data = diamonds, geom = "jitter", alpha = I(1 / 8))

Heat Maps 热度图

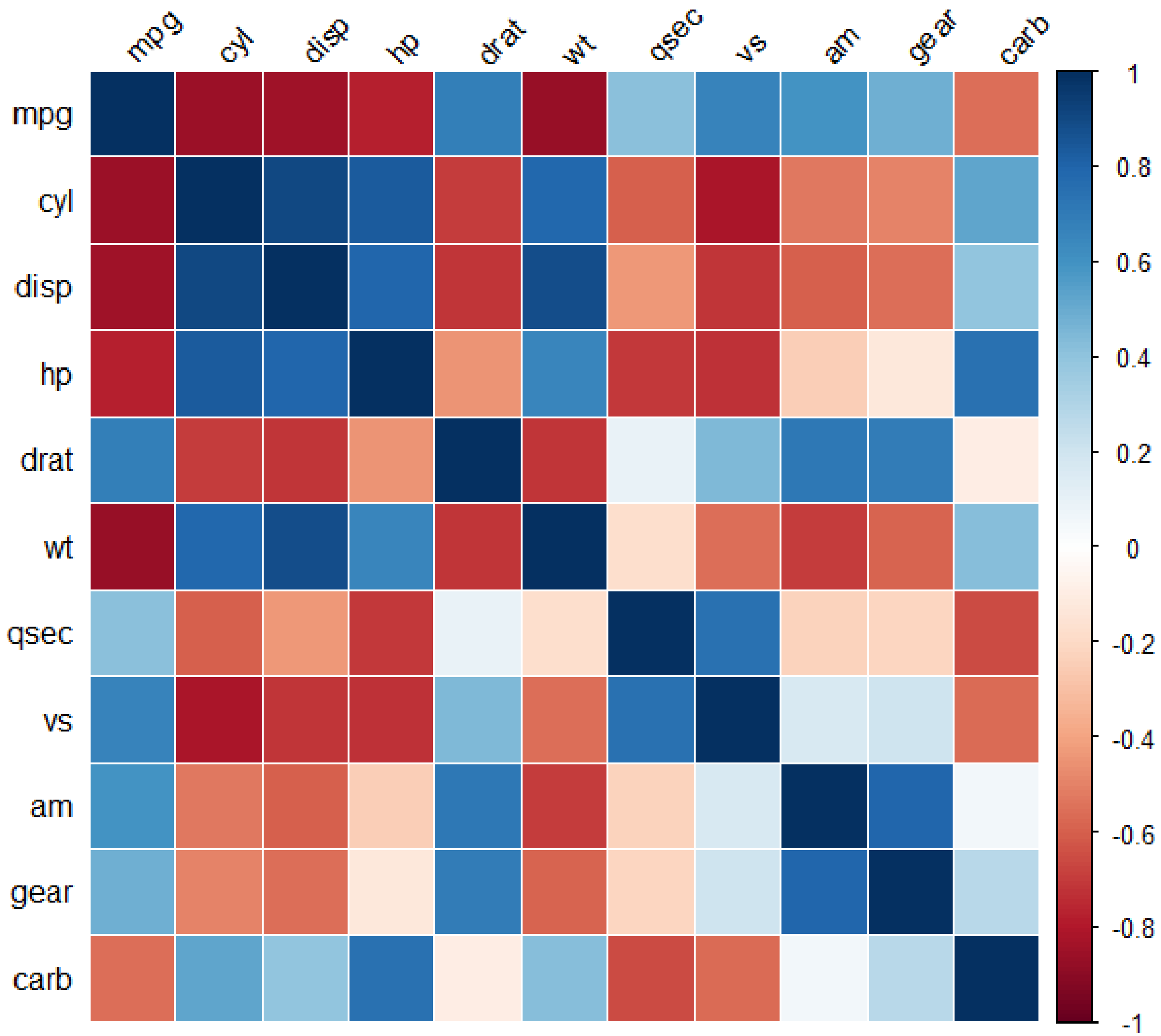
```
>install.packages("corrplot")

>library(corrplot) #也可以在RStudio的Packages中点击选中corrplot包

>mcor <- cor(mtcars) #generate the numerical correlation matrix using cor

>round(mcor, digits=2) # Print mcor and round to 2 digits

>corrplot(mcor)
```

Heat Maps 热度图



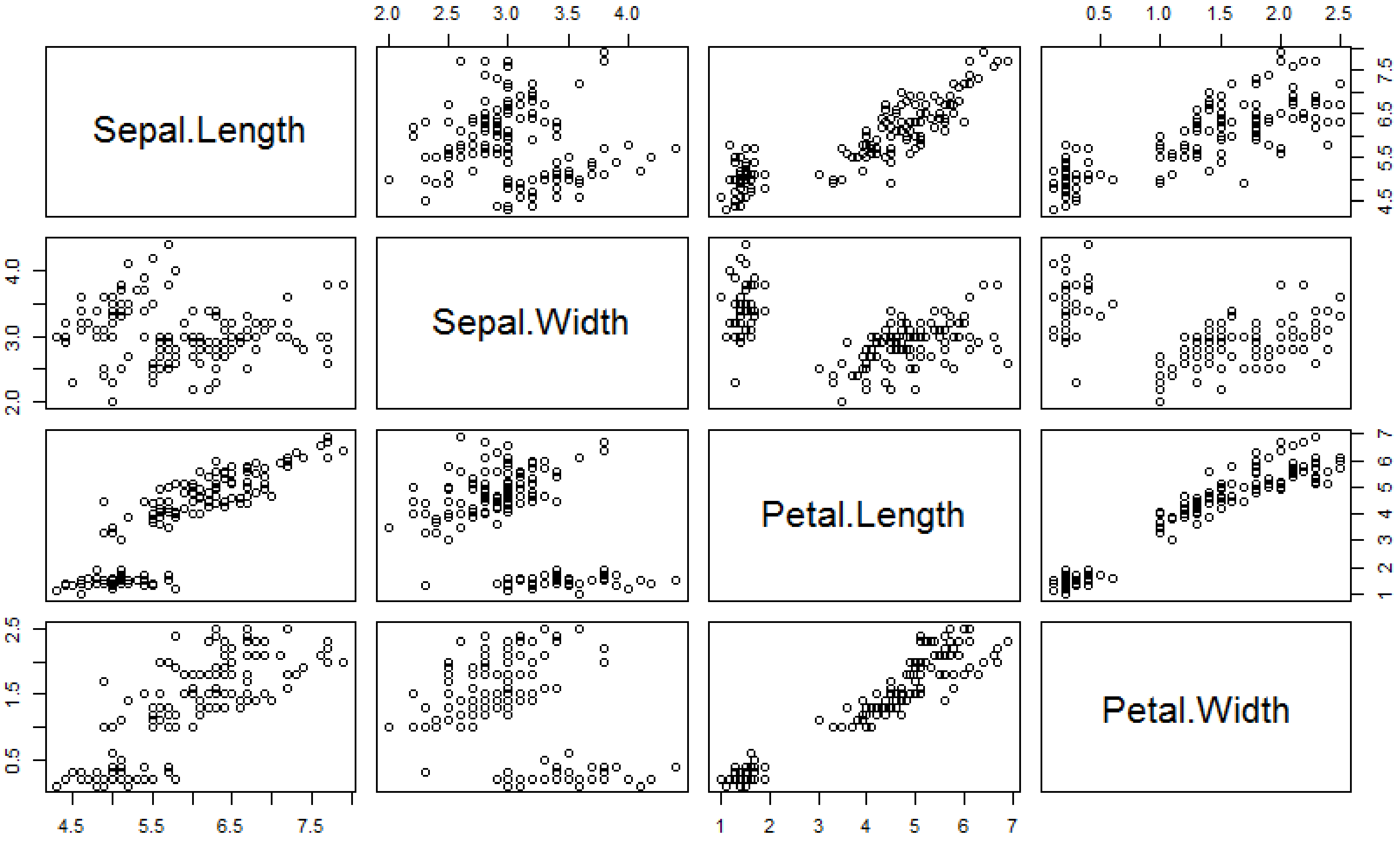
corrplot(mcor)

Heat Maps 热度图



corrplot(mcor, method="shade", shade.col=NA, tl.col="black", tl.srt=45)

Matrix Plot 矩阵图



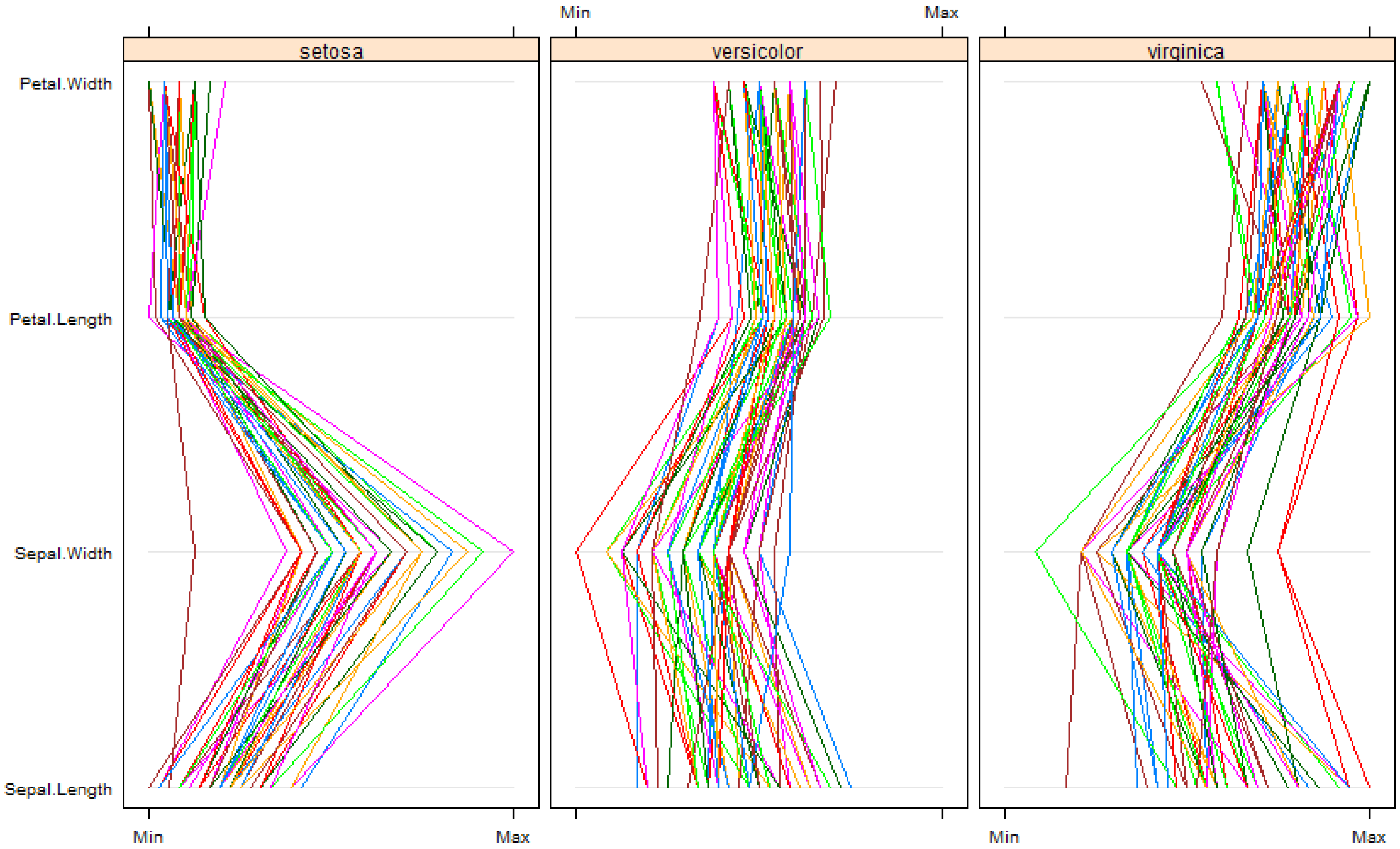
pairs(iris[,1:4])

Matrix Plot 矩阵图

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	1.0000000	-0.1175698	0.8717538	0.8179411
Sepal.Width	-0.1175698	1.0000000	-0.4284401	-0.3661259
Petal.Length	0.8717538	-0.4284401	1.0000000	0.9628654
Petal.Width	0.8179411	-0.3661259	0.9628654	1.0000000

```
cor(iris[,1:4])
```

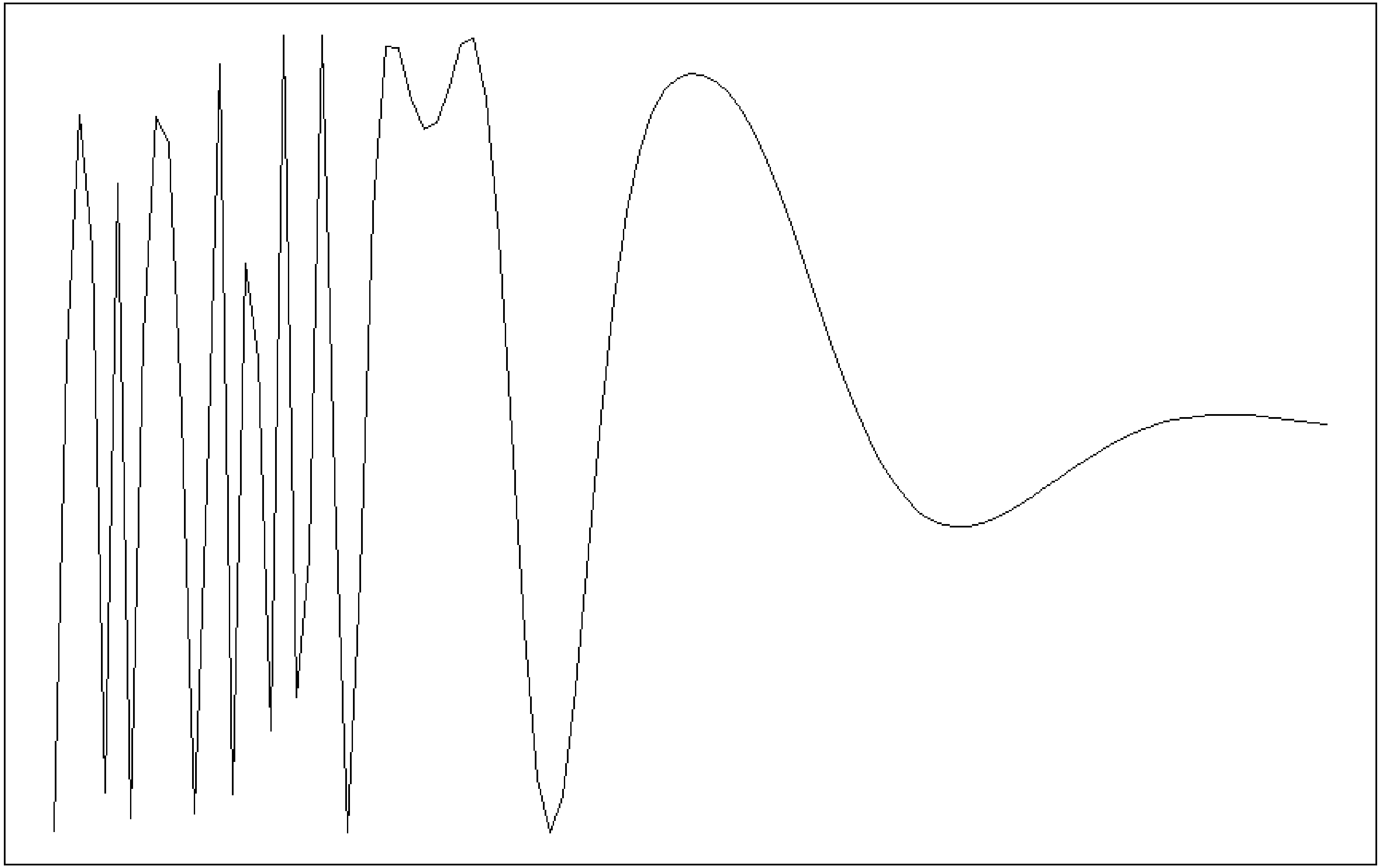
Parallel Coordinate Plot 平行坐标图



library(lattice)

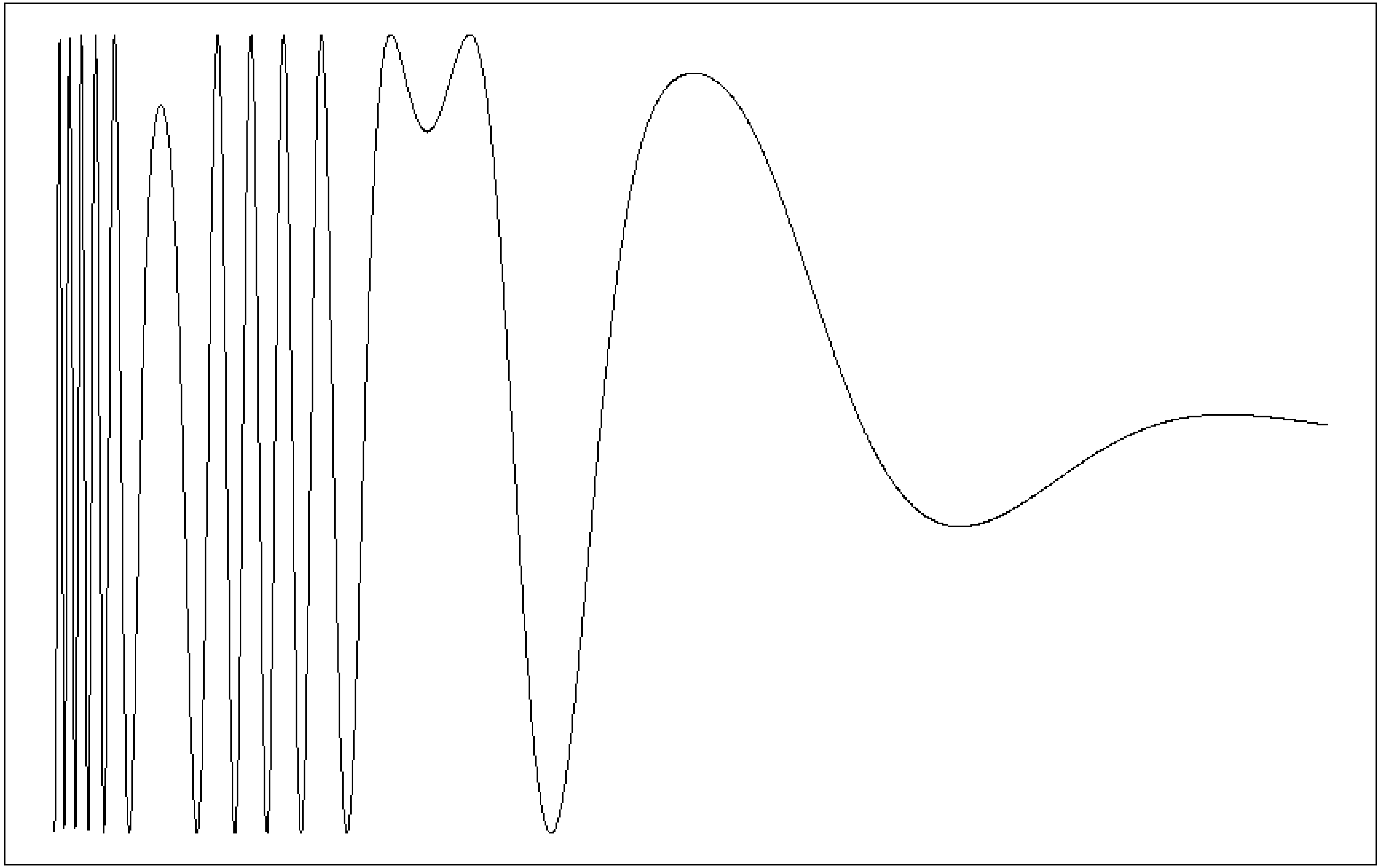
parallelplot(~iris[1:4] | Species, iris)

Plotting a Function 函数画图



```
> chippy <- function(x) sin(cos(x)*exp(-x/2))
> plot (chippy, -8, 7)
```

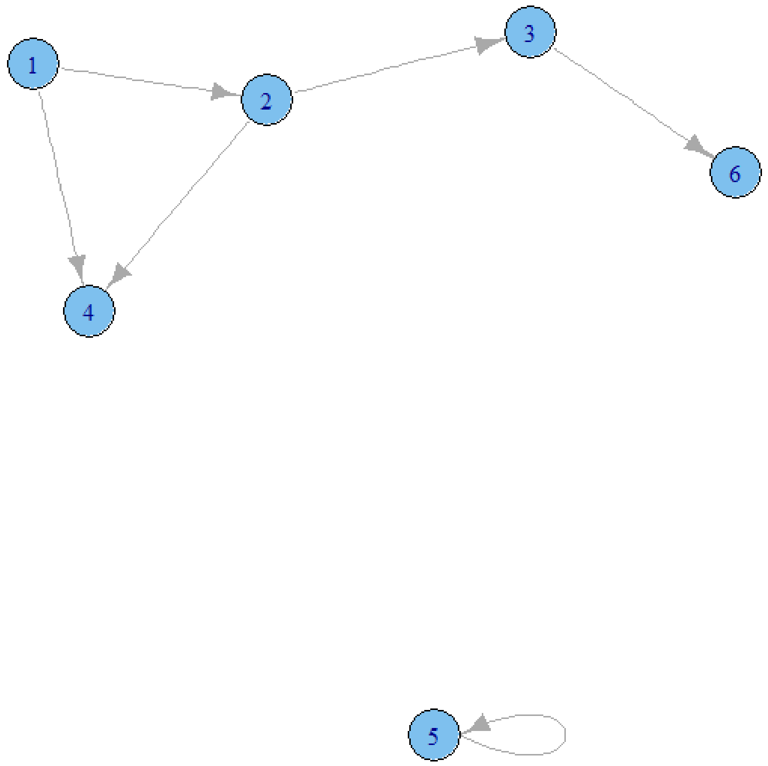
Plotting a Function 函数画图



```
> chippy <- function(x) sin(cos(x)*exp(-x/2))
> curve(chippy, -8, 7, n = 2000)
```

Network Graph 网络图

```
>install.packages("igraph")

>library(igraph)

>gd <- gr

>plot(gd)
```

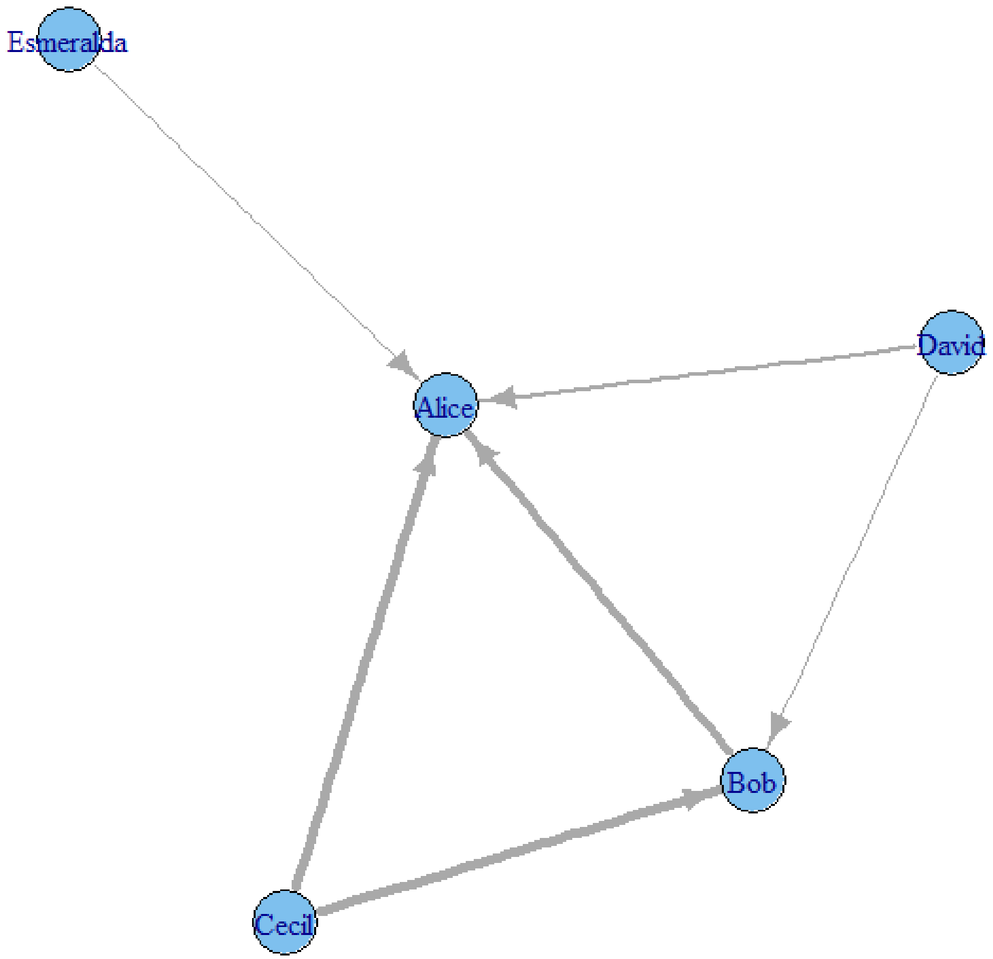


Network Graph 网络图

```
> relations <- data.frame(from=c("Bob", "Cecil", "Cecil", "David", "David", "Esmeralda"),   to=c("Alice", "Bob", "Alice", "Alice", "Bob", "Alice"),      weight=c(4,5,5,2,1,1))

> g <- graph.data.frame(relations, directed=TRUE)

> plot(g, edge.width=E(g)$weight)
```

	from	to	weight
1	Bob	Alice	4
2	Cecil	Bob	5
3	Cecil	Alice	5
4	David	Alice	2
5	David	Bob	1
6	Esmeralda	Alice	1

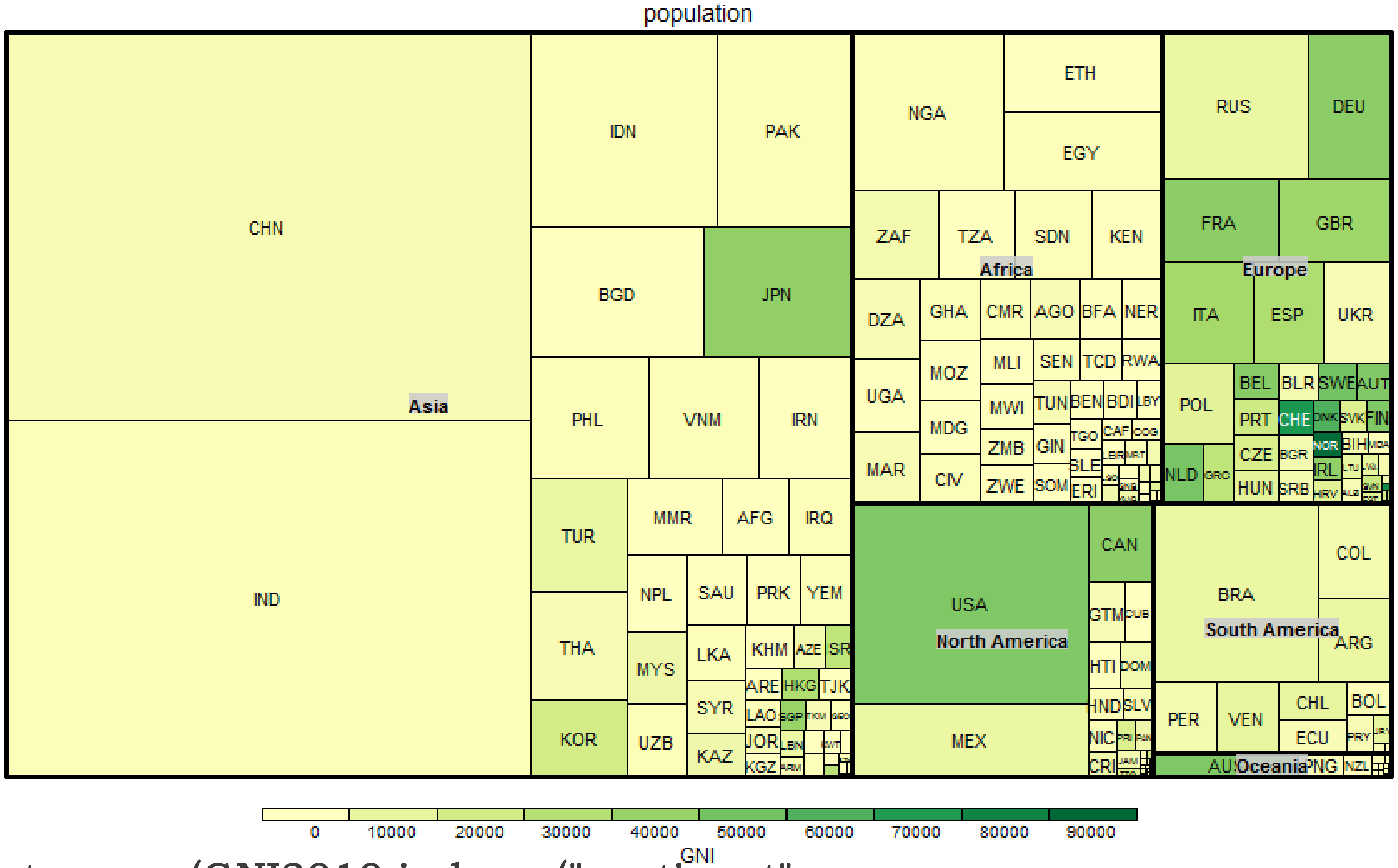
Network Graph 网络图



Treemap 矩阵树图

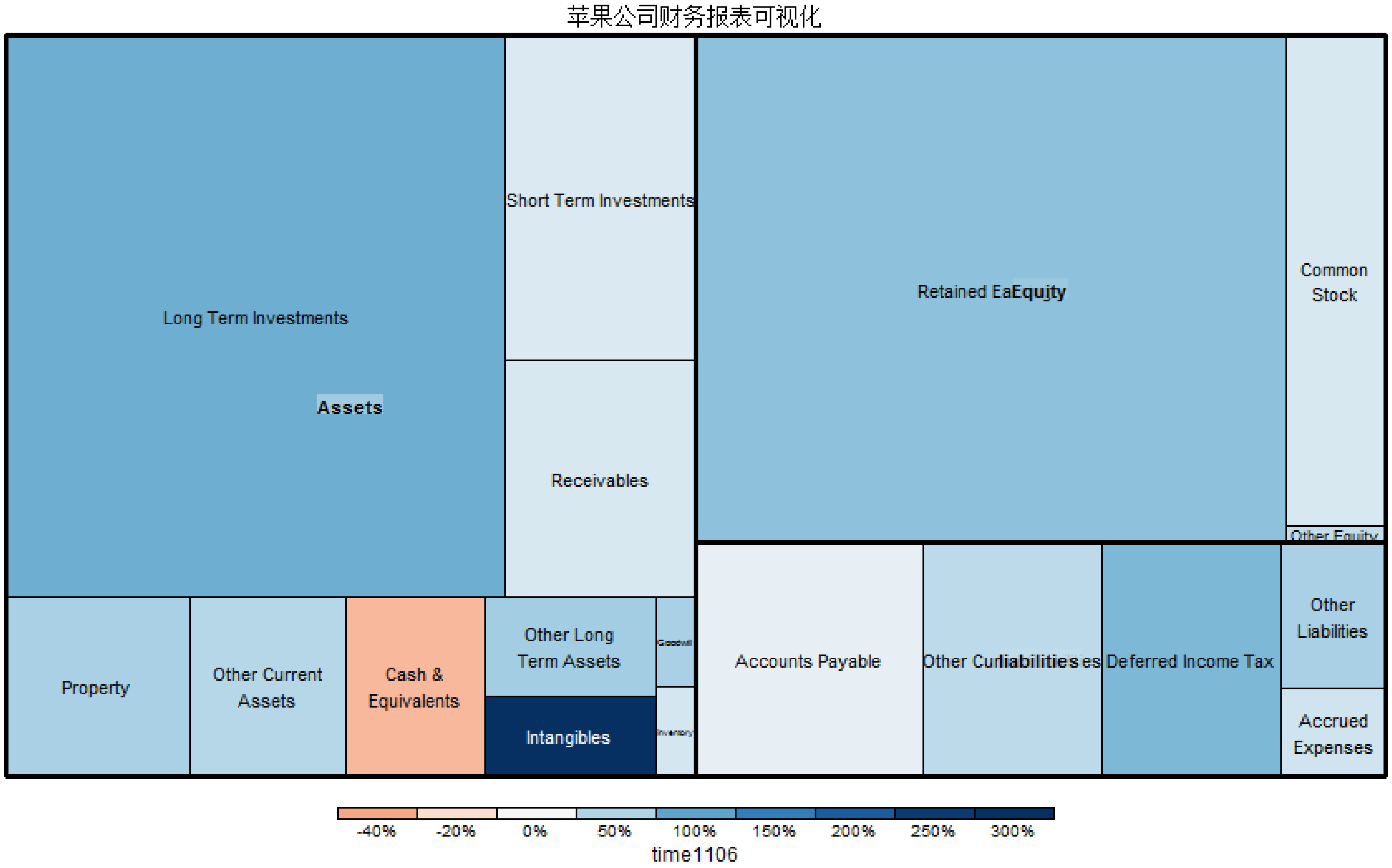
```
>install.packages("treemap")

>library(treemap)

>data(GNI2010)

>treemap(GNI2010,index=c("continent",
"iso3"),vSize="population",vColor="GNI",type="value")
```

Treemap 矩阵树图



```
treemap(GNI2010,index=c("continent",
"iso3"),vSize="population",vColor="GNI",type="value")
```

Treemap 矩阵树图

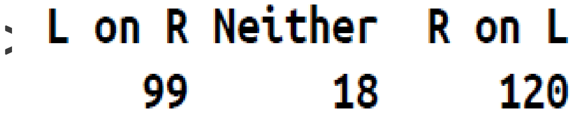


```
data <- read.csv('c:/BA/Visualization/AppleFinance.csv',T)
treemap(data, index=c("item", "subitem"), vSize="time1206", vColor="time1106",
type="comp", title='苹果公司财务报表可视化', palette='RdBu')
```

Pie Chart 饼图

```
>library(MASS) #for dataset
```

```
># Get a table of how many cases are in each level of fold
```

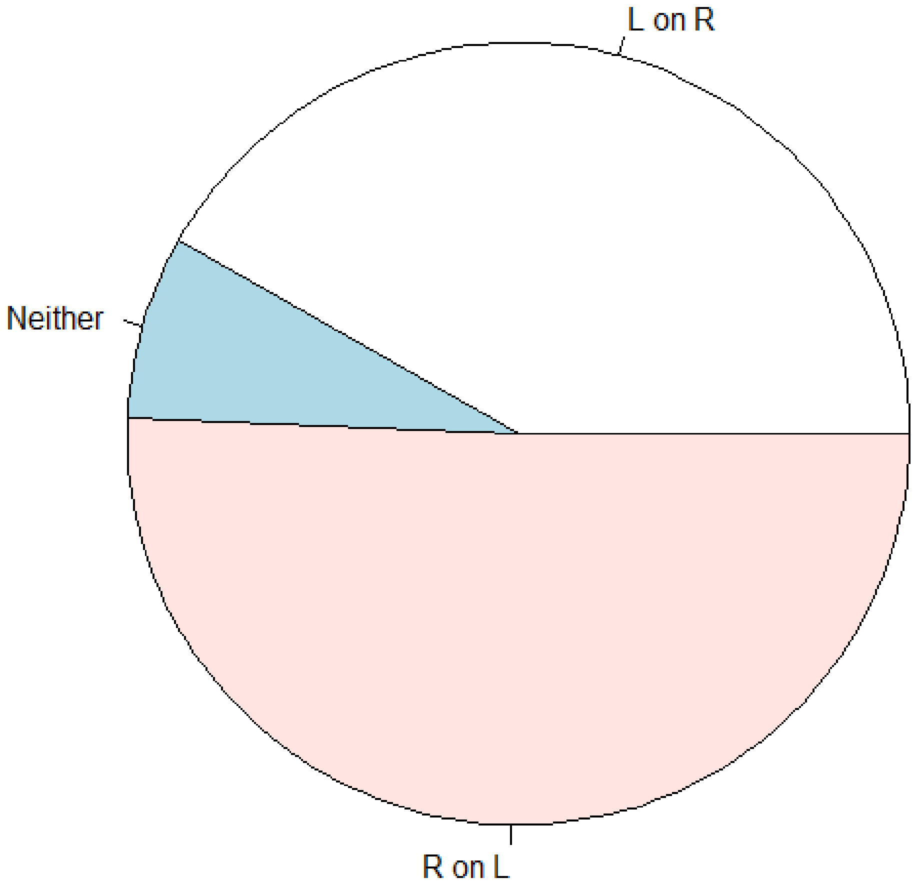


```
L on R Neither   R on L 
    99       18      120 
```

```
y$Fold)
```

```
>fold
```

Pie Chart 饼图

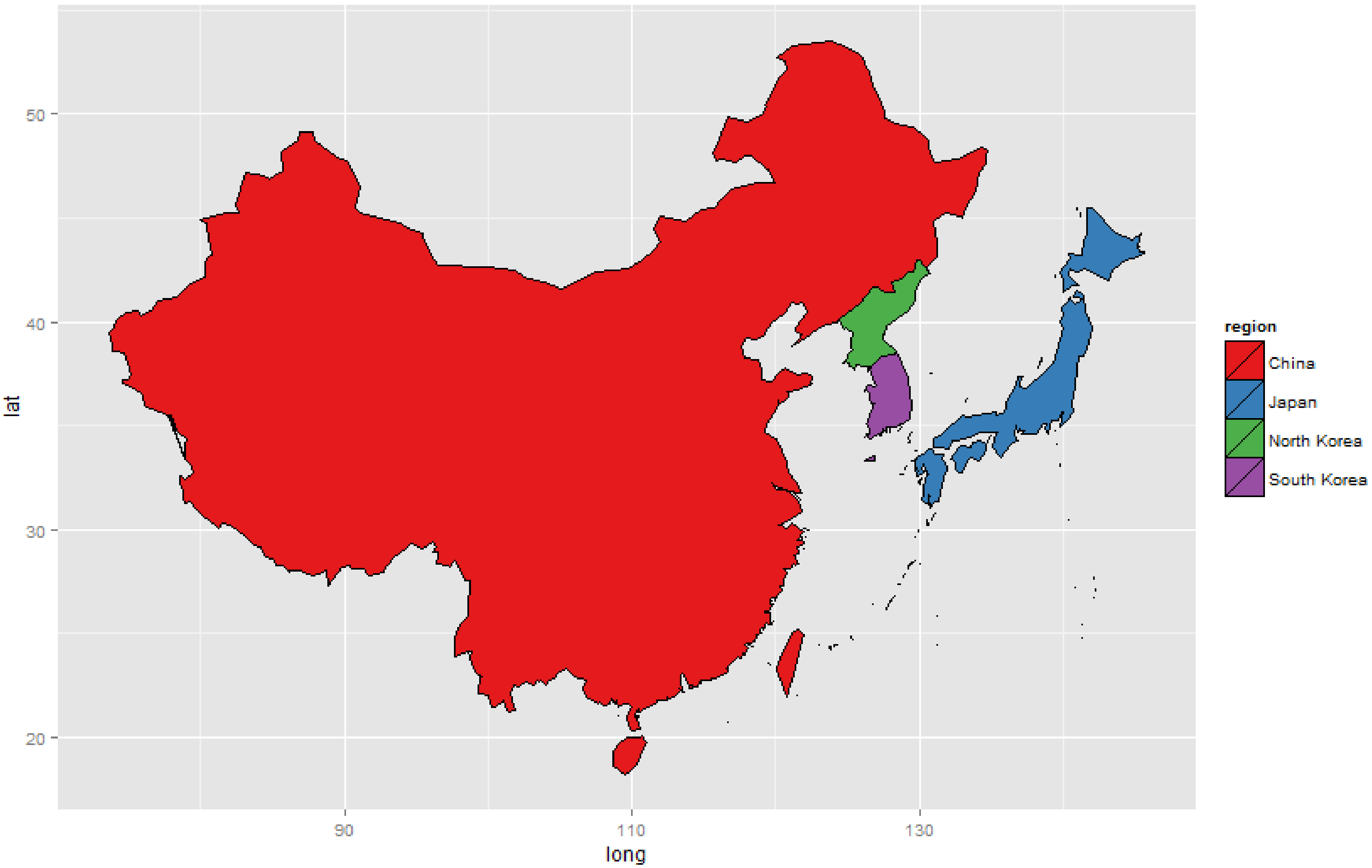


pie(fold)

Creating a Map 绘制地图

```
>install.packages("maps")

>library(maps)

>east_asia <- map_data("world", region=c("Japan",
"China", "North Korea","South Korea")) # Map region
to fill color

>ggplot(east_asia, aes(x=long, y=lat, group=group,
fill=region)) +geom_polygon(colour="black")
+scale_fill_brewer(palette="Set1")
```

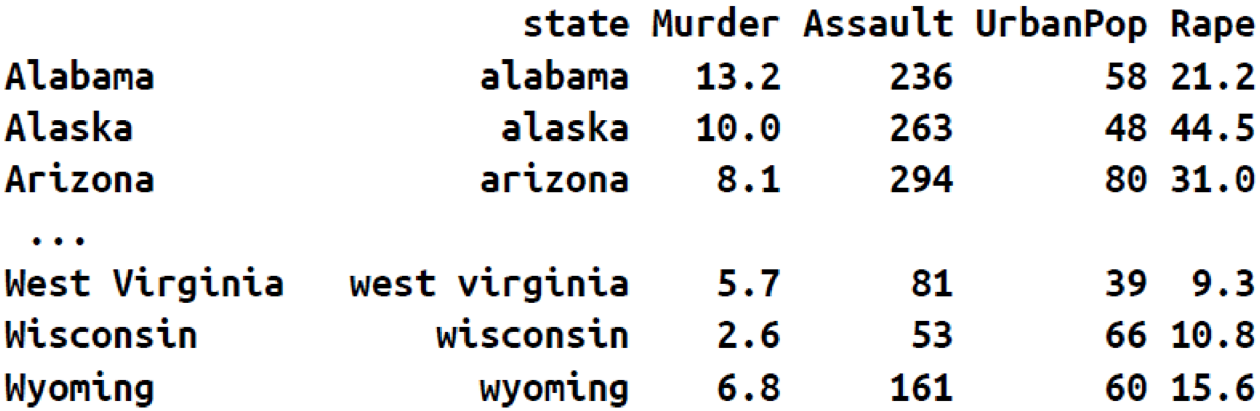
Creating a Map 绘制地图



Creating a Choropleth Map 绘制分区统计图

>crimes <- data.frame(state = tolower(rownames(USArrests)), USArrests) # Transform the USArrests data set to the correct format

>crimes



```
                       state Murder Assault UrbanPop Rape
Alabama              alabama   13.2     236       58 21.2
Alaska                alaska   10.0     263       48 44.5
Arizona              arizona    8.1     294       80 31.0
 ...
West Virginia  west virginia    5.7      81       39  9.3
Wisconsin          wisconsin    2.6      53       66 10.8
Wyoming              wyoming    6.8     161       60 15.6
```

>library(maps) # For map data

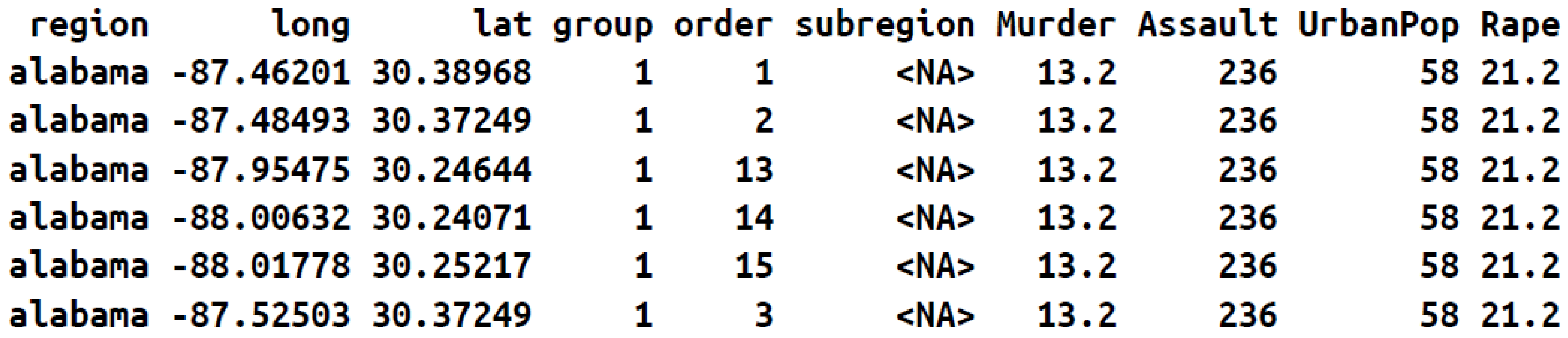
>states_map <- map_data("state") # Merge the data sets together

>crime_map <- merge(states_map, crimes, by.x="region", by.y="state")

# After merging, the order has changed, which would lead to polygons drawn in the incorrect order. So, we sort the data.

Creating a Choropleth Map 绘制分区统计图

>head(crime_map)

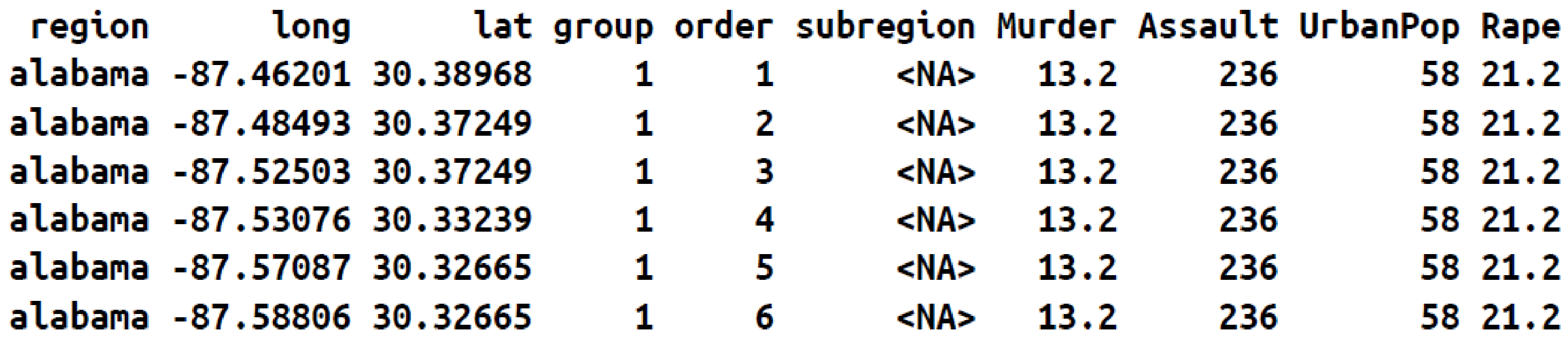


```
 region        long      lat group order subregion Murder Assault UrbanPop Rape
alabama -87.46201 30.38968     1     1      <NA>   13.2     236       58 21.2
alabama -87.48493 30.37249     1     2      <NA>   13.2     236       58 21.2
alabama -87.95475 30.24644     1    13      <NA>   13.2     236       58 21.2
alabama -88.00632 30.24071     1    14      <NA>   13.2     236       58 21.2
alabama -88.01778 30.25217     1    15      <NA>   13.2     236       58 21.2
alabama -87.52503 30.37249     1     3      <NA>   13.2     236       58 21.2
```

>library(plyr) # For arrange() function. Sort by group, then order

>crime_map <- arrange(crime_map, group, order)

>head(crime_map)

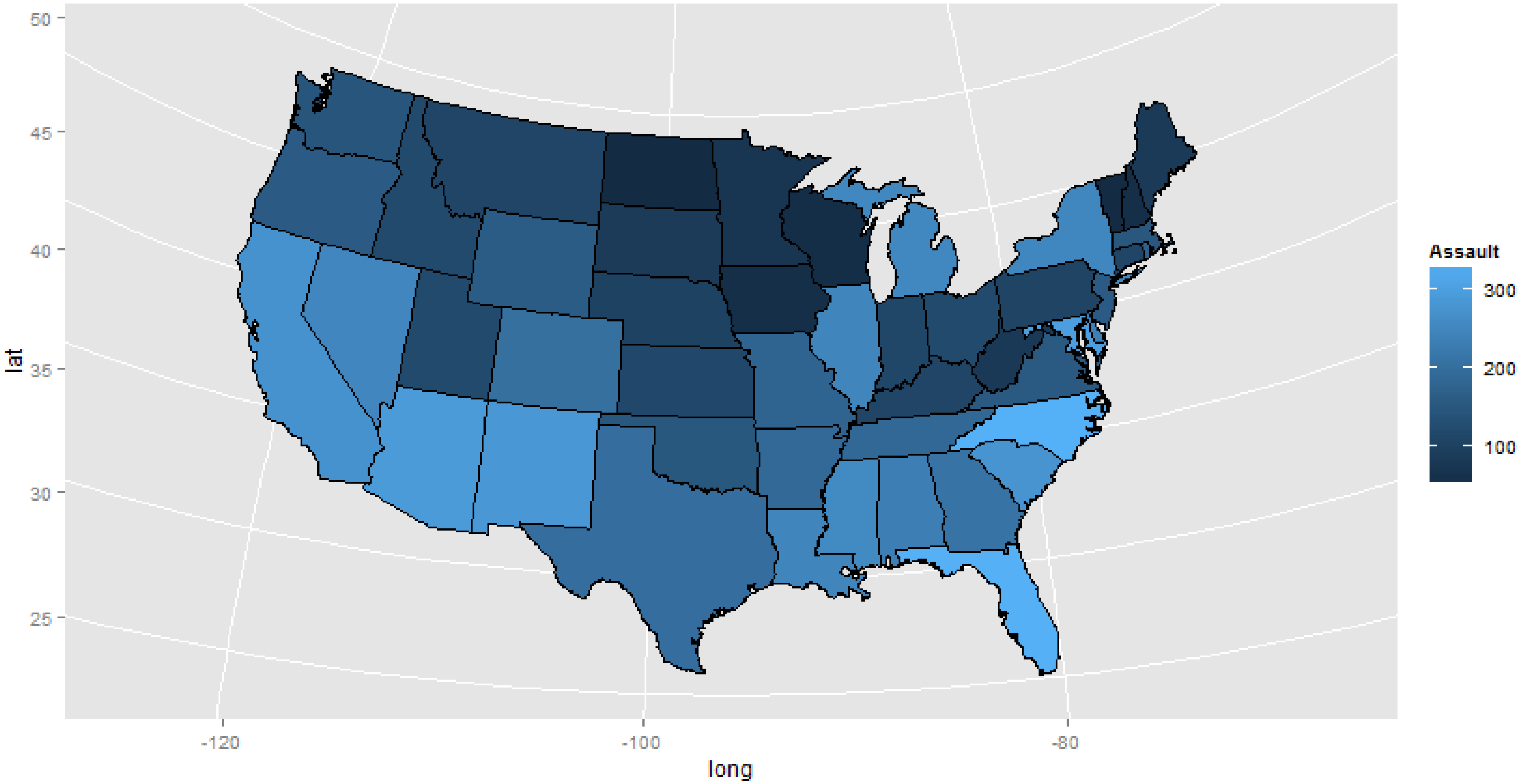


```
 region        long      lat group order subregion Murder Assault UrbanPop Rape
alabama -87.46201 30.38968     1     1      <NA>   13.2     236       58 21.2
alabama -87.48493 30.37249     1     2      <NA>   13.2     236       58 21.2
alabama -87.52503 30.37249     1     3      <NA>   13.2     236       58 21.2
alabama -87.53076 30.33239     1     4      <NA>   13.2     236       58 21.2
alabama -87.57087 30.32665     1     5      <NA>   13.2     236       58 21.2
alabama -87.58806 30.32665     1     6      <NA>   13.2     236       58 21.2
```

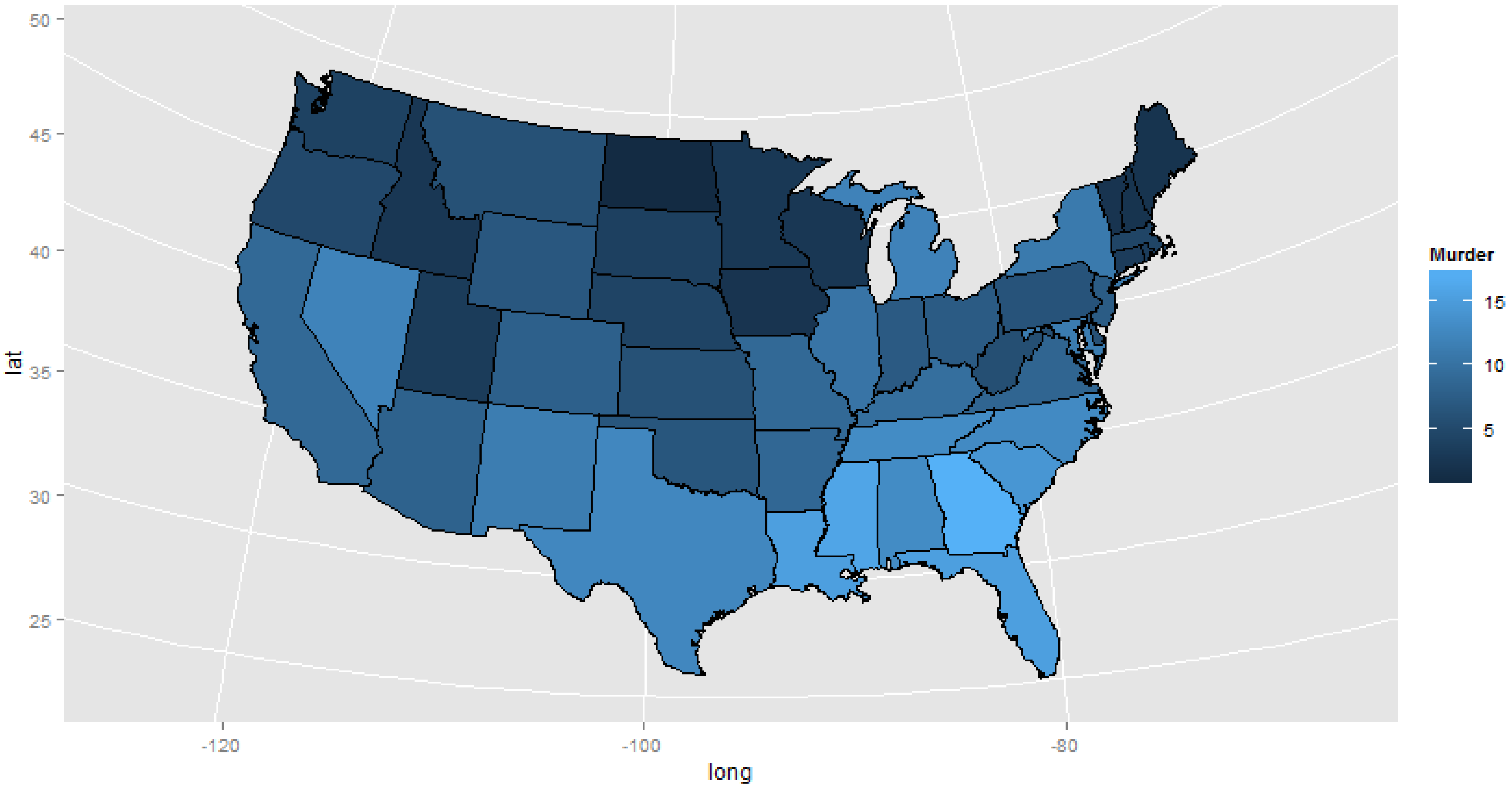
Creating a Choropleth Map 绘制分区统计图

```
>install.packages("mapproj")

>library(mapproj)

>ggplot(crime_map, aes(x=long, y=lat, group=group,
fill=Assault)) +geom_polygon(colour="black")
+coord_map("polyconic")
```

Creating a Choropleth Map 绘制分区统计图



Creating a Choropleth Map 绘制分区统计图


50
45
40
35
30
25
lat
-120
-100
-80
long
Murder
15
10
5