Lesson 9: Data Visualization 1

Introduction to Plotting Systems in R

Chris Qi

R的三大绘图系统

- 基本绘图系统（Base Plotting System）
- Lattice绘图系统（Lattice Plotting System）
- ggplot2 绘图系统（ggplot2 Plotting System）

基本绘图系统（Base Plotting System）

- 人人都是艺术家：绘图开始于空白，一笔一划添加内容
- 需要事先计划好，直观地实时反映绘图和分析数据的逻辑
- 用基本R绘图需要走两步
  - 图
  - 修饰/添加=执行一系列函数
- 适于绘制2D图
- quick, easy, dirty

基本绘图系统

- 绘图函数（graphics包）
  - plot/hist/boxplot/points/lines/text/title/axis
- plot()
  - plot(x,y,...)
  - 重要参数：xlab/ylab/lwd/lty/pach/col
  - ?par
    - 用于设置全局参数（作用于R中的所有plot绘图）
      - bg/mar/las/mfrow/mfcol
      - 这些参数可以在每次plot之前修改

Lattice绘图系统

- 绘图=使用一次函数调用（一次成图）
- 特别适用于观测变量间的交互：在变量z的不同水平，变量y如何随着x变化

Lattice绘图系统

- 绘图函数
  - lattice包
    - xyplot/bwplot/histogram/stripplot/dotplot/splom/levelplot.contourplot
    - 格式xyplot(y~x|f*g,data)，f*g是分类变量
    - panel函数，用于控制每个面板内的绘图
  - grid包
    - 实现了独立于base的绘图系统
    - lattice包是基于grd创建的，很少直接从grid包调用函数

Lattice绘图系统

- xyplot(Temp~Ozone|Month, data=airquality,　　layout=c(5,1))
- Lattice与Base的重要区别
  - base绘图函数直接在图形设备上绘图
  - Lattice绘图函数返回trellis类对象
    - 打印函数真正执行了在设备上绘图
    - 命令执行时，trellis类对象会被自动打印，所以看起来就像是lattice函数直接完成了绘图

ggplot2 绘图系统

- The Grammar of Graphics
  - 图：动词，名词，形容词等
  - 数据映射到几何客体（points/lines/bars)的美学属性（颜色/形状/大小）
  - 基本绘图系统+Lattice绘图系统
  - 自动处理标题/文字说明/空间等，但也允许通过添加注释进行修改

  - complex, flexibility, control details

常用图形

- histogram （柱状图)
- boxplot（箱图）
- scatter（散点图）
- line（线图）

柱状图简单命令

- hist(airquality$Temp)
- ?hist

```
hist(x, ...)

## Default S3 method:
hist(x, breaks = "Sturges",
     freq = NULL, probability = !freq,
     include.lowest = TRUE, right = TRUE,
     density = NULL, angle = 45, col = NULL, border = NULL,
     main = paste("Histogram of" , xname),
     xlim = range(breaks), ylim = NULL,
     xlab = xname, ylab,
     axes = TRUE, plot = TRUE, labels = FALSE,
     nclass = NULL, warn.unused = TRUE, ...)
```

箱图简单命令

- boxplot(airquality$Temp)

```
boxplot(x, ...)

## S3 method for class 'formula'
boxplot(formula, data = NULL, ..., subset, na.action = NULL,
        drop = FALSE, sep = ".", lex.order = FALSE)

## Default S3 method:
boxplot(x, ..., range = 1.5, width = NULL, varwidth = FALSE,
        notch = FALSE, outline = TRUE, names, plot = TRUE,
        border = par("fg"), col = NULL, log = "",
        pars = list(boxwex = 0.8, staplewex = 0.5, outwex = 0.5),
        horizontal = FALSE, add = FALSE, at = NULL)
```

Arguments

`formula`	a formula, such as `y ~ grp`, where `y` is a numeric vector of data values to be split into groups according to the grouping variable `grp` (usually a factor).
`data`	a data.frame (or list) from which the variables in `formula` should be taken.
`subset`	an optional vector specifying a subset of observations to be used for plotting

散点图简单命令

- plot(airquality$Wind, airquality$Temp, type=”p")

`plot(x, y, ...)`

Arguments

`x`	the coordinates of points in the plot. Alternatively, a single plotting structure, function or *any* `R` *object with a* `plot` *method* can be provided.
`y`	the y coordinates of points in the plot, *optional* if `x` is an appropriate structure.
`...`	Arguments to be passed to methods, such as graphical parameters (see `par`). Many methods will accept the following arguments:

`type`
what type of plot should be drawn. Possible types are

- `"p"` for **p**oints,
- `"l"` for **l**ines,
- `"b"` for **b**oth,
- `"c"` for the lines part alone of `"b"`,

线图简单命令

- plot(airquality$Wind, airquality$Temp, type="l")
- 适用于时间序列分析

绘图实践
RStudio!