R数据可视化—gplot2包 第3周

2013.2.8

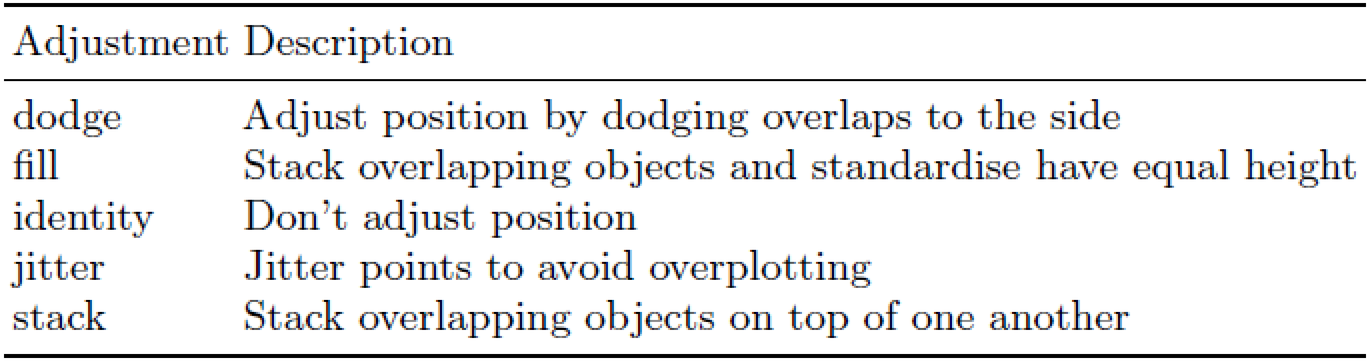
法律声明

【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

Position adjustments



Adjustment	Description
dodge	Adjust position by dodging overlaps to the side
fill	Stack overlapping objects and standardise have equal height
identity	Don't adjust position
jitter	Jitter points to avoid overplotting
stack	Stack overlapping objects on top of one another

例图

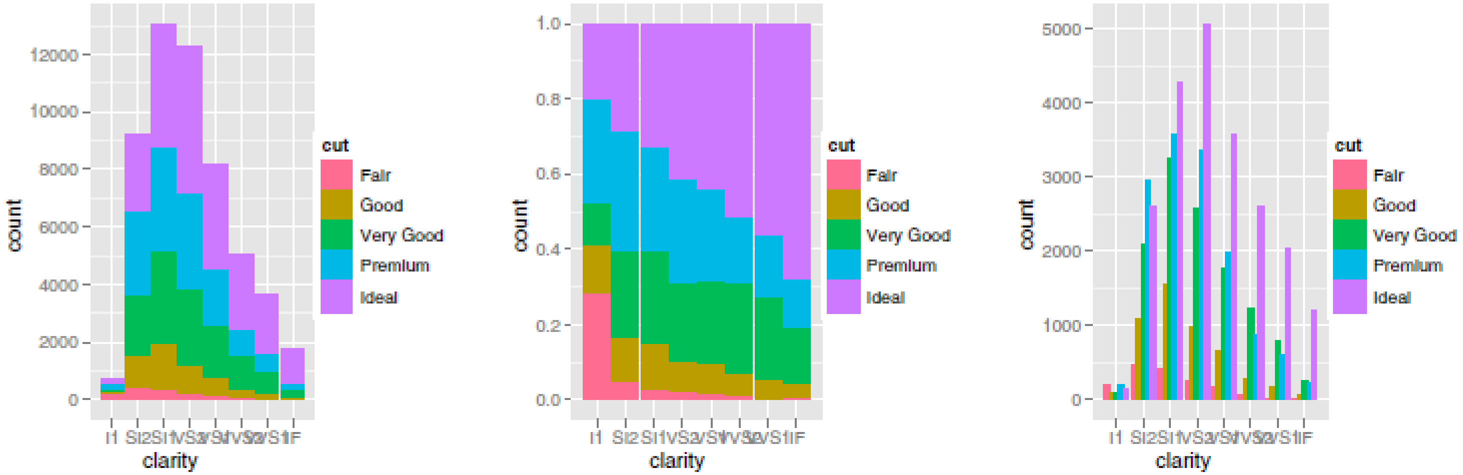
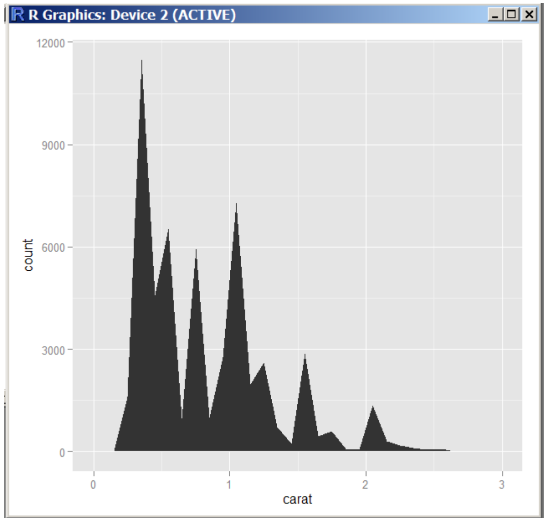


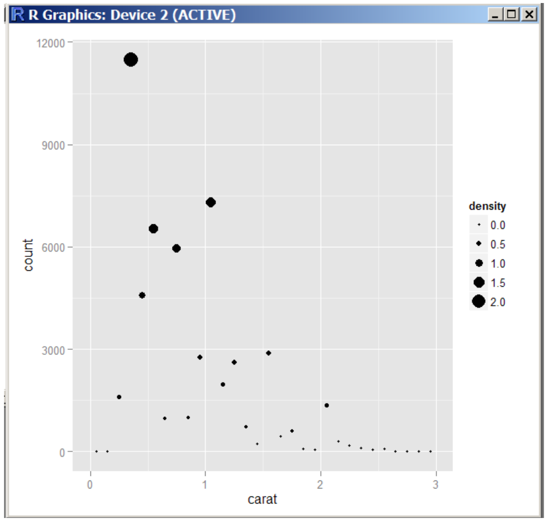
Fig. 4.8: Three position adjustments applied to a bar chart. From left to right, stacking, filling and dodging.

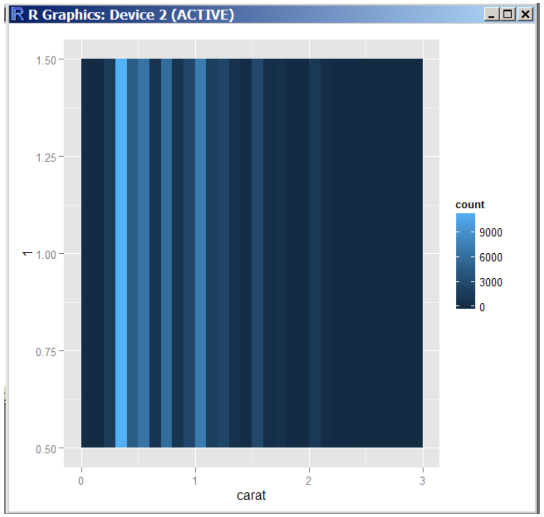
组合geoms和stats

```
d <- ggplot(diamonds, aes(carat)) + xlim(0, 3)
d + stat_bin(aes(ymax = ..count..), binwidth = 0.1, geom = "area")
d + stat_bin(
aes(size = ..density..), binwidth = 0.1,
geom = "point", position="identity"
)
d + stat_bin(
aes(y = 1, fill = ..count..), binwidth = 0.1,
geom = "tile", position="identity"
)
```

例图


R Graphics: Device 2 (ACTIVE)
12000
9000
6000
3000
0
count
0
1
2
3
carat


R Graphics: Device 2 (ACTIVE)
12000
9000
6000
3000
0
count
0
1
2
3
carat
density
0.0
0.5
1.0
1.5
2.0

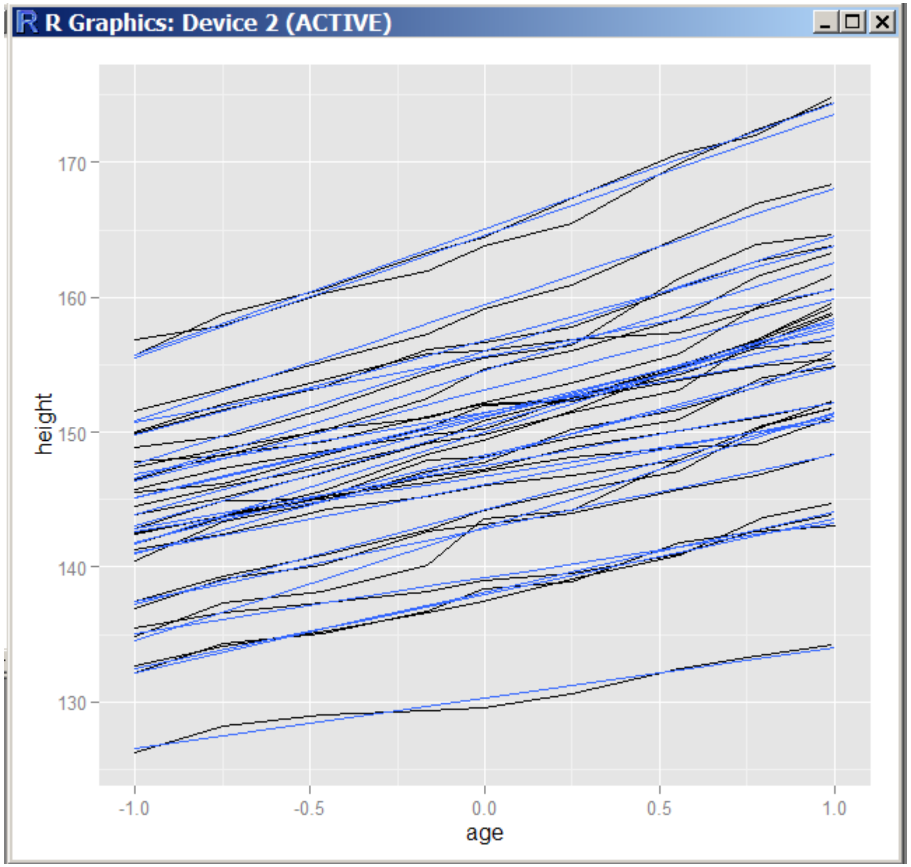

R Graphics: Device 2 (ACTIVE)
1.50
1.25
1.00
0.75
0.50
1
0
1
2
3
carat
count
9000
6000
3000
0

一个复杂例子

```
model <- lme(height ~ age, data = Oxboys, random = ~ 1 + age | Subject)
oplot <- ggplot(Oxboys, aes(age, height, group = Subject)) + geom_line()
age_grid <- seq(-1, 1, length = 10)
subjects <- unique(Oxboys$Subject)
preds <- expand.grid(age = age_grid, Subject = subjects)
preds$height <- predict(model, preds)
```

续

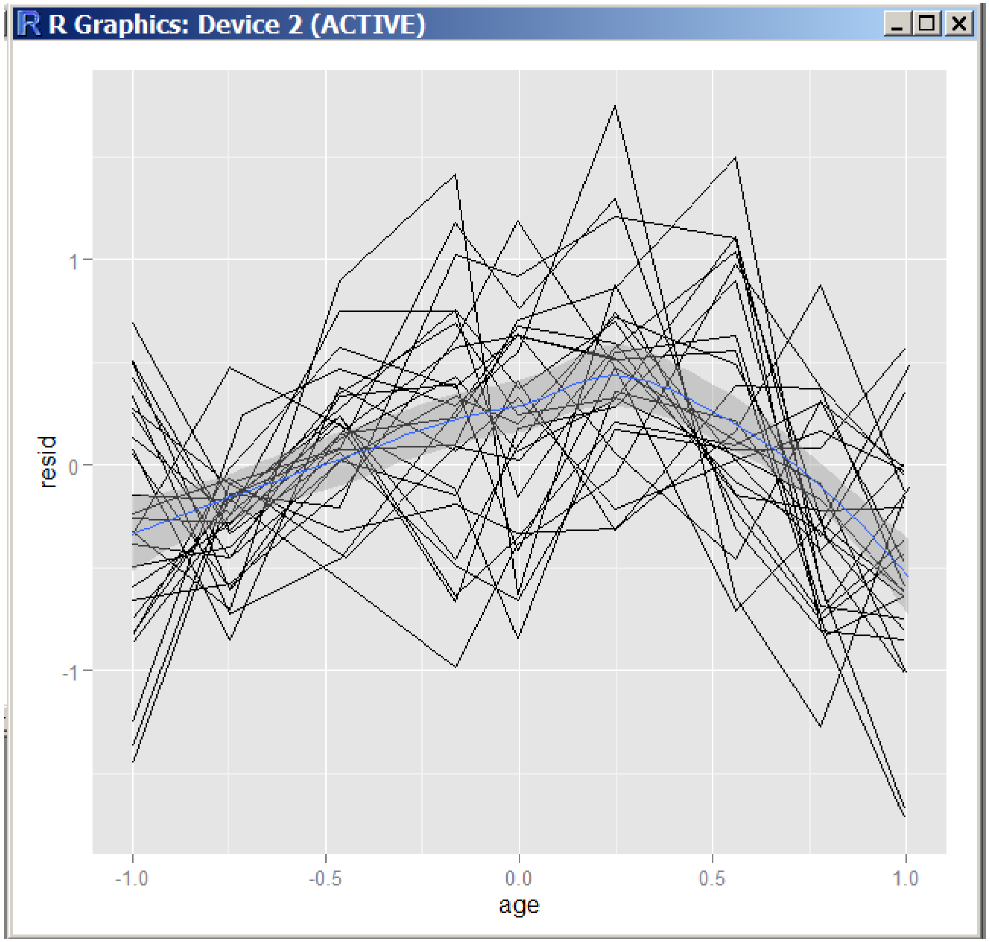
```
oplot + geom_line(data = preds, colour = "#3366FF", size= 0.4)
```



2013.2.8

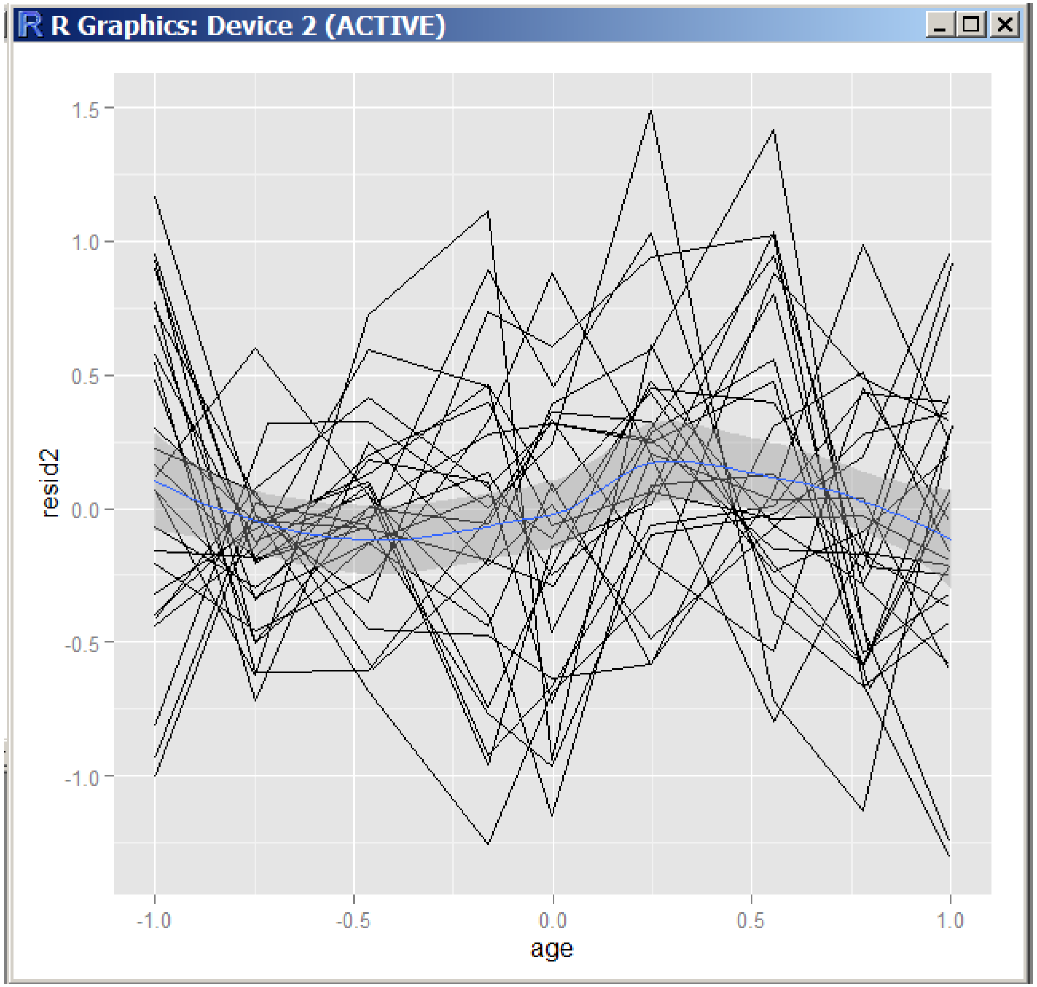
续

```
Oxboys$fitted <-
    predict(model)
Oxboys$resid <-
    with(Oxboys, fitted
    - height)
oplot %+% Oxboys +
    aes(y = resid) +
    geom_smooth(aes(
    group=1))
```



续

```
model2 <- update(model,
    height ~ age + I(age ^ 2))
Oxboys$fitted2 <-
    predict(model2)
Oxboys$resid2 <-
    with(Oxboys, fitted2 -
    height)
oplot %+% Oxboys + aes(y =
    resid2) +
    geom_smooth(aes(group=
    1))
```

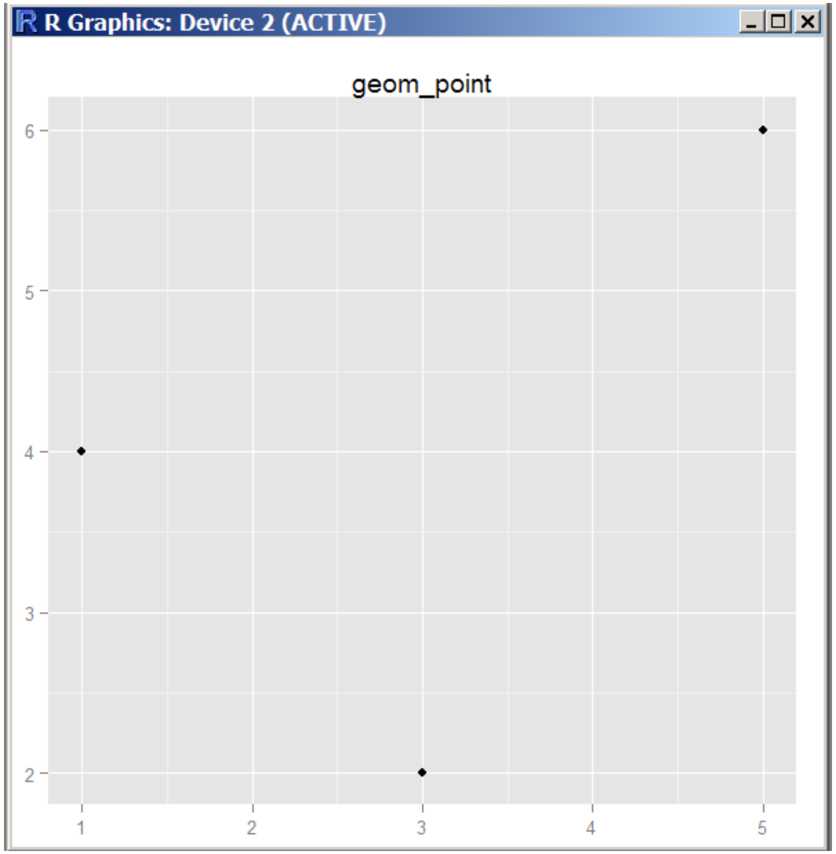


基本作图类型

```
df <- data.frame(
x = c(3, 1, 5),
y = c(2, 4, 6),
label = c("a","b","c")
)
p <- ggplot(df, aes(x, y, label = label)) +
xlab(NULL) + ylab(NULL)
```

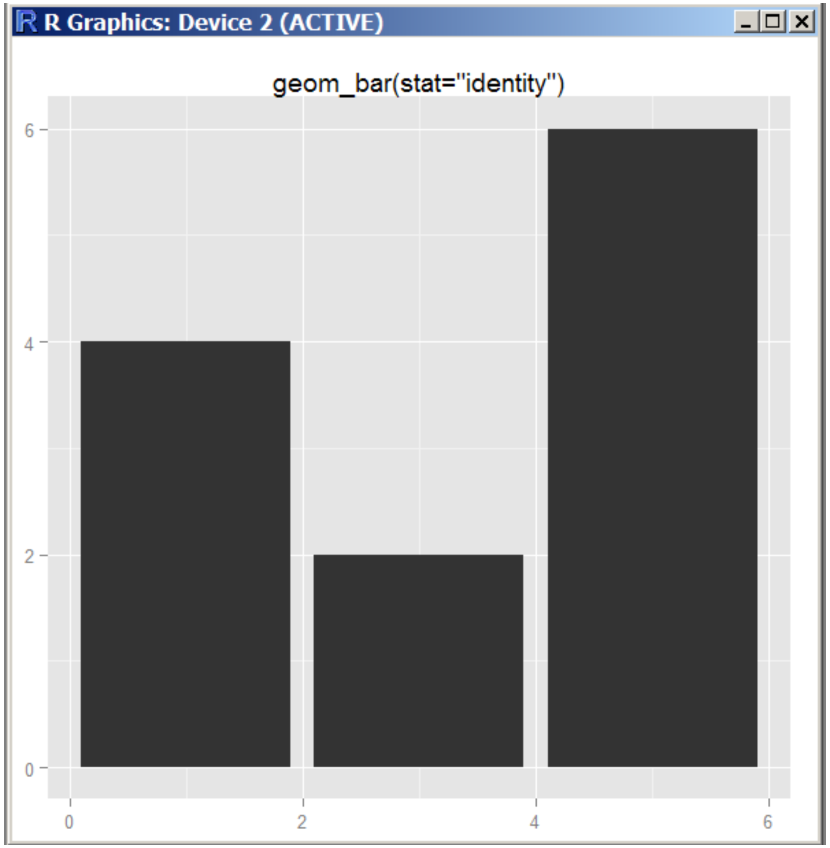
散点图

p + geom_point() + ggtitle("geom_point")



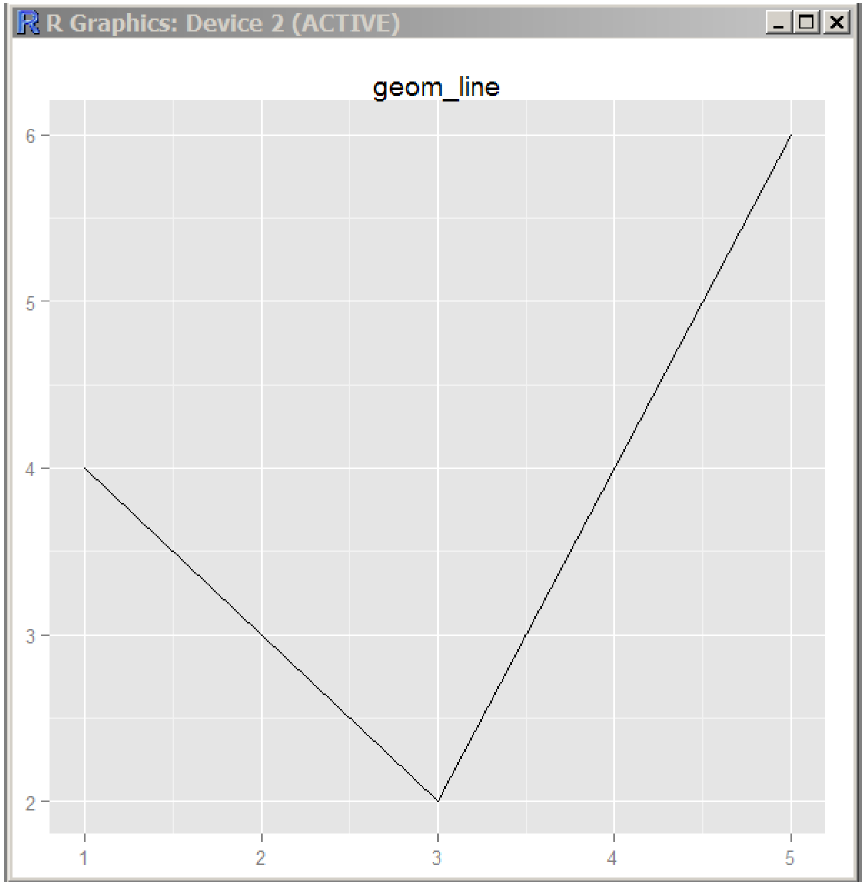
柱形图

```
p + geom_bar(stat="identity") +ggtitle( "geom_bar(stat=\"identity\")")
```



线图

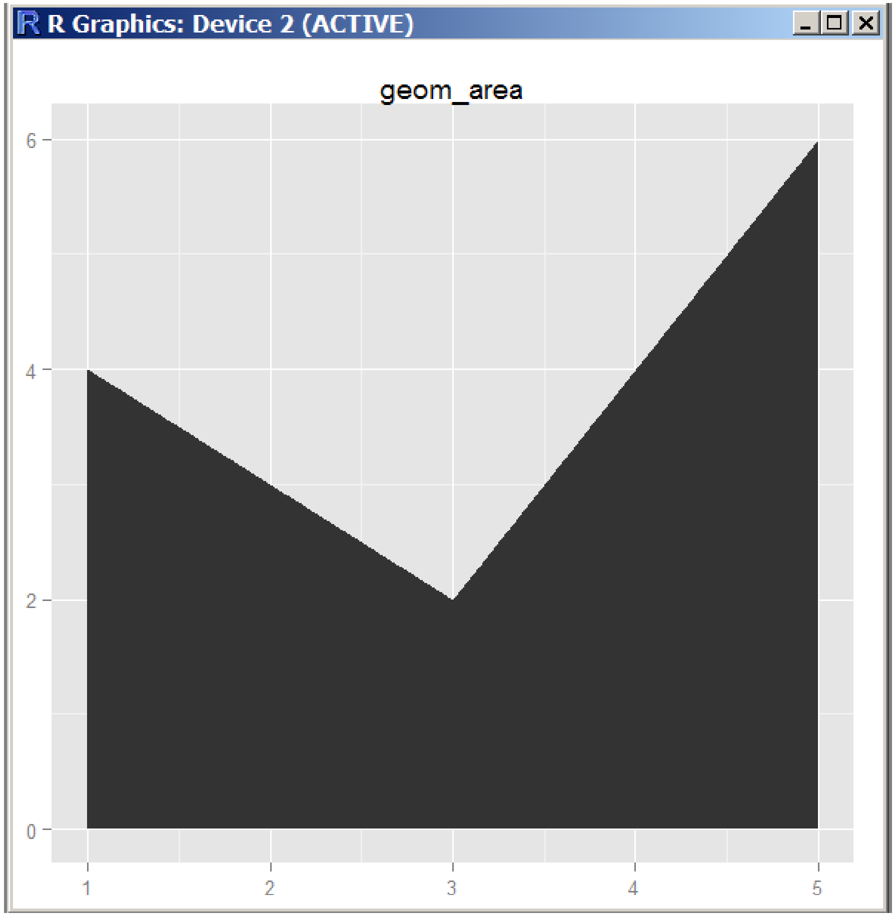
p + geom_line() + ggtitle("geom_line")



2013.2.8

填充图

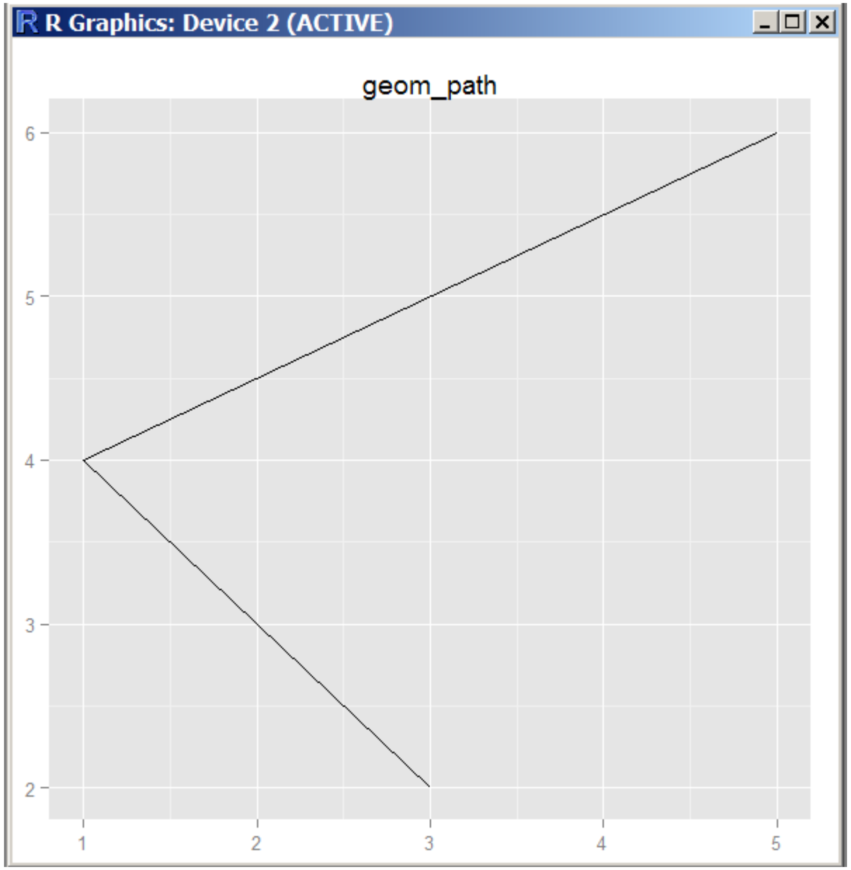
```
p + geom_area() + ggtitle("geom_area")
```



2013.2.8

路径图

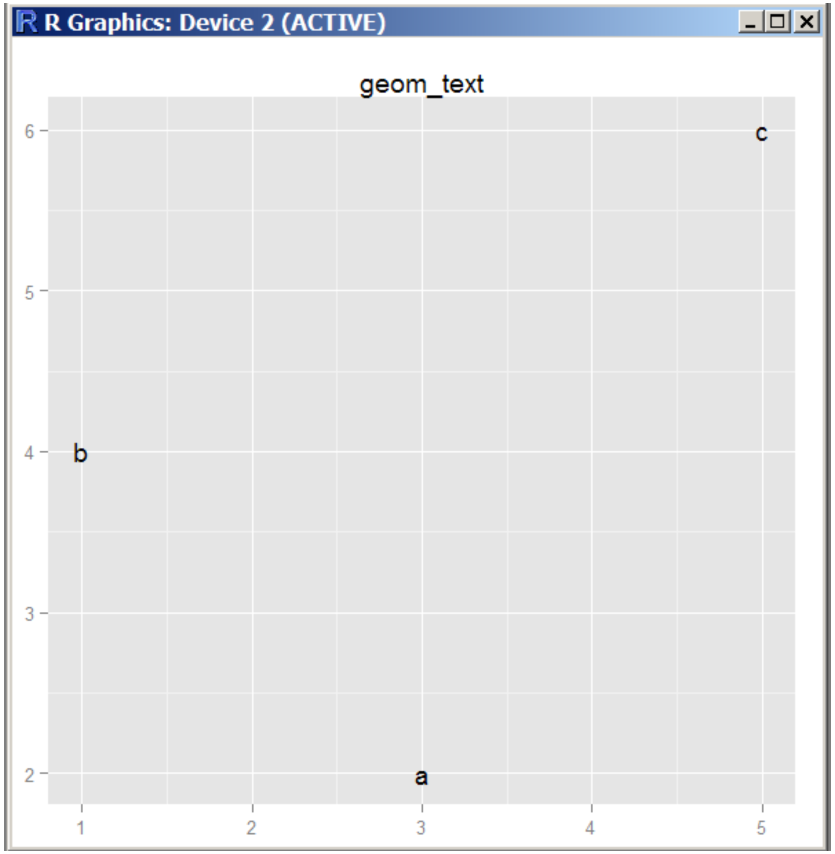
p + geom_path() + ggtitle("geom_path")



2013.2.8

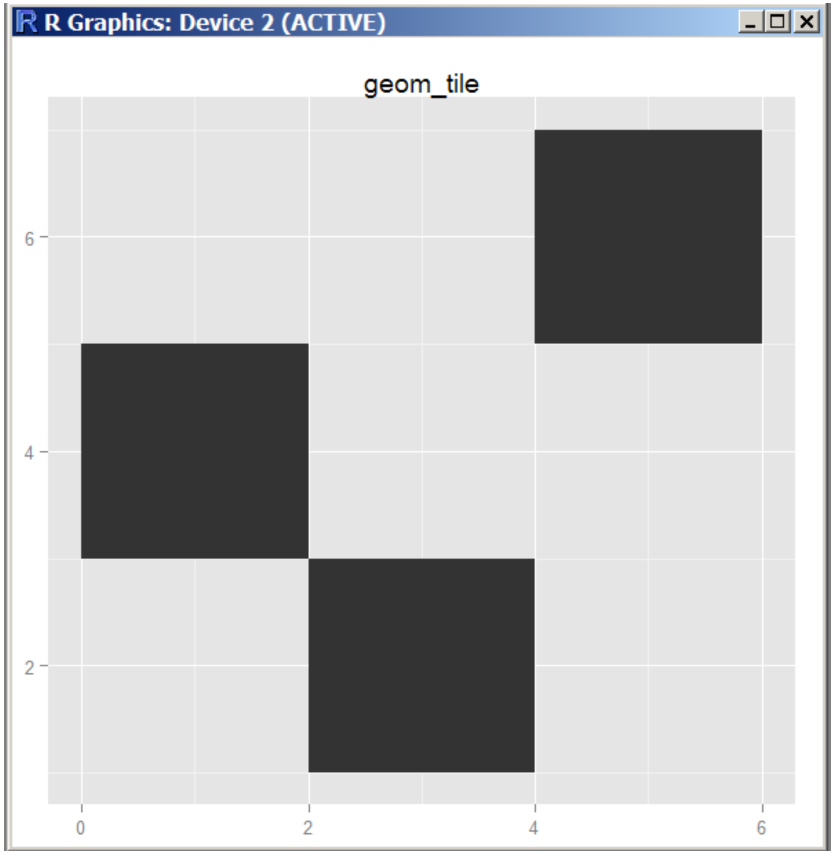
文字标识

```
p + geom_text() + ggtitle("geom_text")
```



Tile plot

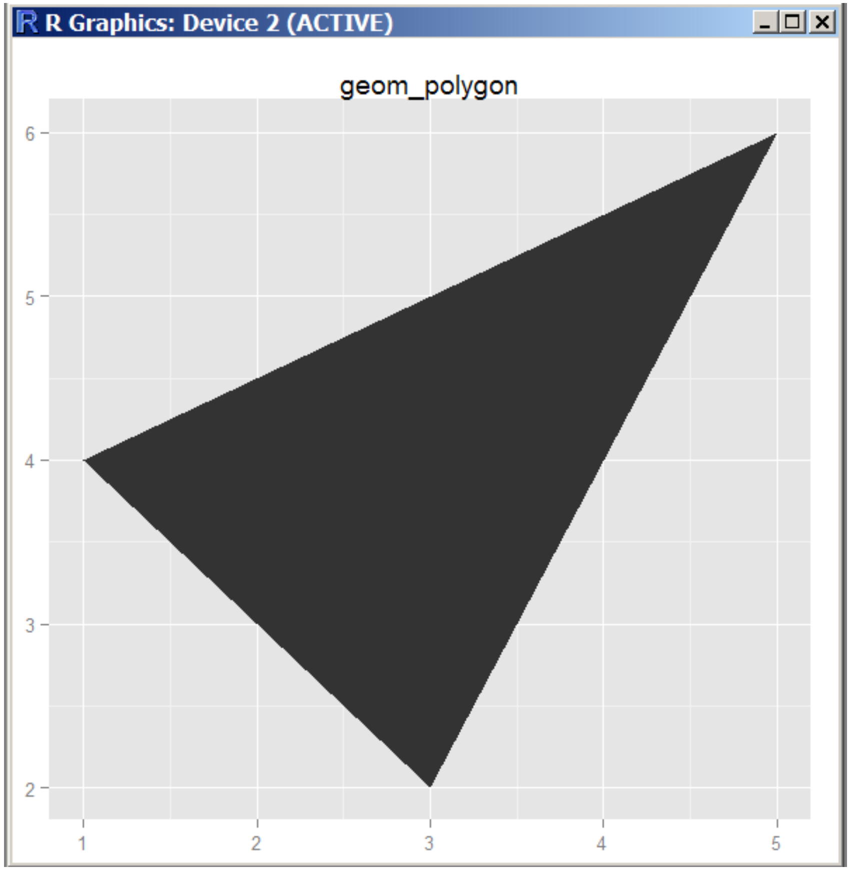
p + geom_tile() + ggtitle("geom_tile")



2013.2.8

Polygon plot

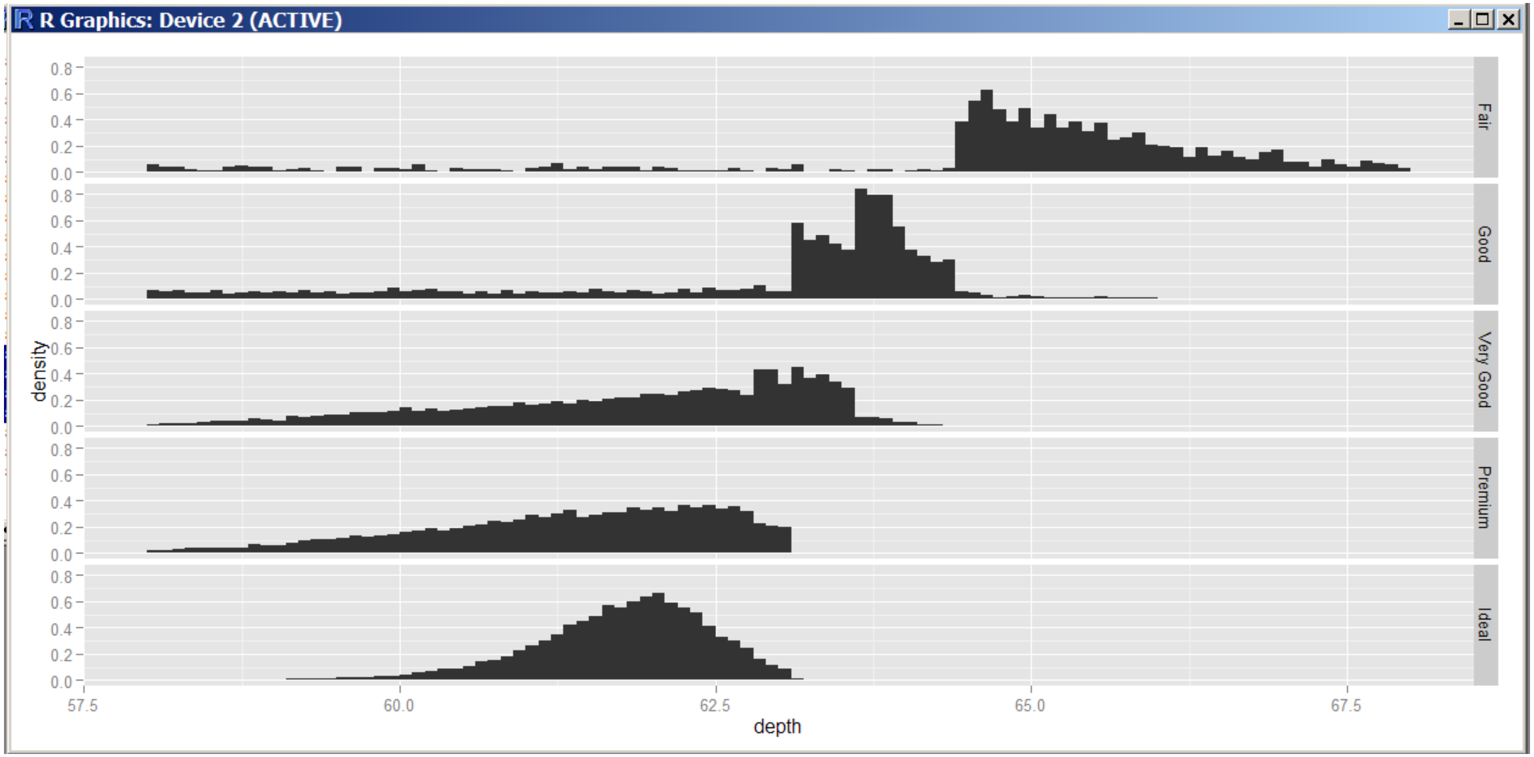
```
p + geom_polygon() + ggtitle("geom_polygon")
```



2013.2.8

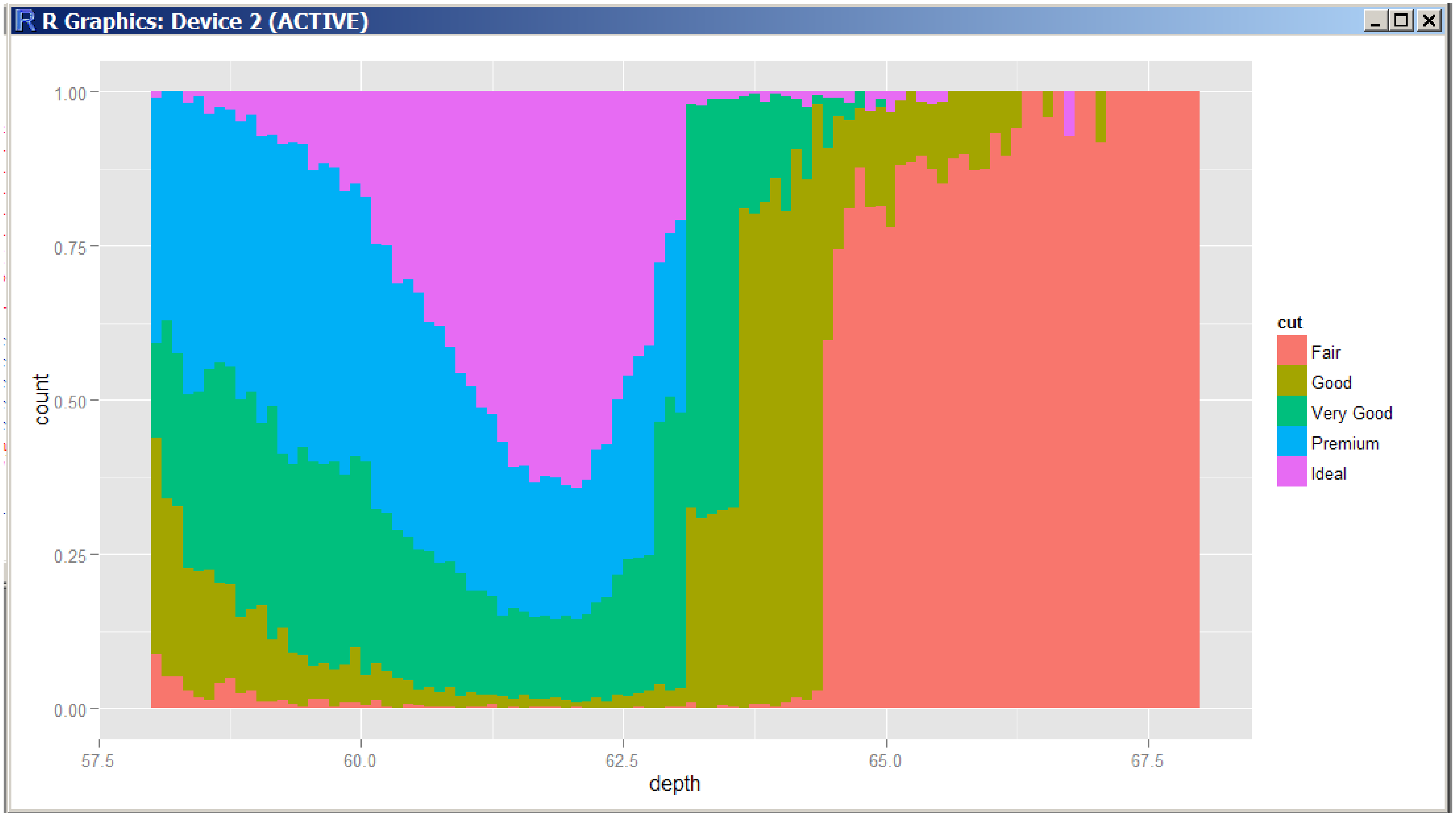
画分布的技巧

```
depth_dist <- ggplot(diamonds, aes(depth)) + xlim(58, 68)
depth_dist +geom_histogram(aes(y = ..density..), binwidth = 0.1) +facet_grid(cut ~ .)
```



分布

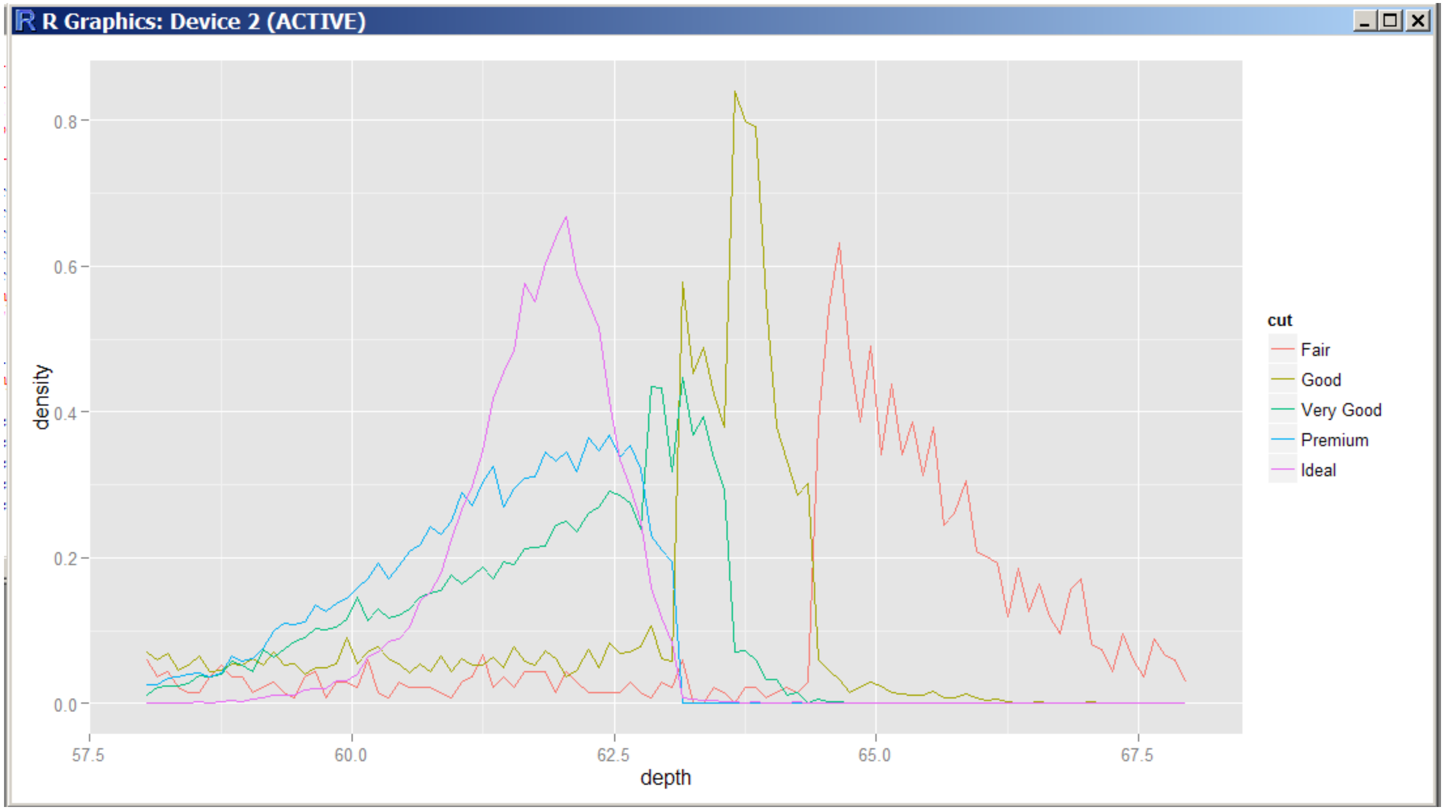
depth_dist + geom_histogram(aes(fill = cut), binwidth = 0.1,position = "fill")



2013.2.8

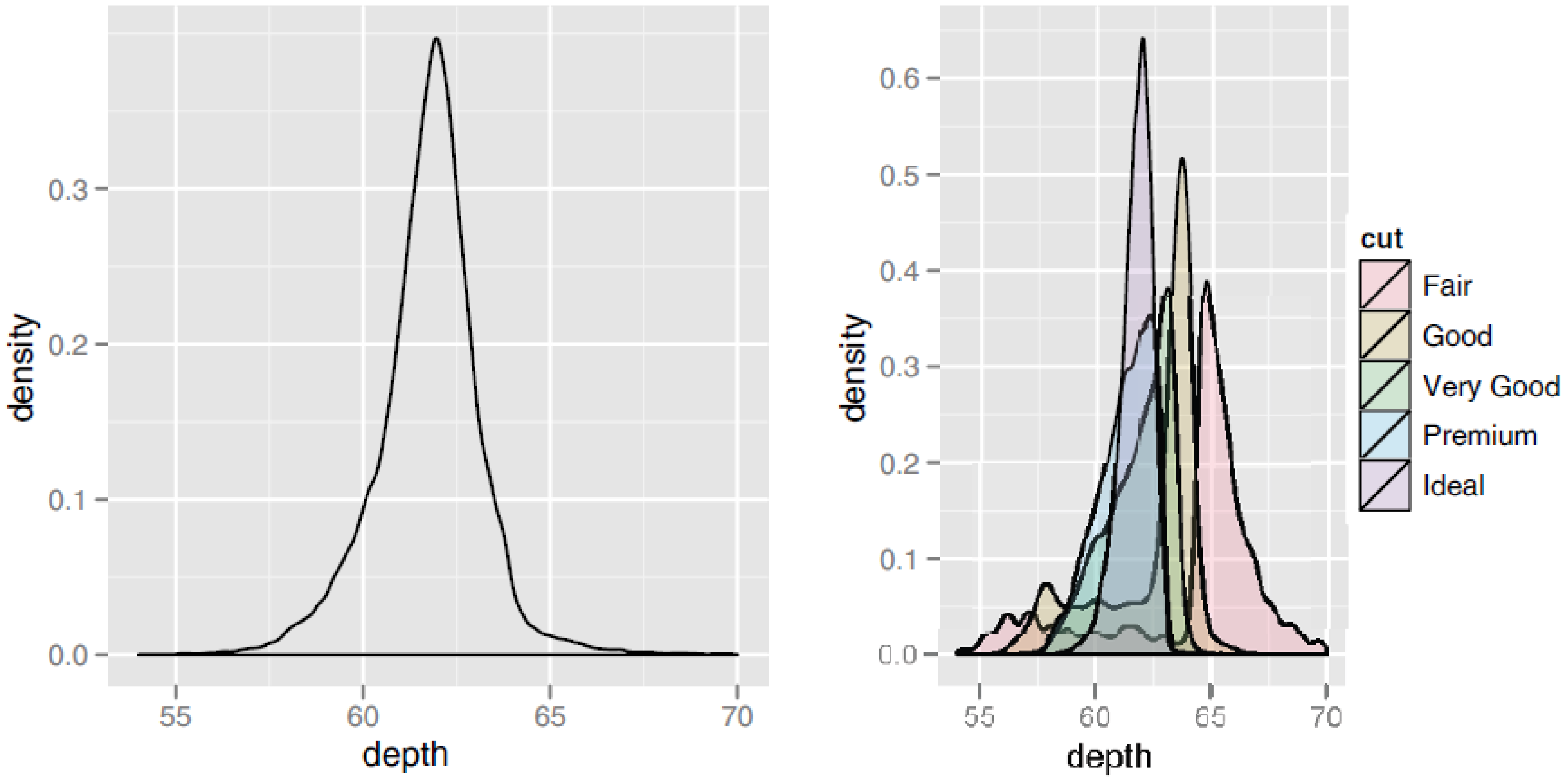
分布

depth_dist + geom_freqpoly(aes(y = ..density.., colour = cut),binwidth = 0.1)



2013.2.8

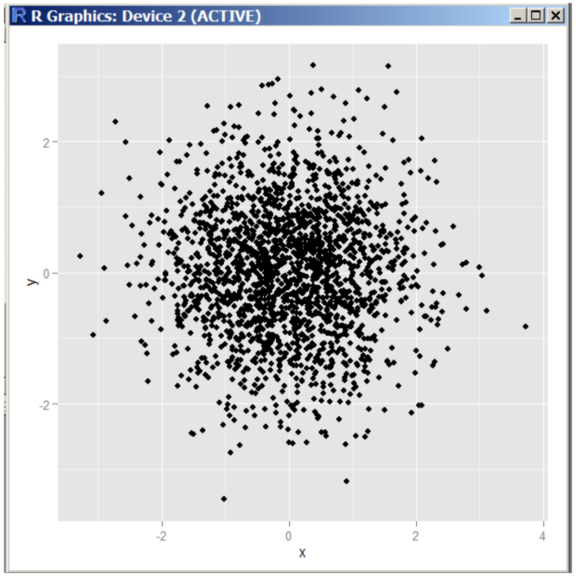
overplotting

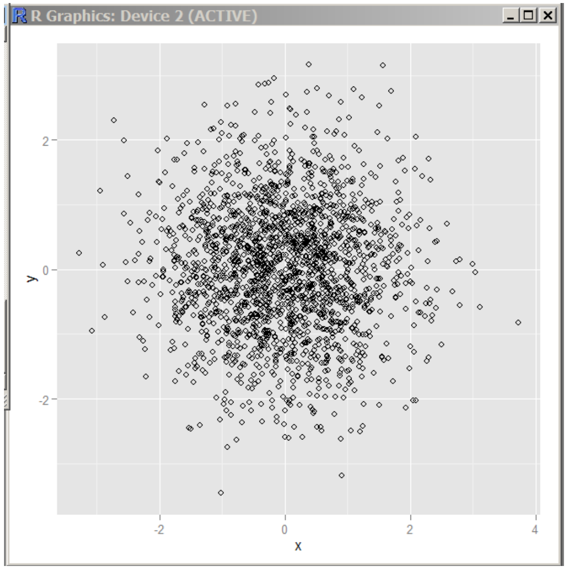


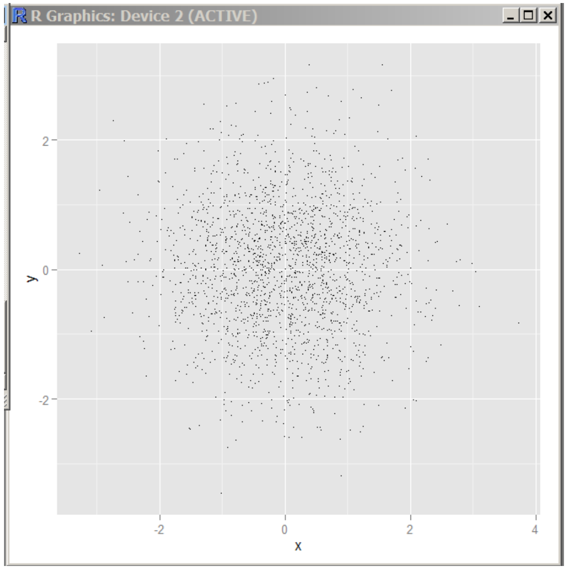
2013.2.8

通过散点形状和大小控制重叠

```
df <- data.frame(x = rnorm(2000), y = rnorm(2000))
norm <- ggplot(df, aes(x, y))
norm + geom_point()
norm + geom_point(shape = 1)
norm + geom_point(shape = ".") # Pixel sized
```

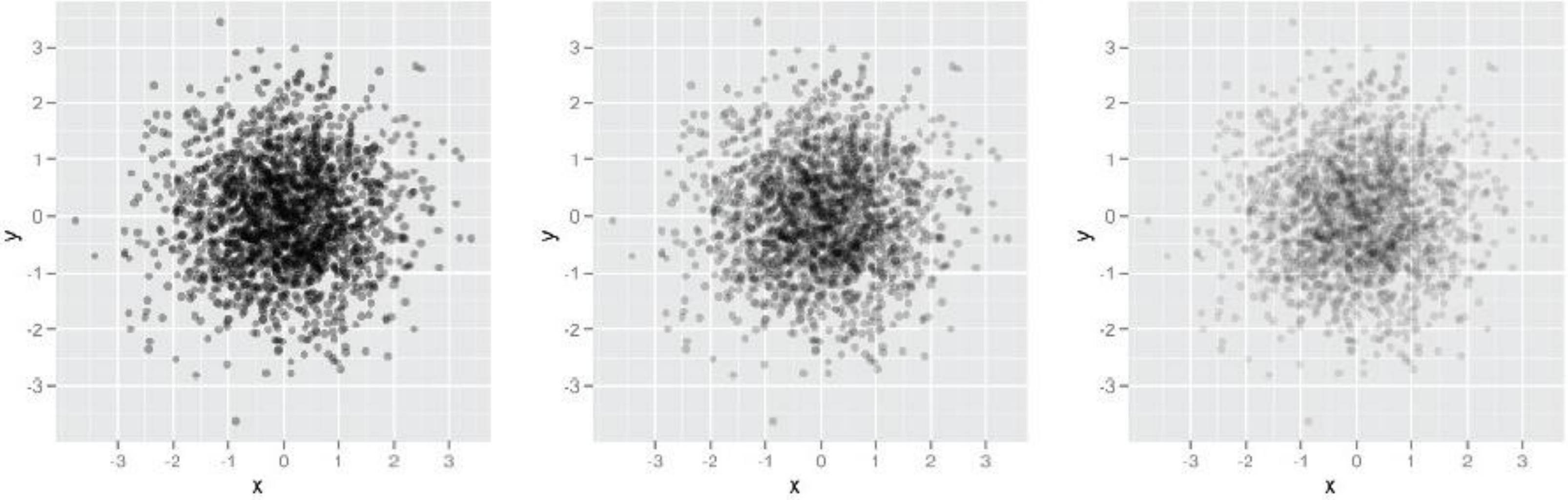






通过透明度控制

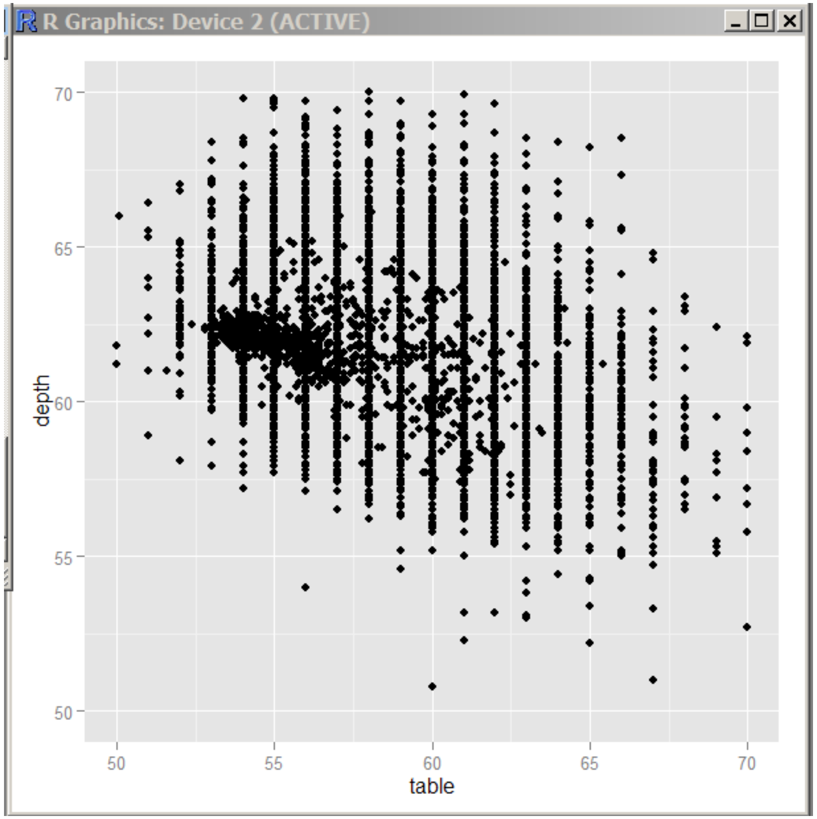
```
norm + geom_point(colour = alpha("black", 1/3))
norm + geom_point(colour = alpha("black", 1/5))
norm + geom_point(colour = alpha("black", 1/10))
```



扰动（jitter）表示法

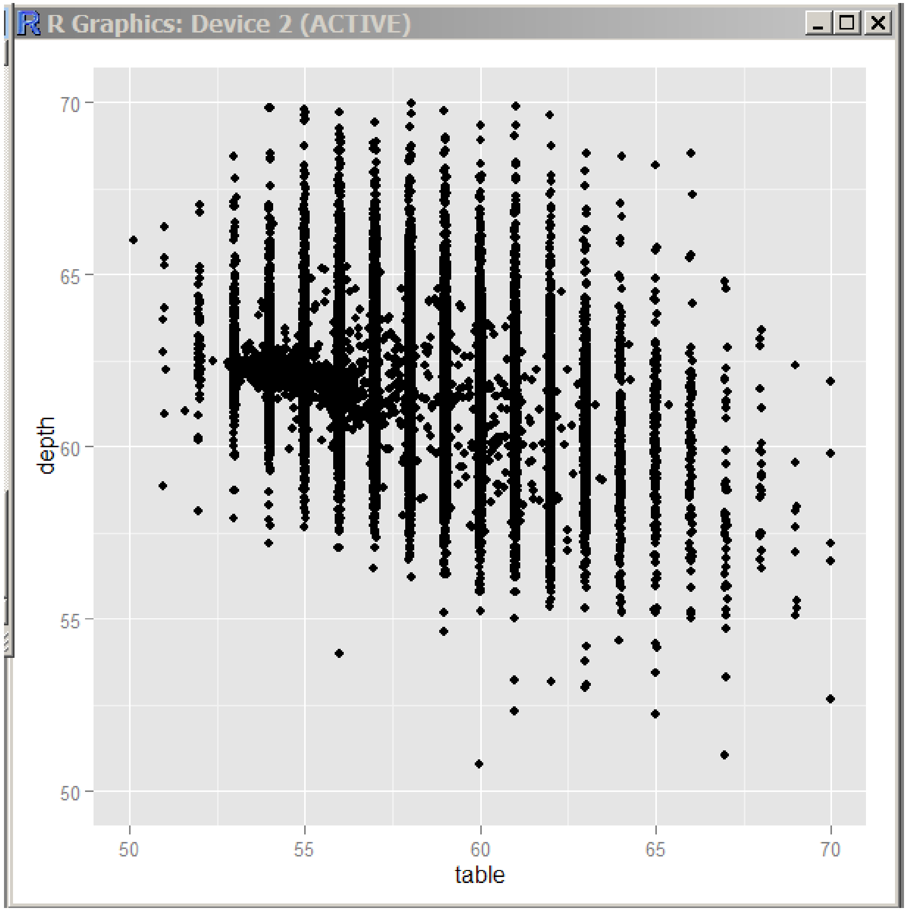
```
td <- ggplot(diamonds, aes(table, depth)) +xlim(50, 70) + ylim(50, 70)

td + geom_point()
```



扰动（jitter）表示法

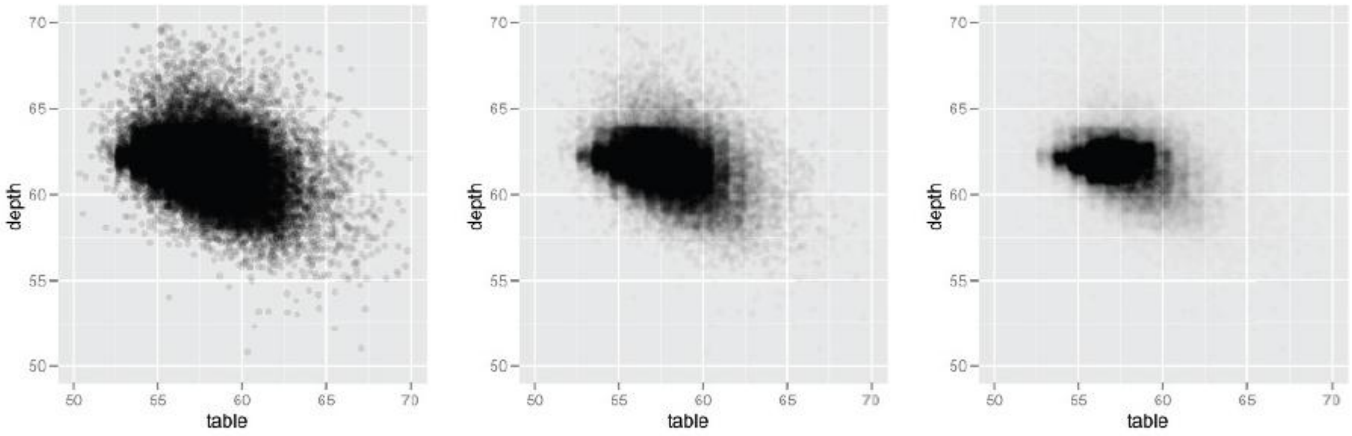
td + geom_jitter()



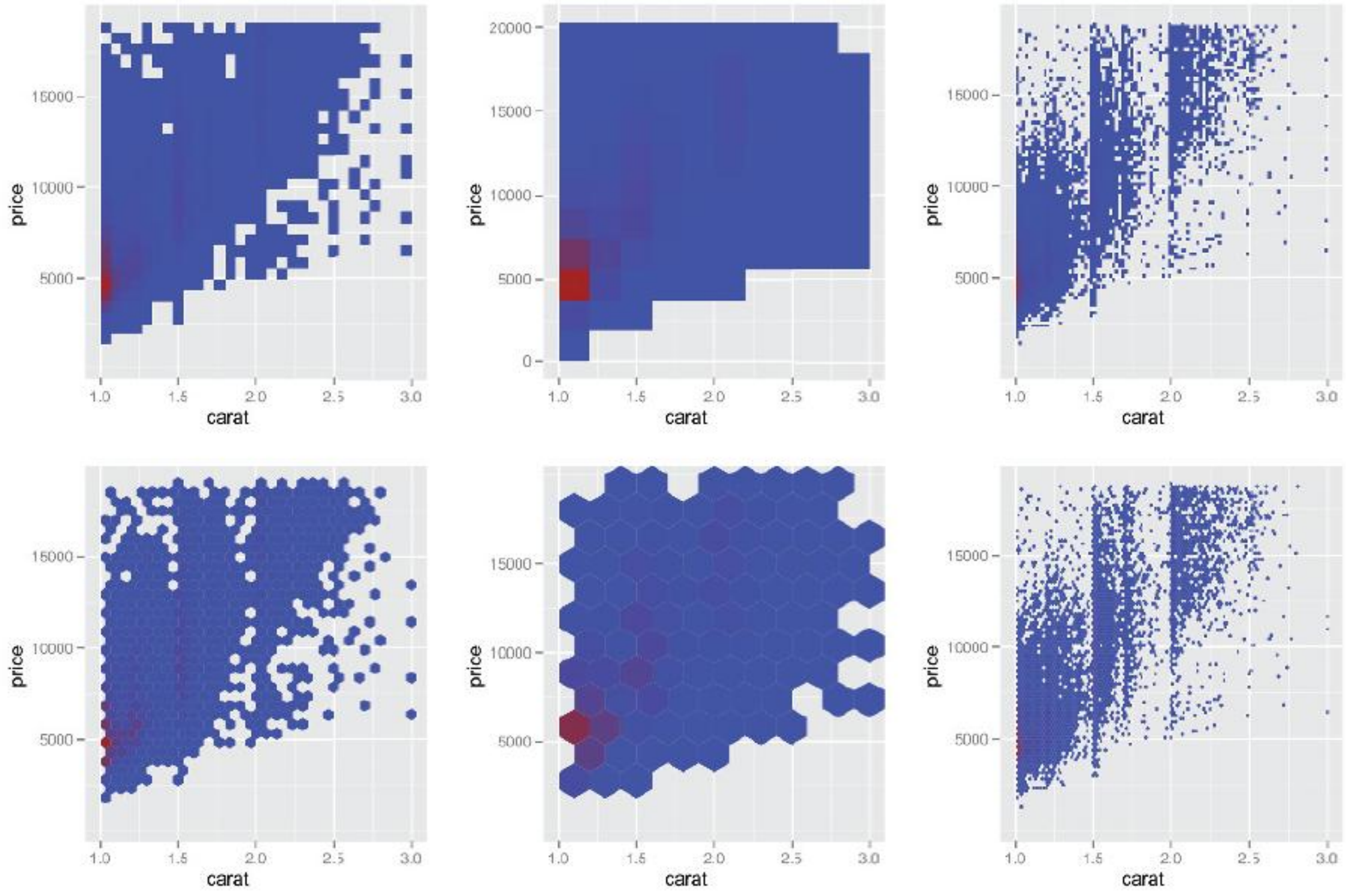
2013.2.8

扰动（jitter）表示法

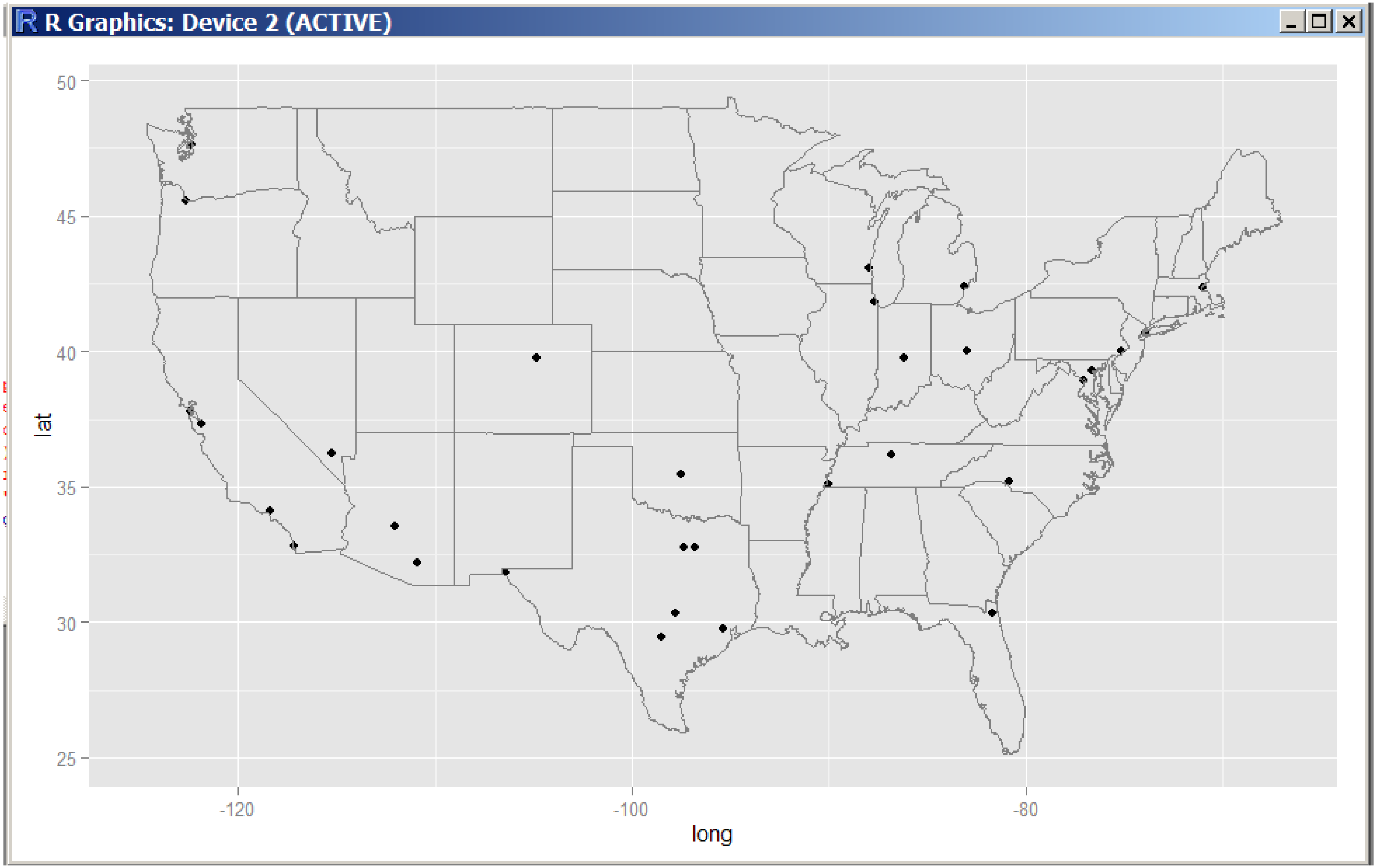
```
jit <- position_jitter(width = 0.5)
td + geom_jitter(position = jit)
td + geom_jitter(position = jit, colour = alpha("black", 1/10))
td + geom_jitter(position = jit, colour = alpha("black", 1/50))
td + geom_jitter(position = jit, colour = alpha("black", 1/200))
```



容器（bin）表示法



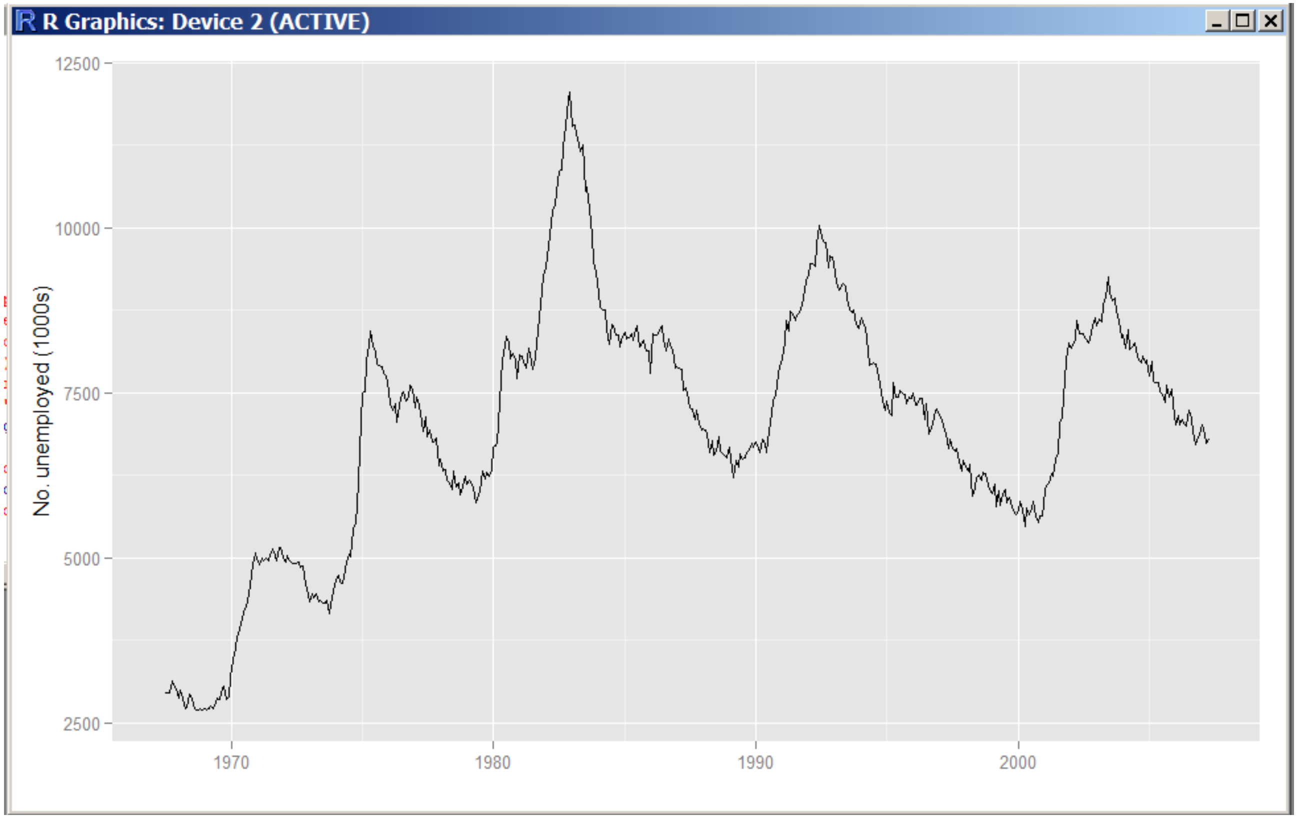
地图



2013.2.8

标记

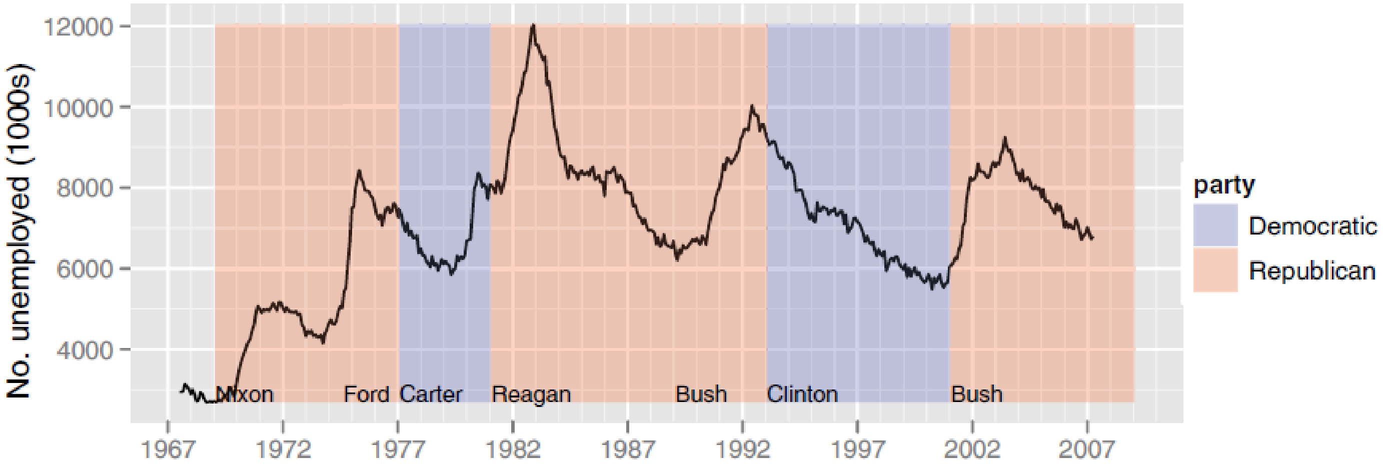
```
(unemp <- qplot(date, unemploy, data=economics, geom="line",xlab = "", ylab =
    "No. unemployed (1000s)"))
```



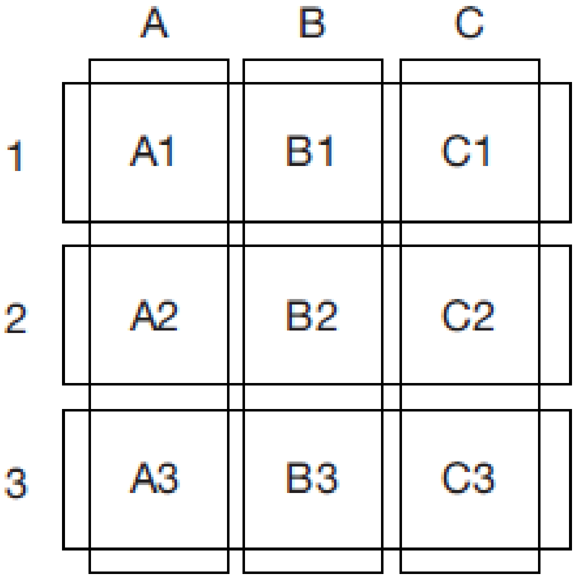
2013.2.8

在图上做标记

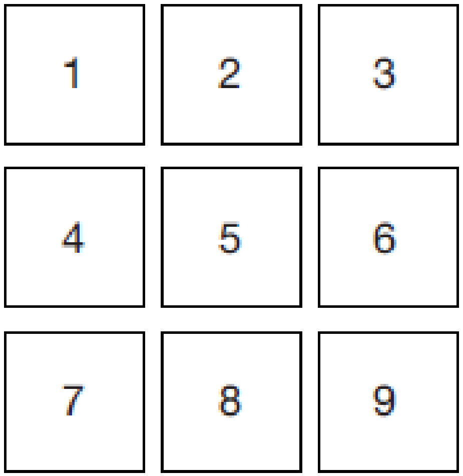
- geom_vline, geom_hline: add vertical or horizontal lines to a plot.
- geom_abline: add lines with arbitrary slope and intercept to a plot.
- geom_rect for highlighting interesting rectangular regions of the plot. geom_rect has aesthetics xmin, xmax, ymin and ymax.
- geom_line, geom_path and geom_segment for adding lines. All these geoms have an arrow parameter, which allows you to place an arrowhead on the line. You create arrowheads with the arrow() function, which has arguments angle, length, ends and type.



分组



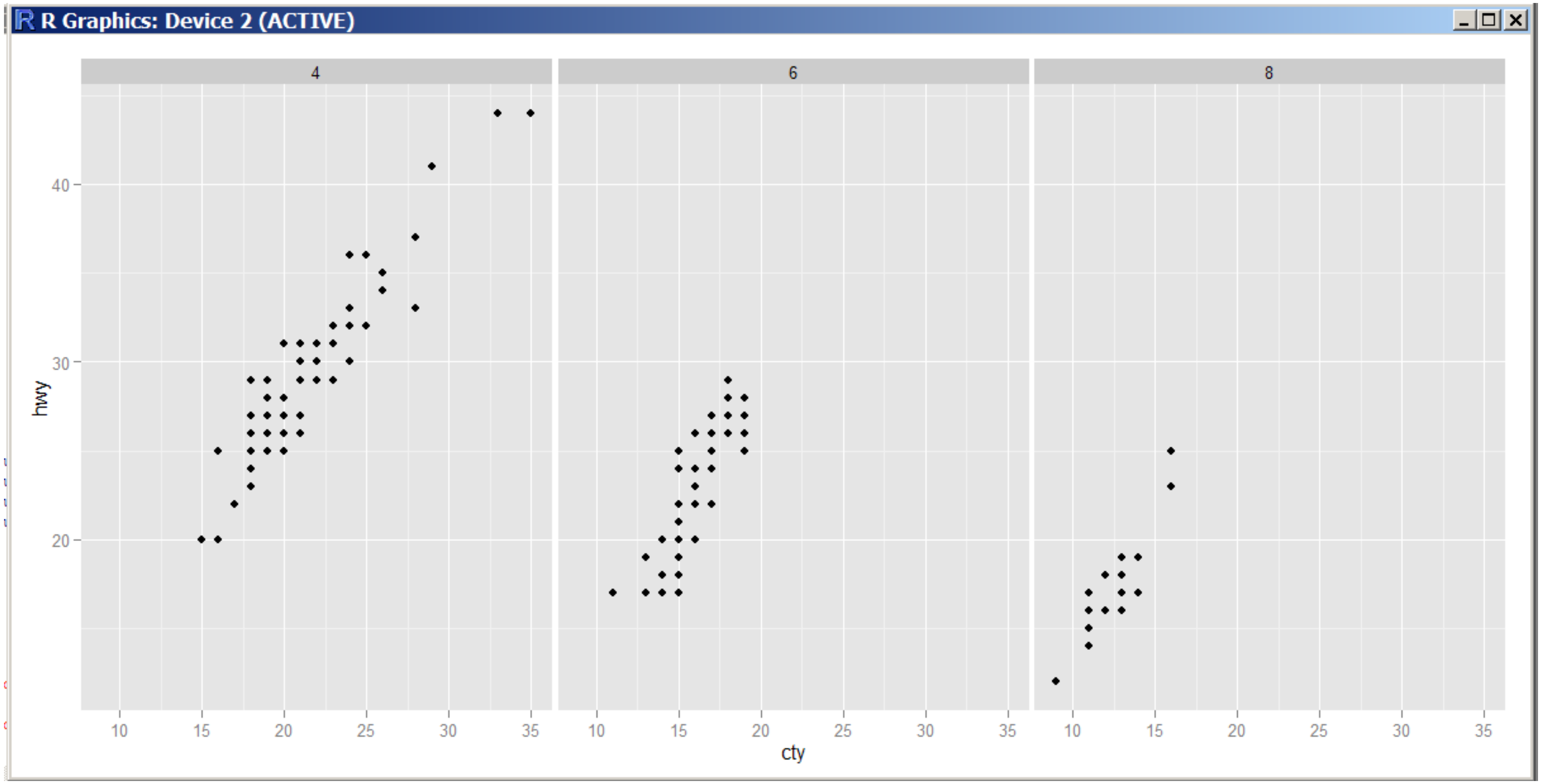
facet_grid



facet_wrap

Facet grid

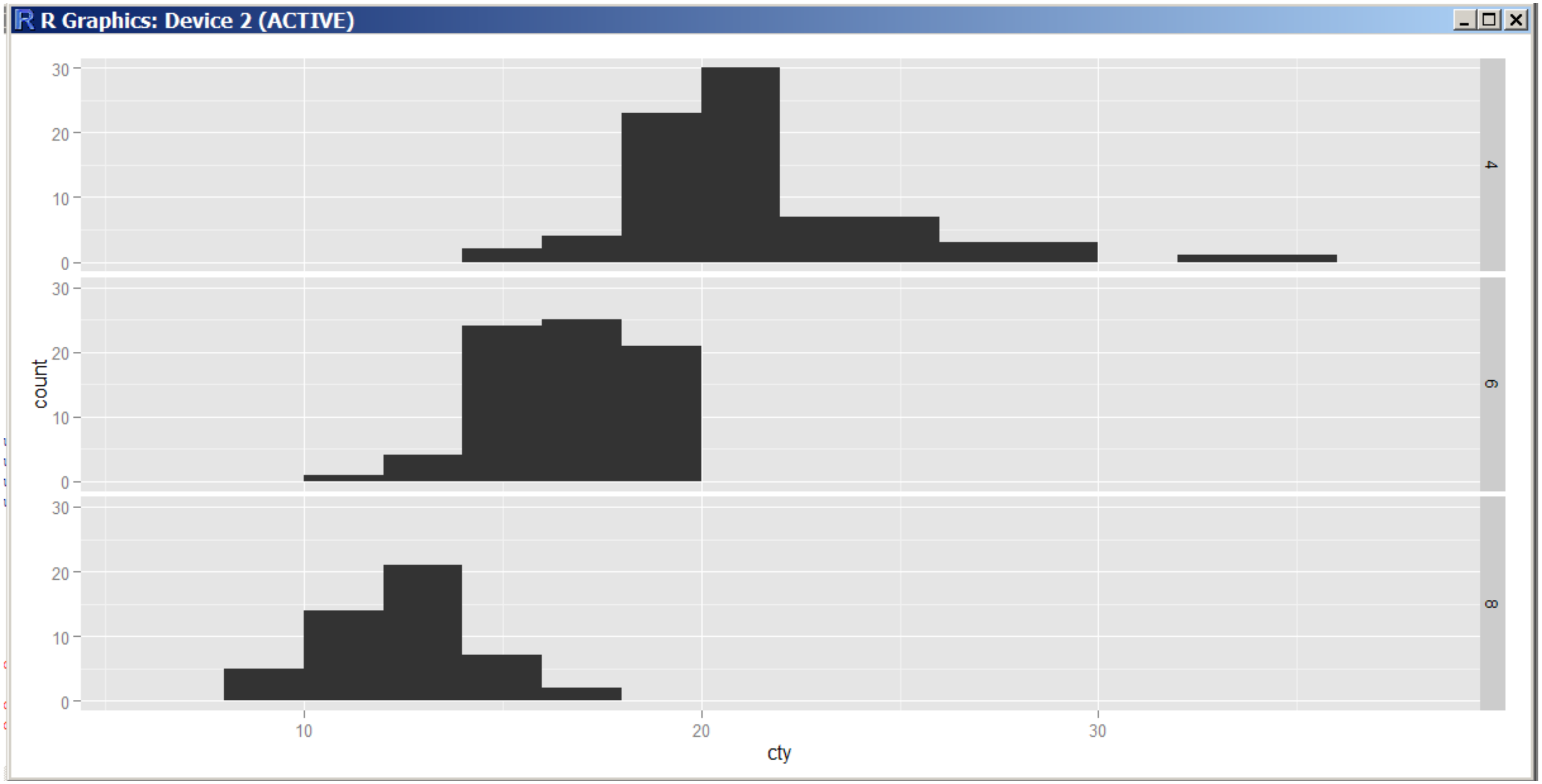
```
mpg2 <- subset(mpg, cyl != 5 & drv %in% c("4", "f"))
qplot(cty, hwy, data = mpg2) + facet_grid(. ~ cyl)
```



2013.2.8

Facet grid

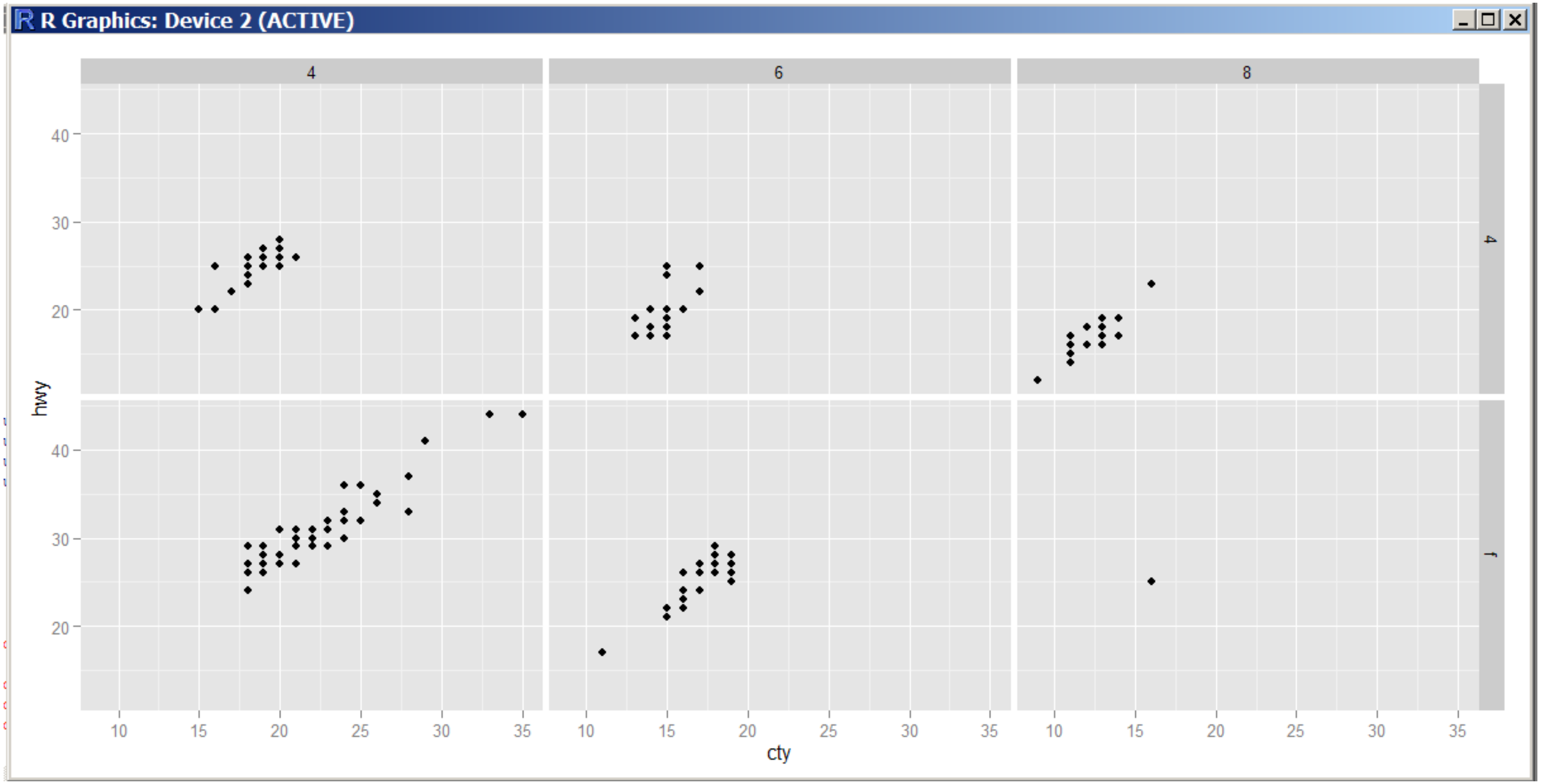
```
qplot(cty, data = mpg2, geom="histogram", binwidth = 2) + facet_grid(cyl ~ .)
```



2013.2.8

Facet grid

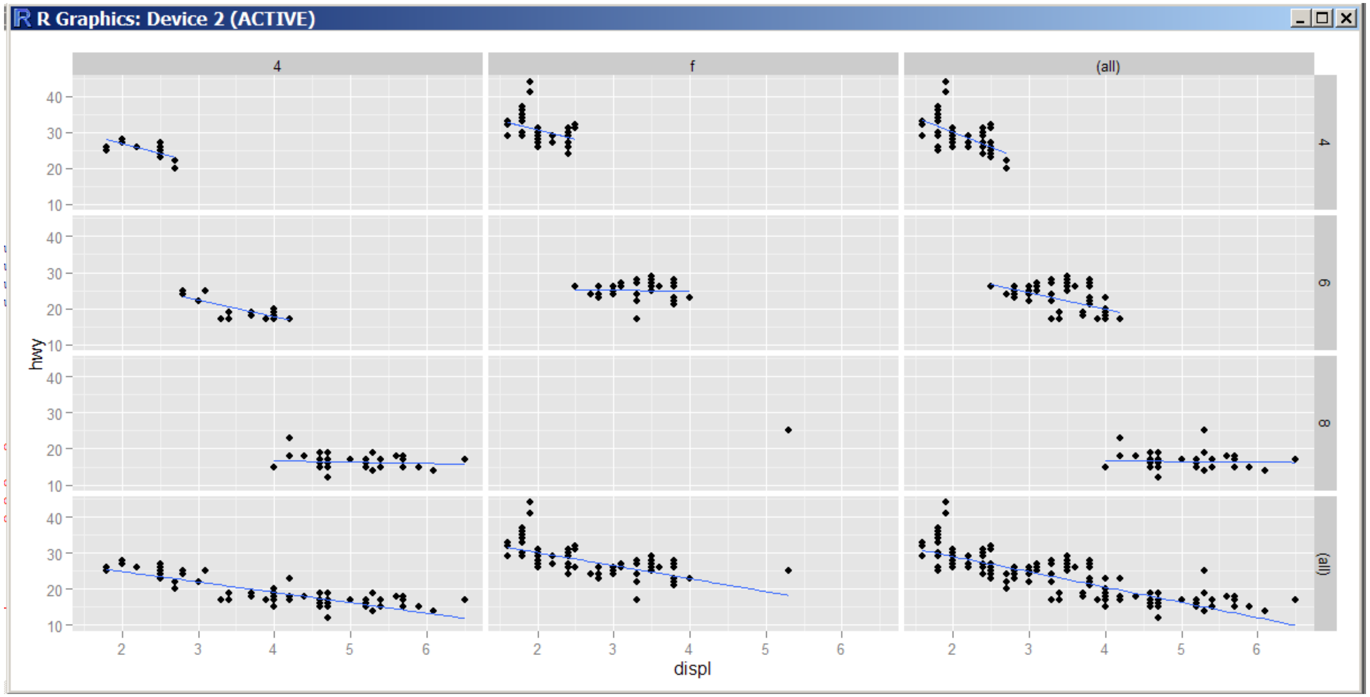
qplot(cty, hwy, data = mpg2) + facet_grid(drv ~ cyl)



2013.2.8

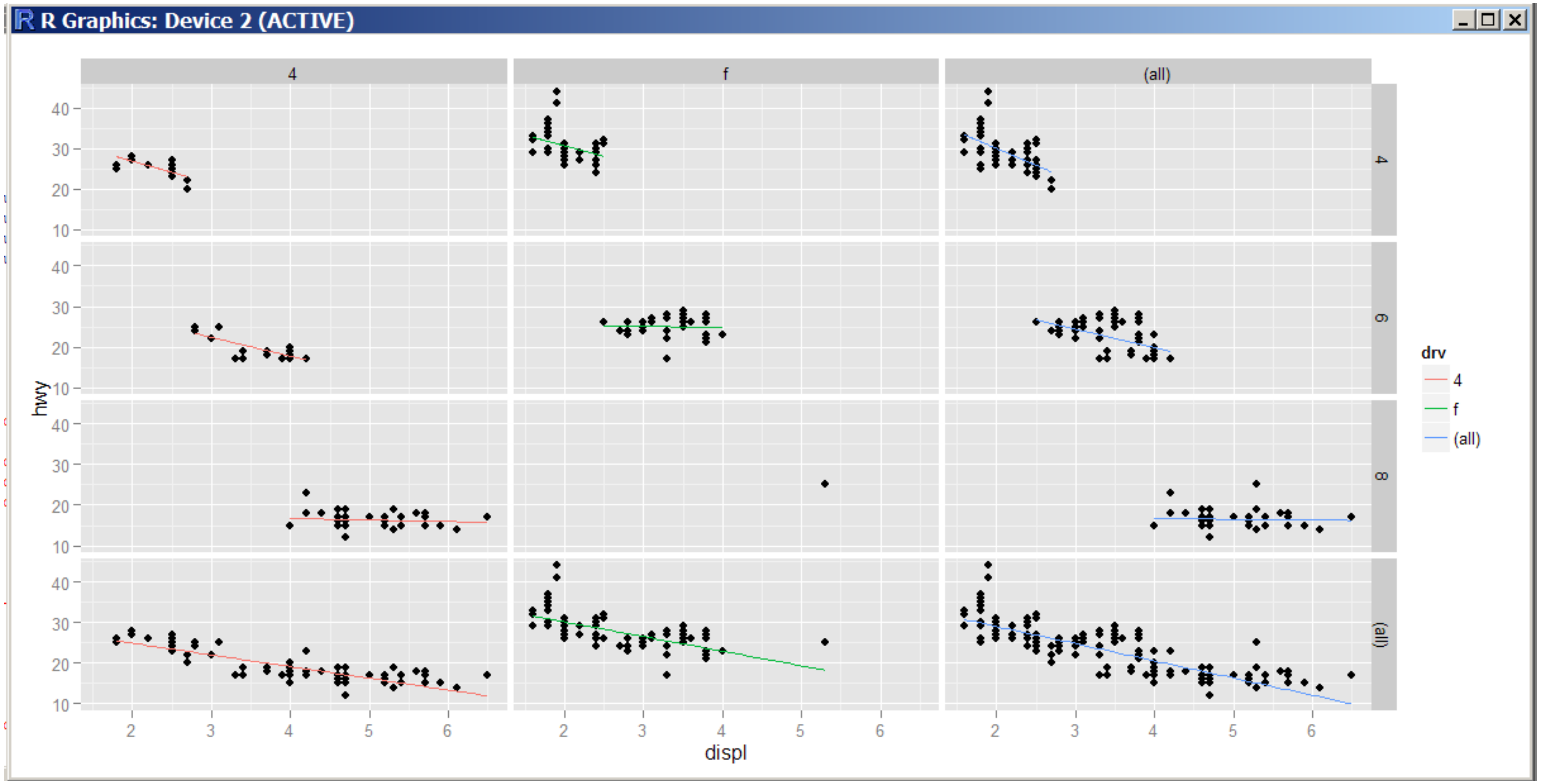
边缘控制

```
p <- qplot(displ, hwy, data = mpg2) +geom_smooth(method = "lm", se = F)
p + facet_grid(cyl ~ drv)
p + facet_grid(cyl ~ drv, margins = T)
```



边缘控制

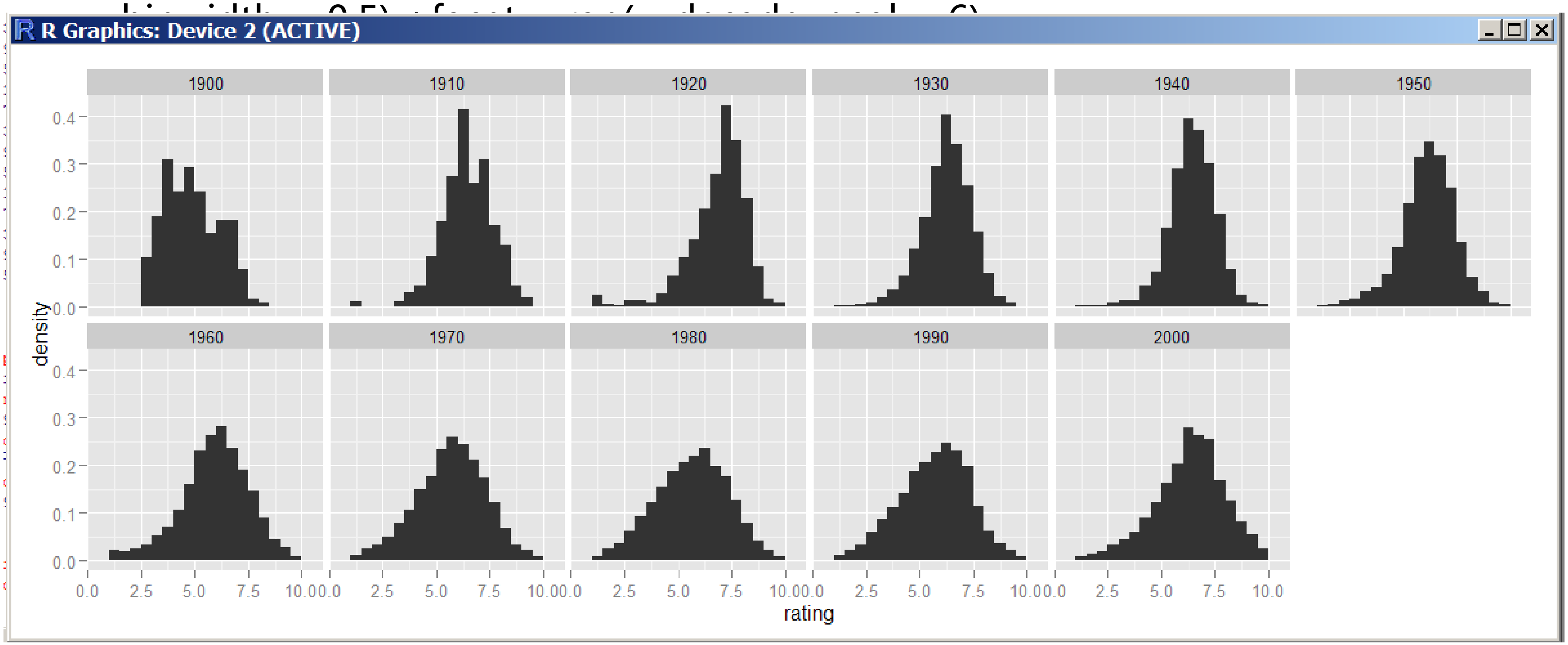
```
qplot(displ, hwy, data = mpg2) + geom_smooth(aes(colour = drv), method = "lm",
    se = F) + facet_grid(cyl ~ drv, margins = T)
```



Facet wrap

```
movies$decade <- floor(movies$year/10)*10

qplot(rating, ..density.., data=subset(movies, decade > 1890),geom="histogram",
    binwidth = 0.5) + facet_wrap(~ decade, ncol = 6)
```



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