

第二章 R语言绘图基础教程

复旦大学附属肿瘤医院 周支瑞





2.1 R 语言绘图初阶

表3-1 病人对两种药物五个剂量水平上的响应情况

剂 量	对药物 A 的响应	对药物 B 的响应
20	16	15
30	20	18
40	27	25
45	40	31
60	60	40

- 字体、颜色、坐标轴、标签等
- 通过函数`par()`来指定选项。以这种方式设定的参数值除非被再次修改，否则将在会话结束前一直有效。其调用格式为：
`par(optionname=value,optionname=name,...)`。
- 不加参数地执行`par()`将生成一个含有当前图形参数设置的列表。添加参数`no.readonly=TRUE`可以生成一个可以修改的当前图形参数列表。

表3-2 用于指定符号和线条类型的参数

参 数	描 述
pch	指定绘制点时使用的符号（见图 3-4）
cex	指定符号的大小。cex 是一个数值，表示绘图符号相对于默认大小的缩放倍数。默认大小为 1，1.5 表示放大为默认值的 1.5 倍，0.5 表示缩小为默认值的 50%，等等
lty	指定线条类型（参见图 3-5）
lwd	指定线条宽度。lwd 是以默认值的相对大小来表示的（默认值为 1）。例如，lwd=2 将生成一条两倍于默认宽度的线条



图3-4 参数pch可指定的绘图符号



图3-5 参数lty可指定的线条类型

表3-3 用于指定颜色的参数

参 数	描 述
col	默认的绘图颜色。某些函数（如 lines 和 pie）可以接受一个含有颜色值的向量并自动循环使用。 例如，如果设定 col=c("red", "blue") 并需要绘制三条线，则第一条线将为红色，第二条线为蓝色，第三条线又将为红色
col.axis	坐标轴刻度文字的颜色
col.lab	坐标轴标签（名称）的颜色
col.main	标题颜色
col.sub	副标题颜色
fg	图形的前景色
bg	图形的背景色

表3-4 用于指定文本大小的参数

参 数	描 述
<code>cex</code>	表示相对于默认大小缩放倍数的数值。默认大小为 1, 1.5 表示放大为默认值的 1.5 倍, 0.5 表示缩小为默认值的 50%, 等等
<code>cex.axis</code>	坐标轴刻度文字的缩放倍数。类似于 <code>cex</code>
<code>cex.lab</code>	坐标轴标签（名称）的缩放倍数。类似于 <code>cex</code>
<code>cex.main</code>	标题的缩放倍数。类似于 <code>cex</code>
<code>cex.sub</code>	副标题的缩放倍数。类似于 <code>cex</code>

表3-5 用于指定字体族、字号和字样的参数

参 数	描 述
font	整数。用于指定绘图使用的字体样式。1=常规，2=粗体，3=斜体，4=粗斜体，5=符号字体（以 Adobe 符号编码表示）
font.axis	坐标轴刻度文字的字体样式
font.lab	坐标轴标签（名称）的字体样式
font.main	标题的字体样式
font.sub	副标题的字体样式
ps	字体磅值（1 磅约为 1/72 英寸）。文本的最终大小为 $ps * cex$
family	绘制文本时使用的字体族。标准的取值为 serif（衬线）、sans（无衬线）和 mono（等宽）

表3-6 用于控制图形尺寸和边界大小的参数

参 数	描 述
pin	以英寸表示的图形尺寸（宽和高）
mai	以数值向量表示的边界大小，顺序为“下、左、上、右”，单位为英寸
mar	以数值向量表示的边界大小，顺序为“下、左、上、右”，单位为英分 ^① 。默认值为 $c(5, 4, 4, 2) + 0.1$

- 可使用title()函数为图形添加标题和坐标轴标签，调用格式： title(main="main title", sub="subtitle", xlab="x-axis label", ylab="y-axis label")
- 你可以使用函数axis()来创建自定义的坐标轴，而非使用R中的默认坐标轴，其格式为： axis(side, at=, labels=, pos=, lty=, col=, las=, tck=, ...)

表3-7 坐标轴选项

选 项	描 述
side	一个整数，表示在图形的哪边绘制坐标轴（1=下，2=左，3=上，4=右）
at	一个数值型向量，表示需要绘制刻度线的位置
labels	一个字符型向量，表示置于刻度线旁边的文字标签（如果为 NULL，则将直接使用 at 中的值）
pos	坐标轴线绘制位置的坐标（即与另一条坐标轴相交位置的值）
lty	线条类型
col	线条和刻度线颜色
las	标签是否平行于（=0）或垂直于（=2）坐标轴
tck	刻度线的长度，以相对于绘图区域大小的分数表示（负值表示在图形外侧，正值表示在图形内侧，0 表示禁用刻度，1 表示绘制网格线）；默认值为-0.01
(...)	其他图形参数

- 函数`abline()`可以用来为图形添加参考线，其使用格式为：`abline(h=yvalues, v=xvalues)`
- 我们可以使用函数`legend()`来添加图例，其使用格式为：`legend(location, title, legend, ...)`

表3-8 图例选项

选 项	描 述
<code>location</code>	有许多方式可以指定图例的位置。你可以直接给定图例左上角的 <code>x</code> 、 <code>y</code> 坐标，也可以执行 <code>locator(1)</code> ，然后通过鼠标单击给出图例的位置，还可以使用关键字 <code>bottom</code> 、 <code>bottomleft</code> 、 <code>left</code> 、 <code>topleft</code> 、 <code>top</code> 、 <code>topright</code> 、 <code>right</code> 、 <code>bottomright</code> 或 <code>center</code> 放置图例。如果你使用了以上某个关键字，那么可以同时使用参数 <code>inset</code> =指定图例向图形内侧移动的大小（以绘图区域大小的分数表示）
<code>title</code>	图例标题的字符串（可选）
<code>legend</code>	图例标签组成的字符型向量
<code>...</code>	其他选项。如果图例标示的是颜色不同的线条，需要指定 <code>col</code> =加上颜色值组成的向量。如果图例标示的是符号不同的点，则需指定 <code>pch</code> =加上符号的代码组成的向量。如果图例标示的是不同的线条宽度或线条类型，请使用 <code>lwd</code> =或 <code>lty</code> =加上宽度值或类型值组成的向量。要为图例创建颜色填充的盒形（常见于条形图、箱线图或饼图），需要使用参数 <code>fill</code> =加上颜色值组成的向量

- 函数text()和mtext()将文本添加到图形上。text()可向绘图区域内部添加文本，而mtext()则向图形的四个边界之一添加文本。使用格式分别为：
 - text(location, "text to place", pos, ...)
 - mtext("text to place", side, line=n, ...)

表3-9 函数text()和mtext()的选项

选 项	描 述
<i>location</i>	文本的位置参数。可为一对 <i>x</i> 、 <i>y</i> 坐标，也可通过指定 <i>location</i> 为 <i>locator(1)</i> 使用鼠标交互式地确定摆放位置
<i>pos</i>	文本相对于位置参数的方位。1=下，2=左，3=上，4=右。如果指定了 <i>pos</i> ，就可以同时指定参数 <i>offset=</i> 作为偏移量，以相对于单个字符宽度的比例表示
<i>side</i>	指定用来放置文本的边。1=下，2=左，3=上，4=右。你可以指定参数 <i>line=</i> 来内移或外移文本，随着值的增加，文本将外移。也可使用 <i>adj=0</i> 将文本向左下对齐，或使用 <i>adj=1</i> 右上对齐

- `par()`函数中使用图形参数 `mfrow=c(nrows, ncols)` 来创建按行填充的、行数为 `nrows`、列数为 `ncols` 的图形矩阵。另外，可以使用 `mfcrow=c(nrows, ncols)` 按列填充矩阵



2.2 R基本图形绘制

- 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。

- 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。

- 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。

- 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。

- 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。

- 饼图可由以下函数创建: `pie(x, labels)`
- 其中`x`是一个非负数值向量, 表示每个扇形的面积, 而`labels`则是表示各扇形标签的字符型向量。

- 直方图通过在x轴上将值域分割为一定数量的组，在y轴上显示相应值的频数，展示了连续型变量的分布。可以使用如下函数创建直方图：hist(x)
- 其中的x是一个由数据值组成的数值向量。参数 freq=FALSE 表示根据概率密度而不是频数绘制图形。参数breaks用于控制组的数量。在定义直方图中的单元时，默认将生成等距切分。

- 在上节中，你看到了直方图上叠加的核密度图。用术语来说，核密度估计是用于估计随机变量概率密度函数的一种非参数方法。虽然其数学细节已经超出了本书的范畴，但从总体上讲，核密度图不失为一种用来观察连续型变量分布的有效方法。绘制密度图的方法（不叠加到另一幅图上方）为：`plot(density(x))`
- 其中的x是一个数值型向量。由于`plot()`函数会创建一幅新的图形，所以要向一幅已经存在的图形上叠加一条密度曲线，可以使用`lines()`函数
- 使用sm包中的`sm.density.compare()`函数可向图形叠加两组或更多的核密度图。使用格式为：`sm.density.compare(x, factor)`

- 箱线图（又称盒须图）通过绘制连续型变量的五数总括，即最小值、下四分位数（第25百分位数）、中位数（第50百分位数）、上四分位数（第75百分位数）以及最大值，描述了连续型变量的分布。箱线图能够显示出可能为离群点（范围 $\pm 1.5 * IQR$ 以外的值，IQR表示四分位距，即上四分位数与下四分位数的差值）的观测。例如：

```
boxplot(mtcars$mpg, main="Box plot", ylab="Miles per Gallon")
```

- 箱线图可展示单个变量或分组变量。使用格式：`boxplot(formula,data=dataframe)` 其中的formula是一个公式，dataframe代表提供数据的数据框（或列表）。一个示例公式为 $y \sim A$ ，这将为类别型变量A的每个值并列地生成数值型变量y的箱线图。公式 $y \sim A*B$ 则将为类别型变量A和B所有水平的两两组合生成数值型变量y的箱线图。
- 添加参数`varwidth=TRUE` 将使箱线图的宽度与其样本大小的平方根成正比。参数`horizontal=TRUE`可以反转坐标轴的方向。
- 箱线图灵活多变，通过添加`notch=TRUE`，可以得到含凹槽的箱线图。若两个箱的凹槽互不重叠，则表明它们的中位数有显著差异

- 小提琴图是箱线图与核密度图的结合。你可以使用vioplot包中的vioplot()函数绘制它。请在第一次使用之前安装vioplot包。
- vioplot()函数的使用格式为：vioplot(x1, x2, ... , names=, col=) 其中x1, x2, ... 表示要绘制的一个或多个数值向量（将为每个向量绘制一幅小提琴图）。参数names是小提琴图中标签的字符向量，而col是一个为每幅小提琴图指定颜色的向量。

- 点图提供了一种在简单水平刻度上绘制大量有标签值的方法。你可以使用 `dotchart()` 函数创建点图，格式为：`dotchart(x, labels=)`
- 其中的 `x` 是一个数值向量，而 `labels` 则是由每个点的标签组成的向量。你可以通过添加参数 `groups` 来选定一个因子，用以指定 `x` 中元素的分组方式。如果这样做，则参数 `gcolor` 可以控制不同组标签的颜色，`cex` 可以控制标签的大小。
- R 中创建散点图的基础函数是 `plot(x, y)`，其中，`x` 和 `y` 是数值型向量，代表着图形中的 `(x, y)` 点。

- 如果将散点图上的点从左往右连接起来，就会得到一个折线图。以基础安装中的 Orange 数据集为例，它包含五种橘树的树龄和年轮数据。

- 【1】 Robert I. Kabacoff 著, 《R语言实战 》(第2版), 人民邮电出版社, 2016
- 【2】 Peter Dalgaard 著, 《R语言统计入门》 》(第2版), 人民邮电出版社, 2014
- 【3】 薛毅 陈立萍 著, 《R语言实用教程》, 清华大学出版社, 2014
- 【4】 张铁军 陈兴栋 刘振球 著, 《R语言与医学统计图形》, 人民卫生出版社, 2018

Thanks!

感谢您的观看!