R Graphics Cookbook

作者：winston 译者：吴红龙

2013 年 5 月 9 日

目录

第一章　前言 **3**

第二章　R 基础 **4**

2.1　R 包安装 5

2.2　R 包加载 5

2.3　固定分割的文本数据加载 5

2.4　Excel 数据加载 5

2.5　SPSS 数据加载 5

第三章　条形图 **6**

3.1　创建散点图 6

3.2　创建线图 9

3.3　创建条形图 10

3.4　创建柱状图 16

3.5　创建盒型图 16

3.6　绘制函数曲线 16

第四章　线形图 **17**

4.1　绘制基本线形图 17

第五章　散点图 **22**

第六章　描述性数据分布 **29**

第七章　注释 **37**

第八章　坐标轴 **42**

第九章　图形全局参数控制 48

第十章　图例 51

第十一章　图层 53

第十二章　颜色的使用 56

第十三章　图形实例 59

第十四章　图形输出格式 62

第十五章　数据处理 63

第十六章　A. ggplot2 介绍 64

第十七章　索引 65

第一章　前言

第二章　R 基础

本章包含的基础知识：安装和使用包以及加载数据。

书中所有的实例需要在你的电脑上安装 ggplot2 和 gcookbook 两个包，如果你想快速开始的话，可以这样做：

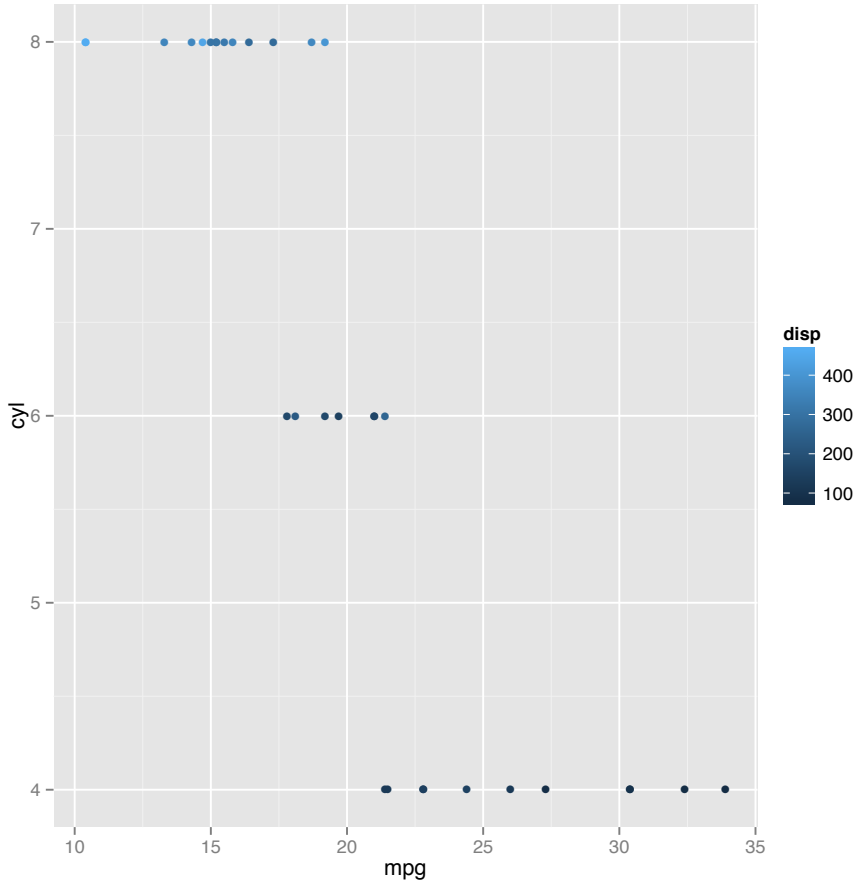
install.packages(c("ggplot2","gcookbook"))

然后在运行本书中例子的过程中，你要加载他们使用如下命令：

library(ggplot2)

library(gcookbook)

```
library(ggplot2)

## Loading required package:  methods

p <- ggplot(data = mtcars, aes(x = mpg, y = cyl, color = disp))
p + geom_point()
```



2.1　R 包安装

2.2　R 包加载

2.3　固定分割的文本数据加载

2.4　Excel 数据加载

2.5　SPSS 数据加载

第三章　条形图

3.1　创建散点图

问题：创建一个散点图

解决方案：用 plot() 绘制散点图，传递 x 值向量与 y 值向量。

```
plot(mtcars$wt, mtcars$mpg)
```

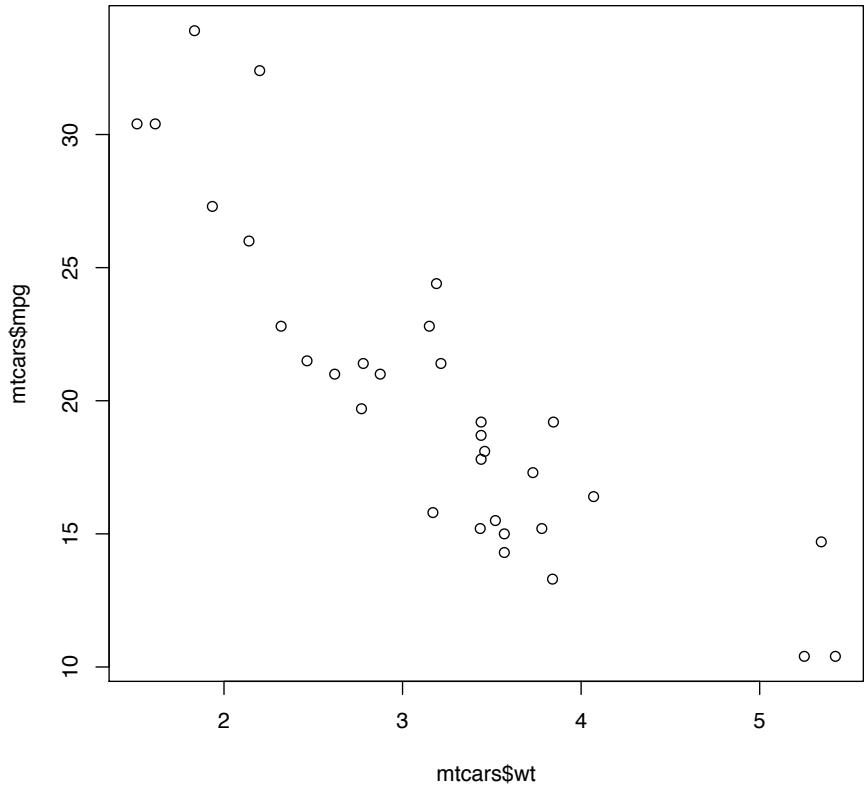
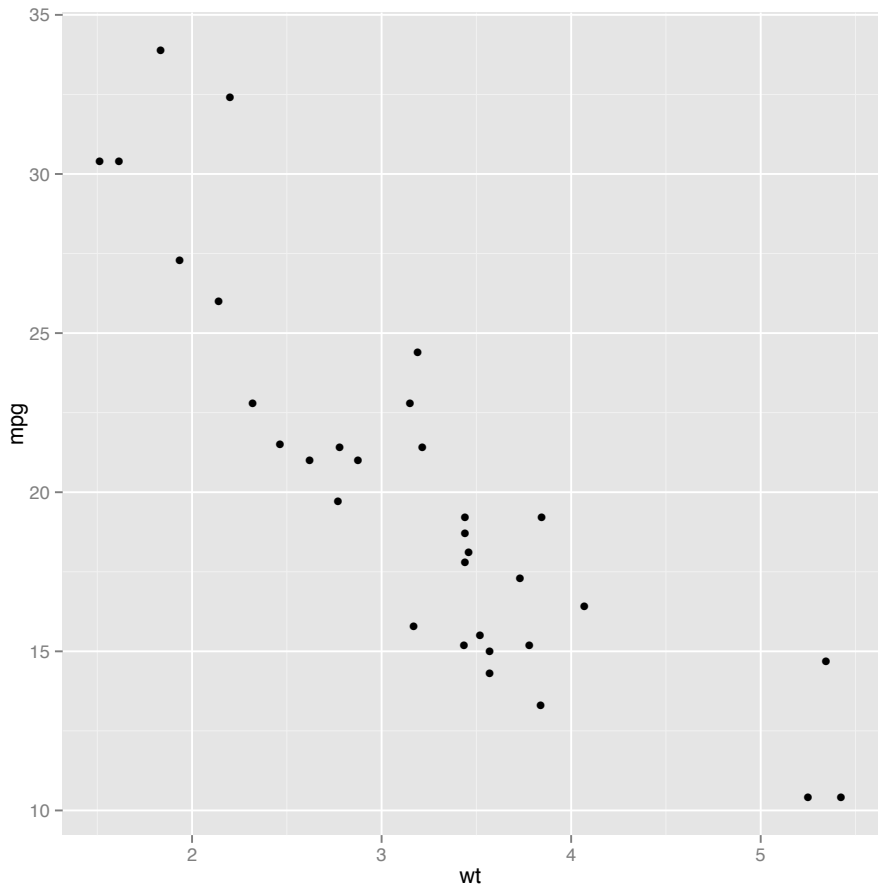
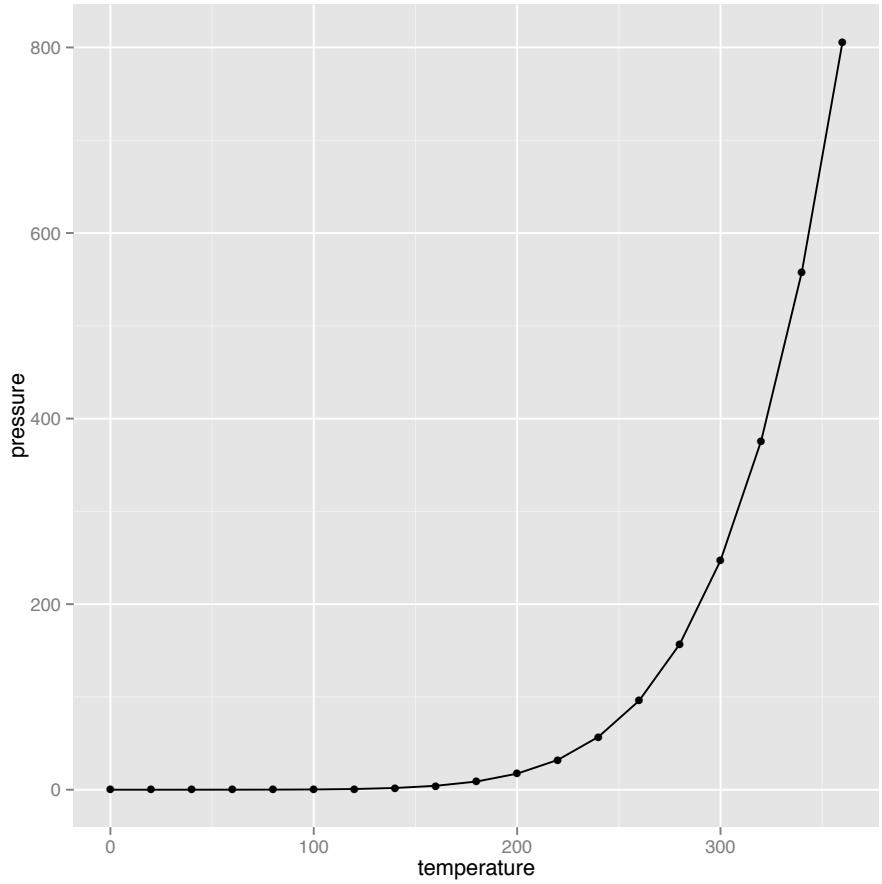


图 2-1. 基本图形函数绘制散点图

```
library(ggplot2)
ggplot(data = mtcars, aes(x = wt, y = mpg)) + geom_point()
```

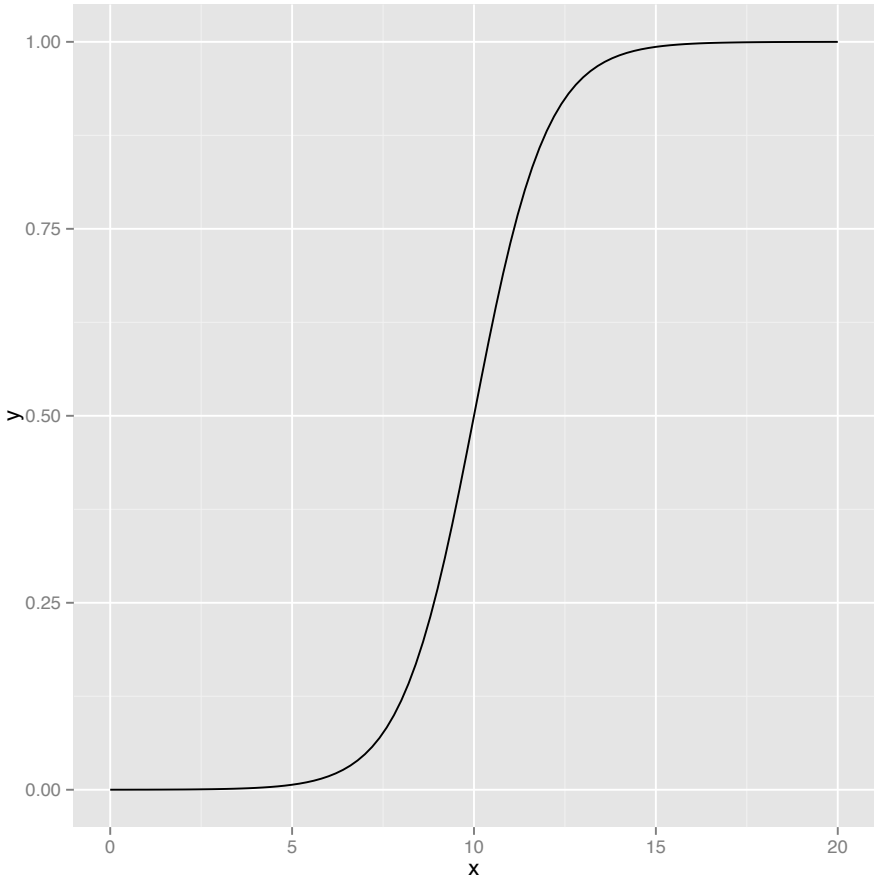



```
ggplot(data = pressure, aes(x = temperature, y = pressure)) + geom_line() +
    geom_point()
```



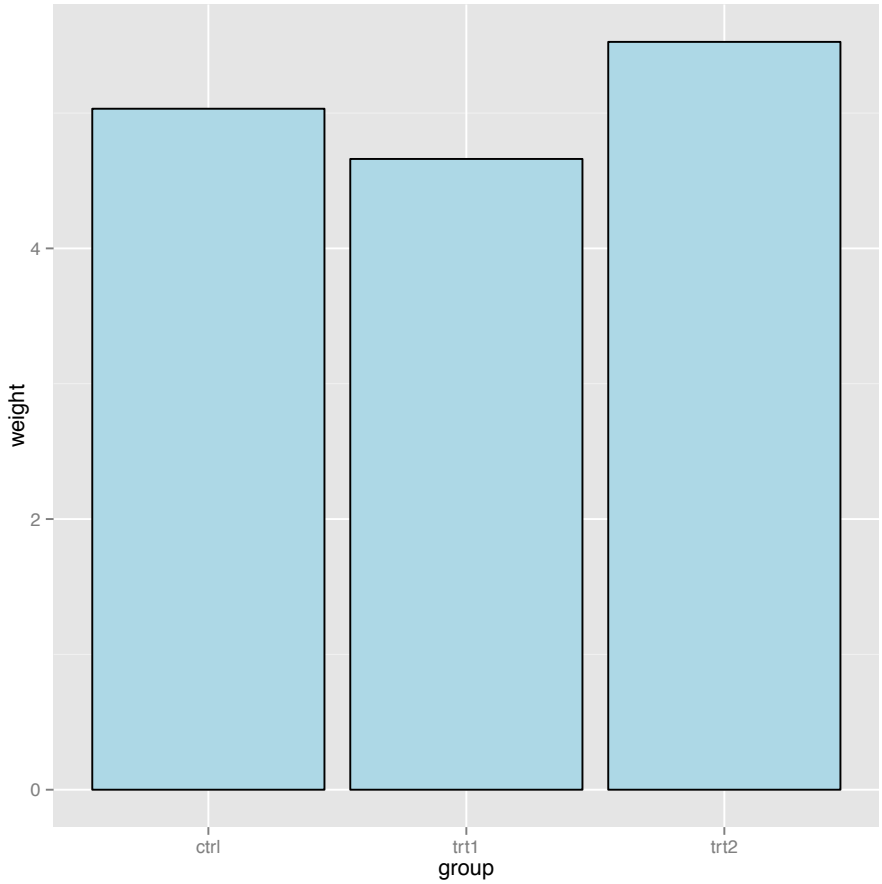
3.2　创建线图

```
myfun <- function(xvar) {
    1/(1 + exp(-xvar + 10))
}
ggplot(data = data.frame(x = c(0, 20)), aes(x = x)) + stat_function(fun = myfun,
    geom = "line")
```

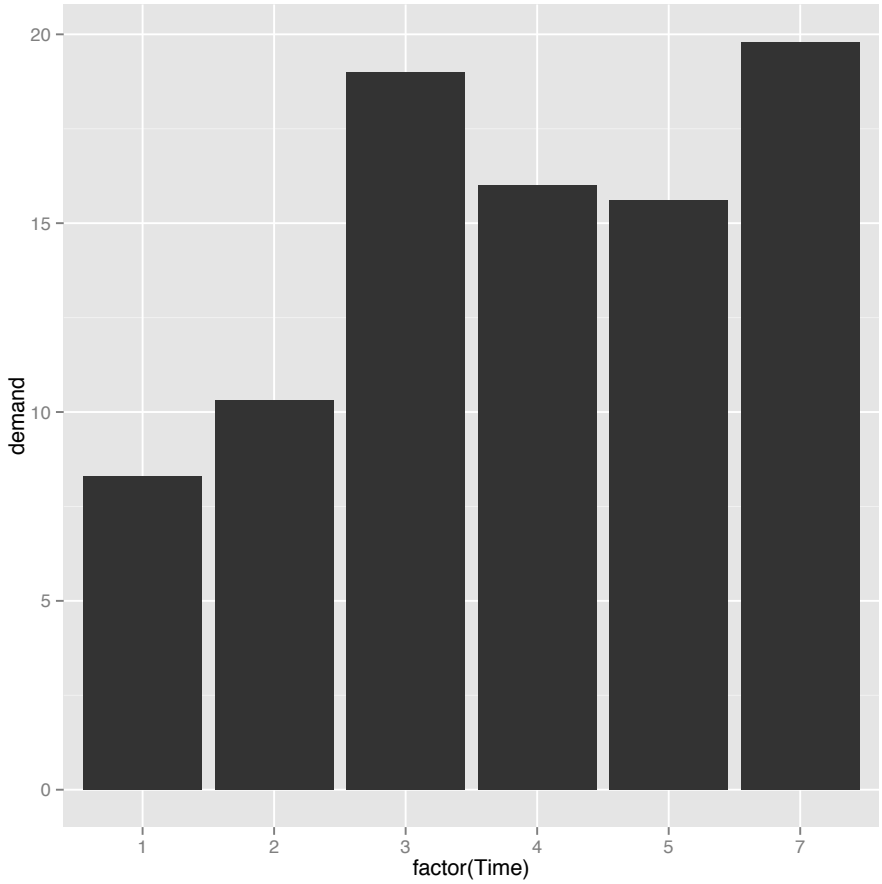


3.3　创建条形图

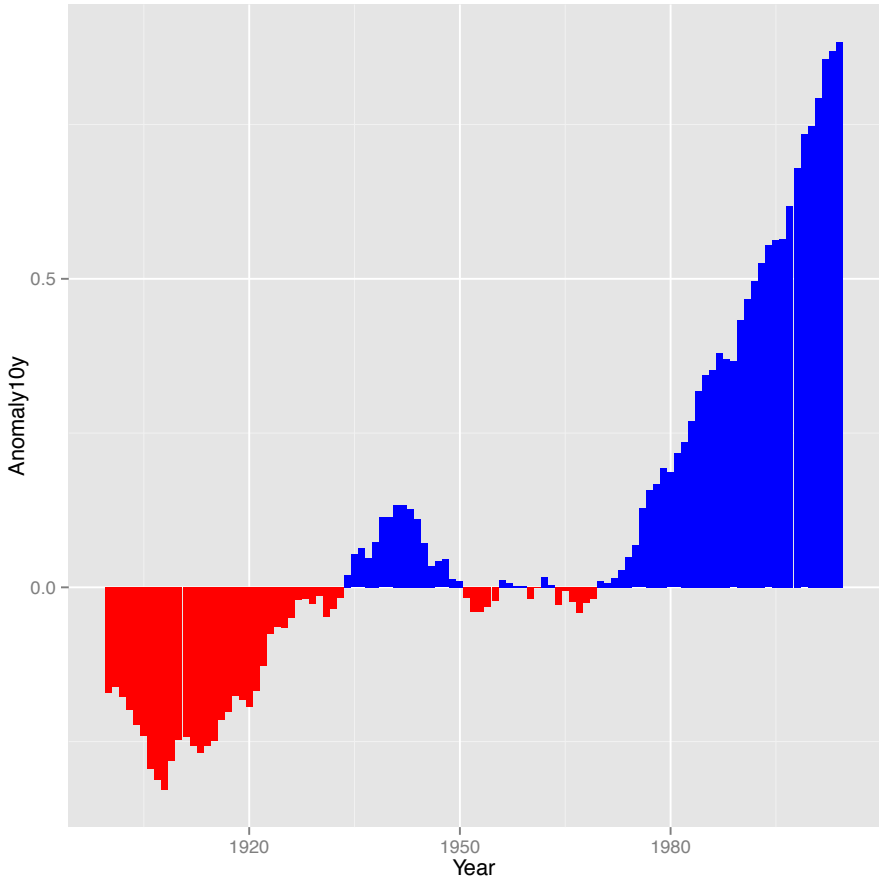
```
library(ggplot2)
library(gcookbook)
ggplot(data = pg_mean, aes(x = group, y = weight)) + geom_bar(stat = "identity",
    fill = "lightblue", color = "black")
```




```
library(ggplot2)
library(gcookbook)
ggplot(data = BOD, aes(x = factor(Time), y = demand)) + geom_bar(stat = "identity")
```

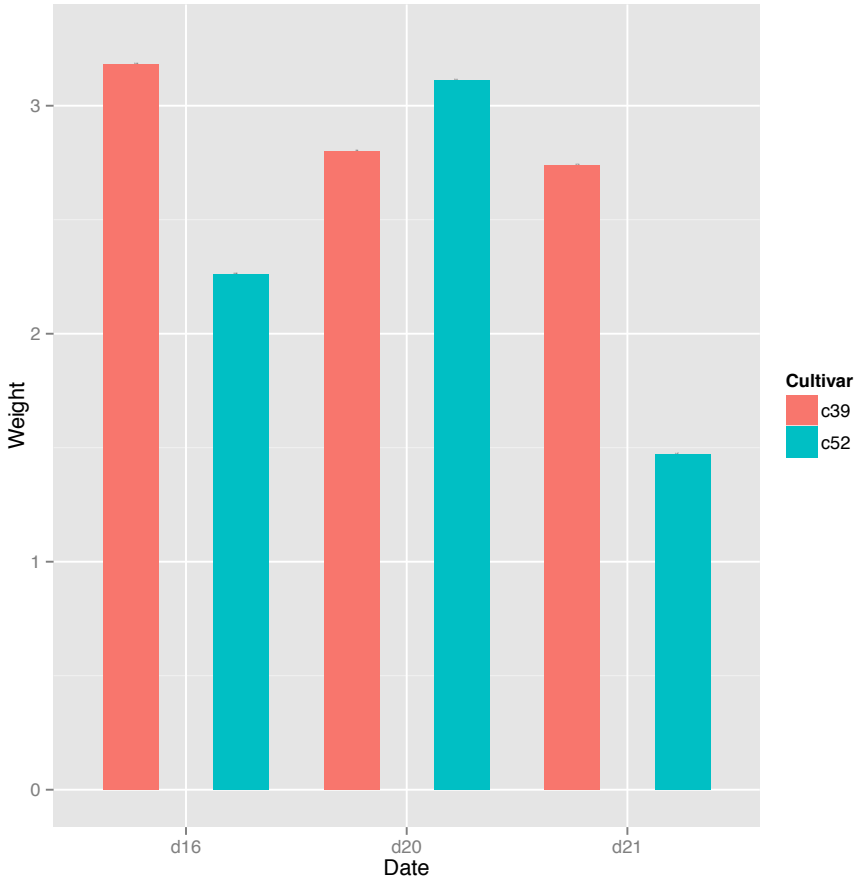



```
library(gcookbook)
csub <- subset(climate, Source == "Berkeley" & Year >= 1900)
csub$pos <- csub$Anomaly10y >= 0
ggplot(csub, aes(x = Year, y = Anomaly10y, fill = pos)) + geom_bar(stat = "identity") +
    scale_fill_manual(values = c("red", "blue"), guide = F)

## Warning:  Stacking not well defined when ymin != 0
```



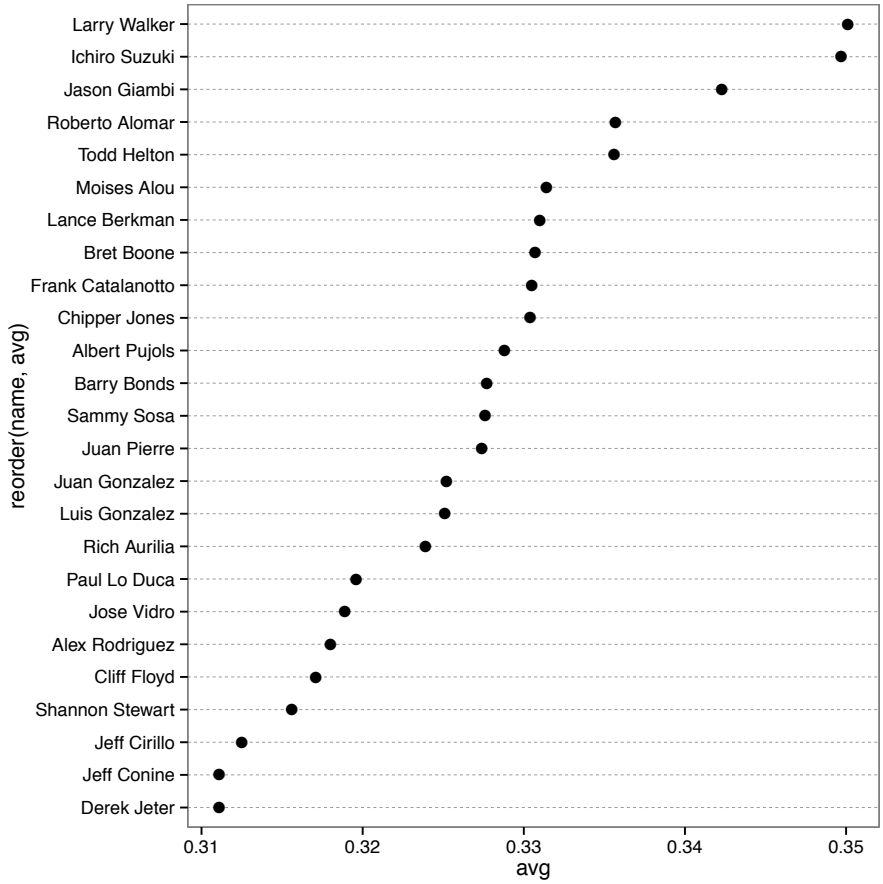
在柱状图上面加标记

```
library(gcookbook)
library(ggplot2)
ggplot(data = cabbage_exp, aes(x = Date, y = Weight, fill = Cultivar)) + geom_bar(stat
    position = position_dodge(1), width = 0.5) + geom_text(aes(label = Weight),
    vjust = -0.2, position = position_dodge(0.9), size = 0.3) + ylim(0, max(cabbage_exp
    0.1)

## ymax not defined:  adjusting position using y instead
```



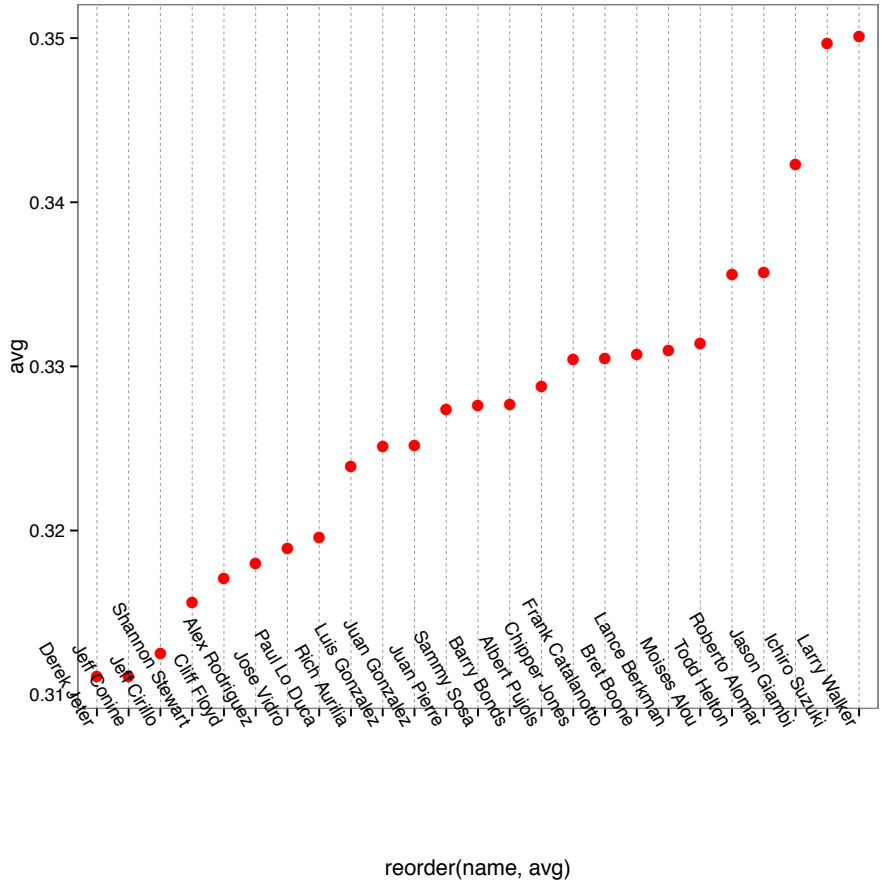
创建点图

```
library(gcookbook)
library(ggplot2)
tophit <- tophitters2001[1:25, ]
ggplot(tophit, aes(x = avg, y = reorder(name, avg))) + geom_point(size = 3) +
    theme_bw() + theme(panel.grid.major.x = element_blank(), panel.grid.minor.x = eleme
    panel.grid.major.y = element_line(colour = "grey60", linetype = "dashed"))
```



反转方向

```
library(gcookbook)
library(ggplot2)
tophit <- tophitters2001[1:25, ]
ggplot(tophit, aes(x = reorder(name, avg), y = avg)) + geom_point(size = 3,
    colour = "red") + theme_bw() + theme(axis.text.x = element_text(angle = -60,
    hjust = 1), panel.grid.major.y = element_blank(), panel.grid.minor.y = element_blan
    panel.grid.major.x = element_line(colour = "grey60", linetype = "dashed"))
```



3.4　创建柱状图

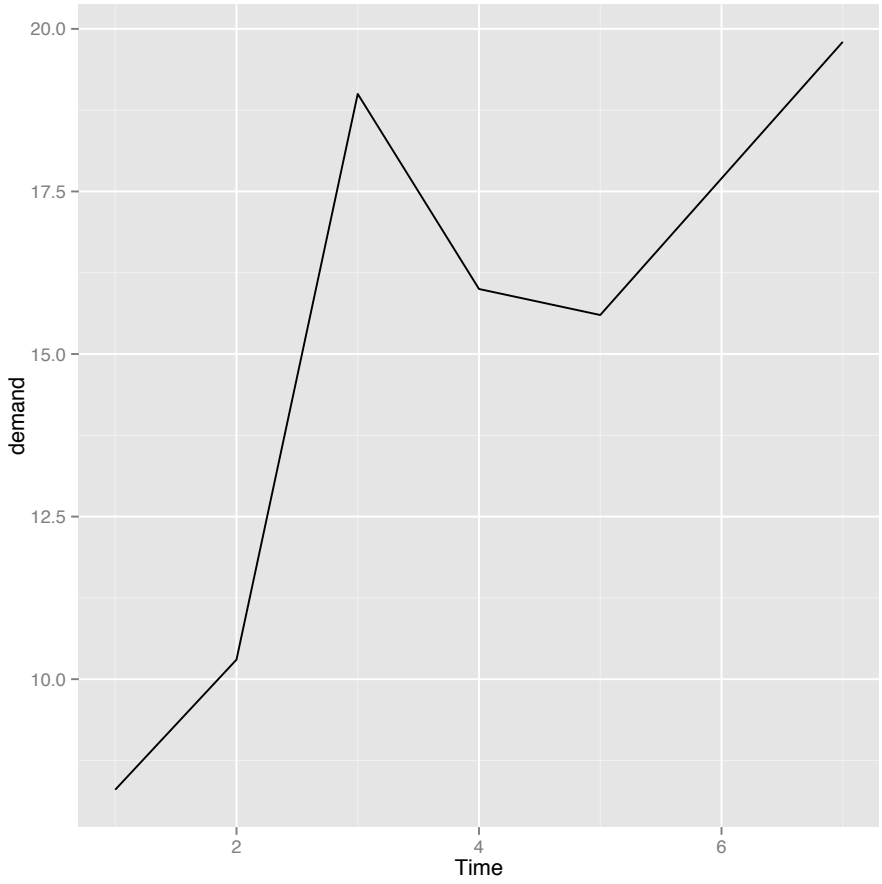
3.5　创建盒型图

3.6　绘制函数曲线

第四章　线形图

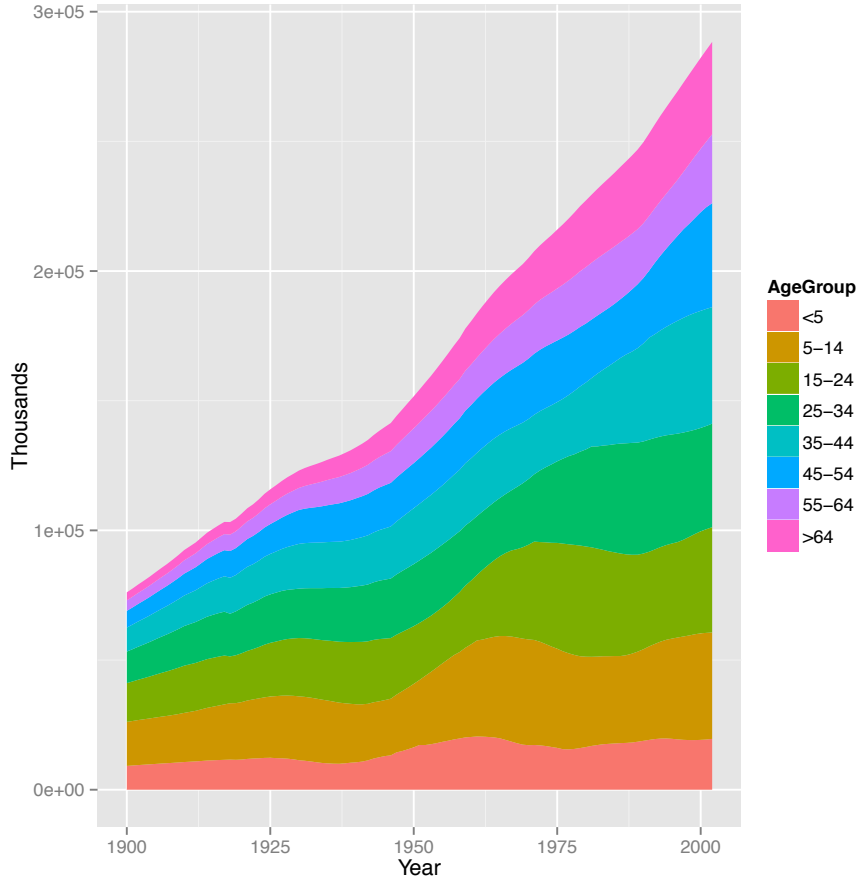
4.1　绘制基本线形图

```
library(ggplot2)
library(gcookbook)
ggplot(data = BOD, aes(x = Time, y = demand)) + geom_line()
```



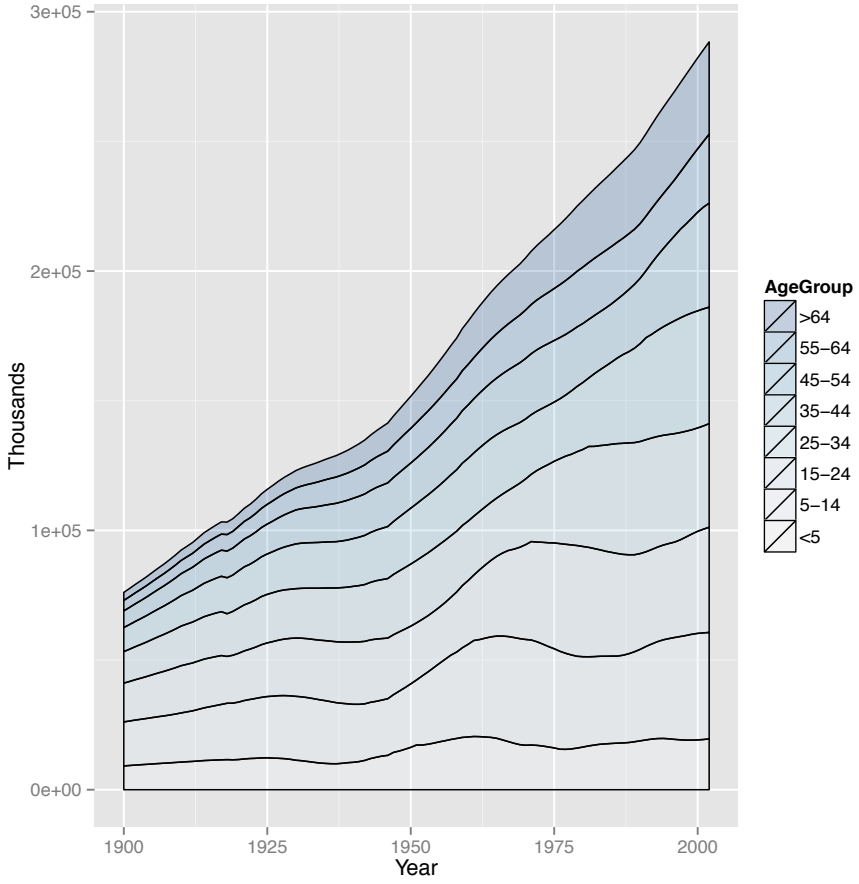
堆积图:

```
library(ggplot2)
library(gcookbook)
ggplot(data = uspopage, aes(x = Year, y = Thousands, fill = AgeGroup)) + geom_area()
```



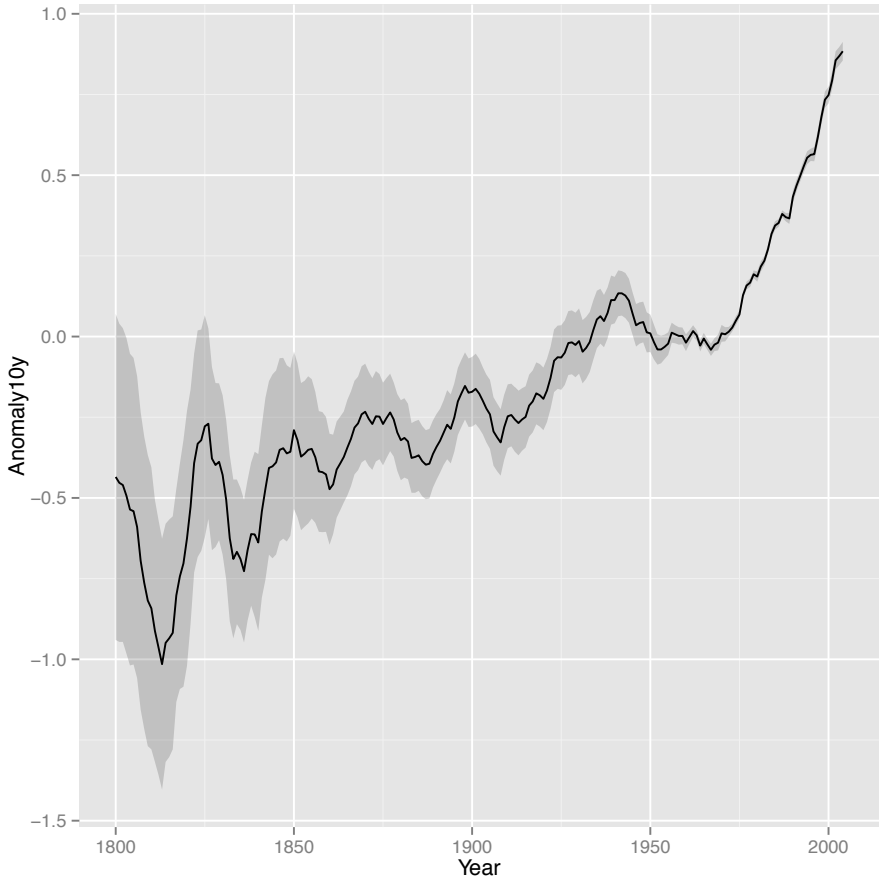
修改颜色配置:

```
library(ggplot2)
library(gcookbook)
ggplot(data = uspopage, aes(x = Year, y = Thousands, fill = AgeGroup)) + geom_area(colo
    size = 0.4, alpha = 0.2) + scale_fill_brewer(palette = "Blues", breaks = rev(levels
```



添加置信区间:

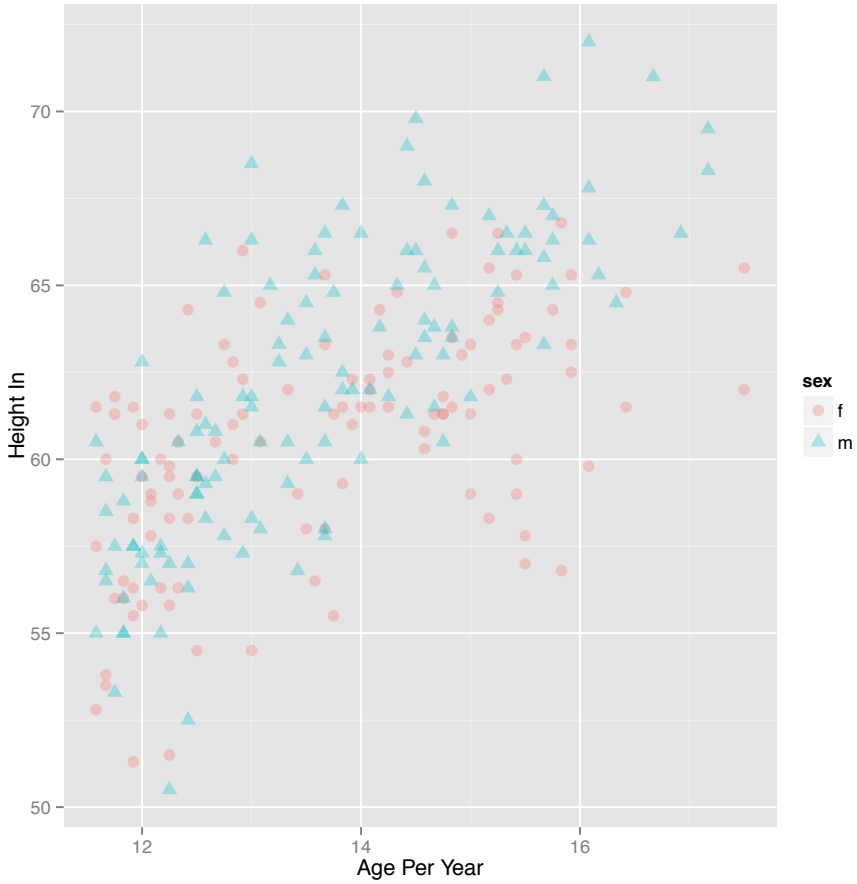
```
library(ggplot2)
library(gcookbook)
clim <- subset(climate, Source == "Berkeley", select = c("Year", "Anomaly10y",
    "Unc10y"))
ggplot(clim, aes(x = Year, y = Anomaly10y)) + geom_ribbon(aes(ymin = Anomaly10y -
    Unc10y, ymax = Anomaly10y + Unc10y), alpha = 0.2) + geom_line()
```


1.0
0.5
0.0
−0.5
−1.0
−1.5
Anomaly10y
1800
1850
1900
1950
2000
Year

第五章　散点图

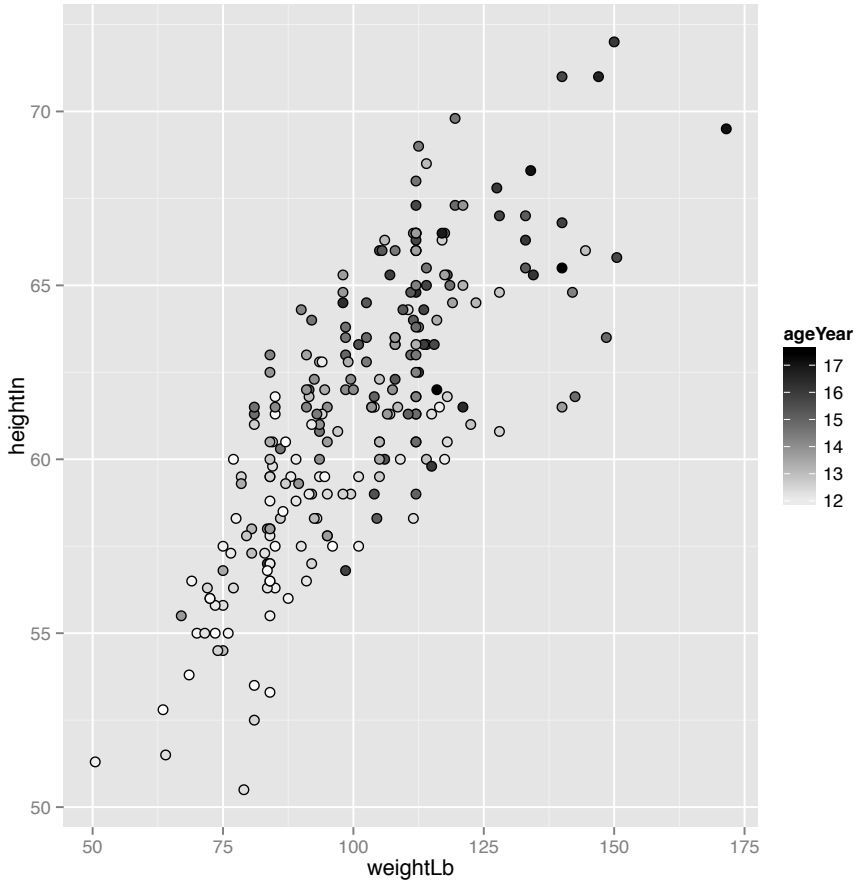
散点图:

```
library(ggplot2)
library(gcookbook)
ggplot(heightweight, aes(x = ageYear, y = heightIn, colour = sex, shape = sex)) +
    geom_point(size = 3, alpha = 0.3) + xlab("Age Per Year") + ylab("Height In")
```



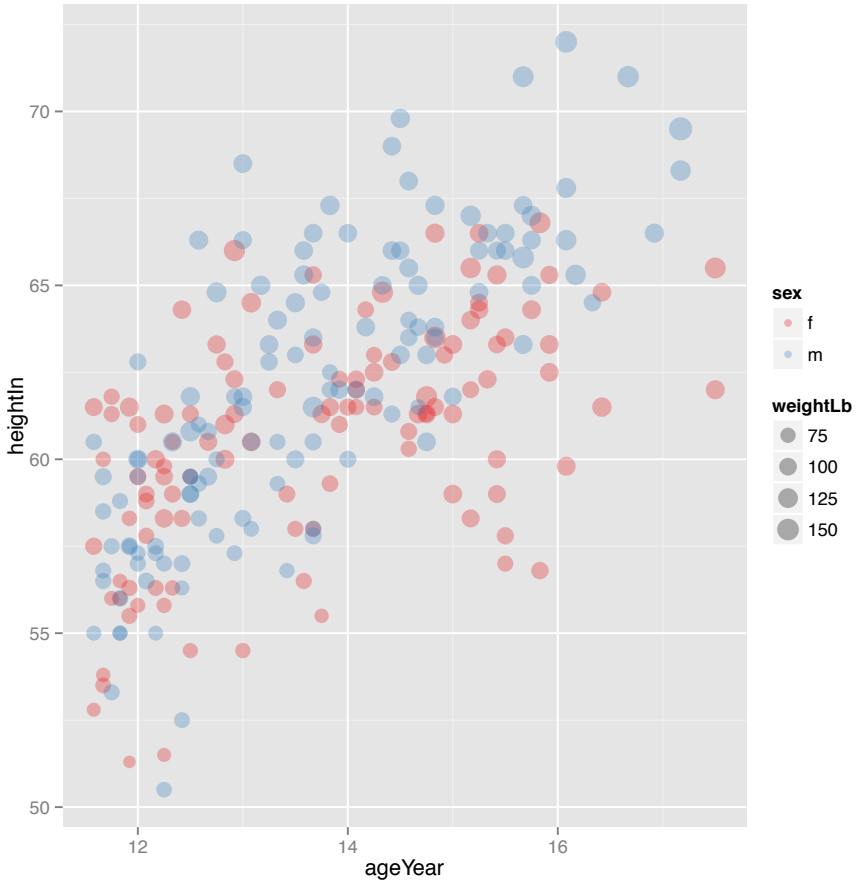
添加梯度:

```
library(ggplot2)
library(gcookbook)
ggplot(heightweight, aes(x = weightLb, y = heightIn, fill = ageYear)) + geom_point(shap
    size = 2.5) + scale_fill_gradient(low = "white", high = "black", breaks = 12:17)
```



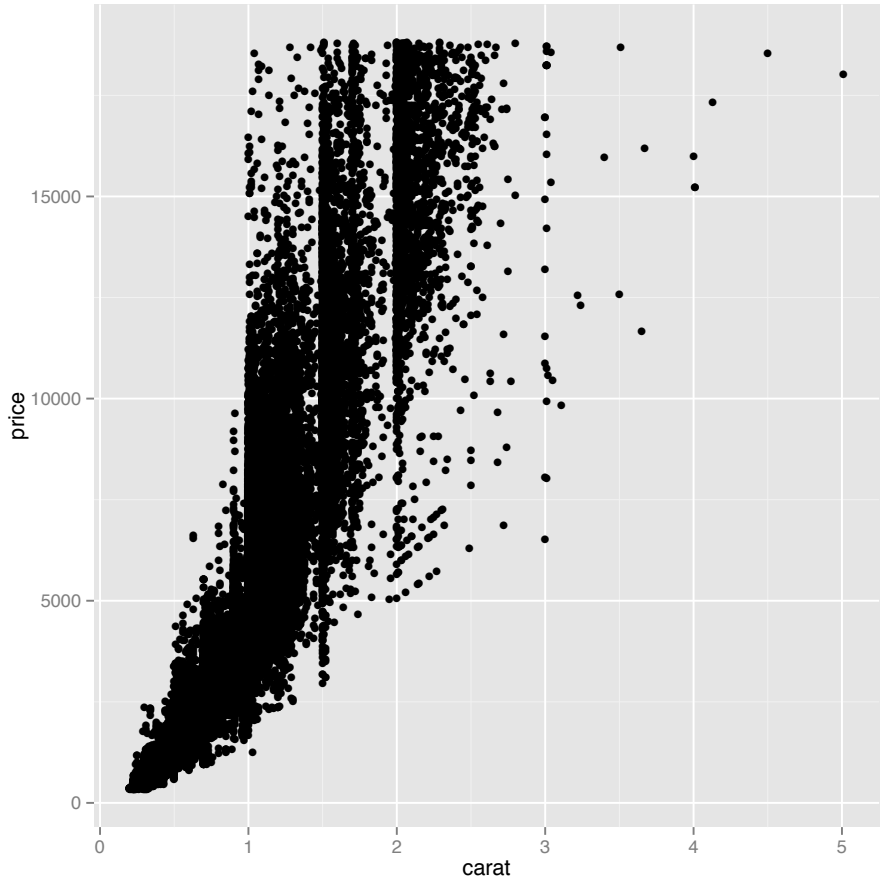
使用 area：

```
library(ggplot2)
library(gcookbook)
ggplot(heightweight, aes(x = ageYear, y = heightIn, size = weightLb, colour = sex)) +
    geom_point(alpha = 0.3) + scale_size_area() + scale_color_brewer(palette = "Set1")
```

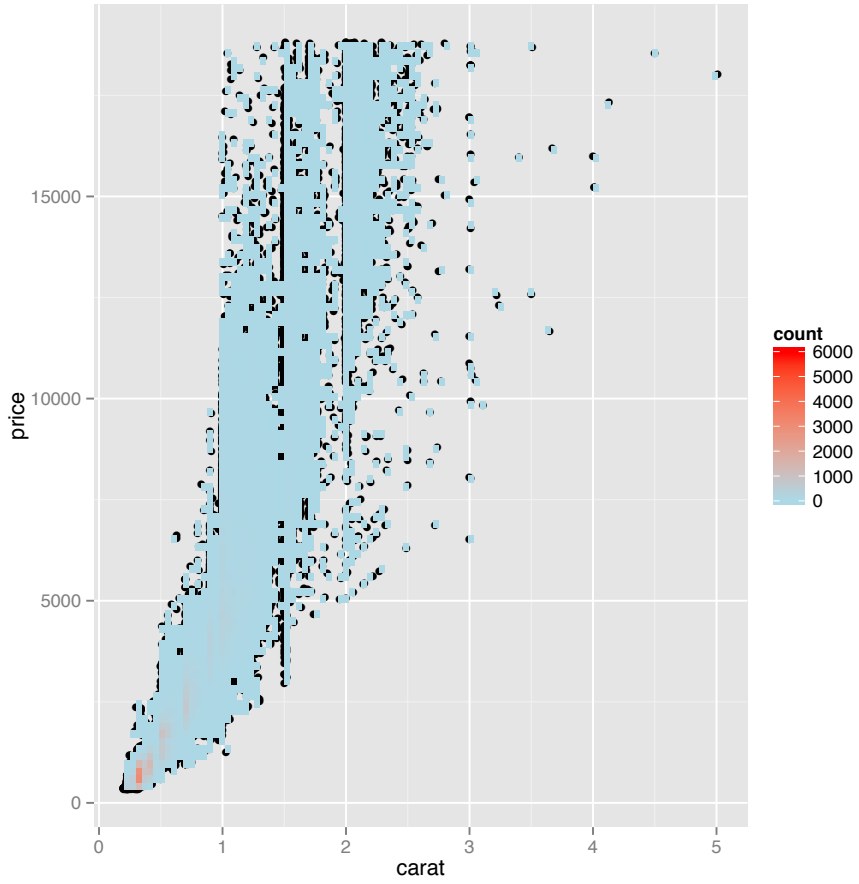


使用 stat_bin2d 避免过多点的 overlap：

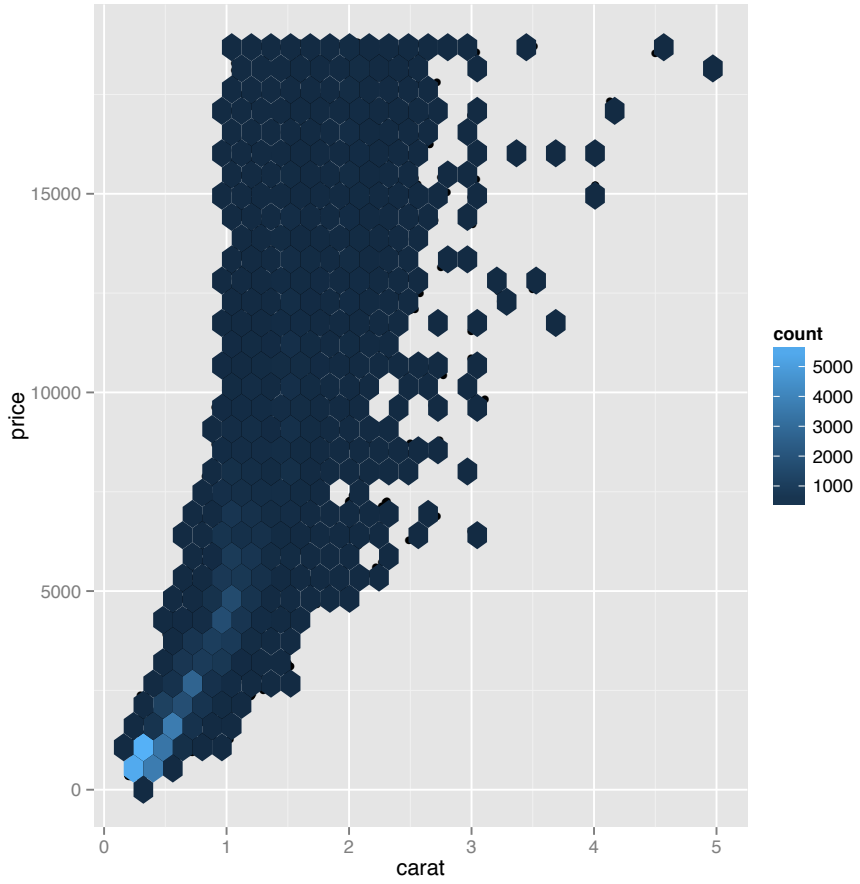
```
library(ggplot2)
library(gcookbook)
a <- ggplot(diamonds, aes(x = carat, y = price)) + geom_point()
a
```




```
b <- a + stat_bin2d(bins = 100) + scale_fill_gradient(low = "lightblue", high = "red",
    limits = c(0, 6000))
b
```

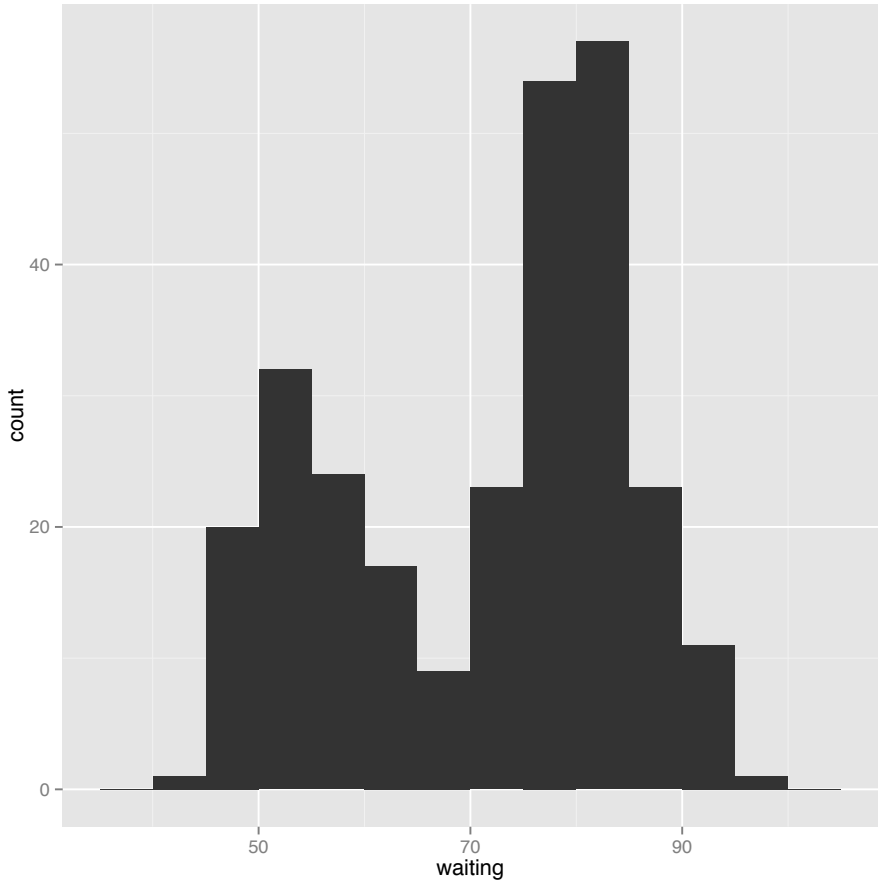



```
c <- a + stat_binhex()
c
```


price
15000
10000
5000
0
0
1
2
3
4
5
carat
count
5000
4000
3000
2000
1000

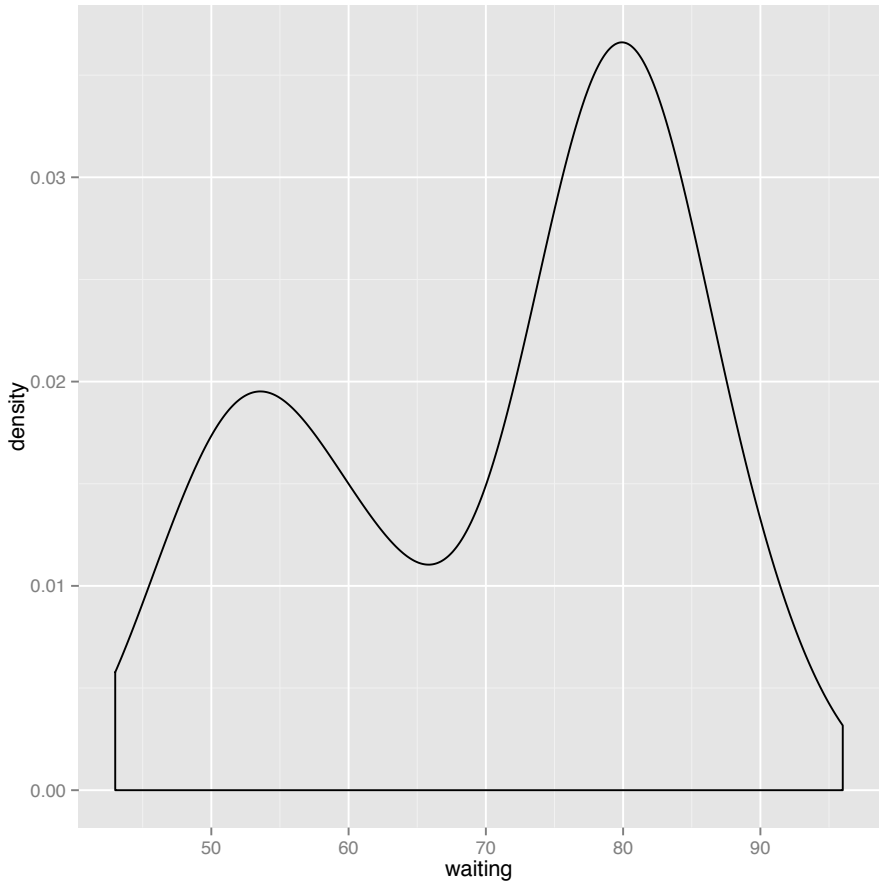
第六章　描述性数据分布

```
library(ggplot2)
library(gcookbook)
ggplot(faithful, aes(x = waiting)) + geom_histogram(binwidth = 5)
```

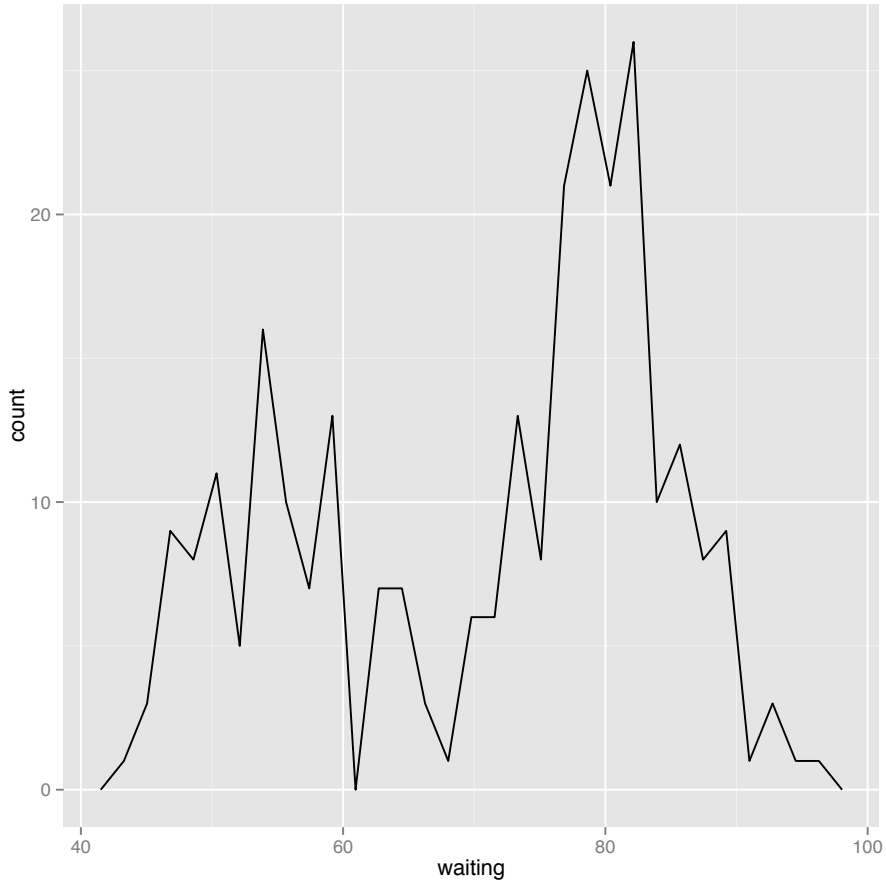


density line 例子:

```
library(ggplot2)
library(gcookbook)
ggplot(faithful, aes(x = waiting)) + geom_density()
```

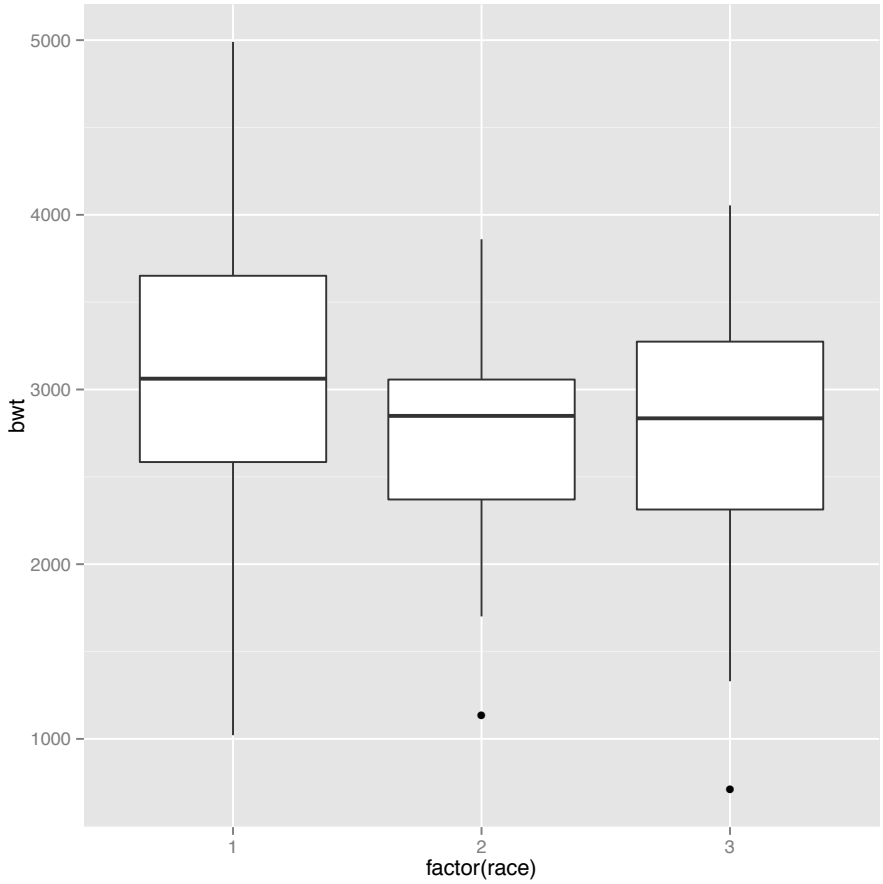



```
b <- ggplot(faithful, aes(x = waiting)) + geom_freqpoly()
b

## stat_bin:  binwidth defaulted to range/30.  Use 'binwidth = x'
to adjust this.
```

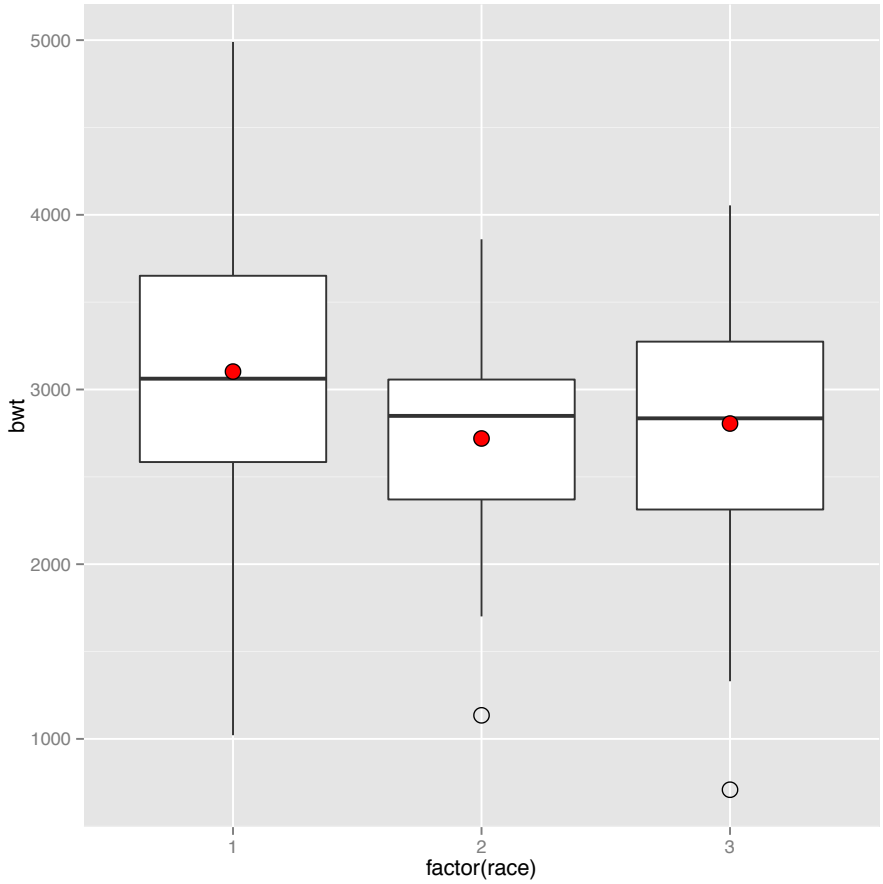


boxplot 例子：

```
library(MASS)
library(ggplot2)
library(gcookbook)
ggplot(birthwt, aes(x = factor(race), y = bwt)) + geom_boxplot()
```

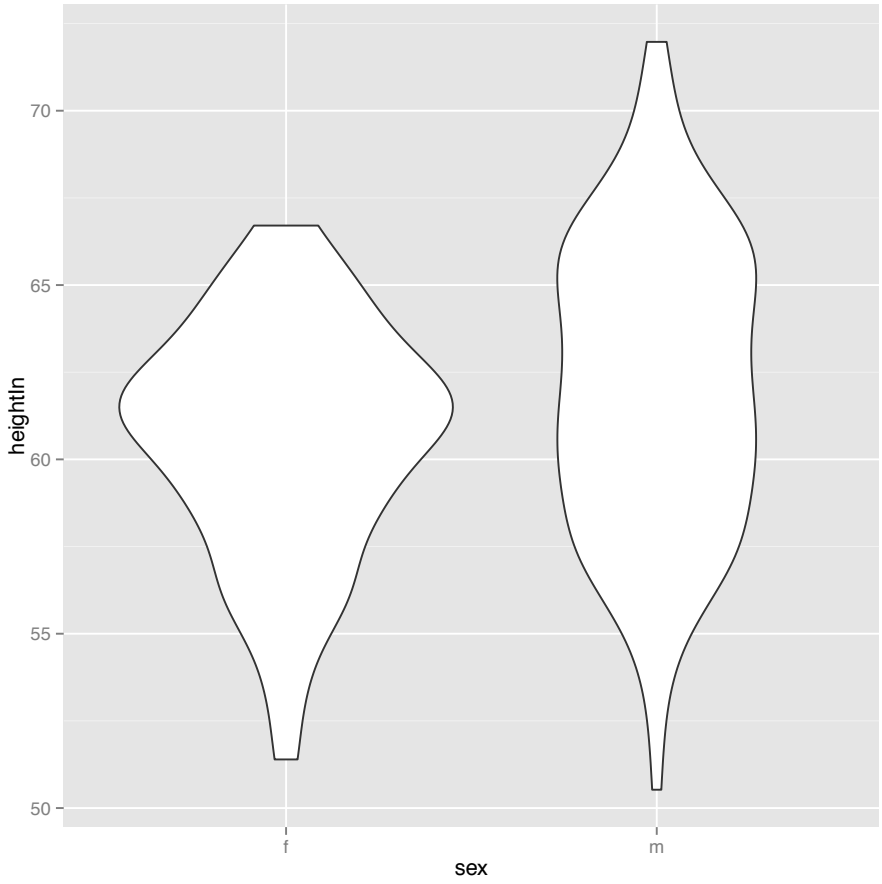



```
a <- ggplot(birthwt, aes(x = factor(race), y = bwt)) + geom_boxplot(outlier.shape = 21,
    outlier.size = 4) + stat_summary(fun.y = "mean", geom = "point", shape = 21,
    size = 4, fill = "red")
a
```



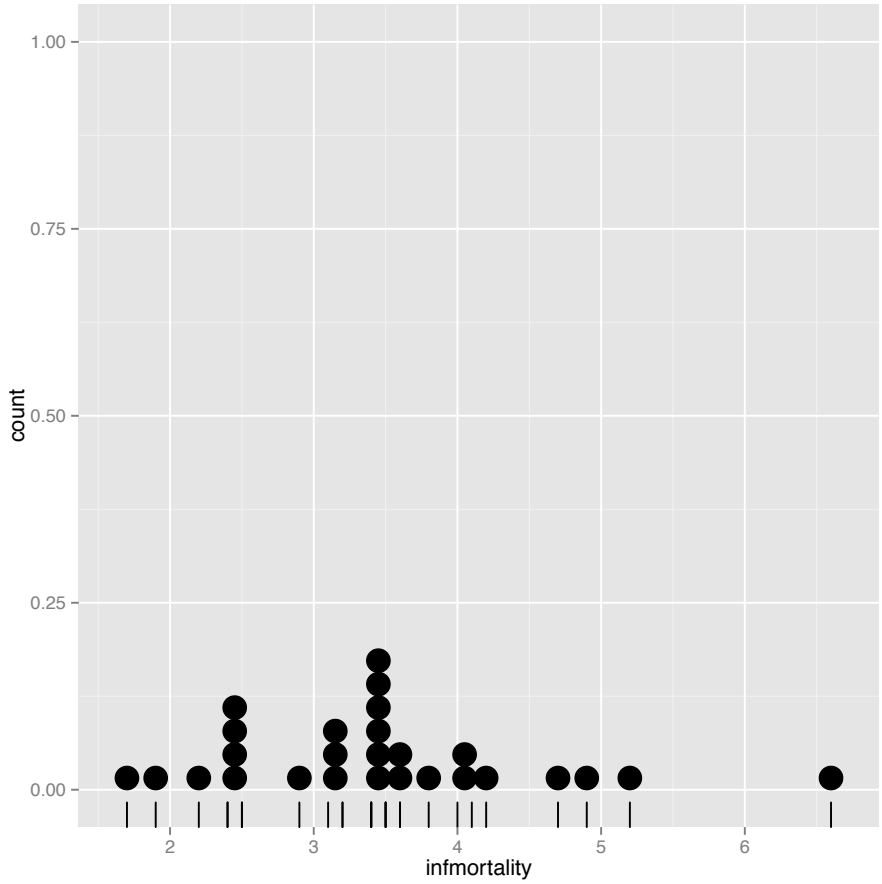
violin 例子：

```
library(ggplot2)
library(gcookbook)
ggplot(heightweight, aes(x = sex, y = heightIn)) + geom_violin()
```



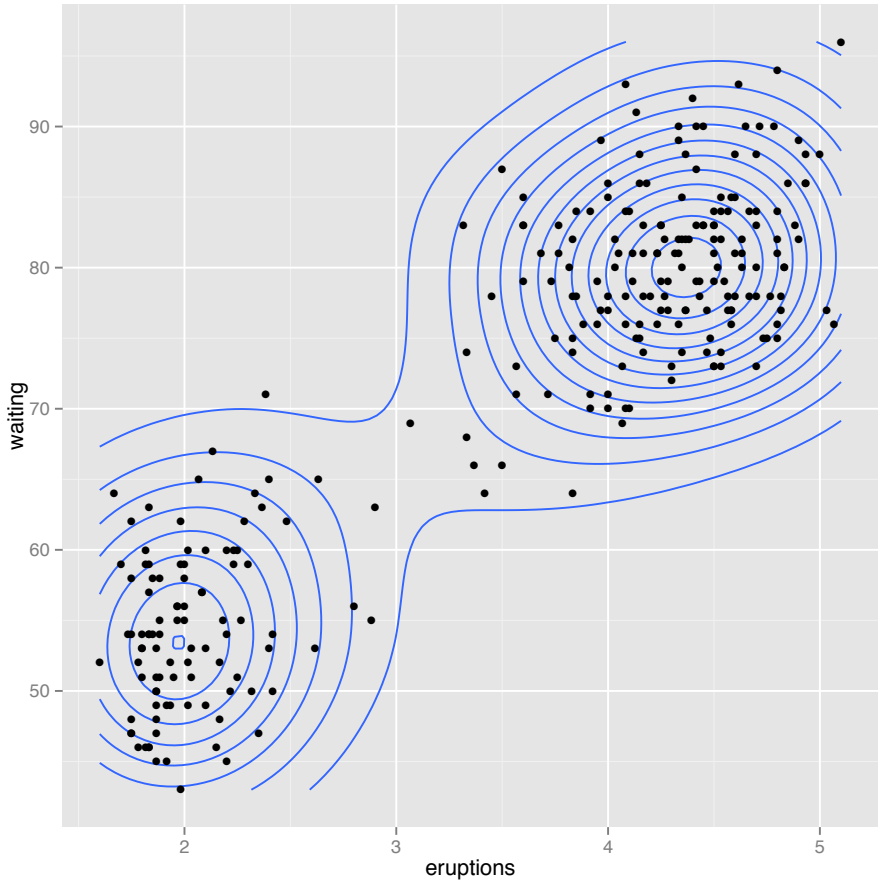
dot 例子：

```
library(ggplot2)
library(gcookbook)
countries2009 <- subset(countries, Year == 2009 & healthexp > 2000)
p <- ggplot(countries2009, aes(x = infmortality))
p + geom_dotplot() + geom_rug()

## stat_bindot:  binwidth defaulted to range/30.  Use 'binwidth
= x' to adjust this.
```



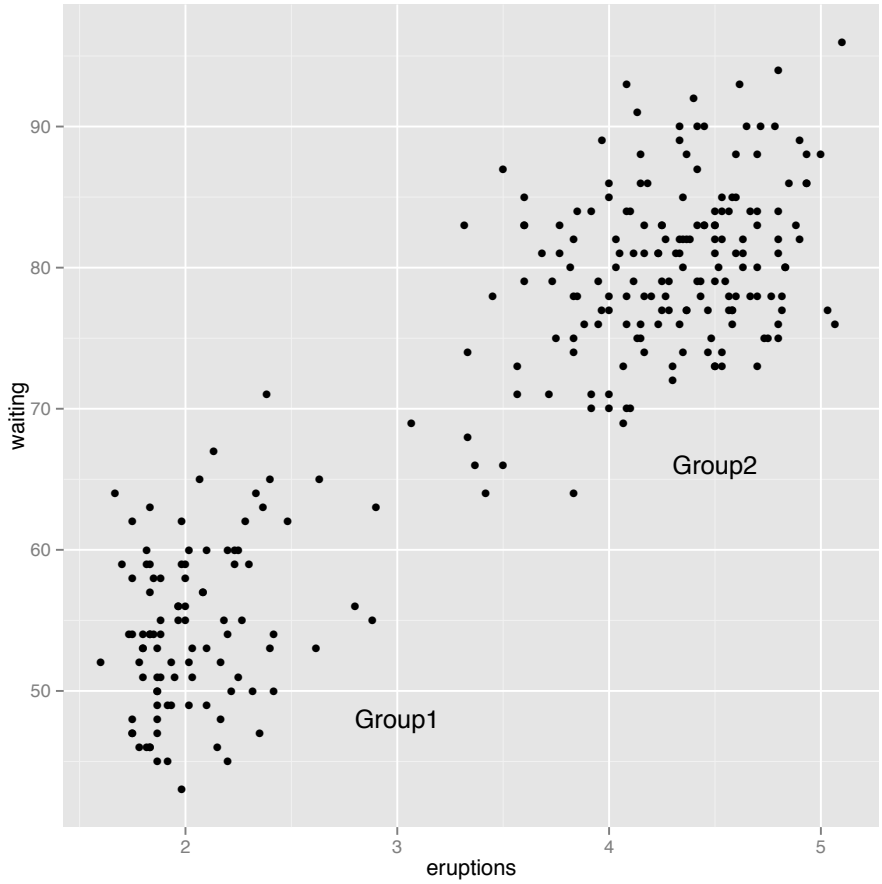
density 例子:

```
library(ggplot2)
library(gcookbook)
a <- ggplot(faithful, aes(x = eruptions, y = waiting)) + stat_density2d() +
    geom_point()
a
```


90
80
70
60
50
waiting
2
3
4
5
eruptions

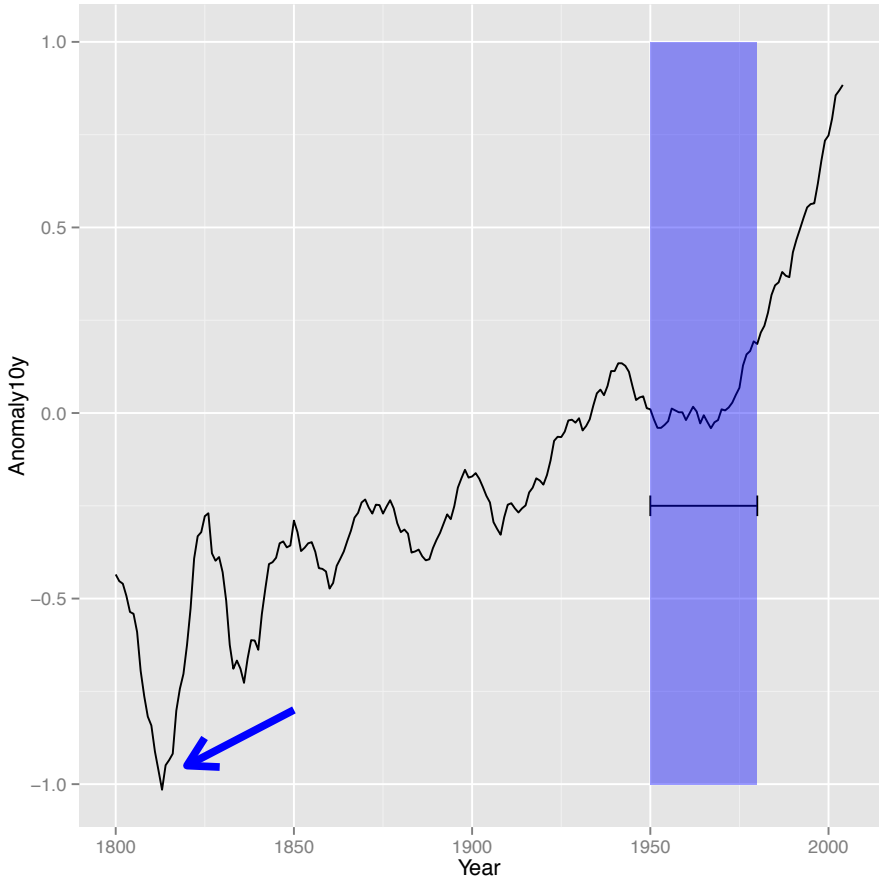
第七章　注释

```
library(ggplot2)
library(gcookbook)
p <- ggplot(data = faithful, aes(x = eruptions, y = waiting)) + geom_point()
p + annotate("text", x = 3, y = 48, label = "Group1") + annotate("text", x = 4.5,
    y = 66, label = "Group2")
```



箭头

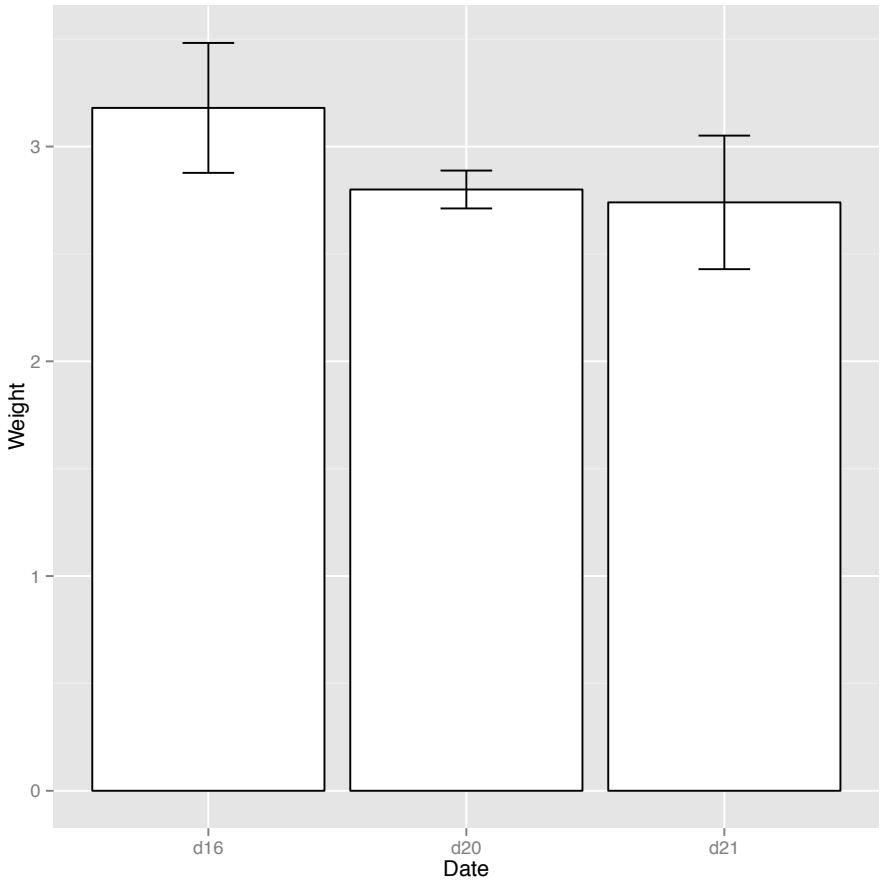
```
library(ggplot2)
library(gcookbook)
library(grid)
p <- ggplot(subset(climate, Source == "Berkeley"), aes(x = Year, y = Anomaly10y)) +
    geom_line()
p + annotate("segment", x = 1950, xend = 1980, y = -0.25, yend = -0.25, arrow = arrow(e
    angle = 90, length = unit(0.2, "cm"))) + annotate("segment", x = 1850, xend = 1820,
    y = -0.8, yend = -0.95, colour = "blue", size = 2, arrow = arrow()) + annotate("rec
    xmin = 1950, xmax = 1980, ymin = -1, ymax = 1, fill = "blue", alpha = 0.4)
```



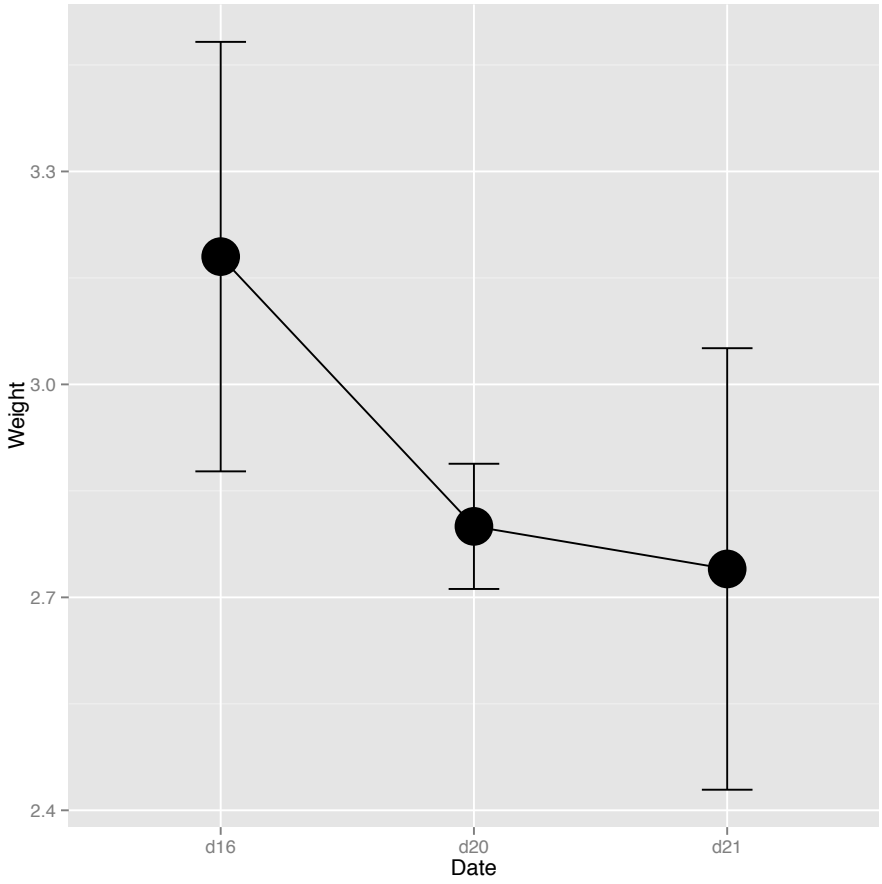
误差线:

```
library(ggplot2)
library(gcookbook)
p <- ggplot(subset(cabbage_exp, Cultivar == "c39"), aes(x = Date, y = Weight))
a <- p + geom_bar(fill = "white", color = "black") + geom_errorbar(aes(ymin = Weight -
    se, ymax = Weight + se), width = 0.2)
a

## Mapping a variable to y and also using stat="bin".
##  With stat="bin", it will attempt to set the y value to the count
of cases in each group.
##  This can result in unexpected behavior and will not be allowed
```

```
in a future version of ggplot2.
##  If you want y to represent counts of cases, use stat="bin" and
don't map a variable to y.
##  If you want y to represent values in the data, use stat="identity".
##  See ?geom_bar for examples.  (Deprecated; last used in version
0.9.2)
```



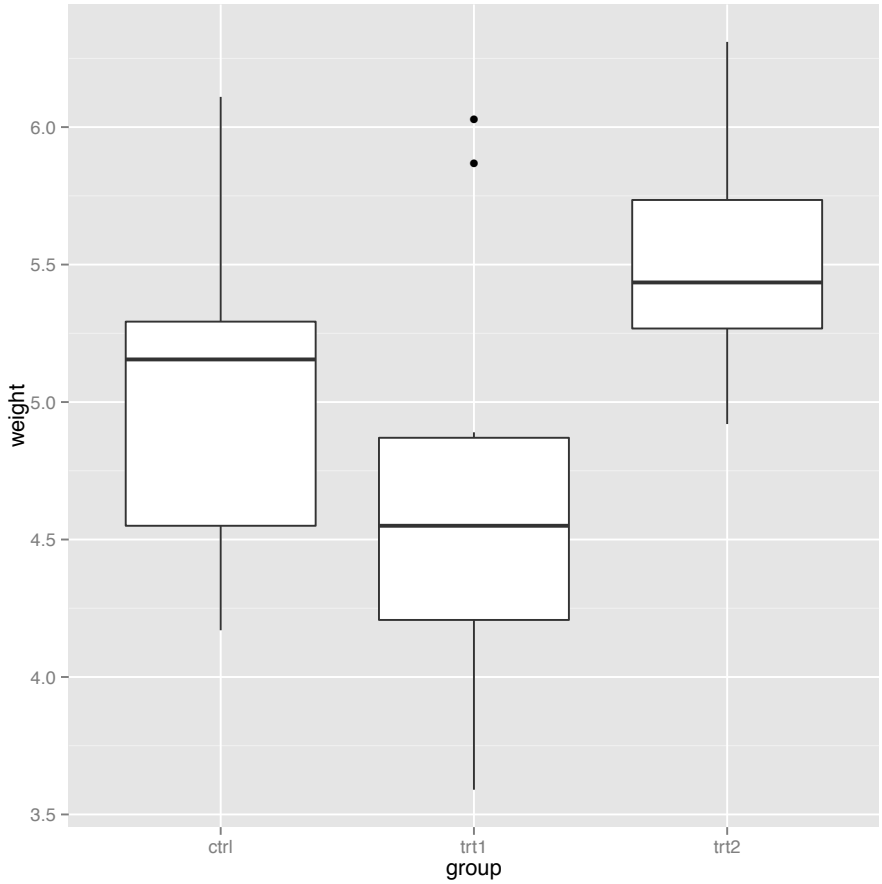

```
b <- p + geom_line(aes(group = 1)) + geom_point(size = 10) + geom_errorbar(aes(ymin = W
    se, ymax = Weight + se), width = 0.2)
b
```


3.3
3.0
2.7
2.4
Weight
d16
d20
d21
Date

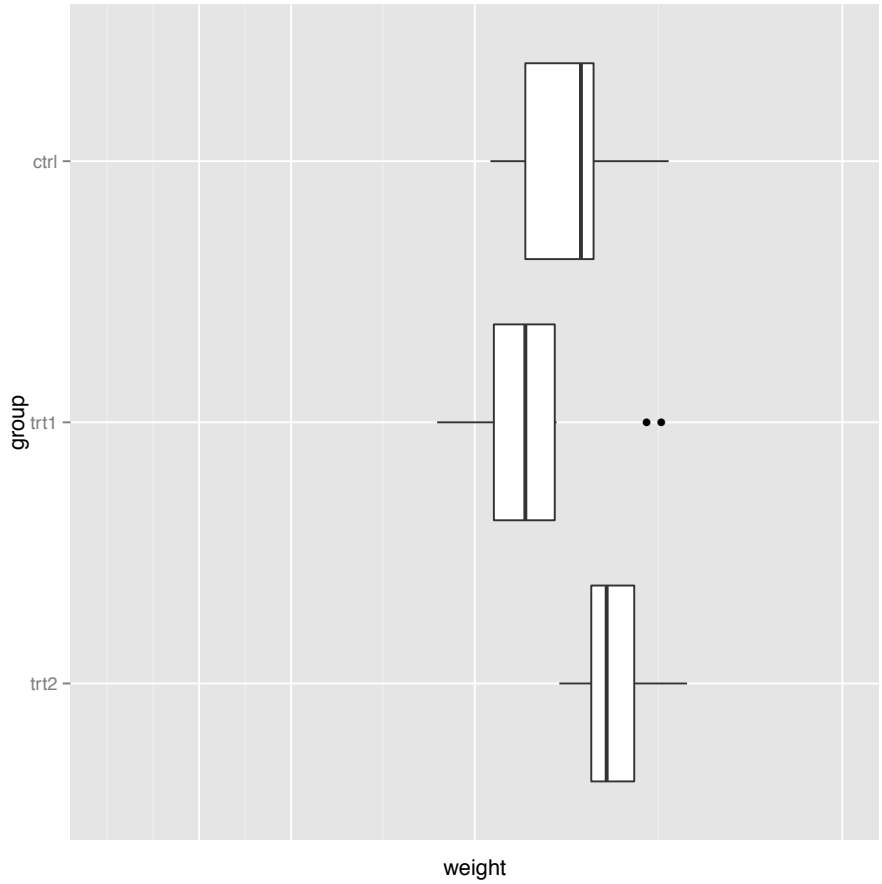
第八章　坐标轴

调整坐标刻度

```
library(ggplot2)
library(gcookbook)
p <- ggplot(data = PlantGrowth, aes(x = group, y = weight))
a <- p + geom_boxplot()
a
```

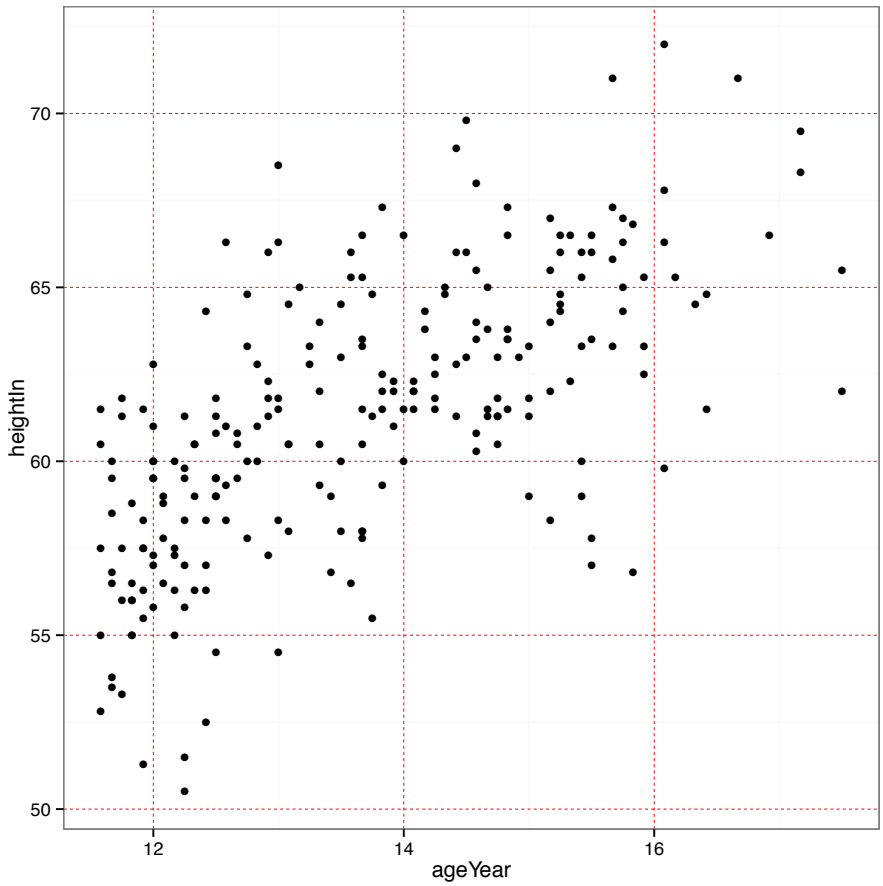



```
b <- p + geom_boxplot() + coord_flip() + scale_x_discrete(limits = rev(levels(PlantGrow
    scale_y_continuous(limits = c(0, 8), breaks = c(1, 2, 4, 8)) + theme(axis.text.x =
    axis.ticks.x = theme_blank())

## 'theme_blank' is deprecated.  Use 'element_blank' instead.  (Deprecated;
last used in version 0.9.1)
## 'theme_blank' is deprecated.  Use 'element_blank' instead.  (Deprecated;
last used in version 0.9.1)

b
```



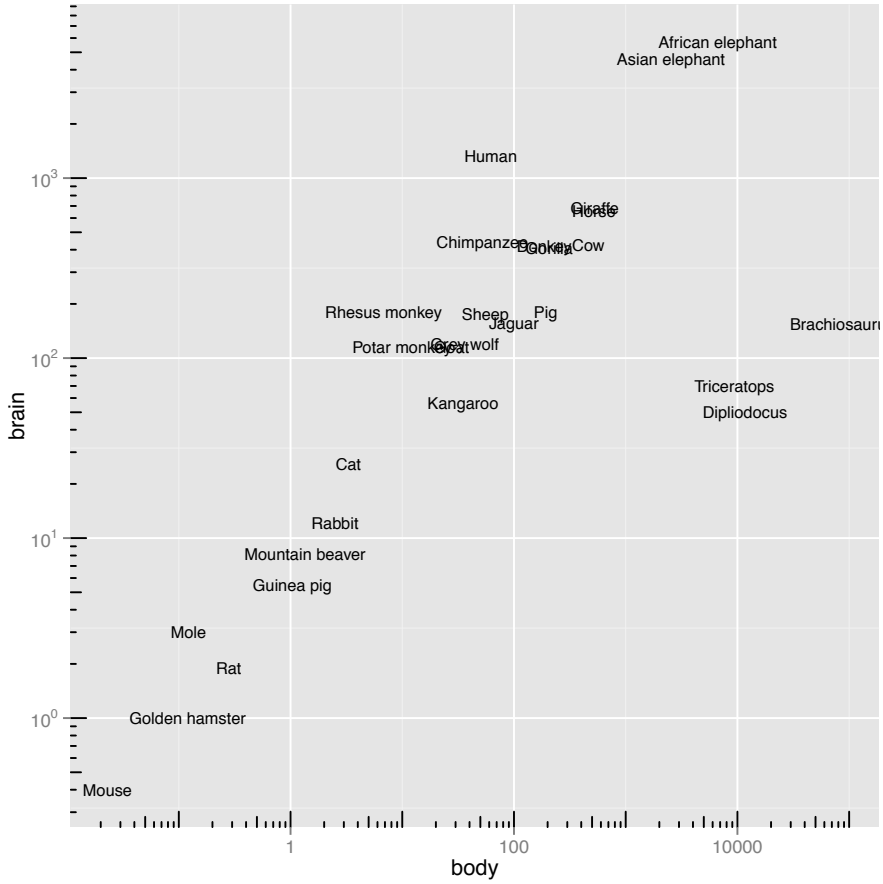
调整刻度线:

```
library(ggplot2)
library(gcookbook)
p <- ggplot(data = heightweight, aes(x = ageYear, y = heightIn)) + geom_point()
p + theme_bw() + theme(panel.grid.major = element_line(color = "red", linetype = "dashe
```



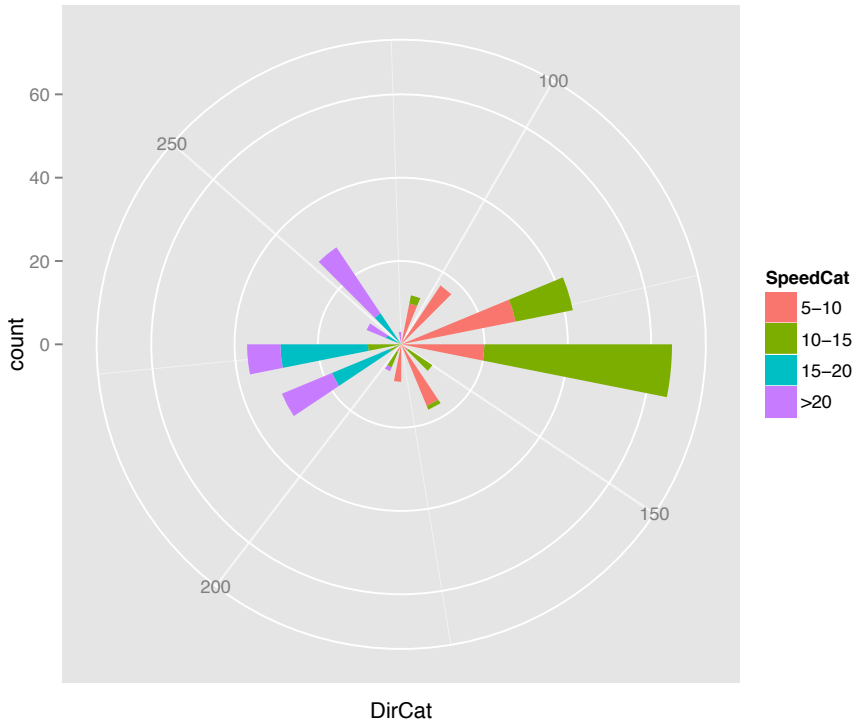
log 坐标刻度线绘制：

```
library(MASS)
library(scales)
library(ggplot2)
library(gcookbook)
ggplot(Animals, aes(x = body, y = brain, label = rownames(Animals))) + geom_text(size =
    scale_x_log10() + scale_y_log10(breaks = trans_breaks("log10", function(x) 10^x),
    labels = trans_format("log10", math_format(10^.x))) + annotation_logticks()
```



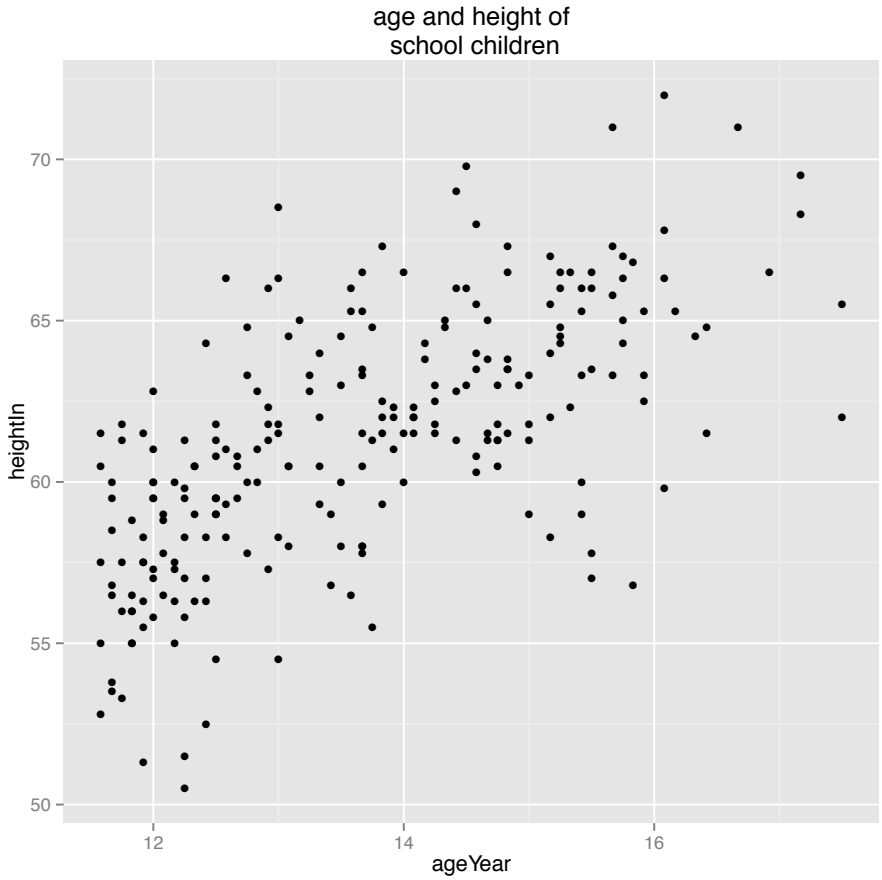
circle 图绘制：

```
library(ggplot2)
library(gcookbook)
ggplot(data = wind, aes(x = DirCat, fill = SpeedCat)) + coord_polar() + geom_histogram(

## stat_bin:  binwidth defaulted to range/30.  Use 'binwidth = x'
to adjust this.
```

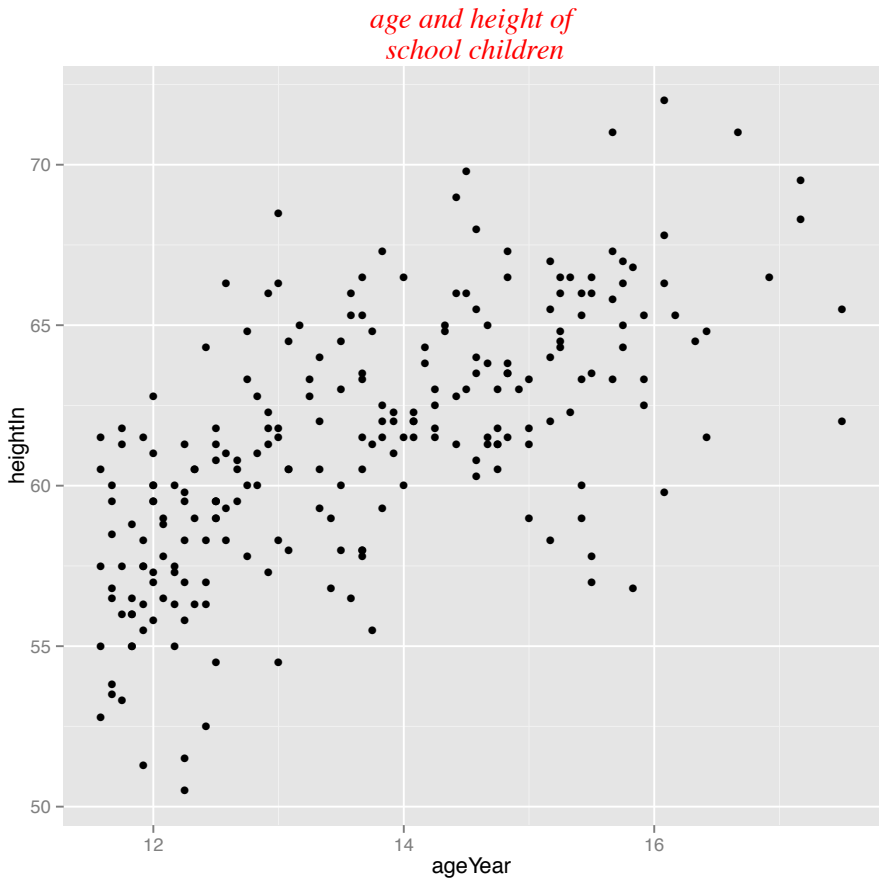

60
40
20
0
count
100
250
150
200
DirCat
SpeedCat
5–10
10–15
15–20
>20

第九章　图形全局参数控制

```
library(ggplot2)
library(gcookbook)
p <- ggplot(data = heightweight, aes(x = ageYear, y = heightIn)) + geom_point()
a <- p + ggtitle("age and height of\n school children")
a
```

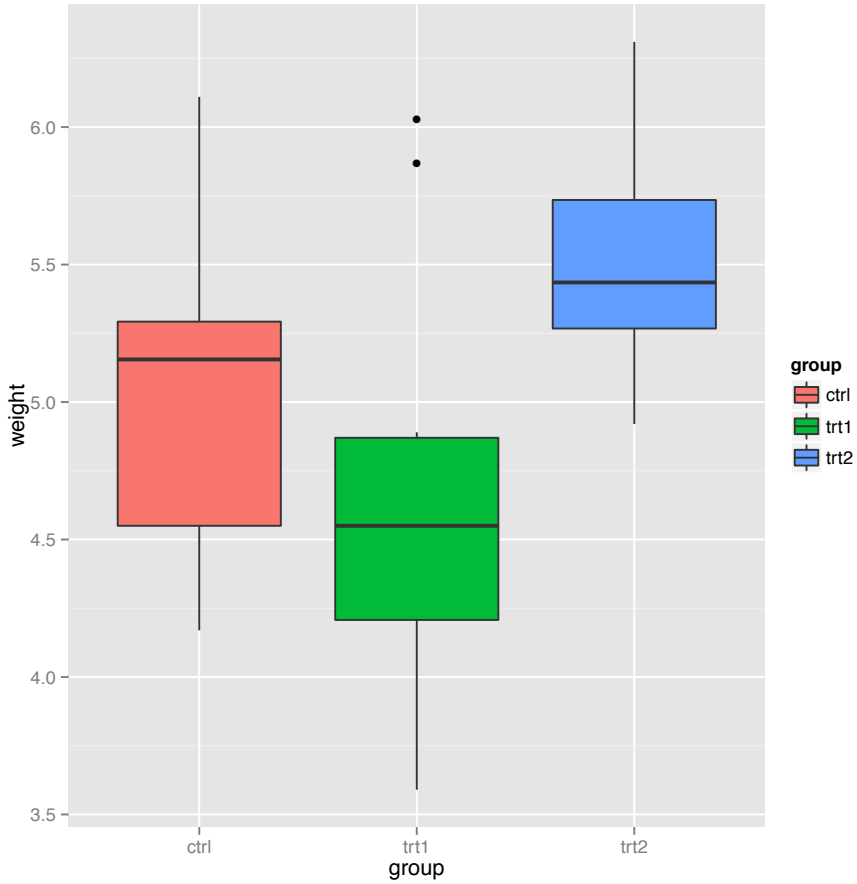



```
b <- a + theme(plot.title = element_text(size = I(16), lineheight = 0.9, family = "Time
    face = "italic", color = "red"))
b
```


age and height of
school children
70
65
60
55
50
heightIn
12
14
16
ageYear

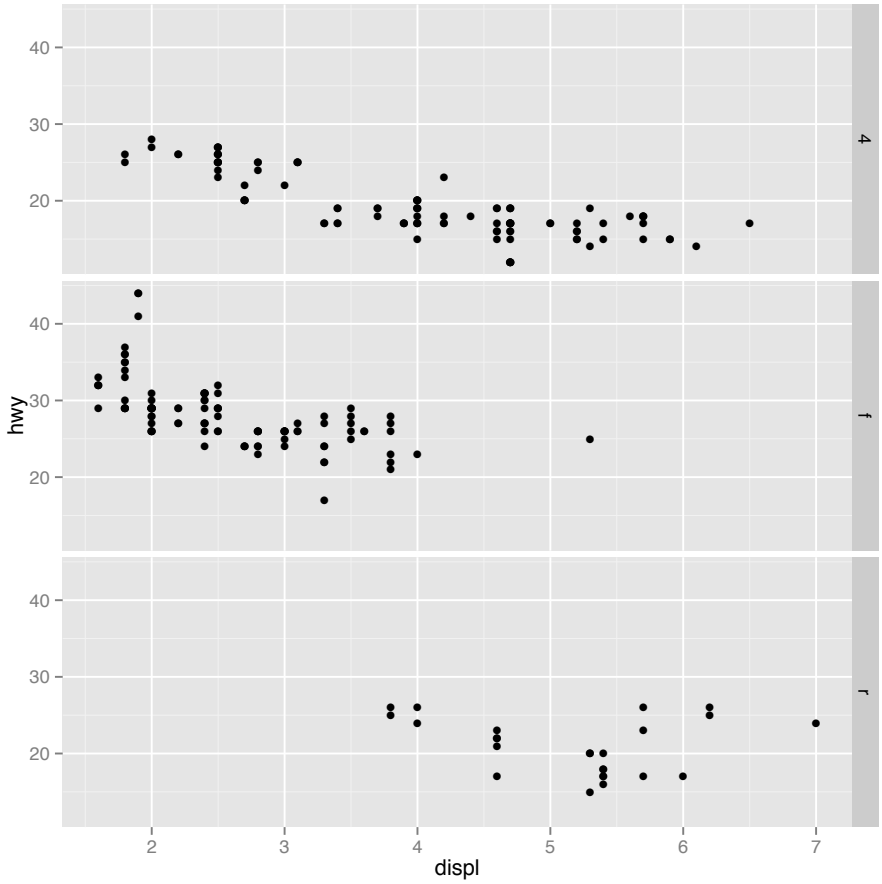
第十章　图例

```
library(ggplot2)
library(gcookbook)
p <- ggplot(PlantGrowth, aes(x = group, y = weight, fill = group)) + geom_boxplot()
p
```

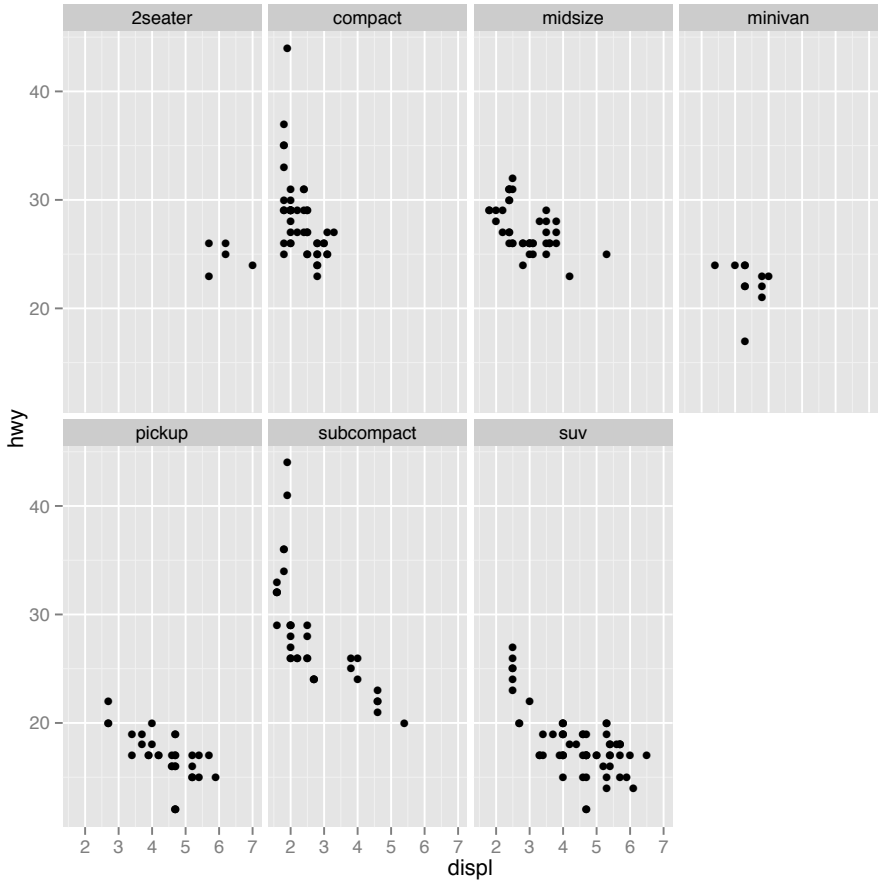

6.0
5.5
5.0
4.5
4.0
3.5
weight
ctrl
trt1
trt2
group
group
ctrl
trt1
trt2

第十一章　图层

```
library(ggplot2)
library(gcookbook)
p <- ggplot(mpg, aes(x = displ, y = hwy)) + geom_point()
a <- p + facet_grid(drv ~ .)
b <- p + facet_wrap(~class, ncol = 4)
a
```

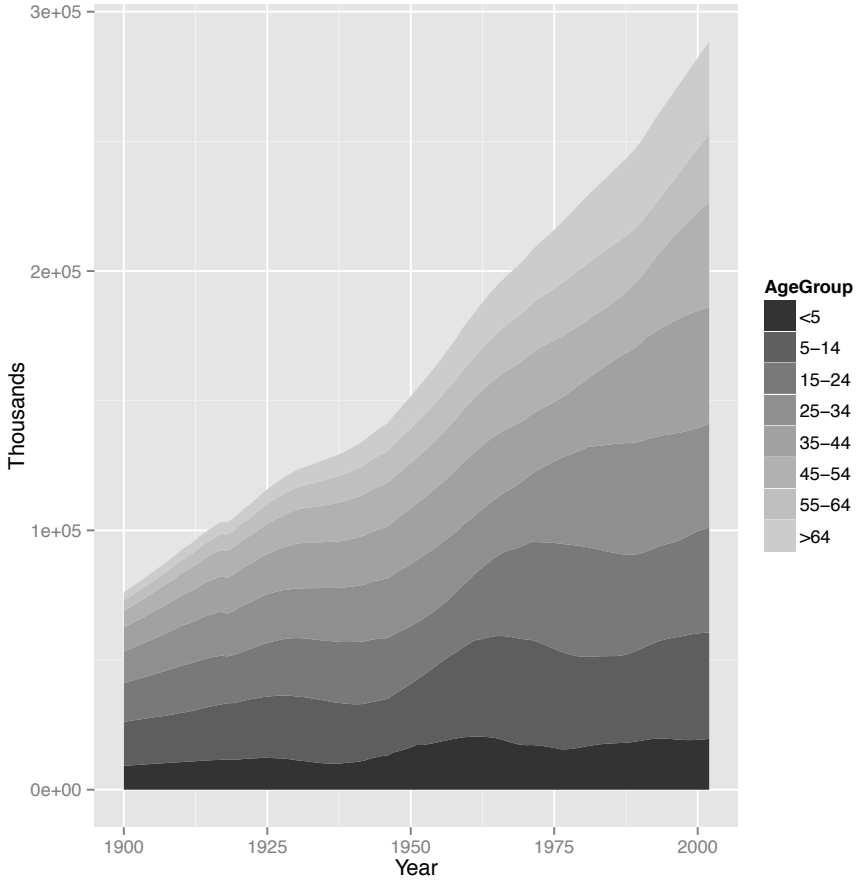



```
b
```

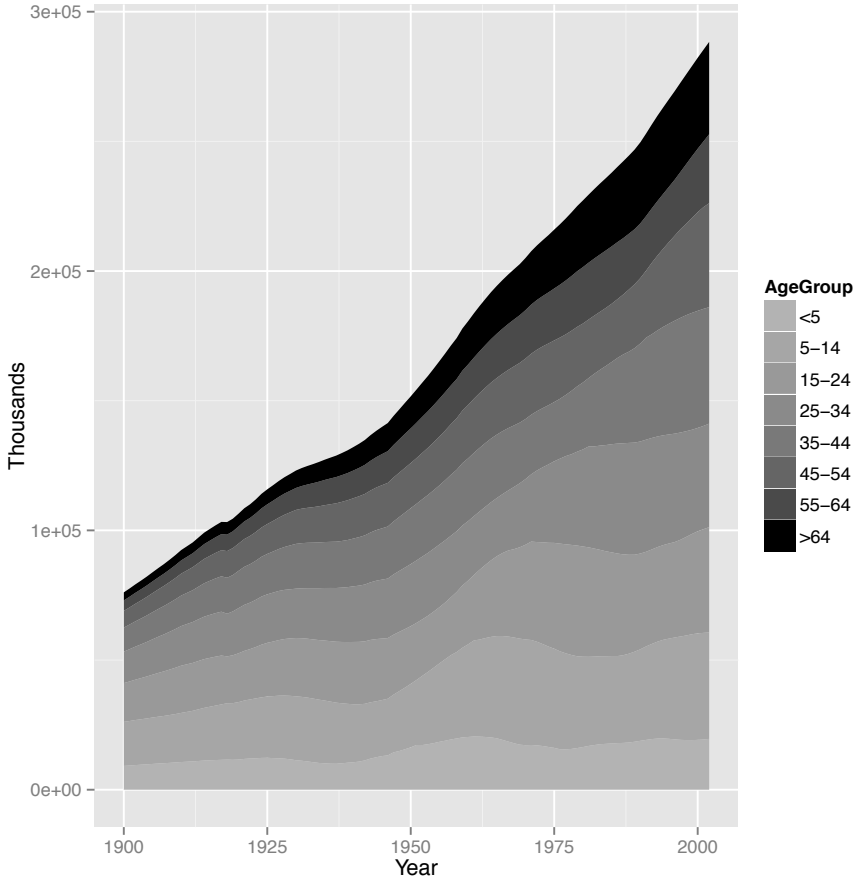

2seater
compact
midsize
minivan
pickup
subcompact
suv
hwy
displ
40
30
20
2
3
4
5
6
7

第十二章　颜色的使用

```
library(gcookbook)
library(ggplot2)
p <- ggplot(data = uspopage, aes(x = Year, y = Thousands, fill = AgeGroup)) +
    geom_area()
a <- p + scale_fill_grey()
a
```

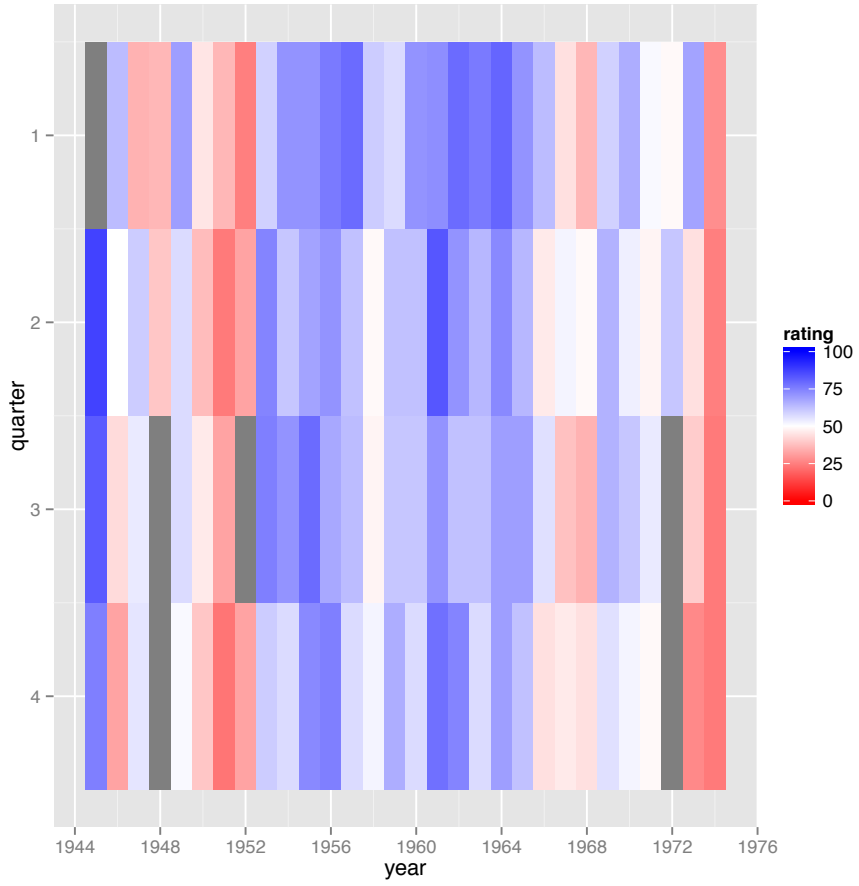



```
b <- p + scale_fill_grey(start = 0.7, end = 0)
b
```


3e+05
2e+05
1e+05
0e+00
Thousands
1900
1925
1950
1975
2000
Year
AgeGroup
<5
5–14
15–24
25–34
35–44
45–54
55–64
>64

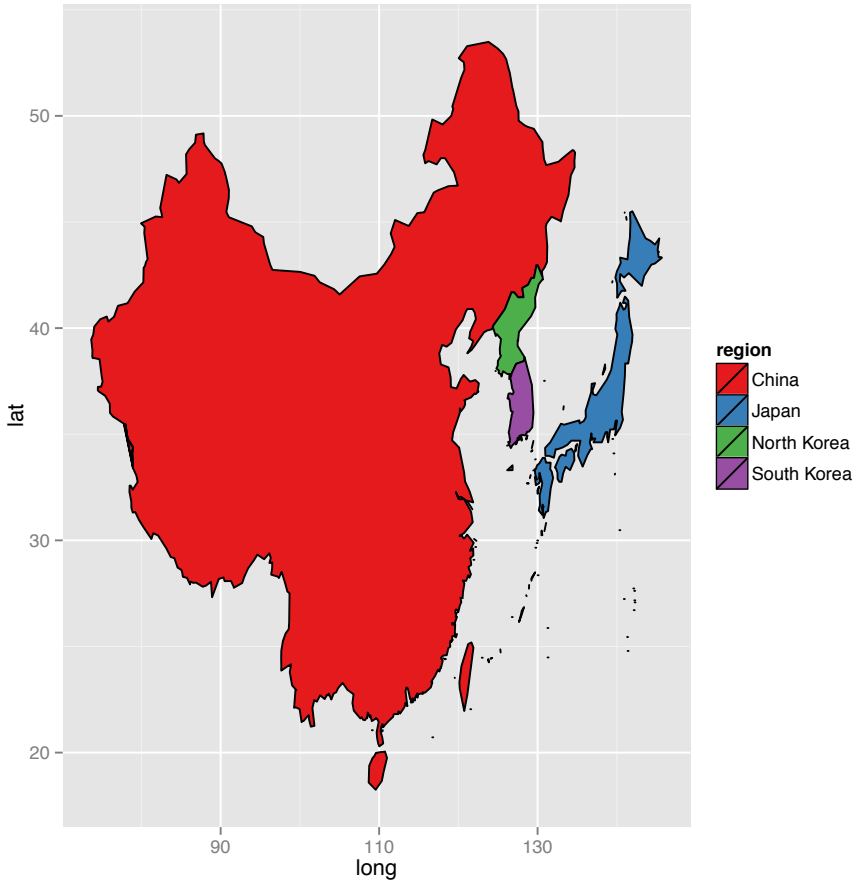
第十三章　图形实例

```
library(ggplot2)
library(gcookbook)
pres_rating <- data.frame(rating = as.numeric(presidents), year = as.numeric(floor(time
    quarter = as.numeric(cycle(presidents)))
p <- ggplot(pres_rating, aes(x = year, y = quarter, fill = rating))
p + geom_tile() + scale_x_continuous(breaks = seq(1940, 1976, by = 4)) + scale_y_revers
    scale_fill_gradient2(low = "red", high = "blue", midpoint = 50, limits = c(0,
        100))
```



地图

```
library(ggplot2)
library(gcookbook)
library(maps)
east_asia <- map_data("world", region = c("China", "Japan", "North Korea", "South Korea
p <- ggplot(east_asia, aes(x = long, y = lat, group = group, fill = region)) +
    geom_polygon(color = "black")
p + scale_fill_brewer(palette = "Set1")
```


50
40
30
20
lat
90
110
130
long
region
China
Japan
North Korea
South Korea

第十四章　图形输出格式

第十五章　数据处理

第十六章　A. ggplot2 介绍

第十七章　索引