

本章内容

- 条形图、箱线图和点图
- 饼图和扇形图
- 直方图与核密度图

我们无论在何时分析数据，第一件要做的事情就是观察它。对于每个变量，哪些值是最常见的？值域是大是小？是否有不寻常的观测？R中提供了丰富的数据可视化函数。本章，我们将关注那些可以帮助理解单个类别型或连续型变量的图形。主题包括：

- 将变量的分布作可视化展示；
- 通过结果变量进行跨组比较。

在以上话题中，变量可为连续型（例如，以每加仑汽油行驶的英里数表示的里程数）或类别型（例如，以无改善、一定程度的改善或明显改善表示的治疗结果）。在后续各章中，我们将探索那些展示双变量和多变量间关系的图形。

在接下来的几节中，我们将探索条形图、饼图、扇形图、直方图、核密度图、箱线图、小提琴图和点图的用法。有些图形可能你已经很熟悉了，而有些图形（如扇形图或小提琴图）可能你比较陌生。我们的目标同往常一样，都是更好地理解数据，并能够与他人沟通这些理解方式。

让我们从条形图开始。

6.1 条形图

条形图通过垂直的或水平的条形展示了类别型变量的分布（频数）。函数`barplot()`的最简单用法是：

```
barplot(height)
```

其中的`height`是一个向量或一个矩阵。

在接下来的示例中，我们将绘制一项探索类风湿性关节炎新疗法研究的结果。数据已包含在随vcd包分发的Arthritis数据框中。由于vcd包并没包括在R的默认安装中，请确保在第一次使用之前先下载并安装它（`install.packages("vcd")`）。

注意，我们并不需要使用vcd包来创建条形图。我们读入它的原因是为了使用Arthritis数据集。但我们需要使用vcd包创建6.1.5节中描述的棘状图（spinogram）。

6.1.1 简单的条形图

若height是一个向量，则它的值就确定了各条形的高度，并将绘制一幅垂直的条形图。使用选项horiz=TRUE则会生成一幅水平条形图。你也可以添加标注选项。选项main可添加一个图形标题，而选项xlab和ylab则会分别添加x轴和y轴标签。

在关节炎研究中，变量Improved记录了对每位接受了安慰剂或药物治疗的病人的治疗结果。

```
> library(vcd)
> counts <- table(Arthritis$Improved)
> counts
  None  Some Marked
   42   14   28
```

这里我们看到，28位病人有了明显改善，14人有部分改善，而42人没有改善。我们将在第7章更充分地讨论使用table()函数提取各单元的计数的方法。

你可以使用一幅垂直或水平的条形图来绘制变量counts。代码见代码清单6-1，结果如图6-1所示。

代码清单6-1 简单的条形图

```
barplot(counts,
  main="Simple Bar Plot",
  xlab="Improvement", ylab="Frequency")
```

← 简单条形图

```
barplot(counts,
  main="Horizontal Bar Plot",
  xlab="Frequency", ylab="Improvement",
  horiz=TRUE)
```

← 水平条形图

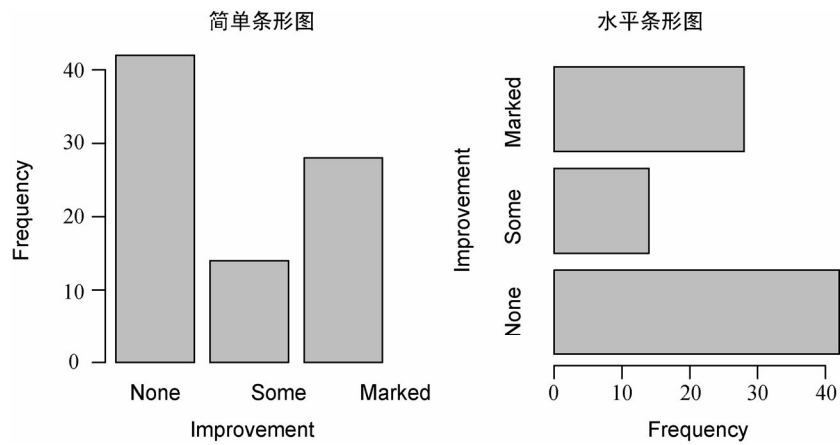


图6-1 简单的垂直条形图和水平条形图

小提示 若要绘制的类别型变量是一个因子或有序型因子，就可以使用函数`plot()`快速创建一幅垂直条形图。由于`Arthritis$Improved`是一个因子，所以代码：

```
plot(Arthritis$Improved, main="Simple Bar Plot",
     xlab="Improved", ylab="Frequency")
plot(Arthritis$Improved, horiz=TRUE, main="Horizontal Bar Plot",
     xlab="Frequency", ylab="Improved")
```

将和代码清单6-1生成相同的条形图，而无需使用`table()`函数将其表格化。

如果标签很长怎么办？在6.1.4节中，你将看到微调标签的方法，这样它们就不会重叠了。

6.1.2 堆砌条形图和分组条形图

如果`height`是一个矩阵而不是一个向量，则绘图结果将是一幅堆砌条形图或分组条形图。若`beside=FALSE`（默认值），则矩阵中的每一列都将生成图中的一个条形，各列中的值将给出堆砌的“子条”的高度。若`beside=TRUE`，则矩阵中的每一列都表示一个分组，各列中的值将并列而不是堆砌。

考虑治疗类型和改善情况的列联表：

```
> library(vcd)
> counts <- table(Arthritis$Improved, Arthritis$Treatment)
> counts
```

	Treatment	
Improved	Placebo	Treated
None	29	13
Some	7	7
Marked	7	21

你可以将此结果绘制为一幅堆砌条形图或一幅分组条形图（见代码清单6-2）。结果如图6-2所示。

代码清单6-2 堆砌条形图和分组条形图

```
barplot(counts,                                     <— 堆砌条形图
        main="Stacked Bar Plot",
        xlab="Treatment", ylab="Frequency",
        col=c("red", "yellow", "green"),
        legend=rownames(counts))

barplot(counts,                                     <— 分组条形图
        main="Grouped Bar Plot",
        xlab="Treatment", ylab="Frequency",
        col=c("red", "yellow", "green"),
        legend=rownames(counts), beside=TRUE)
```

第一个`barplot`函数绘制了一幅堆砌条形图，而第二个绘制了一幅分组条形图。我们同时使用`col`选项为绘制的条形添加了颜色。参数`legend.text`为图例提供了各条形的标签（仅在`height`为一个矩阵时有用）。

在第3章中，我们讲解过格式化和放置图例的方法，以确保最好的效果。请试着重新排布图例的位置以避免它们和条形产生叠加。

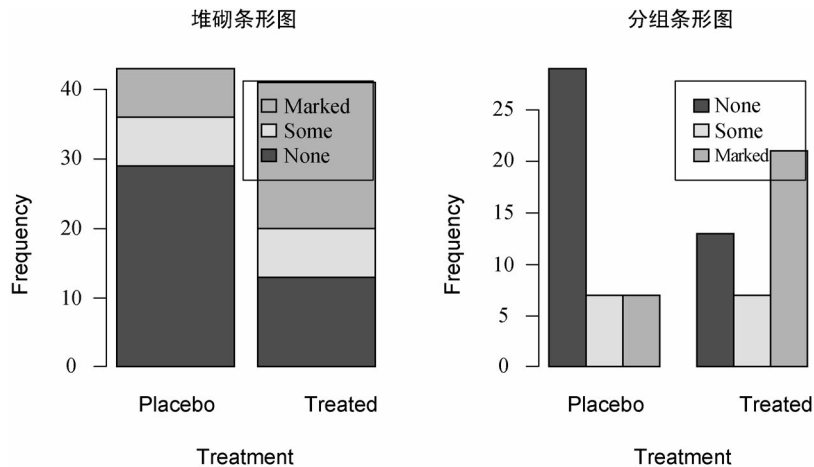


图6-2 堆砌条形图和分组条形图

6

6.1.3 均值条形图

条形图并不一定要基于计数数据或频率数据。你可以使用数据整合函数并将结果传递给 `barplot()` 函数，来创建表示均值、中位数、标准差等的条形图。代码清单6-3展示了一个示例，结果如图6-3所示。

代码清单6-3 排序后均值的条形图

```
> states <- data.frame(state.region, state.x77)
> means <- aggregate(states$Illiteracy, by=list(state.region), FUN=mean)
> means
  Group.1      x
1 Northeast 1.00
2   South 1.74
3 North Central 0.70
4    West 1.02
> means <- means[order(means$x),]
> means
  Group.1      x
3 North Central 0.70
1 Northeast 1.00
4    West 1.02
2   South 1.74
> barplot(means$x, names.arg=means$Group.1)
> title("Mean Illiteracy Rate")
```

① 将均值从小到大排序

② 添加标题

代码清单6-3将均值从小到大排序①。同时注意，使用 `title()` 函数②与调用 `plot()` 时添加 `main` 选项是等价的。`means$x` 是包含各条形高度的向量，而添加选项 `names.arg=means$Group.1` 是为了展示标签。

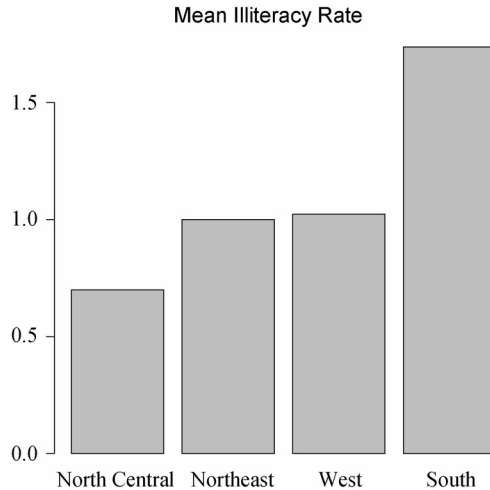


图6-3 美国各地区平均文盲率排序的条形图

你可以进一步完善这个示例。各个条形可以使用 `lines()` 函数绘制的线段连接起来。你也可以使用 `gplots` 包中的 `barplot2()` 函数创建叠加有置信区间的均值条形图。R Graph Gallery 网站 (<http://addictedtor.free.fr/graphiques>) 上的 “barplot2: Enhanced Bar Plots” 页面可以作为一个参考示例。

6.1.4 条形图的微调

有若干种方式可以微调条形图的外观。例如，随着条数的增多，条形的标签可能会开始重叠。你可以使用参数 `cex.names` 来减小字号。将其指定为小于1的值可以缩小标签的大小。可选的参数 `names.arg` 允许你指定一个字符向量作为条形的标签名。你同样可以使用图形参数辅助调整文本间隔。代码清单6-4给出了一个示例，输出如图6-4所示。

代码清单6-4 为条形图搭配标签

```
par(mar=c(5,8,4,2))
par(las=2)
counts <- table(Arthritis$Improved)

barplot(counts,
        main="Treatment Outcome",
        horiz=TRUE, cex.names=0.8,
        names.arg=c("No Improvement", "Some Improvement",
                    "Marked Improvement"))
```

本例中，我们（使用 `las=2`）旋转了条形的标签、修改了标签文本，并（使用 `mar`）增加了y边界的大小，为了让标签更合适，（使用 `cex.names=0.8`）缩小了字体大小。`par()` 函数能够让你对R的默认图形做出大量修改。详情参阅第3章。

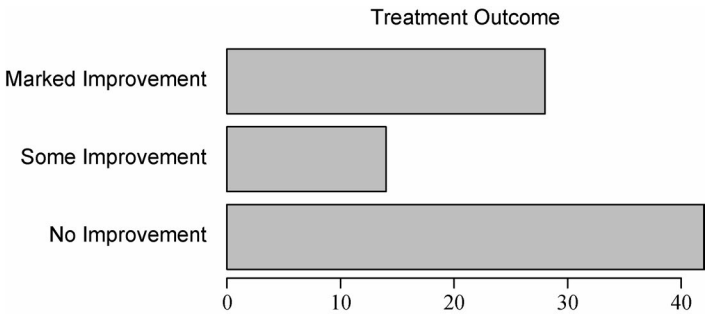


图6-4 微调了标签的垂直条形图

6.1.5 棘状图

在结束关于条形图的讨论之前，让我们再来看一种特殊的条形图，它称为棘状图（spinogram）。棘状图对堆砌条形图进行了重缩放，这样每个条形的高度均为1，每一段的高度即表示比例。棘状图可由vcd包中的函数spine()绘制。以下代码可以生成一幅简单的棘状图：

```
library(vcd)
attach(Arthritis)
counts <- table(Treatment, Improved)
spine(counts, main="Spinogram Example")
detach(Arthritis)
```

输出如图6-5所示。治疗组同安慰剂组相比，获得显著改善的患者比例明显更高。

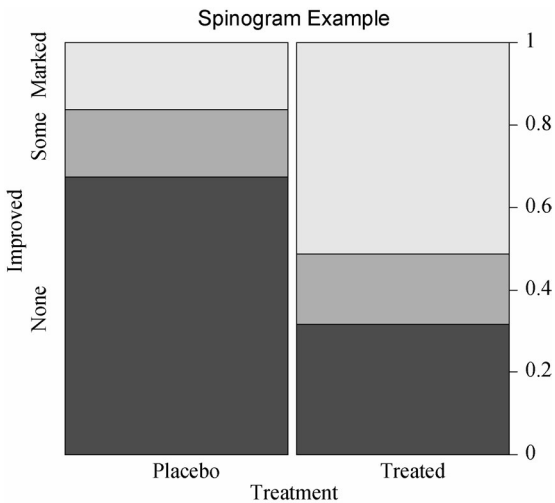


图6-5 关节炎治疗结果的棘状图

除了条形图，饼图也是一种用于展示类别型变量分布的流行工具，接下来讨论它。